# Harvesting the Concrete: Urban Expansion and Agricultural Land Productivity[§]

Hai Hong [a]; Yi Shi [a]; Zenghui Li [a, *]; Kevin Chen [a, b]
([a]. *China Academy of Rural Development, Zhejiang University, Hangzhou 310058, China;* [b]. *International Food Policy Research Institute, Hangzhou 310058, China*)

**Abstract**

Land quality and productivity are fundamental to agricultural development, yet urban expansion poses a growing challenge by reallocating high-quality land away from agriculture for urban development. Focusing on China, a country that features both agricultural land protection and growth-oriented political incentives, this paper examines how land-based urbanization affects agricultural land productivity. Using high-resolution satellite data, we document a significant decline in both the quantity and quality of agricultural land due to urban expansion. To understand the mechanisms behind this decline, we develop a theoretical model that extends the monocentric city framework by incorporating local political incentives and land quality considerations into land allocation decisions. The model predicts that strict farmland protection policies focusing on land quantity, such as China's Agricultural Land Protection Policy (ALPP) introduced in 2010, may inadvertently reduce land quality when local leaders prioritize economic growth. Empirical evidence supports this prediction: we find that ALPP enforcement leads to a decline in agricultural land quality, particularly in regions where political incentives for urban expansion are stronger. Additionally, urban expansion reduces agricultural productivity by decreasing the use of modern agricultural technologies and increasing reliance on fertilizers. These findings underscore the need for policies that balance land conservation with quality considerations to ensure long-term agricultural sustainability.

**Keywords:** Urban Expansion; Land Quality; Agricultural Productivity; Political Incentives.
**JEL Codes:** R11; R12; R14; R52; R58.

## 1. Introduction

Land quality and productivity constitute the bedrock of agricultural development, a proposition tracing back to the classical political economy (Petty, 1662; Ricardo, 1817). However, the inelasticity of land supply creates an inevitable conflict: urban-industrial demand for prime locations collides with agricultural preservation needs (Brueckner, 2000; Brueckner and Largey, 2008; Burchfield et al., 2006; Irwin and Bockstael, 2007; Nechyba and Walsh, 2004). This tension is particularly acute in developing economies balancing food security with growth imperatives (Suri, 2011). Typically, as land quality is unevenly distributed across space, emphasis on developing the urban sector may "cherry-pick" lands with higher quality (e.g., land of lower ruggedness, better geographic location, and infrastructure coverage), which could come at the cost of lower land quality in the agricultural sector and eventually decrease agricultural productivity. Despite its importance, however, there is scant evidence of the effects of urban expansion on agricultural land productivity

[§] Hai Hong, Yi Shi, and Zenghui Li contributed equally to this work and share the senior authorship.
[*] Corresponding author.
*E-mail address:* lizenghui98@zju.edu.cn (Z. Li).

and its implication for agricultural development.

In this paper, we study how urban expansion leads to changes in agricultural land productivity in China. China presents an ideal case for investigating the tension, with two important institutional features standing out. First, as one of the largest developing countries, China has approximately 20% of the global population but only 9% of arable land. To ensure domestic food supply, China underscores the protection of agricultural land and has instituted the world's strictest farmland protection regime since 2006.[1] Second, China has also prioritized economic growth and a primary source of this growth wheel is decentralized land finance and political competition (Li and Zhou 2005; Xu 2011; Han and Kung 2015; Yao and Zhang 2015; Chen and Kung 2016; Xi, Yao, and Zhang 2018). The merit-based political selection creates sufficient incentives for local governments to leverage land finance to achieve targeted economic growth (Yao and Zhang 2015; Xi, Yao, and Zhang 2018; Li et al. 2019). However, such political competitions also result in the over-exploitation of land expropriation as a tool for accelerating economic growth and lead to the over-expansion in the urban size (Deng et al., 2008; Lichtenberg and Ding, 2009; Wang et al., 2020). As both food production and urban development require high-quality land, these features together lead to tensions on how the government allocates productive land between urban and agricultural sectors, creating a relevant and interesting setting for our investigation.

Our investigation comprises three parts. In the first part, we present the general relationship between urban expansion and agricultural land productivity. To measure agricultural land quality, we rely on high-resolution land cover datasets and detailed topographic information to calculate the time-varying terrain ruggedness of agricultural land within each county due to changes in land use types. Higher ruggedness in agricultural land topography thus represents lower quality in agricultural land. Our findings reveal strong evidence that urban outward expansion is associated with agricultural land declines in terms of both quantity and quality. To bolster a causal interpretation, we leverage a shift-share IV design that combines each county's pre-determined economic importance (share) and exogenous national-level infrastructure investment (shift). Examining the potential mechanisms, we show that the expropriation of high-quality agricultural land for industrial and commercial construction is the key driver of the observed correlation.

Next, we exploit a policy change, the Agricultural Land Protection Policy (ALPP) in 2010 that sets clear targets for how much agricultural land each municipality must preserve (see more detailed descriptions of the policy in Section 2), and explore the policy effects on agricultural land productivity. To guide our empirical analysis, we develop a simple theoretical model that extends the traditional monocentric city model by incorporating city leaders' political incentives and land quality into the decision on land allocation between urban and agricultural sectors. In the model, the city leader decides where to construct the new urban suburb (either on agricultural land or undeveloped land) to maximize his/her utility, which we assume is partially altruistic, i.e., taking into account both the city's industrial growth and citizens welfare (the former measures the performance of the city leader while the latter reflects the emphasis on food security). Since land quality is less observable, the model shows that the enforcement of land quantity protection could reduce agricultural land quality if the local leader has higher promotion incentives.[2] We derive several testable predictions and bring the model derivations to data for investigation. Empirically, we show that, the ALPP leads to significant declines in agricultural land quality, and we find consistent evidence that the effects are

[1] As a prominent example, China has established the "1.8 billion mu (120 million hectares) red line" policy in 2006, reiterating the requirement that the country's total area of arable land must never fall below 1.8 billion mu, which is roughly equivalent to two France.

[2] This is in line with the multitasking moral hazard theory proposed in Holmström and Milgrom (1991), that is, when an agent (local government) is faced with an observable quantitative objective (preservation of arable agricultural land) versus an unobservable qualitative objective (land quality), it will strategically reduce qualitative inputs.

much stronger for political leaders with higher political incentives. Heterogenous analysis suggests that the decrease in land quality is more pronounced if (1) the agricultural productivity is higher; (2) the industrial productivity is lower; (3) the construction costs in other undeveloped land are higher; and (4) the enforcement of the ALPP is weaker; all aligns with the model prediction.

Lastly, we investigate the potential consequences of urban expansion. Our findings reveal that urban expansion leads to significant declines in agricultural output per unit of land, suggesting a decrease in agricultural productivity. Moreover, due to the occupation of high-quality agricultural land, urban expansion also decreases the usage of agricultural machinery power and increases fertilizer usage, implying that urban expansion may hinder the adoption of modern agricultural technologies and the sustainability of agricultural development.

Our paper makes three contributions to the literature. First, we speak to studies on urban spatial expansion (Brueckner, 2000; Brueckner and Largey, 2008; Burchfield et al., 2006; Fallah et al., 2011; Glaeser and Kahn, 2004; Irwin and Bockstael, 2007; Nechyba and Walsh, 2004; Oueslati et al., 2019; Song and Zenou, 2006). While most of the existing studies focus on the causes and consequences of urban expansion in developed countries, there is little evidence on the potential costs of urban expansion in developing countries. We contribute to this strand of literature by providing the first empirical evidence on how urban expansion impacts agricultural productivity for one of the largest developing countries. By exploiting the policy change in land use regulation to examine its potential impacts on agricultural land quality, we also contribute to the burgeoning literature that studies the impacts of China's recent land use supervision (Cai et al., 2017; Fan et al., 2023; Tan et al., 2020; Tian et al., 2024).

Second, we contribute to the broad literature on the determinants of land productivity and soil quality (Besley 1995; Bridges and Oldeman 1999; Dubois 2002; Deininger and Jin 2006; Abdulai, Owusu, and Goetz 2011; Amundson et al. 2015; Borrelli et al. 2017; Huntington and Shenoy 2021; Hou, Liu, and Tian 2023; Li and Zhu 2024). Much of the previous research has focused on micro determinants of land quality, e.g., how land property rights affect the farmers' investment in land and the effects on land fertility, and the macro determinants of agricultural land fertility are far less explored. We fill in this gap by studying how the expansion of the urban sector affects agricultural land productivity. In a similar vein, we also contribute to the nascent but growing literature on the consequences of land expropriation (Cheng et al., 2022; Huang et al., 2024; Ma and Mu, 2020; Sha, 2023; Zhao and Xie, 2022). Most of this literature discussed how land expropriation affects individual behaviors (e.g., political trust or social conflicts). None of these studies, however, evaluates the interactions between land expropriation and the quality of land resources themselves. Our paper extends this literature by assessing an important yet largely ignored effect of land expropriation on land productivity.

Finally, our paper adds to the broad literature that studies the behavioral consequences of multitasking agents. Theoretically, multitasking agents tend to exert extra efforts in tasks that are easily observed by their superiors and exert reduced efforts in tasks that are poorly measured (Alesina and Tabellini, 2008, 2007; Baker, 1992; Dewatripont et al., 2000, 1999; Holmstrom and Milgrom, 1991). We contribute to the literature by presenting new evidence on how multitasking local officials sacrifice agricultural land quality for economic growth under the protection mandates of land quantity. In doing so, we join the vast empirical works that explore how local governments respond to multiple tasks and how career concerns further distort the effort allocation of agents (Kung and Chen 2011; Chen, Li, and Lu 2018; Xi, Yao, and Zhang 2018; Li et al. 2019; Wang, Zhang, and Zhou 2020; Hong and Huang 2025).

The remainder of this paper proceeds as follows: Section 2 introduces the data and some related background of the Agricultural Land Protection Policy. Section 3 presents the empirical pattern between urban expansion and agricultural land quality. Section 4 lays out the theoretical model, and Section 5 empirically exploits the policy change to test model predictions. Section 6 discusses the potential consequences of urban expansion. Finally, Section 7 concludes.

# 2. Background and Data

## 2.1 Background

As one of the largest and fastest-growing developing countries, China has long faced trade-offs between agricultural land protection and urban expansion. Official statistics indicate that approximately 70% of newly developed urban land is converted from agricultural land. However, agricultural land resources in China are limited. According to the 2005 National Land Use Change Survey conducted by the (former) Ministry of Land and Resources, China's per capita arable land area was only 1.4 mu, 40% of the world average.

To address concerns over agricultural land loss amid rapid urbanization, the Chinese government introduced the concept of the Favorable Balance of Agricultural Land (FBAL) in 1997 through the *Circular of the State Council of the Central Committee of the Communist Party of China on Further Strengthening Land Management and Effectively Protecting Agricultural Land*. The FBAL policy mandates that local governments ensure a dynamic balance in the total amount of agricultural land. In 1998, the National People's Congress formalized this requirement by incorporating it into the *Land Management Law of the People's Republic of China*. The core principle of the FBAL policy is that any agricultural land converted for non-agricultural purposes (e.g., urban development) must be compensated with an equivalent amount and quality of land through reclamation or new cultivation.

Despite its legal elevation, the overall decline in agricultural land has persisted. This trend is primarily driven by misaligned incentives: while local governments are responsible for enforcing land protection policies, they also have strong motivations to promote urban expansion, and the latter is more important for their promotion. Recognizing that the neglect of agricultural land protection could pose risks to national food security, the central government further reinforced protection measures in 2006. At the end of that year, the State Council suspended the approval of the *Land Use Master Plan* (hereafter, the *Master Plan*) and tasked the Ministry of Land and Resources with drafting a "historical, crisis-oriented, and strategic land use master plan" to ensure that the red line of 1.8 billion mu of agricultural land would not be breached. The *Master Plan*, approved by the State Council in 2008, set binding targets to maintain at least 1.813 billion mu of agricultural land by 2010 and 1.805 billion mu by 2020. These targets were further disaggregated at the provincial level, with provincial governments allocating specific quotas to municipalities, thus establishing a top-down, multi-tiered management system for agricultural land retention.

Once approved, the *Master Plan* became a regulatory framework with enforceable constraints. The central government incorporated agricultural land conservation targets into the performance evaluations of local officials, and violations of land protection requirements triggered official accountability measures. Given that urban construction land primarily comes from the conversion of agricultural land, the *Master Plan* imposed a significant constraint on future urban expansion. However, despite its strict provisions, violations have been widespread. According to the Ministry of Natural Resources, in 2019 alone, unauthorized occupation of agricultural land nationwide totaled approximately 1,142,600 acres, with most violations occurring in urban development projects such as real estate, industrial parks, and landscaping. Some local governments even misreported land protection targets to gain access to additional agricultural land for development—contravening the objectives of the *Master Plan*.

