

n. 1/2008

Sampling design and treatment of products in Istat centralised CPI surveys .

C. De Gregorio, S. Fatello, R. Lo Conte, S. Mosca e F. Rossetti

Le collane esistenti presso l'ISTAT - *Rivista di Statistica Ufficiale*, *Contributi ISTAT* e *Documenti ISTAT* - costituiscono strumenti per promuovere e valorizzare l'attività di ricerca e per diffondere i risultati degli studi svolti, in materia di statistica ufficiale, all'interno dell'ISTAT, del SISTAN, o da studiosi esterni.

La *Rivista di Statistica Ufficiale* accoglie lavori che hanno come oggetto la misurazione dei fenomeni economici, sociali, demografici e ambientali, la costruzione di sistemi informativi e di indicatori, le questioni di natura metodologica, tecnologica o istituzionale connesse al funzionamento dei sistemi statistici e al perseguimento dei fini della statistica ufficiale.

I lavori pubblicati in *Contributi ISTAT* sono diffusi allo scopo di stimolare il dibattito intorno ai risultati preliminari di ricerca in corso.

I *Documenti ISTAT* forniscono indicazioni su linee, progressi e miglioramenti di prodotto e di processo che caratterizzano l'attività dell'Istituto.

Il Comitato di redazione esamina le proposte di lavori da pubblicare nelle tre collane sopra indicate. Quelli pubblicati nei *Contributi ISTAT* e nei *Documenti ISTAT* sono valutati preventivamente dai dirigenti dell'Istituto, mentre i lavori pubblicati nella *Rivista di Statistica Ufficiale* sono subordinati al giudizio di referee esterni.

Direttore responsabile della Rivista di Statistica Ufficiale: Patrizia Cacioli

Comitato di Redazione delle Collane Scientifiche dell'Istituto Nazionale di Statistica

Coordinatore: Giulio Barcaroli

Membri:	Corrado C. Abbate	Rossana Balestrino	Giovanni A. Barbieri
	Giovanna Bellitti	Riccardo Carbini	Giuliana Coccia
	Fabio Crescenzi	Carla De Angelis	Carlo M. De Gregorio
	Gaetano Fazio	Saverio Gazzelloni	Antonio Lollobrigida
	Susanna Mantegazza	Luisa Picozzi	Valerio Terra Abrami
	Roberto Tomei	Leonello Tronti	Nereo Zamaro

Segreteria: Gabriella Centi, Carlo Deli e Antonio Trobia

Responsabili organizzativi per la *Rivista di Statistica Ufficiale*: Giovanni Seri e Carlo Deli

Responsabili organizzativi per i *Contributi ISTAT* e i *Documenti ISTAT*: Giovanni Seri e Antonio Trobia

CONTRIBUTI ISTAT

n. 1/2008

**Sampling design and treatment of products in Istat
centralised CPI surveys**

C. De Gregorio(), S. Fatello(*), R. Lo Conte(*), S. Mosca(*) e F. Rossetti(*)*

(*) ISTAT – Servizio Prezzi

I lavori pubblicati riflettono esclusivamente le opinioni degli autori e non impegnano la responsabilità dell'Istituto

Contributi e Documenti Istat 2008

Istituto Nazionale di Statistica
Servizio Produzione Editoriale

Produzione libraria e centro stampa:
Carla Pecorario
Via Tuscolana, 1788 - 00173 Roma

Abstract

The work describes the methodologies used in Istat CPI centralised surveys and introduced with the innovation process started in the second half of 2003. The focus is posed on survey designs, and in particular on the interaction between supply and demand side aspects of consumer markets and the techniques used to estimate indexes. Several features are considered: the characteristics of samples, the way they are selected, the procedures used for estimates, and the treatment of replacements and missing observations. Replacements, and other related issues such as quality adjustments, have been interpreted not just as stand alone issues but as market specific aspects strictly dependent on the solutions adopted to solve the more general problem of producing estimates of price indexes. A case by case approach has been followed, although the solutions can be classified in a relatively limited set of homogeneous approaches. This classification depends on the main features that characterise the functioning of markets, both from supply and demand side: market structure, characteristics and turnover of the product range, consumption segments, pricing policies, availability of data and market analyses. Five case studies are presented, each representative of a kind of approach to price index estimates.

Sommario

Nel presente lavoro vengono descritti i principali aspetti delle metodologie adottate dall'Istat per la stima degli indici di prezzo al consumo delle posizioni rappresentative sottoposte a rilevazione centralizzata. Tali metodologie sono state introdotte in seguito alla ristrutturazione delle rilevazioni centralizzate avviata nella seconda metà del 2003. L'attenzione viene posta in particolare sull'interazione fra due insiemi di aspetti: da un lato le caratteristiche di funzionamento dei mercati, dal lato della domanda e da quello dell'offerta; dall'altro i disegni di indagine, e in particolare le caratteristiche dei campioni, le procedure di selezione, i metodi di stima e il trattamento delle sostituzioni di prodotto e delle mancate osservazioni. In queste rilevazioni le sostituzioni dei prodotti e le istanze ad esse collegate (come ad esempio gli aggiustamenti di qualità) vengono affrontate come aspetto integrante dello specifico approccio utilizzato per la stima di ciascun indice. Per tali stime è stato adottato un approccio specifico a ciascun mercato, sebbene tali approcci siano classificabili per aree di omogeneità metodologica. Tale classificazione dipende dalle caratteristiche di funzionamento dei mercati, sia dal lato della domanda sia dell'offerta: strutture di mercato, caratteristiche dei prodotti e dinamiche di gamma, segmentazione dei mercati, politiche di prezzo, disponibilità di dati e di analisi di mercato. Vengono inoltre presentati cinque casi di studio, ciascuno rappresentativo di uno dei gruppi di omogeneità metodologica.

Keywords: Consumption price index, Price surveys, Sampling, Replacements

Introduction

The main methodological approaches introduced in the recent restructuring of Istat centralised consumer price index (CPI) surveys are here illustrated. These surveys are characterised by the fact that for technical and economic reasons they are directly and centrally performed by Istat and are not delegated to local or external collection networks. They are used to estimate about 60 item indexes, corresponding to about one fifth of CPI weights. Centralised items are referred to markets generally characterised by peculiar features regarding for example pricing behaviours, market structure, range of products turnover. These features determine, at least partly, the decision to centralise price collection.

The innovation process started in 2003 can be viewed as a sequential process, where the initial start up effort has had two main objectives, given the resource constraint: to provide a reliable and safe monthly estimates for year 2004 unchained indexes; to pave the way for further improvements in the following years to be applied during annual rebasement. This paper describes some aspects of the state of the art which characterise year 2007 estimates cycle. The focus is on the interaction between supply and demand side features of consumer markets on one side and the techniques used to estimate indexes on the other: the characteristics of samples, the way they are selected, the procedures used for estimates, and the treatment of replacements and missing observations. As a matter of fact, a multiplicity of approaches has been followed. These approaches have nonetheless a common reference framework which is represented by the common aspects determined by the existing regulations (in particular the Hicp regulation). They can thus be viewed as market-specific

approaches to a same problem of estimate. A classification of these approaches is also proposed in order to make it evident how market conditions interacted with the technical solutions effectively adopted.

In section 1, some general aspects concerning the choice to centralise surveys are shortly treated. Section 2 deals with the common features that characterise the survey design used for these surveys. Section 3 presents a classification of the markets involved by centralised surveys and the homogeneous approaches followed for estimates. The general aspects concerning the treatment of replacements are discussed in section 4. In section 5 some case studies are presented, one for each group of methodologically homogeneous solutions outlined in section 3. The possibilities at hands for further improvements in the medium and long term are finally described in the concluding section.

1. Istat CPI centralised surveys

Centralised CPI surveys are used by Istat to estimate the monthly price indexes of about 60 representative items, which account for about 10% of the items used in HICP basket and 18% of HICP weights¹. Price collection for these items is performed directly by Istat, while for the remaining items it is delegated to the statistical offices of more than 80 provincial capitals². This technical dichotomy in price collection is reflected in several differences in item indexes estimates. The most evident of these differences stands on the fact that while for centralised items national indexes are directly produced³, in the other cases they are obtained by a weighted aggregation of local indexes. More generally, the two collection techniques bring with themselves different methodological opportunities.

Centralised surveys are widely used by NSI. Their main purpose is to avoid duplications in price collection. The prices of some goods and services are in fact centrally collected because their effective price is fixed by national or regional laws and therefore is uniform at national or regional level. In Italy this happens for example to tobacco, pharmaceutical prescription drugs, lotteries, some transportation services. National uniformity of prices also characterizes some goods and services whose prices are fixed by market mechanisms, although usually within a regulatory framework: for example, telecommunication services, postal services, financial services, books, periodicals. For all these products centralisation is an obvious solution.

For other goods and services centralisation is often the result of a choice based on an evaluation of costs and benefits offered by the two available price collection techniques. In some sectors where there are no intermediaries between producers and consumers, centralisation has in recent year become more efficient due to the growing facilities in remote price collection⁴. This happens for example to university fees, to some transportation services (by air and by sea), to some tourism related services. In these cases centralised surveys, although they do not represent always a theoretical first best, offer quite a convenient solution since they guarantee the possibility to reduce

¹ The weight of centralised surveys on the national CPI index (NIC) is larger and reaches 20%. This difference is mostly determined by the fact that NIC weights include lotteries and national health subsidies, whose price indexes are estimated by means of centralised surveys (Istat, 2006b, 2007a, 2007b). The data on the incidence of centralised surveys are referred to year 2007 baskets; their relative weight has varied very little since 1999.

² See Istat (2007a), par.1.4.

³ In some cases, regional indexes are produced (transports by rail and pharmaceutical drugs).

⁴ They are mainly performed using Internet and, less frequently, by e-mail. Telephone interviews are used only to resolve missing observations. The use of other devices, such as fax and postal forms, has become marginal.

the burden of local statistical offices and the uniformity of the criteria used for price and product treatment at the cost of a limited coverage of the target population⁵.

Centralised surveys are not at all an obvious solution for some other goods and services, characterised by the fact that prices are locally fixed at retail level: for these goods price quotes should in principle be collected by means of local surveys. In the case of durable goods (especially motor vehicles) and hi-tech products the use of centralised surveys is mainly justified by the necessity to control the rapid turnover in product range and characteristics. The choice of centralising can be seen in these cases as determined by structural aspects. An efficient local surveys should be the optimal solution: it requires a high skilled and homogeneous collection network and implies a high burden both for respondents and collectors: a costly solution. On the other hand, product turnover and the evolution of product's characteristics can be more easily monitored centrally: in such a way it is for example possible to guarantee a good knowledge of market mechanisms and a homogeneity in the evaluation of product range turnover.

The choice of centralising surveys is in these cases constrained by the limits and costs of a local delegated survey. As a matter of fact this choice is determined as a solution of a trade off between local surveys (by means of which the local dimension of price variability is monitored but overall producers' market choices are more difficult to be sufficiently monitored) and centralised surveys (where the opposite happens)⁶. For these products, as for many others locally surveyed, the rigid dichotomy (technical, organisational and methodological) between the two kinds of surveys looks increasingly like a limit: a good solution in order to overcome this limit can be found in the integration of these two techniques, concerning data collection, data treatment and estimates⁷.

2. General features of survey design

2.1. Innovations in centralised surveys

The explicit consideration of the specific features of centralised items has had a central role in the recent restructuring of Istat centralised surveys which started in the second half of 2003⁸. In particular, survey design and collection techniques have been defined in order to take into account the functioning of each specific consumer market, in which pricing policies can be strongly influenced by the control of product range and product characteristics by producers or retailer.

The sectors monitored by means of centralised surveys generally identify oligopolies, especially at producer level: the number of competitors ranges between a few units to no more than a few tens. As a consequence, in these sectors market shares are quite concentrated in a few firms: this is more evident at consumption segment level or where regional markets are considered. Competition is thus far from being perfect, and consumers' choice sets are often constrained and strictly controlled by the supply side, for example through price discrimination policies and product range turnover.

⁵ In the case of some accommodation services, where prices are collected by means of the internet, between 75% and 90% of outlets have a web site.

⁶ At the moment Istat CPI surveys are not subdivided between centralised and local surveys according to strictly homogeneous criteria. If some hi-tech products are centrally surveyed (e.g.: telephone handsets, computers) others are not, although their technical and market characteristics are quite similar to the formers (e.g.: television sets, cameras, memory cards and sticks, video and audio recorders and players). Somewhat analogous is the case for tourism services. Some of them are centrally surveyed (e.g.: camping and agritourisms) while others are locally surveyed (hotels, bed & breakfast). This at first sight ambiguous situation is clearly the result of a transition to more structured survey methodologies, eventually based on the joint use of centralised and local surveys.

