

# Assessment of agricultural productivity change at country level: A stochastic frontier approach

Alessandro Magrini

## 1. Introduction

Productivity growth of agriculture is widely recognized as a key resource to meet food demand of the rapidly increasing world population, thus monitoring agricultural productivity change at country level is of core importance for international decision makers. The United States Department of Agriculture (USDA) represents the reference source for agricultural productivity change estimates at country level, covering almost all countries in the world for a long and updated period (from 1961 to 2016). USDA estimates consist of yearly changes in Total Factor Productivity (TFP) based on the growth accounting method, i.e., they are obtained as the ratio between the aggregated output and the sum of the input quantities weighted by their cost shares (Caves *et al.*, 1982). Growth accounting is a widely adopted methodology to assess TFP change due to several advantages, in particular it does not require assumptions on the characteristics of the production processes and allows to consider one decision making unit at a time. However, input cost shares are often partially available, and thus they should be approximated or imputed based on several different sources, like in the case of USDA estimates (see Fuglie, 2015, Table A2), with uncontrollable consequences on the accuracy of estimates. In addition, the growth accounting method has the limitation of assuming that the decision making units operate at their optimal conditions, thus it may overestimate TFP change in presence of technical inefficiency. Frontier-based methods like Data Envelopment Analysis (DEA, Charnes *et al.*, 1978) and Stochastic Frontier Models (SFM, Schmidt & Sickles, 1984) represent valid alternatives because, by estimating the production frontier from the sample of decision making units, they can distinguish between change in technology and change in technical efficiency, and do not require input cost shares. The main difference between DEA and SFMs is that DEA does not make any assumption on the production frontier, but it is unable to account for random shocks independent of production and, as a consequence, all the deviations from the frontier are attributed to technical inefficiency. Instead, SFMs can disentangle technical inefficiency from external shocks, but they require parametric assumptions on the production frontier. Despite their appealing properties, SFMs and DEA have been employed only in some scattered studies (see the review in Kryszak *et al.*, 2021) and, as such, the available estimates are not comparable with USDA ones.

In this paper, we apply a SFM with translog specification to the same data on agricultural output and inputs employed by USDA, and exploit the generalized Malmquist index proposed by Orea (2002) to derive country level measures of agricultural TFP change, which are then compared with USDA estimates. Our preference for SFMs over DEA relies in the opportunity to account for external shocks and to assess differences in technology across countries, interactions among inputs and the trend of returns to scale.

## 2. Data and methodology

In this study, we employ the same data on which USDA estimates of agricultural TFP change are based (USDA, 2019). These data are sourced to Food and Agriculture Organization (FAO)

and to International Labour Organization (ILO), and integrated by modeled estimates. The output variable is the gross agricultural production ( $Y$ , thousand US dollars, 2004–2006 average international prices), while the input variables consist of six measures: land use ( $X_1$ , rain-fed cropland equivalents), labour force ( $X_2$ , economically active adults), livestock ( $X_3$ , cattle equivalents), machinery stock ( $X_4$ , 40 CV tractor equivalents), fertilizer use ( $X_5$ , tonnes of nutrients), animal feed ( $X_6$ , megacalories of metabolizable energy). The data have annual frequency in the period 1961–2016 and cover the almost totality of countries in the world. In particular, the considered countries account for more than 99.7% of FAO’s global gross agricultural output. Some national data have been aggregated to create consistent political units over time (e.g., former Yugoslavia, former Czechoslovakia, Ethiopia plus Eritrea, former Soviet Union) or to avoid very small measurements (e.g., Lesser Antilles, Micronesia), for a total of 170 countries. Further details and descriptive statistics can be found in Fuglie (2015).

Let  $i = 1, \dots, n$  denote the decision making units (countries) and  $t = 1, \dots, T$  the time points (years). Also, let  $y_{i,t}$  be the output level produced by unit  $i$  at time  $t$  and  $x_{i,j,t}$  the level of the  $j$ -th input ( $j = 1, \dots, p$ ) employed by unit  $i$  at time  $t$ . A Stochastic Frontier Model (SFM) has the following general form (Schmidt & Sickles, 1984):

$$y_{i,t} = f(\mathbf{x}_{i,t}; \Theta) \exp(v_{i,t} - u_{i,t}) \quad i = 1, \dots, n; \quad t = 1, \dots, T \quad (1)$$

