Chinese Inter-County Competition and Pro-Business Policy

An Instrumental Variable Approach

Graham Joncas (戈雷)

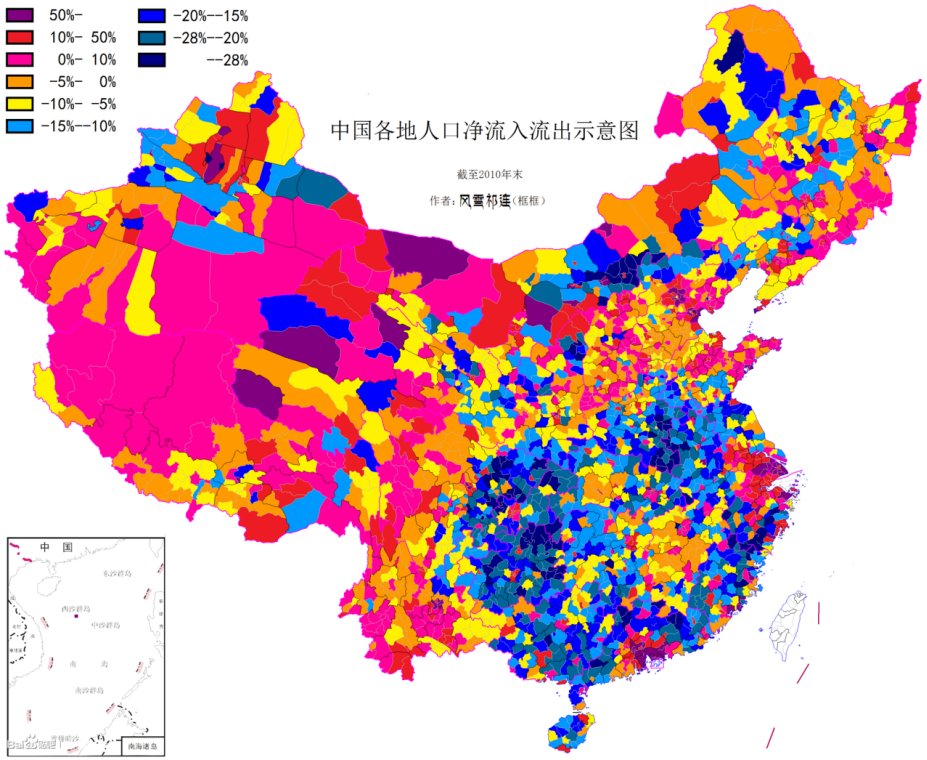
17210680479@fudan.edu.cn

School of Economics
Fudan University

EMA Thesis
Advisor: Liu Yu (刘宇)
May 25, 2018

Outline

Introduction



- Claim: China's success hinges on competition among counties
- More competition $\rightarrow$ more pro-business policies (Cheung, 2014)
- Very little research, since competition is hard to measure

Main Idea

Ideal regression equation

$$\text{pro\_business\_policy} = \beta_0 + \beta_1 \cdot \text{county\_competition} + \alpha \mathbf{X} + \varepsilon$$

- Can use tax enforcement to measure pro-business policies
- Measure effective tax rate: $\frac{\text{tax paid}}{\text{sales}}$ (vs. government tax rate)
- Use county density as proxy for county_competition
- 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

