August 3, 2017

How Tight are Malthusian Constraints?

T. Ryan Johnson
University of Houston

Dietrich Vollrath
University of Houston

Online Appendix

Robustness checks and alternative assumptions for empirical work from the main paper are contained here. Also included are definitions of countries included in regions used in paper.

Contents

A.1 Alternative Mobility Assumptions

Our baseline specifications are built off of a model that assumes agricultural output, as well as labor and capital, are freely mobile across districts within a given province (although we do not require them to be mobile across provinces). However, even within a given province, there may be frictions or limits on the mobility of either output or inputs, or both. If these frictions exist, then our regressions may not be delivering unbiased estimates of β. We outline two alternative assumptions about mobility here, and how our results relate to those.

A.1.1 Immobile Factors

The baseline model assumes capital and labor are free to move between districts within a region. If we make factors immobile, but allow both agricultural and non-agricultural output to move between districts, this changes the specification of the relationship between agricultural productivity and rural density.

The agricultural production function for a district is the same as in (1), and we also need to specify a production function for non-agriculture. We do so as $Y_{Ni} = A_{Ni}K_{Ni}^{\alpha}L_{Ni}^{1-\alpha}$. Capital and labor are assumed to be mobile *within* the district between the two sectors, implying that the return to capital and the return to labor are equalized across different uses. Because of this, the capital/labor ratio in both sectors will be identical, with $K_{Ai}/L_{Ai} = K_{Ni}/L_{Ni} = K_i/L_i$, where K_i/L_i is the district's aggregate capital/labor ratio.

Equality of the return to labor across different sectors implies that

$$p_A(1-\alpha)(1-\beta)\frac{Y_{Ai}}{L_{Ai}} = p_N(1-\alpha)\frac{Y_{Ni}}{L_{Ni}}.$$

Using the condition that the capital/labor ratio will be identical across the two sectors, and re-arranging this relationship, we have that

$$p_A(1-\beta)A_{Ai}\left(\frac{K_i}{L_i}\right)^{\alpha(1-\beta)}\left(\frac{X_i}{L_{Ai}}\right)^{\beta} = p_N A_{Ni}\left(\frac{K_i}{L_i}\right)^{\alpha}$$

Taking logs, are again re-arranging terms, we arrive at

$$\ln A_{Ai} = \beta \ln L_{Ai}/X_i + \ln A_{Ni} + \alpha\beta \ln K_i/L_i + \ln p_N/p_A.$$

This equation shows that the relationship between agricultural productivity, A_{Ai}, and rural density, L_{Ai}/X_i, can still be used to recover an estimate of β. To do this, we must control for the district-specific levels of non-agricultural productivity, A_{Ni}, and capital/labor, K_i/L_i. While we do not have direct measures of those, we believe that our control for night lights will act as a decent proxy for these terms. Finally, the price ratio, p_N/p_A, is the province relative price, as goods are traded freely, so this will be captured by the province level fixed effects.

If our night lights control is not capturing the variation in A_{Ni} or K_i/L_i, then our estimates may be biased if there is a relationship between those variables and rural density. In particular, if rural density is negatively related to A_{Ni} and/or K_i/L_i then we could be under-stating the value of β. It is not clear why this negative relationship would hold only in tropical areas (with small estimate β values), but not in other areas.

A.1.2 Autarkic Districts

If districts are entirely closed, in that neither factors of production nor output can move between districts, then this again changes the specification of our regressions. Here, the crucial remaining assumption is that the value of β is the same across all districts within a given province.

Within each district, let the amount of agricultural output consumed be c_{Ai}, and hence market clearing within the district requires $c_{Ai}L_i = Y_i$ for agricultural output. Using the same production function as in the main section, and again assuming that capital and labor move freely between sectors (non-agriculture and agriculture) so that the capital/labor ratios are equal to the aggregate ratio, we have

$$c_{Ai}L_i = A_i X_i^{\beta} \left(K_i/L_i\right)^{\alpha(1-\beta)} L_{Ai}^{1-\beta}.$$

Taking logs and re-arranging, we have the following

$$\ln A_i = \beta \ln L_{Ai}/X_i - \ln L_{A_i}/L_i - \alpha(1-\beta) \ln K_i/L_i + \ln c_{Ai}.$$

We can recover an estimate of β from the relationship of productivity and rural density, but now must control for the agricultural share of labor, L_{Ai}/L_i, the capital/labor ratio, and the consumption of agricultural goods per capita. For L_{Ai}/L_i, we have this data, and can include it directly in a regression (it is implicitly included in our baseline regression when we use the percent urban). For the capital/labor ratio and consumption of agricultural goods, we believe that the night lights data are a decent proxy for these terms.

