WP/19/179

IMF Working Paper

Crop Selection and International Differences in Aggregate Agricultural Productivity

by Jorge A. Alvarez and Claudia N. Berg

INTERNATIONAL MONETARY FUND

WP/19/179

IMF Working Paper

Crop Selection and International Differences in Aggregate Agricultural Productivity

by Jorge A. Alvarez and Claudia N. Berg

I N T E R N A T I O N A L   M O N E T A R Y   F U N D

 WP/19/179

IMF Working Paper

Research Department

Crop Selection and International Differences in Aggregate Agricultural Productivity*

Prepared by Jorge A. Alvarez and Claudia N. Berg

Authorized for distribution by Chris Papageorgiou

August 2019

Abstract

A large share of cross-country differences in productivity is explained by differences in agricultural productivity. Using a combination of sub-national agricultural statistics and geospatial datasets on crop-specific potential yields, we study the main drivers of this variation from a macroeconomic perspective. We find that differences in geographically-induced crop-specific comparative advantages can explain a substantial share of the variation in yields across the world. Data reveal substantial gaps between potential and observed yields in most countries. When decomposing these within country gaps, we find that crop selection gaps are on average larger than those induced by input usage alone. The results highlight the importance of understanding the interaction of geography and crop selection drivers in assessing aggregate agricultural productivity differences.

JEL Classification Numbers: O11, O13, O18

Keywords: Productivity, Agriculture, Geography

Author's E-Mail Address: JAlvarez@imf.org; CBerg@imf.org

* The authors thank Chris Papageorgiou, Davide Fuceri, Doug Gollin, Brian Blankespoor, Siobhan Murray, and Richard Rogerson, for helpful discussions.

Contents

Tables

Figures

1. Introduction

Agriculture is a source of livelihoods for millions,[1] and thus remains a key component of global welfare, estimated to account for 29.5 percent of global employment and 68.5 percent of employment in low-income countries.[2] Moreover, differences in agricultural productivity account for a large share of aggregate productivity variation across countries: while the GDP per worker of the richest 5 percent of countries is 34 times that of the bottom 5 percent, this same ratio is 78 in the agricultural sector.[3] The prevalence of low agricultural productivity in low income countries has motivated both ambitious policy agendas[4] as well as macro and microeconomic research on the sources of this variation. These efforts have been challenged by the methodological difficulty of accounting for geographical differences accurately, particularly when the subject of interest goes beyond a narrowly defined geographic area. By using exceptionally rich crop-specific global datasets, this paper provides a measure of geographically attainable potential yields and productivity gaps. Moreover, a decomposition of these gaps is presented to shed light on what mechanisms are most quantitatively relevant in explaining global variation in aggregate agricultural productivity.

Explanations behind large agricultural productivity gaps between rich and poor countries have been extensively studied by the development literature. Among these, differences in input usage and technology adoption (fertilizers, pesticides, seeds, machinery, etc.) have been widely highlighted as important sources low yields in low-income countries,[5] and several cross-sectional studies and experiments have been conducted to explore this channel. Duflo, Kremer, and Robinson (2008, 2011), for example, find that low intermediate input usage – and fertilizer in particular – results in low yields for maize farmers in Kenya. Beaman, Karlan, Thuysbaert,

[1] Close to 800 million (or 78 percent) of the world's poor live in rural areas and rely on agriculture for their livelihoods (World Bank 2014). In addition, evidence suggests that growth in agriculture has proven to be more effective at reducing poverty as compared to growth in elsewhere (World Bank 2008).
[2] 2015 figures from ILOSTAT.
[3] There is a large literature emphasizing agricultural productivity differences dating back to Kuznets (1971) and more recently emphasized by Caselli (2005), Restuccia, Yang, and Zhu (2008), Gollin, Lagakos, and Waugh (2014), among others.
[4] For instance, the World Bank Group made $8.3 billion new commitments to agriculture in 2014, the majority of which went towards increasing productivity, food security, and access to markets (World Bank 2014).
[5] For an early and brief summary, see Feder et al. (1985).

and Udry (2013) find similar large potential gains from increasing fertilizer usage for rice farmers in Mali. Other possible explanations that have received less attention in the literature are related to crop selection; that is, factors hampering the choice of optimal crops given land suitability. For instance, Asad (2014) finds that cell phone coverage in rural Pakistan helped correct for a coordination failure between farmers and merchants, thereby decreasing the risk of post-harvest losses and leading farmers to switch to more perishable and lucrative crops. Allen and Atkin (2016) find that falling trade costs in India integrated markets, thereby increasing farmers' revenue volatility if they continued to grow the same crops, and so caused farmers to shift production to crops with less risk. Emran et al. (2012) identify an inverse-U causal relationship between the size of the local market and the pattern of crop specialization within the village economy in Nepal. Micro studies, however, are often limited to a single crop or to a specific region and country. Although they provide a high level of detail into relevant mechanisms at work, the quantitative relevance of these mechanisms at a macroeconomic (and global) level is not entirely clear.

The macroeconomic relevance of these gaps has recently gotten renewed attention. Large gaps have been documented and studied across countries by Gollin, Lagakos and Waugh (2014), Herrendorf and Scholleman (2015), among others, and models explaining gaps between agriculture and other sectors have been developed (Alvarez 2019; Herrendorf and Schoellman 2018; Donovan 2016; Adamopoulos and Restuccia 2014; Lagakos and Waugh 2013; Restuccia et al. 2008; Tombe 2015). These models explain differences in productivity with mechanisms that induce low aggregate investment in intermediate inputs or low quality of labor. While successful in explaining much of the variations, these varying mechanisms do not focus on a careful accounting of the role geographic differences.

This paper attempts to identify quantitatively significant sources of aggregate agricultural productivity variation by using a set of global spatial datasets that account for geographical differences. The closest paper to our work is Adamopoulos and Restuccia (2018), who use similar geospatial data on modeled and potential agricultural production to assess the extent to which geography and land quality account for average yield differences across countries with different income levels. They document that, if countries produced current crops according to

potential yields (by using high inputs), average gaps between the group of richest and poorest countries would nearly disappear. Using different datasets on observed production and prices and a more granular approach,[6] we are able to replicate this result and extend it to explore the variation in within-country yield gaps conditional and unconditional on income levels.

We present three main sets of findings. First, we find that geographical variation plays a substantial role in explaining cross-regional variation in aggregate yields, but substantial gaps remain even after controlling for crop- and region- specific measures of potential. Second, when decomposing the within-country gap between observed and potential production, we find that differences in crop selection are most important, with gaps in inputs playing a lesser role. Third, we show how measured gaps and yields vary systematically across income levels as well as countries with different infrastructure and input availability. Overall, our results highlight the importance of understanding crop selection decisions, given geographical constraints, in explaining aggregate productivity gaps in agriculture.

The rest of this paper is organized as follows. Section 2 presents the data sources and variable construction. Section 3 presents our empirical framework. Section 4 discusses our main findings and documents how our measures of yields and gaps correlate with measures of income, inputs, technology and infrastructure. Section 5 concludes.

2. Data description

The database used in this analysis is constructed from several sources. Information on the harvested land area and observed yield are from Monfreda et al. (2008), henceforth "M3." The M3 data harmonizes national, state, and country-level census statistics and disaggregates them to 5 arc-minute resolution (approximately 10 x 10 km) using a global dataset of cropland from

[6] We construct our database with production estimates from Monfreda et al (2008) and crop potential from the Global Agro-Ecological Zones (GAEZ) project of the Food and Agriculture Organization (FAO). The dataset is at the 5-arc minute resolution, or approximately 10 x 10 km. Unlike Adamopoulos and Restuccia (2018), who relied on international prices and focused on the country level, we take a more granular approach. Specifically, we focus our analysis at the state level and use country-level crop prices from the FAO to calculate crop values. Our main analysis focuses on 2,423 states in 127 countries for which we had price data.

