

Geography and Economic Transition

Global Spatial Analysis at the Grid Cell Level

Mesbah Motamed Raymond Florax Will Masters

Department of Agricultural Economics
Purdue University

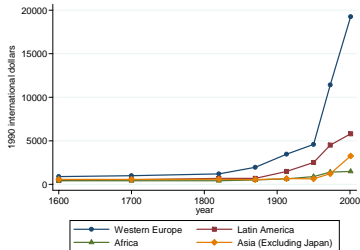
March 2009

Urbanization at the grid cell level

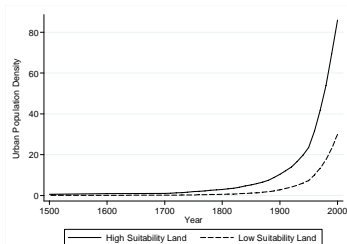
- ▶ What drives the timing of transition from rural to urban activity?
- ▶ We use new global grid cell data on the size and location of cities, matched with data on climate, soils and access to navigable waterways, to test for influence of local farm productivity and transport costs on historical urbanization.
- ▶ Using grid cells allows us to control for country fixed effects and neighbors' urbanization.

Growth regimes and the rural-urban transition

- ▶ A region's income today depends in part on how long ago its modern economic growth began.
- ▶ Economic growth is linked to the formation of cities, which in turn depend on the surrounding countryside.



Regional per capita GDP, 1600-2000 AD, Maddison (2001)



Urban density growth conditional on cultivation suitability level

A two-sector economy

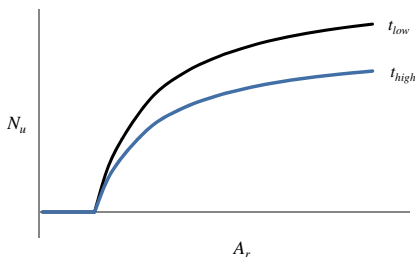
- ▶ Building on Matsuyama (1992),
 - workers are either rural N_r or urban N_u
 - $N_r + N_u = 1$
- ▶ Each sector's output depends on its labor productivity:
 - $Y_r = A_r N_r$ and $Y_u = A_u N_u$
- ▶ Profits depend on relative prices, and for urban firms also on transport to/from rural and other urban areas:
 - $\pi_r = p A_r N_r - w N_r$
 - $\pi_u = (1 - t) A_u N_u - w N_u$
- ▶ A minimum quantity of the rural product is required for survival:
 - $U(C_r, C_u) = \ln(C_r - \underline{r}) + \ln(C_u)$

Equilibrium conditions

- ▶ Labor mobility equalizes wages, subject to subsistence consumption of the rural product
- ▶ In equilibrium, urbanization depends on transport costs and rural productivity:
 - $N_u = \frac{1-t}{2-t} \left(1 - \frac{r}{A_r} \right)$
- ▶ Physical geography could matter for A_r and t
 - A_r : cultivation suitability, frost in winter
 - t : distance to coast, navigable river, elevation

Comparative statics

- ▶ Once subsistence requirements are met, improvements in either rural productivity or transport opportunities drive workers to urbanize:
 - $\partial N_u / \partial A_r > 0$ and $\partial N_u / \partial t > 0$, if $A_r > \underline{r}$



Grid cells as the unit of observation

- ▶ A gridded map of the world
 - cells are 0.5° latitude by 0.5° longitude
 - not subject to endogeneity of administrative boundaries.
- ▶ High resolution data
 - 62,290 observations
 - range from 3000 sq. km. (at equator) to 1000 sq. km. (at arctic circle)
- ▶ Using grid cells allows controls for
 - country fixed effects
 - spatial spillovers from neighboring cells

The timing and degree of urbanization

- ▶ We use Klein-Goldewijk's (2005) historical population maps
 - A consistent time series, 0-1600 (at century intervals) and 1700-2000 (decade intervals)
 - Exploits the identifiable location and persistence of cities
 - Infers rural and urban population for each cell in each year from historical records and growth rates
- ▶ We use these data to identify the year in which a location's urban population reached a series of thresholds:
 - Urban population densities of 1 and 5.67 people per sq. km.
 - Urban population fractions of 10%, 25%, and 50%.

