Food and Agriculture Organization of the United Nations

FAO Statistics Working Paper Series / 19-16

THE NEW FAO GLOBAL DATABASE ON AGRICULTURE INVESTMENT AND CAPITAL STOCK

Working Paper Series

ESS / 19-16

THE NEW FAO GLOBAL DATABASE ON AGRICULTURE INVESTMENT AND CAPITAL STOCK

Marie Vander Donckt and Philip Chan

Food and Agriculture Organization of the United Nations
Rome, 2019

Vander Donckt, M. and Chan, P. 2019. *The new FAO global database on agriculture investment and capital stock.* FAO Statistics Working Paper 19-16. Rome.

ISBN 978-92-5-5-131819-5

Contents

Tables

Figures

Abstract

In this paper, we present the new FAO analytical database on aggregate physical investment flows and capital stock in agriculture, forestry and fishing for 206 countries and territories from 1990 to 2015. We describe the database content as well as the data sources, the methodology used to deal with missing data, and the measurement issues underlying its development. Building on previous research programmes held at the World Bank and at the FAO, we compile long time series of the agricultural investment to value added ratio, which we employ to compute agricultural investment flows. These latter flows are then converted into agricultural capital stock series, by applying a variant of the perpetual inventory method. We improve upon previous research by departing from strictly time series based imputation techniques and allowing for the presence of exogenous regressors in the model. We also extend the database to countries with fully missing data.

Acknowledgements

The paper was drafted by Marie Vander Donckt, of the FAO Statistics Division, and by Philip Chan, of the OECD. Part of this work was done while Philip Chan was a consultant at the FAO Statistics Division. The authors would like to thank Jan Karlsson and Rachele Brivio, who started the "Agricultural Capital Stock" database project. This work would not have been possible without their contribution. The authors are grateful to participants at the Seventh International Conference on Agricultural Statistics (ICAS VII, Rome, October 2016) and at the FAO brown bag seminar, as well as Piero Conforti and Sangita Dubey for very useful comments and discussion. They also wish to thank Silvio Daidone for sharing his dataset on agricultural investment and capital stock. The views expressed in this study are those of the authors and do not necessarily represent those of the Institutions with which they are affiliated.

1 Introduction

Physical capital represents a fundamental input in the production process. Empirical growth models require measures of physical capital to gauge the contribution of this factor of production to real income dynamics. Therefore, estimates of physical capital stock are necessary to gain a deeper understanding of the process of economic growth and its determinants.

Focusing on the agriculture sector, estimates of physical capital and the level of production are crucial, given the role played by this sector in the process of economic development. Notably, productivity growth in agriculture releases resources to other sectors of the economy, providing the basis for successful industrialization. Moreover, the livelihood of rural households builds upon the agricultural sector in many developing countries, with their welfare directly linked to the productivity of the resources at their disposal (Fuglie and Rada, 2013).

Despite the general interest in the estimation of agricultural capital stock, cross-country datasets are still scarce (see, for instance, Larson et al., 2000; Butzer et al., 2010; Daidone and Anríquez, 2011). At the national level, no suitable capital stock data for the agriculture, forestry and fishing industry[1] are available for most countries due to the complex and costly task of compiling such series. At the same time, the lack of internationally comparable capital stock data has been a major obstacle to empirical studies on the contribution of the capital stock in agriculture to the industry's growth and economic development more broadly.

In this paper, we aim to fill this gap by presenting new estimates of the agricultural investment ratio, which are then used to construct a wider database of capital stock in the agriculture sector at the worldwide level, by implementing a variant of the Perpetual Inventory Method (PIM). In this way, we aim at stimulating empirical cross-country research on the role of agriculture in process of international economic development.[2]

The geographical coverage of the Agriculture Capital Stock (ACS) analytical database in-

[1] In the remaining of the text, the term "agriculture" is used to cover the wider ISIC agriculture, forestry and fishing industry.

[2] The dataset is available on the FAOSTAT website at `http://www.fao.org/faostat/en/#data/CS`.

cludes 206 countries and territories. To the best of our knowledge, this is the widest coverage in terms of agricultural GDP at the worldwide level. As for the time dimension, the estimation is carried out for the period 1990–2015. As part of the resulting ACS database, the Food and Agriculture Organization of the United Nations (FAO) publishes country-by-country data on physical investment in agriculture as measured by the System of National Accounts concept of gross fixed capital formation.[3] It should be stressed that the ACS database is an *analytical* database: whenever possible, it integrates official national accounts data harvested from UNSD or OECD. However, if the full set of official data is not available for a specific country, statistical procedures are employed to obtain estimates for the entire time series.

Our purpose in this paper is to present a summary of the research work conducted to deliver the database to the public, explaining how agricultural investment is estimated and how from it the corresponding agricultural capital stock is constructed. Given the highly unbalanced nature of agricultural gross fixed capital formation and capital stock data, we carefully spell out the key methodological issues and, at the same time, describe the imputation techniques we have chosen for dealing with missing data.

Concerning the estimation of the agricultural investment series, we consider three case-scenarios depending on the extent of data missingness, which give rise to different classes of imputation procedures. The first case-scenario concerns countries with a limited number of missing data. In this case, country-specific time series ARIMA models with exogenous variables (ARIMAX) are fitted to the data and missing values replaced. The second case-scenario regards countries for which agricultural gross fixed capital formation data are insufficient to apply any country-specific (univariate) time series techniques. In this case, we exploit the cross-sectional time series dimension of the dataset and we set up unobserved effects panel data models (Wooldridge, 2010) to estimate the missing data. By combining the temporal and cross-sectional dimensions of the dataset, panel data provide a larger amount of information and are particularly useful for imputation purposes when only few observations are available. The third

[3] The terms “investment” and “gross fixed capital formation” are used interchangeably in the paper.

case-scenario relates to countries for which no agricultural capital stock data exist. In this case, we perform pooled OLS estimation of regression models on countries for which some data are available. Then, based on these estimates, the agricultural investment ratio is derived for those countries for which we do not have any agricultural investment data, using the estimated coefficients and the selected covariates.

Lastly, the agricultural gross fixed capital formation series are used to construct new estimates of the capital stock. In order to do so, we apply a variant of the Perpetual Inventory Method, which is a methodology often used by statistical offices to derive estimates of the capital stock data (Berlemann and Wesselhöft, 2014, 2017).[4]

The rest of the paper proceeds as follows. Section 2 reviews the existing literature on the estimation of investment and capital stock in agriculture. Section 3 presents the overall strategy to estimate country-level series on gross fixed capital formation in agriculture, forestry and fishing. To this aim, we introduce the agricultural investment ratio definition and explain how an estimate of this variable can be used to recover an estimate of agricultural investment within the national accounts framework. The section also provides a description of the available input database. Section 4 discusses in more details the three estimation approaches to obtain complete country series on the agriculture investment ratio. Section 5 describes the Perpetual Inventory Method as well as its practical implementation for the derivation of the agricultural capital stock series. Section 6 concludes and suggests directions for further research.

[4] Most attempts, both at national and international level, to construct capital stock data relies on some variant of the PIM approach. Alternative to this methodology is the physical inventory approach, which adds up the sector's components of produced assets. This approach is however very difficult to implement in practice, not the least due to its high cost in terms of data collection. Until 2014, FAO produced estimates of capital stock in agriculture based on the physical inventory approach. However, as discussed in Von Cramon-Taubadel et al. (2009), this approach suffered from a number of weaknesses, including a narrow coverage of fixed assets in farming and the multiplication of data sources and model assumptions needed to produce series on capital stock in agriculture. In addition, the use of constant prices led to a volume index measure that did not account for the age of assets or quality improvements in assets over time. This measure of *gross* capital stock is not suitable for conducting productivity analysis.

2 Related literature

In this section, we briefly discuss the most important approaches used in the literature to estimate agricultural capital stock for a large panel of countries.

One of the first attempts to measure capital in agriculture is by Larson et al. (2000). These authors build a database that includes estimates of three components of agricultural capital (covering fixed capital in agriculture, livestock, and tree stock) as well as capital stock measures for manufacturing and total economy. The dataset covers 62 countries for the years 1967 through 1992. Agricultural capital stock series are estimated based on national accounts investment data, by using a modification of the PIM. In the standard PIM, information on the initial capital stock at the time when the investment time series starts is needed to calculate the current capital stock. Larson et al. (2000) bypass this requirement and generate lengthier investment time series by regressing the logarithm of the investment-to-value added ratio on a time trend, over a sufficiently long sample period. Then, they use the fitted values of the investment-to-value added ratio and multiply them by the value added series to generate the needed missing investment data, in levels.[5]

More recently, Butzer et al. (2010) update the agricultural component of this database from 1970 to 2000 for a subset of 30 countries out of 62. The authors examine the changing composition of agricultural capital, as well as differences in the accumulation of capital across countries. Data suggest that, as economies grow, the composition of agricultural capital changes. Furthermore, the degree of the change differs between high–, middle– and low–income countries.

In an effort to produce a global estimate of the agricultural capital stock consistent with the national accounts framework, FAO has carried on a research programme on the measurement of capital stock and investment in the agriculture sector. In this context, Daidone and Anríquez (2011) construct an extended analytical database of physical investment flows and capital stock in agriculture. These authors build extensively upon Larson et al. (2000) and propose a refined

[5] Additional details are available in the companion paper by Crego et al. (1998), which documents data sources along with the computer programme used to calculate the physical investment and capital stock series.

methodology for the estimation of both agricultural value added and agricultural investment, relying on national accounts data. The resulting dataset features a wider country coverage (80 countries) and a larger time span, from 1970 through 2005.

As in Larson et al. (2000), Daidone and Anríquez (2011) estimate the agricultural capital stock via a variation of the PIM. For those countries for which agricultural investment series are not complete, missing data are imputed by means of a two-step procedure. In the first step, country-specific OLS regressions are estimated on agricultural value added. In the second step, the agricultural investment-to-value added ratio is assumed to be constant, with fluctuations driven by the agricultural value added residuals generated in the first step. Country-level OLS regressions are again estimated and used to recover the agricultural investment flows, by multiplying the fitted investment-to-value added ratio by the official value added series, which is available for most countries. In order to keep some flexibility in the model, structural breaks are allowed in the regressions whenever there are enough degrees of freedom to identify them (i.e., for countries with more than 30 observations, considered as the smallest sample size needed to conduct statistical inference).

Daidone and Anríquez (2011) estimate the agricultural capital stock for a sample of 80 countries. The requirement for being considered in their analysis is the availability of at least some data on both agricultural value added and gross fixed capital formation. The reason is that missing data on gross fixed capital formation in agriculture prevent the calculation of agricultural capital stock. However, 80 countries is hardly representative of the world population. Thus, despite improvements in the estimation of the agricultural investment-to-value added ratio, this geographical coverage bias in their output database warrants a reexamination of their approach. Indeed, Daidone and Anríquez (2011, p. 11) point out that the data only cover “45 percent of the agriculture GDP of middle– and low–income countries”, while representing “95 percent of high-income countries”, and thus advise users of this dataset to account for this selection bias in their analysis.

Our paper improves upon previous contributions by providing estimates of agricultural investment and capital stock on a global scale, for 206 countries and territories covered, over the

1990–2015 time span. In addition, we depart from a strictly time-series imputation approach by also exploiting the panel dimension of the input dataset. In all the estimated models, we account for the influence of exogenous regressors in driving the agricultural investment ratio dynamics. The next section describes in more detail the overall estimation strategy and the data sources.

3 Methodology and input dataset

3.1 Agricultural gross fixed capital formation and capital stock estimation

Our estimation of all capital related variables relies on countries' national accounts series on gross fixed capital formation and capital stock in agriculture. According to the System of National Accounts 2008,[6] capital stock is defined as the value of all fixed assets in use, where fixed assets are described as produced assets (i.e., excluding land) that are used repeatedly in the production process for more than one year. Fixed assets include not only buildings, structures, machinery and equipment but also "cultivated assets yielding repeated products such as animals for breeding, dairy, draught, etc., or perennial tree, crop and plant resources" (SNA 2008).

