学校代码： 10246
学　　号：17210680479

復旦大學

硕 士 学 位 论 文

县与县之竞争对中国亲商业政策的影响

The Impact of Inter-County Competition on Pro-Business Policy in China

院　　　　系：　经济学院
专　　　　业：　应用经济学（中国经济）
姓　　　　名：　戈雷
指 导 教 师：　刘宇讲师

完 成 日 期：　2019 年 03 月 10 日

UNIVERSITY CODE: 10246

STUDENT ID: 17210680479

THE IMPACT OF INTER-COUNTY COMPETITION ON PRO-BUSINESS POLICY IN CHINA

MA THESIS

Presented in Partial Fulfillment of the
Requirement for the Master Degree by the
Graduate School of Fudan University

by

Joncas, Graham

Fudan University

2019

Defense Date: May 14, 2019 EMA Chinese Economy Program

指导小组成员名单

刘　宇	讲　师
沈国兵	教　授
韦　潇	教　授
奚锡灿	讲　师

目　　录

摘　　要

中国是如何在制度不健全的情况下保持经济快速增长的，这一问题是众多“中国谜题”之一。越来越多的人将中国的经济成功归因于其“目标责任制”，也就是说，地方官员的晋升与其所在县的经济增长水平密切相关，这样一来，官员为了晋升不得不想法设法发展经济，提高自己在官场中的竞争力。

各县都希望吸引投资，但企业在投资时对具体地方县却并无明显偏好。因此，各县必须相互竞争以获得企业投资：一个特定地区的县越多（县的密度越高），竞争就越激烈。衡量支持商业政策的指标之一是有效税率，即中央税率税务执行程度。

一个潜在的问题是，可能有一些因素同时影响县的密度和税率，造成虚假的相关性。例如，一个原本较富裕的地区可能有较多的县，经济发展也较快（这将影响税收）。克服内生性问题需要一个与县域密度相关、与经济发展不相关的变量。因此，在本研究中，地理因素将替代县域密度作为工具变量，经济发展则将作为控制变量。

本文共有三个发现。在以往关于地理要素的讨论中有这样一个问题：有利地区（肥沃的土地，平坦的地形）是否会因为当地居民想最大限度地提高他们的利益而拥有更少/更大的县，或者他们是否会因为中央政府想阻止他们独立而拥有更多/更小的县。本文的第一个发现为这一争论提出了新的证据——事实上，更肥沃或者地形更平坦的地区都往往拥有更多的县。

第二个发现是，与 TSLS(0.092%)相比，OLS(0.037%)低估了县的密度对有效税率的影响，在 100 公里半径内，每增加一个邻居，意味着税率降低 0.092%。这一发现即使在许多其他替代性规范的制约下都是有效的，包括以毗邻县而非远邻县测量县域密度，使用替代性的税率，使用人均 GDP 而不是夜间灯光分布来衡量富裕程度。

第三个发现是，贫困县可能并非"一无所长"—他们作为贫困县的同时，很可能地形丰富，农业生产力高。在这类县，肥沃的土地使他们无法转向更有利可图的产业，崎岖的地形使他们相对于生产力高、地势平坦的县失去了比较优势。由于贫困，他们征收的税率高于其他县，阻碍了商业投资，陷入恶性循环。

关键词：中国谜题、县、贫困县、税务执法

中图分类号：F276.6

Abstract

The ‘China puzzle’ is the question of how China could sustain rapid economic growth despite poor institutions. A growing consensus attributes China’s economic success to its ‘target-responsibility system’ where promotion of local officials is tied to their county’s level of growth, leading to fierce competition.

Counties wish to attract investment, but firms are indifferent between locating in one county or another nearby. Thus, counties must compete with one another for firm investment: the more counties in a given area (higher *county density*), the fiercer competition is. One index for measuring pro-business policy is the *effective tax rate*—how much of the federal tax rate is enforced.

A potential problem is that some factor may affect both county density and the tax rate, causing spurious correlation; e.g. an area richer to begin with will have more counties and also more development (which will affect taxes). Overcoming endogeneity requires a variable correlated with county density, but not development. Hence, this study uses geography as an instrument for county density, controlling for development.

This thesis presents three findings. The first weighs in on a geographical debate, of whether favorable areas (fertile land, smooth terrain) will have fewer/larger counties because people living there want to maximize their benefits, or more/smaller counties because the central government wants to stop them from becoming independent. In fact, both more fertile areas tend and areas with smoother terrain tend to have more counties, supporting the latter theory.

The second finding is that OLS highly understates the effect of county density on the effective tax rate (as 0.037%) compared to TSLS, where each additional neighbor within a 100km radius implies a 0.092% lower tax rate. This discrepancy is robust to numerous alternative specifications for county density, tax rate, and wealth.

The third finding is that poverty counties may not simply be worse in all aspects, but there may exist a class of poverty counties with both high geographic variation and high agricultural productivity. In such counties, fertile land keeps them from switching to more profitable industries, but rough terrain makes them lose out relative to high-productivity, smooth-terrain counties. Due to their poverty, they charge higher tax rates than other counties, deterring business investment and leading to a vicious cycle.

Keywords: china puzzle, counties, poverty counties, tax enforcement

Chinese Library Classification Code: F276.6

Introduction

China's rapid development has been attributed to its inter-county competition (Cheung, 2014). The idea is that local governments operate according to a ‘business model’, competing for investment by firms (Li & Zhou, 2005). Firms are largely indifferent between locating in one county or one nearby, so that governments must offer incentives to firms. Thus the more counties are in an area, the stronger each county must compete. Hence, *county density* (number of counties in a unit of area) is a key variable to explain government incentives for firms.

One notable government incentive is tax enforcement. The tax rate is exogenously set by the government, and local governments cannot change it. Yet, they can choose the degree to which they will enforce the tax. Using firm-level data, the *effective tax rate* is measured by taxes paid as a proportion of sales. This indexes pro-business policies, which in turn can be explained by county density.

Yet, a curious problem arises where governors adjust county boundaries for political reasons (Liu, Tang & Zhao, 2017; fig. 1). This could indicate an endogeneity problem, if some historical factor exists that is correlated with both county density and tax enforcement. This would cause a spurious correlation between the two variables. To solve this, this thesis will perform an instrumental variables analysis, using a variable correlated with density but not economic development.

This thesis uses firm-level data spanning 2,780 counties to investigate the effect of county density on tax enforcement. Basic OLS regression shows a small negative effect, but by using geographic variables as instruments, TSLS regression shows an effect 2.5 times larger than the OLS estimate. Specifically, each neighbour within 100km of a county translates into a 0.092% lower effective tax rate. This discrepancy is robust to numerous alternative specifications for county density, tax rate, and wealth.

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

Fig. 1: Firms are administered by new county (red), then register back to original

Chapter 1 – Counties and the China Puzzle

The Formation of Chinese Counties

China's institutional system take the form of a *regionally decentralized authoritarian* regime, combining political centralization with regional economic decentralization (Xu, 2011: 1078). We can picture this graphically as a ‘flat’ level of nodes (all of China's counties), all connected to a single node (the Chinese Communist Party, or CCP). This arrangement has existed from Qin Shi Huang’s unification of China up to the present, and it is worthwhile to ask why China's counties are so numerous, and why they have lasted up to the present day.

In broad strokes, Li (2014) frames China's county system as a result of the shift from the Zhou dynasty's ritual-oriented (礼 *lǐ*) society, to the Qin dynasty's law-oriented (法 *fǎ*) society. Before the rise of the imperial state, the major political entities had been tribal clans, with kinship-based inheritance of political positions (Li, 2014: 73-4). The King of Zhou “exercised his rights in the name of a representative of the clan” (2014: 74), and social norms took the form of conformity to traditional values (ibid., 73) as opposed to coercive force.

Despite failed attempts to reform the political system from hereditary to merit-based, toward the end of the Zhou dynasty the system had degenerated into a form of patronage, in which aristocrats were granted land in exchange for their loyalty (2014: 75). With a fixed amount of land in the realm, such patronage could in the long term only lead the King of Zhou to become weaker, and the local governments of the various fiefdoms to grow stronger.

The key reforms that would later allow for the unification of China were led by Shang Yang (390–338BC) during the Warring States period. Most important was a merit-reward system assigning land and titles to soldiers based on their military exploits, regardless of their social background (2014: 81). Backed by the support of the army, the new system undermined clan powers that could have challenged the emperor. The ruling sovereign thus gained the status of an autocratic monarch, in a position to impose laws (法) backed by state power.[1]

---

[1] The actual word ‘county’ (*xiàn*; simplified 县, traditional 縣) has a convoluted history, changing subtly in meaning at different points of history. The《文源》dictionary writes: “从木，从系，持首” — that is, according to inscriptions on ancient bronze, the original character for 縣 was composed of the characters for wood (木), for department (系), and political capital (首, literally ‘head’). It began as a homophone for *huán* (寰, also written as 還 ‘returning’ or 環 ‘circle’), which denoted “areas surrounding the state capital" (Li, 2014: 83).

Thus, Qin Shi Huang’s unification of China amounted to monopolizing property rights for land, and imposing institutions to tie the population to its native land, (2014: 87). A centralized system for appointing officials, as well as the civil service examinations, displaced hereditary political position, with equality of rights enforced by the military (ibid.). Feudal powers had to submit to the emperor or else lose their position. Thus, the fragmented fiefdoms that led to the Zhou dynasty’s downfall became a political asset, as the large number of atomized local governments could not challenge the center.

Counties & Chinese Economic Growth

The ‘China puzzle’ is the question of how China managed to achieve such high economic growth despite poor institutions. In most developing counties with poor institutions, local governments are characterized by rent-seeking behaviour, whereas Chinese local governments are highly involved in “building local infrastructure, encouraging local businesses and attracting foreign investment” (Li & Zhou, 2005: 1744). Using an ordered probit model, Li & Zhou (2005) find that promotion of local officials is closely related to economic performance. The end outcome is that China's economy is run like a business.

This system has proved successful because of three characteristics of China's economic system. The first is that promotions are entirely in the hands of the central government, who are able to base personnel decisions purely on economic indices as opposed to more personal concerns (2005: 1747). Second, China's economy is structured in a multidivisional way that “makes each leader's performance individually distinguishable and comparable” (ibid.). Third, China's public sector is largely insulated from its private sector, so that government officials have few prospects beyond promotion up the political hierarchy (ibid.).

While this system proved remarkably successful within China's post-1978 institutions, as time advances its flaws have begun to manifest. Thus this section examines the various successes of China's economic growth as led by local governments, but also indicates where, as priorities begin to shift from growth on its own toward broader concerns, the system has begun to work ‘too well’.

Blanchard & Schleifer (2001) offer a simple but helpful model of why China's local governments have promoted growth rather than engaging in rent-seeking. Let a be the share of revenues from growth that go to the local government, so that it receives a total of aY in revenue. Likewise, let b denote private benefits to local government of suppressing growth, e.g. by protecting favored firms. They further denote by p_x the probability that local officials stay in power if they kill growth, and by p_y the probability that they stay in power by fostering growth.

The ratio of the two is denoted as $p = p_y/p_x$, so that a high value of p implies a high p_y and/or low p_x. Note that if officials were elected, rather than appointed by the central government, then p would be less than 1, since p_x exceeds p_y.

Simply put, a local government decides in favor of growth if $p_y \cdot aY > p_x \cdot b$. This notation likewise helps us frame how China's institutions reinforce choosing in favor of growth. By appointing (rather than electing) officials, and rewarding them for fostering growth, we see that $p_y \to 1$ and $p_x \to 0$, so that $p \to \infty$.

Reforms of China's government in favor of decentralization would mean that the central government has less control over appointing local officials, so that p_y decreases and p_x increases, thus creating more incentive for rent-seeking, and causing losses in efficiency. By means of this simple model, then, we can see how any benefits of economic decentralization are reliant upon political centralization.

For our purposes, the most important local political entity is the *county* (*xiàn* 县). The English translation may be somewhat misleading. The average city—which in China are very large—contains 8.6 counties. The average county has an area of 3,000km^2 and population of 450,000, though the variance for both is large; in general, counties in well-developed eastern China are smaller (average 1000km^2), while counties in the west are quite large (Cheung, 2014: 19). Throughout China, there are 2,860 counties in all.[2]

Recall how Xu (2011: 1078) characterizes China's system as 'regionally decentralized authoritarianism', marked by "political centralization and economic regional decentralization." A very important feature is that the modular nature of counties allows regional experimentation with policies, without disrupting the rest of the economy (2011: 1079). This greatly reduces political risks of carrying out reforms, and allows officials to take advantage of local information. As well as explicitly taking place by launching a new policy within a given county, this can also take place implicitly through differential enforcement of nationwide policies imposed by the central government (Cheung, 2014: 31).