⁷ This issue offers a number of possible solutions. Among these, it is worthwhile to mention the use of scanner data and the possibility to integrate local and centralised information in order to estimate price indexes. See Istat (2006).

⁸ De Gregorio et al. (2004); Istat (2006b).

In order to take into account of these distinctive features in price index estimates, two aspects have been judged as strategic in innovating survey designs: the introduction of stratified samples and the sharp increase of the size of samples⁹. These choices had to meet a resource constraint, and in this respect they have to be interpreted as a step forward of an innovation process which went along in last years and is still going on. They were accompanied by more general organisational choices: a software has been designed to treat the data collected through centralised surveys; the use of internet for price collection has been strongly increased; stronger and stable relationships with data providers (both public and private) have been built in order to feed accurate weighting systems; a wider use of the micro and meta data of other Istat structural and short term surveys has been introduced; statistical imputation techniques have been widely introduced for missing observations. Moreover, a quite detailed description of each centralised survey is regularly published on Istat web site in order to meet the information needs of users.

2.2. Consumption segments and stratification

The identification of consumption segments within each representative item has had a primary role in the identification of stratification criteria, following an analogous approach which was gaining ground at the same time at Eurostat¹⁰. The recent proposals of amendment of HICP regulation¹¹ use the notion of consumption segment in order to identify consumption purposes within an exhaustive partition of the set of all product-offers in the statistical universe. Actually, the notion of consumption segment is not a formalised concept and its role meets the need to adequately face the complexity and evolution of product ranges¹². It is unavoidably ambiguous, in particular, the exact definition of what a consumption segment is, and in particular “*the level of aggregation at which it is defined and applied*” (Eurostat, 2007). The main purpose of the use of the notion of consumption segments seems to guarantee that the HICP offers “*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*”.

Under this respect, the use of consumption segments appears at the same time as an innovation inducing device and as a work in progress in CPI methodologies. This notion brings as a consequence a more explicit consideration of sampling techniques in CPI estimates in association with a deeper knowledge of market functioning: a framework and a common methodological ground also for other subjects of CPI methodologies, such as quality adjustment and replacements, which are generally treated as autonomous methodological issues.

The approach adopted in the recent innovation of Istat centralised survey has been to base sample design on market analyses, as accurate as possible, in order to identify stratification criteria to be used for price index sampling and estimates. As a matter of fact, stratification and consumption segments have been used as synonymous. This work clearly implied a case by case approach, depending on market specific features. Notwithstanding these differences, the general approach to price index estimates has been based on a common ground. In particular, for each representative

⁹ It has to be acknowledged the fact that these choices have been also induced by the need to defend CPI surveys from any accusation of inadequacy of samples, in a period (the end of 2003) particularly difficult for official price statistics in the euro area, not only in Italy (see for example the evidence emerged from OECD 2005 seminar “*Inflation Measures: Too High - Too Low - Internationally Comparable?*”; Brunetti et al., 2005).

¹⁰ Eurostat (2007).

¹¹ European Commission (1996).

¹² “*A 'consumption segment by purpose' or '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*” (Eurostat, 2007).

item monthly estimates of price indexes are obtained as a weighted arithmetic mean of stratum indexes base on the following model:

$$I_m = \sum_{h \in H} I_{mh} w_h \quad \langle 2.2.1 \rangle,$$

where

$$\sum_{h \in H} w_h = 1 \quad \langle 2.2.2 \rangle$$

and where I is the unchained index (based December of the preceding year), h is the elementary stratum corresponding to an exhaustive partition H of the market, m is the current month (with $m=1, \dots, 12$) and w is the relative weight of each stratum.

The effective estimate of the index of each representative item is then obtained applying the model $\langle 2.2.1 \rangle$. In particular:

$$\hat{I}_m = \sum_{h \in K \subseteq H} \hat{I}_{mh} \hat{w}_h \quad \langle 2.2.3 \rangle,$$

where the hat “ $\hat{}$ ” on the variables stands for estimates and

$$\sum_{h \in K \subseteq H} \hat{w}_h = 1 \quad \langle 2.2.4 \rangle^{13}$$

where K is the set of sampled strata.

There are obviously two formally distinct sets of estimates implied by $\langle 2.2.3 \rangle$ and $\langle 2.2.4 \rangle$: the estimates of the weighting system, which takes place once a year; and the estimates of monthly price indices at elementary stratum level. As a matter of fact, these two sets of estimates are not at all independent. The former influences and defines the latter, since it contributes to control the variability of final estimates¹⁴. The choice of the stratification implies the definition of the methodologies to be adopted for the estimate of stratum indexes: it reflects the interpretation of markets’ characteristics and behaviours and brings to the identification of market segments, to the definition of products and of the criteria adopted to aggregate lower level indexes (usually weighted arithmetic means or geometric means). Especially for more complex cases, stratification has the further objective to address a well founded statistical imputation technique.

In order to estimate the weighting system, the availability of sufficiently detailed data is of crucial importance. In some cases, the complexity of these yearly estimates is much greater with respect to the corresponding monthly estimates¹⁵. An important role has been attributed to the information available on producers’ and retail markets, which often coincide in the case of centralised items: characteristics of competitive environment, consumption segments, marketing techniques, evolution of the range of products, evolution of their characteristics and its effect on consumers behaviours.

¹³ If the partition K is strictly contained in the exhaustive partition H , then it means that one or more stages in the sampling design have been introduced. More frequently, more than one level of stratification has been used, this in order to make estimates more precise and use lower level stratification for imputation purposes.

¹⁴ De Gregorio et al., 2007.

¹⁵ For tobacco, prescription drugs and school textbooks, for instance, the whole set of price quotes is monthly available through databases, so there is not a meaningful price collection burden and the burden of final estimates is concentrated in the weighting system and in the treatment of replacements.

2.3. Sample size and selection

Only in a few cases sample size and selection have been based on traditional variability analysis: this has been the case for some tourism related services and for school textbooks¹⁶. In many cases, the reasons behind this lack of a statistically founded sample size rely on the fact that samples are exhaustive or nearly exhaustive. This happens because almost all the references available are sampled, or because the cumulated market share of the selected units is very large.

When samples are not strictly exhaustive, usually a mixed approach is followed: units are sampled with probability one above a certain market share threshold and randomly below this threshold. Thresholds have mostly been kept very low so that the market share of sampled units is often well above 90%¹⁷. These selection procedures have been applied also to select one or more sample stages, where present.

In some cases random selection is applied to a subset of the population observed which is mostly defined according to criteria depending on the characteristic of the price collection technique. For instance, in tourism related services price quotes are collected on the internet and therefore only list prices published on the web are observed. There are two potential biases induced by this restriction of the target population: one has to do with the exclusion of whole categories of price makers and the other with the divergence between web prices and effective prices.

In the case of tourism, these two biases (although they have not been measured) may be expected as relatively small: web sites are widely used by firms and there is no a priori sign of a systematic bias; moreover discount policies very often do not meet HICP regulation¹⁸. On the contrary, the solution adopted in the case of hi tech products implies more heroic assumptions, since price quotes are only collected from e-commerce web sites: their market share is growing but still very little and so a possible bias may be generated by the selection of the units. The distance between effective and list prices can instead be a very serious drawback in new cars CPI survey. Some evidence has been collected showing a higher variability of prices at retail level, and it has been traded off with the high cost of a local survey. For other representative items, such as air and maritime transports, where samples are very large, they have been selected mainly by means of judgemental criteria, supported by detailed traffic data.

With the exception of the cases mentioned above, when samples are not exhaustive, random selection is used. This applies also for the replacement of missing observations.

3. Classifying markets

The general framework outlined in model <2.2.1> is common to all centralised representative items. Different approaches have been used for the estimate of stratum indexes. These difference have to do with the sample design and with other features related to the specific nature of each market.

¹⁶ For seasonal products, as in the case of tourism services, variability may be quite different in high and low season and sample size has been studied to meet high season precision objectives (De Gregorio et al, 2005, 2007).

¹⁷ This procedure has been applied, e.g., to tobacco, prescription drugs, periodicals, financial services, university fees (De Gregorio, 2006).

¹⁸ In the case of camping sites, more about three out of four have a web site and this share is even larger in the case of agritourisms are considered. Moreover, although discounts frequently may be applied in accommodation services they mainly discriminate consumers (old customers, longer stays), and are not known in advance: consequently they are not applicable to CPI (De Gregorio et al., 2007).

Table 3.1 synthesizes the main features of centralised items' markets¹⁹. Four distinct headings are considered: nature of producer and retail markets; pricing features; product range; information available (quotes, weighting and market studies). Each heading contains one or more qualitative variable.

Based on these information, representative items have been divided in clusters, identifying relatively homogeneous markets. Each cluster identifies also homogeneous approaches to stratum indexes estimates, concerning in particular survey design and methodology of estimates.

3.1. Market structure

As long as producers' market structures are considered, some items are characterised by national oligopolies: this happens with tobacco, telecommunication services, ICT consumer goods. Similar conditions apply to publishers, where a small number of very large firms is accompanied by a large number of small firms operating mainly on niche markets. Something similar happens to desktop producers' market where small national assemblers and large multinationals coexist.

Market structure is often very important at consumption segment level. In the case of pharmaceutical drugs, for example, although the number of firms operating in the whole market is quite large (some hundreds), those operating in each consumption segment – which can be identified by the chemical substance - are really a few and frequently monopoly or very restricted oligopolies characterise their market structures²⁰. Something similar applies to other items: in new cars market, competition in most consumption segments involves a relatively small number of enterprises and market shares are quite concentrated; the same happens, for example, in air transports or package holidays. Local oligopolies, usually with a very small number of competitors, characterise financial services, transports by sea and some tourism services.

Some of these sectors do not need a retail network: this is the case for telecommunication services, financial services and some tourism activities. In other sectors the retail network is strictly regulated or is mainly dedicated and peculiar. This happens to pharmaceutical drugs, tobacco, vehicles, transport services, package holidays²¹. For hi-tech products, instead, the retail network is widespread, although the market share of larger outlets is prevailing.

3.2. Pricing

As long as pricing behaviours is concerned, the situation appears quite jeopardised. In some cases retail prices are determined by national or regional laws, and they are uniform at national or local level: this happens to pharmaceutical drugs, tobacco, some transportation services. Also in telecommunications services prices are uniform across the country, but in these cases prices are only partly subject to strict regulatory frameworks. The supply in this sector takes the form of tariff plans, that is packages with different combinations of fixed and unit costs associated to the consumption pattern of subscribers: these way of pricing exalts the non linear nature of pricing rules, which presents numerous continuity breaks when intensity and composition of consumption varies. This pricing behaviours also characterise, with minor differences, financial services²².

The coexistence of official list prices, fixed by the producers, and effective prices fixed by the retailer characterises motor vehicles: in this case huge differences between these two prices emerge, especially at the beginning and at the end of the economic life of models. Similar conditions may

¹⁹ Table 3.1 does not include the centralised items whose surveys are trivial.

²⁰ See Agcm (1997).

²¹ In the case of cars and pharmaceutical drugs, the retail network is also regulated by national and European regulations.

²² In financial services pricing condition are usually customised at local level, and decision margins are left at outlet level.

apply to books, where discounts are sometimes applied by retailers: while for cars the gap between the two prices is the rule, in the case of books it is an exception and moreover it cannot exceed a threshold fixed by national legislation. Both for cars and books, list prices change very slowly.

Price dynamics is more variable if we consider hi-tech products, and in particular those related to ICT such as telephones and computers. In these markets price skimming policies are usually applied, where for each model prices tend to decrease with time from its entry to its exit from the market in order to better exploit the expenditure potential represented by the “demand curve”. Also in tourism services variability is very high, but in this case the reason must be searched in the seasonal nature of pricing policies. This happens also to many transportation services: as far as air transports are considered, the variability induced by the particular nature of pricing in this sector adds to seasonal variability²³.

3.3. Segmentation

ICT products are all characterised by a strong variability in the range of the models supplied to consumers with a short average life of models and a rapid evolution of their characteristics. Models are generally standardised and well identified by names and codes: desktop computers are often an exception since customisation can be very deep. New cars market is also characterised by large products range, and although models are clearly coded and named they can be highly customised. A high dynamic of the product range can also be observed in telecommunication services. In other sectors range dynamics is slower, especially within consumption segments where product range is limited: this is the case for pharmaceutical drugs and air transports.