Problem – Endogeneity

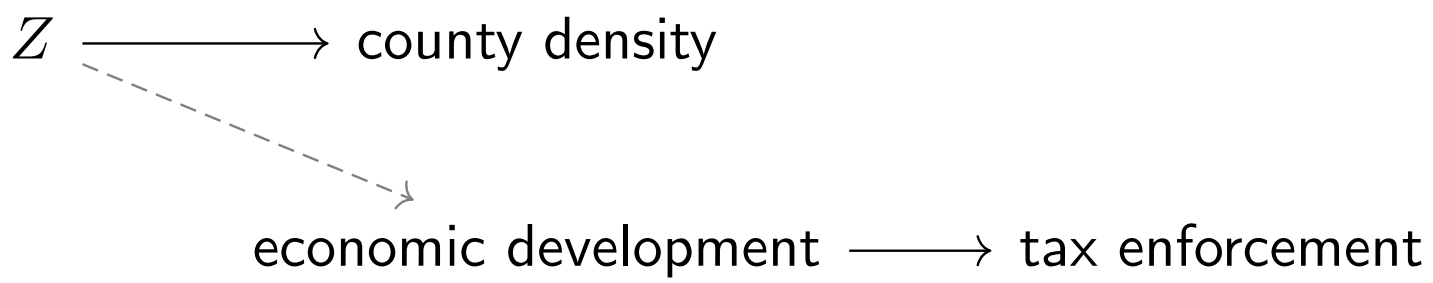
Firm name	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
宝鸡市爱姆食品有限责任公司						渭滨区	金台区	渭滨区	渭滨区	渭滨区
新疆天风发电股份有限公司		新市区	新市区	新市区	天山区	新市区	新市区	新市区	新市区	新市区
常州金马纺织品有限公司				天宁区	天宁区	武进区	天宁区	天宁区	天宁区	天宁区
常州科新永安磁电设备有限公司			新北区	新北区	钟楼区	钟楼区	新北区			
广通机械工程有限公司		黄埔区	黄埔区	黄埔区	黄埔区	黄埔区	萝岗区	萝岗区	萝岗区	黄埔区
宝鸡市热力有限责任公司				渭滨区	渭滨区	金台区	金台区	金台区	渭滨区	渭滨区
内蒙古兴华服装厂	赛罕区	玉泉区	玉泉区	玉泉区	赛罕区	赛罕区	赛罕区	赛罕区	赛罕区	赛罕区
常州天元工程机械有限公司	新北区	新北区	天宁区	天宁区	天宁区	天宁区	天宁区	天宁区	天宁区	新北区

Firms are administered by new county (red), then register back to original

- Governors can adjust county boundaries for political reasons
  $\rightarrow$ e.g. including a certain town to inflate GDP
- Possible endogeneity: some historical factor may affect both county density and tax enforcement $\rightarrow$ spurious correlation
- Need exogenous variable to explain variation in county density

Instrumental Variable

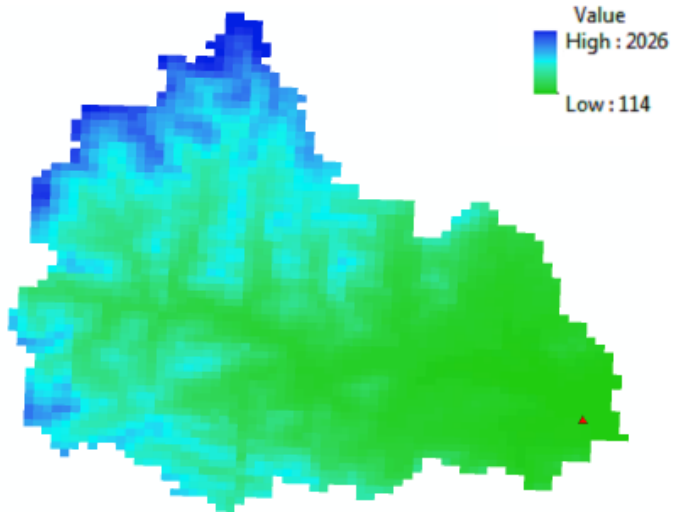
Want IV correlated with county density, but not development



- Most of China's counties have existed since Qin Shi Huang
- Ancestors' concerns: geography, agricultural productivity (Z)
- Geography affects county density, is clearly exogenous
- Can control for economic development using control variables
- TSLS: must have same set of controls in IV & main regression

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)

TSLS Regression

Using ArcGIS data, run the following regression:

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

Where geo_var = geographic variation (variance of land height)
agri_prod = agricultural productivity
$\mathbf{X}$ = control variables (for economic development)

Then, estimates for county density are used to run main regression:

$$\text{tax\_enforcement} = \beta_0 + \beta_1 \cdot \widehat{\text{county\_density}} + \alpha \mathbf{X} + \varepsilon$$