We do not adjust series on agricultural fixed capital to include livestock and tree stock, departing from Larson et al. (2000), Butzer et al. (2010) and Daidone and Anríquez (2011). In our view, there are several reasons for not adjusting our estimated series for cultivated assets. While we acknowledge that cultivated assets may be imperfectly measured in official country data, there is indication suggesting that National Statistics Offices have developed tools to better measure cultivated assets, with much progress achieved over the last two decades. For instance, all OECD countries currently report series on gross fixed capital formation by asset component with data on cultivated biological resources provided separately.[7] In addition, adjusting series

[6] The System of National Accounts 2008 (SNA 2008) is the latest version of the international statistical standard for the national accounts, adopted by the United Nations Statistical Commission (UNSC). For further details please consult the webpage: `https://unstats.un.org/unsd/nationalaccount/sna2008.asp`.

[7] As such, adjusting series for cultivated assets would require reading through National Statistics Offices related documentation – when available – to find out on a country-by-country basis the exact coverage of the agricultural capital stock and, eventually, the necessary adjustment to be undertaken.

for cultivated assets is extremely demanding in terms of data requirements, as annual data on livestock and surface areas of perennial tree, crop and plant resources are needed. Given that FAO has interrupted the collection of information on agricultural capital inventories used by Larson et al. (2000), Butzer et al. (2010) and Daidone and Anríquez (2011) to correct for cultivated assets, this adjustment is currently hard to accomplish.

In order to construct estimates of capital stock, our methodology rests upon the widely used PIM. Specifically, the net capital stock (NCS) for country i at time t is estimated using the following PIM equation:

$$NCS_{i,t} = (1-\delta_i)NCS_{i,t-1} + (1-\delta_i)GFCF_{i,t}, \tag{1}$$

where $GFCF_{i,t}$ denotes gross fixed capital formation for country i at time t and δ_i is a constant country-specific depreciation rate. After repeated substitution, the current capital stock can be represented as a weighted average of past net investments, with exponentially decreasing weights resulting from our assumption of geometric depreciation at a constant rate. As a result, prior to calculating agricultural capital stock, a necessary requirement is to compile relatively long time series on physical investment in agriculture. A more detailed description of our implementation of the PIM is provided in Section 5.

Missing observations on agricultural investment occur frequently. To handle incomplete data, we apply estimation methods for missing data tailored to the degree of data missingness, which varies widely across countries, from just a couple of data points to fully missing data. All the employed estimation methods rest upon first estimating an auxiliary variable, the Agricultural Investment Ratio (AIR), which we define as:

$$AIR_{i,t} = \frac{GFCF_{AFF_{i,t}}}{VA_{AFF_{i,t}}}, \tag{2}$$

where $GFCF_{AFF}$ denotes gross fixed capital formation in agriculture while VA_{AFF} is value added in the same sector. Therefore, the AIR represents gross fixed capital formation as a share of value added in agriculture, gauging how much of the total factor income is reinvested in new fixed assets.

According to equation (2), $GFCF_{AFF_{i,t}}$ can be expressed as the product of $AIR_{i,t}$ and $VA_{AFF_{i,t}}$. This latter variable is fully available for all countries, whereas $GFCF_{AFF}$ is not. Therefore, in order to gauge $GFCF_{AFF}$, our strategy is to take the product of VA_{AFF} and the estimated value of AIR ($\widehat{AIR}$, obtained via regression analysis).

Specifically, we implement the following two-step procedure:

1. Estimate the AIR for country i at time t through regression analysis, i.e.,

$$\widehat{AIR_{i,t}} = f(x_{1,t}^{(i)}, x_{2,t}^{(i)}, \ldots, x_{k,t}^{(i)}) + \varepsilon_{i,t}, \tag{3}$$

   where the hat denotes estimated values and $x_{k,t}^{(i)}$ contains country-specific observation at time t on the k regressors assumed to affect the outcome variable, while $\varepsilon_{i,t}$ is a random disturbance term.

2. Using (2), obtain the estimated $\widehat{GFCF_{AFF_{i,t}}}$ by applying the relation:

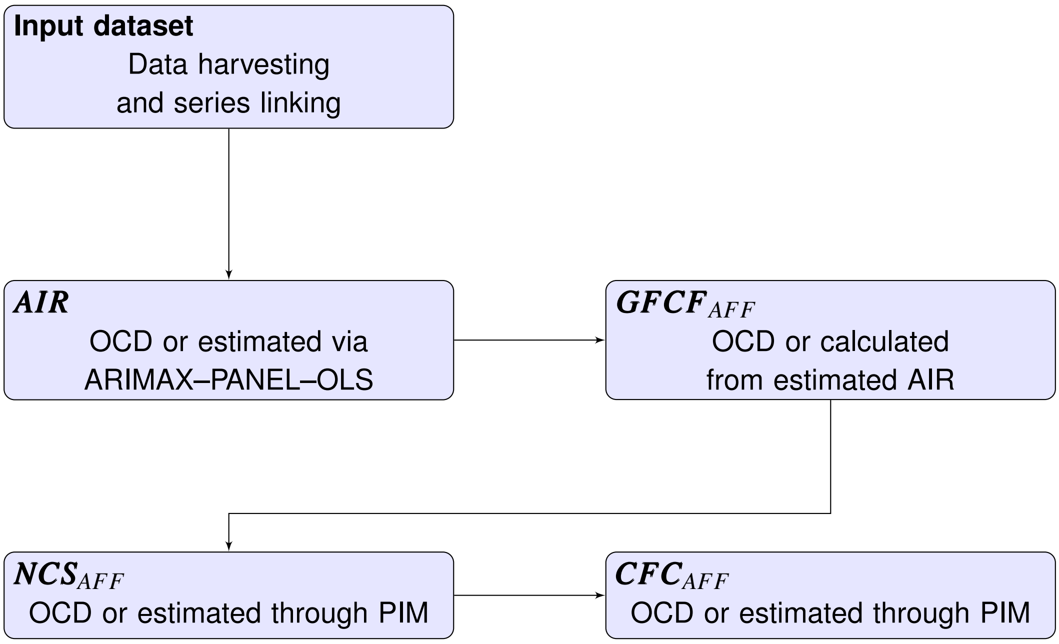
$$\widehat{GFCF_{AFF_{i,t}}} = \widehat{AIR_{i,t}} \times VA_{AFF_{i,t}}, \tag{4}$$

   in which $VA_{AFF_{i,t}}$ is observed.

The complete process of building the FAO ACS analytical database – from data harvesting to calculating NCS and CFC – is illustrated in Figure 1.

The first step requires constructing the input dataset from which we draw to estimate the global agricultural capital stock database. In order to do so, we harvest data from existing international databases such as UNSD Analysis of Main Aggregates (AMA) and OECD National Accounts, before proceeding to series linking to obtain longer time series. This is described in Section 3.2. The AIR is either available from official country data (OCD) or has to be estimated running univariate ARIMAX models, unobserved effects panel data models, or pooled OLS regression models (equation (3)). These estimation methods for missing data will be carefully

Figure 1: FAO ACS analytical database – Complete sequence of data processing



illustrated in Section 4. Then, $GFCF_{AFF}$ is computed as $\widehat{GFCF}_{AFF} = \widehat{AIR} \times VA_{AFF}$ (equation (4)). Once $GFCF_{AFF}$ has been calculated, estimates of the Net Capital Stock in agriculture (NCS_{AFF}) can be derived by relying on the PIM, as detailed in Section 5. Based on the PIM, we also derive Consumption of Fixed Capital (CFC_{AFF}).

3.2 The input dataset

The calculation of AIR requires a host of variables which we compile into an input dataset. Table 1 provides a synthetic view of the database in terms of variable content, source and overall data availability. The input database, which is harvested from existing international databases, has a cross-sectional time series structure with observations on 206 countries and territories over the whole time span from 1970 to 2015 (i.e., maximum of 46 years of observations for each country).

The upper part of Table 1 lists our target variables disseminated in the FAOSTAT agriculture capital stock analytical database: the Agriculture Investment Ratio (AIR), Agriculture Gross Fixed Capital Formation ($GFCF_{AFF}$), Agriculture Net Capital Stock (NCS_{AFF}) and Agriculture Consumption of Fixed Capital (CFC_{AFF}).

Table 1: Input database – Target variables and potential covariates for regression analysis

TARGET VARIABLES	DESCRIPTION	SOURCE	N. COUNTRIES	N. OBSERVATIONS.
AIR	Agriculture Investment Ratio ($GFCF_{AFF}/VA_{AFF}$)	FAO	79	1 962
$GFCF_{AFF}$	Agriculture Gross Fixed Capital Formation	FAO (1)	86	1 875
NCS_{AFF}	Agriculture Net Capital Stock	FAO (1)	38	882
CFC_{AFF}	Agriculture Consumption of Fixed Capital	FAO (1)	95	1 899
COVARIATES	DESCRIPTION	SOURCE	N. COUNTRIES	N. OBSERVATIONS.
VA_{AFF}	Agriculture Value Added	UNSD NAE	206	8 762
GDP	Gross Domestic Product	UNSD NAE	206	8 762
$GDPPC$	GDP per capita	UNSD NAE	206	8 762
$GFCF$	Gross Fixed Capital Formation	UNSD NAE	206	8 762
IR	Investment Ratio ($GFCF/GDP$)	UNSD NAE	206	8 762
POP	Population	UNSD NAE	206	9 407
$EXtoVA_{AFF}$	Export of agricultural products over agriculture value added	FAO, UNCTAD	201	8 043
TOI_{AFF}	Agriculture Trade Openness Index $(EX_{AFF}+IM_{AFF})/VA_{AFF}$	FAO, UNCTAD	201	8 043
$SNA2008$	Dummy SNA 2008	OECD, UNSD, OCD	206	8 762

(1) FAO integration from UNSD National Accounts Official Country Data and OECD Annual National Accounts.

Series for these variables are obtained by integrating official national accounts data harvested from UNSD National Accounts Official Country Data (UNSD OCD) and OECD Annual National Accounts (OECD ANA), with priority given to the latter when data are available from both sources. Series are backward linked across vintages to provide the longest possible time series with scaling factors obtained from the latest overlapping year of two vintages. As a result of this process, the most recent years of data do reflect the latest official country data, while data points from earlier portion of the time series may have been transformed through this linking exercise.[8]

Working with the above-mentioned data sources, we can rely on country data – even though very incomplete in some cases – on some or all of $GFCF_{AFF}$, CFC_{AFF} or NCS_{AFF} for 118 countries to guide us in our data gap filling exercise. For the other 88 countries included in the database, we fully rely on imputation methods.

The bottom part of Table 1 presents the set of candidate explanatory variables that will be considered for inclusion in the regression analysis in the AIR estimation exercise. These variables are presented in their basic form prior to any transformation (e.g., logarithmic, square and ratio transformations).

Among the core variables included in the input database and that will be used in the regression analysis, the first set is formed of SNA based indicators obtained from the UNSD National Accounts Estimates of Main Aggregates (UNSD NAE) database. These include Gross Domestic Product (GDP) and Gross Domestic Product per capita ($GDPPC$), Agriculture Value Added (VA_{AFF}), economy-wide Gross Fixed Capital Formation ($GFCF$) and its associated Investment Ratio (IR), given by the GFCF-to-GDP ratio.

In addition, the input database contains information on the Trade Openness Index for the agriculture industry (TOI_{AFF}), defined as the ratio of total trade in agriculture products over agriculture value added, and on the agriculture Export-to-Value Added ratio, $EXtoVA_{AFF}$, which is total export in agricultural production as a share of VA_{AFF}. For these two variables, the main

[8] The backward linking procedure to obtain the longest possible series from country data is consistent with OECD and UNSD methodology as implemented for the 2017 releases of OECD ANA and UNSD AMA databases.

sources are the FAO Trade Database and the UNCTAD Merchandise trade matrix. On the one hand, the rationale for including trade-related variables to the list of covariates is that we may expect the degree of trade openness in agriculture to influence, on average, the productive capacity in place. On the other hand, this might also introduce a bias towards small-scale countries depending on their specialization, resource base, trade position, etc.

We also include in the dataset a dummy variable that takes on the value 1 for countries implementing the SNA 2008 revision for the compilation of national accounts and 0 otherwise.[9] The motivation is that with the SNA 2008 the capital base of a country is enlarged by treating research and development as investments and not anymore as intermediate consumption. The result of this new accounting principle is also an increase in value added and GDP, as pointed out by van de Ven (2015), who finds that for OECD countries "the increase of GDP due to treating R&D as investment is on average 2.2 percentage points". The dummy is included to account for these differences in data treatment.

