



Nota metodologica
Data Warehouse Ulisse
(siulisse.it)

Ottobre 2015

Nota metodologica Data Warehouse Ulisse

Copyright © 2015 StudiaBo Srl

Ottobre 2015

StudiaBo Srl

via Santo Stefano, 57

40125 Bologna

Italy

Quest'opera è soggetta alla Creative Commons Public License Attribuzione - Non commerciale - Non opere derivate 2.5 Generico (CC BY-NC-ND 2.5) o posteriore. L'enunciato integrale della Licenza in versione 2.5 è reperibile all'indirizzo internet <http://creativecommons.org/licenses/by-nc-nd/2.5/deed.it>.

- Si è liberi di riprodurre, distribuire, comunicare al pubblico, esporre, in pubblico, rappresentare, eseguire e recitare quest'opera alle seguenti condizioni:

Attribuzione Bisogna attribuire la paternità dell'opera nei modi indicati dall'autore o da colui al quale è stata data quest'opera in licenza; in questo caso si tratta di StudiaBo srl.

Non commerciale Non si può usare quest'opera per fini commerciali.

Non opere derivate Non si può alterare o trasformare quest'opera, né usarla per crearne un'altra.

- Ogni volta che si usa o si distribuisce quest'opera, lo si deve fare secondo i termini di questa licenza, che va comunicata con chiarezza.
- In ogni caso si possono concordare con il titolare dei diritti d'autore (StudiaBo srl) usi di quest'opera in deroga da questa licenza.

I nomi commerciali, i loghi, i trademark appartengono ai rispettivi proprietari.

Indice

Presentazione	3
Le fonti dei dati	5
UN Comtrade	5
HS1992	5
HS2007	5
HS2012	6
Altre fonti	6
Eurostat - db Comext	6
UN - db Monthly Comtrade	6
US Census Bureau - db UsaTrade	7
ITC - db Trade Map	7
Congiuntura	8
Normalizzazione	9
Outliers	10
Dati Mancanti	11
Fasce di qualità/prezzo	12
Quantità a prezzi costanti	13

Presentazione

Questo documento descrive le scelte metodologiche che sono state fatte per costruire il Data Warehouse **Ulisse**, riguardante i flussi di commercio estero. I dati di esportazioni e di importazioni dichiarati dai diversi paesi mondiali rappresentano un database ricco di contenuto informativo. Utilizzando questi dati è possibile avere informazioni sulla dinamica dei diversi mercati ad un livello molto elevato di dettaglio. E' possibile inoltre sapere a quale prezzo viene venduto un dato bene su uno specifico mercato e, soprattutto, se quel mercato accetta di pagare o meno un *premium price* per la maggior qualità offerta da un fornitore rispetto agli altri. Infine, consente di valutare la competitività dei diversi paesi, evidenziandone i punti di forza e di debolezza. Spesso, tuttavia, queste informazioni non sono immediatamente fruibili da una semplice lettura dei dati. Le dichiarazioni dei diversi paesi contengono a volte dati che non corrispondono ai flussi effettivi di commercio con l'estero. I motivi principali sono i seguenti:

Procedura di stima. Non sempre i dati sono raccolti nel momento in cui le merci attraversano le varie frontiere doganali. E' il caso, ad esempio, del commercio estero tra i paesi dell'Unione Europea i cui valori sono stimati dai diversi istituti di statistica sulla base delle dichiarazioni Iva effettuate dalle varie imprese. Generalmente, tuttavia, queste procedure di stima sono di elevata affidabilità.

Classificazione merceologica. Non tutti i paesi usano le medesime classificazioni merceologiche e, quindi, flussi relativi a prodotti simili possono essere attribuiti a codici merceologici diversi. Negli ultimi anni, tuttavia,

c'è stato un forte processo di armonizzazione delle diverse classificazioni, grazie soprattutto al lavoro svolto dal World Customs Organization (www.wcoomd.org) che ha sviluppato la classificazione *Harmonized System*, utilizzata ormai da quasi 150 paesi.

Confidenzialità. A volte i flussi di commercio estero per paese e prodotto possono essere segreti perché ritenuti confidenziali. Ogni paese segue le sue regole, con una maggiore incidenza dei flussi segreti per i paesi più piccoli. Tra i paesi della Unione Europea alcuni (v. Danimarca, Finlandia e Austria) presentano un'incidenza dei flussi segreti dal lato delle esportazioni che può superare, nei diversi anni, il 5% del loro export totale.