Consumption segments are generally stable in the medium-long term, but with some meaningful exceptions. In some segments, quite stable base segments are clearly identifiable while upper level consumption segments rapidly evolve. The market of telephone handsets, fixed or mobile, has these features. The car market too, although its evolution is much slower than the former. In most of the remaining ICT sectors, the rapid evolution of product range and consumption segments makes it difficult the identification even of stable base segments.

Within each consumption segment it is usually concentrated the bulk of short term competition among producers. In some cases, anyway, there is also an appreciable elasticity among segments (e.g.: in air transports and tourism services) or among different items (as in the case of telecommunication services).

3.4. Information

The quantity and quality of information and data available on each market are not homogeneous: there are different levels of general knowledge of these markets and different problems to be faced in order to introduce a stratification and to estimate a weighting system. In some sectors the data available are very detailed. For tobacco and pharmaceutical prescription drugs the prices of all the references currently for sale is continuously available accessing to databases; moreover, for these items, a very detailed set of structural (annual) data is available reporting the turnover and the volume of sales for each product. Highly detailed information is also available on motor vehicles concerning both structural data (sales per model) and official list prices. A similar situation takes place for transports by sea and inland waters.

A lower but still quite consistent amount of structural data is also available for a number of other markets: usually this information is directly derived from the elaboration of micro and meta data of other Istat surveys or from other institutions of the national statistical system. In many cases, official prices can be downloaded from web sites, as in the case of books, telecommunication

²³ Agcm (2005).

services, financial services, tourism services, some transportation services. On line purchase simulations are used for air transports services.

For ICT consumer goods, structural information on each market – and consequently on consumption segments - is quite scarce, and they are very scarce for telephones notwithstanding the relatively high weight of this item in family's budget and in CPI baskets. Price quotes for these products are collected from e-commerce web sites.

Some of the markets involved in centralised surveys are under scrutiny of academics, regulation authorities and general public from a long time: this happens, for example, to pharmaceutical drugs, cars and computers. In other cases, attention is much more recent although its level is relatively high: this is the case for telecommunications and financial services. In other cases, only more recent market analyses are available, usually delivered by specialised periodicals.

Table 3.1

Items	Production and retail		Pricing		Range of products		Consumer markets		Data availability				
	Type of oligopoly	Retail network	Price fixing	Pricing	Characteristics	Turnover	Consumption segments	Market competition	Price quotes	Weighting	General knowledge	CPI literature	Official statistics
Tobacco	At country level	Wide and regulated	Uniform at country level	Administrative	Standardised	Very slow	Stable segments	Mainly within each segment	Access to data base	High and detailed	Scarce or absent	Scarce or absent	Very detailed
Pharmaceutical drugs	At segment level	Wide and regulated	Uniform at country or regional level	Administrative	Standardised, small within segments	Medium	Stable segments	Mainly within each segment	Access to data base	High and detailed	Sector specific economic literature	Present and long dated	Very detailed
Telecommunication services	At country level	Direct	Uniform at country level	Non linear packages	Standardised	Medium	Stable base segments and rapidly evolving upper segments	Also with other markets	Availability of official list prices on web sites	Not fully detailed	Sector specific economic literature	Recent	Almost detailed
Financial services	Particularly at regional level	Direct	Retail level but based on producer's advised list	Non linear packages	Officially standardised, customisation	Medium/sLow	Stable segments	Mainly within each segment	Availability of official list prices on web sites	Not fully detailed	Sector specific economic literature	Recent	Almost detailed
Motor vehicles	Particularly at segment level	Wide and regulated	Retail level but based on producer's advised list	Scarce variability	Officially standardised, customisation	Medium/High	Slowly evolving segments	Mainly within each segment	List prices on specialised periodicals	High and detailed	Sector specific economic literature	Present and long dated	Very detailed
Books and periodicals	Highly concentrated but with many small niche competitors	Wide, mainly large outlets	Uniform at country level	Scarce variability	Standardised	Slow	Stable segments	Mainly within each segment	Availability of official list prices on web sites	Not fully detailed	Only dedicated periodicals	Scarce or absent	Almost detailed

Table 3.1 (continued)

Items	Production and retail		Pricing		Range of products		Consumer markets		Data availability				
	Type of oligopoly	Retail network	Price fixing	Pricing	Characteristics	Turnover	Consumption segments	Market competition	Price quotes	Weighting	General knowledge	CPI literature	Official statistics
Telephone handsets	At country level	Wide, mainly large distribution	Retail level	Skimming	Standardised	High	Stable base segments and rapidly evolving upper segments	Mainly within each segment	Availability of price spells on e-comm sites	Scarce or absent	Only dedicated periodicals	Scarce or absent	Scarce or absent
Notebook and peripheral units	At country level	Wide, mainly large distribution	Retail level	Skimming	Standardised	High	Rapidly evolving segments	Mainly within each segment	Availability of price spells on e-comm sites	Not fully detailed	Disseminated market analyses	Only general aspects	Scarce or absent
Desktop	Highly concentrated but with many small local competitors	Wide, mainly large distribution	Retail level	Skimming	Large, not always coded, customisation	High	Rapidly evolving segments	Mainly within each segment	Availability of price spells on e-comm sites	Not fully detailed	Disseminated market analyses	Present and long dated	Scarce or absent
Transport services by air	Particularly at segment level	Wide and regulated	Retail level (agency fees plus producer's price)	Seasonal	Standardised, small within segments	Medium	Stable segments	Also among segments	Availability of price spells on e-comm sites	Not fully detailed	Disseminated market analyses	Scarce or absent	Almost detailed
Transport services by sea	Particularly at local level	Wide and regulated	Retail level (agency fees plus producer's price)	Seasonal	Standardised, very small within segments	Slow	Stable segments	Mainly within each segment	Availability of official list prices on web sites	High and detailed	Scarce or absent	Scarce or absent	Very detailed
Package holidays	Particularly at segment level	Wide and regulated	Long term list prices	Seasonal	Officially standardised, customisation	Slow	Stable segments	Significantly among segments	Availability of official list prices on web sites	Not fully detailed	Disseminated market analyses	Scarce or absent	Only very general aspects
Tourism services	At local level	Direct	Retail level with long term list prices	Seasonal	Very small	Slow	Stable segments	Mainly within each segment	Availability of official list prices on web sites	Not fully detailed	Scarce or absent	Scarce or absent	Only very general aspects

3.5. Clusters

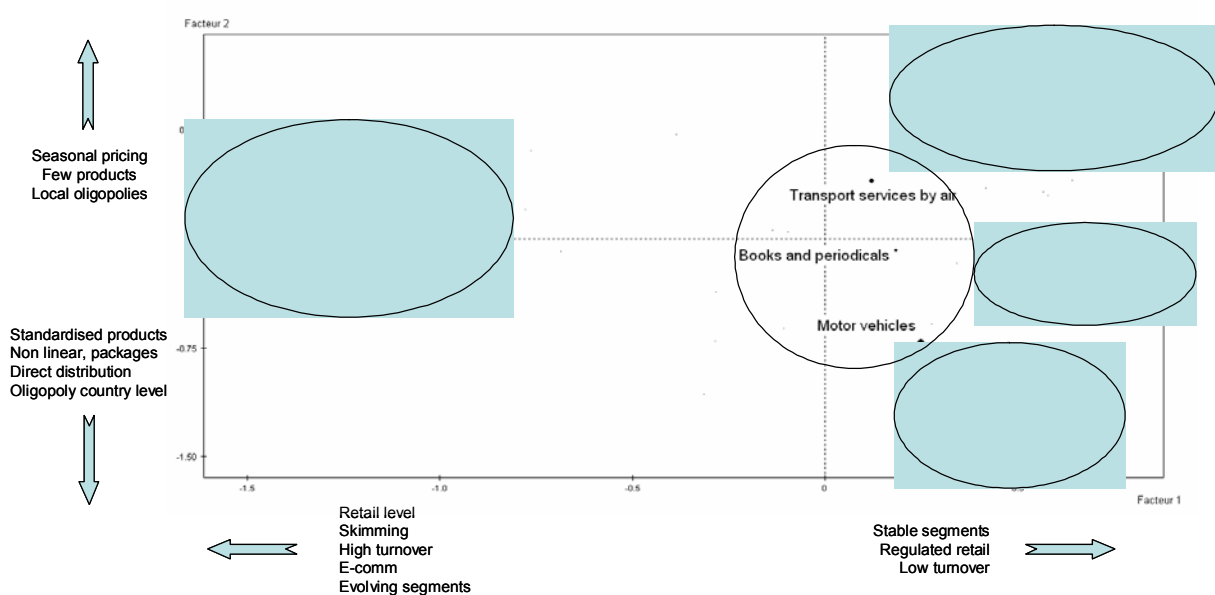
The qualitative information of table 3.1 traces on one side the general characteristics of each market and on the other focuses on aspects more specifically related to pricing policies and to the availability of data for price statistics. All this information has been coded and examined in order to classify centralised items. A correspondence analysis has thus been applied on these data and, sequentially, a clustering algorithm has been applied on the main factors resulting from the analysis. The graph 3.1 shows some results limited to the space determined by the first two factors²⁴: each item is projected on this space and a partition resulting from clustering is also highlighted.

We omit the technical characteristics of this clustering experiment, since the conclusion we derive from it are used simply for ex-post descriptive purposes. More precisely, the objective of this analysis has been to classify markets and to propose in each cluster homogeneous solutions for survey design and estimates. In particular, five clusters resulted from this analysis:

1. ICT consumer goods, strongly characterised by price skimming policies, relevance of the retail network and high range turnover;
2. Telecommunications and financial services, where pricing takes the form of tariff plans (packages), strongly non linear, and where there are no intermediaries. In last years a large literature has studied these sectors, but rarely it has faced the problem of estimating price indexes;
3. The products with administered prices, such as tobacco and pharmaceutical drugs, characterised by a high availability of structural and price data;
4. Tourism services, where seasonal pricing is an important issue, prices are fixed by means of long term (usually annual) list prices, and markets are oligopolies at local level;
5. Other sectors, less homogeneous, with characteristics partly similar to those of group 3 (i.e.: cars and books) and partly to group 4 (i.e.: air transports).

Figure 3.1

A classification of the markets of centralised items



²⁴ The first two factors represent about 40% of total inertia.

4. Survey design and replacements: general aspects

As it has been shown in section 2, estimates of the price indexes of centralised items are all obtained as weighted arithmetic means of strata (i.e.: consumption segments) indexes: on the other hand, the methodologies used to estimate strata indexes differ significantly, although these differences can be classified. In particular, the classifications of the markets summarily described in section 3 can be used in order to classify the methods adopted for stratum estimates. This classification helps to group the solutions according to their methodological homogeneity.

Table 4.1 describes synthetically the main characteristics of each survey design, from the choice of the observed population to the definition of the sample design, the method of estimates, the management of products and the treatment of replacements. Section 5 offers a detailed description of a few case studies, one for each group of items defined in graph 3.1 and in section 3. Some more general remarks can help in order to identify a few aspects of the common background that characterises the approaches followed for stratum estimates.

Survey design, and in particular stratification and sample size, have been given a primary role in order to identify the procedures to manage the main methodological issues in CPI estimates, and more specifically stratum indexes estimates and replacements. Survey design implies an underlying interpretation of the functioning of each market, which is a necessary – and unavoidable – prerequisite in order to manage the more correctly as possible estimates and replacements, given the resource constraints. More specifically the case for replacement has not been interpreted as a sort of random event independent from the normal functioning of a sample in a price index estimate but as a continuous process intrinsic in the functioning of consumer markets. In many cases the rapid evolution of the range of products offered to consumers makes it necessary to take systematically into account the changes occurring in the choice sets available to consumers. In order to represent these aspects, replacements have been considered within survey design.

The reference model that we have adopted as a general framework for centralised surveys interprets the price index of each stratum as a function of the changes which take place in the choice sets available to consumers in the current and in the base period. In particular:

$$I_{h,m} = I(X_{h,m}, X_{h,0}; w_h) \quad <4.1>$$

where X_{hm} represents the choice set available for consumption in the stratum h and in month m ²⁵, while w represents the weighting structure adopted within stratum h ²⁶. The choice sets include all the available combinations of products (e.g.: models) and prices eligible for the estimate of price indexes. Monthly choice sets can have many dimensions determined by the product range and by the variability of the price spells for each product offer. For many centralised items, choice sets are relatively limited, sometimes because product range is small and more frequently because prices are uniform across the country.

