Online Appendix for "Geography and Agricultural Productivity: Cross-Country Evidence from Micro Plot-Level Data"*

Tasso Adamopoulos
York University

Diego Restuccia
University of Toronto
and NBER

July 2021

*Contact: Tasso Adamopoulos, aadamo@yorku.ca; Diego Restuccia, diego.restuccia@utoronto.ca

A Country Sample

Table A.1 lists all 162 countries in our data set, along with the corresponding country code, the number of cells covering the country, and the level of real GDP per capita in 2000.

Table A.1: List of Countries and Other Information

Country	Code	Cell Count	GDP per capita
Afghanistan	AFG	9000	327
Albania	ALB	444	3177
Algeria	DZA	30751	5276
Angola	AGO	14988	2901
Antigua and Barbuda	ATG	5	14522
Argentina	ARG	40080	12519
Armenia	ARM	451	4333
Australia	AUS	100208	30240
Austria	AUT	1447	31574
Azerbaijan	AZE	1311	3722
Bahamas	BHS	160	24593
Bangladesh	BGD	1759	1794
Belarus	BLR	4057	12188
Belgium	BEL	558	29693
Belize	BLZ	271	7910
Benin	BEN	1374	1336
Bhutan	BTN	523	2817
Bolivia	BOL	13284	3346
Bosnia and Herzegovina	BIH	836	5798
Botswana	BWA	7297	7219
Brazil	BRA	101847	8391
Brunei Darussalam	BRN	65	48210
Bulgaria	BGR	1754	6374
Burkina Faso	BFA	3262	1121
Burundi	BDI	312	706
Cambodia	KHM	2184	1764

Continued on next page...

Table A.1 — *Continued from previous page*

Country	Code	Cell Count	GDP per capita
Cameroon	CMR	5470	2448
Canada	CAN	244154	31471
Central African Republic	CAF	7287	918
Chad	TCD	15448	1445
Chile	CHL	11199	14309
China	CHN	136881	4076
Colombia	COL	13318	6620
Congo	COG	4032	3835
Costa Rica	CRI	609	9463
Cote d'Ivoire	CIV	3795	2761
Croatia	HRV	919	9775
Cuba	CUB	1381	7636
Cyprus	CYP	129	20275
Czech Republic	CZE	1419	16044
Democratic Republic of the Congo	ZAR	27327	312
Denmark	DNK	898	30468
Dominican Republic	DOM	598	7559
Ecuador	ECU	2996	4894
Egypt	EGY	13029	4690
El Salvador	SLV	253	5192
Equatorial Guinea	GNQ	314	8820
Eritrea	ERI	1469	668
Estonia	EST	1015	10405
Ethiopia	ETH	13365	892
Fiji	FJI	230	5784
Finland	FIN	9008	26402
France	FRA	9266	27311
Gabon	GAB	3056	8504
Gambia	GMB	132	1289
Georgia	GEO	1099	4310
Germany	GER	6608	29051
Ghana	GHA	2819	1359

Continued on next page...

Table A.1 — *Continued from previous page*

Country	Code	Cell Count	GDP per capita
Greece	GRC	1970	20708
Guatemala	GTM	1326	5530
Guinea	GIN	2908	3235
Guinea-Bissau	GNB	403	657
Guyana	GUY	2475	2457
Haiti	HTI	336	1655
Honduras	HND	1360	3062
Hungary	HUN	1590	13025
India	IND	40163	2687
Indonesia	IDN	22138	4151
Iran (Islamic Republic of)	IRN	22489	8049
Iraq	IRQ	6069	5403
Ireland	IRL	1334	31389
Israel	ISR	285	22356
Italy	ITA	4774	27142
Jamaica	JAM	135	7877
Japan	JPN	5488	28341
Jordan	JOR	1220	4329
Kazakhstan	KAZ	47485	7641
Kenya	KEN	6800	1943
Korea, Republic of	KOR	1434	18597
Kuwait	KWT	225	36146
Kyrgyzstan	KGZ	3098	3310
Lao People's Democratic Republic	LAO	2847	1777
Latvia	LVA	1371	8119
Lebanon	LBN	144	7505
Lesotho	LSO	414	1770
Liberia	LBR	1125	492
Libyan Arab Jamahiriya	LBY	21221	14674
Lithuania	LTU	1325	8566
Luxembourg	LUX	47	63392
Madagascar	MDG	7353	965

Continued on next page...

