Why we like the ECI+ algorithm

Andrea Gabrielli[1], Matthieu Cristelli[1], Dario Mazzilli[2], Andrea
Tacchella[1], Andrea Zaccaria[1], and Luciano Pietronero[*1,2]

[1]Institute for Complex Systems, CNR
[2]Department of Physics, Sapienza University of Rome

August 2, 2017

Abstract

Recently a measure for Economic Complexity named ECI+ has been
proposed by Albeaik et al. We like the ECI+ algorithm because it is math-
ematically identical to the Fitness algorithm, the measure for Economic
Complexity we introduced in 2012. We demostrate that the mathematical
structure of ECI+ is strictly equivalent to that of Fitness (up to normal-
ization and rescaling). We then show how the claims of Albeaik et al.
about the ability of Fitness to describe the Economic Complexity of a
country are incorrect. Finally we hypothesize how the wrong results re-
ported by these authors could have been obtained by not iterating the
algorithm.

1 From the Fitness algorithm to ECI+ with a simple relabeling

Let us call X_{cp} the *extensive* export matrix giving, in a fixed year, the
export expressed in dollars of the product p by the country c. By its
definition we have that

$$X_c = \sum_p X_{cp}$$

is the total export of country c in that year. Analogously the quantity

$$X_p = \sum_c X_{cp}$$

gives the total amount (in dollars) of product p exported in the same year
by all countries. Finally, we call

$$X = \sum_{cp} X_{cp}$$

---

[*]luciano.pietronero@roma1.infn.it

1

the total world export of the considered year.

Let us now recall the fundamental equations of the algorithm [1, 2] from which it is possible to compute the Fitness of countries and the Complexity of products from the COMTRADE data of international export.

In this case the export matrix is binarized by using the Revealed Comparative Advantage (RCA) criterion. The RCA of a country c on a product p is defined as

$$RCA_{cp} = \frac{X_{cp}/X_c}{X_p/X}$$

which can be read as the ratio between the share of product p in the export basket of country c and the share of the same product in the total world export (or equivalently as the ratio between the share of the country c in the total export of product p and the share of the same country in the total world export). RCA_{cp} is generally considered to give a measure of how "good" is a country c in exporting (and therefore producing) a product p: if $RCA_{cp} > 1$ the country c is in average better than the rest of the world to export p. Consequently, the criterion to introduce the *binary* export matrix M_{cp} is simply the following: if $RCA_{cp} > 1$ then $M_{cp} = 1$, while if $RCA_{cp} \leq 1$ then $M_{cp} = 0$.

Through the matrix M_{cp} we can define the fitness-complexity algorithm respectively for the fitness F_c of countries and complexity Q_p of products as [1, 2]

$$
\begin{cases}
F_c^{(N+1)} = \sum_p M_{cp} Q_p^{(N)} \\
\\
Q_p^{(N+1)} = \frac{1}{\sum_c \frac{M_{cp}}{F_c^{(N)}}}
\end{cases}
\tag{1}
$$

with the condition of renormalizing at each step all F_c's and Q_p's by dividing at each iteration their values by the mean values respectively over all c and all p at the same iteration in order to avoid possible divergences due to the hyperbolic nature of the second equation.

We now show in few steps that ECI+ and PCI+ formulas defined in [3] can be simply seen as the version of Eqs. (1) where M_{cp} is simply substituted by the extensive matrix X_{cp} (change that was already discussed in [2]). First, let us substitute the second equation of (1) at iteration N in the first one at iteration $N + 1$:

$$F_c^{(N+1)} = \sum_p M_{cp} \frac{1}{\sum_{c'} \frac{M_{c'p}}{F_{c'}^{(N-1)}}}
\tag{2}$$

If we now substitute X_{cp} to M_{cp} and rename $F_c^{(2N)} = X_c^N$ we get exactly Eq. (10) in [3]:

$$X_c^N = \sum_p X_{cp} \frac{1}{\sum_{c'} \frac{X_{c'p}}{X_{c'}^{N-1}}}
\tag{3}$$

