MODERNIZING PRODUCTS AND SERVICES OF INDUSTRIAL STATISTICS

IN THE CONTEXT OF THE
SUSTAINABLE DEVELOPMENT GOALS

MODERNIZING PRODUCTS AND SERVICES OF INDUSTRIAL STATISTICS

IN THE CONTEXT OF THE SUSTAINABLE DEVELOPMENT GOALS

Vienna 2021

UNIDO ID/448
Sales Number: E.18.II.B.48
ISBN: 978-92-1-106455-1
eISBN: 978-92-1-362857-7

ACKNOWLEDGEMENTS

This publication has been prepared by Valentin Todorov from the Statistics Division at the United Nations Industrial Development Organization (UNIDO) and the UNIDO consultant Ascha Pedersen. The graphical layout and cover design were done by Dina Dragoshinska. Most of the papers in the publication were first presented at the workshop „Modernizing products and services of Industrial statistics in the context of SDG“ conducted 14-16 November 2016 in Vienna, and we are very thankful to the authors who updated these papers with the most recent data available. Two of the chapters (Chapter 5 and Chapter 6) were written specifically for this publication.

Special thanks go to the colleagues and consultants from the Statistics Division who provided excellent support organizing and conducting the seminar: Rita Liang, Jürgen Muth, Raphaela Wirth, Kateryna Gumeniuk, Liliya Kirchberger, Regina McFerren, as well as to the UNIDO Chief Statistician Fernando Cantu Bazaldua for his continuous support.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

FIGURES

TABLES

INTRODUCTION

In response to the rapid globalization in every socio-economic domain, monitoring of empirics on industrial development and industry-related SDGs require specialized and detailed economic, industrial statistics as an indispensable information basis. International statistics for demographic, social, economic, and environmental areas that are collected and disseminated by international data sources have been increasingly demanded for both cross-country and country-specific empirical analysis. Viable strategies, policies, and programmes for sustainable industrial development cannot be formulated unless they are well prepared based on such statistics and analysis. At the same time, the use of the Internet to disseminate those statistical data has resulted in increased user expectations concerning data availability, comparability and timeliness.

This publication sheds light on the dedicated efforts to achieve these ambitions through a compilation of papers that on the one hand exemplifies data collection and processing methods and approaches to the modernization hereof. Some papers, on the other hand, offer economic analyses of national industrial statistics, underlining the importance of data quality. The papers were part of the outcome of a workshop with the same title as this publication and dedicated to the 50th anniversary of the United Nations Industrial Development Organization (UNIDO) in November 2016.

Chapters 2 and 3 features country-specific case studies with a focus on the methods in place for compiling, maintaining and processing industrial statistics at the national statistical offices in Angola and the Republic of Korea. The papers cover, respectively, the construction of an industrial production index while complying with international standards, and the use of administrative survey data to enhance the statistical quality.

Chapters 4 and 5 examine means to advance existing methodologies. A paper from the Trier University reviews the potential benefits of composite indicators as a benchmarking method and analytical tool for the development of country performances according to relevant quantitative sub-indicators - specifically the SDGs in this context. UNIDO focuses on the benefits of big data as a viable solution for missing data treatment; it does so from a general viewpoint but also with particular attention to improving data quality in INDSTAT, a comprehensive international industrial statistical database maintained by the organization.

Finally, Chapters 6 and 7 demonstrate how industrial statistics can be applied to analyse industrial developmental progress. While national statistics is used to give elaborate insights on the status of the Indian food processing industries, UNIDO's Competitive Industrial Performance database is used to examine Kenya's industrial competitiveness compared to immediate/ potential competitors.

2

COMPILING INDUSTRIAL STATISTICS IN THE REPUBLIC OF KOREA USING ADMINISTRATIVE DATA

By Kim Honghee, Statistics Korea, Republic of Korea

2.1 INTRODUCTION

Statistics Korea sources administrative data from government departments to develop statistics and provide statistical services. New legislation in 2007 and 2009 gave central and local governments the authority to request administrative data from public agencies if required to produce official statistics. Since then, the accumulation of administrative data has increased gradually. The need for administrative data, i.e., records collected by public or private organizations to give overviews on registration, transactions, and record-keeping, arose because worsening survey conditions were a burden to respondents. This was the result of having introduced new types of statistics without accompanying them with appropriate integration methods. In contrast, administrative data is easier to procure and is likely to result in higher statistical quality and a reduced budget.

This paper gives a brief overview of the agency's efforts to secure administrative data. Section 2.1 covers the overall processes of using such data, from collecting it to utilizing it for statistics. The Mining and Manufacturing Survey in Section 2.2 is an example of a statistical product stemming from this process. Section 2.3 details the expected outcomes of using administrative data over time. The fourth and final section describes three of Statistics Korea's primary statistics services.

2.2 CURRENT ADMINISTRATIVE DATA USAGE

List of Administrative Data and Information Items

As of 30 September 2016, the Administrative Data Records Database sourced 148 datasets from 63 agencies, which were used to produce and impute 87 types of statistics, and verify and supplement additional 61 statistics.

Administrative Data Processing Flow

Figure 2.1 shows Statistics Korea's current administrative data processing flow, which consists of three stages: Data collection, the establishment of an original administrative database, and a database for statistical uses. The figure provides an overview of the core factors and the specific content of each of these stages.

2.3 FUTURE PLANS FOR ADMINISTRATIVE DATA

Statistics Korea will continue to develop and advance its surveys and administrative data, some of which are depicted in Figure 2.2. The key objective is that the three-stage process shown in Figure 2.1 is maintained and that the symmetrical relation between survey and administrative data is continuously advanced (Figure 2.2).

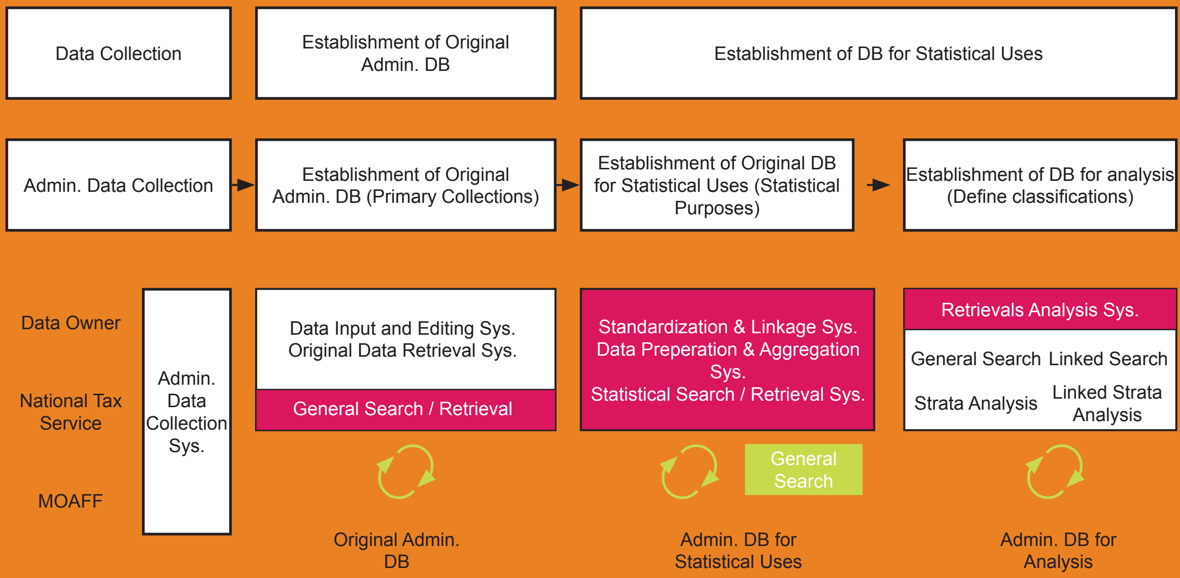

FIG. 2.1 | ADMINISTRATIVE DATA PROCESSING FLOW

SOURCE | STATISTICS KOREA 2016. COMPILING INDUSTRIAL STATISTICS IN KOREA USING ADMINISTRATIVE DATA. INTERNATIONAL WORKSHOP ON MODERNIZING PRODUCTS AND SERVICES OF INDUSTRIAL STATISTICS IN CONTEXT OF SDG, VIENNA, NOVEMBER.

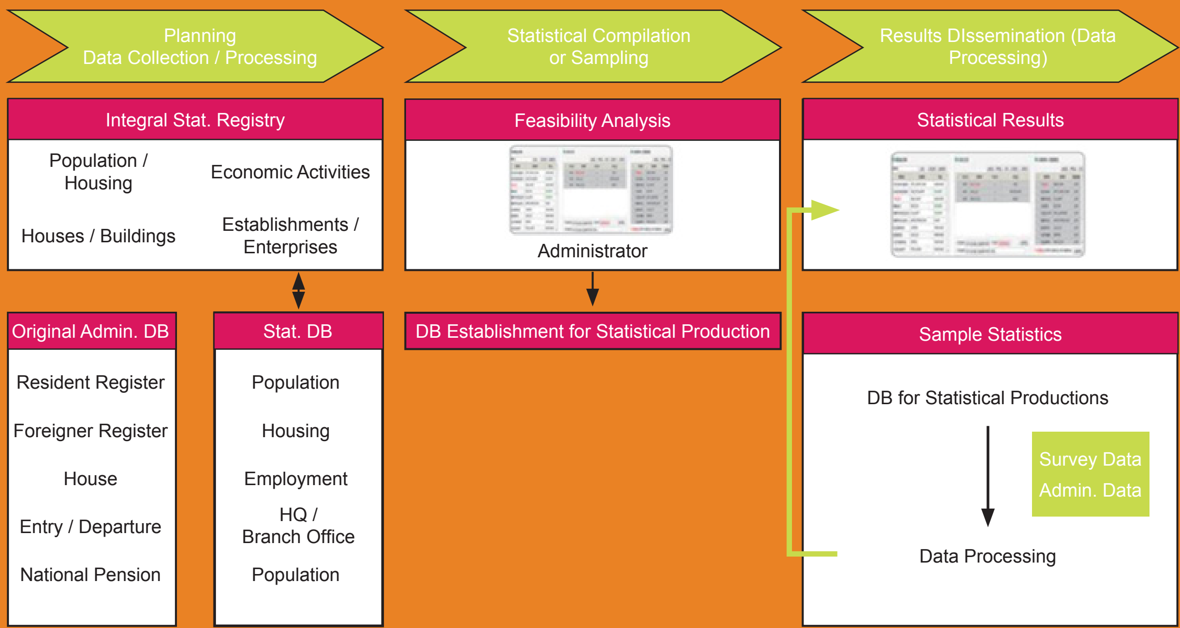

FIG. 2.2 | FUTURE PLANS FOR THE INTEGRAL STATISTICS REGISTRY

SOURCE | SEE FIGURE 2.1.

TABLE 2.1 | SURVEY UNIT AND FREQUENCY

Survey title	Korean Standard Industrial Classification	Survey unit	Survey frequency
Agriculture, Forestry and Fishery Survey	A	Household	Annual[1)]
Mining and Manufacturing Survey	B, C	Establishment	Annual[2)]
Wholesale and Retail Trade Survey	G, I		
Professional, Scientific and Technical Service Industry Survey	M		
Service Industry Survey	E, J, L, N, P, Q, R, S		
Construction Industry Survey	F	Enterprise	Annual
Transportation Survey	H		

NOTES | 1) EXCEPT IN THE YEAR OF AGRICULTURE, FORESTRY AND FISHERY CENSUS, 2) EXCEPT IN THE YEAR OF ECONOMIC CENSUS.
SOURCE | SEE FIGURE 2.1.

TABLE 2.2 | PERFORMANCE OF SURVEY UNITS IMPUTATION / REPLACEMENT

2014			2015		
Target Population (number of establishments)	Replaced Units (number of establishments)	Replacement Ratio (percent)	Target Population (number of establishments)	Replaced Units (number of establishments)	Replacement Ratio (percent)
71,484	**49,676**	**69.5**	**74,386**	**52,559**	**70.7**

SOURCE | SEE FIGURE 2.1.

This will be achieved by establishing the Integral Statistics Registry for Administrative Data and developing a management system for the registry. It will cover key activities such as domain-specific database establishments and functional advancements (e.g., a register-based economic census, a register-based census of the agricultural sector, big data usage). These advancements will establish statistics registries by domain and improve the connections between them, setting up at the same time a test database for statistical production. Furthermore, it will improve the register-based census and expand the Statistics Registries for time-series, longitudinal, and cross-sectional analyses.

2.4 EXAMPLES OF USAGE

Table 2.1 exemplifies Statistics Korea's use of administrative data. Applying the Korean standard industrial classification (KSIC), the agency examines the number of establishments in the Mining and Manufacturing Survey around June every year.

For this purpose, administrative data has mainly been sourced through the National Tax Service. Figure 2.3 illustrates how the relationship between the survey and the administrative data eventually will produce statistics.

Table 2.2 shows the result of using both survey data and administrative data. It presents survey units of businesses in the mining industry and manufacturing industry. It shows that the replacement ratio (of survey units with imputed values) increased from 69.5 percent in 2014 to 70.7 percent in 2015. This is because the use of administrative data increased the number of replaced units and thus increased the replacement ratio. The figures reflect that Statistics Korea mainly used administrative data in these years.

Statistics Korea will enforce the verification of survey data (revenue, business performance items, tangible assets,

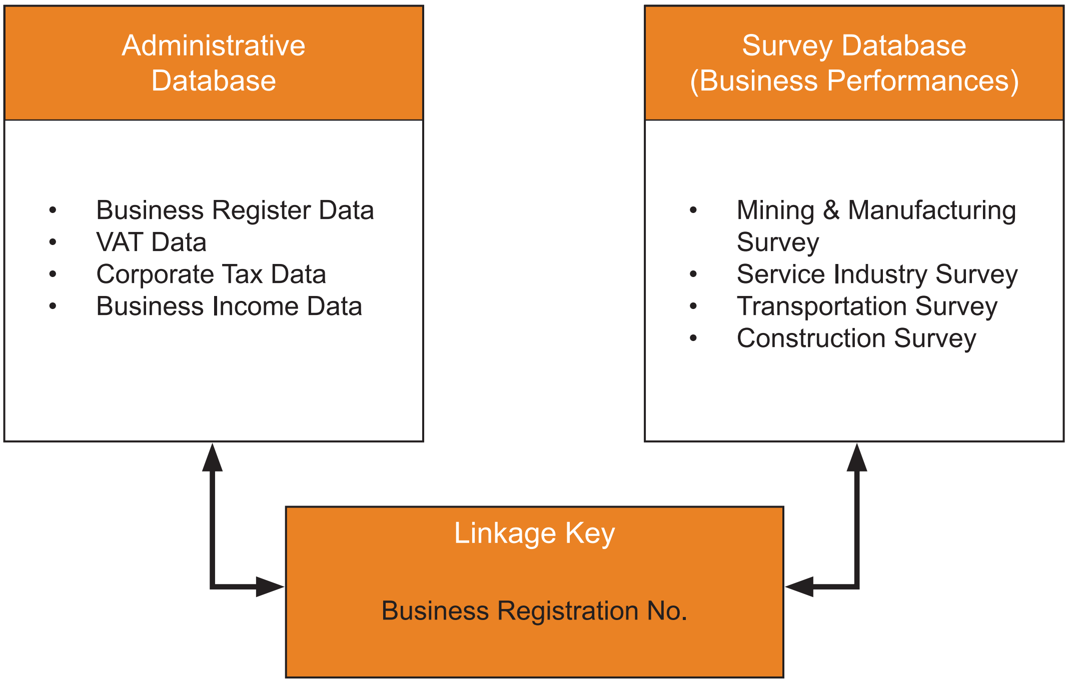

FIG. 2.3 | FUTURE PLANS FOR THE INTEGRAL STATISTICS REGISTRY

SOURCE | SEE FIGURE 2.1.

inventories, etc.) and the imputation of non-response items (business performance items) to set up and advance data usage. Moreover, the agency plans to expand the use of administrative data for the mining and manufacturing survey. In the period 2017-2018, Statistics Korea compiled Trial Statistics on International Trade using 'Import and Export Data by Commodity' produced by Korea Customs Service (KCS). As a result, the statistics 'Trade by Enterprise Characteristics' has been published since 2019. The agency is also currently reviewing the direct imputation and replacement of business performance items for individual establishments and single-corporation establishments.

2.5 EXPECTED OUTCOMES

a. Reduce the burden on respondents and improve usability

- Reduce the burden on respondents by replacing survey items with administrative data;
- Reduce the burden on interviewers and respondents by streamlining surveys, i.e., conduct the surveys at branch offices and not at headquarters;
- Deliver information tailored to specific areas of the economy (i.e., self-employment statistics, franchise statistics) and ensure data timeliness (i.e., re-cycling reference years).

b. Achieve survey efficiency and reduce expenditure

- Complete imputation of 42.3 percent of the business performance items, common items, and asset items;
- No need for separate work on determining enumeration areas;
- Hiring workers at no cost by cooperating with partners.

c. Procure basi draws a map of the economy of the Republic of Korea

- Procuring basic data for the Business Register, which will serve as a sample frame for economic statistics surveys. The Business Register combines survey data and administrative data matched and governed by ‘Business Identification Numbers’;
- Preparing the Introduction of Register-based Economic Census: i)Verify establishments if listed only once in either of the survey data or administrative data, ii) prepare for a transition from a survey-based economic census to a register-based (i.e., administrative data) economic census for the next round of censuses.

2.6 EXAMPLES OF OFFICIAL STATISTICAL SERVICES IN THE REPUBLIC OF KOREA

a. The website of the Korean Statistical Information Service (KOSIS), www.kosis.kr, is the national statistics portal of the Republic of Korea. As of 2015, the KOSIS website featured 700 different statistical services conducted by 200 approved official statistics producing agencies.

b. The E-National Indicators Service, http://www.index.go.kr/ (Korean only), offers 735 key economic and industrial indicators - including employment, industrial production, price indicators, and GDP - to aid the understanding of the current economic situation of the Republic of Korea. It provides detailed information on the meaning of each indicator and relevant policies. Information is carefully communicated to support policymakers in formulating better policies and encourage citizens' awareness.

c. The Statistical Geospatial Information Service (S-GIS), https://sgis.kostat.go.kr/, provides digital map-based small-area statistics (enumeration districts, census tracts) and useful information divided into classifications such as individuals, businesses, government, etc.

3

METHODOLOGICAL SUMMARY AND STEPS TAKEN TO BUILD THE INDUSTRIAL PRODUCTION INDEX IN ANGOLA

By Adão Sebastião Fernando, Instituto Nacional de Estatistica (INE), Angola

3.1 INTRODUCTION

Since 2006, the National Statistics Office of Angola (INE) has released the quarterly publications ‘A Quick Information Sheet on Indices of Industrial Production’, ‘Indices of Working Hours,’ and ‘Indices of Persons Engaged.’ The first series of indices covered the period 2002-2010. It was necessary to update the weights and change the reference year from 2002 to 2010 due to the dynamics of the industrial sector structure in Angola. In the same period, the national Classification of Economic Activities (CAE)—which is harmonized with ‘NACE,’ the industry-standard classification system used in the European Union—was revised, as was the National Classification of Goods and Services. This paper describes the steps taken to build the Industrial Production Index in Angola between 2002 to 2016.

3.2 KEY ELEMENTS FOR BUILDING THE INDUSTRIAL PRODUCTION INDEX IN ANGOLA

General Methodology

An Industrial Production Index was previously published in Angola (1997-2000) based on survey data collected from 57 establishments. From 2006 and onwards, the index has been based on a sample frame of establishments and products that optimally represents the industry's total population. These are selected annually from the Register of Establishments and Product Register, respectively. Until 2015, the sample consisted, on average, of 428 units representing industrial activities in sections C, D, and E of the revised CAE. For various reasons, 100 establishments were not included in the surveys, meaning that the average number of survey establishments averaged 316.

As of 2015, the sample had grown to an average of about 550 establishments with a quarterly response rate of 10-15 percent. It now also represented industrial activities in section B of the second revision of the CAE (in accordance with NACE, Revision 2). The results, which are published quarterly, 45 days after the end of the quarter, are provisional. Definitive results are reviewed and published at the beginning of each year.

Data Collection

The calculated indices are based on the information collected monthly by surveyed establishments, who, in addition to the units produced, report on the number of employees and how many days and hours they have worked, as well as the periodic turnover, sales value, and prices of goods produced. This information is then used to produce and calculate aggregates of the relevant indicators for each industrial activity.

In addition to the monthly questionnaires, data is also collected from the official bodies that comprise the volume production of the entire population of the following industrial activities: Extraction of Oil and Natural Gas, Diamonds, and Electricity. The data collection process (paper-based questionnaires) takes place 15 days after the end of each month. It is conducted by the personnel assigned to the INE unit responsible for the survey and the calculation of the respective indicators.

Calculating the Industrial Production Index

First of all, an elemental index is estimated by the Laspeyres method (volume) using the previous year as the reference period. The basis for calculating the indices is the quantities of the products reported or the hours worked. In some cases, the value of sales of goods produced at fixed prices is used. In case of a total non-response or only a partial response, the corresponding variables are allocated.

Secondly, the influence of prices is eliminated to obtain a reference to the joint evolution of quantity and value. The value of production is estimated for each sampled establishment, taking into account all its products. This is done by multiplying the reported quantities with the product's price in the reference period (base price). The value of production at fixed prices is divided by the average corresponding to the previous year, meaning that the previous year is equal to 100.

The elementary indices of each establishment are then, thirdly, aggregated by calculating the elementary indices by industrial activity. When doing so, the survey coverage is taken into account, and the aggregates are corrected accordingly, hereby adjusting for non-response. The sales value of the goods produced in the reference year is used as a weight for aggregating each establishment‘s indices of the same activity. Indices, calculated based on the hours worked, are corrected with the productivity factor.

Finally, once the elementary indices have been estimated (by industrial activity), they are aggregated at the 2-digit level of the CAE using the corresponding industry's value-added. The published indices are chained for the old series (2002-2014, with 2002 as the reference year) and the new series (2015-onwards, with 2010 as the reference year). The chained elementary indices are calculated quarterly, with the average of the indices corresponding to the previous year's chains.

Furthermore, each year - before calculating the Industrial Production Index - the register is updated, and the sampling frame is set up. The following steps are taken in every quarter of the year:

- Update all variables of the Register of Establishments with the information received during the previous year;
- Update the Product Register with the information received in the previous year;
- Establish the reference file for the calculation of the indexes for the whole year (quantity produced, hours worked, turnover, and prices);
- Establish new weights of the total population at the level of activity;
- Establish productivity and calendar correction factors;
- Exclude establishments in the survey that ended their activity in the previous year but include establishments detected in the previous year that may also be relevant to the industry to which they belong.

3.3 THE STEPS IN DETAIL

Sampling and Coverage

The sample, i.e., the annually selected units from the Register of Establishments and Product Register, considers all the information necessary for the processing and elaboration of the data. It is updated annually and takes into account the change in the structure of turnover and the number of employees of each establishment in each industrial activity in the year preceding the survey.

The sampled establishments are the manufacturing units in each of the industrial activities of sections C, D, and E to 5 digits of the CAE (Revision 1) inspected monthly throughout the national territory. The sample covers at least 80 percent of the volume and 70 percent of employees across all industrial activities.

Collection and Treatment

The data of the respective establishments are reclaimed using the paper questionnaires in a direct interview. In case of non-availability of the information in providing the information to the INE agent, the questionnaire will remain the informant for a maximum of 5 days and then return it to the INE duly filled in. The collected data are reviewed in a standard unit of measure if necessary and typed in the data entry platform (MS Access). This produces several files.

Calculation of Indicators

Depending on data availability, the following alternatives are available for production volume and value to calculate the Industrial Production Index for each sampled establishment:

- Main production indicator: The value of production in a quarter will be sourced from the previous year when the quantities produced;
- Second production indicator: Hours worked in a quarter;
- Third production indicator: Turnover value.

The following formula is used to calculate the elementary indices of each establishment within its activity:

$$I_k^{i,S} = \frac{\sum_{j=1}^{n} p_0^{j,i} q_k^{j,i}}{\sum_{j=1}^{n} p_0^{j,i} \overline{q}_{t-1}^{j,i}}$$

Where

$I_k^{i,S}$ - Is the elementary index (t-1 = 100) of establishment "i" in quarter "k," based on the value of production at fixed prices (eventually hours worked or turnover).

$p_0^{j,i}$ - Price of product "j" of establishment "i" in the reference year.

$q_k^{j,i}$ - Quantities produced in quarter "k" of product "j" in establishment "i"

$\overline{q}_{t-1}^{j,i}$ - The average quantity of product "j" produced in establishment "i" in the reference period (t-1).

The index for activity S (CAE - 5 digits) represented in the sample with a total of M establishments is calculated based on the following formula:

$$I_k^{S} = \frac{\sum_{i=1}^{M} I_k^{i} VV_0^{i}}{\sum_{i=1}^{M} VV_0^{i}}$$

Where

I_k^{S} - Is the elementary index (t-1=100) of activity S in quarter k, based on the value of production at fixed prices (possibly hours worked or turnover).

I_k^{i} - Is the index of the establishment i, i = 1, 2,........, M.

VV_0^{i} - Turnover value of establishment i in the reference period (t-1).

Imputation
In the treatment of non-compliance and extreme values, the following methods are used:

- Missing data are allocated using the average of the previous quarter;
- Missing data are allocated using the average of the establishments within the same activity;
- The hours worked are used in case of extreme values in the production or some activities of difficult measurement of production such as metallic structures, manufacture of footwear and dressing rooms, services related oil extraction and printing and publishing.

3.4 CONCLUSION

The procedures for calculating the production indicators described in this paper have summarized the steps undertaken by INE processes to publish the quarterly Industrial Production Index for Angola - from the process of generating the sample from the Register of Establishments to the calculation procedures of production indicators. INE was supported in developing these procedures by Statistics Norway in 2006. According to the 'International Recommendations for the Index of Industrial Production 2010,' they have been updated over time as recommended by the United Nations Statistics Division.

The information on industrial statistics produced and disseminated on the industry's structure and dynamics will guide the Angolan government in implementing the policies outlined for Angola's development plan for the diversification of the economy. These include building resilient industrial infrastructure (i.e., Industrial Economic Zones) with a focus to facilitate innovation in underdeveloped sectors to promote inclusive and sustainable industrialization. Other policies aim to reduce the volume of imports to boost sustainable economic growth, and to promote full employment, increased productivity, and decent work for all. The indicators for Angola‘s manufacturing industry may enable policymakers to select specific economic policy instruments according to a strategy that contributes to the diversification and sustainable development of economic activities.

