How Tight are Malthusian Constraints?

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

[1] University of Houston

[2] University of Houston

How Tight are Malthusian Constraints?

- **Malthusian constraint**: using a fixed factor (e.g. land) in agricultural production
- **How Tight?**: what is the elasticity of agricultural output w.r.t land?

In this paper

Estimate the elasticity of agricultural output w.r.t land

- Use relationship of rural density and agro-climatic agricultural TFP to estimate the elasticity of output w.r.t land
- Estimates come from *within*-province variation across districts
- Population data from HYDE
- Agro-climatic TFP built from Galor and Özak data on caloric suitability

Advantages of our method

- Not assuming elasticity is same *across* or *within* countries
- Do not need data on inputs other than land and labor

In this paper

We find:

- Elasticities range from 0.1 to 0.4
- Variation is related to agro-climatic conditions, crop suitability
- Temperate, cold, “wheat” suitable ($\sim 0.20 - 0.30$) $\Rightarrow$ tight Malthusian constraints
- Equatorial, hot, “rice” suitable ($\sim 0.10 - 0.15$) $\Rightarrow$ loose Malthusian constraints

Implications

Elasticity determines degree of decreasing returns to mobile factors in agriculture

Higher elasticity $\Rightarrow$

- more sensitive L_A/L is to TFP or population shocks
- more sensitive y is to TFP or population shocks

Variation in elasticity informative (?)

- Modern: slow development of tropical areas?
- Historical: effect of Black Death, “involution” in Asia?

Some Related Literature:

Similar empirical concepts:

- Motamed, Florax, Masters (2014): Pattern/date of urbanization at grid-cell level based on agro-climatic conditions
- Henderson, Squires, Storeygard, Weil (2016): Spatial organization of economic activity, relationship to geographic conditions

Results informative for:

- Malthusian and UGT models
- Structural change
- Agricultural productivity and/or frictions

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

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

- 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

Take logs of L_{Ai}/X_i expression

$$\ln L_{Ai}/X_i = \frac{1}{\beta} \ln A_i + \ln \Gamma, \tag{6}$$

where

$$\Gamma = \frac{L_A}{\sum_{j \in I} A_j^{1/\beta} X_j}. \tag{7}$$

is a *province-specific* term.

- β can be estimated from elasticity of L_{Ai}/X_i w.r.t. A_i.
- $1/\beta$ small (tight), ag. workers spread evenly w/in province
- $1/\beta$ large (loose), ag. workers concentrated on high A_i

Agricultural Labor Allocation

Take logs L_{Ai}/X_i expression

$$\ln L_{Ai}/X_i = \frac{1}{\beta} \ln A_i + \ln \Gamma, \tag{8}$$

where

$$\Gamma = \frac{L_A}{\sum_{j \in I} A_j^{1/\beta} X_j}. \tag{9}$$

is a *province-specific* term.

- Γ is constant for all districts w/in province
- 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:

$$\ln A_{isc} = \alpha + \beta \ln L_{Aisc}/X_{isc} + \gamma_{sc} + \delta' \mathbf{Z}_{isc} + \epsilon_{isc}. \quad (10)$$

- District i, region/state/province s, country c
- γ_{sc}, province/country FE, pick up Γ term
- $\mathbf{Z}_{isc}$ are additional controls
- ϵ_{isc} is error term

Using as a Specification

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

$$\ln A_{isc} = \alpha + \beta \ln L_{Aisc}/X_{isc} + \gamma_{sc} + \delta' \mathbf{Z}_{isc} + \epsilon_{isc}. \tag{11}$$

- Not a causal statement, structural estimate
- Estimating $1/\beta \Rightarrow$ very sensitive to small differences in β

Agricultural Density Data

L_{Aisc} 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

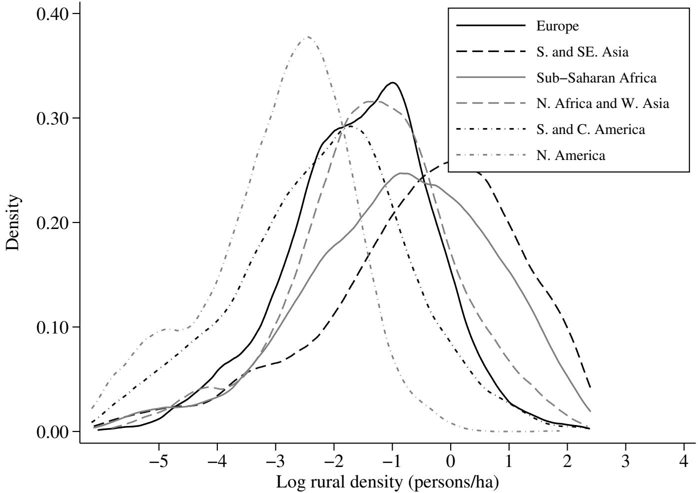
X_{isc} calculated as area of a given district