In order to rank countries the authors of [3] propose a measure of competitiveness of countries, called ECI+, which, a part from the subtraction of an iteration independent offset[1], is given by $\log X_c^\infty$ in strict analogy with

---

[1] This country-dependent offset $\log \frac{\sum_c X_{cp}}{X_p}$ can be seen as obtained by the same formula (3) for X_c^N with $X_c^{N-1} = 1$ for all c. Note that, differently from what written in [3], the

what was done for instance in [4, 5]. Analogously, if (i) we substitute the first equation of Eq. (1) at iteration N in the second one, (ii) substitute X_{cp} to M_{cp} in the same equation, and rename $1/Q_p^{(2N)} = X_p^N$ we get exactly Eq. (13) of [3]:

$$X_p^N = \sum_c X_{cp} \frac{1}{\sum_{p'} \frac{X_{cp'}}{X_p'^{N-1}}} \qquad (4)$$

The reciprocal algebraic relation between $Q_p^{(2N)}$ and X_p^N is recovered in the definition of the metrics called PCI+ in Eq. 16 in [3] as $-\log X_p^\infty = \log(1/X_p^\infty)$, apart from the addition of another iteration independent offset[2] $\log X_p$.

Similarly to the fitness-complexity algorithm, both X_c^N and X_p^N are renormalized at each iteration by dividing by an appropriate mean of their values respectively over all countries and all products in order to avoid divergences due to the non-linear hyperbolic nature of the algorithm. The authors chose for this purpose the geometric mean, probably taking into account the extensive nature of the matrix X_{cp}.

2 Problems and inconsistencies in fitness results

Given the equivalence of the algorithms, the claim reported in [3] that continuous data can be used in ECI+ but not in Fitness is indeed extravagant. In the second part of [3] it is argued that the same algorithm works well when it is named ECI+ but not when it is named Fitness.

The solution of the puzzle is in the different input data used. Clearly, the Fitness algorithm can be used with continuous, discrete, intensive, or extensive data, depending on the objective of the analysis, as already discussed in [2]. Albeaik et al. mix different input data (extensive for ECI+ and intensive for Fitness), and this is used as an erroneous evidence for an apparent difference in the algorithms. Moreover they state that Fitness is strongly correlated with diversity, as it is obvious due to the explicit sum over the products that a country exports. This sum however is weighted by the complexity of products, and this introduces residuals that are strongly informative. However the diversity term is important and cannot be disregarded, as it is a fundamental principle of Economic Complexity. Being defined in the same exact way, ECI+ also has a dependence on country size, which is trivially and explicitly removed by subtracting the term $\log(\sum_p \frac{X_{cp}}{X_p})$, which has a 0.97 correlation with X_c^∞. The Fitness measure as reported in [3] shows an anomalous ranking, in sharp contrast with the established literature [2]. In order to investigate this puzzle we reconstructed the input data used in that paper to the best

---

argument of the logarithm is not the "average share that the country represents in the export of a product", but the *sum* of the shares of all countries in the export of the product p.

[2]We do not understand why the authors of [3] in the definition $[\log X_p - \log X_p^\infty]$ of PCI+ have X_p^∞ which is adimensional while X_p is measured in dollars. Why does to this metrics change if we measure export in euros instead of dollars?

of our ability and ran both algorithms. The results are discussed in the following.

In [3] it is stated that "Fitness Complexity ranks many Southern European countries (such as Spain, Italy, and Portugal) at the top of the ranking, and also, provides very low complexity values for advanced East Asian and European economics, such as South Korea, Switzerland, Finland, Japan, and Singapore". First, it should be noted that country rankings obtained with Fitness have been published in [2], and ignored by Albeaik et al., who report very different results, obtained only by them to the best of our knowledge. Indeed, the real Fitness shows that the top 5 countries by Fitness in 2010 were Germany, China, Italy, Japan and United States; moreover, the 5 countries that are said to have "very low complexity values" are all in the top 20% of the ranking [2]. One should note that these datasets refer only to manufacturing, without taking into account services.