[2] Jin & Zhang (2015: 93) add a helpful note about Chinese county names: "Names of high urbanized areas usually contain the Chinese character for 'city (市 *shi*)' or 'district (区 *qu*)' which are major components of urban areas, and those for low urbanized areas have names that contain the character for 'county (县 *xian*)' or 'banner (旗 *qi*)'."

This freedom to experiment with reforms effectively makes local officials into entrepreneurs (Xu, 2011: 1109). In some instances, officials have taken great personal risk to implement policies that went against central government orders, but were then promoted when the reforms succeeded, and the reform consequently implemented all over China. Key examples are land reform policies, when the central government "did not allow any ownership change to collective farming", and special economic zones, both of which became "endorsed by the central government as national policies" following their initial success (2011: 1098).

Another noteworthy feature of China's system is that regional leaders are frequently rotated to other areas, which occurred in 80% of provincial regions between 1978 and 2005 (2011: 1087). This practice dates back to the Chinese empire, "to prevent regional officials from cultivating strong political power bases within their jurisdictions" (2011: 1094, fn. 28). Likewise, such policies in the present day prevent collusion among officials to engage in rent-seeking and other anti-competitive behaviour (2011: 1099). Rotation is often combined with promotion, such as "promot[ing] mayors of successful municipalities to be governors of other provinces" (2011: 1094), further deepening incentives for officials to work hard and experiment with reforms.

From a mechanism design perspective, 'tournament competition' performs better than other systems when facing unobservable efforts, provided that agents' tasks are similar, and "outside random factors faced by the agents should follow the same distribution" (2011: 1100); otherwise, poor performance can simply be blamed on bad luck.

Another condition, becoming more problematic in recent years, is that such a system only works when all goals fit under one measurable heading (2011: 1129). If economic growth is the only metric that is rewarded, then other priorities will tend to be swept under the rug. One such priority has been the environment, where China's poor performance in this regard is well-known (2011: 1139).

Another neglected metric which is now becoming more important is inequality. China's Gini coefficient is estimated to exceed 0.50, making China (along with others such as South America) one of the most unequal counties in the world (2011: 1135). Among county governments more specifically, in 2003 the richest counties spent 48 times as much as the poorest (2011: 1138). Abetting this trend is 'interregional blockades', i.e. local protectionism, which is legally prohibited by the central government (e.g. banning differential tax rates based on product origin). In most cases this is difficult to prove, since differential pricing may well be based on objective matters such as transportation costs, hence it remains rampant (2011: 1134-5).

Transitioning from single-minded pursuit of growth to a multicriterial system prioritizing metrics such as the environment and inequality, is perhaps the key issue in reforming China's current system. Ways to fix this ‘multi-task problem’ include apportioning narrow, well-defined tasks (especially those with spillover effects) to “specialized ministries, special courts, and specialized regulatory bodies”; delegating monitoring and law enforcement (including regulation) to independent bodies, separate from local governments; and letting firms, rather than local governments, carry out market-based activities (2011: 1139-40).

The problem, of course, is that with multiple metrics, ‘competition’ takes on a new meaning, and Xu (2011: 1141) ventures the bold claim that “if there was no fierce regional competition, corruption would lead to the collapse of the Chinese government and Chinese economy.” The question is therefore how corruption arises, and what institutional changes must take place so that future reforms will not fall prey to it—to which we turn next.

An Institutionalist View

The economist Steven N.S. Cheung (张五常) offers a theoretical account of China's economy based on Ronald Coase's theory of contracts and transaction costs, combined with his own theory of ‘sharecropping’ as a form of rent. While hardly the final word on the subject, Cheung's view provides enough nuance and insight that it is worth dwelling upon in some detail.

Cheung observed circa 1978 that Chinese society was characterized by meticulous and minute gradations of rank that conferred privileges even in mundane matters of life—“[a] certain rank would entitle a comrade to share an automobile, or to an egg every other day, or to the right to buy fish without having to stand in line” (2014: 9-10). He traced the reason for this to China's absence of well-defined and legally-enforced property rights. In the absence of legal rights, socially-reinforced hierarchical rankings filled the void as “contractual restraints essential to reduce rent dissipation under competition” (2014: 10).

For Cheung, the fundamental issue in China’s transition from communism to capitalism was substituting property rights for its previous normative system of comrade-ranking (2014: 13). Particularly of interest to Cheung is the intermediary institutions that arose ‘in between’ these two systems. Corruption, for instance, can be viewed as a way for those previously in a privileged position to transfer their social capital (rights) to economic capital, thereby reducing their resistance to reform (2014: 14).

Thus, corruption is neither idiosyncratic moral aberration, nor a political lubricant worth tolerating in the long term, but instead a way to ‘buy out’ vested interests so they can then drop out of the system, as opposed to perpetuating rent-seeking behaviour. A further important point is that in China's non-democratic political system, high-ranking officials do not engage in corruption because “[p]eople on top want to maintain their hold on power, and corruption is one thing that will most likely destroy this” (2005: 742).

A further intermediate institution is what Cheung refers to as the ‘responsibility contract’ (承包责任合约). A responsibility contract, put as simply as possible, grants usage rights (e.g. of land) in exchange for meeting performance targets (2014: 30). The state retains ownership rights, and thus can revoke usage rights if the targets are not met (2005: 79). These originated in agriculture, but in the late 1980s were extended to industry, and gradually came to characterize China’s entire economic system (2014: 17).

For Cheung, the various layers of Chinese society (country, provinces, cities, counties, towns, villages, households) are “linked vertically by responsibility contracts, but horizontally there are no contractual linkages” (2014: 18). Thus there is horizontal competition among entities in a layer, each with similar responsibilities, but no vertical competition among layers (ibid.).

It is easy to understand in the abstract how strongly counties must compete, but Cheung provides several anecdotes illustrating to what lengths county representatives are willing to go. “A xian with a mere 300,000 in population would often employ 500 investment solicitors” (2014: 24). Business-inviting conferences are held where officials from counties all over China congregate, often attending business dinners several times in one evening (ibid.). One county in Anhui literally held a beauty contest to find women to head such meetings (ibid.). Furthermore, officials will send researchers to other successful counties, for guidance on making investments (2014: 26).

Counties are willing to accept very low cuts, to the point where land prices for incoming firms are often negative—that is, counties will build infrastructure in order to support the firm and may even provide them land free of charge (2014: 23). County officials will perform legal legwork on behalf of firms, such as obtaining a business license or building permit, and employ “worker-recruiting teams, which do the hiring for clients” (2014: 24). The best deals are offered to firms who bring a good image and reputation to the county, and are compatible with its other economic activity (2014: 26).

The immediate incentives this competition include a 17% value-added tax, as well as commission money based on investments in local banks (2014: 26). Cheung (2014: 21) makes much of the fact that this tax can be interpreted as a rent, and thereby accords with Cheung's theory of sharecropping, i.e. when a tenant can use land in exchange for giving the landowner a share of the crops. Cheung views the Chinese system as a form of sharecropping, where the 17% VAT is the 'share' given by counties to the 'landowner' (central government).

He likewise views the relation between counties and industries as a sharecropping system, with the non-classical characteristic of negative land prices (2014: 23). Part of why Cheung stresses the role of rent over tax is that in economics, the two involve widely separate normative interpretations: "tax maximization is routinely criticized, while rent maximization is often endorsed" (2014: 22). Furthermore, from this point of view, the fact that counties have the right to allocate the use of land makes them China's "chief economic power" (2014: 18).

Cheung credits inter-county competition for the concentration of industries within particular locations, where for example (Xu, 2011: 1119, fn. 53):

> Datang township makes one-third of the world's socks; 40 percent of the world's neckties are made in Shengzhou township; more than 70 percent of the buttons for clothes made in China come from Qiaotou township; Songxia township produces 350 million umbrellas every year; and Puyuan township produced 60 percent of China's cashmere sweaters, of which China is the world's largest producer…

Organizing industry in this way greatly cuts down on transaction costs such as transporting supplies, and this efficient division of labor eliminates redundancies such as reduplicating the same facilities in different areas. Cheung likewise claims that competition among counties promoted faster privatization of state-owned enterprises than would otherwise have taken place (2014: 33), since efficiency and growth are prioritized above subsidizing failing firms for the sake of vested interests.

However, Cheung notes an institutional change disrupting the incentive structure described above by Li & Zhou (2005). Since managing a county is essentially managing a business, competent public officials now have job offers from private firms (2014: 28), instead of their sole career option being to advance within the bureaucratic hierarchy.

Cheung makes a convincing case that “intensity of competition among xians is...the chief reason why China was able to sustain rapid growth” (2014: 26)—i.e., the answer to the ‘China puzzle’ is precisely inter-county competition. Yet, while helpful for making sense of anecdotes concerning local governments, his concepts fail to cohere at a higher level. It is unclear how to reconcile county–industry relations as sharecropping, versus as responsibility contracts—which is more akin to a call-option (right to buy) on the county’s part. Put glibly, Cheung’s theory is optimized for writing newspaper articles, not textbooks. Cheung’s institutionalist view helps to frame new questions, but does not itself answer them.

Chapter 2 – Current Problems with Chinese Counties

Counties and Fiscal Policy

In 2009-10, in response to the 2008 financial crisis, China carried out a fiscal stimulus of 4 trillion yuan, or 11% of GDP (Bai, Hsieh & Song, 2016: 129). In China, local governments are legally prohibited from running deficits, but as part of the stimulus they were allowed to create off-balance-sheet companies called local financing vehicles (LFVs). Local governments could then transfer assets (e.g. land) to an LFV, and the LFV could use the asset as collateral for a bank loan (2016: 130).

As is well-known, due to China's poor institutions, private firms can only exist through preferential treatment by public officials (2016: 134). Due to severe budget constraints imposed by the central government, officials could not offer direct financial support to favored firms (2016: 134); nor could they "influence the lending decisions of state-owned banks" on firms' behalf (2016: 130). Instead, they could only favor firms by exempting them from rules (2016: 134).

All this changed once local governments gained the freedom to use local financing vehicles. Out of 4 trillion yuan of stimulus, only 1 trillion shows up on local governments' balance sheets, so that the other three-quarters must have occurred off-balance-sheet—namely, by LFVs (2016: 141).

Thus, Bai, Hsieh & Song examine financial information for LFVs, and find that 60% of this off-balance-sheet stimulus was spent on infrastructure such as municipal construction and transportation (2016: 149). Furthermore, following the end of the stimulus program in 2010, local governments' off-balance-sheet spending via LFVs did not return to its previous level, but instead, "LFVs' total spending in 2014 and 2015 was more than three times larger than the amount spent on the fiscal stimulus in 2009 and 2010" (2016: 148).

They conclude by suggesting that China's recent drop in economic growth has been due to capital misallocation as a result of the stimulus policy. One possible mechanism is simply that local governments' increased spending is crowding out spending by more efficient private firms (2016: 153). Another is that through LFVs, local governments are now able to provide more funding to favored firms (e.g. via bank loans), so that less funds are left over for more efficient, but less connected private firms (2016: 157-9).

Counties & Urbanization

Zou & Lü (2014) examine how counties' level of urbanization influences their economic growth. They posit that urbanization proceeds in stages in the form of a sigmoidal curve (fig. 2). The initial stage (0–19% urbanization) is very slow, reflecting the difficulty of developing a city's initial infrastructure. Once this is in place, however, an acceleration stage (19–50%) occurs. Once such facilities become saturated, there is still a high rate of increase, but in a stage of deceleration (50-81%). Finally, the terminal stage (81–100%) is once again very slow.

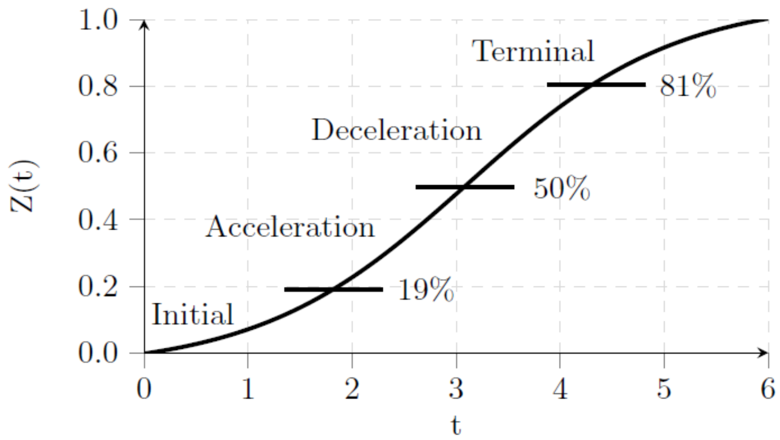

Fig. 2: Urbanization as logistic curve (Zou & Lü, 2014: 390)

Analyzing 44 counties in Liaoning, Zou & Lü (2014) note that county economic growth is stable, but the urbanization rate is not. There was a large spike in 2005 resulting from fiscal policies put forth following the 16th CPC National Congress. Once the policies ended, however, the newly-urbanized farmers returned to the countryside, so that the urbanization rate fell to its previous level (2014: 390). Thus, urbanization in Liaoning remains in the initial stage.