Ramankutty et al. (2008). Data are circa 2000, the average of censuses from 1997-2003. The full data set contains information on 175 crops. We focus our analysis on the 25 crops listed in Table 1. We chose these 25 crops as they are among the main global cash crops. Our selection was guided in part by the M3 database's list of 17 "major crops."[7] We supplemented this list with additional, relatively more perishable commercial crops[8] to arrive at a more balanced selection of high (7), medium (10), and low (8) perishable crops.[9] Our motivation in using this M3 data is that it applies minimal modeling to distribute subnational statistics of yield and harvested area. As such, the downscaling process used was intuitive and transparent. By aggregating this data to the sub-national level, we are able to focus on the less distorted information from which this it was originally constructed.[10]

For data on the potential yield, we turn to the Global Agro-Ecological Zones (GAEZ) data developed by the International Institute for Applied Systems Analysis (IIASA) and the Food and Agriculture Organization (FAO). This dataset compiles climatic, soil, and land cover data to estimate potential production under alternative levels of inputs (low, medium, high). Climatic data includes precipitation, temperature, wind speed, sunshine hours, and relative humidity. Soil data comes from the Harmonized World Soil Database (2009) which contains information on soil nutrient availability, nutrient retention capacity, oxygen availability, toxicity, salinity, and sodicity. Elevation and slope are from the Shuttle Radar Topography Mission (2006). The GAEZ project combines these and estimates the potential yield of different crops within each 5 arc-minute cell of the globe. GAEZ computes attainable yields, given geographic factors, under three different sets of input assumptions (high, medium, low) under both rainfed and irrigated agriculture. High-input assumes advanced management, commercial farming. Production is based on improved or high-yielding seed varieties, is fully mechanized with low labor intensity and uses optimum applications of nutrients and chemical pest, disease and weed control.

[7] These are: barley, cassava, cotton, groundnut, maize, millet, oil palm, potato, rapeseed, rice, rye, sorghum, soybean, sugar beet, sugarcane, sunflower, and wheat.
[8] These are banana, cabbage, carrot, citrus, onion, sweet potato, tomato, and yam.
[9] The perishability rankings used in this paper are guided by field tests in Pakistan described in Asad (2014) and supplemented by authors' judgement.
[10] See Anderson et al. (2014) for comparison and contrast between these modeled production data.

Table 1. List of Crops included in our study

	Crop	Perishability Ranking*	Number of Countries where crop is grown	Harvested Area (1,000 Ha)
1	Banana	High	80	3,536
2	Barley	Low	91	50,788
3	Cabbage	High	109	2,593
4	Carrot	High	102	915
5	Cassava	Medium	61	11,014
6	Citrus	High	70	798
7	Cotton	Low	76	25,813
8	Groundnut	Low	84	18,957
9	Maize	Medium	108	122,232
10	Millet	Low	76	31,398
11	Oil palm	High	33	8,903
12	Onion	Medium	112	2,498
13	Potato	Medium	106	17,937
14	Rapeseed	Medium	69	23,875
15	Rice	Low	91	137,460
16	Rye	Low	60	9,197
17	Sorghum	Low	96	36,026
18	Soybean	Medium	95	72,702
19	Sugar beet	Medium	61	5,865
20	Sugarcane	High	70	17,387
21	Sunflower	Medium	77	18,797
22	Sweet potato	Medium	81	7,991
23	Tomato	High	116	3,212
24	Wheat	Low	98	197,657
25	Yam	Medium	46	3,249

* The perishability rankings are guided by the field tests in Pakistan described in Asad (2014) and supplemented by authors' judgment.

In this analysis, we rely on high inputs, based on the 1961-1990 reference. Under these assumptions, we take potential yield to be the maximum of rainfed and irrigated amounts. Importantly, the measure of potential does not consider profit-maximization, instead capturing the achievable yields under high-input[11] assumptions and agro-ecological conditions. Thus, observed yield in our database is sometimes higher than potential.

Country-specific prices[12] for each crop are sourced from FAOSTAT. We use average crop prices for 2003-2007. Not every crop had a price in every country. To interpolate for missing prices, we assume that the ratio between crop prices are constant. To illustrate, if the price of wheat was missing for a particular country (A), we multiplied its price of maize by the ratio of the international price of wheat by the international price of maize:

$$p_A^{wheat} = p_A^{maize} \frac{p_{Intl}^{wheat}}{p_{Intl}^{maize}} \tag{1}$$

Where p_s^c is the price of crop c in country s. In general, non-perishable crops were used to fill in missing prices for other non-perishable crops. Maize has good price coverage and was used as the benchmark price, followed by potato and sweet potato. In the case of perishable crops, banana was used as the benchmark, followed by tomato, onion, and sunflower. International crop prices were taken from Wood-Sichra et al. (2016, Table 2-11).

Our geographic unit of analysis is the sub-national state level (admin 1 in GADM[13]). To aggregate from the cell to the State level, potential yields from GAEZ were converted to potential production by multiplying it by the total harvested area from the M3 data. M3 observed production was calculated by multiplying the crop-specific yield with the crop-specific harvested area. The crop-specific yield and harvest area data are constructed by Monfreda et al. (2008) using agricultural statistics from various sources and distributing it

[11] Under "high inputs," farming is mainly commercial. Production is based on improved or high yielding seed varieties; is fully mechanized with low labor intensity; and uses optimal applications of fertilizers as well as pest, disease, and weed control (FAO & IIASA 2012).

[12] Farmgate prices would have been preferable, given that these are the ones most likely to influence a farmer's crop and input choices. However, collecting such prices is only available through household surveys and are lacking for much of our study area.

[13] GADM is a vector database of global administrative boundaries.

spatially over a map of cropland from Ramankutty et al. (2008). The cropland map represents the area harvested of each crop.[14]

At the cell-level, there were several cases where positive yield was recorded (from M3) while the potential there was zero (from GAEZ). Possible explanations for such negative productivity gaps include misallocation of production and measurement error of potential. The production data we use may have been misallocated during the downscaling process used in the M3 data (see Monfreda et al. (2008)). We strove to minimize this by redistributing production values from cells with zero potential equally among those cells with positive potential within the country. We focus on the State as the geographic unit of analysis rather than at the cell level, to further minimize the effect of this measurement error.

Out of the 2,423 States we focus on, 749 have a negative gap (more observed than potential production), 1,534 have a positive gap (less observed than potential production), with the remaining 140 having zero observed or potential production.[15]

In a later section of the paper, we investigate the extent to which economic characteristics explain these agricultural gaps at the country level. To this end, we aggregate the main database to the country level and introduce several country-level variables: GDP per capita, road density, fertilizer use, use of modern technology in agriculture, average farm size, subsidies, and a measure of government effectiveness.

GDP per capita and fertilizer use are sourced from the World Bank's World Development Indicators. The average farm size within a country is taken from a 63-cross-country dataset compiled by Vollrath (2007). For the purposes of this analysis, a farm is defined as "an economic unit of agricultural production under single management." Road density (kilometer of road per square kilometer of land) is sourced from Canning (1998), 2007 update. We rely

[14] Given that some crops are harvested several times a year, total harvested area can exceed the physical area of the plot it is grown on.

[15] Exceeding one's potential is not necessarily a good thing if it is driven by market distorting policies. For example, Damania et al (2017) recount how farmers growing thirsty crops such as sugarcane and rice thanks to subsidized irrigation in the desert are worse off overall when a drought hits and irrigation is no longer available.

on paved roads in 2005, the most recent year available in the data. As an indicator of modern technology use in agriculture, we rely on the number of tractors from the Cross-country Historical Adoption of Technology (CHAT) database, see Comin and Hobijn (2010).

3. Empirical Framework

We are interested in decomposing the gap between observed and potential yields into that which is due to input choice and that which is due to crop selection. To that end, we calculate three yield measures: (i) average observed yield within a state, (ii) 'mid-potential' yield obtained from using high inputs and holding crop choice fixed, and (iii) overall potential yield obtained from using high inputs and growing the crop with the highest value using all available cropland.

Observed yield value within a subnational state (y_s) is calculated as follows:

$$y_s = \frac{1}{L_s} \sum_{i=1}^{N} \sum_{c=1}^{C} p^c q_i^c \tag{2}$$

Where L_s is the total harvested land in state *s*, p^c is the price of crop *c* in the country where *s* is located (country subscripts suppressed), and q_i^c is the production of crop *c* in grid cell *i*.

Mid-potential yield holding crop choice fixed (y_s^{mid}) is calculated as follows:

$$y_s^{mid} = \sum_{i=1}^{N} \sum_{c=1}^{C} \omega_i^c p^c \, \overline{q_i^c} \tag{2}$$

Where $\bar{q}_i^c$ is the potential production of crop c in grid cell i if high inputs are used and ω_i^c is the share of harvested area allocated to crop c (i.e. $\omega_i^c = l_i^c / L_i$, where l_i^c is harvested land of crop c and $L_i = \sum_c l_i^c$). The difference between y_s and y_s^{mid} are attributed by the differences between crop-specific observed yields and the corresponding high-input crop-specific potential.

Moreover, overall potential yield is calculated as:

$$y_s^{pot} = \frac{1}{L_s} \sum_{i=1}^{N} \max_c \{p^c \bar{q}_i^c\} \tag{3}$$

Where L_s is the total harvested land within state *s*. This is a measure of revenue maximizing potential across all crops, which is not affected by any choice on the ground and is solely driven by geographical differences between regions.