For subsequent statistical analysis, one way to cluster the data is dividing the countries and territories into three groups: OECD countries (OECD), non-OECD High– and Upper Middle–Income Countries (HMIC), and Lower Middle– and Low–Income Countries (LIC). The country categorization combines OECD membership and World Bank income group classification as rule. The latter two groups (HMIC and LIC) are per 2017 classifications by World Bank, with 4,035 USD Gross National Income (GNI) per capita or less, in 2015 value, being the threshold.

Table 2 illustrates the impact of the changeover to the SNA 2008 on the Agricultural Investment Ratio. The number of available observations on AIR and its average are shown for each income group, distinguishing between SNA2008=1 (SNA 2008 revision) and SNA2008=0 (previous revisions). Within each income class, for those countries for which SNA2008=1 the average AIR is higher than for those using previous SNA revisions in the compilation of national accounts.

[9] It is worth mentioning that country data do reflect national methodologies and therefore may correspond to different versions of the System of National Accounts (SNA) and the International Standard Industrial Classification of All Economic Activities (ISIC). Input data harvested from the UNSD OCD database encompass data in ISIC revisions 3 and 4 as well as in the SNA 1968, 1993, and 2008 revisions depending on the standards currently in use in each country.

Table 2: Data availability and average AIR by SNA and income grouping

SNA2008 DUMMY	AIR	LOW INCOME COUNTRIES	HIGH-& MIDDLE INCOME COUNTRIES	OECD COUNTRIES	ALL COUNTRIES
0	Obs.	264	376	161	801
	Mean	0.1	0.1	0.2	0.1
1	Obs.	64	137	964	1 165
	Mean	0.1	0.2	0.3	0.3
All	Obs.	328	513	1 125	1 966
	Mean	0.1	0.2	0.3	0.2

Finally, Table 3 provides some summary statistics on the input dataset by country grouping. As expected, the agriculture sector contributes to a lower share of GDP in richer countries. Interestingly, it is also evident that agriculture investment ratios (AIR) – both for the agriculture sector and total economy – tend to increase as countries move up along the income categories. It is remarkable that the agriculture investment ratio is higher than its economy-wide counterpart in OECD countries, suggesting high investments flows per unit of value added generated in the sector.

Table 3: AIR by SNA and income grouping – Descriptive statistics

	VARIABLE	OBS.	MEAN	CV	MIN	MAX
Low Income Countries	AIR	328	0.1	0.6	0.0	0.3
	$GDPPC$	3 468	883	0.8	56	4 012
	VA_{AFF}/GDP	3 468	0.3	0.5	0.0	0.8
	IR	3 468	0.2	0.5	0.0	0.7
	$EXtoVA_{AFF}$	3 168	0.2	1.3	0.0	3.2
High & Middle-Income Countries	AIR	513	0.1	0.8	0.0	0.9
	$GDPPC$	3 784	10 478	1.3	150	89 244
	VA_{AFF}/GDP	3 784	0.1	0.9	0.0	0.7
	IR	3 784	0.3	0.4	0.0	1.0
	$EXtoVA_{AFF}$	3 172	1.1	3.8	0.0	57.6
OECD Countries	AIR	1 121	0.3	0.5	0.0	0.8
	$GDPPC$	1 510	25 104	0.6	1 869	88 748
	VA_{AFF}/GDP	1 510	0.05	0.9	0.0	0.3
	IR	1 510	0.2	0.2	0.1	0.4
	$EXtoVA_{AFF}$	1 441	1.0	1.4	0.0	11.9
All Countries	AIR	1 962	0.2	0.7	0.0	0.9
	$GDPPC$	8 762	9 201	1.5	56	89 244
	VA_{AFF}/GDP	8 762	0.2	0.9	0.0	0.8
	IR	8 762	0.2	0.4	0.0	1.0
	$EXtoVA_{AFF}$	7 781	0.7	3.9	0.0	57.6

4 Estimation of the Agricultural Investment Ratio for countries with incomplete series

In this section we discuss different techniques to estimate missing observations on the AIR.

Depending on data availability, we follow three different estimation approaches. The first one applies to countries with limited number of missing data points on the AIR. In that case, univariate ARIMAX models are estimated to ensure that country-specific information is exploited. This approach is described in Section 4.1. The second approach, presented in Section 4.2, applies to countries for which a larger number of observations on the AIR is missing, preventing the use of standard time series models. In this instance, we exploit the cross-sectional time series dimension of the dataset by estimating unobserved effects panel data models for imputation purposes. Finally, for those countries with fully missing data on the AIR, we fit pooled OLS regression models on sub-samples for which data are observed. This approach is explained in Section 4.3.

4.1 The ARIMAX-based approach

4.1.1 Implementing the ARIMAX-based approach

We initially consider countries for which a limited number of data points is missing and long enough time series on the AIR is available to implement time series estimation. In this case, we first fit a battery of ARIMA models with additional exogenous variables to the AIR data. Then, based on standard information criteria, we select the best in-sample specification which is subsequently used to estimate the missing values.

The autoregressive integrated moving average model with exogenous covariates extends the $ARIMA(p,d,q)$ model by including the linear effect that exogenous regressors have on the stationary response series y_t. The general form of the $ARIMAX(p,d,q)$ model estimated sepa-

rately for each country is:

$$\phi_p(L)(1-L)^d(y_t - c - \boldsymbol{x}_t^{'}\boldsymbol{\beta}) \;\; = \;\; \theta_q(L)\varepsilon_t, \;\; \varepsilon_t \sim iid(0, \sigma_\varepsilon^2), \tag{5}$$

where y_t is the agriculture investment ratio (AIR), d is to the order of integration, and the stationary AR polynomial of order p, $\phi_p(1-\phi_1 L - \cdots - \phi_p L^p)$, and the invertible MA polynomial of order q, $\theta_q(1+\theta_1 L + \cdots + \theta_q L^q)$, share no common factors. The parameter c represents the constant term, while $\boldsymbol{x}_t$ is a vector of observed exogenous regressors with coefficients $\boldsymbol{\beta}$.

To estimate the model in (5), we use the STATA built-in *arima* function. Maximum likelihood parameter estimates are obtained via the Kalman filter, after having written the model in state-space form. For nonstationary series, conditions to initialise the filter can be specified as suggested by Hamilton (1994), or by assuming a diffuse prior as suggested by Harvey (1989). See Becketti (2013) for further explanation on STATA's *arima* command.

Our implementation of the ARIMAX approach can be summarised in five steps:

1. Input country-specific data series, check that all series are stationary and transform them appropriately (log-transformation, first differences, etc.) if necessary.

2. For each country i, the endogenous variable is the agriculture investment ratio (AIR), which contains missing values. Within a loop, fit a battery of models and single out the one with the best in-sample fit based on minimization of the BIC criterion. The list of candidate models is defined along two dimensions:

   (a) The autoregressive, moving average, and integration orders of the $ARIMA(p,d,q)$ model, with $p \leq 3$, $d \leq 1$ and $q \leq 2$.

   (b) The following set of exogenous regressors (before transformation): GDP and GDP per capita, agriculture trade openness index, agriculture exports-to-value added ratio, total economy investment ratio and agriculture value-added as a share of GDP.

3. After selecting the country-specific preferred model, construct one-step ahead forecasts to estimate missing data on AIR.

4. Graph forecast against observed value to allow for graphical inspection.

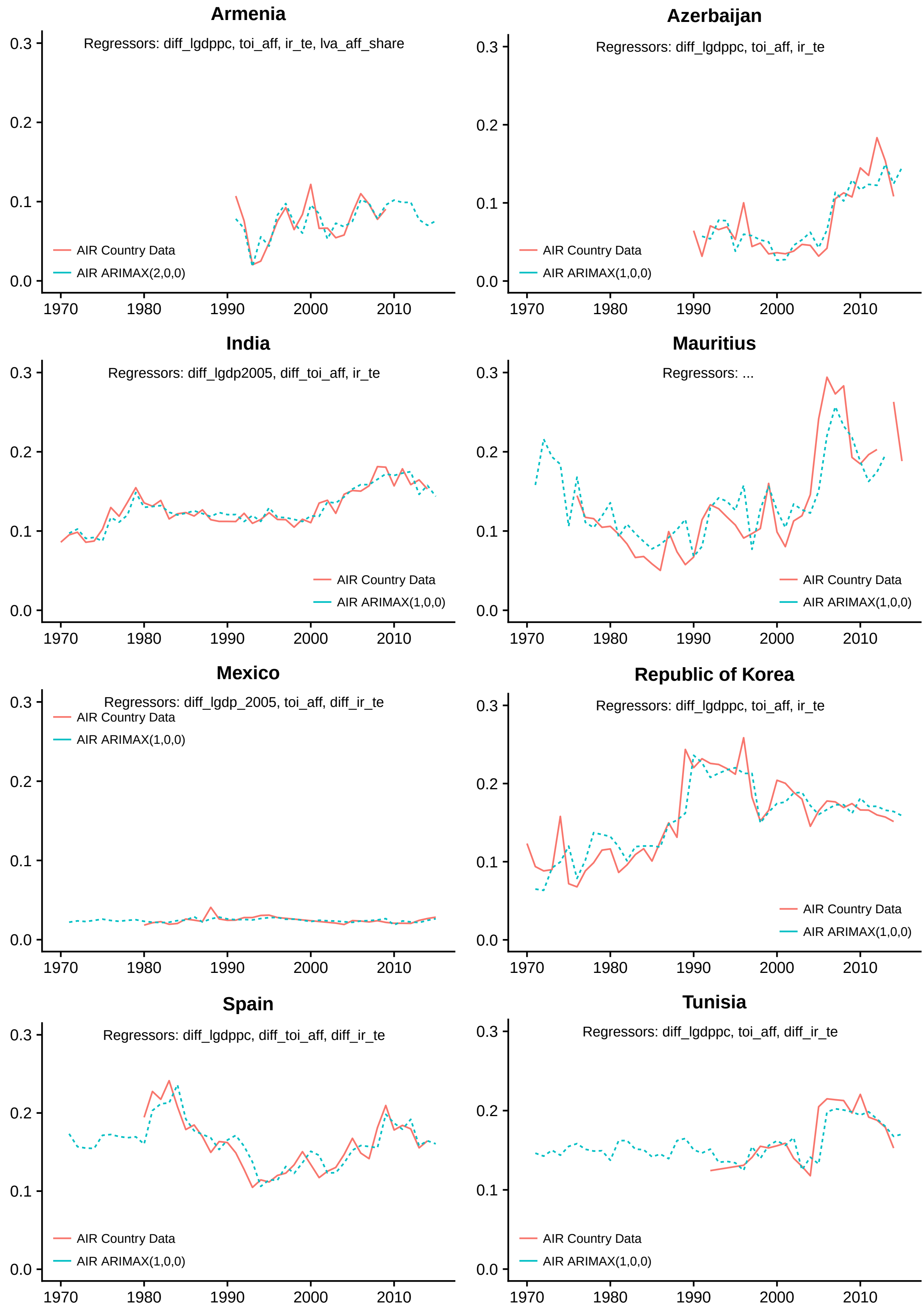
5. Store results and best model specification (with parameter estimates).

If too few observations on y_t are available – less than 15 observations – no time-series-based imputation based on ARIMAX models is performed. For those countries, we use unobserved effects panel data models (Section 4.2).

4.1.2 Results from the ARIMAX-based approach

Out of the 79 countries for which we do have official country data on the AIR ratio, we have employed the ARIMAX procedure to complete the AIR time series for a total of 40 countries. Given that 18 countries have complete time series on AIR (mainly OECD member states) – and therefore no estimation is needed – for 21 countries we have rather implemented the unobserved effects panel data model approach, mostly due to insufficient data availability (less than 15 data points).

Figure 2: ARIMAX model specification and one-step ahead forecasts for AIR, selected countries



Most of the time, $ARIMAX(p,0,q)$ is the preferred specification, with $0 \leq p,q \leq 3$. This is consistent with the fact that, being defined as a ratio of investment to value added, the AIR variable is generally expected to be covariance stationary. Some results from the ARIMAX-based procedure and one-step ahead forecasts are presented in Figure 2.[10] Observed series are plotted in red, while one-step ahead forecasts are in dashed blue.

