UNITED NATIONS
INDUSTRIAL DEVELOPMENT ORGANIZATION

Inclusive and Sustainable Industrial Development Working Paper Series

WP 2 | 2023

Expanded Real Value Added Data for Manufacturing: A New Approach to Measuring Sub-Sectoral Manufacturing Development

DIVISION OF CAPACITY DEVELOPMENT, INDUSTRIAL POLICY ADVICE AND STATISTICS

WORKING PAPER 2/2023

Expanded Real Value Added Data for Manufacturing: A New Approach to Measuring Sub-Sectoral Manufacturing Development

Nobuya Haraguchi
UNIDO

Juergen Amann
OECD

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION

Vienna, 2023

Acknowledgements

The authors are very grateful to the participants of the conference "New Perspectives on Structural Change Causes and Consequences of Structural Change in the Global Economy", Maastricht 21 – 23 February 2018. The work would not have been possible without the precious collaboration of UNIDO colleagues, including Nicola Cantore and Charles Fang Chin Cheng as well as the valuable insights and comments by Christoph Hammer and Emanuele Russo. Niki Rodousakis provided copy-editing assistance. Iguaraya Saavedra supported formatting the working paper.

Contents

List of Figures

List of Tables

Abstract

We propose a new single deflation method to expand real value added data coverage for UNIDO's Industrial Statistics Database (INDSTAT, 2021). Our deflator is consistent with current national accounting practices and self-contained, i.e., it only requires data available via INDSTAT. Furthermore, we discuss various deflator extensions to further real value added coverage, enabling us to derive an extensive data set with notably higher data coverage compared to other cross-country data sets, particularly for low(er) income economies. This allows us to measure the manufacturing sector's performance in unprecedented detail.

Keywords: value added; structural change; developing countries; manufacturing; macroeconomic measurement; manufacturing industry development.

JEL codes: L60; O14; E01; E30.

1 Introduction

Much research is devoted to understanding the mechanisms driving the nexus between industrialisation and economic growth. However, there is a lack of cross-country data providing comparable metrics to measure real value added at the manufacturing industry level,[1] particularly from countries at the low(er) end of the income distribution. This presents a substantial barrier to understanding development trajectories in manufacturing at earlier stages of economic development as this data blind spot disproportionally affects the group of low(er) countries which would benefit most from industrial development.

Academic research on the issue is similarly affected by these constraints. For that matter, most cross-country work on manufacturing development focuses on a selection of already industrialised high-income countries (Timmer et al., 2015; STAN, 2021), is performed at the more aggregated sector level (Palma, 2014) or resorts to estimation techniques to address the lack of real value added data in large industrial databases (Rodrik, 2016).

Furthermore, benchmarking the manufacturing sector's performance to identify and promote long-run, growth-enhancing policy measures remains one of the biggest challenges for policymakers in developing countries, where industrialisation may offer new avenues of sustained economic growth. Without comprehensive cross-country and time-series evidence, it remains unclear for country officials if such dynamics are related to time-specific global market forces and reflect *structural change* dynamics, or if they are reflecting country-specific impairments.

In this paper, we propose a new single deflation method to expand the availability of real value added data at the two-digit ISIC level for UNIDO's Industrial Statistics Database INDSTAT (2021). We derive a single deflator, which is consistent with current national accounting practices and self-contained, i.e., only requires data available via INDSTAT. We also discuss various extensions to further improve data coverage, which we compare to the World Input-Output Database (WIOD) (Timmer et al., 2015) and OECD's Structural Analysis database STAN (2021). Our deflator enables us to build an extensive dataset to analyse structural change dynamics within the manufacturing sector in unprecedented detail, particularly for low(er) income economies. To our knowledge, this is the first study that expands real value added data coverage for a large group of developing countries at the

[1] This paper focuses on the breakdown of the manufacturing sector at the *division* level the International Standard Industrial Classification (ISIC Rev.3.1, 2002), which we refer to as the *industry* or *two-digit* level for the remainder of the document. However, the method outlined in this paper can be applied more broadly, for example, to calculate real value added sequences at the *class* or *four-digit* level of ISIC as well as to older as well as more recent ISIC revisions, as long as the data contain the variables necessary for the construction of the deflator proposed in this paper.

manufacturing sub-sector level.

The remainder of this paper is structured as follows. In the next section, we revisit the discussion on the role of manufacturing in economic development, followed by a primer on current practices and limitations when deflating nominal value added in section 3. In section 4, we derive the single deflation methods for real value added data series. In section 5, we discuss various extensions of the previously derived deflator to further increase the availability of real value added data, as well as a selection mechanism to construct the final database. Finally, we use our novel dataset in section 6 to present two possible avenues for future work, having both a more academic as well as policy-oriented audience in mind. We illustrate how our data can be used to re-visit the earlier structural change literature, which is concerned with structural change within the manufacturing sector as well as a more policy-centred application that concerns the creation of indicators to measure the value added contribution of certain industries. Section 7 concludes.

2 The role of manufacturing for economic development

2.1 Manufacturing as the engine of growth...

There is a wide consensus on the fundamental role manufacturing plays for economic development. Particularly in poor(er) countries, economic growth can at least partially be achieved by transitioning out of agricultural and subsistence farming into a formally integrated economic framework; see, among others, Gollin et al. (2002); Kuznets (1957); Matsuyama (1992). Lewis (1954) shows that developing countries can drive economic development through capital accumulation in a capitalist sector.[2] A sustained process of industrialisation usually accompanies long-term economic growth, i.e. an increase in the share of manufacturing value added (MVA) in gross domestic product (GDP), which is attributed to the sector's higher level of productivity, its linkage effects and demand effects (Kaldor, 1967).[3]

The role of manufacturing as an engine of growth also draws from its strong backward linkages with other sectors (Hirschman, 1958). Manufacturing production generates demand for inputs from all

[2] Lewis (1954) shows that in practice, wages paid to employees in the capitalist sector are higher than those in the subsistence sector due to (i) the higher cost of living in the capitalist sector, which is often located in an urban area, (ii) the psychological cost of moving from a simple life in the subsistence sector to a better organised urban environment or (iii) the recognition that workers demand higher wages after having acquired particular tastes and social prestige associated with urban life.

[3] Besides the 'quantity' of manufacturing in developing countries, several recent studies emphasise that the qualitative characteristic of manufacturing as an engine of growth typically represented in Kaldor's Laws has not changed, either; see, among others, Chakravarty and Mitra (2009); Kathuria and Natarajan (2013); Marconi et al. (2016); Su and Yao (2017); Szirmai and Verspagen (2015).

sectors and creates ripple effects for businesses that are only indirectly linked to the final production stage through supply chains (Timmer et al., 2014). Furthermore, substantial job creation multipliers associated with the manufacturing sectors have been reported for advanced economies (Moretti, 2010; Moretti and Thulin, 2013).

The rapid accumulation of production experiences also enhances learning in many functional areas, including research and development, production and marketing, thus resulting in higher productivity levels of productivity (Thirlwall, 2002; Arrow, 1962; Dalum et al., 1992). For example, Rodrik (2013) highlights the manufacturing sector's unique position of unconditional convergence with the technological frontier. A foothold in the manufacturing sector is therefore likely to lead to continuous productivity increases, regardless of country-specific conditions.[4]

2.2 ...running out of steam?

Structural change, that is, the long-term changes in the composition of economic aggregates, is inherently linked to the theoretical and empirical understanding of the interplay of economic aggregates (Krüger, 2008). Manufacturing contribution to employment and value added creation is acknowledged to follow a hump-shaped pattern as economies move along their income trajectory and transform structurally, driven by market mechanics as well as technological and geopolitical developments.[5]

However, recent studies indicate that the dynamics of the industrial sector, and manufacturing in particular, are not only subject to changes in income level but also seem to have an inter-temporal dimension. For example, Haraguchi (2015) find higher variations in the share of manufacturing along the income trajectory between the 1960s and 1980s than during preceding periods. A similar observation is also made by Palma (2014) as well as Rodrik (2016). These studies find that the hump-shaped relationship between manufacturing-related employment and value added moved down the income scale over time, leading to projected decreases in employment and value added generation at earlier stages of economic development than for earlier industrialisers. Palma (2014) also notes that the hump-shaped relationship for employment seems to disintegrate and level out over time. Such time-dependent patterns of premature deindustrialisation have led researchers to argue that manufacturing-led growth has become a more difficult path for currently developing countries to follow. As (Rodrik, 2016, p. 1)

[4] Such growth-enhancing structural change characterises the experiences made by Asia. At the same time, Latin America and sub-Saharan Africa have primarily witnessed growth-reducing structural change since 1990 (McMillan and Rodrik, 2011), although McMillan et al. (2014) concludes that sub-Saharan Africa has undergone growth-enhancing structural change since 2000.

[5] See, among others, Matsuyama (2009); Van Neuss (2018, 2019); Vu et al. (2021); Sáenz (2022).

puts it, they *'[...] are running out of industrialisation opportunities sooner and at much lower levels of income compared to the experiences of earlier industrialisers'.* Similarly, Felipe et al. (2019) emphasise that the share of manufacturing employment of late developers reaches its tipping point at a much lower per capita income level than was the case for earlier industrialisers. In the same vein, Tregenna (2009) describes deindustrialisation tendencies as identified by a decline in the manufacturing sector's share of employment and value added in the total economy and provides a conceptual framework to determine whether a deindustrialisation process is desirable or not.

2.3 A closer look: Structural change within manufacturing

Concerns about the future relevance of the manufacturing sector are often based on the observed downward shift of manufacturing value added shares in GDP and the share of manufacturing employment in total employment within countries at different income levels.

Addressing agglomeration effects, Haraguchi et al. (2017) shows that the share of aggregate manufacturing value added in GDP, as well as that of aggregate manufacturing employment in total employment across developing countries, has remained constant since 1970, even during the period when the start of deindustrialisation shifted to a lower per capita income level.[6] The difference between country and aggregate averages boils down to the difference between unweighted and weighted country averages. This result underscores the fast-paced manufacturing development of very populous developing countries in the global economy, which seems indicative of the concentration of manufacturing-related production in (a) group(s) of larger, more populous countries. Similar observations were made by Felipe and Mehta (2016).

Addressing within-sector dynamics, earlier research already indicated that structural change dynamics within manufacturing are highly heterogeneous. Starting with the seminal work of Chenery (1960), researchers sought to formalise previous empirical studies by identifying regularities of growth of the manufacturing development pattern through regression estimations, recognising country characteristics such as size (Chenery and Taylor, 1968) as well as non-linear development patterns (Syrquin and Chenery, 1975; Chenery et al., 1986; Syrquin and Chenery, 1989). While these earlier studies have contributed significantly to the understanding of the general patterns of industrialisation relative to other sectors, due to the lack of real value added data, most of the previous studies, including the com-

[6] Haraguchi (2015) calculates the average country-level shares of MVA in GDP as the sum of each country's MVA share in GDP divided by the number of countries, while the aggregate share is measured as the developing countries' total MVA divided by their total GDP.

prehensive work by Syrquin and Chenery (1989), used the share of value added in GDP. This made it difficult to observe the development pattern of individual industries, not as relative to other industries. Furthermore, the periods covered by previous studies focus mainly on the 'heyday' of industrialisation, where there was little concern about premature de-industrialisation or available data for countries at low(er) income levels.

More recently, Haraguchi and Amann (2020, 2021) revisited this strand focusing on heterogeneities of structural development within manufacturing. They highlight that heterogeneities in structural transformation are intrinsic to manufacturing industries; that is, the effect of de-industrialisation and productivity growth vary vastly across industries as well as over time. Furthermore, premature deindustrialisation is a feature of certain industries, in particularly the textiles and wearing apparel industries, while other low-skill industries such as food and beverages's employment and value added growth paths remain largely unaffected by intertemporal dynamics.

3 A primer on deflating nominal value added: Current practices and limitations

3.1 Deflator types

Following international accounting standards (SNA, 2008), nominal value added for period t (V_t) is given by the difference between nominal output (O_t) and intermediate inputs (I_t) :

$$\begin{aligned} V_t &= O_t - I_t \\ &= P_t^O \times Q_t^O - P_t^I \times Q_t^I, \end{aligned} \tag{1}$$

where $\ell_t = P_t^\ell \times Q_t^\ell$ denotes the nominal value of gross output ($\ell = O$) and intermediate inputs ($\ell = I$), with P and Q corresponding to their respective prices and quantities at time t.

The double deflation method derives real value added (V^{DD}), as the difference between deflated current price output and deflated current price intermediate consumption:

$$V_t^{DD} = \frac{O_t}{D_t^O} - \frac{I_t}{D_t^I}, \tag{2}$$

where, $D_t^\ell = (P_t^\ell \div P_b^\ell)$ represents the deflator for output ($\ell = O$) and intermediate inputs ($\ell = I$) and b reflects the reference year, respectively.[7]

The single deflation method derives real value added (V^{SD}) by deflating nominal gross output and nominal intermediate inputs using the same price deflator, in most cases the gross output price deflator:[8]

$$V_t^{SD} = \frac{V_t}{D_t^O}. \tag{3}$$

3.2 Current practices in in cross-country datasets and challenges

General considerations. The double deflation method is sound in theory, but in practice, the estimates are affected by measurement errors in both the output and intermediate consumption volume estimates (SNA, 2008). In processing industries, especially those heavily reliant on imported intermediary inputs, the estimate is highly sensitive to errors. Furthermore, it provides a more volatile estimation of value added during episodes of hyperinflation or rapid changes in product quality (Eurostat, 2014). Because of these issues, many national statistical offices resort to single deflation methods, and the System of National Accounts emphasises that the choice between double and single deflation methods must be based on judgement and country-specific circumstances (SNA, 2008, 15.136).[9]

As national data forms the foundation for many cross-country datasets, single as well as double deflation methods are typically present in the same international cross-country dataset. This is particularly true for sub-sector data, where the data availability on intermediate input price (P_t^I) is often limited, particularly in developing countries.

