Open database for the paper "Participation in global value chains and varieties of development patterns"

Bruno Carballa Smichowski, Cédric Durand, Steven Knauss

Online appendix to the paper "Participation in global value chains and varieties of development patterns"

Bruno Carballa Smichowski – bcarballa@gmail.com
Cédric Durand – cdurand@ehess.fr
Steven Knauss – knauss.steve@gmail.com

This appendix provides further details concerning data treatment issues and additional statistical output that were unable to be included in the main paper due to space issues. It is divided into three parts. The first part refers to the construction of the raw indicators analyzed in section 3 of the main paper and with which the indexes used as variables for the PCA were built. Information about sources, data treatment and all the methodological choices made to build the database are detailed for every raw indicator. The second part offers more statistical output of the PCA that has not been included in the paper. The third section concludes the appendix by providing further information on the impact that our main methodological and theoretical changes had on the different results we obtain.

The underlying database we assembled in order to produce the statistical output in the paper is available upon request from the authors.

1. Raw indicators

1.1. Countries and time period of coverage

As the GVC participation and value capture variables discussed in Section 3 of the main paper are both indispensable to the analysis and dependent on the existence of value-added trade data, our data selection for all variables is necessarily limited to these 61 cases where such data are available from the OECDStat's Trade in Value Added (TiVA) database. As mentioned in Section 4.1 of the main paper, however, ten of these 61 cases

were eliminated from our data set, leaving us with 51 countries overall. Those ten countries – Brunei Darussalam, Bulgaria, Cyprus, Hong Kong, Latvia, Lithuania, Malta, Saudi Arabia, Singapore, and Taiwan – are nonetheless included in Figures 1 and 2 in Section 3 of the paper, as such figures do not require the use of additional variables obtained outside of the TiVA dataset, such as productive investment or social upgrading indicators, but excluded from the econometric analysis in Section 4.

As explained in Section 4.1, all but two of these ten countries were dropped due to the non-availability of data present across more than one indicator, rendering statistical estimation unreliable. The remaining two, Bulgaria and Latvia, were excluded for reasons explained in section 1.5 of the appendix below. The 51 countries retained in our econometric analysis are listed in full in the "Social Indicators Data Table" of section 1.6 of the appendix below.

As mentioned in section 4.1 of the paper, we had to restrict the end point of the time period to 2008 rather than the TiVA database's final year of 2011. Our reason for doing so can be seen in section 1.6 of the appendix below, as the majority of social indicators (median income, Gini index, and labor's share of income) pose significant data unreliability problems beyond the year 2008.

1.2. On the TiVA dataset

There are two drawbacks to be kept in mind about the TiVA dataset.

One potential bias is that transfer prices recorded for trade between foreign affiliates of the same firm may not accurately reflect arms-length market prices (Diewert, Alterman, & Eden, 2005). Of course, the same problem besets any statistical dataset concerning international trade, as statistical offices are only as good as fiscal authorities in accurately capturing such dynamics (Escaith, 2008, p. 25–26).

A second drawback of the TiVA dataset is that it has not yet, at the time of writing, incorporated the new changes brought about in the 5th revision of the classification by Broad Economic Category (BEC), approved in 2016 by the UN Statistical Commission.

These changes introduce a "specific" processed intermediate goods category in order to better disentangle the previous "processed intermediate goods" category that often included what are effectively unprocessed primary products that should not be treated as GVC related trade (UN Statistical Commission, 2016, p. 13). Therefore, there is considerable ground for future precision in measurement according to our conception of GVCs if the TiVA dataset or any other eventually provides value-added trade data based on the new categories of the 5th revision of the BEC. For the moment, of course, the OECD TiVA dataset remains the best practice available.

With this said, we will now discuss in turn, for each variable that entered the main paper's econometric analysis, additional data treatment or data sourcing issues not mentioned in the main text.

1.3. GVC participation

To provide more specific information about the GVC participation measure than offered in the paper, recall that the GVC participation rate is defined as the sum of the non-primary product portion of domestic value added in exports plus intermediate imports over GDP. The formula to determine the GVC Participation rate, as mentioned in sub section 3.1 of the paper, is:

$$\frac{(XDVA) * (1 - ppX) + ipM * (1 - ppM)}{GDP}$$

Where "XDVA" is domestic value added in gross exports, "ppX" is the share of primary products in total exports, "ipM" is gross imports of intermediate products and "ppM" the share of primary products in total imports

Domestic value added content of gross exports and gross imports of intermediate products are taken from OECDStat's Trade in Value Added (TiVA) database, October 2015 version, in US dollars. The shares of primary products in total exports and imports are taken from UNCTADStat (the SITC classification system, as explained in Section 4.1

of the main paper). Primary commodities, precious stones and non-monetary gold (SITC 0 + 1 + 2 + 3 + 4 + 68 + 667+ 971) as total volumes in US dollars of exports and imports for each country were made into shares by dividing by total exports and imports in US dollars from the same database. GDP figures are from the World Bank's "GDP at market prices (current US$)" variable in its World Development Indicators data set, and are given in current prices converted into US dollars through the exchange rate.

For this indicator there were no missing data points, nor was there a need for any additional treatment of the data, with the exception of the treatment of the outlier Cambodia, whose rate of change between 1995 and 2008 was more than 4 times higher than the second highest value in the entire data set. The percentage change between 1995 and 2008 for Cambodia was therefore replaced by the second highest variable in order to not overly skew the results.

1.4. Value capture

Recall that the value capture rate is defined as the non-primary product portion of the domestic value added content of exports over the non-primary product portion of total exports plus intermediate imports. The formula to determine value capture or the GVC gain rate, mentioned in subsection 3.2 of the paper, is:

$$\frac{(XDVA) * (1 - ppX)}{(XDVA) * (1 - ppX) + ipM * (1 - ppM)}$$

Where "XDVA" is domestic value added in gross exports, "ppX" is the share of primary products in total exports, "ipM" is gross imports of intermediate products and "ppM" the share of primary products in total imports

The sources for domestic value added content of gross exports, the share of primary products in total exports and imports, and gross imports of intermediate products are the same as used to construct the GVC participation indicator.

For this indicator there were no missing data points, nor was there a need for any additional treatment of the data.

1.5. Productive investment

As mentioned in Section 4.1 of the main paper, the investment rate measure used is "total investment (percent of GDP)" from the IMF's World Economic Outlook database, defined as "the total value of the gross fixed capital formation and changes in inventories and acquisitions less disposals of valuables for a unit or sector," all in current local currency.

The main limitation with the measure relates to the inclusion of residential investment which could make it difficult to distinguish real capital accumulation from real estate bubbles. Gross capital formation also does not take into account the original level of capital stock which can be assumed to vary enormously between countries. Both of these limitations are obviously to be kept in mind while interpreting the results.

For this indicator there were no missing data points, nor was there a need for any additional treatment of the data, with the exception of the treatment of the outliers Bulgaria and Latvia. Due to an unusually low starting point in 1995, which is highly likely to be an underestimate, Bulgaria would have had a rate of change more than four times the second highest value, Latvia, itself already 41.5 percent higher than any other value in the data set. After trying several methods to deal with these outlying values (replacing by the next highest value of the sample, replacing by the mean; eliminating them and estimating the missing values) that nonetheless continued to over-weight the role of the variable INVESTMENT_INDEX in the PCA and, especially, in the country-composition of the classes that emerged from the clustering based on the PCA, we decided to exclude both Bulgaria and Latvia from the sample.

