



Istituto
nazionale
di statistica

NOTA INFORMATIVA

I nuovi indici del commercio con l'estero (base 2005=100)

L'Istituto Nazionale di Statistica presenta oggi le nuove serie mensili degli indici dei valori medi unitari e dei volumi dei flussi commerciali con l'estero, espressi con anno di riferimento 2005=100.

Si tratta di indici prodotti mediante la tecnica del concatenamento di indici a base mobile, adeguati a cogliere tempestivamente i cambiamenti quantitativi e qualitativi dei prodotti scambiati con l'estero, oltre che a tenere conto delle modificazioni annuali nelle classificazioni delle merci.

Contestualmente vengono fornite informazioni relative alle innovazioni introdotte nelle metodologie e nelle tecniche di calcolo, anche a seguito delle modifiche dei Regolamenti europei relativi agli scambi intracomunitari (Intrastat). Tali innovazioni, oltre ad adeguare gli indicatori al mutato quadro normativo europeo, hanno consentito di migliorare la qualità dell'informazione prodotta.

Le principali novità

Le nuove serie dei numeri indici del commercio con l'estero si distinguono dalle precedenti per i seguenti aspetti:

- la nuova base di riferimento degli indici concatenati diventa l'anno 2005; tutte le serie dal 1996 ad oggi vengono, quindi, pubblicate con riferimento a tale anno;
- i valori medi unitari delle merci vengono calcolati tenendo conto delle unità di misura delle quantità scambiate stabilite dai nuovi Regolamenti comunitari;
- il numero di indici elementari utilizzato per il calcolo degli indici aggregati mensili sale a 220.000 (erano 165.000), a seguito del passaggio ad una più dettagliata stratificazione dei dati di base, ora definita a livello di singolo prodotto della classificazione Nomenclatura Combinata;
- viene utilizzato un nuovo metodo di trattamento dei dati anomali e degli errori di misura per migliorare la qualità delle stime finali;
- vengono adottate nuove soluzioni metodologiche per il calcolo degli indici aggregati previsti dal piano di diffusione dei dati;
- viene introdotto l'aggregato relativo all'UEM a 15 paesi e ricostruita la serie storica progressiva ad esso riferita a partire dal 1996.

I nuovi dati modificano il quadro dell'evoluzione degli indici dei valori medi unitari definito dai vecchi indici. In particolare, il confronto tra le variazioni percentuali annue delle due serie mette in evidenza una più contenuta dinamica dei nuovi indici dei valori medi unitari sia all'importazione sia all'esportazione, con differenze significative verificate soprattutto negli ultimi anni.

Inoltre, a partire dalla metà del 2005 la ragione di scambio calcolata sui nuovi indici si colloca persistentemente al di sotto di quella calcolata sui vecchi indicatori, con una convergenza tra le due serie che si manifesta nella seconda metà del 2007.

Ufficio della comunicazione
Tel. + 39 06 4673.2243-2244

Centro di informazione statistica
Tel. + 39 06 4673.3102

Informazioni e chiarimenti
Statistiche del commercio
con l'estero
Via Balbo 16 - 00184 Roma
Paola Anitori
Tel. + 39 06 4673.2567
Ersilia Di Pietro
Tel. + 39 06 4673.2549

I dati relativi ai nuovi indici del commercio estero sono presenti alla pagina web <http://www.coeweb.istat.it/indici.htm>

Gli indici dei valori medi unitari del commercio estero

I numeri indici dei valori medi unitari del commercio estero misurano le variazioni nel tempo del valore unitario delle merci scambiate sui mercati internazionali. Le loro variazioni incorporano sia le dinamiche dei prezzi dei singoli prodotti scambiati, sia i cambiamenti di composizione merceologica degli acquisti dall'estero, sia, infine, gli effetti dei cambiamenti di qualità dei prodotti.

L'Istat diffonde mensilmente gli indici relativi agli scambi commerciali dell'Italia con gli altri paesi secondo una disaggregazione per gruppi di prodotto e per alcune rilevanti aree geografiche/geoeconomiche¹.

Si tratta di indici mensili che scaturiscono da una sintesi di indici di tipo Laspeyres e indici di tipo Paasche, dando luogo ad indici di tipo Fisher, in cui la base è costituita dall'anno immediatamente precedente a quello di osservazione. Tali indici, detti a base mobile, riferiti a ciascun aggregato previsto dal piano di diffusione dei dati, vengono successivamente concatenati all'anno di riferimento (assimilabile ad un anno "base") attraverso coefficienti di raccordo costituiti, per il medesimo aggregato, dal prodotto delle medie degli indici degli anni precedenti. Ciascun indice a base mobile è una media ponderata di indici elementari, calcolati secondo una stratificazione per tipo di flusso, paese di origine/destinazione e gruppo merceologico.

I pesi utilizzati per le aggregazioni di ordine superiore degli indici elementari sono costituiti dal valore delle merci scambiate relative a ciascuno strato; in particolare, il sistema di ponderazione sarà costituito nel caso dell'indice di Laspeyres dal valore registrato per quello strato nell'anno base e nel caso dell'indice di Paasche dal valore degli scambi nel mese di riferimento.

Nel calcolo degli indici elementari vengono considerate solo le transazioni riferite ai regimi definitivi (esportazioni ed importazioni dirette), escludendo le operazioni temporanee nonché le re-importazioni e le ri-esportazioni: si ritiene, infatti, che la gran parte di questo tipo di transazioni non si basi su effettivi "prezzi" di mercato e che i relativi valori medi unitari riferiti possano incorporare ulteriori elementi spuri.