Table A.1 — *Continued from previous page*

Country	Code	Cell Count	GDP per capita
Malawi	MWI	1425	1032
Malaysia	MYS	3856	14178
Mali	MLI	15355	1108
Malta	MLT	6	19442
Mauritania	MRT	12944	2085
Mexico	MEX	25084	10339
Mongolia	MNG	26562	2008
Montenegro	MNE	214	4877
Morocco	MAR	5529	4574
Mozambique	MOZ	9647	1245
Namibia	NAM	10397	5531
Nepal	NPL	1944	1783
Netherlands	NLD	677	31927
New Zealand	NZL	4206	21437
Nicaragua	NIC	1538	2058
Niger	NER	14499	811
Nigeria	NGA	10772	1275
Norway	NOR	8617	41777
Oman	OMN	3849	23752
Pakistan	PAK	11827	2696
Panama	PAN	888	7124
Papua New Guinea	PNG	5470	2194
Paraguay	PRY	5062	4556
Peru	PER	15324	4975
Philippines	PHL	3538	3955
Poland	POL	5882	10834
Portugal	PRT	1381	19606
Puerto Rico	PRI	113	25955
Qatar	QAT	142	61389
Republic of Moldova	MDA	576	2420
Romania	ROM	3958	6151
Russia	RUS	421168	8305

Continued on next page...

Table A.1 — *Continued from previous page*

Country	Code	Cell Count	GDP per capita
Rwanda	RWA	293	994
Saudi Arabia	SAU	25034	19207
Senegal	SEN	2372	1732
Sierra Leone	SLE	863	1171
Singapore	SGP	7	35424
Slovakia	SVK	858	11844
Slovenia	SVN	341	19043
Solomon Islands	SLB	334	1318
Somalia	SOM	7490	480
South Africa	ZAF	16282	8441
Spain	ESP	7727	24945
Sri Lanka	LKA	793	4603
Sudan	SDN	30052	1546
Suriname	SUR	1706	7490
Swaziland	SWZ	228	6587
Sweden	SWE	11321	27174
Switzerland	CHE	704	34414
Syrian Arab Republic	SYR	2672	2446
Taiwan	TWN	464	21513
Tajikistan	TJK	2120	1902
Thailand	THA	6227	7058
The former Yugoslav Republic of Macedonia	MKD	396	6358
Togo	TGO	682	984
Tunisia	TUN	2186	7572
Turkey	TUR	11699	6428
Turkmenistan	TKM	7077	8716
Uganda	UGA	2834	1094
Ukraine	UKR	10587	5644
United Arab Emirates	ARE	908	38604
United Kingdom	GBR	4857	27032
United Republic of Tanzania	TZA	11088	681

Continued on next page...

Table A.1 — *Continued from previous page*

Country	Code	Cell Count	GDP per capita
United States	USA	160841	39241
Uruguay	URY	2467	11426
Uzbekistan	UZB	6960	1477
Vanuatu	VUT	157	5607
Venezuela	VEN	10758	10553
Viet Nam	VNM	3970	2407
Yemen	YEM	5148	1129
Zambia	ZMB	9045	1038
Zimbabwe	ZWE	4813	4528

Notes: The cell count of each country is from GAEZ (2000), and refers to the number of 5-arc-minute cells covering the country. Real GDP per capita is from Heston et al. (2009).

