Chinese Counties

Geographic Determinants

Graham Joncas (戈雷)

17210680479@fudan.edu.cn

https://github.com/gjoncas

School of Economics
Fudan University

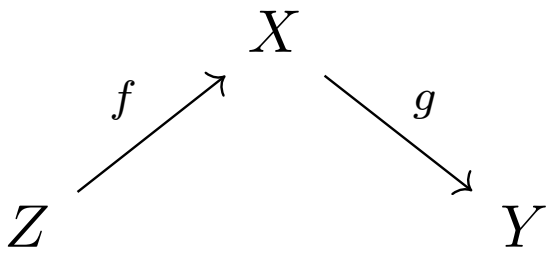
EMA Thesis (Advisor: Liu Yu 刘宇)
April 13, 2018

Outline

Introduction

- Counties compete with each other to attract firm investment
- More counties in a given area $\rightarrow$ stronger competition
- So: pro-business policies can be explained by county density
- Endogeneity: governors can adjust county boundaries for political reasons (e.g. including a certain town to inflate GDP)
- Need exogenous variable to explain variation in county density

Instrumental Variables



Y: independent variable
X: dependent variable

Z is an instrumental variable if:

1. $Z \xrightarrow{f} X$: Z causes X
2. $Z \not\to Y$: Z affects Y only through X
3. $Y \not\to Z$: Z not caused by Y, nor by factors affecting Y

Idea: use geographic variability and agricultural productivity as instrumental variables (Z)

to overcome endogeneity problem in county density (X)

so that others can research competition in policies (Y) among different counties

Previous Research – Michalopoulos (2012)

Cross-country regression

$$ln(\#\text{ of languages}) = \beta_0 + \beta_1 \cdot \text{latitude}_i + \beta_2 \cdot \text{var\_elevation}_i + \beta_3 \cdot \text{var\_land\_qual}_i + \beta_4 \mathbf{X}_i + \varepsilon_i$$

The Origins of Ethnolinguistic Diversity

Aim: uncover common exogenous features behind ethnic diversity

- High land quality, diverse terrain lead to more ethnic groups
- Backs up world-historical narratives with geographic data
- Identifies when we can & cannot treat diversity as exogenous
- Robustness checks likely helpful for other economic research

History of China's Counties (县)

- Most of China's counties have existed since Qin Shi Huang
- Ancestors' concerns: geography, agricultural productivity (Z)
- Want to show: for a broad class of policies Y, Z is exogenous

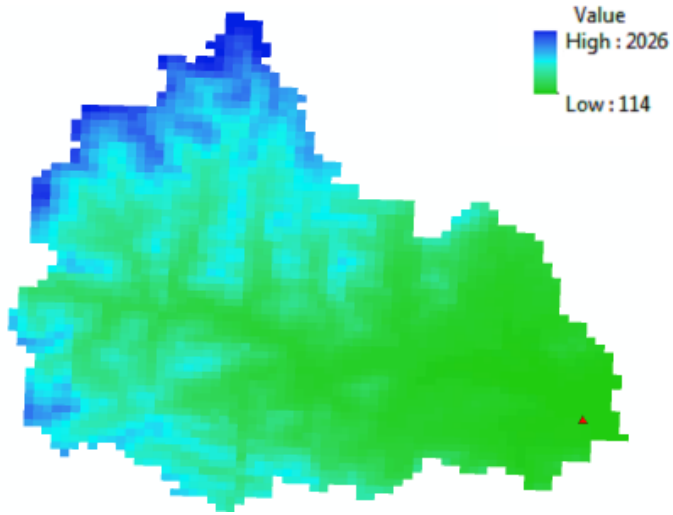
Two (contradictory) theories of how borders are determined:

- 山川形便 – advantages offered by terrain (mountains & rivers)
- 犬牙相错 – not letting local governments have enough geographic advantages that they could become independent

Both theories are consistent with exogeneity, but in different ways.

ArcGIS

(a) Nightlights



(b) Elevation

- Common tool in economics – e.g. nightlights as proxy for GDP
- Various open-source datasets available (e.g. elevation, rivers)
- Collect summary statistics per unit of area (e.g. 100km^2)

Regression Analysis

Using ArcGIS data, run the following regression:

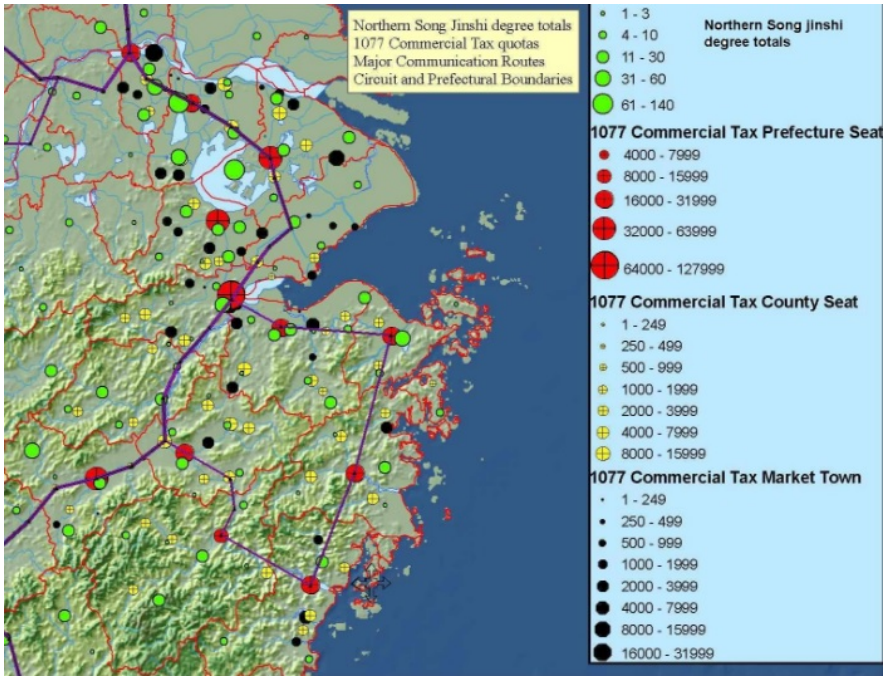
$$\text{county\_density} = \beta_0 + \beta_1 \cdot \text{geo\_var} + \beta_2 \cdot \text{agri\_prod} + \alpha \mathbf{X} + \varepsilon$$

Where geo_var = geographic variation (variance of land height)
agri_prod = agricultural productivity
$\alpha \mathbf{X}$ = other geographical control variables

Expect to see:

- High β_2: high agricultural productivity $\rightarrow$ more counties
- Unclear β_1: more variance $\rightarrow$ hard to govern $\rightarrow$ more counties
  or: more variance $\rightarrow$ lump together $\rightarrow$ fewer counties

Data – China Historical GIS

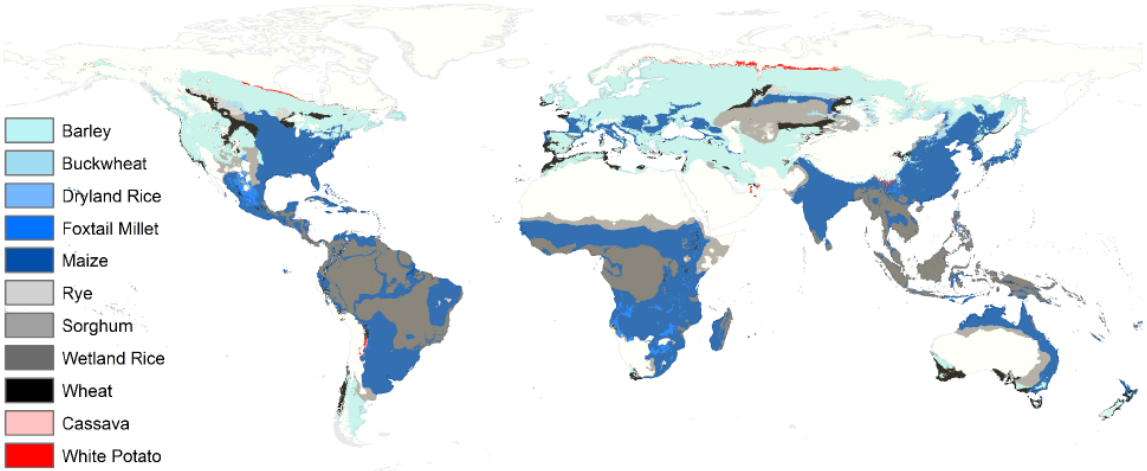


- CHGIS: datasets of administrative system between 221 BC & 1911 AD and major non-administrative towns for 1820 & 1911
- Use digital elevation model (DEM) – constant since 1911
- Calculate variance in elevation over areas of 100km^2

Data – Global Agro-Ecological Zones

- GAEZ: global estimates of various potential crop yields
- Yields (in tons/ha/year) for 11 cereals and 4 roots & tubers
- Two categories of water supply: rain-fed and irrigation
- Three levels of inputs: high, medium, low

- Summarize into general index of agricultural productivity
- USDA National Nutrient Database: convert into calories



Optimal crop in terms of caloric yields among cereals, roots & tubers

Thesis Structure

1. Outline history of Chinese counties (w.r.t. exogeneity)
2. Get summary statistics from geographical datasets
3. Regress county_density on geo_var, agri_prod, etc.
4. Interpret regression results, do robustness checks
5. Outline implications for policy analysis

Next steps:

- Find datasets for control variables, e.g. rivers
- Explore Chinese-language research on 县级边界
- Learn ArcPy — Python scripting for ArcGIS