At its core, these violations stem from the fact that local governments continue to prioritize urban expansion while largely neglecting agricultural development. Additionally, local governments often exploit the compensation mechanism by converting high-quality agricultural land for urban use while offsetting it with lower-quality land.[3] For instance, data from the second and third National Land Surveys indicate that more than 8.8 million acres of forested land with slopes exceeding 25 degrees have been cleared to compensate for agricultural land losses. However, much of this newly designated agricultural land is more fragmented and characterized by poorer farming conditions, undermining the effectiveness of land conservation efforts. In our subsequent analysis, we empirically examine the effectiveness of the *Master Plan* with a focus on how it may potentially affect the quality of agricultural land.

## 2.2 Data

To investigate how urban expansion affects the quality and quantity of agricultural land, we derive a series of variables from satellite observation, government documents, and statistical yearbooks. Our unit of analysis is at the county level, as the county governments face both the pressure of pursuing economic growth and ensuring food production. This subsection briefly introduces the main datasets used in subsequent empirical analysis. The summary statistics are reported in Table 1.

### 2.2.1 Urban expansion

The key explanatory variable in our paper is the extent of urban expansion at the county level. To measure urban expansion, we need detailed data on counties' built-up areas, which is lacking in existing statistical yearbooks. We therefore rely on data from He et al. (2022), who developed the raster file of built-up areas covering the entire territory of China from 1992 to 2020. The resolution of the dataset is 1km×1km. The construction of urban built-up areas is analogous to Harari (2020), who uses nighttime light luminosity to identify the built-up areas for Indian cities. He et al. (2022) advance by additionally considering the Normalized Difference Vegetation Index (NDVI) and surface temperature.[4] Based on the raster file, we calculate the built-up areas for each county from 2001 to 2020.

### 2.2.2 Agricultural land

Our dependent variable of interest is the quality of agricultural land. There are two main challenges when measuring land quality. First, quality itself is hard to observe, making it difficult to find a valid proxy. Second, quality may be affected by other confounders (e.g., farmers' investment decisions), which could introduce additional endogeneity into our estimation. To overcome these challenges, we exploit exogenous agricultural land terrain ruggedness as an indirect measure of land quality. This metric is motivated by two considerations. First, the development of modern agriculture relies on large-scale farming, which is hindered by rugged terrain. This makes terrain ruggedness a relevant measure of land quality and productivity. Second, terrain ruggedness is geographically exogenous and unaffected by farmers' decisions on policy changes. More importantly, since terrain ruggedness is time-invariant, the only time-varying variation in agricultural land quality is from changes in the spatial distribution of agricultural land. As we illustrate further below, the implementation of the ALPP requires the local governments to compensate the same amount of agricultural land if the land is occupied for urban construction. Under ALPP, the expansion of urban built-up areas only involves the location change in agricultural land, and using terrain ruggedness as a measure of land

[3] See http://paper.people.com.cn/rmlt/html/2024-05/31/content_26076362.htm.

[4] For more detailed information on the construction and description of the dataset, see https://www.tpdc.ac.cn/zh-hans/data/3100de5c-ac8d-4091-9bbf-6a02de100c88/.

quality is thus more suitable in our context.

Specifically, to construct the terrain ruggedness of agricultural land, we first create a 30m×30m grid covering the entire territory of China, and for each grid, we calculate its terrain ruggedness using the Digital Elevation Model (DEM) raster file, which has the same resolution as our grid data. We calculate the terrain ruggedness following Nunn & Puga (2012). We then overlay the gridded terrain ruggedness to the land cover raster from the China Land Cover Dataset (CLCD), a remotely sensed product providing nationwide land type classifications at 30-meter resolution from 1990 to 2020 (Yang and Huang, 2021).[5] Finally, we calculate the terrain ruggedness for grids that are identified to be agricultural land by CLCD and average it to the county level for each year between 2001 and 2020. To avoid potential measurement error, when calculating the average terrain ruggedness, we exclude grids that are covered by urban built-up areas in the last year. Similarly, we calculate ruggedness for non-agricultural land outside urban built-up regions to capture the potential urban development costs on non-agricultural land. Using the CLCD, we also calculate the area of agricultural land for each county-year cell to investigate how urban expansion affects the quantity of agricultural land. We also calculate the area of lands other than agricultural land and built-up land in each county (e.g., forest, grassland, wetland, etc.).[6]

### 2.2.3 Agricultural land protection target

To assess the impact of the Agricultural Land Protection Policy (ALPP), we collect agricultural land preservation targets from the *Land Use Master Plans*—or the *Master Plan* for short—published by China's Ministry of Natural Resources and provincial governments in 2010. The *Master Plan* disaggregates the national goal of maintaining 1.8 billion mu (120 million hectares) of agricultural land by 2020 into specific targets for each municipality, requiring them to retain no less than their assigned quota. To strengthen enforcement and supervision, the *Master Plan* is implemented in two phases. In the first phase, it sets agricultural land protection quotas for 2010, mandating that each municipality meet and maintain its assigned target from 2010 onward. In the second phase, the *Master Plan* establishes updated quotas for 2020, requiring municipalities to reach their respective targets by that year.

We construct our measure of agricultural land protection intensity based on the sharp and exogenous changes in protection quotas mandated by the *Master Plan*. Specifically, we first obtained data on each municipality's actual agricultural land area in 2005 from the Ministry of Natural Resources and calculated the differences between mandated and actual agricultural land levels. The first difference is between the 2010 target and the 2005 baseline, and the second is between the 2020 target and the 2010 target. These differences capture variations in agricultural land protection pressures before and after 2010, allowing us to quantify municipalities' exposure to the policy shift. To further refine our identification strategy, we subtract the pre-2010 target difference from the post-2010 target difference. This normalization sets the pre-2010 difference to zero, effectively creating a design analogous to a continuous Difference-in-Differences (DiD) framework, where the period before 2010 serves as the pre-policy period and the period after 2010 as the post-policy period. The treatment intensity is thus captured by the change in agricultural land quotas induced by the policy.

[5] The dataset provides coverage not only for agricultural land but also for other land types such as woodland, grassland, water bodies, etc. For more detailed information on the construction and description of the dataset, see https://essd.copernicus.org/articles/13/3907/2021/.

[6] One potential caveat with our measure of the quality and quantity of agricultural land is that we cannot distinguish whether changes in land quality or quantity are caused by urban expansion, and our dependent variables may thus be measured with error. However, measurement errors in dependent variables only lead to imprecise estimation but are less likely to affect the unbiasedness of the coefficients.

### 2.2.4 County variables

We obtain a set of socio-economic variables from the *County Statistical Yearbooks*. Specifically, we include agricultural GDP, rural per capita income, industrial structure (measured as the ratio of secondary to tertiary industries), and the urban-rural income gap (measured as the ratio of urban to rural income). To mitigate potential endogeneity concerns related to “bad control”, we use only their values from the year 2000—the starting point of our sample period—and interact them with time fixed effects to flexibly account for their time-varying impacts. Additionally, we incorporate a set of geographic variables, including a county’s straight-line distance to the provincial capital, the coastline, and the nearest railway and highway.[7] Since these geographic factors are time-invariant, we interact each variable with a full set of time dummies to allow for differential effects across years. Lastly, we calculate the weighted average of industrial firm TFP, with the weight being the firm’s size. The firm data is derived from the National Tax Survey Dataset (NTSD).[8]

**Table 1 Summary Statistics**

| Variable | Description | Obs | Mean | SD |
|---|---|---|---|---|
| Agricultural Land Quality | Land quality measured by terrain ruggedness[1,2] | 57540 | 96.56 | 118.3 |
| Other Land Quality | Land quality measured by terrain ruggedness[1,2] | 57540 | 179.1 | 150.3 |
| Agricultural Land | County agricultural land areas[2] | 57540 | 440.4 | 732.5 |
| Urban Expansion | County built-up area[2] | 57540 | 53.95 | 97.74 |
| Target (Normalized) | Municipal level agricultural land protection target[3] | 46700 | 0.740 | 0.0400 |
| Agricultural GDP | County agricultural GDP (2000)[3] | 53580 | 14.03 | 314.1 |
| Rural Income | County rural income per capita (2000)[3] | 53580 | 2753 | 3405 |
| Industry Structure | The ratio of secondary to tertiary industries (2000)[3] | 53560 | 61.90 | 1017 |
| Income Disparity 2000 | The ratio of urban to rural income (2000)[3] | 53580 | 4.990 | 6.420 |
| Dist. Railway | Nearest distance to railway (2010)[4] | 53240 | 55.93 | 151.2 |
| Dist. Highway | Nearest distance to the highway (2010)4 | 53240 | 4.990 | 14.63 |
| Dist. Coast | Nearest distance to coastline[5] | 53240 | 605.5 | 623.6 |
| Dist. province | Nearest distance to the provincial capital[5] | 53540 | 132.7 | 77.40 |
| County TFP | Weighted average of industrial TFP[6] | 51440 | 4.040 | 1.100 |

*Notes*: This table presents the summary statistics for the main variables used in the empirical analysis. The superscript denotes data sources from which we construct these variables. 1. China Digital Elevation Model (DEM) Map. 2. China Land Cover Dataset. 3. County Statistical Yearbooks. 4. Open Street Map. 5. County Shapefile. 6. National Tax Survey Dataset (NTSD).

# 3. Empirical Patterns

## 3.1 Specification

This section presents some empirical patterns between urban expansion and the quality and quantity of agricultural land. Our baseline specification exploits the following two-way fixed effects (TWFE) model:

$$y_{it} = \alpha + \beta UE_{it} + \gamma X_i \times T + \delta_i + \zeta_t + \epsilon_{it} \tag{1}$$

Where $y_{it}$ is either the quality or quantity measure of agricultural land in county $i$ at year $t$. $UE_{it}$ is our main explanatory variable, which represents the built-up city area for each county during different years, i.e., the extent of urban expansion. The coefficient $\beta$ measures how additional expansion in the urban

[7] We obtain the 2010 road map from Open Street Map (OSM).

[8] The National Tax Survey Dataset (NTSD) is jointly collected by the State Administration of Taxation and the Ministry of Finance, spanning from 2007 to 2016. We use data from 2007 and follow Levinsohn and Petrin (2003) to calculate each firm’s TFP. See Appendix B for more detailed descriptions of the NTSD.

region predicts the quality/quantity of agricultural land. $X_i$ is a set of socio-economic and geographic control variables detailed in Section 2.2. We interact each covariate in $X_i$ with a full set of year dummies to avoid potential endogeneity and issues related to "bad controls" (Angrist & Pischke, 2008). $\delta_i$ and $\zeta_t$ are county and time fixed effects, respectively. Finally, standard error is clustered at the county level to control for any arbitrary serial correlations within the county.

Coefficients from Equation (1) deliver a casual interpretation only if the unobserved confounder is time-invariant within each county or is a common shock varying across different periods. However, our identification assumption would be violated if there are time-varying omitted variables or the presence of reverse causality. For instance, the observed correlation between urban expansion and the decline in agricultural land quantity may be driven by rural-urban migration. With more rural migrants moving to the urban region, the urban size increases, and rural migrants may abandon their agricultural land. This, however, does not necessarily imply that urban expansion directly leads to the occupation of agricultural land. Another example is that urban expansion may be less/more likely to occur in counties with lower agricultural land quality, as the agricultural land quality itself may affect the decision of direction and speed of urban expansion, resulting in the reverse causality, which could either overestimate or underestimate the true parameter. Finally, since our measure of urban expansion is derived from satellite observations, measurement error could introduce a downward bias in our estimation.

To aid these potential endogeneity concerns, we exploit an instrumental variable approach. Our instrument is akin to a shift-share design (Borusyak et al., 2022; Goldsmith-Pinkham et al., 2020), which combines exogenous shocks in national infrastructure investment with each county's initial contribution to the national GDP. The key insight here is that counties with higher initial GDP should be more exposed to national infrastructure investment, which is a prerequisite for urbanization (Fretz et al., 2022). Specifically, our first stage takes the following form:

$$UE_{it} = \gamma_0 + \gamma_1 GDP_i \times Infras_{t-1} + X_i \times T + \delta_i + \zeta_t + \epsilon_{it} \tag{2}$$

The second stage is then:

$$y_{it} = \alpha_2 + \beta_2 \widehat{UE}_{it} + X_i \times T + \delta_i + \zeta_t + \epsilon_{it} \tag{3}$$

Where $GDP_i$ is the GDP for each county in the year 2000, and $Infras_{t-1}$ is the national road length in the period $t-1$. We use the lagged term for infrastructure investment to avoid issues related to simultaneity. $\widehat{UE}_{it}$ is the predicted urban expansion from the first stage. Other variables follow the same definition as in Equation (1).