In addition to being time- and cost-efficient, the single deflation method is also less sensitive to hyperinflation and production volatility, two problems that may disproportionately affect developing countries. However, Li and Kuroko (2016) show that the single deflation method overestimates real

[7] For the sake of simplicity, we assume D_t is a generic cost-of-goods deflator, such as the Laspeyres price index, $D_t = (P_t \times Q_b) \div (P_b \times Q_b) = P_t \div P_b$, where b denotes the reference year (ILO et al., 2020). Note that the generic representation of the double deflation method in Equation 2 was chosen for expositional purposes only. It does only hold universally true unless $b = t - 1$ as a consequence of the introduction of chain-linked volumes (SNA, 2008) and the non-additivity of chain volumes. However, these technical complications are not relevant to the purpose of this paper which only concerns single deflation.

[8] Some countries use the consumer price index (CPI) as the single deflator even though it only measures prices of goods and services purchased by domestic households for consumption (ILO et al., 2020).

[9] For example, Li and Kuroko (2016) states that China mainly uses the single deflation method and partly uses the quantity extrapolation method.

value added when the price increase of intermediate goods is small relative to that of output. The single deflator method is also subject to a systematic bias stemming from the price differential between output and intermediate price series as described in Appendix A.

Deflation in cross-country sector-level datasets. Researchers have few cross-country data sources available to analyse economic activities within manufacturing. At the ISIC two-digit level, OECD's Structural Analysis database, STAN (2021)[10], and the World Input-Output Database, WIOD by Timmer et al. (2015)[11], are among the most prominent and readily available data sources and are characterised by particularly good data coverage of advanced economies.[12] The use of single deflation methods is widespread in these two datasets. In particular, out of 27 European countries included in the WIOD, the single deflation method was used for over half of the countries (Erumban et al., 2012). Furthermore, one in four countries contained in STAN does not report intermediate consumption deflators necessary for the double deflation.[13]

In turn, INDSTAT (2021)[14] collects manufacturing production data at the industry level for a large country sample going back to the early 1960s, but does not provide an accompanying deflator series to accommodate the need for real value added series by default. This situation generates a trade-off: INDSTAT manufacturing data offers the most comprehensive and detailed overview of manufacturing development (past and present) and, in particular, for countries at lower levels of economic development, but fails to provide a comprehensive tool to deflate value added. Researchers in the past have circumvented this issue by employing econometric techniques. For example, Rodrik (2013) estimates convergence in manufacturing conditional on labour productivity dynamics of a frontier economy (the U.S.) and a common global inflation rate. While such approaches may offer a remedy for analytical work, they do not resolve the issue of a lacking deflator which limits the usability of INDSTAT vis-a-

[10] Data retrieved from `http://www.oecd.org/sti/ind/stanstructuralanalysisdatabase.htm` (last accessed April 2021) are the most prominent contenders. We use the database *SNA93, ISIC Rev. 3 version of STAN (last update: May 2011)* instead of the more recent ISIC Rev. 4 datasets, as the former provides better data coverage for earlier periods and follows the same industry classification as INDSTAT. Next, we analyse STAN data for 29 countries only, as Australia, Ireland and Poland do not report deflator series for the sectors we are interested in; see coverage file available at `http://www.oecd.org/sti/industryandglobalisation/46671527.XLS` (last accessed April 2021) for further information.

[11] We use the July 2014 issue of the Socio-Economic Accounts (SEA) data which contain, among other variables, value added at current and constant prices; see `http://www.wiod.org/database/seas13` for more information, which is the most recent version following the ISIC Revision 3 industry classification.

[12] At the time of writing, updated versions for both the STAN and WIOD dataset were available. However, this study uses the latest instalments of all Rev. 3 classification datasets to retain consistency across datasets, given our focus on deriving rich historical data.

[13] Intermediate consumption deflators are not available for Australia, Canada, Costa Rica, Hungary, Israel, Korea, New Zealand, Spain, Turkey and the United Kingdom according to STAN Country Notes (`https://www.oecd.org/sti/ind/stanstructuralanalysisdatabase.htm`.

[14] Data retrieved from `https://unido.org/researchers/statistical-databases` (last accessed April 2021).

vis the wider fields of application a complete data sources can offer to researchers, policy analysts or practitioners.

4 Extending real value added data in INDSTAT

In what follows, we propose a new single deflation method to expand the availability of real value added data at the two-digit ISIC level for UNIDO's INDSTAT data. We show that our deflator is consistent with current national accounting practices and is self-contained, that is, only requires data available via INDSTAT. Furthermore, we illustrate the much-improved data coverage of this approach relative vis-a-vis WIOD (Timmer et al., 2015) and STAN (2021) datasets, particularly for low(er) income economies.

4.1 A new single deflator for INDSTAT

Consider the representation of the single deflator method in Equation 3 and substitute for the definition of the output deflator:

$$\begin{aligned} V_t^{SD} &= V_t \div D_t^O \\ &= V_t \times (P_b^O \div P_t^O). \end{aligned} \tag{4}$$

Next, expand by Q_t^O and rewrite the expression for nominal gross output in year t as $O_t = P_t^O \times Q_t^O$:

$$\begin{aligned} V_t^{SD} &= V_t \times (P_b^O \div P_t^O) \times (Q_t^O \div Q_t^O) \\ &= V_t \times P_b^O \times (P_t^O \times Q_t^O)^{-1} \times Q_t^O \\ &= V_t \times O_t^{-1} \times P_b^O \times Q_t^O. \end{aligned} \tag{5}$$

Note that Q_t^O is the output quantity in period t is identified by the Index of Industrial Production (IIP), which captured real production growth of industrial activities at the ISIC two-digit level. More precisely, the IIP captures volume changes between a baseline period b and t as:

$$IIP_t = \frac{\tilde{O}_t}{O_b}, \tag{6}$$

where $\tilde{O}_t = O_t \div D_t^O = \left(Q_t \times P_t^O\right) \div \left(P_t^O \div P_b^O\right) = Q_t^O \times P_b^O$, that is, the value of gross output in year t at constant base-year prices b (UNSD, 2013, 2010; Herbel, 2014; Yamada, 2016). Simplifying Equation 6 accordingly and rewriting $O_b = Q_b^O \times P_b^O$ yields:

$$\begin{aligned} IIP_t &= \frac{Q_t^O \times P_b^O}{Q_b^O \times P_b^O} \\ &= \frac{Q_t^O}{Q_b^O}, \\ \Rightarrow Q_t^O &= Q_b^O \times IIP_t. \end{aligned} \tag{7}$$

Substituting Equation 7 into Equation 5 and rewriting the expression for nominal gross output in reference year b as $O_b = P_b^O \times Q_b^O$:

$$\begin{aligned} V_t^{SD} &= V_t \times O_t^{-1} \times P_b^O \times Q_t^O \\ &= V_t \times O_t^{-1} \times P_b^O \times (Q_b^O \times IIP_t) \\ &= V_t \times O_t^{-1} \times (P_b^O \times Q_b^O) \times IIP_t \\ &= V_t \times (O_t^{-1} \times O_b) \times IIP_t \\ &= V_t \times \frac{O_b}{O_t} \times IIP_t. \end{aligned} \tag{8}$$

Consequently, we can derive real value added using the single deflator method by only utilising data contained in the INDSTAT database, namely nominal value added (V) and output (O) as well as the IIP for country i, industry s and periods t, b, respectively.[15] We refer to this deflation method as *Method 1* (*m1*), which we restate for future dispositional purposes below:

$$V_{ist}^{m1} = V_{ist} \times Def_{ist}, \quad Def_{ist} = \frac{O_{isb}}{O_{ist}} \times IIP_{ist}.^{16} \tag{9}$$

[15] The quality of the estimation is dependent on, among other factors, how frequently the weights of the IIPs are updated. Weights reflect the importance of different activities in the industry for which an IIP is constructed (UNSD, 2013). Thus, using outdated weights ignores any structural change within the sub-sector aggregates and introduces bias in the gross output trend. United Nations (2008) surveyed how countries calculate the IIP. The survey included 62 countries, of which 33 were developing countries, including 10 from Africa, 10 from Asia, seven from Europe, five from Latin America and one from Oceania. Among the 33 developing countries, 61% of the countries updated their weights annually or at least every five years. 24% of them had periodical updates with an interval of more than five years. The remaining 15% had irregular updates of the weights without any update for the last five years, giving us confidence in the quality of real value added estimated by the single deflation method for developing countries.

[16] Note that $Def_{ist} = O_{isb} \div O_{ist} \times IIP_{ist} = (P_{isb} \times Q_{isb}) \div (P_{ist} \times P_{ist}) \times (Q_{ist} \div Q_{isb}) = P_{isb} \div P_{ist}$. This corresponds to the inverse of the output price index, complying with the recommendations to use the output deflator for value added deflation Eurostat (2014).

4.2 Assessing data coverage and quality: INDSTAT vs STAN and WIOD

4.2.1 Data coverage

Table 1 compares the data availability of *Method 1* (*m1*) derived in Equation 9 with the two primary data sources for manufacturing at the industry level, i.e. STAN (2021) and WIOD (Timmer et al., 2015). Particularly for countries that are not classified as either 'high income: OECD' or 'Europe & Central Asia', INDSTAT offers a considerable improvement over STAN and WIOD, and the single deflation method provides a substantial increase of real value added data vis-a-vis the other datasets.

4.2.2 Data quality

Merely increasing data coverage to evaluate the performance of *Method 1* (*m1*) is not worth the effort unless we also consider the quality of the newly derived dataset relative to other available data sources between INDSTAT, WIOD and STAN. For this purpose, we offer visual and numerical evidence to verify the data quality of our proposed deflator.

Visual comparison. We compare the nominal value added series of the G7 countries with their respective deflators and real value added time series for *m1* with that in STAN and WIOD. For the sake of brevity, we only provide an illustration of real and nominal value added series and the deflator for the US in Figure 1 but offer more extensive visual evidence for all remaining G7 economies in Appendix B. Furthermore, we provide visualisation for all individual data sequences through a dynamic online tool.[17]

It is worth emphasising that the differences between INDSTAT and STAN/WIOD for nominal data may be explained by UNIDO's international mandate to collect nominal value added data directly from countries following internationally recommended enterprise survey procedures. Therefore, differences in the nominal series do not imply lower data quality of the INDSTAT data, but are the result of different data compilation practices.

Comparing the differences between INDSTAT and STAN/WIOD data, it is evident that discrepancies in the real value added patterns are typically inherited from the respective nominal series. Consequently, this paper's newly proposed value added deflator is not the root of the heterogeneities across datasets.

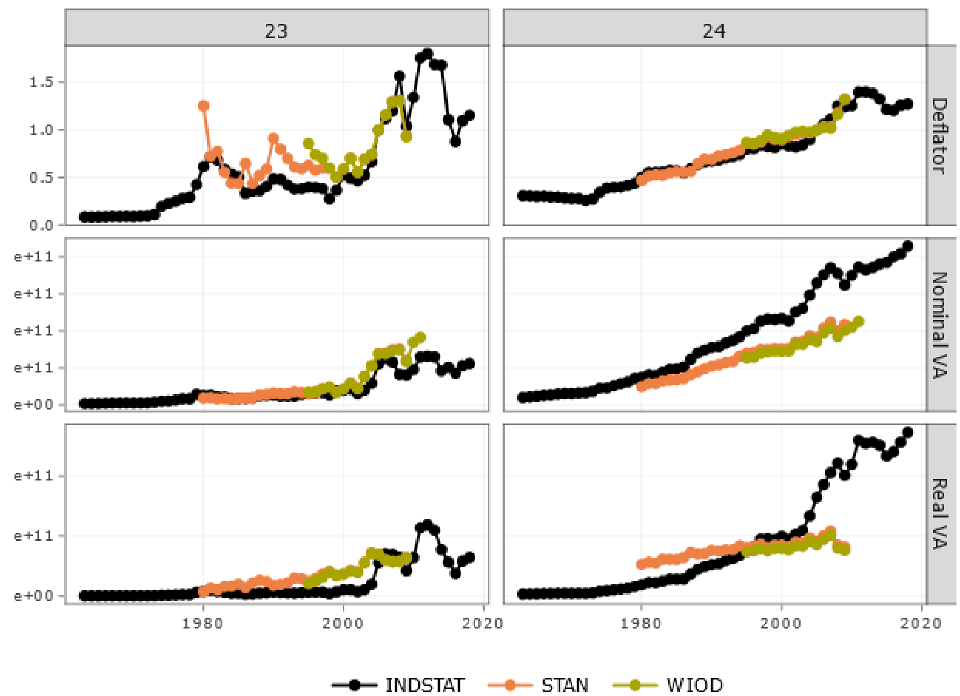
[17] The dynamic online tool is accessible at `https://amannj.shinyapps.io/rVA_Explorer/`.

Table 1: Comparison of data coverage

	No. of observations			No. of countries		
	INDSTAT	STAN	WIOD	INDSTAT	STAN	WIOD
Deflator by income group						
High income	25709	10002	2700	48	28	30
Low income	532	.	.	9	.	.
Lower middle income	6639	.	180	17	.	2
Upper middle income	11911	464	630	28	1	7
Deflator by region						
East Asia & Pacific	6566	1000	450	16	3	5
Europe & Central Asia	21345	8108	2520	44	22	28
Latin America & Caribbean	6765	464	180	12	1	2
Middle East & North Africa	4700	63	90	14	1	1
North America	1894	831	180	2	2	2
South Asia	1067	.	90	2	.	1
Sub-Saharan Africa	2454	.	.	12	.	.
Nominal Value-added by income group						
High income	36777	13485	3060	53	31	30
Low income	4654	.	.	18	.	.
Lower middle income	16051	.	204	34	.	2
Upper middle income	21491	532	714	47	1	7
Nominal Value-added by region						
East Asia & Pacific	12313	1586	510	20	4	5
Europe & Central Asia	26975	10503	2856	46	24	28
Latin America & Caribbean	13291	532	204	26	1	2
Middle East & North Africa	10716	266	102	20	1	1
North America	2251	1130	204	3	2	2
South Asia	3084	.	102	6	.	1
Sub-Saharan Africa	10343	.	.	31	.	.
Real Value-added by income group						
High income	25157	10002	2700	48	28	30
Low income	328	.	.	7	.	.
Lower middle income	6382	.	180	17	.	2
Upper middle income	11192	464	630	28	1	7
Real Value-added by region						
East Asia & Pacific	6560	1000	450	16	3	5
Europe & Central Asia	20025	8108	2520	43	22	28
Latin America & Caribbean	6517	464	180	12	1	2
Middle East & North Africa	4686	63	90	14	1	1
North America	1894	831	180	2	2	2
South Asia	956	.	90	1	.	1
Sub-Saharan Africa	2421	.	.	12	.	.