B GAEZ Data

B.1 Cell-Level Actual Data

The GAEZ (2000) methodology for estimating output and harvested land by crop for each cell in the world uses a downscaling methodology that combines aggregate and cell-level data. GAEZ first estimates for each cell, cultivated land and the split between rainfed and irrigated, using GIS land-cover datasets at the 5-arc minute resolution. The procedure ensures that the land class coverage is consistent with aggregate FAOSTAT (2000) land statistics (arable land) and land cover patterns obtained from remotely sensed data. To allocate crops to cells, GAEZ uses for each country, data on output and harvested area by crop at the national level from the FAO and at the sub-national level (regions, states, provinces, districts, counties, etcetera) from Monfreda et al. (2008). The downscaling procedure employs an iterative optimization algorithm that is initialized by feeding in a prior distribution of crops production allocation to cells that is based on cell-level information on the amount of cultivated land, bio-physical suitability for the production of the different crops, and socio-economic factors such as farming zone system, population density, and distance to market. Then each iteration step determines the discrepancy between statistical totals available at the sub-national (or national) unit level and the respective totals calculated by summing harvested areas and production over cells. The magnitude of these deviations is then used to revise the land and crop allocation and to recalculate discrepancies. The process is continued until all accounting constraints are met, that is, output and harvested land sum up to the aggregates sub-nationally and nationally.

B.2 Input Assumptions for Potential Yields

Low inputs. Under this scenario GAEZ (2000) assumes a subsistence-based farming system, with traditional management, that does not necessarily produce for the market. Production is based on the use of traditional cultivars, labor intensive techniques, no mechanization, and minimum conservation measures. There is no application of nutrients, and no use of chemicals for pest or disease control. The assumed water supply under this scenario is fully rainfed farming.

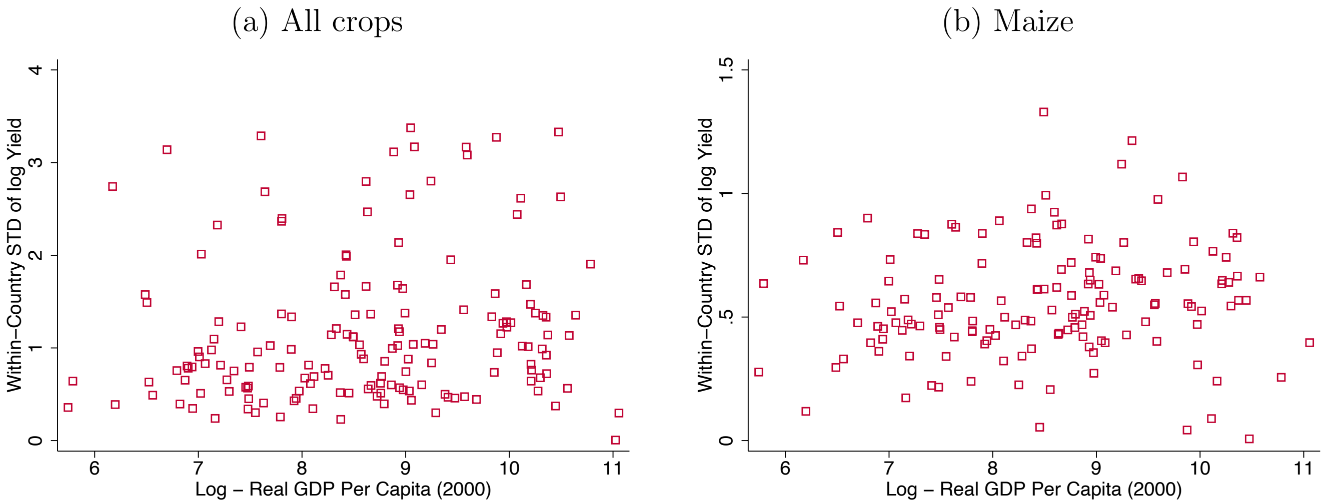
Mixed inputs. Under this scenario GAEZ (2000) assumes that the highest quality land (very suitable and suitable) uses high inputs, the moderately suitable land uses an intermediate level of inputs, and the marginal land uses low inputs. Under intermediate inputs the farming system is partly market oriented involving both subsistence and commercial farming; use of improved varieties; intermediate labor intensity with hand tools and/or animal traction and some mechanization; some fertilizer application; and some chemical pest, disease, and weed control. Under high inputs the there is advanced management; the farming system is primarily market oriented with commercial production; high yield varieties are used; fully mechanized; low labor intensity; optimum application of nutrients and chemical pest, disease and weed control. This scenario covers all land, both under rainfed and irrigated water supply. GAEZ considers the mixed input scenario as a "reasonable reflection of actual agricultural input and management circumstances." (page 97, *GAEZ Model Documentation*, 2012).

