SET SUPPORTING ECONOMIC TRANSFORMATION

SOURCES AND METHODS OF DATA FOR ECONOMIC TRANSFORMATION

SET Data Portal

Dirk Willem te Velde, Jane Kennan

Updated 15 September 2015

Acknowledgements

Thanks to Jakob Engel (Oxford University, PhD) for providing information on the EORA database and Gaaitzen de Vries (University of Groningen) for providing information on the Groningen Growth and Development Centre (GGDC) 10-Sector Database.

TABLE OF CONTENTS

Abbreviations **iii**

Summary **iv**

1. Introduction **1**

2. Data availability **3**

3. Sources and methods **8**

3.1 Economic structures *8*
3.1.1 10-Sector Database, version 2014 – GGDC 8
3.1.2 World Development Indicators (WDI) – World Bank 11
3.1.3 Global Employment Trends (GET) 2014 supporting datasets – International Labour Organization (ILO) 13
3.1.4 United Nations Statistics Division (UNSD) National Accounts Main Aggregates Database 14
3.1.5 ILO World Employment and Social Outlook (WESO) – Trends 2015, supporting datasets 15
3.1.6 Employment using Demographic and Health Surveys 19

3.2 Trade *19*
3.2.1 Basic trade data 19
3.2.2 Measures of trade diversification 22
3.2.3 Data on revealed comparative advantage 23
3.2.4 Trade in value added 23

3.3 Total factor productivity at firm level *29*

3.4 Wages by occupation *32*
3.4.1 Occupational Wages around the World (OWW) Database 32

4. Pros and cons of different data sources **33**

5. Conclusions **37**

Key references **38**

ABBREVIATIONS

AfDB	African Development Bank
CIF	Cost, insurance and freight
DFID	Department for International Development
DHS	Demographic and Health Surveys
DVA	Domestic value added/addition
EIU	Economist Intelligence Unit
ET	Economic transformation
FOB	Free on board
GDP	Gross domestic product
GET	Global Employment Trends
GGDC	Groningen Growth and Development Centre
GNI	Gross National Income
GVA	Gross Value Added
GVC	Global value chain
HS	Harmonised System
ILO	International Labour Organization
IMF	International Monetary Fund
IPOA	Istanbul Plan of Action
ISIC	International Standard Industrial Classification
ITC	International Trade Centre
KILM	Key Indicators of the Labour Market (ILO)
LDC	Least Developed Countries
LFS	Labour force surveys
LIC	Low-Income Country
MIC	Middle-Income Country
MRIO	Multi-Region Input-Output
OECD	Organisation for Economic Co-operation and Development
OWW	Occupational Wages around the World
PPP	Purchasing power parity
RSME	Root mean square error
SET	Supporting Economic Transformation
SITC	Standard International Trade Classification
SNAAMA	System of National Accounts – National Accounts Main Aggregates (UNSD)
TFP	Total factor productivity
TiVA	Trade in value added
UNCTAD	United Nations Conference on Trade and Development
UNIDO	United Nations Industrial Development Organization
UNSD	United Nations Statistics Division
WB	World Bank
WDI	World Development Indicators
WEF	World Economic Forum
WESO	World Employment and Social Outlook (ILO)
WIOD	World Input-Output Database
WTO	World Trade Organization

SUMMARY

This initiative for the Supporting Economic Transformation (SET) programme aims to fill a gap in the availability of data on economic transformation at country level. Whilst there are a range of important international databases (e.g. World Development Indicators (WDI), United Nations Conference on Trade and Development (UNCTAD), International Labour Organization (ILO), United Nations Industrial Development Organization (UNIDO), etc.) that can be used to explore economic transformation (ET), there is currently no space where the data on ET are brought together at the country level. This paper on sources and methods is accompanied by Excel and Word files for 40 variables (Table 1) in 28 countries (of direct interest to the Department for International Development (DFID), see Table 2). It is a preliminary document that aims to foster discussion of data on ET.

In order to understand and measure ET (as used and defined by SET) in countries of interest, the paper explores data on

- economic structures, value addition, employment, labour productivity by sector
- trade, trade diversification and domestic value addition (DVA) in exports in aggregate and by sector
- firm-level total factor productivity (TFP) levels and dispersion within sector
- wage levels by occupation.

For some variables we use data directly or without many calculations (e.g. top exports), for others we compute them, e.g. the domestic value addition in gross exports (which is currently at the forefront of research on trade), dispersion of firm-level total factor productivity within sector (which is currently related to cutting-edge thinking on productivity, such as by Hsieh and Klenow), or within and between sector changes of labour productivity (see McMillan and Rodrik for recent thinking on this).

The emphasis of the work so far has been at the country level. Often a policy-maker or researcher is also interested in comparisons of a variable across countries (e.g. in the same sector). This can be done on the basis of the collected data, with a range of caveats and consideration of issues such as purchasing power parities (PPPs) and bearing in mind the pros and cons of the various databases (Table 8).

What are potential next steps?

We need to better understand the demand and use of ET data. Moreover, the data need to be developed and discussed further. And other databases need to be considered. Once fully checked, the data files could form a useful online datasource (the data interface to be decided). The sources and methods file needs to be reviewed and expanded appropriately. Then, data gaps can be filled, e.g. with respect to certain countries (sector data for the poorest Asian countries; or more specific and granular sector indices of transformation at sector level). Further, the data can be used in analyses of the following type:

- comparing and explaining sector-level productivity across countries and sectors
- exploring the (statistical) impact of the national account rebasing in Ghana, Nigeria, Kenya, Tanzania
- comparing sector-level employment structures and wage data
- comparing and explaining average firm-level TFP (and its dispersion) in a sector across countries and sectors

- comparing and explaining the DVA in gross exports in a sector across countries
- scoring all countries according to some specified benchmark.

1. INTRODUCTION

Many databases exist that can be used to analyse economic transformation (ET) (provided by, e.g., the World Bank (WB), the International Monetary Fund (IMF), United Nations Conference on Trade and Development (UNCTAD) and others). They reveal gaps in the data, and the quality is variable. However, for practitioners and country-level policy-makers concerned with ET, there is a further gap: an easily accessible place where all data on ET using different databases are brought together at the country level. For example, what does a Nigerian policy-maker have at his/her disposal when it comes to analysing data on ET? In a range of accompanying files (in Word and Excel), we have collected data from various sources to describe ET at country level. This draft paper discusses the sources and methods behind the data on ET.

ET (in the SET programme) is defined broadly as moving labour and other resources from lower to higher productive activities. This includes moving between sectors to higher-value activities (for example, from agriculture to manufacturing) and within a sector (for example, from subsistence farming to high-value crops, or from low-productivity firms to high-productivity firms). It usually also involves diversification and increased value addition in export activities.

In order to understand and measure economic transformation in countries of interest, we would therefore like to identify and measure data on the following:

- economic structures, value addition, employment, labour productivity by sector
- trade, trade diversification and domestic value addition in exports
- firm-level TFP levels and dispersion within sector
- wage levels by occupation.

There is not one single source that is sufficient to fully describe these facets of ET. Instead, we focus on a range of data sources to obtain as complete a picture as possible. We use sources ranging from UN Comtrade to the World Bank (WB), ILO, International Monetary Fund (IMF), and Eora. We use a range of basic and slightly more sophisticated techniques to present the data. We also note that the data are of variable quality and availability, and have advantages and disadvantages.

We examine data for 28 countries of interest (the countries in which DFID is interested). We focus initially on each country individually, but depending on the data we can include comparisons across countries later. We provide only a few examples in this paper. At this stage we do not envisage the creation of a composite measure of ET. However, we note that there is a separate work-stream that creates indices of transformation; e.g., Basnett, Keane and te Velde (2014) create an Istanbul Plan of Action (IPOA) index of structural transformation for Least Developed Countries (LDCs).[1] The African Center for Economic Transformation (2014) also creates a composite index.

The structure of this paper is as follows. Section 2 presents the data items and the countries for which we provide data. Section 3 discusses the sources and methods behind

[1] The paper selects a number of indicators based on how we expect variables to contribute to structural transformation, consistent with the literature. In particular we expect an LDC to transform itself structurally when it (compared to a benchmark): becomes more productive in agriculture by achieving a higher cereal yield; increases the share of manufacturing in value addition; increases GDP per capita; increases gross capital transformation; increases the share of information and communications technology (ICT) in services exports; improves its product diversification; increases the number of export markets it trades with; improves health services by decreasing the infant mortality rate; has better telecommunications infrastructure; and has a more developed financial market.

the data. Section 4 provides pros and cons of different databases. Section 5 concludes, and includes suggestions for further analysis.

2. DATA AVAILABILITY

The data files include data for around 40 variables; see Table 1.

Table 1: Variables included in the data files.

Economic structure		
Table	GDP, employment and relative productivity levels	Sectoral structure of GDP and employment over time
Table	Labour productivity levels and changes	Change in labour productivity over time
Figure	Relative productivity and changes in employment shares by sector	Examines which sectors have grown in employment and productivity
Figure	Decomposition of labour productivity change	Examines whether labour productivity has occurred within or between sectors
Figure	Productivity gaps	Difference in maximum theoretical and actual productivity levels
Figure	Total employment by sex and sector	Allows for (gender) breakdown of employment
Figure	Sectoral employment by sex	Allows for (gender) breakdown of employment
Figure	Percentage of workers (age 25+) in agriculture	Examines share of workers in agriculture using DHS data
Trade		
Basic data		
Figure	Total value of trade	Presents simple trends in trade
Figure	Exports by broad HS Section	Examines export shares of certain products
Figure	Change in export share by HS Section	And the changes over time
Figure	Export visualisation	Shows main exports
Figure	Top export products	Exports by product
Figure	Top export markets	Exports by destination
Figure	Imports by broad HS Section	Examines export shares of certain products
Figure	Change in import share by HS Section	And the changes over time
Figure	Import visualisation	Shows main imports
Figure	Top import products	Imports by product
Figure	Top import sources	Imports by source
Diversification		
Figure	Export diversification index	Examines extent of diversification
Figure	Export quality index	Using export prices as a proxy of quality
Figure	Number of export items and markets	Shows diversification by product and destination
Figure	Number of import items and suppliers	Shows diversification by product and source
Revealed comparative advantage		
Table	Revealed comparative advantage by HS Section	Calculates Balassa RCA index
Trade in value added		
Figure	Compound annual growth rate of domestic value added, foreign value added and exports	Shows annual growth rates of domestic and foreign value added in exports
Figure	Domestic and foreign value added content of gross exports as share of gross exports	Shows domestic and foreign value addition as share of gross exports
Figure	Overall value of domestic and foreign value added	Shows value of domestic and foreign value addition
Figure	Compound annual growth rate of DVA embodied in gross exports by sector	Shows annual growth rates of domestic value added in exports by sector
Figure	Sectoral DVA embodied in exports as a share of sectoral gross exports	Shows sectoral domestic value addition as share of sectoral gross exports
Figure	Sectoral DVA as a share of total DVA	Shows the contribution of a sector DVA to total DVA
Figure	Compound annual growth rate of FVA embodied in gross exports by sector	Shows annual growth rates of foreign value addition in exports by sector
Figure	Sectoral FVA embodied in exports as a share of sectoral gross exports	Shows sectoral foreign value addition as share of sectoral gross exports
Figure	Sectoral FVA as a share of total FVA	Shows the contribution of a sector FVA to total FVA
Total factor productivity at firm level		
Figure	Dispersion in productivity across firms by sector	Shows how much productivity differs amongst firms in a sector
Figure	Distribution in productivity – comparing kernel and normal distributions	Shows the distribution of firm productivity in a sector
Wages by occupation		
Figure	Relative wages	Plots relative wages for selected occupations over time

We provide data for these variables for 28 countries, as detailed in Table 2. Not all countries have data for all items. Overall, trade data are available in greater quality and detail and at greater frequency than national accounts or wage data.