Note: Numbers of observations/countries contained in each of the three data sets. *Real value added* for INDSTAT data according to single deflator method *m1* as described in Equation 9. Country classification according to World Bank (2021).
Source: Authors' elaboration based on INDSTAT (2021), STAN (2021) and WIOD (Timmer et al., 2015) data.

Figure 1: Visual comparison *m1* vs STAN and WIOD United States



Note: Figure extracted from `https://amannj.shinyapps.io/rVA_Explorer/`, tab `M1 STAN-WIOD comparison`. Deflator (index, 2005 = 100), nominal and real value added (BUSD), y-axis in log scale. *Real value added* for INDSTAT data according to single deflator method *m1* as described in Equation 9.
Source: Authors' elaboration based on INDSTAT (2021), STAN (2021) and WIOD (Timmer et al., 2015) data.

Numerical comparison. In addition to a visual comparison, we also provide a simple numerical comparison mechanism using Dynamic Time Warping (DTW) to quantify the degree of similarity between the three datasets (INDSTAT, STAN, WIOD) for nominal and real value added as well as the corresponding deflators. We provide a more extensive discussion of the numerical comparison method in Appendix C.

In short, we calculate normalised dynamic time warping distance (nDTW), which quantifies the degree of similarity across the three datasets relative to the similarity of the same data sequence of the baseline pair, i.e., the similarity of the data sequence between STAN and WIOD. Consequently, the closer a sequence pair is to the baseline pair, the closer nDTW is to one, while a high nDTW indicates that the compared pair of sequences is more dissimilar to the baseline pair $j = \{STAN, WIOT\}$. In addition, we also evaluate the similarity of the series with that of a random white noise sequence, WN.

The results in Table 2 show that the nDTWs for the $m1$ deflator is smaller on average than the nominal value added sequences and have less variation. Furthermore, they outperform the random white noise (WN) model unequivocally. Lastly, a more notable variation is reported for the nominal

Table 2: Summary table time-series dissimilarity

	Series	Mean	SD	Percentile 25th	50th	75th
STAN vs. WIOD	Deflator	1.00	0.00	1.00	1.00	1.00
	Nominal VA	1.00	0.00	1.00	1.00	1.00
	Real VA	1.00	0.00	1.00	1.00	1.00
m1 vs. STAN	Deflator	15.51	18.00	5.28	8.85	16.48
	Nominal VA	42.52	51.92	8.46	18.57	68.36
	Real VA	46.63	57.77	10.52	20.77	60.47
m1 vs. WIOD	Deflator	15.94	18.15	5.40	9.62	17.05
	Nominal VA	42.37	51.92	7.70	19.37	67.86
	Real VA	46.50	57.71	9.71	20.43	60.97
m1 vs. WN	Deflator	77.81	60.92	35.52	60.75	103.01
	Nominal VA	32.87	44.24	1.42	11.51	54.04
	Real VA	25.57	38.51	0.85	5.21	33.63
STAN vs. WN	Deflator	79.16	59.36	38.36	61.43	105.92
	Nominal VA	63.57	60.96	22.04	36.52	82.32
	Real VA	53.48	54.20	20.31	33.39	65.71
WIOD vs. WN	Deflator	78.95	59.42h	38.43	60.43	106.05
	Nominal VA	63.30	61.14	20.99	36.61	81.32
	Real VA	53.10	54.01	19.24	32.59	65.41

Note: WN: random white noise. See Appendix C for more information. Real value-added for INDSTAT according to single deflator method *m1* as described in Equation 9. Aggregates obtained by aggregating nDTWs.
Source: Authors' elaboration based on INDSTAT (2021), STAN (2021) and WIOD (Timmer et al., 2015) data.

value added series between INDSTAT, STAN and WIOD. This result, however, is a feature of the raw data compilation of INDSTAT discussed earlier and not the result of using our proposed deflator method.

5 Expansion of country and inter-temporal coverage

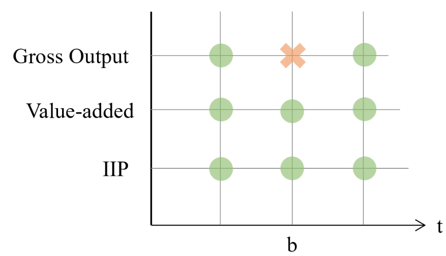
5.1 Extension methods

The single deflation method introduced as *Method 1(m1)* represents the backbone of our analysis. It relies on gross output, nominal value added and IIP sequences to calculate real value added. However, data gaps in either series may result in a partial or complete lack of real value added when applying *m1*. We address this issue by introducing various extension methods below.

Method 2. *Method 2* ($m2$) addresses the case where gross output measures for the reference year are not available. This is a significant problem, as one missing observation alone is enough to make $m1$ infeasible, even if nominal value added and gross output data, as well as the IIP sequences, are readily available for all other years; see Figure 2 for a visual representation. In this scenario, *Method 2* uses the spatial interpolation method co-kriging (CK) to recover the base-year value for gross output $\tilde{O}^j_{isb}$; see Appendix D for more information of the spatial interpolation employed in this paper.[18] Then, real value added can be calculated as:

$$V^{m2,j}_{ist} = V_{ist} \times \frac{\tilde{O}^j_{isb}}{O_{ist}} \times IIP_{ist}. \tag{10}$$

Figure 2: Visual representation of *Method 2*



Method 3. *Method 3* ($m3$) addresses gaps in the data for nominal value added and gross output; see Figure 3. In this case, we assume that the ratio of nominal value added and gross output, $y_{ist} = V_{ist}/O_{ist}$, for a gap of length R, $r = 1, ..., R$, follows a linear progression between the respective observations between the last two known points, i.e. y_{ist} and $y_{is(t+R+1)}$. For any point $r \in R$, where nominal value added and growth output is not observed, the value added to output ration, $y_{is(t+r)} = V_{is(t+r)}/O_{is(t+r)}$, is linearly interpolated as:

$$\frac{y_{is(t+r)} - y_{ist}}{(t+r) - t} = \frac{y_{is(t+R+1)} - y_{ist}}{(t+R+1) - t}.$$

Solving for $y_{is(t+r)}$ we obtain:

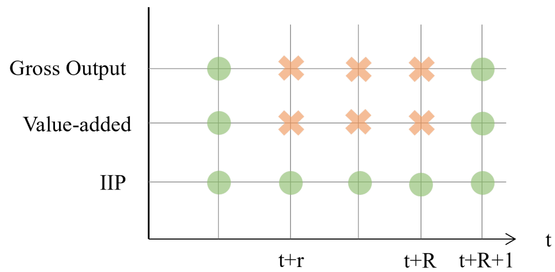
$$y_{is(t+r)} = \frac{y_{ist}((t+R+1) - (t+r)) + y_{is(t+R+1)}((t+r) - t)}{(t+R+1) - t},$$

[18] For the remainder of the paper, any spatially interpolated representation of sequence x_t is designated $\tilde{x_t}$. For any spatially interpolated series, we use CK and one of the following three competing measures of economic proximity. These are *gross capital formation*, *trade*, and *natural resources*, i.e., $j = \{cf,\ td,\ nr\}$, respectively. We obtain the data from the World Development Indicator (WDI) database. See Appendix D for a detailed description of the spatial interpolation technique used throughout this paper.

which we then use to calculate the real value added series for industry s of country i:

$$V^{m3}_{is(t+r)} = y_{is(t+r)} \times O_{isb} \times IIP_{ist}.$$

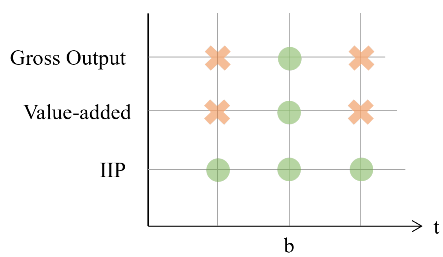
Figure 3: Visual representation of *Method 3*



Method 4. *Method 4 (m4)* addresses the case where gross value added and output is only available in the reference year; see Figure 4. It imposes a stronger assumption on the evolution of the growth output ratio, namely that the ratio of nominal gross value added and gross output remains constant and fixed at the level of the reference year b, i.e., $y_t = y_b \ \forall \ t$:

$$V^{m4}_{ist} = V_{isb} \times IIP_{ist}.$$

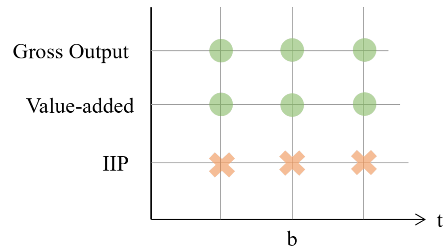
Figure 4: Visual representation of *Method 4*



Method 5. *Method 5 (m5)* is used in the absence of the IIP sequence; see Figure 5. In this case, we use a spatially interpolated IIP to construct an interpolated deflator $\tilde{Def}^{j}_{ist}$ and, eventually, real value added:

$$V^{m5,j}_{ist} = V_{ist} \times \tilde{Def}^{j*}_{ist}, \qquad \tilde{Def}^{j*}_{ist} = \frac{O_{isb}}{O_{ist}} \times \tilde{IIP}^{j}_{ist}. \tag{11}$$

Figure 5: Visual representation of *Method 5*

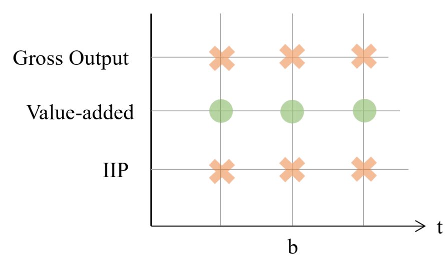


We also generate a set of hybrid forms of Method 4 and Method 5, which slightly deviate from each other.

Method 6. *Method 6 (m6)* addresses the case when only nominal value added is observed for a country/sector combination; see Figure 6. It uses a spatially interpolated deflator:

$$V_{ist}^{m6,j} = V_{ist} \times \tilde{Def}_{ist}^{j}. \tag{12}$$

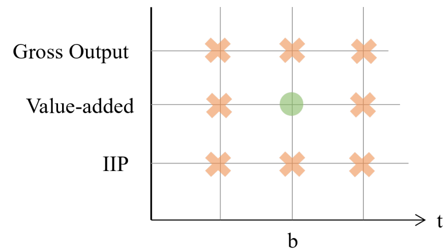
Figure 6: Visual representation of Methods *6*



Method 7. *Method 7 (m7)* addresses the case when the only available data sequence for a sector-country combination is nominal value added for the reference year; see Figure 7. It is the most stringent model as it not only assumes a constant value added-gross-output ratio y but also that the IIP of country i can be recovered from a spatial interpolation process:

$$V_{its}^{m7,j} = V_{isb} \times I\tilde{I}P_{ist}^{j}.$$

Figure 7: Visual representation of *Method 7*



For the sake of clarity, we summarise all discussed models in Table 3, restating their underlying assumptions.

Table 3: Summary methods

Method	Formula	Assumption
m1	$V^{m1}_{ist} = VA_{ist} \times \frac{O_{isb}}{O_{ist}} \times IIP_{ist}$	Single deflation method consistent with **?**.
$m2^j$	$V^{m2,j}_{ist} = nVA_{ist} \times \frac{\tilde{O},j_{isb}}{O_{ist}} \times IIP_{ist}$	The output level in the base year O_{isb} is assumed to adequately be captured through spatial interpolation.
m3	$V^{m3}_{is(t+r)} = y_{it+r} \times O_{isb} \times IIP_{ist}$	The *value-added to output ratio*, $y_t = V_{ist}/O_{ist}$ is assumed to follow a linearly interpolated gap of length $r = 1, \ldots, R$, for which both nominal value-added and growth outputs is not available. The interpolation is based on the last (first) observation before (after) the gap of the respective ratios of nominal value added and output, i.e. y_t and y_{t+R+1}.
m4	$V^{m4}_{ist} = V_{isb} \times IIP_t$	Assumes a constant *value-added to output ratio*, y_t, equivalent to the value of the base year b, i.e. $y_t = y_b \; \forall \; t$.
$m5^j$	$V^{m5,j}_{ist} = V_{ist} \times \frac{O_{isb}}{O_{ist}} \times I\tilde{I}P^j_{ist}$	The IIP is assumed to adequately be captured through spatial interpolation.
$m6^j$	$V^{m6,j}_{ist} = V_{ist} \times D\tilde{e}f^j_{ist}$	The deflator is assumed to adequately be captured through spatial interpolation.
$m7^j$	$V^{m7,j}_{ist} = V_{isb} \times I\tilde{I}P_{ist}.$	A constant ratio of value added and gross output equivalent to the value of the base year b, i.e. $y_t = y_b \; \forall \; t$ and the IIP that is assumed to adequately be captured through spatial interpolation.

Note: Methods ordered in ascending order based on restrictiveness of underlying assumption(s). Spatial interpolation performed using *gross capital formation*, *trade*, and *natural resources*, i.e., $j = \{cf,\ td,\ nr\}$ for co-kriging; see Appendix D.