- Measure of possible agricultural land

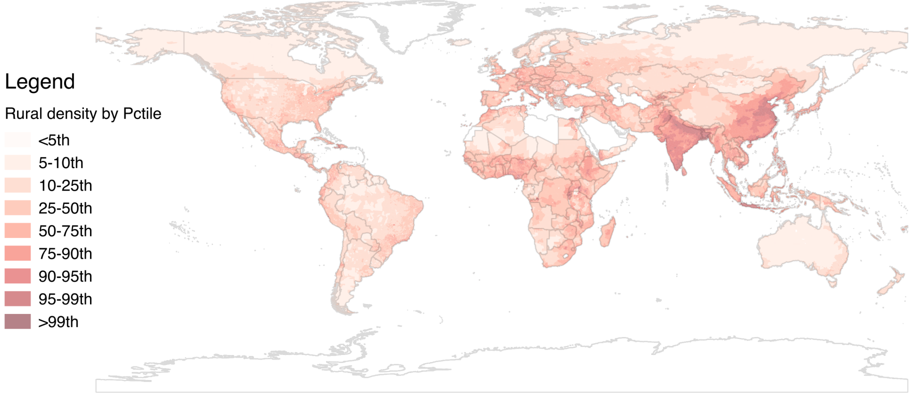
L_{Aisc}/X_{isc} data

- Trim above 99th and below 1st percentiles
- Drop if fewer than 100 total rural residents
- 32,862 total districts

Agricultural Density Data



Agricultural Density Data



Agricultural Productivity Data

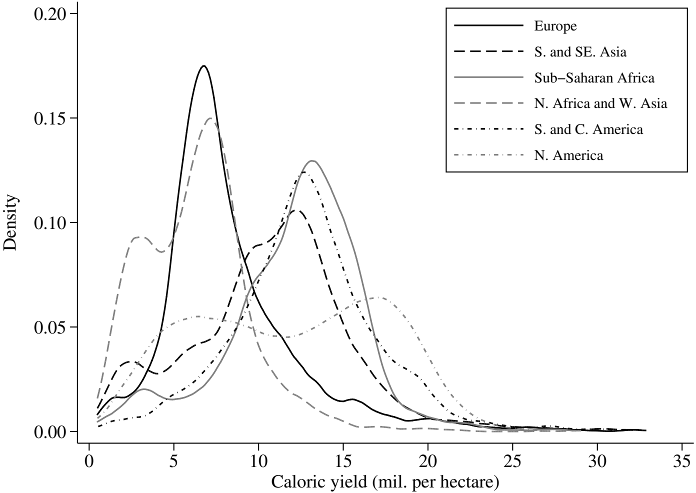
A_{isc} is built from Galor and Özak (2016) caloric suitability index

- 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
- Total max calories across grid cells in district, divide by total area
- As in Galor and Özak, holds technology assumptions constant
- Trim above 99th and below 1st percentile

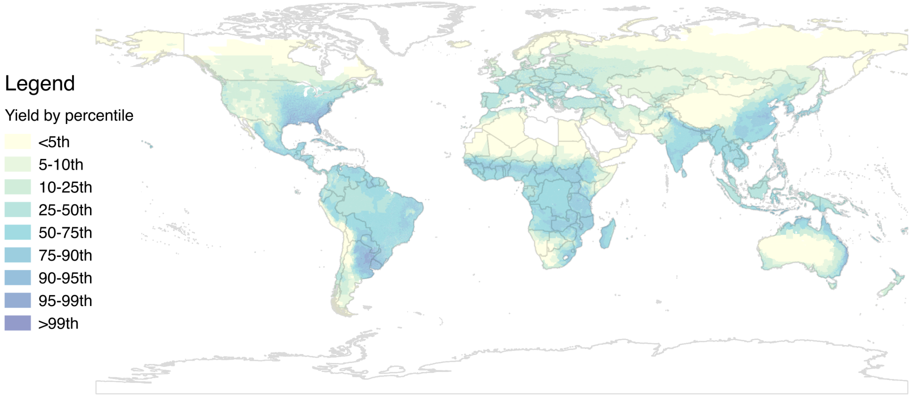
Summary

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
- $Corr(\epsilon_{isc}, \ln L_{isc}/X_{isc}) < 0$

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

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

Spatial Errors and Hypothesis Testing

Assume ϵ_{isc} has spatial auto-correlation. Use Conley s.e. (500km window).

