

istat working papers

N. 8
2011

La rilevazione Istat sui prezzi dei prodotti agricoli acquistati e venduti: innovazioni tecnologiche e metodologiche

Antonietta D'Amore, Orietta Luzi e Franca Silvestri

istat working papers

N. 8
2011

La rilevazione Istat sui prezzi dei prodotti agricoli acquistati e venduti: innovazioni tecnologiche e metodologiche

Antonietta D'Amore, Orietta Luzi e Franca Silvestri

Comitato di redazione

Coordinatore: Giulio Barcaroli

Componenti:

Rossana Balestrino	Francesca Di Palma	Luisa Picozzi
Marco Ballin	Alessandra Ferrara	Mauro Politi
Riccardo Carbini	Angela Ferruzza	Alessandra Righi
Claudio Ceccarelli	Danila Filipponi	Luca Salvati
Giuliana Coccia	Cristina Freguja	Giovanni Seri
Fabio Crescenzi	Aurea Micali	Leonello Tronti
Carla De Angelis	Nadia Mignolli	Sonia Vittozzi

Segreteria:

Lorella Appolloni, Maria Silvia Cardacino, Laura Peci, Gilda Sonetti, Antonio Trobia

Istat Working Papers

**La rilevazione Istat sui prezzi dei prodotti agricoli acquistati e venduti:
innovazioni tecnologiche e metodologiche**

N. 8/2011

ISBN 88-458-1690-7

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

La rilevazione Istat sui prezzi dei prodotti agricoli acquistati e venduti: innovazioni tecnologiche e metodologiche

Antonietta D'Amore, Orietta Luzi e Franca Silvestri¹

Sommario

In questo lavoro vengono illustrate le principali innovazioni tecnologiche e metodologiche della nuova procedura di trattamento delle mancate risposte e calcolo degli indici nella rilevazione dei prezzi dei prodotti agricoli acquistati e venduti condotta dall'Istat. In particolare, è stato adottato un nuovo approccio di stima delle quotazioni mensili mancanti di tipo longitudinale. Anche dal punto di vista tecnologico la nuova procedura presenta elementi innovativi: in essa infatti risultano integrati diversi linguaggi e ambienti operativi e di sviluppo (in particolare, in ambiente Linux Oracle/PHP e in ambiente Windows Oracle/Jsp e R). Nel lavoro sono quindi presentati i risultati di un'applicazione sperimentale per la valutazione dei potenziali effetti distorsivi dovuti alla presenza di prezzi mancanti e alla loro ricostruzione, a livello sia di dati elementari sia di indici sintetici.

Parole chiave: indici dei prezzi, mancate risposte, imputazione.

Abstract

In this paper we describe the main technological and methodological innovations of the new procedure for non responses imputation, and price indices computation for the Istat survey on Prices of Agricultural Products. In particular, a new longitudinal approach to the prediction of the monthly missing information has been adopted. The new statistical production procedure is innovative from a technological point of view, too: different languages and operating and development environments are actually integrated in it (in particular, for Linux Oracle/PHP, and for Windows for Oracle/Jsp and R). The paper also contains the results of an experimental application to the survey data aiming at evaluating the potential biasing effects of missing data and imputation on both elementary data and price indices.

Keywords: price indexes, missing data, imputation.

1. Introduzione

Gli obiettivi di miglioramento continuo della qualità dell'informazione statistica prodotta si traducono sempre più, negli Istituti Ufficiali di Statistica, in un maggior controllo verso tutte le fasi di formazione del dato statistico, da un punto di vista sia operativo, sia tecnologico, sia metodologico.

In questo contesto, si assiste ad una sempre maggiore evoluzione dei processi di rilevazione dei dati verso sistemi informativi strutturati e documentati, in cui il controllo della qualità delle informazioni raccolte avviene sin dalle fasi iniziali di formazione del dato. Il presente lavoro si inserisce in questa ottica produttiva, in cui assume importanza fondamentale la progettazione iniziale dell'intero sistema di indagine, dall'acquisizione al trattamento dei dati alla produzione dei parametri obiettivo. In particolare, in questo lavoro viene illustrato l'approccio adottato per la riprogettazione della rilevazione dei prezzi dei prodotti agricoli acquistati e venduti che ha

¹ A. D'Amore, Istat (damore@istat.it), par. 1,2, 3.2, 3.3 (in collaborazione).

O. Luzi, Istat (luzi@istat.it), par. 3.3 (in collaborazione), 4, 5.

F. Silvestri, Istat (silvestr@istat.it), par. 3.1.

riguardato la procedura di acquisizione, controllo e trattamento dei dati nonché quella per il calcolo degli indici.

L'attività di riprogettazione ha condotto alla creazione di appositi strumenti web per la fase di acquisizione controllata dei dati, la migrazione tecnologica delle precedenti procedure di indagine in nuovi ambienti di sviluppo ed elaborazione, ma anche la revisione e l'ottimizzazione di alcune fasi cruciali del controllo degli errori e del trattamento delle mancate risposte. Da un punto di vista tecnologico, l'applicazione web è stata implementata in ambiente *PHP* e lavora su database Oracle, la procedura di trattamento automatico dei dati e calcolo degli indici è stata sviluppata in linguaggio *R*, mentre per il controllo dei prezzi (variazioni) anomali sono state sviluppate maschere di controllo interattive in ambiente *Jsp*. Da un punto di vista metodologico, è stato adottato un metodo di ricostruzione dei prezzi mancanti di tipo longitudinale che tiene conto sia del livello territoriale, sia delle variazioni medie di prezzo osservate in ogni specifico periodo di riferimento. Il metodo è stato implementato in procedure automatiche in ambiente *R*.

Il lavoro è strutturato come segue. Nel paragrafo 2 sono fornite informazioni generali sulla rilevazione dei prezzi dei prodotti agricoli. Il paragrafo 3 contiene una descrizione sintetica della procedura di acquisizione e trattamento dei dati e di calcolo degli indici, e la descrizione estesa della metodologia adottata per la ricostruzione delle quotazioni mancanti. Nel paragrafo 4 sono descritti i risultati di un'applicazione sperimentale volta a valutare la qualità delle imputazioni ottenute applicando con due versioni alternative della metodologia di imputazione adottata.

2. La rilevazione dei prezzi dei prodotti agricoli venduti e acquistati

Nel contesto delle statistiche sui prezzi, il settore dei prodotti agricoli rappresenta un settore molto peculiare soprattutto per la natura dei fenomeni rilevati e le modalità con cui essi vengono osservati.

La rilevazione dei prezzi dei prodotti agricoli per la produzione dei relativi indici di prezzo congiunturali e tendenziali² è una rilevazione armonizzata a livello Europeo, sia in termini di obiettivi sia di metodologie utilizzate: *“Agriculture is a branch of the European Union’s economy which has long been covered by a common policy. Information on the prices of products and the means of production are indispensable to allow individual targets in the EU agricultural policy to be determined, the necessary measures to be taken and the effects of the policy to be monitored”* (Eurostat, 2008).

L'Istituto Nazionale di Statistica calcola e pubblica due serie di indici dei prezzi per il settore agricolo:

1. *Indice dei prezzi dei prodotti acquistati dagli agricoltori*: indicatore che misura le variazioni dei prezzi dei beni e dei servizi acquistati dagli agricoltori per lo svolgimento della loro attività produttiva;
2. *Indice dei prezzi dei prodotti venduti dagli agricoltori* indicatore che misura le variazioni dei prezzi praticati dagli agricoltori per la vendita dei propri prodotti.

Per quanto riguarda l'indice dei prezzi dei prodotti acquistati dagli agricoltori i prodotti oggetto di osservazione sono 153. Gli Uffici di statistica delle Camere di commercio raccolgono e trasmettono mensilmente 3.326 quotazioni di prezzo, che concorrono al calcolo di 122 indici nazionali di prodotto; per ulteriori 31 prodotti gli indici di prezzo sono ottenuti da altre indagini Istat. Il sistema dei pesi utilizzato è costruito considerando le voci che concorrono alla determinazione dei costi intermedi e degli investimenti del settore Agricoltura così come determinato dalla Contabilità Nazionale per i Conti economici dell'agricoltura relativamente al triennio 2004-2006.

² *Variazione congiunturale*: variazione percentuale rispetto al mese o al periodo precedente.

Variazione tendenziale: variazione percentuale rispetto allo stesso mese o periodo dell'anno precedente.

Per l'indice dei prezzi dei prodotti venduti dagli agricoltori sono rilevati i prezzi di 100 prodotti, per un totale di circa 3.750 quotazioni. I pesi utilizzati sono proporzionali alla media del valore della produzione agricola annua per il triennio 2004-2006. In particolare occorre considerare che gli indici dei prodotti venduti sono calcolati utilizzando un sistema di pesi diverso ogni mese per Frutta e Ortaggi, data l'elevata stagionalità che caratterizza i prezzi dei prodotti ortofrutticoli. Si tratta infatti di prodotti i cui periodi di produzione e commercializzazione possono essere molto brevi. Per ovviare a questo problema per ognuno dei dodici mesi dell'anno viene fissato un paniere variabile per frutta e ortaggi e viene quindi calcolato un diverso sistema di pesi per ciascun mese che resta invariato fino al successivo cambiamento della base di riferimento dell'indice di ribasamento. Gli indici mensili per frutta e ortaggi sono ottenuti aggregando gli indici semplici dei prezzi dei singoli prodotti e utilizzando i pesi che si applicano in ciascun mese, il peso per l'intero gruppo merceologico resta invariato per ciascun mese ma cambia la sua composizione e la percentuale di incidenza dei prodotti al suo interno. Va quindi evidenziato che le variazioni congiunturali degli indici sintetici riflettono sia effettive variazioni di prezzo, sia la stagionalità del sistema di ponderazione impiegato.

I prezzi utilizzati ai fini del calcolo dei due indici sono rilevati con cadenza mensile in 86 province dalle Camere di Commercio e il paniere di prodotti di ciascuna provincia è stato individuato considerando per il triennio a cavallo dell'anno base (anni 2004-2006) la distribuzione territoriale della produzione agricola per i prodotti venduti, e quella dei beni e servizi intermedi e dei beni di investimento per i prodotti acquistati.

A partire da gennaio 2009 gli indici sono calcolati in base 2005=100 così come richiesto dall'ufficio di statistica dell'Unione Europea (EUROSTAT) e la struttura e i livelli di sintesi sono coerenti con la struttura dei Conti Economici dell'Agricoltura.

Gli indici sono calcolati utilizzando la formula di Laspeyres, sono cioè a base fissa con ponderazione riferita all'anno base, e che può essere scritta nel modo seguente:

$$I_t^l = \frac{\sum_{i=1}^k p_t^i q_0^i}{\sum_{i=1}^k p_0^i q_0^i} \times 100 = \sum_{i=1}^k \frac{p_t^i}{p_0^i} \times \frac{p_0^i q_0^i}{\sum_{i=1}^k p_0^i q_0^i} \times 100 \quad (1)$$

dove:

- p_0^i e q_0^i sono rispettivamente il prezzo e la quantità del prodotto i al tempo 0 assunto come base;
- I è l'indice;
- p sono i prezzi dei prodotti;
- q sono le quantità di prodotti venduti (o le quantità acquistate dei mezzi di produzione);
- t è il periodo di osservazione;
- 0 è il periodo base.

Le fasi di calcolo degli indici mensili possono essere così sintetizzate: gli indici elementari di prodotto sono calcolati rapportando i prezzi mensili rilevati in ciascuna piazza al prezzo medio dell'anno base, quest'ultimo ottenuto come media aritmetica dei prezzi mensili dello stesso anno. Gli indici regionali di prodotto sono ottenuti come medie aritmetiche semplici degli indici elementari. L'indice nazionale di prodotto si ottiene come media aritmetica ponderata degli indici regionali. Infine, gli indici nazionali di sintesi, ovvero relativi a raggruppamenti successivi di prodotti, si ottengono come medie aritmetiche ponderate degli indici nazionali. Gli indici trimestrali e annuali di prodotto sono ottenuti mediante medie aritmetiche ponderate degli indici mensili. Infine aggregando gli indici trimestrali o annuali di prodotto, sempre utilizzando una media aritmetica ponderata, si ottengono rispettivamente gli indici trimestrali e annuali per i vari livelli di sintesi. Per gli indici nazionali di prodotto e di sintesi è fornita trimestralmente una stima provvisoria. La versione definitiva viene diffusa congiuntamente a quella della media annuale.

