
Vernon Henderson, LSE
Tim Squires, Brown
Adam Storeygard, Tufts
David N. Weil, Brown

July 2015
Forces Shaping the Spatial Distribution of Population and Economic Activity

• Natural Characteristics: Some places are better to produce or live in than others.
• Agglomeration/Congestion: increasing returns, gains from trade, positive spillovers, but also transport costs for food and local negative spillovers (pollution, etc.). So people agglomerate, but not in just one place.
• Institutions (countries, etc.)
• Persistence (second nature): long-lived capital, equilibrium selection, and political power all lead to inertia in the locations of agglomeration.
First Nature

• Natural characteristics are persistent, but the “prices” attached to them change with technology and development
  – Air conditioning, irrigation
  – Value attached to amenities like pleasant climate
  – Food declines as a share of the consumption basket
  – Falling transport costs reduce the benefit of living in a food-producing area
  – Rising trade opportunities increase value of living near harbor
Agglomeration/Congestion

• Benefits and costs of agglomeration also change with technology and development
  – Clean water, sewage, and antibiotics lower congestion costs
  – Industrial structure and type determines benefits of agglomeration
Persistence

• Cities: we have many examples of city locations today determined by history
  – Mexico City: founded 1325 by Aztecs; population reached 200,000 before Spanish conquest in 1521; today 8.9 million.
  – The factors that attracted the Aztecs may no longer be relevant today

• Other examples: New England Colleges, etc.
Persistence:
Agglomeration vs. Natural Advantage

• Simply observing persistence does not tell us the relative importance of “inertia” vs. persistent natural advantage.

• Literature has looked at various examples to argue for importance of one or the other.
Big Idea

• Because of the importance of agglomeration, there are many possible equilibrium locations cities.
  – Historical persistence can then “select” a particular equilibrium
  – Locations will reflect the “prices” of natural characteristics at the time of agglomeration
  – By looking at current locations and knowing when agglomeration took place, we can estimate how prices attached to fixed natural characteristics have changed over time
Existing Empirical Work

– Cross-country work
  • Difficulty distinguishing geography from institutions
  • Averages across a lot of interesting within-country variation

– Within-country work: various results in different regions
  • Davis and Weinstein: natural advantage in Japan
  • Bleakley and Lin: persistence in the US
  • Michaels and Rauch: persistence in France, persistence overcome in UK

• This paper: looks at whole world, focusing on within-country variation
This paper

• Build a global dataset of geographic determinants of economic activity
  – Agriculture vs. trade
• Outline some facts about their power in explaining the distribution of our economic activity proxy: lights
• Develop a simple model of path dependence
• Provide evidence of this form of path dependence
  – Countries that developed/urbanized early are more affected by geographic features affecting agriculture
  – Countries that developed/urbanized late are more affected by features affecting trade
    • Even though they are more dependent on agriculture
    • Benefits from shaping future urbanization?
Night lights data

• US Air Force weather satellites
• Each satellite observes whole Earth between 8:30 and 10pm local time
• Processing
  – Avoid sunlight, moonlight, aurorae, clouds
  – Remove forest fires, gas flares, squid fishing
• Data we receive – a balanced panel of pixels
  – Annual averages 1992-2012 (sometimes 2 satellites/year)
  – 30 arc-second square pixels
    • slightly less than 1 km x 1 km at equator, decreasing with \( \cos(\text{latitude}) \)
• Uniform data density
Previous work with lights

• ln(lights) are correlated with ln(GDP) at country level
  – Long difference correlation: 0.53 (Henderson, Storeygard and Weil 2012)
  – Cross-sectional correlation: 0.93-0.96

• Subnational evidence with similar elasticities
  – China prefectures panel (Baum-Snow et al 2013, Storeygard 2014)
  – Near-global cross-section of subnational regions (Hodler and Raschky 2014)
Our Empirical Setup

• Unit of analysis: 0.25 x 0.25 degree (longitude/latitude) grid squares
  – 900 pixels (30x30) lights data
  – 777 km² at the equator, decreasing with cos(latitude)
  – Approx. 250,000 observations
    • vs. ~200 million at original lights resolution
  – Limit spurious autocorrelation
  – Limit magnitude of true spillovers
  – Many other variables are coarser anyway
Empirical Setup