Ritardi. Alcuni paesi in via di sviluppo, con una struttura amministrativa meno consolidata, tendono a produrre le statistiche relative ai loro flussi di commercio estero con ritardo, in alcuni casi anche di qualche anno. L'Onu è impegnato nel supportare questi paesi finanziando progetti che hanno l'obiettivo di migliorare le procedure di raccolta, elaborazione e pubblicazione dei dati di commercio estero.

Per poter estrarre dai dati di base informazioni significative ed affidabili è quindi necessario utilizzare una serie di strumenti metodologici finalizzati a distinguere la *misura* del fenomeno dal *rumore statistico*. Appare pertanto fondamentale costruire procedure d'analisi che, sulla base di quanto appena descritto, utilizzino al meglio i numerosi vantaggi insiti nei dati sul commercio internazionale, ovvero:

Elevata Numerosità. La maggior parte dei paesi pubblica mensilmente le statistiche di commercio estero per codice prodotto e per paese partner. Ne consegue che, ogni mese, ciascun paese pubblica dati relativi

vi a vari milioni di flussi commerciali: l'elevata numerosità di dati consente, quindi, di individuare facilmente eventuali dati anomali.

Dati riferiti alla popolazione. Le statistiche annuali di commercio con l'estero riguardano la quasi totalità dei flussi commerciali. Questo consente di evitare tutte le problematiche riguardanti i campioni di rilevazione e gli errori di misura che da questi possono derivare.

Doppia Dichiarazione. Ciascun flusso di commercio estero è dichiarato due volte: dal paese esportatore e dal paese importatore. La doppia dichiarazione, se opportunamente utilizzata, consente di aumentare in modo significativo l'affidabilità delle misure relative ai vari fenomeni.

Queste caratteristiche dei dati consentono di utilizzare con efficacia alcune metodologie di *data mining*, in grado di isolare in modo soddisfacente la *misura* dei diversi fenomeni dal suo *termine di errore*.

In questo documento sono descritte le diverse metodologie utilizzate in fase di costruzione della banca dati Ulisse. Tali metodologie riguardano:

Outlier. Sono metodologie finalizzate all'individuazione di dati inficiati da *errori di misura* significativi e alla loro sostituzione con valori più coerenti con le altre osservazioni;

Dati Mancanti. Sono metodologie che individuano i flussi di commercio per i quali non è presente nella banca dati la relativa misura. Queste metodologie consentono inoltre di stimare un valore per il dato mancante, coerente con le altre informazioni presenti nella banca dati.

Normalizzazione. Sotto questa voce sono raccolte quelle elaborazioni che consentono di

confrontare tra loro misure diverse dello stesso fenomeno (flusso rilevato da paese di origine e paese di destinazione), producendo per ogni variabile una misurazione univoca.

Congiuntura. Riguardano strumenti in grado di produrre una pre-stima di un dato annuale, utilizzando al meglio le informazioni congiunturali disponibili.

Fasce di qualità/prezzo. E' una elaborazione che consente di valutare se il bene in esame può essere oggetto di un'offerta differenziata. In caso di bene differenziabile, le procedure sviluppate da StudiaBo attribuiscono una specifica fascia di qualità/prezzo per ciascun flusso, prendendo in esame il confronto tra il suo prezzo e quello degli altri flussi del medesimo prodotto.

Quantità a prezzi costanti. Le rilevazioni sui flussi di commercio estero riguardano valori e unità di misura fisica. Costruendo il rapporto tra le due variabili, otteniamo quelli che in letteratura vengono chiamati valori medi unitari, spesso identificati con il prezzo. Tali valori medi unitari incorporano le variazioni qualitative a cui il prodotto viene commercializzato. Al fine di qualificare nelle analisi tali modificazioni, StudiaBo ha costruito la variabile *quantità a prezzi costanti*, in modo tale che sia quest'ultima ad includere l'effetto qualità. Gli indicatori di prezzo che ne derivano, dati dal rapporto tra valori e quantità a prezzi costanti, possono quindi essere considerati più attendibili rispetto ai valori medi unitari perchè non inficiati da variazioni qualitative del prodotto.

