C20

FAO's new macro-economic statistics: Agricultural Capital Stock and Agro-Industry Measurement

M. Vander Donckt | FAO, ESS Division | Rome | Italy
S. Dubey | FAO, ESS Division | Rome | Italy

DOI: 10.1481/icasVII.2016.c20d

ABSTRACT

This paper reports on recent work in FAO's Statistics Division (ESS) in developing two new global agricultural statistics databases: an analytical database that estimates agricultural capital stock, and a database on agro-industry measurement (AIM). It describes the data sources, methodology and content of these databases, which provide key national accounts type indicators.The overall approach relies on the use of internationally accepted methodologies and the harvesting of data compiled by multilateral agencies, to minimize burden, duplication of efforts and inconsistencies. Priority is placed, where possible, on the use of official country data reported to these agencies. The paper also presents some key results from our cross-country analysis performed based on the databases. It is set to illustrate how the new databases can help improving our understanding of how countries can benefit from the links between agriculture and agro-industrial development and how agriculture and agro-industry can contribute to economic development. The objective of this paper is to solicit feedbackon the data sources, methodologies, indicators, and overall approach in order to enable improvements and establish key next steps.

Keywords: Agro-Industry Measurement, Capital Stock, Gross Fixed Capital Formation.

PAPER

1. Introduction

As economies grow and develop, the productive landscape undergoes structural transformations, sectors become more interrelated, and determining the importance of any specific sector becomes difficult. This is certainly the case for agriculturewhere agricultural production understates the sector's overall contribution tothe economy through its links to numerous industries such as fertilizer production, food processing and manufacturing, transportation, wholesale, and retail distribution.This raises the questions about the relative roles of these sectors as an economy develops, the impact of agriculture on the entire value-chainand the drivers behind time trends and country-level differences.

To answer these questions and to meet the growing need for consistent statistics to measure the agriculture value-chain, FAO s Statistics Division (ESS) began construction of a new FAO "Macro-Statistics domain" that covers two main areas: the "Agro-Industry Measurement" (AIM) and the "Agriculture Capital Stock" (CS) database. The AIM helps determine agriculture and food-processing industry contributions acrosscountries and time. The CS database aims at providing analytical statistics on capital stock and its components in order to inform questions about investment and productivity.

These databases use the System National Accounts (SNA) framework to ensure data comparability over time and across countries. To reduce burden and duplication, they integrate information from existing international databases from the UN Statistics Division (UNSD), the OECD, the World Input Output Database (WIOD) consortium, and UNIDO. These databases have the advantage of covering a large set of countries, and their use minimizes burden, duplication and effort, while maximizing efficiency and coherence.

The AIM and CS are analytical databases providing provisional indicators. While official country data is the backbone of these indicators, the databases require a significant number of imputations, estimations and assumptions. Data sources and assumptions are well documented, in the hopes that official country statisticians and other experts will help validate or improve the databases and the underlying assumptions, and where possible, provide official country data to replace missing data.

To provide some additional insight to the importance of such data, the rest of this section describes some key trends in the GDP contribution of the agricultural sector and the food processing industry. This rest of this paper presents a technical roadmap to the development of the AIM and CS databases. Section 2 proposes an in-depth description of the database content and the overall approach to database construction. Section 3 discusses the main data and methodological challenges and how they are addressed. Section 4 presents analysis and key findings in order to illustrate some of the potentialities of the database. Section 5 concludes.

2.Database Content and Overall Approach

2.1.Data Coverage

The basic set of variables covered in the AIM and CS databases includeValue-added, Gross Output, Gross Fixed Capital Formation (GFCF), and Net and Gross Capital Stock. These variables are provided in both local currency units (LCUs) and US dollars; current and constant prices for capital stock variables. The built-in indicators include productivity-related indicators, such as the AIR (agriculture GFCF to agriculture value-added); value added shares, such as the Food, beverage and tobacco (FBT) value-added share of manufacturing and of GDP; and the Agriculture Orientation Index (AOI) for capital formation, which normalizes agriculture's physical investmentratio by that of the total economy. Capital related variable are currently limited to the agriculture sector in the current version of the database.