Physical geography for rural productivity

- ▶ Cultivation suitability index due to Ramankutty et al. (2002).
 - Fixed characteristics of soils (carbon density, pH) and climate (growing degree-days, potential and actual evapotranspiration)
 - Measured on 0-1 scale, calibrated to the fraction of each cell under cultivation in 1992
- ▶ Winter frost, following Masters and McMillan (2001)
 - Frequency of frost in winter, after a frost-free summer (temperatures from IPCC 2002)
 - Helps control pests and disease vectors, differentiate temperate from tropical environment
 - Measured as a dummy variable for cells receiving at least 2.11 frost days in winter after frost-free summers (threshold estimated by Funke and Zuo 2003).

Physical geography for transportation costs

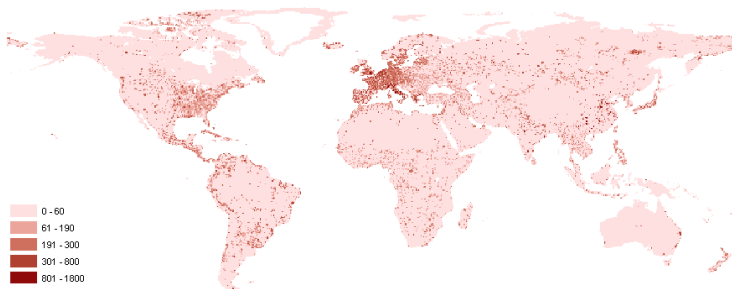
- ▶ Access to the sea
 - distance (in km) from centroid of grid cell to nearest ocean or sea coast
- ▶ Access to a navigable river
 - dummy variable for presence of at least one major river
 - orders defined by number of tributaries, in the Strahler Index
 - use rivers of order four and up, from Vorosmatry et al. (2000)
- ▶ Elevation
 - captures topographical barriers to agriculture and transport.
 - we use average elevation (in m.), from Hijmans et al. (2005)

Summary statistics (n=62,290)

variable	mean	std. dev.	min	max
number of years since the urban population reached:				
1 inhabitant per sq. km.	29.640	109.622	0	2000
5.67 inhabitant per sq. km.	15.859	61.893	0	1600
10% urbanization	34.169	94.903	0	1800
25% urbanization	19.829	53.716	0	1100
50% urbanization	9.690	31.038	0	700
cultivation suitability index	0.264	0.318	0	1
distance to coast (km)	521.984	511.641	1	2514.704
presence of navigable river	0.029	0.168	0	1
presence of frost in winter	0.116	0.320	0	1
land elevation (meters)	594.458	772.698	-76.667	5717.111

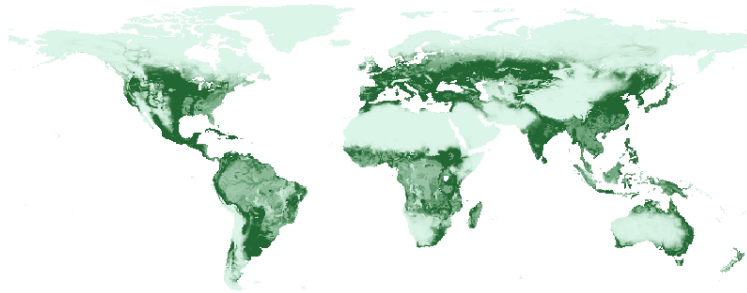
Note: Results include about 10,000 uninhabited cells, the number of which is almost constant over time.

The timing of urbanization



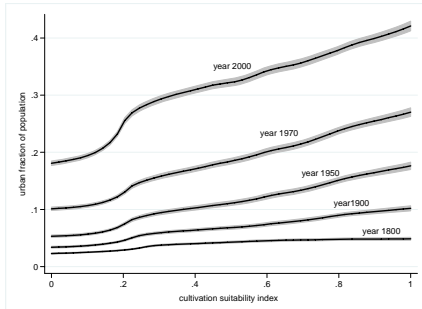
Number of years since transition to a 10% urban population