Jin & Zhang (2015) add more nuance to these findings by examining total factor productivity (TFP) in Chinese counties and cities from 2007–2010. They find that during this period, most of the 2,543 counties, districts and cities experienced TFP decline (2015: 90). Using two separate methodologies, they estimate this rate of decline as from 6.2% to 8.2% per year (2015: 87). Notably, this TFP decline occurred despite economic stimulus measures carried out in response to the 2008 financial crisis (2015: 95).

The few areas with positive TFP growth include well-established provinces such as Liaoning, Shandong and Guangdong, as well as some small, scattered areas in western China with low levels of development, whose growth they attribute to recent policies promoting “Large-scale Development of the Western Region” (2015: 93). While in general, highly urbanized cities act as “TFP growth engines” (2015: 93), there also exist areas such as Fujian and Guangxi where cities’ TFP growth is lower than that of counties (2015: 94).

The authors identify in particular the negative role of the *hukou* system on TFP. While intended as a “valve for flow control” (2015: 98), the *hukou* system instead promotes high turnover rates and flow of labor for migrant workers, leading to social problems that decrease economic efficiency. These problems would be mitigated by allowing migrant workers to settle down and become urban citizens. An important takeaway from this is that an increasing proportion of urban residents does not necessarily imply increased productivity (2015: 98). Conversely, improved TFP in cities cannot simply be attributed to faster urbanization rates, but must be sought out elsewhere (2015: 99).

Land Financing as Resource Curse

Another policy measure with unintended consequences has been relaxed restrictions on land sales by local governments. Since 1998, local governments are given exclusive rights to sell land, entitling them to land conveyance fees which, because they are classified as ‘extra-budgetary revenue’, they need not share with the central government (Chen & Kung, 2016: 87). In 2008, up to 80% of local government finances came from land revenues, versus 10% before 1998 (ibid.).

Mo (2018) frames this largely as a positive development, emphasizing how local governments compete for firm investment by setting low land rents, or use revenue from land sales as collateral for infrastructure projects. He finds that counties with a higher share of land conveyance revenue “can obtain a higher cost of debt, which induces more investment in infrastructures” (2018: 219). This is in part because counties with more revenue have better cash flow, and also because increased infrastructure investment increases the future value of land, strengthening the county's ability to pay debt with future land revenue (ibid.).

Mo notes that stimulus helped to decrease the cost of debt for local governments, which prior to stimulus was “much higher than the risk-free rate” (2018: 219); thus he uses difference-in-difference before and after the stimulus to identify how land revenue affects economic growth. He finds that on average, counties with high initial shares of land revenue grew faster by 1.6–3.5%, versus counties with no land revenue (ibid.).

Therefore, as well as showing how land financing supports the growth of infrastructure investment, Mo’s findings indicate that it is economically rational for banks to lend more to local financing vehicles (rather than a result of coercion) because “urban land is a valuable asset of good quality” (2018: 235).

Yet, Chen & Kung (2016) frame land financing as a form of political ‘resource curse’ that undermines China's performance target system. While historically, promotion is highly correlated to GDP growth, land revenue reduces this correlation, notably for officials with political connections on the prefectural level, or who have “passed the prime age of promotion” (Chen & Kung, 2016: 87).

The authors identify two channels by which land revenue substitutes for GDP growth in determining promotion. First, land revenues tend to be used on “ostentatious public projects”, strategically timed for signaling purposes (2016: 87). Further, using crackdowns on corruption as a form of natural experiment, they find evidence that land revenues promote corruption, in the form of collusion with prefectural officials, and overstaffing as a form of patronage (2016: 99-100).

Zhan (2013) elaborates further on the theme of ‘resource curse’ from a political angle, by comparing two relatively successful counties in Guizhou, one (Weng’an 瓮安) whose economy is based on its mineral resources, the other (Yuqing 余庆) with few resources and an agricultural economy (2013: 80-1).

Literature on resource curses identifies three mechanisms by which abundant resources can bring about social conflict. The first, and most well-known, is that “resource rents provide incentives for rebellion as well as start-up funds for rebel groups” (2013: 84). The second is that unequal distribution of wealth and negative externalities resulting from resources may cause grievances within the population. While the latter came into play for the case studied by Zhan, curiously, crime organizations chose to cooperate with local officials rather than seek independence (2013: 94-5).

The third mechanism leading to a ‘resource curse’ is the most relevant to China: that “resource dependence can lead to bad governance and weak political institutions based on patronage rather than on electoral competition, scrutiny and civil rights” (2013: 85). In brief, resource-rich local governments need not depend on their citizens for tax revenue, and can thus afford to be less receptive to popular preferences (2013: 89). Popular discontent with Weng’an's system manifested itself in mass protests triggered by a seemingly unrelated issue, whereas Yuqing managed to avoid social unrest in a similar scenario (2013: 78-9).

Left-behind Villages

In 1986, China established the State Council Leading Group of Poverty Alleviation and Development (国务院扶贫开发领导小组), which set a national poverty line of 206RMB (~50USD at the time) and set up funds to alleviate poverty. Most important for our purposes, they identified a subset of 'poverty counties' (贫困县) that even today serves as a benchmark for poverty alleviation projects.

The most recent and reliable sources indicate 832 poverty counties in all, calculated according to per capita net income (Li, Long, Tu & Wang, 2015: 16364). Naturally, the process of selection is very political: some non-poor counties try to be listed in order to access government funding, while other genuinely poor counties apply to be removed from the list for reasons of vanity. The most up-to-date list of 832 is also the largest, so we will refer to this hereafter.

Rogers (2014) documents in detail how the same incentives for local governments to promote growth, also encourage 'betting on the strong' in poverty-alleviation policies, which has led to numerous 'left-behind villages' (2014: 205). As the elasticity of poverty alleviation to economic growth decreases, new and "more individualized" forms of poverty have begun to emerge (2014: 204), from urban–rural inequality to inequality among rural areas (2014: 197).

Since the 1980s, poverty investment projects have included subsidized loans (mostly to industry), public works programs (infrastructure), and budgetary grants for agriculture, industry, education and health (2014: 198). As with other policies, poverty-alleviation operates through a target responsibility system, where officials are given growth targets from the central government, and evaluated through reporting and inspections (2014: 204). This creates an incentive for officials to strategically invest in order to "maximise payback" (2014: 203).

The problem is that maximizing growth need not imply allocating gains to the poorest—a classic principal–agent problem. In interviews with a poverty county in Shanxi province, one local government office noted its own selection criteria for projects included "convenient transport, good economic conditions and a centralised population" (2014: 203); for villages not satisfying these conditions, high transaction costs are unlikely to be outweighed by extra growth.

Thus, official statistics tend to underreport the income of richer villages (2014: 203), which have enough of an infrastructural base to take advantage of extra funds for "more visible outcomes, which can be readily measured" (ibid.). With appalling cynicism, one official interviewed by Rogers complained "it was becoming difficult to continue some initiatives as all of the 'rich' villages had already been chosen" (2014: 203).

Projects that do reach the poorest villages tend to be 'political achievement projects' (政绩工程) that look good for inspections and meeting targets, but bring "little tangible improvement to the livelihoods of farmers" (2014: 203). As an example, Rogers outlines a project to lay irrigation pipes in orchards to prevent water shortages. Yet, the project was fundamentally flawed because it drew water "from drought-sensitive water sources (shallow groundwater)," so that in droughts when water is most needed, the pipes do not work (ibid.).

On an ideological level, these deeply misaligned incentives are abetted by the common notion in China that "model villages will lead rural development" (2014: 204). To deal with the problem, Rogers recommends a heavier role for the central government in providing basic rural services (ibid.).

Yep (2008) corroborates this by examining the effects of past fiscal reforms on poverty counties. He notes how inter-jurisdictional spillovers pose another important problem: if public goods benefit residents outside the scope of performance criteria, this creates an incentive to underprovide such goods—though this can be alleviated by compensation from the central government (2008: 235).

This becomes increasingly problematic in light of counties' growing role in providing "basic services such as education and healthcare" (2008: 233). Cash-strapped governments in the poorest regions have been known, in order to maintain basic governance, to resort to borrowing from loan sharks, imposing excessive surcharges, and selling land (2008: 236).

More importantly for our purposes, budgetary grants are frequently used by local governments to finance local priorities instead of basic services (2008: 239). Although throughout China there was a general trend of declining investments in education and health from 1993 to 2002 (2008: 244-5), the data show that even though low-income counties received more budgetary grants, their drop in public expenditures exceeded that of higher-income counties (2008: 250). In short, the "parochial concerns" of local governments are leading to "[d]iversion, withholding or misappropriation" of funds, thereby compromising the central government's redistributive capacity (2008: 251).

Brehm (2013) tackles this problem head-on by analyzing the link between fiscal decentralization and economic efficiency. Despite theoretical predictions of a relation between the two, evidence shows at best an indirect influence, where decentralization influences growth by promoting macroeconomic stability (2013: 92). This lack of direct relation can be explained in part by “asymmetric distribution of incentives...at the sub-provincial level” (ibid.).

More concretely, various functions relegated to county-level governments lead to problems due to myopic incentives. A key example is education expenditure, since investment in human capital leads to spillover effects whose returns county governments cannot internalize (2013: 100); further, quality of education varies greatly among counties as rising salaries for teachers are offset by decreased spending on educational material (2013: 95).

A curious feature of the Chinese system is that although salaries for civil servants are determined by the central government, these must be paid by the counties themselves, so that public sector salaries are often county governments’ largest budget item, and indeed in some counties consume all local revenue (2013: 96). This problem is compounded by arrangements where “committees use appointment as a means to build up patronage” (ibid.), leading to overstaffing. A Chinese government study showed that 85% of redundant public sector positions were at the county level, “even though the county level accounts only for 60% of civil servants” (2013: 96-7).

Infrastructure, as well, can “not only stimulate economic activity, but also reduce transaction costs” (2013: 96), thereby alleviating the barriers discussed above to investment in the poorest counties. However, even given a mandate for infrastructure investment, establishing industrial zones, and thereby attracting new investments, leads to far more favorable evaluations for officials.

Closely linked to such problems is fuzzy budgetary categories, where investing in school buildings “is not categorized as education expenditure but as infrastructure investment,” and under the heading of ‘administrative expenditures’ is grouped both (productive) public goods provided to citizens and (unproductive) salaries paid to bureaucrats (ibid.).

Examining a sample of counties in Zhejiang province, Brehm (2013: 100-1) concludes that “counties featuring a high level of expenditure decentralization and a low level of revenue decentralization operate closest to the technological frontier.” Regarding the latter, it is difficult for central government measures to directly affect livelihood in counties; for instance, value-added tax affects economically advanced counties with more steps in the supply chain, but has little effect on rural regions based largely upon agriculture (2013: 94).

Instead, Brehm (2013: 93) raises the example of Zhejiang's 'province managing county' model, which omits prefectures' role in fiscal policy, and thereby allows policies that promote redistributive policies with spillover effects for neighbouring counties, as opposed to "short-term measurable achievements" (2013: 99) made for careerist purposes.

Chapter 3 – Data & Methodology

Methodology

An idealized regression equation for inter-county competition may look as follows:

$$\mathrm{pro\_business\_policy} = \beta_0 + \beta_1 \cdot \mathrm{county\_competition} + \alpha \mathbf{X} + \varepsilon. \qquad (*)$$

These variables are hard to measure, so the topic is little researched. However, we can use the effective tax rate as an index for pro_business_policy, and county density as a proxy for county_competition. This gives us the regression:

$$\mathrm{tax\_rate} = \beta_0 + \beta_1 \cdot \mathrm{county\_density} + \alpha \mathbf{X} + \varepsilon. \qquad (1)$$

A natural instrumental variable is geography, which is clearly exogenous (policy can't influence it). More importantly, China's counties have been more or less constant since Qin Shi Huang, and these ancestors' concerns in setting counties primarily revolved around favorable terrain and agricultural productivity.

Thus, in two-stage least squares (TSLS) we start with the following regression, where geography stands for geographic variation (variance of land height) and agriculture is agricultural productivity:

$$\widehat{\mathrm{county\_density}} = \beta_0 + \beta_1 \cdot \mathrm{geography} + \beta_2 \cdot \mathrm{agriculture} + \gamma \mathbf{X} + \varepsilon \qquad (2)$$

The estimates for county_density are then used as inputs for regression (1). Importantly, TSLS requires the same control variables in both regressions, so both (1) and (2) contain a vector **X** of controls for contemporary economic development, such as population density and real GDP for each county.

Since we control for economic development, the estimates for county_density should only reflect the effects of geography, thus providing a strong instrumental variable. We use TSLS because there may be historical factors affecting both tax_rate and county_density—for example, an area may have been richer from the start, affecting its economic development over history. If the TSLS estimates are similar to OLS, we can rule out any endogeneity issues, but without confirming via TSLS, our regression (1) could suffer from omitted variable bias.