This is not the place to discuss data quality issues in general. For an overview of economic statistics, see PEAKS topic guide here. For a discussion of data quality especially in relation to gross domestic product (GDP) see Morten Jerven's (2012) book on statistics.

Table 2: Variables / country availability

	Afghanistan	Bangladesh	DR Congo	Ethiopia	Ghana	India	Kenya	Kyrgyz Rep.	Liberia	Malawi	Mozambique	Myanmar	Nepal	Nigeria	Pakistan	Rwanda	Sierra Leone	Somalia	South Africa	South Sudan	Sudan	Tajikistan	Tanzania	Uganda	W. Bank & Gaza	Yemen	Zambia	Zimbabwe
Economic structures																												
GDP, employment and relative productivity levels	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Or, for those countries for which the more detailed data above were not available during the first set of analyses (no longer updated now that more comprehensive data for all countries have become available)																												
Value added (% of GDP)	Y		Y									Y						Y		Y	Y					Y		
Labour productivity levels and changes	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Relative productivity and changes in employment shares by sector	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Decomposition of labour productivity change	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Productivity gaps	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Total employment by sex and broad sector	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Broad sectoral employment by sex	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Sectoral employment by sex	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Percentage of workers (age 25+) in agriculture	N	N	N	Y	Y	N	Y	N	N	Y	Y	N	N	Y	N	Y	N	N	N	N	N	N	Y	Y	N	N	Y	Y
Trade																												
Basic data																												
Total value of trade	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Exports by broad HS Section	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Change in export share by HS Section	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Export visualisation	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y
Top export products	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Top export markets	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Imports by broad HS Section	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Change in import share by HS Section	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Import visualisation	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y

	Afghanistan	Bangladesh	DR Congo	Ethiopia	Ghana	India	Kenya	Kyrgyz Rep.	Liberia	Malawi	Mozambique	Myanmar	Nepal	Nigeria	Pakistan	Rwanda	Sierra Leone	Somalia	South Africa	South Sudan	Sudan	Tajikistan	Tanzania	Uganda	W. Bank & Gaza	Yemen	Zambia	Zimbabwe
Top import products	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Top import sources	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Diversification																												
Export diversification index	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y
Export quality index	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y
Number of export items and markets	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Number of import items and suppliers	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Revealed comparative advantage																												
Revealed comparative advantage by HS Section (and at HS 6-digit product level in the country Excel files only)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Trade in value added																												
Compound annual growth rate of domestic value added, foreign value added and exports, 1996-2011 and 2006-11	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	N
Domestic and foreign value added content of gross exports as share of gross exports, 1996, 2000, 2006 and 2011	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	N
Overall value of domestic and foreign value added, 1996, 2000, 2006 and 2011 (in USD 1,000)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	N
Compound annual growth rate of DVA embodied in gross exports by sector, 1996-2011 and 2006-2011	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	N
Sectoral DVA embodied in exports as a share of sectoral gross exports, 2000 and 2011	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	N
Sectoral DVA as a share of total DVA, 2000 and 2011	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	N
Compound annual growth rate of FVA embodied in gross exports by sector, 1996-2011 and 2006-2011	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	N
Sectoral FVA embodied in exports as a share of sectoral gross exports, 2000 and 2011	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	N
Sectoral FVA as a share of total FVA, 2000 and 2011	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	N

	Afghanistan	Bangladesh	DR Congo	Ethiopia	Ghana	India	Kenya	Kyrgyz Rep.	Liberia	Malawi	Mozambique	Myanmar	Nepal	Nigeria	Pakistan	Rwanda	Sierra Leone	Somalia	South Africa	South Sudan	Sudan	Tajikistan	Tanzania	Uganda	W. Bank & Gaza	Yemen	Zambia	Zimbabwe
Trade in services																												
Exports of goods and services	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	N
Sectoral shares of services exports	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	N
Total factor productivity at firm level																												
Dispersion in productivity across firms by sector	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	Y	N	N	Y	Y	Y	Y	Y	Y	Y
Distribution in productivity – comparing kernel and normal distributions	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	Y	N	N	Y	Y	Y	Y	Y	Y	Y
Wages by occupation																												
Relative wages	N	N	N	N	Y	Y	N	Y	N	Y	N	Y	N	Y	Y	N	N	N	N	N	N	N	N	N	N	N	Y	N

3. SOURCES AND METHODS

This section discusses the sources and methods used to present the data for the variables and countries listed in Section 2:

- data on economic structures (Section 3.1)
- data on trade, trade diversification , value added in trade (Section 3.2)
- data on firm-level productivity (Section 3.3)
- data on wages by occupation (Section 3.4).

3.1 ECONOMIC STRUCTURES

We discuss the following data sources:

- Groningen Growth and Development Centre (GGDC), the 10-Sector Database, version 2014, which includes (updates of) the Africa Sector Database[2] – Section 3.1.1
- World Development Indicators (WDI), Section 3.1.2
- Global Employment Trends (GET) 2014 supporting datasets – International Labour Organization (ILO), Section 3.1.3
- United Nations Statistics Division (UNSD) National Accounts Main Aggregates Database, Section 3.1.4
- ILO World Employment and Social Outlook (WESO) – Trends 2015, supporting datasets, Section 3.1.5
- Demographic and Health Surveys (DHS), Section 3.1.6 (as an example of household-level data).

3.1.1 10-SECTOR DATABASE, VERSION 2014 – GGDC
http://www.rug.nl/research/ggdc/data/10-sector-database

Provider's description[3]

The GGDC 10-Sector Database provides a long-run internationally comparable dataset on sectoral productivity performance in Africa, Asia, and Latin America. Variables covered in the dataset are annual series of value added, output deflators, and persons employed for 10 broad sectors. It gives sector detail to the historical macro data in Maddison (2003) from 1950 onwards. It consists of series for 11 countries in Africa, 11 countries in Asia, 2 countries in the Middle East and North Africa, and 9 in Latin America. For comparison, it has also added data for the US and several European countries.

The GGDC Sector database is constructed on the basis of an in-depth study of available statistical sources on a country-by-country basis and is continuously maintained and expanded. We discuss the contents of the database, the selection procedure of the sources used, and the methods employed to ensure consistency. Compliance with consistency requirements is important to ensure the usefulness of the database in analyses of growth and productivity.

[2] This database, 'The Africa Sector Database – Groningen Growth and Development Centre (GGDC)' can be downloaded from http://www.rug.nl/research/ggdc/data/africa-sector-database. We have used this database for the initial data analyses for 8 countries, but we use the sector database for other countries and will use that in future. It also includes employment by sex data. This document describes only the sector database.

[3] This section is a contribution by Gaaitzen de Vries.

Database sources and methods

See http://www.rug.nl/research/ggdc/data/10sector/10sector_sm_oct2014.pdf.
See also: Timmer et al. (2014)

Contents of the dataset

Table 3 gives an overview of the contents of the GGDC Sector database. The dataset currently includes 11 Asian, 9 Latin American, and 13 African countries. It includes annual data on gross value added (GVA) at current and constant prices from 1950 onwards. In addition, annual data on persons employed is available, which allows the derivation of labour productivity (value added per worker) trends. The database covers the 10 main sectors of the economy as defined in the International Standard Industrial Classification, Revision 3.1 (ISIC rev. 3.1). These 10 sectors cover the total economy.

Table 3: Overview of the GGDC Sector Database

Economic activities distinguished (ISIC rev. 3.1 code):	1. Agriculture, hunting, forestry and fishing (AtB) 2. Mining and quarrying (I) 3. Manufacturing (D) 4. Electricity, gas and water supply (E) 5. Construction (F) 6. Wholesale and retail trade, hotels and restaurants (GtH) 7. Transport, storage, and communication (I) 8. Finance, insurance, real estate and business services (JtK) 9. Government services (LtN) 10. Community, social and personal services (OtP)
Variables included:	Persons engaged Gross value added at current national prices Gross value added at constant 2005 national prices
Countries included:	**Africa**: Botswana, Egypt, Ethiopia, Ghana, Kenya, Malawi, Morocco, Mauritius, Nigeria, Senegal, South Africa, Tanzania, Zambia **Asia**: China, Hong Kong (China), India, Indonesia, Japan, Korea (Rep. of), Malaysia, Philippines, Singapore, Taiwan, Thailand **Latin America:** Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Mexico, Peru, Venezuela **Europe and North America:** Germany, Denmark, Spain, France, United Kingdom, Italy, the Netherlands, Sweden, United States
Time period:	1950 - most recent

Notes: Starting date of time series varies across variables and countries depending on data availability. Typically, the series start in 1950 for Latin American countries, in the 1960s for African countries, and in the 1970s for most Asian countries.

Construction of variables

Gross value added in current and constant prices is taken from the National Accounts of the various countries. As these have all been compiled according to the UN System of National Accounts, international comparability is high, in principle. However, national statistical institutes frequently change their methodologies. Within the National Accounts, GDP series are periodically revised; this includes changes in the coverage of activities (for example after a full economic census has been carried out and “new” activities have been discovered), changes in the methods of calculation (for example the inclusion of software expenditures as investment rather than intermediate consumption), and changes in base year of the prices used for calculating volume growth rates. For sectoral GDP our general approach is to start with GDP levels for the most recent available benchmark year, expressed in that year’s prices, from the National Accounts provided by the National Statistical Institute or Central Bank. Historical national accounts series were subsequently linked to this benchmark year. This linking procedure ensures that growth rates of individual series are retained although absolute levels are adjusted according to the most recent information and methods.

Employment in our dataset is defined as 'all persons employed', thus including all paid employees but also self-employed and family workers. Labour input is normally not available from a country's national accounts as they are not part of the System of National Accounts. Two different primary sources of employment data exist, namely labour force surveys (LFS) with data collected at the household level, and business surveys based on firm-level questionnaires. Both have their advantages and disadvantages as a source for annual sectoral employment trends.

We often use an alternative source based on household questionnaires but with a much larger coverage than the samples of the LFS: the population census. This ensures full coverage of the working population and a much more reliable sectoral breakdown than from the LFS. However, typically population censuses are quinquennial or decennial and cannot be used to derive annual trends. Therefore we use the population census to indicate absolute levels of employment, and we use LFS and business surveys to indicate trends in between. This is the general strategy followed for most countries, but not for all.