For instance, for Armenia which requires estimation for 2010–2015 we have selected an $ARIMAX(2,0,0)$ model with four exogenous regressors (log of GDP per capita in first differences, Agriculture Trade Openness Index, Investment Ratio for total economy, log of Value Added as a share of GDP) as the best specification. This specification corresponds to:

$$\begin{aligned} y_t &= c + \boldsymbol{x}_t^{'}\boldsymbol{\beta} + u_t \\ u_t &= \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t, \end{aligned} \tag{6}$$

where y_t is the AIR at time t. Then, after having fitted the model in (6) to the data, we have computed a sequence of one-step ahead forecasts for 2010 through 2015 (missing data points):

$$\hat{y}_{t+1} = \hat{c} + \boldsymbol{x}_{t+1}^{'}\hat{\boldsymbol{\beta}} + \hat{\phi}_1(y_t - \hat{c} - \boldsymbol{x}_t^{'}\hat{\boldsymbol{\beta}}) + \hat{\phi}_2(y_{t-1} - \hat{c} - \boldsymbol{x}_{t-1}^{'}\hat{\boldsymbol{\beta}}), \tag{7}$$

where "hat" indicates estimated parameters and $\hat{y}_{t+1}$ is the one-step ahead forecast.

The plot shows that the in-sample fit is very good and that the one-step ahead forecasts follow closely the observed series, using the *predict* STATA command. Armenia (as well as Azerbaijan) attained independence from the Soviet Union in 1991, therefore for both countries we have not backdated the AIR series prior to this year.

[10] To save space, we show results for a limited number of countries: Armenia, Azerbaijan, India, Republic of Korea, Spain, and Tunisia.

4.2 Unobserved effects panel data models approach

4.2.1 Implementing the unobserved effects panel data model approach

For countries with insufficiently long time series on AIR (i.e., less than 15 observations), estimation of missing AIR values is done relying on an unobserved effects panel data model approach. In particular, working with the full input dataset described in Section 3.2, we divide the countries into three groups according to their GNI per capita: OECD countries, high and middle-income, and low-income countries. For each of these groups, we use the time-series and cross-section information contained in the input dataset to model and estimate through unobserved effects panel data models the relationship between the AIR and selected independent variables. The fitted model will then serve for imputation purposes.

We first estimate both fixed effects and random effects panel models, then perform the Hausman test to select the most suitable specification. According to the Hausman test, we retain the fixed effects specification for all the three groups of countries.[11] The general form of the panel data fixed effects model is given by:

$$\begin{aligned} y_{i,t} &= c_i + \boldsymbol{x}'_{i,t}\boldsymbol{\beta} + u_{i,t}, \;\; u_{i,t} \sim iid(0, \sigma_u^2), \\ & \quad i = 1, 2, \ldots, N; t = 1, 2, \ldots, T, \end{aligned} \tag{8}$$

where $y_{i,t}$ is the AIR for country i at time t. Moreover, the vector $\boldsymbol{x}_{i,t}$ contains k observed regressors with $\boldsymbol{\beta}$ coefficients. The parameters c_i are the country-specific time-invariant fixed effects, which capture unobserved heterogeneity at the country level. The idiosyncratic errors, denoted by $u_{i,t}$, are assumed to be uncorrelated over time and independent of the regressors. Finally, the fitted values of model (8), including the fitted country-specific fixed effects c_i, are used to estimate the missing data on $y_{i,t}$. This is possible because, for these latter countries, we do observe the vector $\boldsymbol{x}_{i,t}$ all over the 1970–2015 sample ($\forall i, t$).

In the fixed effects model, differently than in the random effects model, the unobserved het-

[11] Estimation results are available from the authors upon request.

erogeneity effect c_i may be correlated with the regressors, and therefore $E(c_i|\boldsymbol{x}_i)$ is allowed to be any function of $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iT})$. Only the assumption of strict exogeneity is required, i.e., $E(u_{i,t}|\boldsymbol{x}_i, c_i) = 0$, $t = 1, 2, \ldots, T$. The vector of coefficients $\boldsymbol{\beta}$ is estimated by implementing the within transformation, which removes the effect of the unobserved time–invariant country characteristics; see Wooldridge (2010) for additional details.

The selection of exogenous regressors is performed taking a combinatorial approach and optimized using a leave-one-out cross-validation scheme. In particular, we use the STATA built-in *combinatorics* command, a data mining procedure that performs batch OLS estimation, out-of-sample validation and leave-one-out cross-validation on all the $2^Z - 1$ non-empty models formed from a set of Z candidate explanatory variables. As outlined in Section 3.2, the list of candidate regressors is made of the variables presented in Table 1 and their transformed counterparts (log-transformation, first differences, square values, ratio values, etc). The final step is to impute the missing AIR data points, using the estimated panel regression coefficients and including the country-specific fixed effects.

4.2.2 Results from the unobserved effects panel data model approach

In this section we present the estimation results of the panel data fixed effects model in (8). As anticipated, we consider three groups of countries: OECD, High- & Middle-Income Countries (HMIC), and Low-Income Countries (LIC). For all OECD member countries, we have applied the ARIMAX-based approach introduced in Section 4.1.1, therefore they will not be considered here. Rather, the focus will be on HMIC (Table 4) and on LIC (Table 5). For HMIC, we work with a sample of 483 observations covering 24 countries, whereas the sample for LIC includes 310 observations with 16 different economies represented.

A summary of the results for HMIC is presented in Table 4, showing that all coefficients are statistically significant at 5 percent level, except the intercept and that of EX/VA_{AFF} (this latter is marginally significant at 10 percent level). The fit of the model is very good, with $R^2 = 0.71$. It is noteworthy that we are only interested in the variables in x_i to the extent that they help increase

Table 4: High- & Middle-Income Countries (HMIC) – Panel fixed effects model

AIR_OCD	EST. COEFFICIENT	STANDARD ERROR	P-VALUE
$LogGDPPC_{USD2005}$	0.17	0.02	0.00
IR	0.32	0.06	0.00
$LogVA_{AFF}/GDP$	0.05	0.02	0.03
EX/VA_{AFF}	0.02	0.01	0.10
TOI_{AFF}	0.03	0.01	0.00
$LogPopulation$	-0.11	0.02	0.00
$Intercept$	0.37	0.33	0.26
σ_u^2	0.27		
σ_e^2	0.06		
ρ (Fraction of variance due to u_i)	0.95		
Number of observations = 483		Number of groups (countries)=24	
F test that all u_i=0: F(23, 453) = 32.18		Prob >F=0.00	
$Corr(u_i, Xb)$ = -0.94		$R^2 = 0.71$	

the model's fit, for interpolation purposes. Possible endogeneity biases in the coefficients are less of a concern here as long as we refrain from causal interpretations of the estimates.

Most coefficients present the expected sign. For instance, the coefficient on the log of GDP per capita ($LogGDPPC_{USD2005}$), capturing the level of economic development across countries, is positive, implying that more economically developed countries have on average a higher investment ratio in agriculture. Similarly for the coefficient on the Investment Ratio ($IR = GFCF/GDP$), which is also positive, reflecting that countries with a higher share of total economy investment over GDP have also on average a higher investment ratio in agriculture. The share of log of value added in agriculture over GDP ($Log(VA_{AFF}/GDP)$), which mirrors the relative size of agriculture sector in total economy, has a positive coefficient as well. On the contrary, the coefficient on log of population ($LogPopulation$) is negative, indicating that larger countries in the considered sample have a lower investment ratio in agriculture. Lastly, whereas the coefficient on export over value added in agriculture (EX/VA_{AFF}) is barely significant only at 10 percent level, that of the agriculture trade openness index (TOI_{AFF}) is positive and significant. The latter indicator adds imports and exports in goods and services in agriculture and divides this sum by agriculture value added. Thus, this seems to suggest that agriculture sectors in countries more exposed to international trade also records a higher agriculture investment ratio, even though in this latter case there might be simultaneity between the two variables.

Table 5: Low-Income Countries (LIC) – Panel fixed effects model

AIR_OCD	EST. COEFFICIENT	STANDARD ERROR	P-VALUE
$LogGDPPC_{USD2005}$	0.11	0.01	0.00
$LogGDP_{USD2005}$	-0.05	0.01	0.00
$\Delta(VA_{AFF})$	-0.05	0.01	0.00
$Intercept$	-0.17	0.05	0.00
σ_u^2	0.09		
σ_e^2	0.03		
ρ (fraction of variance due to u_i)	0.91		
Number of observations = 310		Number of groups (countries)=16	
F test that all u_i=0: F(15, 291) = 29.41		Prob >F=0.00	
$Corr(u_i,Xb)$ = -0.90		$R^2 = 0.74$	

Coming to the results for LIC summarised in Table 5, all coefficients are significant at 1 percent, including the intercept. In this case as well the fit of the model is very good ($R^2 = 0.74$). Similar considerations as above on the coefficients apply. In particular, the coefficient for log of GDP per capita ($LogGDPPC_{USD2005}$) is positive. On the other hand, the coefficient for log of GDP ($LogGDP_{USD2005}$) is negative, probably linked to country-size effects as also observed for log of population in Table 4. Lastly, the coefficient on first difference of value added in agriculture ($\Delta(VA_{AFF})$) is negative. This might be simply related to the fact that, being the AIR defined as investment over value added, a sharp increase in value added generates a reduction in the AIR ratio itself (and vice versa).

We close this section with some considerations on the calculation of standard errors. We have estimated model (8) by using standard error estimates which are consistent in adjusting for heteroscedasticity and autocorrelation. Yet, in our macro-panel dataset there is probably an issue of cross-sectional correlation among countries, which is often a feature observed in macro panels with long time series dimension (Baltagi, 2008). The presence of cross-sectional dependence can be checked by running standard tests such as the Breusch-Pagan (1979) Lagrange multiplier (LM) test of independence or the Pesaran (2004) cross-sectional dependence test. If detected in the sample, then the fixed effect model can be estimated with standard errors corrected for cross-sectional dependence, by following the approach originally proposed by Driscoll

and Kraay (1998) and described in Hoechle (2007, Sect. 4.2).[12] However, we do not expect this to affect our coefficient estimates although their standard errors might be biased. The standard fixed effects estimator continues to be consistent, nonetheless. We plan to elaborate upon this issue in future research.

4.3 Pooled OLS estimation approach

4.3.1 Implementing the pooled OLS estimation approach

For those countries for which we do not have agriculture investment data, the strategy is to estimate with pooled OLS the relationship between the AIR and selected independent variables, using the information from countries for which data on the AIR is available. We then use the estimated regression coefficients to estimate the missing data.

The general representation of the linear regression model is:

$$\begin{aligned} y_{it} &= c + \boldsymbol{x}'_{i,t}\boldsymbol{\beta} + u_{i,t}, \;\; u_{i,t} \sim iid(0, \sigma_u^2), \\ & \quad i = 1, 2, \ldots, N; t = 1, 2, \ldots, T, \end{aligned} \tag{9}$$

where y_{it} is AIR for country i at time t and $\boldsymbol{x}_{i,t}$ is a vector containing k observed regressors, with $\boldsymbol{\beta}$ a vector of unknown coefficients. We assume that the regressors $\boldsymbol{x}_{i,t}$ are uncorrelated with the error term $u_{i,s}$ for all s,t (strong exogeneity assumption). The pooled OLS estimation approach treats all observations of a country as independent with all country-year pairs pooled into a large cross-sectional sample.[13] The fitted model (9) is used to estimate the missing data on $y_{i,t}$ for those countries for which data on the AIR are fully missing. This is possible because we do observe the vector $\boldsymbol{x}_{i,t}$ all over the 1970–2015 sample ($\forall i, t$).

As for the panel regression approach with fixed effects, we treat OECD, HMIC and LIC sep-

[12] Hoechle (2007) developed a STATA routine ($mathsfxtscc$) enabling estimation of both pooled OLS and fixed effects (within) regression models with modified standard errors to account for cross-sectional dependence.
[13] A different estimation strategy might be based on spatial auto-regression models (SAR) with missing variables in the dependent variable; see, for instance, Wang and Lee (2013).

arately, in order to reduce heterogeneity across clusters of countries.[14] For each group of countries, the list of regressors considered for inclusion in the OLS specification is taken from Table 1 together with their transformed versions (logarithmic, square and ratio). The best specification is identified based on a leave-one out cross-validation scheme applied on all the $2^Z - 1$ non-empty models based on a set of Z candidate explanatory variables.

Regarding the data cleaning prior to model selection and estimation, we operate in two stages. First, we identify and eliminate from the samples (i.e., OECD, middle-income and low-income countries) "outlying" countries. Next, we estimate the selected model by performing robust regression techniques, implemented by means of the STATA *rreg* command, which assigns lower weight to the most influential observations.