The validity of our instrument variable hinges crucially on two assumptions. The first is that infrastructure investment should strongly predict subsequent urban expansion, i.e., the relevance assumption. The second is that, conditional on covariates and fixed effects, our instrument is uncorrelated with the error term, i.e., the exclusion restriction. We first provide evidence that our instrument strongly predicts urban expansion and discuss the sensitivity of the exclusion restriction assumption after presenting our main results.

To begin with, in Figure 1, we display a binscatter plot between the residualized value of the urban expansion and the instrument, partialling out the county and year fixed effects. The figure reveals a strong positive correlation between the instrument and our endogenous variable, and the correlation can be well-predicted by a straight line, suggesting that our instrument is a good linear predictor of urban expansion.

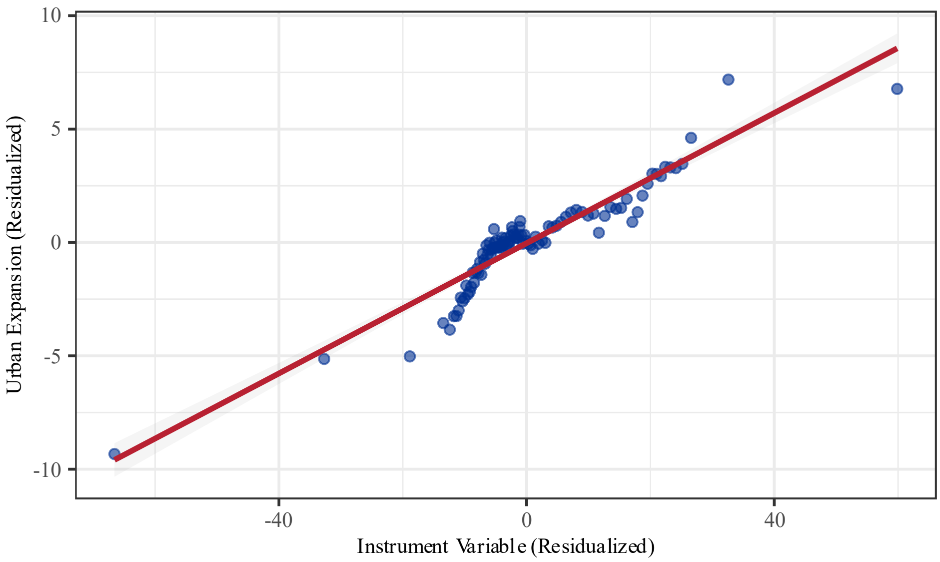



**Figure 1 Binscatter Plot**

*Notes*: This figure presents the binscatter plot for urban expansion versus the instrumental variable, with a fitted line colored in red. The shaded area is the 95% confidence interval.

Formally, Table 2 presents the estimated results from the first-stage regression. In column (1), we perform the most parsimonious specification that only includes county and year fixed effects. We find strong correlations between the instrument and urban expansion. The F-statistic is 33.86, which is higher than the rule-of-thumb value suggested in the literature (Stock and Yogo, 2002), alleviating concerns of weak IV. In column (2), we add our control variables and continue to find the significant predictive power of our instrument. In column (3), we perform a simple placebo test by additionally controlling for the leading term of our instrument. If the observed relationship is caused by spurious correlations, then using the leading term should also have sizeable predictive power on urbanization.[9] However, results from column (3) suggest the opposite. We find no systematic correlations between the leading term of the instrument and urbanization.

Next, in columns (4) and (5), we split our sample into before-2013 and post-2013 subsamples and estimate the specification (1) separately. We do so because much of the road investment at the county level (especially the national roads) occurred after 2013.[10] If our first-stage estimates reflect the impacts of differential exposure to road construction, then we should find larger effects in the post-2013 sample. In line with our expectations, we find a larger coefficient in periods after 2013, suggesting that our instrument effectively captures the variation in changes in national-level road investment.

Our shift-share IV design is essentially analogous to a generalized Difference-in-Difference (DiD) framework. In a nutshell, our IV compares urbanization between counties with higher and lower initial GDP in years when counties receive different infrastructure investment shocks (Borusyak et al., 2025; Goldsmith-Pinkham et al., 2020). To ensure that variations in urbanization are indeed driven by counties' differential

[9] The correlation between our instrument and its leading term is 0.99.

[10] Although the construction of national highway networks was primarily completed before 2013, the highway system is designed to connect major metropolitans, and by 2013, there are still more than half of the counties that are not connected by highway and do not have constructed national roads. See the National Plan for Road Network (2013-2030) at https://zfxxgk.ndrc.gov.cn/web/iteminfo.jsp?id=285. Specifically, the goal National Plan for Road Network is to form a national trunk highway network that is "reasonably arranged, functionally perfect, widely covered, safe and reliable", realizing the radiation of the capital city to the provincial capitals, inter-provincial multi-road connectivity, high-speed access to cities and counties and national highway coverage in counties and counties; within 1,000 kilometers of the provincial capitals can be reached on the same day, and from provincial capitals to cities in the eastern and central regions can be reached on the same day.

exposure to infrastructure investment, we perform three additional tests. First, in column (6) of Table 2, we refine our instrument into a DiD term that interacts with two dummies; the first is whether the county's GDP is above the national average (denoted by $1(HighGDP)_i$), and the second is whether the period is post-2013 (denoted by $1(Post2013)_t$).[11] To avoid concerns that our results may be driven by the co-evolution between urbanization and national infrastructure investment growth, we include in our regression the differential linear time trends for counties in the high and low GDP groups. Consistent with our interpretation, we find that the effects of road investment on urbanization are primarily driven by counties with higher GDP and in periods when the large road investment programs begin.

**Table 2 First Stage Regression**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Dep. Var. | Urban Expansion | | | | | |
| $GDP_i \times Infras_{t-1}$ | 0.050*** | 0.037*** | 0.032** | 0.029*** | 0.074*** | |
| | (0.009) | (0.010) | (0.016) | (0.008) | (0.018) | |
| $GDP_i \times Infras_t$ | | | 0.006 | | | |
| | | | (0.015) | | | |
| $1(HighGDP)_i \times 1(Post2013)_t$ | | | | | | 0.319** |
| | | | | | | (0.129) |
| Control Variables | No | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Period | Full | Full | Full | Before 2013 | After 2013 | Full |
| Observations | 53540 | 53560 | 53560 | 32136 | 21424 | 53560 |
| F-statistic | 33.86 | 15.04 | 10.10 | 16.00 | 15.26 | 34.65 |

*Notes*: This table presents the results from the first stage regression. The dependent variable is the urban built-up area. The observation is at the county-year level. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the county level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

To further support the validity of our first-stage specifications, we conduct an event-study analysis to examine whether there are differential trends between high-versus-low initial GDP counties. The idea is that if the large-scale road construction plan does not begin before 2013, then we should find no statistically significant treatment effects heterogeneity between high-versus-low initial GDP counties before 2013. The event study specification is:

$$UE_{it} = \gamma_0 + \sum_{t=2005, t\neq 2009}^{2020} \gamma_\tau 1(HighGDP)_i \times 1(t=\tau) + X_i \times T + \delta_i + \zeta_t + \epsilon_{it}, \quad (4)$$

where we combine periods before 2008 and periods after 2018 into single periods separately to improve the estimation efficiency. The absence of pre-trends provides additional support to our identification assumption. Figure 2 visualizes the estimated coefficients and the corresponding 95% confidence intervals. Reassuringly, we find insignificant pre-trends before 2013, and the estimated effects are significant only in post-2013 periods.

[11] Formally, the specification takes the following form:

$$UE_{it} = \gamma_0 + \gamma_1 1(HighGDP)_i \times 1(Post2013)_t + X_i \times T + \delta_i + \zeta_t + \epsilon_{it} \quad (F1)$$

Where $1(HighGDP)_i$ denotes whether the county's initial GDP is above the national average, and $1(Post2013)_t$ denotes whether the period is post-2013. Other variables follow the same definition as in other specifications.

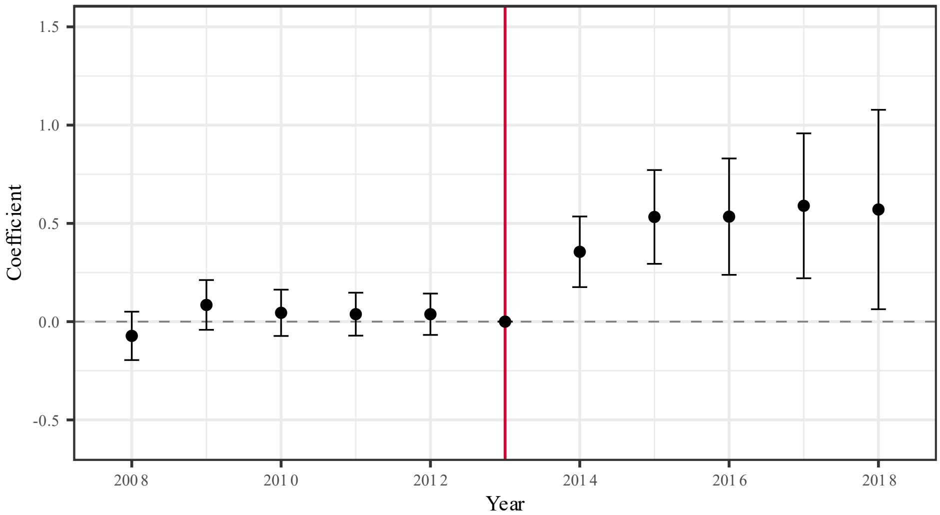



**Figure 2 Event Study Estimates for DiD Instrument**

*Notes*: This figure presents the event study coefficients. Regression specification is presented in Equation (4). The year 2013 is omitted as the reference year. The specification includes county fixed effects and year fixed effects. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the county level. Coefficient estimates and 95% confidence intervals are shown in the figure.

Lastly, we conduct two randomized inferences by randomly assigning the initial GDP values to each county and by randomly assigning road investment to each year. We then construct the placebo instruments by either interacting the placebo initial GDP with the actual road investment or the placebo road investment with the actual initial GDP.[12] Finally, we estimate the placebo coefficients by regressing urban expansion on placebo instruments. We perform both exercises 500 times and visualize the distribution of the estimated coefficients in Appendix Figure A1. We find that the placebo coefficients are very small in magnitude and are far from the true coefficient estimated in Table 2, suggesting that our identified effects are unlikely to be driven by concurrent confounders. Again, this piece of evidence reinforces the validity of our instrument construction.

## 3.2 Main Results

### 3.2.1 Baseline estimates

We now present the estimated effects of urban expansion on the quality and quantity of agricultural land, with Panel A presenting the OLS estimates and Panel B the IV estimates. In columns (1) and (3), we report the coefficients on agricultural land quantity, while in columns (2) and (4), we report the coefficients on agricultural land quality.

Focusing on the effects on the quantity of agricultural land in Panel A, we find that urban expansion leads to significant declines in agricultural land coverage. The estimated coefficients are stable when we include control variables or alternatively use IV estimation in Panel B, suggesting less concern about omitted variable bias. Our coefficient from column (3) in Panel B suggests that a one standard deviation increase in urban built-up area induced by road investment leads to a 0.17 standard deviation decrease in agricultural

[12] The random draw of the initial GDP and road investment follows the same mean and standard deviation as in the actual data.

land area. In columns (1) to (6) of Appendix Table A1, we report the estimated results analogous to Table 3, with the dependent variable replaced by other land cover types (e.g., grassland, forest, and wetland). We find no significant effect of urban expansion on the coverage of these land types. This implies that China's land-based urbanization is primarily through the occupation of agricultural land.

Turning to the estimated effects of urbanization on agricultural land quality, we find that urban expansion leads to significant increases in the terrain ruggedness of agricultural land, suggesting a worsened land quality. The IV estimates are 4 to 5 times larger than the OLS estimates, suggesting a downward bias in OLS regression. Specifically, the coefficient in column (4) of Panel B implies that a one standard deviation increase in urban built-up area leads to a 0.12 standard deviation rise in agricultural land terrain ruggedness, which is equivalent to a 15.1% decrease in land quality relative to the mean. In columns (7) and (8) of Appendix Table A1, we perform the same estimates with the terrain ruggedness of other land (excluding urban and agricultural land) serving as the dependent variable. We find stark contrasts that urban expansion has no significant impact on the quality of other land.