Consistency

In constructing the database, we paid careful attention to three checks on consistency, namely intertemporal consistency, international consistency, and internal consistency.

Our time series of gross value added and employment are consistent over time (that is, intertemporal consistency). Through the linking procedure described above, major breaks in the series have been repaired.

International consistency of the cross-country sectoral data is ensured through the system of national accounts for value added, the employment concept of persons engaged and the use of a harmonised sectoral classification. We classify activities into ten sectors, using ISIC, Revision 3.1. The industrial classification used in the national primary data sources is based on this classification or is directly related to it.

Finally, for the derivation of meaningful productivity measures, the labour input and output measures should cover the same activities (for internal consistency). As we use all persons employed – rather than paid employees – as our employment concept, and base our employment numbers on large-scale surveys, overlap in coverage of the employment statistics and value added from the National Accounts is maximised. However, a notable exception is the own-account production of housing services by owner-occupiers. For this an imputation of rent is made and added to GDP in many countries, according to the System of National Accounts. This imputed production does not have an employment equivalent and preferably should not be included in output for the purposes of labour productivity comparisons. Therefore, the GGDC Sector database presents separate series for imputed rents. In our decomposition analysis we exclude imputed rents.

Limitations of the GGDC Sector database

A note of caution on the data is warranted. Recently scholars have pointed out anew that the statistical foundations underlying GDP and employment estimates in many developing countries, notably but not exclusively Africa and China, are subject to substantial measurement error. The low quality of statistics is related to a weak capacity to collect, manage and disseminate data; inadequate funding of statistical offices; diffuse responsibilities on who is collecting what; and fragmentation in surveys and gathering exercises. Many African countries do not have a well-established statistical system, not even reporting national accounts data on a consistent basis. Therefore it is useful to explore alternative sources of information on national income using demographic and health survey data. Likewise, GDP and employment estimates in large developing countries such as India and China might have substantial measurement errors. However,

most countries included in the Sector database do have a considerable history of collecting national accounts data and in conducting labour and household surveys.

Application in the accompanying analyses

Data from the 10-Sector Database form the basis of the sectoral value added/employment share data and labour productivity analyses for **Nigeria and India.**

The analogous analyses for **Ethiopia, Ghana, Kenya, Malawi, Nigeria,**[4] **South Africa, Tanzania and Zambia** were undertaken using data from the latest available version of GGDC's **Africa Sector Database** – which is identical to the 10-Sector Database except that it covers only the 13 African countries (and it contains information on sectoral employment shares broken down by sex, which is not available in the 10-Sector Database). **The current analyses for these countries will in future be based on using the 10-Sector Database.**

Definitions and methodologies relating to the accompanying analyses

(labour) Productivity level	Calculated by dividing sectoral value added in constant (2005) prices (in local currency) by number of persons engaged.
Relative productivity level	Sectoral productivity level expressed as a ratio of total productivity level
Other industry	Includes construction and public services.
Distribution services	Includes transport services and distributive trade as well as hotels and restaurants.
Finance and business services	Excludes real estate activities.
Other services	Includes other community, personal and household services.
Decomposition of labour productivity change	'Within sector' = individual sector total of annualised growth rate in labour productivity over period x sector share in total employment in earlier year. 'Structural change' = individual sector total of annualised growth rate in labour productivity over period x change in sector share in total employment over same period.

3.1.2 WORLD DEVELOPMENT INDICATORS (WDI) – WORLD BANK

http://databank.worldbank.org/data/views/variableSelection/selectvariables.aspx?source=world-development-indicators

Provider's description

World Development Indicators (WDI) is the primary World Bank collection of development indicators, compiled from officially recognised international sources. It presents the most current and accurate global development data available, and includes national, regional and global estimates.

DFID focus country coverage

All (although the number of indicators/years for which data are available varies widely).

[4] Because of revised data (because of rebasing) in the 10-Sector Database, the analyses for Nigeria have been undertaken twice, using data from each database. Tables/figures based on Africa Sector Database data are labelled (a) in the country file, and those based on 10-Sector Database data are labelled (b).

Indicators used in the accompanying analysis

Agriculture, value added (% of GDP) Industry, value added (% of GDP) Manufacturing, value added (% of GDP) Services, etc., value added (% of GDP) Agriculture, value added (constant 2005 US$) Industry, value added (constant 2005 US$) Services, etc., value added (constant 2005 US$)	Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs. It is calculated without making deductions for depreciation of fabricated assets or depletion and degradation of natural resources. The origin of value added is determined by ISIC, revision 3. Data are in constant 2005 US$. Agriculture corresponds to ISIC divisions 1-5 and includes forestry, hunting and fishing, as well as cultivation of crops and livestock production. Industry corresponds to ISIC divisions 10-45 and includes manufacturing (ISIC divisions 15-37). It comprises value added in mining, manufacturing (also reported as a separate subgroup), construction, electricity, water, and gas. Manufacturing refers to industries belonging to ISIC divisions 15-37. Services correspond to ISIC divisions 50-99 and they include value added in wholesale and retail trade (including hotels and restaurants); transport; and government, financial, professional, and personal services such as education, health care, and real estate services. Also included are imputed bank service charges, import duties, and any statistical discrepancies noted by national compilers as well as discrepancies arising from rescaling. GDP represents the sum of value added by all its producers. Value added is the value of the gross output of producers less the value of intermediate goods and services consumed in production, before accounting for consumption of fixed capital in production. The United Nations System of National Accounts calls for value added to be valued at either basic prices (excluding net taxes on products) or producer prices (including net taxes on products paid by producers but excluding sales or value added taxes). Both valuations exclude transport charges that are invoiced separately by producers. Total GDP is measured at purchaser prices. Value added by industry is normally measured at basic prices.
Employment in agriculture (% of total employment) Employment in industry (% of total employment) Employment in services (% of total employment)	Employees are people who work for a public or private employer and receive remuneration in wages, salary, commission, tips, piece rates, or pay in kind. Agriculture corresponds to division 1 (ISIC revision 2) or tabulation categories A and B (ISIC revision 3) and includes hunting, forestry and fishing. Industry corresponds to divisions 2-5 (ISIC revision 2) or tabulation categories C-F (ISIC revision 3) and includes mining and quarrying (including oil production), manufacturing, construction, and public utilities (electricity, gas, and water). Services correspond to divisions 6-9 (ISIC revision 2) or tabulation categories G-P (ISIC revision 3) and include wholesale and retail trade and restaurants and hotels; transport, storage, and communications; financing, insurance, real estate, and business services; and community, social, and personal services.

Application in the accompanying analyses

The above WDI data, with ILO Global Employment Trends data on total employment numbers (see below), form the basis of the sectoral value added/employment share data and labour productivity analyses for **Afghanistan, Bangladesh, DR Congo, Kyrgyz Republic, Liberia, Mozambique, Myanmar, Nepal, Pakistan, Rwanda, Sierra Leone, Sudan, Tajikistan, Uganda, West Bank & Gaza, Yemen and Zimbabwe** (i.e. countries not covered by either the Africa Sector Database or the 10-Sector Database).

There are insufficient data available for **Somalia** and **South Sudan**.

Definitions and methodologies relating to the accompanying analyses

Sectoral employment figures (in absolute terms) were calculated by applying WDI sectoral employment percentages[5] to ILO employment totals. Sectoral labour productivity figures were then derived by dividing the WDI constant 2005 US$ value added figures by the results, and relative productivity by expressing sectoral productivity as a ratio of total productivity. Decomposition of labour productivity was calculated as described under Africa Sector Database.

3.1.3 GLOBAL EMPLOYMENT TRENDS (GET) 2014 SUPPORTING DATASETS – INTERNATIONAL LABOUR ORGANIZATION (ILO)

http://www.ilo.org/global/research/global-reports/global-employment-trends/2014/WCMS_234879/lang--en/index.htm

Provider's description

The annual Global Employment Trends report provides the latest global and regional estimates of employment and unemployment, employment by sector, vulnerable employment, labour productivity, and working poverty, while also analysing country-level issues and trends.

DFID focus country coverage

All are covered except **South Sudan**.

Indicators included and application to the accompanying analyses

Dataset: 'Employment by status and sex, globally and by region and country'

Status in total employment (thousands) – Total:	Wage and salaried workers Employers Own-account workers Contributing family workers *Vulnerable employment (= own-account workers plus contributing family workers)*
Status in total employment (thousands) – Male	Broken down as above
Status in total employment (thousands) – Female	Broken down as above

The sum of the figures for total wage and salaried workers, employers, own-account workers and contributing family workers was taken as 'total employment', and these totals were then broken down by sector according to the percentages given in WDI for the respective countries/years in order to calculate sectoral labour productivity (as described, and for the DFID focus countries listed, under WDI above).

[5] For some of the countries in some years, the sectoral employment shares for agriculture, industry and services shown in WDI do not add up to 100%. The reason for this is unclear.

Dataset: 'Share of employment by sector and sex, globally and by region and country'

Employment in agriculture	total (%)
	male (%)
	female (%)
Employment in industry	total (%)
	male (%)
	female (%)
Employment in services	total (%)
	male (%)
	female (%)

The male and female percentages for each sector were used to construct the graph 'Total employment by sex and sector' for all DFID focus countries.

Dataset: 'Employment by sector and sex, globally and by region and country'

Total employment in agriculture (thousands)	Male employment in agriculture
	Female employment in agriculture
Total employment in industry (thousands)	Male employment in industry
	Female employment in industry
Total employment in services (thousands)	Male employment in services
	Female employment in services

The male/female employment figures in each sector were calculated as a percentage of total employment in the sector, and the results used to construct the graphs on 'Sectoral employment by sex' for all DFID focus countries not included in the Africa Sector Database.

3.1.4 UNITED NATIONS STATISTICS DIVISION (UNSD) NATIONAL ACCOUNTS MAIN AGGREGATES DATABASE

https://data.un.org/search.aspx?q=gross+value+added+datamart%5bsnaama%5d

Provider's description

The Economic Statistics Branch of the United Nations Statistics Division (UNSD) maintains and annually updates the National Accounts Main Aggregates database. It consists of a complete and consistent set of time series, from 1970 onwards, of the main National Accounts aggregates of all UN Members States and other territories in the world for which National Accounts information is available. Its contents are based on the official data reported to UNSD through the annual National Accounts Questionnaire, supplemented with data estimates for any years and countries with incomplete or inconsistent information.

Estimates are done when no official data are available. The final aggregates are provided in national currency and in United States dollars. The main aggregates in the database include Gross Domestic Product (GDP) by type of expenditure and Gross Value Added by kind of economic activity, both at current and at constant market prices. It also contains Gross National Income (GNI) at current prices and analytical indicators and ratios that reflect economic structure and trends of countries and areas. These analytical indicators are: annual per capita GDP and GNI (calculated using actual population or annual population estimates from the UN Population Division); annual growth rates of GDP; annual shares in total GDP and total Gross Value added of their respective components; and price trends as defined by the implicit price deflators of GDP.