Le merci appartenenti a raggruppamenti particolari quali i Mezzi di trasporto, i Materiali radioattivi e similari, i Veicoli aerospaziali ecc.², per i quali il concetto di valore unitario ha scarso significato economico, vengono invece "sterilizzate" imputando loro il valore medio del gruppo merceologico di appartenenza calcolato escludendo i prodotti in questione.

I numeri indici dei volumi sono calcolati, per ciascun aggregato previsto dal piano di diffusione, a posteriori come rapporto tra l'indice concatenato del valore e l'indice concatenato del valore medio unitario. Gli indici di valore, infine, sono calcolati considerando il totale degli scambi con l'estero, comprensivo, quindi, sia dei regimi esclusi dal computo dei valori medi unitari sia del valore delle merci sottoposte a sterilizzazione. In tal modo l'eventuale perdita di informazione imputabile all'esclusione di tali merci nei valori medi unitari viene incorporata nei dati in volume.

¹ L'articolazione degli indici diffusi mensilmente è la seguente:

	Area geografica o geoeconomica
• ATECO2002 a 2-digit	Totale Mondo, UE, Paesi Terzi
• ATECO2002 a 3-digit	
• RPI	
• Totale merci	Totale Mondo, UE, Paesi Terzi, Germania, Francia, Regno Unito, Spagna, UEM, Russia, OPEC, Usa, EDA, Cina, Giappone

Per approfondimenti si rimanda a <http://www.coeweb.istat.it>

² Si veda Istat (2003) "I nuovi indici del commercio con l'estero". Nota Rapida del 16 luglio 2003.

Il cambiamento della metodologia utilizzata per il calcolo degli indici e le caratteristiche del nuovo impianto di calcolo

Rispetto ai vecchi indici del commercio estero, le nuove serie, pur mantenendone l'impianto di base, descritto in precedenza, rappresentano da un lato un adeguamento a nuovi regolamenti internazionali, dall'altro un avanzamento sul piano delle soluzioni metodologiche adottate.

La metodologia di calcolo fin qui utilizzata prevedeva la costruzione di indici secondo una stratificazione per flusso, paese di origine/destinazione e gruppo merceologico individuato secondo la classificazione "Sistema Armonizzato" (SA) a 6 cifre. Il valore medio unitario veniva costruito come rapporto tra il valore e le quantità, espresse in chilogrammi, riferiti ai mesi dell'anno precedente in cui la merce si è movimentata.

In questo quadro, a partire dal 2005 il sistema Intrastat³ ha eliminato l'obbligo degli operatori commerciali di dichiarare le quantità scambiate, riferite ad un consistente numero di merci indicate nell'Allegato del Regolamento relativo alla Nomenclatura Tariffaria e Statistica ed alla Tariffa Doganale Comune (Nomenclatura Combinata)⁴ in chilogrammi, e di esprimerle secondo l'unità di misura ivi prevista.

Questa disposizione, che interessa oltre il 30% delle merci totali, ha reso necessario il passaggio ad un sistema di calcolo basato sulle nuove unità di misura e su una stratificazione delle variabili per flusso, paese e gruppo di prodotto secondo la classificazione merceologica internazionale "Nomenclatura Combinata" (NC), che ha un livello di dettaglio massimo di otto cifre e fino alla sesta posizione coincide con la classificazione "Sistema Armonizzato" (SA), precedentemente utilizzata.

In tal modo, ciascun indice elementare esprime la variazione del valore medio unitario di ogni merce rispetto all'anno base in funzione dell'unità di misura prevista dal Regolamento Intrastat per quel prodotto, assicurando la coerenza degli indici a base mobile aggregati ed eliminando la necessità di implementare un sistema di imputazione dei dati mancanti, che nel vecchio sistema si era reso necessario come effetto dell'adeguamento delle dichiarazioni degli operatori al nuovo regolamento.

Allo scopo di assicurare l'omogeneità al sistema di calcolo, anche gli indici riferiti ai paesi extracomunitari sono stati calcolati considerando le stesse unità di misura indicate dal Regolamento Intrastat e la medesima stratificazione. L'adeguamento alla nuova metodologia nel caso della rilevazione sugli scambi con i Paesi terzi è stato possibile in quanto il Regolamento riferito a tale rilevazione⁵ ha previsto l'obbligatorietà della trasmissione del dato espresso in specifiche unità di misura per tutti i prodotti fin dalla sua entrata in vigore.

La nuova stratificazione ha implicato, pur con riferimento alla stessa articolazione geografica di calcolo, il passaggio dai circa 165 mila indici elementari mensili del vecchio sistema agli oltre 220 mila del nuovo.

Contestualmente alle innovazioni introdotte in termini di unità di misura e classificazione merceologica, sono state modificate anche le procedure di elaborazione dei dati relativi alle singole transazioni, ed in particolare al controllo dei cosiddetti "errori di misura", cioè quegli errori generatisi nelle fasi di raccolta dei dati (per maggiori dettagli si vedano le Note metodologiche presentate alla fine del documento).