4

QUALITY ASSESSMENT OF COMPOSITE INDICATORS FROM THE SDG FRAMEWORK

By Laura Güdemann and Ralf Münnich, Trier University, Germany

4.1 INTRODUCTION

Composite indicators are a common method used for analyzing the development of country performances according to relevant quantitative sub-indicators, especially in economic and social policy support. They benefit from their superficial ease of interpretation and their ability to summarize complex multidimensional issues. In the course of the Agenda 2030 of Sustainable Development, 17 Sustainable Development Goals (SDGs) with 169 specific targets have been adopted by world leaders as the main basis for the post-2015 intergovernmental process (see Sustainable Development Solutions Framework 2015, p. 5). Each of the 17 SDGs can be represented as a composite indicator since they are constructed by more than one target. Additionally, it is possible to build a composite indicator aggregating the target variables over all 17 SDGs to measure the overall achievement of a country in terms of sustainable development. Such an SDG index has been proposed by the Bertelsmann Stiftung and is reported in the Sustainable Development Report published together with the Sustainable Development Solutions Network (see Sachs et al. 2019, p. 19). Their common use in highly sensitive and important areas urges the need to discuss methodical advantages and disadvantages of using composite indicators as well as their adequacy for performance comparisons. Methods of relative comparisons used for policymaking need to follow high standards regarding the data quality and interpretability of the results. They must be applied carefully to avoid misinterpretation and false conclusions.

This paper aims to showcase the construction of a composite indicator from the SDG framework and to analyze potential causes of quality problems concerning the construction process and its interpretability. In the framework of a composite indicator, variables from different sources and surveys are combined into one single index, often without taking into account the data quality of the single sub-indicators or the data gathering process. This can lead to misinterpretations and false conclusions.

Section 4.2 describes the general aspects and problems arising from the use and the construction process of composite indicators, particularly in connection with the SDG framework. In Section 4.3, the explained issues are put in the context of quality principles that several national and international institutions have proposed to draw attention to the importance of data quality for official statistics. One specific quality issue of official data is accuracy which includes sampling and non-sampling errors. The latter concerns, for example, the existence of missing values in datasets. Missingness and the choice of imputation methods can greatly influence the results of composite indicators - this is further showcased in Section 4.4. Additional sources of uncertainty from the construction process are analyzed in Section 4.5 with an uncertainty analysis. Section 4.6 concludes.

4.2 GENERAL ASPECTS AND THE CONSTRUCTION PROCESS OF COMPOSITE INDICATORS

In general, a composite indicator can be understood as a summary of well-chosen and relevant sub-indicators which are combined or aggregated into a single number using a function to represent a multidimensional construct. Usually, the k sub-indicators (k = 1,...,K) used to measure the multidimensional construct do not have the same measurement scale and, therefore, will be normalized to allow for comparability of the outcomes (see Nhemachena et al. 2018, p. 3). The composite indicator is then evaluated for different regions or countries c (c = 1,...,C) and possibly in different time periods t (t = 1,...,T). Therefore, the composite indicator can be expressed as the function f with

$$CI_{c,t} = f_{c,t}(x^*_{1,c,t}, x^*_{2,c,t}, \ldots, x^*_{K,c,t}) \qquad \textbf{(1)}$$

Where

$x^*_{K,c,t}$ - denotes the value of the k-th normalized sub-indicator for country c at time period t

In the following sections, index t will be omitted since cross-sectional data from the same time period is used for all analyses. The construction of composite indicators encompasses multiple stages in which subjective decisions have to be made. OECD and JRC European Commission (2008) provide an overview of some of the stages and a comprehensive description of different methods within each stage. These are:

a. Understanding and definition of the theoretical framework of the composite indicator;
b. Choice of sub-indicator (variables) and selection of data;

c. Imputation of missing data;
d. Checking the underlying data structure with multivariate analysis;
e. Normalization of the sub-indicators;
f. Weighting and aggregation;
g. Uncertainty and sensitivity analysis;
h. Analyzing the country performances on the sub-indicator level;
i. Checking for links of the composite indicator to other relevant measures;
j. Visualization of the results.

The selection of the single sub-indicators or sub-indicator variables should ideally be made based on considerations about their relevance, analytical soundness, timeliness, and accessibility and with respect to the definition of the theoretical construct to be measured (see OECD and JRC European Commission 2008, p. 23). After choosing the sub-indicators needed for constructing the composite indicator, it can be necessary to impute in case of missing data. The question of which imputation method should be used cannot be answered in general terms but depends on the data structure, the missing value scheme, and the relation structure between variables in the dataset. An overview of single and multiple imputation methods in connection with data for composite indicators can be found, for example, in Münnich et al. (2008a) or OECD and JRC European Commission (2008). Depending on which data is available, it can be distinguished between imputation of micro- or macro-level data. Micro-level data encompasses the data of the individual survey units such as firms in a business survey.

On the other hand, macro-level data refers to data of the calculated indices for country c, such as the manufacturing value-added share in GDP. Section 4.5 will demonstrate how influential the choice of one imputation method can be regarding the variability of the resulting composite indicator values. In this section, the imputation will be done on the macro-level of the sub-indicators. One reason for missing data can be due to the reporting time of the single sub-indicator values, which might be after the composite indicator results are needed. In this case, nowcasting methods can help close the gap due to time delays in the reporting see, for example, Boudt et al. (2009).

Multivariate analysis might be necessary to detect the sub-indicators‘ underlying structure and make informed decisions about later steps of the construction process, such as choosing aggregation and weighting methods (see OECD and JRC European Commission 2008, p. 63ff.). After imputing missing records and possibly the multivariate analysis, the values of the sub-indicators have to be normalized if the measurement units of the variables differ.

There are construction approaches for composite indicators, for example, the benefit of the doubt approach, for which this step is not necessary. Choices of normalization methods can be, for example, standardization, min-max normalization, or distance to reference measures (see OECD and JRC European Commission 2008, p. 27ff. and p. 92). Another choice in the construction process concerns the form of the aggregation function f. The normalized data of the sub-indicators can be aggregated using, for example, linear aggregation methods or forms with which the performances of the sub-indicators cannot fully counterbalance each other, like geometric aggregations (see OECD and JRC European Commission 2008, p. 32f.).

The influence of these choices can be measured using a variance-based sensitivity approach or visualized with the uncertainty analysis explained, for example, in Saisana, Saltelli, and Tarantola (2005) or Saltelli et al. (2008). In Section 4.6, the uncertainty analysis is outlined for a composite indicator of sustainable economic development. The composite indicator and its construction using data from the SDG framework is explained in Section 4.4. To present composite indicator results adequately, it is necessary to also present the country performances on the sub-indicator level, for example, in a dashboard. Additionally, for some applications and to improve policymaking, it might be valuable to check the correlations of the composite indicator with other relevant measures (see OECD and JRC European Commission 2008, p. 20f. and p. 132ff.).

Saisana et al. (2005) and OECD and JRC European Commission (2008) summarize the most important advantages and disadvantages of composite indicators, which exceeds the points mentioned below. Due to the aggregation of several sub-indicators into one number, they can help make multi-dimensional issues more understandable and therefore support policymakers to base their decisions on relevant data. This may, in turn, foster discussion about important topics because the selection of sub-indicators can be subject to political discussion. On the downside, reducing complex problems to a single number may lead to a false impression of simplicity and oversimplified policy conclusion. Since the influences of the single dimensions are not reflected anymore in the value of the composite indicator, it can be challenging to determine which policy actions and policy fields should be supported by actions. This could especially be a problem in the case of a composite indicator based on the SDG framework since the topics of the single goals are very diverse (see OECD and JRC European Commission 2008, p. 13 and Saisana et al. 2005, p. 307f.). Therefore, it could be beneficial to evaluate every sub-indicator‘s influence on the value of the composite indicator.

Besides this, in OECD and JRC European Commission (2008), it is explained that composite indicators can be calculated for consecutive years and hence utilized to assess countries‘ development. This statement has to be discussed in further research and needs evaluation of additional data. Since the values of the sub-indicators are estimated from a sample, changes in the sub-indicators and the aggregated composite indicators could be only due to sampling errors. Therefore, an increase in the value of the composite indicator does not necessarily mean a significant improvement of the countries performance from one point in time to the other (see Münnich and Zins 2011, p. 26ff.). Further evaluation of the data is needed to draw conclusions about the changes in a country's performance. Information about the concerning quality of the sample estimates is required to perform this evaluation. This leads to the following sections in which data quality problems and their impact on the results of a composite indicator based on the SDG framework are discussed and illustrated in more detail.

4.3 DATA QUALITY FOR OFFICIAL STATISTICS

As mentioned before, the simplification of different variables into a composite indicator as one summarising number can become a critical approach, especially if the data quality of the sub-indicators is considerably different amongst the sub-indicators used to build the composite indicator. The issue concerning the need for high-quality official statistical information is summarized in principles that describe the requirements of data quality and concepts of critical steps in the statistical production process. Different institutions propose a framework of several principles to ensure the quality of official statistics and to depict critical steps in the statistical production process. Examples for these frameworks and principles are

- United Nations Fundamental Principles of Official Statistics (United Nations 2015)
- European Statistics Code of Practice (Eurostat 2018)
- African Charter on Statistics from the African Union (African Union 2009)
- Statistics Canada Quality Guidelines (Statistics Canada 2019)
- South Africa Statistical Quality Assessment Framework (Statistics South Africa 2010)
- Statistics Norway's Dissemination Policy (Statistics Norway 2007).

The 10 UN Fundamental Principles encompass the most important aspects of data quality assurance in the data production process. In United Nations (2015), these principles are explained together with implementation guidelines and connected with principles from the frameworks mentioned above. Here, it will be focused on the UN framework since it is vital for constructing a composite indicator using the SDG framework. The UN framework encompasses the aspects of data quality in the following principles:

a. Relevance, Impartiality, and Equal Access
b. Professional Standards, Scientific Principles, and Professional Ethics
c. Accountability and Transparency
d. Prevention of Misuse
e. Sources of Official Statistics
f. Confidentiality
g. Legislation
h. National Coordination
i. Use of International Standards
j. International Cooperation

Out of the 10 Fundamental Principles, three will be discussed in more detail in conjunction with composite indicators. Principle 3 Accountability and Transparency requires the users of the statistics to gain access to information necessary to understand their characteristics and qualities (see United Nations 2015, p. 31f.). This includes information about the survey design, frame, response rates, editing methods, measurement errors, and other methods or procedures used in the data production process.

The information about the quality should encompass both sampling and non-sampling errors to cover the total survey error. Sampling errors arise due to the selection mode of a sample, whereas non-sampling errors originate from mistakes or system deficiencies in the data collection process. Sources for non-sampling errors can be the respondent, the interviewer, refusals to participate, which will result in non-response, or mistakes in the data-entry process (see Biemer and Lyberg 2003, p. 36f.).

It is critical to report the information on the quality assessment because this will increase the transparency of the production process and thus increase the trust and acceptance of the statistical outcome. These two aspects from the user side of official statistics are essential for composite indicators because the chosen sub-indicators are often a result of political discussion. Rosen (1991) points out that to be a helpful tool for discussion and monitoring, composite indicators must gain peer acceptance. Transparency on this note will increase their acceptance.

This information is also necessary to facilitate correct interpretations of the composite indicator results and judge its suitability to represent a specific multidimensional construct (see United Nations 2015, p. 31f.). Commonly used methods for the quality assessment of construction decisions are variance-based sensitivity and uncertainty analysis, as mentioned before. It is also possible to evaluate the quality with unit-by-unit plausibility checks, unit-by-unit checks with previous records, outlier detection, or a comparison of the data at hand with other sources (see Christiansan and Tortora 1995, p. 251f.).

Another method to assess the quality of the sub-indicators is the so-called NUSAP (Numerical, Unit, Spread, Assessment, Pedigree), which helps to clarify if the messages of the sub-indicators are reliable and can be used safely to draw policy conclusion from. A pedigree matrix is used for every sub-indicator to evaluate each step in the statistical production process about its quality. The modes with which the steps are executed will be summarised as a categorical variable in the matrix. The categories will be rated with numerical scores according to their quality. By doing so, the quality of the process can be easily communicated to the statistics user (see Nardo et al. 2005, p. 14f.).

Principle 5: Sources of Official Statistics can also be discussed in connection with composite indicators. Sources of data for the sub-indicators have to be chosen carefully concerning the quality, timeliness, costs, and burden on respondents. This principle aims to ensure that data characteristics and the quality are identified beforehand and governed by implemented rules (see United Nations 2015, p. 31f.). It is important because a composite indicator often summarises data from very different types of surveys. For example, a composite indicator based on the SDGs can encompass data from economic surveys, social surveys, or environmental surveys. Social surveys with households or individuals as respondents and economic surveys in which businesses, institutions, or farms are the units of interest are likely to have very different data structures and data quality issues. For economic surveys, less standardized designs and survey practices are applied than for social surveys, making a comparison between statistics from different survey types difficult. Besides this, economic surveys entail specific characteristics which social surveys do not reflect on to the same extent. Some of these issues concern intensively skewed distributions or rapid rate of changes in the data and estimated sub-indicators. Additionally, more data alternatives for economic surveys are available such as administrative records (see Cox and Chinnappa 1995, p. 2ff.).

Principle 9: For the use of international standards, it is important to promote the utilization of commonly agreed methods and classification systems from statistical agencies. This helps to ensure comparability between statistics from different countries or sources within a country, and it is a dimension of quality that can be communicated via published metadata. Comparability is an important dimension of quality, especially if the statistics are used to compare the performance of countries or regions. Standard methods in the statistical production process can enhance the explanatory power of the comparisons and improve the efficiency of the process (see United Nations 2015, p. 80f.). Though, the level of standardization might be restricted due to country-specific circumstances and particularities. In economic surveys, these can, for example, influence the survey frame. With regard to the data used for a composite indicator based on the SDG framework, the Intern-Agency and Expert Group on SDG indicators (IAEG-SDG) developed criteria and guidelines for regulating the data flow between countries and custodian agencies which are responsible for collecting, analyzing, and reporting the SDGs. These means are aimed to ensure the quality of the reported estimates and the harmonization of the single SDG target variables over the countries.

Gennari and Navarro (2019) discuss three main issues of the guidelines on data flows and global data reporting, which could lead to inconsistencies in the reporting of the SDG target variables. The authors discuss the problem of inconsistencies in the data validation process between national statistical organizations and the custodian organization, the problem that specific provisions of the guidelines are not followed in practice, and the absence of detailed modalities on the data validation process (see Gennari and Navarro 2019, p. 738ff.). This could lead to a situation in which the data quality of countries differs, and comparisons of SDG indicators between countries might not be fair. Further literature, like Thomas et al. (2016) and Sarvajayakesavalu (2015), describes the challenges of countries implementing the statistical infrastructure needed to report quality indicators in more detail.

4.4 A COMPOSITE INDICATOR ON SUSTAINABLE ECONOMIC DEVELOPMENT

Composite indicators using the SDG framework could be generally constructed in two different ways. The SDG index or composite indicator proposed in Sachs et al. (2019) is constructed using a basket of the targets from the 17 SDGs as sub-indicators and aggregating them with equal weighting.

In the case of a composite indicator for countries' sustainable development, this could mean that countries decide on either the single targets of the 17 goals or on the 17 goals themselves to construct the country-specific composite indicator. This composite indicator could be constructed with a combination of targets applicable for every country and targets which are not relevant for all countries (see Melamed and Bergh 2014, p. 4). For example, on the SDG level, goal 14: Life Below Water might not be relevant for landlocked countries.

Additionally, it is possible to use the SDG framework to construct a composite indicator measuring specific aspects of sustainable development. This has been done, for example, by Rickels et al. (2016). The authors build a composite indicator with target variables from SDG 14 to measure oceanic development in the European Union. Another example is the composite indicator on agricultural development by Nhemachena et al. (2018), which uses target variables from SDGs 1, 2, 6, 7, 15.

In alignment with these ideas, the composite indicator used in the following studies is built as a composite indicator on sustainable economic development from a specific set of the 17 SDGs and target variables as sub-indicators. Wu et al. (2018) explain that the 17 SDGs can be split up into three dimensions of sustainable development according to human needs. A different set of SDGs represents each dimension. Sustainable economic development is represented by SDGs 1, 2, 3, 8, and 9. Additionally, the social dimension of sustainable development is captured by SDGs 4, 5, 10, 11, 16, and 17, and the set of SDGs 6, 7, 12, 13, 14, and 15 represents the environmental dimension of sustainable development (see Wu et al. 2018, p. 4).

To construct the composite indicator of sustainable economic development, country-level data on sub-indicators or target variables from the five corresponding goals (SDGs 1, 2, 3, 8, and 9) and from the year 2015 was downloaded from https://

TABLE 4.1 | SUB-INDICATORS OF THE COMPOSITE INDICATOR ON SUSTAINABLE ECONOMIC DEVELOPMENT

SDG	Sub-indicators of the composite indicator of sustainable economic development
1 NO POVERTY	1.1.1 Proportion of population below international poverty line (%) 1.4.1 Proportion of population using basic drinking water services (%)
2 ZERO HUNGER	2.1.1 Prevalence of undernourishment (%) 2.a1 Agriculture orientation index for government expenditures
3 GOOD HEALTH AND WELL-BEING	3.8.1 Universal health coverage (UHC) service coverage index 3.c1 Health worker density, by type of occupation (per 10,000 population)
8 DECENT WORK AND ECONOMIC GROWTH	8.1.1 Annual growth rate of real GDP per capita (growth factor) 8.2.1 Annual growth rate of real GDP per employed person (growth factor) 8.6.1 Proportion of youth not in education, employment, or training (%)
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE	9.2.1 Manufacturing value added as a proportion of GDP (%) 9.3.1 Proportion of small-scale industries in total industry value added (%) 9.b1 Proportion of medium and high-tech industry value added (%) 9.c1 Proportion of population covered by at least a 4G mobile network (%)

unstats.un.org/sdgs/indicators/database/[1]. The sub-indicator variables are shown in Table 4.1, ordered by the sustainable development goal and the corresponding target within each goal. A selection of sub-indicator variables and the year of the data reporting has taken place based on data availability and suitability of the data structure. Data was taken from 20 countries with different economic profiles. Since the following studies refer to data-specific issues and do not try to make any implications about country-specific economic profiles, indices rather than country names will refer to the countries.

Country profiles with regard to their performance on the sub-indicators can be visualized using so-called spider plots or radar charts. This visualization can help gain a quick overview of how balanced the country's performance is across all sub-indicators of interest and compare countries in terms of single sub-indicators. Different aggregation methods handle the balance or imbalance of country-specific performance profiles differently. Hence, it might be more beneficial for a country to have more balanced sub-indicator values under certain construction conditions to achieve a better composite indicator score. This will be explained in more detail in Section 4.6. Figure 4.1 shows a radar chart with the five countries for which a full dataset without missing values was available.

Further explanations on the missing data structure are given in Section 4.5. None of these countries perform equally well

1 Last accessed: 4 July 2020

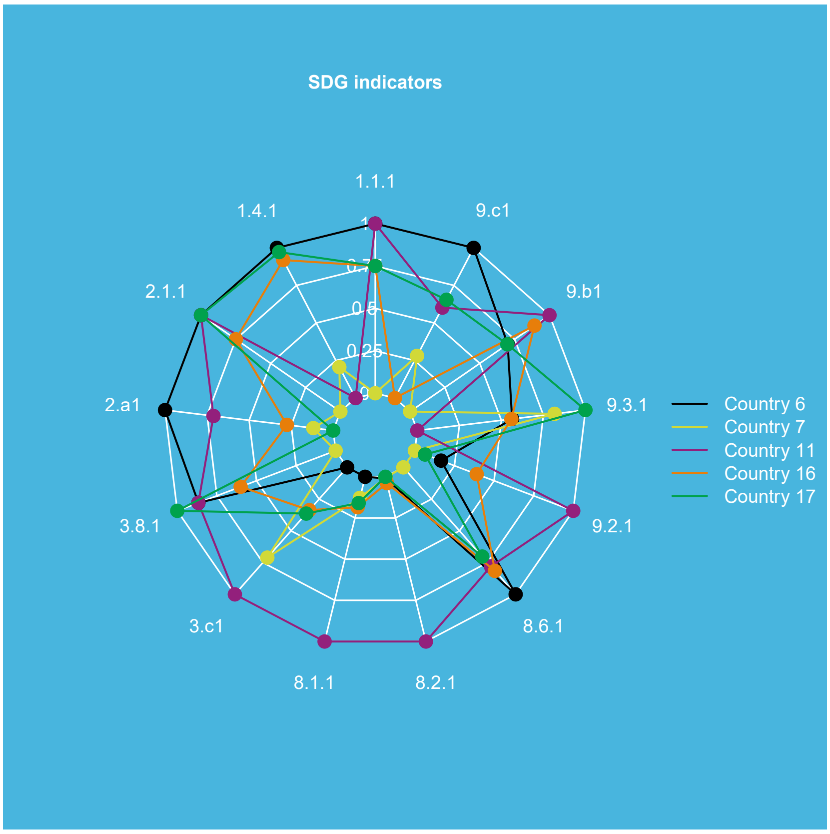

FIG. 4.1 | RADAR CHART FOR THE FIVE COUNTRIES WITH COMPLETE DATASETS

SOURCE | AUTHOR ELABORATION BASED ON https://unstats.un.org/sdgs/indicators/database (LAST ACCESSED: 4 JULY 2020).

across all 13 sub-indicators chosen to construct the composite indicator of sustainable economic development. On the other hand, it can also be stated that none of the five countries perform poorly across all 13 sub-indicators.

As shown in Table 4.1, the sub-indicator variables $x_{k,c}$ are not all measured using the same measurement unit. To make the measurements comparable when building the composite indicator of sustainable economic development, a min-max normalization is applied, using Formula (2)

$$x^*_{k,c} = \frac{x_{k,c} - min(x_k)}{max(x_k) - min(x_k)} \times 100 \qquad \textbf{(2)}$$

Due to Formula (2) and after the aggregation of the sub-indicators, the resulting composite indicator values will range between 0 and 100, similar to Nhemachena et al. (2018) and Sachs et al. (2019). In the normalization step, it is also important to account for the polarization or direction of the sub-indicators. Greater values of positive sub-indicators will indicate greater sustainable economic development, whereas greater values of negative sub-indicators refer to lower sustainable economic development. From Table 4.1, it can be understood that the sub-indicators 1.1.1, 2.1.1, and 8.6.1 are cases of negative sub-indicators. Their scale will be reversed for consistency in the interpretation of the resulting composite indicator values before the aggregation step.

Therefore, greater values of the composite indicator will indicate greater sustainable economic development. To construct the composite indicator of sustainable economic development, the sub-indicators are equally weighted and aggregated using the arithmetic mean, similarly to constructing the composite indicator of agricultural development by Nhemachena et al. (2018). A critical step in the construction process is the handling of missing data, which occurs before the normalization and aggregation of the sub-indicators. Section 4.5 aims to visualize how the choice of imputation method, as a common method to handle missing data, can influence the results of the composite indicators and country comparisons.

4.5 IMPACT OF MISSING DATA ON THE COMPOSITE INDICATOR

In this section, a small case study showing the impact of imputation on the composite indicator results of sustainable economic development is outlined. Figure 4.2 describes the structure of missing values in the data set of 20 countries used subsequently to calculate the composite indicator of sustainable economic development. The left side of Figure 4.2 visualizes the number of missing records for each variable in the dataset. Some of the sub-indicator variables show high frequencies of missing records, such as, for example, sub-indicator 1.1.1 with 45 percent of the values missing and sub-indicator 9.3.1 with 40 percent of the values missing. Sub-indicators 2.1.1 and 9.b1, for example, are recorded completely for all countries in the dataset.

The right side of Figure 4.2 shows all apparent combinations of missing and non-missing values in the sub-indicator variables. From the missing value frequencies of variable pairs, it can be seen that 11 different patterns of missing values can be found in the dataset. Since the variables of sub-indicators 1.1.1 and 9.3.1 have the most missing records, most of the patterns include these two variables. Twenty-five percent of the cases in the dataset has no missing record at all. This is true for the five countries shown in the radar chart above.

OECD and JRC European Commission (2008) introduce different imputation methods for constructing composite indicators and propose rules of thumb to choose between them. Generally, imputation methods are grouped into single imputation and multiple imputation methods. A further overview of imputation methods can be found in Meinfelder (2014), Zhang (2003), or OECD and JRC European Commission (2008). Even though there is no definite answer, the rules of thumb for choosing an imputation method by OECD and JRC European Commission (2008) depend on the scale of the

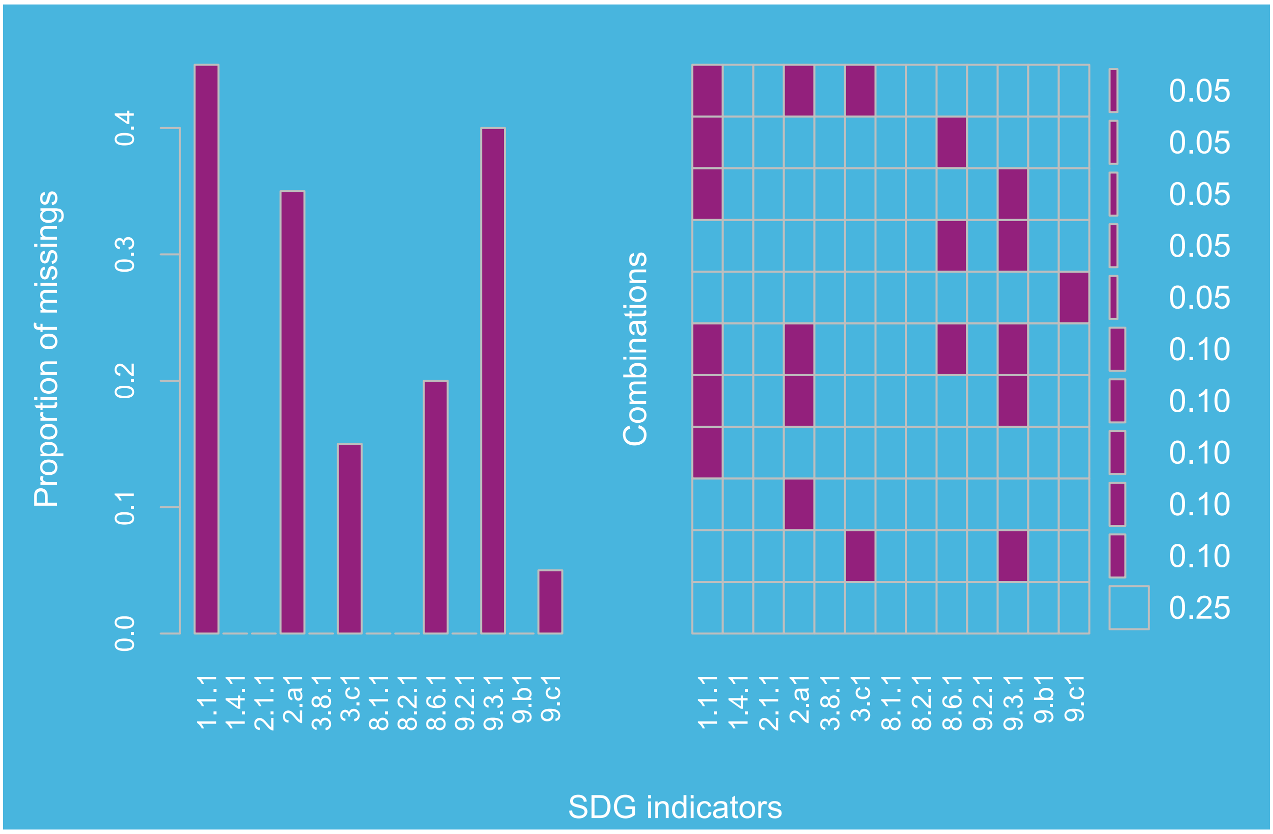

FIG. 4.2 | MISSING DATA PATTERN

SOURCE | AUTHOR ELABORATION BASED ON https://unstats.un.org/sdgs/indicators/database (LAST ACCESSED: 4 JULY 2020).

variables, the amount of missingness in the dataset, and the relationships of the sub-indicator variables with missingness of the corresponding country. The performance of the imputation methods could be examined by applying an in-sample/out-of-sample logic for which the complete part of the dataset is taken, and similar missingness is introduced. After this step, the different considered imputation methods are applied, and their performance is compared, for example, by measuring the correlation between the complete and imputed dataset (see OECD and JRC European Commission 2008, p. 62).