The National Accounts Main Aggregates database is the product of a global cooperation effort between UNSD, international statistical agencies, and the national statistical services

of more than 200 countries and territories. It has been prepared in response to a request by the Statistical Commission during its first session in 1947, and is in agreement with the resolution 48/223 C of the UN General Assembly, which mandates the Committee on Contributions to prepare the Scale of Assessment for UN Member States on the basis of the most recent reliable and verifiable available data on national accounts of as many countries and areas as possible, and that these data be published and disseminated regularly.

DFID focus country coverage
All 28 focus countries are included.

Indicators included and application to the accompanying analyses
Dataset: 'Gross Value Added by Kind of Economic Activity' (at current and constant (2005) US dollars)

Gross Value Added (GVA) (in current or constant (2005) US dollars) is broken down into seven categories of economic activity, classified according to ISIC Revision 3.1:

Sector		ISIC Rev. 3.1 division	
Agriculture, hunting, forestry, fishing		A–B	
Mining, Manufacturing, Utilities		C-E	
Manufacturing		D	
Construction		F	
Wholesale, retail trade, restaurants and hotels		G–H	
Transport, storage and communication		I	
Other Activities		J–P	
i.e.	*Financial intermediation*		*J*
	Real estate, renting and business activities		*K*
	Public administration and defence; compulsory social security		*L*
	Education		*M*
	Health and social work		*N*
	Other community, social and personal service activities		*O*
	Activities of private households as employers and undifferentiated production activities of private households		*P*
Not included:			
Extraterritorial organization and bodies		Q	

Country-specific notes accompanying the dataset indicate variations to the above sectoral composition for some countries in some or all years.

Data from this dataset were used alone to calculate sectoral shares in GVA at current and constant prices, and in conjunction with ILO WESO sectoral employment data (see Section 3.1.5) to calculate labour productivity. An explanation of how data from the two sources were linked follows the description of the ILO WESO dataset.

3.1.5 ILO WORLD EMPLOYMENT AND SOCIAL OUTLOOK (WESO) – TRENDS 2015, SUPPORTING DATASETS

http://www.ilo.org/global/research/global-reports/weso/2015/lang--en/index.htm

Provider's description
Besides other labour market indicators, this report also presents new estimates and projections of detailed sectoral and occupational employment shares, total and by sex. The principal database used for the sectoral and occupational employment shares is the most recent version of the ILO Key Indicators of the Labour Market (KILM), 8th edition

(http://www.ilo.org/kilm). The KILM is a wide-ranging database of 18 labour market indicators from 1980 to the latest available year for about 226 countries. Employment by sector captures the distribution of the employed population across sectors of economic activity. Employment by occupation is an indicator that attempts to categorize the employed population into groups of jobs with similar tasks and duties that are hierarchically organized in a number of levels. The main source of additional data is ILOSTAT, ILO Short-Term Indicators. For India, tabulations based on data from the National Sample Survey Organization survey on the Employment and Unemployment Situation in India were used. Data based on the most recent labour force survey conducted in Bangladesh were also considered.

The groups for occupations are selected in order to be representative of broad levels of skills as defined by educational level required and by type of skills required. The broad occupational categories (skills) are chosen according to the broad level of educational attainment required based on the International Standard Classification of Education; primary, secondary and tertiary educational level. The broad occupational categories (routine) are defined, following Jaimovich and Siu (2012) and Autor et al. (2003).

To produce estimates and projections, data on sectoral value added shares of GDP are taken from the United Nations Statistics Division – System of National Accounts – National Accounts Main Aggregates (UNSD SNAAMA) database. The Economist Intelligence Unit (EIU) database was used to supplement the data from the above sources and also to assist in projecting the value added shares by sector. For only one country (i.e. Taiwan, China) the entire series from EIU was utilized because there were data for this country neither in World Bank WDI nor in UNSD SNAAMA. The demographic variables used in the model come from the United Nations World Population Prospects, the United Nations World Urbanization Prospects and the ILO Economically Active Population Estimates and Projections.

Other sources of data for explanatory variables are the IMF World Economic Outlook database and the World Bank World Development Indicators database. For one variable, the IMF International Financial Statistics was used. The explanatory variables that are considered include: GDP per capita, output per worker, investment, exports of goods, imports of goods, general government final consumption expenditure, gross capital formation, trade in services, real effective exchange rate index, value added by sector, ratio of female-to-male labour force aged 30 to 64 years old, share of urban population in total population, share of population aged less than 15 years old in total population, share of population aged less than 15 years old and population aged above 65 years old in total population, share of wage and salary workers in total employment.

Estimates and projections are produced on the basis of the methodology that proceeds in three steps: (1) run regressions with a set of different combinations of the potential explanatory variables; (2) select the specifications for which the goodness-of-fit is best; and (3) run a bootstrap procedure on those specifications and calculate for each geographical region and each category of employment (i.e. sector and broad occupation) the RSME, based on this procedure. At this point, the RMSE is produced not only for these specifications but also for the average prediction among all the specifications selected, the average among the three best and among the five best performers. Then, for each sector or occupation and for each region, the specification with the lowest RMSE is selected to be used for the final estimates.

Finally, some adjustments are made to the estimates in order to make sure that the sum of shares across all categories equals 100 and that the sum of men and women working in a specific sectoral or occupational category equals the number of the estimate for both sexes.

DFID focus country coverage
Data are available for all except Somalia and South Sudan.

Indicators included and application to the accompanying analyses
Datasets: 'Employment by sector and sex' – thousands and percent

Total employment numbers (for males and females combined, males only and females only) are broken down between 14 economic activities classified according to ISIC Revision 4.

Sector	ISIC Rev. 4 division
Agriculture, forestry, hunting and fishing	A
Mining and quarrying	B
Manufacturing	C
Utilities (electricity, gas, etc.)	D and E
Construction	F
Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods	G
Accommodation and restaurants	I
Transport, storage and communication	H and J
Financial activities	K
Education	P
Health and social work activities	Q
Public administration and defence; compulsory social security	O
Real estate, business and administrative activities	L, M and N
Other services	R, S, T and U
i.e. *Arts, entertainment and recreation*	*R*
Other service activities	*S*
Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	*T*
Activities of extraterritorial organizations and bodies	*U*
Not included:	
Not classifiable by economic activity	X

Linking the UNSD GVA and ILO WESO employment data
These two sets of data were combined by correlating by broad ISIC division – aggregating the ILO data, and in one case disaggregating the UNSD data, as appropriate. The maximum number of sectors for which this could be done was seven (as this is the number covered by the UNSD data). The correlation made was as follows:

UNSD data (ISIC Rev. 3.1)	ILO WESO data (ISIC Rev. 4)	Sectors in our analysis
Agriculture, hunting, forestry, fishing (ISIC A–B)	Agriculture, forestry, hunting and fishing (ISIC A)	Agriculture
Mining, Manufacturing, Utilities (ISIC C–E) *minus* Manufacturing (ISIC D)	Mining and quarrying (ISIC B) *plus* Utilities (electricity, gas, etc.) (ISIC D and E)	Mining and utilities
Manufacturing (ISIC D)	Manufacturing (ISIC C)	Manufacturing
Construction (ISIC F)	Construction (ISIC F)	Construction

UNSD data (ISIC Rev. 3.1)	ILO WESO data (ISIC Rev. 4)	Sectors in our analysis
Wholesale, retail trade, restaurants and hotels (ISIC G–H)	Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods (ISIC G) *plus* Accommodation and restaurants (ISIC I)	Wholesale, retail, hotels
Transport, storage and communication (ISIC I)	Transport, storage and communication (ISIC H and J)	Transport, storage, communication
Other activities (ISIC J–P)	*Sum of:* Financial activities (ISIC K) Education (ISIC P) Health and social work activities (ISIC Q) Public administration and defence; compulsory social security (ISIC O) Real estate, business and administrative activities (ISIC L, M and N) Other services (ISIC R, S, T and U)	Other

A comparison of the sectoral breakdown tables above for the UNSD data and for the ILO WESO data indicates only two disparities:

- the UNSD data do not explicitly include arts, entertainment and recreation as a separate ISIC division, although the elements of this division (newly created in ISIC Rev. 4) were included as ‘classes’ in other divisions in ISIC Rev. 3.1.; and
- the UNSD data definitely do not include activities of extraterritorial organizations and bodies whereas the ILO WESO data do.

The former is considered unimportant (as relevant activities fall under ‘other’ in both datasets); the latter is an incompatibility.

Where feasible, country-specific variations in sectoral composition noted in the UNSD data are taken into account when aggregating the ILO WESO data. For example, where GVA relating to ‘restaurants and hotels’ is noted as included under ‘Other activities’ rather than under ‘Wholesale, retail trade, restaurants and hotels’, employment in the separate ‘Accommodation and restaurants’ division in ISIC Rev. 4 in the ILO WESO data has also been aggregated into ‘Other’. However it is not possible to deal with some other variations noted in the UNSD data in this way (for example, where it is noted that ‘Mining, Manufacturing, Utilities’ in the UNSD data (disaggregated in our analyses into ‘Mining and utilities’ and ‘manufacturing’) ‘Excludes publishing activities, includes irrigation canals’ no corresponding adjustment to the ILO WESO aggregations was possible, as neither publishing activities nor irrigation canals are separately identified in the sectoral employment data). For countries to which such notes apply, therefore, some of the employment data do not correspond exactly to the GVA data.

The correlated and combined data on sectoral GVA (in constant 2005 US dollars) and employment have then been used to calculate labour productivity, relative productivity, labour productivity indices and annualised growth.

3.1.6 EMPLOYMENT USING DEMOGRAPHIC AND HEALTH SURVEYS

We use data from McMillan and Harttgen (2014),[6] who analyse Demographic and Health Surveys (DHS) to analyse changes in employment to complement and examine robustness against other estimates of changes in employment discussed in previous sections.

Demographic and Health Surveys (DHS)[7]
http://www.dhsprogram.com/data/available-datasets.cfm

The DHS data can be used to understand changes in the share of employment in agriculture, by gender, by location, and over time. The above study focuses on African countries, but the link above contains data for other countries as well. However, further work on data cleaning would be required for the non-African countries. This is because the surveys are not all in one place, so they have to be downloaded and combined, and they also contain coding errors, which need to be cleaned. This also makes it difficult to use aggregate numbers available from the website.

The DHS data include spatial information, and hence it is possible, in principle, to examine transformation by geographic location within a country.

The downside of the data is that there are no wages, so it is difficult to discuss issues related to productivity. However, there is a lot of information about health and education and some information about wealth quintiles, so it is possible to say something about changes in welfare (poverty) using the DHS data.

We have used the data contained in McMillan and Harttgen (2014), in particular data on the percentage of workers (age 25+) in agriculture. Economic transformation is generally associated with a decline in employment in agriculture, but this can differ by location or gender. We provide data for men and women.

