How Tight are Malthusian Constraints?

T. Ryan Johnson [1]   Dietrich Vollrath [2]

[1]University of Houston

[2]University of Houston

Translation

What is the elasticity (i.e. how tight) of agricultural output with respect to land (i.e. the Malthusian constraint)?
What we do:

- Estimate the elasticity using district-level variation in rural density
- Estimate elasticity separately for different climate zones, agric. type, etc..
- Show elasticity is about 2x higher for temperate (0.20-0.30) vs. tropics (0.10-0.15)
- Show differences are robust w.r.t. data sources, measurement error

Why do you care?

The elasticity determines the degree of decreasing returns to scale for labor/capital in agriculture.

- **Theory**: The higher the elasticity, the more sensitive are real income and the agric. labor share to shocks in population/productivity.
- **Evidence**: Using epidemiological transition after WWII a la Acemoglu and Johnson (2007), countries with high elasticities had more severe negative effect of rising life expectancy
- **Speculation**: Informative about historical development (Asia vs. Europe) and contemporary development (delayed development in tropics?). Geography as mediator, not cause.

Density and Productivity

Province I contains a set of districts, each denoted by i, with aggregate agricultural production

$$Y_i = A_i X_i^{\beta} \left(K_{Ai}^{\alpha} L_{Ai}^{1-\alpha} \right)^{1-\beta} \tag{1}$$

- A_i is productivity, X_i is land
- K_{Ai} is all other inputs (e.g. capital)
- L_{Ai} is agricultural labor in district i (not a single sector)
- Assume β and α are identical *within* province (but not nec. *across* provinces)

Mobile Factors

The wage and return to capital in each district are given by

$$
\begin{aligned}
w &= \phi_L \frac{Y_i}{L_{Ai}} \\
r &= \phi_K \frac{Y_i}{K_{Ai}}
\end{aligned} \tag{2}
$$

- ϕ_L and ϕ_K are shares of output
- Shares need not equal elasticities
- Shares are identical *within* province (but not nec. *across* provinces)
- Capital and labor are mobile *within* province (but not nec. *across* provinces)

Solving for Labor Allocations

Given mobility of labor and capital within province,

$$\frac{K_{Ai}}{L_{Ai}} = \frac{w}{r}\frac{\phi_K}{\phi_L}. \tag{3}$$

Adding up condition for agricultural labor within province

$$\sum_{i \in I} L_{Ai} = L_A. \tag{4}$$

Solve for allocation of labor (relative to land) to district i

$$\frac{L_{Ai}}{X_i} = A_i^{1/\beta} \frac{L_A}{\sum_{j \in I} A_j^{1/\beta} X_j}. \tag{5}$$

Agricultural Labor Allocation

In log terms:

$$\ln L_{Ai}/X_i = \frac{1}{\beta}\ln A_i + \ln \frac{L_A}{\sum_{j \in I} A_j^{1/\beta} X_j}, \tag{6}$$

where second term is *province-specific*.

- β can be estimated from elasticity of L_{Ai}/X_i w.r.t. A_i.
- β large (tight), $1/\beta$ small, ag. workers spread evenly w/in province
- β small (loose), $1/\beta$ large, ag. workers concentrated on high A_i
- Ag labor relative to total labor (L_A/L) does not enter
- Expression is not unique to heaviliy agricultural provinces (or eras)

More

Using as a Specification

Re-arranging the prior expression and adding some notation:

$$\ln A_{isg} = \alpha_g + \beta_g \ln L_{Aisg}/X_{isg} + \gamma_s + \delta_g' \mathbf{Z}_{isg} + \epsilon_{isg}. \tag{7}$$

- District i, region/state/province s
- g indicates a geographic area (e.g. temperate)
- γ_s, province FE, picks up province-specific term
- $\mathbf{Z}_{isg}$ are additional controls
- ϵ_{isg} is error term

Using as a Specification

We moved productivity, A_{isg} and agric. density L_{Aisg}/X_{isg} to opposite sides:

$$\ln A_{isg} = \alpha_g + \beta_g \ln L_{Aisg}/X_{isg} + \gamma_s + \delta_g' \mathbf{Z}_{isg} + \epsilon_{isg}. \tag{8}$$

- Estimating $1/\beta \Rightarrow$ very sensitive to small differences in β

Agricultural Density Data

L_{Aisg} comes from HYDE 3.1 database (Goldewijk et al, 2011)

- Population counts for 5 degree grid-cells built from administrative data
- We aggregate data back to administrative level (e.g. districts)
- Rural population data (not agricultural)
- Main samples based on year 2000