To show the influence of the choice of the imputation method on the results of the composite indicator of sustainable economic development, mean imputation, hot-deck imputation, and the predictive mean matching imputation as multiple imputation methods are applied using the R package MICE. The predictive mean matching imputation generates five imputed datasets. For each of the five datasets, the composite indicator is calculated as described above for each country. These results are pooled using an arithmetic mean. This method has the advantage that the imputed values are restricted to the observed values and can preserve non-linear relationships (see van Buuren and Groothuis-Oudshoorn 2011, p. 18). For a more detailed explanation of the method, see Little (1988). This is also true for the mean imputation and the hot deck imputation as single imputation methods.

In Figure 4.3, the distribution of the composite indicator results for each country calculated from the imputed datasets is shown by boxplots. The boxplots also include the results of the five imputed datasets from the predictive mean matching method. In addition, the boxplots are ordered according to the pooled results based on the predictive mean matching method. As explained in Section 4.2, ranking and comparisons of the observed regions or countries are often one of the main interests in constructing composite indicators. Higher values of the composite indicator refer to higher sustainable economic development. Therefore, using the predictive mean matching imputation, Country 11 is found to have the highest sustainable economic development, whereas Country 20 shows the lowest sustainable economic development. Results from the three imputation methods are also highlighted with colored squares to inspect if the results had come to similar conclusions regarding the composite indicator values and the country rankings. No squares are shown for countries that have complete datasets as no data were imputed.

First of all, it can be seen that country results differ in their range of resulting composite indicator values. Some of this variation can be easily explained by the amount of missingness for each country, indicated by the percentage numbers above each boxplot. In case of a small amount of missing data, the imputation method might not significantly affect the resulting composite indicators. Some countries with a greater amount of missingness, such as Country 10, also do not show very different results of the composite indicator calculated with different imputation methods. For most of the countries with a greater amount of missing data, on the other hand, it can be seen that the results of composite indicators using mean imputation, hot-deck imputation, or predictive mean matching imputation differ to a larger extent. This may lead to problems in the interpretation of the output; if the chosen imputation methods influence the resulting composite indicator values, country rankings and comparisons will not solely reflect on the country performances but also on possible country-specific choices in the construction process of the composite indicator, here the choice of imputation method in combination with the amount of missing data. Examples for this in Figure 4.3 are Countries 5 and 12, which will change places in the country ranking if the hot deck imputation method is used to calculate the composite indicator rather than predictive mean matching. Another example is Country 14, which will change ranking with Country 3 if mean imputation will be used to impute the missing data before calculating the composite indicator. Ranking might be less distinct, and a country‘s assessed performance will depend highly on the imputation method used on the data.

The yellow squares in Figure 4.3 indicate the results in case of leaving the sub-indicators with missingness out of the calculation of the composite indicator for each respective country. Therefore, the country-specific estimates will be based on different sets of sub-indicators in case of missingness in the data. These results visualize how the country rankings change if different sets of sub-indicators, for example, due to missingness, are used to calculate the composite indicator. Again, we can see that the results differ considerably between the different methods.

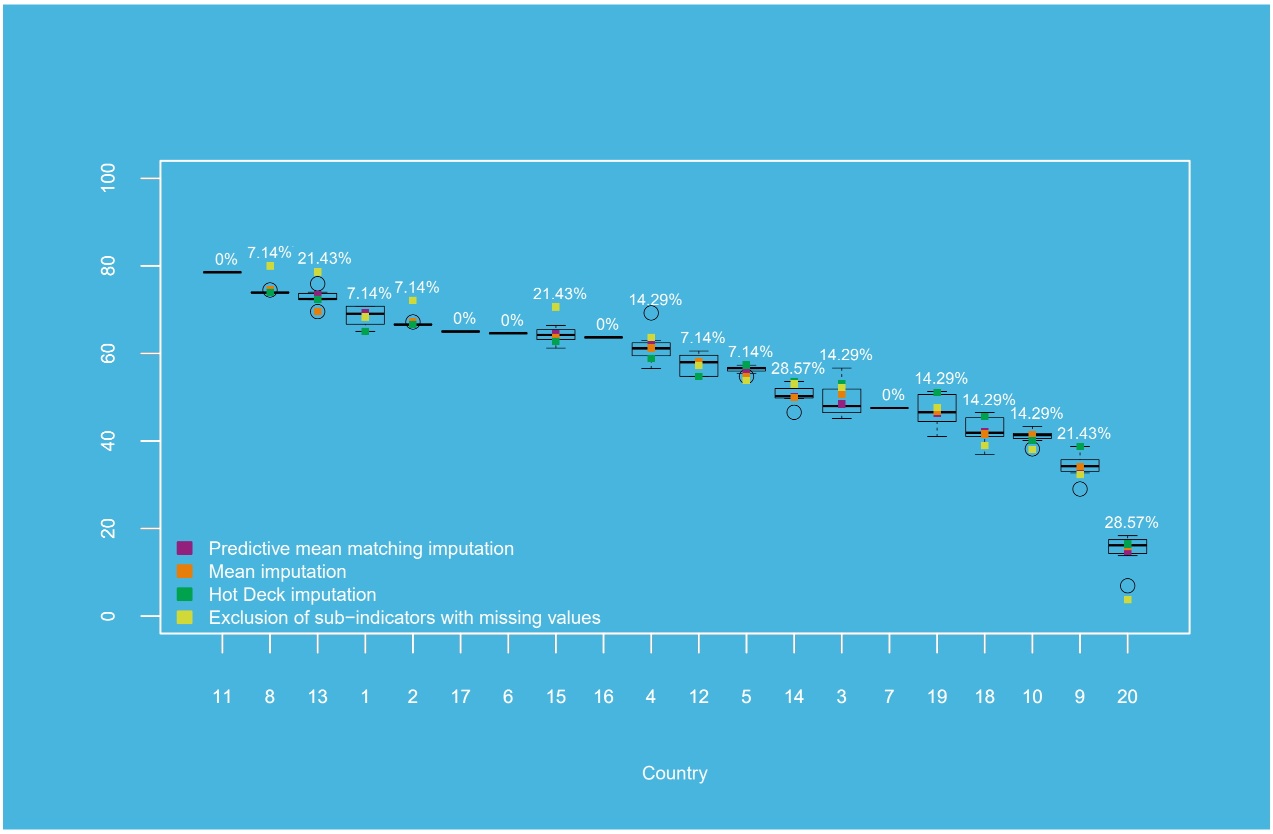

FIG. 4.3 | DISTRIBUTION OF COMPOSITE INDICATOR VALUES WITH AND WITHOUT IMPUTATION

SOURCE | AUTHOR ELABORATION BASED ON https://unstats.un.org/sdgs/indicators/database (LAST ACCESSED: 4 JULY 2020).

It is questionable how meaningful comparisons of performance developments are if different imputation methods or a different set of sub-indicators due to missing values or other considerations are used. Therefore, the imputation method should always be chosen due to discussion and based on the availability of additional data and further knowledge of the sub-indicator's relationships. Further, it is important to report the used methods in the metadata to improve the interpretability of resulting country rankings. For this example, data was imputed on the macro-level because the dataset already included estimates of the sub-indicator variables. In general, it can be helpful to use micro-level data; thereby, additional information and country characteristics can be used to enhance the imputation model.

In the next section, an uncertainty analysis is used to demonstrate how composite indicator results may be affected by the subjective choices made throughout the construction process. This analysis will look at the selection of the sub-indicator set, the normalization method, and the aggregation method.

4.6. UNCERTAINTY ANALYSIS OF THE COMPOSITE INDICATOR

The construction process of a composite indicator includes different stages in which more or less subjective choices have to be made, such as the choice of sub-indicator variables, the normalization method, or the aggregation method, as explained above. Saisana et al. (2005) or Saltelli et al. (2008) describe a variance-based sensitivity analysis and an uncertainty analysis for composite indicators, which evaluate the impact of decisions on the result of an output variable such as the composite indicator values itself or the corresponding country rank. In this setup, choices in the construction process are referred to as input triggers or input variables. The composite indicator values or country rankings are calculated due to every possible combination of input triggers. To conduct a sensitivity analysis that quantifies the impact of each input trigger, the so-called first-order sensitivity indices can be calculated. This approach is based on the decomposition of the total output variance resulting from calculating the composite indicator with each considered construction choice combination. How the first-order sensitivity index is calculated, and a sensitivity analysis is constructed can be understood from Saltelli et al. (2008) and Saisana et al. (2005).

Another more visual approach described by the same authors to assess the influence of the construction choices on the results of the compose indicator is the uncertainty analysis. For this analysis, the distributions of output variables are graphically illustrated by boxplots or by visualizing the relative frequencies of the achieved ranks by country. Generally, country rankings aim to evaluate a country's performance relative to other countries in the observed population. Therefore, it is crucial to analyze how distinct the rankings based on a composite indicator are. Suppose a rank position of a country depends strongly on the methods used to construct the composite indicator and changes considerably under the use of different methods. In that case, it is questionable how meaningful the country's performance can be compared relative to other countries. In such cases, the rank results might not solely depend on the country performance itself.

Below, an uncertainty analysis of the composite indicator of sustainable economic development described above and in Saisana et al. (2005) or Saltelli et al. (2008) will be discussed. For the purpose of this analysis, a complete dataset will be used, which resulted from the application of the hot deck imputation. This uncertainty analysis aims to visualize how strong the influence of the choice of the sub-indicator set, the normalization method, and the aggregation method on the country rankings of the composite indicator values can be.

For the uncertainty analysis, the composite indicator of sustainable economic development is calculated 84 times based on all suitable input trigger combinations. For each calculation, either one or none of the 13 sub-indicators are excluded from the considered set. As a normalization method, either the min-max normalization or standardization is used by applying the function normalise_ci from the R package Compind (see Fusco, Vidoli, and Sahoo 2018). The choice of the aggregation method is also strongly connected to the weighting of the single sub-indicators. Not all sub-indicators are weighted equally when applying different aggregation methods. As in Section 4.5, the first chosen aggregation method is the arithmetic mean, with which all sub-indicators are equally weighted. OECD and JRC European Commission (2008) also suggest using a geometric mean and the aggregation of the sub-indicators in connection with a principal component weighting approach. Using the geometric mean as aggregation will lead countries with higher sub-indicator scores to be rewarded, as it is more difficult to compensate low sub-indicator values with higher sub-indicator values compared to the case of using an arithmetic mean as the aggregating method (see OECD and JRC European Commission 2008, p. 32f.).

Additionally, by using an equal weighting scheme when aggregating the sub-indicators, the weights of the sub-indicators can be calculated using the principal component approach. Thereby, the weights are calculated due to the correlations of the sub-indicators, and the aggregation is done by a weighted arithmetic mean. An in-depth description of how to calculate the weights using the principal component approach can be found in OECD and JRC European Commission (2008). The fourth aggregation method used in this analysis is explained in de Muro, Mazziotta, and Pareto (2012) and is called the Mazziotta-Pareto Index (MPI). This composite indicator started with a linear aggregation and introduced a penalty for unbalanced values of sub-indicator sets using the coefficient of variation. Therefore, this approach can be applied if the set of sub-indicators is considered to be non-substitutable (see de Muro et al. 2012, p. 8ff.). For an in-depth explanation of this composite indicator‘s construction, the reader is referred to De Muro et al. (2012). Not all combinations of normalization methods and aggregation methods are applied in this analysis. The MPI index aggregation is applied using only the standardization method, and the geometric mean aggregation is applied using the min-max method in the step of the normalization. This will lead to 84 combination possibilities of the trigger values for the analysis. A summary of the input triggers is given below in Table 4.2.

TABLE 4.2 | SUMMARY OF THE INPUT TRIGGERS AND POSSIBLE COMBINATIONS FOR THE UNCERTAINTY ANALYSIS

Exclusion of a sub-indicator	Normalization method	Aggregation / weighting scheme
• Calculate CI with all sub-indicators • Calculate CI without fist sub-indicator • Calculate CI without second sub-indicator variable • Calculate CI without the thirteenth sub-indicator variable	• Min-max method • Standardization method	• Arithmetic mean • Geometric mean* • Aggregation with principal component weights • Aggregation as MPI index*

SOURCE | AUTHOR ELABORATION.
NOTE | THE GEOMETRIC MEAN AGGREGATION IS ALWAYS APPLIED USING THE MIN-MAX NORMALIZATION METHOD AND THE MPI INDEX AGGREGATION IS ALWAYS APPLIED USING STANDARDIZED SUB-INDICATOR VALUES. THEREFORE, 28 POSSIBLE COMBINATIONS OF THESE THREE INPUT TRIGGERS ARE NOT CONSIDERED.

The geometric mean aggregation is always applied using the min-max normalization method and the MPI index aggregation is always applied using standardized sub-indicator values. Therefore, 28 possible combinations of these three input triggers are not considered. *

For the uncertainty analysis, the resulting country ranking distributions are analyzed. They are represented in Figure 4.4, which visualizes the relative frequencies of how often each country takes on each rank position over the 84 calculations

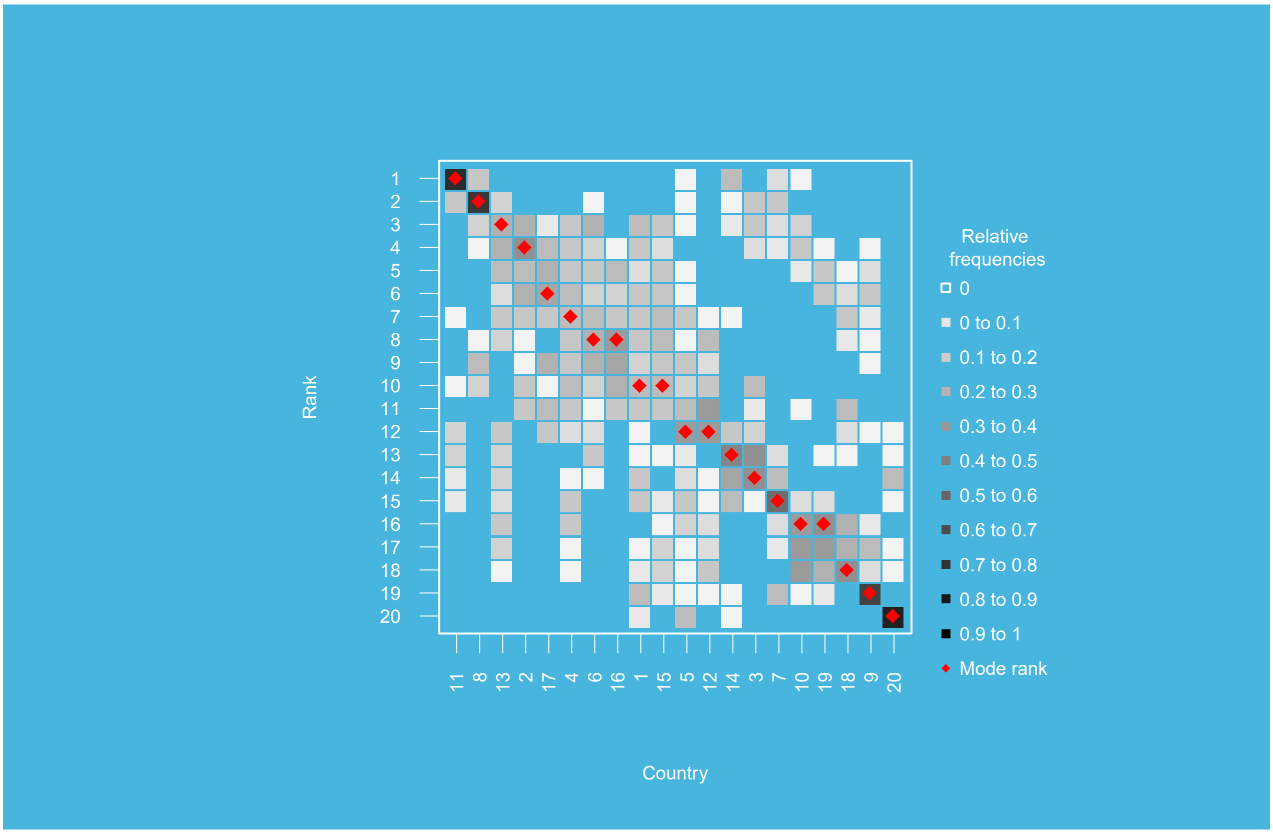

FIG. 4.4 | RELATIVE FREQUENCIES OF COUNTRY RANKS FOR THE UNCERTAINTY WITH ALL THREE INPUT TRIGGERS

SOURCE | AUTHOR ELABORATION BASED ON https://unstats.un.org/sdgs/indicators/database (LAST ACCESSED: 4 JULY 2020).

of the composite indicator. Deeper colors indicate a higher relative frequency of the corresponding rank. The red sign in the figure points out the mode rank for each country, i.e., their most frequently occurring rank. On the x-axis, the countries are ranked by this measure for the readability of the plot.

It can be concluded from Figure 4.4 that due to the construction choices, a comparison based on country rankings has to be viewed critically. Most of the countries take on an extensive range of rank positions across all 84 calculations. This can be

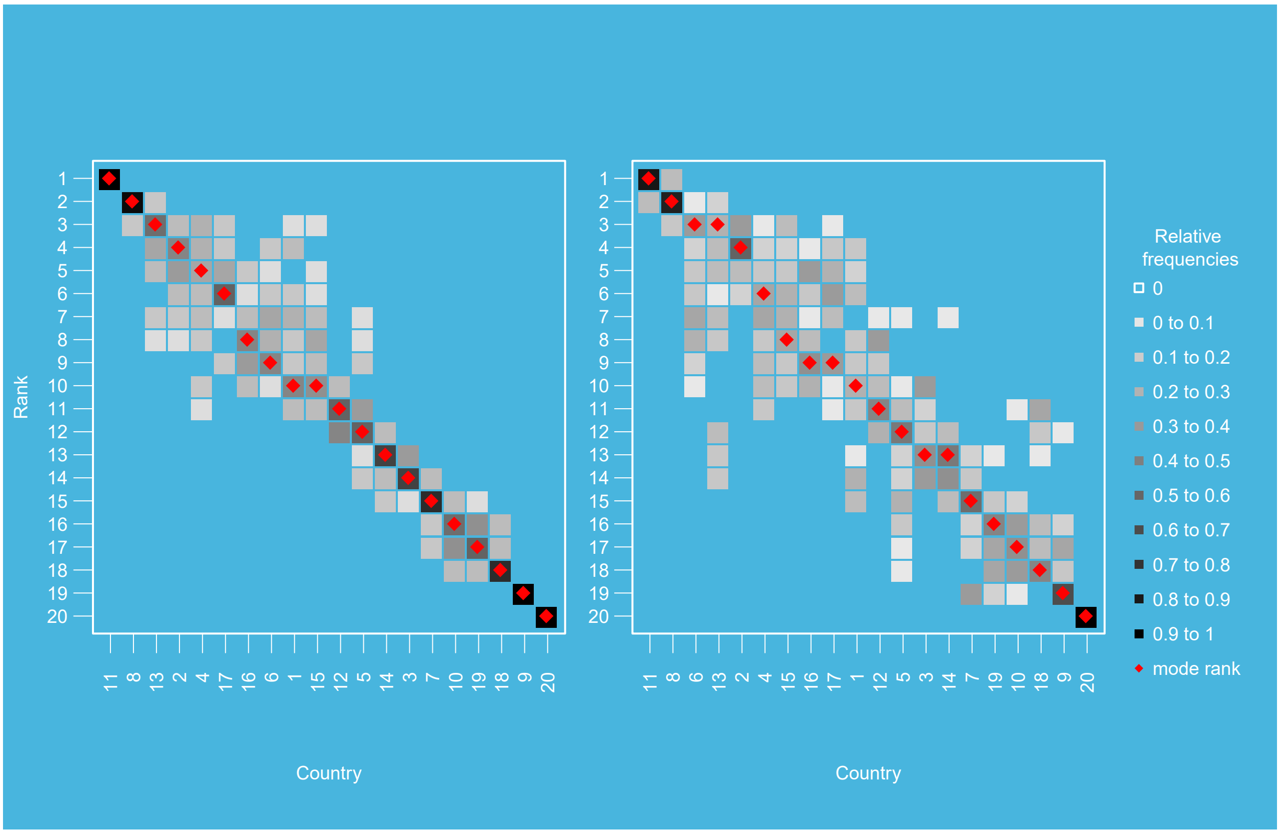

FIG. 4.5A, FIG. 4.5B | RELATIVE FREQUENCIES OF COUNTRY RANKS FOR THE UNCERTAINTY WITH TWO INPUT TRIGGERS

SOURCE | AUTHOR ELABORATION BASED ON https://unstats.un.org/sdgs/indicators/database (LAST ACCESSED: 4 JULY 2020).
NOTE | LEFT SIDE: ARITHMETIC MEAN WAS USED AS AGGREGATION METHOD FOR ALL 28 CALCULATIONS. RIGHT SIDE: THE STANDARDIZATION METHOD WAS USED IN THE NORMALIZATION STEP FOR ALL 42 CALCULATIONS.

seen, for example, for Countries 1 or 5, which take on almost all ranks with similar relative frequencies. Therefore, the country rankings are not distinct. Also, most of the ranks are colored in similar shades, and their mode ranks are colored in light grey, which implies that they are not achieved with high relative frequencies. No country takes on only one rank in this relative comparison across all calculations. However, most countries take on multiple different ranking positions across all calculations with similar frequencies. The ranking of a country very much depends on the construction decisions and not necessarily on the country's performance on the sub-indicators.

For the uncertainty analysis, it is possible to fix one of the input triggers and visualize the effect this decision has on the calculation of the composite indicator of sustainable economic development. The right side of Figure 4.5 shows the results of the uncertainty analysis if the aggregation method is fixed to the arithmetic mean. In this case, the composite indicator was calculated based on 28 different possible combinations of the trigger on the sub-indicator choice and the choice of the normalization method. The left side of Figure 4.5 shows the uncertainty analysis fixing the normalization method to the standardization method. Here, the composite indicator was calculated based on 42 possible combinations from the two other input triggers. As in Figure 4.4, the countries are organized alongside the x-axis by their ranks‘ mode value. Therefore, it is possible that the order in Figure 4.4 and both sides of Figure 4.5 are not the same. From this, it can be seen how counties switch ranks in the relative comparison due to different input trigger choices. Country 15, for example, changes three rank positions between the right and the left side of Figure 4.5 by fixing one of the input trigger decisions.

It is noticeable from both sides of Figure 4.5 that either choice of the aggregation and the normalization method significantly affects the composite indicator values. With one of these choices fixed, most countries achieve their corresponding rank with higher frequency and lower ranks in total overall calculations. With a fixed aggregation method, Countries 11, 9 and 20 will always take on the same ranks in the country comparison. Also, Country 20 will take on the same rank overall calculations if the normalization method is fixed. Overall, the countries achieve lower different ranks with higher frequency, and therefore, the relative comparisons using country ranks get more distinct. In this, the fixing of the aggregation methods seems to achieve the most distinct relative country comparisons, showing how strong the influence of the choice on the aggregation method is.

The results of the uncertainty analysis show that using rankings to compare country performances has to be viewed critically. Very different rankings can result due to different methods used to construct the composite indicator of sustainable economic development. It can be concluded that the normalization and aggregation method has a great impact on the rank positions for most of the countries. Furthermore, these triggers are only examples of critical choices in the construction process. The varying results for the country rankings lead to the problem that countries‘ performances cannot be compared distinctively. It is therefore imperative to analyze the composite indicator values in more detail. This includes evaluating the results regarding their sensitivity towards method choices and the consideration of country performances on single sub-indicators, for example, in dashboards. For a detailed analysis of the results, it is furthermore necessary to account for additional information on the data quality to avoid misinterpretations of the results, such as, for example, the missing data structure and applied imputation methods.

4.7. CONCLUSION

The present paper highlights some problems when drawing conclusions from composite indicators used for informed political decision-making. These concerns comparisons of country performances and measuring development can be especially important for composite indicators constructed from the SDG framework due to its complexity. The primary focus is laid on data quality aspects, such as the impact of missing values and the influence of choices made during the construction process of a composite indicator.

Missing values and the choice of the imputation method, in general, can have a big influence on the outcome of composite indicators, which is shown in an analysis comparing calculation results with imputed datasets using single and multiple imputation methods. Providing information on the missingness for each sub-indicator and the used imputation method is therefore fundamental and can help gain further insight into the data‘s quality. This documentation should go hand in hand with a commonly used process on choosing the most suitable imputation method over the countries of the comparison to ensure consistency.

The uncertainty analysis of the composite indicator of sustainable economic development showed that the results of the country rankings could be changed dramatically by the choice of the normalization and aggregation methods. Different aggregation methods handle the substitutability of sub-indicators differently, and therefore, countries with balanced or unbalanced sub-indicator profiles will benefit differently from the chosen aggregation method. These choices should be discussed intensively and justified considering the insights the analysts intended to gain from the composite indicator. Also, the choice of the sub-indicator set can influence the ranking of a country. This circumstance might be of particular interest for a composite indicator constructed from the SDG framework. Not all SDGs and their targets are important for all countries, or target variables are not available for each country. The extent to which countries can be compared if the composite indicator is calculated based on country-specific sub-indicator sets must be discussed critically. The influence of this choice should not be overlooked.

Results of composite indicators and relative rankings of countries should not be interpreted on their own. An in-depth look into the construction process of the composite indicator with a sensitivity and uncertainty analysis, the country-specific performance on the single sub-indicators using, for example, dashboards and analysis of quality of the data with which the sub-indicators were estimated are vital and make the composite indicator results more reliable.

4.8. REFERENCES

African Union, 2009. African Charter on Statistics. Addis Ababa: African Union.

Biemer, P.P. and Lyberg, L.E., 2003. Introduction to Survey Quality. Hoboken, N.J.: John Wiley & Sons, Inc.

Boudt, K., Todorov, V. and Upadhyaya, S., 2009. Nowcasting Manufacturing Value Added for Cross-Country Comparison. Statistical Journal of the IAOS, 26(1), pp. 15-20.

Christiansan, A. and Tortora, R.D., 1995. Issues in Surveying Business: An International Survey. In: Cox, B.G., Binder, D.A., Chinnappa, B.N., Christianson, A., Colledge, M.J., Kott, P.S. (eds.), Business Survey Methods, pp. 237-256. New York, N.Y.: John Wiley & Sons.

Cox, B. and Chinnappa, B., 1995. Unique Features of Business Surveys. In: Cox, B.G., Binder, D.A., Chinnappa, B.N., Christianson, A., Colledge, M.J., Kott, P.S. (eds.), Business Survey Methods, pp. 237-256. New York, N.Y.: John Wiley & Sons.