3.2 TRADE

We discuss the following data sources for trade-related variables:

- basic trade data (3.2.1)
- measures of diversification (3.2.2)
- data on revealed comparative advantage (3.2.3)
- trade in value added (3.2.4)

3.2.1 BASIC TRADE DATA

The main datasets we use to describe basic data include the UN's Comtrade database and the Atlas of Economic Complexity, which we discuss now.

3.2.1.1 COMTRADE (COMMODITY TRADE STATISTICS) DATABASE (UNITED NATIONS)

http://comtrade.un.org/data/ or via the World Bank's World Integrated Trade Solution (WITS) portal (https://wits.worldbank.org/WITS/WITS/Restricted/Login.aspx, annual data only)

[6] McMillan and Harttgen (2014) *(http://www.nber.org/papers/w20077)*.

[7] See http://www.dhsprogram.com/pubs/pdf/DHSG1/Guide_to_DHS_Statistics_29Oct2012_DHSG1.pdf for a guide. Thanks to Margaret McMillan for insights on use of DHS.

Description
UN Comtrade contains detailed import and export statistics reported by statistical authorities of more than 200 countries or areas. It concerns annual trade data from 1962 to the most recent year. UN Comtrade is considered the most comprehensive trade database available.

All commodity values are converted from national currency into US dollars using exchange rates supplied by the reporter countries, or derived from monthly market rates and volume of trade. Quantities, when possible, are converted into metric units. Commodities are reported in the current classification and revision and are converted all the way down to the earliest classification SITC revision 1. Therefore, if data is received in HS2012 version, it is converted to HS2007, HS2002, HS1996, HS1988, SITC rev. 3, SITC rev. 2, SITC rev. 1 and BEC.

Concepts and definitions
See http://unstats.un.org/unsd/trade/eg-imts/IMTS%202010%20(English).pdf.

DFID focus country coverage
All the DFID focus countries are reporters except **DR Congo, Liberia, Sierra Leone, Somalia, South Sudan,** and **Tajikistan** (none of which has reported in recent years, if ever). All DFID focus countries are included as partners.

Own-reported data from UN Comtrade has been used in the accompanying analyses for all countries other than those listed above. The number of years in which countries have reported, and the latest year in which they have done so, varies widely.

3.2.1.2 TRADE MAP – INTERNATIONAL TRADE CENTRE (ITC)
http://legacy.intracen.org/marketanalysis/default.aspx

Provider's description
Trade Map provides — in the form of tables, graphs and maps — indicators on export performance, international demand, alternative markets and competitive markets, as well as a directory of importing and exporting companies.

Trade Map covers 220 countries and territories and 5,300 products of the Harmonised System. The monthly, quarterly and yearly trade flows are available from the most aggregated level to the tariff line level.

The yearly data in Trade Map for products at 2, 4, and 6-digit level of the Harmonised System are mainly based on UN Comtrade, the world's largest database of trade statistics, maintained by the United Nations Statistics Division (UNSD). This data is complemented by national sources when the information is not available in UN Comtrade. The quarterly and monthly data come from national and regional sources. A footnote shows the source used in the displayed table.

Data are available also for countries that do not report their national trade statistics to UN Comtrade. The trade of these countries has been reconstructed on the basis of data reported by partner countries – mirror data. Mirror data are better than no data at all, but they have a number of shortcomings:

- Mirror data do not cover trade with other non-reporting countries. Consequently they do not cover intra-African trade accurately.
- The trans-shipment issue can hide the true origin of goods.
- Mirror data invert the reporting standards by valuing exports in CIF terms (i.e. including transportation and insurance costs) and imports in FOB terms (i.e. excluding these services).

- The number of reporting countries is different from one year to another. Hence comparisons over time using mirror data need to be interpreted with caution.

DFID focus country coverage
All DFID focus countries are covered except **South Sudan** – although the data for those not covered in the UN Comtrade database are mirror data, which need to be interpreted with caution (see above).

Application in the accompanying analyses
Trade Map mirror data have been used for the basic trade data analyses undertaken for **DR Congo, Liberia, Sierra Leone, Somalia and Tajikistan** (and to augment the single year's data reported to Comtrade by **Myanmar**).

The data above are used to provide data on (i) value in trade (exports and imports), (ii) export by broad HS category (21 product groups), share of total, and changes in such shares.

3.2.1.3 ATLAS OF ECONOMIC COMPLEXITY – CENTER FOR INTERNATIONAL DEVELOPMENT AT HARVARD UNIVERSITY
http://www.atlas.cid.harvard.edu

Provider's description
The Atlas online is an interactive tool that enables users to visualise a country's total trade, track how these dynamics change over time and explore growth opportunities for more than a hundred countries worldwide.

The Atlas is used by investors, entrepreneurs, policy-makers, students and the general public to better understand the competitive landscape of countries around the globe. For any given country, The Atlas shows which products are produced and exported; The Atlas can then use this information to suggest products a country could begin manufacturing in order to fuel economic growth. As a dynamic resource, The Atlas is continually evolving with new data and features to help analyse economic growth and development.

The Atlas can answer questions such as:

- What does a country import and export?
- How has its trade evolved over time?
- What are the drivers of export growth?
- Which new industries are likely to emerge in a given geography? Which are likely to disappear?
- What are the GDP growth prospects of a given country in the next 5-10 years, based on its productive capabilities?

DFID focus country coverage
Bangladesh, Ethiopia, Ghana, India, Kenya, Kyrgyz Republic, Liberia, Malawi, Mozambique, Nigeria, Pakistan, South Africa, Sudan, Tajikistan, Tanzania, Uganda, Yemen, Zambia, Zimbabwe.

Application in the accompanying analyses
Visualisations of exports and imports in 2005 and 2012 are included for each of the above countries.

3.2.1.4 OBSERVATORY OF ECONOMIC COMPLEXITY – MASSACHUSETTS INSTITUTE OF TECHNOLOGY

http://atlas.media.mit.edu/explore/tree_map/hs/export/che/all/show/2012/

DFID focus country coverage

All DFID focus countries are covered except **West Bank & Gaza**.

Application in the accompanying analyses

Visualisations of exports and imports in 2005 and 2012 are included for **Afghanistan, DR Congo, Myanmar, Nepal, Rwanda, Sierra Leone, Somalia and South Sudan** (i.e. all available for which data were not available in The Atlas of Economic Complexity).

3.2.2 MEASURES OF TRADE DIVERSIFICATION

We use two broad measures of diversification: (i) a toolkit provided by the IMF and (ii) own calculation using Comtrade data.

3.2.2.1 DIVERSIFICATION MEASURES FROM IMF TOOLKIT

http://www.imf.org/external/np/res/dfidimf/diversification.htm

Provider's description

Covering 187 countries including most low-income countries (LICs), the toolkit provides indicators on export product diversification and export product quality from 1962-2010. The measures in this toolkit are based on an updated version of the UN–NBER dataset, which harmonises Comtrade bilateral trade flow data at the 4-digit SITC (rev. 1) level. The export diversification and quality database was developed by IMF staff under an IMF–DFID research collaboration.

The Export Diversification Database has three main indicators: the Export Diversification Index, the Extensive Margin, and the Intensive Margin. Extensive export diversification reflects an increase in the number of export products or trading partners. Intensive export diversification considers the shares of export volumes across active products or trading partners. Thus, a country is less diversified when export revenues are driven by only a few sectors or trading partners, even though the country might be exporting many different goods or to many different trading partners. Countries with a more evenly balanced mix of exports or trading partners have a higher level of intensive diversification. Higher values for the all three indices indicate lower diversification.

The Export Quality Database contains export quality measures across different aggregation levels of export products. Higher values for the quality indices indicate higher quality levels.

DFID focus country coverage

All DFID focus countries are covered except **South Sudan**.

Database methods

See: About Diversification Index and About Quality Index

Application to the accompanying analyses

For each of the DFID focus countries, one graph has been produced from the data in each of the databases above. The data used for the graph on export quality have been drawn from the most aggregated dataset – 'Overall and 1-digit level'; more detailed quality indices are available on the link above.

3.2.2.2 DIVERSIFICATION MEASURES BASED ON COMTRADE DATA

We have also computed two simple diversification measures (number of export items at HS6 sub head level and number of destinations), using Comtrade data, for both exports and imports.

3.2.3 DATA ON REVEALED COMPARATIVE ADVANTAGE

Data from the UN's Comtrade database (described under 'Trade – Basic data' above) were used to calculate revealed comparative advantage (the Balassa index) by HS Section. The share of each country's exports in a Section was calculated as a percentage of the country's total exports in HS Chapters 1-97, and these shares were then expressed as a ratio of the analogous shares of world exports. 'World' is defined as Comtrade's 'All countries' aggregate (which represents the sum of the data reported by all countries in any given year).

3.2.4 TRADE IN VALUE ADDED[8]

We use the Eora Multi-Region Input-Output Database to examine the domestic and foreign value addition in exports.

3.2.4.1 EORA MULTI-REGION INPUT-OUTPUT DATABASE[9]

http://www.worldmrio.com/

Description

Trade in value added data is based on the simplified 26-sector Eora Multi-Region Input-Output (MRIO) database (Eora26), which is disaggregated into 189 countries and 26 sectors per country, as well as a rest of world sector that captures other countries and statistical discrepancies. This is the condensed version of Eora with countries that have more than 26 sectors in their input-output or supply-use tables having their accounts simplified (primarily high-income and middle-income countries (HICs and MICs)). However, this does not apply to the vast majority of the DFID priority countries under analysis.

Measuring trade in value added via MRIOs like Eora has a number of clear benefits (see Ahmad, 2013) including the following:

- It provides a crucial tool to understand a country's actual industrial structure and the national and international interlinkages of sectors for developing growth and development strategies, as well as trade and industrial policies.
- It makes clear how particularly non-tariff barriers (including regulatory measures) can affect competitiveness and upstream producers.
- It can help policy-makers better anticipate ex ante the potential impact of macroeconomic shocks.
- It can allow for calculations of the 'job content' of trade.
- Environmentally extended MRIOs, like Eora, allow for an assessment of the impact of trade as it affects ecosystem services.

Methodology

There are now a number of global input-output databases that vary significantly in terms of country, sectoral coverage, time span and approach (see Table 4 below).

[8] The text and calculations are based on work by Jakob Engel.