C Within-country Dispersion of Potential Yields

We document a measure of the variability of land quality across countries. In particular, we compute the standard deviation of log potential yields across cells with agricultural production and report this dispersion across countries by real GDP per capita in Figure C.1. We report two measures of dispersion. The first, in panel (a), for all crops, that is the simple average of potential yields across crops within a cell; the second, in panel (b), only for maize which is the most prevalent crop produced around the world. We use the potential-yield under the rainfed low-input scenario as this most closely reflects land quality productivity.

We find that while the dispersion in land quality within a country differs quite substantially across

Figure C.1: Within-country Dispersion in Land Quality



Notes: Dispersion in potential yields across cells within a country under the rainfed low-input scenario. Panel (a) reports the simple average of yields across all crops within a cell, whereas panel (b) is only for maize in each cell.

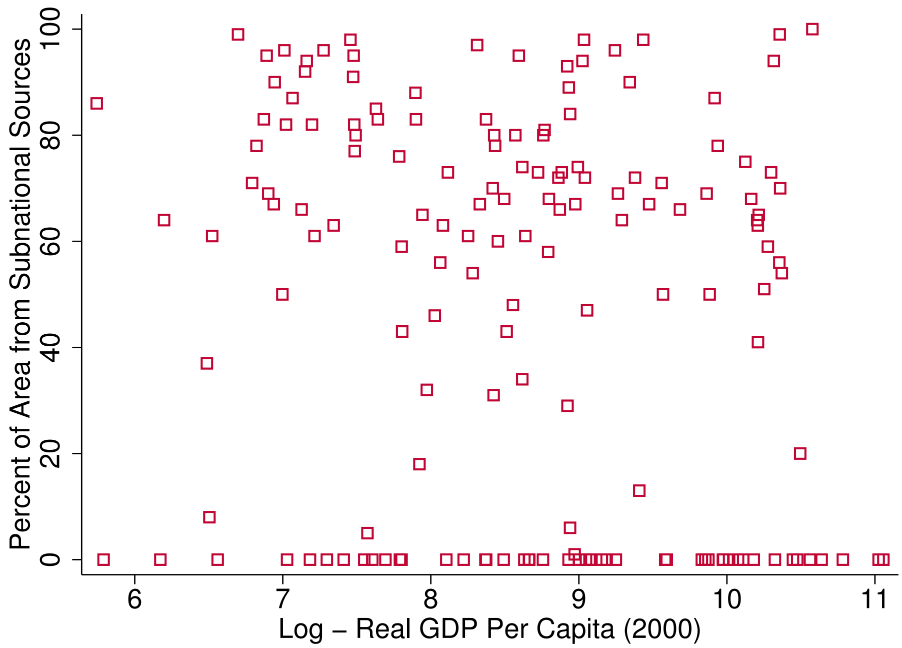
countries, the dispersion is not systematically associated with the level of development, there is only a modest positive association. Hence, land quality endowment in poor countries is not worse than rich countries, not just in terms of averages from aggregate potential yields, but also in terms of the dispersion of land quality within the country. This finding is relevant to the extent that poor countries may not optimize on the location of production or the number of locations with agricultural production.

D Robustness and Validation

Our construction of the country-level potential yields does not rely on actual output or yields at the crop-cell level. In the case of the production-potential counterfactual, however, the aggregate potential yield uses the cell-level land allocation by crop in each country as weights. We examine the bias and reliability of the cell-level land weights, conduct our own external validation, and show the robustness of our results to alternative land weighting schemes.