* * *

Our sample includes 2,780 counties. The main data bottleneck comes from our firm-level data, while our other statistics cover every county in China. Our instrumental variables analysis uses geographical data extracted using the software ArcGIS. Geographical information systems have become increasingly common in economics, e.g. using nighttime light intensity as a proxy for GDP in developing countries with unreliable statistics (Henderson, Storeygard & Weil, 2012). Various open-source databases are available, with information such as land elevation and river proximity. Below we will give a basic description of each variable, along with summary statistics.

Tax Enforcement

County-level data for tax enforcement is available from the China Industry Business Performance Database (中国工业企业数据库) for 2004.[3]

For each of the 271,525 firms in the sample we divide taxes by sales income to get that firm's effective tax rate. Since a *de jure* tax rate of 25% is exogenously imposed from without by the central government, and is the same everywhere, the effective tax rate represents the proportion of taxes that are enforced by county governments.

The main variables relevant to tax enforcement include minor taxes (税金), corporate income tax (应交所得税), value-added tax (应交增值税), and sales (工业销售当). Thus, a simple way to calculate the effective tax rate is to add up these various taxes, as a proportion of firm income:

$$\text{tax\_rate} = \frac{\text{minor\_tax} + \text{corporate\_tax} + \text{value\_added\_tax}}{\text{sales}} \times 100$$

This represents the effective tax rate, i.e. tax proportional to sales. Yet, anomalies such as near-zero sales can lead to outlier tax rates as high as the comically-large 196,473,000,000%, and as low as −110,600,000%.[4]

Thus we drop the top and bottom 0.5% from our sample, leaving 268,809 firms, still representing 2780 counties. Among these, the average tax rate is 4.4%, with a standard deviation of 4.3%. The new minimum tax rate is −4.5, and the new maximum is 37.8%.

It may also be informative to include two other variables that act as further kinds

[3] 中国国家统计局. (2004). 中国工业企业数据库. [数据集]. 北京：中国国家统计局.

[4] There are 3,551 firms who report 0 income, which we set to 0.001 to prevent errors.

of tax: fees for government management (管理费用) and government subsidies (补贴收入), the latter being a 'negative tax'. As well, our formula above uses sales as a proxy for firm size, but another proxy might be the number of employees (全部从业人). These will be examined later as robustness checks.

County Density

The Center for Spatial Sciences at UC Davis hosts a free database of global administrative areas called GADM, with data for: 0) countries, 1) provinces [省], 2) prefectures [地区], and 3) counties [县].[5]

Each county is classified using the ISO 3166-1 alpha-3 country code; for example, Shanghai's Pudong district (浦东新区) is represented as CHN.24.1.8, where each part stands for a different administrative level. This dataset is composed of polygons which can be overlaid onto point datasets to provide summary statistics for 2,408 counties.

A troublesome problem is that our firm-level data is classified using the ISO 3166-2:CN code—where Pudong is classified as 310115, with two digits for each administrative level. ISO 3166-2 gives more detail, so that multiple districts (区) in a city are often represented by a single ISO 3166-1 code. Therefore, for our purposes, a county map based on ISO 3166-2 is more suitable (fig. 3).

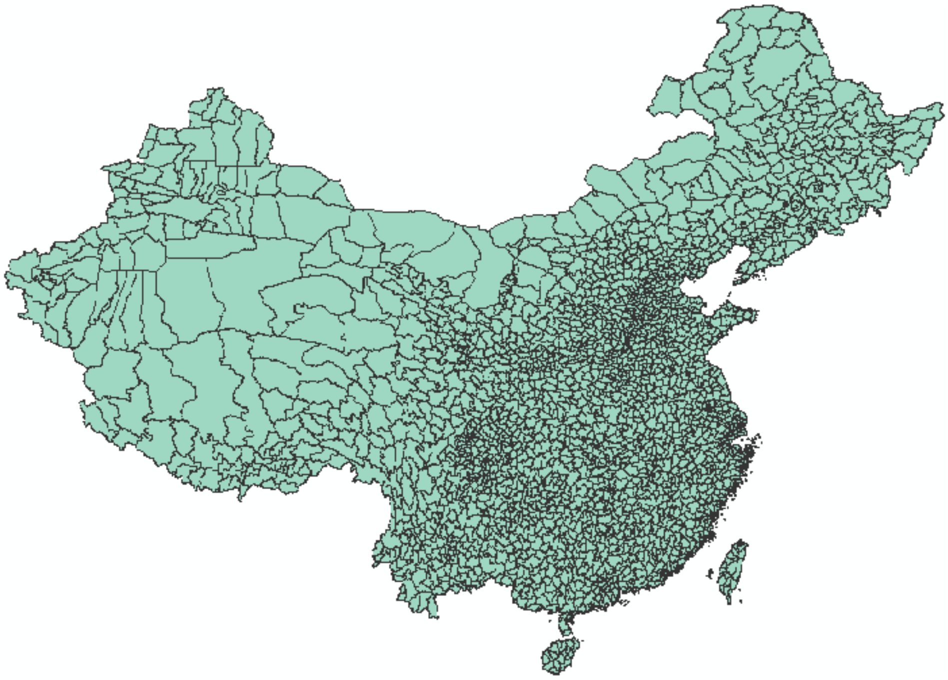

Fig. 3: 2,782 counties in China

[5] University of California, Davis, Center for Spatial Sciences. (2018). Database of Global Administrative Areas (GADM), v. 3.6. [Data set]. Retrieved from https://gadm.org/download_country_v3.html

To get figures for density, we first calculate the centroid of each county, i.e. the arithmetic mean position of all its points. Then, using each county's centroid, its density is the number of other counties whose centroid is within 100km. Well-known formulas exist for finding distance on a sphere given two points' latitudes and longitudes, notably the haversine formula from navigation. A more accurate formula, taking into account the Earth's ellipsoidal shape, gives different results than haversine for 126 counties (half +1, half −1), thus we use the latter instead.[6]

To ease computation, we use adjacency data that for each county records all ‘neighbours’ sharing a border. This lets us define all counties within a given ‘depth’, where immediate neighbours have depth 1, neighbours of neighbours have depth 2, and so on. For measuring counties within 100km, a depth of 3 is optimal.[7]

Geographic Variation

Harvard's Center for Geographic Analysis hosts a free database called the China Historical Geographic Information System (CHGIS), providing a spatial map of China's historical administrative units from 222BC to 1911AD.[8]

Most important for our purposes is its digital elevation model, which records land elevation such as mountains and valleys (fig. 4). Since such topographical features have not noticeably altered since 1911, it is safe to use this for our more up-to-date analysis. For all points in a given county, we can summarize the roughness of the terrain using the variance of elevation, where low variance signifies smooth terrain, while high variance means a mix of high and low terrain.

---

[6] Here I use the distGeo and distHaversine functions from the ‘geosphere’ R package. See Hijmans, R., Williams, E., & Vennes, C. (2017). “Package ‘geosphere’.” Retrieved from https://cran.r-project.org/web/packages/geosphere/geosphere.pdf

[7] Neighbour depth of 4 gives the same results as a depth of 3.

[8] Harvard University, Center for Geographic Analysis. (2016). China Historical GIS (CHGIS), v.6. [Data set]. Retrieved from https://sites.fas.harvard.edu/~chgis

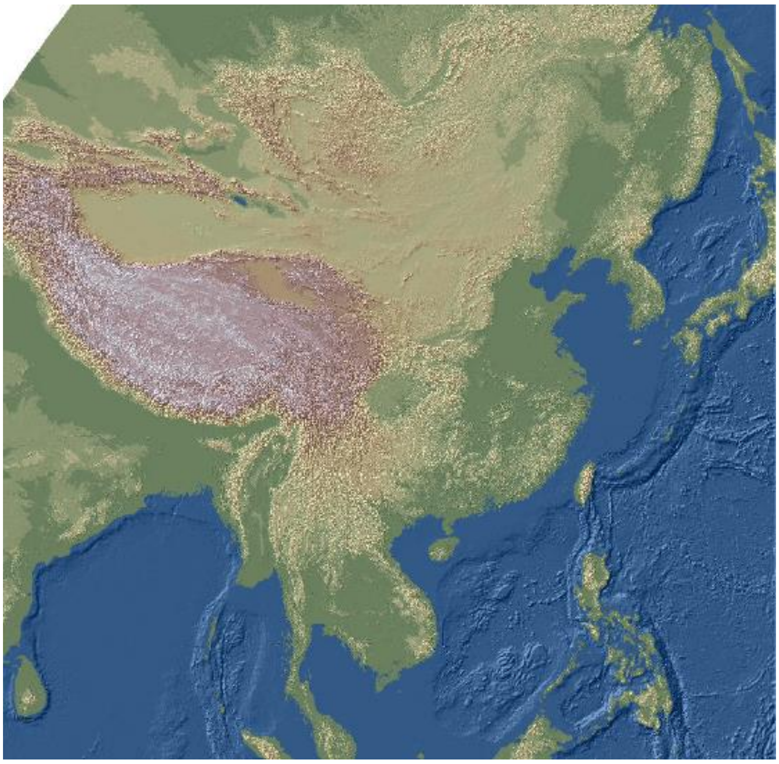

Fig. 4: Digital elevation model of China

Throughout all of China, the average geographic variation per county is 192.4. The county with the highest variation is Qira [策勒县], part of Khotan prefecture [和田地区] in Xinjiang province, with 1802.5. The county with the lowest variation is Yandu district [盐都区], part of Yancheng city [盐城市] in Jiangsu province, with 0.6. Finally, the standard deviation of geographic variation is 204.

Agricultural Productivity

The Food and Agriculture Organization hosts a dataset called Global Agro-Ecological Zones (GAEZ), providing spatial data on suitability of soil for different crops.[9]

GAEZ is a global map that estimates potential crop yields for 11 cereal grains and 4 roots and tubers. These estimates are augmented with counterfactual categories concerning water supply (rain-fed vs. irrigated) and level of inputs by farmers (high, medium, low). To avoid problems of reverse causality, it is safest to consider rain-fed low-input agriculture.

9 Food and Agricultural Organization of the United Nations. (2012). Global Agro-Ecological Zones (GAEZ), v. 3.0.1. [Data set]. Retrieved from http://gaez.fao.org/Main.html. Free sign-up is required to download the data.

Mayshar, Moav, Neeman & Pascali (2018) summarize the GAEZ dataset into a general index of agricultural productivity. They use the USDA National Nutrient Database to convert crop yields into calories, then for each unit of area select the crop with the highest potential caloric yield (2012: 23), then assemble these into a global map. They find that "[c]ereal grains are the highest-yielding crops in approximately 99 percent of the raster points in the sample" (2018: 23).

Since all counties in China have the same optimal crop (2018: 96, fig. E.4), converting the GAEZ index to calories only amounts to a unit change. Thus we can simply define agricultural productivity using the GAEZ index for cereals. For each county, its value of agriculture is the mean of all points in that county.

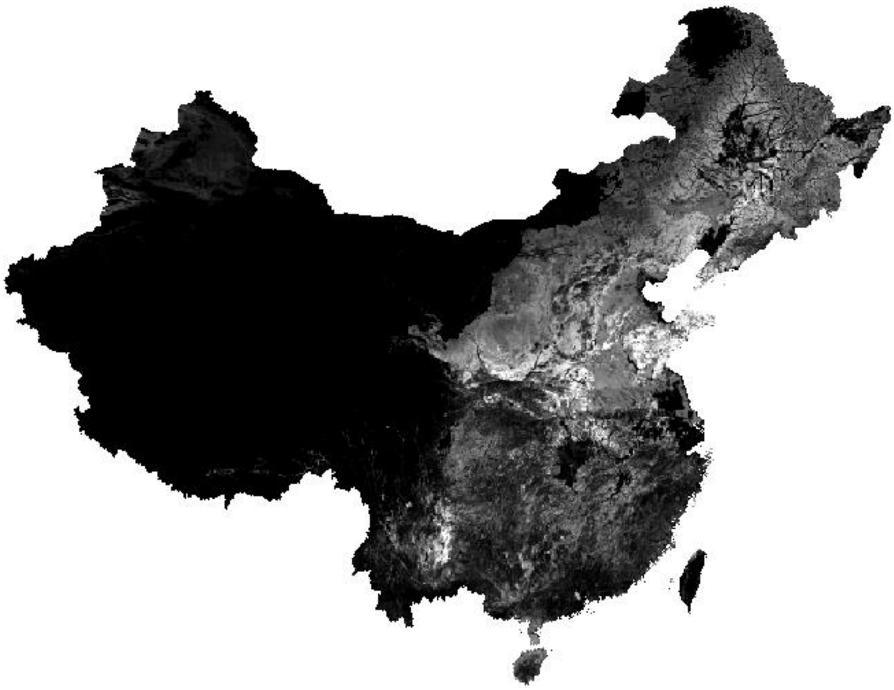

Fig. 5: Agricultural productivity in China, from high (white) to low (black)

Total production capacity for rain-fed low-input cereals throughout China is shown in fig. 5. Low productivity (value 0) is shown as black, while high productivity is lighter. The highest value in the index is 630, but in China the highest value is 369—for Dalian (大连市) in Liaoning province.