De Muro, P., Mazziotta and M., Pareto, A., 2012. Composite Indices of Development and Poverty: An Application to MDGs. Social Indicators Research, 104(1), pp. 1–18.

Eurostat, 2018. European Statistics Code of Practice: For the National and Community Statistical Authorities and Eurostat (EU Statistical Authority). Luxembourg: European Union

Fusco, E., Vidoli, F. and Sahoo, B.K., 2018. Spatial Heterogeneity in Composite Indicator: A Methodological Proposal. Omega, 77(C), pp. 1–14.

Gennari, P. and Navarro, D.K., 2019. Validation of Methods and Data for SDG indicators. Statistical Journal of the IAOS, Preprint, 35(4), pp. 735–741.

Little, R.J.A., 1988. Missing Data Adjustments in Large Surveys (with Discussion). Journal of Business Economics and Statistics, 6(3), pp. 287–301.

Meinfelder, F., 2014. Multiple Imputation: An Attempt to Retell the Evolutionary Process. AStA Wirtschafts- und Sozialstatistisches Archiv, 8(4), pp. 249-267.

Melamed, C. and Bergh, G., 2014. Sustainable Development Goals and Targets: Options for Differentiating between Countries. London: Overseas Development Institute.

Münnich, R., Magg, K., Ohly, D. and Wiegert, R., 2008a. Workpackage 3: Accuracy Measurement of Composite Indicators. Deliverable 3.4. Knowledge Economicy Indicators Project. Trier: Trier University.

Münnich, R. and Zins, S., 2011. Workpackage 3: Variance Estimation for Indicators of Poverty and Social Exclusion. Deliverable 3.2. Advanced Methodology for European Laeken Inicators (AMELI) Project. Trier: Trier University.

Nardo, M., Saisana, M., Saltelli, A. and Tarantola, S., 2005. Workpackage 5: Input to Handbook of Good Practices for Composite Indicators' Development. Deliverable 5.2. Knowledge Economicy Indicators Project. Trier: Trier University.

Nhemachena, C., Matchaya, G., Nhemachena, C.R., Karuaihe, S., Muchara, B. and Nhlengethwa, S., 2018. Measuring Baseline Agriculture-Related Sustainable Development Goals Index for Southern Africa. Sustainability, 10(3), pp. 849–865.

OECD and Joint Research Centre-European Commission, 2008. Handbook on Constructing Composite Indicators: Methodology and User Guide. Paris: OECD publishing.

Rickels, W., Dovern, J., Hoffmann, J., Quaas, M. F., Schmidt, J.O., and Visbeck, M., 2016. Indicators for Monitoring Sustainable Development Goals: An Application to Oceanic Development in the European Union. Earth's Future, 4(5), pp. 252–267.

Rosen, R., 1991. A Review of Life Itself: A Comprehensive Inquiry into Nature, Origin and Fabrication of Life. New York, N.Y.: Columbia University Press.

Sachs, J., Schmidt-Traub, G., Kroll, C., Lafortune, G., and Fuller, G., 2019. Sustainable Development Report 2019. New York, N.Y.: Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN).

Saisana, M., Saltelli, A. and Tarantola, S., 2005. Uncertainty and Sensitivity Analysis Techniques as Tools for the Quality Assessment of Composite Indicators. Journal of the Royal Statistical Society, 168(2), pp. 307–323.

Saltelli, A., Ratto, M., Andres, R., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S., 2008. Global Sensitivity Analysis: The Primer. Chichester: John Wiley & Sons, Ltd.

Sarvajayakesavalu, S., 2015. Addressing Challenges of Developing Countries in Implementing Five Priorities for Sustainable Development Goals. Ecosystem Health and Sustainability, 1(7), pp. 1–4.

Statistics Canada, 2019. Statistics Canada Quality Guidelines. Sixth Edition. Ottawa.

Statistics Norway, 2007. Statistics Norway's Dissemination Policy. Oslo.

Statistics South Africa, 2010. South African Statistical Quality Assessment Framework (SASQAF): Operational Standards and Guidelines. First Edition. Pretoria.

Sustainable Development Solutions Framework, 2015. Indicators and a Monitoring Framework for the Sustainable Development Goals: Launching a Data Revolution. Report to the Secretary-General of the United Nations by the Leadership Council of the Sustainable Development Solutions Network. Paris and New York, N.Y.

Thomas, J.C., Silvestre, E., Salentine, S., Reynolds, H., Smith, J., 2016. What Systems Are Essential to Achieving the Sustainable Development Goals and What Will It Take to Marshal Them? Health Policy and Planning, 31(10), pp. 1445–1447.

United Nations, 2015. United Nations Fundamental Principles of Official Statistics: Implementation Guidelines. New York, NY.

van Buuren, S., Groothuis-Oudshoorn, K., 2011. Mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), pp. 1–67.

Wu, J., Guo, S., Huang, H., Liu, W., and Xiang, Y., 2018. Information and Communications Technologies for Sustainable Development Goals: State-of-the-Art, Needs and Perspectives. IEEE Communications Surveys & Tutorials, 20(3), pp. 2389–2406.

Zhang, P., 2003. Multiple Imputation: Theory and Method. International Statistical Review, 71(3), pp. 581–592.

A NEW POSSIBILITY: BIG DATA FOR MISSING VALUE TREATMENT IN THE STATISTICAL PRODUCTION AT UNIDO

By Shuyi Zhang, Dong Guo, Shyam Upadhyaya, and Valentin Todorov,
United Nations Industrial Development Organization, Austria

5.1 INTRODUCTION

Many international organizations maintain their own databases to enable more detailed cross-country analysis and to promote economic, social, or environmental development in one way or the other. The United Nations Industrial Development Organization (UNIDO) compiles and maintains a thorough international industrial statistical database INDSTAT, which contains time series data on major indicators of industrial statistics for more than 160 countries at different levels of detail (across the databases) according to the International Standard Industrial Classification (ISIC), Revision 3 or Revision 4 (United Nations 2002, 2008). According to Upadhyaya (2014), INDSTAT is primarily populated with data from official national statistics offices (NSOs), and, as a secondary source, whenever national surveys are inadequate, from international organizations such as the Organization for Economic Cooperation and

Development (OECD), the United Nations Statistical Division (UNSD), and the World Bank. Although official statistics are, indisputably, the best source of data, there are various challenges related to its quality. For instance, in many countries, low response rates due to budgetary or political reasons cause a missing value problem in national databases, both spatially and temporally. The gaps may affect, potentially severely, the quality of survey data.

Nowadays, a large portion of data is produced and maintained by non-official organizations and platforms. The growing internet penetration rate, the proliferation of social media, and enhanced sensor deployment have opened up a new information concept: big data. Smith (2018) summarizes how big data can be helpful for official statistics production from four perspectives. Non-official data, along with big data techniques such as data mining, modelling, and forecasting, may provide complementary information to official statistics production to improve its completeness and timeliness, especially in countries and industries where traditional official data and methodologies have known deficits. Among the many potential benefits of big data, this paper focuses on providing better solutions for missing value treatment. Section 5.2 summarizes the missing value problem, how it affects INDSTAT, and discusses how the INDSTAT database could be improved by using big data. Section 5.3 then presents three indicative case studies that investigate how different sources of big data can be turned into such solutions for UNIDO.

5.2 BIG DATA FOR MISSING VALUE TREATMENT

Missing values are a familiar problem at UNIDO when maintaining the organization's statistical databases, including INDSTAT. Currently, UNIDO uses a few external data sources for improving the quality of the reported data. For instance, a common application of the Economist Intelligence Unit[1] (EIU) is to validate data reported from official sources to identify any errors. The benefits of utilizing external big data can go far beyond data validation. The extensive information contained in various data sources can be a very helpful tool to improve the quality of the INDSTAT database by providing better solutions for missing values. This section discusses different types of missing values in the INDSTAT database and points out the importance of big data to improve the current treatment.

As summarized by Todorov (2017), there are six patterns of missing values in the INDSTAT database:

a. Incomplete period coverage;
b. Significant time lag between data reporting and the latest reference year;
c. Incomplete variable coverage;
d. Infrequent conducting of censuses or surveys;
e. Missing data for years when changing the ISIC revision;
f. Data suppressed due to confidentiality reasons (at 2- or 3-digit level).

All types of missing values require extra attention. It is also worth noting that they can be divided into two categories depending on their causes and required treatments. In the case of point 2 above, missing values typically occur at the end of the time series when data is not ready yet. The goal is then to produce so-called nowcasted values up to the current year. Such estimates are considered provisional and are to be replaced as soon as officially reported values to become available. On the

1 Available at: https://www.eiu.com

other hand, missing values belonging to any of the other points above are genuinely missing and require imputed values. The true values will not be attained.

So far, UNIDO's imputation and nowcasting techniques are solely based on the available internal data (Todorov 2017). However, in the academic literature, various projects aiming to estimate economic statistics with big data sources and techniques have already been explored and examined.

There are several potential benefits of utilizing big data resources. Firstly, it is possible that big data can provide UNIDO with more complete estimates, especially when data is very sparse as a result of the country's not yet well-established statistical system. Some developing countries are not equipped with adequate resources to set up and maintain the data collection processes. In this case, the effect of both individual and cross country-based statistical analysis is compromised. For these countries, where data availability is always a well-known issue, bridging the data gaps becomes an even more crucial but complicated challenge. Besides, it is often the same countries that lack data for macro-data and often for the same types of industrial indicators. The current methodologies adopted within UNIDO may not be sufficient due to the high level of data scarcity. Additional relevant data become a necessity for imputations under these circumstances. Big data may help statisticians find out predictors with high correlation with the missing variable(s) and replace the missing values with the outcome of predictive models. By doing so, the effect of budgetary and regularization constraints can be diminished.

Secondly, big data may provide more timely and informative estimates. For years, UNIDO statistical publications have been released only on an annual basis. Yet, data users, especially policymakers and business associations, are mainly interested in more recent or frequent assessments of the overall production growth trends. In 2011, UNIDO started to produce quarterly reports on manufacturing. The source data is monthly or quarterly industrial production indices (IIP) compiled and disseminated by NSOs. However, not all countries produce official estimates in a sufficiently timely manner to meet the schedule.

In many cases, the complied official statistics are also only available significantly later than the reference period. According to Yeats (1990), while developed countries have continuous time series with only a one-year lag, the African economies usually have incomplete data series with a lag of 6 years.[2] The data presented in the INDSTAT database suffers by definition of a time lag of up to three years, which could grow to five, six, or more years in many cases (especially for developing countries). Both the NSOs and UNIDO must find a way to speed up data availability drastically.

Nowcast estimates are always required. However, the current nowcasting model adopted for the INDSTAT database is based on past and some relevant contemptuous indicators, which are still published with a time lag. Nowcasting models built with big data sources are less constrained by the availability of economic indicators since many of the big data sources are available almost real-time. The goal of exploiting real-time data, in this case, is to enhance the prediction accuracy of early estimates and to increase the releases' timeliness. Besides, by capturing the latest trends, the aggregated statistics are able to take into account the influence of major events such as conflicts and financial crises. In such circumstances, the time lag of official statistics reports becomes a great concern for policymakers.

2 See Luken et al. (2020) and the references therein for a more recent review of the data availability in Sub-Saharan Africa.

In summary, unofficial big data may represent a new approach for UNIDO to tackle challenges caused by the limitations of official data. The next step must consider how existing work on big data analytics can be incorporated within the UNIDO statistical production process to provide new opportunities for UNIDO database compilation.

5.3 CASE STUDIES

This section presents three case studies of big data sources that could aid official statistics production. The criteria for selecting the sources are:

a. They must contain geographical information that allows discrimination at least at the country-level;
b. Accessibility to the sources is relatively reliable and consistent;
c. The acquirement and processing of data do not violate any legislative and privacy rules.

Based on these criteria, nightlight satellite images, newspapers, and articles, as well as search query data, are selected. For each case study, the relevant data source is introduced, and the general methodologies used for missing value treatment are summarized. They also discuss how they may be adapted for data estimation in the INDSTAT database. These applications may reveal some possible directions for feasible projects at UNIDO that could be carried out in the future.

Supporting completeness – indicator imputations

Case Study 1- Nightlight satellite images
Satellite images are considered a relatively reliable data source since they are collected digitally without data loss or human intervention. This data source corresponds to category 3 of the big data classification produced by the United Nations Economic

Commission for Europe in 2013.[3] During the last decade, the usefulness of nightlight intensity estimate economic output has been drawing increasing attention. Of particular interest has been applications that range from eliminating the measurement error in official gross domestic product (GDP) (Ghosh et al. 2010; Hu and Yao 2019), regional development analysis (Michalopoulos and Papaioannou 2013) to the evaluation of the accuracy of national income (Chen and Nordhaus 2015; Henderson et al. 2012; Pinkovskiy and Sala-i-Martin 2016). Donaldson and Storeygard (2016) presented a review of applications of satellite data in economics. They summarized three main advantages of such remote sensing data to economic statistical analysis:

a. Access to information is difficult to obtain by other means;
b. Unusually high spatial resolution;
c. Wide geographic coverage.

At the same time, the increasing supply of satellite images from firms and research organizations has provided easier accessibility of such data. Among all types of satellite image sources such as urban/agricultural land use, climate and weather, nightlight images have drawn particular attention due to their simplicity and significance as a proxy for economic activities.

Most images used in the analysis are available at no cost from online repositories such as the Earth Observation Group (EOG) at the National Oceanic and Atmospheric Administration's (NOAA) National Geophysical Data Center (NGDC).[4] Earlier work was primarily concentrated on analysis of the annual images

3 UNECE classification of big data: www1.unece.org/stat/platform/display/bigdata/Classification+of+Types+of+Big+Data.
4 NOAA National Centers for Environmental Information Earth Observation Group, available at: https://ngdc.noaa.gov/eog.

produced by the Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS) instrument[5] from 1992 to 2013 (Figure 5.1). Then DMSP data are not publicly available after 2013. Instead, improved image data from the Visible Infrared Imaging Radiometer Suite (VIIRS) onboard the NOAA's Suomi National Polar-orbiting Partnership (NPP) satellite[6] launched in October 2011 are used for later research. In addition to more stable image quality with less noise, another significant advantage of VIIRS is that monthly composite images are produced instead of yearly. The increased amount of data is beneficial for modelling. Researchers have also sought to integrate these two datasets to observe a consistent temporal trend (Li et al. 2020).

In terms of estimating the economic activity of production, the underlying assumption of work in this direction is that there is a positive correlation between the amount of light, and the amount of production within a region. Galimberti (2017) evaluated the usefulness of night-time satellite images for to predict of annual GDP growth across a sample of 172 countries from 1992 to 2017 with DMSP/OLS images. The author tested two sets of model specifications, namely, a panel and country-individual specifications. Individually estimated models tend to outperform the pooled specifications. The author asserted that the use of night lights data is advantageous for GDP growth forecasting, particularly with individually estimated models, which achieve in-sample accuracy improvements ranging from 2.9 percent to 7.2 percent (cross-country weighted averages) relative to the benchmark model without nightlight indicators.

5 DMSP-OLS Nighttime Lights images are available at: https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html
6 VIIRS Nighttime Lights images are available at: https://ncc.nesdis.noaa.gov/VIIRS/

Nonetheless, for out-of-sample estimation, the performance of the individual specifications deteriorates, resulting from the small sample size used for model building (only nine data points available in the period 1992-2000). Interestingly, wealthier countries tend to be less prone to the effects of such estimation uncertainty and adding indicators of night lights still benefits. A similar GDP estimation task was carried out by Debbich (2019). The author employed VIIRS images to estimate the oil and non-oil real GDP for 72 countries from the MENAP (Middle East, North Africa, Afghanistan, Pakistan), CCA (Caucasus, Central Asia) regions, and sub-Saharan Africa from 2012 to 2017. This work demonstrated a strong positive relationship between nightlights intensity and real GDP level for the selected countries (see Figure 5 in Debbich (2019)). This result is coherent with the underlying assumption mentioned earlier.

Based on the estimation results, the author concluded that in countries where infrastructure for official statistical production is poor or countries with a conflict-setting, satellite images act as a precious tool to assess and characterize economic developments.

A methodological framework of how to turn the images into such a tool can be summarized. Firstly, pre-processing of the images starts with relatively direct conversion of sensed quantities for individual pixels to physical quantities of interest. The intensity of the night light's radiance is converted into 6-bit digital number values for analysis. Figure 5.1 shows an illustrative snapshot (left) and the transferred graph represented by digital number values ranging from 0 to 63 (right). Afterward, the data preparation procedures usually involve the following steps:

a. Inter-calibrate the light intensity over time to re-scale the parameters across the satellite-year composites;
b. Take the average if multiple images of the same region are produced by different satellites;

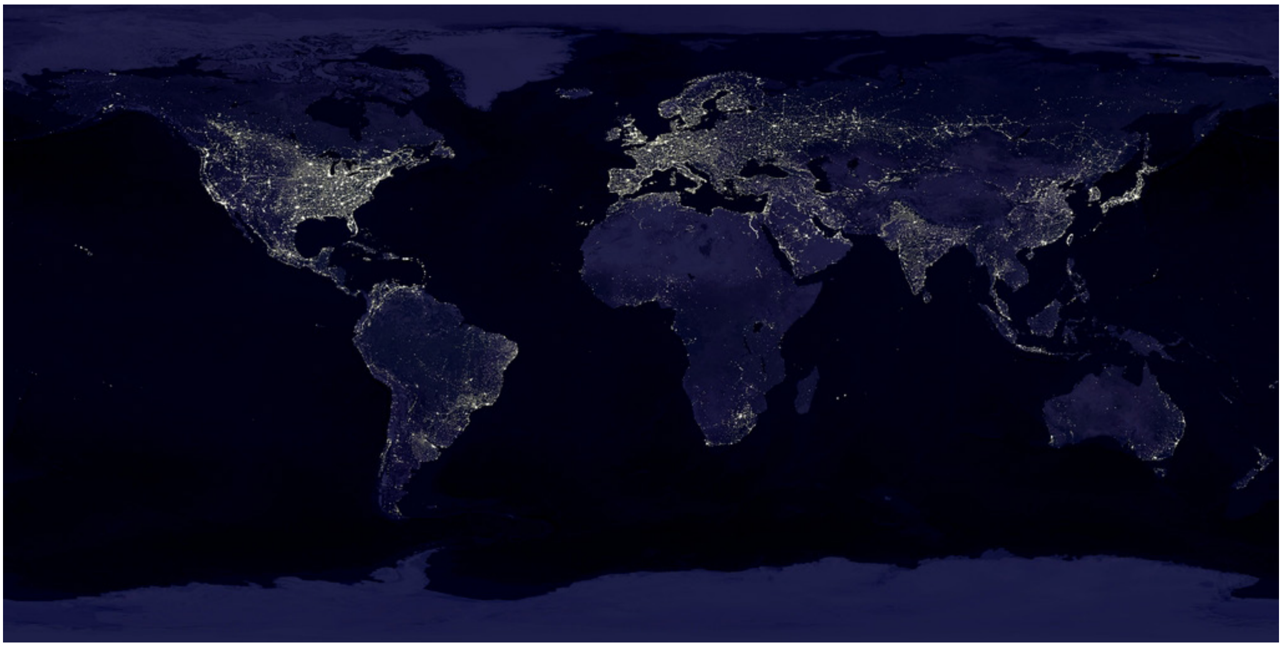

FIG. 5.1 | WORLD NIGHT LIGHT INTENSITY: AN EXAMPLE OF NIGHT-TIME LIGHTS IMAGERY

SOURCE | IMAGE AND DATA PROCESSING BY NOAA NATIONAL CENTER. DMPS DATA COLLECTED BY THE US AIR FORCE WEATHER AGENCY.

c. Re-scale the light intensity based on latitude due to Earth's curvature;
d. Draw countries or territories borders to split the images;
e. Limit the noise of data by, for example, removing regions with no urban centers or no human activity such as oil production plants, and eliminate ephemeral lights stemming from forest fires and so on;
f. Extract nightlight indicators;
g. Detect and remove outliers due to country-specific disruptive events;
h. Build predictive regression models, e.g., AR(1) based on the designed predictors.

Possible nightlight indicators that can be used in a model include:

a. Aggregated indicators:
   - e.g., Sum of Light (SoL): the sum of light intensity observed within a country;
b. Distribution-based indicators:
   - Digital number median, kurtosis, skewness, the spatial Gini coefficient;
c. Location-based to decompose a country's SoL into pixels:
   - Pixel-by-pixel time series of light intensity changes;
   - Grading of the intensity of each pixel compared with its surroundings;
d. Other variables
   - Fragile States Index;
   - Indicators from other sources.

When implementing these approaches, one should be mindful that the extent to which night-time lights are useful as a proxy for real economic activity may differ over time and across countries. Some preliminary work evaluating the underlying assumption asserting a positive correlation between the amount of nightlight and the desired response variable other than GDP growth will have to be carried out. Although current work mainly focuses on predicting GDP growth, extensions to predictions beyond a country's GDP growth are possible. Whether and how nightlight images can be used to capture the countries' industrial development can be seen as a direction for future work. As a starting point, similar models can be built with gross output growth or value added growth as the target variable since they are known to be strongly connected to GDP. Additional predictors that may improve the imputation of missing values can also be added where appropriate. For instance, a predictive model for gross output or value added can be constructed by optimally combining the available official statistics and extracted information from nightlights and geographic location.

Case study 2: Documents and texts

Qualitative information can also be used as a proxy for economic activity. Nowadays, the medium of publishing news and events has become faster. Immense amounts of articles become available online every minute of every day. Articles from reputable sources are good ways of evidencing public opinion. This source of data corresponds to category 1 of the 2013 UNECE classification of big data. Natural language processing (NLP) tools can extract such information to carry out rigorous economic statistical analysis. In particular, sentiment analysis is one of the most common applications in the NLP's research domain, which helps gauge the tone of texts pooled from various publication venues.

When identifying a suitable data source, it is always good to avoid bias by looking at reports from alternative sources on the same subject. Economic and industrial related news articles and national surveys can all be utilized as the data set for carrying out sentiment analysis. More specific data collection range may be country dependent. For instance, the Inflation Expectations Survey of Households (IESH) conducted by the Reserve Bank of India and the Indian daily newspaper "The Hindu,"[7] along with other forms of information are selected by Kumar (2018) to predict inflation trend in India. Tetlock (2007) evaluates the sentiment of Wall Street Journal articles, while Uhl (2010) uses sentiment data of the Washington Post, USA Today, the Houston Chronicle, the New York Times, and The Wall Street Journal to forecast private consumption in the United States.

7 Available at http://www.thehindu.com/archive/, containing archives from January 2000 onwards.

There have also been some efforts made on industrial production prediction. Ulbricht et al. (2017) selected the most influential German general and economic newspapers (Frankfurter Allgemeine Zeitung, Handelsblatt, and Süddeutsche Zeitung) as well as a business journal (WirtschaftsWoche) to carry out real-time out-of-sample forecasting based on word count indices and sentiment indices from January 2001 to April 2014. A gain in prediction accuracy was observed for 10- to 12- month horizon forecasts. Ardia and Bluteau (2017) also designed a similar framework to predict industrial production in the United States' from January 2001 to December 2016. Experiments have shown incorporating of qualitative predictors improves the prediction accuracy for 9- and 12-month horizons compared with models built on purely quantitative economic and financial indicators. The "time lag" effect is reasonable as it takes time for the sentiment to affect economic behaviors and become visible in the published figures. One of the major advantages of this later work is that the framework adapts itself to the changing environment.

A step-by-step methodology for predictor extraction through text-mining can be summarized as below:

a. Choose a number of reputable article providers to select articles from. Decide the relevant topics based on expert opinion, choose a subset of articles from each topic from each provider;
b. Choose the lexicons that will be used[8];
c. Compute the sentiment $\{SENTI_{i,t,l}\}$ for each text i at time t based on lexicon l by calculating the appearance-frequency

8 A lexicon is a list of word associated with a pre-defined sentiment orientation (positive/negative) and a sentiment strength.

of words that are labeled as positive, negative, or neutral in the lexicon;

d. For each topic k, form an aggregated sentiment vector $\{SENTIMENT_{k,t}\}$ based on $\{SENTI_{i,t,l}\}$;
e. Normalize the aggregated sentiment matrix. Calibrate the matrix where appropriate. e.g., weigh the source of the article providers according to their credibility;
f. Build a predictive model.

One key benefit of utilizing sentiment analysis based on NLP is that there are many readily-available technical software and packages. Along with their research paper, Ardia and Bluteau (2017) also provide a R[9] package sentometrics, which implements all steps described, hoping to encourage practitioners to use it. There are plenty of other useful tools for sentiment analysis as well. Many lexicons are available through R package lexicons (e.g., Tyler 2019), including some field-specific ones such as the Loughran-McDonald Master Dictionary (Loughran and McDonald 2011) specifically designed for financial and economic discourse presented in Table 5.1. When necessary, researchers can even generate a unique domain-specific lexicon with probabilities and information-theoretic techniques. For instance, Shapiro et al. (2018) developed a sentiment-scoring model, which includes a new lexicon explicitly built to capture the sentiment in economic news articles.

UNIDO can adopt similar methodologies to aid the prediction of value added and gross output with sentiment information when these figures are missing. The combination of such qualitative information and any already-available quantitative information

9 R is a programming language widely used among statisticians and data miners for developing statistical software and data analysis.

is the most desired. It may be challenging for UNIDO that the imputation has to be carried out for a large number of countries. In this case, language complexity and inconsistency may become an issue. Although there are also lexicons in other languages such as Spanish (Rangel et al. 2014) and German (Waltinger 2010), it is possible that context-specific lexicons have to be built from scratch for less developed countries. Usually, several lexicons are used for one set of texts in order to increase the model's generalization power. To make accurate predictions, the generation and mainteinance of various lexicons require expert knowledge.

TABLE 5.1 | LEXICON EXAMPLE

Word lists	Number of words	Examples
Negative	2337	litigation, termination, discontinued, penalties, unpaid, felony, investigation, misstatement, misconduct, forfeiture, serious, allegedly, noncompliance, deterioration
Positive	353	achieve, attain, efficient, improve, profitable, upturn
Uncertainty	285	approximate, contingency, depend, fluctuate, indefinite, uncertain, variability
Strong modal	19	always, highest, must, will
Weak modal	27	could, depending, might, possibly

SOURCE | AUTHORS' ELABORATION BASED ON THE LOUGHRAN MCDONALD MASTER DICTIONARY (LOUGHRAN AND MCDONALD 2011).

Improving timeliness – indicator nowcasts

Case Study 3: Google search query

Another interesting big data source is search query data, which illustrates the public's interest in the demand for information. This source of data corresponds to category 2 of the 2013 UNECE classification of big data. Google Trends[10] is a platform that provides a public view of internet search volumes and interests. It offers real-time data from more than 100-billion searches through the engine. The search query data are available globally from 2004 onwards. Since launching Google Trends, researchers from various fields have been keen to explore what can be learned from it. A plethora of studies has also proved that search trends data can be a supportive tool for carrying out effective economic studies.