1.6. Social Indicators

We will now discuss all additional issues for the four social variables described in section 4.1

Employment rate

The employment rate indicator was taken from the "Labor force participation rate, total (% of total population ages 15+)" data provided by the World Bank, defined as "Labor force participation rate is the proportion of the population ages 15 and older that is economically active: all people who supply labor for the production of goods and services during a specified period" (World Bank, 2016).

As mentioned in the main text (Section 4.1), this indicator was chosen instead of a simple unemployment rate figure because of the huge problem of informality in developing countries, making a cross-country comparison of unemployment rates of dubious value. To take an example, India has a mere 4.1 percent unemployment rate in the same World Bank database for 2008, and 4 percent for 1995, performing significantly above the average in both years as well as for the percentage change between the two years. Yet the LABORSTA (ILO) database has 84 percent of India's non-agricultural workforce in informal employment in 2009. While a reliable measure of total formal employment as a percentage of the working age population would therefore be ideal, the lack of sufficient data in the ILO database or elsewhere makes the World Bank's "labor force participation rate" the best available indicator (India scores, more realistically, significantly below average with this indicator).

There were no missing data points with this indicator, nor was there a need for any additional treatment of the data.

Median income

Median income was chosen as a more precise measure of the average individual's situation than average income, which is subject to relative variations in top incomes.

PovcalNet data came from the query of the database by Dykstra, Dykstra and Sandefur (2014) where "the population of each country is divided into 10,000 equal-size groups, where each group represents 0.01% of a country's population. The groups are ranked by per capita income or consumption" (Kochhar 2015). The median income score was taken as the 50th percentile of these 10,000 equal-size groups, which is not a precise median at the individual level but the closest thing available for many developing countries (Kochhar 2015). For greater comparability the Luxembourg Income Study data was treated in the same way, splitting the distribution into 10,000 equal size groups and taking the 50th percentile of these groups. Since the PovcalNet data is given at the level of individuals rather than households or equivalized scales, the Luxembourg Income Study data was also taken at the individual level by dividing each observation's disposable income total (which is at the household level) by the variable "number of household members" in order to transform the individual household observation into multiple numbers of individual observations (as many as exist in a given household) with the same individual income level (that of their household divided by its number of members). As with the methodology of the Pew Research Center, this method is not able to capture economies of scale inside households, and thus likely overestimates the real income gap between wealthy countries with smaller family units and poorer countries with larger ones (Kochhar 2015). This is the main data limitation for the median income variable, along with the potential incompability of taking consumption data in the countries where it was presented and income data in others (all LIS data is disposable household income data while Povcal is either income or consumption data depending on the survey year and country). Nonetheless, combining income and consumption data for cross country comparisons in this manner is common in the literature (Birdsall, 2010; Hellebrandt and Mauro 2015). Furthermore, since both of the listed limitations barely enter the picture when the rate of change inside one country is measured between two years, rather than the comparison between countries at a given year's values, these limitations are strongly mitigated in our analysis since the variable that eventually entered our principal component analysis is 50 percent determined by the rate of change in a given country between 1995 and 2008 (see Equation 3 in Section 4.2 of the paper).

In all cases for both the median incomes (and the gini indicators discussed below), data was taken as available as close as possible to the years 1995 and 2008, up to 3 years before or after the benchmark years (i.e. from 1992 to 1998 and from 2005 to 2010). For the median income, if the reported data came from one of the surrounding years that was not the benchmark year, it was extrapolated to the benchmark year following Kochhar's method of assuming an annual rate of change equal to 70 percent of the change in real household consumption expenditures, with the data taken from the World Bank's "household final consumption expenditure per capita growth (annual %)" variable in its Data Catalog.

All data was put in 2005 constant local currency prices in order to compare the rate of change between 1995 and 2008, and the 1995 values were put into 2011 international PPP dollars in order to have a comparable figure between countries for the 50 percent weight with beginning values that entered the principal component analysis. These conversions were done using the World Bank's International Comparison Program 2005 PPP to local currency convertors where applicable (that is, for all PovcalNet data since it is reported in 2005 PPPs) and the World Bank data catalog's "consumer price index (2010=100)" both to put non-PovcalNet data in 2005 local currency units and to bring all 1995 data to 2011 local currency units in order to use the 2011 International Comparison Program PPP (Table R3, "individual consumption expenditure by households") convertors to put the values into 2011 PPP dollars. Where data was missing from the World Bank's consumer price index, the IMF's World Economic Outlook Database (updated on January 19, 2016) was used for the same purpose (this was the case for Argentina, Chile, and China).

In addition to these adjustments, other adjustments that were necessary included multiplying the PovcalNet figures by 12 (they are reported as monthly estimates) to get a yearly estimate comparable with the Luxembourg Income Study data, and using Eurostat's "former euro area national currencies vs. euro/ECU – annual data" convertors in cases where the reported data of countries was done in the old national currency of countries who now use the Euro (and thus the 2011 PPP convertor is in Euros). This was the case for the 1995 values for Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, and Spain, in order to put them in 1995 local Euros

before bringing them to 2005 Euros to make the rate of change calculation. It would have been done for the 1995 value with Slovenia as well, but the result gave an unreliable estimate (a 30% decrease in median income between 1995 and 2008), so the rate of change was estimated instead (see below Table 1) while the Euro reported 2008 value was deemed reliable. Late Euro adopters Estonia and Slovakia were treated differently, since not only their 1995 values but also their 2008 values were reported in their pre-Euro local currency. Their 1995-2008 rates of change were therefore calculated by leaving both years in 2005 pre-Euro local currencies, while their 1995 beginning values were converted to Euros in order to make the 2011 PPP estimate.

Section 4.1 of the main paper states that if reported values for any countries were available for both 1995 and 2008 from both the Luxembourg Income Study and the PovcalNet database, the LIS data was chosen as more consistent with the rest of the dataset. The only exception here is Mexico, where there was a large difference between the numbers given by the LIS and the PovcalNet database concerning the rate of change between 1995 and 2008 – an astronomical 122 percent increase with LIS data versus a below the mean 25 percent increase with PovcalNet data. Given what is known about Mexican real income stagnation during this period (CONEVAL 2014: 51), the LIS data was deemed unreliable (in all likelihood the LIS 1995 figure is far too low, since the 2008 values themselves in 2011 PPP dollars are not very different, PovcalNet's are even a bit higher -- $2,556 PPP versus $2,492 PPP). The PovcalNet data were therefore used uniquely in Mexico's case when both PovcalNet and Luxembourg Income Study data were available.

Again, as stated in Section 4.1 of the main paper, where values were not available for both years in either the Luxembourg Income Study or the PovcalNet database, the best alternative estimate consistent between the two years was sought, and statistical estimation was used in the cases where nothing could be found (some combination of finding other sources or using statistical estimation by the NIPALS method occurred for only 9 out of 51 countries).

Table 1 below describes the data used for every country for both median incomes, gini values, and labor's share of the income, along with the years of the reported data (the

column is for the median income and gini variables, where there was more variation, while any variation from 1995 or 2008 for labor's share is noted inside the labor's share column itself) and whether there were any complications and, if so, how they were handled.