Taken together, our results from Table 3 reveal the significant yet unintended costs of urban expansion on both the quality and quantity of agricultural land, and additional results from Appendix Table A1 further confirm that the costs are exclusively borne by agricultural land. This result is not surprising, as agricultural land tends to be of higher quality than other land (e.g., forests that tend to cover mountainous areas), which makes it more profitable for local governments to occupy agricultural land for urban development.[13]

**Table 3 The Effects of Urban Expansion on Agricultural Land**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. Var. | Land Quantity | Land Quality | Land Quantity | Land Quality |
| ***Panel A. OLS estimates*** | | | | |
| Urban Expansion | -0.929*** | 0.048*** | -1.131*** | 0.037** |
| | (0.145) | (0.015) | (0.168) | (0.015) |
| Control Variables | No | No | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Observations | 57540 | 57540 | 53560 | 53560 |
| ***Panel B. IV estimates*** | | | | |
| Urban Expansion | -0.786*** | 0.228*** | -1.372** | 0.149** |
| | (0.266) | (0.062) | (0.536) | (0.076) |
| Control Variables | No | No | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Observations | 53580 | 53580 | 53560 | 53560 |
| KP-F statistics | 33.856 | 33.856 | 15.038 | 15.038 |

*Notes*: This table presents the results on the effects of urban expansion on the quality/quantity of agricultural land, with Panel A presenting the OLS estimates and Panel B the IV estimates. The endogenous variable is the urban built-up area, instrumented by the interaction between initial county GDP and national road investment. The observation is at the county-year level. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the county level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

### 3.2.2 Robustness

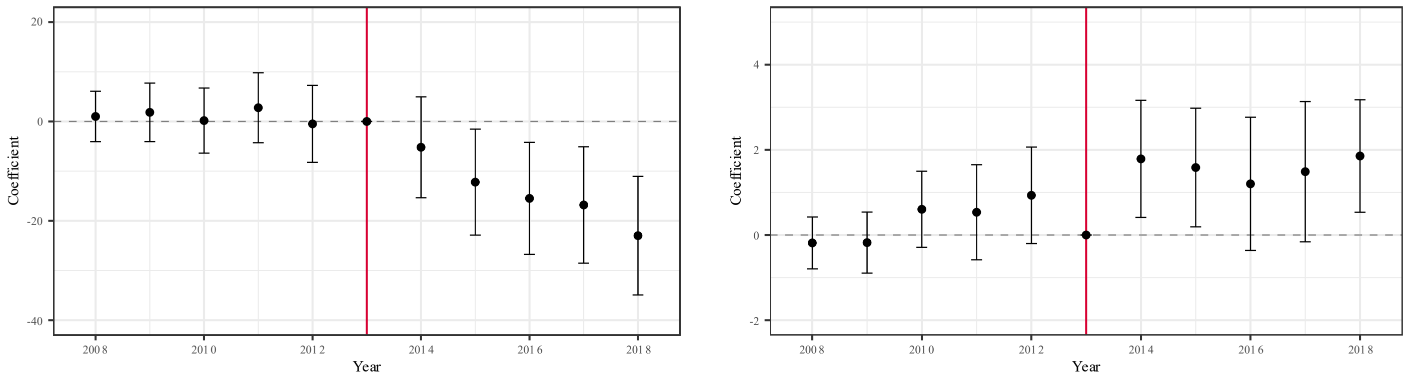
We now briefly discuss the robustness of our baseline results. Specifically, we show that our results are

[13] Our theoretical model presented in the next section rationalizes this result.

robust to exploit an alternative instrument, which additionally allows us to examine the exogeneity assumption by using an event study specification to test the parallel trend assumption (Borusyak et al., 2025). We also examine the sensitivity of the exclusion restriction and provide evidence that our estimated effects suffer less concern of the exclusion restriction assumption. Lastly, we show that our results are robust to alternative model specifications (e.g., controlling for alternative fixed effects and adjusting the clustering levels).

**Alternative Instrument.** We first exploit an alternative instrument described in Section 3.1, i.e., the interaction between two dummies indicating whether the county's GDP is above the national average and whether the time period is post-2013. As shown in Figure 2, the effects of large-scale road construction on urban expansion are insignificant prior to 2013. If our instrument indeed captures the variations in aggregate infrastructure investment changes and the shares (initial GDP) are as good as randomly assigned, then we should find only small effects in reduced-form analyses before 2013 as well. Figure 3 presents the reduced form of event study regressions using the specification outlined in Equation (4), with the dependent variables replaced by agricultural land quantity (Panel A) and quality (Panel B). Aligning with our expectations, we find statistically insignificant estimates prior to 2013 and larger effects on agricultural land quantity/quality after 2013. This piece of evidence again confirms that our shift-share design can effectively identify the causal impacts of urban expansion on the quantity and quality changes of agricultural land. In Appendix Table A2, we report the estimated coefficients from the reduced-form and IV regressions with the DiD term $1(HighGDP)_i \times 1(Post2013)_t$ serving as the alternative instrument. Again, we find strong evidence that urban expansion leads to significant decreases in both the quality and quantity of agricultural land.

**Exclusion Restriction.** We then examine the sensitivity of the exclusion restriction assumption. The exclusion restriction assumption requires that our IV should affect agricultural land only through urban expansion but not through other channels. To provide support, we first perform a simple exercise by directly controlling for the instrument in the OLS specification. If the exclusion restriction assumption holds, then the effects of the instrument on dependent variables should be muted, conditional on the level of urban expansion. Results from Appendix Table A3 show that, additionally controlling for the instrument does not have independent effects on the quantity/quality of agricultural land, supporting the exclusion restriction assumption. We should note that, however, one caveat of such specification is that including the instrument in the OLS specification could exacerbate the potential endogeneity. Therefore, the estimated results from Table A3 should be interpreted with caution.




Panel A Land Quantity

Panel B Land Quality

**Figure 3 Event Study Estimates for Agricultural Land Quantity/Quality**

*Notes*: This figure presents the event study coefficients. Regression specification is presented in Equation (4), with the dependent variables replaced by agricultural land quantity (Panel A) and quality (Panel B). The year 2013 is omitted as the reference year. The specification includes county and year fixed effects. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the county level. Coefficient estimates and 95% confidence intervals are shown in the figure.

To lend further credit to the exclusion restriction assumption, we next directly investigate whether our instrument would affect other potential channels that are correlated with agricultural land. One concern is that our instrument could operate through changes in rural-urban migration and employment, as the construction of roads could stimulate local employment (Chatterjee et al., 2025; Chaurey and Le, 2022). In Appendix Table A4, we regress a set of population and employment indicators on the instrument to examine whether they are affected by road investment. Reassuringly, we find that the instrument is not correlated with any of the employment or population changes (e.g., agricultural and industrial employment, rural and urban population), which suggests that our instrument is unlikely to affect the quantity/quality of agricultural land through population changes.

Another concern is that the instrument may directly affect agricultural land quantity and quality. For instance, the investment in road infrastructure could also directly affect agricultural activities by, for example, improving the market access of agricultural goods (Adamopoulos, 2025; Chan, 2022; Sotelo, 2020), which could have an impact on agricultural land. Relatedly, a county's initial GDP could also be correlated with historical agricultural or geographic endowment (Dall Schmidt et al., 2018; Litina, 2016), which may also indirectly affect current land quality and quantity. To ensure that our IV estimates are not driven by the violation of exclusion restriction, in Appendix Table A5, we additionally control for a set of variables through which our instrument could affect agricultural land. Typically, we consider time-varying agricultural output, agricultural firm entry,[14] and exogenous predetermined agricultural productivity (measured by the Global Agro-Ecological Zones (GAEZ) model).[15] The last variable is interacted with full sets of year dummies to allow for time-varying effects. Our estimated effects remain significant and qualitatively the same after the inclusion of these additional controls, again suggesting that the instrument is not likely to affect changes in agricultural land through channels other than urban expansion.

Our final approach to examine the exclusion restriction assumption is to use the "Plausibly Exogenous" estimation developed by Conley, Hansen, and Rossi (2012), which allows for an IV estimation with a certain degree of violation of the exclusion restriction. That is, we allow the IV to enter the second stage linearly with coefficient $\gamma$.[16] In conventional IV estimates, the exclusion restriction assumption explicitly assumes that $\gamma$. is zero in the second stage, but by assigning different values to $\gamma$, we are able to evaluate how the 2SLS coefficient is affected, which allows for an inspection of the sensitivity of our estimates to the potential violation of exclusion restriction. To choose suitable values of $\gamma$, we rely on estimates from Appendix Table A3, where we directly include the instrument in the OLS regression. Specifically, point estimates in Table A3 suggest that the direct effect of the instrument should be no larger than 0.004 and no smaller than -0.009. We therefore set the maximum and the minimum of possible $\gamma$ values to be 0.01 and -0.01, respectively. Figure 4 visualizes the coefficients from the 2SLS estimations by allowing $\gamma$ to take different values

[14] The data on the county-level agricultural firm entry is calculated using the Business Enterprise Registration Dataset, which is derived from the National Enterprise Credit Information Publicity System, https://www.gsxt.gov.cn/index.html8.

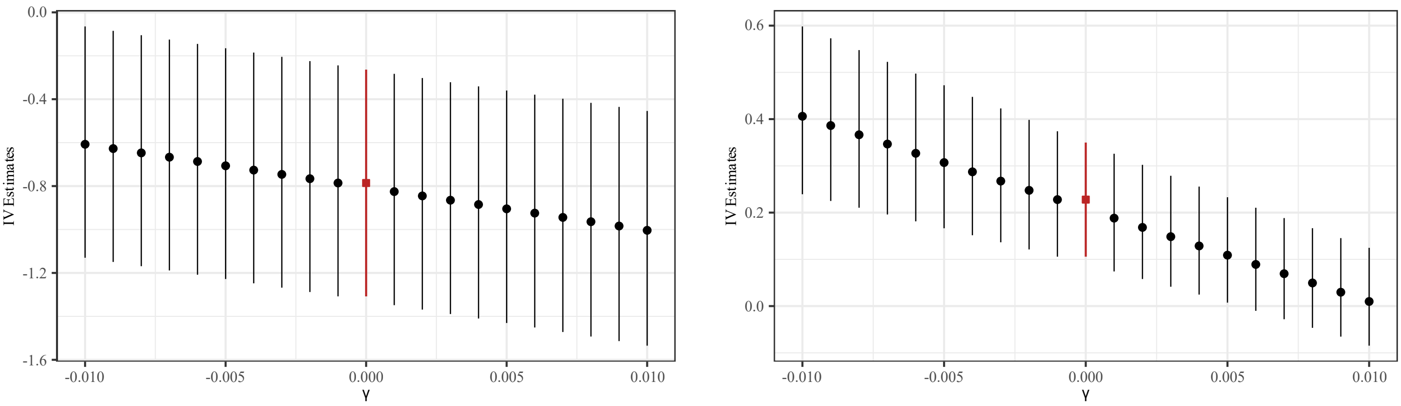
[15] The data is available at https://www.resdc.cn/data.aspx?DATAID=261.

[16] Following Conley, Hansen, and Rossi (2012), the specification here is:

$$y_{it} = \beta UE_{it} + \gamma GDP_i \times Infras_{t-1} + \theta X_i \times T + \delta_i + \zeta_t + \epsilon_{it}.$$

Where the interaction between the county's initial GDP and national road investment, as the instrument, enters in the second stage linearly with coefficient $\gamma$. All other variables share the same definition as in Equation (1).

between -0.01 and 0.01, where the estimates colored with red represent the conventional 2SLS coefficients.[17] Our results in Figure 4 indicate that even allowing for a certain degree of violation of the exclusion restriction, the 2SLS regression can still confirm the negative effects of urban expansion on the quantity and quality of agricultural land.




Panel A Land Quantity    Panel B Land Quality

**Figure 4 2SLS Estimates for Plausibly Exogenous IV**

*Notes*: This figure presents the coefficients from the 2SLS estimation by allowing for plausibly exogenous IV, with the dependent variables being agricultural land quantity (Panel A) and quality (Panel B). We allow for the coefficient $\gamma$ to take different values between -0.1 and 0.1, which covers the point estimates from Appendix Table A3. The estimates colored with red are the conventional 2SLS regression coefficients ($\gamma = 0$). The specification includes county and year fixed effects. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the county level. Coefficient estimates and 95% confidence intervals are shown in the figure.

**Alternative Model Specifications.** We also provide supportive evidence that our baseline estimates are robust to alternative model specifications. Typically, in Appendix Table A6, we augment our baseline specification by additionally including province-specific year trends and province-year fixed effects to account for unobserved aggregate shocks specific to each province. Though the inclusion of these additional trends and fixed effects absorb substantial variations in the first-stage regressions (as the KP-F Statistics reduce), the estimated effects remain significant at the conventional level, and the magnitude is still comparable to the baseline coefficients. This alleviates concerns that our results may be confounded by unobserved provincial heterogeneities.