Rivers

Rivers and other water basins are a staple of geographic analysis, so ArcGIS includes datasets for rivers in China. One dataset entitled china_hydro covers smaller rivers (thin lines), while another dataset entitled china_rivers (part of a larger WorldRivers distribution) includes the more famous major rivers such as the Yangtze and Yellow River.

Fig. 6: Rivers in China

Our rivers dataset sums the total lengths of rivers in each county, and divides them by the total area of that county. Taking the logarithm of this gives our variable for river density. Because some of these logged percentages are very low, several counties have negative river densities, with the minimum being −6.14, in Inner Mongolia's Sonid Left Banner (苏尼特左旗). The county with the highest river density is Huangpu district (黄浦区) in Shanghai, with 3.9. Throughout China, average river density is 1.29, with a standard deviation of 0.85.

Control Variable: Nightlights

In response to the problem of unreliable statistics in developing counties, it has become common in the development literature to use night lights as a proxy for wealth (Henderson, Storeygard & Weil, 2012). Furthermore, Chinese statistics in particular are known for often being manipulated. For our purposes, night lights can be thought of as a composite index of GDP and population density.

Our data for nightlights comes from NASA's satellite VIIRS (Visible Infrared

Imaging Radiometer Suite), which orbits the Earth collecting visible and infrared imagery of land, clouds, and oceans.[10]

The data, in the form of a high-resolution TIFF file (excerpted in fig. 7), was collected from April to October 2012, and has been corrected for phenomena such as gas flares and aurora borealis.

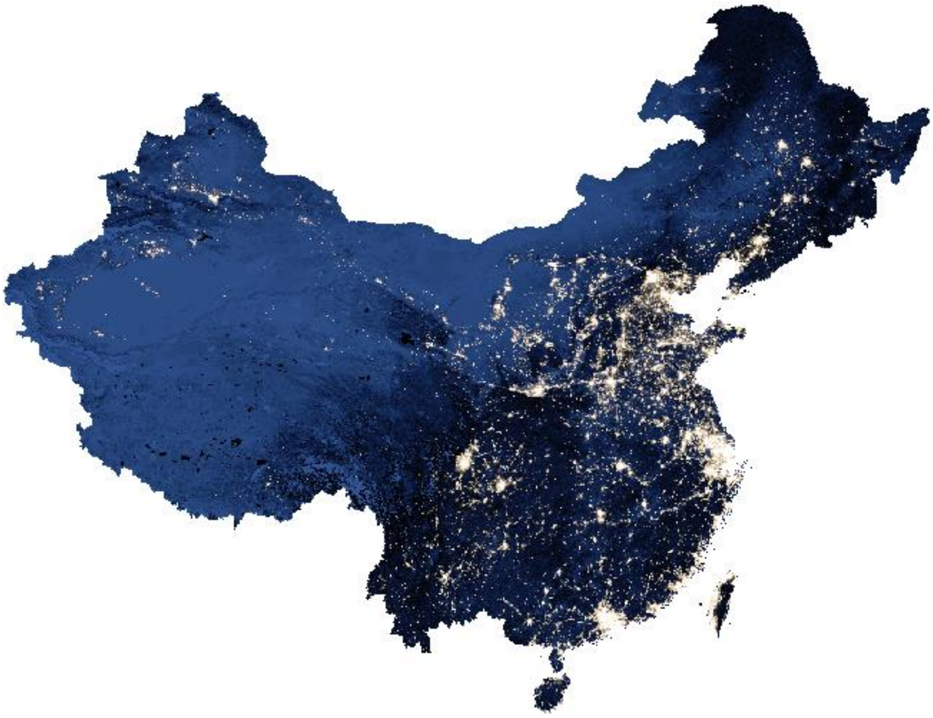

Fig. 7: Nightlights in China

Each pixel has a value ranging from 0 (darkest) to 255 (lightest). Throughout all of China, the average level of night lights per county is 46.4, with a standard deviation of 58. The brightest county is Shanghai city (上海市区), at 255. The darkest county is Huzhong county [呼玛县], part of Daxing'anlin prefecture [大兴安岭地区] in Heilongjiang province, at 2.67.

We include this control variable in both stages of our TSLS regression so as to rule out any endogenous factors that may have existed during the formation of Chinese counties, and persisted up to the present day as reflected in GDP. Thus it is best not to interpret its coefficients too literally, especially in TSLS.

Since a subset of counties in our sample have GDP data, as a robustness check we can compare the results of nightlights versus GDP for this subset.

[10] Suomi National Polar-orbiting Partnership. (2012). Visible Infrared Imaging Radiometer Suite. [Data set]. Retrieved from https://www.earthobservatory.nasa.gov/images/79765/night-lights-2012-flat-map

Chapter 4 – Results

Basic Hypotheses

We expect a negative coefficient for density, reflecting how higher inter-county competition leads to lower tax enforcement. Still, it is unclear whether this is good or bad for economic development. On the one hand, we can imagine multiplier effects of successful businesses spreading to the rest of the county. On the other, corporate tax is a crucial means for poor counties to collect government funds, which can be used to build infrastructure and other social goods. The magnitude of the coefficient may provide some insight on this front, thus stimulating further research into microfoundations of inter-county competition.

Of auxiliary interest, the coefficients on the first-stage regression (2) give some insight into the geo-economics of Chinese counties. We expect to see a high value for agriculture, since high agricultural productivity leads to more counties. However the sign for geography is unclear *a priori*. Higher variance of terrain makes an area harder to govern, and so may lead to more counties (positive coefficient), or more variance may incline the central government to lump these areas together, leading to fewer counties (negative coefficient).

The latter issue is reflected in an ongoing debate among Chinese geographers, who have two competing theories of how terrain influences county density.

The first theory is 山川形便 – that counties wish to maximize the advantages offered by terrain. This theory predicts that counties with favorable geographic conditions (high agricultural productivity, low variance of terrain) will be relatively larger, and thus that county density in a unit of area will be lower.

The second theory is 犬牙相错 – that the central government will not let local governments have enough geographic advantages that they can become independent. This theory predicts that areas with favorable geographic conditions will have higher county density, since the emperor directly limited their size.

Our regression (2) gives the sign and magnitude for terrain variance's effect on county density *irrespective of wealth effects*, thereby contributing to the debate.

Regression Analysis

Given our specification of tax_rate, our coefficients have the intuitive meaning of adding β percent to the effective tax rate for each unit of X. Since counties have many firms, we use sample weighting in inverse proportion to the number of firms in each county.[11] Simple OLS regression of equation 1 yields the coefficients shown in table 1.

Table 1: OLS regression – Tax rate

	(1)	(2)	(3)
density	−0.04 (0.0045)		−0.037 (0.004)
lights		−0.005 (0.0005)	−0.002 (0.0005)
R^2	0.010	0.006	0.011

Note: All variables significant at 1%; errors clustered by county. ($N = 268,809$)

In line with our basic prediction, we see that higher county density implies less tax enforcement; likewise, higher GDP implies lower tax enforcement. Our regressors are all highly significant, though R^2 is only 1.1%. The magnitude is small but non-negligible—especially considering that the amount of neighbours ranges from 1 to 54.

Results of TSLS regression are shown in table (2), with the first stage in the top panel, second stage in the middle, and effective F-statistic at the bottom.[12]

Table 2: TSLS regression – Various specifications (2 variables)

	(1)	(2)	(3)	(4)	(5)	(6)
geography	−0.021 (0.001)	−0.017 (0.001)	−0.024 (0.001)			
agriculture	0.036 (0.003)			0.047 (0.003)	0.053 (0.003)	
lights		0.063 (0.003)		0.075 (0.003)		0.075 (0.003)
rivers			2.89 (0.26)		3.74 (0.30)	2.80 (0.25)
density	−0.10 (0.01)	−0.193 (0.019)	−0.12 (0.009)	−0.028 (0.013)	−0.046 (0.008)	−0.064 (0.019)
lights		0.010 (0.002)		−0.003 (0.001)		−0.002 (0.0016)
F-statistic	86.6	67.6	57.1	1.24	22.9	19.4

Note: Top is first-stage regression (dep. var: density), bottom is TSLS (dep. var: tax rate) Standard errors are clustered at the county level. Number of observations: 268,809.

[11] In Stata, we add [pweight=1/firms] to our regression; firms is a variable for the number of firms in each county.

[12] In Stata, this is implemented using the weakivtest command. In this test, as a general rule of thumb, an F-statistic less than 3 indicates a weak instrument. See Pflueger, C. & Wang, S. (2014). "A Robust Test for Weak Instruments in Stata." Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2323012

Table 3: TSLS regression – Various specifications (3+ variables)

	(1)	(2)	(3)	(4)	(5)
geography	−0.012 (0.001)	−0.019 (0.0013)	−0.016 (0.001)		−0.010 (0.0009)
agriculture	0.037 (0.003)	0.036 (0.003)		0.046 (0.003)	0.038 (0.003)
lights	0.064 (0.001)		0.07 (0.003)	0.07 (0.003)	0.06 (0.003)
rivers		2.96 (0.27)	2.33 (0.24)	2.68 (0.25)	2.40 (0.24)
density	−0.092 (0.013)	−0.090 (0.008)	−0.15 (0.015)	−0.038 (0.011)	−0.082 (0.011)
lights	0.002 (0.001)		0.007 (0.001)	−0.002 (0.001)	0.001 (0.001)
F-statistic	23.1	58.8	28.9	14.5	22.8

Note: Top is first-stage (dep. var: density), bottom is TSLS (dep. var: tax rate)
Standard errors clustered by county. Number of observations: 268,809.

In the first stage we see that, all else equal, areas with higher geographic variation tend to have lower density, areas with high agricultural productivity tend to be more dense, areas with high GDP (lights) tend to be more dense, and areas with more rivers tend to be more dense.

Recall the geographic theories from earlier about areas with favorable conditions (low geographic variance, high agricultural productivity). The 山川形便 theory predicts that favorable areas have fewer counties (low density), but 犬牙相错 predicts more counties. Throughout tables 2 and 3 we find that higher agricultural productivity and lower geographic variation both imply more counties, supporting 犬牙相错.

This likely reflects an upper limit to the number of people that can be effectively governed. If an area can grow more food, it can support more people; likewise, a flatter area can fit more people, so that it hits this boundary sooner and the county is smaller.

Controlling for wealth (2.1 vs. 3.1) dampens the effect of both geography and rivers on county density but increases the effect of agriculture, whereas controlling for rivers (2.1 vs. 3.2) has little effect on either geography or agriculture. It is worth noting here that single-variable regression shows that rivers are negatively correlated to both geography (which accords with common sense) and agriculture. The latter makes more sense once we recall how agricultural productivity means optimality for growing cereals (the crop with the highest potential yield in calories), as opposed to rice, where rivers come more in handy for irrigation.

As seen in the second stage, TSLS regression consistently shows a far stronger role of county density than OLS. Overidentification tests indicate that rivers is an unreliable instrument, so hereafter we omit it from our regressions.[13] Thus, the most reliable estimate would seem to be (3.1), or 0.092% less taxes per neighbour in TSLS, as opposed to 0.037% with OLS (1.3), or about 2.5 times larger. This implies some form of omitted variable bias artificially lowering the OLS results.

Suppose for instance that a certain county is richer to start with, for non-geographic reasons. The effects of this initial endowment could carry on over time, possibly creating spillover effects with neighbouring provinces, so that it is wealthier in the present day than counties with a smaller initial endowment. Since it (and possibly its neighbours) are already wealthy, they need not compete as much for business investment, and so need not lower their effective tax rate. By using TSLS and controlling for wealth, we are considering county density in its purely geographic aspect, omitting any such legacy effects from initial non-geographic endowments.

Robustness Checks

Adjacent Neighbours

A possible objection to calculating county density within a 100km radius is that only directly adjacent counties (sharing a border) matter for competition, while non-adjacent neighbours matter much less.[14] If two large counties share a border, for instance, their centroids may not be within 100km of each other, so using a 100km radius as a metric will tend to discount such cases. On the other hand, in an area very dense with counties, only using adjacent counties as a metric will discount many other counties who are close in absolute distance.

Recall that for neighbours in a 100km radius, density ranges from 1 to 54, with a mean of 20 and standard deviation of 11.7. For adjacent neighbours, density ranges from 1 to 20, with a mean of 5.75 (s_x = 1.85). Table 4 compares the results of our previous TSLS regressions to new TSLS regressions that define density by counting adjacent neighbours.