Two hypothesis tests:

- Is the land constraint binding?
  - $H_0 : \beta = 0$ vs. two-sided alt
- Is the land constraint the same in two samples (e.g. Europe and Sub-Saharan Africa)?
  - $H_0 : \beta = \beta^{Ref}$ vs. two-sided alt
  - β^{Ref} from ad hoc “reference” sample
  - Implemented with interaction regression combining given and reference sample

Interaction

Results by Crop Family

Define the following:

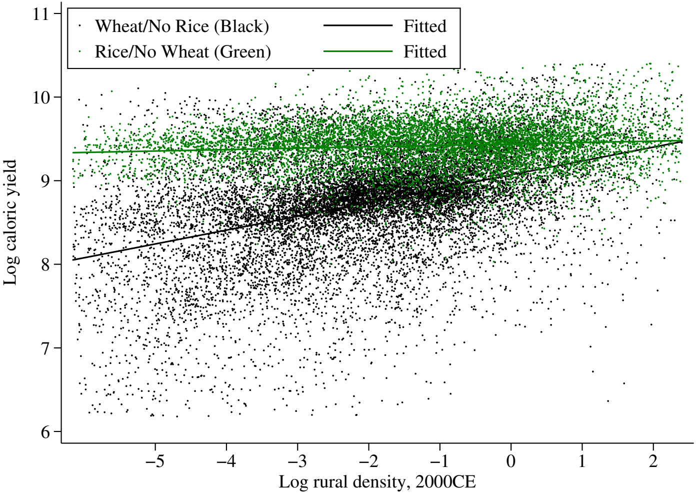
- **Wheat family**: barley, buckwheat, rye, oats, white potatoes, wheat
- **Rice family**: cassava, cowpeas, pearl millet, sweet potatoes, paddy rice, yams

Create sub-samples of districts based on

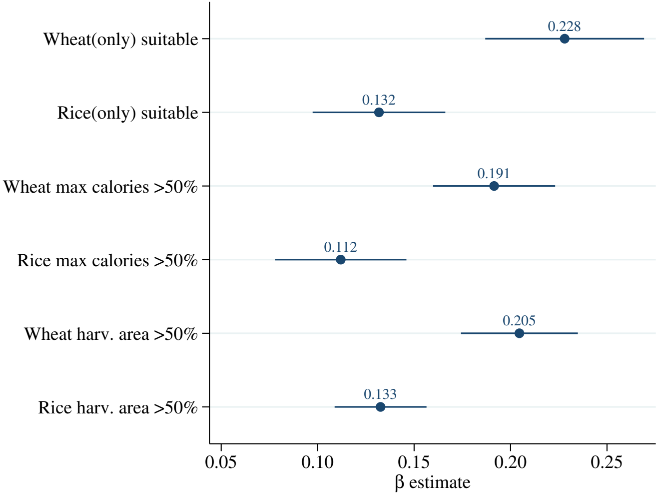
- GAEZ suitability indices for a crop family (0 to 100)
- Source of maximum calories in our A_{isc} measure
- Actual harvested area for a crop family