5.2 Selection rule of extension methods

The different extension methods vary notably in their underlying assumptions and pose a selection challenge. On the one hand, methods based on fewer and less stringent assumptions of the underlying model may guarantee high data quality but add fewer observations. On the other hand, more assumption-based methodologies may be most useful for improving data coverage but may do so at the cost of deluding data quality. To assess this trade-off, we benchmark the various extension methods against $m1$ using the same tools we employed to assess the similarity between m1 and the STAN and WIOD datasets in the previous section.

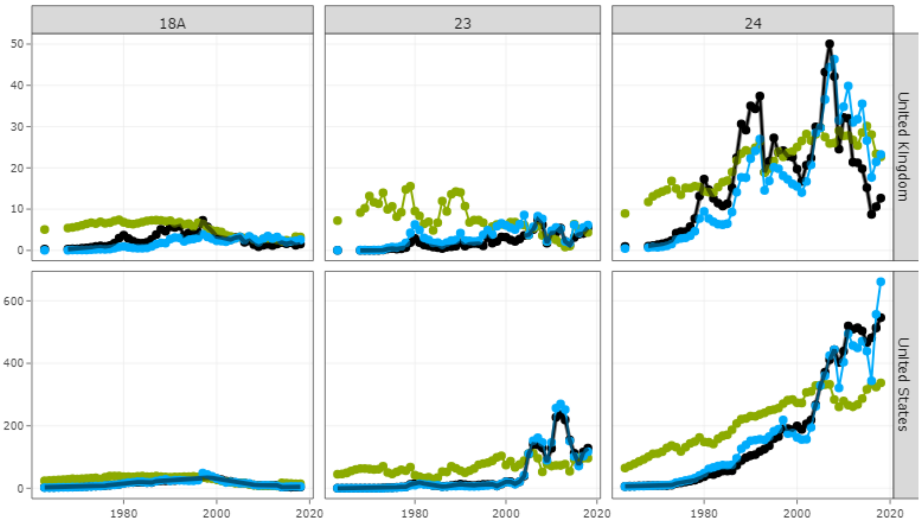
Visual comparison. Figure 8 visualises real value added series based on the different methods for the United Kingdom. As before, we provide access to the various extension methods to the raw dataset through a dynamic online.[19] The results presented in this figure are representative in that the observed patterns seem to follow through for most country-sector combinations. In most cases, there is a very close relationship between $m1$ and $m2$, which is no surprise as these two methods are most closely related. At the same time, some of the less conservative methods allow us to collect significantly more observations over time. They also introduce some more volatility, however. Some initial visual evidence suggests that $m2$ tends to overestimate historic real value added relative to the $m4.5Def$ family, which seems to follow $m1$ much more closely.

Numerical comparison. To arrive at a quantifiable hierarchical order of the various deflation procedures, we resort to Hierarchical Clustering (HC), which allows us to rank-order real value added sequences based on their degree of similarity; see Appendix E for a more detailed description of the analysis. Figure 9 indicates that, on average, the most similar real value added sequences relative to m1 are those that belong to the *m6* family. Here, the series with the interpolated deflator using trade openness performs best in that it is most similar to $m1$ as demonstrated by the smallest centred absolute average distance.

Data coverage across real value added sequences. Table 4 summarises the coverage of the different extension methods indicating how many additional data points they add over *m1*, broken down by region and income groups. There is a notable difference how well the different methods perform regarding data coverage. For example, *m2* does not provide additional data on top of *m1*.

[19] See `https://amannj.shinyapps.io/rVA_Explorer/`, tab `rVA extension methods`, where all country-sector-method can be visualised and compared interactively.

Figure 8: Real value added series based on different methods, selected countries.



Note: Figure extracted from `https://amannj.shinyapps.io/rVA_Explorer/`, tab `rVA extension methods`, for methods *m1* (black), *m2.td* (green) and *m6.td* (blue) as summarised in Table 3.
Source: Authors' elaboration based on INDSTAT (2021).

Table 4: Conditional data coverage

		Additional observations over m1													
	m1	m2.cf	m2.rs	m2.td	m3	m4	m6.cf	m6.rs	m6.td	m7.cf	m7.rs	m7.td	m5.cf	m5.rs	m5.td
Region															
East Asia & Pacific	7797	0	0	0	1010	851	3257	3092	3257	5675	3226	3445	2605	2454	2605
Europe & Central Asia	21599	0	0	0	2477	2870	5376	4720	5376	14810	4720	5376	3005	2634	3005
Latin America & Caribbean	6955	0	0	0	1526	1123	5867	4985	5867	3967	4991	5873	1377	1350	1377
Middle East & North Africa	4677	0	0	0	817	583	5289	4640	5289	4718	4641	5320	2747	2489	2747
North America	1894	0	0	0	107	21	68	68	68	10	68	68	0	0	0
South Asia	993	0	0	0	910	11	2087	1839	2087	15	1839	2091	4	4	4
Sub-Saharan Africa	2847	0	0	0	494	480	6921	6284	6961	5912	6582	7279	3105	2917	3109
Income group															
High	17933	0	0	0	1921	1372	4895	4896	4895	5359	5009	5009	3075	3075	3075
Upper middle	11284	0	0	0	1153	949	3895	3895	3895	3102	4001	4001	1414	1414	1414
Lower middle	8892	0	0	0	1471	1134	7100	7100	7100	5505	7166	7158	3433	3441	3433
Low	2480	0	0	0	513	422	5687	5687	5687	4879	5764	5737	2762	2782	2766

Note: Country classification according to World Bank (2021).
Source: Authors' elaboration based on INDSTAT (2021).

Given the relatively high average absolute distance to *m1*, this indicates that *m2* is inappropriate for extending real value added coverage. Lastly, the *m6* family extends the existing data most notably.[20]

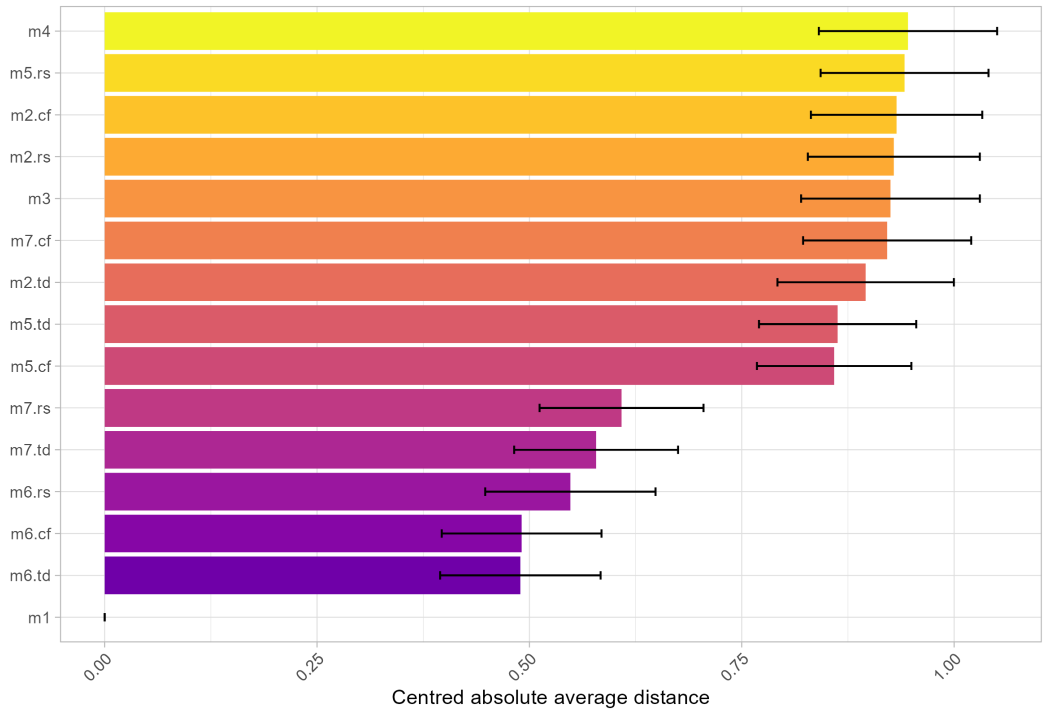
5.3 Final selection rule

Based on this information, we compile the final real value added dataset following the selection procedure below:

1. For a particular country/sector combination, we select *m1* data if this sequence contains obser-

[20]Note that the conditional data generating capacity as highlighted in Table 4 does *not* correspond to the final real value added data frequency, which is the result of the final data selection we introduce in the next section. Rather, it counts the number of observations each method may add to *m1*, when applying the final selection criterion.

Figure 9: Hierarchical clustering of extension methods



Note: Central absolute average distance by real value added sequence of methodologies summarised in Table 3 based on hierachical clustering as described in Appendix E.
Source: Authors' elaboration based on INDSTAT (2021).

vations for at least two periods.

2. If not, we check whether *m6.td* data exist for at least two periods for that particular country-sector combination.

3. If so, we impute the *m6.td* sequence for the particular sector-country combination.

4. If not, no imputations are carried out, and the particular sector-country sequence remains empty.

Using this procedure, we end up with a final dataset, which extensively improves the data coverage of $m1$. This is demonstrated in Table 5, which summarises the number of observations (by region, income group and sector) for which real value added data have become available after employing our methodological extensions vis-a-vis $m1$ only.[21]

[21] Note that the numbers of observations for *m1* in Table 4 and Table 5 do not necessarily correspond to the counts in Table 1. The former two conditions on the final selection criterion (1), that is, we select m1 for a particular country/sector combination if m1 data exist for at least two periods in time for that particular country/sector combination. In turn, Table 1 merely counts the number of observations produced by applying *m1* to the raw INDSTAT data without imposing any further requirements on the individual series. This is because Table 1 attempts to illustrate the superior data coverage

Table 5: Data coverage extended data set

	INDSTAT†			
	m1	*final*	STAN	WIOD⋆
Region				
East Asia & Pacific	7797	10212	1000	450
Europe & Central Asia	21428	25715	7853	2430
Latin America & Caribbean	6937	12159	464	180
Middle East & North Africa	4677	9204	63	90
North America	1894	1957	831	180
South Asia	993	2743	0	90
Sub-Saharan Africa	2847	7780	0	0
Income group				
High	17933	19869	9747	2700
Upper-middle	2480	4612	464	540
Lower-middle	8892	11062	0	180
Low	11284	14728	0	0
ISIC 2-digt sector aggregate				
(15) Food and beverages	2174	2898	457	0
(16) Tobacco	1677	2333	377	0
(17) Textiles	2931	3969	562	0
(18A) Wearing apparel	2857	3990	642	0
(20) Wood products	3113	4297	763	600
(21) Paper and paper products	2854	3850	612	0
(22) Printing and publishing	2715	3616	637	0
(23) Coke, petroleum and nuclear	2022	2776	566	600
(24) Chemicals	2800	3919	673	600
(25) Rubber and plastic	2870	3923	680	600
(26) Non-metallic minerals	3066	4241	808	600
(27) Basic metals	2499	3399	636	0
(28) Fabricated metals	2800	3745	619	0
(29C) Machinery	2775	3696	691	0
(31A) Computer and electronics	2563	3450	474	0
(33) Precision instruments	1551	2073	481	0
(34A) Motor vehicles	2728	3749	788	600
(36) Furniture and n.e.c.	2578	3607	0	0

Note: † based on the proposed procedure: column *m1* - data coverage of *Method 1 (m1)* following the representations in Equation 9 (see also Table 4). column *final* - *m1* plus *m6.td*, following the selection rule in section 5.3. ⋆ ISIC sector classification at more aggregated level; see Socio Economic Accounts documentation (Timmer et al., 2015). Manufacturing 2-digit industry classification according to Table 7. Country classification according to World Bank (2021).
Source: Authors' elaboration based on INDSTAT (2021).

6 Analysing development patterns within manufacturing

The availability of reliable data on real value added for many countries over the years is crucial for measuring a country's performance in relation to others and to the historical patterns of manufacturing development. Furthermore, not only can the level of manufacturing performance be assessed at different income levels, but the speed of development over the years can also be measured and compared across countries. Using real value added data allows countries to accurately assess the performance and progress of specific manufacturing industries. This is in contrast to using indicators like manufacturing value added per capita and manufacturing share in GDP, which are commonly used to assess a country's overall level of manufacturing development and are also used as indicators for SDG 9 (industry). A more detailed analysis of manufacturing performance at a sub-sector level enables policymakers to develop targeted industrial policies that can boost industrialisation and facilitate structural transformation.

6.1 Case study 1: Analysing the level of manufacturing development

Setup. First, to assess the level of industrial development, i.e. whether the industry's value added is high or not, a country needs to have benchmarks for comparison. A benchmark can be a country from the same region or one that shares similar geographic or development characteristics. Apart from cross-country comparisons, it is useful to see the historical average pattern of development because each manufacturing industry follows its own distinct development trajectory. For example, certain labour-intensive industries like textiles and wearing apparel tend to show an inverted U pattern of development. Thus, a decrease in real value added at a high-income level is quite normal and does not point to inferior performance.