[13] In Stata this is implemented with the ivreg2 command. The relevant test is Hansen's J-statistic, where if p > 0.05 our instrument is valid. Omitting rivers (and pweight) gives p = 0.69, while alternative specifications fail the test.

[14] In our prior terminology, this means using a neighbour depth of 1, rather than 3 as in the standard regression.

Table 4: Tax rate – $100km^2$ vs. adjacent ($'$)

	OLS	OLS$'$	IV	IV$'$
density	−0.037 (0.004)	−0.036 (0.025)	−0.092 (0.013)	−0.27 (0.146)
lights	−0.002 (0.0005)	−0.006 (0.0005)	0.002 (0.001)	−0.004 (0.0005)

Note: IV$'$ does not use weights, which give poor results
Errors are clustered by county. ($N = 268,809$)

In all the regressions lights differs very little, except in IV where it switches signs. However, density is more erratic, going from roughly the same in OLS' than OLS, to much higher (though not significant) in IV' than IV. In both versions, TSLS is larger than OLS. Most important to note is how the ranges for both specifications differ: a 100km radius allows many more neighbours in a dense area (up to 54), while it is unlikely that one large county will exist in an area with many smaller counties contiguous to it, so that the upper bound for adjacent counties is 20.

Therefore, we would expect the coefficient on IV' to be higher, since its value for influence on competition is ‘spread out’ among fewer counties. While this is indeed the case, the variance is too high, making 100km neighbours the more reliable choice.

Table 5: Density – $100km^2$ vs. adjacent ($'$)

	density	density$'$
geography	−0.012 (0.001)	−0.0002 (0.0002)
agriculture	0.037 (0.003)	−0.0001 (0.0005)
lights	0.064 (0.001)	−0.076 (0.0005)
R^2	0.39	0.09

Note: Clustered by county. ($N = 268,809$)

This is further corroborated by table 5, which shows how our geographic variables relate to density under both specifications. For density with adjacent counties, both the explanatory variables are insignificant, and the R^2 is a poor fit of 9%, whereas for 100km-radius counties the R^2 is a strikingly high 39%. Since the regression in table 5 is the first stage of TSLS, this gives us more confidence in the IV result in table 4, so that overall, using a 100km radius seems to be the better metric.

Tax Specifications

To gauge the effective tax rate, we used tax (minor tax + corporate income tax + VAT) proportional to sales (income). Yet, several other variables can be seen as forms of tax, which should arguably be specified in the effective tax rate.

Table 6 reports coefficients and standard errors for IV regressions with each tax specification. Here, a management fee is a murky category (which can include bribes to government officials) that acts as a positive tax, and subsidy adds a 'negative' tax. Specifications using the number of employees (e.g. tax per employee) give nonsensical results (e.g. an effect of –16% per neighbour), so we omit these results from our table.

We should note from the outset that these variables can be quite noisy—management ranges from −250,264 to 5,219,374 with a mean of 3,524 and standard deviation of 34,622. Likewise, subsidy ranges from −4,096 to 700,685 with a mean of 211 and standard deviation of 4,650.

Thus it makes sense to trim outliers for each specification, cutting off values above the 99th percentile and below the 1st percentile. The danger here is that different firms will be cut off for each specification, possibly even excluding whole counties. Further, since some numbers (such as 0) occur often, this leads to some variation in the number of firms and counties covered by each regression.

Recall that for the full sample, the average tax rate was 4.4% (s_x = 5.3). For our trimmed regressions, the tax rate including management has a mean of 9.8% (s_x = 5.3), while the tax rate including subsidy has a mean of 4% (s_x = 3.8).

Results for these trimmed regressions are shown in table 6.

Table 6: Specifications for tax enforcement

	#1	#2	#3	#4
manage		✓		✓
subsidy			✓	✓
density	−0.092 (0.013)	−0.68 (0.05)	−0.072 (0.013)	−0.67 (0.05)
lights	0.002 (0.001)	0.04 (0.004)	0.001 (0.001)	0.04 (0.004)
firms	268,809	267,673	266,093	266,646
counties	2780	2780	2777	2780

Note: Standard errors are clustered at the county level

Adding management drastically increases the effect of density on tax enforcement, while adding subsidy dampens the effect but stays within the same order of magnitude. Importantly, management is highly sensitive to the quantiles being used: table 6 trims the top and bottom 1%; trimming the top and bottom 5% (not shown) gives a coefficient of –0.311 (s_x = 0.024) for density in regression #2; and trimming the top and bottom 10% gives a more realistic but still quite large coefficient of –0.185 (s_x = 0.015).

If subsidies were given as a further way of reducing the tax rate for incoming firms, we would expect to see a larger magnitude for density once we include subsidy; instead we see the opposite. Thus, the smaller magnitude would make sense if subsidies tend to be given to firms that already pay high taxes, rather than those with low taxes.

In a similar way, treating management as a proxy for bribes, the higher magnitude for density would make sense if the majority of bribes are paid by firms in uncompetitive areas. If a firm is forced to pay too many bribes by the local government, it can relocate to a nearby county with fewer bribes, but if there is less competition with nearby counties, local officials can adopt rent-seeking behaviour and firms will have no choice but to pay exorbitant fees.

GDP vs. Nightlights

GDP data is only available for 1,959 counties out of the 2,780 in our sample.[15] Because of this, on top of the well-known unreliability of statistics in China, we used nightlights data as a proxy. Still, it may be informative to re-run our regression using this limited sample (128,234 firms), and compare both results.

Recall that with the full sample the average density was 20 (σ = 11.7) and average brightness was 62.8 (σ = 74.15). Within the sample for which we have GDP data, mean density is 18.5 (s_x = 11.6), and mean brightness is 30 (s_x = 34.6). The mean of per capita GDP is 101,790 (s_x = 202,737), and after taking logarithms the mean is 10 with a standard deviation of 1.67.

We expect GDP to be upward-biased, so since the control variable's coefficient in regression 2 is positive, the GDP statistics will likely lessen this result.

[15] CEIC. Gross Domestic Product: per Capita: County Level Region. In China Premium Database [Online]. Available: Euromonitor Institutional Investor.

Table 7: Control variables – lights vs. GDP (′)

	OLS	OLS′	IV	IV′
density	−0.037 (0.006)	−0.047 (0.005)	−0.072 (0.014)	−0.077 (0.01)
lights	−0.006 (0.002)		0.001 (0.002)	
log(gdp)		−0.074 (0.032)		−0.078 (0.032)

Note: Standard errors are clustered at the county level.
Number of observations: 128,234 (1,959 counties).

The results of table 7 confirm both control variables' similarity: the coefficients for density are within two standard deviations of each other. Simple OLS shows the following linear relationship between the two controls:

$$\text{lights} = \underset{(1.29)}{-3} + \underset{(0.13)}{7.82} \cdot \log(\text{gdp}) \tag{3}$$

Regression 3 has an R^2 of only 2.6%, and in this light it is almost surprising that they give such similar results. In any case, lights here is only a control variable, so we should not take its magnitude too literally.

Table 8: Density – lights vs. GDP (′)

	density	density′
geography	−0.009 (0.001)	−0.015 (0.001)
agriculture	0.05 (0.003)	0.06 (0.004)
lights	0.15 (0.008)	
log(gdp)		−0.308 (0.11)
R^2	0.49	0.32

Note: All variables are significant at 1%
Clustered by county. ($N = 268,809$)

Table 8 compares the two control variables' effects on density, for the sample of 1,959 counties for which GDP data is available. Except for the control variables, the coefficients are about the same. Further, the R^2 using lights is larger than for log(gdp). Since this is the case, and since it is unclear what kind of sampling bias may influence availability of GDP data, it seems fair to prefer using night lights as a proxy for wealth.

Policy Implications: Poverty Counties

Ever since the term was first introduced, the criteria for inclusion as a 'poverty county' (贫困县) have been unclear, and therefore easy to game for political purposes. From a certain point of view, geographic variables may be thought of as unhelpful for economics because they cannot be changed for policy purposes. Yet, this lack of change becomes an asset in searching for objective differences between poverty and non-poverty counties. If anything can help us distinguish between true poverty counties, poverty counties that should not be labeled as such, and non-poverty counties that *should* be poverty counties, geography would seem to be a prime candidate.

Conveniently, the most up-to-date list of 832 is also the largest, so this is what we use below.[16] To analyze how poverty counties fit into inter-county competition, we use a dummy variable labeled '1' for poverty counties, '0' otherwise. Among the counties for which we have firm-level data, 791 are considered poverty counties. Of our 268,809 non-outlier firms, 13,106 (5%) are registered in one of these counties. A simple but informative exercise is to re-run our previous regressions while including this poverty dummy as a control variable.[17]

Recall that among all counties, average geographic variation is 170 (s_x = 198.6), average agricultural productivity is 112 (s_x = 78), average density is 20.2 (s_x = 11.7), average brightness is 63 (s_x = 74), and average river density is 1.2 (s_x = 0.6).

Among the sample of poverty counties, average geographic variation is 281.6 (s_x = 220.6), average agricultural productivity is 89.6 (s_x = 68), average density is 13.6 (s_x = 8.8), average brightness is 14.6 (s_x = 14.5), and average river density is 1.3 (s_x = 0.85).

Conversely, for non-poverty counties, average geographic variation is 125.9 (s_x = 170), average agricultural productivity is 120.7 (s_x = 80), average density is 22.8 (s_x = 11.7), average brightness is 82 (s_x = 79.4), and average river density is 1.3 (s_x = 0.94).

Comparing these, we see that poverty counties tend to have rougher terrain, less fertile land, and fewer neighbours compared to non-poverty counties.

[16] The list of poverty counties in China is quite confusing, with government documents and Chinese-language news outlets all citing different numbers, and English news outlets awash with vacuous articles about *n* counties being removed from the list. The next most reliable figure is 592, which appears in government documents circa 2012, and may help as a robustness check if it excludes counties that shouldn't be listed.

[17] Although several government websites cite the amount as 832, none contain a full list, so I use the following source: 谷枫 & 杨坪. (2016 年 12 月 2 日). "股转系统发布 832 家贫困县名单，云南最多 (附名单)." 21 世纪经济报道. Retrieved from http:// 21jingji.com/article/20161202/herald/a524550b4261dcef8de2c1c1705e03e4.html

Table 9: OLS with poverty county dummy

	(1)	(2)	(3)
density	−0.037 (0.005)	−0.033 (0.004)	−0.031 (0.004)
lights	−0.002 (0.0005)		−0.0009 (0.0005)
poverty		0.825 (0.12)	0.786 (0.13)
R^2	0.011	0.015	0.015

Note: Clustered by county. ($N = 268,809$)
Dependent variable: tax rate

Table 9 shows results for OLS with a dummy variable for poverty counties. Recall that for the entire sample the average tax rate is 4.4%, with a standard deviation of 4.3%. Firm data shows that poverty counties have an average tax rate of 5.8% (s_x = 5.5), while non-poverty counties have an average tax rate of 4.3% (s_x = 4.2), and OLS shows that this discrepancy holds even after controlling for wealth. In comparison, table 10 re-runs TSLS using this dummy variable as well as an interaction term.

Table 10: TSLS with poverty county dummy

	(1)	(2)	(3)	(4)
geography	−0.012 (0.001)	−0.011 (0.001)	−0.012 (0.001)	−0.010 (0.001)
agriculture	0.038 (0.003)	0.036 (0.003)	0.038 (0.003)	0.035 (0.003)
lights	0.064 (0.003)	0.059 (0.003)	0.063 (0.003)	0.058 (0.003)
poverty		−2.45 (0.38)		−3.74 (0.73)
pov*light			−0.02 (0.02)	0.08 (0.05)
density	−0.092 (0.013)	−0.077 (0.013)	−0.092 (0.013)	−0.079 (0.013)
lights	0.002 (0.001)	0.002 (0.001)	0.003 (0.001)	0.002 (0.001)
poverty		0.58 (0.13)		0.53 (0.17)
pov*light			0.016 (0.005)	0.003 (0.006)
F-statistic	23.1	21.7	23.0	21.4

Note: Top is first-stage (density), bottom is TSLS (tax rate)
Standard errors clustered by county. ($N = 268,809$)

The most straightforward regression to interpret is 10.2, where poverty counties tend to charge 0.58% more in taxes. It would seem that, having little ability to compete with wealthier neighbours, officials in poverty counties adopt a rent-seeking approach. This dissuades other businesses from locating there, perpetuating their poverty.

Also curious is the first stage of TSLS 10.2, where poverty counties tend to have 2.45 fewer neighbours within 100km than other counties. Removing lights from this regression (not shown) raises this to 5.5 fewer ($s_x = 0.4$).