Le metodologie utilizzate hanno consentito di produrre una banca dati altamente in grado di *raccontare* le caratteristiche del fenomeno *Commercio Estero* nei suoi minimi dettagli.

Le fonti dei dati

UN Comtrade

La banca dati Ulisse StudiaBo è stata costruita a partire dalle informazioni rese disponibili da diverse fonti di analisi economica, prima fra tutte la Divisione Statistica delle Nazioni Unite. Tale organismo ha costruito e mantiene aggiornato il database Comtrade, accessibile sia via cd-rom che attraverso il sito <http://comtrade.un.org/db/>. Comtrade riunisce le dichiarazioni annuali di più di 170 paesi verso i propri partner commerciali, dettagliate a livello di codice prodotto in classificazione merceologica HS a 6 cifre. La Divisione Statistica delle Nazioni Unite stima così di riuscire ad esplicitare più del 95% del commercio mondiale. Per ogni flusso di commercio internazionale (inteso come dichiarazione di un paese verso il proprio partner commerciale in un dato anno), Comtrade riporta il valore in dollari e la quantità commercializzata, espressa in kg e/o in un'unità di misura supplementare, che varia da prodotto a prodotto. UN Comtrade, inoltre, propone dal 2005 agli utenti un proprio progetto volto ad estendere la copertura dell'informazione in quantità, che potrà essere oggetto di comparazione grazie al processo in atto di omologazione verso i kg. Tale intervento di analisi statistica è identificabile dalla variabile "Flag", esplicitata con il fine di distinguere i dati stimati da ciò che effettivamente è oggetto delle dichiarazioni per paese. Le classificazioni doganali con cui Comtrade rende disponibile l'informazione sono essenzialmente due:

- Harmonised System (HS): classificazione sviluppata dalla World Customs Organization (WCO) e sottoposta a revisione ogni 5 anni (l'ultima volta nel 2012), che, come già detto precedentemente, propone un livello di

disaggregazione massima a 6 cifre. Dato l'alto livello di dettaglio nelle dichiarazioni per prodotto, tale classificazione è stata oggetto della procedura StudiaBo per la costruzione della banca dati Ulisse;

- Standard International Trade Classification (SITC): classificazione sviluppata a partire dal 1951 e periodicamente sottoposta a revisione, che ordina le merceologie secondo una struttura gerarchica arrivando ad un livello di disaggregazione a 5 digits.

HS1992

La versione datata 1992 della classificazione merceologica HS consente agli utenti di disporre di una banca dati caratterizzata da un vasto arco temporale, arrivando a coprire anche le dichiarazioni più recenti dell'ultimo anno trascorso. Tramite l'apposita tavola di conversione formulata dall'ONU, infatti, la Divisione Statistica, a partire da revisioni moderne della classificazione, si è impegnata nel ricostruire le precedenti versioni, tra cui HS 1992. L'importanza che le serie storiche ricoprono nell'analisi delle dinamiche riguardanti il commercio internazionale ha determinato il successo di questa banca dati, utilizzata anche da importanti organizzazioni come il Cepii, principale istituto francese di economia internazionale.

HS2007

Nel 2007, la WCO ha dato vita ad una sostanziale revisione della classificazione HS: nuovi codici doganali hanno permesso di distinguere merceologie dalle caratteristiche più moderne (ed accorpate altre), consentendo un'analisi di statistica-economica che tenga maggiormente conto dei fenomeni di innovazione produttiva, con particolare riguardo per il settore ICT. Tale classificazione, tuttavia, copre l'informazione sui flussi di com-

mercio estero per un arco temporale relativamente ridotto, che va dal 2007 al 2013.

HS2012

Il passare degli anni ha portato il WCO a sviluppare nel corso del 2012 la quinta revisione periodica della classificazione HS. Attualmente è questa la classificazione vigente. Per questo motivo, in un'ottica di continuo ammodernamento degli strumenti d'analisi a disposizione, StudiaBo ha ritenuto opportuno adottarla, nonostante l'esiguo arco temporale coperto (2012-2013). Dall'altro lato, per non perdere le preziose informazioni derivanti dall'analisi delle serie storiche, la procedura StudiaBo ha identificato 2 anni "cardini", ovvero il 2012 ed il 2008, che hanno consentito, tramite l'utilizzo dei tassi di variazione, la ricostruzione dinamica dei flussi espressi in revisioni ormai superate.