Since the variation of our outcome variables is primarily driven by changes in land use patterns, which may have inter-temporal correlations (for instance, the conversion of agricultural land for urban construction may take years to complete, while the decision could be made years ago), the quantity/quality of agricultural land and urban expansion may be affected by past outcomes. Although the exploitation of an instrumental strategy can largely alleviate such concerns, it is still possible that past outcomes could be correlated with the county's initial economic development. To ensure that this is not the case, in columns (1) and (2) of Appendix Table A7, we additionally control for the lagged quantity/quality of agricultural land as a control variable. Again, we find that our results are robust to the inclusion of the lagged dependent variable.[18]

Our instrument exploits variations in lagged national-level road investment for identification, as the

[17] The analyses are conducted with the Stata command "plausexog", based on the approach of the "union of confidence interval."

[18] The results, however, should be interpreted with caution, as the inclusion of lagged dependent variable in panel regressions can lead to the so-called "Nickell Bias" (Nickell, 1981).

effects of road investment on urban expansion may also have inter-temporal correlations. Therefore, the instrument may affect future urban expansion, which could be correlated with past agricultural land conversion. To alleviate such concerns, in columns (3) and (4) of Appendix Table A7, we use the lagged term of the instrument as an alternative IV and re-estimate the 2SLS regression. We find that the estimated effects are similar to the baseline estimates in terms of both magnitude and significance, again providing valid support to our identification strategy.

Finally, we explore whether our estimates are sensitive to different clustering levels. Specifically, in columns (1) and (2) of Appendix Table A8, we two-way cluster the standard errors at the county and province-year level that account for both within-county temporal correlations and within-province cross-sectional correlations of the error term, whereas in columns (3) and (4), we cluster the standard errors at the city level to allow for any arbitrary correlations within the city, as the land use pattern may be correlated within the same city. As expected, the estimates become more imprecise when allowing for more dimensions of correlations in the error term, but the coefficients remain significant at the 10% level.

### 3.2.3 Mechanism

After establishing the robust causal relations between urban expansion and declines in agricultural land quality/quantity, we now investigate the potential mechanisms. Since we have shown that urban expansion is primarily through the occupation of agricultural land but not through other types of land, there is no need to discuss how urban expansion decreases agricultural land coverage. We thus exclusively focus on the mechanism of the decline in agricultural land quality. Intuitively, as the supply of high-quality land is inelastic, to decrease agricultural land quality, urban expansion must occupy much of the high-quality agricultural land. To verify this channel, we leverage data from land transactions, which record detailed information on the transacted land parcels, including the exact location of the parcel, transaction price, and the industry to which it was transacted (see Appendix B for more detailed descriptions).[19]

We first geocode all land parcels to acquire their longitude and latitude based on their detailed location. Secondly, we match all parcels to our 30m terrain ruggedness grid cells and assign the terrain ruggedness of the grid cells to each parcel. Next, we combine the land cover dataset to determine whether the land parcel was agricultural land before being transacted. Finally, we distinguish whether the land is transacted for industrial, residential/commercial, or other usage and calculate the ratio of the weighted average terrain ruggedness between agricultural land and non-agricultural land, within each county-year cell for all different usages.[20] A lower terrain ruggedness ratio implies that the transacted agricultural land is more flattened than non-agricultural land and that the transacted agricultural land is of higher quality.

Column (1) of Table 4 reports the OLS and IV estimates with the dependent variable replaced by the ratio of terrain ruggedness. We find strong evidence that urban expansion leads to more occupation of high-quality agricultural land relative to non-agricultural land. This suggests that quality-biased land occupation amid urbanization is a key contributor to the declining quality of agricultural land. In columns (2) to (4), we investigate the potential heterogeneity in land occupation across different land use types. We show that the quality-biased agricultural land occupation is more pronounced for industrial land parcels and residential/commercial land parcels. The effects are much smaller if the land parcel is occupied for other uses (e.g., public infrastructure). These patterns further suggest that the occupation of high-quality agricultural land is for promoting the growth of the urban economy.

---

[19] The data is derived from the China Land Market Website at https://www.landchina.com/#/. The data is widely used for analyzing China's land market; see, for example, Han and Kung (2015) and Chen and Kung (2016; 2019).

[20] We use the land area of the transacted parcel as the weight.

**Table 4 Urban Expansion and Land Transaction**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. Var. | The Ratio of Terrain Ruggedness between Agricultural and Non-agricultural Land | | | |
| | All | Industrial | Residential/Commercial | Other |
| ***Panel A. OLS estimates*** | | | | |
| Urban Expansion | -0.767*** | -0.346*** | -0.333*** | -0.088*** |
| | (0.112) | (0.054) | (0.058) | (0.018) |
| Adjusted R-squared | 0.508 | 0.418 | 0.499 | 0.442 |
| ***Panel B. IV estimates*** | | | | |
| Urban Expansion | -3.094*** | -1.235*** | -1.684*** | -0.175*** |
| | (0.722) | (0.252) | (0.468) | (0.063) |
| Control Variables | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Observations | 53560 | 53560 | 53500 | 53500 |
| F-statistic (first stage) | 15.038 | 15.038 | 15.038 | 15.038 |

*Notes*: This table presents the results on the effects of urban expansion on the occupation pattern of agricultural land, with Panel A presenting the OLS estimates and Panel B the IV estimates. The dependent variable is the ratio of terrain ruggedness between agricultural and non-agricultural land, calculated from the land transaction data. Column (1) focuses on all transacted land parcels, while in the remaining three columns, we focus on parcels that are transacted for industrial, residential/commercial, and other usages. The endogenous variable is the urban built-up area, instrumented by the interaction between initial county GDP and national road investment. The observation is at the county-year level. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the county level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

# 4. Theoretical Model

Given that urban expansion can lead to substantial declines in agricultural land productivity (measured by both quality and quantity), we now explore whether the implementation of the Agricultural Land Protection Policy (ALPP) can mitigate such adverse effects. To guide our empirical analysis, we first develop a simple theoretical model that incorporates the policy constraints into the urban expansion decision of local government. We then derive several comparative statics, focusing on the role of political incentives, for empirical tests.

## 4.1 Set up

Our theoretical model is a simple two-sector model that characterizes the effects of urban expansion on agricultural land productivity. We discuss the issue under the scenario that the central government requires a balance between occupying and compensating for agricultural land. Our model assumes a monocentric city with the linear form as in Duranton and Puga (2015) and Wang, Zhang, and Zhou (2020), where we extend the model by considering the land quality (both urban land and agricultural land) in our model and consider how the decision of outward expansion made by the local leader would result in differential productivity and welfare consequences.

Specifically, we divide the real line into three parts: urban, rural, and undeveloped regions. Within the urban and rural regions, land productivity decreases as the distance to the city center increases (for the urban region) or the rural-urban fringe (for the rural region, as in Oueslati, Salanié, and Wu (2019). Figure 5 delivers a visual representation, where $S_0$ and $S_1$ are the boundaries of urban and rural areas, respectively. The urban region lies in the line between 0 and $S_0$, and the rural region lies in the line between $S_0$ and $S_1$.

Beyond $S_1$, it's undeveloped land.

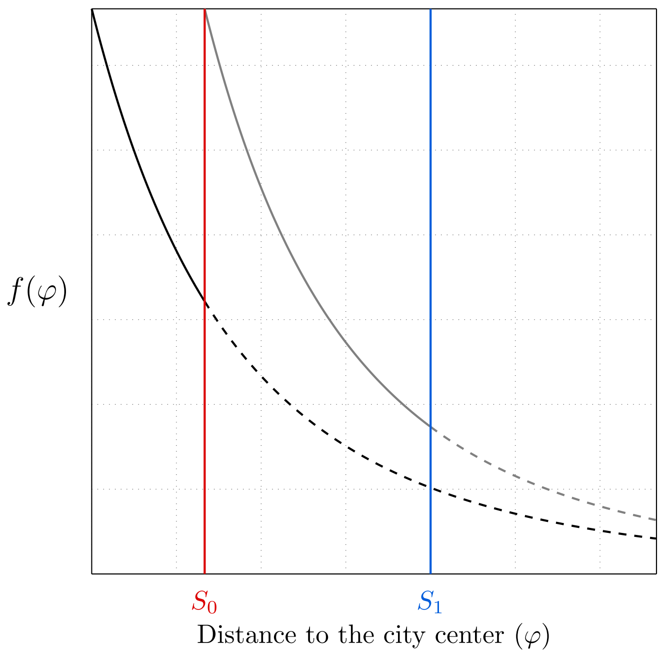
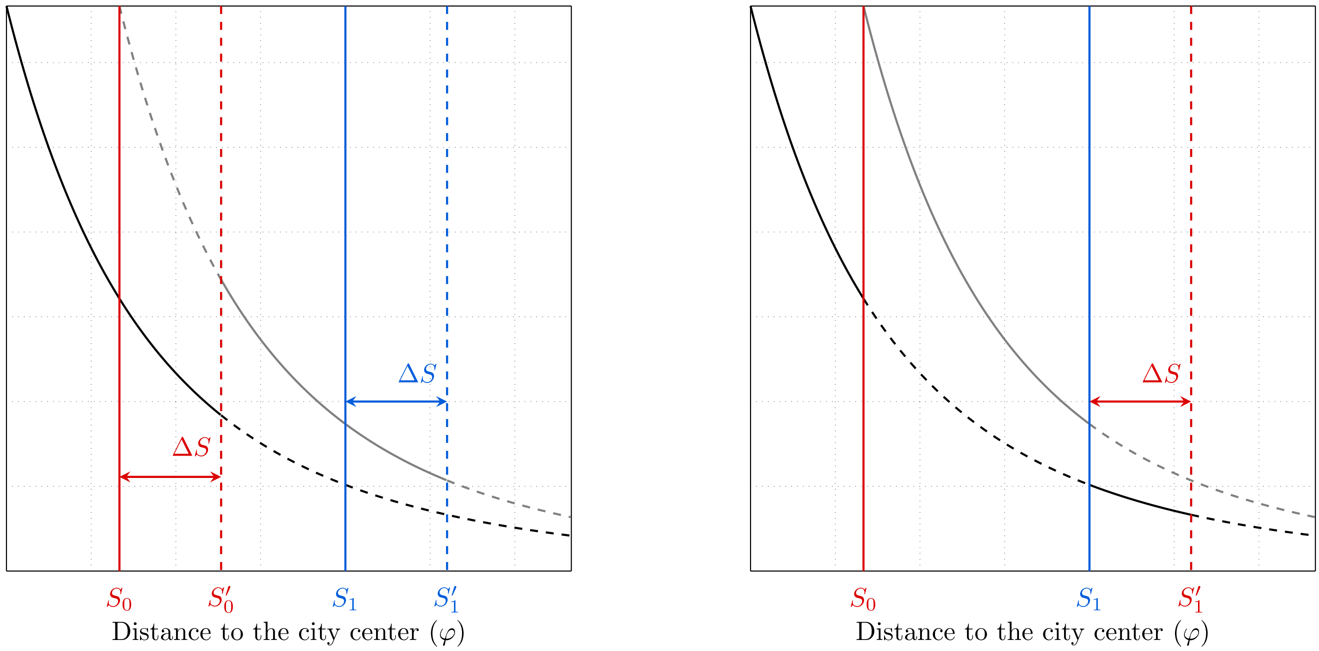



**Figure 5 Graphical Illustration of Land Productivity Distribution**

*Notes*: This figure presents a graphical illustration of land productivity distribution within the monocentric city structure, featuring both urban and rural areas, where $S_0$ and $S_1$ are the boundaries of urban and rural areas.

Given the city structure, the city leader now wants to expand the urban region by $\Delta S$ and decides where the newly constructed urban area should take place (i.e., on agricultural land or undeveloped land). Under the case where the occupation of agricultural land is strictly compensated, if the city leader decides to occupy the agricultural land, then it has to compensate for the same amount of land created in undeveloped land. Otherwise, the city leader occupies undeveloped land to construct new urban areas and keep agricultural land intact. This can be visualized in Figure 6. In Panel A, we plot the case when occupying agricultural land. Outward expansion by $\Delta S$ requires the local government to compensate another $\Delta S$ amount of agricultural land from undeveloped land. This leads to higher urban land productivity but potentially reduces agricultural land productivity. In another case, the local government instead decides to occupy the undeveloped land and keep the amount of agricultural land unchanged, which is shown in panel B of Figure 6. In this case, the outward expansion leads to lower urban land productivity but maintains high agricultural land productivity. This introduces trade-offs between occupying agricultural land and undeveloped land in the decision of outward expansion. We now present the formal model to characterize such trade-offs and explore how changes in key parameters (e.g., enforcement stringency and political incentives) can alter the behavioral change of local officials and how these changes affect agricultural land quantity and quality.




Panel A Occupy Agricultural Land

Panel B Occupy Undeveloped Land

**Figure 6 Urban Expansion Under APPL**

*Notes*: This figure presents how urban and agricultural land productivity changes under different occupation schemes. Panel A illustrates the scenario where the local leader decides to occupy agricultural land, whereas Panel B illustrates the scenario where the local leader decides to occupy undeveloped land.