Including the log of L_{Ai}/L_i explicitly as a control is possible, and the results of this are shown in Table A.11 at the end of this Appendix.

The results in Table A.11 may still be biased, however, if the night lights proxy does not pick up the variation in consumption or the capital/labor ratio. If the capital/labor ratio is positively related to the rural density, then we would be under-estimating the true value of β. The small estimated values of β in tropical areas may be because of this relationship, although it is not clear why rural density would be positively related to capital/labor ratios only in tropical areas. Alternatively, if consumption of agricultural goods is negatively related to rural density, and we are not controlling for it with night lights, then we may be under-estimating β. This could possibly be true only in tropical areas if they are relatively poor, whereas this relationship no longer holds in richer, temperate areas. This is clearly a possibility, although recall that this would only be a problem if we believe that districts are *autarkic*, which may be an extreme assumption.

A.2 Solving for Labor Share and Real Income

In section 4.1 we solved for L_A/L and y, the agricultural labor share and real income, respectively. The algebra leading to equations (12) and (13) is as follows.

Based on the district-level production functions from (1) total agricultural supply in province I can be written as

$$Y_A = \sum_{i \in I} A_i X_i^{\beta} \left(K_{Ai}^{\alpha} L_{Ai}^{1-\alpha} \right)^{1-\beta}. \tag{A.1}$$

We know each L_{Ai} from (4). By a similar logic used for labor we can establish that the allocation of capital to any individual location i is

$$K_{Ai} = A_i^{1/\beta} X_i \frac{K_A}{\sum_{j \in I} A_j^{1/\beta} X_j} \tag{A.2}$$

where K_A is the aggregate allocation of capital to agriculture. Combine (11) and (A.2) with the expression in (A.1) and we can solve for

$$Y_A = A_A \left(\frac{K_A}{L_A} \right)^{\alpha(1-\beta)} L_A^{1-\beta}$$

where

$$A_A = \left(\sum_{j \in I} A_j^{1/\beta} X_j \right)^{\beta}$$

is the measure of aggregate agricultural total factor productivity for the province.

With the assumption that land earns no return, and the share earned by capital is ϕ_K is both sectors, and for labor the share is ϕ_L in both sectors, it follows that the capital/labor ratio in both sectors is equal to the aggregate capital labor ratio,

$$\frac{K_A}{L_A} = \frac{K_N}{L_N} = \frac{K}{L} = \frac{w}{r} \frac{\phi_K}{\phi_L}.$$

Using the equilibrium condition on wages across sectors from (2), we can solve for

$$\frac{p_A}{p_N} = \frac{Y_N}{L_N} \frac{L_A}{Y_A}. \tag{A.3}$$

Noting that $Y_N = c_N L$ and $Y_A = c_A L$, we can rearrange this be

$$\frac{p_A c_A}{p_N c_N} = \frac{L_A}{L_N}, \tag{A.4}$$

which shows that the relative amount of labor employed in agriculture and non-agriculture is equal to the relative expenditures on those goods. With the adding up conditions $L_A + L_N = L$ and $p_A c_A + p_N c_N = M$, it follows that in log terms

$$\ln L_A/L = \ln p_A c_A/M. \tag{A.5}$$

Turning to the demand function from (10), we can re-arrange that to

$$(1-\epsilon)\ln p_A c_A/M = \ln\theta_A + (\epsilon-\gamma)(\ln p_N - \ln p_A) - \epsilon \ln c_A.$$

Using the relationships in (A.3) and (A.5), as well as the fact that $c_A = (Y_A/L_A)(L_A/L)$, we can substitute here to find

$$(1-\epsilon)\ln L_A/L = \ln\theta_A + (\epsilon-\gamma)\left(\ln Y_N/L_N - \ln Y_A/L_A\right) - \epsilon\left(\ln Y_A/L_A + \ln L_A/L\right).$$

Collecting terms we have

$$\ln L_A/L = \ln\theta_A + (\epsilon-\gamma)\ln Y_N/L_N - \gamma \ln Y_A/L_A.$$

Using the production functions in (9) and (A.1), we can write this as

$$\ln L_A/L = \ln\theta_A + (\epsilon-\gamma)\ln\left(A_N(K/L)^{\alpha}\right) - \gamma\ln\left(A_A(K/L)^{\alpha(1-\beta)}L_A^{-\beta}\right) - \gamma\beta\ln L + \gamma\beta\ln L,$$

where we've added and subtracted the term involving L. At this point, what remains is to separate the productivity and capital terms using the logs, and then straightforward algebra to arrive at

$$\ln L_A/L = \ln\theta_A + \frac{\beta\gamma}{1-\beta\gamma}\ln L - \frac{\gamma}{1-\beta\gamma}\ln A_A + \frac{\gamma-\epsilon}{1-\beta\gamma}\ln A_N + \frac{\alpha(\beta\gamma-\epsilon)}{1-\beta\gamma}\ln K/L.$$

Exponentiating this, we arrive at (12) from the main text.