It should also be noted that median income was one variable influencing our need to restrict the end date of the period used in our overall analysis to 2008, rather than the TiVA database's end point of 2011, as we did not want to go too far away from having internally consistent data. Even with 2008 as an end point, there are 9 countries without data in either of our two combined sources, thus necessitating the use of other sources to find estimates, as mentioned in the previous paragraph. But there are a further 13 countries that do not pose data problems for 2008, but would, with a 2011 end point, necessitate either using 2008 or 2009 data or seeking 2011 data in other sources that would further weaken the internal consistency of the data.

Table 1 : Social indicators data table

Country	Country code	Median income source	Type of data / complication	GINI Source	Years of reported data	Labor's Share Source
Argentina	ARG	PovcalNet	Income	Same as median	1995, 2008	Trapp 2015 (2008 not available, 2007 used)
Australia	AUS	Luxembourg Income Study	Income	Same as median	1995, 2008	OECD Stat
Austria	AUT	UNU Wider – WIID 3.3	Income (Available for 1995 only in the LIS, but not taken since it is available for both years from the same source from the European Commission gathered by the WIID database, which makes the rate of change calculation more reliable)	Luxembourg Income Study is available for 1995 but not for 2008. For 2008, it is listed at 27.8 by two different data sets gathered by the "all the Ginis" (Milanovic) database, with a close number from yet another in 2005 and no far away values, so the 27.8 figure was taken.	1995, 2006 for median income. 1995, 2008 for GINI.	OECD Stat
Belgium	BEL	UNU Wider – WIID 3.3	Income (Available for 1995 only in the LIS, but not taken since it is available for both years from the same source from the European Commission gathered by the WIID database, which makes the rate of change calculation more reliable)	Luxembourg Income Study is available for 1995 but not for 2008. For 2008, it is listed at 28.5 by two different data sets gathered by the "all the Ginis" database, with a close number from yet another in 2005 and no far away values, so the 28.5 figure was taken.	1995, 2006 for median income. 1995, 2008 for GINI.	OECD Stat
Brazil	BRA	PovcalNet	Income	Same as median	1995, 2008	Trapp 2015
Cambodia	KHM	PovcalNet	Consumption	Same as median	1994, 2008	No data available
Canada	CAN	Luxembourg Income Study	Income	Same as median	1994, 2007	OECD Stat
Chile	CHL	PovcalNet	Income	Same as median	1994, 2009	Trapp 2015
China	CHN	PovcalNet	Consumption	Same as median	1996, 2008	Trapp 2015
Colombia	COL	PovcalNet	Income	Same as median	1996, 2008	Trapp 2015
Costa Rica	CRI	PovcalNet	Income	Same as median	1995, 2008	Trapp 2015
Croatia	HRV	PovcalNet	Consumption	Same as median	1998, 2008	Trapp 2015 (1995 and 2008 not available; 1996 and 2007 used)
Czech Republic	CZE	Luxembourg Income Study	Income	Same as median	1996, 2007	Trapp 2015 (2008 not available, 2005 used)
Denmark	DNK	Luxembourg Income Study	Income	Same as median	1995, 2007	OECD Stat
Estonia	EST	Luxembourg Income Study / PovcalNet	1995 value from PovcalNet, consumption; 2008 value from Luxembourg Income Study, income	Same as median	1995, 2007	Trapp 2015 (2008 not available, 2005 used)
Finland	FIN	Luxembourg Income Study	Income	Same as median	1995, 2007	OECD Stat
France	FRA	Luxembourg Income Study	Income	Same as median	1994, 2010	OECD Stat
Germany	DEU	Luxembourg Income Study	Income	Same as median	1994, 2007	OECD Stat
Greece	GRC	Luxembourg Income Study	Income	Same as median	1995, 2007	OECD Stat (1995 value is an estimated value by OECD Stat)
Hungary	HUN	Luxembourg Income Study	Income	Same as median	1994, 2007	Trapp 2015 (2008 not available, 2006 used)
Iceland	ISL	Luxembourg Income Study	Income (to calculate the final 2008 value. However, as no data is available for the years surrounding 1995, the rate of change between 1995 and 2008 needed to be estimated).	The final 2008 value could be calculated with the 2008 LIS data. However, due to the lack of a 1995 estimate anywhere, the rate of change between 1995 and 2008 was estimated.	2007	OECD Stat (1995 is an estimated value by OECD Stat)
India	IND	PovcalNet	Consumption	Same as median	1993, 2009	Trapp 2015
Indonesia	IDN	PovcalNet	Consumption	Same as median	1996, 2008	No data available
Italy	ITA	Luxembourg Income Study	Income	Same as median	1995, 2008	OECD Stat
Ireland	IRL	Luxembourg Income Study	Income	Same as median	1995, 2007	OECD Stat
Israel	ISR	Luxembourg Income Study	Income	Same as median	1997, 2007	OECD Stat
Japan	JPN	UNU Wider – WIID 3.3	Income. However, given that the data provided was at the household level without available information on the micro level regarding persons per household, the figure was divided by the 1995 average household size in Japan (NIPPSR 1998) and the 2010 average household size in Japan as a proxy for 2008 (Gu et al. 2015).	The GINI is available to calculate from LIS for 2008 only. For 1995, the "all the Ginis" database provided 4 sources with very close estimates for 1993 and one outlier for 1994. An average of the 1993 estimates was taken.	1995, 2006 for median income data. 1993, 2008 for the GINI calculation.	OECD Stat (1995 is an estimated value by OECD Stat)
Luxembourg	LUX	Luxembourg Income Study	Income	Same as median	1994, 2007	OECD Stat
Malaysia	MYS	PovcalNet	Income	Same as median	1995, 2007	No data available
Mexico	MEX	PovcalNet	Income	LIS data was available for both years, the only country with both years available in both the PovcalNet and LIS databases. LIS data was deemed reliable for the income dispersion to calculate GINIs. But PovcalNet was more reliable for the absolute levels of median income.	1994, 2007 for median income. 1994, 2008 for GINI.	Trapp 2015
Netherlands	NLD	Luxembourg Income Study	Income	Same as median	1993, 2007	OECD Stat