Changes in the product range cannot be in general limited to the identification of a one-to-one link between a product offer which exits the market and one that enters. A comparison limited to these two entities cannot offer a general solution to replacements and index estimates. The effect of product turnover should in principle be measured taking into account its overall implications on the choice sets available for consumption. Practical solutions effectively adopted to stratum estimates

²⁵ Choice sets can vary continuously (i.e.: daily, hourly). Conventionally we refer to monthly choice sets as the sets available in the date of collection, which for centralised surveys is scheduled one day per month.

²⁶ Base period is identified with θ .

depend on one side on the nature of the choice sets and on the other on the information available²⁷. In general, replacements have been managed within the conceptual framework adopted in each sample design.

Each sample design identifies, in particular, a specific representation of choice sets alterations. Two distinct sets of approaches have been adopted. For some items the whole range of products is monthly surveyed and sample size can thus be variable. In these cases, index estimates derive from the evaluation of the whole range of products currently traded: replacements take place automatically as product range changes.

In the case of telecommunications and financial services choice sets are relatively limited. Prices are uniform across the country and the range of monthly offers is clearly identified within each stratum and it is exhaustively surveyed. In these surveys sample size can thus be monthly variable and replacements are completely automatic. In these cases no one-to-one relation is established between exits and entries in the choice sets, and price indexes are estimated throughout an evaluation of the whole current and previous supply on a sample of consumption profiles. Base and current month choice sets are directly compared.

A somewhat similar approach is also followed for ICT goods, where price spells are determined at outlet level. Also for these goods the entire range of products is surveyed within each stratum²⁸. Sample size varies monthly and replacements are automatically managed. Also in these cases choice sets are directly compared.

In other cases, sample size is fixed, and replacements take place only when the turnover of product range involves the samples. This approach can be envisaged as a second best choice, constrained by resource and data availability: this solution in fact implies an asymmetry. Stratum indexes are estimated by means of the sampled products until they keep on being available to consumers. In this context, fixed size samples work as range turnover signals, but these signals only concern the exit of products in the sample, while new entries in the range are taken into account only eventually, in substitution of old products and in a one-to-one correspondence with the exiting product. The impact of this asymmetry can be lowered through the use of very large samples.

In the case of the group of items characterised by administered prices (pharmaceutical drugs and tobacco) very large and almost exhaustive fixed size samples are used. Prices are uniform at country level and the whole set of price quotes is available. Turnover in product range is negligible in the case of tobacco, but it is relatively more frequent for pharmaceutical drugs. For this item, a very detailed stratification is adopted identifying a consumption segment within each chemical substance. Replacements are managed by means of a decision tree and a direct comparison between products moving “out” and “in” is the more frequent solution (see section 5). The selection of substitutes is random within the products available within the stratum. Replacements do not take place when no substitutes are available, and missing observations are imputed with stratum or upper level stratum average rates of change.

Also in the cases of tourism services fixed sample size are used, with a few hundreds of outlets in each sample. In these cases the sampled fraction is generally far from 100%, especially for accommodation services. For these items it is not straightforward to establish an exact meaning for changes in the product range. Usually, replacements are not induced by market events (like the default of an outlet, i.e.: a camping) but by technical events, mainly determined by the impossibility

²⁷ Obviously a general constraint is represented by the resources available for price collection and methodological support.

²⁸ Entry and exit of products is in these cases asymmetric, since exits from the official product range generally do not imply the immediate and definitive exit of products from the market: exits can be in fact very slow and it is evaluated through the results obtained from price collection.

to collect prices (i.e.: a web site no more available). Replacements for these items are limited to the use of reconstructed overlap, with the chaining of the price index of the new outlet to the price index of the old one.

Other solution are adopted for the remaining items, all characterised by a fixed sample size. A decision tree is used to manage replacements in the case of new cars, where the number of sampled models is very large and almost exhaustive. Direct substitutions are the main solution effectively adopted for replacements. Replacements are quite rare in the case of books, where sample is in absolute quite large but relatively small if compared to the available supply: they are mainly due to change in the series. Direct comparison is also adopted in this case: it is systematically applied in the best sellers stratum. For air transport direct substitution is applied within each stratum.

Some case studies are examined in the following sections, relating to each of the items classified in section 3.

Table 4.1

Sampling and Replacements	Pharmaceutical drugs	Tobacco	Tele-communication services	Financial services	Telephone handsets	Computers and peripheral units	Package holidays	Campings/ Agritourism	Transport by sea and inland waters	Transport by air	Cars	Books
Observed phenomenon	Whole	Whole	Whole	Whole (official supply)	Sales on e-comm	Sales on e-comm	List prices	Outlets with a web site	Whole	Sales on Internet	List prices	Whole (List prices)
Quotes collected	4.500	240	60 (variable)	160 (variable)	1300 (variable)	3000 (variable)	260	200/140	110	800	240	200
Final sample	4.500	240	60 (variable)	160 (variable)	220 (variable)	400 (variable)	260	400/140	110	113	240	250
Sampled item	Single drug	Packet	Package	Package	Models	Models	Single offer	Accommodation	Route (Origin-Destination), Company	Flight (Origin-Destination)	Version	Book (isbn code)
Sampling fraction	Very high at stages	Very high	Exhaustive	Very high	Very high at brand level	Very high at brand level	Low	Low	Very high	Medium	Very high at model level	Low
Consumption segment	Chemical	Brand	Operator and consumption profile	Bank and consumption profile	Segment/Brand	Segment/Brand	Destination (mainly Country)	Geographical area and type of location	Traffic segment	Traffic segment	Segment/Brand	Subject
Stratification levels	4 (1st level ATC, 5th level ATC, type of drug, drug)	1 (Brand)	2 (Operator and profile)	2 (Region, bank, profile)	2 (Segment, Brand)	2 (Segment, Brand)	2 (Geographical area, Destination)	2 (Geographical area and type of location)	Geographical area and type of route	Geographical area and type of flight	3 (Segment, Brand, Model)	1 (Subject)
Sampling stages	2 (5th level ATC, drug)	1 (Brand)	2 (Operator and profile)	2 (Bank and profile)	1 (Brand)	2 (Segment, Brand)	1 (Destination)	no	no	no	2 (Brand, Model)	no
Sampling selection	Exhaustive above and random below threshold	Exhaustive above and random below threshold	Cut off for operators	Exhaustive above and random below threshold for banks, Exhaustive on packages	Cut off for brands. Exhaustive for models	Cut off for brands and segments. Exhaustive for models	Cut off for destinations. Random for single offers	Random	Cut-off	Cut-off	Exhaustive above and random below threshold, judgemental for versions	Random
Cases of replacements	Frequent	Rare	Frequent	Frequent	Frequent	Frequent	Few	Few	Not applicable	Frequent	Frequent	Rare
Replacements	Decision tree (mainly direct)	Decision tree (mainly direct)	Direct	Direct	Direct	Direct	Mainly reconstructed overlap	Mainly reconstructed overlap	no	Direct	Decision tree (mainly direct and mixed)	Direct
Always replaced?	no	no	yes	yes	no	no	yes	yes		Yes	no	yes
Replacement selection	Mainly random	Mainly random	Automatic	Automatic	Automatic	Automatic	Mainly random	Mainly random		Automatic	Judgemental	Mainly random
Replacement procedure	Partly automatic	Manual	Automatic	Automatic	Automatic	Automatic	Manual	Manual		Automatic	Manual	Manual

5. Some case studies

5.1. Administered prices: the case of prescription drugs

Survey design

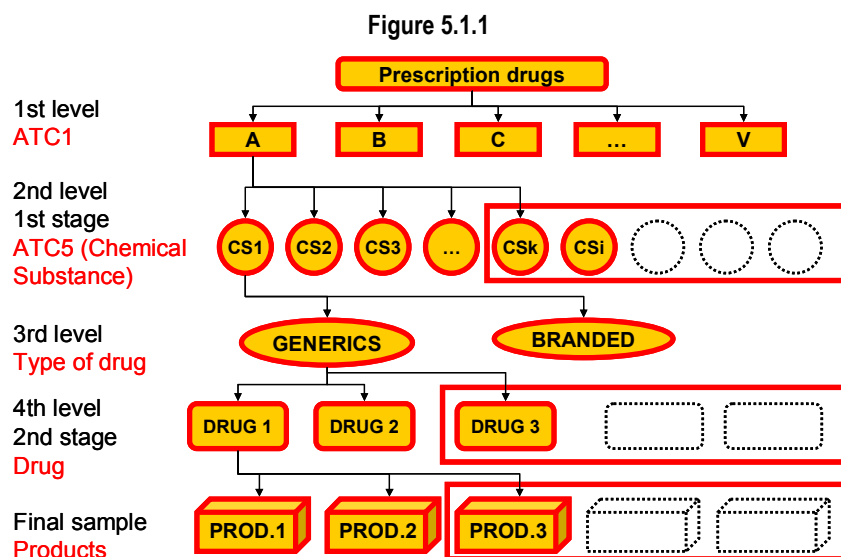
A large fixed size sample is used for this survey, with four stratification layers and two intermediate sampling stages. Replacements are managed by means of a decision tree²⁹.

The first stratification level is defined by the first digit of ATC classification (figure 5.1.1)³⁰. A second level of stratification is identified by the chemical substance, corresponding to the fifth digit of ATC classification. A first sampling stage takes place at this level: chemical substances with higher expenditure are all included until 90% of upper stratum weights are represented, and a random selection of the remaining substances is added in order to represent the residual market share. The probability of selection π of the chemical substance i can be represented as follows:

$$\pi_i = \begin{cases} 1 & \text{if } F_i > \alpha \text{ and } F_{i-1} \leq \alpha \\ x & \text{if } F_i > \alpha \text{ and } F_{i-1} > \alpha \end{cases} \quad <5.1.1>$$

where x is a constant (with $0 < x < 1$), $i \in ATC1$, with $f_i \geq f_{i+1}$ e $F_i = \sum_{k=1}^i f_k$. In particular, f_i is the market share of the chemical substance i and F_i marks the cumulated market share of the first i substances ordered by decreasing turnover.

Within each chemical substance, a third stratification level is identified by the type of drugs (generic/non generic): this stratification identifies consumption segments. Within each segment, a sample of drugs is selected, identified by a producer and a name, and a final sample of 4.500 references is finally selected: both samples are selected according to <5.1.1>.



Elementary stratum indexes are obtained as weighted arithmetic means of the price indexes of each reference:

²⁹ More details on this survey are available from De Gregorio (2006).

³⁰ The Anatomic Therapeutic Chemical classification is the official World Health Organisation classification for pharmaceutical drugs (see <http://www.whooc.no/atcddd/>).

$$\hat{I}_{h,m} = \sum_{i \in h} \frac{P_{im}}{P_{i0}} \hat{w}_i = \sum_{i \in h} I_{h,m,i} \hat{w}_i \quad \langle 5.1.2 \rangle,$$

where $\sum_{i \in h} \hat{w}_i = 1$ and h identifies the stratum.

Replacements

Although sample size is fixed, a complete monitoring of product range is anyway possible through the access to a continuously updated database: the number of new references entering the market each year is around 2-3% of the total, while exits involve a similar share of products. Total turnover is about 5%.