Case study 3 shows that timeliness is a key advantage of search trends data compared with the data sources introduced in the above two case studies. Economic researchers have applied the popularity of web searches as an indicator of contemporaneous economic activity before the official data become available and/ or are revised (Choi and Varian 2012). For instance, plenty of work has been carried out on nowcasting of the labor market. The novel work by Askitas and Zimmermann (2009) shows the strong correlations between specific Google keyword searches and monthly unemployment rates in Germany.

Studies inspired thereof have been made for unemployment rate nowcasting based on Google search volumes in many other

10 Google Trends data are available at https://trends.google.com/trends/.

countries such as the United States (D’Amuri and Marcucci 2017; Tuhkuri 2015), the United Kingdom (McLaren and Shanbhogue 2011), Israel (Suhoy 2009), Italy (D‘Amuri 2009), Norway (Ellingsen 2017), Turkey (Chadwick and Gönül 2015), Brazil (Lasso and Snijders 2016) and Portugal (Ferreira 2014). The results are quite promising. For instance, Chadwick and Gönül (2015) claim that the model containing the search query data is 47.7 percent more accurate in-sample and 38.4 percent more accurate out-of-sample in terms of relative root mean square errors compared to a benchmark autoregressive model for the non-agricultural unemployment rate in Turkey. The extensive line of studies in this area confirms the assumption that search query data are informative indicators for nowcasting.

Most of the studies mentioned above employ a similar methodology, which can be summarized as follows:

a. Select a range of search terms that might describe the conditions of a labor market. Most state-of-the-art systems adopt keywords that are typically identified by a domain expert. However, researchers are also investigating more sophisticated ways to identify more topic-specific keywords for nowcasting various economic variables (Combes and Bortoli 2016; Ross 2013).
b. Combine keywords with operators “and” or “or” when appropriate;
c. Define the specific time period (e.g., month) and geographic area for data extraction;
d. Extract the search query data directly or the respective index time series computed as the total query volume for a specific keyword that was searched within a certain geographic area divided by the total number of queries in the same area;
e. Cross-sectional mean values may be used to reconstruct the series to get rid of the sampling noise;
f. Dimensionality reduction tool is employed to retain only the most significant components;
g. Remove seasonal variation with seasonal adjustment tools;

h. Build a dynamic factor model or a linear regression model for nowcasting.

Note that when a specific country is of interest, the translated version of the search keywords is considered. For instance, some common keywords used to describe the labor markets in the United Kingdom, the Netherlands, and Portugal are presented in Table 5.2. It is clear that by merely using native words and country-specific governmental bodies, any particular search query's coverage is automatically narrowed down. When a language is shared by more than one country, the geographic information from the IP address where the search is initiated can be explored to reveal the situation of the country in question. These characteristics are beneficial for UNIDO because country-specific nowcast can be adapted without much extra effort. Although 'Number of Employees', one of the key indicators within the INDSTAT database, is a different criterion to the unemployment rate, they are both used to describe the labor market. Therefore, an extension to nowcast 'Number of Employees' is worth investigating.

TABLE 5.2 | EXAMPLE OF SEARCH QUERY LIST FOR DIFFERENT COUNTRIES

United Kingdom (McLaren and Shanbhogue 2011)	"job", "unemployment", "unemployed", "unemployment benefit", "unemployment insurance", "Jobseeker's Allowance", "JSA"
Netherlands (te Brake 2017)	"UWV" (Institute for Employee Insurance), "Uitkering" (unemployment benefit), "Bijstand" (income support), "Werkloosheid" (unemployed) and "WW" (unemployment law)
Portugal (Ferreira 2014)	"Desemprego" (unemployment), "net empregos" (a popular Portuguese website for searching job offers), "ofertas emprego" (job offer), "subsidio desemprego" (unemployment benefits)

SOURCE | AUTHORS' ELABORATION BASED ON MCLAREN AND SHANBHOGUE (2011), TE BRAKE (2017), AND FERREIRA (2014)

5.4 CONCLUSION

As data-fuelled insights continue to highlight new efficiencies in data generation and collection, today, non-official big data is regarded as an innovative tool to provide comprehensive and timely insights for economic and industrial analysis. UNIDO considers big data resources to be a useful complement for official statistics production when information is unavailable or is fragmented. With appropriate information extraction methodologies and models in place, big data analytics may improve database quality and coverage and thus facilitate international comparisons in various aspects. This paper aims to inspire UNIDO statisticians on how certain big data sources may be used to provide more accurate imputation values and nowcasting values.

Three indicative case studies with different data sources have been presented to highlight possible future work in this direction. Before engaging in costly and time-consuming investments, the suitability, acceptability, feasibility, and sustainability of such should be carefully examined with some pilot experimental projects. Hammer et al. (2017) from the International Monetary Foundation have claimed that: “Sound partnerships, legal issues, and the right skills and technologies are as important as statistical expertise, data representativeness and methodological accuracy, and effective collaboration between data scientists and subject matter economists.” To fully and efficiently explore the strength of big data, participation and concerted action are needed from various parties. Coordination efforts are key.

5.5 REFERENCES

Ardia, D., Bluteau, K. and Boudt, K., 2019. Questioning the News About Economic Growth: Sparse Forecasting Using Thousands of News-Based Sentiment Values. International Journal of Forecasting, 35(4), pp. 1370-1386.

Askitas, N. and Zimmermann, K.Z., 2009. Google Econometrics and Unemployment Forecasting. Discussion Paper No. 4201. Bonn: Institute of Labor Economics (IZA).

Chadwick, M.G., and Şengül, G., 2015. Nowcasting the Unemployment Rate in Turkey: Let‘s Ask Google. Working Paper No: 12/18. Ankara: Central Bank of the Reublic of Turkey.

Chen, X., and Nordhaus, W., 2015. A Test of the New VIIRS Lights Data Set: Population and Economic Output in Africa. Remote Sensing, 7(4), pp. 4937-4947.

Choi, H., and Varian, H., 2012. Predicting the Present with Google Trends. Economic Record, 88(1), pp. 2-9.

Combes, S. and Bortoli, C., 2016. Nowcasting with Google Trends, the More is Not Always the Better. First International Conference on Advanced Research Methods and Analytics, CARMA16, July 2016, Valencia.

D‘Amuri, F., 2009. Predicting Unemployment in Short Samples with Internet Job Search Query Data. Munich Personal RePEc Archieve (MPRA) Paper No. 18403. Munich.

D’Amuri, F. and Marcucci, J., 2017. The Predictive Power of Google Searches in Forecasting US Unemployment. International Journal of Forecasting, 33(4), pp. 801-816.

Donaldson, D. and Storeygard, A., 2016. The View from Above: Applications of Satellite Data in Economics. Journal of Economic Perspectives, 30(4), pp. 171-198.

Debbich, M., 2019. Assessing Oil and Non-Oil GDP Growth from Space: An Application to Yemen 2012-17. IMF Working Paper 19/221. Washington, D.C.: International Monetary Fund.

Ellingsen, J., 2017. Let‘s Google It. Can Google Search Indices Nowcast Norwegian Retail Sales and Unemployment Rate? Master’s Thesis. Oslo: University of Oslo.

Ferrara, L. and Simoni, A., 2019. When Are Google Data Useful to Nowcast GDP? An Approach via Pre-Selection and Shrinkage. Working Paper 717. Paris: Bank of France [online] https://www.oecd.org/naec/new-economic-policymaking/Ferrara_Simoni.pdf

Ferreira, P., 2014. Improving Prediction of Unemployment Statistics with Google Trends:Part 2. Eurostat Working Paper.

Galimberti, J. K., 2017. Forecasting GDP Growth from Outer Space. KOF Working Papers, No. 427. Zurich: Swiss Economic Institute.

Ghosh, T., Powell, R.L., Elvidge, C.D., Baugh, K.E., Sutton, P.C., and Anderson, S., 2010. Shedding Light on the Global Distribution of Economic Activity. The Open Geography Journal, 3, pp. 148-161.

Hammer, C.L., Kostroch, D.C., Quirós, G. and STA Internal Group. 2017. Big Data: Potential, Challenges, and Statistical Implications. IMF Staff Discussion Note 17/06. Washington, D.C.: International Monetary Fund.

Henderson, J.V., Storeygard, A. and Weil., D.N., 2012. Measuring Economic Growth from Outer Space. American Economic Review, 102(2), pp. 994-1028.

Hu, Y. and Yao, J., 2019. Iluminating Economic Growth. IMF Working Paper 19/77. Washington, D.C.: International Monetary Fund.

Kumar, S., 2018. Using Machine Learning and Sentiment Analysis in Official Statistics. SSRN Electronic Journal, doi: 10.2139/ssrn.3193263.

Lasso, F. and Snijders, S., 2016. The Power of Google Search Data; An Alternative Approach to the Measurement of Unemployment in Brazil. Student Paper. Rotterdam: Erasmus University Rotterdam.

Li, X., Zhou, Y., Zhao, M. and Zhao, X., 2020. A Harmonized Global Nighttime Light Dataset 1992-2018. Scientific Data, 7(1), pp. 1-9.

Loughran, T. and McDonald, B., 2011. When iIs a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. The Journal of Finance, 66(1), pp. 35-65.

Luken, R., Moerec, U. and Mainert, T., 2020. Data Quality and Feasibility Issues with Industry-related Sustainable Development Goal Targets for Sub-Saharan African Countries. Sustainable Development, 28(1), pp. 91-100.

McLaren, N. and Shanbhogue, R., 2011. Using Internet Search Data as Economic Indicators. Bank of England Quarterly Bulletin, 51(2), pp. 134-140.

Michalopoulos, S. and Papaioannou, E., 2013. Pre-colonial Ethnic Institutions and Contemporary African Development. Econometrica, 81(1), pp. 113-152.

Pinkovskiy, M. and Sala-i-Martin, X., 2016. Lights, Camera ... Income! Illuminating the National Accounts-Household Surveys Debate. The Quarterly Journal of Economics, 131(2), pp. 579-631.

Rangel, I.D., Sidorov, G. and Guerra, S.S., 2014. Creación y Evaluación de un Diccionario Marcado con Emociones y Ponderado para el Español. Onomázein, 29, pp. 31-46.

Ross, A., 2013. Nowcasting with Google Trends: A Keyword Selection Method. Fraser of Allander Economic Commentary, 37(2), pp. 54-64.

Shapiro, A.H., Sudhof, M. and Wilson, D., 2020. Measuring News Sentiment Journal of Econometrics, in press.

Smith, H., 2018. Big Data for Official Statistics. New York, N.Y.: United Nations Statistics Division.

Suhoy, T., 2009. Query Indices and a 2008 Downturn: Israeli Data. Discussion Paper No. 2009.06. Bank of Israel Working Papers. Kiryat HaMemshala: Bank of Israel.

te Brake, G., 2017. Unemployment? Google It! Analyzing the Usability of Google Queries in Order to Predict Unemployment. Master's Thesis. Barcelona: University of Barcelona.

Tetlock, P.C., 2007. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. The Journal of Finance, 62(3), pp. 1139-1168.

Todorov, V., 2017. R in the Statistical Office: The Case of UNIDO. New Techniques and Technologies for Statistics 2017, 14-16 March, Brussels.

Tuhkuri, J., 2015. Big Data: Do Google Searches Predict Unemployment? Master's Thesis. Helsinki: University of Helsinki.

Tyler, R., 2019. lexicon: Lexicons for Text Analysis. Database. Available at: https://cran.r-project.org/web/packages/lexicon/index.html

Uhl, M.W., 2010. Explaining US Consumer Behavior and Expectations with News Sentiment. KOF Working Papers, No. 263. Zurich: Swiss Economic Institute.

Ulbricht, D., Kholodilin, K.A. and Thomas, T., 2017. Do Media Data Help to Predict German Industrial Production? Journal of Forecasting, 36(5), pp. 483-496.

United Nations, 2002. International Standard Industrial Classification of All Economic Activities (ISIC). Revision 3.1. Statistical Papers Series M NO. 4,REV.3.1. New York, N.Y.

United Nations, 2008. International Standard Industrial Classification of All Economic Activities. Revision 4. Statistical Papers Series M NO. 4,REV.4, New York, N.Y.

Upadhyaya, S., 2014. Use Of Non-official Sources for Transforming National Data Into an International Statistical Product – UNIDO's Experience. European Conference on Quality in Official Statistics (Q2014), Special Session on Measures for Enhancing the Quality of International Statistics: Serving Policy Making with International Statistics (Organized by the Committee for the Coordination of Statistical Activities (CCSA)), Vienna, June.

Waltinger, U., 2010. GermanPolarityClues: A Lexical Resource for German Sentiment Analysis. Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May, Valetta.

Yakovleva, K., 2018. Text Mining-based Economic Activity Estimation. Russian Journal of Money and Finance, 77(4), pp. 26-41.

Yeats, A.J., 1990. On the Accuracy of Economic Observations: Do Sub-Saharan Trade Statistics Mean Anything? The World Bank Economic Review, 4(2), pp. 135-156.

6

KENYA‘S INDUSTRIAL COMPETITIVENESS

By Nelson Correa and Valentin Todorov

6.1 INTRODUCTION

This paper analyses Kenya‘s industrial competitiveness. Kenya is an Eastern African country bordering Ethiopia, Somalia, South Sudan, the United Republic of Tanzania, and Uganda. It is a resource-based economy that has undertaken several efforts to advance on the path to industrialization. In 2008, the government launched the Kenya Vision 2030, a long-term development blueprint that acts as a national development strategy and roadmap. Its objective is to create a globally competitive and prosperous nation with a high quality of life by 2030. It aims to increase job and wealth creation in the manufacturing sector by increasing its contribution to GDP. In 2012, the Republic of Kenya published its National Industrialization Policy Framework for Kenya 2012–2030, with the objective of „transforming Kenya into a globally competitive regional industrial hub.“ More recently, in 2015, the Ministry of Industrialization and Enterprise Development unveiled „Kenya‘s Industrial Transformation Programme, „a strategic, comprehensive, and integrated program to guide the country on its path to industrialization (Republic of Kenya, 2008, 2012 and 2015).

The industrial competitiveness analysis presented here aims to examine the role and influence of Kenya‘s manufacturing sector, focusing on identifying the country‘s position in terms of competitiveness and potential. Specifically, Kenya‘s manufacturing production performance, export performance, level of technological upgrading and deepening, and global ranking will be reviewed using the most recent data from UNIDO databases. A number of indices related to manufacturing will be presented, especially in terms of market share, and revealed comparative advantages.

Kenya‘s competitiveness will also be assessed and analyzed by comparing it with that in three other countries: Côte d‘Ivoire, Ethiopia, and Sri Lanka. A comparison between Kenya and these three countries may be interesting in itself but only acquires true meaning when the reasons behind this selection of comparators are understood.

6.2 SELECTION OF COUNTRY COMPARATORS

The selection of country comparators is a delicate matter because evaluating what makes a ‘good’ (or ‘bad’) comparator country is intrinsically subjective and depends on individual perceptions. However, some questions can provide guidance in this regard. For example: Can the comparators offer useful information? For which activities can the comparators provide valuable input? What is a manageable number of comparators? Are these comparators immediate/potential competitors or, rather, role models? The answers to these questions may not result in us choosing a particular comparator but bearing these questions in mind during the selection process is recommended.

A more pragmatic approach is to use the widespread practice of comparing a country with its neighbors. This is often done due to the geographic adjacency and similarity in socio-economic structures. Moreover, neighboring countries often trade and compete with each other. In this regard, competition may take different forms, ranging from gaining market share in particular niches to compete for foreign investment when transnational companies try to gain access to their region.

Following this common practice, Ethiopia was chosen as the first comparator country for Kenya. Ethiopia is also a resource-based economy, with a similar contribution of its manufacturing sector to gross domestic product (GDP). The countries' levels of GDP and manufacturing value added (MVA) are comparable, which indicates that the size of the countries' internal markets could be comparable. Moreover, it suggests that these countries may be facing similar challenges in terms of economies of scale. Despite these similarities, it should be mentioned that Ethiopia is far more populous than Kenya and therefore—given their similar GDP and MVA levels—Kenya is considerably richer in terms of GDP per capita and more advanced in terms of industrial development when measured as MVA per capita.

The second comparator was also chosen on the basis of its geographic location, GDP per capita and productive structure. Bearing in mind that Ethiopia is more populous and has a lower GDP per capita and MVA per capita level, it is desirable to choose a similar African country that matches these differences in the opposite direction. Thus, Côte d'Ivoire was chosen as the second comparator because it is also an African country with a slightly higher income per capita than Kenya, it is less populous than Kenya and Ethiopia, but it is also a resource-based economy.

The development path of a country's industrial sector depends heavily on what it produces. In this regard, the fact that these three African economies (Côte d'Ivoire, Ethiopia, and Kenya) are resource-based economies may still be too broad. This is because their future industrial development path could depend on sector-specific aspects that could differ considerably, even among resource-based industries. Some examples of these differences in sector-specific characteristics could be production requirements (natural resources, labor force skills, capital or technology), market structures and integration in their global value chains, consumer demands and exposure to international trade, etc.

When looking at Kenya's main export products, tea stands out as a central product in the country's export mix. The list of world leaders (or the world's top competitors) in the production and export of tea includes China, India, and Sri Lanka. Sri Lanka is therefore selected as the third comparator country, given the size of the Chinese and Indian economies, which enjoy various benefits from their internal market and their economies of scale to the extent that is simply unattainable for the Kenyan economy. While Côte d'Ivoire and Ethiopia can be classified as immediate/potential competitors of Kenya, considering that they compete in several resource-based products, Sri Lanka is far more closed to be a role model. It has a higher GDP per capita and MVA per capita and a higher contribution of its manufacturing sector to the economy.

Table 1 summarizes general statistics. The first five columns on GDP, MVA, and population provide an idea of the relative size of the economies as well as how they have developed over time; the columns on GDP per capita and MVA per capita offer a static indication of the levels of economic and industrial development in these countries. The last column on the share of MVA in GDP indicates the manufacturing sector's relative contribution to the economy.

Table 1 also provides another piece of information: the only country with a manufacturing sector that grows significantly faster than the rest of Ethiopia's economy. Furthermore, Ethiopia's economy also recorded higher GDP and MVA growth rates from 1990 to 2019. Sri Lanka is second-best in GDP and MVA growth; both variables rose quickly and at a similar rate. Finally, the manufacturing sectors seem to have difficulties

TABLE 6.1 | GENERAL DATA IN 2019, SELECTED COUNTRIES

Economy	GDP (billions)	GDP annual growth rate (1990-2019, percent)	MVA (billions)	MVA annual growth rate (1990-2019, percent)	Population (millions)	GDP per capita (dollars)	MVA per capita (dollars)	Share of MVA in GDP (percent)
Côte d'Ivoire	44.5	3.4	6.0	3.2	25.7	1,729.5	235.2	13.6
Ethiopia	87.2	6.9	6.5	8.5	112.1	778.4	57.8	7.4
Kenya	79.8	3.9	6.7	2.5	52.6	1,517.8	126.6	8.3
Sri Lanka	92.2	5.1	15.0	5.2	21.3	4,325.0	705.4	16.3

SOURCE | ELABORATION BASED ON THE MANUFACTURING VALUE ADDED 2020 DATABASE (UNIDO 2020A).
NOTE | THE VALUES OF GDP AND MVA WERE MEASURED IN 2015 CONSTANT $ DOLLARS.

keeping up with GDP growth in Côte d'Ivoire and Kenya. Their GDP growth is still significant, but their MVA growth is sluggish, particularly in Kenya.

6.3 KENYA'S RANKING IN UNIDO'S COMPETITIVE INDUSTRIAL PERFORMANCE INDEX

Every two years, UNIDO publishes its Competitive Industrial Performance (CIP) report. The CIP index provides country measures of industrial competitiveness that enable cross-country comparisons. It measures how successful a country's industries are at producing and selling their goods on domestic and foreign markets, and consequently, how much they contribute to structural change and development.

The CIP index uses six indicators that cover three principal dimensions. These dimensions are i) the capacity to produce and export manufactured goods, ii) technological deepening and upgrading, and iii) world impact. The higher the scores in any of the three dimensions, the higher the country's industrial competitiveness and its CIP index.[1] Figure 1 provides a graphic explanation of how the CIP index is built.

Sri Lanka's manufacturing industry is the most competitive of the four case economies, occupying position 75 in the CIP ranking. Following Sri Lanka is Côte d'Ivoire in position 105, Kenya in 115, and finally, Ethiopia in position 134. These are the

1 The CIP report 2016 provides more information on the definitions, data sources and others for each of the components, as well as a detailed description of the methodology used to deal with missing values and to calculate the CIP index (UNIDO 2017).

FIG. 6.1 | DIMENSIONS OF THE COMPETITIVE INDUSTRIAL PERFORMANCE INDEX

SOURCE | UNIDO 2017.

countries' ranks in the current 2020 CIP edition based on data from 2018. It provides no further information on the development of these countries' industrial competitiveness over time.

Figure 2 presents the global CIP ranks for the selected countries between 1990 and 2018. It confirms the existing differences in these economies' industrial competitiveness and shows that the order between these countries in the industrial competitiveness ranking has not changed over the last three decades. Furthermore, it reveals that Sri Lanka and Ethiopia managed to achieve some progress and move up in the CIP global ranking during this period, while Côte d'Ivoire and Kenya registered the opposite trend, losing 5 and 10 positions, respectively. These

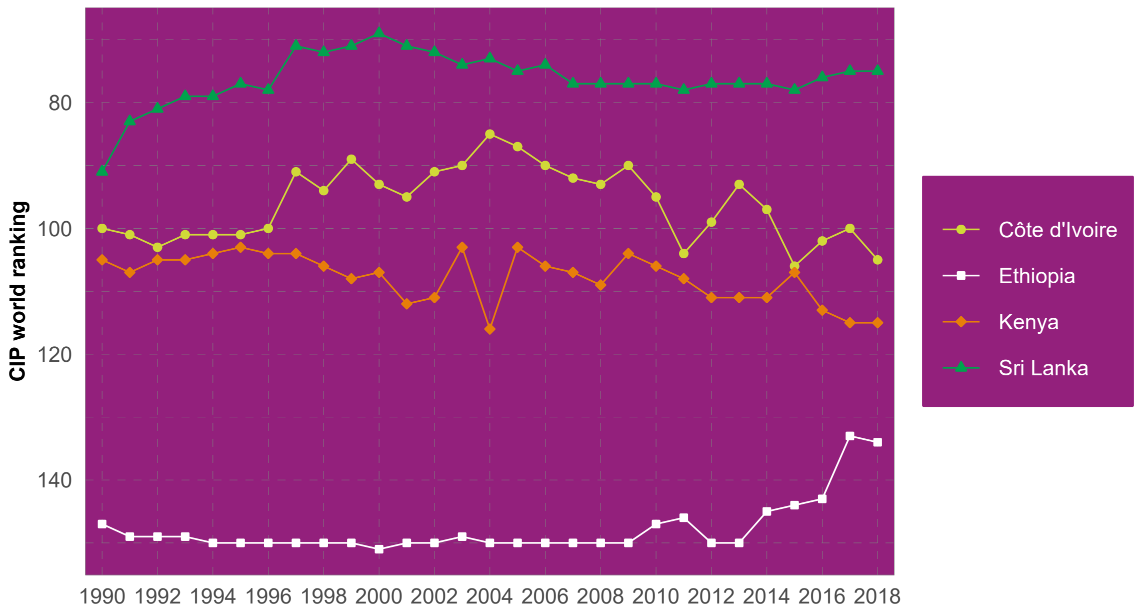

FIG. 6.2 | THE TREND OF CIP RANKING OF THE SELECTED COUNTRIES, 1990–2018

SOURCE | ELABORATION BASED ON THE COMPETITIVE INDUSTRIAL PERFORMANCE INDEX 2020 DATABASE (UNIDO 2020B).

trends have not, of course, been exempt from volatility, as demonstrated by the constant fluctuation of these countries in the global ranking.

6.4 CAPACITY TO PRODUCE

As already mentioned, the capacity to produce is one of the pillars of the CIP index and is critical for industrial competitiveness. High competitiveness requires a high capacity to produce a suitable amount of quality products within a specific to meet the requirements of domestic and foreign markets.

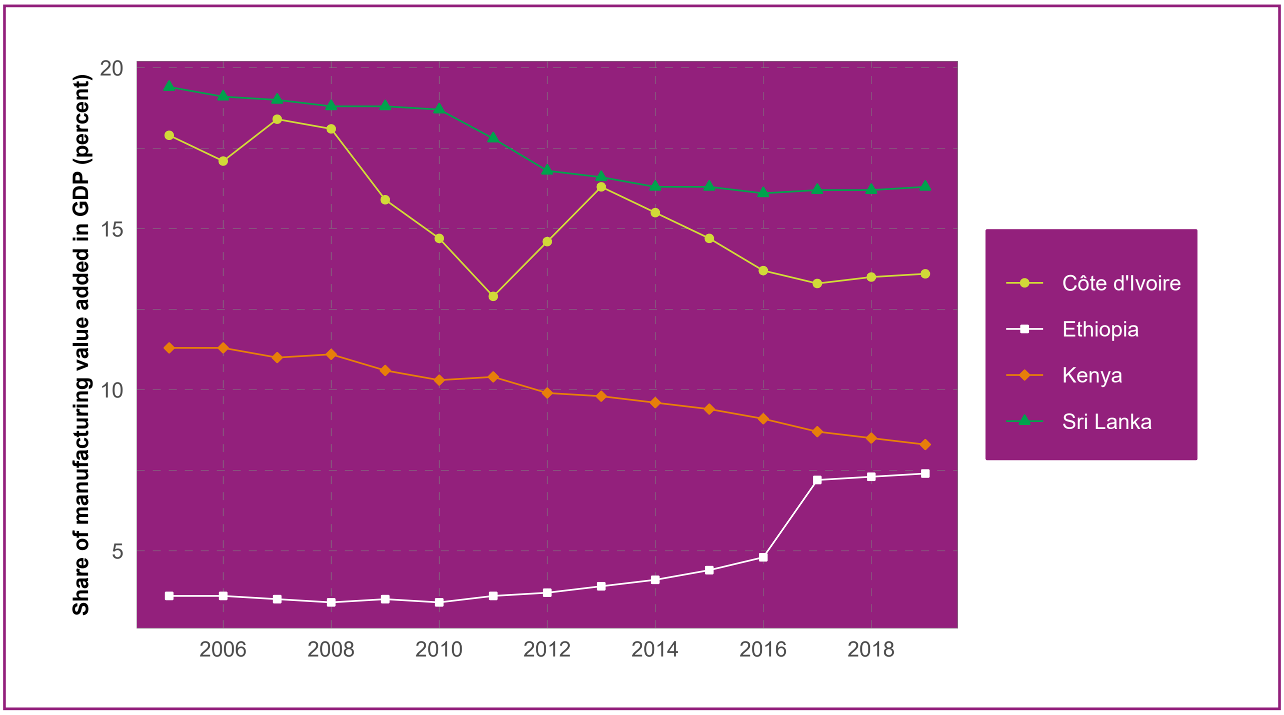

FIG. 6.3 | SHARE OF MVA IN GDP OF SELECTED COUNTRIES, 2005–2019

SOURCE | ELABORATION BASED ON THE MANUFACTURING VALUE ADDED 2020 DATABASE (UNIDO 2020A).
NOTE | THE UNDERLYING VALUES OF GDP AND MVA WERE MEASURED IN 2015 CONSTANT $ DOLLARS.