Altre fonti

L'obiettivo di disporre di una banca dati in grado di supportare l'analisi anche dei fatti più recenti ha reso necessario lo sviluppo di una procedura per l'aggiornamento periodico della banca dati Ulisse StudiaBo. Tale risultato è stato possibile avvalendosi delle informazioni congiunturali messe a disposizione dalle seguenti fonti:

- Banca dati Comext sul commercio con l'estero mensile dei paesi UE ed Efta, prodotta da Eurostat (<http://epp.eurostat.ec.europa.eu/>);
- Banca dati Comtrade a livello mensile prodotta dalla Divisione Statistica delle Nazioni Unite (<http://comtrade.un.org/monthly/>);

- Banca dati UsaTrade sul commercio con l'estero mensile degli Stati Uniti, prodotta dal U.S. Census Bureau (<https://usatrade.census.gov/>);
- Banca dati Trade Map a livello mensile, prodotta da International Trade Center (<http://www.trademap.org/Index.aspx>).

Allo stato attuale il campione considerato conta circa 70 paesi¹, che copre circa l'80% del commercio mondiale.

Eurostat - db Comext

Eurostat Comext riporta per i 28 paesi UE e 4 paesi Efta le dichiarazioni di commercio estero pervenute, raccolte e divulgate in modo strutturato ogni mese sia tramite dvd che attraverso sito internet. I paesi membri dell'Unione Europea utilizzano un livello di classificazione ancor più dettagliata rispetto alla HS, denominata "Nomenclatura Combinata" a 8 digits (NC8), di validità annuale. La NC8 consiste in un'ulteriore specifica della HS: ogni capitolo HS identificato da un codice a 6 digits è costituito dalla somma di più sottocapitoli NC8. Per ogni flusso mensile di commercio internazionale Comext riporta il valore in euro e la quantità commercializzata espressa in kg e/o in unità di misura supplementare. I dati vengono aggiornati mensilmente a scadenze fisse, con un ritardo di 6 settimane rispetto al mese concluso per i flussi con partner extra-europei e 10 settimane per i flussi con partner intra-europei. Diverso il caso dei 4 paesi Efta, i cui flussi di commercio internazionale sono dichiarati in classificazione HS a 6 cifre (ultima revisione) e riportati in valore espresso in euro ed in kg.

UN - db Monthly Comtrade

Simile alla raccolta dei dati Efta, risulta anche la banca dati Monthly Comtrade, che per ogni flusso

¹Per il dettaglio dei paesi facenti parte del campione Congiuntura Mondiale Ulisse, si faccia riferimento alla sezione "Documenti/Classificazione Paesi Ulisse" di siulisse.it

di commercio estero ne esplicita il valore in dollari e in kg. Come afferma la stessa Divisione Statistica delle Nazioni Unite, trattandosi di un progetto sperimentale iniziato nel 2010, Monthly Comtrade non è in grado di fornire informazioni sulla revisione della classificazione utilizzata dagli stati nelle loro dichiarazioni, nè di esplicitare informazioni in unità di misura supplementare. Pertanto, nell'ambito della costruzione della banca dati Ulisse, StudiaBo ha deciso di avvalersi delle dichiarazioni di paesi che hanno dimostrato di aderire con continuità al progetto Comtrade.

US Census Bureau - db UsaTrade

U.S. Census Bureau veicola le statistiche di commercio con l'estero degli Stati Uniti d'America tramite il portale UsaTrade Online (<https://usatrade.census.gov/>). Questa fonte contiene le dichiarazioni mensili delle imprese USA relative a tutti i paesi partner del commercio con l'estero degli Stati Uniti, secondo la classificazione doganale Harmonized System a 10 cifre (HS10). HS10 consiste in una specificazione ulteriore rispetto alla classificazione doganale Harmonized System a 6 cifre (HS6). Ogni capitolo HS6 è costituito dalla somma di uno o più sottocapitoli HS10.

