



Nota metodologica
Data Warehouse
Congiuntura Paesi UE

Gennaio 2015

Progetto Congiuntura Paesi UE

Copyright © 2015 StudiaBo Srl

Gennaio 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
Costruzione serie storiche	5
La fonte dei dati	5
La Teoria dei grafi	5
Armonizzazione	7
Rettifica	7
Outliers	8
Pre-stime trimestre in corso	9

Presentazione

Il presente documento descrive le scelte metodologiche effettuate nell'ambito della creazione del Data Warehouse **Congiuntura Paesi UE**, riguardante i flussi di commercio estero dei paesi UE verso 150 paesi mondiali, di fonte Eurostat. I dati inerenti le esportazioni e le importazioni dichiarati dai paesi UE rappresentano un database ricco di contenuto informativo: l'aggiornamento strutturato mese a mese con cui vengono divulgati, infatti, permette l'analisi anche dei fatti più recenti, con un livello di dettaglio merceologico particolarmente elevato. E' possibile, inoltre, valutare la competitività dei diversi paesi UE, 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. Nel caso dei paesi UE, con l'avvento del Mercato Unico Europeo e la rimozione delle formalità doganali (fonte tradizionale di informazioni statistiche sul commercio internazionale) si è rafforzata l'adozione di una nuova modalità di *data collection*, denominata Intrastat. Nel dettaglio, i flussi di commercio estero sono stimati dai diversi istituti nazionali di statistica, sulla base delle dichiarazioni Iva effettuate dalle varie imprese. Generalmente, tuttavia, queste procedure di stima sono di elevata affidabilità.

Confidenzialità. A volte i flussi di commercio estero possono essere segreti, perchè ritenuti di carattere confidenziale. I paesi più

piccoli, in particolare, possono presentare una significativa incidenza di flussi segreti: è il caso, ad esempio, di Austria e Danimarca. Nei diversi anni, tali stati presentano una quota di flussi segreti dal lato delle esportazioni che può arrivare a superare il 5% del loro export totale. L'alto livello di dettaglio sia dal punto di vista della classificazione merceologica sia dal punto di vista dei paesi oggetto dell'interscambio commerciale, infatti, può dar luogo a misurazioni volte a salvaguardare le strategie competitive dei diversi paesi. Per questa ragione, la confidenzialità concerne sia i prodotti oggetto di commercio internazionale sia l'identificazione dei partner commerciali.

Rotture delle serie storiche. Come si è detto, i paesi membri dell'Unione Europea utilizzano un livello di classificazione merceologica particolarmente dettagliato, denominato "Nomenclatura Combinata" a 8 digits (NC8). La NC8 è sottoposta a revisione ogni anno, con il fine di garantire la massima coerenza tra gli effettivi sviluppi di prodotto e la classificazione stessa. Cambiamenti così frequenti, tuttavia, rendono difficile la lettura dei dati storici, a causa di possibili *rottture* temporali nel passaggio tra una revisione e l'altra. Questo è sicuramente il più grande ostacolo ad una fruizione efficace dei dati di commercio con l'estero di fonte Eurostat.

Per poter estrarre dai dati di base informazioni significative, affidabili ed aggiornate, 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. Il progetto Eurostat Comext presenta un alto livello di strutturazione ed integrazione, tale per cui tutti i paesi UE fanno pervenire mensilmente all'istituto di Statistica Europeo i dati di commercio estero per codice prodotto e per paese partner. Ne consegue che, ogni mese, ciascun paese pubblica dati relativi a vari milioni di flussi commerciali: l'elevata numerosità di informazioni consente, quindi, di individuare facilmente eventuali dati anomali.

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

Le caratteristiche appena descritte 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*.

Il presente documento delinea le diverse metodologie applicate in fase di costruzione del Data Warehouse Congiuntura Paesi UE. Tali metodologie riguardano:

Costruzione delle serie storiche. Come si è detto, i frequenti cambiamenti a cui è sottoposta la classificazione NC8 rendono difficile la lettura dei dati storici, a causa di possibili *rottture* temporali nel passaggio tra una revisione e l'altra. StudiaBo, avvalendosi della *Teoria dei grafi*, ha costruito uno strumento informativo ad hoc che, tenendo conto dei cambiamenti intervenuti nella classificazione NC8, è in grado, a partire dalla classificazione vigente, di ricostruire interamente l'evoluzione del commercio con l'estero dei diversi prodotti.

Armonizzazione. Dati storici e dati recenti possono differire significativamente in termini di rilevazione statistica: i dati storici presentano un alto grado di affidabilità perchè già *certificati* dagli Stati Membri e da Eurostat. I dati più recenti, invece, come è noto, sono passibili di modifiche e revisioni. StudiaBo, tramite l'individuazione di un anno cardine le cui dichiarazioni sono presenti a database Eurostat sia nella loro forma provvisoria che definitiva, ha effettuato un'ipotesi di rettifica anche delle informazioni più recenti.