The sector's and industries cover the following:

- Agriculture, forestry and fishing (ISIC Rev. 3 divisions 01 to 05);
- Food-processing activities covered by ISIC Rev. 3 divisions 15 and 16, which includes the FBT industry sector.

The current geographical coverage includes over 200 countries and territories, though for the food-processing variables, indicators have been constructed for a reduced sample of 46 countries captured in the OECD and WIOD databases. Subject to resources, the intention is to expand geographic coverage of the food processing sector, subject to data availability, and to expand variable coverage to include among others trade variables. At update, the targeted time span for data coverage is 1970 to year-2.

2.2. Overall Approach to Database Construction

The methodology in compiling both databases followed similar steps. A first step was to identify and bridge data across sources. The second was to estimate missing values for countries that had reported at least some official data, using established statistical estimation and imputation methods. A third step was to calculate final indicators, which included current price versions in both local currency units (LCUs) and US dollars; constant 2005 price indicators in LCUs and USD for capital stock variables; and the above-mentioned set of built-in indicators.

The original data sources used in compiling AIM and CS were as follows:

- United Nations Statistics Division, National Accounts Estimates of main aggregates
- (NAE), and official country tables;
- OECD, STAN and National Accounts databases;
- WIOD, Socioeconomic accounts;
- World KLEMS; and
- UNIDO, INDSTAT database.

Section 3. Data and methodological challenges

3.1. The Agriculture Value Chain and the Input-Output Model

The original motivation for the AIM project was to develop a database tomeasure the farm-to-fork contribution of the agricultural sector. In its widest concept, the agriculture value chain would coverall economic operations involved in the production and distribution of products that originate from or are used in the production of agriculture output. The first challenge was to define a framework for measuring the agriculture value chain. TheInput-Output (IO) model,developed by W. Leontief in the 1930s, was acknowledged as one of the most powerful tools to measure a value-chain as well as a sector's full contribution to an economy through its inter-sectoral linkages. In particular, the IO model shows how output of one sector is used as an input into another sector, showing sectoral interdependenceby identifying buyer of outputs and supplier of inputs. Due to the resource and time intensity of this approach, limited IO data available for developing countries, and lack of timeliness in the IO data that is available, this approach was discarded in favour of the approach used: to integrate and bridge existing national accounts data/indicators across multiple data sources. However, the absence of an IO framework is also creating challenges in measuring the agriculture value-chain, and determining where to start.

3.2. Agro-Industry, Definition Issue

Because the agriculture value chain includes all the suppliers of inputs to agriculture, and all the users of its outputs, direct and indirect, the project needed to first measure a narrow portion of this value-chain. FAO value chain experts recommended the project begin by focusing on the Agro-industry, starting with the food-processing industry. This was recommended both due to constraints on data availability, and due to the absence of international statistical definitions that covered the breadth of the "agro-industry."

An important future step for FAO will be to help define "agro-industry," a request frequently received by both FAO and theUN Statistics Division (UNSD).

3.3. Bridging of Series for AIM and CS Databases

In the actual compilation of data, an important first important database construction stepis to identify and bridge data acrosssources and across varying methodological and classification versions in order toensure time-series continuity and comparability. For both databases, series are rescaled whenever necessary to match the ISIC Rev. 3.1 classification of economic activities and the 1993 version of the System of National Accounts (SNA).Continuity of long time trends is particularly important in constructing indicators such as GFCF, the basis to our capital stock estimations. In bridging data, the first step is to bridge within an input data source, such as between the SNA and ISIC revisionswithin the UNSD National Accounts Official Country Data.The second step is to bridge across data sources.