### 4.1.1 Preference

The representative agent consumes agricultural goods, $Y_a$, and manufacturing goods, $Y_m$, with the following Cobb-Douglas utility function:

$$U_c = \alpha^{-\alpha}(1-\alpha)^{1-\alpha} Y_a(\varphi)^{\alpha} Y_m(\varphi)^{1-\alpha},$$

with the budget constraint given by

$$w = P_a Y_a(\varphi) + P_m Y_m(\varphi)$$

where the price of agricultural goods and manufacturing goods are given by $P_a$ and $P_m$, respectively. The term $\alpha^{-\alpha}(1-\alpha)^{1-\alpha}$ is a convenient normalization. We assume both rural and urban regions are endowed with one unit of immobile labor, and labor supply is inelastic. Solving the consumer problem, we can derive the indirect utility $V$ as

$$V = \frac{w}{P_a^{\alpha} P_m^{1-\alpha}}$$

### 4.1.2 Production

The production function of manufacturing goods is given by

$$Y_m = z_m(\varphi) L_m,$$

where $z_m$ is the productivity, which is jointly decided by the amount of land and quality of land supplied to the manufacturing firms. We assume $z_m = \int_0^{S_0} \varphi f_m(\varphi)\, d\varphi$, where $f_m(\varphi)$ is the probability density function of the Pareto distribution with the following form:

$$f_m(\varphi) = \theta A_m^{\theta} \varphi^{-\theta-1}, \qquad \varphi \geq A_m$$

Similarly, the production function of agricultural goods is given by

$$Y_a = z_a(\varphi)L_a,$$

where $z_a$ is the agricultural productivity, which we write as $z_a = \int_{S_0}^{S_1} \varphi f_a(\varphi)d\varphi$, and we have

$$f_a(\varphi) = \theta A_a^\theta \varphi^{-\theta-1}, \qquad \varphi \geq A_a$$

For simplicity, we assume the shape parameter, $\theta$, is the same for $f_a(\varphi)$ and $f_m(\varphi)$.[21] By construction, we have $S_0 = A_a$, and we denote $S_1$ as $A_u$ in our subsequent analyses for the ease of exposition.

Given the distribution of land productivity, we can derive $z_m$ and $z_a$ as functions of boundaries, that is,

$$z_m = \int_{A_m}^{A_a} \varphi\theta A_m^\theta \varphi^{-\theta-1}d\varphi = \frac{\theta}{\theta-1}A_m\left[1-\left(\frac{A_m}{A_a}\right)^{\theta-1}\right]. \tag{5}$$

And

$$z_a = \int_{A_a}^{A_u} \varphi\theta A_a^\theta \varphi^{-\theta-1}d\varphi = \frac{\theta}{\theta-1}A_a\left[1-\left(\frac{A_a}{A_u}\right)^{\theta-1}\right]. \tag{6}$$

Combining (5) and (6), we have the relative productivity as

$$\frac{z_a}{z_m} = \left(\frac{A_a}{A_m}\right)\left(\frac{1-\left(\frac{A_a}{A_u}\right)^{\theta-1}}{1-\left(\frac{A_m}{A_a}\right)^{\theta-1}}\right).$$

We assume $z_a/z_m < 1$ as the productivity of urban land is generally higher than that of agricultural land. Finally, under perfect competition, we can write the price of goods with respect to wage and productivity

$$P_a = \frac{w}{z_a(\varphi)} \qquad P_m = \frac{w}{z_m(\varphi)}$$

Substituting into the indirect utility function, we derive

$$V = z_a(\varphi)^\alpha z_m(\varphi)^{1-\alpha}.$$

### 4.1.3 Local government

The utility of local government, $G$ is comprised of two parts: the growth of urban economy $Y_m$ and the utility of residents $V$, where the relative weight on urban economic growth is $\gamma$, that is, $G = \gamma Y_m + (1-\gamma)V$. Given the above city structure, the local government now decides to expand the urban region by $\Delta S$. The occupation of agricultural land and undeveloped land involves occupation cost, which we denote as $c_a$ and $c_u$, respectively.[22] Since productivity is generally higher in cropland than in undeveloped land, we assume $c_u > c_a$. That is, the cost of transforming undeveloped land into urban land is higher than the cost of converting agricultural land. We analyze how the decision of whether to occupy agricultural land or undeveloped land changes with the incentive of promoting economic growth ($\gamma$) and the relative cost of occupying different types of land.

We first consider the case of occupying cropland. Note that in equilibrium, the change in manufacturing output and agricultural output is solely decided by the change in productivity (i.e., $\Delta Y_m = \Delta z_m$ and $\Delta Y_a = \Delta z_a$). We therefore have

$$\Delta z_m = \int_{S_0}^{S_0+\Delta S} \varphi\theta A_m^\theta \varphi^{-\theta-1}d\varphi = \frac{\theta}{\theta-1}A_m^\theta S_0^{1-\theta}\left[1-\left(1+\frac{\Delta S}{S_0}\right)^{1-\theta}\right] \tag{7}$$

Under the Agricultural Land Protection Policy, the change in agricultural productivity is given by

[21] Although allowing for different shape parameters will essentially have little impact on our theoretical derivation but only increase the complexity of the model.

[22] We assume $c_a$ contains both the cost of occupying agricultural land and the cost of compensating cropland.

$$\Delta z_a = \int_{S_1}^{S_1+\Delta S} \varphi\theta A_a^\theta \varphi^{-\theta-1} d\varphi - \int_{S_0}^{S_0+\Delta S} \varphi\theta A_a^\theta \varphi^{-\theta-1} d\varphi,$$

which we can simplify and write this as

$$\Delta z_a = \frac{\theta}{\theta-1} A_a^\theta S_1^{1-\theta}\left[1-\left(1+\frac{\Delta S}{S_1}\right)^{1-\theta}\right] - \frac{\theta}{\theta-1} A_a^\theta S_0^{1-\theta}\left[1-\left(1+\frac{\Delta S}{S_0}\right)^{1-\theta}\right] \quad (8)$$

Applying Taylor's first-order approximation to (7) and (8), we can simplify the change in land productivity as

$$\Delta z_m = \theta A_m^\theta S_0^{-\theta} \Delta S \qquad \Delta z_a = \theta A_a^\theta \left(\delta S_1^{-\theta} - S_0^{-\theta}\right)\Delta S$$

The change in the utility of residents can be expressed as

$$\Delta V = \alpha\left(\frac{z_a}{z_m}\right)^{\alpha-1} \Delta z_a + (1-\alpha)\left(\frac{z_a}{z_m}\right)^{\alpha} \Delta z_m$$

Under ALPP, the change in the utility of local government is

$$\Delta G = \gamma \Delta z_m + (1-\gamma)\Delta V - c_a$$

Which can be written as

$$\Delta G = \theta A_m^\theta \left[\gamma S_0^{-\theta} + (1-\gamma)\alpha\left(\frac{z_a}{z_m}\right)^{\alpha-1}\left(S_1^{-\theta} - S_0^{-\theta}\right) + (1-\gamma)(1-\alpha)\left(\frac{z_a}{z_m}\right)^{\alpha} S_0^{-\theta}\right]\Delta S - c_a.$$

Now, consider the case when the local government instead chooses to occupy undeveloped land. Under such circumstances, we have $\Delta z_a = 0$, and the change in utility is thus

$$\Delta G' = \gamma \Delta z_m + (1-\gamma)\Delta V - c_u = \theta A_m^\theta \left[\gamma S_1^{-\theta} + (1-\gamma)(1-\alpha)\left(\frac{z_a}{z_m}\right)^{\alpha} S_1^{-\theta}\right]\Delta S - c_u$$

Subtract these two terms, we have

$$\Delta\hat{G} \equiv \Delta G - \Delta G' = \theta A_m^\theta \left[\gamma - (1-\gamma)\alpha\left(\frac{z_a}{z_m}\right)^{\alpha-1} + (1-\gamma)(1-\alpha)\left(\frac{z_a}{z_m}\right)^{\alpha}\right]\left(S_0^{-\theta} - S_1^{-\theta}\right)\Delta S - (c_a - c_u)$$

The choice of whether to occupy cropland or undeveloped land is thus decided by the sign of $\Delta\hat{G}$. An immediate conclusion from the above expression is that, when the cost of occupying undeveloped land is higher, for instance, undeveloped land is more rugged and more distant from the city center, then the local government is more likely to occupy agricultural land.

## 4.2 Comparative Statics

We now consider how the change in political incentives affects the decision of the city leader. Take the partial derivative of $\Delta\hat{G}$ with respect to $\gamma$, we have

$$\frac{\partial \Delta\hat{G}}{\partial \gamma} = \theta A_m^\theta \left[1 + \alpha\left(\frac{z_a}{z_m}\right)^{\alpha-1} - (1-\alpha)\left(\frac{z_a}{z_m}\right)^{\alpha}\right]\left(S_0^{-\theta} - S_1^{-\theta}\right)\Delta S > 0$$

This implies that, as the incentives for promoting economic growth increase, local leaders are more willing to occupy agricultural land, which may further decrease its productivity. We can also consider how an increase in agricultural productivity (i.e., $z_a$) can alter the choice of local government. Take the partial derivative with respect to $z_a$, we have

$$\frac{\partial \Delta\hat{G}}{\partial z_a} = \theta\alpha(1-\gamma)(1-\alpha) A_m^\theta \left(z_a^{\alpha-2} z_m^{1-\alpha} + z_a^{\alpha-1} z_m^{-\alpha}\right)\left(S_0^{-\theta} - S_1^{-\theta}\right)\Delta S > 0$$

We find that higher agricultural productivity results in a higher probability of occupying agricultural land. Since occupying agricultural land with higher productivity later translates into higher industrial productivity, it provides sufficient incentives for local government to occupy high-productivity cropland and compensate for land with relatively low productivity.

An analogous expression can be derived for the industrial productivity, which is

$$\frac{\partial \Delta \hat{G}}{\partial z_m} = -\theta\alpha(1-\gamma)(1-\alpha)A_m^\theta\left(z_a^{\alpha-1}z_m^{-\alpha} + z_a^\alpha z_m^{-\alpha-1}\right)\left(S_0^{-\theta} - S_1^{-\theta}\right)\Delta S < 0$$

This implies that cities with higher industrial productivity are less likely to occupy agricultural land for urban construction. The intuition behind this is that the local government needs to balance between urban development and citizens' welfare, as higher industrial productivity guarantees the local government to achieve urban economic growth, which in turn leads to more emphasis on promoting citizens' welfare, through which the local government can optimize their utility level.

# 5. The Effects of Agricultural Land Protection Policy

## 5.1 Specification

We now empirically investigate whether the implementation of the ALPP can effectively prevent the further decline in agricultural land quantity and quality. Specifically, we exploit the difference in agricultural land protection targets before and after 2010 to construct each city's exposure intensity to the ALPP. Our specification takes the following form:

$$y_{ict} = \alpha + \beta ALPPTarget_{ct} + \gamma X_i \times T + \delta_i + \zeta_t + \epsilon_{ict} \tag{9}$$

Where $y_{ict}$ is the measure of either agricultural land quality or quantity. $ALPPTarget_{ct}$ is our key independent variable, which represents the differences in agricultural land protection targets and initial agricultural land quantity across cities in different years. Typically, we compute two policy-driven variations: (1) between the mandated 2010 agricultural land quotas and observed 2005 levels, and (2) between the 2020 target quotas and their 2010 counterparts. These sequential differentials serve as time-varying variations in agricultural land preservation pressures, enabling quantification of municipal-level policy exposure intensity across distinct regulatory phases (pre-2010 vs. post-2010).[23] Other variables share the same definition as in Equation (1). Since our exposure is city-specific, we cluster the standard error at the city level to allow for arbitrary correlations within the city.