We expect that countries with a greater capacity to produce manufactured goods will also exhibit higher shares of MVA in GDP, as well as a higher MVA per capita. If manufacturing is indeed the engine of growth for a specific country, then the growth rate of MVA should be higher than the rest of the economy, which would imply an increasing share of MVA in GDP, together with a rapidly growing MVA per capita.

Figure 13 depicts the share of MVA in GDP for the four case countries. From 2005 to 2018, the contribution of Kenya's manufacturing industry to its economy fell from 11.3 per cent to 8.3 percent, by 27 percent. This decline is only comparable with that of Côte d'Ivoire, whose share fell by 24 percent over the

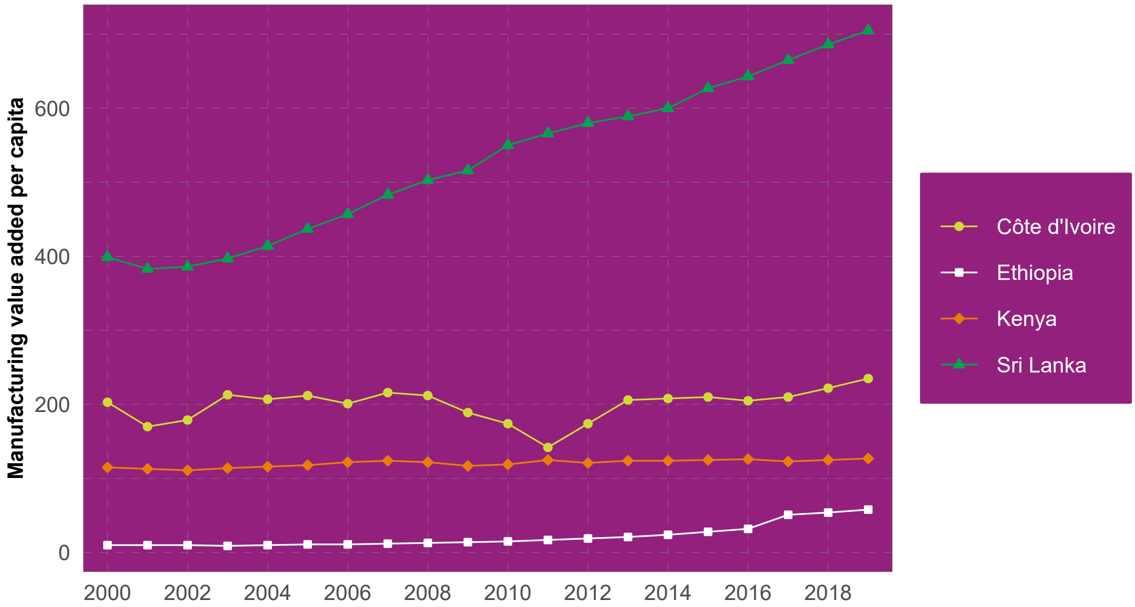

FIG. 6.4 | TREND OF MVA PER CAPITA IN SELECTED COUNTRIES, IN CONSTANT 2015 DOLLARS

SOURCE | ELABORATION BASED ON THE MANUFACTURING VALUE ADDED 2020 DATABASE (UNIDO 2020A).
NOTE | THE UNDERLYING VALUES OF GDP AND MVA WERE MEASURED IN 2015 CONSTANT $ DOLLARS.

same period. Sri Lanka also recorded a considerable decrease, with its share of MVA in GDP dropping by 16 percent. The exception to this negative trend was Ethiopia, which doubled its share over the last decade. In sum, according to this indicator (MVA in GDP), the decline in Kenya's share increased the gap between the country's production capacity and those countries with a more advanced industrial competitiveness (Côte d'Ivoire and Sri Lanka). The opposite occurred in Ethiopia, where the gap with Kenya narrowed considerably.

Another useful indicator for examining a country's production capacity is MVA per capita. This indicator allows us to compare the production capacities of economies with different population sizes. Figure 4 depicts the development of this indicator for the four selected economies.

The difference between this figure and the previous one is immediately visible. While Figure 3 indicates that the share of MVA in GDP declined in 3 out of 4 economies, Figure 4 illustrates that all economies registered slight or significant increases in MVA per capita. Does this mean that the two indicators contradict each other? The short answer is no. They simply provide different information.

The declining MVA share in GDP in Côte d'Ivoire, Kenya, and Sri Lanka does not necessarily imply that their manufacturing sector is producing less as time goes by. It means that their manufacturing sector is growing at a slower pace than the rest of the economy. In other words, the manufacturing sector cannot keep up with faster growth in different sectors of the economy. The opposite trend is visible in Ethiopia, where the manufacturing sector acts as the engine of economic growth.

MVA per capita increased in all four economies during 2000–2019, providing evidence of the manufacturing sector's growth relative to the countries' population size. While Côte d'Ivoire and Kenya only registered marginal increases, Sri Lanka and Ethiopia's manufacturing industries exhibited strong growth.

The positive trends in Sri Lanka and Ethiopia require particular attention. These economies have shown high and sustained industrial growth. Given the limited dimension of their internal markets, it is plausible to assume that international demand has played a tremendous role in their industrial development. This possibility is explored in the next subsection.

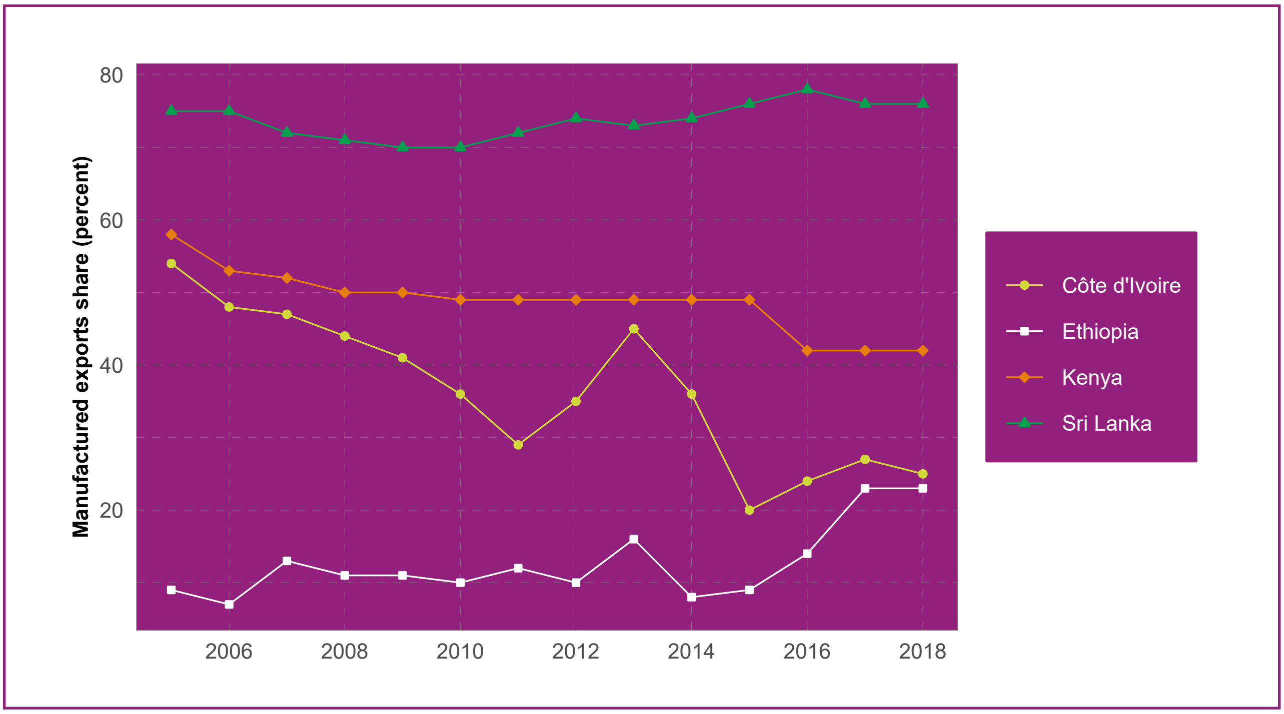

FIG. 6.5 | SHARE OF MANUFACTURED EXPORTS IN TOTAL EXPORTS IN SELECTED COUNTRIES, 2005-2018

SOURCE | ELABORATION BASED ON THE COMPETITIVE INDUSTRIAL PERFORMANCE INDEX 2020 DATABASE (UNIDO 2020B).
NOTE | THE UNDERLYING VALUES OF MANUFACTURED AND TOTAL EXPORTS WERE MEASURED IN CURRENT $ DOLLARS.

6.5 CAPACITY TO EXPORT

The capacity to export manufactured goods is another pillar of industrial competitiveness and reflects the domestic manufacturing industry's capacity to meet foreign demand. One widely used indicator to measure a country's capacity to export is its share of manufactured exports in total exports. The higher the manufacturing contribution to the country's total exports, the higher its capacity to export and its relevance for the economy in terms of GDP, trade balance, and foreign currency inflows.

Figure 5 shows the share of manufactured exports in total exports for all case countries. Once again, there appears to

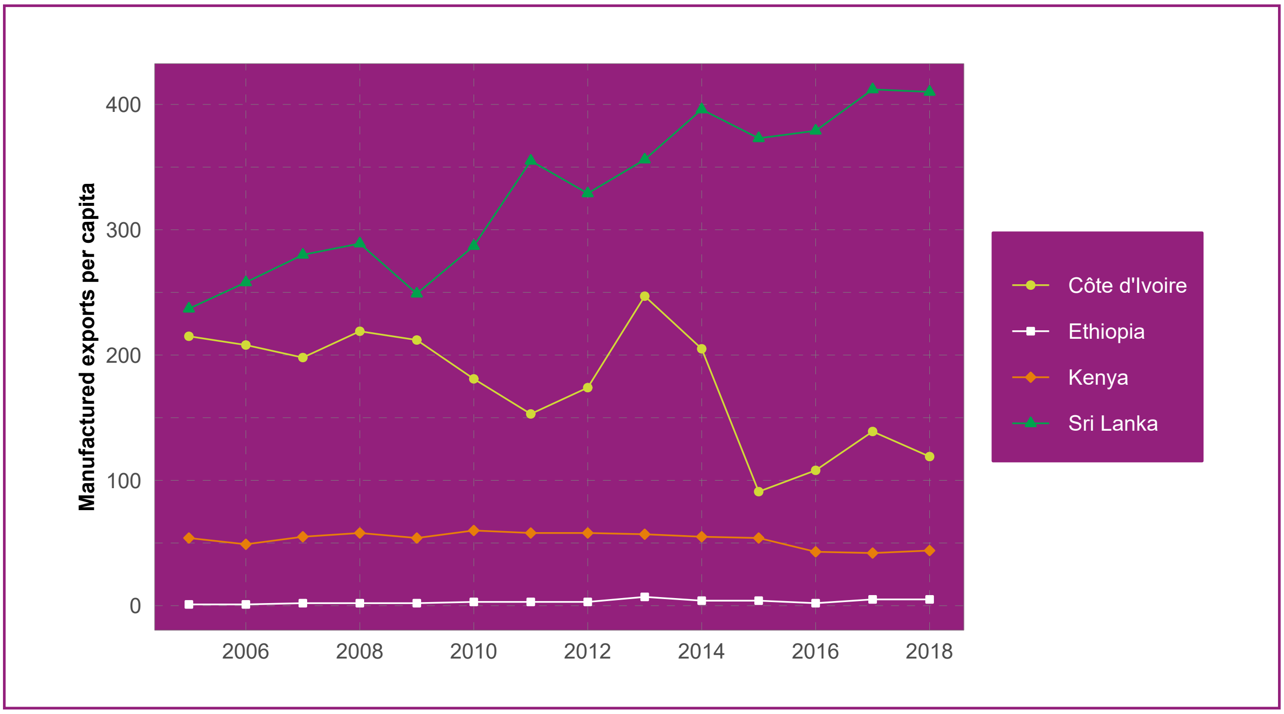

FIG. 6.6 | TREND OF MANUFACTURED EXPORTS PER CAPITA FOR THE SELECTED COUNTRIES

SOURCE | ELABORATION BASED ON THE COMPETITIVE INDUSTRIAL PERFORMANCE INDEX 2020 DATABASE (UNIDO 2020B).

be divergent trends between Côte d'Ivoire and Kenya and the other two economies. The share of Côte d'Ivoire and Kenya's manufactured exports in total exports declined significantly, plunging from 54 percent and 58 percent, respectively, in 2005 to 25 percent and 42 percent, respectively, in 2018. Sri Lanka's manufactured exports registered a high and stable contribution of around 75-76 percent to total exports. Ethiopia recorded a remarkable increase from 9 percent to 23 percent over the same period.

While the share of manufactured exports in total exports gives us an idea of how important the manufacturing sector is for the country's international trade, the indicator 'manufactured exports per capita' provides insights into the development of the country's export performance. This indicator is used to control the effects of population on the capacity to export manufactured goods.

Figure 6 reveals a familiar pattern. Once again, Côte d'Ivoire and Kenya show a decreasing trend, which is very different from Sri Lanka and Ethiopia's. Manufactured exports per capita fell in Côte d'Ivoire from $215 in 2005 to $119 in 2018. Kenya registered a similar decrease, from $54 to $44 for the same period. Sri Lanka recorded a sustained increase in manufactured exports per capita from $237 in 2005 to $410. Finally, Ethiopia's manufactured exports per capita rose from $1 to $5.

Furthermore, it is also interesting to note that in 2005, the value of manufactured exports per capita in Côte d'Ivoire was not particularly different from Sri Lanka's. Yet, 13 years later, their trajectories evolved in completely different directions, creating a considerable gap between these economies. The expansion of this gap in manufactured exports per capita indicates that the two economies took different paths in their approach to meet foreign demand for their manufactured products. These differences should be reflected in the technology upgrading and deepening of their exports, as well as in their market shares.

6.6 TECHNOLOGICAL DEEPENING AND UPGRADING

As mentioned in the previous sections, the capacity to move up the technological ladder is a pillar of industrial competitiveness. To examine the progress of the four case countries along the technological ladder, it is necessary to look at their export structure by technology group.

Table 2 describes the export structure by technology group in the four selected countries. Based on this information, the countries' diverse technological trajectories can immediately

TABLE 6.2 | STRUCTURE OF EXPORTS IN SELECTED COUNTRIES, 2005–2018 (IN PERCENT)

Technology group	Côte d'Ivoire		Ethiopia		Kenya		Sri Lanka	
	2005	2018	2005	2018	2005	2018	2005	2018
Primary products	44.8	66.5	85.5	71.7	40.8	56.5	22.8	23.1
Total manufacturing	54.5	25.3	9.2	23.3	58.1	42.0	75.1	76.3
Resource-based	31.9	16.1	5.1	3.4	32.7	18.0	16.0	17.7
Low-technology	2.7	3.3	4.0	10.1	16.7	15.0	53.2	50.5
Medium-technology	11.4	5.5	0.1	6.6	7.2	6.7	2.9	6.7
High-technology	8.4	0.5	0.1	3.1	1.6	2.3	3.0	1.3
Other transactions	0.8	8.2	5.3	5.0	1.1	1.5	2.1	0.7
Total exports	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

SOURCE | ELABORATION BASED ON THE COMPETITIVE INDUSTRIAL PERFORMANCE INDEX 2020 DATABASE (UNIDO 2020B).

be identified. Additionally, the information contained in Table 2 confirms the results presented in Figure 5: Sri Lanka is the only country in which manufactured products represent the bulk of their export mix. The share of manufactured products in Sri Lanka's total exports has been high and stable over the years, at around 75-76 percent. Low-technology products make up the biggest share of Sri Lanka's exports, representing around half of Sri Lanka's total exports.

Côte d'Ivoire and Kenya have taken a very different path. In the mid-2000s, manufactured products were the main contributor to the countries' total exports; however, this share fell considerably (as shown in Figure 4). Table 2 provides more insightful information, suggesting that the decrease in the share of manufactured exports in total exports is attributable to the decline in resource-based manufactured goods and an increase in the share of primary products.

Ethiopia is located at the other extreme, as it has not had a significant share of manufactured exports in total exports. Primary products have undoubtedly been its main source of exports, as demonstrated by the high share of primary products in total exports. Yet, it should be highlighted that Ethiopia has undertaken major efforts to improve its export mix and climb the technological ladder. Therefore, although the country's manufacturing share is still meager, it has grown considerably due to increases in low-, medium- and even high-technology manufactured products.

6.7 EXPORT MARKET SHARES

Table 3 presents the participation of the four selected economies in the world market by export market share by technology group. The last row of Table 3 clearly shows that their participation is minimal, as none of their total export market shares reached a level of 0.1 percent in world exports. Therefore, instead of addressing these countries' impact on the world market, this subsection will instead focus on the countries' export market shares and the underlying comparative advantages.

Among these countries, Côte d'Ivoire leads in terms of participation in world exports, with an export market share of 0.07 percent in world exports. This is not a minor achievement, mainly because Côte d'Ivoire is the smallest economy in the group when measured in total GDP. However, most of Côte d'Ivoire's export market share consists of primary products, and the only type of manufactured goods that achieved significant values were resource-based manufactured products. Even those experienced a sharp decline between 2005 and 2018. According to the data in Table 3, Côte d'Ivoire appears more

TABLE 6.3 | EXPORT MARKET SHARE BY TECHNOLOGY GROUP, IN PERCENT

Technology group	Côte d'Ivoire		Ethiopia		Kenya		Sri Lanka	
	2005	2018	2005	2018	2005	2018	2005	2018
Primary products	0.20	0.30	0.05	0.06	0.09	0.12	0.09	0.10
Total manufacturing	0.05	0.02	0.00	0.00	0.03	0.02	0.06	0.06
Resource-based	0.15	0.06	0.00	0.00	0.07	0.03	0.06	0.07
Low-technology	0.01	0.02	0.00	0.01	0.04	0.03	0.23	0.23
Medium-technology	0.03	0.01	0.00	0.00	0.01	0.01	0.01	0.01
High-technology	0.04	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Other transactions	0.01	0.09	0.01	0.01	0.01	0.01	0.02	0.01
Total exports	**0.07**	**0.07**	**0.01**	**0.01**	**0.03**	**0.03**	**0.06**	**0.06**

SOURCE | ELABORATION BASED ON THE COMPETITIVE INDUSTRIAL PERFORMANCE INDEX 2020 DATABASE (UNIDO 2020B).

integrated into the global economy than the other three countries. Still, the quality of this integration is relatively modest as most of its participation is based on products at the bottom of the technological ladder.

Sri Lanka ranks second in terms of participation in world exports, as indicated by its total export market share in world exports, which was stable at around 0.06 percent from 2005–2018. Contrary to Côte d'Ivoire and the other countries, Sri Lanka's largest market share is in low-technology products. As mentioned earlier, Sri Lanka is a world leader in the production of tea (primary product), but its textile industry (low-tech) is also substantial. Hence, even though Sri Lanka's export market share in primary products and resource-based manufactured goods is still significant, its participation in the global market is not exclusively based on the export of raw materials.

Kenya's market share in total exports trails far behind Côte d'Ivoire and Sri Lanka's, its share being only half of theirs. Additionally, the quality of Kenya's integration in global markets is also modest. As is the case for Côte d'Ivoire, Kenya's biggest export market share is also at the very bottom of the technological ladder, namely in the export of unprocessed natural resources (primary products). Although the country undertook significant efforts in the mid-2000s to add value by processing these resources, the effects of these efforts partially disappeared within a decade, as demonstrated by the decline in the market share of resource-based manufactured goods, which dropped from 0.07 percent in 2005 to 0.03 percent in 2018.

Ethiopia's economy is the least integrated into world trade. Despite being the most populous and the second-largest economy of the group measured in total GDP, Ethiopia has the lowest market share in total exports. Similar to Côte d'Ivoire and Kenya, Ethiopia's most prominent market share is in primary products. Despite the relative improvement in its export structure, the data in Table 10 indicates that Ethiopia is so specialized in primary products that all the other technology groups look almost irrelevant in comparison.

As in the previous sections, the analysis of Kenya's industrial competitiveness and its comparators continues with an evaluation of the revealed comparative advantage (RCA) indexes for each economy and technology group. In addition to the four countries' RCA indexes, Table 4 includes two additional columns on annual growth in world export by technology group to measure world demand. As in the previous sections, this information has been added to determine whether these countries have developed a comparative advantage in those sectors with growing international demand.

The annual growth rate from 2005–2018 provides an idea about the developments during that period. World exports of resource-based manufactured products grew at 5.2 percent per year from 2005 to 2018. After "other transactions,"[2] resource-based manufactured goods were the technology group with the biggest growth in international demand. This may come as a surprise, as technologically more advanced sectors usually exhibit higher growth, and yet, the reader should bear in mind that manufactured products such as food, refined oil, basic metals, and other similar products represented the core of the so-called commodity boom that started at the beginning of 21st century and lasted for over a decade. The commodity boom had a significant influence over the period 2005–2018, which means that choosing another period would most probably show a different growth ranking for these sectors.[3] High-technology manufactured goods recorded the second-highest growth rate following resource-based products, followed by low- and medium-tech product. Finally, primary products registered the lowest growth during this period.

Since it is unlikely that another commodity boom will soon come around, more recent growth rates might better serve as estimators of future demand. Table 4, therefore, presents the calculations of the annual growth rates over the last five years, that is, from 2013 to 2018. The order of the technology groups by growth rate is very similar to that presented in the previous sections. High-technology manufactured products rank higher than medium- and low-technology goods, trailed by resource-based goods and primary products, both of which recorded negative growth.

2 The analysis of the technology group "other transactions" has been excluded for the same reasons mentioned in previous sections.
3 In the previous sections, the growth rates were calculated after the independence of South Sudan, from 2012 to 2019, and the results were much more predictable: High-tech products registered the highest growth rate, followed by medium- and low-tech goods; they were followed by the resource-based group and finally by primary products, which registered the lowest growth rate.

In sum, this implies that while having a comparative advantage in resource-based goods in the period 2005–2018 was beneficial due to the fast growth in prices, it is now recommended to specialize in high- and medium-technology manufactured products, as the most recent growth rates, show that the commodity boom is over. Moreover, primary products and resource-based manufactured products are affected by high volatility in prices, which increases the country‘s risk and vulnerability to external shocks (Boly 2013).

In terms of revealed comparative advantages, it can be assumed - based on the previous tables - that all four economies will exhibit a comparative advantage in primary products. Yet, some interesting nuances emerge. For instance, contrary to the rest of the economies, Ethiopia did not reinforce this comparative advantage, as its RCA index fell from 5.2 to 4.9.

In the case of total manufacturing, Sri Lanka was the only economy that did not have a comparative disadvantage, as its RCA index remained constant and equal to 1, indicating that the country neither has a comparative advantage nor a comparative disadvantage in the export of manufactured products as a whole. It is also worth mentioning that while Côte d‘Ivoire and Kenya‘s comparative disadvantage increased, Ethiopia is the only country that moved in the opposite direction.

Interestingly, resource-based goods did not provide a clear comparative advantage in the countries considered, which is a huge missed opportunity as they did not manage to take full advantage of the commodity boom. Côte d‘Ivoire and Kenya suffered significant declines in their RCA indexes, thus eroding most of their comparative advantage in this particular technology group. Unfortunately, the technological trajectory for these countries is clear: they regressed from a situation in which they were adding some value to their natural resources to a new situation in which they export pure commodities without processing them.

TABLE 6.4 | REVEALED COMPARATIVE ADVANTAGE BY TECHNOLOGY GROUP

Technology group	Annual growth in world exports		Côte d'Ivoire		Ethiopia		Kenya		Sri Lanka	
	2005-2018	2013-2018	2005	2018	2005	2018	2005	2018	2005	2018
Primary products	3.9	(5.0)	2.7	4.6	5.2	4.9	2.5	3.9	1.4	1.6
Total manufacturing	4.8	1.4	0.7	0.3	0.1	0.3	0.7	0.5	1.0	1.0
Resource-based	5.2	(0.9)	2.0	1.0	0.3	0.2	2.0	1.1	1.0	1.0
Low-technology	4.7	1.4	0.2	0.2	0.3	0.7	1.2	1.0	3.7	3.5
Medium-technology	4.6	2.0	0.4	0.2	0.0	0.2	0.2	0.2	0.1	0.2
High-technology	4.9	2.9	0.5	0.0	0.0	0.2	0.1	0.1	0.2	0.1
Other transactions	5.9	1.6	0.1	1.3	1.0	0.8	0.2	0.2	0.4	0.1
Total exports	**4.7**	**0.3**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**

SOURCE | ELABORATION BASED ON THE COMPETITIVE INDUSTRIAL PERFORMANCE INDEX 2020 DATABASE (UNIDO 2020B).

Sri Lanka's biggest revealed comparative advantage is in low-technology products. RCA values reveal the importance of the low-technology industry (and particularly, the textile industry) for the country. This industry is also important for Kenya, and even though this technology group has lost some ground, Kenya has neither an advantage nor a disadvantage within it. Côte d'Ivoire and Ethiopia show a clear revealed comparative disadvantage. Yet, by opening to new markets and expanding their low-technology exports, Ethiopia is countering its comparative disadvantage in this technology group.

Regrettably, the RCA values indicate that there are very few opportunities for these countries to generate exports in medium- and high-technology products, given the fact that all these countries have a significant comparative disadvantage in industries that are at the top of the technological ladder.

6.8 CONCLUSION

Just like the rest of Africa, Kenya has been gradually industrializing over the last decade and a lot of work is still ahead. The 2020 CIP index ranks Kenya in the last quarter of the CIP global ranking, at position 115 out of 152 economies. The case study reveals some problems with Kenya‘s industrial competitiveness. In terms of its production and export capacity, Kenya exhibits somewhat negative signs. For example, Kenya‘s MVA share in GDP and its manufacturing share in total exports are decreasing. The results are mixed for technology, market share, and revealed comparative advantage.

On the one hand, Kenya demonstrates relatively high levels of manufactured exports in total exports, market share, and RCA indexes in resource-based and low-technology products. On the other hand, these levels have been decreasing considerably over time, and the only technology group that has reinforced its comparative advantage is primary products. In this sense, while the situation is not yet critical, Kenya's technological trajectory points towards a deterioration of its technological capabilities as activities that were adding value to its natural resources have been removed.

From the comparator countries‘ perspective, these results help explain why - according to the CIP index - Ethiopia is catching up with Kenya in terms of industrial competitiveness, while Kenya's gap is expanding with Côte d‘Ivoire and Sri Lanka.

Further research would be necessary to analyze Kenya‘s industrial competitiveness comprehensively. Sectoral MVA data for Kenya and its comparator countries would be needed to examine production and exports patterns. Additionally, an analysis of revealed comparative advantages and the growth in international demand would have been more meaningful at a more disaggregated level, i.e., replacing the technology groups with Kenya‘s most important exports.

6.9 REFERENCES

Boly, A., 2013. Industrial Development in Least Developed Countries. Development Policy, Statistics and Research Branch Working Paper 05/2012, UNIDO, Vienna: United Nations Industrial Development Organization.

Ministry of Industrialization, Trade and Enterprise Development (MoITED), 2012. Sessional Paper No. 9 of 2012 on the National Industrialization Policy Framework for Kenya 2012 – 2030. Nairobi.

