July 22, 2017

How Tight are Malthusian Constraints?

T. Ryan Johnson
University of Houston

Dietrich Vollrath
University of Houston

Abstract

We provide a methodology to estimate the elasticity of agricultural output with respect to land - the Malthusian constraint - using variation in rural densities across different locations. We use district-level data from around the globe on rural densities and inherent agricultural productivity to estimate the elasticity for various sub-samples. We find the elasticity is highest in areas suitable for temperate crops (e.g. wheat and rye), and lowest in areas suitable for sub-tropical crops (e.g. cassava and rice). We show theoretically that a higher elasticity results in greater sensitivity of non-agricultural employment and income per capita to shocks in population size and productivity, and confirm this with evidence from the post-war mortality transition.

JEL Codes: O1, O13, O44, Q10

Keywords: land constraints, Malthusian stagnation, agriculture

Contact information: 201C McElhinney Hall, U. of Houston, Houston, TX 77204, devollrath@uh.edu. We thank Francesco Caselli, Martin Fiszbein, Oded Galor, Nippe Lagerlöf, Debin Ma, Stelios Michalopolous, Nathan Nunn, Ömer Özak, Enrico Spolaore, Joachim Voth, and David Weil, as well as seminar participants at the London School of Economics and the Brown Conference on Deep-rooted Determinants of Development for their comments. All errors remain our own.

Table 1: Summary Statistics for District Level Data, 2000CE

			Percentiles:				
	Mean	SD	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

Notes: A total of 32,862 observations for each variable (these come from 2,471 provinces in 154 countries). Caloric yield, A_{isc} calculated by the authors using data from **?**. Rural density, L_{Aisc}/X_{isc} calculated by the authors using data from **?** for rural population. Both caloric yield and rural density were trimmed at the 99th and 1st percentiles of their raw data prior to calculating the summary statistics in this table. Urbanization rate taken from **?**. Log mean light density derived from the Global Radiance Calibrated Nightime Lights data provided by NOAA/NGDC, as in **?**.

Table 2: Estimates of Malthusian Tightness, β, by Crop Suitability, 2000CE

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

Notes: 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. The coefficient estimate on rural population density indicates the value of β, see equation (**??**). Rural population is from HYDE database (**?**), and caloric yield is the author's calculations based on the data from **?**. Inclusion of districts in the regression is based on the listed criteria related to crop families. See text for all crops included in the wheat and rice families, and for details of the inclusion criteria.

Table 3: Estimates of Malthusian Tightness, β, by Köppen-Geiger Zone, 2000CE

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	0.120	0.156	0.172	0.236
	(0.016)	(0.030)	(0.020)	(0.032)
p-value $\beta = 0$	0.000	0.000	0.000	0.000
p-value $\beta = \beta^{Equa}$		0.276	0.033	0.001
Countries	79	55	93	40
Observations	10600	2533	12748	5936
Adjusted R-square	0.11	0.10	0.15	0.19

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

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

Notes: 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. The coefficient estimate on rural population density indicates the value of β, see equation (**??**). Rural population is from HYDE database (**?**), and caloric yield is the author's calculations based on the data from **?**. Inclusion of districts is based on whether they have more than 50% of their land area in the given Köppen-Geiger zone. See text for details.

Table 4: Estimates of Malthusian Tightness, β, by Regions, 2000CE

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

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

Panel C

	All China (1)	Temperate China (2)	Sub-Tropical China (3)	Japan (4)	North & South Korea (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

Notes: 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. See appendix for lists of exact countries included in each region. The coefficient estimate on rural population density indicates the value of β, see equation (**??**). Rural population is from HYDE database (**?**), and caloric yield is the author's calculations based on the data from **?**. The countries included in each region can be found in the appendix.

Table 5: Panel Estimates of Effect of Population Change, by Tightness of Malthusian Constraint

	Dependent Variable:					
	Log GDP per capita		Log GDP per worker		Log population	
	β <Median (1)	β >Median (2)	β <Median (3)	β >Median (4)	β <Median (5)	β >Median (6)
	Panel A:					
Mortality rate	0.620	0.573	0.560	0.638	-0.446	-0.505
	(0.277)	(0.147)	(0.272)	(0.157)	(0.188)	(0.155)
p-value $\theta = 0$	0.027	0.000	0.042	0.000	0.019	0.001
p-value $\theta = \theta^{Loose}$	.	0.882	.	0.804	.	0.809
Countries	16	17	16	17	16	17
Observations	126	136	126	136	126	136
	Panel B:					
Log life expectancy	-0.204	-1.665	-0.196	-1.721	1.578	1.869
	(0.425)	(0.259)	(0.405)	(0.267)	(0.214)	(0.249)
p-value $\theta = 0$	0.632	0.000	0.629	0.000	0.000	0.000
p-value $\theta = \theta^{Loose}$	.	0.004	.	0.002	.	0.377
Countries	16	17	16	17	16	17
Observations	120	129	120	129	120	129
	Panel C:					
Log population	-0.374	-0.740	-0.383	-0.733		
	(0.135)	(0.069)	(0.130)	(0.065)		
p-value $\theta = 0$	0.006	0.000	0.004	0.000		
p-value $\theta = \theta^{Loose}$	.	0.016	.	0.017		
Countries	16	17	16	17		
Observations	126	136	126	136		

Notes: Robust standard errors are reported in parentheses. All regressions include both year fixed effects and country fixed effects. The value of β for each country was found by estimating equation (**??**) separately for each, including province-level fixed effects. Countries are then included in a regression here based on how their β compares to the median from the 34 countries. The mortality rate used as an explanatory variable in Panel A is the mortality rate from 15 infectious diseases, as documented by **?**. All data on GDP per capita, GDP per worker, population, and life expectancy is also taken directly from those authors dataset. The p-value of $\theta = \theta^{Loose}$ is from a test that the estimated coefficient in a column (with β over the median) is equal to the coefficient in the column immediately preceding it (with β under the median).