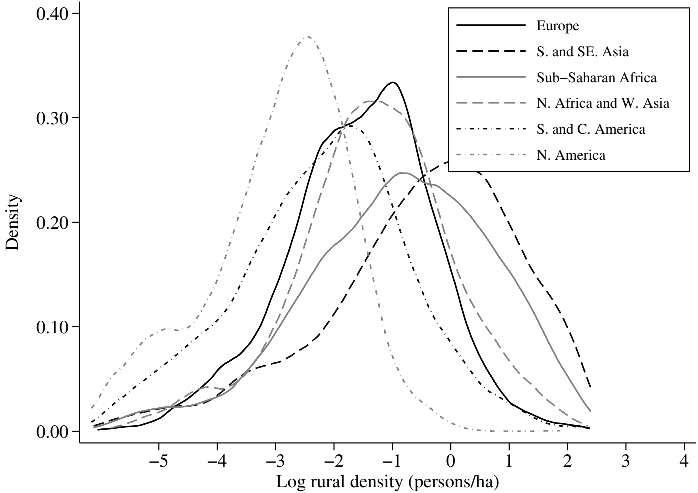
X_{isg} calculated as area of a given district

- Measure of possible agricultural land

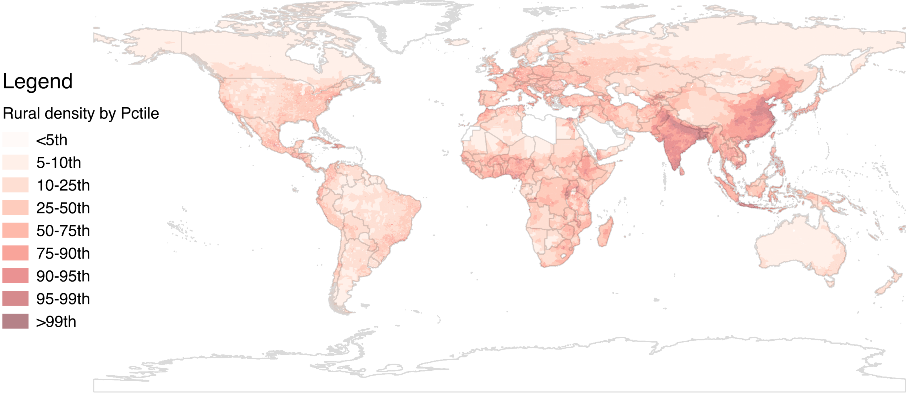
L_{Aisg}/X_{isg} data

- Trim above 99th and below 1st percentiles
- Drop if fewer than 100 total rural residents
- 35,451 total districts from around world

Agricultural Density Data



Agricultural Density Data



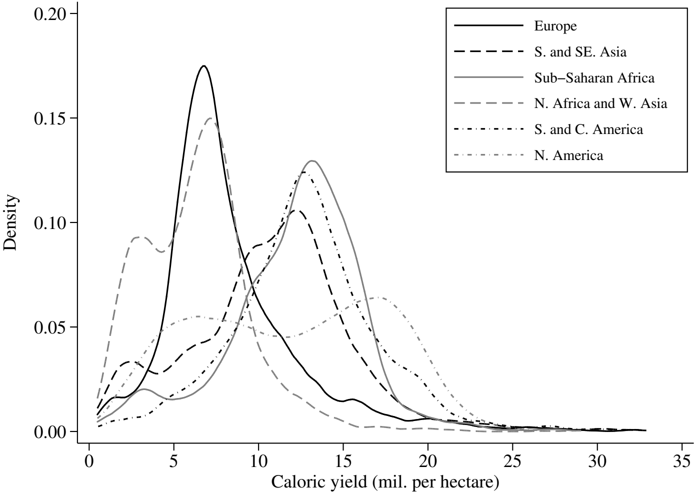
Agricultural Productivity Data

A_{isg} is built from Galor and Özak (2016) caloric suitability index

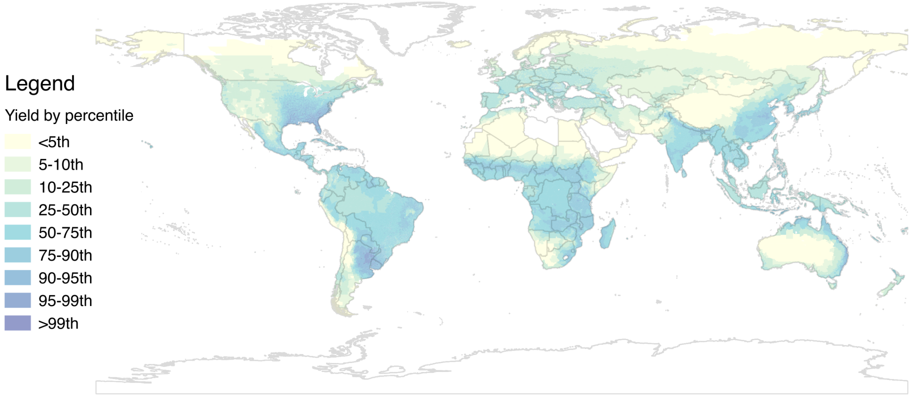
- Data from GAEZ on agro-climatic possible yield (in raw tons) for each crop
- Combine with nutritional information by crop (total calories per raw ton)
- For each grid cell, determine max calories across all crops
- Sum max calories across grid cells in district, divide by total area
- Holds technology assumptions constant
- Trim above 99th and below 1st percentile

Summary
Crops

Agricultural Productivity Data



Agricultural Productivity Data



Control Variables

Henderson et al (2016) on spatial distribution of economic activity

- Urban activity correlated with (caused by?) high agricultural productivity (in some places)
- Low rural density because of urban activity
- $Corr(\epsilon_{isg}, \ln L_{isg}/X_{isg}) < 0$

Include two controls at the district level in $\mathbf{Z}_{isg}$ for urban/economic activity:

- Night lights density: follows Henderson et al (2016) using Global Radiance Calibrated data
- Urban percent of population: from HYDE

“Temperate” versus “Tropical”

How do we define the geographic types g?