3.4. Agricultural CS: New Methodology based on the Perpetual Inventory Method

In developing a new agricultural CS database, two approaches were considered. The first is to estimate capital stock using the physical inventory approach, which adds up the sector s components of produced assets. Physical investments in capital, or capital formation, can then be calculated as changes in capital stock between two time periods. In the case of agricultural capital stock, produced assets include land development, machinery and equipment, farm structures, livestock, and orchards. This approach, which had been previously used in a FAOSTAT agriculture capital stock series, was evaluated and abandoned due to several factors:

- Data quality issues in the underlying datasets: This arose in part from low and declining response rates to the FAO questionnaire sent to countries, and low rates of questionnaire completeness. This problems also resulted in limited country coverage.

- Methodological issues in component calculations: A large number of assumptions were required to calculate capital stock components, such as estimating land improvements from land value.

- Exclusive focus on the narrow agriculture sector: the approach excluded forestry and fisheries, making it more difficult to benchmark the capital stock estimates even against those published by some countries in their national accounts, covering the broader "agriculture" sector.

The second approach, adopted by FAO, is to use the internationally accepted perpetual inventory method (PIMs). In this approach, capitalstock in one period is the sum of capital stock in the previous period, plus the current periodflow of capital investments (GFCF) minus the consumption of fixed capital. The central equation to compile net capital stock (K) is

$$K_t \;=\; K_{t-1} \;+\; GFCF_t - \delta * \left(\frac{GFCF_t}{2}\right) - \delta * K_{t-1} \;+\; X_t, \qquad (1)$$

where $K_t$is the net capital stock at the end of year t;$GFCF_t$ is gross fixed capital formation in year t; the term $\delta*(GFCF_t/2+K_{t-1})$ measures the consumption of fixed capital in year t, or total depreciation;and X_t is the "other changes in volumes" of the group of assets and it is set equal to zero[1].If series on GFCF are available, this method requires assumptions about the initial stock of capital in $t_0$as well as the depreciation rate.

Depreciation rates, δ.

The first step towards computing the net stock above is to select a rate of consumption of fixed capital, δ, also known as the depreciation rate. In the absence of good information about rates of depreciation, δ can be set in reference to other countries' depreciation rates of similar types of assets or other countries' service lives of similar types of assets.

Following Hulten and Wykoff (1996), we apply the declining balance model that converts an average service life of a group of assets, TA, into a depreciation rate, δ. They propose a two-step procedure based on the 'declining balance' formula:

$$\delta = r/T^A, \qquad (2)$$

where parameter$r$ is an estimated declining-balance rate.

In order to fix the depreciation rate and assuming the declining balance model of depreciation in the FAO capital stock database, we need thus to make assumptions on the value of the declining balance parameter, r, and on the average service life of agriculture, T^A.

[1] Other flows are changes in the value of assets and liabilities that do not result from transactions. Examples are losses due to natural disasters and the effect of price changes on the value of assets and liabilities. For more, see SNA2008, paragraph 3.99.

Our review of practices by national statistical offices that apply the declining balance model tend to support a value of r greater than one. In the FAO capital stock database, we assume a unique declining balance parameter common to all countries for which agriculture capital stock has to be estimated that we arbitrarily fix to 1.5.

As for the average service life of assets used in agriculture, it is clear that service lives differ depending on the type of assets considered with structures having the longest average service lives. The average service lives varies also considerably among countries.

We arbitrarily fix T^A to be equal to 25 years. Combined with a declining balance parameter value of 1.5, the derived depreciation rate used in the FAO capital stock database is δ = 1.5/25 = 0.06.

This depreciation rate is consistent with implicit average depreciation rates retrieved using country series on NCS and GFCF for select countries.

Initial capital Stock.