Not estimating a crop-specific production function

Raw Correlation for Rice/Wheat Families

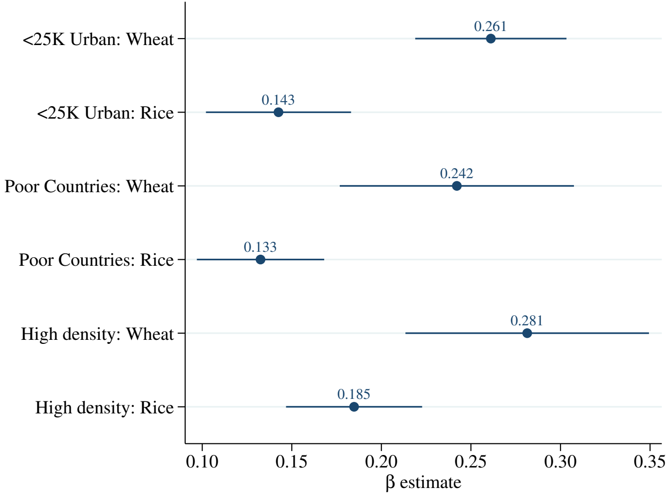


Results by Crop



Table

Results by Crop

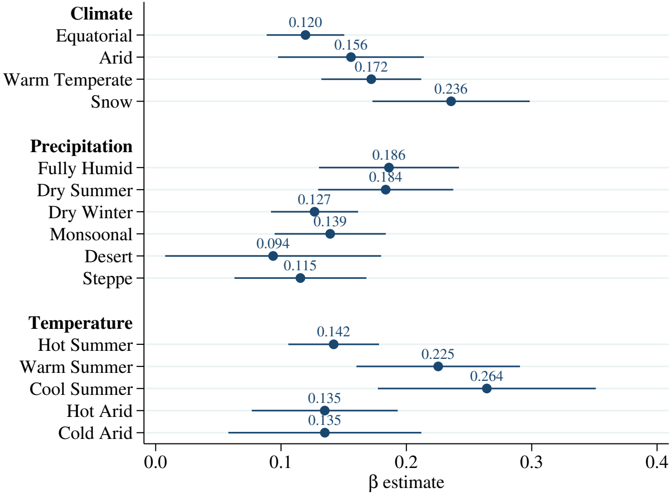


Results by Climate Zone

Crop suitability based in part on climate conditions.

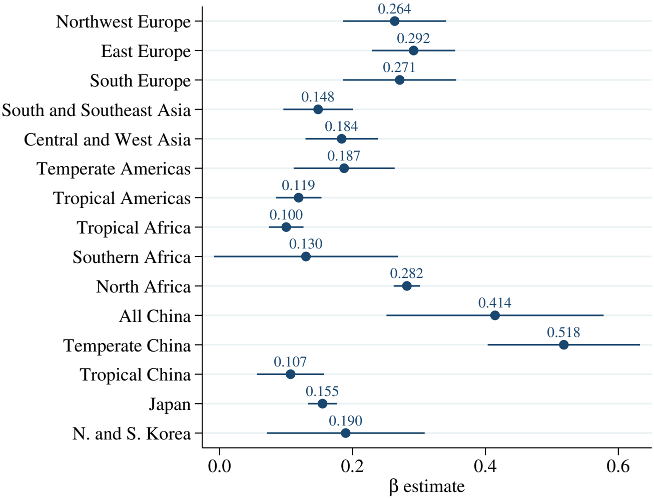
- Create samples based on Köppen-Geiger zones
- Three layers: Climate, Precipitation, Temperature
- Each layer has multiple types (e.g. Climate is Equatorial, Arid, ...)
- Create samples where districts have >50% of land in a given type
- $\Rightarrow$ heterogeneity within countries in β

Results by Climate Zone



Table

Results by Sub-Region



Table

Explanations?

Evidence suggests:

- Tight constraints: wheat family, temperate/snow, warm/cool summers
- Loose constraints: rice family, equatorial/arid, hot summer/arid

Why looser constraints some 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?

Robustness and extensions

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

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?

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?

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

Back to the model

Aggregate production of agricultural good

$$Y_A = A_A \left(\frac{K_A}{L_A} \right)^{\alpha(1-\beta)} L_A^{1-\beta}, \tag{12}$$

where

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

and non-agricultural good

$$Y_N = A_N \left(\frac{K_N}{L_N} \right)^{\alpha} L_N. \tag{13}$$

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

$$p_A \phi_L \frac{Y_A}{L_A} = p_N \phi_L \frac{Y_N}{L_N}. \tag{14}$$

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

$$\ln c_A = \ln \theta_A + (1-\epsilon)\ln M + (\gamma - 1)\ln p_A + (\epsilon - \gamma)\ln p_N \tag{15}$$

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

$$\frac{L_A}{L} = \theta_A \left(\frac{L^{\beta\gamma}}{A_A^{\gamma} A_N^{\epsilon-\gamma} \hat{k}^{\alpha(\epsilon-\beta\gamma)}} \right)^{\frac{1}{1-\beta\gamma}} \tag{16}$$