Lower and middle income countries do not tend to have less detailed data. GAEZ uses the most spatially disaggregated data available from Monfreda et al. (2008), at the sub-national level. One natural concern is that the prevalence of spatially disaggregated agricultural statistics that GAEZ uses is disproportionate for developed countries than for lower and middle income countries. Using sub-national data availability from Monfreda et al. (2008) (from their Table 2), in Figure D.2 we plot the percentage of land covered by sub-national data across countries against their GDP per capita. There is no systematic relationship of sub-national data availability and the level of income. There are rich and poor countries with substantial sub-national coverage and there rich and poor countries with no sub-national coverage. We conclude that GAEZ does not have more detailed information for the richer countries than for the poorer countries, that can inherently bias their downscaling methodology of attributing national and sub-national statistics to cells across countries.

Figure D.2: Subnational Data Availability



GAEZ validation. A general limitation of cell-level estimated data is that because they require substantial amounts of crop statistics and other data at the finest level of disaggregation, there is not much "out-of-sample" data remaining in order to externally validate the results. Although, not in a systematic fashion, GAEZ has done some testing of the reliability of its actual agricultural estimates: "Tests in China and Brazil by comparing downscaled results based on statistics available on national level, with detailed sub-national statistics on county and micro region level revealed strong correlations between downscaled national statistics and county/micro-region level statistics of harvested areas, yields and crop production." (GAEZ-FAO website, under "FAQs," `www.fao.org/nr/gaez/faqs/en/`.)

Our own validation. We conduct our own validation exercise to confirm that the crops GAEZ attributes as being produced to the different cells are actually produced there. In particular, we use survey farm-level data from UNPS (2009), the World Bank's Living Standards Measurement Study (LSMS) for Uganda for the earliest year with GPS data in the survey of 2009. To make the comparison we use the GPS coordinates of the locations of farms in the LSMS data to assign farm households to pixels in the GAEZ rasters (5 arc minute resolution). Given that the LSMS data are survey data, the observations are sparse with some pixels having multiple households, others very few, and most none. For the pixels for which there is any LSMS household we compute a series of dummies, one for each of the 18 GAEZ crops and for each of the LSMS and GAEZ. In the set of dummies for LSMS (GAEZ), the dummy in a pixel takes the value of 1 if the crop is produced in LSMS (GAEZ). Then we create a third dummy that takes the value of 1 for a crop in a given pixel if the crop is produced in both GAEZ and LSMS. Finally we compute the fraction of the aligned crops in total GAEZ crops for every pixel. For the median pixel 60 percent of all the crops GAEZ attributes to a cell are also confirmed to be produced according to the LSMS. If we focus only on maize, one of the most widely produced crops in Uganda 77 percent of the cells that GAEZ attributes maize production is also confirmed by the LSMS data. This is remarkable given that the LSMS data are survey data, are spatially sparse, and are 9 years later than the GAEZ data. For these reasons we could not make more direct comparisons of land allocations and yields.

Our conclusions do not change if we weigh crops within cells equally. Please refer to Robustness Section 4.4 in the main text.

Our conclusions do not change if we use crop-cell land allocations from IFPRI's Harvest Choice project. The International Food Policy Research Institute's (IFPRI) Harvest-Choice (2012) is another project that uses crop statistics at the national and subnational levels across the world for the year 2005 to estimate crop yields and harvested land at the 5 arc-minute grid cell level. Their model for downscaling agricultural statistics, called "Spatial Production Allocation Model (SPAM)," uses a cross-entropy optimization approach that uses information on cropland surface, location of irrigated areas, crop suitability, rural population densities, production systems and crop prices. In this sense, even though distinct from GAEZ, Harvest-Choice (2012)'s disaggregation methodology is similar in nature and uses similar types of disaggregate information to estimate the same cell-level resolution as GAEZ. Harvest-Choice (2012) however uses data that reflect the year 2005 (rather than 2000 in GAEZ), and includes a finer set of crops. The underlying crop suitability surfaces that SPAM uses are from GAEZ, but uses a more updated cropland surface, and has made an effort to use more district-level agricultural statistics within countries. While the cell-level data from Harvest-Choice (2012) are also estimates at the 5 arc-minute resolution, and involve a five year gap from GAEZ, we compare the cell-level land allocations across crops (which determine the cell-level weights in our methodology) between GAEZ (2000) and Harvest-Choice (2012). The cell-by-cell correlation of harvested land for all crops between GAEZ and *Harvest Choice* for the entire world is 0.71. The same correlation for the three most popular crops, wheat, rice, and maize is 0.69. The correlation for rice alone is 0.84. While these correlations are high, we go a step further and re-compute each country's potential yield with our methodology, aggregating GAEZ potential yields, but using the cell-level land allocations across crops from *Harvest Choice* instead (rather than the GAEZ ones). Just as in the alternative with equal weights across all produced crops above, with the Harvest Choice weights our conclusions are the same.