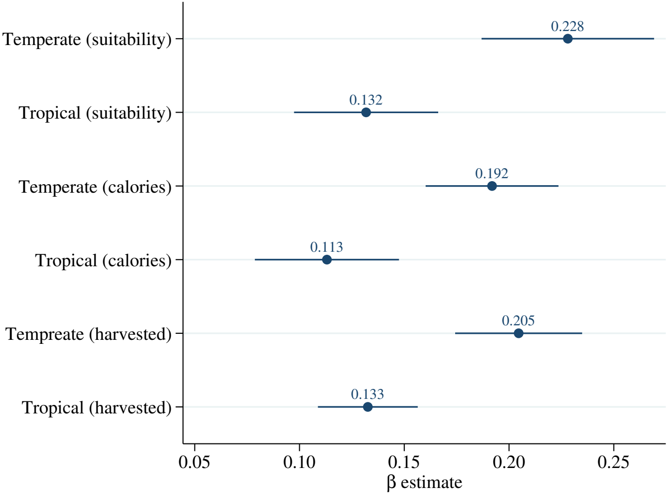
- **Temperate**: suitable for barley, buckwheat, rye, oats, white potatoes, and/or wheat but zero suitability for Tropical crops
- **Tropical**: suitable for cassava, cowpeas, pearl millet, sweet potatoes, paddy rice, and/or yams but zero suitability for Temperate crops

Measure suitability several ways

- GAEZ suitability indices (0 to 100), OR
- Source of maximum calories in our A_{isg} measure, OR
- Actual harvested area

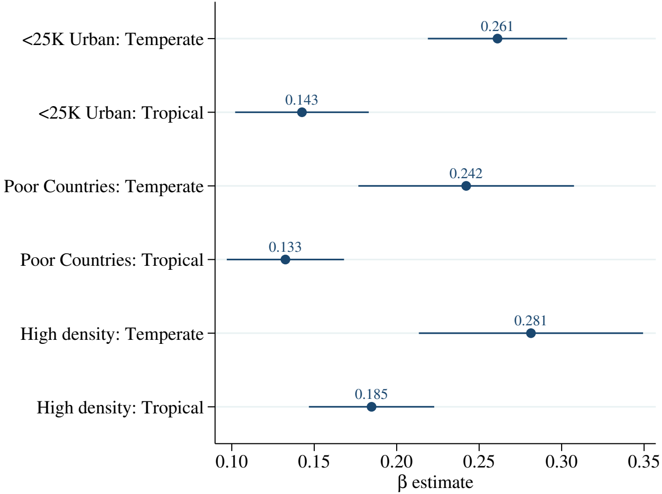
Definition of geographic type is specific to a province, not a country, so heterogeneity of β within country

Results by Temperate/Tropical



Table

Results by Temperate/Tropical

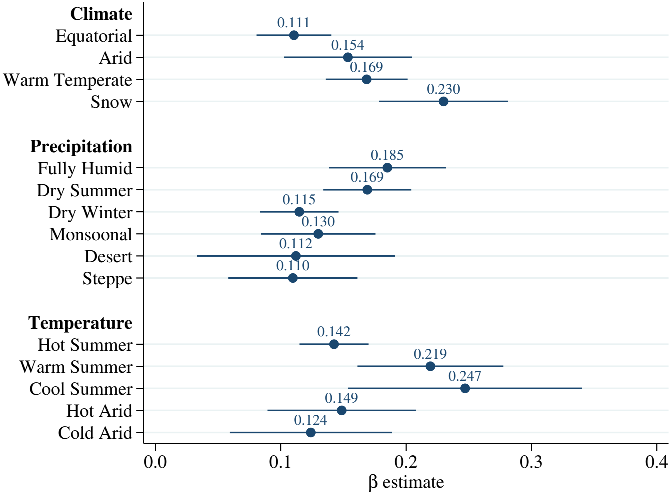


Results by Climate Zone

Define g by looking at specific climate factors, rather than crops grown

- Create samples based on Köppen-Geiger zones
- Three layers: Climate, Precipitation, Temperature
- Each layer has multiple types (e.g. Climate is Equatorial, Arid, ...)
- Create samples where districts have >50% of land in a given type
- $\Rightarrow$ heterogeneity within countries in β