• Radiance calibrated 2010 lights
  – Not topcoded anymore
  – Still bottom-censored: 36% of observations non-zero

• Dependent variable:

\[ y = \begin{cases} 
  \ln \left( \frac{\text{lights}}{\text{landarea}} \right) & \text{if } \text{lights} > 0 \\
  \ln(0.0043) & \text{if } \text{lights} = 0 \end{cases} \]

  – \text{landarea} adjusts for water and latitude
Measures of “First Nature”

• “Agriculture”
  – Binary: 12 biome dummies
  – Continuous: elevation, ruggedness (Nunn and Puga), temperature, precipitation, growing days, land suitability (Ramankutty et al.), abs(latitude)

• “Trade”
  – Binary: Coastal dummy, dummies for within 25 km of a natural harbor, ocean-navigable major river, or large lake
  – Continuous: distance to a coast

• All aligned on same 0.25 degree grid as lights
Country Fixed Effects

• Some of the effect of climate variables on lights may be due to Europeans bringing their institutions/human capital/etc. to some regions and not others.

• Including country fixed effects picks up the effect of nature holding this effect constant.

• Surprisingly, country F.E. don’t change the coefficients on nature terms by much.
## Basic First Nature Results

<table>
<thead>
<tr>
<th>Covariate Set</th>
<th>R-Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture and Trade</td>
<td>0.457</td>
</tr>
<tr>
<td>County Fixed Effects</td>
<td>0.336</td>
</tr>
<tr>
<td>Country FE, Agriculture, and Trade</td>
<td>0.565</td>
</tr>
<tr>
<td>Agriculture</td>
<td>0.435</td>
</tr>
<tr>
<td>Trade</td>
<td>0.056</td>
</tr>
<tr>
<td>Country FE, Agriculture</td>
<td>0.551</td>
</tr>
<tr>
<td>Country FE, Trade</td>
<td>0.354</td>
</tr>
</tbody>
</table>
Predicted calibrated lights in 2010
Baseline specification without country fixed effects
Extensive vs. Intensive Margins
(Fraction lit = 0.36)

<table>
<thead>
<tr>
<th>Margin</th>
<th>Ag&amp;Trade</th>
<th>Country fixed effects</th>
<th>Ag&amp;Trade&amp;FE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.457</td>
<td>0.336</td>
<td>0.565</td>
</tr>
<tr>
<td>Extensive</td>
<td>0.384</td>
<td>0.269</td>
<td>0.474</td>
</tr>
<tr>
<td>Intensive</td>
<td>0.280</td>
<td>0.246</td>
<td>0.375</td>
</tr>
</tbody>
</table>

- Table shows the R-Squared from each specification.
- Independent variables are full set of agriculture and trade variables.
- Geography variables explain more than country FEs alone on both margins.
- Geography explains the extensive margin better than the intensive margin.
How the *Path* of Development Affects Spatial Distribution

- In a world without spatial inertia:
  - Fraction of the population urbanized depends on agricultural productivity
  - Number of cities depends on agglomeration/congestion as well as trade costs
  - Location of cities depends on trade costs

- In a world with spatial inertia, the history of these things matters as well
Model

- Two goods: agriculture (f) and manufacturing (m)
- Two identical regions labelled coast (c) and interior (i)
- Each region can have at most one (congestible) city for manufacturing (to come: more cities)
- Fixed total population: $L = L_{fc} + L_{fi} + L_{mc} + L_{mi}$
- Free migration (long run), costly trade
Production/Preferences

• Agriculture (numeraire)
  – Per-worker production decreasing returns in labor: $A_f L_f$
  – Fixed absolute consumption, $\gamma$

• Manufacturing (prices $p_{mc}$ and $p_{mi}$)
  – Subject to both positive agglomeration effects and negative congestion effects
  – Per-worker production: $A_m L_m (1 + tL_m)^{1+}$

• Iceberg trade costs: $p_{mc} = p_{mi} (1)$
Agglomeration/Congestion

Manufacturing Productivity vs. City Population
Equilibrium

• Individuals move between regions and occupations such that utility (consumption of manufactured good) equalized in all.