Moreover, Albeaik et al. state that "for the Fitness measure, the economy of Greece is ranked higher than that of Japan, Sweden, or China." This is again wrong and inconsistent with the results we published in [2], where Greece is ranked 34th, while Japan, Sweden, and China are 4th, 14th, and 2nd respectively. We point out that, given the strong weight the Fitness measure gives to diversification, it is really hard to believe a dataset exists such as China is ranked below Greece.

The strong differences between the rankings published in [2] and those reported in [3] can not be explained by a difference in the starting dataset alone. We base our following analysis on the BACI dataset [6], which we filter following the prescriptions given in [3]. On such dataset we recompute Fitness and ECI+ up to convergence, following the prescriptions given in [7] (we give the number of the performed iterations for reproducibility: 200). Curiously enough, the Fitness rankings reported in [3] are extremely similar to those that one would obtain on the same dataset by iterating the algorithm *just once*, which appears totally unreasonable. If one iterates the Fitness algorithm for 200 steps, the rankings appear much more reasonable and very different from those reported in the ECI+ paper, as well as coherent with those reported in [2].

We found also very strange that, in order to have Spain at the top of the Fitness ranking, as reported by Albeaik et al, after 1 iteration we have to set $F_c^0 = k_c^0$ and $Q_p^0 = k_p^0$ instead of the usual constant initial conditions used for Fitness ($F_c = 1 \forall c$ and $Q_p = 1 \forall p$). While the starting point of the iteration procedure becomes irrelevant when the algorithm is iterated up to convergence (and in this case Spain is never at the top), obviously it becomes more important if the number of iterations is reduced. In order to understand which misconceptions lead Albeaik et al. to such Fitness ranking, we tried to reproduce their results by visually comparing our computations of the Fitness and ECI+ with theirs. In particular, in fig. 1 we reproduced the original comparison between ECI+ and Fitness as presented in [3] and compare it with the one recomputed by us[3]. The upper figure is the original Fitness vs ECI+ graph taken from [3], the center

---

[3]In order to reproduce the figures in [3] we standardized the Fitness. We point out that the correct procedure is to take its logarithm.

figure is our best reproduction of that plot, that is done by performing only one (1) iteration of Fitness and with the initial conditions mentioned above. In the lower panel we present the comparison between ECI+ and the logarithm of the value of Fitness at convergence, obtained with 200 iterations. While the authors of [3] compute the logarithm of X after iterating, they omit to compute the logarithm of Fitness as it is usually done in the literature [1, 2, 4, 5] to compare it to other macroeconomic intensive indicators. One can easily realize that the best reproduction of the results presented by Albeaik et al. is obtained if the Fitness algorithm is iterated only once, which is clearly a mistake. On the contrary, if the same algorithm is iterated up to convergence, the two measures correlate more, given the mathematical equivalence of the algorithms. However, since the input matrices differ, some deviations are still present.

Conclusions

In summary, the paper of Albeaik et al. [3] does not introduce any new algorithm but just renames the Fitness one as "ECI+". In this respect one may also note that ECI+ has nothing to do with the original ECI [8]. A detailed discussion of the problems of ECI and the reasons to introduce the Fitness can be found in [2]. Also in that occasion the authors of [8] took inspiration from our work and learned, without ever citing, that the linear calculation of the ECI can be solved exactly by computing an eigenvector rather than with 18 iterations.

The numerical results reported in [3] for the Fitness are incorrect and even embarassing in view of the mathematical equivalence of ECI+ and Fitness. Albeaik et al. [3] present a totally distorted view of the situation, from both a mathematical and numerical point of view.