Country	Country code	Median income source	Type of data / complication	GINI Source	Years of reported data	Labor's Share Source
New Zealand	NZL	UNU Wider – WIID 3.3	Income (Not available near the years desired in either LIS or PovcalNet, data comes from OECDStat gathered by the WIID database). Due to its consistency it was used to calculate the rate of change. However, given that it was household reported income and not individual income, the 2008 income level was estimated in order to construct the composite variable of 2008 final value + rate of change between 1995 and 2008	Due to the consistency of the source reporting the GINIs for 1995 and 2008 in the WIID database, coupled with the larger variation around 1995 in the different sources reported in the "all the Ginis" database, the WIID estimates were taken for both 1995 and 2008.	1995, 2008	OECD Stat
Norway	NOR	Luxembourg Income Study	Income	Same as median	1995, 2007	OECD Stat
Philippines	PHL	PovcalNet	Consumption	Same as median	1994, 2009	Trapp 2015
Poland	POL	Luxembourg Income Study	Income	Same as median	1995, 2007	Trapp 2015
Portugal	PRT	UNU Wider – WIID 3.3	Income (Not available near the years desired in either LIS or PovcalNet, data comes from the European Commission gathered by the WIID database)	Due to the consistency of the source reporting the GINIs for 1995 and 2008 in the WIID database, coupled with the non-availability of data in the "all the Ginis" database, the WIID estimates were taken for both 1995 and 2008.	1995, 2006	OECD Stat
Romania	ROU	PovcalNet	Consumption	Same as median	1998, 2008	Trapp 2015
Russia	RUS	PovcalNet	Consumption	Same as median	1996, 2008	Trapp 2015
Slovakia	SVK	Luxembourg Income Study	Income	Same as median	1996, 2007	Trapp 2015 (2008 not available, 2006 used)
Slovenia	SVN	Luxembourg Income Study	Income (However, the rate of change was estimated, because there seems to have been a problem with the 1995 data for Slovenia in LIS, thus making a rate of change calculation difficult although the 2008 data is reliable for the final 2008 value).	Luxembourg Income Study (the problem with the 1995 data for Slovenia relates to the absolute magnitude of the 50th percentile value and not to the relative dispersal of income among the whole population, so it was still considered reliable to calculate the GINI.	1997, 2007	OECD Stat
South Africa	ZAF	PovcalNet	Consumption	Same as median	1995, 2008	Trapp 2015
South Korea	KOR	Data not available	Estimated	The GINI is available to calculate from LIS for 2008 only. For 1995, the "all the Ginis" database provides two estimates with the same figure for 1998, one of which also provides a 1993 estimate. The 1993 estimate was therefore taken.	1993, 2006 for GINI.	OECD Stat
Spain	ESP	Luxembourg Income Study	Income	Same as median	1995, 2007	OECD Stat
Sweden	SWE	Luxembourg Income Study	Income	Same as median	1995, 2005	OECD Stat
Switzerland	CHE	Report – "L'évolution des inégalités de revenus en Suisse"	Income (Available for 1995 only from the LIS, but not taken since it is available for both years from the same source as an individualized income through the household equivalence scale, the best available data to calculate an internally consistent rate of change)	Luxembourg Income Study is available for 1995 but not for 2008. For 2008, it is listed at 32.3 by two different data sets gathered by the "all the Ginis" database, without any far away values from other data sets, so the 32.3 figure was taken.	1998, 2006 for median income. 1992, 2008 for GINI.	OECD Stat
Thailand	THA	PovcalNet	Consumption	Same as median	1994, 2008	Trapp 2015
Tunisia	TUN	PovcalNet	Consumption	Same as median	1995, 2010	Trapp 2015
Turkey	TUR	PovcalNet	Consumption	Same as median	1994, 2008	Trapp 2015 (2008 not available, 2006 used)
United Kingdom	GBR	Luxembourg Income Study	Income	Same as median	1995, 2007	OECD Stat
United States	USA	Luxembourg Income Study	Income	Same as median	1994, 2007	OECD Stat
Vietnam	VNM	PovcalNet	Consumption	Same as median	1998, 2008	No data available

Gini coefficient

As with median income, there are a few additional details that can be mentioned for the treatment and sourcing of the gini variable.

In terms of the calculations to get the gini coefficients, these were in the vast majority of cases (see Table 1 above) computed with the same income distribution data that provided the median income values (principally the Luxembourg Income Study and PovcalNet data – thus the gini variable posed a similar weight as median income in the need to restrict the end point to 2008).

Disposable household income presented in the LIS was divided by the number of household members in order to use "per capita" data, consistent with the way the data from PovcalNet is presented. As mentioned in Section 4.1 of the main paper, all ginis were double checked against all sources with data available for the given years in the "All the Ginis" database. The "All the Ginis" dataset lists all ginis available from any of eight original sources for all countries for all years from 1950 to 2012. The problem is one of direct comparability between the sources, with no single source providing estimates for the desired years for most countries. The direct calculations from the Luxembourg Income Study and PovcalNet were therefore more desirable for the majority of the countries (see above Table 1). The method of double checking the LIS and PovcalNet calculated ginis against the "All the Ginis" dataset was to compare the calculated value with all values given within 3 years of the benchmark year from any source reporting in the "All the Ginis" dataset. The direct calculations used from the LIS and PovcalNet data were in all cases within a 20 percent variation range of the majority of available estimates within three years of the benchmark year in the "All the Ginis" database, with the exception of Russia and Romania for 1995, although both of the latter corresponded to the PovcalNet estimate reported in the "All the Ginis" database and thus were kept.

For the eight cases where LIS or PovcalNet data were not available for a gini calculation for either one year or both (see above Table 1), the most consistent number appearing closest to the benchmark years from the "All the Ginis" was taken or, in the cases of New Zealand and Portugal, where there was no consistent figure, the UNU Wider WIID 3.3

database estimate was taken which corresponded to the median income figure and was thus internally consistent. As can be seen in the above Table 1, Iceland needed to be estimated for 1995 to calculate the rate of change, since none of the above options provided a reliable figure.

Labor's share of income

The above Table 1 lists the source for all countries, in addition to mentioning if a surrounding year was used rather than the benchmark year in cases of incomplete data, and the three cases where the 1995 value from OECDStat was an estimation on their behalf (all in the labor's share source column).

As mentioned in Section 4.1 of the main paper, Van Treeck's data were taken in the eight cases where data were available in both data sets used for the labor share indicator: this was the case for Czech Republic, Estonia, Hungary, Mexico, Poland, Romania, Slovakia, Turkey. Of these eight cases, there was only one case where the 1995 values from the different data sets differed by more than 20 percent, and only 2 cases out of 8 where this was the case for the 2008 values. In the case of four countries (Cambodia, Indonesia, Malaysia, Vietnam) where data was not available in either data set, the values were estimated using the NIPALS method.

Labor's share is another variable leading us to restrict the end year of the time period to 2008, as in addition to the 4 countries needing to be estimated in 2008, a 2011 end point would lead to using 2008 or pre-2008 values for eight other countries, or searching for 2011 values for such countries from other data sources, further weakening internal consistency.

2. Principal component analysis

In this second part of the annex we present some statistical output of the principal component analysis that have been excluded from the paper and can interest some

readers that would like to enter further into the details. When pertinent to the appraisal of the methodological choices explained in the paper, information regarding the axis that has not been retained (axis F4) has been included. This additional output can be seen in Tables 2 through 7 below.

We retained three axes for the following three reasons. First, following the Kaiser criterion (Kaiser, 1960), axes F1, F2 and F3 were the ones with eigenvalues above 1. Second, while the decrease in eigenvalues from F1 to F3 where smooth, the decrease from axis F3's eigenvalue to axis F4's was sharp, which indicates that axes F1 to F3 should be retained (see the appendix). Third, taking axes F1, F2 and F3 provided significant correlations (over 0.76) between axes and at least one of the variables. Had we taken only two, the variable SOCIAL_INDEX, would have been virtually uncorrelated with the axes while it is very strongly correlated (0.97) with axis F3.

No rotations were applied because the information concentrated in the first three axes showed virtually no increase when rotations where tested.

Regarding the clustering, we tested two methods. The first one is the agglomerative hierarchical clustering (AHC) method, which, when applied to the coordinates of each observation in the three axes retained, shows the number and (country) composition of classes for which classes can be considered homogeneous. The result obtained was that three classes was the best choice. We then tested the k-mean clustering method for purpose of robustness: That is, we performed a cluster analysis using the k-means method on the coordinates of the observations in axes F1, F2 and F3 of the PCA with an open range of classes from 1 to 5. Contrary to the AHC method, the k-mean clustering method implies choosing the number of classes beforehand to obtain their (country) composition. After having tried several number of classes, the conclusion was also that 3 classes is the most solid choice.