3. La nuova procedura di acquisizione dei prezzi, trattamento dei dati e calcolo degli indici

In questo paragrafo viene descritta la struttura generale del nuovo sistema di produzione degli indici dei prezzi agricoli, della procedura di acquisizione dei prezzi, e della nuova procedura di trattamento dei dati e di calcolo degli indici realizzate nel corso del 2010. Vengono quindi descritti gli algoritmi di imputazione adottati per la ricostruzione delle quotazioni mensili mancanti.

3.1. Struttura e caratteristiche generali del nuovo sistema di produzione degli indici dei prezzi agricoli

Lo schema generale del nuovo sistema di produzione degli indici dei prezzi dei prodotti agricoli venduti e acquistati è riportato nella Figura 1.

Parliamo di sistema in quanto il nuovo processo di produzione è il risultato dell'integrazione di diversi strumenti e ambienti di sviluppo: l'acquisizione dei dati è sviluppata in ambiente LINUX/APACHE/PHP, i moduli per il calcolo degli indici, per il controllo dei dati anomali e degli errori, e per l'imputazione dei dati mancanti sono stati sviluppati in ambiente Windows con linguaggio *R* (versione 2.10.1),³ il controllo interattivo dei dati avviene sulla base di maschere sviluppate in ambiente Windows/*Jsp*; tutte le funzioni operano su una base dati *Oracle* (*DB Oracle* nel seguito). Inoltre, nel nuovo sistema risultano integrati passi interattivi (in PHP o in *Jsp*) e automatici (in *R*).

Come meglio descritto nel paragrafo 3.2., il *DB Oracle*, contenente le quotazioni di prezzo e che risiede sul server Creta, (Figura 1) viene aggiornato mensilmente dalla Camere di Commercio con l'aiuto di una procedura di acquisizione controllata (vedere paragrafo 3.2). L'insieme dei dati aggiornati nel *DB Creta* viene trasferito nel *DB Spagna* (*DB Spagna* nella Figura 1) che costituisce l'insieme di partenza per il calcolo dei micro indici una volta che tutte le informazioni sono state validate. Il *DB Spagna* viene infatti sottoposto ad una serie di elaborazioni volte alla verifica della qualità del materiale raccolto, per poi procedere alla stima dei valori mancanti, al calcolo degli indici di sintesi (o aggregati). I moduli *Micro.R*, *Imputaz.R*, *SintesiX.R*,⁴ effettuano rispettivamente tali operazioni, e sono richiamati mediante un'Applicazione *Jsp/R* accessibile dagli operatori autorizzati. In ogni momento del processo, i microdati possono essere visualizzati attraverso maschere. La visualizzazione è offerta per prodotto o per provincia.

La fase di elaborazione vera e propria dei dati ha inizio avviando le funzioni del modulo *Micro.R*, il quale effettua il calcolo degli indici elementari su cui fa perno l'individuazione dei dati anomali da sottoporre a revisione interattiva e dei casi errati.

I prezzi anomali per ogni prodotto previsto nel mese di riferimento sono individuati sulla base della variazione percentuale (*V*) dell'indice elementare al mese *t* rispetto all'indice elementare del mese *t-1*: ad ogni quotazione osservata viene associato un *flag* (*f*) tale che:

$$f = \left\{ \begin{array}{ll} 1 \text{ se} & 15 < V \leq 25 \\ 2 \text{ se} & V > 25 \\ 3 \text{ se} & -25 \leq V < -15 \\ 4 \text{ se} & V < -25 \\ 0 & \text{altrimenti} \end{array} \right.$$

I valori del *flag* possono essere utilizzati per selezionare l'insieme dei microdati su cui procedere al controllo interattivo per verificare se si tratti di dati anomali.

³ Utilizzando i package RODBC (<http://cran.r-project.org>), xlsReadWrite e reshape.

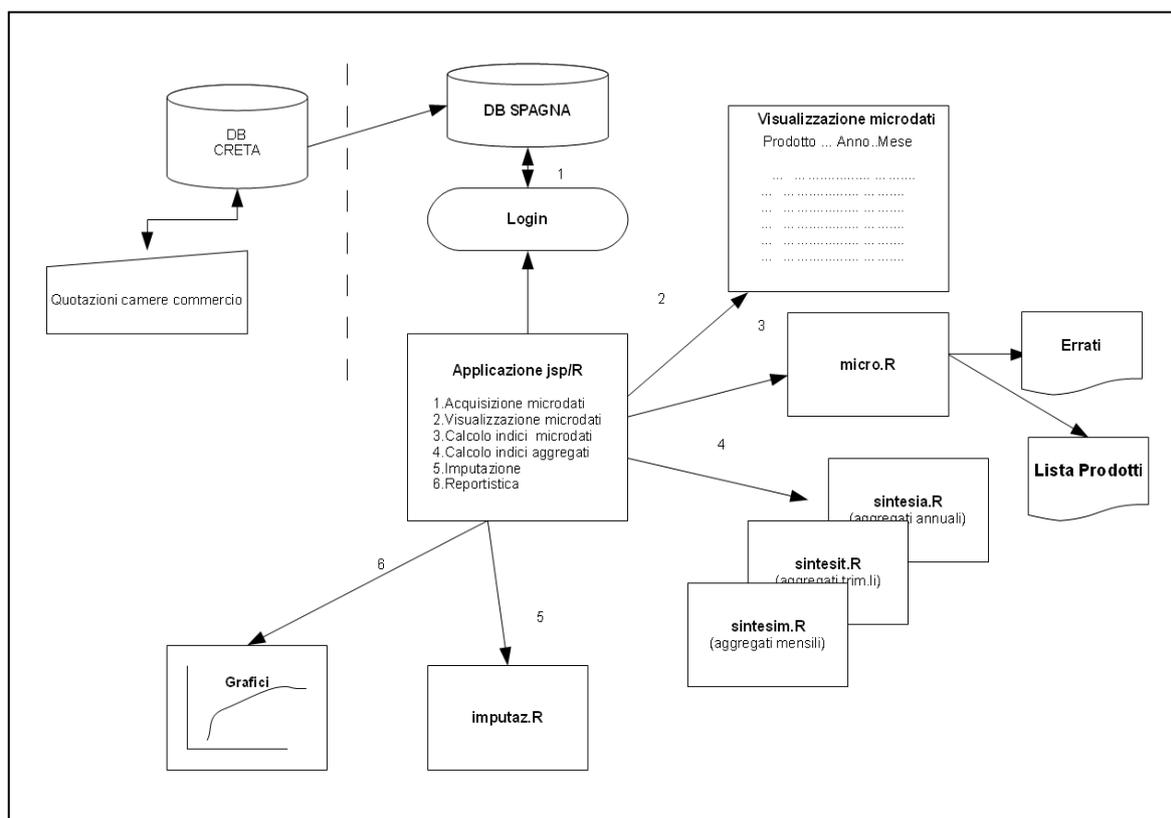
⁴ X=a (aggregati annuali), X=m (aggregati mensili), X=t (aggregati trimestrali).

Il modulo Micro.R effettua inoltre un controllo sulla validità dei dati, ad esempio vengono classificati come errori tutti i casi in cui l'indice è ≥ 999 oppure ≤ 10 . Questi casi vengono segnalati in uno specifico report (Errati in Figura 1).

In fase di verifica interattiva, effettuata attraverso opportune maschere di controllo (modulo Visualizzazione microdati in Figura 1), i prezzi classificati come anomali potranno essere confermati dall'operatore (e pertanto non subiranno modifiche nelle fasi successive del processo), ovvero corretti manualmente, o infine affidati ad un algoritmo di imputazione automatica (ed associati ad un valore $f=5$). A valle di questo processo viene prodotta una specifica reportistica (Lista Prodotti in Figura 1) contenente, per ogni prodotto, informazioni su:

- numero di quotazioni osservate ($f=1\div 4$);
- numero di quotazioni non osservate (da sottoporre a imputazione automatica);
- numero di quotazioni per valori di f dopo la fase interattiva (quindi inclusi valori con $f=5$, cioè prezzi anomali da sottoporre a imputazione automatica);
- numero di prezzi anomali confermati.

Figura 1 - La nuova procedura di trattamento dati e calcolo degli indici sintetici dei prezzi



La previsione dei valori non osservati e dei valori anomali non confermati e non corretti manualmente (cioè aventi $f=5$), è realizzata (modulo Imputaz.R in Figura 1) attribuendo alle quotazioni mancanti la variazione percentuale osservata sulle quotazioni disponibili nel mese di riferimento (l'algoritmo di imputazione utilizzato è descritto in dettaglio nel paragrafo 3.3).⁵ Al termine di tale processo, viene prodotta della reportistica (Grafici in Figura 1) contenente informazioni sull'entità delle imputazioni effettuate, quindi viene automaticamente effettuato il ricalcolo degli indici di sintesi.

⁵ Una quota di valori potrebbe non risultare ricostruibile a causa di errori residui nella base dati.

3.2. Il nuovo sistema di acquisizione dati

La registrazione e la trasmissione mensile dei dati delle indagini sui prezzi agricoli da parte delle Camere di Commercio avviene attraverso una pagina web (<https://indata.istat.it/preagr>) che garantisce l'autenticità del sito e la protezione dei dati trasmessi mediante il protocollo SSL.

Il nuovo ambiente *user-friendly* è stato sviluppato in linguaggio *PHP*. A partire da una struttura ad albero che raggruppa l'insieme dei prodotti facenti parte del paniere di ogni singola provincia, si accede ad una maschera riepilogativa in cui sono riportati i dati inviati negli ultimi tre anni, consentendo così una rapida verifica anche longitudinale dell'informazione raccolta.

Selezionando il mese di rilevazione si accede alla maschera di registrazione del singolo prezzo su cui sono riportate alcune informazioni caratterizzanti il prodotto: varietà, unità di misura rilevata, quantità rilevata. Nella finestra di dialogo un campo specifico prevede la segnalazione di cambiamenti di caratteristiche qualitative e/o commerciali del prodotto per evitare che siano considerate come variazioni di prezzo modificazioni dovute a cambiamenti della qualità dei prodotti o delle loro condizioni di vendita o di acquisto, in questo modo sarà possibile calcolare un fattore di correzione per concatenare la nuova serie di indici con quella precedente.

Nel corso della fase di acquisizione dei dati sono attivi alcuni controlli finalizzati alla segnalazione di quei dati che presentano una eccessiva variazione rispetto al periodo precedente: l'inserimento di opportune note giustificative consente il proseguimento delle operazioni.

L'insieme dei dati così ottenuto è archiviato nella banca dati Oracle (DB Creta nella Figura 1) alimentata dalle sole quotazioni mensili delle Camere di Commercio e costituisce l'insieme di partenza per la costruzione degli indici mensili dei prezzi dei prodotti acquistati e dei prodotti venduti dagli agricoltori.

3.3. La nuova procedura di imputazione delle quotazioni mancanti

La presenza di osservazioni mancanti nei dati forniti mensilmente dalle Camere di Commercio implica l'adozione, nell'ambito del processo di elaborazione e sintesi dei dati, di metodi di previsione (*imputazione*) per il completamento della base informativa su cui sono calcolati gli indici di prezzo ai vari livelli di sintesi (Little and Rubin, 1987).

Nel contesto della rilevazione sui prezzi dei prodotti agricoli Eurostat (2008) adotta la seguente definizione di osservazioni mancanti: *Member States often fail to collect the intended number of representative prices (in other words, when there is weight but the corresponding price data is not available), which leads to "missing observations"*.

La procedura di ricostruzione delle informazioni mancanti è in parte automatica, in parte interattiva. Quest'ultima soluzione si rende necessaria in caso di prodotti particolarmente critici, soprattutto in termini di numero medio mensile di quotazioni disponibili.