To further operationalize our identification framework, we implement a normalization procedure whereby we calculate the net policy effect by subtracting the pre-2010 target differential from the post-2010 differential. This adjustment achieves two analytical objectives: first, it establishes the pre-2010 period as the counterfactual baseline (normalized to zero), and second, it generates a continuous treatment intensity metric reflecting quota adjustments attributable to the policy intervention.[24] The resultant empirical design nests in a generalized Difference-in-Differences (DiD) framework, with policy shocks in the year 2010, and the treatment intensity is parameterized as the policy-induced variation in agricultural land allocation target. Formally, the DiD specification is:

$$y_{ict} = \alpha + \beta TargetIntensity_c \times Post2010_t + \gamma X_i \times T + \delta_i + \zeta_t + \epsilon_{ict} \tag{10}$$

Where $TargetIntensity_c$ is the normalized treatment intensity calculated through the procedures described above. $Post2010_t$ is a dummy variable indicating whether the period is after 2010. One additional benefit of using Equation (10) for estimation is that we can perform the event study analysis that examines the treatment effects dynamics and the presence of potential pre-trends, which can provide further support

[23] Since the central government's requirements for agricultural land protection vary from place to place, for example, some cities can continue to occupy agricultural land after 2010, while others need to replenish agricultural land after 2010 in order to fulfill the assessment target. For the convenience of coefficient interpretation, we normalize the explanatory variable by scaling it to a variable that takes values between 0 and 1. The exact formula for normalization is:

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{F2}$$

[24] Similarly, we normalize the treatment intensity following Equation (F2).

to our identification assumption. The event study is specified as follows:

$$y_{ict} = \alpha + \beta_t \sum_{t=2005,t\neq 2009}^{t=2020} TargetIntensity_c \times I_t + \gamma X_i \times T + \delta_i + \zeta_t + \epsilon_{ict} \tag{11}$$

Where we replace the single dummy $Post2010_t$ with a set of year dummies $I_t$, and estimate the coefficients of the treatment intensity at different years. If the parallel trends assumption holds, then we should estimate insignificant effects for $\beta_t$s where $t$ is prior to 2010. Following Marcus and Sant'Anna (2021), the period 2009 and periods before 2005 are set to be the reference periods.

## 5.2 Main results

### 5.2.1 Baseline estimates

Table 5 reports the estimated effects of the ALPP on agricultural land quality, with Panel A reports the coefficients estimated from Equation (9) whereas Panel B reports the coefficients estimated from Equation (10). Column (1) presents the estimated coefficients from the parsimonious specification that only includes county and year fixed effects, whereas in column (2) we include the set of controls to examine the stability and sensitivity of the main coefficients (Oster, 2019). Since the protection quota is mandated by the provincial government, it may result in correlations in protection quotas within the same province. In column (3), we further control for the province-year fixed effects to absorb for any time-varying effects at the province level. Through different specifications, we consistently estimate positive coefficients which are statistically significant at least at the 5% level, indicating that the ALPP leads to a substantial decline in agricultural land quality. Specifically, our estimate in column (3), Panel B of Table 5 suggests that a one unit increase in treatment intensity of the ALPP would result in 11.472 units increase in terrain ruggedness of agricultural land, which corresponds to an 11.9% increase relative to the mean.

**Table 5 The Effects of ALPP on Agricultural Land Quality**

| Dep. Var. Agricultural Land Quality | (1) | (2) | (3) |
|---|---|---|---|
| ***Panel A: OLS specification*** | | | |
| *ALPPTarget* | 12.830*** | 11.106** | 18.221*** |
| | (3.819) | (4.753) | (4.987) |
| Observations | 46,700 | 46,380 | 46,360 |
| Adjusted R-squared | 0.998 | 0.998 | 0.998 |
| ***Panel B: DiD specification*** | | | |
| *TargetIntensity×Post2010* | 7.744*** | 6.643** | 11.472*** |
| | (2.366) | (2.977) | (3.127) |
| County FE | Yes | Yes | Yes |
| Year FE | Yes | Yes | No |
| Province by Year FE | No | No | Yes |
| Controls | No | Yes | Yes |
| Observations | 46,700 | 46,380 | 46,360 |
| Adjusted R-squared | 0.998 | 0.998 | 0.998 |

*Notes*: This table presents the results on the effects of the ALPP on the quality of agricultural land, with Panel A presenting the OLS specification (Equation (9)) and Panel B the DiD specification (Equation (10)). The observation is at the county-year level. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the city level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1%

level.

The causal interpretation of the estimated coefficients requires the parallel trends assumption of the continuous DiD design (Callaway et al., 2024a). Specifically, the underlying identifying assumption of Equation (10) requires that, in the absence of changes in treatment intensity, counties with higher treatment intensity and counties with lower treatment intensity should have similar or parallel trends in agricultural land quality (Callaway et al., 2024b). To provide partial support for the identification assumption, we leverage Equation (11) to examine the presence of pre-trends. To estimated coefficients are presented in Figure 7.

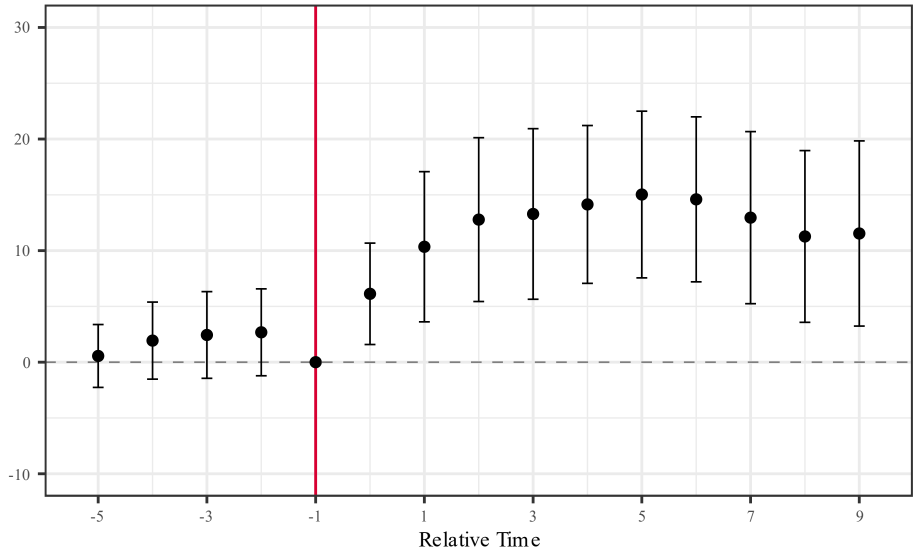



**Figure 7 Event Study Estimates on the Effects of ALPP on Agricultural Land Quality**

*Notes*: This figure presents the event study coefficients. Regression specification is presented in Equation (11), with the dependent variables replaced by agricultural land quality. The year 2009 and years before 2005 are omitted as the reference year (period -1 and periods before -5 in the figure). The specification includes county and year fixed effects. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the city level. Coefficient estimates and 95% confidence intervals are shown in the figure.

We observe two salient features from Figure 7. First, prior to the intensification of agricultural land protection in 2010, counties with higher and lower treatment intensities did not show different trends in agricultural land quality, and none of the estimated coefficients prior to 2010 were significantly different from zero, indicating that there were no significant pre-trends. Second, after the implementation of the ALPP, we see a substantial decline in agricultural land quality, measured by heightened terrain ruggedness. Moreover, the policy effects are observed across the sample periods, suggesting that our estimated effects are not driven by shortened adjustments in agricultural land location, but rather the persistent deterioration of the land quality. Although the decline in land quality is observable throughout the entire post-policy period, we note that the estimated magnitude shows an upward and then a downward trend (an inverted U shape). This may be due to the fact that the central government has strengthened the regulation of agricultural land quality protection after 2015. However, compared to quantity, quality regulation is more difficult to implement and measure, so even with the emphasis on regulating the quality of arable land, we find that the quality of agricultural land still declines.

### 5.2.2 Robustness

This subsection briefly discusses the robustness of our estimation of the effects of the ALPP, with

Appendix Table A9 summarizing the results of our robustness checks.

**Alternative Dependent Variable.** The terrain ruggedness of agricultural land is only one dimension that reflects the quality of the land. To argue that the ALPP has indeed led to a decline in agricultural land quality, we further use the agricultural potential yield (APY) as another measure of land quality. The construction of the measure of APY on agricultural land is similar to the measure of terrain ruggedness. Specifically, we match the land cover dataset with the APY raster data to calculate the average APY on agricultural land for each county.[25] We then re-estimate Equations (9) and (10) with the dependent variable replaced by the APY measure.

The corresponding estimated results are reported in Column (1) of Appendix Table A9, where Panel A reports the coefficients of Equation (9), and Panel B reports the coefficients of Equation (10). It can be seen that the average APY shows a significant decrease after the implementation of ALPP. In particular, our estimated coefficients indicate that for every unit increase in treatment intensity of the ALPP, the average APY of agricultural land decreases by 0.17-0.28 tons per hectare approximately, which implies a 4-7% decrease in APY relative to the mean. This result further confirms our above findings that the strengthening of the regulation of the quantity of agricultural land unexpectedly leads to a decline in land quality.

**Adjusting the Clustering Level.** In our baseline estimation, we cluster the standard error at the city level to control for the within-city correlations in both spatial and temporal dimensions. However, as the protection quotas are specified by the provincial government, there may be concerns that the protection quotas are correlated within the province. To control for such correlation structure in the error term, in column (2) of Appendix Table A9, we adopt the two-way clustering that clusters the standard error at both the city and province-year level. This additionally accounts for the spatial correlations within each province-year cell. In column (3), we further cluster the standard error at the provincial level. To overcome issues related to small clustering size, we adopt the Wild Cluster Bootstrap method (Cameron et al., 2008). Reassuringly, we find no significant changes in the statistical significance of our estimated coefficients.

**Control for Additional Trends/Fixed Effects.** Another concern related to our econometric design is that our estimated effects may be confounded by the differential trends at the city level. Although we have included a set of pre-determined county controls that interacted with year fixed effects in our regression, there could still be cases where our results may be driven by city-level unobserved variations that potentially correlate with the intensity of ALPP. To ensure that this is not the case, in column (4) of Table A9, we further control for city-specific year trends to absorb any time-varying city-level unobservables that follow a linear time trend. The estimated coefficients remain statistically significant, though the estimated magnitude declines to some extent, as the inclusion of city-specific trends absorbs a large fraction of variations in the explanatory variable.

Although our event-study estimates in Figure 7 detect no significant pre-trends, this could also be the case that event-study estimates are of low statistical power, and may not be able to detect differential pre-trends (Roth, 2022). Recent econometric studies on the continuous DiD model find that the model is essentially comparing the difference between groups with above-mean treatment intensity (high treatment group) and groups with below-mean treatment intensity (low treatment group), before and after the policy shock (Callaway et al., 2024a). This provides a basic idea for empirical studies to address the issue of potential undetected pre-trends. Specifically, we categorize the sample into high and low-treatment-intensity groups based on treatment intensity and include the group-year fixed effects to control for possible trend deviation

[25] The data on the agricultural potential yield raster is from the Resource and Environment Science Data Platform, see more detailed description of the dataset at https://www.resdc.cn/DOI/doi.aspx?DOIid=43. The dataset measures the agricultural potential yield for years from 1970 through 2010, at a ten-year interval. To avoid potential endogeneity concerns, we use the agricultural potential yield measured in 2000.

across different groups. The corresponding results are reported in columns (5) of Table A9. We see that the inclusion of additional fixed effects does not significantly change the estimated coefficients, suggesting that our effects are not driven by the potential violation of the parallel trend assumption.

**Control for Other Confounding Policy.** In addition to the ALPP, there are other policies that restrict land supply during our sample periods, which could have an impact on our estimation results. One prominent example is the National Main Functional Zone Planning (NMFZP). The NMFZP was implemented in 2010, which divides the national land into four categories: optimized development zones, key development zones, restricted development zones, and prohibited development zones. Typically, the establishment of prohibited development zones may also constrain the sources of land supply for local governments, making them more likely to resort to occupying agricultural land for urban development, thus leading to a decline in the quality of agricultural land. To rule out the potential effect of this policy, we calculate the area of prohibited development zones for each county in 2010, and then interact it with time fixed effects and include this additional fixed effects in the regression. The estimation results are reported in column (6) of Table A9. We find that the estimated coefficients from the baseline regression remain statistically and economically significant after controlling for the size of prohibited development zones, indicating that our results are not driven by the potential confounding policy.

**Randomized Inference.** Finally, we conduct two randomized inferences as in Figure A1 to ensure that our results are not driven by other unobserved factors that systematically correlate with the protection target. In the left panel of Appendix Figure A2, we randomly assign the time-varying ALPP target to each city, whereas in the right panel of Figure A2, we randomly assign the treatment intensity and interact with the post-2010 dummy. We run each specification 500 times and plot the distribution of the coefficients on the figure. The baseline coefficients are on the far right side of the placebo distribution, suggesting our results are not driven by unobserved random shocks.

### 5.2.3 The effects on land quantity

Our above results consistently show that the implementation of the ALPP leads to a decline in agricultural land quality. A related question is, whether the decrease in land quality is attributed to the newly increased agricultural land, that is, whether the implementation of the ALPP results in increases in agricultural land. To do so, in columns (1) and (2) of Table 6, we replace the dependent variable with the agricultural land quantity. We find no evidence that the ALPP increases agricultural land quantity. The estimated coefficients are insignificant and have a negative sign. This result rules out the possibility that the decline in agricultural land quality is caused by increased land quantity.