The method used to estimate the pattern of industrial development follows the methodological approaches outlined in recent literature on structural change (Haraguchi and Amann, 2021). More specifically, we estimate a panel fixed effects model to analyse the development patterns for $s = 18$ manufacturing industries and focus on ten dominant sub-sectors; see Table 7 for a complete list and description of the manufacturing sub-sectors. In addition to the derived real value added sequences from INDSTAT, our empirical analysis uses real GDP data taken from the Penn World Table database (Feenstra et al., 2015). For each sub-sector s, we run a separate regression which takes the form:

of INDSTAT when employing *m1* alone. In contrast, tables 4 and 5 illustrate the extent to which the methodologies described in this paper can help further extend the coverage of INDSTAT in the most sensible way.

$$y_{it} = \boldsymbol{\beta}\mathbf{X}_{it} + \boldsymbol{\gamma}\mathbf{Z}_{it} + \epsilon_{it}, \tag{13}$$

where y_{ist}, the independent variable, is the log of real value added per capita, which we construct based on the discussion in the previous section. Furthermore, $\mathbf{Z}$ includes a set of controls, including time and country fixed effects for each country and industry. The explanatory variables in $\mathbf{X}_{it}$ contain the logs of real GDP per capita. They are added in their linear, quadratic and cubic representation, i.e. $\{log(rGDP_{it}),\ log(rGDP_{it})^2,\ log(rGDP_{it})^3\}$. Consequently, $\boldsymbol{\beta}_s$ contains the industry-level coefficients for each regression and ϵ_{it} denotes the idiosyncratic error component. We use a fixed-effects (FE) setup to capture endogeneity arising from variations in time-invariant firm attributes, regional and industry-specific market conditions and technology-related developments (Mundlak, 1978; Wooldridge, 2005). To this end, the error term in Equation 13 has the following structure:

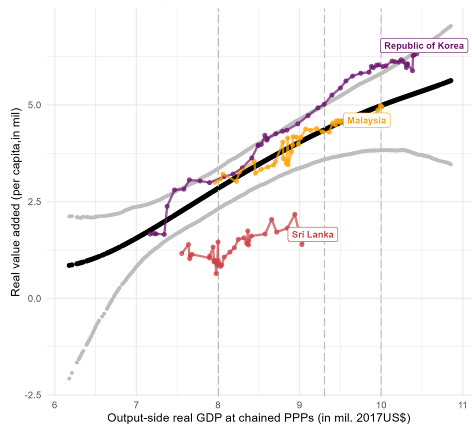
$$\varepsilon_{it} = \alpha_i + D_{st} + e_{it}, \tag{14}$$

where α_i denotes industry-specific intercepts while industry-year interactions (D_{st}) capture industry-specific market conditions and technology-related developments.

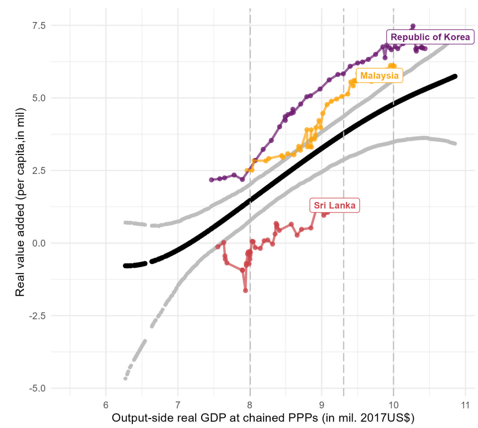
Results. Figure 10 shows the value added per capita levels in a log scale for the Republic of Korea, Malaysia, and Sri Lanka, and their changes across income levels for low-tech, medium-tech, and high-tech industries against the estimated patterns based on Equation 13 for selected sectors.

Overall, all three countries seem to follow the estimated trajectories of development depicted in bold black lines. Both the Republic of Korea and Malaysia have reached an income level high enough to exhibit the full development pattern of so-called "early industries," such as textiles and wearing apparel, which are typically dominant industries at low- and middle-income stages and decline at a high-income level. While the countries tend to move along the patterns, they do so at different levels. The Republic of Korea consistently has the highest value added among the three countries, regardless of income levels. Malaysia falls in the middle, while Sri Lanka generally has lower levels of value added, except in the textiles and wearing apparel industries where Sri Lanka's performance is comparable to, and in some cases superior to, that of Malaysia.

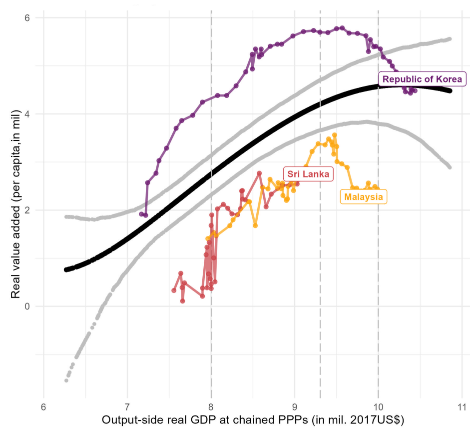
Figure 10: Absolute and relative convergence: a cross-country example



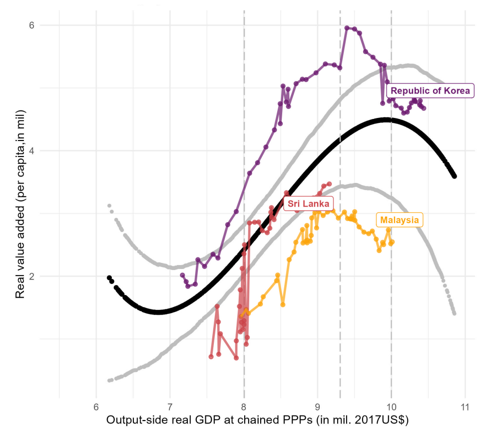
(a) (24) Chemicals



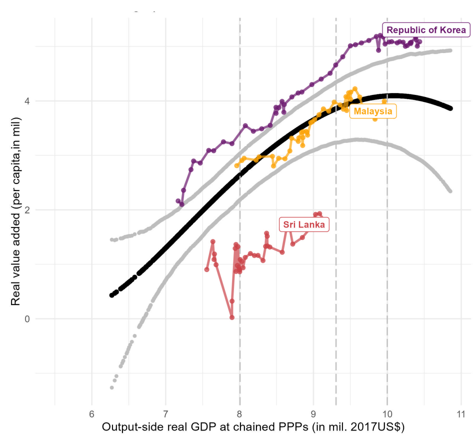
(b) (31A) Computer



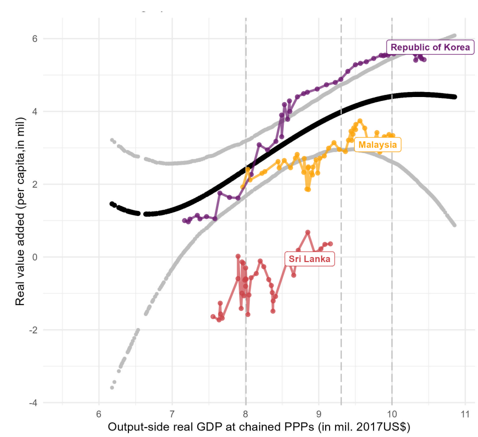
(c) (17) Textiles



(d) (18A) Wearing apparel



(e) (26) Fabricated metals



(f) (27) Basic metals

Note: X- and y-axis in logs. Estimated patterns and confidence intervals based on a subset of large countries (average population over sample period ≥ 18 mil). Residual bootstrap with $R = 1,000$ repetitions. Models include time- and county-fixed effects. *Source:* Authors' elaboration based on INDSTAT (2021).

As demonstrated in this example, utilising real value added data allows for the measurement of not just the present performance of a country's manufacturing industry, but also the comparison of its performance to that of any country in the past, including a successfully industrialised country at the same income level. Furthermore, looking at the development trajectory of an industry, one could assess whether the industry's progress is on or off track relative to that of a selected country or the average development pattern of the industry.

Lastly, the results may also be used for cross-country benchmarking purposes. Applying the framework to the preselected country sample, the Republic of Korea, Malaysia and Sri Lanka tend to follow the expected average patterns of large countries to which they belong in terms of their population size. The Republic of Korea and Malaysia have reached an income level high enough to exhibit the industries' full development patterns. Both countries show an inverted U-curve for low-tech industries (textiles and wearing apparel) and a gradual slowdown for medium-tech industries (non-metallic and basic metals) in line with the estimated patterns. Also, as expected, they exhibit sustained growth for high-tech industries (chemicals and electrical machinery). While the countries tend to move along the patterns, they do so at different levels. The Republic of Korea consistently maintains the highest value added among the three countries across income levels, while Malaysia remains in the middle and Sri Lanka at lower levels.

6.2 Case study 2: Analysing the speed of manufacturing development

Setup. Even if the industrial production of two countries is at a similar level, one could still be seen as having better development prospects if its rate of development is faster. For this purpose, we can measure the speed of manufacturing development, which is expressed as the difference of the respective sector level value added per capita for a country between one GDP per capita level to another divided by the number of years it has taken the respective country to move through the two GDP per capita levels. In the case of the comparisons of the Republic of Korea, Malaysia, and Sri Lanka, the speed of sub-sector development between USD 3,000 and USD 7,500 per capita (where all three countries overlap as seen in Figure 10) can be formally expressed as follows:

$$Speed_{is}^{GDP=\{USD\ 7500,\ USD\ 3000\}} = \frac{\left(V_{ist}^{GDP=USD7500} - V_{ist}^{GDP=USD3000}\right)}{\sum_t \mathbf{1}(GDP_{it} \leq USD\ 7500\ \&\ GDP_{it} \geq USD\ 3000)},$$

where $\mathbf{1}(\cdot)$ is an indicator function which is equal to 1 whenever the GDP level of country i is between USD 3,000 and USD 7,500, and zero otherwise.[22]

Results. The three countries differ not only in terms of their level of manufacturing development at a given income level but also in the speed of their manufacturing development. Table 6 indicates how fast the three countries' industries developed across income levels of between USD 3,000 and USD 7,500. The speed of the manufacturing industry's development, expressed by value added increase per year, was much higher in the Republic of Korea than in the other two countries. In turn, Malaysia's industries grew more quickly than Sri Lanka's, except for food and beverages, textiles and wearing apparel. Countries may have similar development patterns, but their performance differs in terms of the level and speed of manufacturing development at a given income level.

Table 6: Speed of manufacturing development

	Speed of Manufacturing Convergence		
Industry	Republic of Korea	Malaysia	Sri Lanka
(15) Food and beverages	6.88	1.26	1.82
(17) Textiles	11.74	0.51	0.84
(18A) Wearing apparel	10.19	0.77	1.50
(24) Chemicals	4.17	2.36	0.32
(26) Minerals	2.54	0.77	0.21
(27) Basic metals	6.37	0.55	0.10
(28) Fabricated metals	7.72	0.31	0.11
(29C) Machinery	10.23	0.76	0.14
(31A) Electrical machinery	10.97	2.15	0.07
(34A) Motor vehicles	9.11	0.75	0.14

Note: The speed is expressed as the difference of the respective sector level value added per capita figure for a country with a GDP per capita level between USD 3,000 and USD 7,500 divided by the number of years it has taken the respective country to move through this income corridor. Sector classification according to Table 7.
Source: Authors' elaboration based on INDSTAT (2021).

As these examples demonstrates, the availability of real value added data for numerous countries and years improves the ability to benchmark manufacturing performance. First, countries can assess whether the development trajectories of their manufacturing industries are following the expected patterns. Secondly, they can evaluate whether their industries' performance is better or worse than the historical patterns or that of other countries in the same size group and can measure how much better or worse they are faring. Finally, the speed of manufacturing development across countries can be compared to determine how quickly industries are climbing the development curve relative to

[22]On the basis of the real value added sequences, further statistical indicators of industrial performance, such as the ones discussed in UNIDO (2010, chapter 6), could easily be calculated with the goal to expand the analytical reach of industry country diagnostics.

others. While not discussed in this paper, the real value added dataset could also expand productivity analyses for developing countries at the sub-sector level. None of these assessments would be possible if only nominal value added data or real value added data from a few countries were available.

7 Conclusion

Despite the significant role industrialisation plays in economic development, the limited availability of a comprehensive cross-country database for manufacturing real value added at the manufacturing sub-sector level has prevented many developing countries from benchmarking their performance and evaluating their industrial development trajectories. Similarly, academic research on the issue has remained equally affected by these constraints, with most cross-country work on manufacturing development either focusing on a selection of already industrialised high-income countries (Timmer et al., 2015; STAN, 2021), performed at the more aggregated sector level (Palma, 2014), or based on estimation techniques to address the lack of real value added data in large industrial databases (Rodrik, 2016).

In this paper, we proposed a new single deflation method to expand the availability of real value added data at the two-digit ISIC level for UNIDO's Industrial Statistics Database INDSTAT (2021). To our knowledge, this is the first study that seeks to expand data coverage of the real value added of a large group of developing countries at the manufacturing sub-sector level.

We derive a single deflator which is consistent with current national accounting practices and self-contained, i.e., it only requires data available via INDSTAT. We also discuss various extensions to further improve data coverage. We illustrate the much-improved data coverage of our approach by comparing the data coverage of our derived data set with that of the WIOD (Timmer et al., 2015) and STAN (2021). Our method increases data availability by approximately 5 and 18 times relative to the data currently available in the STAN and WIOD databases, respectively.[23]

Our approach enables us to build an extensive dataset to analyse structural change dynamics within the manufacturing sector in unprecedented detail. This significant increase in data availability opens up a new avenue for research on the manufacturing performance of specific countries and on patterns of structural change and manufacturing development. Aside from the examples we provide in this paper, future applications could analyse countries' comparative advantages, country-specific

[23] Comparison of data coverage of the extended dataset (Table 5) with that of STAN and WIOD as broken down in Table 1.

and time-specific effects on industrial development, deindustrialisation tendencies, and changes in the development patterns of industries.

References

Arrow, K., 1962. The Economic Implications of Learning by Doing. American Economic Review 29 (3), 155–173.

Berndt, D. J., Clifford, J., 1994. Using dynamic time warping to find patterns in time series. In: Proceedings of the 3rd international conference on knowledge discovery and data mining. pp. 359–370.

Chakravarty, S., Mitra, A., 2009. Is industry still the engine of growth? An econometric study of the organized sector employment in India. Journal of Policy Modeling 31 (1), 22–35.

Chenery, H., 1960. Patterns of industrial growth. The American Economic Review 50 (4), 624–654.

Chenery, H., Robinson, S., Syrquin, M., Feder, S., 1986. Industrialization and growth. Oxford University Press New York.

Chenery, H. B., Taylor, L., 1968. Development Patterns: Among Countries and Over Time. Source: The Review of Economics and Statistics 50 (4), 391–416.

Dalum, B., Johnson, B., Lundvall, B.-A., 1992. Public Policy in the Learning Society. In: National systems of innovation: Towards a theory of innovation and interactive learning. Pinter Publishers, London, pp. 296–317.