For real income, in agricultural terms we have

$$y = \frac{M}{p_A} = c_A + \frac{p_N}{p_A}c_N.$$

Using (A.4) we can write this as

$$y = c_A + \frac{p_N c_N}{p_A c_A}c_A = c_A\left(1 + \frac{L_N}{L_A}\right) = c_A\frac{L}{L_A}.$$

Noting that $c_A = Y_A/L$, we have that

$$y = \frac{Y_A}{L_A} = A_A(K/L)^{\alpha(1-\beta)}(L_A/L)^{-\beta}L^{-\beta},$$

where the second equality follows from (A.1). At this point, we can use (12) to plug in for L_A/L in the above equation, and solve for

$$\ln y = \frac{1}{1-\beta\gamma}\ln A_A - \frac{\beta}{1-\beta\gamma}\ln L + \frac{\beta(\epsilon-\gamma)}{1-\beta\gamma}\ln A_N + \frac{\alpha(1-\beta)+\alpha\beta(\epsilon-\gamma)}{1-\beta\gamma}\ln K/L.$$

Exponentiating, we arrive at (13) in the main text.

A.3 Definitions of regions

Regions: Countries are included as follows:.

- **Central and West Asia**: Afghanistan, Azerbaijan, Bhutan, Georgia, Iran, Iraq, Jordan, Kazakhstan, Kyrgyzstan, Lebanon, Oman, Pakistan, Palestina, Russia (Asia), Syria, Tajikistan, Turkey, Uzbekistan
- **Eastern Europe**: Belarus, Bulgaria, Czech Republic, Hungary, Poland, Romania, Russia (Europe), Slovakia, Ukraine
- **North Africa**: Algeria, Egypt, Morocco, Sudan, Tunisia
- **Northwest Europe**: Austria, Belgium, Denmark, Estonia, Finland, France, Germany, Isle of Man, Latvia, Lithuania, Luxembourg, Netherlands, Norway, Sweden, Switzerland, United Kingdom
- **South Africa**: Botswana, Namibia, South Africa, Swaziland
- **South and Southeast Asia**: Bangladesh, Brunei, Cambodia, India, Indonesia, Laos, Malaysia, Myanmar, Philippines, Sri Lanka, Thailand, Timor-Leste, Vietnam
- **Southern Europe**: Albania, Bosnia and Herzegovina, Croatia, Greece, Italy, Portugal, Serbia, Slovenia, Spain
- **Temperate Americas**: Argentina, Canada, Chile, United States, Uruguay
- **Tropical Africa**: Angola, Benin, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Cte d'Ivoire, Democratic Republic of the Congo, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Liberia, Madagascar, Malawi, Mali, Mauritania, Mozambique, Niger, Nigeria, Republic of Congo, Reunion, Rwanda, Senegal, Sierra Leone, Somalia, South Sudan, So Tom and Prncipe, Tanzania, Togo, Uganda, Zambia, Zimbabwe
- **Tropical Americas**: Bolivia, Brazil, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, French Guiana, Guadeloupe, Guatemala, Guyana, Haiti, Honduras, Martinique, Mexico, Nicaragua, Panama, Paraguay, Peru, Suriname, Venezuela

For China-only regressions: We exclude Tibet, Xinjiang, Gansu, and Qinghai entirely, given that their climates do not fit well into the temperate versus sub-tropical distinction we make in the regressions.

- **Temperate provinces**: Hebei, Heilongjiang, Jilin, Liaoning, Nei Mongol, Ningxia Hui, Shaanxi, Shanxi, Tianjin, Sichuan, Shandong, Yunnan
- **Sub-tropical provinces**: Guangxi, Guangdong, Fujian, Jiangxi, Hunan, Guizhou, Chongqing, Hubei, Anhui, Zhejiang, Henan, Jiangsu, Hainan

Russian provinces: We split Russia into separate Asian and European sections for inclusion in the regions. That breakdown takes place at the province level

- **Russia(Asia)**: Altay, Amur, Buryat, Chelyabinsk, Gorno-Altay, Irkutsk, Kemerovo, Khabarovsk, Khakass, Khanty-Mansiy, Krasnoyarsk, Kurgan, Novosibirsk, Omsk, Primor'ye, Sakhalin, Sverdlovsk, Tomsk, Tuva, Tyumen', Yevrey, Zabaykal'ye