Figure D.3: Potential Yields with Harvest Choice Weights

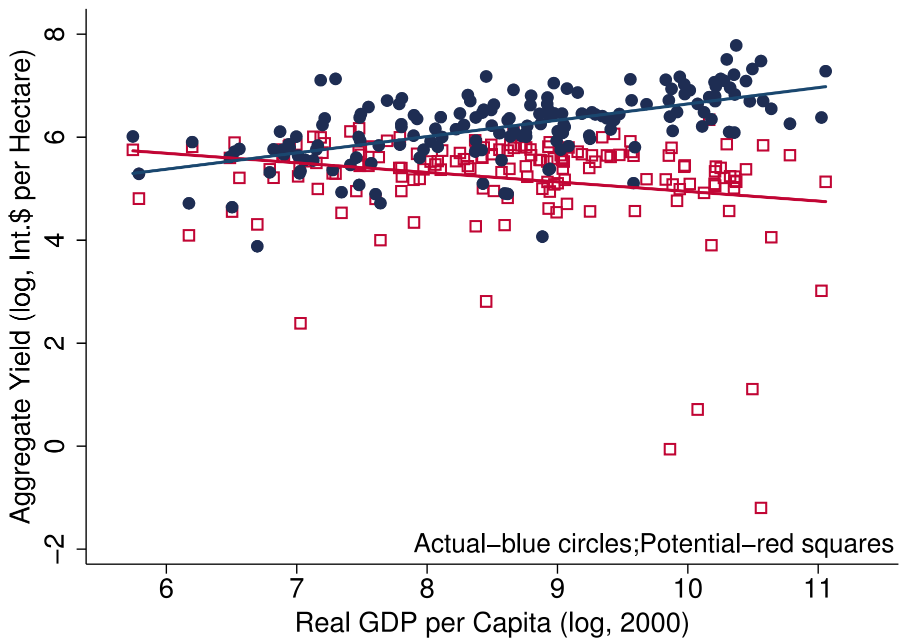


Figure D.3 displays the aggregate potential yield for each country (against GDP per capita) using the cell-by-cell Harvest-Choice (2012) weights. The cell-level potential yields from GAEZ used are for low inputs and rainfed water supply. The lack of a systematic relationship of aggregate potential yields with income is true here as well.

References

FAOSTAT (2000). Production statistics. *Food and Agricultural Organization (FAO)*, (Value of Agricultural Production, Area Harvested). http://www.fao.org/faostat/en/#data (accessed January 20, 2021).

GAEZ (2000). Global agro-ecological zones, version 3.0. *Food and Agricultural Organization (FAO) and International Institute for Applied Systems Analysis (IIASA)*. https://www.gaez.iiasa.ac.at/ (accessed March 11, 2018).

Harvest-Choice (2012). Global spatially-disaggregated crop production statistics data for 2005 version 3.2. *International Food Policy Research Institute (IFPRI) and International Institute for Applied Systems Analysis (IIASA)*, (Harvard Dataverse, V9, 10.7910/DVN/DHXBJX). https://doi.org/10.7910/DVN/DHXBJX.

Heston, A., Summers, R., and Aten, B. (2009). Penn world table, version 6.3. *Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania*. https://www.rug.nl/ggdc/productivity/pwt/pwt-releases/pwt-6.3?lang=en.

Monfreda, C., Ramankutty, N., and Foley, J. A. (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).

UNPS (2009). Uganda bureau of statistics, national panel survey (unps) 2005-2009. *World Bank*, (Ref. UGA_2005-2009_UNPS_v01_M.). https://microdata.worldbank.org/index.php/catalog/1001/get-microdata (accessed June 15, 2020).