where  $f$  is the production frontier, representing the maximum output level technically feasible based on a given combination of the inputs  $\mathbf{x}_{i,t} = (x_{i,1,t}, \dots, x_{i,j,t}, \dots, x_{i,p,t})$  and a given technology  $\Theta$ , while  $v_{i,t} \in \mathbb{R}$  and  $u_{i,t} \in \mathbb{R}^+$  are two random errors representing the deviation from the production frontier  $f$  due to shocks, respectively, independent of the producer and related to the production. As such, the maximum feasible output may differ from the maximum output level technically feasible due to the occurrence of either favourable or unfavourable events beyond the control of producers. Specifically, the maximum feasible output for unit  $i$  at time  $t$  is equal to  $y_{i,t}^* = f(\mathbf{x}_{i,t}; \Theta) \exp(v_{i,t})$ , thus technical efficiency is  $\text{TE}_{i,t} = y_{i,t}/y_{i,t}^* = \exp(-u_{i,t})$ . We employ the following translog specification for  $f$ :

$$f(\mathbf{x}_{i,t}; \Theta) = \exp \left( \alpha_i + \delta t + \gamma t^2 + \sum_{j=1}^p \beta_j \log x_{i,j,t} + \sum_{j=1}^p \sum_{k=i}^p \beta_{j,k} \log x_{i,j,t} \log x_{i,k,t} + \sum_{j=1}^p \lambda_j t \log x_{i,j,t} + \sum_{j=1}^p \eta_j t^2 \log x_{i,j,t} \right) \quad (2)$$

This formulation is identical to the most commonly adopted one in the literature (see the review in Laureti, 2006, Chapter 3, and in Magrini, 2021), with the difference that we added parameters  $\eta_1, \dots, \eta_p$  to allow output elasticities to vary in time according to a quadratic trend, rather than to a linear one. The frontier specification in (2) leads to the following SFM:

$$\log y_{i,t} = \alpha_i + \delta t + \gamma t^2 + \sum_{j=1}^p \beta_j \log x_{i,j,t} + \sum_{j=1}^p \sum_{k=i}^p \beta_{j,k} \log x_{i,j,t} \log x_{i,k,t} + \sum_{j=1}^p \lambda_j t \log x_{i,j,t} + \sum_{j=1}^p \eta_j t^2 \log x_{i,j,t} + v_{i,t} - u_{i,t} \quad (3)$$

with  $\varepsilon_{i,t} = v_{i,t} - u_{i,t}$ . We complete the specification of the SFM by assuming:

$$\begin{aligned} v_{i,t} &\sim_{i.i.d.} \mathbf{N}(0, \sigma_V^2) \\ u_{i,t} &= \phi_i t + \psi_i t^2 + U_i \quad U_i \sim_{i.i.d.} \mathbf{N}^+(0, \sigma^2) \\ \text{Cov}(v_{i,t}, U_i) &= 0 \quad \forall i, t \end{aligned} \quad (4)$$

where parameters  $\phi_i$  and  $\psi_i$  regulate the second order polynomial trend of the logarithmic technical efficiency of country  $i$ , ‘i.i.d.’ stands for ‘independent and identically distributed’,  $N(\cdot)$  and  $N^+(\cdot)$  denote the Normal and the half Normal distribution, respectively. This specification for  $u_{i,t}$  is the same as in Battese & Coelli (1995) with the addition of the quadratic term.

In order to account for technological gaps among countries with different level of development, we specify four separate models according to the WESP 2020 classification (United Nations, 2020): ‘industrialized’ (28), ‘transition’ (22), ‘developing’ (42), ‘least developed’ (78). Before estimating the parameters, the time variable is coded as the year minus 1961, thus  $t = 0, 1, \dots, 55$ , and the input variables are divided by their respective sample mean. This allows first order coefficients  $\beta_1, \dots, \beta_p$  to be interpreted as the output elasticity of each input evaluated at the sample mean and at the first time point (year 1961), and makes the output elasticity of the  $j$ -th input evaluated at the sample mean and at year  $s$  equal to  $\beta_j + \lambda_j(s - 1961) + \eta_j(s - 1961)^2$ .