• Stability: We consider
  – Moving people from one region to another (migration)
  – Moving people from one sector to the other within a region

• An equilibrium is stable if the people who are moved are worse off than those who didn’t
Inertia

• We don’t model formally model inertia in locations, but assume that the past can determine which of multiple equilibria is selected.
Technological Change

• We consider two dimensions of technology
  – $A$ is agricultural productivity
  – $\tau$ is inter-regional trade costs

• In the “pre-development” era:
  – $A$ is low $\Rightarrow$ most people work in agriculture
  – $\tau$ is high $\Rightarrow$ no inter-regional trade

• In the “developed” era:
  – $A$ is high $\Rightarrow$ most people work in manufacturing
  – $\tau$ is low $\Rightarrow$ inter-regional trade
Equilibrium with low $A$ and high $\tau$
Equilibrium with high $A$ and low $\tau$

Possibility #1: Symmetric
Equilibrium with high $A$ and low $\tau$

Possibility #2: Corner

- **Coast**
- **Interior**

Manufacturing Productivity

City Population

City Population
Political Economy Note
Regarding the Model

• “Trade” can include rent extraction, tribute, etc.
  – Anything that flows from cities to farms in return for food
Path dependence

- Initial: Low τ, low A
- Early agglomerators
- Late agglomerators
- Final: High A, high τ
Path #1: $A$ rises before $\tau$ decreases (early agglomerators)

**Coast**

- Initial
- After $A$ rises but before $\tau$ decreases
- Final

**Interior**

- Initial
- After $A$ rises but before $\tau$ decreases
- Final
Path #2: $\tau$ decreases before $A$ rises (late agglomerators)

- City Population
- Manufacturing Productivity

**Coast**
- Initial
- After $\tau$ decreases but before $A$ rises
- Final

**Interior**
- Symmetry is unstable
Model summary

• In countries that agglomerated early
  – $A$ rose before $\tau$ decreased
  – Agglomeration took place where it was good to grow food (path #1)
  – Current distribution of population/economic activity reflects agricultural productivity

• In countries that agglomerated later
  – $\tau$ decreased before $A$ rose
  – Agglomeration took place where it was good to manufacture/trade (path #2)
  – Current distribution of population/economic activity reflects trade possibilities

• Non-obvious (counter-intuitive?) since poor countries have a larger fraction of their population in agriculture
Applying the Model to the Data

- “agriculture” variables: proxy for “stuff that mattered more for early agglomeration”
- “trade” variables: proxy for “stuff that mattered more for late agglomeration”
- Model says that agriculture variables should determine locations relatively more in “early agglomeration” countries than in “late agglomeration” countries
Applying the Model to the Data

• How do we determine which countries are “early agglomerators” vs. “late agglomerators”?
• We use two measures:
  – Education in 1950 (proxy for overall technology)
  – Urbanization in 1950 (scope to change distribution)
  – In both cases, low value indicates later agglomeration
• In each case, we let the data choose a cutoff between the “early” and “late” groups (à la Durlauf and Johnson, 1995)
  – Loop through different division points and find the one that minimized sum of squared residuals
Cumulative distributions

- Average years of schooling in 1950
- Fraction of urban population in 1950

Cumulative fraction of the world population in 2010

- educ1950
- urban1950
Dividing Sample by Education without country fixed effects

Education: Total SSR (no country FE)

Cutoffs of average years of schooling in 1950

Total SSR

- 1800000
- 1790000
- 1780000
- 1770000
- 1760000
- 1750000
- 1740000
- 1730000
Dividing Sample by Education with country fixed effects

**Graph:**

Title: Education: Total SSR (with country FE)

Y-axis: Total SSR

X-axis: Cutoffs of average years of schooling in 1950

The graph shows a trend where the Total SSR decreases as the cutoffs of average years of schooling increase, reaching a minimum around the cutoff of 2.8. Beyond this point, the Total SSR increases with increasing cutoffs.
## Differential Contribution to R-Squared from Agriculture vs. Trade