Ulteriori innovazioni hanno riguardato anche la metodologia di calcolo degli indici aggregati, modificata allo scopo di limitare la volatilità di breve periodo della dinamica dei valori medi unitari. Uno dei metodi per ottenere tale obiettivo è quello di calcolare gli indici aggregati utilizzando medie troncate (*trimmed mean*): queste scaturiscono dall'eliminazione dal computo di ciascun indice aggregato di una certa percentuale di osservazioni esterne ad un intervallo definito secondo criteri statistici, e commisurato alle caratteristiche strutturali della specifica aggregazione di prodotti. L'obiettivo è quello di migliorare la robustezza dei segnali informativi forniti dagli indici. In pratica, il metodo selezionato viene applicato alle distribuzioni degli indici elementari, "pesate" con i valori dell'interscambio dello specifico prodotto.

L'impatto complessivo dei cambiamenti introdotti (adeguamento al Regolamento Intrastat per quanto riguarda le unità di misura e le classificazioni delle merci, adozione di nuovi metodi per il trattamento dei dati anomali, nuova metodologia di calcolo degli indici aggregati) ha determinato cambiamenti rispetto a

³ Per approfondimenti si vedano Commission Regulation n° 1915/2005 (OJ L 307), Regulation n°493/2005 (OJ L 82) e successive modificazioni.

⁴ In precedenza, infatti, la dichiarazione della cosiddetta unità supplementare era facoltativa mentre l'informazione in kg. era obbligatoria.

⁵ Si vedano Commission Regulation n° 1172/95 (OJ L 118), n° 1917/2000 (OJ L 229/14) e successive modificazioni.

quanto misurato dalle serie degli indici dei valori medi unitari e dei volumi finora pubblicate. In particolare, le nuove serie mostrano una dinamica meno accentuata delle precedenti, pur mantenendo tassi di crescita molto significativi.

La dinamica dei nuovi indici dei valori medi unitari e dei volumi

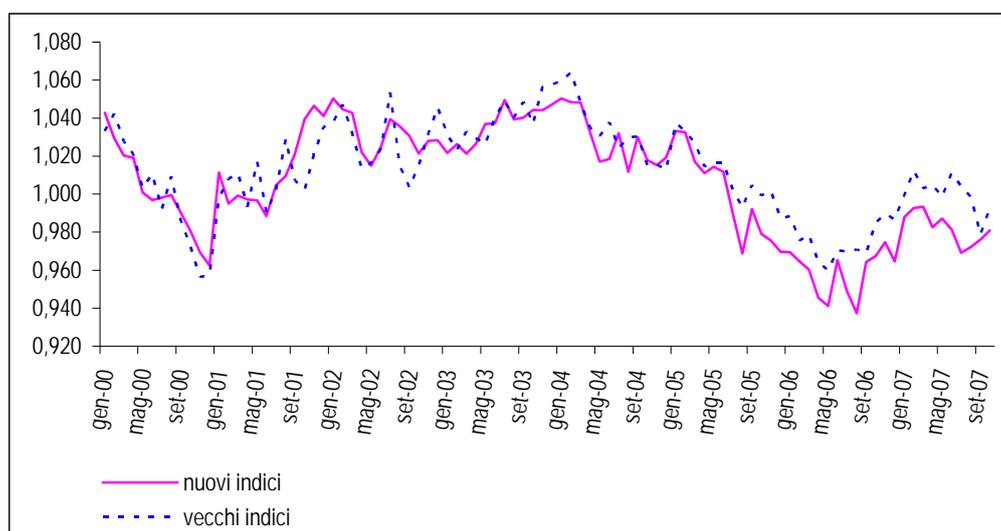
Il confronto tra le variazioni percentuali annue dei nuovi e dei vecchi indici, riportati alla stessa base (2000=100) e riferiti all'interscambio del totale dei prodotti verso il mondo, mette in evidenza una più lenta dinamica dei nuovi indici dei valori medi unitari sia all'importazione sia all'esportazione (Tabella 1). In particolare, per le esportazioni la divergenza tra la dinamica dei vecchi e nuovi indici è stata, nel 2004 e nel 2005, pari ogni anno a circa 1,5 punti percentuali. La differenza massima tra le due serie si rileva comunque nel corso del 2007 (periodo gennaio-ottobre). Di conseguenza, sia per le importazioni sia per le esportazioni i nuovi indici di volume risultano più dinamici dei precedenti.

Tabella 1 - Indici dei valori medi unitari e dei volumi. Totale mondo (base 2000=100). Anni 2000-2007 (tassi di variazione annui)

ANNI	VALORI MEDI UNITARI				VOLUMI			
	Nuovi indici		Vecchi indici		Nuovi indici		Vecchi indici	
	Importazioni	Esportazioni	Importazioni	Esportazioni	Importazioni	Esportazioni	Importazioni	Esportazioni
2001	0.8	2.0	2.7	3.6	1.4	3.1	-0.5	1.5
2002	-1.0	0.9	-0.5	1.4	0.0	-2.3	-0.6	-2.8
2003	-0.7	-0.3	-0.3	0.8	1.0	-1.7	0.7	-2.6
2004	3.3	2.5	4.8	4.2	5.1	4.7	3.6	3.0
2005	7.9	4.8	8.7	6.3	0.7	0.6	-0.1	-0.9
2006	9.5	5.1	10.5	6.6	4.0	5.5	2.3	2.7
gen-ott 2007	2.0	4.8	5.7	8.6	3.0	4.7	0.8	2.7

Fonte: ISTAT- Statistiche del commercio con l'estero

Figura 1 - Ragioni di scambio (*) degli indici dei valori medi unitari. Totale mondo (base 2000=100). Anni 2000-2007 (livelli)



(*) Rapporto tra indici dei valori medi unitari delle esportazioni e delle importazioni

L'andamento delle ragioni di scambio (rapporto tra indici dei valori medi unitari delle esportazioni e delle importazioni), calcolate sulla base dei vecchi e dei nuovi indici (Figura 1), fornisce una rappresentazione sintetica dell'effetto complessivo delle innovazioni introdotte.