Erumban, A. A., Gouma, R., de Vries, G., de Vries, K., Timmer, M., 2012. Sources for National Supply and Use Table Input files: WIOD.

Eurostat, 2014. Essential SNA - Building the basics. Publications Office of the European Union. URL `https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-14-008`

Feenstra, R. C., Inklaar, R., Timmer, M. P., oct 2015. The Next Generation of the Penn World Table. American Economic Review 105 (10), 3150–3182.

Felipe, J., Mehta, A., 2016. Deindustrialization? a global perspective. Economics Letters 149, 148–151.

Felipe, J., Mehta, A., Rhee, C., 2019. Manufacturing matters... but it's the jobs that count. Cambridge Journal of Economics 43 (1), 139–168.

Gollin, D., Parente, S., Rogerson, R., 2002. The role of agriculture in development. American Economic Review 92 (2), 160–164.

Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-temporal interpolation using gstat. The R Journal 8, 204–218.
URL https://journal.r-project.org/archive/2016/RJ-2016-014/index.html

Haraguchi, N., 2015. Structural change and. In: Routledge Handbook of Industry and Development. Routledge, Ch. Patterns of structural change and manufacturing development, pp. 38–64.

Haraguchi, N., Amann, J., 2020. The importance of manufacturing development and structural change for pro-poor growth. In: Designing Integrated Industrial Policies Volume I. Routledge, pp. 95–132.

Haraguchi, N., Amann, J., 2021. Sectoral dynamics. In: New Perspectives on Structural Change: Causes and Consequences of Structural Change in the Global Economy. Vol. 1. Oxford University Press, pp. 240–269.

Haraguchi, N., Cheng, C. F. C., Smeets, E., 2017. The importance of manufacturing in economic development: Has this changed? World Development 93, 293–315.

Hastie, T., Tibshirani, R., Friedman, J. H., Friedman, J. H., 2009. The elements of statistical learning: data mining, inference, and prediction. Vol. 2. Springer.

Herbel, N., 2014. Methods and Practical Experiences in Compiling the IIP. Regional Seminar on Indices of Industrial Production (IIP).
URL https://www.unido.org/sites/default/files/2014-07/Workshop-Paper_Herbel_EN_0.pdf

Hirschman, A. O., 1958. The strategy of economic development.

ILO et al., 2020. International Labour Office (ILO) and International Monetary Fund (IMF) and Organisation for Economic Co-operation and Development (OECD) and European Union (EU) and United Nations (UN) and World Bank. Consumer price index manual: Concepts and methods.
URL https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/publication/wcms_761444.pdf

IMF, 2017. Measure up: A better way to calculate GDP. International Monetary Fund.
URL https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2017/02/02/Measure-up-A-Better-Way-to-Calculate-GDP-44598

INDSTAT, 2021. INDSTAT 2 based on International Standard Industrial Classification (ISIC), Revision 3. UNIDO Statistics.
URL https://stat.unido.org/database/INDSTAT%202%202020,%20ISIC%20Revision%203

ISIC Rev.3.1, 2002. nternational Standard Industrial Classification of All Economic Activities (ISIC) Revision 3.1.
URL https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/Isic31_English.pdf

Kaldor, N., 1967. Strategic Factors in Economic Development. Ithaca, New York.

Kathuria, V., Natarajan, R., 2013. Is Manufacturing an Engine ofGrowth in India in the Post-Nineties? Journal of South Asian Development 8 (3), 385–408.

Krüger, J. J., 2008. Productivity and structural change: a review of the literature. Journal of Economic Surveys 22 (2), 330–363.

Kuznets, S., 1957. Quantitative aspects of the economic growth of nations: II. industrial distribution of national product and labor force. Economic development and cultural change 5 (S4), 1–111.

Lewis, W. A., 1954. Economic Development with Unlimited Supplies of Labour. The Manchester School 22 (2), 139–191.

Li, J., Kuroko, M., 2016. Single Deflation Bias in Value Added: Verification Using Japanese Real Input–Output Tables (1960–2000). Journal of Economics and Development Studies 4 (1), 16–30.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2022. cluster: Cluster Analysis Basics and Extensions. R package version 2.1.4 — For new features, see the 'Changelog' file (in the package source).
URL https://CRAN.R-project.org/package=cluster

Marconi, N., Reis, C. F. d. B., de Araújo, E. C., 2016. Manufacturing and economic development: The actuality of Kaldor's first and second laws. Structural Change and Economic Dynamics 37, 75–89.

Matsuyama, K., 1992. Agricultural productivity, comparative advantage, and economic growth. Journal of economic theory 58 (2), 317–334.

Matsuyama, K., 2009. Structural change in an interdependent world: A global view of manufacturing decline. Journal of the European Economic Association 7 (2-3), 478–486.

McMillan, M., Rodrik, D., 2011. Globalization, structural change and productivity growth. Making Globalization Socially Sustainable February, 49–85.

McMillan, M., Rodrik, D., Verduzco-Gallo, V., 2014. Globalization, Structural Change, and Productivity Growth, with an Update on Africa. World Development 63, 11–32.

Moretti, E., 2010. Local multipliers. American Economic Review 100 (2), 373–377.

Moretti, E., Thulin, P., 2013. Local multipliers and human capital in the united states and sweden. Industrial and Corporate Change 22 (1), 339–362.

Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica: journal of the Econometric Society, 69–85.

OECD, 2011. ISIC Rev. 3 Technology Intensity Definition. OECD Directorate for Science, Technology and Industry; Economic Analysis and Statistics Division.
URL https://www.oecd.org/sti/ind/48350231.pdf

Palma, J. G., 2014. De-industrialisation,'premature'de-industrialisation and the dutch-disease. Revista NECAT-Revista do Núcleo de Estudos de Economia Catarinense 3 (5), 7–23.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/

Ratanamahatana, C. A., Keogh, E., 2004. Everything you know about dynamic time warping is wrong. In: Third workshop on mining temporal and sequential data. Vol. 32. Citeseer.

Rodrik, D., 2013. Unconditional convergence in manufacturing. Quarterly Journal of Economics 128 (1), 165–204.

Rodrik, D., 2016. Premature deindustrialization. Journal of Economic Growth 21 (1), 1–33.

Sáenz, L. F., 2022. Time-varying capital intensities and the hump-shaped evolution of economic activity in manufacturing. Journal of Macroeconomics 73, 103429.

Sardá-Espinosa, A., 2017. Comparing time-series clustering algorithms in r using the dtwclust package. R package vignette 12, 41.

SNA, 2008. System of National Accounts 2008.

STAN, 2021. STAN STructural ANalysis Database. OECD.
URL https://www.oecd.org/industry/ind/stanstructuralanalysisdatabase.htm

Su, D., Yao, Y., 2017. Manufacturing as the key engine of economic growth for middle-income economies. Journal of the Asia Pacific Economy 22 (1), 47–70.

Syrquin, M., Chenery, H., 1975. Accumulation and Allocation Processes.

Syrquin, M., Chenery, H., 1989. Patterns of Development, 1950 to 1983. World Bank Washington, DC.

Szirmai, A., Verspagen, B., 2015. Manufacturing and economic growth in developing countries, 1950-2005. Structural Change and Economic Dynamics 34, 46–59.

Thirlwall, A. P., 2002. The nature of economic growth: An alternative frame-work for understanding the performance of nations. Edward Elgar Publishing, Cheltenham.

Timmer, M. P., Dietzenbacher, E., Los, B., Stehrer, R., De Vries, G. J., 2015. An illustrated user guide to the world input–output database: the case of global automotive production. Review of International Economics 23 (3), 575–605.

Timmer, M. P., Erumban, A. A., Los, B., Stehrer, R., De Vries, G. J., 2014. Slicing up global value chains. Journal of economic perspectives 28 (2), 99–118.

Tregenna, F., 2009. Characterising deindustrialisation: An analysis of changes in manufacturing employment and output internationally. Cambridge Journal of Economics 33 (3), 433–466.

UNIDO, 2010. Industrial Statistics. Guidelines and Methodology.
URL https://www.unido.org/sites/default/files/2012-07/Industrial%20Statistics%20-%20Guidelines%20and%20Methdology_0.pdf

United Nations, 2008. Country Practices for the Collection and Calculation of the Index of Industrial Production.

UNSD, 2010. International Recommendations for the Index of Industrial Production. United Nations Statistics Division.
URL https://unstats.un.org/unsd/statcom/doc10/bg-industrialstats.pdf

UNSD, 2013. International Recommendations for the Index of Industrial Production.
URL https://unstats.un.org/unsd/industry/docs/f107_edited.pdf

Van Neuss, L., 2018. Globalization and deindustrialization in advanced countries. Structural Change and Economic Dynamics 45, 49–63.

Van Neuss, L., 2019. The drivers of structural change. Journal of Economic Surveys 33 (1), 309–349.

Vu, K., Haraguchi, N., Amann, J., 2021. Deindustrialization in developed countries amid accelerated globalization: Patterns, influencers, and policy insights. Structural Change and Economic Dynamics 59, 454–469.

Wooldridge, J. M., 2005. Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. Review of Economics and Statistics 87 (2), 385–390.

World Bank, 2021. Country and lending groups.
URL https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

Yamada, T., 2016. Methodological recommendations for the compilation of the Index Numbers of Industrial Production (IIP). UN-ESCWA Training Workshop on Short-term Economic Indicators.
URL https://archive.unescwa.org/sites/www.unescwa.org/files/u593/iip.pdf

Appendix

A Bias of the single deflation method

The single deflation method is subject to a bias stemming from the differential between output and intermediate price movements. To find an expression for this bias, consider the single deflator method as described in Equation 3, expand by $(I_t \div D_t^I)$ and find the expression for the double-deflator given in Equation 2:

$$
\begin{aligned}
V_t^{SD} &= \frac{(O_t - I_t)}{D_t^O} \\
&= \frac{O_t}{D_t^O} - \frac{I_t}{D_t^O} \\
&= \left(\frac{O_t}{D_t^O} - \frac{I_t}{D_t^I} \right) + \left(\frac{I_t}{D_t^I} - \frac{I_t}{D_t^O} \right) \\
&= V_t^{DD} + I_t \times \left(\frac{1}{D_t^I} - \frac{1}{D_t^O} \right).
\end{aligned}
\tag{15}
$$

The bias is then given by:

$$
\begin{aligned}
Bias = V_t^{SD} - V_t^{DD} &= I_t \times \left(\frac{1}{D_t^I} - \frac{1}{D_t^O} \right) \\
&= I_t \times \left(\frac{D_t^O - D_t^I}{D_t^I \times D_t^O} \right).
\end{aligned}
\tag{16}
$$

This result is identical to the representation in IMF (2017) and indicates that the single deflator method is structurally biased unless $(D_t^O - D_t^I) = 0$, that is, the intermediate input and output price movements are identical. Conversely, if $D_t^O > D_t^I$, $(D_t^O < D_t^I)$, the single deflation method is upward-(downward-)biased.