Replacement is managed with a decision tree, which is activated when a product which is in the sample ceases to be available (Figure 5.1.2). The decision tree selects the action depending on the nature of the events which involve the change in the product range. The possible solutions are quite numerous although more frequently outgoing products are replaced with identical products offered by the same or by other producers and selected randomly. In these cases, direct comparison is applied as follows:

$$I_{h,m,i} = I_{h,m-1,i} \frac{P_{im}^{new}}{P_{i,m-1}^{old}} \quad \langle 5.1.3 \rangle.$$

If this solution is not applied, replacements are searched among identical products with different doses: in this case the expenditure necessary to reply the dose of the outgoing reference is directly compared with its price. Further layers of the decision tree produce as possible solutions reconstructed overlaps.

$$I_{h,m,i} = I_{h,m-1,i} \frac{P_{im}^{new}}{P_{i,m-1}^{new}} \quad \langle 5.1.4 \rangle.$$

If no replacement is possible within the third level stratum then no replacement is made and the missing observation is imputed using the average monthly rate of change estimated for the elementary stratum to which the observation belongs:

$$\hat{p}_{h,m,i}^{miss} = p_{h,m-1,i}^{old} (1 + \hat{x}_{h,m}) \quad \langle 5.1.5 \rangle,$$

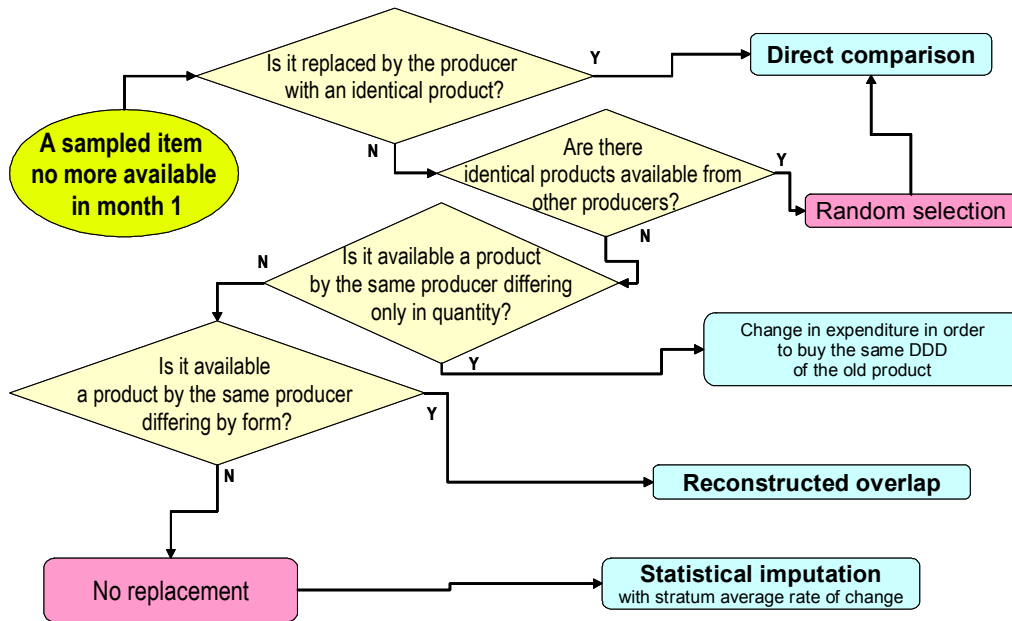
where $\hat{x}_{h,m}$ is the average monthly rate of change estimated at elementary stratum level. In particular,

$$\hat{x}_{h,m} = \sum_{i \in (h \cap R_m)} I_{h,m,i} \hat{w}_i \quad \langle 5.1.6 \rangle,$$

where R_m is the set of reference whose price has been effectively collected in month m . Since samples are revised yearly, imputation can be maintained until the end of the year.

Further improvements in this survey, and in particular in replacement techniques, seem to imply a deeper specialist insight of the pharmaceutical products, their markets and of their regulatory framework.

Figure 5.1.2



5.2. Tariff plans and non linear pricing: the case mobile tlc services³¹

Survey design

A monthly variable sample of tariffs is observed, corresponding to the whole supply of the largest operators. Product replacements are automatically managed, through the simulation of the consumption behaviour of more than 300 consumption profiles³².

Survey design is based on the stratification of the market by operator (Figure 5.2.1). This stratification corresponds with a sampling stage. The market is very concentrated and a very small number of large operators is present³³. A cut off sampling approach has thus been followed. Selection probabilities are as follows:

$$\pi_i = \begin{cases} 1 & \text{if } f_i > \alpha \\ 0 & \text{if } f_i \leq \alpha \end{cases} \quad <5.2.1>.$$

The cumulated market share of the sampled operators in the consumer segment is above 90%.

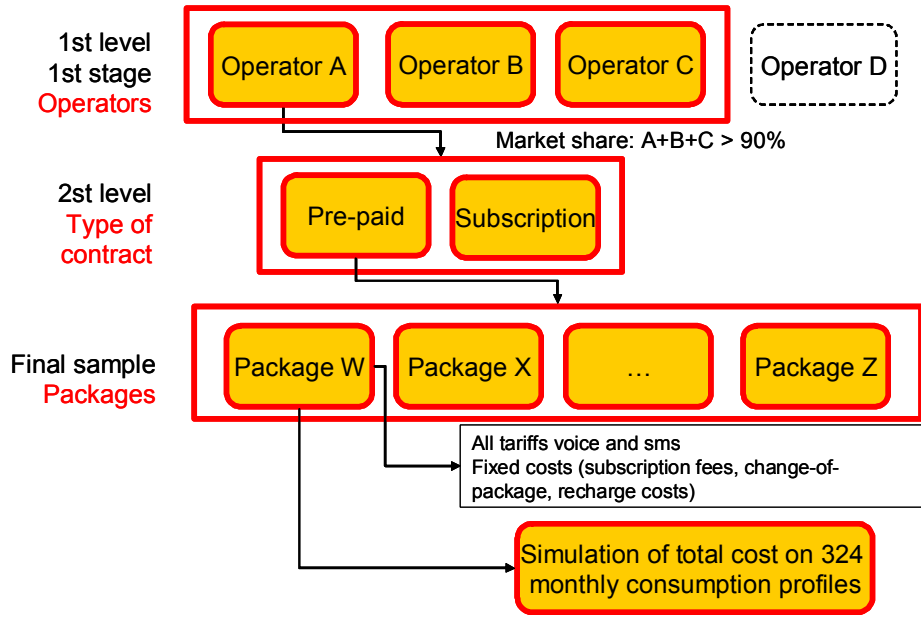
For each operator, a supplementary stratification level has been defined distinguishing the two types of contracts available: subscriptions and pre-paid contracts. This level identifies the consumption segments. Within each segment, all available supply is exhaustively sampled: it is composed by tariff plans (packages), which can be interpreted as vectors of elementary fixed and variable prices (set up and duration) for each type of traffic (voice and data) and destinations (local, national). Monthly fixed costs, recharge and change-of-plan costs are also collected.

³¹ The authors wish to thank Gerolamo Giungato and Angelo Montani, working at Istat Structural business statistics unit, for their useful ideas and suggestions.

³² The use of consumption profiles to estimate price indexes for telecommunication services has been recently spurred by the publication of the Nera report (1999), commissioned by the British sector regulation authority (Ofcom). Similar approaches have also been suggested, for example by Lacroix et al. (2001), Vassiliou (2001) and Le Gallo et al. (2003).

³³ Agcom (2007).

Figure 5.2.1



Each tariff plan can be represented by a vector of prices t_{hj} where h identifies the tariff plan and j identifies the type of fare (by type of connection, with $j=1$ to m). In month 1 k_1 tariff plans are currently offered, part of which is available to all consumers (k_{1a}), while the remaining ($k_{1b}=k_1-k_{1a}$) are only available to former users. The whole set of the k_1 tariff plans can be represented by a matrix as follows:

$$T_1^S = [t_{hj}]_{k_1 \times m} \quad <5.2.2>$$

where S identifies the segment.

The monthly cost associated to the adoption of each tariff plan is simulated on n consumption profiles which have been estimated using the data available on traffic distribution, intensity and frequency of usage. Profiles may be represented as matrix:

$$\hat{C} = [\hat{c}_{ij}]_{n \times m} \quad <5.2.3>$$

whose generic element \hat{c}_{ij} represent the consumption of the type of connection j made by profile i .

Let

$$\hat{F} = [\hat{f}_{ih}]_{n \times k} \quad <5.2.4>$$

be the matrix of fixed costs, whose generic element represents the fixed costs (reported on a monthly bases) incurred by profile i while choosing the tariff plan h . They include fixed charges and a discounted portion of change-of-plan costs.

The set of tariff plans available in the current month can be divided in two subsets, one corresponding to the plans available to all consumers and the other to the plans only available to former users. Then, a matrix

$$\hat{S} = [\hat{s}_{ih}]_{n \times k} \quad <5.2.5>$$

identifies the choice set available for each profile, its generic element being equal to 1 when availability occurs and it is set to missing otherwise.

Finally, referred to period 1, a matrix with all the simulated costs associated to each combination of profile-available tariff plan is derived from the following:

$$\hat{V}_1^s = \left(\hat{C}_{n \times m} T_{1, m \times k_1}^s + F_{1, n \times k_1}^s \right) \times S_{1, n \times k_1}^s = \left[\hat{v}_{ij,1}^s \right] \quad <5.2.6>.$$

In particular, the product **CT** identifies costs associated to each combination of profile and tariff plan. Matrix **F** adds fixed costs and matrix **S** deletes the not available associations profiles and tariff plans³⁴.

Replacements

It is assumed that each profile selects monthly the cheapest tariff plan among those which are part of his current choice set. In particular, it is hypothesized that each profile simulates the behaviour of a consumer who opts monthly for the tariff plan less expensive among those in the choice set available to him, taking into account all the fixed costs associated with this choice: the consumption story of profiles is thus tracked monthly. More precisely, this index measures the dynamic of the costs associated to the most convenient tariff plans monthly available for consumption and defined for each consumption profile.

From matrix **V** a vector **P** is derived as follows:

$$\hat{P}_{1, n \times 1}^s = \left[\min_j \left(\hat{v}_{ij,1}^s \right) \right] \quad <5.2.7>.$$

Finally the stratum index is derived as the ratio of the weighted average costs incurred by profiles in period 0 and period 1: in particular:

$$\hat{I}^s = \frac{\hat{P}_{1, 1 \times n}^s \hat{W}_{n \times 1}}{\hat{P}_{0, 1 \times n}^s \hat{W}_{n \times 1}} \quad <5.2.8>.$$

where the matrix **W** reports the fixed weight associated to each consumption profile.

Within this survey design replacements are completely automatic and they are managed by means of the hypothesis in profile behaviour implicit in <5.2.7>. The turnover of tariff plans is quite high, and the choice set available to all consumers varies almost completely during the year. The sample of tariff plans, identified by the matrix **T**, as a consequence varies monthly.

This approach puts all the emphasis on current choice sets, independently of the effective consumption choices of consumers: as a matter of fact, a high (infinite) elasticity of substitution is implied. Other approaches base consumer choices on more articulated behavioural models, which usually dilute them with smoothed passages from less to more convenient packages³⁵.

³⁴ In particular matrix S deletes the combinations profile-tariff plan, where the plan is no more available to new subscriber and the profile was not among the former subscribers. Moreover, matrix S constraints the choice sets of a small share of randomly selected profiles (10% on an annual basis) only to currently traded plans: they are used to simulate consumer turnover (“deaths” and “births”).

³⁵ Nera (1999). These approaches need a starting point and backward simulations are then needed for at least a couple of years. The use of consumer behaviour models requires a more detailed information on consumer markets, which implies stronger relationships with operators and market authorities.

5.3. ICT goods: the case of mobile phones

Survey design

For mobile phones variable samples are used, including all the currently available models supplied by a sample of the largest brands. Replacement is automatically managed.

Economic literature has given a quite large space to ICT goods price indexes, although this attention has been concentrated mainly on computers³⁶. Notwithstanding its recent fast growing relevance, the market of telecommunications apparatuses has been almost neglected by official literature³⁷: the weight of mobile and fixed telephone handsets on Italian CPI basket is 40% higher than personal computers and peripheral units.

The Italian market of mobile handsets is almost saturated and it is mostly a substitution market. Notwithstanding its relevance, the information available on this market is very scarce, and although the number of producers is very small, no certified data exist even on their market shares: some data referred to the whole European market are available from private sources. Consequently, there is no official and certified data available concerning consumption segments (whatever their definition is) and type of outlet.

In order to match this lack of information a supply side approach has been followed in order to identify consumption segments³⁸. The whole range of models belonging to the largest producers has been thus classified according to a large set of characteristics concerning communication features and other functions. This data set has been analysed with a correspondence analysis and clustering methods have been applied in order to classify the available supply. As a result three distinct segments have been identified: a base segment, with models having the strictly necessary functions for mobile communications, an intermediate and an upper segment characterised by larger sets of functions.