Per quanto riguarda l'imputazione di tipo automatico, sempre Eurostat (2008) suggerisce alcuni metodi di ricostruzione fra quelli maggiormente utilizzati dagli Stati Membri:

- a) ripetizione dell'ultima quotazione osservata. In caso di un elevato tasso di variazione, può essere appropriato adattare l'ultima quotazione osservata;
- b) ripetizione dell'ultima quotazione osservata applicando ad essa il normale andamento stagionale;
- c) imputazione delle variazioni di prezzo sulla base dei prezzi osservati in altri mercati (*domini*) per lo stesso prodotto.

Nel caso specifico della rilevazione Istat, si è scelto di adottare un approccio di tipo *misto*, in cui ogni quotazione mancante viene predetta "stimando" la corrispondente quotazione mediante il calcolo della variazione degli indici elementari per prefissati domini territoriali. Tali domini variano a seconda del contesto informativo specifico del singolo prodotto nello specifico periodo di riferimento. Nella definizione dell'algoritmo di previsione, inoltre, si è fatto riferimento alla suddivisione dei prodotti agricoli oggetto della rilevazione in due sottogruppi: *Ortofrutta*, *Altri prodotti*. Ai fini dell'imputazione, infatti, la differenza fra i due gruppi è che i prodotti appartenenti alla classe *Altri* sono inclusi nel paniere per tutti i 12 mesi dell'anno, e il loro peso non cambia nel

corso dell'anno stesso; per questi beni la ricostruzione delle eventuali quotazioni mancanti può essere effettuata quindi a partire dal prezzo del mese precedente o comunque dall'ultimo prezzo osservato disponibile. I prodotti appartenenti alla classe *Ortofrutta*, invece, oltre ad avere pesi variabili nel corso dell'anno, sono caratterizzati da *stagionalità*, cioè sono inclusi nel paniere solo in periodi particolari dell'anno (*campagne*) (vedere allegato 1). Per la ricostruzione delle eventuali quotazioni mancanti, in questo caso è pertanto necessario tener conto di questa peculiarità.

Nei paragrafi seguenti gli algoritmi di ricostruzione delle quotazioni mancanti sono descritti separatamente per i prodotti appartenenti alle suddette categorie.

3.3.1. Metodi di ricostruzione delle quotazioni mancanti per la classe "Altri" prodotti

Supponiamo di essere nella situazione di assenza del prezzo di una varietà j di un prodotto k , nella provincia P al mese t . Si abbia la serie delle 12 quotazioni precedenti per quella stessa varietà j :

$t-12$	$t-11$	$t-10$	$t-9$	$t-8$	$t-7$	$t-6$	$t-5$	$t-4$	$t-3$	$t-2$	$t-1$	t
p_{t-12}^j	p_{t-11}^j	p_{t-10}^j	p_{t-9}^j	p_{t-8}^j	p_{t-7}^j	p_{t-6}^j	p_{t-5}^j	p_{t-4}^j	p_{t-3}^j	p_{t-2}^j	p_{t-1}^j	$p_t^j = 0$

Nello schema seguente riportiamo la situazione generale al tempo t per un prodotto k avente $j=(1,2,3)$ tre diverse varietà:

Varietà	t	$t-1$	I_t	I_{t-1}
1		0	p_{t-1}^1	$I_{t-1}^1 = \frac{p_{t-1}^1}{p_0^1}$
2		p_t^2	p_{t-1}^2	$I_t^2 = \frac{p_t^2}{p_0^2}$ $I_{t-1}^2 = \frac{p_{t-1}^2}{p_0^2}$
3		p_t^3	p_{t-1}^3	$I_t^3 = \frac{p_t^3}{p_0^3}$ $I_{t-1}^3 = \frac{p_{t-1}^3}{p_0^3}$

Nello schema, p_0^j rappresenta il prezzo base della varietà j . Ai fini della ricostruzione del prezzo mancante della varietà $j=1$ al mese corrente t , è necessario distinguere fra i casi illustrati di seguito.

Caso 1: sono state osservate più di due varietà del prodotto k nella stessa Provincia⁶

L'osservazione mancante p_t^j si può ottenere applicando il metodo (iii) indicato da Eurostat, che consiste nell'adattamento del prezzo p_{t-1}^j della serie relativa a j del mese precedente, con la variazione media (*media geometrica*)⁷ da $t-1$ a t , degli indici delle altre varietà $l \in k$ ($l \neq j$):

$$\hat{p}_t^j = p_{t-1}^j \times \frac{\left[\prod_{l=1}^{n_{pk}} \frac{p_t^l}{p_0^l} \right]^{1/n_{pk}}}{\left[\prod_{l=1}^{n_{pk}} \frac{p_{t-1}^l}{p_0^l} \right]^{1/n_{pk}}} = p_{t-1}^j \times \frac{\left[\prod_{l=1}^{n_{pk}} I_t^l \right]^{1/n_{pk}}}{\left[\prod_{l=1}^{n_{pk}} I_{t-1}^l \right]^{1/n_{pk}}} \quad (2)$$

⁶ Col termine *osservate* si intende sia la presenza di altre varietà del prodotto per il mese di riferimento t , sia valori diversi da zero del loro prezzo.

⁷ In questa sezione si farà riferimento all'uso della media geometrica. Formule analoghe valgono nel caso di uso della media aritmetica dei microindici.

dove:

- $I_{t-1}^l = \frac{P_{t-1}^l}{P_0^l}$ e $I_t^l = \frac{P_t^l}{P_0^l}$, con P_0^l = prezzo base della varietà l ;
- n_{pk} è il numero delle altre varietà del prodotto k per le quali è disponibile un prezzo $\neq 0$ nella provincia P .

Caso 2: è stata osservata una sola altra varietà del prodotto k nella stessa Provincia

Se esiste una sola altra varietà l per la quale $P_t^l \neq 0$ nella provincia P al mese t , il prezzo p_t^j verrà stimato aggiustando il prezzo precedente sulla base della variazione del solo rapporto relativo ad l :

$$\hat{p}_t^j = p_{t-1}^j \times \frac{I_t^l}{I_{t-1}^l} = p_{t-1}^j \times \frac{\frac{P_t^l}{P_0^l}}{\frac{P_{t-1}^l}{P_0^l}} \quad (3)$$

Da un punto di vista computazionale, questo caso è incluso nel precedente. Un'alternativa che potrà essere oggetto di valutazione consiste nell'utilizzare, in questo caso, tutte le quotazioni p_t^j osservate a livello regionale (vedi caso 3), per evitare di "propagare" possibili variazioni "anomale" di singole quotazioni. Inoltre, in presenza di più prezzi mancanti per lo stesso prodotto nella stessa provincia, questi verrebbero imputati tutti con la stessa variazione, producendo un "appiattimento" in termini di variabilità. Tuttavia, essendo i comportamenti a livello provinciale in generale molto omogenei, l'uso del dato territoriale potrebbe risultare preferibile.

Caso 3: non sono disponibili osservazioni di altre varietà del prodotto k nella stessa Provincia, ma sono disponibili osservazioni di due o più varietà nella Regione

Se non sono disponibili osservazioni di altre varietà del prodotto nella stessa provincia P , il coefficiente di aggiustamento del prezzo precedente viene calcolato a livello regionale (metodo (iii) precedente), ossia come *media geometrica* dei micro-indici relativi alle n_{Rk} varietà del prodotto k con prezzo osservato $\neq 0$ disponibili nella stessa regione R :

$$\hat{p}_t^j = p_{t-1}^j \times \frac{\left[\prod_{l=1}^{n_{Rk}} I_t^l \right]^{\frac{1}{n_{Rk}}}}{\left[\prod_{l=1}^{n_{Rk}} I_{t-1}^l \right]^{\frac{1}{n_{Rk}}}} = p_{t-1}^j \times \frac{\left[\prod_{l=1}^{n_{Rk}} \frac{P_t^l}{P_0^l} \right]^{\frac{1}{n_{Rk}}}}{\left[\prod_{l=1}^{n_{Rk}} \frac{P_{t-1}^l}{P_0^l} \right]^{\frac{1}{n_{Rk}}}} \quad (4)$$

Caso 4: non esistono osservazioni di altre varietà del prodotto k nella stessa Provincia, ma è stata osservata una sola altra varietà del prodotto nella stessa Regione

Se nella stessa regione è disponibile un unico prezzo osservato, diciamo per la varietà l , l'imputazione avviene mediante la trasformazione [2] (computazionalmente inclusa nel caso 3 precedente). Anche in questo caso, vale l'alternativa di passare, in questa situazione, al livello territoriale successivo (caso 5): le due opzioni saranno valutate sperimentalmente.

Caso 5: non sono state osservate altre varietà del prodotto k né nella stessa provincia, né nella stessa regione, ma sono state osservate almeno due varietà nella Ripartizione Territoriale

In questo caso, il coefficiente di variazione del prezzo precedente viene calcolato a livello di ripartizione territoriale. Se indichiamo con $RT(g)$ la *Ripartizione Territoriale* in esame⁸ l'integrazione del dato mancante avverrà come segue:

$$\hat{p}_t^j = p_{t-1}^j \times \frac{I_{RT(g);k}^t}{I_{RT(g);k}^{t-1}} \quad (5)$$

dove:

$$a) \quad I_{RT(g);k}^t = \frac{\sum_{R \in RT(g)} w_{R \in RT(g)} I_{R \in RT(g);k}^t}{\sum_{R \in RT(g)} w_{R \in RT(g)}} \quad \text{è la media aritmetica ponderata degli indici}$$

regionali $I_{R \in RT(g);k}^t$ (per le sole regioni R appartenenti alla ripartizione territoriale $RT(g)$) del prodotto k al mese t , definiti come:

$$I_{R;k}^t = \left(\prod_{l=1}^{n_{Rk}} \frac{p_l^t}{p_l^0} \right)^{\frac{1}{n_{Rk}}} = \left(\prod_{l=1}^{n_{Rk}} I_l^t \right)^{\frac{1}{n_{Rk}}}, \quad \text{dove per ogni regione } R \text{ appartenente alla ripartizione } RT(g), \text{ e } n_{Rk} \text{ è il numero di varietà } l \text{ del prodotto } k \text{ aventi prezzo } \neq 0 \text{ in essa disponibili;}$$

$$b) \quad w_{R \in RT(g)} \text{ è il peso di ogni regione } R \text{ appartenente alla ripartizione } RT(g) \text{ (} R=1, \dots, N_{R \in RT(g)} \text{)}.$$

Analoghe formule valgono per gli indici al mese $t-1$.

Caso 6: non sono state osservate altre varietà del prodotto k né nella stessa provincia, né nella stessa regione né nella stessa Ripartizione Territoriale

In questo caso, il coefficiente di variazione da applicare al prezzo precedente viene calcolato a livello Nazionale, come segue:

$$\hat{p}_t^j = p_{t-1}^j \times \frac{I_{Naz;k}^t}{I_{Naz;k}^{t-1}} \quad (6)$$

⁸ $g=1$ Nord-ovest, $g=2$ Nord-est, $g=3$ Centro, $g=4$ Sud e Isole

dove:

$$a) I_{Naz;k}^t = \frac{\sum_{R=1}^{N_R} w_R I_{R;k}^t}{\sum_{R=1}^{N_R} w_R}$$

è la *media aritmetica ponderata* degli indici regionali $I_{R;k}^t$ del

prodotto k al mese t , definiti come:

$$I_{R;k}^t = \left(\prod_{l=1}^{n_{Rk}} \frac{p_l^t}{p_l^0} \right)^{\frac{1}{n_{Rk}}} = \left(\prod_{l=1}^{n_{Rk}} I_t^l \right)^{\frac{1}{n_{Rk}}}, \text{ dove } R=1, \dots, N_R, \text{ e } N_R \text{ è il numero di regioni}$$

considerate per il prodotto k al mese t , dove n_{Rk} è il numero di varietà l del prodotto k aventi prezzo $\neq 0$ disponibili nella regione R ;

b) w_R è il peso della regione R ($R=1, \dots, N_R$).