Results by Climate Zone



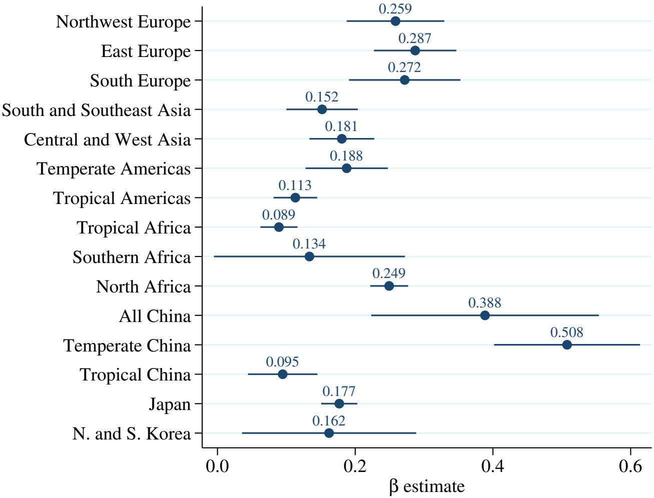
Table

Results by Political Region

Define g by grouping countries based on common usage

- Assumes homogeneity of β for provinces within countries
- General overlap of geographies and these political regions

Results by Region



Table

Explanations?

Why looser constraints in tropics compared to temperate areas?

- Positive(?): Multiple cropping, longer growing periods, more sun, more rain during growing periods $\Rightarrow$ land area isn't binding?
- Negative(?): Soil leaching, lack of frost $\Rightarrow$ land is less useful?
- Neutral(?): Crops have unique requirements for labor inputs versus land inputs?

Working on new project that uses actual crop yields to back out elasticities *by crop* within a given grid cell, holding geographic conditions constant.

Robustness and questions

- Use province level data (with country FE) Results
- Use rural density for 1900 from HYDE Results
- Use IPUMS for agricultural population Results
- Use GRUMP for population data Results
- Estimate β for individual provinces Results
- Use cultivated area of land Results
- Maize and soy Results
- Exclude districts with few harvest Ha Results
- Workers not mobile between districts? Slides
- Districts autarkic? Slides
- Elasticity of substitution? Slides
- Measurement error? Slides
- Factor shares? Slides

Back to the model

Aggregate production of agricultural good

$$Y_A = A_A \left(\frac{K_A}{L_A}\right)^{\alpha(1-\beta)} L_A^{1-\beta}, \tag{9}$$

where

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

and non-agricultural good

$$Y_N = A_N \left(\frac{K_N}{L_N}\right)^{\alpha} L_N. \tag{10}$$

Factor shares and mobility

Land earns zero return

- No effect of β on factor share
- Let ϕ_L and ϕ_K be factor shares of labor and capital in both sectors

Mobility of labor between sectors

$$p_A \phi_L \frac{Y_A}{L_A} = p_N \phi_L \frac{Y_N}{L_N}. \tag{11}$$

where p_A and p_N are nominal prices of agric. and non-agric.

Mobility of capital implies $K_A/L_A = K_N/L_N = K/L$.

Preferences

Following Boppart (2014), there exists a utility function such that

$$\ln c_A = \ln \theta_A + (1-\epsilon)\ln M + (\gamma - 1)\ln p_A + (\epsilon - \gamma)\ln p_N \tag{12}$$

is the demand for c_A.

- θ_A is a preference parameter
- M is nominal income
- $0 < \epsilon < 1$ to capture Engel's Law
- $\epsilon > \gamma$ means willingness to substitute between c_A and c_N

Solving

Agricultural labor share is

$$\frac{L_A}{L} = \theta_A \left(\frac{L^{\beta\gamma}}{A_A^{\gamma} A_N^{\epsilon-\gamma} \hat{k}^{\alpha(\epsilon-\beta\gamma)}} \right)^{\frac{1}{1-\beta\gamma}} \tag{13}$$

while real income (in agricultural terms, M/p_A)

$$y = \left(\frac{A_A A_N^{\beta(\epsilon-\gamma)} \hat{k}^{\Omega}}{L^{\beta}} \right)^{\frac{1}{1-\beta\gamma}} \tag{14}$$

where $\hat{k} = (\phi_K K/\phi_L L)$, and $\Omega = \alpha(1-\beta) + \alpha\beta(\epsilon-\gamma)$

Elasticities

The elasticities of the agricultural labor share (L_A/L) and real income (y) with respect to various shocks,

(a) Agricultural productivity (A_A):

$$\frac{\partial \ln L_A/L}{\partial \ln A_A} = -\frac{\gamma}{1-\beta\gamma} \qquad \frac{\partial \ln y}{\partial \ln A_A} = \frac{1}{1-\beta\gamma} \tag{15}$$

(b) Non-agricultural productivity (A_N):

$$\frac{\partial \ln L_A/L}{\partial \ln A_N} = -\frac{\epsilon-\gamma}{1-\beta\gamma} \qquad \frac{\partial \ln y}{\partial \ln A_N} = \frac{\beta(\epsilon-\gamma)}{1-\beta\gamma} \tag{16}$$

(c) Population (L):

$$\frac{\partial \ln L_A/L}{\partial \ln L} = \frac{\beta\gamma}{1-\beta\gamma} \qquad \frac{\partial \ln y}{\partial \ln L} = -\frac{\beta}{1-\beta\gamma} \tag{17}$$

are all increasing in absolute value with β.