A clearer picture of poverty counties' geographical endowment is given by single-variable regressions with geography and agriculture as regressands and poverty as regressor (Table 11). As we expect, geography gives a positive coefficient, meaning that poverty counties tend to have rougher terrain. To our surprise, so does agriculture, suggesting that (at least some) poverty counties tend to have higher agricultural productivity. Adding geography into this regression (not shown) does not change this effect, but controlling for lights flips the sign to negative, which is what we expected in the first place.

Table 11: Single-variable OLS with poverty dummy as regressor

	geography	agriculture	density	lights	rivers
poverty	142.5 (1.45)	18.3 (0.57)	−11.6 (0.08)	−123 (0.22)	−0.73 (0.22)
R^2	0.076	0.002	0.05	0.09	0.0002

Note: All variables are significant at the 1% level

The 'obvious' picture of poverty counties is simply as counties that are worse off in everything—more rugged terrain, less productive soil. Yet, noting the high variance of agriculture (mean 89.6, $s_x = 68$), these results seem to suggest that there is a subset of poverty counties with an odd combination of fertile land but highly variable terrain.

Perhaps productivity is high enough to dissuade these counties from switching to more favorable industries (as low-productivity counties would), but rough terrain makes them lose out relative to high-productivity, low-variation counties.

Such poverty counties would be less able to offer amenities such as infrastructure to incoming firms, and also less able to afford cutting taxes. Due to rough terrain they cannot compete with better-endowed counties, but their (relatively) high agricultural productivity implies high opportunity costs of allocating resources to industry.

Geographic factors are likely well-known to policymakers, but the role of county density is perhaps less so. While China's central government has been generous in its support for poverty counties, our analysis emphasizes network effects—not merely a problem of absolute disadvantage, but instead one of relative disadvantage, coupled with high opportunity costs of re-allocating capital.

结　　论

This thesis was structured in four parts. The first part outlined theories of Chinese inter-county competition, from the origin of counties in imperial China, to counties' role in China's economic growth, and finally to how counties answer the 'China puzzle'.

The second part examined current problems in China's economy linked to counties. Examples included capital misallocation from local governments abusing stimulus policies to benefit favored firms, land financing as a 'resource curse' that undermines China's target-responsibility system, and poverty counties being left behind as China's growth-centric system fails to account for new priorities such as inequality.

The third part outlined summary statistics as well as our econometric approach. To attract businesses, county governments offer lower (enforced) taxes, so the effective tax rate acts as an index for pro-business policy. Further, county density within a given area can proxy for inter-county competition. To control for omitted variable bias, we use geographic variables as instruments for density, and also control for present-day GDP.

The fourth step was to run our regression and do robustness checks, experimenting with adjacent neighbours instead of neighbours within a set distance, with various specifications for the tax rate, and with per capita GDP instead of night lights.

Our first finding was geographical. In the debate whether geographically favorable areas will have fewer, larger counties (to maximize advantages) or many smaller counties (limited by the emperor), we found that both more fertile areas and areas with smoother terrain tend to have more counties, supporting the latter theory.

Our second finding was methodological. We found that OLS understates the effect of county density on the effective tax rate (as 0.037%), compared to TSLS with geographic IVs, where each additional neighbor within 100km implies a 0.092% lower tax rate. Further, this discrepancy was robust to numerous alternative specifications.

Our third finding was practical. Taken as a whole, China's poverty counties tend to have rougher terrain, fewer neighbours, and less fertile land than average. Yet, there is some evidence to suggest a class of poverty counties with both high agricultural productivity and relatively high geographic variation—where fertile land keeps them from switching to more favorable industries, but rough terrain makes them lose out relative to high-productivity, smooth-terrain counties. Further, due to their poverty, they must charge a higher effective tax rate than other counties, leading to a vicious cycle.

参考文献

[1] Bai, C., Hsieh, C., Song, Z. (2016). “The Long Shadow of China's Fiscal Expansion.” *Brookings Papers on Economic Activity* 2016(2), pp. 129-65

[2] Blanchard, O. & Schleifer, A. (2001). “Federalism with and without political centralization: China versus Russia.” IMF Staff Papers 48, pp. 171-9

[3] Brehm, S. (2013). “Fiscal Incentives, Public Spending, and Productivity: County-Level Evidence from a Chinese Province.” *World Development* 46, pp. 92-103

[4] Cheung, S. (2005). *Economic Explanation: Selected Papers of Steven N.S. Cheung*. Hong Kong: Arcadia Press.

[5] Cheung, S. (2014). “The Economic System of China.” *Man and the Economy* 1(1), pp. 1-49

[6] Henderson, J., Storeygard, A., Weil, D. (2012). “Measuring Economic Growth from Outer Space.” *American Economic Review* 102(2), pp. 994-1028

[7] Jin, F. & Zhang, Q. (2015). “Chinese regional productivity and urbanization: a county-level study in 2007-2010.” *Journal of Chinese Economic and Foreign Trade Studies* 8(2), pp. 82-105

[8] Li, R. (2014). “The Era of Prefectures and Counties: An Inquiry into the Power Structure and State Governance in Ancient Chinese Society.” *Journal of Chinese Humanities* 1, pp. 67-87

[9] Li, H. & Zhou, L. (2005). “Political turnover and economic performance: the incentive role of personnel control in China.” *Journal of Public Economics* 89, pp. 1743-62

[10] Li, T., Long, H., Tu, S., Wang, Y. (2015). “Analysis of Income Inequality Based on Income Mobility for Poverty Alleviation in Rural China." *Sustainability* 7, pp. 16362-78

[11] Liu, Y., Tang, Y., & Zhao, X. (2017). “Connections and Corporate Tax Avoidance: Evidence from Chinese County Boundary Adjustments." 7th International Conference on Transition and Economic Development.

[12] Mayshar, J., Moav, O., Neeman, Z., Pascali, L. (2018). “The Emergence of Hierarchies and States: Productivity vs. Appropriability.” Working Paper. Retrieved from http://warwick.ac.uk/fac/soc/economics/staff/omoav/mmnp9_1_2018.pdf

[13] Mo, J. (2018). “Land financing and economic growth: Evidence from Chinese

counties.” *China Economic Review* 50, pp. 218-39

[14] Rogers, S. (2014). “Betting on the strong: Local government resource allocation in China's poverty counties.” *Journal of Rural Studies* 36, pp. 197-206

[15] Xu, C. (2011). “The fundamental institutions of China's reforms and development.” *Journal of Economic Literature* 49(4), pp. 1076-151

[16] Yep, R. (2008). “Enhancing the redistributive capacity of the Chinese state? Impact of fiscal reforms on county finance.” *The Pacific Review* 21(2), pp. 231-55

[17] Zhan (2013). “Natural Resources, Local Governance and Social Instability: A Comparison of Two Counties in China." *The China Quarterly* 213, pp. 78-100

[18] Zou, J. & Lü, J. (2014). “The Research of County Urbanization and County Economy Development: Based on 44 Agricultural Counties in Liaoning Province." 2014 Fourth International Conference on Instrumentation and Measurement, Computer, Communication and Control, pp. 388-92

Appendix – R Code

County Density

```
#Using data on county adjacency + centroids, find county density
library("readxl")         #NB: I can probably bypass using Excel files later on
library("writexl")
library("geosphere")    #more accurate calculation of distance on earth

#1. Get data from county_centroids + county_neighbour
#2. Make a list of neighbours for each county
#3. For each county, get its neighbours of neighbours of ... (to reduce search space)
#4. Use distGeo formula to calculate distances between centroids in set
#5. If a county j's centroid is <100km from county i, add 1 to i's density

neighbours =
read_excel("C:/Users/Graham/Desktop/Thesis/County_Borders/county_neighbour.xls",
col_types = c("skip","text", "text","skip","skip"))
colnames(neighbours) = c("src_id","nbr_id")

centroids <-
read_excel("C:/Users/Graham/Desktop/Thesis/County_Borders/county_centroids.xls",
              col_types = c("skip","skip","skip","skip","text","numeric","numeric"))
#counties of HK (810000) and Macau (820000) have same county_id -- take average
hk = subset(centroids, county_id=="810000")
hk =
data.frame(county_id=hk[[1]][1],Centroid_X=mean(hk[[2]]),Centroid_Y=mean(hk[[3]]))
mc = subset(centroids, county_id=="820000")
mc =
data.frame(county_id=mc[[1]][1],Centroid_X=mean(mc[[2]]),Centroid_Y=mean(mc[[3]]))
#
centroids <- centroids[!centroids$county_id %in% c("810000","820000"),]
centroids <- rbind(centroids,hk)
centroids <- rbind(centroids,mc)
  counties <- centroids$county_id                    #2853 unique counties
rm(hk,mc)

#makes vector of neighbours for a county
beMyNeighbour <- function(county) {
     nbs.i <- which(neighbours$src_id==county)      #where the county shows up in the vector
     adj.v <- neighbours$nbr_id[nbs.i]                   #vector with neighbours of county i
     return(adj.v)                                                  #to try: binary search + bug-prevention
}

#list: for each county, gives vectors of neighbours
getNeighbours <- function() {
   nbrs.l <- list()
   n <- length(counties)

   for (i in (1:n)) {
      nbrs.l[i] <- list(beMyNeighbour(counties[i]))
      names(nbrs.l)[i] <- counties[i]
   }
   return(nbrs.l)
```

```
}

adjacent.l = getNeighbours()        #NB: not in nbrs.sq b/c gets called over & over in c.density

#recursive function to find neighbours to the nth degree (neighbours 'squared')
nbrs.sq <- function(county,n) {
   if (n<=0) {return(county)} else
   nbrs.v = NULL
      for (i in county) {
      cnty.s <- as.character(i)        #without this, interprets i as literal index value
      allnbr <- lapply(adjacent.l[[cnty.s]],beMyNeighbour)
      unique <- unique(unlist(allnbr))
      nbrs.v <- c(nbrs.v,unique)
   }
   county <- unique(nbrs.v)            #re-using varname 'county' to emphasize recursion
   return(nbrs.sq(county,n-1))
}


#NB: distGeo function comes from geosphere package; can enter distHaversine as test
distance = function(county1,county2,FUN=distGeo) {
   i <- which(centroids$county_id==county1)
   coord1 <- c(centroids$Centroid_X[i],centroids$Centroid_Y[i])
   j <- which(centroids$county_id==county2)
   coord2 <- c(centroids$Centroid_X[j],centroids$Centroid_Y[j])
   d = FUN(coord1,coord2)
   return(d/1000)      #distance in km
}
#length(which(densities!=test)) == ___
#in __, haversine > geo.Dist; in __, geo.Dist > haversine

#use to compare mean of nbrs.sq("110101",2) to mean of nbrs.sq("110101",3), etc.
test.dist <- function(county,n,FUN=mean) {
   test.v <- NULL
   neighbours <- nbrs.sq(county,n)
   for (i in neighbours) {test.v <- c(test.v,distance(county,i))}
   return(FUN(test.v))
}
#test.dist("110101",2)==80.50643          #test.dist("110101",2)==147.3078


#densities of one county
c.density <- function(county,n,limit=100) {
   neighbours <- nbrs.sq(county,n)
   density <- 0
   for (i in neighbours) {
      if (distance(county,i) <= limit) {density = density + 1}
      else next
   }
   return(density)
}

#loops over all counties in vector
getDensities <- function(counties,n,limit=100) {
   densities <- NULL
```

```
    for (i in counties) {densities <- c(densities,c.density(i,n,limit))}
    return(densities)
}

densities <- getDensities(counties,3,100)      #n=2: testing, n=3: robustness check


#Robustness check: does it make a difference whether neighbour depth is n or n+1?
test.nbr.depth <- function(d1,d2) {
    density1 <- getDensities(counties,d1)
    density2 <- getDensities(counties,d2)
    diff.dense <- length(which(density1!=density2))
    return(diff.dense)
}
#test.nbr.depth(2,3)           #robustness check -- actually, difference is 0!

#county_density <- data.frame(county_id=counties,density=densities)
#tempfile <- write_xlsx(county_density)

#get histogram of densities
hist(densities, main=NULL, xlab="County density",
        cex.lab=1.55, cex.axis=1.55, cex.main=1.55, cex.sub=1.55)           #makes font larger
```

