Supporting Information: A New Metrics for Countries' Competitiveness and Products' Complexity

Andrea Tacchella[1,2], Matthieu Cristelli[2,1], Guido Caldarelli[3,2,4], A. Gabrielli[2,3], Luciano Pietronero[1,2,4]

[1] *"Sapienza", Università di Roma, Dip. Fisica, P.le A. Moro 2, 00185, Roma, Italy*

[2] *ISC-CNR, Dip. Fisica "Sapienza", Università di Roma, P.le A. Moro 2*

[3] *IMT, Institute for Advanced Studies, Piazza S. Ponziano, 6, 55100 Lucca, Italy*

[4] *London Institute for Mathematical Sciences, South Street 22, Mayfair London, UK*

1 Conversion to HS2007 and improvement of the quality of data.

The strong non-linearity which characterizes our iterative method makes it sensible to certain types of errors which we have found to be present in the BACI dataset. This dataset has been built regarding at applications of a very different kind, mostly as a source for statistical analysis of trade. In this perspective, a missing or innacurate entry does not affect in any way the robustness of the information contained in the rest of the database. Conversely, for the purpose of our work even a single missing or inaccurate entry can represent a big issue, since this wrong information propagates non-linearly as the method is iterated. As reported in the documetation provided with to the BACI dataset[1], the raw data, consisting in the original reports of the custom offices of every considered country, have undergone a rigorous reconciliation procedure. Nevertheless the data is not yet sufficiently accurate for our purposes: an example should help clarify the point.

---

[1] Available at http://www.cepii.fr/anglaisgraph/bdd/baci.htm

Codes from 8004 to 8006: some products made of tin   Between 1995 and 2006 the total traded value for these 3 categories, as reported in the BACI database, oscillated between 80 and 150 millions of US dollars per year. In 2007-2008 this value went down to approximately 30 millions and finally fell to just 4 millions in 2009. There was no real reason for a sudden drop of the production of these items and, in fact, it didn't happen. But as a consequence of these erroneous reports, the Complexity of these products would greatly increase, without any real justification. In 2007 the World Custom Organization (WCO) released a new version of their Harmonized System categorization, the HS2007: in HS2007 the 3 codes 8004, 8005 and 8006 don't exist anymore as they are merged in the new 8007 code. Presumably, some countries started reporting these products correctly with the new code whereas some other continued using the old categorization. This clearly explains the sudden variation in the volume.

To handle situations like the one given as example we decided to convert the BACI dataset (wich is released following the HS1992 categorization) to the HS2007 categorization. Referring again to our example, all the volumes reported as 8004, 8005 and 8006 are simply added up in the new 8007 category, and the 3 old codes are cancelled.

In some cases an exact recasting of the data in HS2007 is not possible, due to the fact that a single code can be split into more than one new category, with unknown proportions. Such rare cases are treated one by one in the spirit of keeping the information as genuine as possible.

Some sudden jumps and intermittencies are still present after the conversion to HS2007. To further improve the quality of the data we use a simple regularization procedure which helps in

obtainig suitably continuous Fitness and Complexity values.

2   Statistical Validation of Results

One of the problems of the analysis described is that it is impossible a simple comparison or a statistical validation with an independent ranking of competitiveness. For instance, the metrics introduced in this paper cannot be compared to standard monetary indicator such as GDP. Indeed, one of our thesis is that the deviations with respect to standard monetary rankings are a key element which is informative on the potentiality of growth of a country. A widely know measure of competitiveness is the Global Competitiveness Index (GCI), issued by the World Economic Forum. Neither a comparison with this indicator would give a clear result because the GCI measures the perception of a country interviewing top managers to evaluate services, life quality, education quality, etc. By looking at the ranking of the GCI it is easy to realize that, in summary, the countries are ranked according to the quality of life: Switzerland, Scandinavian countries, etc. are in the very first positions. Our metrics does not measure the quality of life of countries and does not include services in its evaluation. It is instead aimed at assessing the present and future competitiveness of a country from its productive capacity of which export is assumed to be a proxy. In essence, our metrics is aimed to measure how good you are in the economic competition, to be compared with how much money you earn. While the two things are to some extent correlated (on average), it is also clear that (for instance) oil producers have a large wealth without being particularly fit in the economic competition. We expect such difference (when present) to be a key element in the prediction of future growth. Even a direct comparison with the GDP growth rate is rather com-

plicated. Comparison through standard regressions gives inconsistent results. The reason is that we expect a sort of selective predictability of our metrics, that is a heterogeneous dynamics of the growth of countries. A systematical economic validation of our fundamental analysis requires a new scheme of predictability in terms of the dynamics of the country competitiveness, in order to select those regime in which the fitness variable is the dominant driving force of the development of a country. For instance, even though the fitness of sub-Saharian countries can be measured, the political instabilities of this region, the famine and the persistent civil wars are clearly dominant factors which lock these countries in poverty traps, independently on their fitness level. This clarifies why a simple regression with the GDP growth rate leads to poor results. We are currently working on this systematic validation by defining suitable benchmark which will be presented in a future work.