Analoghe formule valgono per gli indici regionali e nazionale al mese $t-1$.⁹

3.3.2. Metodi di ricostruzione delle quotazioni mancanti per i prodotti della classe "Ortofrutta"

In questo caso, gli algoritmi di ricostruzione delle quotazioni mancanti sono sostanzialmente analoghi a quelli descritti per la classe di prodotti precedente.

La particolarità, per i prodotti di tipo *Ortofrutta*, sta nella necessità di selezionare il mese precedente $t-v$, fissato il quale è poi possibile calcolare i valori medi per il calcolo dei coefficienti di "variazione" del prezzo p_{t-v}^j nelle espressioni [1]-[5].

Come si vede dal prospetto riportato nell'Allegato 1, i prodotti stagionali sono caratterizzati da un mese di inizio campagna, da un mese di fine campagna, e da mesi intermedi. Il mese di riferimento $t-v$ corrisponderà, in tutti e tre i casi, all'ultimo mese precedente in cui è prevista la rilevazione del prodotto k , con l'avvertenza che:

- se il mese t è un mese di inizio campagna, il mese $t-v$ corrisponde al mese di fine campagna dell'ultima campagna precedente;
- se il mese t è un mese intermedio o un mese di fine campagna, il mese $t-v$ corrisponde al mese precedente.

I metodi di previsione del prezzo mancante saranno quindi analoghi a quelli illustrati nella sezione 3.1.1, sostituendo a $t-1$ il riferimento temporale $t-v$.

4. Una valutazione sperimentale

4.1 La sperimentazione

Al fine di valutare l'accuratezza del metodo di imputazione delle quotazioni mancanti adottato, rispetto ad altre possibili soluzioni alternative, è stato effettuato uno studio sperimentale su un sottoinsieme di prodotti rilevati nell'indagine. Questo studio ha l'obiettivo non solo di verificare i potenziali effetti distorsivi dovuti all'imputazione delle quotazioni mancanti, ma anche e soprattutto di individuarne le possibili cause, in modo da poter intervenire sugli eventuali elementi strutturali del processo di produzione statistica (in particolare, acquisizione e trattamento dati).

⁹ L'indice nazionale ai tempi t e $t-1$ per un certo prodotto k naturalmente va calcolato sulle stesse regioni.

Sono stati sottoposti a valutazione due metodi di imputazione alternativi:

- a) misura della variazione media delle quotazioni ad ogni specifico livello territoriale come rapporto fra *medie aritmetiche* degli indici;
- b) misura della variazione media delle quotazioni ad ogni specifico livello territoriale come rapporto fra *medie geometriche* degli indici.

In ciascuna delle due alternative, le variazioni medie sono state calcolate a partire dal livello provinciale, aggregando via via ai vari livelli territoriali superiori (*Regione, Ripartizione Territoriale, Nazione*) in caso di non disponibilità di quotazioni osservate a questo livello (e via via a quelli successivi) per i prodotti oggetto di imputazione.

La valutazione comparativa dei due metodi è stata effettuata sperimentalmente quantificando, per ciascuno di essi:

- la capacità di corretta ricostruzione dei prezzi mancanti (*livello micro*)
- la capacità di preservazione degli indici di sintesi (*livello aggregato*).

La valutazione è stata basata su un processo di simulazione controllata di prefissate percentuali di prezzi mancanti in un insieme di dati rilevati validi, e sul confronto fra questi ultimi e i corrispondenti prezzi stimati mediante ciascuno dei due metodi di imputazione proposti. Il tasso di non risposta è simulato in modo da riprodurre quello effettivo della rilevazione per ogni specifico prodotto/mese.

A livello *micro*, per ogni prodotto o gruppo di prodotti (*prod*), la valutazione della qualità della ricostruzione può essere misurata in termini di distanza relativa fra prezzi elementari originariamente osservati (*ril*) e imputati (*imp*) utilizzando la funzione (Eurostat, 2007):

$$ER^{prod} = \frac{\sum_{i=1}^{n_{prod}} |P_{ril,i} - P_{imp,i}|}{\sum_{i=1}^{n_{prod}} P_{ril,i}} \times 100 \quad (7)$$

detto *Errore di imputazione relativo*, dove n_{prod} è il numero di quotazioni originariamente rilevate nel mese di riferimento per il prodotto (o gruppo di prodotti) *prod* e annullate artificialmente. L'indice ER^{prod} fornisce un'indicazione sintetica dell'entità delle differenze dovute all'uso dello specifico modello di imputazione per lo specifico sottoinsieme di dati. L'indice potrebbe anche essere calcolato per diversi livelli territoriali al fine di valutare la rilevanza del fattore territorio sulla qualità delle imputazioni.

A livello *aggregato*, per ogni strategia di imputazione la valutazione dell'impatto sugli indici di sintesi può essere effettuata misurando la diversità fra gli indici di sintesi calcolati sui dati rilevati e su quelli imputati. In particolare, si può utilizzare la distanza relativa:

$$RB_{liv}^{prod} = \frac{I_{ril}^{prod(liv)} - I_{imp}^{prod(liv)}}{I_{ril}^{prod(liv)}} \times 100 \quad (8)$$

detta *Errore relativo di stima ("distorsione") da imputazione*, dove $I_{ril}^{prod(liv)}$ è l'indice di sintesi di (gruppo di) prodotto calcolato sui prezzi rilevati ai diversi livelli territoriali

(provinciale, regionale, ripartizionale, nazionale), e $I_{imp}^{prod(liv)}$ è il corrispondente indice calcolato sui prezzi imputati.

Per una valutazione più accurata degli effetti di ogni modello di imputazione, a livello sia *micro* sia *aggregato*, è necessario iterare per un certo numero m di volte l'operazione di imputazione e valutare la corrispondente distorsione attraverso la *media* degli indici [7] e [8] ottenuti nelle m iterazioni.

A livello *micro*:

$$ER^{mprod} = \frac{1}{m} \sum_{j=1}^m ER_j^{prod} \quad (9)$$

detto *Errore di imputazione relativo medio*, dove ER_j^{prod} è l'*errore di imputazione relativo* ottenuto alla j -esima iterazione.

Per quanto riguarda il *livello aggregato*, è stato considerato per semplicità il solo livello di sintesi nazionale degli indici, ma naturalmente la valutazione può essere effettuata anche ad altri livelli di sintesi per una verifica più dettagliata degli effetti dell'imputazione. Per la valutazione a livello aggregato sono state dunque utilizzate le seguenti misure:

$$RB_{naz}^{prod} = \frac{1}{m} \sum_{j=1}^m \frac{(I_{ril}^{prod(naz)} - I_{imp-j}^{prod(naz)})}{I_{ril}^{prod(naz)}} \times 100 \quad (Relative\ Estimation\ Error) \quad (10)$$

assimilabile ad una misura di distorsione, dove $I_{imp-j}^{prod(naz)}$ è l'indice di sintesi a livello nazionale calcolato sui prezzi imputati all'iterazione j -esima;

$$RMSE_{naz}^{prod} = \sqrt{\frac{1}{m} \sum_{j=1}^m \frac{(I_{ril}^{prod(naz)} - I_{imp-j}^{prod(naz)})^2}{(I_{ril}^{prod(naz)})^2}} \times 100 \quad (Relative\ Root\ Mean\ Squared\ Error) \quad (11)$$

La sperimentazione è stata condotta su tutti i mesi degli anni 2008 e 2009. Ai fini della sperimentazione sono stati presi in esame i seguenti gruppi di prodotti:¹⁰

1. latte (3 prodotti)
2. frutta (8 prodotti)
3. bovini (3 prodotti)
4. mangimi (22 prodotti)
5. frumento (2 prodotti)

¹⁰ Va sottolineato che nella procedura effettiva, ogni serie di prodotto viene trattato separatamente ai fini dell'imputazione. I risultati sono pertanto da interpretare alla luce di tale circostanza.

Per ogni mese/anno considerato e per ogni prodotto la procedura di valutazione si articola nei seguenti passi:

1. *selezione delle sole osservazioni non originariamente mancanti e non soggette ad alcuna modifica nel corso della procedura di verifica/validazione dei dati (unità di misura, cambi, ecc.);*
2. *simulazione sull'insieme di dati di cui al punto precedente di una proporzione di prezzi nulli (determinata sulla base della percentuale di quotazioni nulle sul mese/anno corrispondente) secondo un meccanismo completamente casuale;¹¹*
3. *imputazione dei prezzi nulli sulla base dei modelli di imputazione alternativi;*
4. *iterazione dei passi 1÷3 per $m=100$ volte;*
5. *valutazione dell'effetto della non risposta e dell'imputazione sulla base delle misure introdotte.*

4.2 I risultati

Prima di procedere all'applicazione sperimentale dei metodi di imputazione descritti nei paragrafi precedenti, è stata fatta un'analisi esplorativa delle frequenze delle quotazioni di prezzo mancanti per gli anni, i mesi e i gruppi di prodotti considerati (paragrafo 4.1). Ciò al fine non solo di verificare l'entità del fenomeno da trattare, in modo da avere un elemento di valutazione aggiuntivo in fase di analisi dei risultati, ma anche ai fini della generazione dei dati mancanti nello studio sperimentale. Nelle Tavole 1 e 2 seguenti, per ogni gruppo (o *categoria*) di prodotto e mese di riferimento sono riportati il numero di quotazioni osservate e le relative percentuali di prezzi mancanti, rispettivamente negli anni 2008 e 2009. Come si può osservare, passando dal 2008 al 2009, per alcuni prodotti la frequenza di quotazioni mancanti diminuisce, anche notevolmente (in particolare per *Latte*, *Bovini*, *Mangimi*), a conferma di una maggiore collaborazione delle Camere di Commercio alla rilevazione. Fa eccezione il gruppo *Frutta*, per il quale si osservano frequenze medie di prezzi mancanti molto elevate (intorno al 50-60%) in entrambi gli anni.

Il problema della frequenza di prezzi mancanti, come è noto, è cruciale in termini di effetto distorsivo sui parametri oggetto di pubblicazione (in questo caso, indici sintetici di prezzo), per cui il problema principale per specifici prodotti e/o per specifici mesi diventa comprenderne le eventuali cause strutturali. Ciò vale appunto per il gruppo di prodotti *Frutta*, per il quale si ritiene necessaria una valutazione più approfondita della natura delle non risposte prima di procedere ad approfondimenti metodologici legati alla loro ricostruzione statistica. Più in generale, l'analisi dei tassi di mancata risposta dovrebbe tener conto del fattore territoriale, al fine di identificare un'eventuale dipendenza del fenomeno da questo elemento. Si rimanda a studi specifici successivi su caratteristiche e distribuzione delle mancate risposte sui vari gruppi di prodotto coinvolti nella rilevazione, in particolare sulla categoria *Frutta*.

Come già detto, la valutazione della distorsione potenziale sugli indici sintetici, effettuata con riferimento all'indice nazionale e per alcuni specifici prodotti, è stata effettuata simulando, per ogni anno, mese e prodotto percentuali di dati mancanti analoghe alle corrispondenti frequenze effettive. I risultati ottenuti sulle $m=100$ iterazioni per i prodotti presi in considerazione sono riportati nelle tabelle 3÷8 seguenti. In ogni tabella, per ogni anno e per ogni mese sono riportati i valori degli indicatori ER^{mprod} (*Errore di imputazione relativo medio*, *ER* nel seguito per semplicità), RB_{naz}^{prod} (*Relative Estimation Error*, *RB* nel seguito), e $RMSE_{naz}^{prod}$ (*Relative Root Mean Squared Error*, *RMSE* nel seguito) calcolati per ognuno dei due metodi di imputazione analizzati (sintesi degli indici elementari mediante *media aritmetica – metodo A*, oppure mediante *media geometrica – metodo G*).

¹¹ Analisi per meccanismi diversi potranno essere effettuate sulla base di ulteriori verifiche della dipendenza delle mancate risposte da elementi "strutturali", come ad esempio la localizzazione territoriale.