Expect to see $\beta_1 < 0$: higher county density leads to less taxes

Auxiliary Results

Geo-economics of county density

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

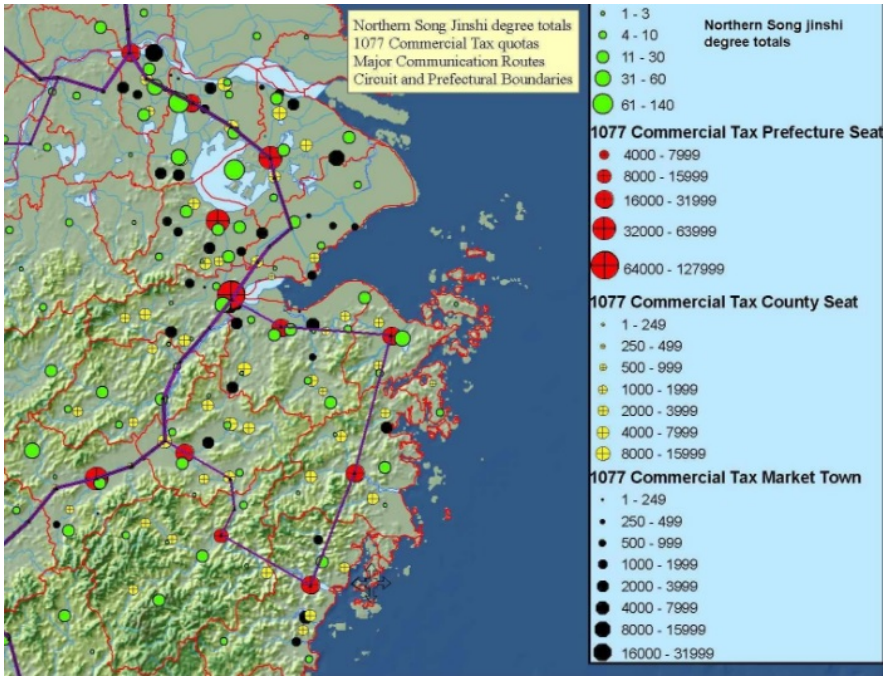
Two (contradictory) theories of how county density is determined:

- 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

First theory predicts $\beta_1 > 0$: rough terrain makes an area harder to govern, thus leads to more counties

Second theory predicts $\beta_1 < 0$: areas with smooth terrain have higher county density, since emperor limits their size

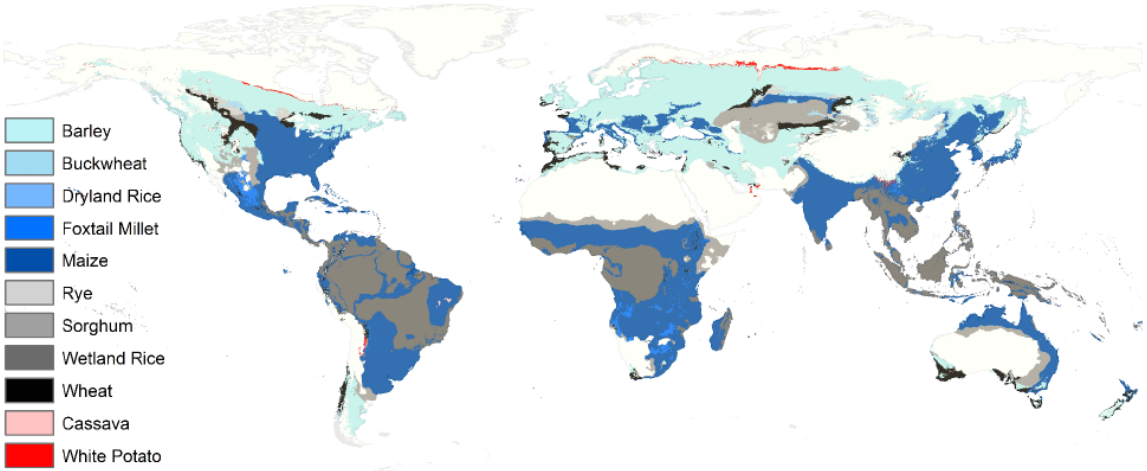
Data – China Historical GIS



- 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

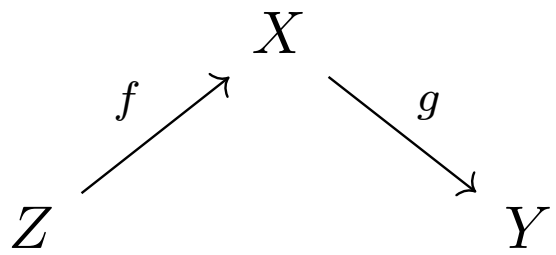
Thesis Structure

1. Outline theories of Chinese inter-county competition
2. Get summary statistics from geographical datasets
3. Instrumental variable analysis (e.g. strength of IV)
4. Regress tax_enforcement on county_density + controls
5. Interpret regression results, do robustness checks

Next steps:

- Find datasets for control variables, e.g. county GDP
- Explore Chinese-language research on 县与县之间的竞争
- Learn ArcPy — Python scripting for ArcGIS

Summary



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
4. $\nexists W$ s.t. $W \to Z$ & $W \to Y$: no W causes both Z and Y

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

which overcomes endogeneity problems in county density (X)

to measure pro-business policies (Y) among competing counties