B Visual value-added comparison for the G7

Figure 11: Database comparison Canada

Figure 12: Database comparison France

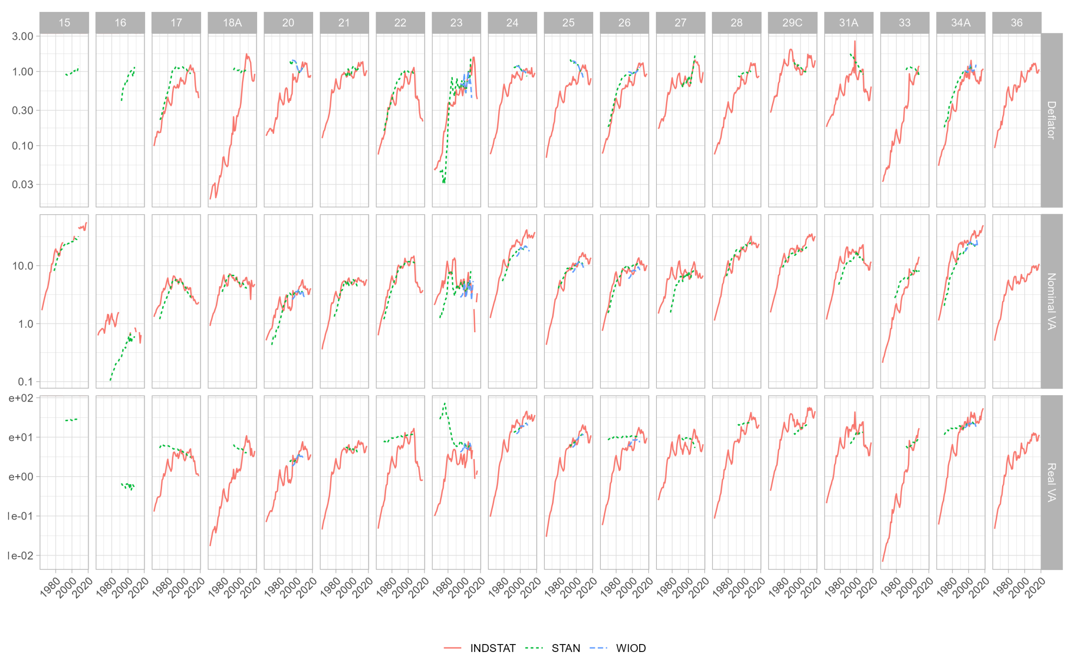


Figure 13: Data base comparison Germany

Figure 14: Data base comparison Japan

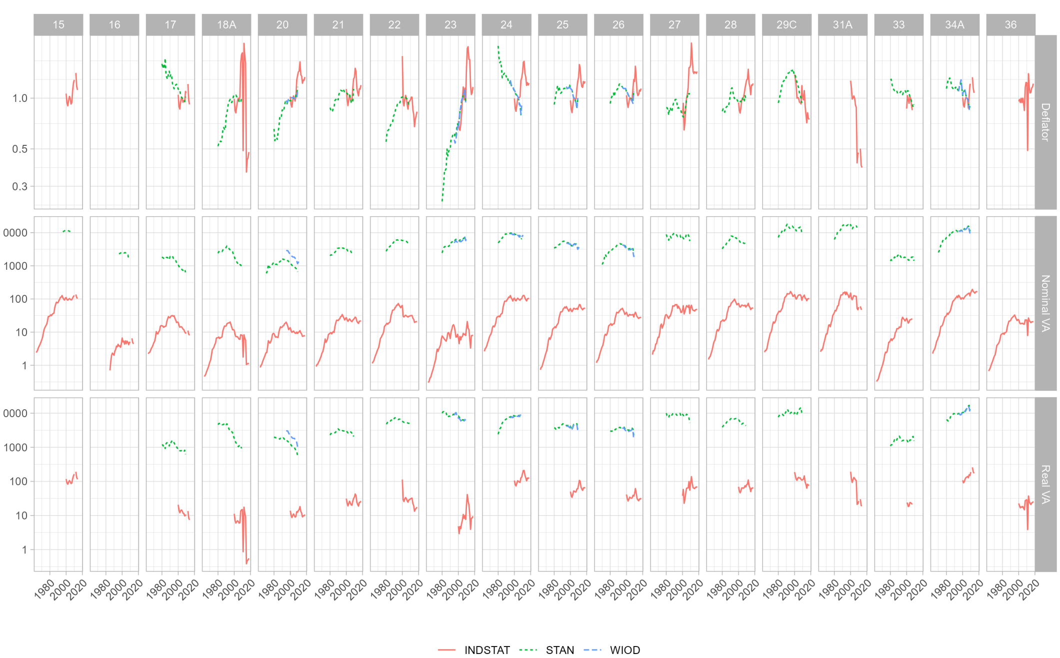


Figure 15: Data base comparison United Kingdom

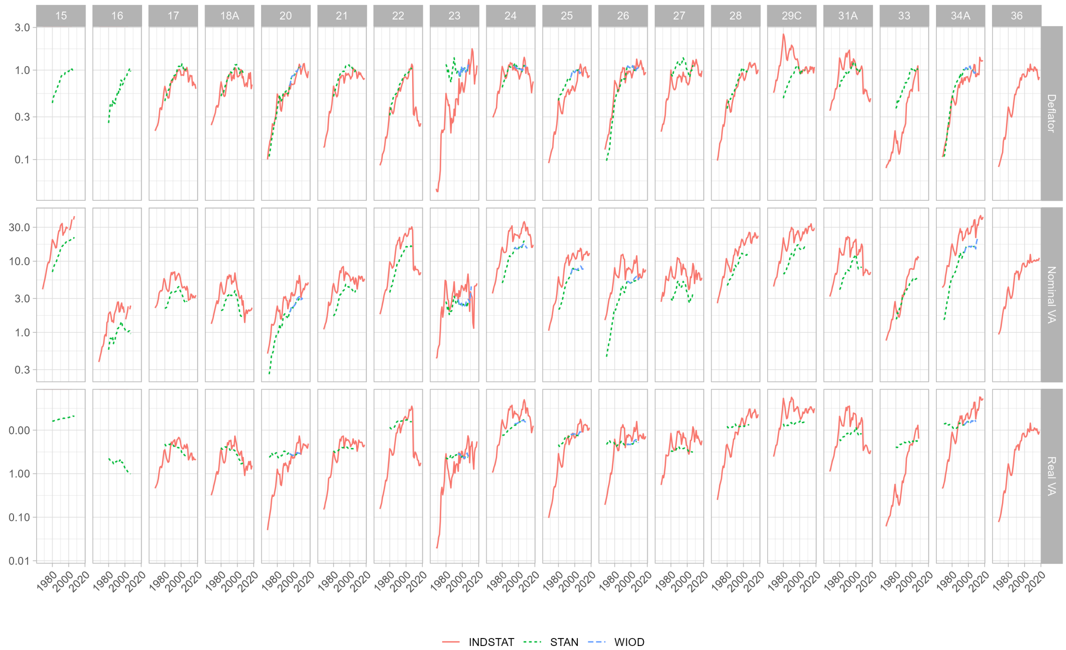
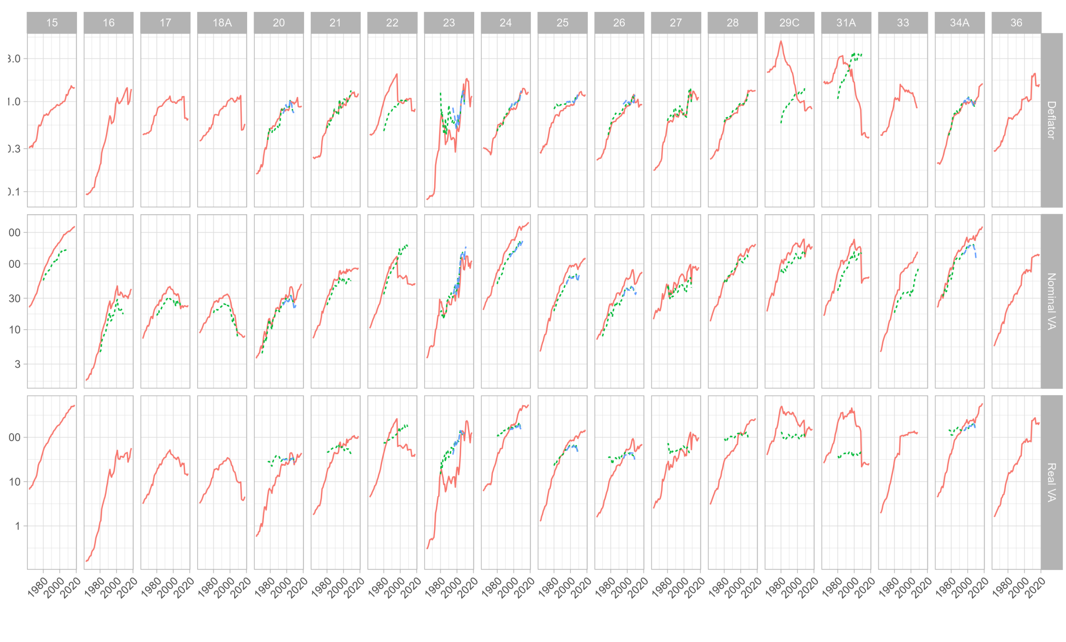


Figure 16: Data base comparison United States



C Similarity analysis using Dynamic Time Warping (DTW)

We employ Dynamic Time Warping (DTW) to compare two separate time series; see Ratanamahatana and Keogh (2004) and Berndt and Clifford (1994) as well as Sardá-Espinosa (2017) for software implementation in R (R Core Team, 2021). We resort to DTW to compare the time series of nominal and real value-added and output as well as the single deflation method proposed in Equation 9 vis-a-vis the corresponding data sequences contained in STAN and WIOD. Following the notation in Sardá-Espinosa (2017), in a first step, DTW creates a local cost matrix (LCM) for every pair of series, x, and y, we want to compare. The set of combinations is given by:

$$\{(x^d, y^d)\} = \{deflator^d, nVA^d, rVA^d\}$$

with $\{d\} \in \{INDSTAT, STAN, WIOD\}$, For each input pair, (i, j) the LCM obtains the l_p norm between x_i and y_i is calculated as:

$$LCM(i,j) = \left(\sum_v |x_i^v - y_j^v|^p \right)^{1/p}.$$

Next, the DTW algorithm finds the path that minimises the alignment between (x, y) by iterating through the LCM sequences, where we define any combination of (i, j) as $\phi = \{(1, 1), \ldots (n, m)\}$, and n (m) corresponds to the length of variable x (y). The final distance is given by:

$$DTW_p(x,y) = \left(\sum \frac{m_\phi LCM(k)^p}{M_\phi} \right), \quad \forall\ k \in \phi.$$

Since we are not interested in the absolute DTWs but the performance of m1 vis-a-vis the respective STAN and WIOD series, we calculate the normalized dynamic time warping distance as

$$nDTW_p(x,y)^d := \frac{DTW_p(x^d, y^d)}{DTW_p(x^{STAN}, y^{WIOD})}.$$

Note that for $d = WN$, we perform the time series validation on a random white noise (WN) process, which we compare with each of the individual series and data sets. The normalised dynamic time warping distance (nDTW) for $d = \{STAN, WIOT\}$ is always going to be equal to 1. Conversely, a high (low) nDTW indicates that the compared pair of sequences is more dissimilar (similar) to the pair $d = \{STAN, WIOT\}$.

D Spatial interpolation

Application. This section describes the spatial interpolation technique used in methods $m2$, $m5$, $m6$ and $m7$. They all employ spatial interpolation techniques of variations of the IIP/deflator. Below, we provide a general discussion of the interpolation.

Procedure. We want to produce a spatially interpolated counterpart of an arbitrary series x for country i contained in an arbitrary data set of I countries, $i \in I$, at time t and sector s, denoted x_{ist}. We define its interpolated version $\tilde{x}_{ist}$. This counterpart is calculated as the weighted average of a set of proxy countries I', with arbitrary member i', which is an proxy for i such that $i' \in I' \in I$ and $i \neq i'$, as $\tilde{x}_{ist} = \sum_{i' \in I, i \neq i} w_{i'} x_{i'st}$. We find the weights $w_{i'}$ as follows:

1. For all countries in I, we obtain the spatial distance between its capital and all other capitals of countries as well as the corresponding income group classification from World Bank (2021).

2. For each country $i \in I$, we set the following criteria for any country to become a possible proxy country i', $i \neq i'$:

   (a) The capital of i' must lie within a certain radius (of 5,000 km) of that of i.

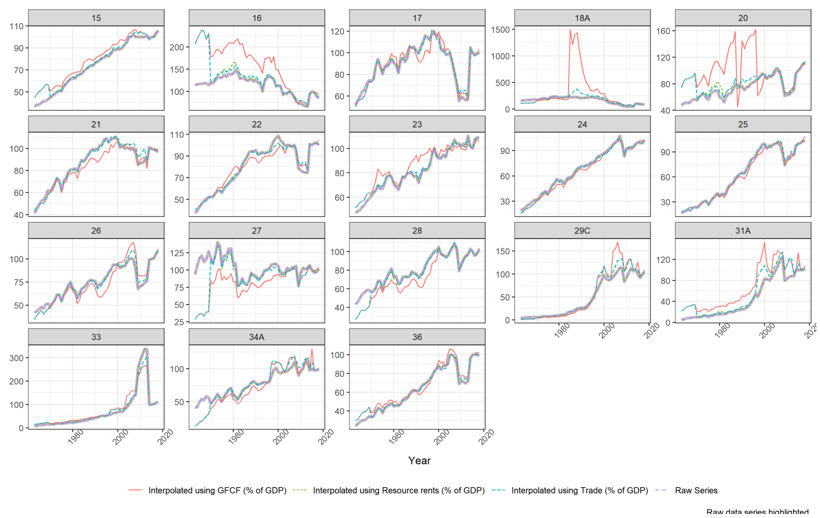
   (b) There cannot be more than a maximum of 20 members of I'_i for each i. In other words, we restrict the analysis to the closest 20 most common economies.

   (c) All potential proxies must share the same income group in the same year.

   The rationale behind (c) is that potential proxy countries should be selected based on the current development pattern of the economy and not on its final stage of development. To ensure this, we rely on the historical World Bank income group classification (World Bank, 2021), which provides time series between 1987 and 2019 for the Analytical Classification (as presented in World Development Indicators) as well as more extensive classifications following the Bank Operational Lending Categories, which extend back to 1970. Using both classifications, it is possible to provide sensible income group thresholds up to 1970. For a more extensive discussion on the World Bank historic income group classification used in this paper, see Appendix F.

3. Given this selection criterion, a list of proxy countries I' for each country i is obtained. Note that at this stage, the only selection criteria are those listed in (a) (b) and (c) above.

Figure 17: Interpolated IIPs for different economic indicators United States



4. Next, we evaluate whether any potential proxy country i' reports the series $x_{i'st}$, which we would like to interpolate for the same year and sector. For the case of $m2$, this would constitute the output level in the base year O_{isb} and the deflator (IIP) and for methods $m4$ and $m7$ ($m6$), respectively. See Table 3 for a summary of the various extension methods.

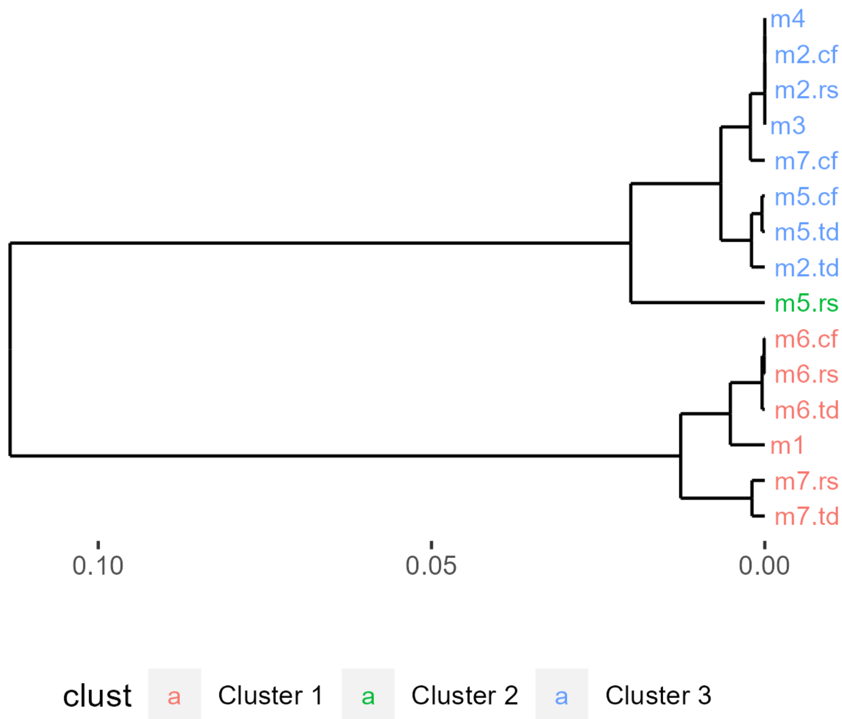
5. For all countries i' reporting $x_{i'st}$, we then employ co-kriging (CK) using the R (R Core Team, 2021) package `gstat` (Gräler et al., 2016) by estimating:

$$\tilde{x}_i(u) = \sum_{\alpha_1=1}^{n1(u)} w_{\alpha_1} Z_1(u_{\alpha_1}) - \sum_{\alpha_2=1}^{n(u)} w_{\alpha_2} \left[Z_2(u_{\alpha_2} - m_2 + m1]\right),$$

where m denotes the means and w the weights of the primary and secondary variables, $\{1, 2\}$ at location (u) and n the number of measured values of all pre-selected proxy countries in I' used for the estimation of the neighbourhood of u. We use both spatial data locations (coordinates) as well as a set of economic variables used in the spatial interpolation to account for 'economic similarities'. For the economic series, we use data from the WDI for *gross capital formation*, *trade* and *natural resources* as a percentage of total GDP. We weigh each possible proxy country from the above list by its spatial proximity and 'economic similarity' to the target country.