Table 8 shows the country composition of each class along with the number of countries in each, the sum of weights, within-class variance, and minimum, average and maximum distance to centroid for each class.

Table 2: Eigenvalues, variability and cumulative variability for axes F1 to F4

	F1	F2	F3	F4
Eigenvalue	1,225	1,097	1,004	0,674
Variability (%)	30,626	27,427	25,088	16,859
Cumulative %	30,626	58,053	83,141	100,000

Table 3: Eigenvectors for each variable for axes F1 to F4

	F1	F2	F3	F4
PART_INDEX	0,689	0,347	0,203	-0,603
VALCAPT_INDEX	0,023	0,867	-0,150	0,474
INVESTMENT_INDEX	0,705	-0,349	0,037	0,616
SOCIAL_INDEX	-0,168	0,075	0,967	0,177

Table 4: Squared cosines of the variables for axes F1 to F4

	F1	F2	F3	F4
PART_INDEX	**0,581**	0,132	0,041	0,246
VALCAPT_INDEX	0,001	**0,825**	0,023	0,152
INVESTMENT_INDEX	**0,609**	0,134	0,001	0,256
SOCIAL_INDEX	0,035	0,006	**0,938**	0,021

Values in bold correspond for each variable to the factor for which the squared cosine is the largest

Table 5: Factor scores of each observation for axes F1 to F3

Observation	F1	F2	F3
ARG	-0,033	-0,562	-0,428
AUS	-1,148	-2,651	0,300
AUT	-0,416	1,263	0,818
BEL	-0,096	0,020	0,873
BRA	-0,867	0,046	0,006
CAN	-1,045	-0,584	0,355
CHE	-0,046	1,914	1,046
CHL	-0,541	-1,800	-0,306
CHN	2,477	0,573	-0,763

COL	-1,084	-0,096	-1,844
CRI	1,245	0,770	-0,661
CZE	1,342	-0,278	0,025
DEU	-0,437	1,942	0,203
DNK	-0,340	-0,129	0,992
ESP	-0,036	-0,534	0,838
EST	0,603	0,279	-0,093
FIN	0,086	0,255	0,457
FRA	-0,704	0,367	-0,066
GBR	-2,051	1,132	0,360
GRC	-0,901	-1,134	0,179
HRV	1,565	-0,790	-0,871
HUN	0,752	-0,090	0,305
IDN	-0,082	-1,583	-0,388
IND	1,782	-0,302	-1,551
IRL	0,707	0,687	2,251
ISL	-0,247	-1,036	2,856
ISR	-1,295	1,526	-0,925
ITA	-1,009	0,531	-0,183
JPN	-0,061	2,160	-1,243
KHM	-0,150	2,064	0,317
KOR	1,325	-0,189	0,395
LUX	1,648	0,404	1,069
MEX	-0,056	-0,628	-1,726
MYS	1,853	-0,324	-0,131
NLD	-1,539	1,512	0,810
NOR	-1,189	-1,617	0,505
NZL	-1,506	-0,958	0,801
PHL	-1,114	1,257	-1,490
POL	0,797	0,070	-0,964
PRT	-0,940	-0,254	0,271
ROU	0,716	-1,588	-0,374
RUS	-1,233	-1,932	0,088
SVK	1,415	-0,107	0,534
SVN	1,161	0,211	0,385
SWE	-0,292	0,768	0,723
THA	1,605	-0,210	0,334
TUN	0,176	0,270	-0,951
TUR	-1,047	-0,138	-1,571
USA	-1,741	0,193	-0,102
VNM	1,891	-0,209	1,234
ZAF	0,101	-0,491	-2,697

Table 6: Contributions of the observations to axes F1 to F3 (in percentage points)

	F1	F2	F3
ARG	0,002	0,565	0,358
AUS	2,111	12,560	0,176
AUT	0,277	2,853	1,307
BEL	0,015	0,001	1,489
BRA	1,202	0,004	0,000
CAN	1,747	0,609	0,246
CHE	0,003	6,550	2,137
CHL	0,468	5,792	0,183
CHN	9,817	0,587	1,137
COL	1,881	0,016	6,643
CRI	2,482	1,060	0,853
CZE	2,882	0,138	0,001
DEU	0,306	6,744	0,081
DNK	0,185	0,030	1,922
ESP	0,002	0,509	1,371
EST	0,581	0,139	0,017
FIN	0,012	0,116	0,409
FRA	0,793	0,241	0,008
GBR	6,732	2,291	0,253
GRC	1,300	2,300	0,063
HRV	3,918	1,115	1,483
HUN	0,906	0,014	0,182
IDN	0,011	4,479	0,294
IND	5,083	0,162	4,701
IRL	0,800	0,844	9,899
ISL	0,098	1,919	15,939
ISR	2,683	4,162	1,673
ITA	1,630	0,503	0,066
JPN	0,006	8,340	3,020
KHM	0,036	7,611	0,196
KOR	2,810	0,064	0,305
LUX	4,347	0,291	2,232
MEX	0,005	0,706	5,821
MYS	5,497	0,188	0,034
NLD	3,790	4,084	1,281
NOR	2,262	4,674	0,499
NZL	3,630	1,640	1,252
PHL	1,985	2,822	4,339
POL	1,017	0,009	1,814
PRT	1,415	0,116	0,143
ROU	0,821	4,504	0,273
RUS	2,435	6,674	0,015
SVK	3,205	0,021	0,558
SVN	2,157	0,080	0,289
SWE	0,136	1,054	1,022

THA	4,123	0,079	0,217
TUN	0,050	0,130	1,766
TUR	1,754	0,034	4,825
USA	4,852	0,066	0,020
VNM	5,722	0,078	2,973
ZAF	0,016	0,432	14,214

Table 7: Squared cosines of the observations for axes F1 to F4

	F1	F2	F3	F4
ARG	0,001	0,146	0,085	**0,768**
AUS	0,153	**0,816**	0,010	0,021
AUT	0,068	**0,627**	0,263	0,042
BEL	0,012	0,001	**0,986**	0,002
BRA	**0,830**	**0,002**	**0,000**	0,168
CAN	**0,622**	0,194	0,072	0,112
CHE	0,000	**0,728**	0,217	0,055
CHL	0,070	**0,780**	0,022	0,127
CHN	**0,563**	0,030	0,053	0,354
COL	0,255	0,002	**0,737**	0,006
CRI	**0,525**	0,201	0,148	0,127
CZE	**0,906**	0,039	0,000	0,055
DEU	0,045	**0,893**	0,010	0,052
DNK	0,094	0,014	**0,796**	0,096
ESP	0,001	0,147	0,363	**0,488**
EST	**0,497**	0,107	0,012	0,385
FIN	0,015	0,130	0,420	**0,435**
FRA	**0,581**	0,158	0,005	0,256
GBR	**0,748**	0,228	0,023	0,001
GRC	0,370	**0,586**	0,015	0,029
HRV	0,355	0,091	0,110	**0,444**
HUN	0,175	0,002	0,029	**0,794**
IDN	0,002	**0,696**	0,042	0,261
IND	**0,551**	0,016	0,418	0,015
IRL	0,079	0,075	**0,801**	0,045
ISL	0,006	0,112	**0,850**	0,032
ISR	0,345	**0,479**	0,176	0,000
ITA	**0,762**	0,211	0,025	0,002
JPN	0,001	**0,749**	0,248	0,003
KHM	0,005	**0,964**	0,023	0,008
KOR	**0,729**	0,015	0,065	0,191
LUX	0,196	0,012	0,082	**0,710**
MEX	0,001	0,112	**0,844**	0,044
MYS	**0,795**	0,024	0,004	0,176
NLD	**0,358**	0,346	0,099	0,197