Once a depreciation rate is selected, the next step is to estimate a starting stock of capital. According to Kholi (1982), a simple approximation can be used when geometric age-price profiles apply for thenetstock at the beginning of the benchmark year *t0*:

$$K_{t0} \approx [GFCF_{t0-1} + (1-\delta)\,GFCF_{t0-2} + (1-\delta)^2 GFCF_{t0-3} + \ \ldots] \qquad (3)$$

Assuming that the long-run growth in the volume investment is equal to the long-run growth rate of volume agriculture value-added, θ: $GFCF_t = GFCF_{t-1} * (1+\theta)$; and inserting this relation into the equation above, the initial capital stock at the beginning of period *t0* is approximated by :

$$K_{t0}(geometric) \approx \frac{GFCF_{t0}}{(\delta+\theta)} \qquad (4)$$

3.5. Estimating the GFCF Series

Calculating capital stock in periods after t0requiresa relatively long and bridged time series of national accounts variables, particularly on GFCF for agriculture ($GFCF_{AFF}$), but also on auxiliary variables used to estimate missing GFCF variables. Imputation and estimation of missing valuesused a combination of hot deckimputation methods (carrying forward data from previous years for a country within a data series), cold deckimputation(using data from a “nearest neighbour”, in terms of structure and level of economicdevelopment, measured by GDP/capital), and linear interpolation.

The Agriculture Investment Ratio (AIR) was used to construct the $GFCF_{AFF}$ series in the FAO Capital Stock database, where $AIR = GFCF_{AFF}/VA_{AFF}$. The following two-steps procedure was used:

- Estimate the AIR in time *t* for country*i*, through regression analysis;
- Compile the level of investment in agriculture by applying the AIR to value-added:

$$GFCF_{AFFi,t} = AIR_{i,t}^{estimated} * VA_{AFFi,t}. \qquad (5)$$

Two approaches were used to compile AIR series: the first for countries with fully missing AIR data; the second for countries with some but incomplete AIR series.

3.5.1. Estimation of AIR for countries with fully missing series.

The estimation of the AIR for countries with fully missing series is based on panel data analysis, derived from an input datasetcovering 86 countries thathave AIR observations. Data in the input dataset model and estimate the AIR, our dependent variable, based on one or more independent variables. The fitted model infers an AIR series for those countries with missing agriculture investment data.

The approach follows the followingsteps.

1) **Model pre- selection**: Potential predictors considered for inclusion in the model were: log of GDP per capita; log of real GDP (measured in 2005 USD); the trade openness index in agriculture; the total economy investment ratio; the annual GDP growth rate; total population; and VA in agriculture as a

share of GDP. The list of explanatory variables was determined, in part, by data availability. For instance, employment in agriculture and VAAFF over employment in agriculture would have been interesting variables to include and to test for in the regression analysis, however due to lack of consistent data, these variables haven't been included in the analysis so far.

Countries were divided into two sets: low income countries and middle- and high-income countries (separation criterion set at 10,000 USD of GDP per capita, 2011 value).The best models for each set wereselected based on stepwiseregression and on the leaps-and-bounds algorithm (Furnival and Wilson, 1974).

2) **Regression method**: Once the benchmark model was chosen, different regression methods were assessed: pooled OLS, panel fixed-effects (FE); and panel random-effects (RE). The FE method was selected based on the following results:

a. OLS vs. RE: Breush-Pagan Lagrange multiplier. The null hypothesis of no panel effect (no significant difference across units) was rejected, and the RE model selected.

b. FE vs. RE (and OLS): Hausman test. Reject the null hypothesis of a RE. By transitivity, the OLS specification is also rejected (it can be shown that the RE is equivalent to an OLS generalized for panel effects).