In particolare, a partire dalla metà del 2005 la ragione di scambio calcolata sui nuovi indici si colloca persistentemente al di sotto di quella calcolata sui vecchi indicatori, con una convergenza tra le due serie

che si manifesta solo nella seconda metà del 2007.

Il confronto tra la dinamica dei vecchi e dei nuovi indici, riferito alle due principali aree geografiche di origine e destinazione delle merci (Ue a 27 paesi ed Extra-Ue), consente ulteriori valutazioni (Tabella 2).

I nuovi indici dei valori medi unitari delle esportazioni mostrano una dinamica più lenta di quelli precedenti soprattutto per quanto riguarda i flussi verso i paesi non comunitari; nel corso del 2007, tuttavia, anche la differenza riscontrata per l'area Ue appare rilevante.

Tabella 2 - Indici dei valori medi unitari e dei volumi per area⁶ (base 2000=100). Anni 2000-2007 (tassi di variazione annui)

ANNI	UNIONE EUROPEA (a)				PAESI TERZI				
	Nuovi indici		Vecchi indici		Nuovi indici		Vecchi indici		
	Importazioni	Esportazioni	Importazioni	Esportazioni	Importazioni	Esportazioni	Importazioni	Esportazioni	
Valori medi unitari									
2001	1.7	1.5	4.7	3.1	-0.5	2.8	-0.8	4.5	
2002	0.9	1.3	0.7	1.7	-3.7	0.3	-2.4	1.2	
2003	0.6	0.8	0.8	1.2	-2.6	-1.9	-2.1	-0.3	
2004	2.0	3.0	4.0	4.5	5.1	1.8	6.1	3.9	
2005	3.0	4.1	4.9	5.1	14.9	5.9	14.9	8.3	
2006	5.4	4.6	6.7	4.9	14.7	6.0	15.7	9.4	
gen-ott 2007	3.5	4.9	7.5	8.2	0.4	4.6	3.5	9.3	
Volumi									
2001	2.0	2.9	-1.0	1.5	0.1	3.4	0.4	1.7	
2002	-0.5	-3.0	-0.2	-3.4	0.4	-1.4	-1.0	-2.3	
2003	0.2	-0.6	0.1	-0.9	2.1	-3.4	1.5	-5.0	
2004	4.3	3.3	2.3	1.8	6.6	6.8	5.6	4.8	
2005	0.8	0.1	-1.1	-0.9	1.2	1.3	1.2	-1.2	
2006	4.5	6.0	1.1	2.9	3.9	4.7	3.3	1.9	
gen-ott 2007	1.2	2.5	-0.2	2.5	5.3	8.0	1.9	2.9	

Fonte: ISTAT - Statistiche del commercio con l'estero

(a) a 27 paesi

La Tabella 3 riporta il dettaglio settoriale dei nuovi indici dei valori medi unitari.

Dal lato delle esportazioni, i dati relativi agli ultimi anni confermano una tendenza alla crescita dei valori medi unitari generalizzata alla gran parte dei settori. Per il complesso dei manufatti, il tasso di crescita risulta pari al 4,7% nel 2005, al 5% nel 2006 ed al 4,8% nei primi dieci mesi del 2007. Al di là di settori, come quello petrolifero e quello dei metalli e prodotti in metallo, caratterizzati da un marcato ciclo inflazionistico su scala internazionale, dinamiche elevate si riscontrano soprattutto per il settore del cuoio e dei prodotti in cuoio ed in quello degli articoli in gomma e materie plastiche. Dal lato opposto, andamenti relativamente contenuti si registrano per la carta e l'editoria, per il settore alimentare, per quello tessile e dell'abbigliamento.

Sul fronte delle importazioni, negli anni più recenti dinamiche intense si riscontrano, oltre che per il comparto energetico e per quello dei metalli, per il cuoio e prodotti in cuoio, legno e prodotti in legno, gomma e materie plastiche, prodotti agricoli. In diminuzione risultano invece gli indici dei valori medi unitari delle macchine e apparecchi elettrici.

Tabella 3 - Indici dei valori medi unitari per settore da e verso il mondo. Anni 2000-2007 (tassi di variazione annui)

⁶ Date le caratteristiche degli indici concatenati (Allen R.G.N, 1975 "Index numbers theory and practice" Mc Millan Press, London), gli indici riferiti all'interscambio da e verso il mondo (cfr. tav.1) potrebbero violare la condizione di internalità e risultare esterni all'intervallo costituito dagli indici delle due sub-aree.