To decompose the agricultural productivity gap, note that:

$$y_s^{pot} - y_s = \left(y_s^{pot} - y_s^{mid}\right) + \left(y_s^{mid} - y_s\right)$$

The first component of the gap is the gap is due to crop selection while the second is the gap due to gaps between the high input assumption of the database and inputs used. In what follows, we express the productivity gaps as a percentage of potential production. That is:

$$gap = 100 \times \frac{y_s^{pot} - y_s}{y_s^{pot}}$$

$$gap_{crop} = 100 \times \frac{y_s^{pot} - y_s^{mid}}{y_s^{pot}}$$

$$gap_{in} = 100 \times \frac{y_s^{mid} - y_s}{y_s^{pot}}$$

Where gap, gap_{crop}, and gap_{in} are the total gap, gap due to crop choice, and gap due to inputs, respectively.

In a later section of this paper, we aggregate our data to the country level and document the correlation of yields and gaps with several economic characteristics. We report regressions of the form:

$$\ln(gap_k) = \gamma \ln(X_k) + \theta_r + \varepsilon_k$$

where gap_k is the gap for country k and X_k is a vector of available cross-country measures that are informative of mechanisms highlighted by the literature on agricultural productivity. In particular, we include GDP per capita as a measure of a country's level of income; road

density as an indicator of a country's connectivity and infrastructure development;[16] the use of fertilizer as a proxy for key inputs determining agricultural production;[17] tractor use as an indicator of machinery use in agriculture[18] and average farm size as a scale of production indicator.[19] We also include regional fixed effects (θ_r) for East Asia and Pacific; Europe and Central Asia; Latin American and the Caribbean; Middle East and North Africa; North America; South Asia; and Sub-Saharan Africa.

4. Results

We now turn to the spatial data to analyze the roles played by geography, crop selection, and input choice in determining agricultural yield gaps between and within countries.

The role of geography in crop-specific yield potentials

Not all land is created equal. Not only does yield potential for each crop vary significantly across space, but geographical patterns also vary significantly by crop. A clear illustration of this pattern is shown in Figure 1, which depicts potential yields of wheat and sugarcane, at the 5 arc-minute resolution across the world. As expected, potential yields vary significantly for each of these crops when comparing different of latitudes of the planet. What is most striking, however, is that these maps are nearly mirror images of each other. While wheat grows well in the temperate regions of Europe and North America, sugarcane favors the tropical areas in South America, Africa and Asia. In this extreme example, the potential yields of these crops

[16] There is evidence that better transport networks facilitate farmers' access to agricultural inputs and to markets where they can sell their produce (see for example, Damania et al 2016). Furthermore, transportation costs tend to be higher for perishable crops, particularly over longer distances (given the need for coordinating harvesting, refrigeration). Because of these hurdles, trade in perishable goods is often restricted to local markets. This interaction with perishability is important as perishable crops carry a higher risk for farmers. The basic argument is that if a farmer cannot sell their produce quickly enough, post-harvest losses can potentially be quite high (Asad 2014, Kaminski and Christiaensen 2014, FAO 2014, IFPRI 2013, Grolleaud 2002, and Oehmke 1992).

[17] This is motivated by the literature highlighting low usage of fertilizer and high-yielding seeds as the main explanation behind large agricultural productivity gaps. (See for example, Beaman et al 2013, Comin and Hobijn 2010, Feder et al 1985.)

[18] Tractors and other heavy machinery are far more efficient than manual labor. Tractor use has been linked to increased returns to scale in agriculture (Takeshima, Houssou, and Diao 2018).

[19] Theory and related literature suggest that land distribution is important in explaining the variation in agricultural productivity. Notably, the average farm size of the richest 20 percent of countries is 34 times that of the poorest 20% (Restuccia 2011).

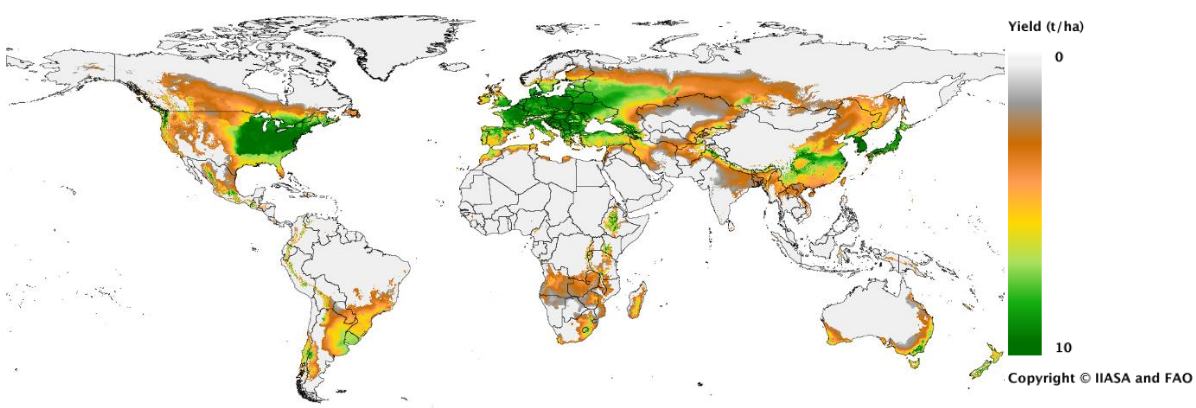
are negatively correlated across space. Since every crop has its own unique geographical pattern, there are multiple examples where potential yields across crops are far from being perfectly correlated. This lack of synchronicity in geographic variation of crop-specific yields underscores the importance of crop-selection optimization in fully exploiting geographical comparative advantages.

The role of geography in cross-country yield variation

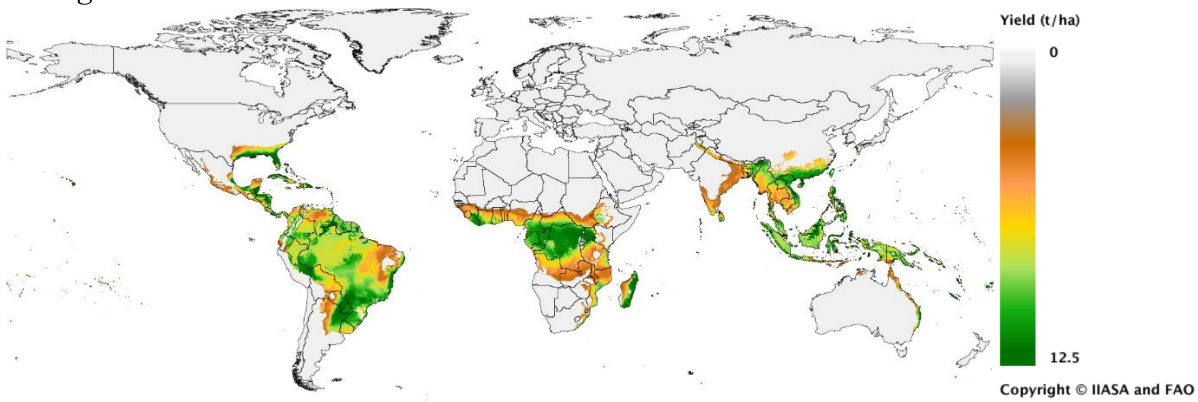
Aggregating across crops, there is substantial geographical variation in yields across the world. In particular, the magnitude of variation in our geography-driven measures of potential is not much smaller than the variation in observed yields. Figure 2 plots the distribution of observed yields (blue), mid-potential yield using high inputs while holding crops fixed (green), and potential yield under both optimal crops and high inputs (red) for all sub-national regions in the sample. Unsurprisingly, the overall distribution of potential yield is to the right of the other two measures. The exception is at the upper end of the distribution, where observed yield exceeds potential. This is explained by the use of technologies that go beyond the high-input rainfed assumptions behind our potential measures. Perhaps more surprisingly, the variation in potential yields (variance of 0.99) is of the same order of magnitude of that of observed yields (1.13). Recall that variation in geography is the sole driver of potential yield variation. Thus, this high variance emphasizes the importance of geography in shaping observed yield variation across the globe.

Figure 1: Geographic distribution of rainfed wheat and sugarcane potential yield

1a. Wheat



1b. Sugarcane



Source: IIASA and GAEZ
Notes: Maps depict agro-climatically attainable yield for high-input, rainfed crops, under baseline (1961-1990) assumptions.