Speculative implications:

Three settings where the Malthusian constraint might matter

- **Effect of Black Death:** Large effects on European development (Voigtländer and Voth, 2013a,b) due to tight constraint? Similar epidemics in Asia w/o major changes due to loose constraint?
- **Involution:** Higher densities and output, but not living standards, in response to productivity (Geertz, 1963; Huang, 1990) due to loose constraint?
- **Response to agric. technology/inputs:** Necessary increase to match rich countries in TFP/inputs is larger with loose constraint (Eberhardt and Vollrath, 2016a,b)

Empirical implications:

Acemoglu and Johnson (2007) analysis of epidemiological transition:

- WHO (and other) interventions lower mortality rates from many tropical diseases
- They find higher resulting population was negative for income per capita
- According to our theory, the negative effect should be *bigger* for places with *tighter* land constraints
- We divide countries in AJ into “loose” and “tight” groups based on country-specific elasticities, run AJ analysis separately for those groups

Empirical implications:

Estimate β separately for each country in Acemoglu and Johnson sample (assumes β is homogenous within country). Divide into two groups based on those estimated β values.

- "Loose": below-median β
- "Tight": above-median β

Run panel regressions separately for both groups of form

$$y_{it} = \alpha + \theta x_{it} + \gamma_i + \delta_t + \epsilon_{it} \tag{18}$$

for country i and time (decade) t. x_{it} is mortality rate, (log) life expect, or (log) population size. y_{it} is (log) GDP per capita, (log) GDP per worker, or (log) population.

Prediction is that $\theta_{Tight} > \theta_{Loose}$.

Response to mortality rate changes

	Dependent Variable:					
	Log GDP per capita		Log GDP per worker		Log population	
	Loose (1)	Tight (2)	Loose (3)	Tight (4)	Loose (5)	Tight (6)
	Panel A:					
Mortality rate	0.333	0.723	0.284	0.776	-0.361	-0.597
	(0.271)	(0.136)	(0.262)	(0.145)	(0.186)	(0.152)
p-value $\theta = 0$	0.220	0.000	0.281	0.000	0.054	0.000
p-value $\theta = \theta^{Loose}$	.	0.199	.	0.102	.	0.327
Countries	16	16	16	16	16	16
Observations	128	128	128	128	128	128

Response to life expectancy changes

	Dependent Variable:					
	Log GDP per capita		Log GDP per worker		Log population	
	Loose (1)	Tight (2)	Loose (3)	Tight (4)	Loose (5)	Tight (6)
	Panel B:					
Log life expectancy	0.067	-1.864	0.051	-1.876	1.520	2.008
	(0.419)	(0.226)	(0.399)	(0.236)	(0.228)	(0.223)
p-value $\theta = 0$	0.873	0.000	0.899	0.000	0.000	0.000
p-value $\theta = \theta^{Loose}$	.	0.000	.	0.000	.	0.128
Countries	16	16	16	16	16	16
Observations	122	121	122	121	122	121

Response to population change

	Dependent Variable:					
	Log GDP per capita		Log GDP per worker		Log population	
	Loose (1)	Tight (2)	Loose (3)	Tight (4)	Loose (5)	Tight (6)
	Panel C:					
Log population	-0.380 (0.125)	-0.776 (0.067)	-0.383 (0.121)	-0.763 (0.062)		
p-value $\theta = 0$	0.003	0.000	0.002	0.000		
p-value $\theta = \theta^{Loose}$	.	0.006	.	0.006		
Countries	16	16	16	16		
Observations	128	128	128	128		

Conclusion

- Estimate Malthusian constraint - elasticity of agricultural output w.r.t. land - from variation in rural density within provinces
- Constraint is “tight” (0.20-0.30) in temperate areas (N. China, Europe, US/Canada, S. Africa)
- Constraint is “loose” (0.10-0.15) in tropical areas (S. China, SE Asia, C. Africa, S/C America)
- Constraint affects the sensitivity of L_A/L and living standards to population and productivity
- Evidence from epidemiological transition is consistent with our findings and theory
- Implications for the study of historical and contemporary development

Interaction Regression

Combine a given sample with the reference sample (denoted by *Ref*). Run the following regression with interaction terms

$$\begin{aligned}\ln A_{isg} = \beta \ln L_{Aisg}/X_{isg} + (\beta_g^{Ref} - \beta_g) \ln L_{Aisg}/X_{isg} \times I(Ref) \\ + \gamma_s + \delta_g' \mathbf{Z}_{isg} + (\delta_g^{Ref} - \delta_g)' \mathbf{Z}_{isg} \times I(Ref) + \epsilon_{isg}.\end{aligned} \tag{19}$$

where $I(Ref)$ is an indicator for the reference region. Our hypothesis test is $H_0 : \beta_g^{Ref} - \beta_g = 0$, the coefficient on the interaction term for rural density.