NOR	0,328	**0,607**	0,059	0,006
NZL	**0,591**	0,239	0,167	0,003
PHL	0,238	0,303	**0,426**	0,033
POL	0,283	0,002	**0,413**	0,302
PRT	**0,831**	0,061	0,069	0,040
ROU	0,151	**0,740**	0,041	0,068
RUS	0,267	**0,655**	0,001	0,077
SVK	**0,541**	0,003	0,077	0,379
SVN	**0,677**	0,022	0,074	0,226
SWE	0,066	**0,456**	0,405	0,073
THA	**0,942**	0,016	0,041	0,001
TUN	0,029	0,067	**0,832**	0,073
TUR	0,285	0,005	**0,642**	0,068
USA	**0,979**	0,012	0,003	0,005
VNM	**0,661**	0,008	0,281	0,049
ZAF	0,001	0,029	**0,867**	0,103

Values in bold correspond for each observation to the factor for which the squared cosine is the largest

Table 8: Country composition of the classes found using the k-clustering and statistical results by class (World Bank country abbreviations)

Class	1	2	3
Objects	14	16	21
Sum of weights	14	16	21
Within-class variance	1,559	1,886	1,995
Minimum distance to centroid	0,402	0,517	0,371
Average distance to centroid	1,109	1,272	1,225
Maximum distance to centroid	2,398	2,006	2,740
	Argentina	Austria	China
	Australia	Brazil	Costa Rica
	Belgium	Switzerland	Czech Republic
	Canada	Colombia	Estonia
	Chile	Germany	Finland
	Denmark	France	Croatia
	Spain	United Kingdom	Hungary
	Greece	Israel	India
	Indonesia	Italy	Ireland
	Iceland	Japan	South Korea
	Norway	Cambodia	Luxembourg
	New Zealand	Netherlands	Mexico
	Portugal	Philippines	Malaysia
	Russian Federation	Sweden	Poland
		Turkey	Romania
		United States	Slovakia
			Slovenia
			Thailand
			Tunisia
			Viet Nam
			South Africa

3. Measuring the impact of author's methodological choices in GVC participation and value capture indicators

In Section 3 of the paper we proposed new indicators to measure GVC participation and value capture. Each of these indicators presents a series of methodological differences in their construction in respect to traditional ones. The objective of this Annex is to measure, for each indicator, the impact each methodological change had on the different results we obtain when compared to those of traditional indicators.

We will measure this impact in three ways. Firstly, we will look at how much each methodological change affected the country ranking of GVC participation and value capture respectively. Secondly, we will calculate how each methodological change separately affected the correlation between the author's indicators and the traditional ones. Finally, we will do the PCA and cluster analyses we had carried on in Section 4 of the paper using the traditional GVC participation and value capture indicators in order to show the impact our methodological innovations in calculating these two indicators affect the conclusions that can be reached regarding the link between GVC participation and developmental outcomes.

3.1 GVC participation

As shown in Table 1 of the main paper, the author's GVC participation indicator presents three methodological differences compared to the traditional indicator:

1. While the traditional indicator includes primary commodities, the author's does not

2. The denominator of the traditional indicator is gross exports, while the denominator of the author's indicator is GDP

3. The traditional indicator applies the 'two borders rule', while the author's does not.

Difference number 3 means that, when calculating the traditional GVC participation indicator, only if a good has been exported from a country A to a country B and then to a third country C (i.e. only when the good has crossed two borders) that trade is considered to be GVC-related and therefore measured in the traditional GVC participation indicator. In terms of measurement, this implies that in order to measure a country's GVC participation, the numerator of the indicator considers only the imports of re-exported intermediate inputs (VS) and the exports of intermediate inputs that are re-exported by the importer (VS1). For the reasons detailed in subsection 3.1 of the main paper, the author's indicator does not follow the two borders rule. Therefore, in order to measure the GVC participation of a given country, it considers the imports of all intermediate inputs regardless of the fact that they are re-exported or not, while it

excludes the imports of finished products. Regarding exports, it includes both the exports of all intermediate inputs and final products.

In order to measure the impact each of these changes had on the country ranking of GVC participation in respect to the traditional indicator, we calculated this ranking for a series of 'intermediate indicators' in which we introduce only one of the methodological changes into the traditional indicator at a time. Then we measure the average absolute difference in terms of rankings with the traditional indicator in order to quantify the impact each of the three above-mentioned methodological changes had in terms of country rankings.

Another way to measure the impact each methodological change had in respect to the traditional GVC participation indicator is to calculate the Pearson correlation coefficient between each intermediate indicator and the traditional indicator. Lower correlations would indicate a larger departure from the results expected from the traditional indicator, which translates into a higher impact of the methodological change in question.

The results are presented in Tables 9 and 10.

Table 9 : Absolute difference in country rankings between author's GVC participation indicator, intermediate indicators and the traditional GVC participation indicator for 1995 and 2011

Year	Author's GVC participation indicator	Standard GVC participation indicator without commodities	Standard GVC participation indicator with GDP in the denominator	Standard GVC participation indicator without the two borders rule
1995	7.8	9.3	7.1	17.2
2011	9.2	9.9	7.0	16.9

NB: All comparison are made in respect to the traditional GVC participation indicator

Table 9 shows that, among the three methodological differences the author's GVC participation presents in respect to the traditional one, the non-inclusion of the two borders rule is the one that has the higher impact regarding the changes in the ranking. In effect, when the author's indicators is used in 1995, a country shifts in average 7.8 positions in the country ranking of GVC participation in comparison to the ranking that would be obtained using the traditional indicator. If only the two borders rule was lifted from the traditional indicator, each country would switch in average 17.2 positions in 1995.

Table 10 : Pearson correlation coefficient between author's value capture indicator with and without commodities and the traditional value capture indicator for 1995 and 2011

Year	Author's GVC participation indicator	Standard GVC participation indicator without commodities	Standard participation indicator with GDP in the denominator	Standard GVC participation indicator without the two borders rule

1995	0.81	0.58	0.83	0.12
2011	0.72	0.53	0.77	0.21

NB: All correlations are calculated with the traditional GVC participation indicator

The results of Table 10 go in the same direction as those of Table 9. When only the two borders rule is lifted from the traditional indicator, its correlation with the traditional indicator drops to 0.12 in 1995 and 0.21 in 2011. Moreover, when only commodities are removed from the traditional indicator the correlation with the latter drops to 0.58 and 0.53 in 1995 and 2011 respectively, while in other cases they remain above 0.7. This shows that the elimination of the two border rules is the methodological change that has the higher impact in changing the results obtained with the traditional GVC participation indicator both in terms of country ranking and correlation.