Tavola 1 - Frequenze assolute di quotazioni osservate e frequenze % di quotazioni mancanti per mese e (gruppo di) prodotto - Anno 2008

MESE	Prodotto/Gruppo di prodotto											
	Latte		Frutta		Bovini		Mangimi		Concimi		Frumento	
	Tot.	%Mis.	Tot.	%Mis.	Tot.	%Mis.	Tot.	%Mis.	Tot.	%Mis.	Tot.	%Mis.
Gennaio	55	25,45	125	49,6	269	17,47	506	26,28	634	16,56	80	8,75
Febbraio	55	23,64	125	51,2	269	18,22	506	26,28	634	17,82	80	12,50
Marzo	55	21,82	118	51,69	269	18,22	506	24,51	634	17,51	80	12,50
Aprile	55	20,00	118	54,24	269	21,56	506	25,69	634	20,03	80	15,00
Maggio	55	20,00	120	53,33	269	17,47	506	28,85	634	21,14	80	18,75
Giugno	55	20,00	355	62,82	269	16,36	506	27,47	634	17,82	80	36,25
Luglio	55	27,27	394	66,75	269	15,99	506	25,89	634	18,45	80	20,00
Agosto	55	21,82	353	62,61	269	16,73	506	28,06	634	18,30	80	17,50
Settembre	55	20,00	382	58,64	269	15,99	506	25,89	634	19,24	80	10,00
Ottobre	55	18,18	213	45,54	269	16,73	506	25,69	634	19,87	80	10,00
Novembre	55	21,82	213	56,34	269	19,70	506	28,46	634	21,77	80	11,25
Dicembre	55	18,18	213	60,09	269	17,10	506	26,28	634	19,09	80	8,75
Media		21,52		56,07		17,63		26,61		18,97		15,10

Tavola 2 - Frequenze assolute di quotazioni osservate e frequenze % di quotazioni mancanti per mese e per (gruppo di) prodotto - Anno 2009

MESE	Prodotto/Gruppo di prodotto											
	Latte		Frutta		Bovini		Mangimi		Concimi		Frumento	
	Tot.	%Mis.	Tot.	%Mis.	Tot.	%Mis.	Tot.	%Mis.	Tot.	%Mis.	Tot.	%Mis.
Gennaio	55	7,27	125	50,40	269	12,27	506	8,50	634	14,51	80	10,00
Febbraio	55	10,91	125	52,00	269	13,75	506	8,70	634	13,72	80	10,00
Marzo	55	10,91	118	55,93	269	12,27	506	8,89	634	13,41	80	10,00
Aprile	55	9,09	118	56,78	269	10,04	506	8,70	634	12,62	80	8,75
Maggio	55	10,91	120	56,67	269	12,64	506	8,89	634	12,93	80	11,25
Giugno	55	10,91	355	59,44	269	9,29	506	9,88	634	11,83	80	42,50
Luglio	55	12,73	394	64,47	269	9,29	506	9,09	634	12,78	80	22,50
Agosto	55	12,73	353	62,04	269	12,27	506	9,29	634	12,62	80	16,25
Settembre	55	10,91	382	57,07	269	9,67	506	8,10	634	10,09	80	8,75
Ottobre	55	10,91	213	50,23	269	9,67	506	9,09	634	10,41	80	8,75
Novembre	55	10,91	213	53,05	269	9,29	506	8,50	634	10,09	80	8,75
Dicembre	55	10,91	213	60,56	269	9,29	506	11,66	634	12,15	80	10,00
Media		10,76		56,55		10,81		9,11		12,26		13,96

In termini di effetto differenziale del metodo di imputazione (media aritmetica piuttosto che media geometrica per la sintesi dei microindici) non si registrano differenze rilevanti in nessuno dei tre indicatori e per nessuno dei due anni considerati. Il tipo di sintesi dei microindici nell'ambito del metodo di imputazione non risulta quindi rilevante in termini di distorsione prodotta né sui dati elementari, né sull'indice di prezzo a livello nazionale.

Alla luce di questo risultato, fissato metodo di imputazione (A , G), nel confronto fra i due anni di riferimento si osservano tendenzialmente valori più contenuti dei tre indicatori per il 2008 rispetto ai corrispondenti valori del 2009. Si riscontrano comunque un numero di situazioni diverse per alcuni prodotti che saranno discusse nel seguito.

Sempre in generale, a parità di metodo di imputazione (A , G) e di anno di riferimento, si osservano valori di RB ed $RMSE$ generalmente contenuti per la maggior parte dei prodotti.

Nel seguito i vari gruppi di prodotti sono analizzati separatamente.

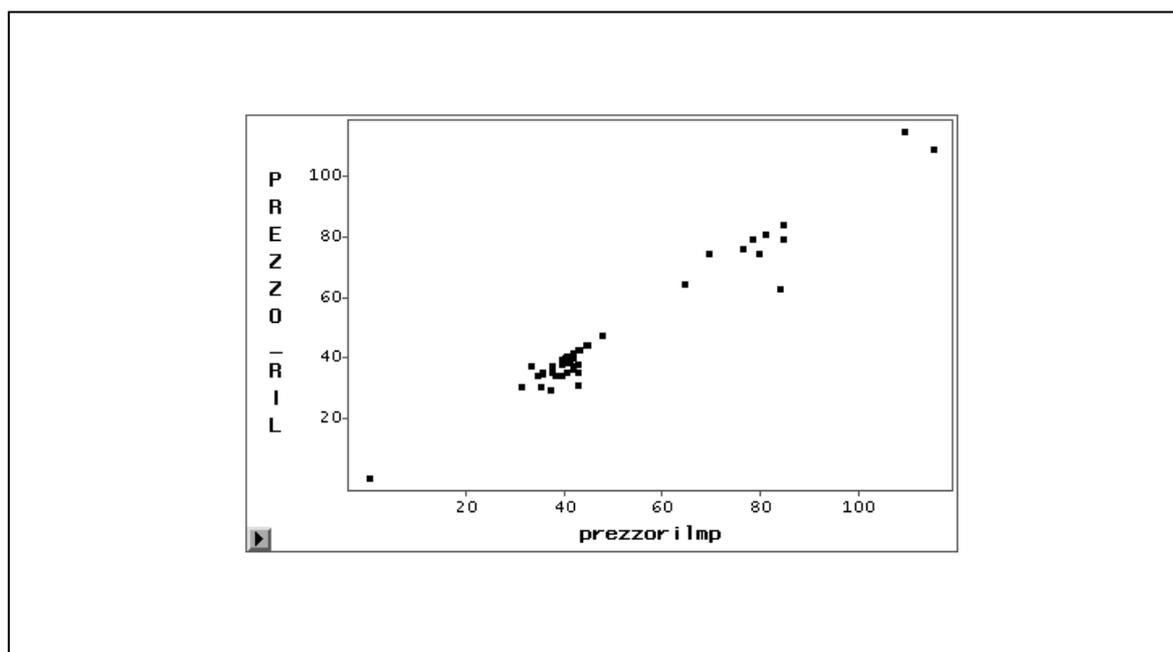
Per quanto riguarda il gruppo di prodotti *Latte*, la Tavola 3 riporta i valori degli indicatori di qualità per tutti i mesi dei due anni presi in esame. In termini di RB , si osserva un effetto distorsivo sull'indice sintetico generalmente trascurabile o comunque contenuto nella maggior parte dei mesi di entrambi gli anni (ad eccezione dei mesi di gennaio e ottobre 2009).

Tavola 3 - Latte: valori degli indicatori di qualità per anno e mese

LATTE	2008						2009					
	ER_A	ER_G	RB_A	RB_G	RMSE_A	RMSE_G	ER_A	ER_G	RB_A	RB_G	RMSE_A	RMSE_G
Gennaio	4,46	4,46	0,04	0,04	0,71	0,71	1,75	1,76	1,86	1,86	3,38	3,38
Febbraio	2,07	2,06	0,00	0,00	0,41	0,41	1,53	1,52	0,04	0,04	0,17	0,17
Marzo	1,33	1,33	0,09	0,09	0,24	0,24	1,44	1,47	0,08	0,09	0,24	0,25
Aprile	3,08	3,12	0,15	0,15	0,34	0,34	2,68	2,69	0,05	0,07	0,30	0,30
Maggio	2,27	2,27	0,07	0,06	0,25	2,57	0,77	0,79	-0,04	-0,05	0,25	0,25
Giugno	2,78	2,77	-0,12	-0,12	0,48	0,48	1,85	1,87	-0,03	-0,03	0,35	0,35
Luglio	1,47	1,46	-0,05	-0,05	0,41	0,41	1,98	2,01	-0,03	-0,02	0,18	0,19
Agosto	1,92	1,92	0,04	0,04	0,42	0,42	3,00	3,02	0,08	0,08	0,40	0,40
Settembre	1,89	1,89	0,10	0,10	0,27	0,27	0,99	0,99	-0,05	-0,05	0,25	0,25
Ottobre	1,12	1,10	-0,05	-0,05	0,24	0,24	0,58	0,56	3,01	3,01	4,27	4,27
Novembre	1,61	1,62	0,10	0,09	0,71	0,71	1,90	1,91	-0,12	-0,12	0,79	0,79
Dicembre	3,10	3,11	0,07	0,07	0,38	0,38	3,78	3,78	0,10	0,09	1,08	1,08

Ad esempio, considerando il mese di gennaio 2009, nella Figura 2 sono rappresentati i prezzi del mese in esame ($l_prezril$) rispetto al mese precedente (l_prezmp) (scala logaritmica).

Figura 2 - Latte: scatter plot dei prezzi osservati nel mese di gennaio 2009 vs i prezzi osservati nel mese precedente (scala logaritmica)



I tre cluster di valori corrispondono sostanzialmente ai 3 prodotti del gruppo considerati congiuntamente nella sperimentazione (Latte per uso alimentare, Latte per uso industriale, Latte di pecora) e determinano variabilità aggiuntiva nel processo di ricostruzione dei prezzi mancanti. Inoltre, si osserva (in particolare, nel cluster più consistente) la presenza di una sistematica differenza fra prezzi osservati nel mese corrente e i corrispondenti prezzi rilevati nel mese precedente. Questi fattori sono possibili cause degli effetti distorsivi del processo di imputazione sia a livello di indice sintetico, sia naturalmente in termini di RE (la differenza media in questo mese fra prezzi osservati nel mese corrente e nel mese precedente è pari al 3%). Valori consistenti del RE si osservano in effetti per vari mesi sia del 2008 sia del 2009: seguendo le considerazioni appena fatte, questo tipo di distorsione può in conclusione essere dovuto sia alla frequenza di prezzi mancanti per questi prodotti (Tavole 1 e 2), sia alla eterogeneità dei prodotti considerati congiuntamente, sia alla variazione dei