- **Russia(Europe)**: Adygey, Arkhangel'sk, Astrakhan', Bashkortostan, Belgorod, Bryansk, Chechnya, Chuvash, City of St. Petersburg, Dagestan, Ingush, Ivanovo, Kabardin-Balkar, Kaliningrad, Kalmyk, Kaluga, Karachay-Cherkess, Karelia, Kirov, Komi, Kostroma, Krasnodar, Kursk, Leningrad, Lipetsk, Mariy-El, Mordovia, Moscow City, Moskva, Nizhegorod, North Ossetia, Novgorod, Orel, Orenburg, Penza, Perm', Pskov, Rostov, Ryazan', Samara, Saratov, Smolensk, Stavropol', Tambov, Tatarstan, Tula, Tver', Udmurt, Ul'yanovsk, Vladimir, Volgograd, Vologda, Voronezh, Yaroslavl'

A.4 Alternative Population Data

As mentioned in the main text, given possible issues with the HYDE data on population, we use two alternative sources of population data. These show that our results are not contingent on the particular data from HYDE.

A.4.1 GRUMP Data

We accessed both the gridded population map, and the urban extents grid. The GRUMP data is provided at 30 arc-second grids (rougly 1km squares), a higher resolution than the HYDE data (which has 5 arc-minute grids, or roughly 10km squares). We overlay the urban extents grid on the population map, and retrieve only the population count of grid cells that are not part of an urban extent. We then sum up the population count of grid cells within each district. The district definitions are from GADM, identical to those used with the HYDE data, so we can compare the counts directly.

Because GRUMP counts zero rural residents in an area that is part of an urban extent, and all population in non-urban locations as rural, the variation in rural residents across districts is more severe than with HYDE. Some districts in GRUMP are entirely covered by urban extents, and so have zero rural residents. Hence the GRUMP data leads to 28,471 districts with data on rural density, compared to 35,451 using HYDE. For those 28,471 districts, the correlation of (log) rural density across the two datasets is 0.81, significant at less than 1%.

The following table shows the results using the GRUMP data. The results are consistent with our baseline findings. The elasticity for wheat family suitable districts is about 0.20, while for rice family suitable areas it is only 0.115, a difference of about 0.10 that is significant at less than 1%. Panel B shows that these results hold up if we restrict the sample due to urban size, country development, or the density of rural workers.

The only discrepancy with the baseline results is in Panel A, columns (5) and (6), which distinguish samples by their harvested area. In this case, the estimated β values are similar (statistically and practically). This appears to be due to a set of outliers in the GRUMP data that have very low reported rural density relative to the equivalent HYDE areas. We can identify these outliers by regressing the GRUMP (log) rural density on the HYDE (log) rural density, and looking for observations whose residuals from that regression are in the bottom 1% (meaning GRUMP's estimate is well below the HYDE estimate). If we eliminate those from the GRUMP data, then the results for columns (5) and (6) are much closer to the baseline results. The underlying issue here is that these districts have many cells coded as "urban" by GRUMP, but HYDE's data indicates large rural populations. Without better information, we do not know if this is an error of HYDE's or GRUMP's methodologies.

A.4.2 IPUMS Data

We use 39 countries that have both geographic identifier data (the GEOLEV2 variable from IPUMS) as well as information on individual industry of employment. We create a 0/1 variable indicating whether an individual is an agricultural worker (meaning they are reported as in the workforce). We then aggregate this variable (weighted by their IPUMS provided sampling weight) across individuals within a geographic area to get a count of the total agricultural workers. Using a similar method, we are also able to count the number of urban residents, which allows us to measure the percent urban within a geographic area. We end up with a total of 8,393 geographic areas.

Before we run regressions, the IPUMS data is useful in assessing how good of an approximation rural population (including workers and non-workers) is for the number of agricultural workers. The correlation of (log) rural residents and (log) agricultural workers across the areas is 0.91, significant at less than 1%. There are a few outliers where the number of agricultural workers is high relative to rural residents, which likely represents agricultural processing work in urban areas, or urban farmers with small plots. Our results are not affected by excluding these outliers.

The geographic areas provided by IPUMS in the GEOLEV2 variable are in many cases agglomerations of the districts we use from GADM. This is because IPUMS aggregates districts with fewer than 25000 observations (to protect anonymity) or districts whose boundaries have changed over time (so that the agglomerations are comparable over time for a given country). This means the IPUMS geographic areas are not directly comparable to our districts. Because the IPUMS agglomerations are much larger than districts, it is not practical to use province/state fixed effects, as most of these have only one or two IPUMS areas within them. Hence we run our regressions only with country fixed effects. Because the GEOLEV2 areas are different than the districts in our baseline specifications, we create new GEOLEV2 level versions of our caloric suitability index, night lights data, and other crop suitability measures.