TFP change is assessed through the generalized Malmquist index proposed by Orea (2002), which allows to account for variable returns to scale. Based on this index, the TFP change between two time points  $s$  and  $t$  (TFPC $_{s,t}$ ) is decomposed into technological change (TC $_{s,t}$ ), technical efficiency change (EC $_{s,t}$ ), and scale change (SC $_{s,t}$ ):

$$\text{TFPC}_{s,t} = \text{TC}_{s,t} \cdot \text{EC}_{s,t} \cdot \text{SC}_{s,t} \quad (5)$$

Orea (2002) showed that, under a translog production frontier, these three terms equate to:

$$\begin{aligned} \text{TC}_{s,t} &= \exp \left[ \frac{1}{2} \left( \frac{\partial \log y_{i,s}}{\partial s} + \frac{\partial \log y_{i,t}}{\partial t} \right) \right] \\ \text{EC}_{s,t} &= \exp [\mathbb{E}(u_{i,s} | \varepsilon_{i,s}) - \mathbb{E}(u_{i,t} | \varepsilon_{i,t})] \\ \text{SC}_{s,t} &= \exp \left[ \frac{1}{2} \sum_{j=1}^p \left( \frac{\sum_{j=1}^p e_{i,j,s} - 1}{\sum_{j=1}^p e_{i,j,s}} e_{i,j,s} + \frac{\sum_{j=1}^p e_{i,j,t} - 1}{\sum_{j=1}^p e_{i,j,t}} e_{i,j,t} \right) \frac{\log x_{i,j,t}}{\log x_{i,j,s}} \right] \\ e_{i,j,s} &= \frac{\partial \log y_{i,s}}{\partial \log x_{i,j,s}} \quad e_{i,j,t} = \frac{\partial \log y_{i,t}}{\partial \log x_{i,j,t}} \end{aligned} \quad (6)$$

### 3. Results

We performed maximum likelihood estimation of model (3) for each group of countries using the R package `frontier` (Coelli & Henningsen, 2020). Parameter estimates imply significant and positive output elasticities at the sample mean for the almost totality of time points in all the four models, suggesting consistency with the economic theory. Also, the quadratic component of the trend of output elasticities (parameters  $\eta_j$ ,  $j = 1, \dots, p$ ) and of logarithmic technical inefficiencies (parameters  $\psi_i$ ,  $i = 1, \dots, n$ ) are significant, respectively, for most inputs and countries, supporting the adequacy of our model formulation.

Figure 1 displays the time series of the estimated overall elasticity at the sample mean, equal to the sum of all output elasticities at the sample mean by time point. Since the overall elasticity is almost always significantly lower than one for all groups of countries, we deduce that returns to scale are decreasing (and not constant) in the considered period. Based on this result, the assumption of constant returns to scale made by many authors appears just a simplification and not a real property of the production processes of the various countries.

Based on the estimated models, we computed TFP and its components (TC, EC and SC) with  $s = t - 1$  (chained index numbers) and with  $s = 1961$  (index numbers with base year 1961). Table 1 reports average annual percentage changes averaged by group of countries, while Figure 2 displays the time series of index numbers with base year 1961 for a selection of countries. We see that USDA estimates of TFP change are greater in absolute value than ours for most

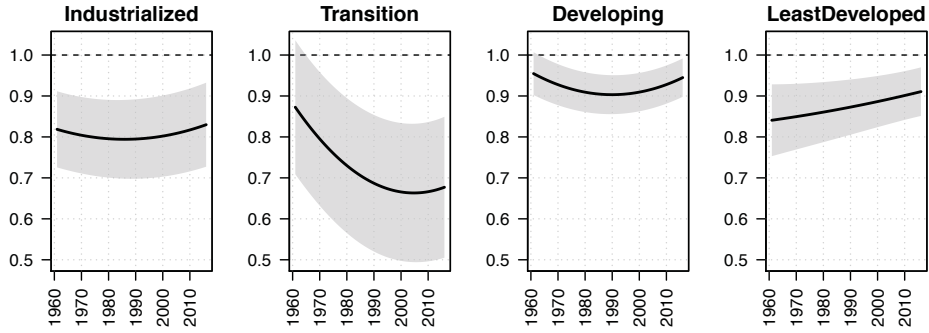


Figure 1: Time series of the estimated overall elasticity at the sample mean. Shaded areas indicate 95% confidence intervals.

Table 1: Average annual percentage variation of our and USDA's estimates of TFP change averaged by geographical region. The region 'Africa, sub-Saharan' does not include South Africa, while the region 'Oceania' does not include Australia and New Zealand.