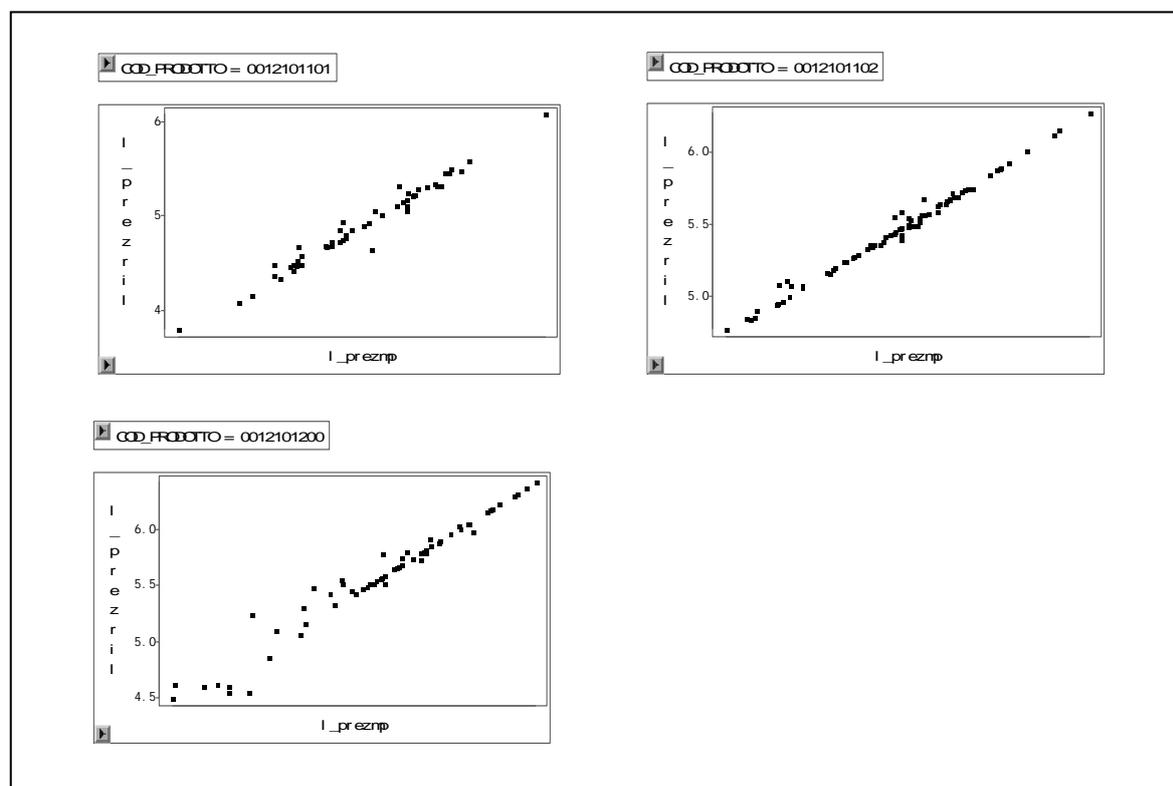
prezzi dal mese precedente al mese corrente. Al fine di verificare la presenza di un effetto variabilità interna dei prezzi per prodotto, per questo come per altre categorie di prodotti, verranno effettuate in seguito sperimentazioni relative ai singoli prodotti del gruppo.

Per quanto riguarda la voce *Bovini* (Tavola 4), si osservano valori contenuti di tutti gli indicatori per entrambi gli anni: *RB* e *RMSE* sono sempre inferiori all'1%, e il *RE* è generalmente inferiore al 2,5%. Fanno eccezione alcuni mesi del 2008 (*Luglio, Agosto, Settembre e Dicembre*), con valori compresi fra il 3% e il 4%, per un effetto combinato della frequenza di prezzi mancanti (molto maggiore della corrispondente frequenza nel 2009), e della variabilità interna dei prezzi: relativamente a questo ultimo aspetto, se si considera ad esempio il mese di *Settembre* 2008, dalla Figura 3 si può osservare come la distribuzione dei prezzi del mese corrente vs le corrispondenti quotazioni del mese precedente (sempre in scala logaritmica) presentino una forte variabilità, soprattutto per uno dei tre prodotti trattati congiuntamente.

Tavola 4 - Bovini: valori degli indicatori di qualità per anno e mese

BOVINI	2008						2009					
	ER_A	ER_G	RB_A	RB_G	RMSE_A	RMSE_G	ER_A	ER_G	RB_A	RB_G	RMSE_A	RMSE_G
Gennaio	1,41	1,43	0,39	0,40	0,80	0,80	1,69	1,70	0,60	0,60	1,08	1,08
Febbraio	1,21	1,24	0,53	0,53	0,99	0,98	1,45	1,45	-0,01	-0,01	0,08	0,08
Marzo	1,66	1,69	0,05	0,04	0,11	0,11	2,56	2,66	0,04	0,03	0,07	0,07
Aprile	1,23	1,25	0,35	0,35	0,90	0,90	2,78	2,86	0,11	0,10	0,17	0,16
Maggio	1,36	1,37	0,54	0,55	1,06	1,06	2,34	2,46	-0,01	-0,02	0,08	0,09
Giugno	1,65	1,63	-0,04	-0,04	0,11	0,11	1,93	1,92	0,04	0,04	0,09	0,09
Luglio	3,11	3,06	-0,02	-0,02	0,09	0,09	1,80	1,82	-0,03	-0,02	0,18	0,17
Agosto	3,52	3,64	-0,07	-0,05	0,14	0,13	0,99	0,99	0,05	0,05	0,08	0,08
Settembre	4,86	5,04	0,00	-0,01	0,13	0,13	1,23	1,26	0,27	0,27	0,82	0,82
Ottobre	1,93	2,02	0,01	0,00	0,10	0,09	1,11	1,16	0,04	0,04	0,16	0,16
Novembre	2,50	2,49	-0,02	-0,02	0,09	0,09	1,69	1,70	-0,06	-0,06	0,10	0,10
Dicembre	2,98	2,96	-0,12	-0,13	0,16	0,16	1,07	1,05	-0,03	-0,03	0,07	0,07

Figura 3 - Bovini: prezzi osservati nel mese di Settembre 2008 vs i prezzi osservati nel mese precedente (scala logaritmica)



Anche per la voce *Mangimi* (Tavola 5), si osservano in entrambi gli anni effetti sostanzialmente trascurabili dell'imputazione in termini di *RB* e *RMSE* (sempre inferiori al 2%). Anche in termini di *RE* si registrano valori generalmente contenuti per il 2009 (tra lo 0,5 e il 3%), anche collegati alla forte riduzione della frequenza di prezzi mancanti in questo anno. Per la maggior parte dei mesi del 2008 si osservano invece valori del *RE* piuttosto elevati (in particolare per il mese di *Ottobre 2008*).

Questo dato, in parte spiegabile dalla elevata frequenza di prezzi mancanti, può essere anche dovuto al fatto che i prodotti appartenenti al gruppo, che nella simulazione sono considerati congiuntamente, sono caratterizzati da prezzi medi e da andamenti mese precedente/mese corrente piuttosto divergenti (vedere Figure 4 e 5). Nel complesso, si noti che la differenza media relativa fra prezzi osservati a *Ottobre 2008* e prezzi osservati nel mese precedente è pari a 7,5%, a fronte di un rapporto fra le medie degli indici elementari, calcolate come descritto nel paragrafo 3.3.1, pari a 0,95: in altre parole, nella determinazione di un valore imputato, in questo caso incide maggiormente la variazione da *t* a *t-v* delle singole quotazioni, rispetto alla variazione media degli indici elementari.

Tavola 5 - *Mangimi*: valori degli indicatori di qualità per anno e mese

MANGIMI	2008						2009					
	ER_A	ER_G	RB_A	RB_G	RMSE_A	RMSE_G	ER_A	ER_G	RB_A	RB_G	RMSE_A	RMSE_G
Gennaio	3,99	4,14	1,23	1,25	1,42	1,43	1,10	1,12	1,11	1,11	1,39	1,39
Febbraio	4,13	4,03	0,05	0,05	0,14	0,14	2,30	2,31	0,08	0,08	0,18	0,18
Marzo	0,81	0,82	2,12	2,12	2,43	2,43	1,65	1,64	0,08	0,08	0,14	0,14
Aprile	1,97	1,97	0,03	0,02	0,10	0,10	0,40	0,42	1,13	1,13	1,44	1,44
Maggio	2,43	2,43	1,06	1,06	1,21	1,21	3,02	3,00	0,03	0,03	0,19	0,19
Giugno	3,11	3,11	0,08	0,09	0,33	0,33	2,17	2,16	0,25	0,25	0,39	0,39
Luglio	2,01	2,03	1,23	1,23	1,38	1,39	2,42	2,43	0,06	0,06	0,18	0,18
Agosto	1,86	1,86	0,14	0,13	0,29	0,29	0,48	0,48	1,16	1,16	1,57	1,57
Settembre	1,96	1,98	1,99	1,98	2,31	2,31	3,09	3,09	0,04	0,04	0,17	0,17
Ottobre	8,68	8,74	0,10	0,10	0,34	0,34	2,54	2,57	0,08	0,08	0,23	0,23
Novembre	4,48	4,47	0,10	0,09	0,28	0,28	2,20	2,27	0,11	0,11	0,20	0,20
Dicembre	3,64	3,60	0,19	0,19	0,29	0,28	1,76	1,76	0,07	0,06	0,15	0,15

Figura 4 - *Mangimi*: prezzi osservati nel mese di Ottobre 2008 vs prezzi osservati nel mese precedente (scala logaritmica)

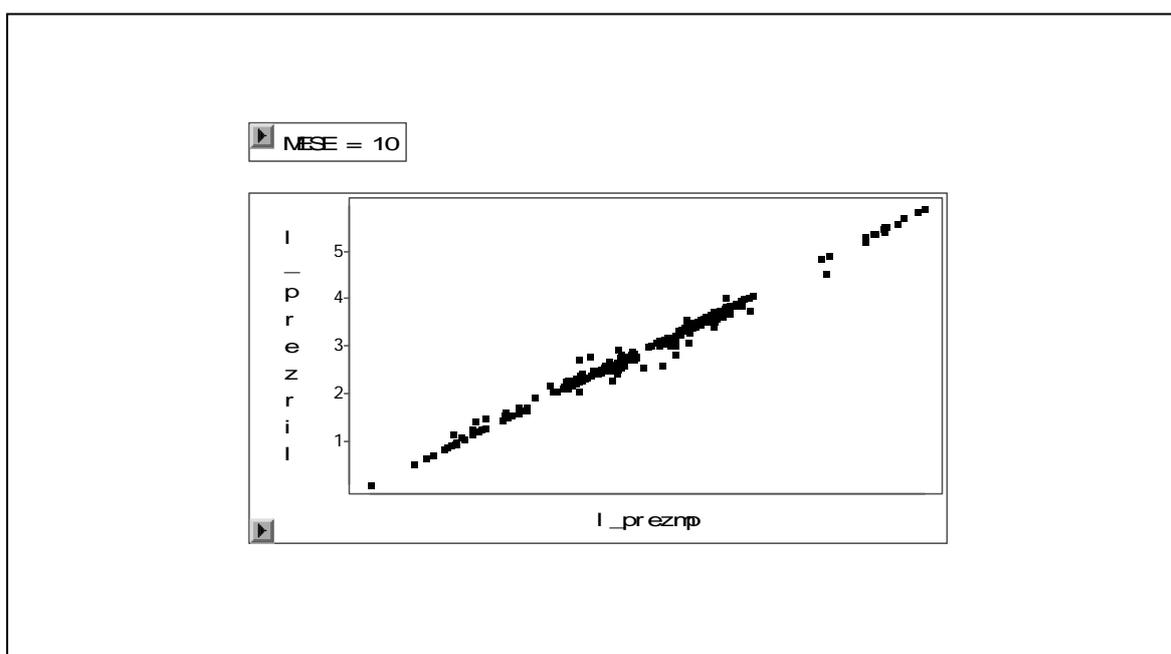
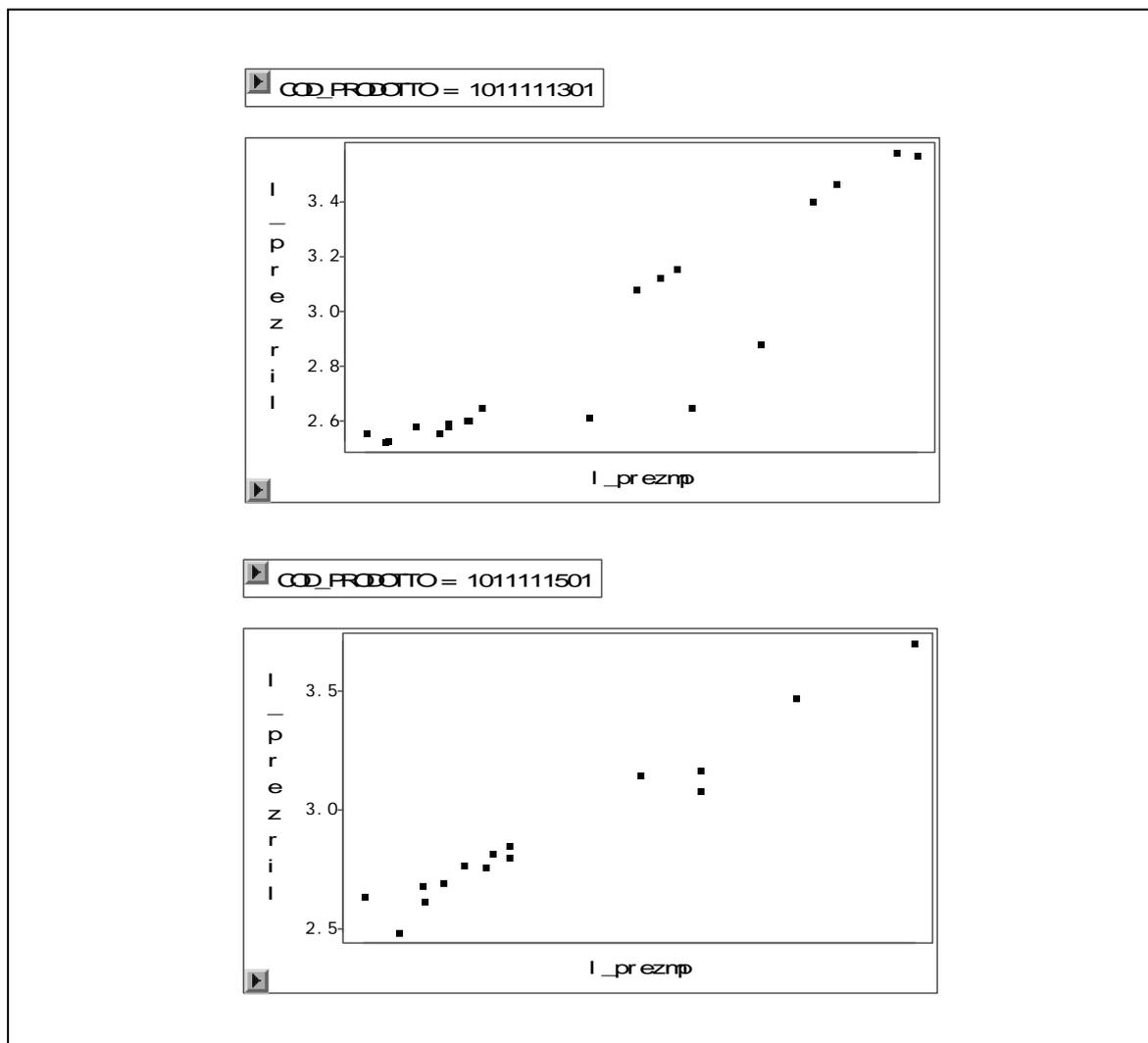