Return

Summary statistics

	Mean	SD	Percentiles:				
			10th	25th	50th	75th	90th
Rural density (persons/ha)	0.68	1.32	0.02	0.07	0.21	0.62	1.75
Caloric yield (mil cals/ha)	10.65	4.89	4.64	7.01	10.52	13.74	16.79
Urbanization rate	0.34	0.34	0.00	0.00	0.28	0.66	0.85
Log light density	-2.71	3.06	-6.42	-3.81	-2.33	-0.66	0.57

Return

Crops used in productivity calculation

alfalfa, banana, barley, buckwheat, cassava, chickpea, cowpea, drypea, flax, foxtail millet, greengram, groundnut, indica rice, maize, oat, pearl millet, phaselous bean, pigeon pea, rye, sorghum, soybean, spring wheat, sweetpotato, rape, wet/paddy rice, wheat, winter wheat, white potato, and yams

Return

Crop 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 (0.021)	0.132 (0.018)	0.192 (0.016)	0.113 (0.018)	0.205 (0.015)	0.133 (0.012)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta_{Temp}$		0.000		0.001		0.000
Countries	91	81	83	71	74	84
Observations	10661	9088	10768	8113	10708	7564
Adjusted R-square	0.24	0.20	0.21	0.18	0.20	0.18

Return

Crop Results

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

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 (0.022)	0.143 (0.021)	0.242 (0.033)	0.133 (0.018)	0.281 (0.035)	0.185 (0.019)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta_{Temp}$		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

Climate Results

Dependent Variable in all panels: Log caloric yield (A_{isc})					
Panel A: Climate Zones					
	Equatorial (1)	Arid (2)	Temperate (3)	Snow (4)	
Log rural density	0.142 (0.014)	0.219 (0.030)	0.247 (0.048)	0.149 (0.030)	0.124 (0.033)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Hot}$		0.004	0.035	0.850	0.576
Countries	61	84	26	43	27
Observations	9312	9858	540	1582	1160
Adjusted R-square	0.16	0.22	0.16	0.14	0.15

Return

Climate Results

Panel B: Precipitation Zones	Fully Humid (1)	Dry Summer (2)	Dry Winter (3)	Monsoon (4)	Desert (5)	Steppe (6)
Log rural density	0.185	0.169	0.115	0.130	0.112	0.110
	(0.024)	(0.018)	(0.016)	(0.023)	(0.040)	(0.026)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.005	0.000
p-value $\beta = \beta^{Fully}$		0.594	0.013	0.090	0.120	0.034
Countries	98	45	74	42	30	54
Observations	17327	3150	9224	1816	364	2339
Adjusted R-square	0.21	0.20	0.18	0.20	0.20	0.19

Climate Results

Panel C: Temperature Zones					
	Hot Summer (1)	Warm Summer (2)	Cool Summer (3)	Hot Arid (4)	Cold Arid (5)
Log rural density	0.142	0.219	0.247	0.149	0.124
	(0.014)	(0.030)	(0.048)	(0.030)	(0.033)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Hot}$		0.004	0.035	0.850	0.576
Countries	61	84	26	43	27
Observations	9312	9858	540	1582	1160
Adjusted R-square	0.16	0.22	0.16	0.14	0.15

Region Results

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

Panel A

				Excl. China, Japan, Korea	
	North & Western Europe (1)	Eastern Europe (2)	Southern Europe (3)	South & Southeast Asia (4)	Central & West Asia (5)
Log rural density	0.259	0.287	0.272	0.152	0.181
	(0.036)	(0.031)	(0.041)	(0.026)	(0.024)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{NWEur}$		0.539	0.791	0.017	0.072
Countries	16	9	9	13	18
Observations	1684	4821	1137	4312	2878
Adjusted R-square	0.22	0.32	0.28	0.18	0.20

Return

Region Results

Dependent Variable in all panels: Log caloric yield (A_{isc})					
Panel B					
	Temperate Americas	Tropical Americas	Tropical Africa	South Africa	North Africa
Log rural density	0.188	0.113	0.089	0.134	0.249
	(0.030)	(0.016)	(0.014)	(0.071)	(0.014)
p-value $\beta = 0$	0.000	0.000	0.000	0.059	0.000
p-value $\beta = \beta^{NWEur}$	0.133	0.000	0.000	0.116	0.827
Countries	5	22	39	4	5
Observations	3796	9373	3181	198	1220
Adjusted R-square	0.20	0.10	0.14	0.20	0.24

Region Results

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

Panel C

	All China	Temperate China	Sub-Tropical China	Japan	North & South Korea
	(1)	(2)	(3)	(4)	(5)
Log rural density	0.388	0.508	0.095	0.177	0.162
	(0.084)	(0.054)	(0.026)	(0.013)	(0.065)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.012
p-value $\beta = \beta^{NWEur}$	0.156	0.000	0.000	0.034	0.192
Observations	289	134	155	1198	326
Adjusted R-square	0.26	0.28	0.22	0.26	0.22

Results by Province

Baseline results assume β constant within larger sub-samples.