[9] Lenzen, M., Kanemoto, K., Moran. D. and Geschke, A. (2012) 'Mapping the structure of the world economy', *Environmental Science & Technology* 46(15): 8374-8381. DOI: 10.1021/es300171x

Lenzen, M., Moran, D., Kanemoto, K. and Geschke, A. (2013) 'Building Eora: A Global Multi-regional Input-Output Database at High Country and Sector Resolution', *Economic Systems Research* 25(1): 20-49. DOI:10.1080/09535314.2013.769938

Table 4: Overview of main MRIO databases

Name	Countries	Type	Detail (I x p)[a]	Time	Extensions	Approach
Eora	World (over 180)	MR SUT/IOT	Variable (20-500)	1970-2010	Various, especially environmental	Create initial estimate, gather all data in original formats, formulate constraints; detect and judge inconsistencies; set routine; calculate global MR SUT/IOT
EXIOPOL/ CREEA	World (43 countries + RoW)	MR SUT	129 x 129	2000 and 2007	30 emissions, 60 IEA, energy carriers, water, land, 80 resources	Create SUT; split use into domestic and imported use; detail and harmonise SUTs; use trade shares to estimate implicit exports; confront with exports in SUT; RAS out differences; add extensions
WIOD	World (40 countries + RoW)	MR SUT	35 x 59	1995-2009, annually	Detailed socio-economic and environmental satellite accounts	Harmonise SUTs; create bilateral trade database for goods and services; adopt import shares to split use into domestic and imported use; trade information for RoW is used to reconcile bilateral trade shares; add extensions
GTAP-MRIO	World (129 countries)	MR IOT	57 x 57	1990, 1992, 1995, 1997, 2001, 2004, 2007	5 (GWP), land use, energy volumes, migration	Harmonise trade; use IOTs to link trade sets; IOT balanced with trade and macroeconomic data
GRAM	World (40 countries)	MR IOT	48 x 48	2000, 2004	Various	Use harmonised OECD IOTs; neglect differences like *i* x *i* and *p* x *p*; use OECD bilateral trade database to link trade
IDE-JETRO	Asia-Pacific (8: 1975; 10: 1985-2005)	MR IOT	56 x 56 (1975); 78 x 78 (1985-1995); 76 x 76 (2000, 2005)	1975 - 2005	Employment matrices (2000, 2005)	Harmonise IOTs based on cross-country survey information; link via trade; manual balancing to reduce discrepancies within certain bounds

[a] *i – number of industries, p – number of products; SUT – supply-use tables, IOT – Input-Output tables*
Source: Tukker and Dietzenbacher (2013)

The Eora database, much like the World Input-Output Database (WIOD) and the OECD/WTO's trade in value added (TiVA) database, uses available information to produce measures of trade in value added for all countries, in order to produce a contiguous and continuous dataset, interpolating for countries that do not have the necessary data.

The Eora MRIO dataset has recently been used for global value chain (GVC)-related analyses in a number of international reports including the *African Economic Outlook* (AfDB, 2014) and the *World Investment Report* (UNCTAD, 2013). It has several advantages over other databases (see Lenzen et al., 2012), including the following:

- It is disaggregated into 189 countries, providing important advantages for assessing impacts of consumption and production on relatively poor countries.
- It has a historical time series spanning 1990-2011 (soon to be extended to 1970-2011 and updated with an approximate two year delay).
- It includes tables of basic prices, as well as two margins (taxes on products and subsidies on products).
- To clarify levels of uncertainty, standard deviation estimates have been calculated for all MRIO events.

- It is publicly available at www.worldmrio.com.

Developing an MRIO like Eora requires combining multiple data sources to ensure reasonably accurate estimates, with domestic input-output tables seen as the most reliable source. The construction of Eora was based on the principle of changing the structure of the original data sources as little possible for the sake of transparency. Its matrices are based on the use of the following types of raw data (in order of assumed accuracy):

- input-output (I-O) tables and main aggregates data from national statistical offices where these are available (these are generally not available for low-income countries (LICs)
- I-O compendia from Eurostat, IDE-JETRO, and OECD
- the UN National Accounts Main Aggregates Database
- the UN National Accounts Official Data
- the UN Comtrade international trade database
- the UN Service Trade international trade database.

Eora has a historical time series spanning 1990-2011 based on an iterative process using an initial year estimate for 2000, overlaying estimates for 1999 and 2001, respectively with new data, and then re-balancing. This makes it well suited for analyses of smaller developing countries not included in other datasets;[10] however, it should be complemented by more nuanced sectoral analysis drawing on alternative methodologies.

In terms of key indicators, this report draws on the methodologies first developed by Hummels, Ishii and Yi (2001) in measuring vertical specialisation and in turn formalised by Koopman, Powers, Wang and Wei (2011) (and later Koopman, Wang and Wei 2014) to derive some of the most commonly used trade-in-value-added indicators, including domestic and foreign value added.

Of the 26 available sectors, the 15 that were selected for more detailed analysis are as follows:

- Agriculture
- Electrical and Machinery
- Financial Intermediation and Business Activities
- Fishing
- Food & Beverages
- Hotels and Restaurants
- Metal Products
- Mining and Quarrying
- Other Manufacturing
- Petroleum, Chemical and Non-Metallic Mineral Products
- Post and Telecommunications
- Textiles and Wearing Apparel
- Transport
- Transport Equipment
- Wood and Paper

[10] In comparative analysis (UNCTAD, 2013) with the WIOD dataset, Eora was found to provide broadly similar results when calculating foreign and domestic value added, albeit with a slight upward bias (which is to be expected because of the greater number of highly heterogeneous developing countries, many of which have been subsumed in WIOD's rest of world matrix).

This is intended to capture the most important tradable sectors in MICs and LICs, but omits other economic sectors such as public administration or recycling. These are captured in the 'Other' category.

Benefits

MRIO tables allow for analysis of value contribution along supply chains, enabling the analysis of sources and destinations of value that flow through GVCs. In most cases Eora results provide a reasonably accurate estimate for key indicators of GVC competitiveness at the country level, and of relative performance of sectors in relation to other sectors within an economy and in relation to comparator countries. Further, they are likely to provide a reasonably reliable approximation of the sectoral decomposition of value added and the direction of value-added trade, as well as trends for these indicators over a 15-year time period (especially as all four data-points are derived from the same source and methodology). Thus, in the absence of national Input-Output Tables (IOTs) and Supply-Use Tables (SUTs), and despite substantial uncertainties (particularly at the sectoral level), Eora provides the best available and internationally comparable dataset for calculating key value-added trade indicators, though they are best complemented by more nuanced sectoral analysis drawing on alternative methodologies (e.g. producer surveys, firm-level analysis, and case studies).

Limitations

However, a few caveats relating particularly to the accuracy of the Eora data are in order. First, it is important to bear in mind that Eora's MRIO tables were modelled based on existing sources when national input-output or supply-use tables were not available. As mentioned above, this includes national accounts data, and Comtrade import and export data, among others. The availability of data for 189 countries comes at the expense of accuracy. While WIOD and TiVA relied primarily on national input-output and/or supply-use tables for the countries analysed, these are not available for most developing countries: among the countries of interest, only India (1993, 1998, 2003 and 2006), Kenya (2003), Kyrgyzstan (2001) and South Africa (1993, 1998, 1999, 2000, 2002, and 2005) have input-output and/or supply-use tables available. As a result, the following data sources were used:

Data sources	Data	Number of categories	Price
National Accounts	final demand	4	purchasers' prices
Main Aggregates	value added	7	basic prices
Database	imports	1	f.o.b.
	exports	1	f.o.b.
UN National	gross output	1	basic prices
Accounts Official	intermediate demand	1	purchasers' prices
Data	final demand	6	purchasers' prices
	value added	18	basic prices
UN Comtrade	exports	About 5000 (HS 6-digits)	f.o.b.
	Imports	About 5000 (HS 6-digits)	c.i.f.

Source: Lenzen et al. 2013

However, when looking at trade data, even in the presence of a national IOT, none of these databases give a complete picture of trade, as national import matrices provide no information on the exporting country, and trade databases have no information on the using sector (Lenzen et al. 2013, p. 14). This is even further complicated in the absence of reliable import matrices.

Further, in order to achieve the MRIO's overall balancing requirements, raw data on imports and exports have been at times misrepresented, with the overall focus on representing large data items and fulfilling balancing conditions for large countries. Eora's optimisation approach attempts to strike a balance between the frequent conflicts between country-level total exports and imports and trade balances, but this has likely led to

substantial uncertainties, particularly for small values (such as those of LICs and MICs that have not developed national IOTs). [11]

Given these slight imbalances, total gross exports are at times marginally larger or smaller than the value added constituting them, and the sum of domestic and foreign value added as a share of gross exports is generally not exactly equal to 100% at the national and particularly sectoral level. Currently total gross exports (calculated as a country's contribution to all other countries' transaction and final demand matrices) have been left as the denominator, but it would also be possible to take the sum of exported and domestic value added as the denominator in order to have value added shares that reflect the relative importance of foreign and domestic contributions to exported value added (and so that these shares equal 1, as they mathematically should).[12]

Second, the calculation of Eora-derived indicators for this project was based on the reduced Eora26 version, with no subsequent corrections carried out. As reported by the Eora developers, there are likely to be discrepancies between the full Eora database and Eora26, which is the product of some aggregation. As a consequence of this, results may differ slightly from those in other publications such as the UNCTAD *World Investment Report 2013* or the AfDB *African Economic Outlook* 2014.

Some limitations also go beyond just Eora and affect all MRIO tables. They are subject to a degree of uncertainty – as is common with any applications using accounts data and trade flow data – which is augmented when examining developing counties, where statistical capacities tend to be substantially lower than in developed countries. They furthermore are subject to two assumptions that contribute to their uncertainty. First, they assume that all products (for export and domestic use) have the same import content (proportionality assumption – see Winkler and Milberg, 2012). This is particularly relevant for analysing trade flows from developing countries (OECD, 2012). Second, they assume a uniform use of inputs among all firms in a sector (homogeneity assumption).

Country coverage

All SET countries are included in the database, but data for Zimbabwe for 2011 and for South Sudan and Sudan prior to 2011 are highly flawed, so these have been omitted. Among the Occupied Palestinian Territories, only data for the Gaza Strip are provided.

Explanation of indicators and further extensions[13]

Domestic value added (DVA) embodied in exports: The potential benefits from GVC participation are most clearly demonstrated by the evolution of its DVA embodied in gross exports over time. Put simply, increasing DVA embodied in gross exports over time signifies greater value addition within the country itself. At the industry level, DVA consists of value added created in a specific industry itself, value added created in other domestic sectors supplying this industry, as well as previously exported intermediates re-imported from abroad for use in a given industry.

[11] One obvious example of this is that some countries – especially smaller developing countries and those where trade exceeds GDP – are not well balanced (i.e. the ratio of gross national expenditure + exports versus GDP + imports is not equal to 1) – see worldmrio.com/EoraFAQ.pdf.

[12]The Eora Frequently Asked Questions provide an explanation for the causes of these imbalances: '*data on country-wise total exports and imports fundamentally conflict with global trade balances. One cannot achieve a balanced global multi-region input-output table whilst at the same time respecting data on exports and imports. This means that in a real MRIO table, either balancing conditions must be violated or raw data mis-represented'. Under ideal balancing conditions, national ratios of Gross National Expenditure + exports versus Gross Domestic Product + imports should be 1, i.e. in an IO table the total of all inputs to a given sector (i.e. the column values) should equal the total value of that sector's outputs (i.e. the row values). However, due to data conflicts, this is in most cases a few percentage points more or less than 1.*

[13] See also OECD 2012

Domestic value added embodied in exports as a share of exports: This indicator can be a relatively good proxy of value chain participation, with a decline in DVA as a share of gross exports indicative of participation in longer and more sophisticated value chains where more imported value added is in turn being re-exported. However, it could also suggest the declining significance of the services economy, which tends to have short value chains and high values of exported DVA as a share of exports.