Figure 2: Distributions of Observed, Mid-Potential, and Overall Potential Yields, sub-national level

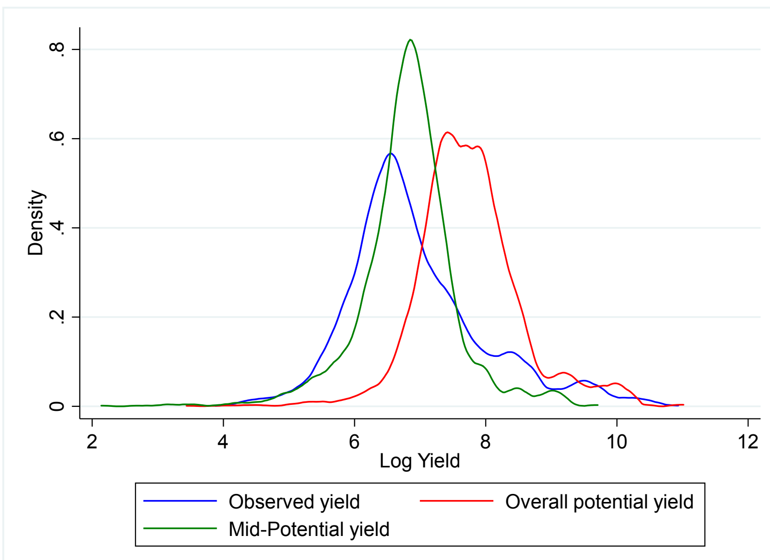
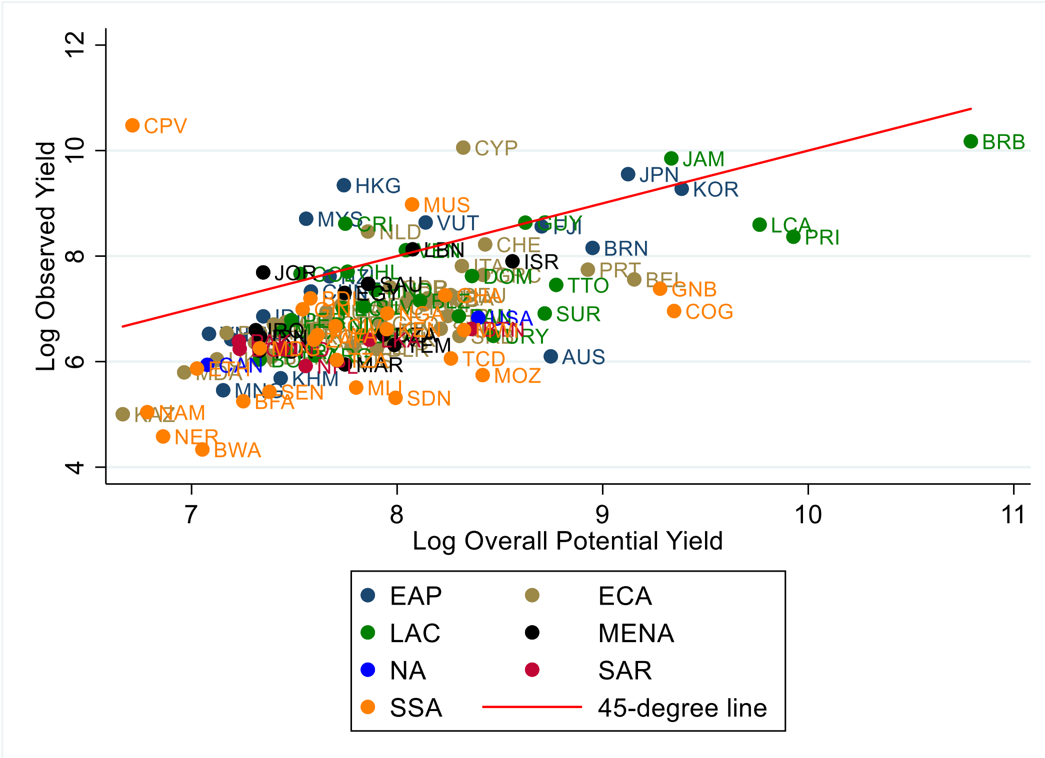


Figure 3: Observed vs. Potential Yield at the Country Level



Note: EAP =East Asia & Pacific, ECA = Europe & Central Asia, LAC = Latin America & the Caribbean, MENA = Middle East & North Africa, NA = North America, SAR = South Asia, SSA= Sub-Saharan Africa.

Figure 3 compares observed and potential yields, at the country-level, for all countries in the sample. Several features are worth noting in this cross-country comparison. First, mirroring the sub-national pattern, variation in overall potential yields at the country level is significant with a variance of 0.73, compared with a variance of 1.25 in observed yields. The high variance of our measure of potential yield—which does not take into account any non-geographic information—is present within all world regions: Sub-Saharan Africa has a variance in log potential of 0.51 log points, Latin America and the Caribbean of 0.58 log points, Middle East & North Africa of 0.95 log points, North America of 1.22, East Asia & Pacific of 0.75, and Europe & Central Asia of 0.44 log points.

Second, observed and potential yields are positively correlated. The correlation between the two measures (in logs) is 0.35, with potential yield explaining 13 percent of the cross-sectional

variation in observed yields. Table 3 shows the share of variation in country-level yields explained by potential yields (R^2) within geographical regions. The relationship is weakest in the Middle East and North Africa, an area dominated by deserts, and relationships in other regions are affect by the presence of small islands and states,[20] where irrigation and other technological peculiarities are likely more pronounced. For instance, the variation in observed yield that can be explained by potential yield is also relatively weak in Europe and Central Asia ($R^2 = 0.08$), but this becomes stronger ($R^2 = 0.32$) once Cyprus is excluded. Similarly, the share of observed yield variation in Sub-Saharan Africa that is explained by potential yields is 0.06, which rises to 0.31 once Cape Verde is excluded. The explanatory power of potential yield is higher in the East Asia and Pacific ($R^2 = 0.46$), and Latin America and Caribbean ($R^2 = 0.46$) regions. The high variation in potentials and their positive correlation with yields highlight the importance of incorporating geography in the analysis of country-level agricultural productivity dispersion.

The role of crop selection in cross-country variation

In addition to potential and observed yield variation, Figure 2 also illustrates the variation in yields under the counterfactual scenario with high-input usage across all regions and crops, thus only allowing crop-selection to vary across space. The mid-potential yield under fixed crops selection has a reduced variance of 0.88, compared to the variance of 0.99 and 1.13 in potential and observed yields, respectively. This suggests that since countries respond to geography-specific comparative advantages between crops, crop selection mitigates the variation in geography. However, the combined variation in crop selection and geography still fails to explain the right-hand fat-tail of the distribution of observed yields. The right-hand tail seems to be a product of high-input usage in high-yield countries.

[20] Because of local contexts, these countries tend to grow crops for which they have poor natural advantages and thus require higher inputs than are reflected in our data.

Table 2. Mean observed, yields, potential yields, and gaps

	Mean Observed Yield	Mean Mid-Potential Yield	Mean Overall Potential Yield	Mean Gap due to Crops	Mean Gap due to Inputs	Mean Overall Gap
East Asia & Pacific	3,995	1,883	3,838	0.505	0.138	0.468
Europe & Central Asia	1,722	969	2,962	0.628	0.140	0.630
North America, Latin America & the Caribbean	3,370	1,676	6,243	0.643	0.086	0.432
Middle East & North Africa	1,364	1,137	2,499	0.504	0.244	0.592
South Asia	559	928	2,211	0.546	0.170	0.716
Sub-Saharan Africa	2,310	1,048	2,981	0.579	0.228	0.762
World	2,399	1,267	3,710	0.591	0.167	0.642

Note: Observed and Potential averages are calculated based the raw values and are measured in tons per hectare.

Table 3. Share of observed and mid-potential yield variance explained by overall potential yield

	(1) Obs. Yield (y_s)	(2) Mid-Pot. Yield (y_s^{pot})
East Asia & Pacific	0.390	0.659
Europe & Central Asia	0.430	0.290
North America, Latin America & the Caribbean	0.579	0.718
Middle East & North Africa	0.203	0.330
South Asia	0.324	0.748
Sub-Saharan Africa	0.040	0.413
World	0.321	0.525

Note: (1) R^2 calculated based on OLS regressions of observed on potential yield. (2) R^2 calculated based on OLS regressions of mid-potential on overall potential yield.

Agricultural productivity gaps

Comparing observed with potentials yields allows us to compute productivity gaps within countries. As shown in Figure 2 at the country-level, we find that yield falls short of potential for most countries as shown by points below the 45-degree line. The exception, as mentioned before, are countries dominated by deserts, and small islands and states. When looking at variation across regions, we find the largest gap to be present in Sub-Saharan African countries, with an average gap of 0.73. These are followed by East Asia and Pacific with an average gap of 0.56 and Latin America and the Caribbean with an average gap of 0.55.[21] Figure 4 shows the map of these productivity gaps for all countries available.