The following table shows the results using the IPUMS data. For the main results using the suitability for wheat family and rice family crops, the results show a similar pattern to what we found in our baseline. The estimated β in wheat-suitable areas is 0.213, compared to 0.025 in rice-suitable areas, and the difference is statistically significant.

In columns (3) and (4), the samples are selected on which crops we find deliver the maximum calories within an area. A slight modification is made here to the baseline selection criteria, because the IPUMS areas are so large. Column (3) includes areas that have any of their maximum calories derived from the wheat family, and *none* from the rice family. Column (4) excludes areas where *none* of the maximum calories from the wheat family, and includes areas with *any* maximum calories derived from the rice family. Here, we again see that wheat family areas have an elasticity around 0.20, while the rice family elasticity is estimated to be zero, and the difference is statistically significant. In this case, we cannot reject the hypothesis that there is no land constraint in these rice-family dominated areas. In columns (5) and (6), based on actual harvested area, we again get results similar to our baseline specification using HYDE data, and again the rice family coefficient is estimated to be very small.

We cannot replicate the second panel of our baseline results, for several reasons. First, the areas from IPUMS are so large that they almost universally contain large cities, so there is no point in removing areas with large cities. Second, the 39 countries are dominated by poor countries, so eliminating "rich" countries does not change the results (it eliminates only the U.S., Spain, and Austria).

Finally, the 39 countries included from IPUMS are, with the census date listed: Argentina (2001), Austria (2001), Bolivia (2001), Brazil (2000), Cambodia (1998), Cameroon (2005), Chile (2002), Colombia (2005), Costa Rica (2000), Ecuador (2001), El Salvador (2007), Fiji (1996), Ghana (2000), Greece (2001), Haiti (2003), India (1999), Iran (2006), Iraq (1997), Jordan (2004), Kyrghzstan (1999), Malawi (1998), Mexico (2000), Morocco (2004), Mozambique (1997), Panama (2000), Peru (2007), Sierra Leone (2004), South Africa (2001), Spain (2001), South Sudan (2008), Sudan (2008), Turkey (2000), Uganda (2002), Egypt (1996), Tanzania (2002), United States (2000), Burkina Faso (1996), Venezuela (2001), Zambia (2000)

A.5 Robustness Tables

Each table that follows is a replica of Table 2 from the main paper, which estimates β, the land elasticity, for sub-samples of districts distinguished by their suitability or production of different crops.

Here we list the baseline assumptions behind each table, rather than replicating the same footnotes over and over again. In each case, these are the baseline assumptions, and the individual table may change or drop the assumption, as will be noted in each table in bold.

Conley standard errors, adjusted for spatial auto-correlation with a cutoff distance of 500km, are shown in parentheses. All regressions include province fixed effects, a constant, and controls for the district urbanization rate and log density of district nighttime lights. Rural population is from HYDE database, and caloric yield is the author's calculations based on the data from Galor and Ozak (2016), see the main paper for an explanation of the construction of both.

A list of tables is provided for convenience. The first table (A.1), replicates the baseline results for comparison purposes.

List of Tables

Table A.1: Baseline results

Dependent Variable in all panels: Log caloric yield (A_{isc})

Panel A: Samples defined by crop family (wheat vs. rice):

	By suitability:		By max calories:		By harvest area:	
	Wheat Only (1)	Rice Only (2)	Wheat $> 33\%$ (3)	Rice $> 33\%$ (4)	Wheat $> 50\%$ (5)	Rice $> 50\%$ (6)
Log rural density	0.228	0.132	0.191	0.112	0.205	0.133
	(0.021)	(0.018)	(0.016)	(0.017)	(0.015)	(0.012)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.000		0.001		0.000
Countries	91	81	83	71	74	84
Observations	10661	9088	10786	8217	10708	7564
Adjusted R-square	0.24	0.20	0.21	0.18	0.20	0.18

Panel B: Samples with other restrictions (using suitability to distinguish crop families)

	Urban Pop. $< 25K$:		Ex. Europe/N. Amer.:		Rural dens. $>$ 25th P'tile:	
	Wheat Only (1)	Rice Only (2)	Wheat Only (3)	Rice Only (4)	Wheat Only (5)	Rice Only (6)
Log rural density	0.261	0.143	0.242	0.133	0.281	0.185
	(0.022)	(0.021)	(0.033)	(0.018)	(0.035)	(0.019)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.000		0.003		0.015
Countries	83	75	24	70	89	77
Observations	7648	6662	824	8826	7237	7082
Adjusted R-square	0.29	0.24	0.19	0.14	0.27	0.22

Baseline results

Table A.2: Using GRUMP Population Data

Dependent Variable in all panels: Log caloric yield (A_{isc})

Panel A: Samples defined by crop family (wheat vs. rice):

