Conjuncture of EU Contries' foreign trade:

Methodological note

September 2018

Methodological note of ***Conjuncture EU Countries***

September 2018

StudiaBo srl
via Santo Stefano, 57
40125 Bologna Italy

Table of Contents

Presentation

This document describes the methodological choices made in the creation of the database **Conjuncture - EU Countries**, dealing with the quarterly foreign trade flows of EU countries collected by Eurostat. These data represent a rich database of information content for the following main reasons:

- they allow an analysis of more recent events, at a very high product detail;
- they allow to assess the competitiveness of the different EU countries, highlighting their strengths and weaknesses.

Often, however, this information is not immediately accessible from a simple reading of the data. The statements of different countries sometimes contain data that does not correspond to the actual flows of foreign trade. The main reasons for that are as follows:

Estimation procedures : in the case of EU countries, the advent of the Single European Market and the removal of customs formalities (traditional source of statistical information on international trade) caused the adoption of a new method of data collection called Intrastat. In detail, foreign trade flows are estimated by the different national statistical institutes, based on the VAT returns made by the various companies, and not collected when the goods arrive at the different customs borders. Generally, however, these estimation procedures are of high reliability.

Confidentiality : sometimes foreign trade flows by country and product can be secreted because they are considered confidential. Each country follows its own rules, with a higher incidence of secreted flows for smaller countries. Among the countries of the European Union some ones (i.e. Denmark, Finland, Austria) present a ratio of secreted flows more than 5% of their total exports.

Breaks in time-series : EU countries use a very detailed level of product classification, called Combined Nomenclature, at 8-digits (CN8). The CN8 classification is subject to review each year, in order to ensure maximum consistency between the classification and the actual products' developments. Such frequent changes, however, make it difficult to read the historical data, because of possible temporal breaks in the passage between a revision and the other. This is definitely the biggest obstacle to an effective use of the data of foreign trade collected by Eurostat.

In order to extract meaningful, reliable and up-to-date information from the Eurostat data base, it is therefore necessary to use a variety of methodological tools designed to distinguish the *measure* of the phenomenon from the *statistical noise*. It is therefore essential to build analytical procedures that, on the basis of what has just been described, best use the many advantages inherent in the international trade data, namely:

High Numerousness : Eurostat Comext project has a high level of structuring and integration, such that all EU countries send their monthly foreign trade data by product code and by partner country. It follows that, each month, each country publishes data related to several million trade flows: the high number of information thus makes it possible to easily identify possible outliers.

Data refer to the population : foreign trade statistics affect almost all trade flows in EU. This allows to avoid all the problems regarding sampling and measurement errors that can arise from their use.

The characteristics just described allow to make effective use of some data mining methodologies, able to isolate in a satisfactory way the *measure* of different phenomena from its *error term*. This document outlines the different methods applied in the construction of the database **Conjuncture - EU Countries**.

Construction of time-series

Data sources

Eurostat disseminates in a structured manner each month the foreign trade statements of the EU countries (http://bit.ly/2NY1hsk).
For each monthly flow of foreign trade Eurostat returns the value in euros and the quantity expressed in kilograms and / or in another supplementary unit measure.
The data is updated monthly at a fixed time, generally with a delay of 6 weeks compared to the month ended for the flows with non-European partners and of 10 weeks for the flows with intra-European partners.
The EU countries use a very detailed level of product classification, called Combined Nomenclature at 8 digits codes (CN8) of annual validity. The CN8 consists in a specific branch of the Harmonized System at 6 digit codes (HS6), developed by the World Customs Organization (WCO) and also used by the Statistics Division of the United Nations as part of its project to build and continuously update the Comtrade database. The HS6 is subject to review every five years, the last time in 2017. Each HS6 chapter is, therefore, the sum of several CN8 codes.
As mentioned above, the Combined Nomenclature is revised every year, in order to ensure maximum consistency between the classification itself and the actual products' developments. Such frequent changes, however, make it difficult to read the historical data, because of possible temporal breaks in the passage between a revision and the other.
The issue is well known in the economic literature, so that, in the case of UN Comtrade database, the Statistics Division of United Nations has provided both the publication of the information in the original classification codes used by reporting countries, both a conversion of newly classification codes in the previous classifications. On the contrary, Eurostat disseminates data only in the current product classification codes.
StudiaBo, using the so called *Graph Theory*, has built a specific information tool, which, taking into account all the historical changes, is able to rebuild the entire time-series backwards from the classification in force.
The methodology introduced by StudiaBo is particularly ambitious because, unlike what has been achieved by the UN that updates data on product classification codes of origin, it aims to rebuild backwards the time-series of current product classification codes.