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

$$y = \left(\frac{A_A A_N^{\beta(\epsilon-\gamma)} \hat{k}^{\Omega}}{L^{\beta}} \right)^{\frac{1}{1-\beta\gamma}} \tag{17}$$

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):

$$\frac{\partial \ln L_A/L}{\partial \ln A_A} = -\frac{\gamma}{1-\beta\gamma} \qquad \frac{\partial \ln y}{\partial \ln A_A} = \frac{1}{1-\beta\gamma} \tag{18}$$

(b) Non-agricultural productivity (A_N):

$$\frac{\partial \ln L_A/L}{\partial \ln A_N} = -\frac{\epsilon-\gamma}{1-\beta\gamma} \qquad \frac{\partial \ln y}{\partial \ln A_N} = \frac{\beta(\epsilon-\gamma)}{1-\beta\gamma} \tag{19}$$

(c) Population (L):

$$\frac{\partial \ln L_A/L}{\partial \ln L} = \frac{\beta\gamma}{1-\beta\gamma} \qquad \frac{\partial \ln y}{\partial \ln L} = -\frac{\beta}{1-\beta\gamma} \tag{20}$$

are all increasing in absolute value with β.

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)

Conclusion

- Estimate Malthusian constraint from variation in rural density within provinces
- Constraint is “tight” (0.20-0.30) in temperate areas (N. China, Europe, US/Canada, S. Africa) suitable for wheat family crops
- Constraint is “loose” (0.10-0.15) in tropical areas (S. China, SE Asia, C. Africa, S/C America) suitable for rice family crops
- Constraint dictates the sensitivity of L_A/L and living standards to population and productivity

Interaction Regression

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

$$\begin{aligned} \ln A_{isc} = {} & \beta \ln L_{Aisc}/X_{isc} + (\beta^{Ref} - \beta) \ln L_{Aisc}/X_{isc} \times I(Ref) \\ & + \gamma_{sc} + \delta' \mathbf{Z}_{isc} + (\delta^{Ref} - \delta)' \mathbf{Z}_{isc} \times I(Ref) + \epsilon_{isc}. \end{aligned} \tag{21}$$

where $I(Ref)$ is an indicator for the reference region. Our hypothesis test is $H_0 : \beta^{Ref} - \beta = 0$, 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)	0.191 (0.016)	0.112 (0.017)	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 $\beta = \beta^{Wheat}$		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

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 $\beta = \beta^{Wheat}$		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

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	0.225	0.264	0.135	0.135
	(0.018)	(0.033)	(0.044)	(0.030)	(0.039)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.001
p-value $\beta = \beta^{Hot}$		0.006	0.010	0.831	0.848
Countries	61	84	25	42	25
Observations	8495	9452	438	1505	957
Adjusted R-square	0.15	0.21	0.15	0.12	0.14

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	0.186	0.184	0.127	0.139	0.094	0.115
	(0.028)	(0.027)	(0.018)	(0.023)	(0.044)	(0.027)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.033	0.000
p-value $\beta = \beta^{Fully}$		0.947	0.073	0.190	0.078	0.072
Countries	97	44	74	42	29	53
Observations	16216	2978	8503	1655	330	2093
Adjusted R-square	0.19	0.19	0.17	0.19	0.19	0.18

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	0.225	0.264	0.135	0.135
	(0.018)	(0.033)	(0.044)	(0.030)	(0.039)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.001
p-value $\beta = \beta^{Hot}$		0.006	0.010	0.831	0.848
Countries	61	84	25	42	25
Observations	8495	9452	438	1505	957
Adjusted R-square	0.15	0.21	0.15	0.12	0.14

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	0.264	0.292	0.271	0.148	0.184
	(0.040)	(0.032)	(0.043)	(0.027)	(0.028)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{NWEur}$		0.569	0.884	0.016	0.099
Countries	16	9	9	13	18
Observations	1628	4772	1114	3921	2762
Adjusted R-square	0.21	0.31	0.26	0.16	0.18

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	0.187	0.119	0.100	0.130	0.282
	(0.039)	(0.018)	(0.013)	(0.071)	(0.010)
p-value $\beta = 0$	0.000	0.000	0.000	0.066	0.000
p-value $\beta = \beta^{NWEur}$	0.170	0.001	0.000	0.099	0.654
Countries	5	22	39	4	5
Observations	3183	8730	3032	178	1147
Adjusted R-square	0.18	0.10	0.14	0.19	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	0.414	0.518	0.107	0.155	0.190
	(0.083)	(0.058)	(0.026)	(0.011)	(0.061)
p-value $\beta = 0$	0.000	0.000	0.000	0.000	0.002
p-value $\beta = \beta^{NWEur}$	0.102	0.000	0.001	0.008	0.309
Countries	1	1	1	1	2
Observations	266	130	136	1039	311
Adjusted R-square	0.25	0.26	0.21	0.21	0.21

Results by Province

Baseline results assume β constant within larger sub-samples.