<table>
<thead>
<tr>
<th>Educ</th>
<th>No country FE</th>
<th>Agriculture</th>
<th>Trade</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;= 3.6</td>
<td></td>
<td>0.537</td>
<td>0.065</td>
</tr>
<tr>
<td>&lt; 3.6</td>
<td></td>
<td>0.315</td>
<td>0.162</td>
</tr>
</tbody>
</table>
Differential Contribution to R-Squared from Agriculture vs. Trade with County Fixed Effects

<table>
<thead>
<tr>
<th></th>
<th>Just Country Fixed Effects</th>
<th>Fixed Effects plus Agriculture</th>
<th>Fixed Effects plus Trade</th>
<th>Increase in R-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educ&gt;3</td>
<td>0.350</td>
<td>0.629</td>
<td>0.393</td>
<td>0.279</td>
</tr>
<tr>
<td>Educ&lt;3</td>
<td>0.285</td>
<td>0.443</td>
<td>0.396</td>
<td>0.158</td>
</tr>
</tbody>
</table>
Differential coefficients

\[ y_{ic} = (1 + A \text{Late}_c) A X_{Aic} + (1 + T \text{Late}_c) T X_{Tic} + \text{Late}_c + u_{ic} \]

- For cell \( i \) in country \( c \),
  - \( X_A \) are agriculture variables;
  - \( X_T \) are trade variables
  - \( \text{Late} \) is a late agglomerator dummy
  - \( y \) is \( \ln(\text{lights/area}) \)

- Estimate with nonlinear least squares

- Model prediction: \( A < 0 \)
  \( T > 0 \)
Differential coefficients

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>below cut × agriculture</td>
<td>-0.254***</td>
<td>-0.165***</td>
</tr>
<tr>
<td></td>
<td>(0.0105)</td>
<td>(0.0184)</td>
</tr>
<tr>
<td>below cut × trade</td>
<td>3.137***</td>
<td>2.478***</td>
</tr>
<tr>
<td></td>
<td>(0.288)</td>
<td>(0.218)</td>
</tr>
</tbody>
</table>

Split based on Education
Country FEs No
Observations 228,690
R-squared 0.482

Education Yes
Observations 228,690
R-squared 0.570
Conclusion

• Physical geography accounts for a substantial fraction of global spatial variation in economic activity

• Factors affecting agriculture matter relatively more in countries that urbanized early

• Factors affecting trade matter relatively more in countries that urbanized late
  – Scope for influencing optimality of location choice of relatively less urbanized countries?

• Future work
  – Growth
  – Counterfactuals
  – Optimality?
Implications for Public Policy
Why Spatial Economics is Different

• Because of agglomeration effects and persistence, the spatial pattern of population and economic activity is *inherently* characterized by multiple equilibria.

• It is possible for some of these equilibria to be better than others.

• There is a potential role for policy in selecting which equilibrium is observed.

• This can be good (opportunity to do it right) or bad (a further way that bad institutions can mess things up).
What is a Good Spatial Distribution?

• Are rich countries a good model for what a good spatial equilibrium looks like?
  – Simple theory/narrative suggests maybe not
  – Our analysis of data seems to confirm this

• This paper is a step toward understanding what the best spatial distribution would be
## Selected Biome Coefficients

<table>
<thead>
<tr>
<th>Biome</th>
<th>% of obs</th>
<th>Coefficient</th>
<th>Coeff (w/FEs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tropical moist forest</td>
<td>11.6</td>
<td>-0.118</td>
<td>-0.303***</td>
</tr>
<tr>
<td>tropical dry forest</td>
<td>2.2</td>
<td>-0.870***</td>
<td>0.132*</td>
</tr>
<tr>
<td>temperate broadleaf</td>
<td>10.4</td>
<td>1.786***</td>
<td>1.329***</td>
</tr>
<tr>
<td>temperate conifer</td>
<td>3.3</td>
<td>0.713***</td>
<td>0.111</td>
</tr>
<tr>
<td>boreal forest</td>
<td>16.7</td>
<td>-0.546***</td>
<td>-1.366***</td>
</tr>
<tr>
<td>tropical grassland</td>
<td>12.0</td>
<td>-1.091***</td>
<td>-0.258***</td>
</tr>
<tr>
<td>temperate grassland</td>
<td>7.7</td>
<td>0.738***</td>
<td>0.961***</td>
</tr>
<tr>
<td>montane grassland</td>
<td>3.3</td>
<td>0.600***</td>
<td>0.720***</td>
</tr>
<tr>
<td>tundra</td>
<td>12.4</td>
<td>-1.095***</td>
<td>-1.630***</td>
</tr>
<tr>
<td>Mediterranean forest</td>
<td>2.4</td>
<td>0.788***</td>
<td>1.361***</td>
</tr>
<tr>
<td>mangroves</td>
<td>0.4</td>
<td>0.360**</td>
<td>-0.0771</td>
</tr>
<tr>
<td>desert</td>
<td>17.6</td>
<td>reference</td>
<td>reference</td>
</tr>
</tbody>
</table>
## Other coefficients