3) **Sensitivity analysis and MSE based evaluation**: We next perform an in-sample Mean Square Error (MSE) analysis of the AIR estimated under our alternative model specifications and we select the final benchmark model that will be used for the underlying estimation of GFCFAFF for countries with fully missing series based on MSE and the by-country Symmetric Relative Efficiency measure, following Tratar and Toamn (2014), which is defined as:

$$SREM_{model\ 1/model\ 2} = \begin{cases} 1 - \frac{MSE_{model\ 1}}{MSE_{model\ 2}} & for\ MSE_{model\ 1} < MSE_{model\ 2} \\ \frac{MSE_{model\ 2}}{MSE_{model\ 1}} - 1 & for\ MSE_{model\ 1} \geq MSE_{model\ 2} \end{cases} \quad (6)$$

3.5.2. Imputation of AIR for countries with incomplete series

In this sub-section we discuss alternative techniques to fill incomplete time series on AIR. While, the targeted time coverage for publication in the capital stock database is 1970-2014, most countries present incomplete time series with missing data for the most recent years that have be imputed. The length of "missingness" varies across countries from just a couple of years to over two decades.

Approach 1: Constant ratio

The first approach imputes a missing AIR as an average of the last three to five years of available AIRs. This technique is naïve in the sense that it does not identify and, therefore, fails to project, any trend or cyclical component embedded in the time series.

Approach 2: Exponential Smoothing

Forecasts produced by exponential smoothing models are exponentially weighted moving average of past observations. As such, these models extract information endogenous to the historical time series regarding its four components – level, trend, seasonality, and noise – and use it to produce forecasts.

In the second approach, the ETS package in R (Hyndman et al., 2008) was used to forecast missing values. The ETS statistical framework allows selection of the best exponential smoothing model based on the AIC information criteria. In total, 30 model combinations are possible in the ETS framework depending on the type of trend and seasonal components considered (e.g. additive/multiplicative/damped additive/damped multiplicative trend, etc.) as well as on the underlying statistical model (additive vs. multiplicative errors).

This approach is better able to capture trend and/or cyclical components contained in the historical profile of the series. However, the information in the series is insufficient to address the structural break in countries' agricultural investment profiles that originated from the financial crisis.

In the next two sections we present the results obtained from applying time series regression techniques that allow for the inclusion of exogenous variables: the ARIMAX and, its generalization, the autoregressive distributed lag model (ARDL). These methods are preferred as they make use of both the information embedded in the historical series of the indicator under consideration and the knowledge of relationships between the indicator and external factors so that special events as the crisis can be account for.

Approach 3: ARIMAX

In the third approach, we fit ARIMA class models with exogenous variables before using the best model for forecasting purposes. The autoregressive integrated moving average model including exogenous covariates, ARIMAX(p,d,q), extends the

ARIMA(p,d,q) model by including the linear effect that exogenous series have on the stationary response series y_t. The general form of the ARIMAX(p,d,q) model is

$$\phi(L)y_t = c^* + \ x_t{}'\beta + \ \theta^*(L)\varepsilon_t \tag{7}$$

where$c^* = c/(1-L)^d$ and $\theta(L)^* = \theta(L)/(1-L)^d$. Vector x_t'holds the values of the r exogenous, time-varying predictors at time t, with coefficients denoted β. We use this model to check if a selection of exogenous variables has an effect on a linear time series.

It should be noted that ARIMAX models have the same stationarity requirements as ARIMA models. If the response series y_t is not stable, we need to difference it to form a stationary ARIMA model with d degrees of integration. In addition, it is also required to assess in the pre-estimation phase whether the predictor series x_t are stationary and to difference them if necessary. If $x_t$is nonstationary, then a test for the significance of β can produce a false negative. As usual, the practical interpretation of β changes if you take a transformation (difference) the predictor series.

For each country, we perform a best model fitting exercise looping through a maximum of 144 models (8 combinations of explanatory variables times 18 possible combinations of ARIMA(p,d,q) specifications). The preferred model is selected based on the BIC minimization criterion, we then backcast/forecast the country series using preferred model.

Approach 4: Auto-regressive distributed lag model (ARDL)

The fourth approach tested relies on the autoregressive-distributed lag (ARDL) model, which takes the following general form:

$$y_t = \ \alpha + x_{i.t-q}^{'}\beta_i + \ y_{t-p}^{'}\gamma + \ \varepsilon_t, \tag{8}$$

whereE_t is a random disturbance term that is assumed to be serially independent. The model is autoregressive in the sense that it contains one or more lagged values of the dependent variable, yt..The model also has a distributed lag component, in the form of successive lags of the explanatory variables, xt. A model containing p lags of y and q lags of x is denoted ARDL(p,q).