Foreign value added (FVA): This is an indicator of backward integration. Increases in FVA in exports are indicative of a greater share of foreign inputs being exported by the respective country. Increasing backward integration tends to be particularly important for developing countries as it links to a number of measures of structural transformation.

The country files include the country-level analyses. A range of comparative analysis can be undertaken, and here we include just two. Figure 1 shows the domestic value added as % of gross exports by country for both 1996 and 2011. It shows that (i) this percentage varies amongst countries; (ii) this percentage declined from 1996 to 2011, and (iii) the dispersion has increased, which means that there is wider variety in the extent to which countries use foreign markets in their exports.

Figure 1: Domestic value added in exports as % of gross exports (2011 vs 1996) by country

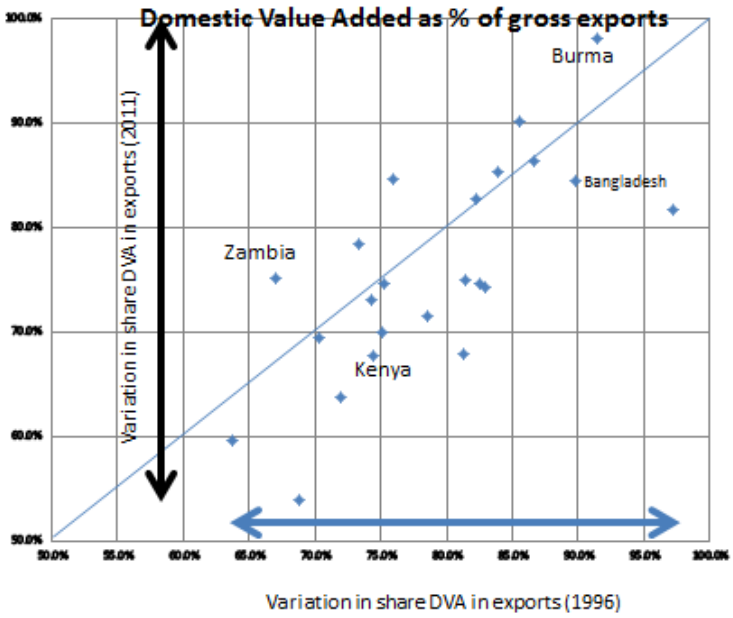

Table 5 examines the domestic value addition (as a percentage of gross exports) by sector for the average of 24 countries. It shows how the percentage varies by sector, and also how it has declined over time. Moreover, with the exception of the food and beverages sector, all sectors have experienced greater variation in the percentage. Thus countries and sectors have become increasingly diverse in their experiences with value addition.

Table 5: DVA content of gross exports as ratio of gross exports

Sector	Mean		Standard deviation	
	1996	*2011*	*1996*	*2011*
Financial Intermediation and Business	0.87	0.86	0.08	0.11
Post and Telecommunications	0.86	0.84	0.07	0.09
Agriculture	0.82	0.82	0.10	0.11
Hotels and Restaurants	0.77	0.76	0.12	0.15
Food & Beverages	0.75	0.73	0.16	0.14
Wood and Paper	0.75	0.69	0.14	0.17
Petroleum, Chemical and Non-Metallic Minerals	0.73	0.67	0.13	0.16
Electrical and Machinery	0.73	0.66	0.10	0.14
Metal Products	0.74	0.66	0.12	0.16
Textiles and Wearing Apparel	0.74	0.66	0.15	0.2
Fishing	0.69	0.65	0.22	0.24
Other Manufacturing	0.74	0.65	0.12	0.18
Transport	0.66	0.64	0.16	0.17
Transport Equipment	0.69	0.61	0.12	0.16
Mining and Quarrying	0.60	0.60	0.22	0.23

Source: Data analysis based on Eora data for 24 countries (excludes ETH, SUD, ZWE, SDS owing to data outliers)

3.3 TOTAL FACTOR PRODUCTIVITY AT FIRM LEVEL

We also use productivity measures at the firm level. Economic transformation occurs when resources are shifted from low-productivity to high-productivity activities. This happens, for example, when resources are shifted from low-productivity firms to high-productivity firms within a sector. Generally the scope for such shifts is greater in developing countries than in developed countries, because there is less pressure and competition (e.g. more protection) in developing countries and hence fewer penalties for being less productive.

We can examine total factor productivity at firm level using WB enterprise survey data (data available from http://www.enterprisesurveys.org/).

We follow Saliola and Seker (2012) in estimating productivity: Productivity is measured as the residual term of estimating a standard production function as follows

*Ln (value-added) = a*capital + b*labour + residual*

Where value-added is measured as firm sales (database code: d2) minus cost of raw materials and intermediate materials (n2e), capital as value of property, machinery and land (n6a + n6b), labour as total compensation (n2a n4a).

It is possible to estimate different production functions, but we have not yet done this. It is also possible to compare productivity levels across countries (e.g. similar sectors), but this requires exchange rate conversions (using market exchange rates or PPPs) as well as base year conversions (using GDP or CPI deflators). We have not yet included comparisons.

The files include two charts per country using the following:

- **Dispersion in productivity across firms by sector**, which shows two types of density distribution – histogram (bar chart) and kernel (the line). The higher the peak around zero, the more firms are centred around average productivity. The lower the peak and the wider the distribution, the more dispersion in productivity is present in a sector, which indicates large opportunities for improving productivity within a sector.
- **Distribution in productivity by country – comparing kernel and normal distributions**. For many countries, the kernel density peaks higher and before the normal distribution. This means that a large number of firms have a

productivity that is just below the average level productivity and then with a long tail to the right, which means very few firms have very high levels of productivity (although this could also pick up sector effects).

Table 6 shows country-level estimates underlying the productivity estimations. Overall, the estimations suggest that country-level production functions tend to be Cobb-Douglas with a labour share of between around 2/3 to 5/6.

To examine whether exporters have greater productivity, we estimate:

*Residual = a*exporter+residual2*

Table 6: Production function estimates based on firm-level data

	No of obs.	Coefficient for capital stock	Coefficient for labour	Coefficient on exporter status for productivity
Afghanistan 2014	14	0.41 (0.31)	0.32 (0.39)	-
Bangladesh 2013	1090	0.19 (0.02)	0.78 (0.02)	0.38 (0.06)
DR Congo 2013	142	0.16 (0.06)	0.83 (0.07)	0.37 (0.39)
Ethiopia 2011	98	0.11 (0.08)	0.65 (0.09)	0.23 (0.39)
Ghana 2011	126	0.26 (0.05)	0.65 (0.06)	0.29 (0.31)
Kenya 2013	220	0.11 (0.05)	0.89 (0.05)	0.07 (0.17)
Kyrgyz Republic 2013	44	0.23 (0.09)	0.87 (0.16)	-0.73 (0.32)
Liberia 2009	136	0.31 (0.07)	0.79 (0.09)	-
Malawi 2009 (no capital stock data)				-
Mozambique 2007	277	0.03 (0.03)	1.08(0.05)	0.43 (0.16)
Myanmar 2014	135	0.15 (0.05)	0.71 (0.06)	0.59 (0.19)
Nepal 2013	179	0.12 (0.06)	1.26 (0.08)	-0.26 (0.24)
Nigeria 2007	913	0.14 (0.01)	0.91 (0.02)	0.26 (0.13)
Pakistan 2007	110	0.36 (0.06)	0.66 (0.07)	0.02 (0.24)
Rwanda 2006	51	0.14 (0.06)	0.95 (0.09)	0.43 (0.25)
Sierra Leone 2009	143	0.38 (0.06)	0.59 (0.10)	-0.10 (0.52)
Somalia				-
South Africa 2007	652	0.07 (0.01)	0.94 (0.01)	0.01 (0.04)
South Sudan				-
Sudan				-
Tajikistan	44	0.07 (0.05)	0.92 (0.07)	0.36 (0.37)
Tanzania 2013	124	0.28 (0.06)	0.64 (0.07)	0.17 (0.14)
Uganda 2013	79	0.13 (0.07)	0.61 (0.10)	0.75 (0.37)
West Bank and Gaza 2013	66	0.16 (0.06)	0.62 (0.09)	-
Yemen 2013	66	0.53 (0.10)	0.22 (0.13)	1.11 (0.82)
Zambia 2007	288	0.17 (0.02)	0.76 (0.03)	0.23 (0.11)
Zambia 2013	142	0.10 (0.03)	0.93 (0.05)	0.87 (0.27)
Zimbabwe 2011	350	0.12 (0.04)	0.73 (0.03)	0.15 (0.14)

OLS estimation, standard errors between parentheses.

Table 7: Productivity distributions across firms, moments (standard deviation, skewness and kurtosis)

	No of obs.	Standard deviation	Skewness	Kurtosis
		Is there a wide variation in productivity?	*Is there a long tail on the right (positive), i.e. is distribution right skewed?*	*Are there extreme deviations and frequent modestly sized deviations? (the higher the value, the more the kurtosis)*
Afghanistan 2014	14	1.45	0.09	1.99
Bangladesh 2013	1090	0.99	1.18	5.37
DR Congo 2013	142	1.45	2.17	8.95
Ethiopia 2011	98	1.53	0.97	4.96
Ghana 2011	126	1.16	0.62	3.78
Kenya 2013	220	1.31	0.96	3.88
Kyrgyz Republic 2013	44	1.05	1.77	9.33
Liberia 2009	136	1.80	1.56	5.21
Mozambique 2007	277	1.08	5.69	54.6
Myanmar 2014	135	0.91	0.83	3.03
Nepal 2013	179	0.96	0.37	2.65
Nigeria 2007	913	0.61	0.84	4.32
Pakistan 2007	110	1.33	0.84	4.19
Rwanda 2006	51	0.71	-0.08	4.16
Sierra Leone 2009	143	1.26	1.78	7.70
South Africa 2007	652	0.49	1.02	4.67
Tajikistan	44	0.82	1.17	3.69
Tanzania 2013	124	1.00	2.04	8.30
Uganda 2013	79	1.54	0.8	4.40
West Bank and Gaza 2013	66	0.96	0.54	2.68
Yemen 2013	66	2.54	-2.40	17.5
Zambia 2013	142	0.99	0.60	3.20
Zimbabwe 2011	350	0.97	1.38	5.84

Which country has the greatest variation in productivity?

- Top 5 countries for greatest variation: Yemen, Liberia, Uganda, Ethiopia, DR Congo.
- Bottom 5 countries for smallest variation: South Africa, Nigeria, Rwanda, Tajikistan, Myanmar.

Which country has the greatest right tail in productivity?

- Top 5 countries for greatest skewness: Mozambique, DR Congo, Tanzania, Sierra Leone, Kyrgyz Republic.
- Bottom 5 countries for smallest skewness: Yemen, Rwanda, Afghanistan, Nepal, West Bank.