The Graph Theory

In mathematics, the Graph theory is the study of discrete objects, named graphs, that allow to analyse a variety of situations and processes in quantitative and algorithmic terms. A graph refers to a structure consisting of:

- simple objects, called nodes;
- links between nodes, called arcs or edges.

The Graph theory is applied in various fields where the analysis environment can be intuitively represented by a network: for instance, in the analysis of social dynamics, in the regulation of road traffic, or in biology. The Graph theory applied to the foreign trade data is of particular importance when one considers the importance of the length of time-series

for analysis purpose. In the case of database **Conjuncture - EU Countries**, the graph represents the history of the individual product codes of the Combined Nomenclature (about 9000 items). The nodes consist of the product codes in the single years, while the arcs consist of the changes that such codes have suffered with various updates. From this information, with the help of the bridging tables year to year provided by Eurostat, StudiaBo has rebuilt the time-series backwards of NC8 2018 product codes.
According to the present methodology, the backwards reconstruction stops in the moment in which, in the transition from year t-1 to year t, occurs a breakdown that cannot bring in a univocal way the product classification code existing at year t-1 to the classification of year t. The codes that do not match a corresponding code backwards are put into some residual codes specifically built by StudiaBo with the aim of ensuring sectoral analyses consistent over time.

Normalisation

Adjusted and non-adjusted figures

Once built the graph that summarizes the history of a single product code, the StudiaBo methodology takes into account the updating status of information available each year. The most recent data (i.e. related to years 2017 and 2018), in fact, are still subject to changes (*adjustments*), especially due to the fact that European companies provide information with different timing, depending typically on their size. StudiaBo therefore build up an hypothesis of adjustments, through the use of a characteristic ratio between adjusted and non-adjusted figures in a pivotal year.

Outliers

A special case, and generally difficult to encode, of any discrepancy between statements of imports and exports in quantity is the presence of outliers.
Outliers are measurement errors which, if not treated, might affect the understanding of the economic phenomenon. StudiaBo, therefore, submit the statements collected by Eurostat (whether expressed in kilograms or in any supplementary unit measure) to some control filters.
The filtering methodology used by StudiaBo is based primarily on the construction of the following key ratios:

- average unit value per flow, given by the ratio between values and kilograms;
- conversion factor per flow, given by the ratio of kilograms and the supplementary measure.

On the basis of the orderly distribution of these key ratios, a *range of reliability* is defined (based on the first and ninth decile of the distribution) within which the observations are considered to be reliable. Observations excluded from this range are therefore recalculated to be in the range of validity, resulting in a *smoothing* of the distribution.

Estimation of Quarter-End

Foreign trade data may be a very powerful tool for measuring and analysing economic phenomena. While meeting monthly updating of its database, Eurostat provides, however, different timings for European and non-European partner countries. With the aim to exploit all the information available at a certain time, StudiaBo has developed ARMA forecasting methodologies, which are particularly reliable tools to make estimations in the short term. In particular, the model proposed by StudiaBo is ARMA (1,1), that is to say autoregressive with a moving average of order 1. This econometric modeling is based on the study of the *autocorrelation* (i.e. the measurement of the *tie* between an economic variable and its past). We can speak of unconditional forecasts, because independent of external assumptions concerning the macroeconomic scenario and economic policies in place. As can be seen from the equation below, the model pays great attention to the study of the temporal structure of the data:

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

where X_t represents the value of exports at time t, as a variable which depends on the constant c, on a autoregressive component constituted by exports in the past t-1, and on a moving average component, formed by random shocks at present (Z_t) and in the past (Z_{t-1}).

In detail, the StudiaBo procedure calculates seasonally adjusted monthly time series for single product, considering the logarithmic transform, to which the operator *first difference* of order 12 is applied. Once formulated the hypothesis on the pattern generator of trend rates, the StudiaBo procedure estimates the coefficients of the model to predict short-term forecasts [1]. Output of the procedure is the change rate expected for each month. The estimate is then used to determine the intra-EU flows that have not yet been declared, to be added to the statements regarding non-EU partner countries. This procedure allows to make an hypothesis particularly robust for closing a quarter well in advance of the publication of the data. Once official monthly statements are fully available, these replace the figures estimated by ARMA models.

[1] As is well known in statistics, the first difference of the logarithm of a number approximates the rate of change.