The cultivation suitability index

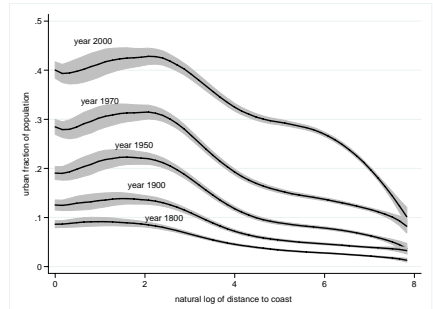


Worldwide distribution of land suitability for cultivation

Urbanization and physical geography in selected years, 1800-2000



Local polynomial regressions of urbanization on cultivation suitability index in each year



Local polynomial regressions of urbanization on natural log of distance to coast in each year

Econometric specification

- ▶ Strategy: regress the time since a grid cell reached each urbanization threshold on its geographic characteristics.

$$\ln T_{i,j} = \beta_0 + \beta_1 \ln cultiv_{i,j} + \beta_2 \ln coast_{i,j} + \beta_3 river_{i,j} \\ + \beta_4 frost_{i,j} + \beta_5 \ln elevation_{i,j} + \delta_j + \epsilon_{i,j}$$

- ▶ T_{ij} is number of years elapsed since grid cell i in country j reached the threshold.
- ▶ $cultiv_{ij}$ is the cultivation suitability index
- ▶ $coast_{ij}$ and $river_{ij}$ are navigable waterway variables
- ▶ $frost_{ij}$ and $elevation_{ij}$ are winter frosts and average elevation
- ▶ δ_j captures country fixed effects

Results from OLS

	(1) 1 inhabitant per sq. km.	(2) 5.67 inhabitants per sq. km.	(3) 10% urban	(4) 25% urban	(5) 50% urban
<i>cultiv</i>	0.329	0.229	0.303	0.246	0.158
<i>coast</i>	-0.130	-0.253	-0.131	-0.149	-0.148
<i>river</i>	0.570	0.493	0.913	0.809	0.647
<i>frost</i>	0.847	0.686	0.821	0.736	0.544
<i>elevation</i>	-0.089	-0.084	-0.097	-0.045	-0.008
<i>constant</i>	2.116	0.991	1.878	0.552	-0.735
R^2	0.355	0.354	0.274	0.248	0.196

$n = 62, 290$. All coefficients are significant at the $p < 0.01$ level.

All regressions use country fixed effects (not shown).

Results from Tobit

	(1) 1 inhabitant per sq. km.	(2) 5.67 inhabitants per sq. km.	(3) 10% urban	(4) 25% urban	(5) 50% urban
<i>cultiv</i>	1.732***	1.769***	1.187***	1.152***	1.157***
<i>coast</i>	-0.480***	-0.743***	-0.499***	-0.638***	-0.882***
<i>river</i>	1.699***	2.121***	2.134***	2.361***	2.856***
<i>frost</i>	1.884***	2.195***	1.787***	1.873***	2.070***
<i>elevation</i>	-0.109***	-0.151***	-0.150***	-0.036	0.080*
<i>constant</i>	4.180***	2.880***	2.769***	-0.721***	-8.464***

Note: $n = 62, 290$

Significance levels indicated by *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

All regressions use country fixed effects (not shown).

Marginal effects from Tobit

	(1) 1 inhabitant per sq. km.	(2) 5.67 inhabitants per sq. km.	(3) 10% urban	(4) 25% urban	(5) 50% urban
<i>cultiv</i>	0.388***	0.309***	0.330***	0.280***	0.217***
<i>coast</i>	-0.107***	-0.130***	-0.139***	-0.155***	-0.166***
<i>river</i>	0.414***	0.406***	0.659***	0.639***	0.599***
<i>frost</i>	0.456***	0.414***	0.534***	0.488***	0.415***
<i>elevation</i>	-0.024***	-0.026***	-0.042***	-0.009	-0.015*

Note: $n = 62, 290$

Significance levels indicated by *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

All results shown control for country fixed effects.