3.2 Value capture

As explained in subsection 3.2 of the paper, two methodological changes were introduced in the author's value capture indicator in respect to the traditional one (domestic value added content of gross exports), namely:

1. While the traditional value capture indicator considers commodities, the authors' does not

2. While the traditional value capture indicator has gross exports in the denominator, the author's has what we consider 'GVC related trade': domestic value added in exports and imports of all intermediate inputs.

In order to calculate the impact the change of the denominator we introduced (difference number 2) had in terms of country ranking and correlation with the traditional indicator, we calculate an intermediate indicator in which only the denominator is changed in respect to the traditional one. The same could not be done regarding difference number 1 because, in order to exclude primary commodities from both the numerator and the denominator of the original indicator, we would have to multiply both by the share of non-primary-commodities in exports of the country, which would result in the indicator remaining unchanged. Therefore, the separate effect of having removed primary commodities from the traditional value capture indicator is to be assessed by comparing overall changes between the traditional and the author's indicator and change between the traditional and the intermediate value capture indicator mentioned above.

The results of the same calculations done in subsection 3.1 of this Appendix are presented in Tables 11 and 12 regarding value capture indicators.

Table 11: Absolute difference in country rankings between author's value capture indicator with and without commodities and the traditional value capture indicator for 1995 and 2011

Year	Traditional value capture indicator with author's GVC-related trade in the denominator	Author's value capture indicator (excluding commodities)
1995	10.4	21.1
2011	6.8	20.6

NB: comparison are made with the traditional value capture indicator: domestic value added content of gross exports

Table 11 shows that the difference in terms of country ranking when the author's indicator is used is considerable. In average, countries switch 21.1 positions in 1995 and 20.6 positions in 2011 in comparison with the positions they would occupy in the ranking if the traditional value capture indicator was used. This difference drops to 10.4 and 6.8 for 1995 and 2011 respectively if only the change of the denominator is introduced. This shows that both of the changes introduced in respect to the traditional value capture indicator are relevant to explain the shifts in country rankings, although the relevance of having excluded commodities seems to grow over time, which is consistent with the timing of the commodity boom.

Table 12: Pearson correlation coefficient between author's value capture indicator, the traditional value capture indicator with author's GVC-related trade in the denominator and the traditional value capture indicator for 1995 and 2011

Year	Traditional value capture indicator with author's GVC-related trade in the denominator	Author's value capture indicator (excluding primary commodities)
1995	0.73	-0.09
2011	0.82	-0.18

When analyzing the impact of the two above-mentioned methodological changes in respect to the traditional value capture indicator in terms of correlation with the latter, the exclusion of primary commodities appears as having an enormous impact. Indeed, when only the denominator is changed from gross exports to author's GVC-related trade, the correlation of that indicator and the traditional value capture indicator remains high (0.73 and 0.82 in 1995 and 2011, respectively). On the contrary, if the author's indicator is used, the correlation is insignificantly negative for both 1995 and 2011. Therefore, of the two changes introduced, the exclusion of primary commodities is the one that had the higher impact.

3.3 PCA and cluster analysis using traditional indicators of GVC participation and value capture

We have seen in subsections 3.1 and 3.2 of the main paper that using authors' GVC participation and value capture indicators instead of the traditional ones alters considerably country rankings. In subsections 3.1 and 3.2 of this appendix we have

analyzed, for each of the authors' indicators, the separate effect each methodological change in respect to the traditional indicator had in terms of country ranking and correlation with the traditional indicator. Nevertheless, one could think that, although there are considerable variations in terms of country ranking when the authors' indicators are used, the effect of using the authors' indicators instead of the traditional ones would be minor in the PCA and cluster analysis carried on in the article, since correlations between authors' and the traditional GVC participation and value capture indicator remain reasonably high.

In order to see if the use of alternative GVC participation and value capture indicators had an important effect on the PCA and the cluster analysis carried on in the article, we present in this subsection of the Appendix the results of the same analyses using the traditional GVC participation and value capture indicators, while keeping the other variables and parameters intact.

3.3.1 PCA using traditional GVC participation and value capture indicators

In this subsection we will compare the factor loadings of each variable on axes F1, F2 and F3 and the square cosines of each variable in each axes when traditional and authors' GVC participation and value capture are used.

Table 13: Factor loadings of each variable for axes F1, F2 and F3 using authors' GVC participation and value capture indicators

	F1	F2	F3
PART_INDEX	**0.76**	0.36	0.20
VALCAPT_INDEX	0.03	**0.91**	-0.15
INVESTMENT_INDEX	**0.78**	-0.37	0.04
SOCIAL_INDEX	-0.19	0.08	**0.97**

NB: the highest factor loading of each variable is presented in bold

Table 14: Factor loadings of each variable for axes F1, F2 and F3 traditional GVC participation and value capture indicators

	F1	F2	F3
PART_INDEX	0.534	**0.590**	0.273
VALCAPT_INDEX	0.424	-0.516	**0.719**
INVESTMENT_INDEX	-0.425	**0.637**	0.466
SOCIAL_INDEX	**0.710**	0.246	-0.356

NB: the highest factor loading of each variable is presented in bold

When comparing tables 13 and 14 we can see that in both PCAs the association between variables remains the same: GVC participation and investment seem to be correlated, while value capture and social index seem to be independent, each being located at the

extremes of separate axes. Nevertheless, it should be noted that correlations with the axes are in all cases higher when authors' indicators are used, which indicates that the conclusions to be drawn from that PCA are more trustworthy than those obtained with a PCA that uses traditional GVC and value capture indicators.

Table 15: Squared cosines of the variables for axes F1, F2 and F3 using authors' GVC participation and value capture indicators

	F1	F2	F3
PART_INDEX	**0.581**	0.132	0.041
VALCAPT_INDEX	0.001	**0.825**	0.023
INVESTMENT_INDEX	**0.609**	0.134	0.001
SOCIAL_INDEX	0.035	0.006	**0.938**

Table 16: Squared cosines of the variables for axes F1, F2 and F3 using traditional GVC participation and value capture indicators

	F1	F2	F3
PART_INDEX	0.286	**0.348**	0.075
VALCAPT_INDEX	0.180	0.266	**0.517**
INVESTMENT_INDEX	0.181	**0.406**	0.217
SOCIAL_INDEX	**0.503**	0.061	0.127

When we examine the squared cosines of the variables for axes F1, F2 and F3 for both PCAs, the differences between the two widen. In the PCA that uses authors' GVC participation indicator, the share of information of each variable contained in the axis to which it is associated is considerable, never below 58% (cf. Table 13). On the contrary, in the PCA that uses traditional GVC and value capture indicators that share of information never goes *beyond* 52%. Therefore, the conclusions to be drawn from the PCA that uses the traditional GVC participation and value capture are to be interpreted with cautiousness, while those of the authors' PCA are robust.

3.3.2 Cluster analysis using traditional GVC participation and value capture indicators

We now compare the country composition of the three classes that arise from the k-means cluster analyses in both PCAs.