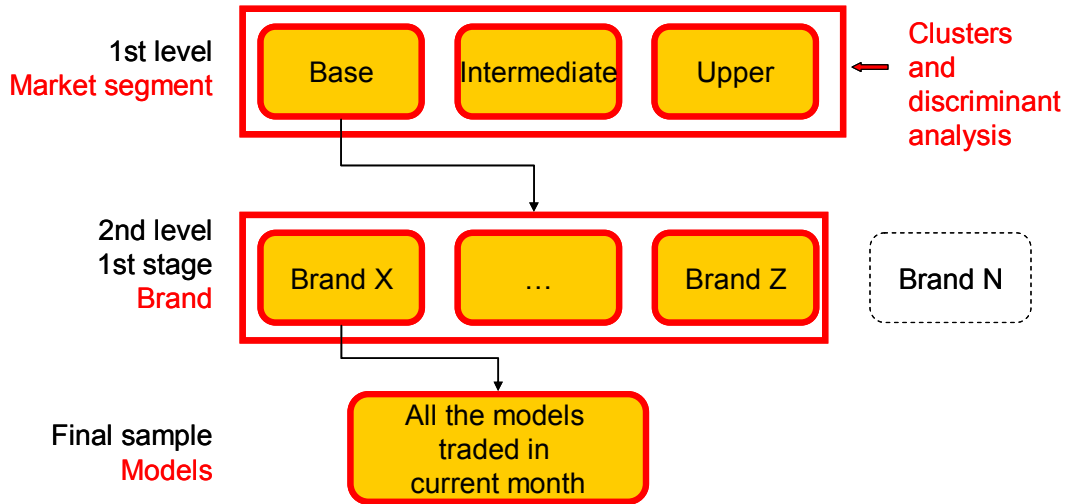
These consumption segments coincide with the first stratification level. Within each segment, a sample of producers has been selected: in particular all the largest producers present in Italian market have been selected. Elementary strata are then identified by crossing consumption segments and brand (Figure 5.3.1). For each brand, all the models' range is surveyed. When new models start being traded, all their characteristics are recorded and, by means of a discrimination analysis algorithm, they are associated to one of the three consumption segments. Models that disappear from the producers' official range of products keep on being sampled until their presence in the market, as resulting from price collection, keeps on being meaningful. Within each elementary stratum the sample of models is then exhaustive, and its size changes monthly.

³⁶ See for example Triplett (2004), Eurostat (2004), Berndt et al. (2001). Less frequently attention has been devoted to hifi goods (Kokoski, 2001).

³⁷ A relevant but isolated exception can be found in the work of Hausmann (1997).

³⁸ Mazzoni et al. (2004).

Figure 5.3.1



Replacements

Within each elementary stratum (h), call R_m the range of product available in month m and N the number of e-commerce web sites monthly visited for price collection. The sample of products in period t (the set C_t) consists of all the models officially included in the product range plus all the models out of range but sampled in period $t-1$ and offered in at least a $\alpha\%$ of e-commerce web sites in the preceding three months. In formulas:

$$C_m \{i \mid i \in R_m\} \cup \{i \notin R_m \mid i \in T_{m-1} \cup T_{m-2} \cup T_{m-3}\} \quad <5.3.1>$$

where i identifies each model, and $T_m = \{i \mid n(p_{i,m})/N > \alpha\}$ where $n(p_{i,m})$ is the number of price spells observed for model i in month m .

Price collection brings to a monthly price quote matrix:

$$P_t = [p_{ij,t}] \quad <5.3.2>$$

whose generic element identifies the price spell of model i in the e-commerce outlet j . Obviously, while the number of outlets (that is the columns of the matrix) is blocked, the number of models in the sample (the rows) varies on a monthly basis.

From the P matrixes referred to current and base period, the elementary stratum price index is then estimated as follows:

$$I_{h,m,i} = \frac{\min_{i \in h_t} [\text{geomean}_{j=1}^N (p_{ij,m})]}{\min_{i \in h_0} [\text{geomean}_{j=1}^N (p_{ij,0})]} \quad <5.3.3>$$

In other words, the stratum price index is derived from the ratio of its reference prices in current and base period. Reference prices are derived in two steps: first, an average price is estimated for each model as a geometric mean of the price spells collected; secondly, the model with the lowest average price is selected as representative of the stratum.

Further improvements in this approach to estimates can be soundly achieved with better market data and information, which can help both to ameliorate weighting and to improve market segmentation. At the moment the large amount of models and prices sampled in last years gives the possibility of a deeper analysis of time series, tracing the price history of single models and its connections with their characteristics and functions.

5.4. Tourism services: the case of package holidays³⁹

Survey design

For these item, a fixed stratified sample is used, with two intermediate sampling stages. Replacements are managed using bridged overlaps.

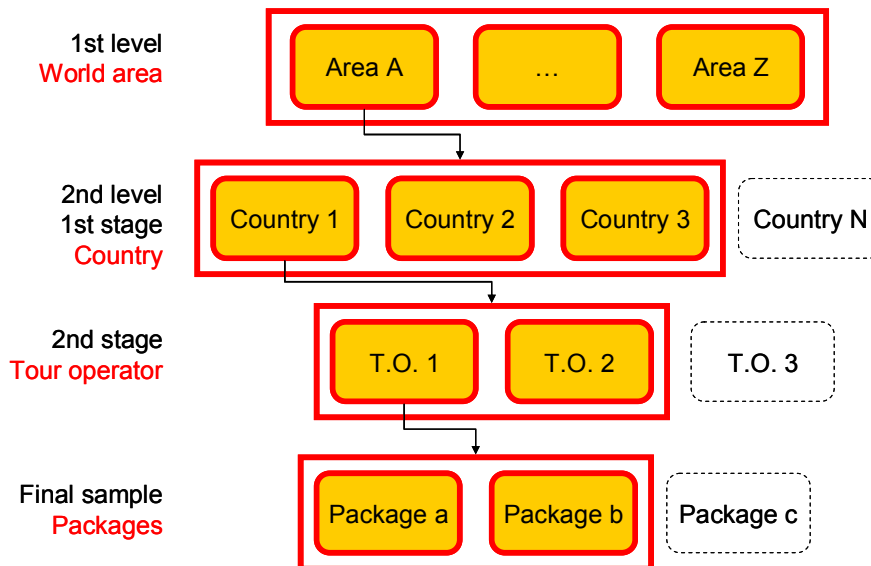
The Italian market of package holidays has a relatively large number of competitors (a few tens), although their number is sharply reduced if specific destinations are considered. Quite detailed data on the volume of passengers by destination is available from the largest Italian tour operators association. No data is available on tour operators' market shares.

The sample has been first stratified exhaustively by geographical world areas (continents). Within each first level stratum, a stratification by country has been introduced. This stratification correspond with a first sampling stage: countries are selected according to a cut off criterion with a sampling probability defined as follows:

$$\pi_i = \begin{cases} 1 & \text{if } f_i > \alpha \\ 0 & \text{if } f_i \leq \alpha \end{cases} \quad <5.4.1>$$

where f_i are country market shares and α is a threshold. Within each country a further sampling stage has been introduced corresponding to tour operators. They have been selected randomly and their number is proportional to country weights. A final sample of about 250 package offers is finally randomly selected among those supplied by each tour operator.

Figure 5.4.1



Replacements

Sample size is fixed. The estimate of price indexes at consumption segment (i.e.: country) level is obtained as follows:

$$\hat{I}_{h,m} = \prod_{i \in h} \frac{P_{i,m}}{P_{i,0}} \quad <5.4.1>$$

where h identifies the segment and i the sampled offers.

Price quotes are collected from tour operators catalogues, available from their web sites. Catalogues vary during the year, depending on the kind of destination: seasonal editions of catalogues are then

³⁹ For a detailed description of the price index for some accommodation services see De Gregorio et al. (2007).

quite frequent and product linkages between old and new catalogues is normally made between identical offers. In these cases - which cannot be properly defined as replacements - direct comparisons are used to compare the prices in different catalogues. Otherwise, should a sampled offer disappear from catalogues, replacement is managed through a reconstructed overlap: a new offer is selected, in the same stratum and from the same tour operator, whose current and previous month reference prices are used to estimate the monthly rate of change of the elementary index. In particular, as in the case of <5.1.4.>:

$$I_{h,m,i} = I_{h,m-1,i} \frac{P_{im}^{new}}{P_{i,m-1}^{new}} \quad <5.4.2>.$$

No other replacements techniques are used. The use of reconstructed overlap derives from a resource constrained decision: notwithstanding its recent enlargement⁴⁰, sample size is in fact too small to make it possible and safe a comprehensive evaluation of the effects of the changes introduced in new catalogues. For the same destinations, the packages supplied by each tour operator can in fact be quite numerous and articulated, depending on the length of the stays, on the kind of activities, on the consumer target. In these cases, direct substitutions (<5.1.3>) are then quite difficult to be managed and consistently applied, since they usually imply a comparison among quite different products.

In order to improve and generalise the replacement approach it is needed a deeper knowledge of each tour operator range of offers and a further segmentation of markets. Both imply a complex and time consuming task, which represents the long term objective of the restructuring started in 2003: the actual configuration of the survey can be considered a step forward that objective.

5.5. Other sectors: the case of new cars

Survey design

A fixed size sample is used to estimate the new cars price index, with a three layers stratification. The sampling design is defined by two intermediate sampling stages. Replacements are managed by means of a decision tree.

The market for new cars can count on a great deal of statistical information concerning the volume of sales by model and the producers' market shares⁴¹. Sample design is structured in three stratification levels and two sampling stages (figure 5.5.1)⁴². The first stratification level is defined by the market segment and it is quite articulated (Table 5.5.1)⁴³.

⁴⁰ More than 250 packages were included in year 2007 sample.

⁴¹ This is also a market deeply analysed by economic literature. The rediscovery of hedonic methods dates half a century (Griliches, 1961). The first works on quality adjustments and hedonic models in the first half of 20th century were financed by General motors. Quality adjustment keep on being one of the most treated subjects in the literature on car price indexes (see for example Tomat, 2002; Raff et al. 1995). More recently attention has moved also towards segmentation analysis and pricing (see for example Lutz, 2004; Zettelmeyer et al., 2003).

⁴² An upper layer of the stratification is represented by the type of engine (gasoline or diesel).

⁴³ The definition of consumption segments in new cars market is not backed by an official source. Several classifications are available and they frequently vary during years. We based our stratification on the approach used by the specialised monthly magazine "Quattroruote".

Table 5.5.1. Car market segments

First level	Second level
MINI CARS	MICROCARS
	CITY CARS
SMALL CARS	SMALL CARS
SEDANS (SALOONS)	SMALL SEDANS
	MEDIUM SEDANS
	MEDIUM-LARGE SEDANS
	LARGER SEDANS
STATION WAGONS (ESTATES)	COMPACT STATION WAGONS
	LARGE STATION WAGONS
SPORT CARS	COUPES
	CONVERTIBLES
	HIGH PERFORMANCE CARS
OFF-ROAD AND SPORT UTILITY VEHICLES	OFF-ROAD AND SPORT UTILITY VEHICLES
MINIVANS AND MULTI PURPOSE VEHICLES	SMALL MINIVANS
	COMPACT MINIVANS
	LARGER MINIVANS
	MPVs

A second level of stratification is identified by the brand. A first sampling stage takes place at this level. In particular, brands with higher expenditure are all included in the sample until a cumulated market share threshold value is reached. This brands represent each a separate stratum within the segment. A residual stratum is added, containing a random selection of the remaining brands. The probability of selection π is thus:

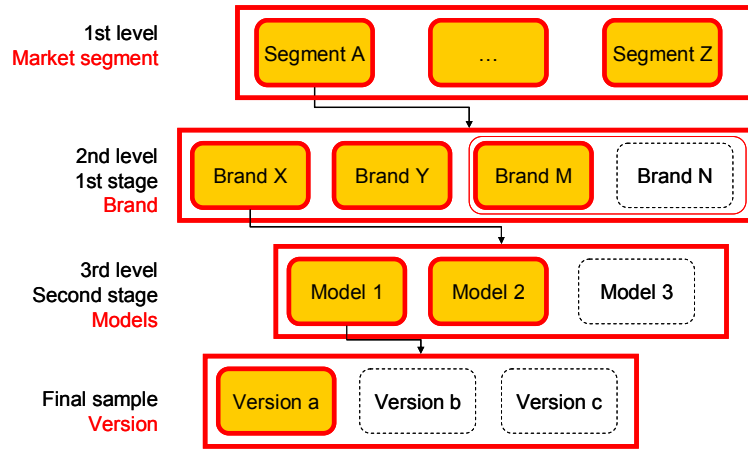
$$\pi_i = \begin{cases} 1 & \text{if } F_i > \alpha \text{ and } F_{i-1} \leq \alpha \\ x & \text{if } F_i > \alpha \text{ and } F_{i-1} > \alpha \end{cases} \quad <5.5.1>$$

where $0 < x < 1$, i identifies brands, with $f_i \geq f_{i+1}$ e $F_i = \sum_{k=1}^i f_k$. In particular, f_i marks the market share of brand i and F_i marks the cumulated market share of the first i brands in terms of expenditure.

Within each brand, a third stratification level is identified by the models: this level coincides with sampling stage. Models are selected with a cut off criterion defined according to a threshold market share value⁴⁴. Models are usually presented in the market with several versions: only one version is selected for each model, judgementally chosen as the most marketed one. The final sample is made by 240 versions of new cars.

⁴⁴ Most frequently, brands are present in each market segments with only one model.