1961–1975						1991–2005					
Region	USDA	TFPC	TC	EC	SC	Region	USDA	TFPC	TC	EC	SC
America, north	+0.56	+1.36	+1.35	+0.17	-0.16	America, north	+2.39	+1.48	+1.51	+0.08	-0.11
America, central	+1.05	-0.02	+1.48	-1.31	-0.17	America, central	+0.93	+0.58	+0.96	-0.29	-0.08
America, south	+1.11	+0.22	+1.26	-0.90	-0.13	America, south	+0.67	+1.11	+1.05	+0.21	-0.14
Europe, north	-0.20	+0.65	+1.38	-0.74	+0.02	Europe, north	+1.07	+0.82	+0.59	+0.14	+0.10
Europe, west	+1.14	+0.74	+1.06	-0.07	-0.26	Europe, west	+0.99	+0.81	+0.66	+0.01	+0.15
Europe, south	+1.13	+0.26	+1.34	-0.81	-0.26	Europe, south	+1.33	+0.74	+0.43	+0.27	+0.05
Europe, east	-0.63	-1.10	+0.93	-0.83	-1.18	Europe, east	+0.56	+2.24	+0.93	-0.07	+1.37
Asia, west	+0.07	+0.06	+1.00	-0.72	-0.21	Asia, west	+2.79	+0.69	+0.97	-0.20	-0.07
Asia, central	+0.10	+0.16	+1.14	-0.37	-0.59	Asia, central	+1.58	+1.36	+1.29	-0.20	+0.27
Asia, east	+1.04	+0.10	+0.71	-0.23	-0.38	Asia, east	+0.90	+0.46	+0.91	-0.27	-0.18
Africa, north	+2.92	+0.48	+1.15	-0.45	-0.21	Africa, north	+0.77	+1.49	+0.98	+0.61	-0.11
Africa, sub Saharan	-0.41	+0.05	+0.81	-0.61	-0.14	Africa, sub Saharan	+0.46	+0.61	+1.15	-0.42	-0.10
South Africa	-0.65	+0.32	+0.89	-0.34	-0.22	South Africa	+2.95	+1.62	+1.33	+0.25	+0.04
Australia-New Zealand	+1.18	+1.02	+0.98	+0.40	-0.36	Australia-New Zealand	+1.77	+1.06	+1.49	-0.14	-0.27
Oceania	-1.17	-0.54	+0.99	-1.54	+0.03	Oceania	+0.27	-0.90	+0.54	-1.32	-0.11
1976–1990						2006–2016					
Region	USDA	TFPC	TC	EC	SC	Region	USDA	TFPC	TC	EC	SC
America, north	+1.42	+1.55	+1.41	+0.13	+0.01	America, north	+1.18	+1.73	+1.59	+0.05	+0.09
America, central	+0.24	+0.24	+1.16	-0.80	-0.11	America, central	+1.03	+0.96	+0.82	+0.15	-0.02
America, south	+1.17	+0.58	+1.11	-0.35	-0.17	America, south	+2.02	+1.65	+0.98	+0.70	-0.03
Europe, north	+1.19	+0.88	+1.06	-0.31	+0.13	Europe, north	+1.82	+0.74	+0.17	+0.52	+0.05
Europe, west	+1.73	+0.86	+0.91	-0.03	-0.02	Europe, west	+1.47	+0.47	+0.37	+0.04	+0.06
Europe, south	+1.97	+0.36	+0.93	-0.27	-0.29	Europe, south	+0.71	+0.62	-0.20	+0.75	+0.08
Europe, east	+0.36	-0.35	+0.39	-0.45	-0.28	Europe, east	+1.95	+1.70	+1.52	+0.27	-0.09
Asia, west	+2.34	+0.08	+0.99	-0.47	-0.43	Asia, west	-0.07	+0.82	+0.96	+0.04	-0.17
Asia, central	+0.30	+0.46	+1.02	-0.29	-0.26	Asia, central	+1.18	+0.77	+1.41	-0.12	-0.51
Asia, east	+1.15	+0.33	+0.87	-0.26	-0.27	Asia, east	+1.32	+0.60	+1.08	-0.29	-0.21
Africa, north	+1.94	+0.94	+1.06	+0.07	-0.19	Africa, north	+1.66	+2.11	+1.07	+1.09	-0.05
Africa, sub Saharan	+0.11	+0.32	+0.95	-0.52	-0.10	Africa, sub Saharan	+0.21	+0.80	+1.30	-0.33	-0.16
South Africa	+2.11	+1.01	+1.10	-0.04	-0.05	South Africa	+1.55	+1.99	+1.45	+0.51	+0.02
Australia-New Zealand	+0.95	+1.29	+1.16	+0.13	+0.00	Australia-New Zealand	+0.53	+1.13	+1.60	-0.38	-0.09
Oceania	-0.90	-0.86	+0.81	-1.43	-0.22	Oceania	-1.36	-0.89	+0.43	-1.22	-0.09