As described in the previous section, we can decompose these gaps into sub-components. Figure 5 illustrates the distributions of the total gap (red), relative to the gap due to crops (green), and that due to inputs (blue) as described in the previous section. Consistent with the patterns shown in Figure 2, the gap due to crop choice is still very similar to the overall gap, again suggesting that changing the crop mix is necessary to close the total yield gap. Gaps from inputs are considerably smaller and more dispersed. The log gap from inputs averages -2.09 with a variance of 1.28, compared with log crop selection gaps averaging -0.57 with a variance of 0.15, and overall gap in logs averages -0.54 with a variance of 0.49. Moreover, there are differences in how input and crop selection gaps vary in the cross-sectional countries, as shown in Figures 6 and 7. For instance, although input driven gaps appear to be larger in many Sub-Saharan African countries compared to the rest of the world, this pattern does not hold when looking at crop selection gaps. Both because of the higher means in crop selection gaps as well as the heterogeneity across different economies, the results suggest that studying crop selection differences, in addition to inputs, is required to reach a full understanding of aggregate agricultural productivity gaps across all world regions.

[21] North America has a large gap at the country level, but it only includes two countries (Canada and the USA). At the sub-regional level, the average gap is considerably smaller (54).

Figure 4: Total Agricultural Productivity Gap

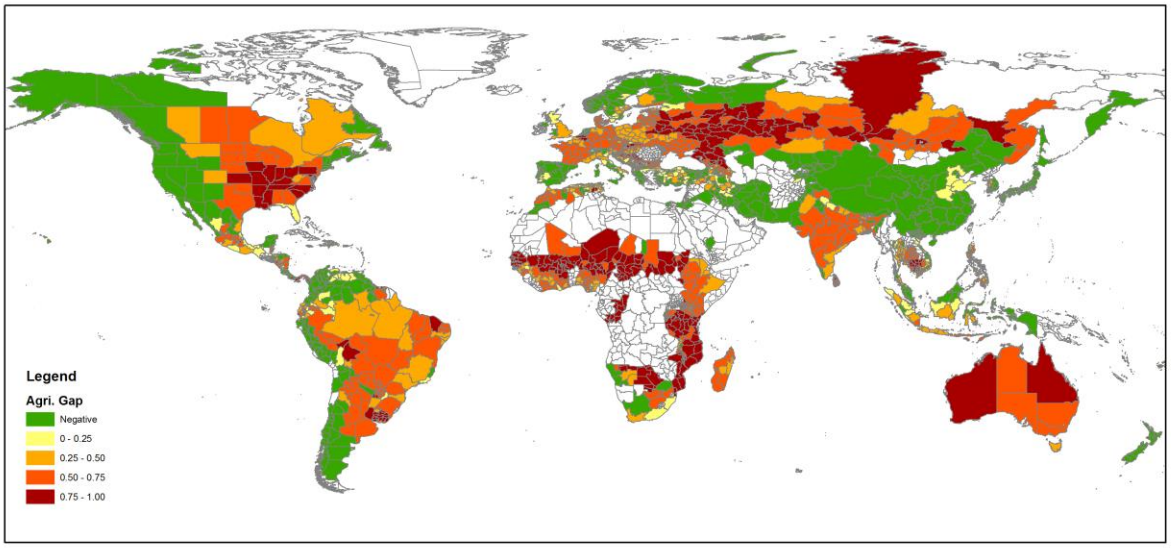


Figure 5: Total and Decomposed Agricultural Productivity Gaps

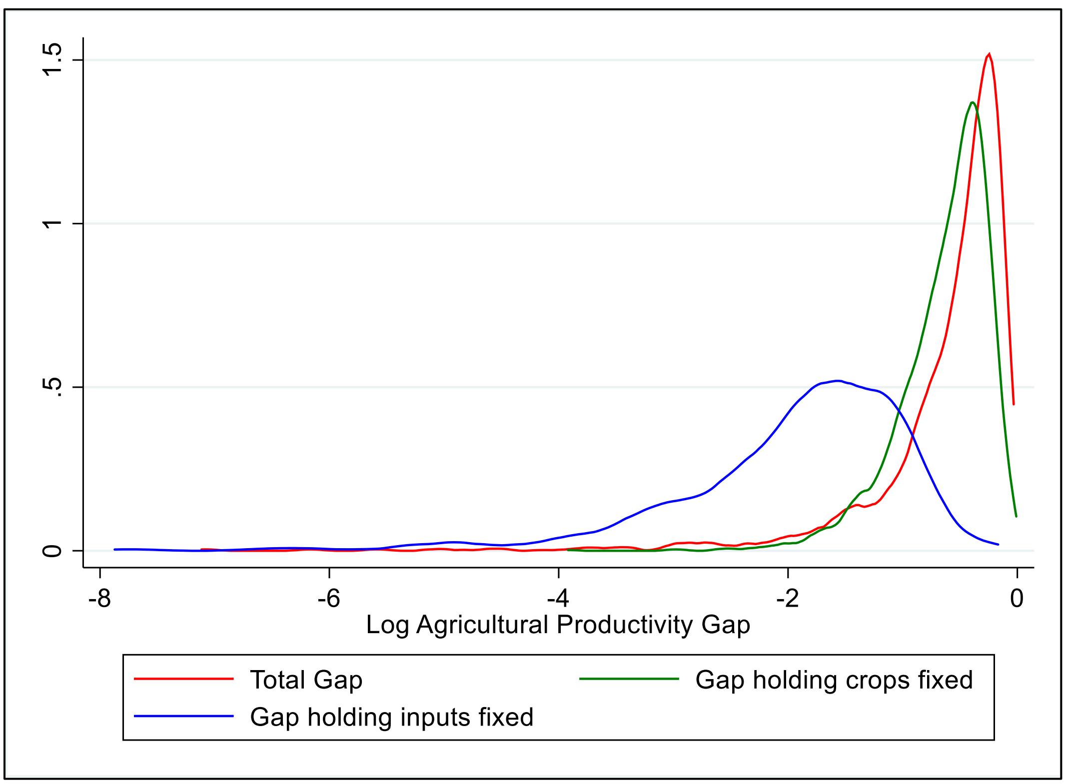


Figure 6: Agricultural Productivity Gap Due to Crop Selection by Country

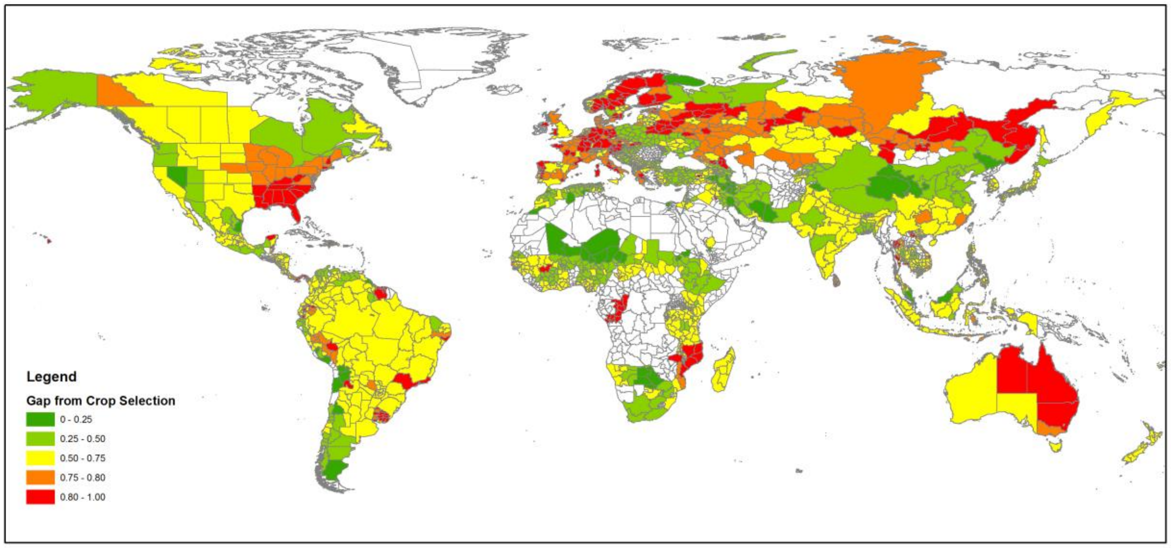
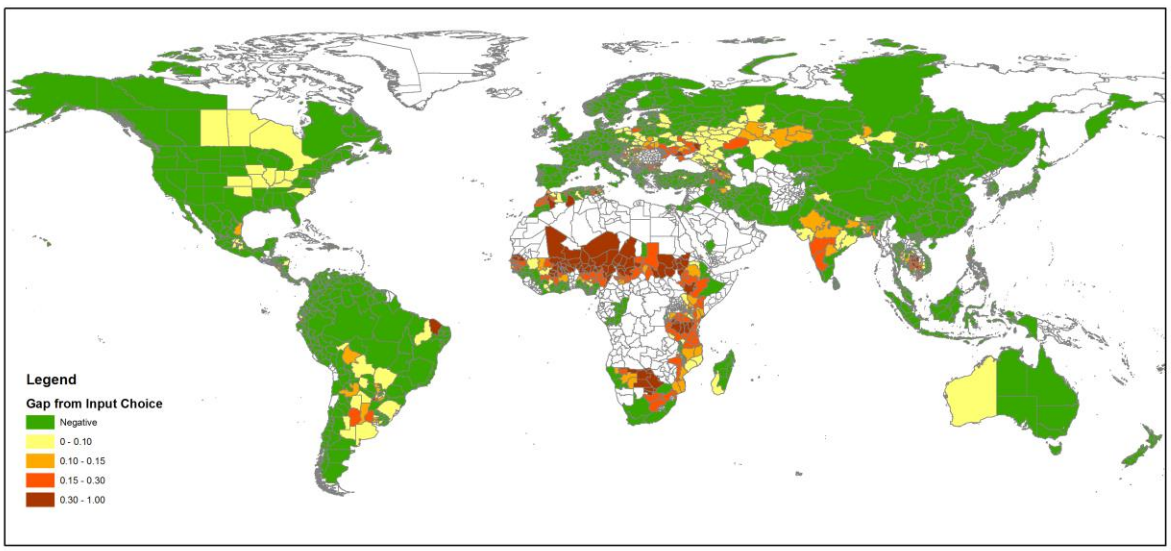