Figure 5.5.1



Sample size is fixed. Elementary stratum indexes are obtained as weighted arithmetic means of the price indexes of each reference:

$$\hat{I}_{h,m} = \sum_{i \in S} \frac{P_{im}}{P_{i0}} \hat{w}_i \quad <5.5.2>$$

where $\sum_{i \in h} \hat{w}_i = 1$ and h identifies the stratum.

Replacements⁴⁵

Although the sample size is fixed, a complete monitoring of product range is anyway possible through the access to complete price lists. Replacements are quite frequent and annually involve more than 10% of the models sampled. As in the case of package holidays and pharmaceutical drugs, the sample works as a change-of-range signalling device. Replacement is managed with a decision tree, which is activated when a sampled version ceases to be available (Figure 5.5.2). Through the decision tree actions are selected depending on the nature of the change in the product range. A substitute is searched among the currently available versions of the same model. Usually a substitute version is clearly identified and direct comparison is applied (<5.1.3>).

Less frequently no direct substitute exists and a new version is selected in order to represent the stratum. In this case, the monthly rate of change is imputed using a direct comparison between the base versions of current and former month range of versions. In particular:

$$I_{h,m,i} = I_{h,m-1,i} \frac{P_{im}^{base}}{P_{i,m-1}^{base}} \quad <5.5.3>$$

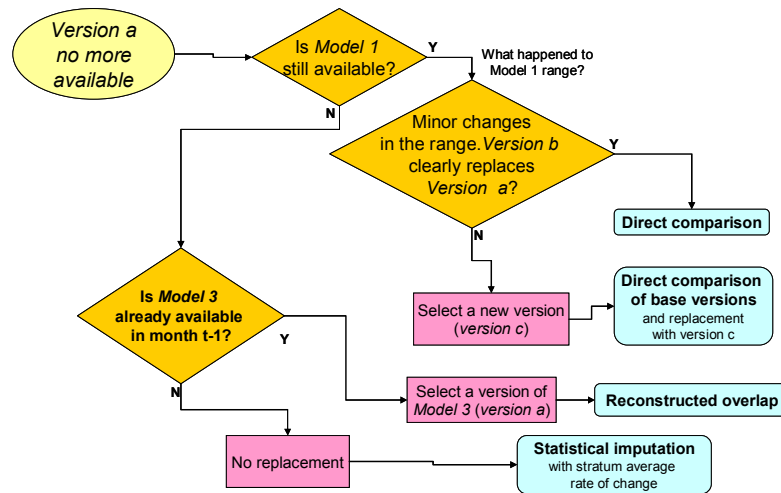
while in the following months the price of the substitute version is used:

$$I_{h,m+1,i} = I_{h,m,i} \frac{P_{im+1}^{new}}{P_{i,m}^{new}} \quad <5.5.4>$$

If the sampled model is no more available, then another model has to be selected among those available from the same brand and segment. In this case, reconstructed overlap is used (see <5.1.4>). Should there not be any substitute model available, the observation is left missing and its rate of change is imputed with the stratum average (see <5.1.5> and <5.1.6>).

⁴⁵ A detailed description of the replacement procedures used for motorcycles can be found in Mosca (2007).

Figure 5.5.2



This index measures list price changes and this is its main drawback, since effective prices are different both in level and dynamics, although they require high collection costs. Further improvements in the replacement technique can thus be considered, at this stage, of minor importance and little more than an academic exercise. A deeper insight on car markets is anyway needed in order to represent more precisely segments and interpret more soundly the evolution of consumers choice sets.

Concluding remarks

In the recent restructuring of Istat CPI centralised surveys the management of replacements has been interpreted as a by-product of survey design. Replacements, and other related issues such as quality adjustments, have been interpreted not just as stand alone issues, but as particular aspects not distinguishable from the more general problem of producing estimates of price indexes. More precisely, the improvements which have been achieved in the samples used to estimate centralised indexes have been used in order to pave the way for a methodological background to replacements. As a consequence the innovations which have been introduced represent a first step forward further improvements: the large data sets of price quotes and of technical characteristics of products collected in these years can in fact help to propose more methodologically founded survey designs.

Consumption segments have been adopted as a reference concept and - although it is still relatively imprecise – it has been the main tool by means of which surveys have been designed. Consumption segments can in fact interact positively with the aim of obtaining more founded statistical designs for surveys: given the constraint posed by existing Hicp rules, solution have been founded joining traditional statistical approaches to economic and market analyses.

Centralised surveys are particularly suited for this task. If compared to local surveys, centralised ones are generally easier, samples can be brought near to be exhaustive, market mechanisms and pricing policies can be more easily studied and traced. Sectors are usually oligopolies, highly concentrated, characterised by peculiar features. A case by case approach has consequently been followed, although the solutions can be classified in a relatively limited set of homogeneous approaches.

These solutions are generally strictly constrained by the limited amount of resources available for price collection and for the methodological work. Medium and long term improvements can be generally found in deeper market analyses and more detailed data useful for weighting and sample selection. The investments made in these years have been made in order to meet both the short term

objective to increase the reliability of samples and the medium and longer term objective to go on with the innovation process.

References

- Agcm - Autorità garante della concorrenza e del mercato (1997), Indagine conoscitiva Nel settore farmaceutico. *Indagine conoscitiva* n. 14.
- Agcm - Autorità garante della concorrenza e del mercato (2005), Dinamiche tariffarie del trasporto aereo passeggeri. *Indagine conoscitiva* n. 24.
- Agcom - Autorità per le garanzie nelle comunicazioni (2007), Relazione annuale sull'attività svolta e sui programmi di lavoro.
- Berndt E. R., Rappaport N. J. (2001), Price and Quality of Desktop and Mobile Personal Computers: A Quarter Century Historical Review. *American Economic Review*, 91(2), May, 268-273.
- Brunetti A., De Gregorio C., Monducci R., Nuccitelli A., Polidoro F., Ricci R., (2005), *Official CPI and users: recent developments in Istat strategy*. OECD seminar "Inflation Measures: Too High - Too Low - Internationally Comparable?", Paris, 21-22 June.
- De Gregorio C. (2006), Il nuovo impianto della rilevazione centralizzata del prezzo dei medicinali con obbligo di prescrizione. *Contributi Istat*, n.16.
- De Gregorio C., Fatello S. (2005), L'indice dei prezzi al consumo dei testi scolastici nel 2004. *Documenti Istat*, n.12.
- De Gregorio C., Fatello S., Mosca S., Rossetti F. (2004), Improving the quality of Istat consumer price index: the case of the centralised surveys on prices. Associated paper at the *European conference on quality and methodology in official statistics (Q2004)*, Mainz, 24-26 May.
- De Gregorio C., Munzi C., Zavagnini P. (2007), *Problemi di stima, effetti stagionali e politiche di prezzo in alcuni servizi di alloggio complementari: alcune evidenze dalle rilevazioni centralizzate dei prezzi al consumo*. Mimeo, Istat
- European Commission (1996), *Regulation (EC) No 1749/96 on initial implementing measures for Council Regulation (EC) No 2494/95 concerning harmonised indices of consumer prices*
- Eurostat (2004), Final report on computers and software. *Task Force Quality Adjustment and Sampling*.
- Eurostat (2007), Draft Commission Regulation amending Regulation no 1749/96 on initial implementing measures for Council Regulation no 2494/95 concerning Harmonised indices of consumer prices.
- Goolsbee A. (2000), Competition in the computer industry: online versus retail. University of Chicago.
- Griliches Z. (1961), Hedonic Prices for Automobiles: An Econometric Analysis of Quality Change. In: NBER, *The Price Statistics of the Federal Government*. New York.
- Hausmann J. (1997), Cellular telephone: new product and the CPI. *NBER working paper*, n.5982.
- Istat (2006a), *Innovazioni introdotte e proposte nel sistema di indici dei prezzi al consumo*. Commissione di studio sui numeri indici dei prezzi al consumo. December
- Istat (2006b), *Come si rilevano i prezzi al consumo*, <http://www.istat.it/prezzi/precon/aproposito/altro/metodologia2006.pdf>

- Istat (2007a), *Come si rilevano i prezzi al consumo*, <http://www.istat.it/prezzi/precon/aproposito/metodologia2007.pdf>
- Istat (2007b), The characteristics of Istat CPI centralised surveys. Mimeo, Istat.
- Kokoski M., Waehrer K., Rozaklis P. (2001), Using Hedonic Methods for Quality Adjustment in the CPI: The Consumer Audio Products Component. *Bls working paper* n.344.
- Lacroix T., Magnien F. (2001), *Methodology for constructing a price index for mobile-telephony services*. Paper presented at the Sixth Meeting of the International Working Group on Price Indices Canberra, Australia, 2-6 April
- Le Gallo F., Magnien F. (2003), *Measuring price change in mobile-telephony services: an arduous task*. Ottawa Group International Working Group on Price Indices, Seventh Meeting, Paris, May 27-29
- Lutz M. (2004), Pricing in Segmented Markets, Arbitrage Barriers, and the Law of One Price: Evidence from the European Car Market. *Review of International Economics*, 12(3), 456–475.
- Mazzoni C., Addeo F., Castaldi L. (2004), Multidimensional segmentation of the cell phone market: preliminary results of an empirical survey. Quaderni del Dipartimento di Strategie Aziendali e Metodologie Quantitative, Seconda università di Napoli.
- Mosca (2007) Ciclomotori e motocicli: le sostituzioni di prodotto. Mimeo, Istat.
- Nera (1999), *A price index for mobile telephony. A report for Oftel*.
- Raff D.M.G., Trajtemberg M. (1995), Quality adjusted prices for the american automobile industry: 1906-1940. *NBER working paper* n. 5035.
- Tomat G. M. (2002), Durable goods, price indexes and quality change: an application to automobile prices in Italy, 1988-1998. *ECB working paper* n.118.
- Triplett J. (2004), Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products, *OECD Science, Technology and Industry Working Papers*, 2004/9, OECD.
- Vassiliou S. (2001), *Consumer Price Indexes for Telecommunications Services Based On A Sample of Customer Bills: Summary Of Progress*. Room Document at the Sixth Meeting of the International Working Group on Price Indices. Canberra, Australia, 2-6 April
- Zettelmeyer F., Scott Morton F., Silva-Risso J. (2003), Inventory Fluctuations and Price Discrimination: The Determinants of Price Variation in Car Retailing. *Management Science*.

Contributi ISTAT(*)