geographical regions and periods. Exceptions include North America, Sub-Saharan Africa, East Europe, Central Asia, North Africa and Australia-New Zealand, where our estimates are greater in absolute value than USDA ones, or even discordant, for at least half the periods shown in Table 1. From TFP changes at country level, we note that our and USDA's estimates are in substantial agreement for United States, France, United Kingdom, Australia, South Africa and India, while USDA estimates are very higher than ours for Germany, Italy, Japan, China and Brazil, and moderately higher for Russian Federation and former Yugoslavia. Instead, our estimates are fairly higher than USDA ones for Canada, Afghanistan and Somalia.

The difference between our and USDA's estimates may be due to the presence of techni-

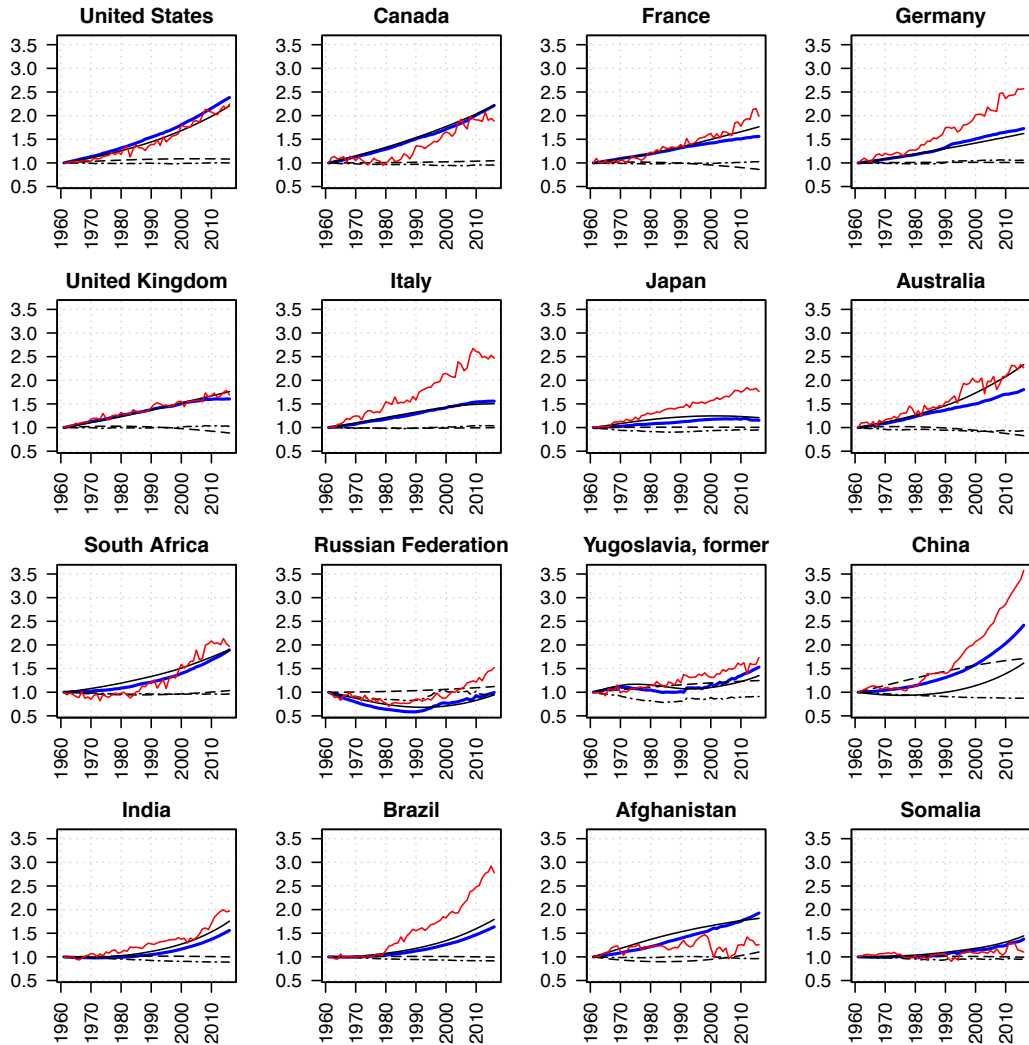


Figure 2: Our and USDA's estimates of TFP change for a selection of countries (indices, 1961=1). The time series of TFPC is shown in blue (TC, EC and SC denoted respectively by straight, dashed and dash-dotted black lines), while USDA estimates are shown in red.

cal inefficiency, that can be taken into account only by stochastic frontier models, but also to inaccuracies in USDA's input cost shares and/or in our model specification. Furthermore, our model is able to detect changes in input use through the term SC, which appears generally non-negligible, coherently with the evidence found in favour of decreasing returns to scale.

To provide an overall assessment on the agreement between our and USDA's estimates, we computed the Person correlation by country and found a median equal to 0.857, with first and third quartile equal to 0.554 and 0.943, respectively. These correlations emphasize that our and USDA's methodology provide different results but in substantial agreement, thus confirming the different theoretical foundations and suggesting the empirical validity of both of them. Full results are available at <https://github.com/alessandromagrini/agrTFP>.

## 4. Concluding remarks

We have estimated agricultural TFP change at country level based on the same data employed by the United States Department of Agriculture (USDA) using a stochastic frontier model instead of the growth accounting method. This work has the value to provide, for the first time in the literature, a comparison between agricultural TFP changes estimated with different methodologies, and an additional data source that can be employed in a large variety of longitudinal economic analyses at country level.