4.3.2 Results from the pooled OLS estimation approach

In this section we present the results obtained estimating the model in (9) by pooled OLS. Tables 6 and 7 refer to HMIC and LIC, respectively. For HMIC, we work with a sample of 440 observations, and for LIC, 276 observations.[15]

All estimated coefficients in Table 6 are statistically significant at conventional levels. Among the regressors, we introduce the $SNA2008$ dummy capturing a newer definition (the System of National Accounts 2008, which replaced the 1993 version of the system) of capital stock only used by some countries in the sample. The corresponding coefficient is positive, meaning that countries adopting the newer national accounting definition of capital stock have on average a higher AIR. The remaining coefficients bear the expected signs. In particular, the coefficient on the log of GDP per capita ($Log(GDPPC_{USD2005})$), which proxies the level of economic development, is positive, while the coefficient on the square of the same variable is negative, pointing to a possible nonlinearity in the estimated relation. The coefficient on export over value added

[14] In principle, we might try to work with a number of clusters larger than three, in order to further reduce heterogeneity across country groups. A data-led technique for strata delimitation might be used for this purpose. This is left for future research.

[15] As mentioned in Section 4.3.1, we lose some observations due to the removal of potential outliers prior to model selection and estimation.

Table 6: High- & Middle-Income Countries (HMIC) – pooled OLS estimation

AIR_OCD	Estimated Coeff.	Standard Error	P-value
$Log(GDPPC_{USD2005})$	0.14	0.06	0.02
$(Log(GDPPC_{USD2005}))^2$	-0.01	0.00	0.02
$Log(VA_{AFF}/GDP)$	-0.04	0.01	0.00
EX/TOI_{AFF}	0.09	0.01	0.00
IR	0.14	0.04	0.00
$SNA2008$	0.05	0.01	0.00
$Intercept$	-0.69	0.25	0.01
Number of observations = 440		F(6, 433) = 41.96	
$R^2 = 0.34$		$Adjusted\ R^2 = 0.33$	
Variable	VIF	1/VIF	
$Log(GDPPC_{USD2005})$	334.97	0.00	
$(Log(GDPPC_{USD2005}))^2$	347.78	0.00	
$Log(VA_{AFF}/GDP)$	3.55	0.28	
EX/TOI_{AFF}	1.51	0.66	
IR	1.03	0.97	
$SNA2008$	1.25	0.80	
Mean VIF	115.02		

in agriculture (EX/TOI_{AFF}) is positive, suggesting that AIR is larger for data points with higher export shares. On the contrary, the coefficient on the share of log of value added in agriculture over GDP ($Log(VA_{AFF}/GDP)$) is negative, indicating that in the sample under consideration there is a negative relation between AIR and the share of agriculture in the economy (in terms of value added). Lastly, the coefficient on the Investment Ratio (IR) is positive, meaning that observations for which the share of total economy investment over GDP is higher are associated, on average, with a higher investment ratio in agriculture.

The bottom part of Table 6 reports the variance inflation factor (VIF) associated to each coefficient. The VIF measures the impact of collinearity among the explanatory variables in the estimated regression model. The reciprocal of the VIF is the tolerance. VIF ranges from 1 to infinity, whereas the tolerance is bounded between 0 and 1. If a variable is highly correlated with other explanatory variables, then there is an issue of multicollinearity. As a rule of thumb, multicollinearity is present for a given variable if the corresponding $VIF(\beta_i) > 10$. In Table 6, all the variables report a VIF coefficient smaller than 10, apart from $Log(GDPPC_{USD2005})$ and its squared transformation $(Log(GDPPC_{USD2005}))^2$. Overall, multicollinearity is not exerting an undue influence on the results.

Table 7: Low-Income Countries (LIC) – pooled OLS estimation

AIR_OCD	EST. COEFFICIENT	STANDARD ERROR	P-VALUE
$Log(GDPPC_{USD2005})$	0.06	0.01	0.00
$Log(GDP_{USD2005})$	0.01	0.02	0.00
$Log(VA_{AFF}/GDP)$	0.20	0.03	0.00
$(Log(VA_{AFF}/GDP))^2$	0.05	0.01	0.00
IR	0.04	0.03	0.26
$SNA2008$	0.05	0.01	0.00
$Intercept$	-0.24	0.04	0.00
Number of observations = 301		F(6, 294) = 75.08	
$R^2 = 0.51$		$Adjusted\ R^2 = 0.50$	
VARIABLE	VIF	1/VIF	
$Log(GDPPC_{USD2005})$	3.79	0.26	
$Log(GDP_{USD2005})$	3.29	0.30	
$Log(VA_{AFF}/GDP)$	51.24	0.02	
$(Log(VA_{AFF}/GDP))^2$	45.58	0.02	
IR	1.05	0.95	
$SNA2008$	3.41	0.29	
Mean VIF	18.06		

Lastly, Table 7 illustrates the estimated model for LIC. The GDP per capita variable is again positively related to the AIR variable as is the GDP in level ($Log(GDP_{USD2005})$). Differently from HMIC countries, the estimated coefficient on $Log(VA_{AFF}/GDP)$ is positive, suggesting that countries with a larger contribution of agriculture to GDP tend to display a higher AIR. This positive association is reinforced through a non-linear effect captured by the positive coefficient on the square of the same variable ($(Log(VA_{AFF}/GDP))^2$). The coefficient on the Investment Ratio (IR) is also positive, as is for HMIC countries: observations with a larger total economy investment-to-GDP ratio tend to be associated with a higher investment ratio in agriculture. Lastly, the $SNA28$ dummy variable takes on a positive coefficient, confirming that countries which adopted the newer national accounting definition of capital stock have on average a higher AIR. The bottom part of Table 7 shows the VIF and tolerance values for each variable. Also in this case, we conclude that multicollinearity is not a concern in the estimated model.

5 Capital stock measurement

In this section, we present and discuss a model-based exercise which relies on the PIM for constructing aggregate series on agriculture capital stock.

The two common valuation methods for capital stock series are net and gross. The gross capital stock measure represents the value of all the fixed assets in use valued at the price of new assets regardless of their age. That is, *gross* stock measures ignore the depreciation of fixed assets and consider past investments as new, with retirement only being taken into account. By contrast, *net* capital stock corresponds to the value of gross capital stock minus the cumulative value of consumption of fixed capital or, more generally, depreciation. The decline in value of fixed assets can occur because of physical and economic deterioration (the latter may be due to obsolescence for instance). In what follows, we will focus our discussion on the derivation of net capital stock series in volume (i.e., at 2005 constant prices), which is the preferred measure for productivity analysis.

5.1 The Perpetual Inventory Method

The PIM produces estimates of capital stock and consumption of fixed capital from time series of gross fixed capital formation. In particular, net capital stock is modelled as a sum of past investments in fixed assets still in use after correcting for depreciation.[16]

The net capital stock for country i at the end of period t, $NCS_{i,t}$, can be written as a function of the net capital stock at the end of previous period, $NCS_{i,t-1}$, gross investment in current period, $GFCF_{i,t}$, and consumption of fixed capital, $CFC_{i,t}$:

$$NCS_{i,t} = NCS_{i,t-1} + GFCF_{i,t} - CFC_{i,t}. \quad (10)$$

Assuming a constant age-independent geometric depreciation rate δ, we can rewrite the net

[16] Our PIM representation constitutes a simplified variant of the PIM in that different types of assets are considered jointly. Indeed, full implementation of the PIM would require relatively long time series of gross fixed capital formation, broken down by type of assets, which are not available.

capital stock in (10) as:[17]

$$NCS_{i,t} = (1-\delta)NCS_{i,t-1} + (1-\delta)GFCF_{i,t}. \tag{11}$$

After repeated substitution, the PIM equation becomes:

$$NCS_{i,t} = \sum_{j=1}^{\infty}(1-\delta)^{j}GFCF_{i,t-(j-1)}. \tag{12}$$

The basic idea of the PIM clearly shows through the above model representation: investment flows add up cumulatively to build up the capital stock after adjustment for depreciation. Once fixed assets enter the capital stock, they virtually remain there forever (hence the "perpetual" term) and provide productive services at a diminishing rate.

However, an infinite series of past investment flows is never available. Therefore, working with long but finite time series on $GFCF$ of length T, the PIM equation in (12) can be rewritten as:

$$NCS_{i,t} = (1-\delta)^{T}NCS_{i,t-T} + \sum_{j=1}^{T}(1-\delta)^{j}GFCF_{i,t-(j-1)}, \tag{13}$$

where $NCS_{i,t-T}$ is the initial capital stock. Accordingly, the following information is required for PIM implementation:

1. Long time series on $GFCF_t$ measured at constant prices;

2. An estimate of the depreciation rate δ of the existing capital stock;

3. Estimates of the initial net capital stock.

Based on the estimation exercise presented in Section 4, we have built long time series on

[17] Our assumption of a geometric depreciation profile is reasonable in regards to the findings of Hulten and Wykoff (1996). These authors find that the depreciation pattern of aggregated stock of fixed assets tend to approach a geometric form, even if the narrowly-defined asset groupings present highly non-geometric profiles. That is, when working with aggregate data on fixed assets, the depreciation profile is well approximated by the geometric model (see also Kamps, 2006).

$GFCF_t$ at constant prices, for 46 data points covering the 1970–2015 time span. The next two sections discuss the other pieces of information needed, i.e., the depreciation rate δ and the initial capital stock.

5.2 The Depreciation Rate

Within the PIM framework, fixing the depreciation rate is a central step to produce accurate estimates of net capital stock series and thereby to correctly assess an industry's productive capacity. For instance, productivity measures such as multi-factor productivity estimates depend on the evolution of the capital stock at disposal within an industry. Yet, limited deviations in the assumed depreciation profiles can lead to different statistical measures of the amounts of capital in fixed assets that is being applied in the production process.

Much attention has been given in the literature to estimating depreciation patterns, with concordance of results yet in sight. Generating accurate depreciation profiles is a difficult task for several reasons. While the preferred approach would be to estimate depreciation profiles for narrowly-defined groups of fixed assets on an industry-by-industry basis, often such level of disaggregated information is not available (see van Rooijen-Horsten et al., 2008). In addition, there exists different approaches towards deriving depreciation rates in practice (OECD, 2009). These always involve a number of model-based assumptions and a choice among different estimation techniques (see Statistics Canada, 2007).

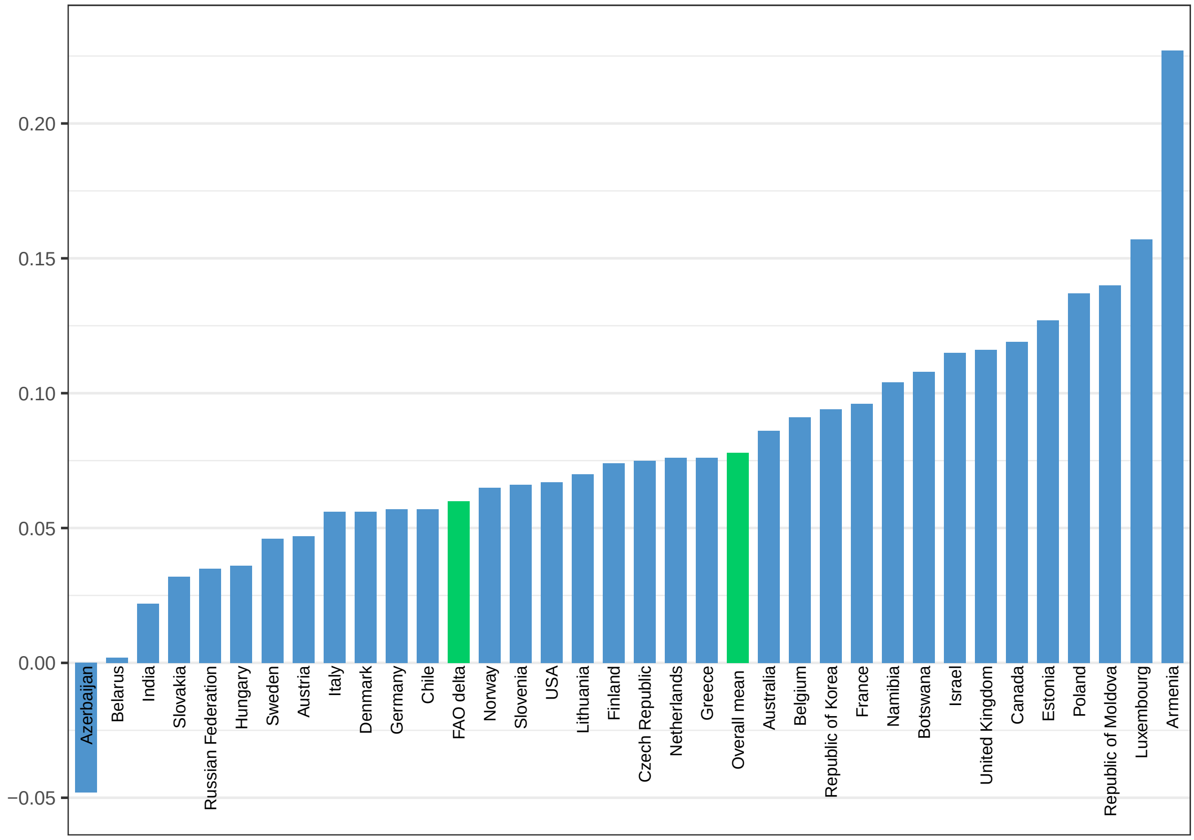
For our purposes, we need a depreciation rate that applies to aggregate measure of investments in fixed assets for the agriculture sector for each country. However, in the absence of econometric estimates of such depreciation rates, we consider a constant annual depreciation rate of 6.0 percent that identically applies to all countries.