Outlier. I dati di commercio internazionale riguardanti quantità fisiche sono stati trattati secondo metodologie finalizzate all'individuazione di eventuali *errori di misura* e alla loro sostituzione con valori più coerenti con le altre osservazioni;

Prestime. Con il fine di utilizzare tutta l'informazione EUROSTAT a disposizione e sviluppare analisi sempre aggiornate, StudiaBo si avvale di modelli econometrici detti ARMA (Auto Regressive Moving Average) per effettuare prestime di chiusura del trimestre in corso.

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

Costruzione serie storiche

La fonte dei dati

Eurostat Comext riporta per i 27 paesi UE le dichiarazioni di commercio estero pervenute, raccolte e divulgate in modo strutturato ogni mese sia tramite dvd che attraverso sito internet (<http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing?sort=1&dir=comext>).

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 merceologica particolarmente dettagliato, denominato "Nomenclatura Combinata" a 8 digits (NC8), di validità annuale. La NC8 consiste in una specifica del "Harmonized System" (HS), ovvero la classificazione sviluppata dalla World Customs Organization (WCO) e utilizzata anche dalla Divisione Statistica delle Nazioni Unite nell'ambito del suo progetto per la costruzione e l'aggiornamento continuo del database Comtrade. La HS è sottoposta a revisione ogni 5 anni, l'ultima volta nel 2012, e propone un livello di disaggregazione massima a 6 cifre. Ogni capitolo HS è costituito dalla somma di più sottocapitoli NC8. Come si è detto, la "Nomenclatura Combinata" è sottoposta a revisione ogni anno, con il fine di garantire la massima coerenza tra gli effettivi sviluppi di prodotto e la classificazione stessa. Cambiamenti così frequenti, tuttavia,

rendono difficile la lettura dei dati storici, a causa di possibili *rottture* temporali nel passaggio tra una revisione e l'altra. StudiaBo, avvalendosi della *Teoria dei grafi*, ha costruito uno strumento informativo ad hoc che, tenendo conto dei cambiamenti, è in grado, a partire dalla classificazione vigente, di ricostruire interamente lo storico dei prodotti.

La Teoria dei grafi

In matematica, la *Teoria dei grafi* si occupa di studiare oggetti discreti, detti appunto grafi, che permettono di schematizzare una grande varietà di situazioni e di processi, consentendone così l'analisi in termini quantitativi e algoritmici. Per grafo si intende una struttura costituita da:

- oggetti semplici detti nodi;
- collegamenti tra i nodi denominati archi o spigoli.

La *Teoria dei grafi* è applicata nei più disparati campi: nell'analisi delle dinamiche sociali, nella regolamentazione del traffico stradale o nella biologia, ad esempio nello studio delle reti di interazioni tra proteine, cioè in tutte quelle situazioni dove l'ambiente di analisi può essere intuitivamente rappresentato mediante una rete. Nel caso del Data Warehouse Congiuntura Paesi UE, il grafo rappresenta la *storia* dei singoli codici prodotto di Nomenclatura Combinata (circa 9000 item). I nodi sono costituiti dai codici prodotto dei singoli anni, mentre gli archi dai *movimenti* annuali che tali codici hanno subito con i vari aggiornamenti. A partire da queste informazioni, con l'ausilio delle tabelle di raccordo anno ad anno fornite da Eurostat, StudiaBo ha costruito a ritroso la serie storica dei codici prodotto NC8 del 2014. Detto in altri termini, a ciascun prodotto classificato secondo lo schema doganale vigente, StudiaBo ha associato gli anni della serie storica. Quindi, per ciascun anno, ha individuato i codici prodotto

NC8 confluiti nel codice NC8 del 2014. Secondo la presente metodologia, la ricostruzione a ritroso si interrompe nel momento in cui, nel passaggio dall'anno $t-1$ all'anno t , si verifica una disaggregazione tale da non poter ricondurre in maniera univoca il codice prodotto esistente al tempo $t-1$ a quello classificato secondo la NC8 dell'anno t . I codici che non trovano un corrispettivo confluiscono in codici residuali costruiti appositamente da StudiaBo, con il fine di realizzare analisi settoriali coerenti nel tempo.

La Teoria dei grafi, applicata ai dati sul commercio estero, assume particolare rilievo se si considera l'importanza che le serie storiche ricoprono nelle analisi in dinamica degli interscambi commerciali. Il tema è ben noto in letteratura economica, tanto che, nel caso del database Comtrade, la stessa Divisione Statistica delle Nazioni Unite ha provveduto sia a pubblicare le informazioni nella classificazione originaria utilizzata dai paesi dichiaranti, sia ad effettuare una conversione nelle classificazioni precedenti fino alla HS datata 1992. Eurostat, al contrario, riporta per ogni anno i dati nella classificazione vigente, rendendo ostiche le elaborazioni relative ad un arco temporale più ampio. Pertanto, la metodologia StudiaBo è in grado di apportare un elevato valore aggiunto rispetto alle informazioni presenti in Comext. La metodologia introdotta da StudiaBo risulta particolarmente ambiziosa, perché, a differenza di quanto realizzato dall'ONU che aggiorna i dati relativi alle classificazioni merceologiche di origine, essa mira ad *attualizzare* alla nomenclatura vigente le serie storiche dei flussi di commercio con l'estero.