References

[1] A. Tacchella, M. Cristelli, G. Caldarelli, A. Gabrielli, L. Pietronero, Scientific Reports **2**, 723 (2012).

[2] M. Cristelli, A. Gabrielli, A. Tacchella, G. Caldarelli, L. Pietronero, PLoS ONE **8**, e7072 (2013).

[3] S. Albeaik, M. Kaltenberg, M. Alsaleh, C. A. Hidalgo, https://arxiv.org/abs/1707.05826v3 (2017).

[4] M. Cristelli, A. Tacchella, L. Pietronero, PLoS ONE **10**(2), e0117174 (2015).

[5] E. Pugliese, G.L. Chiarotti, A. Zaccaria, L. Pietronero, PloS ONE **12**(1), e0168540 (2017)

[6] G. Gaulier, and S. Zignago. "Baci: international trade database at the product-level (the 1994-2007 version)." (2010).

[7] E. Pugliese, A. Zaccaria, and L. Pietronero. The European Physical Journal Special Topics 225.10: 1893-1911. (2016)

[8] C.A. Hidalgo and R. Hausmann. PNAS 106.26: 10570-10575 (2009).
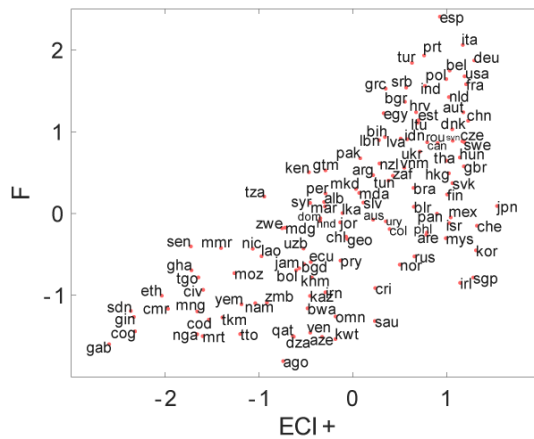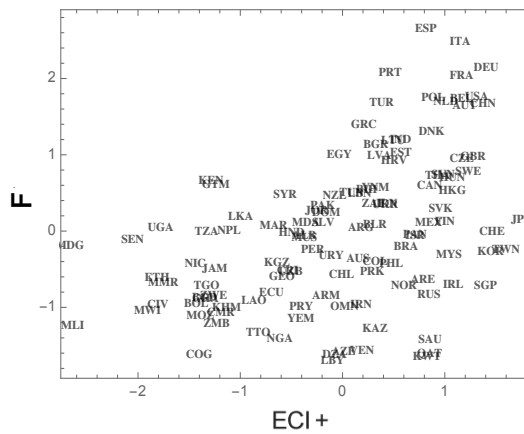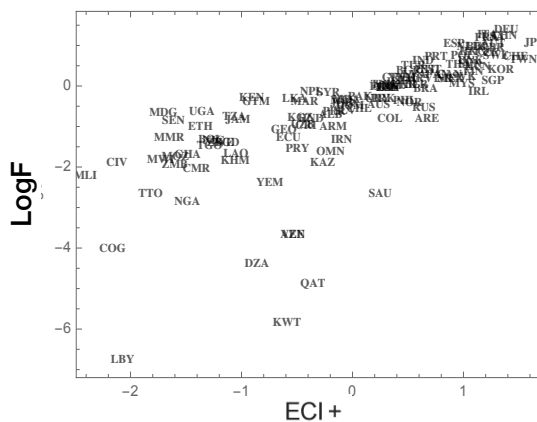
Figure 1: Upper figure: the original comparison between Fitness and ECI+, as presented in [3]. Central figure: our best attempt to reproduce the above results: the Fitness algorithm is iterated only once. Lower figure: when the Fitness algorithm is iterated up to convergence, the two measures became highly correlated, as one may expect from the identical mathematical structure of the algorithms. The residual differences are due to the different input matrix and the normalizations.