MoITED, 2015. Kenya‘s Industrial Transformation Programme. Nairobi.

Ministry of State for Planning, National Development and Vision 2020, 2008. First Medium Term Plan, 2008-2012. Kenya Vision 2030: A Globally Competitive and Prosperous Kenya. Nairobi.

United Nations Industrial Development Report (UNIDO), 2017. Competitive Industrial Performance Report 2016. Volume I. Vienna.

UNIDO, 2020a. Manufacturing Value Added 2020. Database. Available at: https://stat.unido.org/database/MVA%20 2020,%20Manufacturing [Accessed December 2020].

UNIDO, 2020b. Competitive Industrial Performance Index, 2020 Edition. Database. Available at: https://stat.unido.org/database/CIP%202020 [Accessed December 2020].

FOOD PROCESSING INDUSTRIES IN INDIA

By Bivas Chaudhuri, Government of India, India
Valentin Todorov, United Nations Industrial Development Organization, Austria

7.1 INTRODUCTION

India's food processing sector has earned the denotation 'sunrise sector' due to its rapid growth in recent years and the expectation that it will gain prominence in the future. The country's diverse range of topography, soil quality, and climatic conditions make for advantageous conditions for the production of many agricultural products, both crops, and non-crops, and has made it possible to establish a large agricultural base. (Ghosh et al. 2013) While agriculture remains the primary source of livelihood for a large share of the country's population, the food processing sector is indispensable for the overall development of the Indian economy as it provides a vital linkage and synergy between the agricultural and industrial sectors. It helps diversify and commercialize farming and enhances their income. And it creates markets for the export of agri-foods and generates more employment opportunities. Such industries enable a broader range of food products to be sold and open up export opportunities. (Adukia 2012) If utilized properly, the food processing sector has the potential to become a driver of economic growth in India.

Food processing is commonly defined as a process of value addition to the agricultural, horticultural, plantation, animal husbandry or fishery produce by various methods like grading, sorting, and packaging" (source). Such processing methods enhance the quality of food and prolong shelf-life. Due to new technologies and emerging market channels, food processing is no longer confined to simple preservation (e.g., salting, curdling, drying, pickling). It now includes ready-to-eat food products, beverages, processed and frozen fruit and vegetable products, marine and meat products. India's many food processing industries are classified according to Divisions 10 and 11 of the 2008 National Industrial Classification (harmonized with the United Nations' International Standard Industrial Classification, Revision 4). Packaging and transportation are also no longer the primary services within the sector. Today, it also encompasses "the establishment of post-harvest infrastructure for the processing of various food items like cold storage facilities, food parks, packaging centers, value-added centers, irrigation facilities, and modernized abattoir". (Majumdar 2013)

In addition to the abundant supply of raw materials, the food processing sector has been stimulated by rising demand for (better) food products and public fiscal incentives. However, even though the sector in recent years has exhibited faster growth than agriculture and accounts for 32 percent (as of 2019) of the country's total food market, it is still widely considered an untapped opportunity for sustainable growth and resilience, and improvement in livelihood for the millions of people employed in the sector (IBEF 2017).

This paper provides an overview of the structure and economic significance of the food processing sector in India. Section 7.2

considers the sector's overall contribution to the Indian economy, including its international competitiveness. Section 7.3 details structural characteristics in terms of gender distribution, the sector's geographical presence across the country, and differences between organized and unorganized enterprises. The economic potential of the food processing sector is underlined through a glance at India's vast but underutilized agricultural sector in Section 7.4. Section 7.5 then outlines a selection of key industrial policies in place to further the development of food processing industries in the country. Section 7.6 concludes.

TABLE 7.1 | CONTRIBUTION OF THE FOOD PROCESSING SECTOR TO THE INDIAN ECONOMY
A | Gross value added (trillion Rupees)

	2011	2012	2013	2014	2015	2016	2017	2018
All economic activities	81,070	85,470	90,640	97,120	104,920	113,280	120,740	128,030
Manufacturing	14,100	14,870	15,610	16,840	19,040	20,550	21,910	23,170
Agriculture, forestry and fishing	15,000	15,240	16,090	16,060	16,160	17,260	18,280	18,720
Food processing	1,000	1,330	1.30	1,340	1,610	1,790	1,910	2,080

NOTES | ALL VALUES ARE IN 2011 RUPEES.
SOURCE | NATIONAL ACCOUNTS STATISTICS 2020 (MOSPI 2020A) AND MOFPI 2018 AND 2020.

7.2 SIGNIFICANCE OF THE FOOD PROCESSING SECTOR IN THE INDIAN ECONOMY

Gross value added

The food processing sector has emerged as an important segment of the Indian economy. In less than a decade, the sector's gross value has doubled from one trillion Rupees in 2011 to slightly more than 2 trillion Rupees in 2018 (see Table 7.1, panel a). Not only has the sector returned double-digit growth rates, but it almost performed as well as the overall economy in terms of annual average growth in the period (panel b). It Is particularly noticeable how it has grown faster than both the manufacturing and the agricultural sector since 2015. The sector's share of gross value added of all economic activities has also increased steadily from that point on (panel c). As of 2018, the food processing

TABLE 7.1 | CONTRIBUTION OF THE FOOD PROCESSING SECTOR TO THE INDIAN ECONOMY

B | Growth in gross value added (percent)

	2012	2013	2014	2015	2016	2017	2018	Average annual growth rate
All economic activities	5.43	6.05	7.15	8.03	7.97	6.59	6.04	6.75
Manufacturing	5.46	4.98	7.88	13.06	7.93	6.62	5.75	7.38
Agriculture, forestry, and fishing	1.53	5.58	-0.19	0.62	6.81	5.91	2.41	3.24
Food processing	-9.52	-2.26	3.08	20.15	11.18	6.70	8.90	5.46

NOTES | ALL VALUES ARE IN 2011 RUPEES.
SOURCE | NATIONAL ACCOUNTS STATISTICS 2020 (MOSPI 2020A) AND MOFPI 2018 AND 2020.

TABLE 7.1 | CONTRIBUTION OF THE FOOD PROCESSING SECTOR TO THE INDIAN ECONOMY

C | Share in gross value added of total economic activities (percent)

	2011	2012	2013	2014	2015	2016	2017	2018
Manufacturing	17.39	17.46	17.22	17.34	18.15	18.14	18.14	18.10
Agriculture, forestry, and fishing	18.53	17.83	17.75	16.54	15.40	15.24	15.14	14.62
Food processing	1.81	1.55	1.44	1.38	1.53	1.58	1.58	1.62

NOTES | ALL VALUES ARE IN 2011 RUPEES.
SOURCE | NATIONAL ACCOUNTS STATISTICS 2020 (MOSPI 2020A) AND MOFPI 2018 AND 2020.

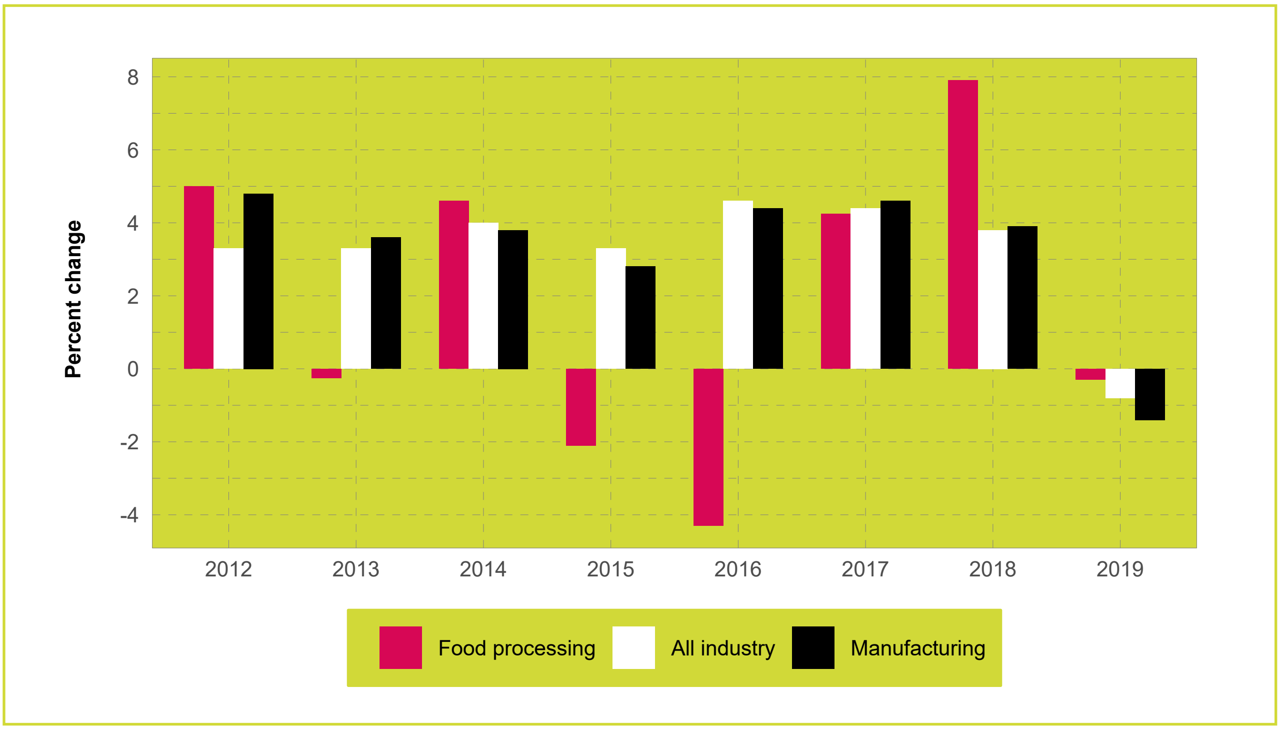

FIG. 7.1 | GROWTH IN THE INDEX OF INDUSTRIAL PRODUCTION FOR SELECTED SECTORS

SOURCE | AUTHOR ELABORATION BASED ON THE INDEX OF INDUSTRIAL PRODUCTION (MOSPI 2020B).

TABLE 7.2 | EMPLOYMENT IN REGISTERED AND UNREGISTERED FOOD PROCESSING FACTORIES

Sector	Food processing sector (thousands)	All economic activities (thousands)	Share (percent)
Registered (2017)	1,933	15,614	12.4
Un-registered (2015)	5,111	36,041	14.2

NOTES | SECTION 2(M) IN THE FACTORIES ACT, 1948 DEFINES A REGISTERED FACTORY AS ANY PREMISES THAT EMPLOY TEN OR MORE WORKERS AND UNDERTAKES A MANUFACTURING PROCESS WITH THE AID OF POWER OR EMPLOYS TWENTY OR MORE WORKERS AND UNDERTAKES A MANUFACTURING PROCESS WITHOUT THE AID OF POWER.
SOURCE | AUTHOR ELABORATION BASED ON THE 2017-2018 ANNUAL SURVEY OF INDUSTRIES (MOSPI 2018) AND NSSO 2018.

sector had grown to more than a tenth of the agricultural sector's size. However, compared to the economy overall, the impact of the sector seems less pronounced.

These trends are also reflected in the Index of Industrial Production (see Figure 7.1) as calculated by the National Statistical Office (MOSPI 2020b).

Employment

Food processing, particularly of fresh food, remains one of the most labor-intensive manufacturing sectors. As of 2017, 12.4 percent of employment generated in India's registered factories was in food processing, followed by the textile and wearing apparel sector (MOFPI 2017), see Table 7.2 and Figure 7.2. The share is slightly greater among unregistered factories at 14.2 percent. As is the case in many developing economies, India's unregistered sector is significant and is critical for the employment of the country's sizeable unskilled labor force. Whereas the total

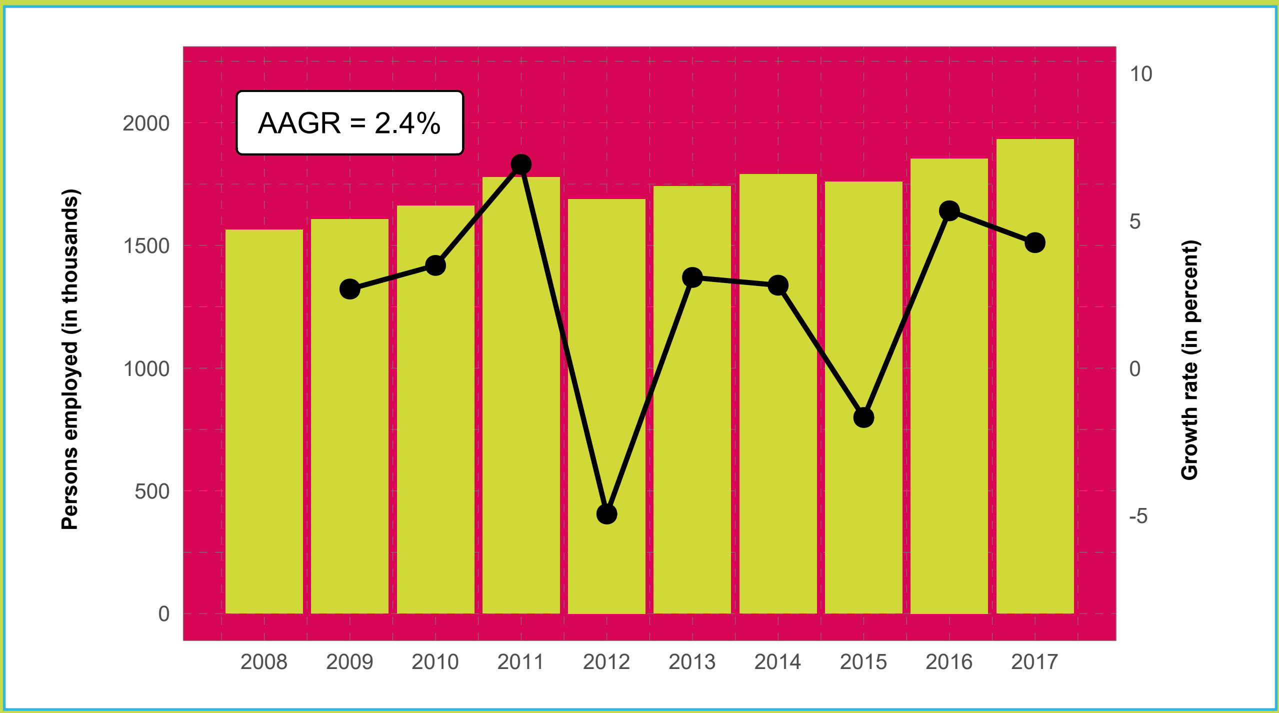

FIG. 7.2 | PERSON EMPLOYED IN REGISTERED FOOD PROCESSING FACTORIES

SOURCE | AUTHOR ELABORATION BASED ON THE ANNUAL SURVEY OF INDUSTRIES (MOSPI 2018).
NOTE | AAGR IS THE ANNUAL AVERAGE GROWTH RATE IN THE PERIOD 2008-2017.

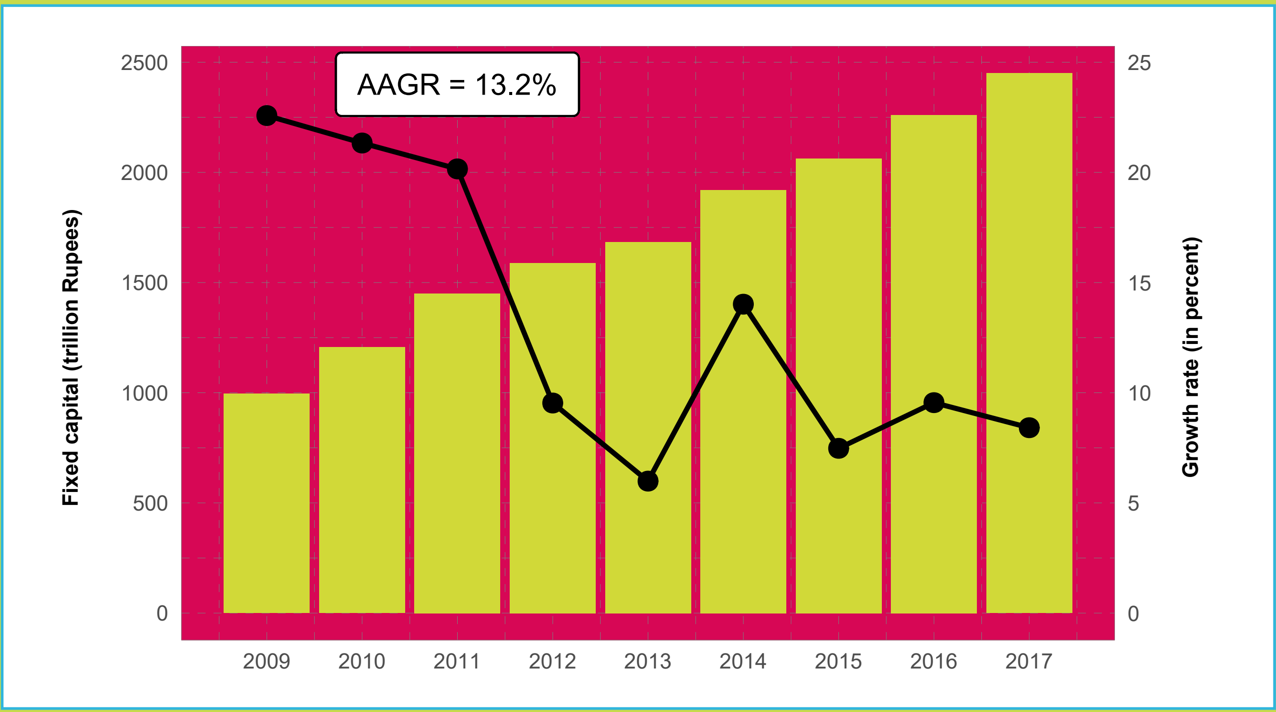

FIG. 7.3 | FIXED CAPITAL IN REGISTERED FOOD PROCESSING FACTORIES

SOURCE | AUTHOR ELABORATION BASED ON THE 2017-2018 ANNUAL SURVEY OF INDUSTRIES (MOSPI 2018).
NOTE | ALL VALUES ARE IN BILLION RUPEES. FIXED CAPITAL IS DEFINED AS THE DEPRECIATED VALUE OF FIXED ASSETS OWNED BY THE FACTORY.

number of unregistered factories in India was 2.3 times higher than registered ones, the ratio is even higher for the food processing sector at 2.6.

Fixed Capital

In terms of investment in fixed capital, registered factories in the food processing sector have seen an average annual growth rate of 9.1 percent between 2009 and 2018.

Capital Efficiency

An industry's capital efficiency is measured in terms of GVA per gross fixed capital formation in a particular year. It indicates how efficient an economy employs its capital to produce a product or a service. As evident from data in the Annual Survey of Industries for the period 2014-2017, there are considerable differences in performance within the organized food processing sector. Table 7.3 shows how some industries face vast year-to-year fluctuations. Others have seen their capital efficiency increasing progressively over the period. These improvements may be explained by modernizations of machinery and technological know-how, as well as increased utilization of existing capital. It is expected that many such measures will further impact the food processing industries in the near future.

Foreign Direct Investments

The strong potential of India's food processing industries has not gone unnoticed. In less than a decade, foreign direct investments (FDI) have increased with a factor of

TABLE 7.3 | CAPITAL EFFICIENCY

		2014	2015	2016	2017	Average annual growth (percent)
		Ratio				
1010	Processing and preserving of meat	3.68	3.35	3.78	4.87	8.2
1020	Processing and preserving of fish, crustaceans, and mollusks and products thereof	4.72	3.49	3.58	3.60	-5.7
1030	Processing and preserving of fruit and vegetables	3.04	5.62	2.99	3.30	12.1
1040	Manufacture of vegetable and animal oils and fats	5.50	5.59	3.78	5.61	4.4
1050	Manufacture of dairy products	2.01	2.26	2.83	2.21	3.9
1061	Manufacture of grain mill products	3.76	4.28	5.77	5.54	11.2
1062	Manufacture of starches and starch products	1.97	1.11	2.46	2.59	20.8
1071	Manufacture of bakery products	4.06	6.41	4.75	6.73	18.4
1072	Manufacture of sugar	1.63	2.19	3.77	2.93	21.1
1073	Manufacture of cocoa, chocolate, and sugar confectionery	1.55	2.52	15.47	5.72	128.4
1074	Manufacture of macaroni, noodles, couscous, and similar farinaceous products	4.50	14.30	7.30	1.35	21.8
1075	Manufacture of prepared meals and dishes	24.71	4.80	28.29	6.64	83.1
1079	Manufacture of other food products n.e.c.	4.95	4.47	5.07	5.73	4.2
1080	Manufacture of prepared animal feeds	5.52	3.00	3.15	6.55	16.8
1101	Distilling, rectifying, and blending of spirits; ethyl alcohol production from fermented materials	4.12	5.67	3.75	5.22	10.7
1102	Manufacture of wines	4.42	4.55	7.53	10.85	28.1
1103	Manufacture of malt liquors and malt	3.65	3.14	3.49	3.31	-2.0
1104	Manufacture of soft drinks; production of mineral waters and other bottled waters	2.18	3.19	3.42	3.64	15.0

SOURCE | AUTHOR ELABORATION BASED ON THE 2017-2018 ANNUAL SURVEY OF INDUSTRIES (MOSPI 2018).

6.5 from 189 million dollars in 2010 to 905 million dollars in 2019 (see Table 7.4). Although FDI policies have been rapidly liberalized since 1991, there has been some resistance to opening up the food processing sector to foreign investments. However, in 2016, landmark legalization effectively permitted 100 percent FDI for retail trading (including e-commerce) of food products that have been manufactured and/or processed within India. (MOFPI 2017) Other policy changes and amendments to regulations have attracted global retail giants, who today have a strong presence in the Indian market.

TABLE 7.4 | INFLOW OF FOREIGN DIRECT INVESTMENTS

	FDI (Million Rupees)	FDI (Million dollars)
2010	8.580	189
2011	8.262	170
2012	21.937	402
2013	251.068	3,983
2014	31.647	517
2015	33.174	506
2016	48.659	727
2017	58.356	905
2018	44.304	628
2019	64.147	905

SOURCE | DEPARTMENT OF INDUSTRIAL POLICY AND PROMOTION (2010-2020).

Exports

From 2011 to 2018, India saw impressive growth in the exports of processed food products at 12.2 percent, surpassing that of the manufacturing sector (7.9 percent); see Table 7.5. Despite the sector's rapid growth, its share in manufacturing exports has remained mostly stable in this period. India's position in the international trade of processed food products is also unsubstantial, accounting for just 2.3 percent as of 2018 (International Trade Centre 2020).

TABLE 7.5 | EXPORTS IN THE FOOD PROCESSING SECTOR

	Food processing sector (million dollars)	Total merchandise exports (million dollars)	Share (percent)
2011	31,456	305,964	10.3
2012	35,898	300,401	12.0
2013	38,051	314,405	12.1
2014	36,172	310,338	11.7
2015	29,672	262,291	11.3
2016	30,872	275,852	11.2
2017	35,318	303,526	11.6
2018	35,303	330,069	10.7
Average annual growth rate (percent)	2.2	1.4	

SOURCE | AUTHOR ELABORATION BASED ON DGCI&S (2020).

TABLE 7.6 | EXPORTS IN 2018 ACCORDING TO PRODUCT GROUPS (HS SECTIONS) OF PROCESSED FOOD AND BEVERAGES

HS Section	Product	Million dollars	Share of total (percent)
02	Meat and edible meat offal	3,723	10.5
03	Fish and crustaceans, mollusks, and other aquatic invertebrates	6,257	17.7
04	Dairy produce; birds' eggs; natural honey; edible prod. of animal origin, not elsewhere spec. or included	539	1.5
07	Edible vegetables and certain roots and tubers	1,300	3.7
08	Edible fruit and nuts; peel of citrus fruit or melons	1,616	4.6
09	Coffee, tea, mate, and spices	3,200	9.1
10	Cereals	8,162	23.1
11	Products of the milling industry; malt; starches; inulin; wheat gluten	321	0.9
12	Oilseeds and olea. fruits; misc. grains, seeds, and fruit; industrial or medicinal plants; straw and fodder	1,641	4.6
13	Lac; gums, resins, and other vegetable saps and extracts	1,057	3.0
15	Animal or vegetable fats and oils and their cleavage products; pre. edible fats; animal or vegetable waxes	1,097	3.1
16	Preparations of meat, fish or crustaceans, mollusks, or other aquatic invertebrates	433	1.2
17	Sugars and sugar confectionery	1,629	4.6
18	Cocoa and cocoa preparations	193	0.5
19	Preparations of cereals, flour, starch, or milk; pastrycooks products	535	1.5
20	Preparations of vegetables, fruit, nuts, or other parts of plants	589	1.7
21	Miscellaneous edible preparations	769	2.2
22	Beverages, spirits, and vinegar	325	0.9
23	Residues and waste from the food industries; prepared animal fodder	1,918	5.4
	Total	**35,304**	

SOURCE | AUTHOR ELABORATION BASED ON DGCI&S (2020).

TABLE 7.7 | DISTRIBUTION OF REGISTERED ENTERPRISES ACROSS STATES AND UNION TERRITORIES IN 2016

	Number of registered enterprises	Percentage share of total
Andhra Pradesh	5,861	14.7
Arunachal Pradesh	30	0.1
Andaman and Nicobar Island	5	0.0
Assam	1,409	3.5
Bihar	881	2.2
Chandigarh	19	0.0
Chhattisgarh	1,309	3.3
Dadar &Nagar Haveli	8	0.0
Daman & Diu	32	0.1
Delhi	166	0.4
Goa	98	0.2
Gujarat	2,240	5.6
Haryana	918	2.3
Himachal Pradesh	193	0.5
Jammu and Kashmir	176	0.4
Jharkhand	228	0.6
Karnataka	2,251	5.7
Kerala	1,629	4.1
Lakshadweep	0	0.0
Madhya Pradesh	876	2.2
Maharashtra	2,808	7.1
Manipur	28	0.1
Meghalaya	26	0.1
Mizoram	0	0.0
Nagaland	21	0.1
Odisha	1,127	2.8
Puducherry	60	0.2
Punjab	2,906	7.3
Rajasthan	883	2.2
Sikkim	19	0.0
Tamil Nadu	5,077	12.8
Telangana	3,969	10.0
Tripura	95	0.2
Uttar Pradesh	2,068	5.2
Uttarakhand	372	0.9
West Bengal	1,960	4.9
Total	**39,748**	

SOURCE | AUTHOR ELABORATION BASED ON THE 2016-2017 ANNUAL SURVEY OF INDUSTRIES (MOSPI 2017).

The majority of India's exports are low-value, raw, or semi-processed. Typically, the products are bulk marketed and shipped off for processing in other countries. The share of high-value and agri-products in agricultural exports comes down to just 15 percent - a considerable difference to China with its 49 percent (CII 2019).

Table 7.6 breaks down the sector's exports figures into 19 product groups, comprising 463 processed food and beverage products. Identified by the World Bank, the products are classified according to the six digit-level of the Harmonized System (HS) of the International Trade Classification, which is equivalent to Chapters 10 & 11 of India's 2008 National Industrial Classification (NIC) used by the Ministry of Food Processing Industries. In 2018, the top three exported commodities were cereals constituting 23.1 percent of total exports of processed food and beverages, followed by fish and invertebrates (17.7 percent) and meat products (10.5 percent).