Reproducible Research

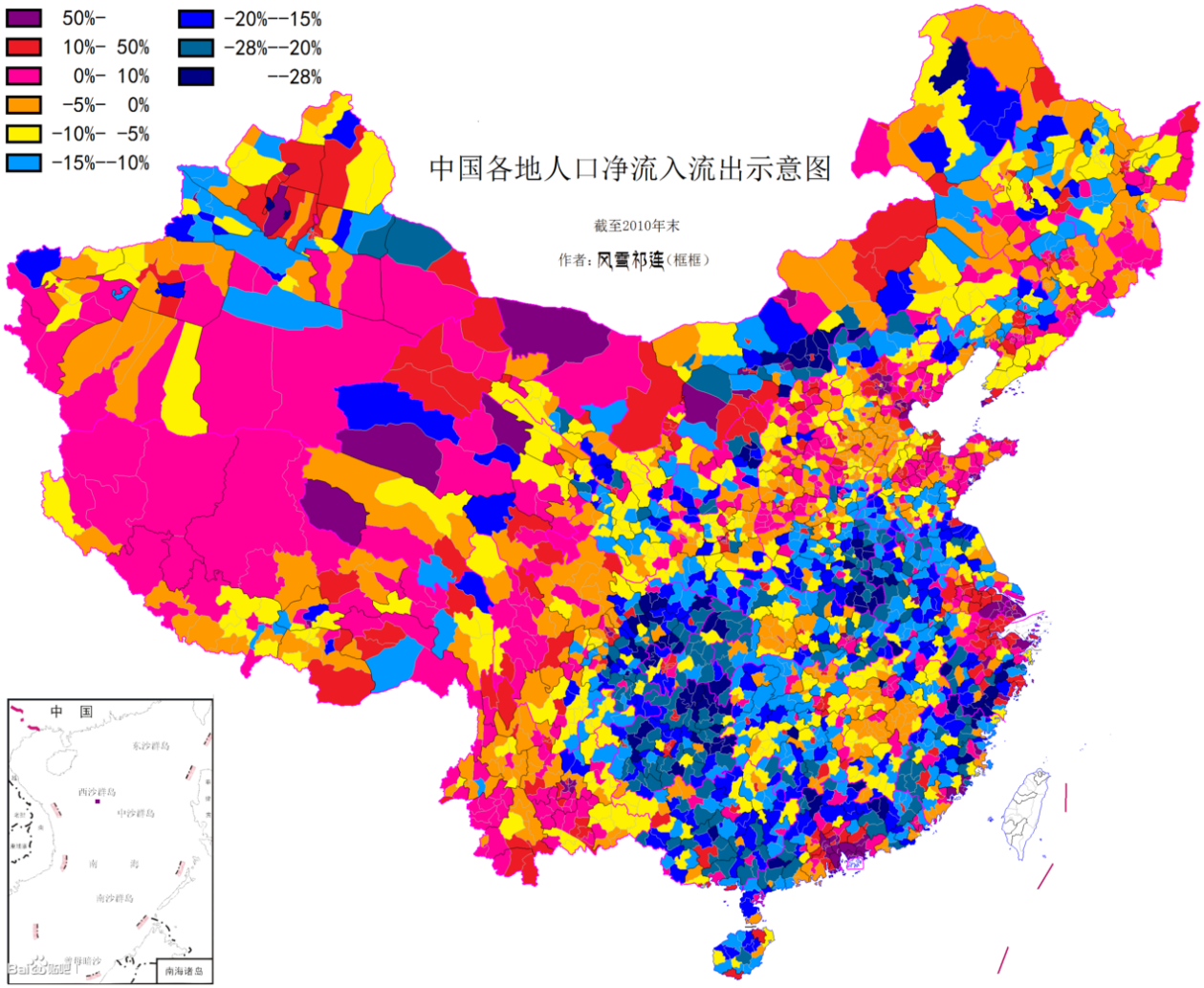
Gentzkow & Shapiro (2014)

1. Automate everything that can be automated
2. Write one script that executes all code from beginning to end

- Want results to be easily replicable by other researchers
- Many tools from computer science, e.g. version control (Git)
- ArcGIS (data) $\rightarrow$ Python (regressions) $\rightarrow$ LaTeX (text)

Gentzkow, M. & Shapiro, J. (2014). "Code & Data for the Social Sciences: A Practitioner's Guide." University of Chicago mimeo. Retrieved from `http://web.stanford.edu/~gentzkow/research/CodeAndData.pdf`

The End


50%–
10%– 50%
0%– 10%
–5%– 0%
–10%– –5%
–15%––10%
–20%––15%
–28%––20%
––28%
中国各地人口净流入流出示意图
截至2010年末
作者：风雪祁连（框框）
中 国
南海诸岛