SETTORI	Importazioni							Esportazioni						
	2001	2002	2003	2004	2005	2006	Gen-ott 2007	2001	2002	2003	2004	2005	2006	Gen-ott 2007
Prodotti dell'agricoltura, della caccia, della silvicoltura e della pesca	-2.0	-1.1	-0.6	1.3	2.2	4.3	4.8	6.8	4.3	3.3	-1.4	3.4	5.9	3.3
Minerali energetici e non energetici (a)	-0.8	-8.3	0.8	10.1	35.6	24.9	-6.2
Prodotti trasformati e manufatti	1.2	0.2	-0.9	2.2	3.5	6.3	3.9	1.9	0.8	-0.3	2.6	4.7	5.0	4.8
Prodotti alimentari, bevande e tabacco	3.6	-1.8	-0.9	1.5	1.9	4.0	2.6	2.5	1.8	1.7	0.4	-0.2	2.3	1.9
Prodotti delle industrie tessili e dell'abbigliamento	4.4	1.1	-0.7	0.1	1.5	5.3	3.1	4.7	3.3	0.7	1.9	4.1	3.9	3.2
Cuoio e prodotti in cuoio, pelle e similari	9.4	3.9	-5.8	-0.2	2.6	6.6	4.7	9.4	8.2	2.5	1.4	4.8	6.6	8.1
Legno e prodotti in legno	-1.9	-0.3	-1.5	-1.6	1.2	4.2	4.8	-2.7	0.2	-1.1	2.7	1.4	3.9	3.8
Pasta da carta, carta e prodotti di carta; prodotti dell'editoria e della stampa	-4.3	-6.8	-4.6	-2.7	0.6	4.5	2.1	1.1	-3.5	-2.8	-2.1	0.0	0.8	1.3
Coke, prodotti petroliferi raffinati e combustibili nucleari	-9.5	-1.4	5.3	8.3	32.0	21.6	-0.7	-6.6	-3.4	4.8	16.0	37.0	18.2	-1.8
Prodotti chimici e fibre sintetiche e artificiali	0.5	-3.4	-0.3	4.7	7.1	5.7	3.7	-0.1	-2.9	-2.7	1.1	5.1	5.6	2.8
Articoli in gomma e materie plastiche	1.6	-2.3	-0.5	1.1	2.8	4.6	5.1	0.9	-0.9	-1.2	0.5	4.1	5.2	4.2
Prodotti della lavorazione di minerali non metalliferi	4.0	-0.2	-0.6	0.7	2.6	3.2	3.9	3.2	0.6	0.7	1.9	3.6	4.7	3.4
Metalli e prodotti in metallo	-2.1	-0.8	-0.4	15.0	9.4	17.2	12.8	-0.4	-1.8	-0.2	9.7	6.7	9.3	12.7
Macchine ed apparecchi meccanici	3.2	2.6	-1.4	-0.6	2.2	4.4	1.9	1.8	0.6	-0.5	1.7	3.3	2.6	4.8
Macchine elettriche ed apparecchiature elettriche, elettroniche ed ottiche	0.9	1.3	-5.3	-3.6	-2.6	-1.0	-3.4	1.1	0.3	-3.1	0.1	2.0	3.8	3.7
Mezzi di trasporto	3.7	3.1	3.2	0.2	0.1	3.8	3.0	2.6	1.8	0.8	2.8	2.0	4.1	4.7
Altri prodotti delle industrie manifatturiere	1.6	1.1	-2.7	0.5	3.4	6.0	2.5	2.3	2.7	0.6	1.9	4.5	4.0	3.9
Energia elettrica, gas e acqua (a)	2.2	-3.8	-7.6	13.0	8.6	31.9	9.3
Totale	0.8	-1.0	-0.7	3.3	7.9	9.5	2.0	2.0	0.9	-0.3	2.5	4.8	5.1	4.8

Fonte: ISTAT - Statistiche del commercio con l'estero

(a) Per la forte erraticità delle serie storiche e l'esiguo valore degli aggregati, le relative variazioni degli indici non sono diffuse

Note metodologiche

La nuova procedura di trattamento dei dati di base

Da esperienze empiriche consolidate risulta verificato⁷ che l'informazione di fonte doganale sulle quantità è di solito qualitativamente inferiore al dato in valore. C'è inoltre da considerare che le classificazioni ufficiali del commercio internazionale sono riferite, anche al più fine livello di dettaglio, a "merci" che rappresentano nella maggior parte dei casi veri e propri gruppi di beni i quali, pur essendo molto omogenei tra loro possono risultare, di fatto, prodotti significativamente diversi. Da un punto di vista dinamico ciò implica che, tra due periodi di osservazione, la composizione interna di una "merce" può modificarsi se alcuni dei prodotti che la rappresentano o non si movimentano affatto, o vengono movimentati ad intervalli irregolari e non prevedibili o, più spesso, vengono sostituiti con beni del tutto simili ma di contenuto qualitativo diverso.

In questo quadro, la metodologia adottata per il trattamento degli errori di misura, in linea con le raccomandazioni internazionali, risulta pienamente coerente anche con gli obiettivi di controllo e di contenimento della variabilità interna di ciascuno strato.

Il controllo statistico dell'errore di misura è qui ottenuto mediante l'implementazione di un nuovo metodo di trattamento dei dati anomali che, operando sui livelli dei valori medi unitari delle singole transazioni registrate per ciascuno strato (incrocio tra tipo di flusso, paese di origine/destinazione e singolo prodotto identificato dalla NC a otto cifre), esegue un controllo delle distribuzioni utilizzando metodi statistici non parametrici per distribuzioni univariate e asimmetriche. Si tratta di metodi robusti che discriminano le informazioni anomale sulla base dei parametri di posizione, adattandosi alla forma della distribuzione e minimizzando la perdita complessiva dell'informazione. In particolare, attraverso tali metodi viene definito un intervallo di tolleranza al di fuori del quale le transazioni vengono segnalate e, nel caso specifico, escluse dal calcolo dell'indice elementare del proprio strato.