	By suitability:		By max calories:		By harvest area:	
	Wheat Only (1)	Rice Only (2)	Wheat $> 33\%$ (3)	Rice $> 33\%$ (4)	Wheat $> 50\%$ (5)	Rice $> 50\%$ (6)
Log rural density	0.207	0.115	0.176	0.100	0.166	0.140
	(0.041)	(0.021)	(0.033)	(0.017)	(0.028)	(0.020)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.045		0.041		0.442
Countries	86	75	81	69	71	82
Observations	8734	6769	8585	6230	8922	5844
Adjusted R-square	0.19	0.16	0.15	0.13	0.14	0.13

Panel B: Samples with other restrictions (using suitability to distinguish crop families)

	Urban Pop. $< 25K$:		Ex. Europe/N. Amer.:		Rural dens. $>$ 25th P'tile:	
	Wheat Only (1)	Rice Only (2)	Wheat Only (3)	Rice Only (4)	Wheat Only (5)	Rice Only (6)
Log rural density	0.264	0.129	0.195	0.115	0.298	0.191
	(0.047)	(0.024)	(0.019)	(0.022)	(0.048)	(0.031)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.010		0.002		0.056
Countries	81	71	23	67	80	68
Observations	6431	5050	777	6697	6078	5440
Adjusted R-square	0.24	0.20	0.13	0.11	0.24	0.21

Using GRUMP to measure rural density in place of HYDE.

Table A.3: Using IPUMS Population Data

Dependent Variable in all panels: Log caloric yield (A_{isc})

Panel A: Samples defined by crop family (wheat vs. rice):

	By suitability:		By max calories:		By harvest area:	
	Wheat Only (1)	Rice Only (2)	Wheat/ No Rice (3)	Rice No Wheat (4)	Wheat $> 50\%$ (5)	Rice $> 50\%$ (6)
Log ag. worker density	0.213	0.025	0.200	0.000	0.223	0.034
	(0.067)	(0.016)	(0.056)	(0.017)	(0.030)	(0.014)
p-value $\beta = 0$	0.004	0.124	0.002	0.993	0.000	0.021
p-value $\beta = \beta^{Wheat}$		0.006		0.000		0.000
Countries	23	24	24	23	21	26
Observations	1104	2416	1595	2389	1207	1427
Adjusted R-square	0.50	0.54	0.39	0.56	0.37	0.51

Clustered standard errors at the country level are shown in parentheses. All regressions include country fixed effects, and controls for (log) night lights and urban percent. Agricultural worker density is from IPUMS, and caloric yield is the author's calculations based on the data from Galor and Ozak (2016), see the main paper for an explanation of the construction.

Table A.4: Conley SE cutoff of 1000km

Dependent Variable in all panels: Log caloric yield (A_{isc})

Panel A: Samples defined by crop family (wheat vs. rice):

	By suitability:		By max calories:		By harvest area:	
	Wheat Only (1)	Rice Only (2)	Wheat $> 33\%$ (3)	Rice $> 33\%$ (4)	Wheat $> 50\%$ (5)	Rice $> 50\%$ (6)
Log rural density	0.240	0.143	0.200	0.114	0.220	0.126
	(0.028)	(0.020)	(0.022)	(0.023)	(0.021)	(0.014)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.004		0.006		0.000
Countries	91	79	82	71	74	84
Observations	9922	8396	10142	7411	9929	6810
Adjusted R-square	0.24	0.20	0.21	0.17	0.20	0.17

Panel B: Samples with other restrictions (using suitability to distinguish crop families)

	Urban Pop. $< 25K$:		Ex. Europe/N. Amer.:		Rural dens. $>$ 25th P'tile:	
	Wheat Only (1)	Rice Only (2)	Wheat Only (3)	Rice Only (4)	Wheat Only (5)	Rice Only (6)
Log rural density	0.279	0.156	0.253	0.143	0.289	0.188
	(0.027)	(0.023)	(0.041)	(0.021)	(0.039)	(0.024)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.001		0.018		0.028
Countries	83	74	24	69	89	74
Observations	7046	6117	785	8168	6807	6606
Adjusted R-square	0.29	0.24	0.18	0.14	0.26	0.22

Use 1000km to form cutoffs for Conley standard errors

Table A.5: Province-level data

Dependent Variable in all panels: Log caloric yield (A_{isc})

Panel A: Samples defined by crop family (wheat vs. rice):

