

How Similar Are International Economic Relations of EU Member States? Comparing Trade, Investment and Political Behavior

Matteo Fiorini Miklós Koren Filippo Santi Gergő Závecz

RESPECT Closing Conference

Motivation

Data

Trade, Investments, and additional controls

Export data come from COMEXT (Eurostat 2019). We use the chapter-level product distribution of bilateral exports between EUMS and their Extra-EU partners, measured between 2001 and 2017 (although most analysis uses the years 2015-17 due to unavailability of other political measures discussed below). Because we do not have access to shipment-level data, we approximate the number of shipments by dividing the value of exports by EUR 12,000, following estimates in Hornok and Koren (2015a,b).

Investment data come from the fDIMarket database (Financial Times, 2019). We aggregate single Greenfield FDI transactions at origin-destination-sector level over the period 2003-2018, focusing on the flows that occurred between EUMS and the rest of the world. Also in this case, we divide the value of larger-than-average investments by the average investment value (computed on strictly positive transactions only). In other words, we are assuming that a factor- n larger investment is comparable to n average size FDI. This assumption is required to reduce data sparsity.

Political Variables

To study the behavior of Member States, we turn to media mentions of state visits and similar events. For what concern our main variable of interest, we extract information on the number of events about economic or diplomatic cooperation between two state actors for the period 2015-17 from the Global Database of Events, Language and Tone (The GDELT Project 2020).

In particular, we limit our attention to positive, cooperative events (as coded by GDELT) in which government agencies or decision makers from EUMS are considered as the initiating actors''. There are two groups of cooperative events: one classified as intent'' (intent to cooperate) and one classified as ''visits.'' In this latter category we include state visits, formal negotiations, signing of agreements, and material cooperation. With reference to any member state, we tally all such events happening in, or being related to, any potential partner country.

As an example, let us focus on a single member state, say France. Then, the French foreign minister visiting Turkey could be one event,

Methods

Export composition

The exports of country i to country j in year t is hence characterized by a sequence of shares, $s_{ijt1}, s_{ijt2}, \dots, s_{ijtP}$, where P is the overall number of products, with the shares summing to one, $\sum_{p=1}^P s_{ijtp} = 1$. These value shares completely characterize the trade structure of a pair of countries for our purposes. We also control for the overall volume of trade.

Comparing country i to the EU average (denoted by $*$) amounts to comparing two sets of shares, $\{s_{ijtp}\}$ and $\{s_{*jtp}\}$. Our goal is to ask if country i 's trade shares are different from the EU average, and if so, to quantify the magnitude of the difference.

Industry similarity indexes between regions and countries have been proposed in other contexts by Finger and Kreinin (1979), Krugman (1991), and Fontagné et al (2018). In contrast to these, our proposed index of similarity is based on an economic choice model (Anderson et al. 1992).

Kullback-Leibler Divergence

Our preferred measure of difference between country-specific and EU trade shares is the Kullback-Leibler divergence (Kullback 1987, KLD henceforth), defined as

$$\text{KLD}_{ijt} = \sum_{p=1}^P s_{ijtp} \ln(s_{ijtp}/s_{*jtp}). \quad (1)$$

This is a measure of distance between the two distributions, only taking the value zero if all the products have the same share, and positive otherwise. As mentioned above, a key benefit of this index is that it is based on utility maximizing decision model. More specifically, take a consumer with logit preferences (a standard assumption in discrete choice models) whose ideal consumption shares are given by s_{ijtp} . If this consumer instead consumes the products in shares s_{*jtp} , her utility will be reduced by a magnitude proportional to the KLD between s and s^* .

The problems with sparsity

In practice, the KLD index will never be zero, as no two countries have exactly the same product composition of exports. In order to quantitatively judge what constitutes a significant gap between the trade composition of two countries, we test whether the KLD is significantly different from zero. This is important because the KLD index will be biased upwards in small samples.

To test for statistical significance and to mitigate small sample bias, we conduct the following procedure. Let x_{ijtp} denote the number of export shipments from country i to country j in product p in year t . Shipments are the basic units of observation in trade statistics, and a small number of shipments can lead to small-sample bias (Armenter and Koren 2014). The total number of shipments between a pair of countries $n_{ijt} = \sum_p x_{ijtp}$ is taken as given.

The Polya Index

The *null hypothesis* is that all countries' shipments are distributed according to the same distribution. We chose the multivariate Polya distribution (Eggenberger and Pólya 1923) as the parametric distribution that best suites this application.

More specifically, $\{x_{ijtp}\}$ is assumed to be distributed according to the Polya distribution with expected product shares $\{\pi_{jtp}\}$ and a precision T_{jt} . We estimate the expected shares and the precision parameter with maximul likelihood, separately for each partner country j and year t .

Under this null hypothesis, the KLD index has a distribution F_{ijt} :

$$\Pr(\text{KLD}_{ijt} \leq x | n_{ijt}) = F_{ijt}(x). \quad (2)$$

Computing this distribution in closed form is possible, but requires prohibitively many combinatorial steps. We would have to compute the probability of each possible allocation of shipments, for thousands of shipments. With a 100 product categories, even just 1,000 shipments could be distributed about 10^{39} different ways.

Computation

Instead, we approximate $F()$ with its empirical distribution function. We simulate the distribution with 10,000 Monte Carlo draws and define $\hat{F}_{ijt}(x)$ as the fraction of draws in which the simulated KLD index is smaller or equal to x .

We then define the *Polya Index* as the tail probability of the empirical distribution, evaluated at the actual KLD,

$$\text{Polya}_{ijt} \equiv 1 - \hat{F}_{ijt}(\text{KLD}_{ijt}). \quad (3)$$

The Polya Index captures the probability that we would observe the measured KLD index or higher, conditional on all countries' trade structures being the same in expectation. Formally, it is the statistical size of a one-sided test of the null hypothesis that all countries have a KLD of zero (with their shares generated from the Polya distribution).

Interpretation

The Polya Index is an index of similarity. When the product distribution of the country is statistically indistinguishable from the rest of the countries, Polya_{ijt} is very close to one. By contrast, low levels of the Polya Index mean that we can reject the null hypothesis of similarity. It is important to note, however, that a large Polya Index does not necessarily mean a full alignment of the country's trade structure with that of the EU. It can also arise when we have too few transactions to statistically differentiate the two trade structures. A low Polya Index, on the other hand, surely indicates significant differences.

With the due differences, the Polya Index can also be applied to bilateral investments, although the sparsity of investment transactions might make harder for statistically significant differences to emerge in investment portfolios (compared to trade). Hence, we expect the Polya Index for investment to be larger than its trade counterpart.

Findings