Figure 7: Agricultural Productivity Gap Due to Inputs by Country



Aggregate yields and gaps across income groups

There are large differences in aggregate observed yields between rich and poor countries, with the richest 10 percent of countries in our sample having yields that are 4.5 times larger than those of the poorest 10 percent (by GDP per capita). As documented by Adamopoulos and

Restuccia (2018), the correlation weakens significantly when considering our mid-potential measure—where cross-country variation comes exclusively from differences in crop-selection. In our sample, variation from crop-selection alone reduces the gap between the richest and poorest 10 percent of countries to from 4.5 to 1.2. Geographical variation alone, as measured by the potential measure, has a slightly higher ratio of 1.3 (see Table 4).

Moreover, Table 5 documents the correlation of GDP with within country gaps. It shows that GDP per capita is negatively correlated with total and input gaps, but not with crop selection gaps. That is, wealthier countries appear to have a geographical advantage in terms of potential yield, but their crop selection does not lead to systematically better yields, on average.

These average differences, however, mask substantial heterogeneity within the different income groups. Figure 8 shows the scatter plots correlating our yield measures with GDP per capita. Note that systematic differences in yield variance across income groups are not evident. Within each income group, variances seem to replicate the patterns documented in Figure 2. Similarly, Figure 9 plots yield gaps by income group and shows substantial variation in gaps within all income groups, with outliers among higher income countries driven by overperformance in input gaps.

Table 4. Production by Yield Measure

	Yield		
	Observed	Mid-Potential	Potential
GDP per capita			
Top 10%	2,750	1,350	3,834
Bottom 10%	610	1,138	2,874
Ratio	4.51	1.19	1.33
Road Density			
Top 10%	4,815	2,050	7,868
Bottom 10%	810	911	3,337
Ratio	5.94	2.25	2.36
Fertilizer			
Top 10%	5,468	1,784	5,079
Bottom 10%	561	697	2,676
Ratio	9.75	2.56	1.90
Farm Size			
Top 10%	2,675	1,001	2,635
Bottom 10%	1,838	1,268	4,147
Ratio	1.46	0.79	0.64
Technology			
Top 10%	7,024	2,156	7,298
Bottom 10%	1,520	1,224	4,284
Ratio	4.62	1.76	1.70

Note: Farm Size calculations exclude farm size = 0.

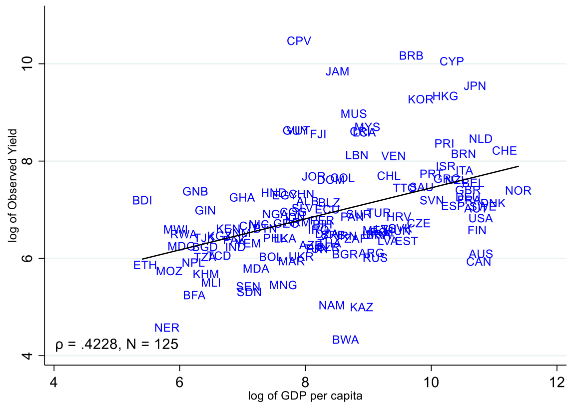
Table 5. Bivariate Regressions between our gap measures and each market characteristic

	(1) ln GDP pc	(2) ln Road Density	(3) ln Fertilizer	(4) ln Farm Size	(5) ln Tech use
A. *ln(total gap)*					
Market Characteristic	-0.046*	-0.059**	-0.080***	0.019	0.001
	(-1.80)	(-2.21)	(-2.71)	(1.43)	(0.13)
Constant	1.299***	0.765***	1.233***	0.881***	0.899***
	(6.34)	(9.64)	(11.53)	(23.45)	(7.11)
Observations	121	117	114	100	111
R-squared	0.093	0.174	0.230	0.139	0.039
B. $ln(gap_{crop})$					
Market Characteristic	0.015**	0.002	-0.000	0.003	0.000
	(2.57)	(0.37)	(-0.08)	(0.70)	(0.09)
Constant	0.820***	0.952***	0.950***	0.947***	0.944***
	(16.22)	(89.30)	(44.06)	(110.83)	(29.71)
Observations	125	121	117	102	112
R-squared	0.181	0.109	0.107	0.221	0.117
C. $ln(gap_{in})$					
Market Characteristic	-0.094**	-0.098**	-0.129**	0.030	0.007
	(-2.33)	(-2.30)	(-2.56)	(1.29)	(0.44)
Constant	1.403***	0.377***	1.138***	0.568***	0.553***
	(4.40)	(2.93)	(6.27)	(8.16)	(2.83)
Observations	121	117	114	100	111
R-squared	0.102	0.163	0.210	0.121	0.030

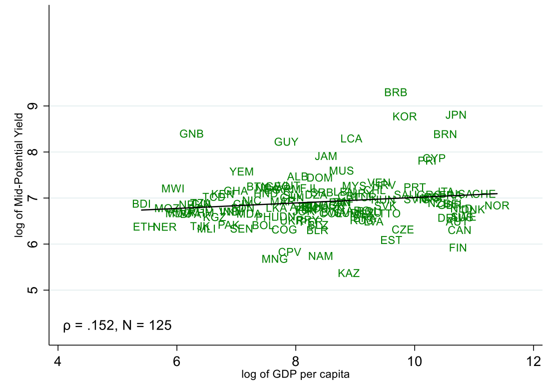
Note: Regional dummies are included. Robust t-statistics in parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%.