Due to too short a time series and because of multicollinearity in the data, the ARDL approach fails in most cases. We therefore prefer the ARIMAX based approach.

3.5. Food-processing industry.

The AIM database uses national accounts (NA) variables for the FBT when available, as for many OECD countries, and the INDSTAT2 database from UNIDO, otherwise. Since INDSTAT2 data are based on national industry surveys, which do not correspond to NA data for coverage and conceptual reasons, the INDSTAT2 series is re-scaled and estimated to national accounts levels. Re-scaling is performed using composition ratios, such and manufacturing to total economy value-added.

Since indicators relating to the food-processing industry are jointly compiled with UNIDO, and UNIDO provides a complete series, see Boudt et. al. for the methods used to impute missing data.

4. Analysis and findings

In this section, we present and comment a few results extracted from the new FAO macro-statistics database. We first look at the relative contribution of agriculture and food processing.Looking at select countries, Figure 5 shows that agriculture as a share of GDP tended to decline between 1970 and 2013, while food processing remained relatively stable.

Figure 6 presents the value-added share of food and beverages processing in total manufacturing against the contribution of manufacturing to total GDP, and shows the resilience of the FBT sector even during crises such as the 2009 Financial Crisis. This can be interpreted simply by the fact that the population still needs to eat, that is there is a low income elasticity of demand for food and beverage products.

Figure 1 - Contribution of Agriculture and Food-processing to GDP.

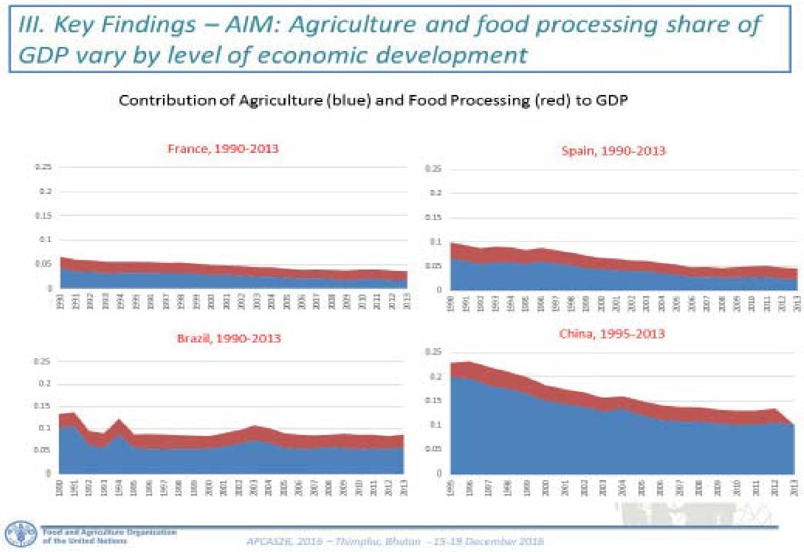
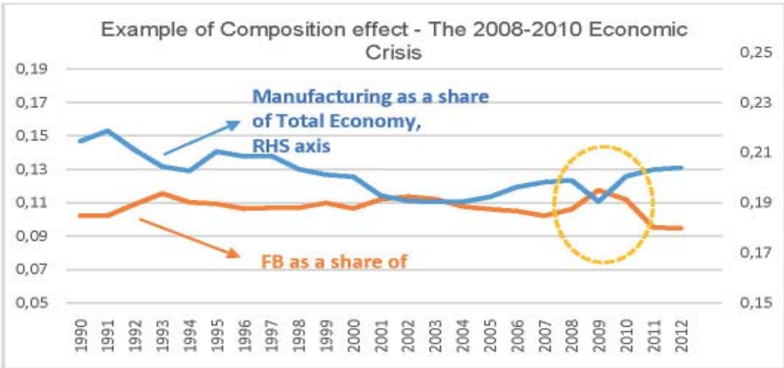


