



Nota metodologica

Data Warehouse

Congiuntura Mondiale

Aprile 2015

Progetto Congiuntura Mondiale

Copyright © 2015 StudiaBo Srl

Aprile 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	4
Eurostat - db Comext	4
UN - db Monthly Comtrade	4
US Census Bureau - db UsaTrade	4
ITC - db Trade Map	4
Principali tecniche di Data Mining utilizzate	5
Dichiarazioni periodiche mancanti	5
Misure mancanti	5
Outliers	6
Pre-stime ultimo trimestre	7
Paesi UE	7
Paesi extra-UE	7
Quantità a prezzi costanti	8

Presentazione

Il presente documento descrive le scelte metodologiche effettuate nell'ambito della creazione del Data Warehouse **Congiuntura Mondiale**, riguardante i flussi trimestrali di commercio estero dichiarati da un campione di circa 70 paesi¹ verso 150 paesi partner mondiali, a partire dal 2011. Tali dati, inerenti le esportazioni e le importazioni, rappresentano un database altamente rappresentativo degli scambi con l'estero a livello internazionale: con la creazione del Data Warehouse Congiuntura Mondiale, StudiaBo stima di esplicitare circa l'80% del commercio internazionale. StudiaBo intende, peraltro, ampliare progressivamente il novero dei paesi dichiaranti considerati nel Data Warehouse Congiuntura Mondiale, a mano a mano che si renderanno disponibili informazioni affidabili in merito².

¹Per il dettaglio dei paesi facenti parte del campione Congiuntura Mondiale Ulisse, si faccia riferimento alla sezione *Classificazione Paesi Ulisse* del portale Ulisse Analytics (<http://uda.studiabo.it/informazioni/#sezClassificazionePaesi>).

²Gli organismi internazionali sono sempre più attivi nella disseminazione dei dati congiunturali di commercio estero dei diversi paesi del mondo. Questa attività sarà continuamente monitorata per verificare la possibilità di integrare e rendere sempre più complete le informazioni del Data Warehouse Congiuntura Mondiale.

Le fonti dei dati

Nel presente capitolo vengono descritte le diverse fonti alla base del Data Warehouse Congiuntura Mondiale.

Eurostat - db Comext

Eurostat Comext (<http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing?sort=1&dir=comext>) riporta le dichiarazioni mensili di commercio estero dei paesi UE e dei paesi EFTA.

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.

I paesi membri dell'Unione Europea utilizzano un livello di classificazione doganale particolarmente dettagliato, denominato Nomenclatura Combinata, a 8 cifre (NC8). NC8 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 NC8.

UN - db Monthly Comtrade

Dal 2010 la Divisione Statistica delle Nazioni Unite porta avanti un progetto sperimentale denominato Monthly Comtrade, per la costruzione di un database online, accessibile all'indirizzo <http://comtrade.un.org/monthly/>, sui flussi di commercio estero mensili dichiarati dai paesi ONU.

Gli interscambi commerciali sono, in questo caso,

esplicitati in dollari e kg, per ogni prodotto della classificazione doganale Harmonized System a 6 cifre (HS6).

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).

Principali tecniche di Data Mining utilizzate

Nel presente capitolo vengono descritte le principali tecniche di *data mining* impiegate nella costruzione del Data Warehouse Congiuntura Mondiale, al fine di poter estrarre dai dati di base informazioni significative, affidabili ed aggiornate.

Dichiarazioni periodiche mancanti

I dati doganali possono essere, a volte, affetti da *buchi* nelle dichiarazioni periodiche nei database di origine.

Per gestire questa casistica di dichiarazioni, nell'ambito della costruzione del Data Warehouse Congiuntura Mondiale, StudiaBo ha utilizzato una tecnica di *riempimento*, basata sull'interpolazione delle serie storiche, previa individuazione - per ogni paese dichiarante oggetto di riempimento - dei paesi partner più frequenti, con cui presumibilmente si potrebbero essere realizzati degli interscambi anche nel periodo di dichiarazione non disponibile.

Misure mancanti

I dati doganali possono presentare delle misure mancanti (codificate nei database di origine come “not available”), soprattutto con riferimento alle quantità.