Figure 8. Correlations between yield measures and GDP per capita across countries

8a. Observed Yield



8b. Mid-potential yield



8c. Potential Yield

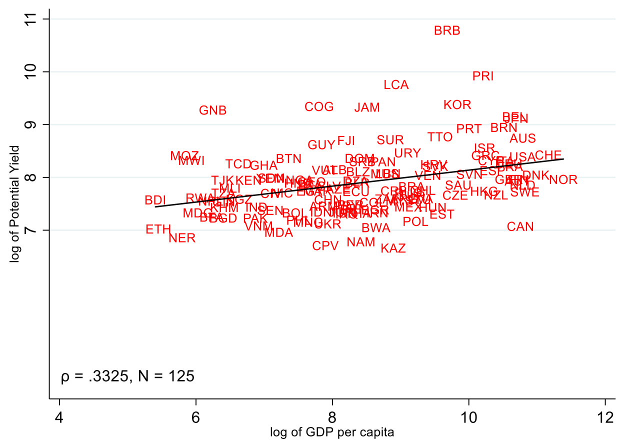
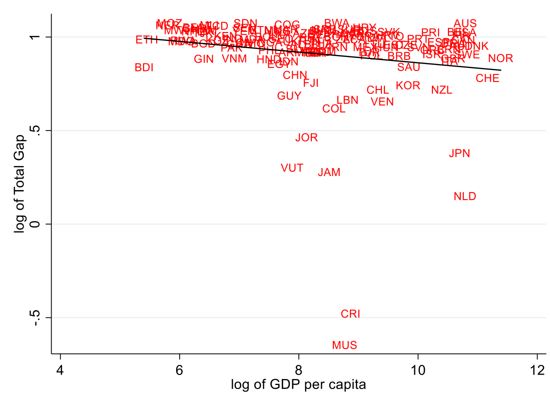
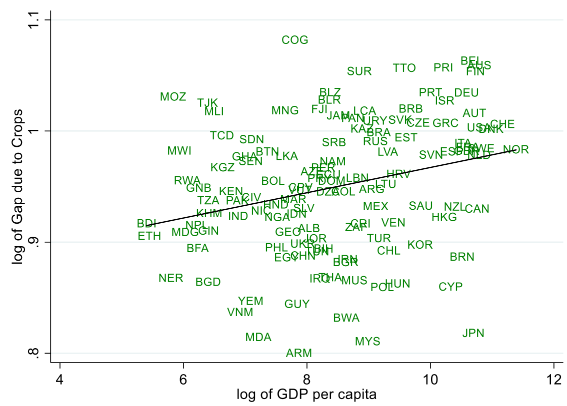


Figure 9. Correlations between gap measures and GDP per capita across countries

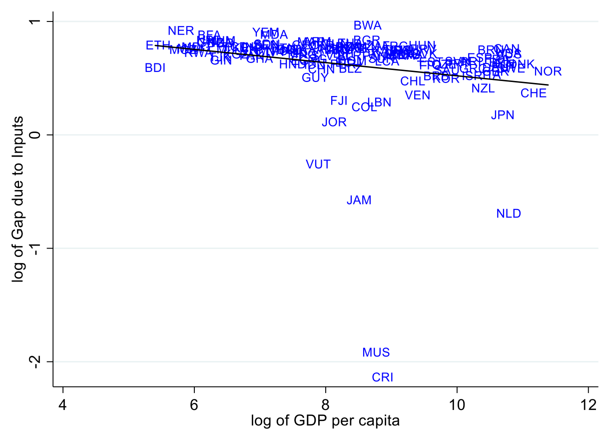
9a. Total Gap



9b. Gap due to Crops



9c. Gap due to Inputs



Correlation of yields and gaps with infrastructure and inputs

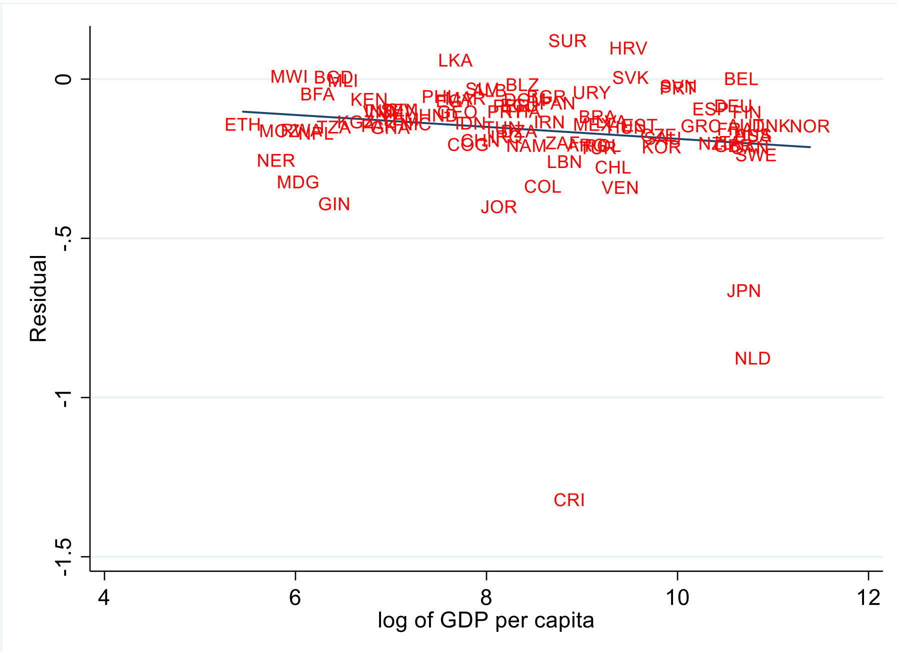
The negative relationships between gaps and GDP per capita, documented above, appears to be largely accounted for by differential use of technology and road infrastructure, as this relationship significantly weakens once these factors are accounted for at the country-level (Table 7b and Figure 10). Motivated by this, we document correlations between our gap measures and available cross-country measures of input usage, farming characteristics, and infrastructure availability. The goal here is not to infer causality, but to simply document systematic variation between documented input and crop selection gaps on the one hand, and explanations highlighted by the literature on the other. In particular, we document correlations

of our gap measure with road density, fertilizer use, use of modern technology in agriculture, and average farm size.

Table 6 reports bivariate correlations between yield gaps and these variables: panel A reports the correlations with the overall gap, panel B on the gap due to crops, and panel C on the gap due to inputs. Results with and without conditioning by GDP per capita are presented. We find that road density and fertilizer use are each negatively correlated with the overall gap. When decomposing this effect, this relationship appears to stem from the gap due to inputs. Table 7a shows similar results when regressing our gap measures with all input and infrastructure variables together. The correlation is weaker for farm sizes and modern technology usage. Similar results are obtained after conditioning by GDP per capita (Table 7b).

Overall, these cross-country results show both the systematic variation in gaps and yields across income groups as well as how variation within income groups is associated with road density and fertilizer use. In fact, the 10 percent of countries with the highest road density have an average yield that is 5.9 times greater than those from the bottom 10 percent. The equivalent ratio among the top and bottom users of fertilizer is 9.8. These ratios are higher than differences in potential yields —which are driven entirely by geography— where the ratio of averages is below 2.4, as well as higher than the equivalent ratio in yields between the top and bottom income countries of 4.5 (Table 4).

Figure 10. Correlation between gaps and GDP per capita, controlling for country characteristics



Note: The vertical access plots the residual from a regression of the total yield gap on country characteristics.

Table 6. Correlations between the agricultural gap measures and market characteristics, with and without conditioning on GDP per capita

	A. Total Gap		B. Gap due to Crop Selection		C. Gap due to Inputs	
	Cond.	Uncond.	Cond.	Uncond.	Cond.	Uncond.
Road density	-0.056**	-0.059**	-0.005	0.002	-0.085**	-0.098**
Fertilizer use	-0.076**	-0.080***	-0.009*	0.000	-0.112**	-0.129**
Farm size	0.026*	0.019	0.000	0.003	0.046*	0.030
Tractor use	0.022	0.001	-0.005	0.000	0.047	0.007

Note: Unconditional refers to the coefficients from a bivariate regression of the gap measure on each of the market characteristics in logs (road density, fertilizer use, farm size, and tractor use). Conditional refers to multivariate regression of the gap measure on each of the market characteristics including GDP per capita as a control. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 7. Cross-country regressions of agricultural gap on market characteristics

7a. Unconditional: Omitting GDP per capita

	(1) *ln(total gap)*	(2) *ln(gap, crops)*	(3) *ln(gap, inputs)*
ln Road density	-0.043*	0.001	-0.069**
	(-1.97)	(0.21)	(-2.03)
ln Fertilizer	-0.066**	-0.001	-0.106**
	(-2.46)	(-0.18)	(-2.15)
ln Farm size	0.015	0.004	0.021
	(1.32)	(1.04)	(0.92)
ln Tech use	0.011	-0.001	0.024
	(1.01)	(-0.23)	(1.29)
Constant	0.952***	0.956***	0.632***
	(7.58)	(19.05)	(3.20)
Observations	88	89	88
R-squared	0.324	0.267	0.264

Note: log gaps are calculated as ln(gap+2). Robust t-statistics in parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%.

7b. Conditional: including GDP per capita

VARIABLES	(1) *ln(total gap)*	(2) *ln(gap, crops)*	(3) *ln(gap, inputs)*
ln GDP pc	-0.025	0.030***	-0.078***
	(-1.45)	(3.90)	(-2.81)
ln Road density	-0.037*	-0.006	-0.050
	(-1.82)	(-0.96)	(-1.61)
ln Fertilizer	-0.062**	-0.005	-0.097**
	(-2.39)	(-0.77)	(-2.05)
ln Farm size	0.020	-0.001	0.034
	(1.49)	(-0.23)	(1.38)
ln Tech use	0.015	-0.005	0.035*
	(1.22)	(-1.54)	(1.67)
Constant	1.125***	0.752***	1.161***
	(7.79)	(11.51)	(5.28)
Observations	88	89	88
R-squared	0.333	0.414	0.289

Note: log gaps are calculated as ln(gap+2). Robust t-statistics in parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%.