L'algoritmo utilizzato è stato calibrato sulle distribuzioni non pesate dei valori medi unitari, sulla base dell'ipotesi che l'errore di misura è indipendente dal peso relativo dell'osservazione nello strato⁸. Tuttavia, le esclusioni non costituiscono una perdita completa di informazione; infatti, nel calcolo degli indici dei valori e dei volumi, il peso di ciascuno strato includerà anche le transazioni segnalate ed escluse dal computo dei valori medi unitari.

Il numero di transazioni escluse dal calcolo è risultato compreso tra il 2% e il 18% a seconda dello strato. Le distribuzioni sottoposte a controllo sono state sia quelle relative ai valori medi unitari mensili che determinano il numeratore dell'indice, sia quelle dei valori medi unitari dell'anno base, che ne rappresenta il denominatore.

L'innovazione metodologica relativa al trattamento dei dati anomali ad un così elevato livello di dettaglio è stata possibile anche grazie all'evoluzione dei sistemi informatici utilizzati, che consentono di controllare un numero totale di transazioni pari a circa 33 milioni ogni anno.

Tra i diversi metodi non parametrici il più adatto alle distribuzioni osservate è risultato l'*Asymmetric Fences Method (AFM)*, "integrato" e calibrato *ad hoc* per tenere conto delle specificità della variabile di studio. Data la forte asimmetria riscontrata nelle distribuzioni mensili dei livelli dei valori medi unitari, i dati riferiti alle singole transazioni registrate in ciascuno strato determinato dall'incrocio "prodotto NC-paese di origine/destinazione-flusso" sono stati sottoposti preventivamente a trasformazione logaritmica⁹ in modo da assicurare maggiore robustezza dei risultati.

Nella sua formulazione generale, dato un *set* di osservazioni $\{x_1, \dots, x_n\}$, il metodo *AFM* segnala il dato x_i come anomalo se è verificata una delle seguenti condizioni:

⁷ UNSD (1981) "Strategy for price and quantity measurement in external trade - A technical report", *Statistical papers n°69, series M, NY*. Eurostat (2006) "Statistics on the trading of goods - User guide", *Methods and nomenclatures. Theme External trade, Luxembourg*.

⁸ Controlli a campione hanno consentito di verificare la presenza di errori di misura anche nel caso di transazioni di alto valore riferite ad un medesimo operatore.

⁹ Tra tutte le trasformazioni ottenibili applicando l'algoritmo di Box-Cox (Box, G.E.P, Cox D.R. 1964) allo scopo di rendere più simmetriche le distribuzioni originali, la trasformazione più appropriata è risultata la quella logaritmica.

$$q_1 - x_i > k_{AFM} * \max(q_2 - q_1, c * |q_2|) \quad (A1)$$

oppure

$$x_i - q_3 > k_{AFM} * \max(q_3 - q_2, c * |q_2|) \quad (A2)$$

dove q_1 , q_2 e q_3 sono rispettivamente il primo quartile, la mediana e il terzo quartile della distribuzione di x mentre k_{AFM} e c sono due costanti da determinare¹⁰. In particolare, la prima costante definisce l'ampiezza dell'intervallo all'esterno del quale cadono i valori segnalati come anomali; la seconda evita che nei casi in cui la distanza tra il quartile e la mediana sia troppo piccola (ad esempio se i valori unitari delle singole transazioni hanno una variabilità eccezionalmente bassa per cui i quartili e la mediana sono molto vicini tra loro), il metodo identifichi un numero eccessivo di *outlier*. Nella pratica, i valori delle costanti k_{AFM} e c sono stabiliti sulla base di sperimentazioni empiriche che, nel caso in questione e dopo diverse verifiche effettuate su sottocampioni dei dati originali, hanno portato a fissare $k_{AFM} = 1$ e $c = 0.05$.

L'*AFM* ha un *break-down point*, ossia la percentuale massima di *outlier* che possono essere trattati senza che lo stimatore ne risulti distorto¹¹, pari al 25% e, poiché l'intervallo di tolleranza è più ampio dal lato in cui la distanza del quartile dalla mediana è maggiore, il metodo rispetta la naturale asimmetria della distribuzione dei dati. Tuttavia, esso risulta inefficace quando la numerosità delle osservazioni è esigua¹². Nel caso dei valori medi unitari diverse prove empiriche hanno suggerito che la soglia minima più adeguata a rispettare le capacità discriminatorie dell'*AFM* è quella che stabilisce almeno 100 osservazioni per cella.

Ne consegue che in tutti i casi in cui il numero n di transazioni per cella è inferiore a 100 è stato necessario individuare, sempre nell'ambito della famiglia dei metodi non parametrici, un metodo alternativo a garanzia della robustezza dei risultati, soprattutto per valori di n molto piccoli. Poiché, come noto, in presenza di un numero limitato di osservazioni il parametro che meglio sintetizza le caratteristiche di una distribuzione è la mediana, il metodo prescelto per integrare l'*AFM* utilizza come riferimento la mediana degli scarti di ciascuna transazione x_i dalla mediana della distribuzione dello strato, ossia il MAD (*Median Absolute Deviation*),

$$MAD_i = mediana_{i \in nc8}(|x_i - q_2|) \quad (A3)$$

e stabilisce che la generica transazione x_i sia segnalata come anomala se le seguenti condizioni sono soddisfatte simultaneamente:

$$\begin{cases} |x_i - q_2| > k_{MAD} * MAD \\ \frac{|\exp(x_i) - \exp(q_2)|}{|\exp(q_2)|} > A \end{cases} \quad k_{MAD} = 2 \quad e \quad A = 0.1 \quad (A4)$$

dove le costanti A e k_{MAD} sono state stabilite a seguito di diverse prove su opportuni sottoinsiemi dei dati iniziali. Il metodo basato sul *MAD* è in generale più robusto dell'*AFM* in quanto ha un *breakdown point* di circa il 50%; pur tuttavia, esso ha mostrato un potere discriminatorio non eccessivo se la distribuzione è unimodale o se nelle celle vi sono poche decine di osservazioni.

