An Updated Database of Revealed Factor Intensity Indices[[1]](#footnote-1), 1962-2007

This short note describes an updated version of Revealed Factor Intensity (RFI) Database for the period 1962-2007. The database contains three indices of Revealed Factor Intensity: Revealed Physical Capital Intensity Index, Revealed Human Capital Intensity Index and revealed Natural Resource Intensity Database. The database also includes the updated country endowment data (capital stocks per worker, average years of schooling and arable land per worker) for a cross section of 137 countries (countries that have all three endowments) for the period 1961-2007.

Using the latest updated versions of different databases we have re-constructed the country endowments' database and recalculated our database on RFI for the period from 1962 to 2007[[2]](#footnote-2). To update the country-endowment data we have used (a) the Penn World Table database - version 6.3[[3]](#footnote-3) (August 2009) for constructing capital stocks per worker, (b) Barro and Lee (2010) for yearly data on average years of schooling[[4]](#footnote-4) and (c) Arable Land per person was taken from WDI[[5]](#footnote-5). We have also used the updated version of World Bank data on agricultural distortions (version of April 2009) to correct for distortions in agricultural prices.

The dataset is available both in Stata and Excel format. For each format the dataset consists of four folders, titled (a) HS, (b) SITC, (c) 1994 and 2000 and (d) Endowment.

Folders - HS and SITC[[6]](#footnote-6)

The HS folder contains the RFI indices for products classified at the HS 6 digit level (version HS 88/92) and the year coverage is from 1988 to 2007. The SITC folder contains the RFI indices for products classified at the finest disaggregated level[[7]](#footnote-7) of the SITC (Revision 1) and the year coverage is from 1962 to 2007. Though the finest level of the SITC is a 5 digit, not every 4 digit is divided into five digits. Therefore our SITC database consist s of a mix of 4 and 5 digits in order to cover all the products.

In constructing the database, we faced a trade-off between “width” and “consistency” in country-endowment data. On the one hand, we are interested in having indices for as many countries as possible, to give a width to the database. On the other hand, to track the evolution of RFI indices for each good over several years, we need to have a complete (i.e. balanced) panel of data on endowments of a given set of countries for the same length of years, to ensure that the indices are constructed in comparable ways. However, if there is systematic bias in the selection of countries in the panel (say, if low-income countries are underrepresented in the data), RFI indices will be biased against factors of which low-income countries are poorly endowed. This may not necessarily alter the ranking of goods by RFI, but will affect the relative intensities. In order to minimize this bias, the wide-coverage (unbalanced) panel includes, each year, all the countries for which data are available in that year.

For those indices calculated from a panel or "balanced" dataset[[8]](#footnote-8), the filename is "balanced\_hs\_year\_indices" or "balanced\_sitc\_year\_indices". Table 1 (A and B) provide the description of the variables of the files in HS and SITC folders.

Table 1A. HS



Table 1B. SITC



Finally, in order to deal with the endogeneity problem that might arise from using the RFI indices in explaining trade flows, we have also calculated the Country Specific RFI indices for each individual country by excluding the country exports from the calculation. The Country Specific RFI indices are given in a separate folder, called "CountrySpecific"[[9]](#footnote-9).

Folder - 1994 and 2000

The RFI indices calculated using additional data on natural resources from the World Bank are given in a folder called "1994 and 2000". Though the database on natural resource from the World Bank covers only two years (1994 and 2000), it provides us with the most complete measure of natural resource endowments to date. Natural resource capital in the database consists of non-renewable resources subsoil assets (including oil, natural gas, coal, and mineral resources), cropland, pastureland, forested areas (including areas used for timber extraction and non-timber forest products), and protected areas. The list of the variables and their descriptions are given in Table 2 (A-B).

Table 2A. RFII database for 1994 and 2000, HS



Table 2B. RFII database for 1994 and 2000, SITC



Folder - Endowment

In addition, we have attached our newly constructed database on country endowment. (called "endowments61-07.dta" and "endowment1994\_2000.dta"; the same file name for Excel). The Table 3 (A-B) presents the list of variables and the description.

Table 3A. Endowment database for the period 1961-2007



Table 3A. Natural resource endowment database, 1994 and 2000



Reference:

**Alan Heston, Robert Summers and Bettina Aten, Penn World Table Version 6.3, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, August 2009.**

**Bolormaa Tumurchudur Klok, Miho Shirotori and Olivier Cadot, "Revealed Factor Intensity Indices at the Product Level", UNCTAD Policy Issues in International trade and Commodities Series No. 44, 2010.**

**Kym Anderson and Ernesto Valenzuela, "Estimates of Global Distortions to Agricultural Incentives, 1955 to 2007", World Bank, Washington DC, April 2009.**

**Robert Barro and Jong-Wha Lee, April 2010,** [**"A New Data Set of Educational Attainment in the World, 1950-2010."**](http://www.nber.org/papers/w15902) **NBER Working Paper No. 15902.**

1. For robustness check we have calculated the indices by excluding a country with the biggest share of export for each product and, on average, the correlation coefficients with the original indices were as high as 94% for RHCI indices, 93% for RCI indices and 80% for RNRI\_land at HS 6 digit level. (the same numbers were slightly lower for SITC classification). [↑](#footnote-ref-1)
2. For the details of the methodology of RFII construction, please see the paper on the UNCTAD website (http://siteresources.worldbank.org/INTEEI/214578-1110886258964/20748034/All.pdf) [↑](#footnote-ref-2)
3. The version 6.3 of PWT has 2005 as the base year and provides aggregate investment data for 189 countries for the period 1950-2007. [↑](#footnote-ref-3)
4. We have interpolated the five-yearly (quinquennial) estimates of average years of schooling from Barro & Lee (2010) to obtain yearly measures of human capital for 146 countries from 1950 to 2010. [↑](#footnote-ref-4)
5. Using Arable Land per person from the World Bank's World Development Indicators (for 205 countries for the period 1961-2007), we have calculated Arable Land per worker. [↑](#footnote-ref-5)
6. The RFI indices have been calculated by using two different classification schemes: Revision 1 of the United Nations Standard International Trade Classification (SITC 5-digit) and the Harmonized System (HS88/92 6-digit). Each has its own advantages and disadvantages. SITC provides longer years of trade statistics (since 1962) with fewer revisions than the HS, thus has the advantage of giving maximum comparability over the sample period. HS gives us a more disaggregated product classification, at the 6-digit level, than SITC. Whereas there are only over 1,000 products at the 4-5 digit products of the SITC classification, there are over 5,000 products at the HS 6-digit level. [↑](#footnote-ref-6)
7. [↑](#footnote-ref-7)
8. Our balanced database covers 94 countries for SITC (1962-2007) and 117 countries for HS (1988-2007) classification. [↑](#footnote-ref-8)
9. As the Country Specific RFI indices' files are considerably large, the files are only available in Stata format. [↑](#footnote-ref-9)