In practice we would expect to have country-specific depreciation rates, owing to the composition of the aggregate stock of capital by type of assets, which is likely to differ across countries. Besides, external factors such as weather conditions or different practices in the way capital stock is used can influence the pattern of depreciation and decay of fixed assets used for agriculture production. Along similar lines, the composition of fixed assets is expected to evolve over

time so that aggregate rates of depreciation are unlikely to remain constant over time. However, in the absence of further statistical information, as in Kamps (2006) and Berlemann and Wesselhöft (2014, 2017), we decide to work with a common and time-invariant depreciation rate in order not to introduce non evidence-based distortion in cross-country analysis.

Note that our assumed depreciation rate is somewhat higher than that in Berlemann and Wesselhöft (2014, 2017), where the aggregate depreciation rate is assumed to range from 3.4 percent to almost 4.5 percent. However, these authors construct series on economy-wide aggregate capital stock so that residential fixed assets – which have estimated depreciation well below 2 percent – may reduce the overall depreciation rate. In Figure 3, we present implicit depreciation rates calculated for those countries for which we do have time series on agriculture capital stock. It appears from this graph that our assumed depreciation rate of 6 percent is somewhat below the sample average implicit depreciation rate, despite being higher than in Berlemann and Wesselhöft (2014, 2017).

Figure 3: Average implicit depreciation rate

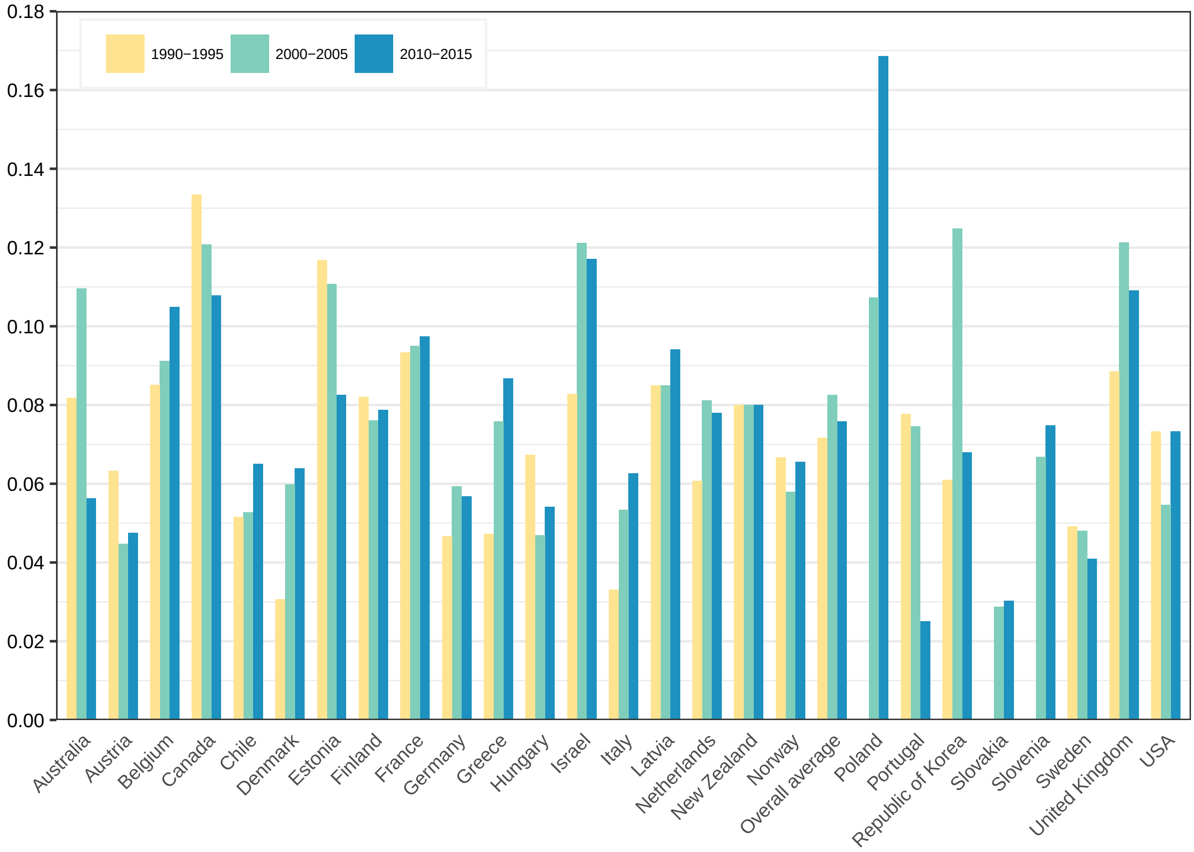


Source: Authors' calculations on official country data on ACS.

Another approach to calculate the depreciation rate is provided by the declining balance method. This method is sometimes used to establish a depreciation rate δ in the absence of econometric estimates, on the basis of information about average service lives of a group of assets (OECD, 2009). The "declining balance" formula is $\delta = DBR/L_{sl}$, where L_{sl} is the length of service life and DBR is an estimated declining-balance rate. This parameter determines the curvature of the geometric depreciation profile, with larger values associated with higher depreciation of the fixed assets earlier in the service life. Assuming an average-declining balance rate for aggregate agricultural capital stock of 1.5 and a value for δ equal to 0.06 gives an average service life of 25 years.[18]

[18] This coincides with the empirical work by Hulten and Wykoff (1996), which produced an average value of DBR

Figure 4: Implicit capital stock depreciation rate, 5-year average



Source: Authors' calculations on official country data on ACS.

We display in Figure 4 country-average implicit depreciation rates for three different periods. We see that while the depreciation rate is not constant over time, no clear regularity could be observed across countries either. This further supports our choice of using a time-invariant depreciation profile.

5.3 The Initial Capital Stock

Once a selection for the depreciation rate has been made, an estimate of the initial capital stock at time t_0 is needed. We follow the description in Berlemann and Wesselhoft (2014) of the steady-state approach for approximating the initial capital stock. This approach, first introduced

smaller than 2.

by Harberger (1978), assumes that the industry value added grows at the same rate θ as the capital stock (steady-state assumption based on the neoclassical growth theory), yielding:

$$\theta_{VA_{AFF}} = \theta_{NCS_{AFF}} = \frac{NCS_{i,t_0} - NCS_{i,t_0-1}}{NCS_{i,t_0-1}} = \frac{(1-\delta)GFCF_{t_0}}{NCS_{i,t_0-1}} - \delta, \tag{14}$$

where the last equality holds from equation (11). Solving the above equation for NCS_{i,t_0-1} leads to:

$$NCS_{i,t_0-1} = \frac{(1-\delta)GFCF_{i,t_0}}{(\delta+\theta)}. \tag{15}$$

where $GFCF_{i,t_0}$ is the first period for which there is information on agriculture investment flows.[19]

From this expression, and assuming that the economy is in steady state, it is sufficient to have information on the level of investment, the depreciation rate and the growth rate of the industry value added in order to initialize the agriculture capital stock series.

We close this section by providing some final remarks regarding our practical implementation of the above strategy. First, according to equation (15), the initial capital stock depends on the first observed value of $GCFC_i$. In order to produce more reliable and stable initial capital stock estimates, we use three-year averages of $GFCF_i$ instead of a single observation (Berlemann and Wesselhöft, 2014, 2017). Second, for a small handful of countries, we need to set $\theta = 0$ when the calculated long-run real growth rate is negative, in order to avoid implausibly large initial stocks. Note that the approximation error in our measure of the initial capital stock value will tend to zero as the initial base period is left further behind. This stems from the PIM equation (13) where the weight associated to the initial stock decreases as T increases. For this reason, although we start compiling agricultural capital stock series from 1970, FAO's published net capital stock series do not start before 1990.

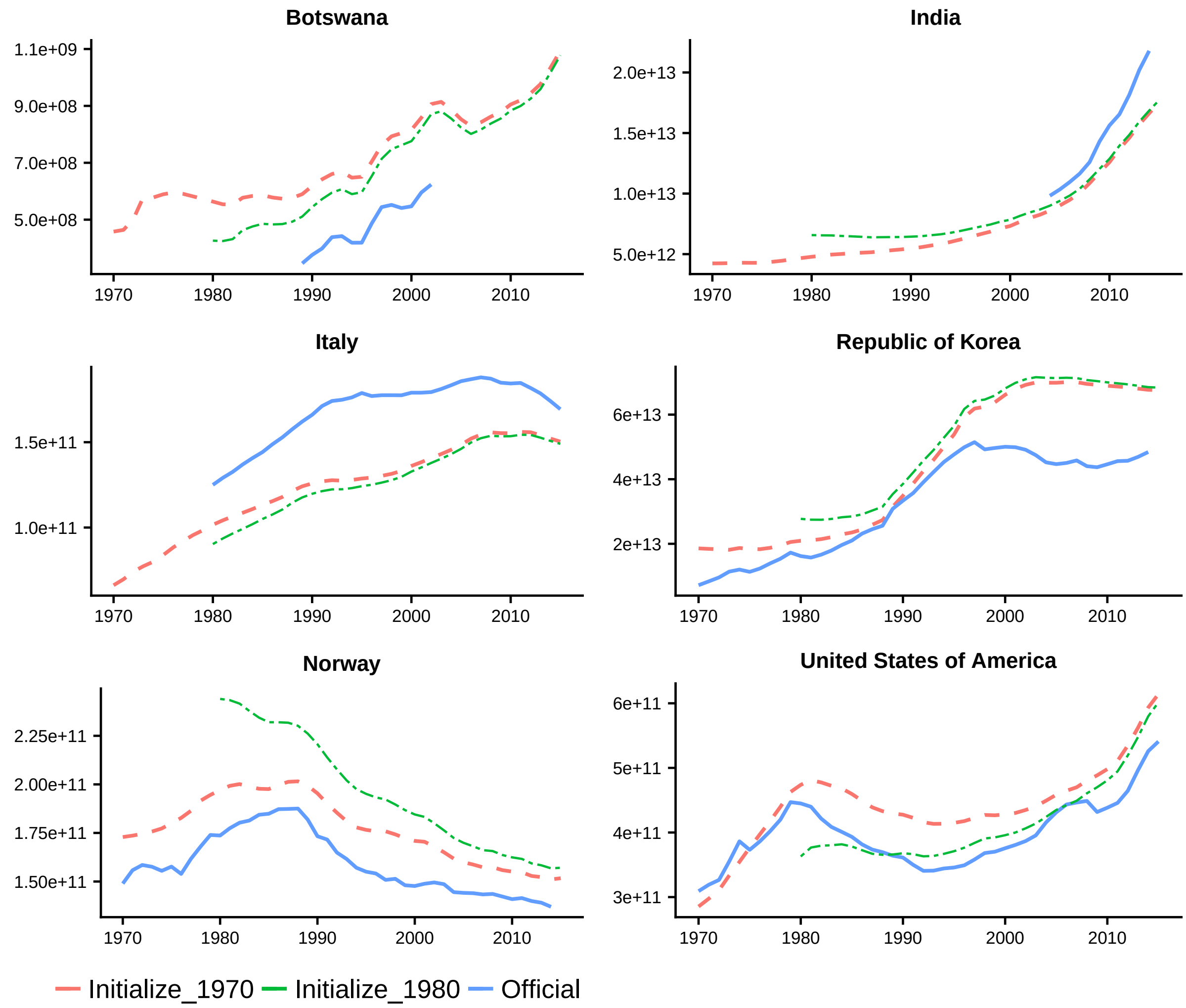
[19] That is, the initial capital stock can be approximated by the level of investment expenditure at time t_0, $GFCF_{i,t_0}$ (i.e., the first period for which there is information on investment expenditure in fixed assets, and a combination of parameters on the long-term growth rate on the industry output, θ, and our assumed depreciation rate, δ).