Le transazioni che risultano aver fallito il test *AFM* o quello basato sul *MAD* vengono segnalate come valori anomali ed escluse dal calcolo dell'indice elementare mensile di strato che sarà, quindi, definito a partire dai soli dati non segnalati in valore $val_{i,m}^{*,t}$ e in quantità $qua_{i,m}^{*,t}$, con riferimento sia al mese

¹⁰ ILO (2004 "Consumer price index manual: theory and practice". Chapter 9.

<http://www.ilo.org/public/english/bureau/stat/guides/cpi/>

¹¹ Geyer C.J. (2006) "Breakdown point theory notes", Class notes (nonparametrics).

www.stat.umn.edu/geyer/5601/notes/break.pdf

¹² Nel caso limite in cui $n = 5$, la presenza di un solo *outlier* avrebbe effetto sulla posizione dei quartili rendendolo, dunque, inefficace.

osservato m dell'anno t (numeratore dell'indice) sia all'anno base $t-1$. Anche il *set* di dati mensili riferiti all'anno base, infatti, è sottoposto a controllo attraverso il metodo *AFM* "integrato" con il *MAD*.

In sostanza, l'indice elementare di strato ${}_{t-1}I_m^{nc,t}$ (omettendo per semplicità la notazione riferita al paese p e al flusso f) sarà dato da:

$${}_{t-1}I_m^{nc,t} = \frac{vmu_m^{nc,t}}{vmu_{t-1}^{nc,t}} = \frac{\sum_{i \in nc} val_{i,m}^{*t}}{\sum_{i \in nc} qua_{i,m}^{*t}} / \frac{\sum_{i \in nc} val_{i,m}^{*t-1}}{\sum_{i \in nc} qua_{i,m}^{*t-1}} \quad (A5)$$

Tutte le distribuzioni sottoposte a test sono non pesate; per tanto, l'ipotesi alla base dell'esclusione dei valori anomali dal computo della relazione (A5) è che l'errore di misura incida sulla variabile a prescindere dal peso relativo (in valore) della transazione nello strato.

Innovazioni nel calcolo degli indici aggregati: l'utilizzo di medie troncate

La media troncata di una distribuzione di osservazioni è il valore medio ottenuto eliminando dal calcolo un numero di elementi appartenenti alle code della distribuzione stessa. Il ricorso a tale tipo di stimatore viene giustificato dalla necessità di fornire una stima del fenomeno preservata per quanto possibile da fattori di instabilità di breve periodo.

Dato il piano di diffusione degli indici dei valori medi unitari, l'intervallo di troncamento per il calcolo degli indici aggregati è stato calibrato sulle distribuzioni degli indici elementari per gruppo di prodotti secondo la classificazione CPA a due cifre, area di origine/destinazione (Ue, Paesi Terzi) e flusso. In aggiunta, per garantire un'omogeneità di trattamento alle distribuzioni mensili degli indici elementari delle aggregazioni scelte, i punti di troncamento sono stati stimati annualmente e in ciascun arco temporale annuale sono stati tenuti fissi per ogni mese. Gli indici aggregati riferiti al totale dei prodotti dell'interscambio da e verso il mondo, infine, sono stati calcolati per aggregazione diretta degli indici elementari ammessi al calcolo degli aggregati riferiti alle due sub-aree, Ue e Paesi terzi.

Dal punto di vista statistico, la media troncata è uno stimatore robusto¹³ e, anche in presenza di asimmetria delle distribuzioni, può rappresentare uno strumento dall'elevata capacità informativa.

In linea generale, la media troncata dipende dal criterio scelto per l'identificazione dei punti di troncamento (*cutting points*) che definiscono l'ampiezza dell'intervallo rispetto al quale l'osservazione che risulta esterna viene esclusa dal calcolo. In tal senso, la stima può mostrare una certa sensibilità a diversi elementi, quali il livello di disaggregazione dei dati originari, il grado di asimmetria delle distribuzioni, la lunghezza del periodo rispetto al quale viene calibrata la stima ecc.

Per quanto riguarda la determinazione di intervalli "ottimali" è possibile utilizzare informazioni ausiliarie (ad esempio, una serie storica relativa a dati altamente correlati con la serie da stimare; una serie storica pregressa che possa rappresentare un "*benchmark*" ecc.) ma, più spesso, i criteri vengono definiti in base ai risultati ottenuti in termini di distanza della distribuzione finale da quella originaria o, ancora, sulla base della minimizzazione della variabilità residua, o l'asimmetria residua, delle distribuzioni finali.

Come criterio generale, le distribuzioni utilizzate per calibrare i punti di troncamento sono state le

distribuzioni degli indici elementari $\frac{vmu_m^{nc,g}}{vmu_{t-1}^{nc,g}}$ del generico strato g -individuato dal gruppo di prodotti

secondo la classificazione CPA a due cifre, l'area geoeconomica (Ue e Paesi terzi) e il tipo di flusso, "pesate" con i valori mensili dell'interscambio. In tal senso, si è scelto di non considerare le variazioni

mensili degli indici $\frac{vmu_m^{nc,g}}{vmu_{m-1,t}^{nc,g}}$ in quanto da prove empiriche effettuate su sottoinsiemi specifici è risultato

¹³ Si tratta di una robustezza "relativa" in quanto, utilizzando un numero maggiore di informazioni, risulta sempre meno robusta della mediana.

che il metodo avrebbe determinato una perdita eccessiva di informazione, risultando troppo severo.