5. Conclusion

Differences in geographically-induced, crop-specific comparative advantages can explain a substantial share of the variation in yields across the world. This does not imply fatalistic geographic determinism, however, as large gaps between potential and actual yields suggest that there are large potential gains to be had from improving crop and input choices. When decomposing these gaps, we find that crop selection plays a relatively larger role in explaining gaps as compared to input usage, even if the latter are more strongly correlated with GDP per capita levels. This highlights the need to understand the determinants of crop selection if we are to fully uncover the macro-relevant sources of aggregate productivity gaps and formulate appropriate policies to address them.

Although our analysis provides a broad view of relevant sources of aggregate yield variation, a few caveats are important to acknowledge before landing on specific policy conclusions. First, the above results are subject to measurement limitations. Ideally, our analysis would have used observed yields and land allocation at the plot level. In the absence of such a precise dataset, we instead rely on modeled data whereby national statistics are downscaled to the 5-arcmin cell level following a number of assumptions described in section 2. This might miss important features of how land is allocated to different crops within each country and introduce biases in our country-level potential yield measures. Second, our cross-country analysis suggests that road infrastructure and access to better inputs such as fertilizer are associated with lower agricultural yield gaps. These are simple correlations and drawing causal implications is beyond the scope of this paper.

Our results, nonetheless, do signal that the interaction of crop selection and market accessibility is a mechanism that is quantitatively important in explaining aggregate yield variation. The evidence encourages further work on policies that not only improve input usage in current crops, but also facilitate crop diversification either directly, or indirectly through infrastructure and investments in new technologies.

References

Adamopoulos, Tasso (2011) "Transportation costs, agricultural productivity, and cross-country income differences," *International Economic Review*, 52(2): 489-521.

Adamopoulos, Tasso and Diego Restuccia (2018) "Geography and Agricultural Productivity: Cross-Country Evidence from Micro Plot-Level Data," working paper.

Adamopoulos, Tasso and Diego Restuccia (2014) "The Size Distribution of Farms and International Productivity Differences," *American Economic Review*, 104(6): 1667-97.

Alvarez, Jorge (2019) "The Agricultural Wage Gap: Evidence from Brazilian Micro-data," *American Economic Journal: Macroeconomics*, *forthcoming*.

Anderson, Weston, Liangzhi You, Stanley Wood, Ulrike Wood-Sichra, and Wenbin Wu (2014) "A comparative Analysis of Global Cropping Systems Models and Maps," *IFPRI Discussion Paper* 01327.

Asad, Saher (2014) "The Crop Connection: Impact of Cell Phone Access on Crop Choice in Rural Pakistan," unpublished manuscript, George Washington University.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Chris Udry (2013) "Profitability of Fertilizer: Experimental Evidence from Female Rice Farmers in Mali," *American Economic Review*, 103(3): 381-386.

Buys, Piet, Uwe Deichmann, and David Wheeler (2010) "Road Upgrading and Overland Trade Expansion in Sub-Saharan Africa," Journal of African Economies, 19(3): 399-432. https://doi.org/10.1093/jae/ejq006

Canning, David (1998) "A Database of World Stocks of Infrastructure: 1950-1995," *The World Bank Economic Review*, 12(3): 529-548.

Comin, Diego and Bart Hobijn (2010) "An Exploration of Technology Diffusion," *American Economic Review*, 100(5): 2031-2059.

Damania, Richard, Claudia Berg, Jason Russ, A. Federico Barra, John Nash, and Rubaba Ali (2016) "Agricultural Technology Choice and Transport," *American Journal of Agricultural Economics*, 99(1): 265-284.

Donovan, Kevin (2016) "Agricultural Risk, Intermediate Inputs, and Cross-Country Productivity Differences," Unpublished manuscript, University of Notre Dame.

Duflo, Esther, Michael Kremer, and Jonathan Robinson (2008) "How High Are Rates of Return to Fertilizer? Evidence from a Field Experiment in Kenya," *American Economic Review*, 98(2): 482-88.

Duflo, Esther, Michael Kremer, and Jonathan Robinson (2011) "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya," *American Economic Review*, 101(6): 2350-90.

Emran, M. Shahe and Forhad Shilpi (2012) "The extent of the market and stages of agricultural specialization," *Canadian Journal of Economics*, 45(3): 1125-1153.

FAO (2014). *The State of Food and Agriculture 2014: Innovations in Family Farming.* Food and Agriculture Organization.

FAO & IIASA (2012) "GAEZ: Definitions and Variables"
www.fao.org/fileadmin/user_upload/gaez/docs/Definitions_EN.pdf

Feder, G, R Just, and D Zilberman (1985) “Adoption of Agricultural Innovations in Developing Countries: A Survey,” *Economic Development and Cultural Change*, 33(2): 255-299.

Gollin, Douglas, Stephen Parente, and Richard Rogerson (2002) “The Role of Agriculture in Development,” *American Economic Review*, 92(2): 160-164.

Grolleaud, M. (2002) *Post-Harvest Losses: Discovering the Full Story,* Rome: Food and Agriculture Organization.

Herrendorf, B. and T. Schoellman (2015). “Why is measured productivity so low in agriculture?” *Review of Economic Dynamics,* 18(4), 1003-1022.

Herrendorf, B. and T. Schoellman (2018). “Wages, Human Capital, and Bariers to Structural Transformation” *American Economic Journal: Macroeconomics,* 10(2), 1-23.

Hayami, Y. (1969). Sources of agricultural productivity gap among selected countries. *American Journal of Agricultural Economics*, *51*(3), 564-575.

IFPRI (2013). *Global Food Policy Report*. Washington, D.C.: International Food Policy Research Institute.

Kaminski, J. and L. Christiaensen (2014) “Post-Harvest Loss in Sub-Saharan Africa: What do farmers say?” *Global Food Security*, 3(3-4): 149-158.

Kauffman, Daniel and Aart Kraay (2002) “Growth without Governance,” *World Bank Policy Research Working Paper* 2928.

Lagakos, David and Michael E. Waugh (2013) “Selection, Agriculture, and Cross-Country Productivity Differences,” *American Economic Review*, 103(2): 948-80.

Monfreda, C, N. Ramankutty, and J. A. Foley (2008) "Farming the Planet 2: Geographic Distribution of Crop Areas, Yields, Physiological Types and Net Primary Production in the Year 2000," *Global Biogeochemical Cycles*, 22(1): 1-19.

Oehmke, J. F. (1992). *Technology, Impact and Agricultural Transformation: Lesson Learned from Impact Assessments*. Washington, D.C.: USAID.

Ramankutty, N, A T Evan, C Monfreda, and J A Foley (2008) "Farming the Planet: 1. Geographic Distribution of Global Agricultural Lands in the Year 2000." *Global Biogeochemical Cycles*, 22(1): 1-19.

Restuccia, Diego (2011) "Recent Developments in Economic Growth," *FRB Richmond Economic Quarterly,* 97(3): 329-357.

Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu (2008) "Agriculture and aggregate productivity: A quantitative cross-country analysis," *Journal of Monetary Economics*, 55(2): 234-250.

Solt, Frederick (2019) "Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database." SWIID Version 8.0, February 2019.

Takeshima, H., Houssou, N., & Diao, X. (2018). Effects of tractor ownership on returns-to-scale in agriculture: Evidence from maize in Ghana. *Food Policy*, *77*, 33-49.

Tombe, Trevor (2015) "The missing food problem: Trade, agriculture, and international productivity differences," *American Economic Journal: Macroeconomics*, 7(3): 226-258.

Uchida, Hirotsugu and Andrew Nelson (2008) "Agglomeration Index: Towards a New Measure of Urban Concentration," Background Paper for World Development Report 2009.

Wood-Sichra, Ulrike, Alison B. Joglekar, and Liangzhi You (2016) "Spatial Production Allocation Model (SPAM) 2005: Technical Documentation," Harvest Choice, IFPRI.

World Bank (2008) *World Development Report: Agriculture for Development*, Washington, DC.

World Bank (2014) "For Up to 800 Million Rural Poor, a Strong World Bank Commitment to Agriculture," (November 12, 2014) http://www.worldbank.org/en/news/feature/2014/11/12/for-up-to-800-million-rural-poor-a-strong-world-bank-commitment-to-agriculture