Instead, estimate β individually for each province

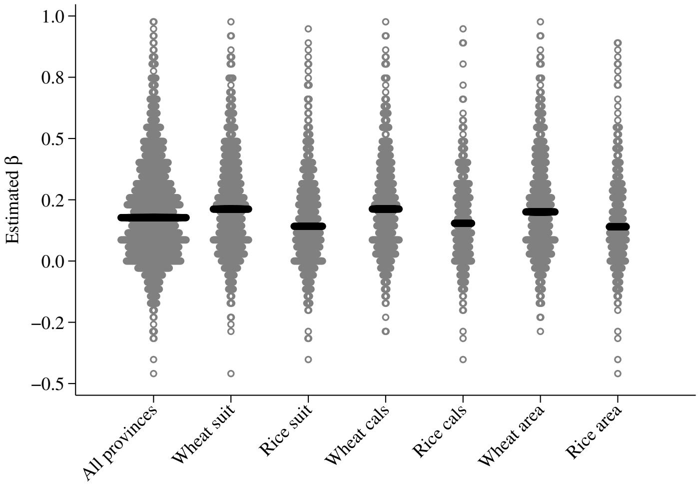
- Only provinces with 6 or more districts (1,260 provinces)
- ... so really big SE on any individual estimate
- Then look at pattern of β's for each sub-sample

Results by Province

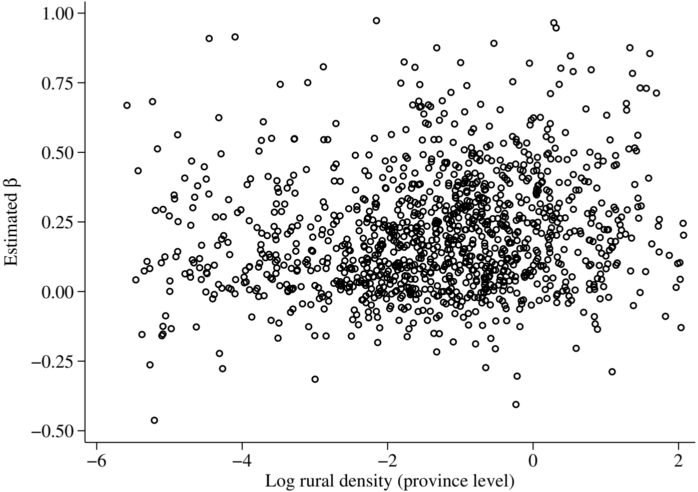
				Percentiles:				
Sample	Prov.	Mean	SD	10th	25th	50th	75th	90th
All provinces	1,183	0.21	0.23	-0.02	0.05	0.18	0.33	0.49
Wheat Suitable	619	0.23	0.23	-0.00	0.08	0.21	0.36	0.50
Rice Suitable	469	0.17	0.24	-0.04	0.03	0.15	0.28	0.43
Wheat cals>33%	457	0.23	0.22	-0.00	0.08	0.21	0.36	0.50
Rice cals>33%	307	0.18	0.21	-0.04	0.03	0.15	0.29	0.41
Wheat area>50%	485	0.22	0.23	-0.01	0.07	0.20	0.36	0.53
Rice area>50%	297	0.18	0.23	-0.04	0.03	0.14	0.29	0.47
Northwest Europe	79	0.26	0.31	0.00	0.08	0.22	0.46	0.62
Eastern Europe	173	0.24	0.20	0.01	0.09	0.23	0.38	0.50
Southern Europe	60	0.27	0.17	0.08	0.16	0.25	0.37	0.50
South and S. East Asia	248	0.20	0.23	-0.03	0.04	0.16	0.31	0.49
Central and West. Asia	163	0.20	0.20	-0.02	0.06	0.16	0.33	0.45
Temperate Americas	87	0.14	0.25	-0.13	0.02	0.10	0.26	0.40
Tropical Americas	195	0.18	0.26	-0.02	0.06	0.17	0.28	0.39
Tropical Africa	118	0.16	0.24	-0.09	-0.01	0.09	0.25	0.51
Southern Africa	11	0.15	0.21	-0.11	-0.05	0.17	0.33	0.34
Northern Africa	49	0.32	0.20	0.06	0.22	0.31	0.42	0.66

Return

Results by Province



Relationship of β to rural density, by province



Return

Labor and capital not mobile

Factors cannot move within province, but output can.

Changes relationship of density and productivity to

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

- Night lights provide proxy for A_{Ni} and K_i/L_i?
- p_N/p_A is province-specific (FE)
- Is correlation of A_{Ni} and K_i/L_i with L_{Ai}/X_i different by climate zone?