5.4 Agriculture capital stock estimation results

Implementation of the PIM methodology requires assumptions regarding the capital stock depreciation profile and its initialization. In this section, we present results obtained from applying the above-described PIM framework and discuss how the two main assumptions, namely the selection of the depreciation rate and of the initialization period, affect the estimated agriculture capital stock series. In other words, we assess the sensitivity of the capital stock estimates to different choices of the depreciation rate and of the initialization period.

First, we evaluate how working with different initial periods affects our estimated series. In Figure 5, official country series on agriculture capital stock are plotted (blue line) against two series derived through the PIM methodology. The red dashed line corresponds to the capital stock series obtained by applying our regular assumptions, i.e., using a depreciation rate $\delta = 0.06$ and fixing the initial period to 1970, while the dashed green line is obtained using 1980 for initialization. From equation (15), we know that the level of the stock for the initial year depends on the observed level of the investment flow in the previous year, explaining the difference between the two estimated curves in the first portion of the time span as displayed in Figure 5. Also, the distance between the two PIM-based curves reduces over time to almost vanishing at the end of the period, consistent with equation (13). The more distant we get from the initial period, the lower the weight associated to the initial level in the estimated stock as it is progressively replaced by new investment flows.

Figure 5: Agriculture Capital Stock – Official Country Data vs PIM-based series for different initial periods

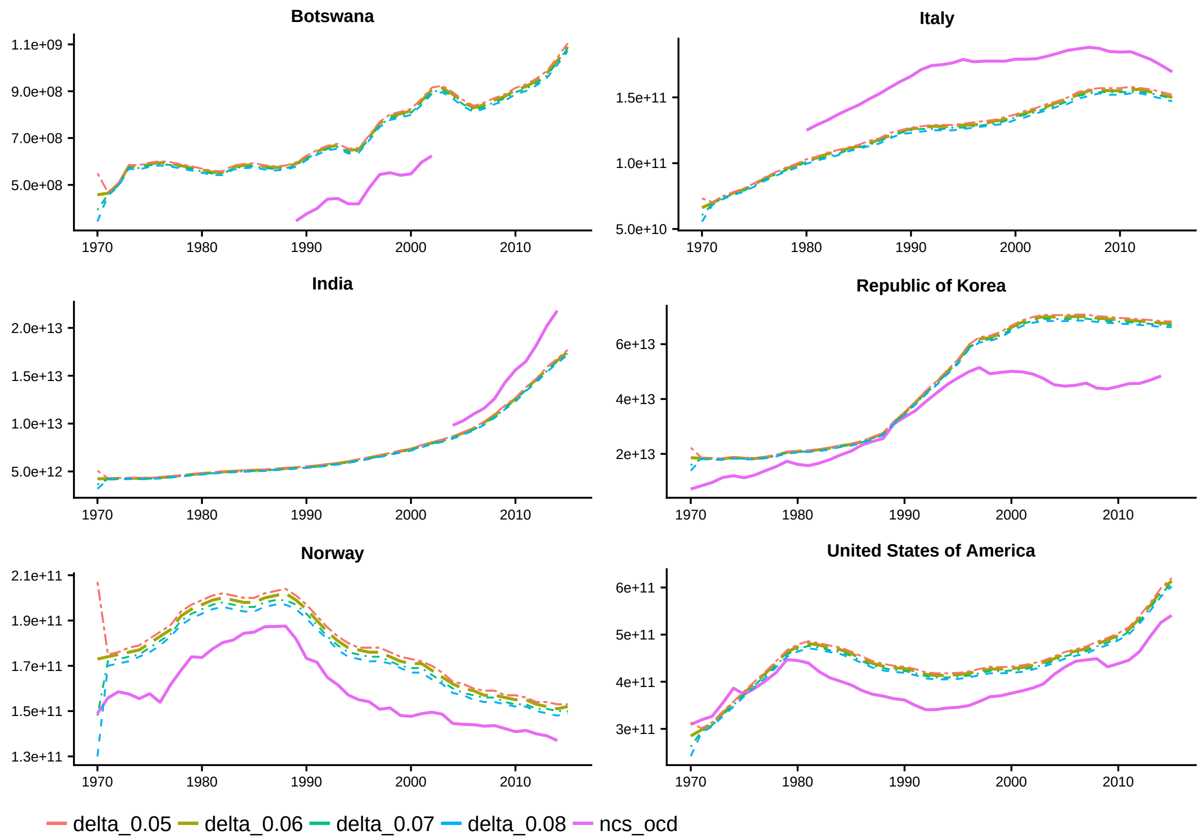


All series are expressed in local currency, 2005 constant price.

We next consider the effect of the chosen depreciation rate δ on the estimated series. To this end, along with the capital stock series obtained from our benchmark PIM (i.e., with $\delta = 0.06$), three additional series for different values of the depreciation rate are estimated and presented in Figure 6 for a selection of countries. As expected, increasing the depreciation rate – in this case from 0.05 to 0.08 – shifts the capital stock series downward but without modifying its overall time profile. This directly results from the assumption of a time-invariant depreciation rate, impacting the level of capital stock only, while the dynamics is determined by the profile of the series on

$GFCF_{AFF}$.

Figure 6: Agriculture Capital Stock – Official Country Data vs PIM-based series for different values of the depreciation rate



All series are expressed in local currency, 2005 constant price.

We conclude this section by commenting on the differences between official country series and the estimated series via our benchmark PIM. At least for the reference countries presented in the above figures, the *time profile* of our estimated capital stock series is consistent with the dynamics observed in the official country series and seems relatively robust to some variation in the depreciation rate and in the initialization period used. As for the latter, a 1970 initialization usually facilitates an estimated capital stock series whose dynamics is similar to that of official country series, as can be seen for Republic of Korea, Norway and the USA. The fit could be explained, at least in part, by the fact that the geometric depreciation rate is a reasonable ap-

proximation of the true depreciation rate, when working with an aggregate capital stock series that combines different types of asset categories (Hulten and Wyckoff, 1981).

However, our estimated series do not align *in levels* to the observed country series. Several reasons may cause this gap in levels, one being that countries usually construct official capital stock series at a more disaggregated level. That is, capital stock series by asset categories are first estimated and then aggregated up into an industry-level capital stock series. In doing so, statisticians can fine-tune the set of assumptions used in their capital stock model (including the depreciation rates and profiles, the expected service life length, etc.). This bottom-up approach, which better accounts for asset composition effects over-time, is absent from our analysis. This discussion also points to the importance of having good estimates of gross fixed capital formation series, as its accuracy is expected to have an even stronger influence on the goodness of approximation of capital estimates than the depreciation rate or the initialization period.

5.5 Global and regional trends in agriculture capital stock

In this section we present an overview of the main trends of agricultural gross fixed capital formation and capital stock at the global and regional level. There is empirical evidence suggesting that increased investment in physical capital in agriculture is one possible driver of the long-term growth in agriculture value-added. Over the 1995–2015 period, global physical investment flows in agriculture – as measured by gross fixed capital formation in agriculture – doubled, going from 212 to 436 billion in constant 2005 USD. Real global agriculture value-added rose as well, from 1.1 trillion in 1995 to 2.0 trillion in 2015.

Yet, very different patterns emerge among regions, as reported in Table 8. Over the 20 year time span, real physical investments rose on average at a global annual grow rate of 3.7 percent, led by Asia (6.5 percent), Africa (4.3 percent) and Northern America (4.1 percent). Latin America & Caribbean followed with an average annual growth rate of 2.6 percent. This rate was 1.0 percent for Europe and dipped slightly into the negative for the Other Developed Countries (-0.5 percent).

Given that investment flows add up to build capital stocks after adjustment for depreciation,

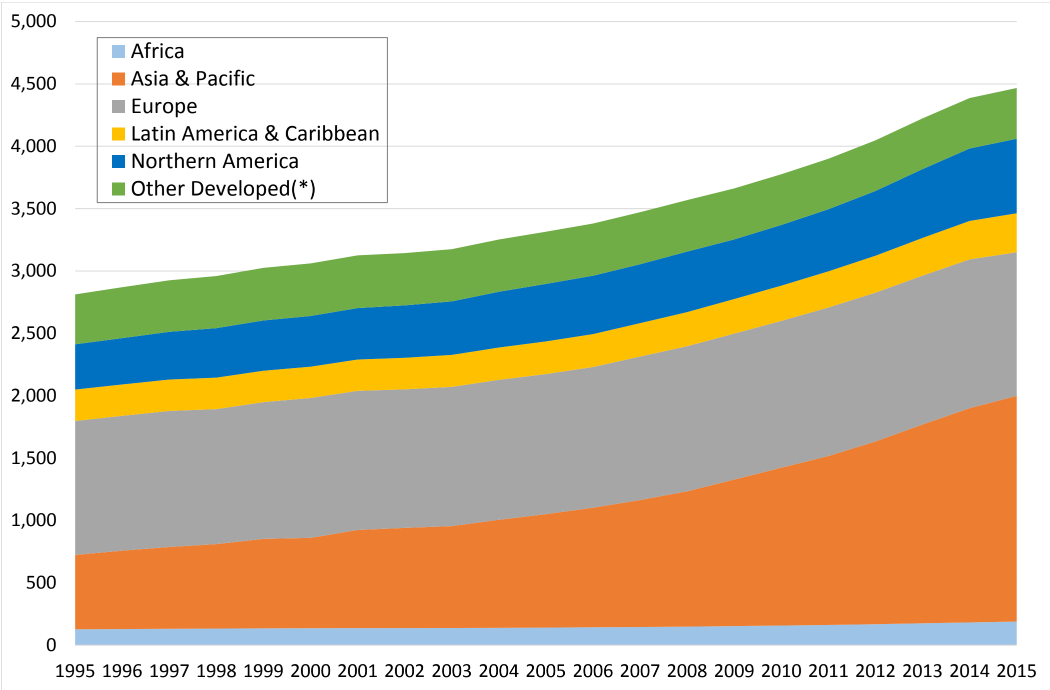
Table 8: Average annual growth rate in agriculture investment flows, 1995–2015

	1996–2000	2001–2005	2006–2010	2011–2015	1996–2015
Africa	3.1%	3.6%	4.5%	5.7%	4.3%
Asia & Pacific	2.4%	6.3%	9.8%	5.0%	6.5%
Europe	2.1%	1.2%	0.4%	-1.6%	1.0%
Latin America & Caribbean	0.5%	4.9%	4.6%	-1.5%	2.6%
Northern America	1.5%	5.0%	2.4%	5.9%	4.1%
Other Developed(*)	-1.0%	-0.9%	-3.5%	4.1%	-0.5%
Global (Total)	1.6%	3.5%	4.4%	3.2%	3.7%

*Other Developed includes Australia, Japan and New Zealand.