Figure 2 - Resilience of the Food-Processing sector

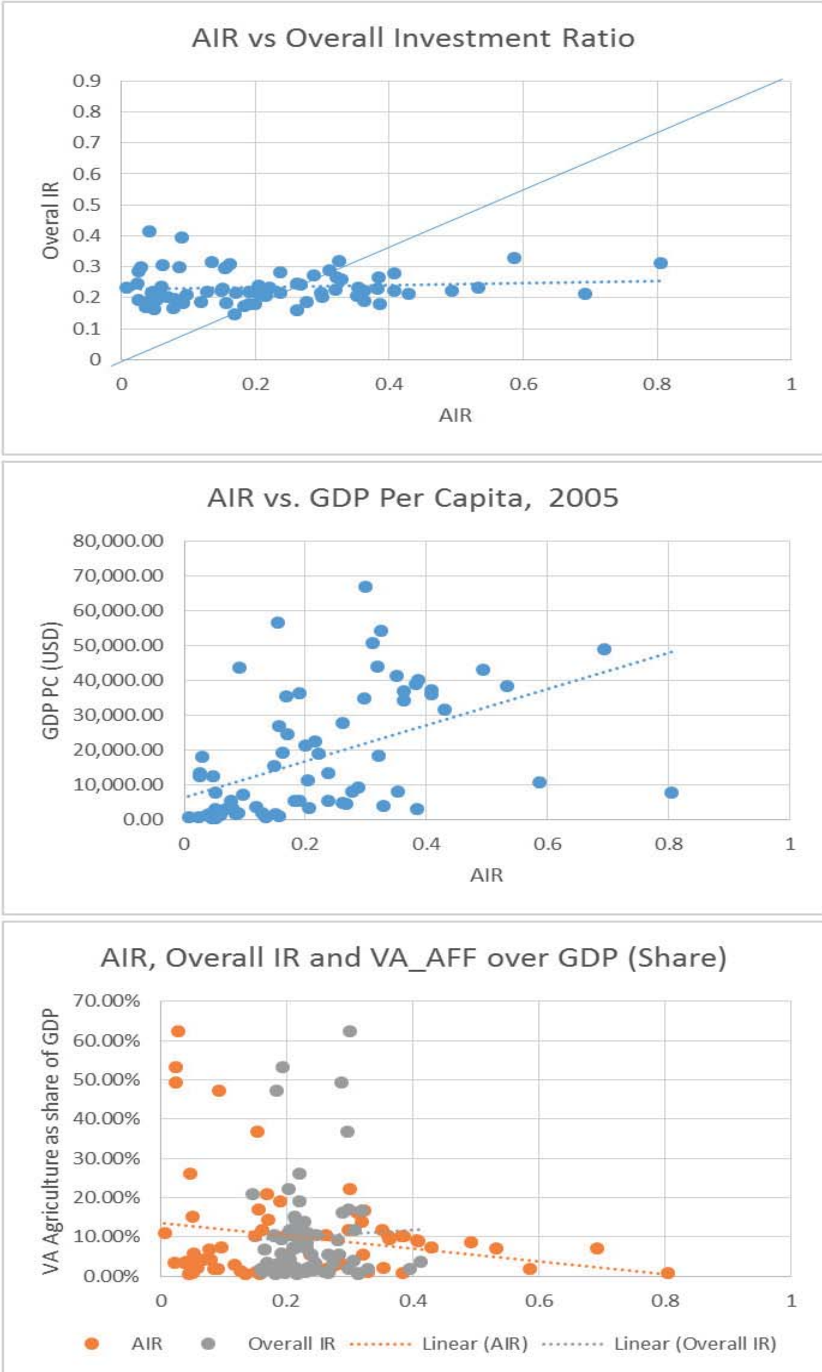


Another possible line of analysis bears on capital stock related variables in the agriculture sector.[2] The following set of graphs present the AIR as defined in Section 3 against a set of economic variables.