Armonizzazione

Rettifica

Una volta costruito il grafo che sintetizza la storia delle merceologie, la metodologia StudiaBo tiene conto del diverso livello di aggiornamento delle informazioni per anno. I dati più recenti, relativi al 2013 e 2014, infatti, riguardano dichiarazioni ancora passibili di modifiche, soprattutto in ragione del fatto che, tipicamente, le imprese europee forniscono informazioni con tempistiche diverse, a seconda della loro dimensione.

StudiaBo pertanto, effettua un'ipotesi di rettifica e validazione, attraverso l'utilizzo di un anno cardine. Nella banca dati EUROSTAT Comext i dati relativi al 2012 sono riportati sia con i loro valori originari (non rettificati), sia con valori rettificati in un secondo momento. Attraverso la costruzione del rapporto caratteristico tra il dato 2012 rettificato ed il dato 2012 non rettificato, StudiaBo ha definito i nuovi valori relativi agli anni 2013 ed al 2014. Il rapporto per flusso, infatti, viene applicato ai dati più recenti, per attribuire ad esso un valore che possa essere considerato definitivo e quindi confrontabile con i dati storici.

Outliers

Un caso particolare, e generalmente difficile da codificare, di discrepanza tra dichiarazioni di import e di export in quantità consiste nella presenza di outliers. Gli outliers sono errori di misurazione che, se non trattati, vanno ad inficiare la comprensione del fenomeno economico. StudiaBo, pertanto, sottopone le dichiarazioni presenti in Eurostat Comext (siano esse espresse in kg o nell'unità di misura supplementare) ad alcuni filtri di controllo. La metodologia di filtraggio utilizzata si basa in primo luogo sulla costruzione dei seguenti rapporti caratteristici:

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

Tramite la distribuzione ordinata di tali rapporti caratteristici, StudiaBo stabilisce 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 appianamento (*smoothing*) della distribuzione.

Pre-stime trimestre in corso

I dati di commercio estero possono essere uno strumento molto potente di misurazione ed analisi dei fenomeni congiunturali. Pur provvedendo mensilmente all'aggiornamento del suo database, Eurostat prevede, tuttavia, tempistiche diverse a seconda che le dichiarazioni abbiano come partners paesi europei o paesi extra-europei. Con il fine di sfruttare l'informazione a disposizione per sviluppare analisi congiunturali, StudiaBo si avvale di metodologie di previsione applicate alle serie storiche che risultano particolarmente affidabili nel misurare i fenomeni di breve periodo. In particolare, il modello proposto da StudiaBo si definisce ARMA (1,1), ovvero autoregressivo a media mobile di ordine 1. 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 condizionali, perchè indipendenti da ipotesi esterne riguardanti lo scenario internazionale e la politica economica. Come si evince dall'equazione sottostante, il modello dedica grande enfasi allo studio della struttura temporale dei dati:

$$X_t = c + aX_{t-1} + Z_t - bZ_{t-1} \quad (1)$$

dove X_t rappresenta il valore delle esportazioni al tempo t , quale variabile che dipende dalla costante c , da una componente autogressiva costituita dalle esportazioni al tempo passato $t-1$ e da una componente a media mobile, formata dagli shock casuali presenti (Z_t) e passati (Z_{t-1}).

Nel dettaglio, StudiaBo destagionalizza le serie storiche mensili per prodotto, considerandone la trasformata logaritmica, a cui applica l'opera-

tore *differenza prima* di ordine 12. Una volta formulata l'ipotesi sul modello generatore dei tassi di variazione tendenziali, la procedura StudiaBo stima i coefficienti del modello per effettuare la previsione di breve periodo¹. Output di tale operazione è il tasso di variazione previsto mese per mese. La stima viene quindi utilizzata per determinare i flussi intra UE non ancora dichiarati, da sommare alle dichiarazioni che hanno come partner paesi extra UE. Tale operazione consente di effettuare un'ipotesi particolarmente robusta di chiusura dell'ultimo trimestre in anticipo rispetto alla pubblicazione dei dati. Una volta completamente disponibili le dichiarazioni mensili ufficiali, queste vanno a sostituire quanto stimato dai modelli ARMA, in un'ottica di costante aggiornamento del Data Warehouse Congiuntura Paesi UE.

¹Come è noto a livello statistico, la differenza prima del logaritmo di una serie ne approssima il tasso di variazione.