	By suitability:		By max calories:		By harvest area:	
	Wheat Only (1)	Rice Only (2)	Wheat $> 33\%$ (3)	Rice $> 33\%$ (4)	Wheat $> 50\%$ (5)	Rice $> 50\%$ (6)
Log rural density	0.399	0.070	0.248	0.016	0.368	0.052
	(0.058)	(0.020)	(0.030)	(0.013)	(0.043)	(0.021)
p-value $\beta = 0$	0.000	0.000	0.000	0.199	0.000	0.014
p-value $\beta = \beta^{Wheat}$		0.000		0.000		0.000
Countries	60	65	70	63	69	73
Observations	417	587	768	617	797	721
Adjusted R-square	0.39	0.27	0.29	0.26	0.35	0.30

Panel B: Samples with other restrictions (using suitability to distinguish crop families)

	Urban Pop. $< 25K$:		Ex. Europe/N. Amer.:		Rural dens. $>$ 25th P'tile:	
	Wheat Only (1)	Rice Only (2)	Wheat Only (3)	Rice Only (4)	Wheat Only (5)	Rice Only (6)
Log rural density	0.505	0.057	0.038	0.073	0.193	0.047
	(0.127)	(0.022)	(0.134)	(0.020)	(0.058)	(0.021)
p-value $\beta = 0$	0.000	0.012	0.780	0.000	0.001	0.023
p-value $\beta = \beta^{Wheat}$		0.001		0.797		0.019
Countries	13	28	6	59	49	61
Observations	28	89	11	557	234	470
Adjusted R-square	0.54	0.34	-0.09	0.05	0.13	0.06

Using provinces as the units of observation, with country fixed effects. Night lights and urban percent controls are at the province level.

Table A.6: Using cultivated area to measure density

Dependent Variable in all panels: Log caloric yield (A_{isc})

Panel A: Samples defined by crop family (wheat vs. rice):

	By suitability:		By max calories:		By harvest area:	
	Wheat Only (1)	Rice Only (2)	Wheat $> 33\%$ (3)	Rice $> 33\%$ (4)	Wheat $> 50\%$ (5)	Rice $> 50\%$ (6)
Log rural density	0.229	0.144	0.191	0.113	0.207	0.142
	(0.024)	(0.020)	(0.020)	(0.021)	(0.020)	(0.015)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.006		0.006		0.010
Countries	90	76	82	68	74	81
Observations	9871	8295	10100	7343	9911	6749
Adjusted R-square	0.20	0.17	0.17	0.15	0.16	0.15

Panel B: Samples with other restrictions (using suitability to distinguish crop families)

	Urban Pop. $< 25K$:		Ex. Europe/N. Amer.:		Rural dens. $>$ 25th P'tile:	
	Wheat Only (1)	Rice Only (2)	Wheat Only (3)	Rice Only (4)	Wheat Only (5)	Rice Only (6)
Log rural density	0.277	0.161	0.248	0.146	0.262	0.170
	(0.021)	(0.024)	(0.038)	(0.021)	(0.032)	(0.023)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.000		0.014		0.019
Countries	82	72	23	67	90	75
Observations	7000	6025	778	8092	6263	7175
Adjusted R-square	0.26	0.22	0.17	0.14	0.21	0.18

Rural density measured using rural population per hectare of cultivated land. Also includes a control for cultivated land as a percent of total land.

Table A.7: Using population from 1900CE

Dependent Variable in all panels: Log caloric yield (A_{isc})

Panel A: Samples defined by crop family (wheat vs. rice):

	By suitability:		By max calories:		By harvest area:	
	Wheat Only (1)	Rice Only (2)	Wheat $> 33\%$ (3)	Rice $> 33\%$ (4)	Wheat $> 50\%$ (5)	Rice $> 50\%$ (6)
Log rural density	0.240	0.143	0.200	0.114	0.220	0.126
	(0.025)	(0.018)	(0.021)	(0.018)	(0.020)	(0.013)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.001		0.002		0.000
Countries	91	79	82	71	74	84
Observations	9922	8396	10142	7411	9929	6810
Adjusted R-square	0.24	0.20	0.21	0.17	0.20	0.17

Panel B: Samples with other restrictions (using suitability to distinguish crop families)

	Urban Pop. $< 25K$:		Ex. Europe/N. Amer.:		Rural dens. $>$ 25th P'tile:	
	Wheat Only (1)	Rice Only (2)	Wheat Only (3)	Rice Only (4)	Wheat Only (5)	Rice Only (6)
Log rural density	0.279	0.156	0.253	0.143	0.289	0.188
	(0.023)	(0.021)	(0.044)	(0.019)	(0.038)	(0.020)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.000		0.019		0.018
Countries	83	74	24	69	89	74
Observations	7046	6117	785	8168	6807	6606
Adjusted R-square	0.29	0.24	0.18	0.14	0.26	0.22

Rural density measured using population data from 1900CE from HYDE database.