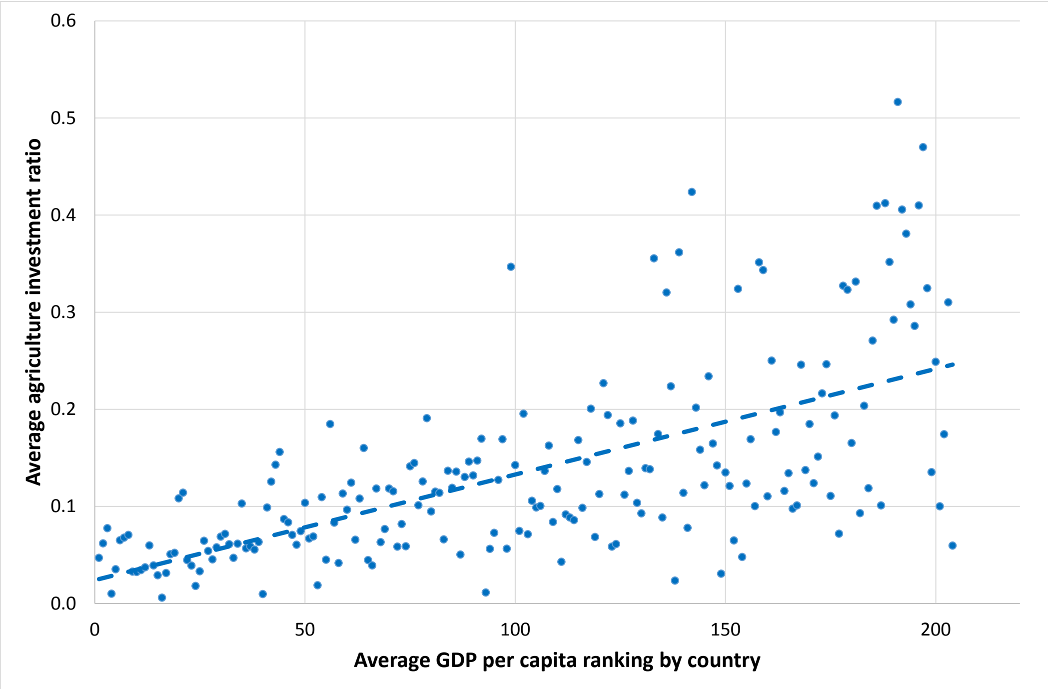
the diverging regional trends in investment directly contribute to differences in regional agricultural capital stock data (Figure 7). It is worth emphasizing that the two regions which saw a significant increase in their contribution to global agriculture value-added – i.e., Asia and Africa – also experienced the highest average growth rate in physical investment flows and capital stock, confirming the central role of physical capital as a key driver of long-term growth through enhanced productivity.

Figure 7: Global Agriculture Net Capital Stock and its Regional Distribution, constant 2005 USD (billions), 1995–2015



Despite lower average growth rates in physical investment flows, developed economies continue to present higher investment ratios – as measured by the share of physical investment over value-added in agriculture. This is evidenced in Figure 8, in which the average agriculture investment ratio (AIR) over the 1995–2015 period is plotted against the GDP per capita ranking by country. The positive relationship indicates that in countries with higher GDP per capita physical investment flows represent a larger share of agriculture value-added, pointing to a more highly mechanised agriculture sector compared to developing countries.

Figure 8: Agriculture Investment Ratio against GDP per capita ranking by country, 1995–2015 average



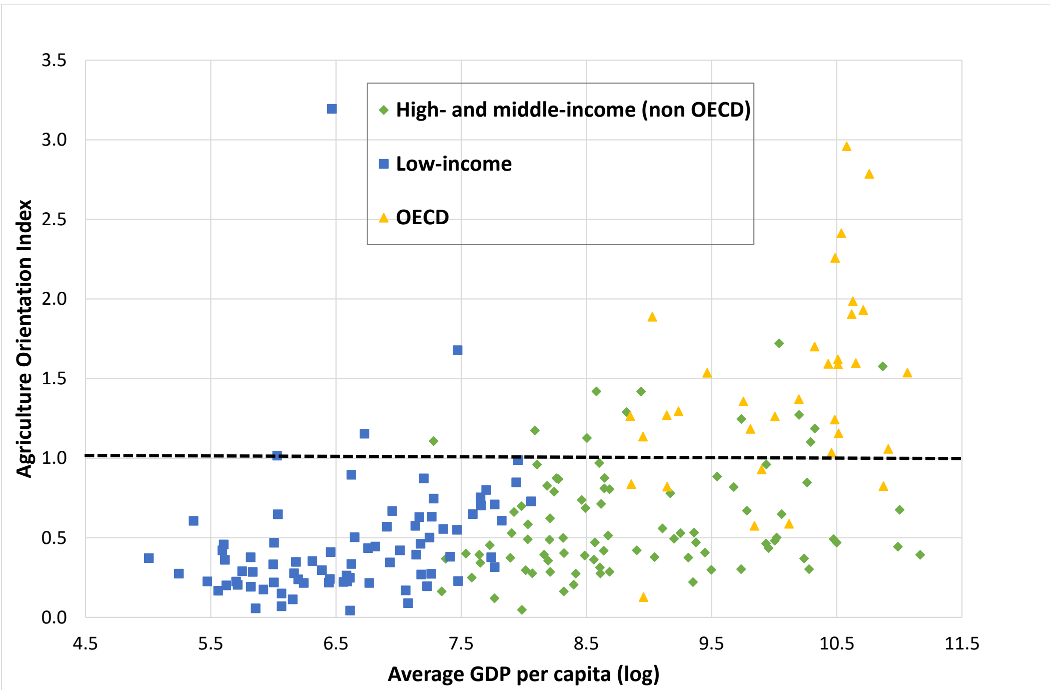
Additional insights can be gained by considering the Investment Ratio Agriculture Orientation Index, which provides a measure of how the investment intensity in agriculture compares to that of the total economy. More specifically, the Agriculture Orientation Index for the Investment Ratio (IR_AOI) for country i at time t is defined as follows:

$$IR\_AOI_{i,t} = \frac{(GFCF_{AFF}/VA_{AFF})_{i,t}}{(GFCF/GDP)_{i,t}}. \tag{16}$$

Countries with a higher investment intensity in agriculture compared to overall economy have an IR_AOI index greater than 1, indicating that on average a larger share of each unit of value-added is spent on GFCF in agriculture with respect to the other sectors of the economy. Interestingly, the OECD average agriculture IR_AOI was above 1.4 over the 1995–2015 period,

while it remained at 0.6 and 0.4 respectively in high- and middle- income countries (non-OECD) and low-income countries, over the same time span. That is, in countries where agriculture has become less important as a contributor to GDP, the investment intensity is even higher than for the other non-agricultural sectors taken as a whole, suggesting that in these countries the agricultural sector is highly mechanised (Figure 9).

Figure 9: Agriculture Orientation Index against GDP per capita. Country and income group averages, 1995–2015



6 Concluding remarks

In this paper we have presented an overview of the new FAO analytical database of physical investment flows and capital stock in the agriculture, forestry and fishery sector. This database includes data (in local currency units and US dollars, both in current and constant value) for 206

countries around the globe and extends from 1990 through 2015. To the best of our knowledge, this is the widest coverage in terms of agricultural GDP at the worldwide level. In this way, we aim at stimulating empirical cross-country research on the role of agriculture in the process of economic development. Whenever available, the database reports official country series on investment and capital stock in agriculture. However, such industry-level data are not available for many countries and have to be estimated through statistical techniques as detailed in the present work. In these cases, the resulting capital stock series in our analytical database do not represent official data on national accounting released by the government or statistical agencies. This is an important feature of the database the data user should be aware of.

We have carefully described the sources and the methodology adopted for the construction of the database, which are consistent with the System of National Accounts framework. Building on previous research programmes held at the World Bank and at the FAO, we have compiled long time series of the agricultural investment-to-value added ratio, which we have then employed to derive agricultural investment flows. The estimated investment series are then converted into agricultural capital stock by applying a variant of the PIM. Our work improves upon previous efforts by departing from strictly time-series-based imputation techniques and allowing for the presence of exogenous regressors. We have also extended the database to countries with fully missing data on gross fixed capital formation.

In terms of future developments, we would like to explore new ways to deal with cross-country heterogeneity in the agricultural capital stock. For instance, we might work with a larger number of strata in order to reduce cross-country heterogeneity. At the same time, we might look into more robust statistical methods for strata delimitation, in addition to the separation criterion we have used (OECD, low, and high and middle-income countries). In terms of econometric models, we might explore spatial auto-regression models with missing data in the dependent variable. For this class of models, a nonlinear least squares method has been proposed by Wang and Lee (2013). We leave all these issues for future research.

References

[1] Baltagi, B. H. (2008). *Econometric Analysis of Panel Data*. 4th Edition. John Wiley & Sons Inc.

[2] Becketti, S. (2013). *Introduction to Time Series Using Stata*. College Station, TX: Stata Press.

[3] Berlemann, M., and J. E. Wesselhöft (2014). Estimating aggregate capital stocks using the PIM: A survey of previous implementations and new empirical evidence for 103 countries. *Review of Economics*, Vol. 65(1), pp. 1–34

[4] Berlemann, M., and J. E. Wesselhöft (2017). Aggregate capital stock estimations for 122 countries: An update. *Review of Economics*, Vol. 68(2), pp. 75–92.

[5] Breusch, T. S., and A. R. Pagan (1979). A simple test for heteroskedasticity and random coefficient variation. *Econometrica*, Vol. 47(5), pp. 1287–1294.

[6] Butzer, R., Y. Mundlak, and D. F. Larson (2010). Measures of fixed capital in agriculture. Policy Research Working Paper 5472. The World Bank, Washington, DC.

[7] Crego, A., D. Larson, R. Butzer, and Y. Mundlak (1998). A new database on investment and capital for agriculture and manufacturing. Policy Research Working Paper 2013. DRG, World Bank, Washington, D.C.

[8] Daidone, S., and G. Anríquez (2011). An extended cross-country database for agricultural investment and capital. ESA Working Paper No. 11-16, Food and Agriculture Organization of the United Nations, Rome.

[9] Driscoll, J. C., and A. C. Kraay. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics*, Vol. 80(4): pp. 549–560.

[10] Fuglie, K. O., and N. E. Rada (2013). Resources, policies, and agricultural productivity in Sub-Saharan Africa. United States Department of Agriculture, Economic Research Report Number 145, February.

[11] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

[12] Harberger, A. C. (1978). Perspectives on Capital and Technology in Less Developed Countries. In: M. J. Artis and A. R. Nobay (Eds.): *Contemporary Economic Analysis*, London, pp. 42–72.

[13] Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

[14] Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *The Stata Journal*, Vol. 7(3), pp. 281–312.

[15] Hulten, C. R., and F. C. Wykoff (1981). The estimation of economic depreciation using vintage asset prices: An application of the Box-Cox power transformation. *Journal of Econometrics*, Vol. 15(3), pp. 367–396.

[16] Hulten, C. R., and F. C. Wykoff (1996). Issues in the measurement of economic depreciation. Introductory remarks. *Economic Inquiry*, Vol. 34(1), pp. 10–23.

[17] Kamps, C. (2006). New estimates of government net capital stocks for 22 OECD countries, 1960–2001. Staff Papers, International Monetary Fund, Vol. 53(1), pp. 120–150.

[18] Larson, D. F., R. Butzer, Y. Mundlak, and A. Crego (2000). A cross-country database for sector investment and capital. *World Bank Economic Review*, Vol. 14, pp. 371–391.

[19] OECD (2009). *Measuring Capital. OECD Manual*. OECD, Paris.

[20] OECD Statistics - National Accounts Division (2017). *Annual National Accounts in OECD.STAT*. OECD, Paris.

[21] Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. University of Cambridge, Faculty of Economics, Cambridge Working Papers in Economics No. 0435.

[22] Statistics Canada, Micro-economic Analysis Division (2007). Depreciation Rates for the Productivity Accounts. The Canadian Productivity Review. No. 5.

[23] United Nations Statistics Division (2017). *Methodology for the National Accounts Main Aggregates Database*. New York.

[24] van de Ven, P. (2015). New standards for compiling national accounts: What's the impact on GDP and other macro-economic indicators? OECD Statistics Brief. No. 20.

[25] van Rooijen-Horsten, M., van den Berg, D., de Heij, R., and M. de Haan (2008). Service life and discard patterns of capital in the manufacturing industry, based on direct capital stock observations, the Netherlands. Discussion paper 08011, Statistics Netherlands.

[26] Von Cramon-Taubadel, S., G. Anríquez, H. de Haen, and O. Nivyevskiy (2009). Investment in developing countries' food and agriculture: Assessing agricultural capital stocks and their impact on productivity. Notes prepared for FAO Expert meeting on How to Feed the World in 2050, 24-26 June 2009.

[27] Wang, W. and L. Lee (2013). Estimation of spatial autoregressive models with randomly missing data in the dependent variable. *The Econometrics Journal*, Vol. 16(1), pp. 73–102.

[28] Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Second Edition. The MIT Press, Cambridge, Massachusetts.

Contact:

Statistics Division (ESS)
FAO-statistics@fao.org
www.fao.org/economic/ess/workingpapers

The Food and Agriculture Organization of the United Nations Viale delle Terme di Caracalla
00153 Rome, Italy

ISBN 978-92-5-131819-5

9 789251 318195
CA6133EN/1/09.19