Spatial diagnostics and unobserved spillovers

- ▶ So far, we have used only each cell's own characteristics
- ▶ There could be distance-dependent interactions between cells
- ▶ Moran's I statistic on the errors of the OLS regression reveal significant spatial dependence
- ▶ LM tests suggest the spatial process occurs in the error terms

	(1) 1 inhabitant per sq. km.	(2) 5.67 inhabitants per sq. km.	(3) 10% urban	(4) 25% urban	(5) 50% urban
Moran's I	0.402	0.179	0.207	0.172	0.123
LM lag	126,015	55,539	75,567	55,901	31,176
LM error	134,991	83,057	110,694	76,783	39,390

Note: Moran's I statistics are significant at the $p < 0.001$ level.

Modeling neighborhood effects

- ▶ To account for spatial dependence in the error term, we use a spatial error model.

$$y = \beta X + \mu$$

$$\mu = \lambda W \mu + \epsilon$$

- ▶ W is a $n \times n$ matrix of weights indicating neighboring grid cells
 - Spatial range of the weights matrix is 5 orders of contiguity, implying an 11×11 neighborhood.

Results from the spatial error model

- ▶ Dependent variable is number of years since transition.

	(1) 1 inhabitant per sq. km.	(2) 5.67 inhabitants per sq. km.	(3) 10% urban	(4) 25% urban	(5) 50% urban
<i>cultiv</i>	0.315	0.231	0.300	0.254	0.176
<i>coast</i>	-0.044	-0.074	-0.050	-0.088	-0.097
<i>river</i>	0.635	0.516	0.630	0.617	0.522
<i>frost</i>	0.226	0.262	0.210	0.211	0.237
<i>elevation</i>	-0.168	-0.155	-0.139	-0.094	-0.074
λ	0.900	0.882	0.897	0.872	0.815

$n = 62, 290$. All coefficients are significant at the $p < 0.01$ level.

- ▶ Results are as predicted; note declining importance of cultivation and elevation, but rising importance of coasts

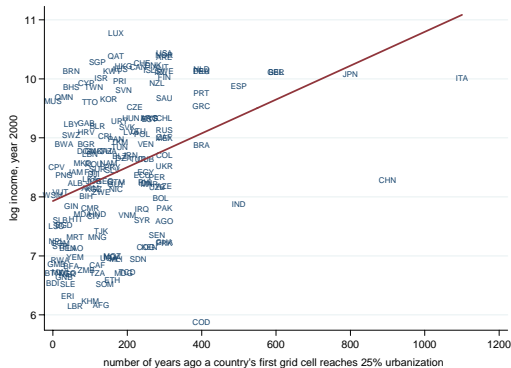
Results from the spatial error model

► Country fixed effects



Country effects from the spatial error model (25% urbanization)

Linking historic date of transition to year 2000 income



Country-level per capita income in year 2000 and country's first transition date (25% urbanization)

Linking historic date of transition to year 2000 income

- ▶ We test this relationship in a regression setting while accounting for cross-country spatial dependence using the spatial lag model.

$$\ln(\text{income}_i) = \beta_{0i} + \rho W \ln(\text{income}_i) + \beta_{1i} \text{trans}_i + \epsilon_i$$

Results from the spatial lag model

	(1) 1 inhabitant per sq. km.	(2) 5.67 inhabitants per sq. km.	(3) 10% urban	(4) 25% urban	(5) 50% urban
<i>trans</i>	0.0003	0.0009	0.0007	0.0015	0.0023
<i>constant</i>	1.80	1.88	1.83	1.95	1.94
ρ	0.777	0.76	0.76	0.740	0.746

$n = 169$. All coefficients but one are significant at the $p < 0.01$ level. *trans* in (1) is significant at $p < 0.1$ level.

Marginal effects

	1 inhabitant per sq. km.	5.67 inhabitants per sq. km.	10% urban	25% urban	50% urban
<i>trans</i>	0.0015	0.0037	0.0031	0.0058	0.0090

- ▶ Reaching 50% urbanization one year earlier raises modern incomes by nearly 1%.

Conclusions and implications

- ▶ A region's historical date of urbanization is associated with its physical geography
 - Earlier urbanization occurred in regions more suitable for cultivation, with winter frosts, closer to the ocean, with navigable rivers and with lower elevations
 - At higher levels of urbanization, coefficients are smaller for cultivation and elevation, but higher for distance to coast
- ▶ A country's date of first urbanization influences income today
- ▶ Looking forward, new technology could overcome old geography
 - to raise rural productivity
 - to lower transport costs