Figure 17 illustrates the interpolated IIPs for the three different economic indicators, *gross capital formation*, *trade* and *natural resources*, for the United States, including the initial raw IIP series (Raw series). We provide a complete set of visualisations online at `http://u.pc.cd/9Pm`.

Figure 18: Dendrogram USA, sector 15



E Hierarchical clustering of deflators

We employ Hierarchical Clustering (HC) (Hastie et al., 2009) to create a hierarchy of ordered sequences to obtain clusters of time sequences evaluated by their degree of similarity using the `agnes` function available in the R (R Core Team, 2021) `cluster` (Maechler et al., 2022) package.

Figure 18 is a visual representation of the created hierarchy for sector 31 in the United States. In the visualisation, the height of each node is proportional to the value of the inter-group dissimilarity between its two daughter nodes. In this particular case, $m1$ shares the highest degree of similarity with the $m6$ family. In a similar vein, we can identify a second prominent group containing $m2$, $m3$ and $m5$ deflators, and which are notably dissimilar to $m1$ for this particular country- and sector configuration. Figure 18 only serves expositional purposes. To rank order the respective real value-added sequences based on their similarity relative to $m1$. For any country-sector combination, we record the aggregate inter-group dissimilarities (height) across sectors and countries to arrive at an aggregate hierarchical clustering displayed in Figure 9.

F Extension of World Bank Historic Income Groups

The historic World Bank income group classification (World Bank, 2021) provides time series between 1987 and 2019 for the Analytical Classification (as presented in World Development Indicators) as well as more extensive classifications following the Bank Operational Lending Categories, which extend back to 1970 as illustrated in Figure 19a. We illustrate the respective threshold cut-offs in Figure 19a. The upper-middle to high (UM) income group threshold is only available from the late 1980s onwards. This is a noteworthy limitation, as it prevents us from correctly mapping the transition from upper-middle to high income countries for around half of our data sample. To extend the UM threshold cut-off to earlier periods, we calculate the *threshold ratio* between the upper-middle to high (UM), lower-middle to upper-middle (LM) and low to lower-middle (L) income group threshold cut-offs in period t as:

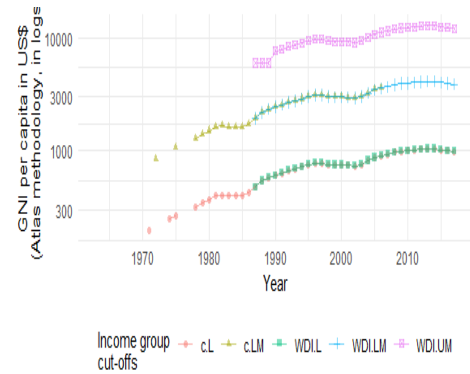
$$Threshold\ ratio_t = UM_t/(LM_t - L_t) \tag{17}$$

The *threshold ratio* remains very stable over time and close to a value of 4.14 as seen in Figure 19b. We identify previous threshold values for UM by setting the $threshold\ ratio = 4.14$ and solving for the missing upper-middle threshold in Equation 17:

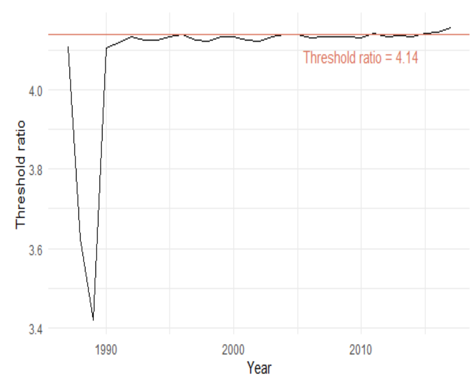
$$UM_t^* = (LM_t - L_t) \times 4.14. \tag{18}$$

We also test an alternative approach by back-casting the high-income threshold by approximating the series using its ARIMA representation and the values for the two remaining threshold values as additional explanatory variables (also known as ARIMAX). A simple ARIMAX(1,0,0) model can be written as $z_t = \alpha + \phi z_{t-1} + \theta\epsilon_{t-1} + \gamma x_t + \epsilon_t$, where, x_t represents the exogenous variable with θ containing the respective coefficients. The remaining model specification was identified by employing and testing various auto-regressive models, moving average representations of the initial data process, and selecting the one with minimal AIC and/or SIC. In this case, an auto-regressive model of order one has been selected. In Figure 20, we compare the UM threshold cut-off for the pre-1987 period obtained by employing the constant threshold ratio and the ARIMAX estimation. The difference between the

Figure 19: World Bank historic income group thresholds cut-offs



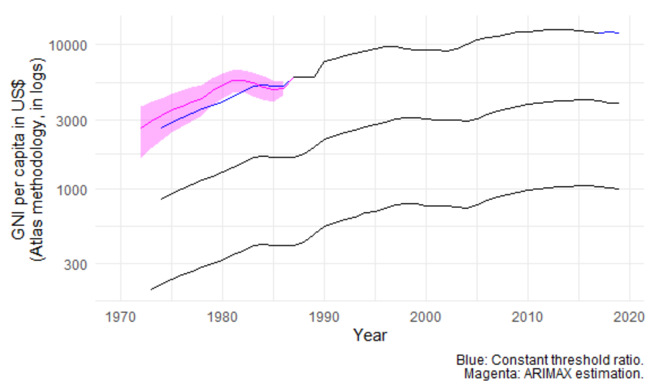
(a) Income threshold cut-offs over time



(b) Thresholds ratio over time

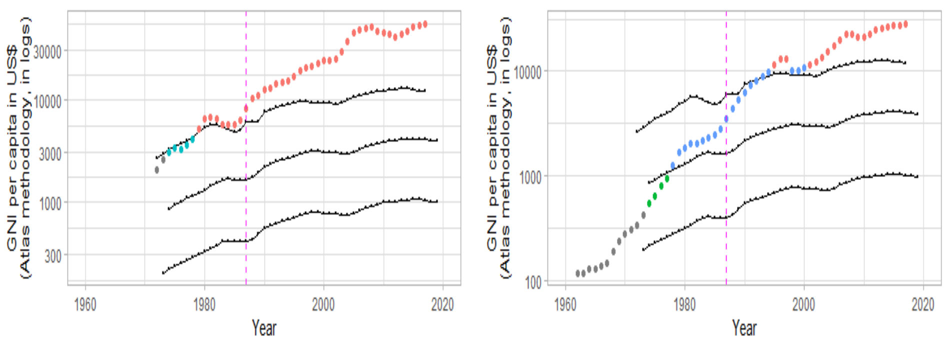
In Figure 19a, variables 'c.L' and 'c.LM' correspond to Bank Operational Lending Categories thresholds for low and lower-middle income countries. Variables 'WDI.L', 'WDI.LM' and 'WDI.UM' correspond to Analytical Classification thresholds for low to lower-middle (L), lower-middle to upper-middle (LM) and upper-middle to high (UM) income group cut-offs (World Bank, 2021). The threshold ratio in Figure 19b follows the definition in Equation 17.

Figure 20: Extended historic World Bank income group threshold ratio



two measures remains moderate. We use the constant threshold ratio cut-off to determine the UM threshold cut-offs for the pre-1987 period. For illustrative purposes, Figure 21 presents selected country examples where the imputed UM threshold is used for classification left of the dashed line.

Figure 21: Examples of onterpolated IIP series, selected economies



(a) Ireland

(b) Republic of Korea

Income group: Low (L), lower-middle (LM) and upper-middle (UM) and high (H) income according to World Bank (2021) (right of dashed line) and extended income group cut-offs based on Equation 18 (left of dashed line). NA: No world Bank income group classification available.

G Data classification and aggregation

Classification of economic activity The industry sector level classification in this paper follows the *International Standard Industrial Classification* (ISIC). The ISIC combinations chosen for this report are presented in Table 7. Regarding the technology classification of industries, all manufacturing industries are further classified by their technology intensity following the technology classification of the *Organization for Economic Co-operation and Development* (OECD), which is based on Research and Development (R&D) intensity relative to value-added and gross production statistics (OECD, 2011).

Aggregation to ISIC Rev. 3 combinations While many countries report manufacturing data according to the ISIC industry aggregation in Table 7, the majority of countries report manufacturing data in INDSTAT at the level of individual industries. A simple summation across industries can obtain data to arrive at a consistent ISIC industry aggregate for nominal value-added and output. For example, for industry *(18A) wearing apparel*, this implies a summation of value-added and output for sectors *(18) Manufacture of wearing apparel; dressing and dyeing* and *(19) Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear*, respectively. To obtain the corresponding IIP series for an arbitrary industry s derived from aggregating a set of arbitrary set of raw industries $\mathfrak{s}, \mathfrak{s} \in s$, we compute:

$$IIP_{ist} = \sum_{\mathfrak{s} \in s} \varphi_{i\mathfrak{s}t} \times IIP_{i\mathfrak{s}t}, \quad \varphi_{i\mathfrak{s}t} = \frac{V_{i\mathfrak{s}t}}{\sum_{\mathfrak{s} \in s} V_{i\mathfrak{s}t}}$$

where $V_{i\mathfrak{s}t}$ corresponds to the manufacturing sector-level value-added of industry $\mathfrak{s}$, which is to be aggregated to industry s. We use the same weighting approach to aggregate ISIC industry aggregates for the STAN data.

Table 7: Manufacturing industry classification

ISIC Industry Aggregation		ISIC Industry Classification			Data Set Coverage		
(Code) Abbreviation	Rev. 3 Combination	Rev. 3 Code	Rev. 3 Industry Description	Technology Group	INDSTAT	STAN	WIOT
(15) Food and beverages	15	15	Manufacture of food products and beverages	Low	15	15	15t16**
(16) Tobacco	16	16	Manufacture of tobacco products	Low	16	16	15t16**
(17) Textiles	17	17	Manufacture of textiles	Low	17	17	17t18**
(18A) Wearing apparel	18 + 19	18	Manufacture of wearing apparel; dressing and dyeing of fur	Low	18A	18*	17t18**
(18A) Wearing apparel	18 + 19	19	Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear	Low	18A	19*	19
(20) Wood products	20	20	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	Low	20	20	20
(21) Paper and paper products	21	21	Manufacture of paper and paper products	Low	21	21	21t22**
(22) Printing and publishing	22	22	Publishing, printing and reproduction of recorded media	Low	22	22	21t22**
(23) Coke, petroleum and nuclear	23	23	Manufacture of coke, refined petroleum products and nuclear fuel	Medium	23	23	23
(24) Chemicals	24	24	Manufacture of chemicals and chemical products	High	24	24	24
(25) Rubber and plastic	25	25	Manufacture of rubber and plastics products	Medium	25	25	25
(26) Non-metallic minerals	26	26	Manufacture of other non-metallic mineral products	Medium	26	26	26
(27) Basic metals	27	27	Manufacture of basic metals	Medium	27	27	27t28**
(28) Fabricated metals	28	28	Manufacture of fabricated metal products, except machinery and equipment	Medium	28	28	27t28**
(29C) Machinery	29 + 30	29	Manufacture of machinery and equipment n.e.c.	High	29C	29*	29
(29C) Machinery	29 + 30	30	Manufacture of office, accounting and computing machinery	High	29C	30*	30t33**
(31A) Computer and electronics	31 + 32	31	Manufacture of electrical machinery and apparatus n.e.c.	High	31A	31*	30t33**
(31A) Computer and electronics	31 + 32	32	Manufacture of radio, television and communication equipment and apparatus	High	31A	32*	30t33**
(33) Precision instruments	33	33	Manufacture of medical, precision and optical instruments, watches and clocks	High	33	33	30t33**
(34A) Motor vehicles	34 + 35	34	Manufacture of motor vehicles, trailers and semi-trailers	High	34A	34*	34A
(34A) Motor vehicles	34 + 35	35	Manufacture of other transport equipment	High	34A	35*	34A
(36) Furniture and n.e.c.	36	36	Manufacture of furniture; manufacturing n.e.c.	Low	36	36	36t37**

* ISIC Rev. Industry aggregated to designated Rev. 3 combination.

** Higher level of aggregation not mapped onto respective ISIC Rev. 3 combinations. ISIC Rev. industry (37) Recycling not considered. Technology classification based on OECD (2011).

Note: ISIC Rev. 3 technology group classification according to OECD (2011).

Data coverage: INDSTAT (2021) (1963-2018), STAN (2021) (1970-2009), Timmer et al. (2015) (1995-2011)

UNITED NATIONS
INDUSTRIAL DEVELOPMENT ORGANIZATION

Vienna International Centre · P.O. Box 300 9 · 1400 Vienna · Austria
Tel.: (+43-1) 26026-0 · E-mail: info@unido.org
www.unido.org