Table 17: Comparison of country composition of the classes using the author's and the standard GVC participation and value capture indicators

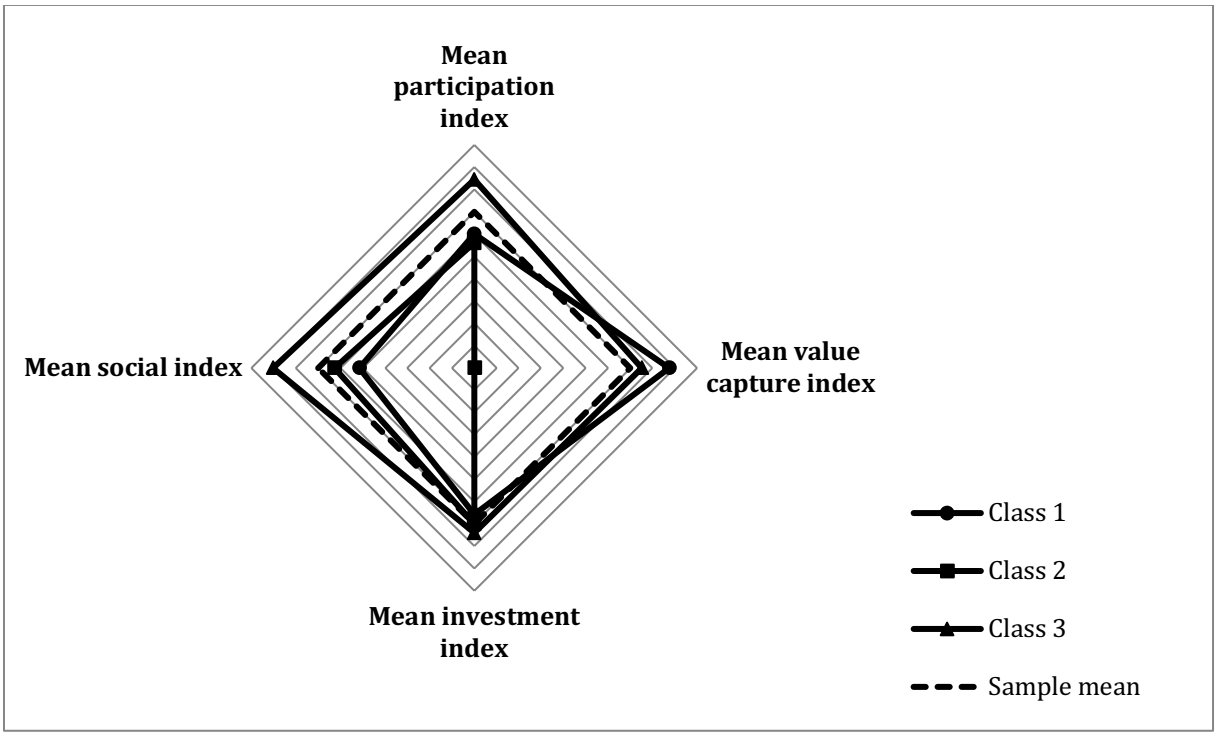
Class 1 of the authors' grouping	Class 2 of the authors' grouping	Class 3 of the authors' grouping
Argentina	Brazil	Czech Republic
Canada	Philippines	Finland
Greece	United States	Hungary
Indonesia	**Colombia**	Ireland
Portugal	**Germany**	South Korea
Russian Federation	**France**	Luxembourg
Australia	**United Kingdom**	Malaysia
New Zealand	**Israel**	Slovakia
Belgium	**Italy**	Slovenia
Chile	**Japan**	Thailand
Denmark	**Turkey**	Viet Nam
Spain	**Austria**	**Costa Rica**
Iceland	**Switzerland**	**Estonia**
Norway	**Cambodia**	**Croatia**
	Netherlands	**India**
	Sweden	**Mexico**
		Poland
		South Africa
		China
		Romania
		Tunisia
Percentage of countries having switched classes	57%	81%	48%
Number of countries added/subtracted from the authors' original class	7	-8	1

NB: Countries in bold are those that switched classes when traditional GVC participation and value capture indicators were used to perform the PCA and the cluster analysis

As Table 17 shows, when the cluster analysis is performed using the PCA that includes the standard GVC participation and value capture indicators the country composition of the three classes, as well as the length of classes 1 and 2, are considerably altered.

Finally, if we analyze the characteristics of each class in terms of the values each variable take for each of them we find a profile very different to that found when the authors' GVC participation and value capture indicators were used.

Figure 1: Mean value of each variable by class and for the sample when traditional GVC participation and value capture indicators are used



As shown in Figure 1, when traditional GVC participation and value capture indicators are used the profiles of each group differ radically from those obtained using the authors' indicators. Class 1 and class 3 have no distinct profile. All the variables are close to the sample mean. The same applies to class 2, with the exception of value capture, which takes a very low value, making this the only distinctive trait to be found in the profile of the three classes.

Appendix references

Birdsall, Nancy. 2010. "The (Indispensable) Middle Class in Developing Countries; or, the Rich and the Rest, Not the Poor and the Rest." Center for Global Development Working Paper No. 207. Available at SSRN: http://ssrn.com/abstract=1693899

CONEVAL (Consejo Nacional de Evaluación de la Politica de Desarrollo Social). 2014. "Informe de Evaluación de la Política de Desarrollo Social en México 2014." Available at http://www.coneval.org.mx/Informes/Evaluacion/IEPDS_2014/IEPDS_2014.pdf

Diewert, W.E., Alterman, W., Eden, L., 2005. Transfer prices and import and export price indexes: theory and practice. Available SSRN 734883.

Dykstra, S., Dykstra, B. and Sandefur, J. 2014. "We Just Ran 23 Million Queries of the World Bank's Website", Center for Global Development, Working Paper 362.

Escaith, H., 2008. Measuring trade in value added in the new industrial economy: statistical implications. Available SSRN 1189162.

Gu, Danan, Feng, Quishi, Wang, Zhenglian, and Yi Zeng. 2015. "Recommendation to consider the crucial impacts of trends in smaller household size on sustainable development goals." Global Sustainable Development Report Brief. Available at https://sustainabledevelopment.un.org/content/documents/7021Recommendation%20to%20consider%20the%20crucial%20impacts%20of%20trends%20in%20smaller%20household%20size%20on%20sustainable%20development.pdf

Hellebrandt, Tomas, and Paolo Mauro. 2015. "The Future of Worldwide Income Distribution." Peterson Institute for International Economics Working Paper(15-7).

ILO, n.d. LABORSTA.

Kochhar, Rakesh. 2015. "Appendix: Methodology and Data Sources." http://www.pewglobal.org/2015/07/08/appendix-methodology-and-data-sources/#fnref-33279-53

Kuhn, Ursina and Christian Suter (2015), "L'évolution des inégalités de revenus en Suisse. Social Change in Switzerland." N° 2. Retrieved from http://socialchangeswitzerland.ch

Milanovic, Branko L. All the Ginis Dataset, World Bank Group. Available online at http://data.worldbank.org/data-catalog/all-the-ginis

NIPSSR (National Institute of Population and Social Security Research). 1998. "Overview of Household Projections for Japan." Available at http://www.ipss.go.jp/pp-ajsetai/e/hprj_98/hprj_98.html

UN Statistical Commission, 2016. Classification by Broad Economic Categories, Revision 5.

UNU-WIDER, "World Income Inequality Database (WIID3c)", September 2015, https://www.wider.unu.edu/project/wiid-world-income-inequality-database

World Bank, 2016. World Development Indicators.