Instead, estimate β individually for each province

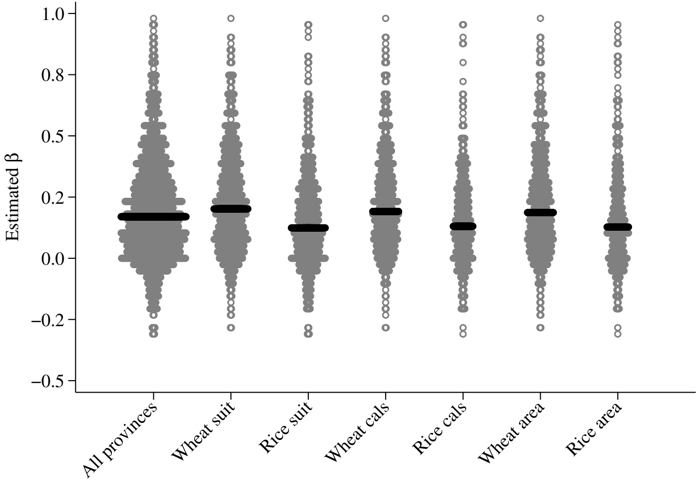
- Only provinces with 6 or more districts (1,260 provinces)
- ... so really big SE on any individual estimate
- Then look at pattern of β's for each sub-sample

Results by Province

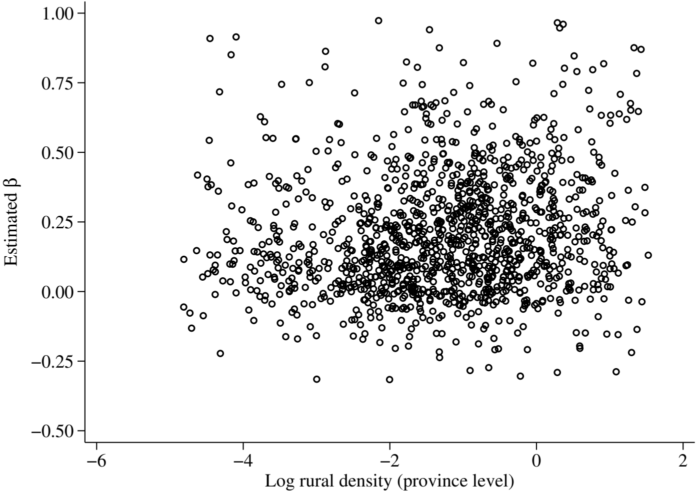
				Percentiles:				
Sample	Prov.	Mean	SD	10th	25th	50th	75th	90th
All provinces	1,260	0.21	0.31	-0.04	0.04	0.17	0.33	0.52
Wheat Suitable	640	0.24	0.25	-0.01	0.08	0.21	0.38	0.55
Rice Suitable	514	0.16	0.28	-0.06	0.01	0.12	0.28	0.45
Wheat cals>33%	484	0.23	0.26	-0.01	0.08	0.19	0.37	0.55
Rice cals>33%	328	0.16	0.23	-0.06	0.02	0.13	0.28	0.41
Wheat area>50%	511	0.24	0.27	-0.01	0.07	0.19	0.37	0.55
Rice area>50%	315	0.17	0.26	-0.06	0.00	0.12	0.29	0.50

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

$$\ln L_{Aisc}/X_{isc} = \ln L_{Aisc}/X^C_{isc} + \ln X^C_{isc}/X_{isc} \tag{22}$$

- 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 (0.024)	0.144 (0.020)	0.191 (0.020)	0.113 (0.021)	0.207 (0.020)	0.142 (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 (0.058)	0.070 (0.020)	0.248 (0.030)	0.016 (0.013)	0.368 (0.043)	0.052 (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	0.240	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

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

	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)
Dependent Variable in all panels: Log caloric yield (A_{isc})						
Samples defined by suitability for each crop:						
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