Figura 5 - *Mangimi*: prezzi osservati nel mese di Ottobre 2008 vs prezzi osservati nel mese precedente per due prodotti del gruppo (scala logaritmica)



Ricordando ancora una volta che la procedura di imputazione corrente calcola separatamente le quotazioni di prezzo relative ai singoli prodotti, gli effetti dell'imputazione per la voce *Mangimi* possono essere considerati un limite superiore (*rischio massimo*) associato a questa fase del trattamento dei dati di questo gruppo di prodotti. L'effetto più consistente a livello di indice sintetico che si osserva in alcuni mesi del 2008 (*Marzo* e *Settembre*) è in genere imputabile alla presenza di alcune quotazioni anomale di cui andrebbe approfondita la natura.

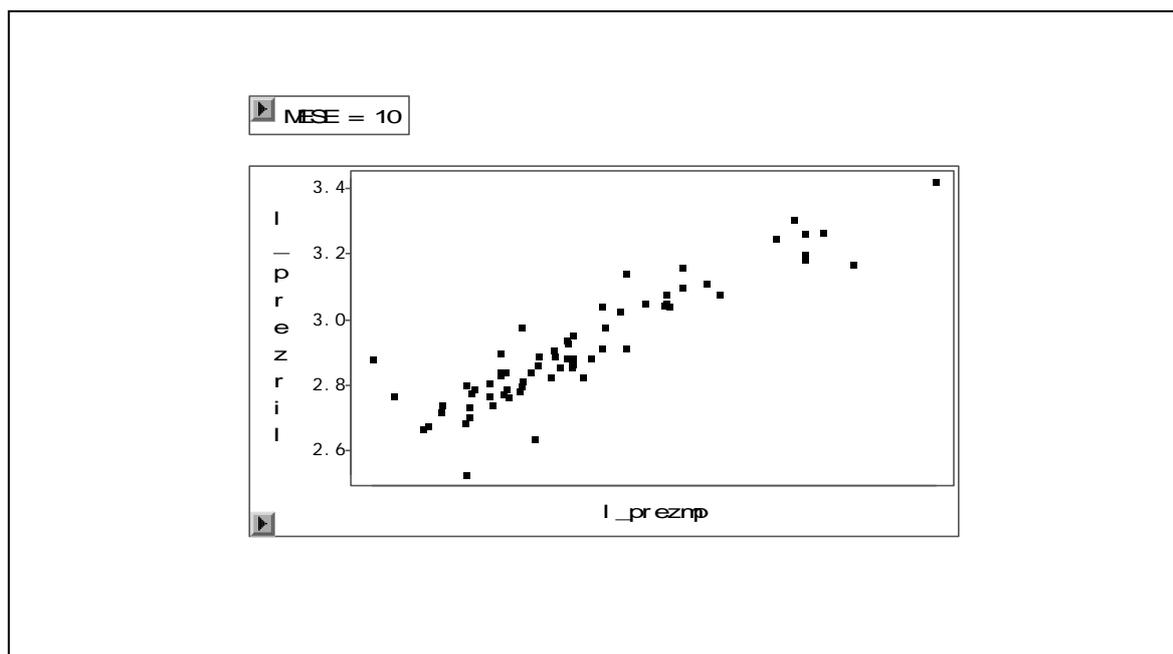
Per quanto riguarda la voce *Frumento* (Tavola 6), a livello di indice sintetico non si registra un significativo effetto distorsivo della previsione dei prezzi mancanti per nessuno dei due anni considerati. Tuttavia, si osservano valori di *RE* piuttosto elevati in alcuni mesi del 2008 e nella maggior parte dei mesi del 2009. In questi casi, la frequenza di prezzi mancanti può essere considerata una delle cause, da associare però ad altri fattori. Uno di questi può verosimilmente essere collegato da un lato alla variazione dei prezzi mese precedente/mese corrente del *Frumento*, dall'altro al rapporto fra medie degli indici elementari.

Tavola 6 - Frumento: valori degli indicatori di qualità per anno e mese

MANGIMI	2008						2009					
	ER_A	ER_G	RB_A	RB_G	RMSE_A	RMSE_G	ER_A	ER_G	RB_A	RB_G	RMSE_A	RMSE_G
Gennaio	1,30	1,32	-0,02	-0,02	0,14	0,14	3,23	3,22	0,08	0,07	0,22	0,22
Febbraio	1,26	1,25	0,01	0,01	0,08	0,08	3,78	3,80	0,02	0,02	0,26	0,26
Marzo	0,76	0,77	0,01	0,02	0,05	0,05	3,38	3,40	-0,05	-0,05	0,24	0,24
Aprile	2,06	2,10	0,00	-0,01	0,17	0,17	2,15	2,11	-0,01	-0,01	0,11	0,11
Maggio	2,21	2,23	-0,04	-0,05	0,16	0,17	2,86	2,84	-0,04	-0,04	0,19	0,19
Giugno	2,76	2,75	-0,07	-0,08	0,27	0,27	3,02	3,07	0,00	0,01	0,33	0,34
Luglio	5,84	5,83	0,00	-0,02	0,33	0,33	4,13	4,14	0,05	0,04	0,30	0,30
Agosto	1,45	1,42	0,69	0,69	1,33	1,33	4,89	4,91	-0,08	-0,08	0,45	0,45
Settembre	4,73	4,74	-0,03	-0,03	0,17	0,17	2,23	2,20	0,06	0,06	0,13	0,13
Ottobre	6,16	6,16	-0,03	-0,02	0,22	0,22	1,19	1,18	-0,01	-0,01	0,08	0,08
Novembre	2,42	2,41	-0,02	-0,02	0,17	0,17	1,42	1,42	0,05	0,05	0,09	0,09
Dicembre	2,44	2,51	0,22	0,23	0,79	0,79	0,70	0,69	-0,04	-0,04	0,06	0,06

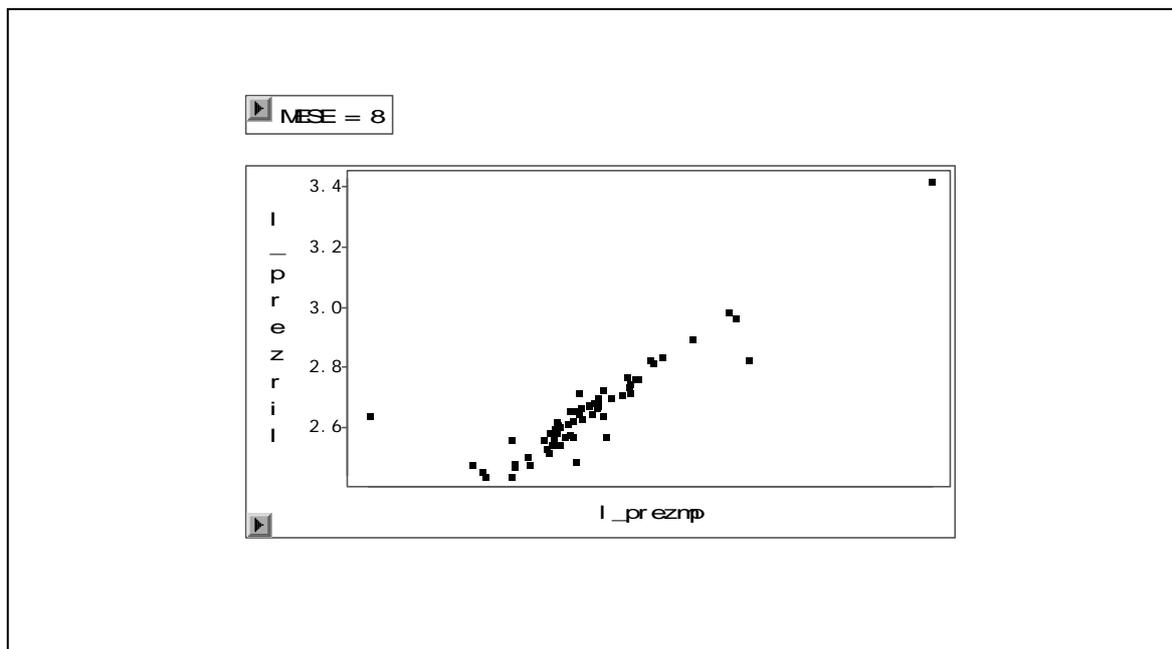
Se consideriamo ad esempio il mese di *Ottobre* 2008, nella Figura 6 osserviamo una significativa variazione dei prezzi dal mese precedente al mese corrente, con una differenza media delle quotazioni mese/mese pari al 9,8%. Tale distanza ovviamente si ripercuote sulla qualità delle stime dei prezzi elementari (tenuto conto che il rapporto fra le medie degli indici elementari è per questo mese e per questa voce pari a 0,92).

Figura 6 - Frumento: prezzi osservati nel mese di Ottobre 2008 vs prezzi osservati nel mese precedente (scala logaritmica)



Anche se consideriamo il mese di *Agosto* 2009 (Figura 7) osserviamo una significativa variabilità dei prezzi mese precedente/mese corrente (con una differenza che in media è del 4,5%), che si ripercuote sulla qualità delle previsioni. Inoltre, alcuni valori della distribuzione lasciano margine a ulteriori analisi sulla variabilità e/o correttezza delle quotazioni osservate.

Figura 7 - Frumento: prezzi osservati nel mese di Agosto 2009 vs prezzi osservati nel mese precedente (scala logaritmica)



5. Conclusioni e prospettive future

Nel lavoro sono descritte le caratteristiche del nuovo sistema di produzione degli indici dei prezzi dei prodotti agricoli venduti e acquistati. Il sistema, che comprende una nuova acquisizione dei prezzi, una nuova procedura di trattamento dei dati e di calcolo degli indici dei prezzi, presenta elementi innovativi da un punto di vista sia tecnologico (uso integrato di ambienti e linguaggi di sviluppo, anche di tipo *open-source*), sia metodologico (metodi per l'imputazione automatica dei valori mancanti).