Al fine di gestire questa casistica di dichiarazioni, StudiaBo ha sviluppato la seguente tecnica, vol-

ta a stimare le misure “not available” (siano esse espresse in kg o in unità di misura supplementare):

- costruzione, per ogni flusso di commercio estero³, di un rapporto caratteristico, denominato Valore Medio Unitario, espresso come rapporto tra valore monetario e quantità espressa in chilogrammi;
- costruzione di un *benchmark* del rapporto caratteristico del flusso, dato dalla media geometrica di due distribuzioni marginali relative alle dimensioni paese dichiarante (valore medio unitario espresso dal paese dichiarante) e paese partner (valore medio unitario insito nelle dichiarazioni aventi un dato paese partner);
- calcolo della serie storica del Valore Medio Unitario per singolo codice doganale, ovvero la distribuzione più *al margine*;
- nei casi in cui il rapporto caratteristico del flusso non risulti disponibile, questo viene ricostruito tenendo conto delle tre componenti di cui sopra, ovvero: dinamica temporale, valore medio unitario espresso dal dichiarante, valore medio unitario caratteristico del partner.

Si noti che, in seconda battuta, tale metodologia si applica anche ad un secondo rapporto caratteristico del flusso, espresso dal rapporto tra le misure di quantità in chilogrammi e in unità di misura supplementare. Appare opportuno sottolineare come in questo secondo ambito le distribuzioni al margine siano caratterizzate da una maggiore stabilità rispetto al singolo evento [export/import del dichiarante verso il partner nella data unità di tempo], proprio in virtù del fatto che rappresentano valori medi, calcolati pertanto su un maggior numero di osservazioni.

La stabilità delle informazioni al margine è

³Un flusso (o evento) è definito come combinazione delle dimensioni export/import di un paese dichiarante verso un paese partner per un dato codice doganale, in una data unità di tempo.

garantita, inoltre, da StudiaBo tramite tecniche specifiche di individuazione e gestione degli *outliers*.

Outliers

Gli outliers sono errori di misurazione che, se non trattati, vanno ad inficiare la comprensione del fenomeno economico. Nella letteratura riguardante i flussi di commercio estero, è noto come tali errori siano più frequenti per le misure in quantità. Per questa ragione StudiaBo, successivamente alla costruzione dei rapporti caratteristici del flusso di cui sopra, sottopone le serie storiche ottenute ad alcuni filtri di controllo. Tramite la distribuzione ordinata di tali rapporti caratteristici, StudiaBo stabilisce dei *range* (sulla base del primo e nono decile della distribuzione) 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 appianamento (*smoothing*) della distribuzione delle osservazioni.

Pre-stime ultimo trimestre

Le dichiarazioni di commercio estero possono essere uno strumento molto potente di misurazione ed analisi 'in tempo reale' dei fenomeni congiunturali. Un ostacolo in tal senso è costituito dalla diversa tempistica con cui queste dichiarazioni possono essere rese disponibili.

Al fine di consentire di monitorare anche i fatti più recenti, StudiaBo ha ritenuto utile mettere a fattore comune l'insieme delle informazioni congiunturali a disposizione, effettuando una pre-stima del trimestre in corso. Si noti come, a causa delle diverse caratteristiche delle banche dati di partenza, anche le metodologie applicate siano diverse. Per descrivere tali metodologie, appare pertanto opportuno distinguerle in funzione della tipologia di paesi dichiaranti.

Paesi UE

Nel caso delle dichiarazioni riguardanti i paesi UE, sfruttando il vasto arco temporale di riferimento della banca dati di origine (dal 2000), StudiaBo si avvale di una modellistica econometrica applicata alle serie storiche, detta ARMA, che risulta particolarmente affidabile nel misurare fenomeni di breve periodo. Tale modellistica econometrica si basa sullo studio dell'*autocorrelazione*, ovvero sulla misurazione del *legame* tra una variabile economica ed il proprio passato. Si parla così di previsioni *non condizionate*, perché indipendenti da ipotesi esterne riguardanti lo scenario internazionale e la politica economica. Una volta completamente disponibili le dichiarazioni ufficiali, queste vanno a sostituire quanto stimato dai modelli ARMA.