ITC - db Trade Map

StudiaBo ha recentemente avviato un progetto di ampliamento del Data Warehouse Congiuntura Mondiale, finalizzato ad integrare le informazioni presenti nella banca dati Trade Map sviluppata dall'International Trade Centre (ITC). ITC è l'agenzia nata nel 1964 per volontà congiunta della World Trade Organization (WTO) e dell'ONU, per migliorare la competitività dei paesi in via di sviluppo sul mercato globale. Trade Map (<http://www.trademap.org/Index.aspx>) è la banca dati che ITC dedica ai flussi di commercio estero, secondo la classificazione doganale Harmonized System a 6 cifre (HS6).

Congiuntura

Come anticipato nel capitolo precedente, i dati congiunturali di commercio estero sono stati utilizzati con l'obiettivo di sviluppare, nell'ambito della costruzione del Data Warehouse **Ulisse**, una procedura in grado di supportare l'analisi anche dei fatti più recenti. StudiaBo, infatti, segue con attenzione e recepisce in tempi brevi le dichiarazioni annuali di commercio con l'estero dei diversi paesi raccolte nel database Comtrade. Come già accennato, tuttavia, la Divisione Statistica delle Nazioni Unite non pone termini per il trasferimento delle informazioni; di conseguenza i flussi commerciali relativi all'ultimo anno trascorso possono spesso risultare carenti. Con il fine di ovviare a questo inconveniente, StudiaBo ha deciso di effettuare una propria ipotesi di chiusura d'anno sulla base delle dichiarazioni mensili disponibili, relative alle banche dati Eurostat Comext, UN Monthly Comtrade, US Census Bureau - UsaTrade, ITC Trade Map. ².

Tale procedura di stima consiste, in primo luogo, nell'isolare l'ultimo biennio trascorso. Quindi, a partire dagli n mesi già oggetto di dichiarazione, per ogni flusso (paese di origine - paese di destinazione) sono stati calcolati i tassi di variazione tendenziali relativi a valori e quantità. Tale operazione consente di stimare le informazioni di commercio estero dei $12-n$ mesi ancora assenti nel database e da qui di ricavare una prima ipotesi di chiusura d'anno: la procedura StudiaBo, infatti, applica alle dichiarazioni annuali Comtrade dell'anno precedente la dinamica espressa dai diversi database mensili. Le informazioni annuali Comtrade, espresse in classificazione HS 2012, sono state così ricostruite applicando al penultimo

anno il tasso di crescita ricavato dalle due banche dati mensili.

Si noti come, nel caso di paesi non UE che non aderiscono a UN Monthly Comtrade, la procedura di stima risulti ancor più "imponente". Dato un flusso paese di origine - paese di destinazione, StudiaBo ha, infatti, isolato le seguenti casistiche:

- se, a fronte della costruzione di un campione chiuso per l'ultimo biennio, è stato possibile calcolare sia un tasso di variazione per esportatore che per importatore, al flusso non ancora esplicitamente dichiarato viene applicato un tasso di variazione pari alla media tra i 2;
- nel caso in cui le informazioni a disposizione consentano di ricavare unicamente il tasso di variazione dell'esportatore o dell'importatore, si utilizza il solo rapporto caratteristico (tasso di variazione) presente per costruire i valori dell'anno più recente;
- quando nel biennio in questione non è stato possibile calcolare nessun tasso di variazione per un dato flusso, a quest'ultimo viene applicato il rapporto tra le esportazioni totali del prodotto nell'ultimo anno storico e le esportazioni totali del prodotto nel penultimo anno.

²Si noti come tale ipotesi sia comunque sempre coerente con quanto riportato da istituti internazionali quali il Fondo Monetario Internazionale e l'olandese Cpb World Trade Monitor.

Normalizzazione

Come anticipato nel capitolo “Presentazione”, le informazioni sul commercio estero si caratterizzano per la doppia dichiarazione: ciascun flusso viene dichiarato sia dal paese esportatore che dal paese importatore, con possibili margini di differenza. Come noto, infatti, i valori dichiarati in import risentono dell’incidenza dei costi di assicurazione e trasporto (CIF: cost, insurance and freight) e risultano pertanto generalmente più alti rispetto ai flussi dichiarati in export (FOB: free on board). A tal proposito la procedura StudiaBo, basandosi sulla metodologia che in letteratura economica si definisce “Analisi delle *Statistiche Mirror*”, ha individuato e trattato in maniera diversa 3 tipi di casistiche:

- Il flusso commerciale viene dichiarato solo dal lato dell’esportatore o solo dal lato dell’importatore (es. per motivi di segretezza commerciale o per esigua importanza economica del flusso). Nella condizione appena descritta, StudiaBo, in accordo con la letteratura economica che stima i costi di assicurazione e trasporto come un fattore proporzionale al valore totale compreso tra 3 e 8%, ha ipotizzato uno scostamento in valore tra dichiarazioni CIF e FOB pari al 5%.
- Il flusso commerciale è dichiarato sia in import che in export e le dichiarazioni risultano tra loro coerenti³. In questo caso nessun aggiustamento viene apportato dalla procedura.
- Il flusso commerciale è dichiarato sia in import che in export, ma le dichiarazioni risultano tra loro incoerenti. Nei casi in cui si

verifica quest’ipotesi, la procedura StudiaBo ricalcola i valori FOB e CIF del flusso utilizzando la seguente media semplice:

$$FOB = (V_{exp} + V_{imp}/1.05)/2 \quad (1)$$

$$CIF = (V_{exp} * 1.05 + V_{imp})/2 \quad (2)$$

dove V_{exp} e V_{imp} rappresentano rispettivamente il valore di export dichiarato dal paese di origine ed il valore di import dichiarato dal paese di destinazione.

La tecnica appena descritta consente di estrapolare il più alto livello informativo dai dati resi disponibili da Comtrade, con particolare riguardo all’elemento di base delle dichiarazioni, ovvero i valori in dollari. L’analisi in quantità, al contrario, richiede nella procedura StudiaBo un ulteriore passaggio a livello metodologico tramite l’identificazione degli outliers e la stima dei dati mancanti, che vedremo successivamente.

³StudiaBo ha ipotizzato di poter considerare coerenti i flussi le cui dichiarazioni in import ed export non differiscano in valore assoluto più del 20%.

Outliers

Una possibile causa di discrepanza tra dichiarazioni di import e dichiarazioni di export in quantità consiste nella presenza di outliers, ovvero errori di misurazione che, se non trattati, vanno ad inficiare la comprensione del fenomeno economico. Eventuali anomalie nelle dichiarazioni, siano esse espresse in kg o nell'unità di misura supplementare, vengono pertanto sottoposte ad alcuni filtri. In primo luogo, StudiaBo ha deciso di eliminare l'informazione in quantità stimata dalla Divisione Statistica delle Nazioni Unite, ritenendo più opportuno far riferimento ad una metodologia interna ben codificata. Quindi, sulla base delle dichiarazioni doganali rimaste disponibili, vengono costruiti i seguenti rapporti caratteristici:

- Valore medio unitario per flusso, dato dal rapporto tra valore monetario e kg;
- Fattore di conversione, dato dal rapporto tra kg e unità di misura supplementare.

A partire dalla distribuzione ordinata di tali rapporti caratteristici, si stabiliscono dei *range* (sulla base del primo e nono decile) entro i quali le osservazioni vengono considerate attendibili. Le osservazioni escluse da tali range sono conseguentemente ricalcolate per rientrare nell'intervallo di validità, determinando un appiattimento (o *smoothing*) della distribuzione. Tale metodologia viene applicata per tutti i beni oggetto di commercio con l'estero che siano stati individuati preventivamente come *differenziabili* ⁴.

Nel caso in cui, invece, l'analisi preliminare abbia concluso la sostanziale omogeneità di un bene,

il range di attendibilità dei rapporti caratteristici viene ristretto attorno alla mediana.

⁴Con la finalità di distinguere beni standard da beni differenziabili qualitativamente, StudiaBo ha costruito un modello econometrico Panel data ad effetti fissi in cui viene stimata l'elasticità delle quote di export alle variazioni di prezzo.