- 1/2004 – Marcello D’Orazio, Marco Di Zio e Mauro Scanu – *Statistical Matching and the Likelihood Principle: Uncertainty and Logical Constraints*
- 2/2004 – Giovanna Brancato – *Metodologie e stime dell’errore di risposta. Una sperimentazione di reintervista telefonica*
- 3/2004 – Franco Mostacci, Giuseppina Natale e Elisabetta Pugliese – *Gli indici dei prezzi al consumo per sub popolazioni*
- 4/2004 – Leonello Tronti – *Una proposta di metodo: osservazioni e raccomandazioni sulla definizione e la classificazione di alcune variabili attinenti al mercato del lavoro*
- 5/2004 – Ugo Guarnera – *Alcuni metodi di imputazione delle mancate risposte parziali per dati quantitativi: il software Quis*
- 6/2004 – Patrizia Giaquinto, Marco Landriscina e Daniela Pagliuca – *La nuova funzione di analisi dei modelli implementata in Genesee v. 3.0*
- 7/2004 – Roberto Di Giuseppe, Patrizia Giaquinto e Daniela Pagliuca – *MAUSS (Multivariate Allocation of Units in Sampling Surveys): un software generalizzato per risolvere il problema dell’allocazione campionaria nelle indagini Istat*
- 8/2004 – Ennio Fortunato e Liana Verzicco – *Problemi di rilevazione e integrazione della condizione professionale nelle indagini sociali dell’Istat*
- 9/2004 – Claudio Pauselli e Claudia Rinaldelli – *La valutazione dell’errore di campionamento delle stime di povertà relativa secondo la tecnica Replicazioni Bilanciate Ripetute*
- 10/2004 – Eugenio Arcidiacono, Marina Briolini, Paolo Giuberti, Marco Ricci, Giovanni Sacchini e Giorgia Telloli – *Procedimenti giudiziari, reati, indagati e vittime in Emilia-Romagna nel 2002: un’analisi territoriale sulla base dei procedimenti iscritti nel sistema informativo Re.Ge.*
- 11/2004 – Enrico Grande e Orietta Luzi – *Regression trees in the context of imputation of item non-response: an experimental application on business data*
- 12/2004 – Luisa Frova e Marilena Pappagallo – *Procedura di now-cast dei dati di mortalità per causa*
- 13/2004 – Giorgio DellaRocca, Marco Di Zio, Orietta Luzi, Emanuela Scavalli e Giorgia Simeoni – *IDEA (Indices for Data Editing Assessment): sistema per la valutazione degli effetti di procedure di controllo e correzione dei dati e per il calcolo degli indicatori SIDI*
- 14/2004 – Monica Pace, Silvia Bruzzone, Luisa Frova e Marilena Pappagallo – *Review of the existing information about death certification practices, certificate structures and training tools for certification of causes of death in Europe*
- 15/2004 – Elisa Berntsen – *Modello Unico di Dichiarazione ambientale: una fonte amministrativa per l’Archivio delle Unità Locali di Asia*
- 16/2004 – Salvatore F. Allegra e Alessandro La Rocca – *Sintetizzare misure elementari: una sperimentazione di alcuni criteri per la definizione di un indice composto*
- 17/2004 – Francesca R. Pogelli – *Un’applicazione del modello “Country Product Dummy” per un’analisi territoriale dei prezzi*
- 18/2004 – Antonia Manzari – *Valutazione comparativa di alcuni metodi di imputazione singola delle mancate risposte parziali per dati quantitativi*
- 19/2004 – Claudio Pauselli – *Intensità di povertà relativa: stima dell’errore di campionamento e sua valutazione temporale*
- 20/2004 – Maria Dimitri, Ersilia Di Pietro, Alessandra Nuccitelli e Evelina Paluzzi – *Sperimentazione di una metodologia per il controllo della qualità di dati anagrafici*
- 21/2004 – Tiziana Pichiorri, Anna M. Sgamba e Valerio Papale – *Un modello di ottimizzazione per l’imputazione delle mancate risposte statistiche nell’indagine sui trasporti marittimi dell’Istat*
- 22/2004 – Diego Bellisai, Piero D. Falorsi, Annalisa Lucarelli, Maria A. Pennucci e Leonello G. Tronti – *Indagine pilota sulle retribuzioni di fatto nel pubblico impiego*
- 23/2004 – Lidia Brondi – *La riorganizzazione del sistema idrico: quadro normativo, delimitazione degli ambiti territoriali ottimali e analisi statistica delle loro caratteristiche strutturali*
- 24/2004 – Roberto Gismondi e Laura De Sandro – *Provisional Estimation of the Italian Monthly Retail Trade Index*
- 25/2004 – Annamaria Urbano, Claudia Brunini e Alessandra Chessa – *I minori in stato di abbandono: analisi del fenomeno e studio di una nuova prospettiva d’indagine*
- 26/2004 – Paola Anzini e Anna Ciammola – *La destagionalizzazione degli indici della produzione industriale: un confronto tra approccio diretto e indiretto*
- 27/2004 – Alessandro La Rocca – *Analisi della struttura settoriale dell’occupazione regionale: 8° Censimento dell’industria e dei servizi 2001 7° Censimento dell’industria e dei servizi 1991*
- 28/2004 – Vincenzo Spinelli e Massimiliano Tancioni – *I Trattamenti Monetari non Pensionistici: approccio computazionale e risultati della sperimentazione sugli archivi INPS-DM10*
- 29/2004 – Paolo Consolini – *L’indagine sperimentale sull’archivio fiscale modd.770 anno 1999: analisi della qualità del dato e stime campionarie*
- 1/2005 – Fabrizio M. Arosio – *La stampa periodica e l’informazione on-line: risultati dell’indagine pilota sui quotidiani on-line*
- 2/2005 – Marco Di Zio, Ugo Guarnera e Orietta Luzi – *Improving the effectiveness of a probabilistic editing strategy for business data*
- 3/2005 – Diego Moretti e Claudia Rinaldelli – *EU-SILC complex indicators: the implementation of variance estimation*
- 4/2005 – Fabio Bacchini, Roberto Iannaccone e Edoardo Otranto – *L’imputazione delle mancate risposte in presenza di dati longitudinali: un’applicazione ai permessi di costruzione*
- 5/2005 – Marco Broccoli – *Analisi della criminalità a livello comunale: metodologie innovative*
- 6/2005 – Claudia De Vitiis, Loredana Di Consiglio e Stefano Falorsi – *Studio del disegno campionario per la nuova rilevazione continua sulle Forze di Lavoro*
- 7/2005 – Edoardo Otranto e Roberto Iannaccone – *Continuous Time Models to Extract a Signal in Presence of Irregular Surveys*

(*) ultimi cinque anni

- 8/2005 – Cosima Mero e Adriano Pareto – *Analisi e sintesi degli indicatori di qualità dell'attività di rilevazione nelle indagini campionarie sulle famiglie*
- 9/2005 – Filippo Oropallo – *Enterprise microsimulation models and data challenges*
- 10/2005 – Marcello D' Orazio, Marco Di Zio e Mauro Scanu – *A comparison among different estimators of regression parameters on statistically matched files through an extensive simulation study*
- 11/2005 – Stefania Macchia, Manuela Murgia, Loredana Mazza, Giorgia Simeoni, Francesca Di Patrizio, Valentino Parisi, Roberto Petrillo e Paola Ungaro – *Una soluzione per la rilevazione e codifica della Professione nelle indagini CATI*
- 12/2005 – Piero D. Falorsi, Monica Scannapieco, Antonia Boggia e Antonio Pavone – *Principi Guida per il Miglioramento della Qualità dei Dati Toponomastici nella Pubblica Amministrazione*
- 13/2005 – Ciro Baldi, Francesca Ceccato, Silvia Pacini e Donatella Tuzi – *La stima anticipata OROS sull'occupazione. Errori, problemi della metodologia attuale e proposte di miglioramento*
- 14/2005 – Stefano De Francisci, Giuseppe Sindoni e Leonardo Tininini – *Da Winci/MD: un sistema per data warehouse statistici sul Web*
- 15/2005 – Gerardo Gallo e Evelina Palazzi – *I cittadini italiani naturalizzati: l'analisi dei dati censuari del 2001, con un confronto tra immigrati di prima e seconda generazione*
- 16/2005 – Saverio Gazzelloni, Mario Albisinni, Lorenzo Bagatta, Claudio Ceccarelli, Luciana Quattrocchi, Rita Ranaldi e Antonio Toma – *La nuova rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*
- 17/2005 – Maria Carla Congia – *Il lavoro degli extracomunitari nelle imprese italiane e la regolarizzazione del 2002. Prime evidenze empiriche dai dati INPS*
- 18/2005 – Giovanni Bottazzi, Patrizia Cella, Giuseppe Garofalo, Paolo Misso, Mariano Porcu e Marianna Tosi – *Indagine pilota sulla nuova imprenditorialità nella Regione Sardegna. Relazione Conclusiva*
- 19/2005 – Fabrizio Martire e Donatella Zindato – *Le famiglie straniere: analisi dei dati censuari del 2001 sui cittadini stranieri residenti*
- 20/2005 – Ennio Fortunato – *Il Sistema di Indicatori Territoriali: percorso di progetto, prospettive di sviluppo e integrazione con i processi di produzione statistica*
- 21/2005 – Antonella Baldassarini e Danilo Birardi – *I conti economici trimestrali: un approccio alla stima dell'input di lavoro*
- 22/2005 – Francesco Rizzo, Dario Camol e Laura Vignola – *Uso di XML e WEB Services per l'integrazione di sistemi informativi statistici attraverso lo standard SDMX*
- 1/2006 – Ennio Fortunato – *L'analisi integrata delle esigenze informative dell'utenza Istat: Il contributo del Sistema di Indicatori Territoriali*
- 2/2006 – Francesco Altarocca – *I design pattern nella progettazione di software per il supporto alla statistica ufficiale*
- 3/2006 – Roberta Palmieri – *Le migranti straniere: una lettura di genere dei dati dell'osservatorio interistituzionale sull'immigrazione in provincia di Macerata*
- 4/2006 – Raffaella Amato, Silvia Bruzzone, Valentina Delmonte e Lidia Fagiolo – *Le statistiche sociali dell'ISTAT e il fenomeno degli incidenti stradali: un'esperienza di record linkage*
- 5/2006 – Alessandro La Rocca – *Fuzzy clustering: la logica, i metodi*
- 6/2006 – Raffaella Cascioli – *Integrazione dei dati micro dalla Rilevazione delle Forze di Lavoro e dagli archivi amministrativi INPS: risultati di una sperimentazione sui dati campione di 4 province*
- 7/2006 – Gianluca Brogi, Salvatore Cusimano, Giuseppina del Vicario, Giuseppe Garofalo e Orietta Patacchia – *La realizzazione di Asia Agricoltura tramite l'utilizzo di dati amministrativi: il contenuto delle fonti e i risultati del processo di integrazione*
- 8/2006 – Simonetta Cozzi – *La distribuzione commerciale in Italia: caratteristiche strutturali e tendenze evolutive*
- 9/2006 – Giovanni Seri – *A graphical framework to evaluate risk assessment and information loss at individual level*
- 10/2006 – Diego Bellisai, Annalisa Lucarelli, Maria Anna Pennucci e Fabio Rapiti – *Feasibility studies for the coverage of public institutions in sections N and O*
- 11/2006 – Diego Bellisai, Annalisa Lucarelli, Maria Anna Pennucci e Fabio Rapiti – *Quarterly labour cost index in public education*
- 12/2006 – Silvia Montagna, Patrizia Collesi, Florinda Damiani, Danila Fulgenzio, Maria Francesca Loporcario e Giorgia Simeoni – *Nuove esperienze di rilevazione della Customer Satisfaction*
- 13/2006 – Lucia Coppola e Giovanni Seri – *Confidentiality aspects of household panel surveys: the case study of Italian sample from EU-SILC*
- 14/2006 – Lidia Brondi – *L'utilizzazione delle surveys per la stima del valore monetario del danno ambientale: il metodo della valutazione contingente*
- 15/2006 – Carlo Boselli – *Le piccole imprese leggere esportatrici e non esportatrici: differenze di struttura e di comportamento*
- 16/2006 – Carlo De Gregorio – *Il nuovo impianto della rilevazione centralizzata del prezzo dei medicinali con obbligo di prescrizione*
- 1/2007 – Paolo Roberti, Maria Grazia Calza, Filippo Oropallo e Stefania Rossetti – *Knowledge Databases to Support Policy Impact Analysis: the EuroKy-PIA Project*
- 2/2007 – Ciro Baldi, Diego Bellisai, Stefania Fivizzani, e Marina Sorrentino – *Production of job vacancy statistics: coverage*
- 3/2007 – Carlo Lucarelli e Giampiero Ricci – *Working times and working schedules: the framework emerging from the new Italian lfs in a gender perspective*
- 4/2007 – Monica Scannapieco, Diego Zardetto e Giulio Barcaroli – *La Calibrazione dei Dati con R: una Sperimentazione sull'Indagine Forze di Lavoro ed un Confronto con GENESEES/SAS*
- 5/2007 – Giulio Barcaroli e Tiziana Pellicciotti – *Strumenti per la documentazione e diffusione dei microdati d'indagine: il Microdata Management Toolkit*
- 6/2007 – AA.VV. – *Seminario sulla qualità: l'esperienza dei referenti del sistema informativo SIDI - 1ª giornata*
- 7/2007 – Raffaella Cianchetta, Carlo De Gregorio, Giovanni Seri e Giulio Barcaroli – *Rilevazione sulle Pubblicazioni Scientifiche Istat*
- 8/2007 – Emilia Arcaleni, e Barbara Baldazzi – *Vivere non insieme: approcci conoscitivi al Living Apart Together*
- 9/2007 – Corrado Peperoni e Francesca Tuzi – *Trattamenti monetari non pensionistici metodologia sperimentale per la stima degli assegni al nucleo familiare*
- 10/2007 – AA.VV. – *Seminario sulla qualità: l'esperienza dei referenti del sistema informativo SIDI - 2ª giornata*

11/2007 – Leonello Tronti – *Il prototipo (numero 0) dell'Annuario di statistiche del Mercato del Lavoro (AML)*

12/2007 – Daniele Frongia, Raffaello Martinelli, Fernanda Panizon, Bruno Querini e Andrea Stanco – *Il nuovo Sistema informatico Altri Servizi. Progetto di reingegnerizzazione dei processi produttivi delle indagini trimestrali di fatturato degli altri servizi*

1/2008 – Carlo De Gregorio, Stefania Fatello, Rosanna Lo Conte, Stefano Mosca, Francesca Rossetti – *Sampling design and treatment of products in Istat centralised CPI surveys*