Paesi extra-UE

Nel caso dei paesi extra-UE, data la disponibilità di un arco temporale più ristretto (dal 2010) rispetto ai paesi UE, tali dichiarazioni risultano meno adatte alla trattazione con modellistica ARMA. Per questa ragione, le dichiarazioni dei paesi extra-UE vengono stimate tramite l'utilizzo dei tassi di variazione. Tale procedura di stima consiste, in primo luogo, nell'isolare l'ultimo biennio trascorso. A partire dagli n mesi già oggetto di dichiarazione, per ogni flusso si calcolano i tassi di variazione tendenziali relativi a valori e quantità. Tale operazione consente di stimare le informazioni di commercio estero dei mesi ancora assenti nel database e da qui ricavare una prima ipotesi di chiusura del trimestre. Ovviamente la pre-stima effettuata presuppone in partenza l'ipotesi che i mesi non ancora oggetto di dichiarazione siano stati caratterizzati da una dinamica uguale ai mesi già dichiarati.

Quantità a prezzi costanti

Come si è detto, le rilevazioni sui flussi di commercio estero riguardano valori monetari (espressi, nel DW Congiuntura Mondiale, dalla misura V) e quantità (esprese, nel DW Congiuntura Mondiale, dalle misure K e U , che indicano rispettivamente il peso e l'unità di supplementare relativa al prodotto).

Una trattazione poco approfondita delle suddette misure può portare a distorsioni nell'interpretazione dei fenomeni economici; infatti è noto che:

- il valore monetario risente delle variazioni di prezzo nel tempo; pertanto, valori più elevati di commercio estero possono essere potenzialmente riconducibili solo ad aumenti nei prezzi, a parità di quantità scambiate;
- le misure fisiche dipendono dal prodotto: pertanto, per panieri di beni non omogenei⁴, può non avere significato economico sommare le quantità in peso⁵ o esprese nell'unità di misura supplementare.

Al fine di rendere possibili le analisi sugli aggregati economici, è opportuno introdurre una misura

di *quantità economica*. A tal fine statistica economica ha introdotto il concetto di **quantità a prezzi costanti**.

Nel Data Warehouse Congiuntura Mondiale sviluppato da StudiaBo è stata costruita, quindi, la misura *Quantità a prezzi costanti* (Q). In sostanza la misura Q racchiude un'operazione di *deflazionamento*⁶, in cui la serie storica dei valori monetari (V) è stata trasformata in un'analogia serie di valori espressi a prezzi costanti, con un riferimento a un anno determinato, detto anno base (nel caso di DW Congiuntura Mondiale: il 2011).

In altri termini, la metodologia StudiaBo ha deflazionato rispetto al 2011 tutti i valori correnti, secondo la seguente equazione:

$$Q_{flusso,data} = V_{flusso,data} / IP_{data}$$

dove $Q_{flusso,data}$ identifica il valore a prezzi costanti nel dato periodo congiunturale per una data combinazione [paese di origine - paese di destinazione], $V_{flusso,data}$ individua il valore monetario nel dato periodo congiunturale per una data combinazione [paese di origine - paese di destinazione], IP_{data} è l'indice di prezzo costruito come segue:

$$IP_{data} = P_{data} / P_{2011}$$

Il numeratore dell'indice di prezzo P_{data} è dato dal rapporto V_{data} / K_{data} ovvero il valore medio unitario registrato in ogni unità di tempo; il denominatore è dato dal valore medio unitario calcolato nel 2011.

Così facendo, StudiaBo ha ottenuto una serie sto-

⁴La non omogeneità di beni economici può essere dovuta ai seguenti fattori:

- differenze tecniche/qualitative dei beni;
- presenza di servizi aggiunti ai diversi prodotti;
- disponibilità del bene in luoghi diversi;
- disponibilità del bene in tempi diversi.

⁵Si pensi a tal proposito ad una analisi complessiva del commercio mondiale: appare evidente come una aggregazione espressa in kg che comprenda beni diversi tra loro per livelli di qualità (come, ad esempio, frutta, diamanti, cascami in metallo, ecc.) risulti assolutamente priva di senso.

⁶In statistica economica il *deflazionamento* assume grande importanza nel consentire confronti inter-temporali, ossia quando si vogliono analizzare le variazioni reali (di quantità) subite da un certo aggregato. Infatti, le variazioni nel tempo dei valori correnti sono il risultato di variazioni complesse che interessano sia i prezzi che le quantità. Mediante il *deflazionamento*, le differenze di valore fra aggregati in periodi diversi sono "depurate" dalle variazioni dei prezzi, poiché gli aggregati stessi sono espressi sempre nei prezzi del periodo base.

rica della *quantità economica* del fenomeno considerato, in grado di esprimere modificazioni non solo delle quantità fisiche dei beni considerati, ma anche della loro qualità.