Which country has the highest peaks?

- Top 5 countries for highest kurtosis: Mozambique, Yemen, Kyrgyz Republic, DR Congo, Tanzania.
- Bottom 5 countries for lowest kurtosis: Afghanistan, Nepal, West Bank, Myanmar, Zambia.

3.4 WAGES BY OCCUPATION

3.4.1 OCCUPATIONAL WAGES AROUND THE WORLD (OWW) DATABASE
http://www.nber.org/oww/

This database contains occupational wage data for 161 occupations in 171 countries from 1983 to 2008. The occupational wage data are derived from the ILO October Inquiry database (http://laborsta.ilo.org) by calibrating the data into a normalised wage rate for each occupation.

The ILO October Inquiry is the most far-ranging survey of wages by occupation around the world. It started in 1924 in 15 countries for 18 occupations and has covered 161 occupations since 1983. The 1983-2008 ILO October Inquiry is mostly available in electronic format (http://laborsta.ilo.org), except for data for 45 occupations which are only available in hardcopy format. We have therefore scanned the 1983-1984 ILO October Inquiry data, which was not available in electronic format, creating a complete electronic version of the ILO October Inquiry for the 1983-2008 period.

The normalised wages refer to average hourly or monthly wage rates for adult workers. We use monthly wages (US$) for a few representative occupations and express them as a ratio of the average monthly wage – such as for a plantation supervisor, a labourer, a computer programmer, an office clerk, a primary school teacher, a plumber, an accountant, an automobile mechanic, etc. The advantage of OWW is that this database provides rich sources on wages by occupation (a key measure of productivity and economic transformation), but the disadvantage is the limited cover across occupations, time and countries.

Oostendorp (2012) provides details of the data cleaning and interpolations needed to obtain estimates. Many countries report more than one wage for a single occupation. Some give hourly wage rates and average earnings. Others give wages for men and wages for women. Others give wages for one gender and for both genders. Many contain multiple wage figures. All this can help with the calibration of the data into a standardised format, but it makes it difficult to use the raw data in cross-country comparisons, particularly since different countries report pay differently.

For monthly wages, Oostendorp provides the following data:

<u>Monthly wages in US$:</u>
mw1us: wage reported in standard format in raw data (type 1)
mw2wuus: wage with country-specific calibration (type 2, uni weighting)
mw3wuus: wage with country-specific calibration (type 3, uni weighting)
mw4wuus: wage with country-specific and uniform calibration (type 4, uni weighting)
mw2wlus: wage with country-specific and uniform calibration (type 2, lex weighting)
mw3wlus: wage with uniform calibration (type 3, lex weighting)
mw4wlus: wage with uniform calibration (type 4, lex weighting)

We have selected the mw3wuus variable in the Stata file that can be downloaded from the OWW link.

4. PROS AND CONS OF DIFFERENT DATA SOURCES

Table 8 discusses pros and cons of the use of different databases for analysing ET. Some examples include the following:

- There is good availability of trade statistics in general (though for some countries we need to use mirror trade data), including at developing-country level and at detailed product/sector levels.
- There are a range of methods to explore input-output tables; some are of high quality, others offer excellent country coverage.
- Sector national accounts are often of poor quality, and it usually takes a number of consistency checks to get to the data. There are gaps; e.g. poor Asian countries do not tend to be included in cross-country databases apart from WDI databases, which are weak in sector coverage.
- Wages by occupation are very weak in comparability and coverage.
- There is good country coverage of WB enterprise surveys, but the number of firms interviewed is limited.

We will further develop the characteristics and the pros and cons of these databases.

Table 8: Pros and cons of different data sources

Area of transformation	Data sources	Pros	Cons	Uses
Economic structure				
	Groningen productivity database (10-Sector Database)	10 sectors; employment includes informal sector. 33 countries – in Africa (11), North Africa (2), Asia (11) and Latin America (9), plus USA and 8 EU countries	Few DFID focus countries included (8 in Africa, 1 in Asia).	Sectoral shares in value added and employment; labour productivity by sector.
	World Development Indicators	Wide country/time coverage.	Few sector details; needs to be updated after rebasing; no employment level estimates.	Sectoral shares in value added and employment; labour productivity by sector.
	UNSD National Accounts Main Aggregates Database	GDP by type of expenditure and GVA by kind of activity (7 sectors). Very wide country/region coverage over long time period (since 1970).	Extent of estimated data not apparent.	Sectoral shares in value added (and, in conjunction with ILO WESO data, labour productivity by sector).
	ILO Global Employment Trends data	Wide country coverage since 1991. Includes formal and informal levels of employment (total and by sex and broad sector).	Sectoral breakdowns cover agriculture, industry and services only.	Employment by sex and broad sector (and, in conjunction with WDI data, labour productivity by broad sector)
	ILO World Employment and Social Outlook supporting datasets	Wide country coverage since 1991. Detailed sectoral (14 sectors) and occupational employment shares, total and by sex.	Extent of estimated data not apparent.	Employment by sex and sector (and, in conjunction with UNSD data, labour productivity by sector)
	UNIDO Manufacturing database	Level of detail for manufacturing sub-sectors; does not include informal sector.	Manufacturing only; includes formal sector.	Sectoral shares in value added and employment; labour productivity by sector.
	Demographic and Health surveys	Micro basis.		Employment/occupation by sector, gender, age.
	Household surveys – LSMS	The DHS data include spatial information, and hence it is possible, in principle, to examine transformation by geographic location within a country.	There are no wages, so it is difficult to discuss issues related to productivity.	

Area of transformation	Data sources	Pros	Cons	Uses
Trade				
(basic trade data)	UN Comtrade	Detailed data provided by over 200 reporting countries, many since 1962.	Max. HS 6-digit level; some DFID focus countries are not reporters.	Analysis of top export/import products and markets/suppliers; changes over time; number of products traded and number of markets/suppliers over time; revealed comparative advantage.
	ITC Trade Map	Detailed data (from UN Comtrade in most cases), available at national tariff line level for many countries; mirror data compiled for non-reporting countries.	Covers 2001 onwards only.	Analysis of top export/import products and markets/suppliers; changes over time; number of products traded and number of markets/suppliers over time.
	Atlas of Economic Complexity	Trade data and complexity visualisations for c. 125 countries over time.		Export/import visualisations.
	Observatory of Economic Complexity	Trade data and complexity visualisations for c. 145 countries over time.		Export/import visualisations.
(trade diversification)	*DFID–IMF Diversification Toolkit: Export Diversification and Quality Databases.*			Indices of export quality and diversification.
(trade in value added)	World Input-Output Model	High quality data.	Few developing countries; 40 countries included.	Contribution of domestic value addition in gross exports.
	Global Trade and Analysis Project (GTAP)	Input-output tables for 129 countries and 57 industries.	Not an official dataset; basis not fully clear; limited years over time.	Trade analysis.
	EORA www.worldmrio.com	Disaggregated into 189 countries, historical time series over 1990-2011. It includes tables of basic prices, as well as two margins (taxes on products and subsidies on products).	Eora's MRIO tables were modelled based on existing sources – national accounts data, Comtrade import and export data, among others – when national input-output or supply-use tables were unavailable. Other limitations go beyond just Eora and affect all MRIO tables; e.g. they are unable to accurately assess the services. They are subject to two	Contribution of domestic value addition in gross exports (used by UNCTAD, World Bank and other users).

Area of transformation	Data sources	Pros	Cons	Uses
			assumptions: all products (for export and domestic use) have the same import content (proportionality assumption) and they assume uniform use of inputs among all firms in sector.	
Productivity at firm level				
	WB enterprise survey data	Number of countries, comparability across countries.	Number of firms covered in each country can be low.	Total factor productivity level; dispersion across sectors.
Wages by occupation				
	ILO–OWW database	Historical data on wages.	Data availability in recent years. Measurement issues and variability across time.	

5. CONCLUSIONS

This initiative aims to fill a gap in the availability of data on ET at country level. We have taken initial steps to provide data for 30-40 variables (Table 1) in 28 countries (of direct interest in DFID, see Table 2), specifically in the following areas:

- economic structures, value addition, employment, labour productivity by sector
- trade, trade diversification and domestic value addition in exports in aggregate and by sector
- firm-level TFP levels and dispersion within sector
- wage levels by occupation.

There are likely to be many gaps, which we would like to explore. What further datasets and analyses can we use? What is of most use for policy-makers?

The emphasis of this paper and the data was on country-level data. Often, a policy-maker or researcher is also interested in comparisons[14] across countries (e.g. in the same sector). The following analytical extensions could be considered:

- comparing and explaining the DVA in gross exports in a sector across countries
- comparing and explaining average firm-level TFP in a sector across countries and sectors
- comparing and explaining sector-level productivity across countries and sectors
- scoring all countries according to some specified benchmark.

14 Of course, a range of further issues need to be discussed when comparing across countries and over time, such as the rebasings of the national accounts or the use of purchasing power parity. See e.g. Penn world table for PPP factors http://www.rug.nl/research/ggdc/data/penn-world-table Penn world table 8.0

KEY REFERENCES

AfDB/OECD/UNDP (2014) 'African Economic Outlook 2014: Global Value Chains and Africa's Industrialisation'. Tunis: African Development Bank.

Ahmad, N. (2013) 'Estimating trade in value-added: why and how?' in Elms, D. and Low, P. (eds.) *Global Value Chains in a Changing World*. Geneva: WTO.

Autor, D.H., Levy, F. and Murnane, R. (2003). 'The skill content of recent technological change: An empirical exploration', *Quarterly Journal of Economics* 118(4): 1279–1333.

Hummels, D., Ishii, J., & Yi, K. M. (2001). 'The nature and growth of vertical specialization in world trade', *Journal of International Economics* 54(1): 75-96.

Jaimovich, N. and Siu, H. (2012). 'The trend is the cycle: Job polarization and jobless recoveries', *NBER Working Paper* No. 18334. Cambridge, MA: National Bureau of Economic Research.

Koopman, R., Wang, Z., & Wei, S. J. (2012). 'Tracing value-added and double counting in gross exports (No. w18579)'. National Bureau of Economic Research.

Lenzen M., Kanemoto, K., Moran, D. and Geschke, A. (2012) 'Mapping the structure of the world economy', *Environmental Science & Technology* 46(15): 8374-8381.

Lenzen, M., Moran, D., Kanemoto, K. and Geschke, A. (2013) 'Building Eora: A Global Multi-regional Input-Output Database at High Country and Sector Resolution', *Economic Systems Research* 25(1): 20-49.

OECD (2012) 'Trade in Value-Added: Concepts, Methodologies and Challenges'.

M.P. Timmer, G.J. de Vries, and K. de Vries (2014). 'Patterns of Structural Change in Developing Countries'. GGDC research memorandum 149.

Tukker, A. and Dietzenbacher, E. (2013) 'Global multiregional input-output frameworks: an introduction and outlook', *Economic Systems Research* 25(1): 1-19.

UNCTAD (2013) *World Investment Report 2013.* Geneva: United Nations Conference on Trade and Development.