Our methodology overcomes the limitation of USDA estimates which rely on approximated and imputed input cost shares, and of the growth accounting method in general, which ignores technical inefficiency. However, the accuracy of estimates based on a stochastic frontier are sensitive to model specification. For this reason, we employed a more flexible specification than those adopted in the literature, but, since it is based on deterministic trends, it may be inadequate for long periods, like the one considered in this study. We also paid attention to account for heterogeneity in technology among the various countries by specifying four separate models based on the level of development.

In the future, we plan to improve our methodology by introducing autoregressive coefficients to represent stochastic trends, and by specifying latent classes to account for heterogeneity in technology among the various countries.

## References

- Battese, G.E., Coelli, T.J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data *Empirical Economics*, **20**: 325-332.
- Charnes, A., Cooper, W.W., Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, **2**(6): 429-444.
- Caves, D.W., Christensen, L.R., Diewert, W.E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica*, **50**: 1393-1414.
- Coelli, T., Henningsen, A. (2020). frontier: stochastic frontier analysis. R package version 1.1-8. <https://CRAN.R-Project.org/package=frontier>
- Fuglie, K.O. (2015). Accounting for growth in global agriculture. *Bio-based and Applied Economics*, **4**(3): 221-254.
- Kryszak, L., Swierczynska, K., Staniszewski, J. (2021). Measuring total factor productivity in agriculture: A bibliometric review. *International Journal of Emerging Markets*, **16**(5).
- Laureti, T. (2006). L'efficienza rispetto alla frontiera delle possibilità produttive: Modelli teorici ed analisi empiriche. Firenze University Press, Florence, IT.
- Magrini, A. (2021). A stochastic frontier model to assess agricultural eco-efficiency of European countries in 1990–2019. *International Journal of Statistics and Probability*, **10**(4): 138-156.
- Orea, L. (2002). Parametric decomposition of a generalized Malmquist productivity index. *Journal of Productivity Analysis*, **18**: 5-22.
- Schmidt, P., Sickles, R.C. (1984). Production frontiers and panel data. *Journal of Business & Economic Statistics*, **2**(4): 299-326.
- United Nations (2020). World economic situation and prospects (WESP 2020). Department of Economic and Social Affairs, United Nations. <https://www.un.org/development/desa/publications/wesp-2020.html>
- USDA (2019). International agricultural productivity. Economic Research Service (ERS), United States Department of Agriculture (USDA). <https://www.ers.usda.gov/data-products/international-agricultural-productivity/>