Dati Mancanti

Nei casi in cui le informazioni in quantità risultino “not available” oppure stimate dalla Divisione Statistica ONU, StudiaBo si impegna a ricostruire i dati sulla base della seguente metodologia interna di interpolazione:

- In primo luogo si costruisce la serie storica annuale per prodotto dei valori medi unitari (dati dai rapporti caratteristici tra valori e kg di commercio mondiale del prodotto per ciascun anno). Tale serie rappresenta la dinamica di base seguita dai prezzi della merceologia oggetto d’analisi ed è caratterizzata dalla sola dimensione annuale. Si noti, inoltre, come, nel caso siano disponibili poche informazioni per l’ultimo anno trascorso, il rapporto caratteristico di riferimento è posto pari a quello del penultimo anno. Nel caso, invece, le informazioni a disposizione risultino sufficientemente consistenti, il valore medio unitario dell’ultimo anno viene calcolato sulla base del campione chiuso di osservazioni presenti nel biennio.
- Se non risulta possibile attribuire un prezzo annuale ad una merceologia, StudiaBo effettua una prima interpolazione utilizzando come riferimento gli anni a disposizione.
- Data la serie storica annuale dei valori medi unitari, si costruisce la matrice a 2 dimensioni dove la discriminante dei prezzi non è più solo l’anno ma anche il dato per esportatore. Si noti come, per ragioni legate alla *robustezza* dell’informazione, i valori all’interno di tale matrice vengono esplicitati soltanto per quei paesi che registrano in un anno base (es. 2010) una quota di incidenza nel commercio

mondiale per prodotto pari almeno all’1% del totale. Nel caso, invece, di esportatori meno rilevanti, il rapporto caratteristico di riferimento coinciderà con quello della serie storica per anno. Nel caso che, in un dato anno e per un dato esportatore, non si abbia disponibilità di informazioni in quantità, StudiaBo di nuovo ricorre all’interpolazione, utilizzando l’indice dato dal rapporto caratteristico per esportatore ed anno rispetto al rapporto caratteristico per anno.

- Da ultimo si costruisce il rapporto cartesiano, che individua, per l’arco temporale 1995-2012, tutte le combinazioni possibili di flussi tra paese di origine e paese di destinazione. Quindi per ogni combinazione esportatore ed importatore si effettua l’interpolazione dell’indice dato dal rapporto caratteristico per flusso sul rapporto caratteristico per esportatore.
- Giunti al massimo livello di completezza di dati a disposizione, l’informazione in kg viene ricostruita a partire dai valori medi unitari, ovviamente solo per i flussi che hanno un corrispettivo in valore.
- Conseguentemente alla ricostruzione dell’informazione in kg, tale procedimento è ripetuto per una seconda tipologia di rapporto caratteristico, ovvero il fattore di conversione, che esplicita per ogni merceologia il peso per unità di misura supplementare.

Fasce di qualità/prezzo

Con il fine di focalizzare l'attenzione sul livello qualitativo caratterizzante le merceologie oggetto d'analisi, la procedura StudiaBo attribuisce ad ogni flusso una variabile "R" (*Range*), costruita a partire dalla distribuzione ordinata dei valori medi unitari (dati dal rapporto tra valori e kg). Nel dettaglio, i valori di commercio internazionale vengono deflazionati tramite un indice di prezzo alla produzione (a base 2010), costruito ad hoc per circa 200 famiglie di beni. Tale indice sintetizza gli effetti dei seguenti fattori produttivi:

- costo del lavoro;
- prezzo dei materiali.

In particolare, il costo del lavoro è calcolato a partire dai dati sul salario orario per paese, forniti da U.S. Bureau of Labor Statistics. Il salario orario medio viene pesato per le esportazioni dei 150 paesi Ulisse e quindi "depurato" di una percentuale fissa imputabile alla produttività.

A sua volta il prezzo dei materiali incorpora sia il petrolio (quale proxy del consumo energetico) sia il prezzo della commodity principale intervenuta nel processo produttivo.

Il deflazionamento dei valori medi unitari permette di costruire una distribuzione di prezzi tra loro confrontabili, perchè tutti relativizzati all'anno base 2010. Una volta identificati i decili della distribuzione, questi vengono impiegati come discriminanti per l'identificazione di 5 fasce di prezzo. Successivamente è possibile attribuire una dinamica a dati intervalli: per ogni anno, infatti, vengono individuati i nuovi decili, applicando il tasso di variazione, incorporato nell'indice di prezzo alla produzione, ai decili del 2010, secondo

l'equazione:

$$Decile_{n,year} = Decile_{n,2010} * Indice_{year}/100 \quad (3)$$

Dove $Decile_{n,year}$ rappresenta il decile che nel dato anno ricopre la posizione n , $Decile_{n,2010}$ è il suo corrispettivo calcolato nel 2010 e $Indice_{year}$ rappresenta il valore espresso dall'indice sintetico dei prezzi alla produzione nel dato anno. Pertanto, gli intervalli individuati dinamicamente vanno a distinguere le seguenti fasce di qualità/prezzo:

- *LL* identifica la fascia bassa (low-low), ovvero valori medi unitari al di sotto del II decile della distribuzione ordinata dei prezzi a cui il prodotto viene commercializzato;
- *LM* identifica la fascia medio-bassa (low-medium), compresa tra il II e IV decile, a cui il prodotto viene commercializzato;
- *MM* identifica la fascia media (medium-medium), compresa tra il IV e VI decile, a cui il prodotto viene commercializzato;
- *HM* identifica la fascia medio-alta (medium-high), compresa nell'intervallo VI - VIII decile, a cui il prodotto viene commercializzato;
- *HH* identifica la fascia alta (high-high), riguardante valori medi unitari superiori al VIII decile della distribuzione dei prezzi a cui il prodotto è commercializzato.

Tale procedura di attribuzione, ovviamente, è applicata solo nei casi in cui le merceologie oggetto d'analisi siano state preliminarmente identificate come *differenziabili*, attraverso la stima economica del modello in Panel data.

Nel caso di beni *non differenziabili*, tali per cui ai paesi produttori risulta impossibile ottenere un *premium-price* in ragione di superiori caratteristiche merceologiche, StudiaBo, in fase di costruzione della variabile "R", attribuisce a tutti i flussi una fascia di qualità/prezzo media (MM).

Quantità a prezzi costanti

Come si è detto, le rilevazioni sui flussi di commercio estero riguardano essenzialmente valori ed unità di misura fisiche, generalmente kg. Costruendo, quindi, il rapporto tra valori e quantità, si ottengono quelli che in letteratura economica sono detti *valori medi unitari*, spesso considerati quali proxy dei prezzi. I valori medi unitari sono stati ampiamente utilizzati nell'ambito della procedura StudiaBo per la costruzione del Data Warehouse **Ulisse**. Va tuttavia sottolineato che, se assimilati ad indicatori di prezzo, i valori medi unitari hanno la particolarità di riflettere anche variazioni qualitative del prodotto. Al fine di elaborare analisi che considerino tali modificazioni, StudiaBo ha preferito scorporare gli effetti qualitativi dai prezzi, per includerli in una nuova variabile "Q", che misura la quantità a prezzi costanti. Così facendo è possibile ottenere in seconda battuta un indicatore di prezzo, dato dal rapporto tra valori a prezzi correnti e quantità a prezzi costanti, più *veritiero*, perché non inficiato da modificazioni qualitative. Nel dettaglio, la variabile "Q" è stata costruita con riferimento all'anno base 2010, distinguendo la quantità sulla base della fascia di qualità/prezzo a cui il flusso paese esportatore - paese importatore appartiene. In tale anno la procedura StudiaBo costruisce sia il valore medio unitario specifico per ogni combinazione [paese di origine - paese di destinazione - fascia di prezzo] sia il prezzo medio per fascia di qualità/prezzo. Nel caso in cui, per un dato flusso, si verifichi un aumento qualitativo rispetto al 2010, tale da essere identificato da una diversa fascia di qualità/prezzo, ciò si traduce nell'abbandono del valore medio unitario specifico del 2010 in favore del prezzo medio associato

alla fascia più alta. Ciò determina un conseguente aumento della quantità espressa in "Q", come esplicitato dalla seguente equazione:

$$Q_{flusso,year} = K_{flusso,year} * VMU_{2010} \quad (4)$$

Dove $Q_{flusso,year}$ identifica la quantità a prezzi costanti nel dato anno per una data combinazione [paese di origine - paese di destinazione - fascia di prezzo], $K_{flusso,year}$ identifica i kg oggetto di interscambio commerciale nel dato anno per una data combinazione [paese di origine - paese di destinazione - fascia di qualità/prezzo], VMU_{2010} rappresenta il valore medio unitario associato nel 2010 alla fascia di qualità/prezzo espressa dal flusso.