Table A.8: Using population from 1950CE

Dependent Variable in all panels: Log caloric yield (A_{isc})

Panel A: Samples defined by crop family (wheat vs. rice):

	By suitability:		By max calories:		By harvest area:	
	Wheat Only (1)	Rice Only (2)	Wheat $> 33\%$ (3)	Rice $> 33\%$ (4)	Wheat $> 50\%$ (5)	Rice $> 50\%$ (6)
Log rural density	0.240	0.143	0.200	0.114	0.220	0.126
	(0.025)	(0.018)	(0.021)	(0.018)	(0.020)	(0.013)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.001		0.002		0.000
Countries	91	79	82	71	74	84
Observations	9922	8396	10142	7411	9929	6810
Adjusted R-square	0.24	0.20	0.21	0.17	0.20	0.17

Panel B: Samples with other restrictions (using suitability to distinguish crop families)

	Urban Pop. $< 25K$:		Ex. Europe/N. Amer.:		Rural dens. $>$ 25th P'tile:	
	Wheat Only (1)	Rice Only (2)	Wheat Only (3)	Rice Only (4)	Wheat Only (5)	Rice Only (6)
Log rural density	0.279	0.156	0.253	0.143	0.289	0.188
	(0.023)	(0.021)	(0.044)	(0.019)	(0.038)	(0.020)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.000		0.019		0.018
Countries	83	74	24	69	89	74
Observations	7046	6117	785	8168	6807	6606
Adjusted R-square	0.29	0.24	0.18	0.14	0.26	0.22

Rural density measured using population data from 1950CE from HYDE database.

Table A.9: Dropping districts under 25th percentile in production

Dependent Variable in all panels: Log caloric yield (A_{isc})

Panel A: Samples defined by crop family (wheat vs. rice):

	By suitability:		By max calories:		By harvest area:	
	Wheat Only (1)	Rice Only (2)	Wheat $> 33\%$ (3)	Rice $> 33\%$ (4)	Wheat $> 50\%$ (5)	Rice $> 50\%$ (6)
Log rural density	0.226	0.140	0.186	0.111	0.213	0.125
	(0.025)	(0.020)	(0.017)	(0.021)	(0.018)	(0.013)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.008		0.005		0.000
Countries	82	65	77	58	70	72
Observations	7568	6092	7540	5374	8400	5704
Adjusted R-square	0.22	0.18	0.19	0.16	0.19	0.16

Panel B: Samples with other restrictions (using suitability to distinguish crop families)

	Urban Pop. $< 25K$:		Ex. Europe/N. Amer.:		Rural dens. $>$ 25th P'tile:	
	Wheat Only (1)	Rice Only (2)	Wheat Only (3)	Rice Only (4)	Wheat Only (5)	Rice Only (6)
Log rural density	0.272	0.149	0.243	0.141	0.271	0.183
	(0.027)	(0.023)	(0.046)	(0.020)	(0.044)	(0.023)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.001		0.043		0.082
Countries	73	64	18	62	78	63
Observations	5093	4127	582	6036	5156	4982
Adjusted R-square	0.27	0.23	0.15	0.13	0.23	0.19

Drops all districts below the 25th percentile of total tonnes of staple crops produced across all districts. Raw tonnes are used, unadjusted for calorie content.

Table A.10: Using log rural percent of population as a control

Dependent Variable in all panels: Log caloric yield (A_{isc})

Panel A: Samples defined by crop family (wheat vs. rice):

	By suitability:		By max calories:		By harvest area:	
	Wheat Only (1)	Rice Only (2)	Wheat $> 33\%$ (3)	Rice $> 33\%$ (4)	Wheat $> 50\%$ (5)	Rice $> 50\%$ (6)
Log rural density	0.254	0.148	0.213	0.120	0.231	0.136
	(0.024)	(0.019)	(0.021)	(0.020)	(0.020)	(0.015)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.001		0.001		0.000
Countries	91	79	82	71	74	84
Observations	9922	8396	10142	7411	9929	6810
Adjusted R-square	0.25	0.21	0.22	0.18	0.21	0.18

Panel B: Samples with other restrictions (using suitability to distinguish crop families)

	Urban Pop. $< 25K$:		Ex. Europe/N. Amer.:		Rural dens. $>$ 25th P'tile:	
	Wheat Only (1)	Rice Only (2)	Wheat Only (3)	Rice Only (4)	Wheat Only (5)	Rice Only (6)
Log rural density	0.286	0.159	0.288	0.149	0.299	0.194
	(0.025)	(0.022)	(0.041)	(0.020)	(0.036)	(0.020)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.000		0.002		0.012
Countries	83	74	24	69	89	74
Observations	7046	6117	785	8168	6807	6606
Adjusted R-square	0.30	0.25	0.21	0.15	0.27	0.22

Include log rural percent of the population as a control, consistent with a model of districts being autarkic.