The upper part of Figure 8 indicates that there is no clear positive relationship between the overall investment ratio (calculated for the total economy) and the AIR. It also shows a larger variability of the AIR relative to the overall economy investment ratio. When presenting the AIR against GDP per capita, we instead find a clear positive relationship; and when presenting against the agriculture share of GDP, we find a negative relationship (i.e. countries with a high agriculture share of GDP tend to have a low AIR).

[2] The insertion of capital stock variable for the food processing sector will constitute an extension of the current AIM database.

Figure 3 - AIR against overall investment ratio, agriculture VA and GDP (2005 data)



An interesting pattern (Table 1) emerges between the AIR and the overall investment ratio when we divide the countries into two groups, low-income countries versus middle- and high-income countries. While countries in both groups present a similar average overall investment ratio, the average AIR is much lower in low-income countries, indicating that in those countries – where agriculture often remains an important contributor to GDP – the primary sector is behind in terms of investment in physical capital with respect to the other sectors of the economy. On the other hand, industrialized countries tend to have much more mechanized agriculture sector.

Table 1 - Average AIR: Total economy vs. Agriculture

(all years)	All countries	Low income countries	Middle- and high-income countries
Avg. AIR	0.214	0.109	0.255
Avg. Overall IR	0.231	0.222	0.235

5. Conclusion

The aim of this paper is to present a new macro-statistics database developed by the statistics division FAO, in order to meet the growing need for consistent statistics measuring agricultural capital stock and the agriculture value-chain.

While largely based on secondary data, the database construction implies a large number of data challenges to be overcome, from integrating series based on different SNA versions and classification systems, to imputing missing values, to making assumptions about depreciation rates. This leaves wide scope for further improvements, as country experts provide better information and replace FAO estimates with their own, as well as extensions of the database in terms of geographical, variable and subsector coverage.

It is hoped that presenting the work and making available the databases in 2016 will facilitate the type of feedback that will lead to an even more useful evidence-base for agriculture and food policies.

References

[1] Boudt, Kris, Valentin Todorov, and Shyam Upadhyaya (2000). Nowcasting Manufacturing Value Added for Cross-Country Comparison.Statistical Journal of the IAOS, Vol. 26, Nos. 1-2, pp. 15-20, 2009.

[2] Claessens, Stijn, Hui Tong and Shang-Jin Wei (2011). From the Financial Crisis to the Real Economy: Using Firm-Level Data to Identify Transmission Channels.NBER Working Paper 17360, August 2011.

[3] Furnival, G. M., and R. W. Wilson (1974). Regression by leaps and bounds. Technometrics 16: 499–511.

[4] Gros, Daniel and Cinzia Alcidi (2010). The Crisis and the Real Economy. Intereconomics.

[5] Hulten, Charles R. and Frank C. Wykoff (1996). "Issues in the Measurement of Economic Depreciation," Economic Inquiry, Vol. XXXIV, No. 1 (January): 10-23.

[6] Kolhi, Ulrich (1982); "Production Theory, Technological Change, and the Demand for Imports:

Switzerland 1948-1974"; European Economic Review 18, pp. 369-86.

[7] Ksantini, Majdi, and YounesBoujelbène (2014).Impact of Financial Crises on Growth and Investment: An Analysis of Panel Data.Journal of International and Global Economic Studies, 7(1), June 2014, 32-57

[8] Lindsey, Charles and Simon Sheather (2010). Variable selection in linear regression.The Stata Journal (2010), 10:4, pp. 650–669

[9] TratarLiljana, Blaž Mojškerc, and Aleš Toman (2016), "Demand forecasting with four-parameter exponential smoothing", unpublished manuscript.

[10] Williams, Richard (2015), Lecture Notes on Multicollinearity, University of Notre Dame.