Oltre a garantire minori tempi e costi di produzione rispetto alla situazione precedente, la nuova procedura fornisce un contributo al miglioramento del livello di accuratezza dei dati e quindi dei risultati prodotti.

Nel lavoro sono inoltre illustrati i risultati di uno studio sperimentale avente lo scopo di verificare gli eventuali effetti distorsivi, a livello di dati elementari e di indici di sintesi, del processo di ricostruzione delle quotazioni mancanti, valutando al contempo la performance comparata di due versioni alternative del metodo di imputazione proposto. I risultati della sperimentazione hanno evidenziato come più che gli aspetti metodologici (scelta della media - aritmetica o geometrica - per la sintesi dei microindici nel metodo di imputazione), la distorsione a livello di dati elementari e a livello di indici di sintesi (almeno per il livello nazionale) può essere tenuta sotto controllo garantendo che la base dati utilizzata nel processo di imputazione sia sufficientemente completa (cioè con percentuali ridotte di valori mancanti) e accurata (cioè con frequenze contenute di errori di misura). Per questo motivo, l'ottimizzazione della fase di controllo dei dati potrebbe essere supportata efficacemente dalla predisposizione di indicatori di qualità del materiale raccolto (ad esempio, frequenza dei prezzi mancanti, anche ai diversi livelli territoriali, frequenza dei prezzi fuori *range* e loro localizzazione geografica, ecc.). Questo consentirebbe sia di re-distribuire le risorse necessarie per la revisione interattiva di situazioni particolarmente critiche, sia di monitorare in modo più puntuale il comportamento delle Camere di Commercio, che naturalmente rivestono un ruolo fondamentale nell'economia della rilevazione. Più in generale, da un punto di vista più operativo/tecnologico, sembrerebbe opportuna una ulteriore revisione del flusso di dati e delle elaborazioni sul *DB*, per l'eliminazione di alcune ridondanze ancora presenti nel processo (ad esempio, iterazioni non necessarie del calcolo degli indici, revisioni dei dati non

guidate da indicatori di qualità dei dati) al fine di una ulteriore ottimizzazione dei tempi di produzione degli indicatori e della qualità della rilevazione.

Da un punto di vista più strettamente metodologico, si prevede di procedere ad ulteriori approfondimenti su diversi aspetti ancora non sufficientemente esplorati per l'ulteriore miglioramento della nuova procedura di ricostruzione delle quotazioni mancanti.

Innanzitutto, è necessario approfondire l'analisi longitudinale delle variazioni degli indici di prodotto per la definizione di valori soglia differenziati per l'ottimizzazione della fase di individuazione dei casi anomali e/o di altri errori non campionari.

Inoltre, è necessario proseguire nell'analisi delle caratteristiche e della natura dei meccanismi di mancata risposta che agiscono nello specifico contesto d'indagine, al fine di verificarne l'eventuale dipendenza da aspetti strutturali (ad esempio, il livello territoriale) o di altra natura. Questo aspetto assume una particolare rilevanza nel caso di prodotti orto-frutticoli, in cui i tassi di mancata risposta risultano particolarmente elevati.

Vanno comunque approfondite le analisi di alcuni beni particolarmente problematici (ad esempio in termini di stagionalità come i prodotti ortofrutticoli ma anche altre categorie di prodotti) al fine di migliorare il metodo di imputazione, ad esempio utilizzando periodi di riferimento (per il calcolo della variazione temporale nell'ambito del metodo di imputazione) diversi da quelli attualmente adottati.

Al fine di avere un quadro completo degli effetti potenziali del processo di imputazione sugli aggregati oggetto di calcolo nell'ambito della rilevazione, è necessario inoltre procedere alla valutazione sperimentale della distorsione degli indici di sintesi a livelli territoriali diversi da quello nazionale.

Infine, allo scopo di depurare le valutazioni sperimentali dalla presenza di valori precedentemente imputati (quindi non corrispondenti a situazioni effettivamente osservate), sarebbe opportuno effettuare ulteriori test di valutazione dei potenziali effetti distorsivi sugli indici di prodotto e di sintesi utilizzando le variazioni di prezzo per le sole quotazioni effettivamente fornite dalle Camere di Commercio, escludendo cioè quelle che nel periodo di riferimento ($t-v$) risultassero a loro volta imputate.

Riferimenti Bibliografici

- Eurostat. 2008. Handbook for EU Agricultural Price Statistics. Version 2.0, March 2008.
- Eurostat. 2007. *Recommended Practices for Editing and Imputation in Cross-Sectional Business Surveys*.
http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/RPM_EDIMBUS.pdf.
- ISTAT. 2010. Indici dei prezzi dei prodotti agricoli IV Trimestre 2010,
http://www.istat.it/salastampa/comunicati/non_calendario/20110401_01/.
- Little R.J.A. and D.B. Rubin. 1987. *Statistical Analysis with Missing Data*. John Wiley & Sons, Inc, New York.

Allegato 1 - Stagionalità dei prodotti appartenenti alla categoria Ortofrutta

		Gen.	Feb.	Mar.	Apr.	Mag.	Giu.	Lug.	Ago.	Set.	Ott.	Nov.	Dic.
0011152101	PATATE PRIMATICCE	0	0	0	1	1	1	0	0	0	0	0	0
0011211101	CIPOLLE	1	1	1	1	0	0	1	1	1	1	1	1
0011211201	POMODORI in piena aria	0	0	0	0	0	1	1	1	1	1	0	0
0011211202	POMODORI in serra	1	1	1	1	1	0	0	0	0	0	1	1
0011211203	PISELLI freschi	0	0	0	0	1	1	0	0	0	0	0	0
0011211204	FAGIUOLI freschi	0	0	0	0	0	0	1	1	1	0	0	0
0011211205	FAGIOLINI	0	0	0	0	0	1	1	1	1	0	0	0
0011211206	MELONI	0	0	0	0	0	0	1	1	0	0	0	0
0011211207	COCOMERI	0	0	0	0	0	0	1	1	0	0	0	0
0011211301	CAVOLFIORI	1	1	1	0	0	0	0	0	0	1	1	1
0011211302	CAVOLI VERZE	1	1	1	1	0	0	0	0	1	1	1	1
0011211304	INDIVIE	1	1	1	1	0	0	0	0	1	1	1	1
0011211306	CARCIOFI	1	1	1	1	1	0	0	0	0	0	0	0
0011211307	ASPARAGI	0	0	0	0	1	1	0	0	0	0	0	0
0011211308	MELANZANE in piena aria	0	0	0	0	0	0	1	1	1	1	0	0
0011211309	MELANZANE in serra	1	1	1	1	1	1	0	0	0	0	1	1
0011211310	ZUCCHINE in piena aria	0	0	0	0	0	1	1	1	1	1	0	0
0011211311	ZUCCHINE in serra	1	1	1	1	1	0	0	0	0	0	1	1
0011211312	SEDANI COSTE	1	1	1	1	1	0	0	0	1	1	1	1
0011211314	PEPERONI in piena aria	0	0	0	0	0	0	0	1	1	0	0	0
0011211315	PEPERONI in serra	1	1	1	1	1	1	1	0	0	1	1	1
0011211316	SPINACI	1	1	1	1	1	1	0	0	1	1	1	1
0011211317	FINOCCHI	1	1	1	1	1	0	0	0	0	1	1	1
0011222101	PIANTE di CICLAMINO	1	1	1	1	0	0	0	0	1	1	1	1
0011222102	PIANTE di IMPATIENS	0	0	1	1	1	1	1	0	0	0	0	0
0011222103	PIANTE di PELARGONIA (o gerani)	0	0	1	1	1	1	0	0	0	0	0	0
0011222104	PIANTE di PETUNIA	0	0	0	1	1	1	0	0	0	0	0	0
0011222105	PIANTE di POINSETTIA	0	0	0	0	0	0	0	0	0	0	0	1
0011222106	PIANTE di PRIMULA	1	1	0	0	0	0	0	0	0	1	1	1
0011222108	PIANTE DI BEGONIE	0	1	1	1	1	1	1	1	1	0	0	0
0011222205	FRESIE	1	1	1	1	1	0	0	0	0	0	0	0
0011222207	TULIPANI	1	1	1	1	0	0	0	0	0	0	0	1
0011311100	UVA DA TAVOLA	0	0	0	0	0	0	0	1	1	1	0	0
0011311201	UVA da vino comune	0	0	0	0	0	0	0	0	1	1	0	0
0011311202	UVA da vino DOC e DOCG	0	0	0	0	0	0	0	0	1	1	0	0
0011311203	UVA da vino ad indicaz. Geog.o tipico	0	0	0	0	0	0	0	0	1	1	0	0
0011322401	OLIVE da mensa	1	0	0	0	0	0	0	0	0	1	1	1
0011322402	OLIVE per oleificazione	1	0	0	0	0	0	0	0	0	1	1	1
0011332201	ARANCE	1	1	1	0	0	0	0	0	0	0	1	1
0011332202	MANDARINI	1	1	1	0	0	0	0	0	0	0	0	1
0011332203	CLEMENTINE	1	0	0	0	0	0	0	0	0	0	1	1
0011342302	PERE	0	0	0	0	0	0	1	1	1	1	1	1
0011342303	ALBICOCCHIE	0	0	0	0	0	1	1	0	0	0	0	0
0011342304	CILIEGIE	0	0	0	0	0	1	0	0	0	0	0	0
0011342305	PESCHE	0	0	0	0	0	1	1	1	1	0	0	0
0011342306	NETTARINE	0	0	0	0	0	1	1	1	1	0	0	0
0011342307	SUSINE	0	0	0	0	0	1	1	1	1	0	0	0
0011342308	FRAGOLE in piena aria	0	0	0	0	1	1	0	0	0	0	0	0
0011342309	FRAGOLE in serra	1	1	1	1	0	0	1	1	1	1	1	1
0011342403	MANDORLE	1	1	1	1	1	0	0	0	1	1	1	1
0011342404	NOCCIOLE	1	1	1	1	0	0	0	0	1	1	1	1
0011342405	NOCI	1	1	0	0	0	0	0	0	1	1	1	1

Informazioni per gli autori

La collana è aperta ad autori dell'Istat e del Sistema statistico nazionale, e ad altri studiosi che abbiano partecipato ad attività promosse dal Sistan (convegni, seminari, gruppi di lavoro, ecc.). Da gennaio 2011 essa sostituirà Documenti Istat e Contributi Istat.

Coloro che desiderano pubblicare sulla nuova collana dovranno sottoporre il proprio contributo alla redazione degli Istat Working Papers inviandolo per posta elettronica all'indirizzo iwp@istat.it. Il saggio deve essere redatto seguendo gli standard editoriali previsti, corredato di un sommario in italiano e in inglese; deve, altresì, essere accompagnato da una dichiarazione di paternità dell'opera. Per la stesura del testo occorre seguire le indicazioni presenti nel foglio di stile, con le citazioni e i riferimenti bibliografici redatti secondo il protocollo internazionale 'Autore-Data' del *Chicago Manual of Style*.

Per gli autori Istat, la sottomissione dei lavori deve essere accompagnata da una mail del proprio dirigente di Servizio/Struttura, che ne assicura la presa visione. Per gli autori degli altri enti del Sistan la trasmissione avviene attraverso il responsabile dell'ufficio di statistica, che ne prende visione. Per tutti gli altri autori, esterni all'Istat e al Sistan, non è necessaria alcuna presa visione. Tutti i lavori saranno sottoposti al Comitato di redazione, che valuterà la significatività del lavoro per il progresso dell'attività statistica istituzionale. La pubblicazione sarà disponibile su formato digitale e sarà consultabile on line.

Gli articoli pubblicati impegnano esclusivamente gli autori, le opinioni espresse non implicano alcuna responsabilità da parte dell'Istat. Si autorizza la riproduzione a fini non commerciali e con citazione della fonte.