Assemble Data

```
#Assemble different data from spreadsheets into one data frame
library("readxl")
library("writexl")

#variables: tax_enforcement, density, agri_prod, geo_var, lights

#- drop problem counties so that each variable has 629 elements
#- geo_var & agri_prod drop too many -- down to 473. Find out why.
#- then countrol drops even more, down to 462
#- use rle() to repeat elements for each firm in the firm-level data
#- make into a great big data.frame, export to Excel (use in Stata)
#- county dataset only has 2408 counties; LY's has 2853

firm_data <- read_excel("C:/Users/Graham/Desktop/Thesis/Regression/firm_data.xlsx",
                                          col_types = c("text","skip","skip","skip","numeric", "numeric",
                                        "numeric","numeric","numeric","numeric","numeric","numeric"))
firm_data = na.omit(firm_data)                                                       #276474 obs
firm_data = firm_data[order(firm_data$county_id),]
#firm_data = subset(firm_data, county_id %in% agri_prod$county_id) #276,474 --> 271,525

agri_prod = read_excel("C:/Users/Graham/Desktop/Thesis/Regression/agri_prod.xls",
                                    col_types = c("skip","text","skip","skip","skip","numeric")) #2782 obs
lost_agri <- read_excel("C:/Users/Graham/Desktop/Thesis/Regression/missing_agri.xlsx",
                                                col_types = c("text","numeric"))
lost_agri = na.omit(lost_agri)                              #still missing: c("210402", "340702")
agri_prod = rbind(agri_prod,lost_agri)                #27 firms in "210402", 28 firms in "340702"
agri_prod = agri_prod[order(agri_prod$county_id),]    #now: 2851 obs
rm(lost_agri)
```

```
# problem: firm_data has 2849 counties, but IDs don't all match IDs in agri_prod
#                  69 counties in firm_data but not agri_prod, 71 in agri_prod but not firm_data
#solution: check for counties with close IDs, change firm_data IDs into close ones
full = unique(firm_data$county_id)                                    #2849 counties
missing = full[which(! full %in% agri_prod$county_id)]               #69 counties
#miss2 = agri_prod$county_id[which(! agri_prod$county_id %in% full)] #71 counties= 69+2

#given a missing observation, tries to find any county_id within 10 points
salvage <- function(obs,n=10) {
   lost = as.numeric(obs)
   try1 = lost + c(1:n)           #vector from obs+1 to obs+n
   try2 = lost - c(1:n)
   if (any(try1 %in% agri_prod$county_id)) {
      sup = min(try1[which(try1 %in% agri_prod$county_id)])    #least upper bound
   }
   else sup = FALSE
   if (any(try2 %in% agri_prod$county_id)) {
      inf = max(try2[which(try2 %in% agri_prod$county_id)])    #greatest lower bound
   }
   else inf = FALSE

   if (sup==FALSE & inf==FALSE)    {return(obs)} else        #if no match, returns original
   if (sup==FALSE) {return(as.character(inf))} else
   if (inf==FALSE) {return(as.character(sup))} else
   if ((sup - lost) < (lost - inf)) {return(as.character(sup))}      #gives value closest to obs
   else return(as.character(inf))          #NB: using < (over <=) means favoring lower id in a tie
}

find.agri <- function(missing,n=10) {
   found = NULL
   for (i in missing) {found = c(found,salvage(i,n))}
   return(found)
}
#length(which(missing==find.agri(missing)))         #how many counties not salvaged



geo_var = read_excel("C:/Users/Graham/Desktop/Thesis/Regression/geo_var.xls",
                              col_types = c("skip","text","skip","numeric"))
colnames(geo_var) = c("county_id","elevation")
geo_var = geo_var[order(geo_var$county_id),]

rivers = read_excel("C:/Users/Graham/Desktop/Thesis/Regression/county_river.xls",
                           col_types = c("text","numeric","numeric"))    #only 2615 obs (omits 0s)
lost_rivs = agri_prod$county_id[(which(! agri_prod$county_id %in% rivers$county_id))]
lost_rivs = data.frame(lost_rivs,1e-08,1)              #for each river, sets length=1e-08 & area=1
colnames(lost_rivs) = c("county_id","length","area")
rivers = rbind(rivers,lost_rivs)
rivers = rivers[order(rivers$county_id),]
rivers$r.density = log(100000*rivers$length / rivers$area)
rivers$r.density[which(rivers$area==1)] = 0

density = read_excel("C:/Users/Graham/Desktop/Thesis/Regression/county_density.xlsx",
                              col_types = c("text", "numeric"))
```

```
density = density[order(density$county_id),]


lights = read_excel("C:/Users/Graham/Desktop/Thesis/Regression/nightlights.xls",
                    col_types = c("skip", "text", "skip", "skip", "skip", "numeric"))
lights = lights[order(lights$county_id),]

poverty = read_excel("C:/Users/Graham/Desktop/Thesis/Regression/poverty_counties.xls",
                     col_types = c("text", "skip", "numeric", "numeric", "numeric"))
poverty = poverty[order(poverty$county_id),]


#get same number of observations (2780) for each variable
agri_prod <- subset(agri_prod, county_id %in% firm_data$county_id)#2851 obs --> 2780 obs
lights = subset(lights, county_id %in% agri_prod$county_id)          #2853 obs --> 2780 obs
geo_var = subset(geo_var, county_id %in% agri_prod$county_id)     #2853 obs --> 2780 obs
density = subset(density, county_id %in% agri_prod$county_id)        #2853 obs --> 2780 obs
rivers = subset(rivers, county_id %in% agri_prod$county_id)          #2831 obs --> 2780 obs
poverty = subset(poverty, county_id %in% agri_prod$county_id)      #2873 obs --> 2780 obs
firm_data <- subset(firm_data, county_id %in% agri_prod$county_id) #276474   --> 271525

#define tax_enforcement variable
firm_data$income = firm_data$income + 0.001      #avoid dividing by zero
firm_data$employ = firm_data$employ + 0.001
tax_rate = 100*(firm_data$taxes + firm_data$corp_tax + firm_data$vat) / firm_data$income

#use run length encoder to get # of repetitions (firms) in each county
reps.v <- rle(firm_data$county_id)$lengths

#copies values for each firm in the county
#template: rep(c(1,2,3),c(3,2,4)) == c(1,1,1,2,2,3,3,3,3)
dsty = rep(density$density,reps.v)
geov = rep(geo_var$elevation,reps.v)
agri = rep(agri_prod$MEAN,reps.v)
lite = rep(lights$MEAN,reps.v)
rvrs = rep(rivers$r.density,reps.v)
pvty = rep(poverty$poverty832,reps.v)       #this is the 832 list; leave out 591 for now

regression = data.frame(firm_data$county_id,tax_rate,dsty,geov,agri,lite,rvrs,pvty)
colnames(regression) <- c("county","tax_rate","density","geo_var",
                          "agri_prod","lights","rivers","poverty")
rm(tax_rate,dsty,geov,agri,lite,rvrs,pvty)
regression$county = as.character(regression$county)         #convert from factor to char

#drop outliers from tax_enforcement (NB: need original version for robustness checks)
top.pct = quantile(regression$tax_rate,0.995)      # 37.8
bot.pct = quantile(regression$tax_rate,0.005)      #-4.53
regression = subset(regression, tax_rate>=bot.pct & tax_rate<=top.pct)        #268809 obs
rm(top.pct,bot.pct)
#length(which(regression$tax_rate > top.pct))      #1358 for both --> 2716 firms dropped (1%)
#tempfile <- write_xlsx(regression)          #get from C:\Users\Graham\AppData\Local\Temp

reps <- rle(regression$county)$lengths
regression = data.frame(regression,rep(reps,reps))
colnames(regression)[9] = "weight"
```

Robustness Checks

```
#import functions from density.R
#Robustness check #1 - Immediate neighbours only
adj.l = subset(adjacent.l, names(adjacent.l) %in% agri_prod$county_id)
adj.l = adj.l[order(names(adj.l))]

#count neighbours for each county
n = length(adj.l)
adj.nbr <- vector(length=n)
   for (i in (1:n)) {
      adj.nbr[i] = length(adj.l[[i]])
   }
rm(i,n)

#repeat densities for firms in each county
reps.v = rle(regression$county)$lengths
adj.nbr = rep(adj.nbr,reps.v)
reg.adj = regression                    #NB: don't actually need separate file
reg.adj$density = adj.nbr
#tempfile <- write_xlsx(reg.adj)


#Robustness check #2 - Different tax_enforcement specifications
#NB: use original regression data (without quantiles)
firm_data = firm_data[order(firm_data$county_id),]      #make sure they're in order
regression = regression[order(regression$county),]

tax2 = 100*(firm_data$taxes + firm_data$corp_tax + firm_data$vat +
            firm_data$mgmt_fee) / firm_data$income
tax3 = 100*(firm_data$taxes + firm_data$corp_tax + firm_data$vat - firm_data$subsidy) /
            firm_data$income
tax4 = 100*(firm_data$taxes + firm_data$corp_tax + firm_data$vat + firm_data$mgmt_fee -
            firm_data$subsidy) / firm_data$income

#put them in full regression, *then* drop
tax2.r = data.frame(regression,tax2)
tax3.r = data.frame(regression,tax3)
tax4.r = data.frame(regression,tax4)

tax2.r = subset(tax2.r, tax2>=quantile(tax2.r$tax2,0.01) & tax2<=quantile(tax2.r$tax2,0.99))
tax3.r = subset(tax3.r, tax3>=quantile(tax3.r$tax3,0.01) & tax3<=quantile(tax3.r$tax3,0.99))
tax4.r = subset(tax4.r, tax4>=quantile(tax4.r$tax4,0.01) & tax4<=quantile(tax4.r$tax4,0.99))

reps2 <- rle(tax2.r$county)$lengths
tax2.r = data.frame(tax2.r,rep(reps2,reps2))
colnames(tax2.r)[10] = "weight"

reps3 <- rle(tax3.r$county)$lengths
tax3.r = data.frame(tax3.r,rep(reps3,reps3))
colnames(tax3.r)[10] = "weight"

reps4 <- rle(tax4.r$county)$lengths
tax4.r = data.frame(tax4.r,rep(reps4,reps4))
colnames(tax4.r)[10] = "weight"
```

```
#length(unique(tax2.r$county))
#tempfile <- write_xlsx(tax2.r)


#Robustness check #3 - For sample with gdp data, compare with results using nightlights

gdp.ctrl = read_excel("C:/Users/Graham/Desktop/Thesis/Regression/controls_yearly.xls")
gdp.ctrl = subset(gdp.ctrl, year==2005)      #2013
gdp.ctrl = subset(gdp.ctrl, county_id %in% agri_prod$county_id)
gdp.ctrl = na.omit(gdp.ctrl)        #134 obs

pc_gdp = read_excel("C:/Users/Graham/Desktop/Thesis/Regression/pc_gdp.xlsx",
              col_types = c("text","skip","numeric","skip","skip","skip","skip","skip"))
colnames(pc_gdp) = c("county_id","gdp2005")
pc_gdp = pc_gdp[order(pc_gdp$county_id),]
pc_gdp = subset(pc_gdp, county_id %in% agri_prod$county_id)        #2076 obs --> 1410 obs
pc_gdp = na.omit(pc_gdp)                                            #1410 obs --> 1331 obs

#NB: must do na.omit on pc_gdp and gdp.ctrl beforehand
#NB: for 6 obs in pc_gdp, a county_id is used twice
mixgdp = data.frame(agri_prod$county_id,NA)
colnames(mixgdp) = c("county_id","gdp")
mixgdp$county_id = as.character(mixgdp$county_id)
for (i in 1:2780) {
   if (mixgdp$county_id[i] %in% pc_gdp$county_id) {
      mixgdp$gdp[i] = pc_gdp$gdp2005[which(pc_gdp$county_id==mixgdp$county_id[i])]
      } else
      if (mixgdp$county_id[i] %in% gdp.ctrl$county_id) {
         mixgdp$gdp[i] = gdp.ctrl$gdp[which(gdp.ctrl$county_id==mixgdp$county_id[i])]
         } else next
}

gdp = rep(mixgdp$gdp,reps.v)
gdp.reg = data.frame(regression,gdp)
gdp.reg = na.omit(gdp.reg)
#tempfile <- write_xlsx(gdp.reg)
```

后　　记

This thesis is dedicated to my grandfather, Bernard Desjardins (1935-2017).

Thanks to Liu Yu for help choosing a topic, finding data, and piquing my interest.

Thanks to Terry Sicular for her classes on China’s economy and development economics.

Statement of Originality

The thesis is independently written by the author under the direction of the advisor. In addition to what is specially labeled and contained in the acknowledgement, the thesis does not include anything published or written by other persons or organizations. Inspirations and contributions by other people have been clearly stated and appreciated in the thesis.

Author's Signature: Date:

Authorization Statement for Thesis Use

The author fully understands the provisions of Fudan University on keeping and using degree theses, namely: the school reserves the right to retain a copy of the thesis and allow it to be searched and read by others. The school may publish all or part of the thesis and retain it by photocopying, microprinting or other reproduction means. Confidential theses Comply with these provisions after the decryption.

Author's Signature: Advisor's Signature: Date:

论文独创性声明

本论文是我个人在导师指导下进行的研究工作及取得的研究成果。论文中除了特别加以标注和致谢的地方外，不包含其他人或其它机构已经发表或撰写过的研究成果。其他同志对本研究的启发和所做的贡献均已在论文中作了明确的声明并表示了谢意。

作者签名：__________　　日期：____________

论文使用授权声明

本人完全了解复旦大学有关保留、使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其它复制手段保存论文。保密的论文在解密后遵守此规定。

作者签名：_________　导师签名：_________　日期：__________