Return

Districts are autarkic

Factors and output are immobile within province.

Changes relationship of density and productivity to

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

- Can control for L_{Ai}/L_i using HYDE data
- Night lights provide proxy for K_i/L_i?
- c_{Ai} doesn't vary much and/or proxied by night lights?

Return

Crop 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	0.254 (0.024)	0.148 (0.019)	0.213 (0.021)	0.120 (0.020)	0.231 (0.020)	0.136 (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

Using Cultivated Area

Cultivated area, X^C_{isc}, available from GAEZ. Rural density is

$$\ln L_{Aisc}/X_{isc} = \ln L_{Aisc}/X^C_{isc} + \ln X^C_{isc}/X_{isc} \tag{20}$$

- Regress $\ln A_{isc}$ on both terms on the right hand-side
- Coefficient on $\ln L_{Aisc}/X^C_{isc}$ gives similar results for β
- Controls for percent of land actually being cultivated

Return

Using Cultivated Area

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

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

Return

Population Data from 1900

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	0.294	0.170	0.239	0.149	0.258	0.182
	(0.032)	(0.025)	(0.023)	(0.025)	(0.026)	(0.017)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Wheat}$		0.002		0.007		0.014
Countries	91	81	83	71	74	84
Observations	10644	9081	10774	8213	10689	7561
Adjusted R-square	0.30	0.25	0.25	0.22	0.24	0.22

Return

Above 25th Percentile Harvested Area

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

Return

Maize and Soy Results

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

Samples defined by suitability for each crop:

	Maize suitable AND:			Soy suitable AND:		
	Wheat or Rice (1)	Rice Only (2)	Wheat Only (3)	Wheat or Rice (4)	Rice Only (5)	Wheat Only (6)
Log rural density	0.135	0.142	0.209	0.136	0.144	0.216
	(0.015)	(0.018)	(0.035)	(0.015)	(0.018)	(0.034)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{All}$		0.760	0.011		0.721	0.004
Countries	116	77	78	117	78	61
Observations	14499	8365	6781	14486	8220	6311
Adjusted R-square	0.12	0.12	0.15	0.12	0.12	0.15

Return

Measurement Error

Measurement error $\Rightarrow$ attentuation bias

- Population data from HYDE may not be accurate for districts
- Is measurement error more pronounced in some places (e.g. tropical areas) and driving results?
- Is true variance of $\ln L_{Aisc}/X_{isc}$ one-third of measured variance?
- Is rural population mis-stated by factor of > 2 or < 0.5?

Return

Elasticity of Substitution?

What if land and labor do not have elasticity of subs. equal to one?

- Elasticity of output w.r.t. land depends on rural density L_A/X
- With EOS more than one, higher density, lower elasticity
- Do results fit this?
  - South/SE Asia, some SS Afr are high density, low β
  - ..but C/S America, other SS Afr are low density, low β
  - ..but N America lowest density, not highest β

Density? Return

Factor shares

Our β does not correlate well with factor share data

- Fuglie (2010) reports share for "land and structures"
  - 0.22-0.25 share for Brazil, India, Indonesia $>$ our estimates
  - 0.22 share for China, aggregate number?
  - 0.17-0.26 for US, ex-Soviet $\approx$ our estimates
- Hayami, Ruttan, Southworth (1979) for Asia
  - 0.3-0.4 shares for Taiwan, Japan, Korea, Philippines $>$ our estimates
- Clark (2002) for England, long run
  - 0.3-0.35 shares $>$ our estimate for Northwest Europe

Factor Shares

Possible sources of difference

- Biased estimates of β
- Market frictions/wedges for agricultural inputs
- Mis-reporting of land versus labor income
- Difference of aggregate elasticity from farm-specific elasticity/share

Return

Using IPUMS Population Data

39 countries in IPUMS with geographic identifiers for individuals at the "district" level (agglomerations of districts with constant boundaries).

- IPUMS gives industry/occupation and labor force status, so we can measure *agricultural worker density*, not just rural density. In practice, rural density and ag worker density correlated at 91%, sig at less than 1%
- Counts of residents in each district are presumably more accurate that HYDE or GRUMP, as they are drawn from census.
- Cost is limited coverage of countries, and fewer districts
- Rebuild data on A_{isc} and L_{Aisc} at the level of the IPUMS districts

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	Rice Only	Wheat/ No Rice	Rice No Wheat	Wheat $> 50\%$	Rice $> 50\%$
	(1)	(2)	(3)	(4)	(5)	(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

Return

Using GRUMP Population Data

GRUMP (Global Rural-Urban Mapping Project) also provides grid-cell population estimates.

- Define urban/rural differently than HYDE. They denote specific grid-cells as "urban", and any population in that cell is assumed to be urban.
- Inclusion of cells in districts and provinces is identical to HYDE.

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	Rice Only	Wheat/ No Rice	Rice No Wheat	Wheat $> 50\%$	Rice $> 50\%$
	(1)	(2)	(3)	(4)	(5)	(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

Return