In funzione del grado di asimmetria della distribuzione sono stati messi a confronto diversi metodi per l'individuazione dei punti di troncamento: a fronte del criterio più semplice basato su un intervallo di troncamento fisso per ogni anno della serie e scelto indipendentemente dalla forma della distribuzione, sono stati comparati il criterio di *Tukey* e l'*AFM*, entrambi nelle formulazioni relative a distribuzioni con e senza trasformazione logaritmica. In questo contesto, il metodo di *Tukey* e l'*AFM* sono stati utilizzati nella loro formulazione standard.

Nello specifico, l'algoritmo di *Tukey* è un metodo simmetrico basato sui quartili q_1 e q_3 della distribuzione di un insieme di osservazioni $\{x_1, \dots, x_n\}$ e definisce il seguente intervallo di accettazione:

$$[q_1 - k(q_3 - q_1); q_3 + k(q_3 - q_1)] \quad (A6)$$

dove il parametro k è pari a 1,5 e definisce i cosiddetti “*inner fences*” mentre per $k=3$ definisce gli “*outer fences*”¹⁴. Nel caso venga applicata la trasformazione logaritmica ai dati originari, l'intervallo di accettazione diviene il seguente:

$$[\log(q_1) - k(\log(q_3) - \log(q_1)); \log(q_3) + k(\log(q_3) - \log(q_1))] \quad (A7)$$

Dato il *set* di indici elementari dello strato, dunque, la media troncata è calcolata come media aritmetica ponderata degli indici elementari dello strato che sono inclusi nell'intervallo (A6) o (A7). Poiché l'algoritmo di *Tukey* individua un filtro simmetrico, il metodo è appropriato per distribuzioni simmetriche o lievemente asimmetriche.

L'algoritmo *AFM*¹⁵, nella formulazione standard, è invece un filtro asimmetrico che individua i seguenti intervalli, rispettivamente senza trasformazione logaritmica

$$[q_1 - k(q_2 - q_1); q_3 + k(q_3 - q_2)] \quad (A8)$$

e con trasformazione logaritmica

$$[\log(q_1) - k(\log(q_2) - \log(q_1)); \log(q_3) + k(\log(q_3) - \log(q_2))] \quad (A9)$$

Nel caso di distribuzioni simmetriche, in corrispondenza di $k=3$ si ritrovano gli *inner fences* del metodo di *Tukey* mentre per $k=6$ si individuano le corrispondenti *outer fences*. Tale algoritmo è più appropriato per distribuzioni mediamente o fortemente asimmetriche.

Per garantire una certa omogeneità di trattamento alle distribuzioni mensili sottoposte a filtro, tuttavia, è stato necessario assicurare che il calcolo dei quartili non risentisse dell'effetto di eventi di breve periodo capaci di alterare in modo anomalo l'ampiezza dell'intervallo. A tale riguardo, i parametri di posizione di ciascuno strato sono stati prima calcolati annualmente e poi sintetizzati in una media aritmetica semplice sui tre anni immediatamente precedenti a quello di osservazione t . Tale media “mobile” dei quartili delle distribuzioni annuali è stata quindi utilizzata per la definizione degli intervalli di troncamento. Ciò ha consentito di attenuare il possibile disturbo sui livelli dei quartili generato da eventi subitanei pur consentendo, allo stesso tempo, di incorporare parte dell'effettiva dinamica degli indici elementari negli indici aggregati.

Una volta calcolati i parametri di posizione sono stati applicati gli algoritmi ritenuti più adatti alle distribuzioni sottostanti in funzione del loro grado di asimmetria, in modo da individuare i *cutting points*; in funzione di questi ultimi sono state calcolate sia le medie troncate di strato sia, con riferimento alle distribuzioni *ex post* (prive cioè delle osservazioni esterne alle *fences*), alcuni indicatori di variabilità residua, di asimmetria e curtosi residue, nonché il peso dell'informazione esclusa dal computo dell'indice aggregato.

¹⁴ Tukey J.W. (1977) “Exploratory data analysis”, Addison-Wesley, NY.

¹⁵ Qui l'*AFM* è espresso in una forma leggermente modificata per agevolare la comparazione con il metodo di *Tukey*. Nella formulazione originale l'intervallo è dato da $[q_2 - c(q_2 - q_1); q_2 + c(q_3 - q_2)]$ dove $c = k+1$.

Il metodo di troncamento più adeguato è risultato quello in grado di minimizzare i valori degli indicatori nelle distribuzioni residue. In estrema sintesi, per distribuzioni simmetriche o lievemente asimmetriche, nella maggior parte dei casi sono risultati più adatti l'algoritmo di *Tukey* o l'*AFM* senza trasformazione logaritmica; per le distribuzioni fortemente asimmetriche è risultato più adatto l'*AFM* con trasformazione logaritmica, mentre per le distribuzioni mediamente asimmetriche, nei casi più problematici è stato necessario modificare leggermente l'ampiezza degli intervalli (attraverso il parametro k) e valutare anche il peso dell'informazione esclusa dal calcolo della media troncata.