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Coefficient</th>
<th>Coeff (w/FEs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>elevation (km)</td>
<td>0.602</td>
<td>0.575***</td>
<td>0.156***</td>
</tr>
<tr>
<td>ruggedness</td>
<td>2.7</td>
<td>-0.0101***</td>
<td>-0.0190***</td>
</tr>
<tr>
<td>temperature</td>
<td>9.97</td>
<td>0.187***</td>
<td>0.139***</td>
</tr>
<tr>
<td>precipitation</td>
<td>60.7</td>
<td>-0.00909***</td>
<td>-0.0115***</td>
</tr>
<tr>
<td>growing days</td>
<td>139</td>
<td>0.0103***</td>
<td>0.00877***</td>
</tr>
<tr>
<td>land suitability</td>
<td>0.273</td>
<td>2.730***</td>
<td>2.218***</td>
</tr>
<tr>
<td>abs(latitude)</td>
<td>38.34</td>
<td>0.131***</td>
<td>0.0519***</td>
</tr>
<tr>
<td>1(coast)</td>
<td>0.098</td>
<td>0.436***</td>
<td>0.447***</td>
</tr>
<tr>
<td>dist. to coast (km)</td>
<td>485</td>
<td>-0.000693***</td>
<td>-0.000674***</td>
</tr>
<tr>
<td>1(harbor)</td>
<td>0.027</td>
<td>1.685***</td>
<td>1.454***</td>
</tr>
<tr>
<td>1(river)</td>
<td>0.027</td>
<td>0.786***</td>
<td>0.718***</td>
</tr>
<tr>
<td>1(big lake)</td>
<td>0.127</td>
<td>0.335***</td>
<td>0.192***</td>
</tr>
</tbody>
</table>
% urbanized, India
% urbanized, USA
Global real freight index

* excludes periods including world war years
Dividing Sample by Urbanization without country fixed effects

Urban population fraction in 1950
Dividing Sample by Urbanization with country fixed effects

Urban population fraction in 1950
### Differential coefficients

<table>
<thead>
<tr>
<th></th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>below cut × agriculture</td>
<td>-0.252***</td>
<td>-0.198***</td>
</tr>
<tr>
<td></td>
<td>(0.00959)</td>
<td>(0.0171)</td>
</tr>
<tr>
<td>below cut × trade</td>
<td>2.821***</td>
<td>2.095***</td>
</tr>
<tr>
<td></td>
<td>(0.280)</td>
<td>(0.191)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Split based on</th>
<th>Urbanization</th>
<th>Urbanization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country FEs</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>243,661</td>
<td>243,661</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.472</td>
<td>0.569</td>
</tr>
</tbody>
</table>
Differential Contribution to R-Squared from Agriculture vs. Trade (using urbanization cut)

<table>
<thead>
<tr>
<th></th>
<th>No country FE</th>
<th>With country FE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agriculture</td>
<td>Trade</td>
</tr>
<tr>
<td>Urb &gt;= 0.44</td>
<td>0.527</td>
<td>0.071</td>
</tr>
<tr>
<td>Urb &lt; 0.44</td>
<td>0.350</td>
<td>0.152</td>
</tr>
</tbody>
</table>
Overall motivation

• How does geography affect the distribution of population and economic activity? Theory
  – Natural advantage/physical geography
  – Agglomeration and congestion
  – History, persistence
  – Value of natural characteristics varies with technology
    • Temperature after air conditioning
    • Rivers after trains
  – Same for value/cost of agglomeration and congestion: antibiotics, IT, refrigerated freight, etc.