7.3 STRUCTURE OF INDIA'S FOOD PROCESSING SECTOR

Geographical distribution of registered factories
According to the 2016-2017 Annual Survey of Industries, India had 39,748 registered enterprises in the food processing sector in 2016. As shown in Table 7.7, more than half of all activities in the registered sector are centered in the five states Andhra Pradesh (14.7 percent), Tamil Nadu (12.8 percent) and Telangana (10.0 percent), Maharashtra (7.1 percent) and Punjab (7.3 percent).

The organized and unorganized sectors of the Indian economy
As previously mentioned, the food processing industries continue to be more prevalent in the unorganized and

home-based sector of India's economy. Table 7.8 shows the distribution of food processing enterprises and their value-added in the country's unorganized and organized sectors. Size-wise, in terms of the number of enterprises, the latter is but a fraction of the former, although the ratio declined between 2000 and 2015. This represents an enormous employment potential in the unorganized sector. On the other hand, while value added in the organized sector also surpasses that of the unorganized sector and has grown progressively in the period 2000-2015, the value addition per enterprise is but a fraction. With transformative

TABLE 7.8 | ENTERPRISES AND VALUE-ADDED IN THE ORGANIZED AND UNORGANIZED SECTORS

	Enterprises			Value added (billions)		
	Organized sector (1)	Unorganized sector (2)	Ratio (2/1)	Organized sector (1)	Unorganized sector (2)	Ratio (2/1)
2000	21,649	3,011,300	139	165	47	0.3
2005	23,734	2,602,807	110	235	154	0.7
2010	30,253	2,241,195	74	552	221	0.4
2015	33,359	2,459,929	74	1,039	2,648	2.5
2016	34,709	NA	-	1,214	NA	-
2017	34,560	NA	-	1,307	NA	-

SOURCE | AUTHOR ELABORATION BASED ON NSSO (2018) AND THE 2016-2017 ANNUAL SURVEY OF INDUSTRIES (MOSPI 2017).

industrial policies in place, the food processing sector could become a driver of economic growth.

The significant earning potential of food processing industries in the organized sector is illustrated in Figure 7.4. From 2000 to 2017, the registered income in enterprises increased by a factor of 8.5 or 15.1 percent in terms of average annual growth.

Table 7.9 details the performance of the organized food processing sector. Similar to the operating enterprises' income, the total value of production has also increased significantly. However, the table also reveals that the ratio

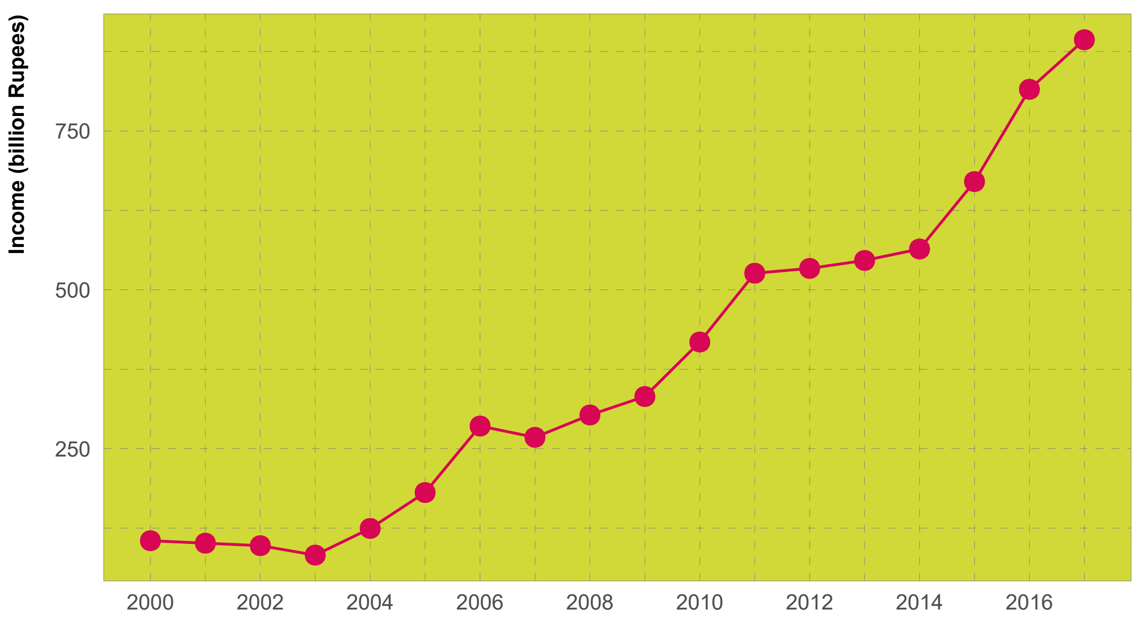

FIG. 7.4 | INCOME OF ENTERPRISES IN FOOD PROCESSING INDUSTRIES IN THE ORGANIZED SECTOR

SOURCE | AUTHOR ELABORATION BASED ON THE 2016-2017 ANNUAL SURVEY OF INDUSTRIES (MOSPI 2017).

TABLE 7.9 | THE ORGANIZED FOOD PROCESSING SECTOR

	Total enterprises in operation	Employment (thousands)	Total value of production	Materials consumed	Income
			(trillion Rupees)		
2000	21,649	1,333	1,355	1,075	105
2002	22,395	1,311	1,304	1,038	97
2003	22,490	1,311	1,570	1,292	83
2004	23,471	1,343	1,754	1,445	125
2005	23,734	1,392	2,013	1,626	181
2006	23,951	1,488	2,460	1,950	286
2007	24,616	1,502	2,967	2,432	268
2008	25,788	1,564	3,547	2,926	303
2009	25,915	1,611	3,855	3,17.5	332
2010	30,253	1,662	5,076	4,228	418
2011	30,331	1,788	6,387	5,256	526
2012	30,889	1,699	6,441	5,276	534
2013	32,068	1,741	7,168	5,930	546
2014	33,275	1,774	7,989	6,654	564
2015	33,358	1,776	8,196	6,749	670
2016	34,712	1,854	9,426	7,755	816
2017	34,564	1,933	10,313	8,411	894

SOURCE | AUTHOR ELABORATION BASED ON THE 2016-2017 ANNUAL SURVEY OF INDUSTRIES (MOSPI 2017).

between the production value and materials consumed has slightly declined over time, suggesting that efficiency measures could benefit the sector.

Urban versus rural distribution

Many enterprises in the unorganized sector are home-based and are therefore often located in the countryside. In 2015, 66 percent of all unorganized factories were rural, as was the majority of workers (59 percent) - a decline from 78 percent and 76 percent, respectively, since 2000 (see Table 7.10). It is most likely the share of home-based enterprises that are responsible for such high rural numbers. In the same period, urban enterprises grew in both numbers and employees, although not enough to prevent a decline in the total enterprise count (18 percent) and workforce (25 percent) of the unorganized food processing sector. It is unclear whether these changes reflect a rural to urban relocation of factories, a reclassification from unorganized to organized, or simple closure of businesses. If any of the two latter is the case, they may partly explain the increased value addition of the food processing sector witnessed in recent years.

Gender distribution

Table 7.11 suggests that men largely dominate the food processing sector. As of 2017, only one industry has a majority of women, specifically Processing and preservation of fish. The industries with the highest male-to-female ratio were Manufacture of sugar, Manufacture of vegetable and animal oils and fats, and Manufacture of starches and starch products. Between 2010 and 2017, the gender distribution saw little positive change.

An encouraging story highlights the prospects of female empowerment in the food processing sector. In 1959, seven illiterate, poor women borrowed 80 Rupees to start a 'papad' (a type of fine bread made of gram) business. From a meager turnover of 6,196 Rupees in the first year, it had an annual turnover of 8 billion Rupees and exports of 2.9 billion in 2018. Over time, the company Shri Mahila Griha Udyog Lijjat Papad, popularly known as Lijjat, established itself as an Indian women's worker cooperative manufacturing various fast-moving consumer goods. The organization's main objective is the empowerment of women by providing them employment opportunities. As of 2018, it employed around 43,000 women across the country, primarily in their own homes. It is acknowledged as one of the most remarkable entrepreneurial initiatives by women in India and a beacon for female empowerment. (Nitin 2020)

TABLE 7.10 | FOOD PROCESSING INDUSTRY IN THE UNORGANIZED SECTOR

	Number of enterprises			Number of workers		
	Rural	Urban	Total	Rural	Urban	Total
2000	2,360,700	650,600	3,011,300	5,171,100	1,670,900	6,842,000
2005	2,075,265	527,542	2,602,807	4,963,717	1,382,051	6,345,768
2010	1,545,962	695,233	2,241,195	3,090,362	1,702,200	4,792,562
2015	1,622,071	837,857	2,459,928	3,034,231	2,076,865	5,111,097

SOURCE | AUTHOR ELABORATION BASED ON NSSO (2018).

TABLE 7.11 | GENDER DISTRIBUTION IN SELECTED FOOD PROCESSING INDUSTRIES IN THE ORGANIZED SECTOR

	2010		2014		2017	
	Male	Female	Male	Female	Male	Female
Processing and preservation of meat	56	5	68	9	56	8
Processing and preservation of fish	16	23	27	49	39	64
Processing and preservation of fruits and vegetables	18	8	15	11	15	9
Manufacture of vegetable and animal oils and fats	19	1	19	2	20	2
Manufacture of dairy product	34	3	35	3	33	4
Manufacture of grain mill products	8	1	8	1	9	1
Manufacture of starches and starch products	16	4	16	3	23	3
Manufacture of bakery products	31	4	34	5	40	8
Manufacture of sugar	180	3	177	1	189	1
Manufacture of cocoa, chocolate, and sugar confectionery	30	4	36	5	34	6
Manufacture of macaroni, noodles, couscous, and similar farinaceous products	32	5	29	8	28	13
Manufacture of prepared meals and dishes	45	17	47	8	49	9
Manufacture of other food products	29	42	26	36	24	21
Manufacture of prepared animal feed	24	2	30	4	35	5
Distilling, rectifying, and blending of spirits, ethyl alcohol production from fermented materials	48	6	48	8	54	11
Manufacture of wines	31	7	23	12	19	6
Manufacture of malt liquors and malt	78	12	75	9	61	10
Manufacture of soft drinks, production of mineral waters and other bottled waters	17	1	15	2	14	2

NOTES I NUMBERS REFLECT THE AVERAGE NUMBER OF MALE AND FEMALE WORKERS, RESPECTIVELY, FOR ALL ENTERPRISES PER FOOD PROCESSING INDUSTRY. THE FIGURES ONLY COVER 'DIRECTLY EMPLOYED' WORKERS, HEREBY EXCLUDING WORKERS EMPLOYED THROUGH CONTRACTORS.
SOURCE | AUTHOR ELABORATION BASED ON THE 2016-2017 ANNUAL SURVEY OF INDUSTRIES (MOSPI 2017).

TABLE 7.12 | INDIA'S 2017 RANK IN WORLD PRODUCTION OF AGRICULTURAL AND FOOD PRODUCTS (TONES)

Product	Rank	India	World	Percent share
Castor oil seed	1	1,568,000	1,846,409	84.9
Ghee, of buffalo milk	1	2,863,000	3,605,209	79.4
Pigeon peas	1	4,870,000	6,807,551	71.5
Milk, whole fresh buffalo	1	86,261,680	123,031,341	70.1
Okra	1	6,003,000	9,641,284	62.3
Chickpeas	1	9,075,000	14,792,454	61.4
Ghee, butteroil of cow milk	1	137,550	240,558	57.2
Anise, badian, fennel, coriander	1	1,529,000	2,153,312	51.6
Areca nuts	1	723,000	1,439,280	50.2
Papayas	1	5,940,000	13,169,443	45.1
Chilies and peppers, dry	1	2,096,000	4,939,834	42.4
Meat, buffalo	1	1,635,153	4,201,780	38.9
Millet	1	11,560,000	30,456,356	38.0
Mangoes, mangosteens, guavas	1	19,506,000	55,590,973	35.1
Nutmeg, mace and cardamoms	1	43,000	130,046	33.1
Milk, whole fresh goat	1	6,165,500	18,894,731	32.6
Ginger	1	1,070,000	3,621,248	29.6
Bananas	1	30,477,000	125,341,716	24.3
Dry beans	1	6,390,000	32,739,471	19.5
Pulses*	1	29,260,978	161,023,088	18.2
Cauliflowers and broccoli	2	8,557,000	36,434,147	23.5
Cashew nuts, with shell	2	745,000	3,971,258	18.8
Dry onions	2	22,427,000	122,207,474	18.4
Rice, paddy	2	168,500,000	984,087,842	17.1
Green peas	2	5,345,000	33,290,590	16.1
Sugar cane	2	306,069,000	1,946,321,421	15.7
Lentils	2	1,220,000	7,762,648	15.7
Tea	2	1,325,050	8,574,503	15.5
Eggplants (aubergines)	2	12,510,000	85,217,885	14.7
Pumpkins, squash and gourds	2	5,142,812	35,500,978	14.5
Groundnuts, with shell	2	9,179,000	64,247,617	14.3
Lemons and limes	2	2,364,000	19,571,407	12.1
Milk, whole fresh cow	2	83,633,570	706,393,439	11.8
Wheat	2	98,510,000	906,059,207	10.9
Potatoes	2	48,605,000	487,396,255	10.0
Tomatoes	2	20,708,000	241,928,300	8.6
Cabbages and other brassicas	2	8,807,000	105,494,481	8.4
Goat meat	2	499,673	8,146,114	6.1
Garlic	2	1,693,000	50,381,020	3.4
Coconuts	3	11,469,837	61,100,124	18.8
Sesame seed	3	751,000	5,899,028	12.7
Oranges	3	7,647,000	81,998,902	9.3
Rapeseed	3	7,917,000	89,512,351	8.8
Lettuce and chicory	3	1,090,770	42,027,374	2.6
Green beans	3	675,188	43,515,135	1.6

NOTE I THE PRODUCTS ARE LISTED ACCORDING TO THE QUANTITY WITHIN EACH RANKING SCORE.
SOURCE | AUTHOR ELABORATION BASED ON FAOSTAT (2019).

7.4 THE POTENTIAL OF PROCESSING INDIA'S VAST AGRICULTURAL OUTPUT

India persists as a global agricultural powerhouse, and production output continues to increase on an annual basis. The diversity of the country's topography, soil quality, and climatic conditions makes for advantageous conditions for producing many agricultural products, both crops and non-crops (Ministry of Agriculture 2015). Table 7.12 shows how the country performs in producing selected agricultural and processed food products compared to the rest of the world. In 2017, India was the world's largest producer of castor seed oil, buffalo milk and ghee produced thereof, pigeon- and chickpeas, and okra. It was the second-largest producer of cauliflower and broccoli, cashew nut, dry onions, rice paddy, green peas, sugarcane, and lentils. Its production of coconuts and sesame seeds also ranks the third highest in the world.

An abundant supply of raw materials and a rising demand for food products supported by public policy incentives have positively impacted the food processing sector. However, only a narrow fraction of India's vast agricultural output is processed. According to a study by the Institute of Economic Growth, the extent to which a representative bundle of agro-products in India was processed (processed quantity of a product as a share of total agricultural production) was just 6.8 percent in 2010. Table 7.13 shows the extent of processing in 2005 and 2010 in both the organized and unorganized food processing sectors.

According to the Food and Agriculture Organization (2011), more than 40 percent of food losses in developing countries occur at post-harvest and processing levels. There are no country-specific estimates for India, but during a 2013 parliamentary hearing, a former Minister of Agriculture

TABLE 7.13 | AVERAGE EXTENT OF FOOD PROCESSING OF AGRO-PRODUCTS

	Weighted by production quantity		Weighted by production value	
	2005	2010	2005	2010
Total food processing sector	5.3	5.4	6.4	6.8
Organized sector	5.0	5.2	6.0	6.4
Unorganized sector	0.3	0.3	0.4	0.3

NOTES I THE CALCULATION IS BASED ON THE OUTPUT FROM 12 SUB-GROUPS, SPECIFICALLY MILLED COARSE CEREALS, MILLED RICE, MILLED WHEAT, MILLED PULSES, FRUITS & VEGETABLES, SPICES, GROUNDNUT, SOYBEANS, MEAT, FISH, MILK, AND EGGS.
THE QUANTITY-WEIGHTED AVERAGE IS THE TOTAL QUANTITY PROCESSED DIVIDED BY THE TOTAL NET QUANTITY PRODUCED IN AGRICULTURE. QUANTITY IS MEASURED IN WEIGHT. THE WEIGHTED AVERAGE IS CALCULATED AS THE AVERAGE OF THE EXTENTS OF PROCESSING OF EACH SUB-GROUP WEIGHTED BY THE SUB-GROUP'S SHARE IN THE TOTAL OUTPUT PRODUCED. ALL VALUES ARE AT 2005 PRICES.
SOURCE | GHOSH (2014).

suggested that the country sees a similar share of its total value of annual product wasted (Hindustan Times 2013).

However, studies by the Central Institute of Post-Harvest Technology and Engineering (CIPHET) have provided a quantitative assessment of the harvest and post-harvest losses for 46 agricultural produces across 106 randomly selected districts in 2010 and 45 produces across 107 districts in 2015. The wastage considered in these studies was produce deemed “unfit for human consumption” and assessed according to different stages of processing, i.e., harvesting, collection, threshing, grading/sorting, winnowing/cleaning, drying, packaging, transportation, and storage depending upon the commodity. Harvest and post-harvest losses were estimated at 44.1 billion Rupees (at 2009 wholesale prices) in 2010 and 92.3 billion Rupees (at 2012 at 2014 wholesale prices) in 2015.

In both studies, most wastage occurred for fruits and vegetables. Table 7.14 summarizes the findings for major produces and Table 7.15 for selected horticultural and cereal crops. In the case of cereals, the majority of losses occur at the farm level during harvesting, collection, and threshing, while storage losses only are 0.8-1.2 percent. High losses also occur in farm-level operations for fruits. Adequate processing facilities would reduce much of this wastage, thus providing remunerative prices to producers and ensuring greater supply to consumers (MOFPI 2018).

7.5 INDUSTRIAL POLICIES TO PROMOTE THE FOOD PROCESSING SECTOR

The Indian government has long acknowledged the significant potential of the food processing sector. In addition to establishing a conducive policy ecosystem, an umbrella of initiatives and policies has been implemented with the overall objective to position India as the 'Food Basket.' The sector is a key pillar in one of India's largest national programs, "Make in India," which is designed to foster innovation, enhance skill development, and build best-in-class manufacturing infrastructure in the country.[1]

To address one of the biggest impediments to the sector's growth and export potential, the Ministry of Food Processing Industries has rolled out the Mega Food Park Scheme with "the aim of creating modern infrastructure facilities for food processing along the value chain from farm to market with strong forward and backward linkages through a cluster-based approach" (MOFPI 2019, p. 9). The scheme offers projects within a Park financial assistance of up to 500

1 See more at www.makeinindia.com.

TABLE 7.14 | ESTIMATED CUMULATIVE WASTAGE FOR MAJOR PRODUCES (PERCENT)

	2010	2015
Cereals	3.9 – 6.0	4.7 – 6.0
Pulses	4.3 – 6.1	6.4 – 8.4
Oilseeds	2.8 – 10.1	3.1 – 10.0
Fruits and Vegetables	5.8 – 18.0	4.6 – 15.9
Milk	0.8	0.9
Fisheries (inland)	6.9	5.2
Fisheries (marine)	2.9	10.5
Meat	2.3	2.7
Poultry	3.7	6.7

SOURCE | AUTHOR ELABORATION BASED ON NANDA ET AL. (2012) AND JHA ET AL. (2015).

TABLE 7.15 | ESTIMATED CUMULATIVE WASTAGE FOR SELECTED HORTICULTURAL AND CEREAL CROPS (PERCENT)

	2010	2015
Horticultural crops		
Guava	18.6	15.9
Mango	12.7	9.2
Apple	12.3	10.4
Grapes	8.3	8.6
Papaya	7.3	6.7
Banana	6.6	7.8
Cereal crops		
Wheat	5.9	4.9
Paddy	5.2	5.5
Bajra	4.8	5.2
Maize	4.1	4.7

SOURCE | AUTHOR ELABORATION BASED ON NANDA ET AL. (2012) AND JHA ET AL. (2015).

million Rupees. As of 2019, 42 Parks were in the pipeline, and 11 were already operational. Another scheme reaches enterprises outside of the Parks. The Scheme for Agro-Marine Processing and Development of Agro-Processing Clusters set to allocate 60 billion Rupees for investments in modern infrastructure to make supply chain management more efficient - from farm gate to retail outlet - in the period 2016-2020.

Food processing units and infrastructure supportive thereof are also attractive tax incentives (MOFPI 2019). It is noticeable that cold chain and post-harvest storage have been recognized as an infrastructure sub-sector. Major incentives include:

- **Exemption or reduction of the national Goods and Service Tax:** 36 percent of all food items are fully tax-exempt, and 37 percent are only subject to a 5 percent rate. As these food items cover the bulk of raw materials that undergo further value addition, therefore, it is expected that the production cost of processed food will decline. Tax exemption is also given to certain pre-cold storage services, specifically pre-conditioning, pre-cooling, ripening, waxing, retail packing, and labeling of fruits and vegetables.

- **Income tax:** For a period of five years, enterprises involved in the processing, preservation, and packaging of fruits and vegetables, meat, poultry, marine and dairy products, along with enterprises in the integrated business of handling, storage, and transportation of food grains are eligible for 100 percent deduction of the profits and gains derived from such activities. After that, the enterprises are subject to a 25 percent deduction in up to 10 years. Furthermore, 150 percent of expenditures incurred on investments in and operation of a cold chain facility or warehouse facility for storage of agricultural produce are allowed. (MOFPI 2019)

- **Financing:** The food processing sector has been granted Priority Sector Lending. As enterprises in this sector often do not receive timely or adequate lending, the Reserve Bank of India requires banks in the country to provide a specified portion of the bank lending. Moreover, the National Bank for Agriculture and Rural Development provides credit at affordable rates to boost the sector through a special 20 billion Rupees fund.

A National Food Processing Policy has been in the works for years but is yet to materialize. There is an increasing pressure to align central and state policies and procedures (Ambwani 2020). Moreover, particular emphasis has recently been given to the adoption of food safety and quality assurance mechanisms such as Total Quality Management (TQM), including ISO 9000, ISO 22000, Hazard Analysis and Critical Control Points (HACCP), Good Manufacturing Practices (GMP) and Good Hygienic Practices (GHP). Compliance with such standards would improve India's competitiveness. Two certification schemes, "IndiaGHP" and "IndiaHACCP," that are based on globally accepted Codex Standards, allow enterprises to demonstrate compliance to global standards as an alternative to costly and time-consuming application processes for mandated foreign certifications. (MOFPI 2019)

7.6 REFERENCES

Adukia, R.S., 2012. Food Processing Industry in India. Mumbai: Adukia & Associates.

Ambwani, M.V., 2020. Industry Urges Centre to Formulate National Food Processing Policy. The Hindu BusinessLine. Available at: <www.thehindubusinessline.com/economy/industry-urges-centre-to-formulate-national-food-processing-policy/article30550946.ece>.

CII (Confederation of Indian Industry), 2019. Indian Food Processing Sector: Trends & Opportunities. New Delhi.

Department of Industrial Policy and Promotion, 2010-2020. Quarterly Fact Sheet on Foreign Direct Investment. FDI Statistics. Available at: <https://dipp.gov.in/publications/fdi-statistics> (Accessed January 2020).

DGCI&S (Directorate General of Commercial Intelligence and Statistics), 2020. Balance of Trade. Database. Available at: <www.dgcisanalytics.in/dgcis/EXIM-Analytics>.

Diwakar, GD, Gupta, OP, and Singh, DV (1983). A study of estimation of losses in foodgrains caused by rats. Journal of Indian Society of Agricultural Statistics, 31(1), pp. 76-78.

FAOSTAT, 2019. Production: Crops and Processed Crops. Database. Available at:<http://faostat.fao.org/>.

FAO (Food and Agricultural Organisation), 1980. Assessment and Collection of Data on Post-harvest Food Grain Losses. Rome.

--------, 1986. The ICS Users’ manual: Interlinked Computer Storage and Processing System of Food and Agricultural Commodity Data. Rome.

-------, 1996. Definition and Classification of Commodities. Draft Report. Rome.

Ghosh, N., Bhandari, B.S. and Sharma, S., 2013. The Extent of Processing Agricultural Products: Towards Creating a Statistical Database in India. The Journal of Industrial Statistics, 2(1), pp. 24-36.

Ghosh, N., 2014. An Assessment of the Extent of Food Processing in Various Food Sub-sectors. New Delhi: Institute of Economic Growth.

IBEF (India Brand Equity Foundation), 2017. Food Processing. New Delhi.

International Trade Centre, 2019. Trade Map. Database. Available at: <www.trademap.org>. (Accessed January 2021).

Jha, S.N., Vishwakarma, R.K, Ahmad, T., Rai, A. and Dixit, A.K., 2015. Report on Assessment of Quantitative Harvest and Post-harvest Losses of Major Crops and Commodities in India. ICAR - All India Coordinated Research Project on Post-Harvest Technology. Ludhiana: CIPHET (Central Institute of Post-harvest Engineering and Technology).

Majumdar, K., 2013. Export Performance of Processed Food in India. Global Journal of Management and Business Studies, 3(3), pp. 261-270.

Ministry of Agriculture, 2015. Agricultural Situation in India. New Delhi.

MOFPI (Ministry of Food Processing Industries, 2012. Data Bank on Economic Parameters of the Food Processing Sector. New Delhi.

--------, 2016. 2015-2016 Annual Report. New Delhi.

--------, 2017. Food Processing: A Ready Reckoner for FDI in India. New Delhi.

--------, 2018. 2017-2018 Annual Report. New Delhi.

--------, 2019. Draft National Food Processing Policy 2019. New Delhi.

MOSPI (Ministry of Statistics and Programme Implementation), 2017. Annual Survey of Industries 2016-2017. Database. Available at: <http://mospi.nic.in/asi-summary-results>.

--------, 2018. Annual Survey of Industries 2017-2018. Database. Available at: <http://mospi.nic.in/asi-summary-results>.

--------, 2020a. National Account Statistics 2020. Database. Available at: <http://mospi.nic.in/publication/national-accounts-statistics-2020>.

--------, 2020b. Index of Industrial Production 2020. Database. Available at:<http://mospi.nic.in/iip>.

Nanda, S.K., Vishwakarma, R.K., Bathla, H.V.L., Rai, A., and Chandra, P. Harvest and Post-harvest Losses of Major Crops and Livestock Produce in India. ICAR - All India Coordinated Research Project on Post-Harvest Technology. Ludhiana: CIPHET (Central Institute of Post-harvest Engineering and Technology).

Nitin, S.N., 2020. Women Empowerment through Cooperative Society - A Study on Lijjat Papad. Conference Paper, Co-operatives in a Globalised Economy: Relevance, Issues & Challenges, Goa, January 2020.

NSSO (National Sample Survey Organisation), 2018. India: Unincorporated Non-Agricultural Enterprises (Excluding Construction), July 2016-June 2016, 73 Round. New Delhi.

UNIDO