One of the main reasons for this result is that, although the ALPP strictly limits the agricultural land targets, local governments have little incentive to protect agricultural land under an incentive scheme that emphasizes economic development as the main goal. Therefore, the ALPP only serves to restrict the expropriation of agricultural land and hardly serves to increase the quantity of agricultural land. To illustrate, we further examine the effect of ALPP on urban land in columns (3) and (4) of Table 6. Again, we do not find that the ALPP has the effect of curbing the expansion of the urban land, suggesting that even under the constraint of protecting agricultural land, the local government is still in pursuit of the rapid development of the urban economy. We will further test the mechanism of political incentives in the subsequent sections.

**Table 6 The Effects of ALPP on Land Quantity**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. Var. | Agricultural Land Quantity | | Urban Land Quantity | |
| *ALPPTarget* | -31.184 | | 3.815 | |

| | | | | |
|---|---|---|---|---|
| | (52.563) | | (5.657) | |
| *TargetIntensity*×*Post2010* | | -18.257 | | 2.383 |
| | | (33.108) | | (3.537) |
| Observations | 46,360 | 46,360 | 46,360 | 46,360 |
| County FE | Yes | Yes | Yes | Yes |
| Province by Year FE | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes |
| Adjusted R-squared | 0.985 | 0.985 | 0.994 | 0.994 |

*Notes*: This table presents the results on the effects of the ALPP on the quantity of agricultural land and urban land, with odd columns presenting the OLS specification (Equation (9)) and even columns the DiD specification (Equation (10)). The observation is at the county-year level. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the city level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

## 5.3 Mechanism and Heterogeneity

We now examine the potential mechanism through which the ALPP leads to decreased agricultural land quality and empirically investigate the propositions we derived from our theoretical model. We first examine the political incentive mechanism. Since the ALPP has not fundamentally changed the incentives and evaluation scheme of local governments, local officials will still focus most of their efforts on promoting local economic growth while resorting to less effort in accomplishing the goals of the protection quota. As predicted by the theoretical model in the previous section, we expect that the greater the incentives for local governments to promote economic growth, the more the quality of agricultural land will decline.

As incentives to promote economic growth are largely related to the appraisal and promotion of local officials, we utilize officials' promotion incentives to reflect local governments' incentives to promote economic growth. Specifically, we manually collect biographical information of mayors in prefecture-level cities, including officials' date of birth, gender, education, and work experience. Following the existing literature, since the promotion of officials is largely determined by the age of local officials, and the probability of promotion of officials is significantly different before and after the age of 57, we set the incentives for promotion to be stronger for officials younger than 57, and weaker for officials older than 57 (Axbard and Deng, 2024; Yao and Zhang, 2015).

Table 7 reports the subsample regression where we divide our sample into the high incentive group (city leader whose age is below 57) and the low incentive group (city leader whose age is above 57). We show that the negative effects of the ALPP on agricultural land quality are entirely driven by local officials with higher promotion incentives. For local leaders with lower promotion incentives, we find that the estimated coefficients are small in magnitude and are insignificant. This is in line with our theoretical prediction that the occupation of agricultural land for urban construction is more likely to take place if the city leader has higher political incentives, which would lead to a substantial decline in agricultural land quality. Whereas for city leaders with lower political incentives, since the reward for occupying agricultural land is relatively small, they are more likely to occupy undeveloped land for urban construction, which would have no effect on agricultural land quality.

**Table 7 The Role of Political Incentives**

| Dep. Var. Agricultural Land Quality | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | High Incentive | Low Incentive | High Incentive | Low Incentive |

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| *ALPPTarget* | 15.020*** | 0.095 | | |
| | (4.685) | (7.017) | | |
| *TargetIntensity* × *Post2010* | | | 9.612*** | -1.330 |
| | | | (2.920) | (4.032) |
| Observations | 40,348 | 5,666 | 40,348 | 5,666 |
| County FE | Yes | Yes | Yes | Yes |
| Province by Year FE | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes |
| Adjusted R-squared | 0.998 | 0.996 | 0.998 | 0.996 |

*Notes*: This table presents the mechanism of political incentives, with odd columns presenting the results on the high incentive sample and even columns on the low incentive sample. Columns (1) and (2) present the OLS specification (Equation (9)) and columns (3) and (4) present the DiD specification (Equation (10)). The observation is at the county-year level. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the city level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

We then move on to examine other predictions derived from our theoretical model. First, our model predicts that, when the city has higher agricultural productivity, the local government is more likely to occupy agricultural land for urban development, which decreases agricultural land quality. On the other hand, when the city has higher industrial productivity, the local government is more likely to keep agricultural land intact and occupy undeveloped land for urban construction. The intuition behind this is, that since local governments need to balance the trade-off between local economic growth and residents' welfare, it is more profitable for them to sacrifice some of their residents' welfare to promote faster urban economic growth when agricultural productivity is high, while the opposite is true when industrial productivity is high, and local governments have greater incentives to sacrifice some of their economic growth to ensure that residents' welfare does not fall too fast.

Panel A and Panel B of Figure 8 examine the above predictions. We note that, since there is no suitable measure of agricultural productivity at the county level, we measure it using the per capita rural income. To calculate industrial productivity at the county level, we leverage detailed firm-level data sourced from the National Tax Survey Dataset (see Appendix B for data description). Specifically, we first calculate the total factor productivity (TFP) for each manufacturing firm and then aggregate to the county level by taking the weighted average of firm TFP, with the weight being the industrial output of each firm.[26] To examine how the treatment effect varies with rural income per capita or industrial productivity, we first generate the tercile of the two variables and estimate a quantile regression of our baseline specifications at each tercile. The estimated results in Panels A and B of Figure 8 confirm our model derivation. Specifically, we show that the treatment effect of the ALPP on agricultural land quality increases when the rural income per capita increases, and that the treatment effects are more pronounced in counties with lower industrial productivity.

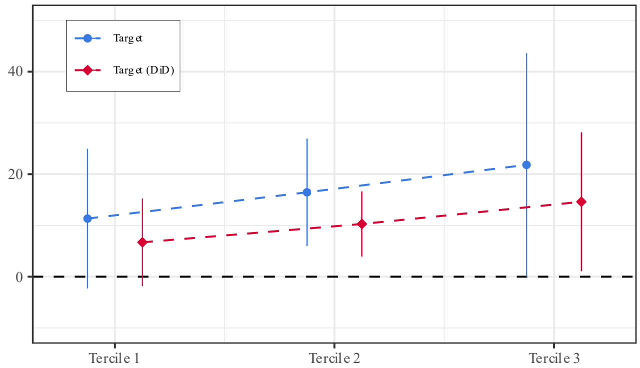
These results implicitly imply misallocations in the promotion-incentive-driven urbanization, where the agricultural land quality declines more in counties with higher agricultural productivity, which suggests substantial efficiency loss. Similarly, as the occupation of agricultural land is more salient in counties with lower industrial productivity, this implies an over-expansion of urban regions, which again is inefficient.

Besides these results, our model also predicts that the decline in agricultural land quality would occur in counties with higher construction costs on undeveloped land. We examine this prediction in Panel C of

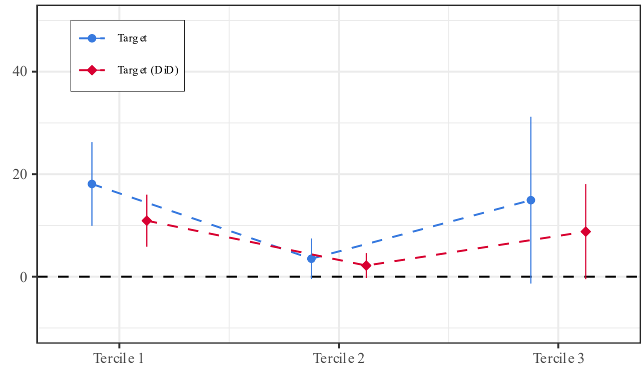
[26] To avoid potential endogeneity, we only exploit ex ante variables. Specifically, we use 2000 rural per capita income to measure agricultural productivity in each county. Since the NTSD were first available in 2007, we use the weighted average productivity in 2007 to measure industrial productivity.

Figure 8. Aligning with our expectation, we find consistent evidence that the decline in agricultural land quality is primarily driven by counties with higher construction costs on undeveloped land.

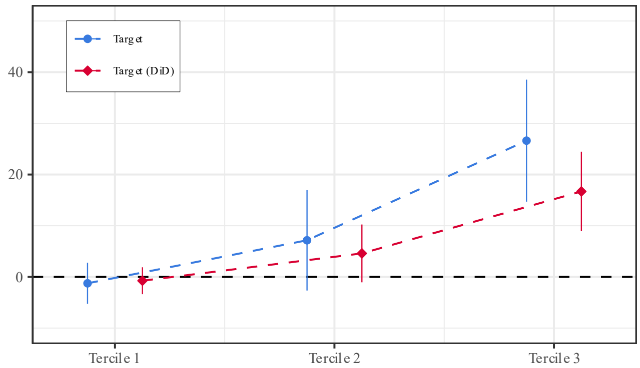
Finally, though not explicitly modeled, an expected result of our model is that the decline in agricultural land quality is more prevalent in counties that face lower policy enforcement intensity. To measure the enforcement intensity of the ALPP, we follow Fan et al. (2023) and use the distance to the provincial capital as an indirect measure of land use supervision intensity. Longer distance implies weaker enforcement intensity. The results from Panel D of Figure 8 reveal that the decline in agricultural land quality is more likely to occur as the distance to the provincial capital increases, suggesting that the lack of effective supervision is another important channel that the ALPP leads to decreased agricultural land quality. Although the ALPP stresses both the protection of agricultural land quality and quantity, as the quality is a hard-to-observe metric, the local government still has sufficient incentives to occupy high-quality land and compensate for low-quality land, which unintendedly results in the decline in land quality.



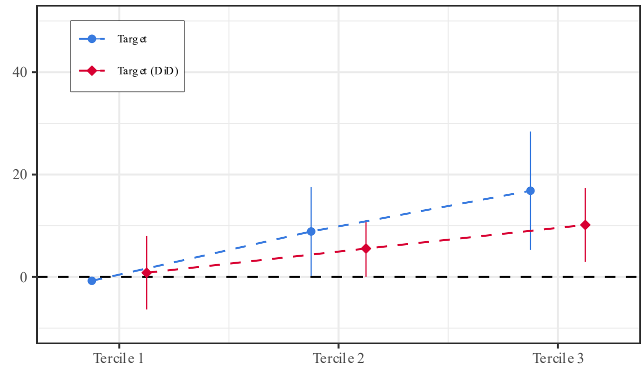

Panel A Rural Income




Panel B Industrial Productivity




Panel C Construction Costs on Undeveloped Land




Panel D Distance to the Provincial Capital

**Figure 8 Heterogeneity in the Effect of ALPP on Agricultural Land Quality**

*Notes*: This figure displays point estimates and the corresponding 95% confidence intervals from the estimation of the quantile treatment effects. Coefficients colored with blue are estimated from Equation (9) while coefficients colored with red are estimated from Equation (10). The x-axis represents the specific variable that is exploited in heterogeneous analysis, while the y-axis denotes the corresponding coefficients. All regressions include county fixed effects and province-by-year fixed effects. Control variables interacted with either a full set of polynomial functions of year trend or year fixed effects. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard errors are clustered at the city level.

## 6. Further Discussion

In the last part of our empirics, we investigate the potential consequences of China's land-based urbanization. Given that we have shown that urban expansion negatively affects agricultural land quality, we ask whether this negative effect would threaten the development and sustainability of agricultural production. Theoretically, urban expansion could have multiple impacts on agricultural production. The first channel is its direct impact on agricultural productivity through the reduction of land quality. Moreover, since our measure of land quality is terrain ruggedness, urban expansion may also affect factor input adjustment. A closely related input is physical capital such as mechanization. As terrain ruggedness increases, agricultural land may be less suitable for mechanization, hindering the efficiency of agricultural production. Finally, as land productivity decreases, farmers may adjust to apply more fertilizer and pesticide input, reducing the sustainability of agricultural production.

In Table 8, we examine the impacts of urban expansion on agricultural production and input adjustment. In columns (1) to (3), we report the OLS estimates, whereas in columns (4) to (6), we report the IV estimates to avoid potential endogeneity. The dependent variable in columns (1) and (4) is county's grain output, in columns (2) and (5), its agricultural machinery power, and in columns (3) and (6), its pesticide and fertilizer usage.[27] Our results indicate that urban expansion significantly reduces the grain output and agricultural machinery power within the county, while increasing pesticide and fertilizer input.[28] Taken together, these results highlight the potential consequences of urban expansion on the development and sustainability of agricultural production.

**Table 8 The Effects of Urban Expansion on Agricultural Production and Input Adjustment**

| | OLS estimates | | | IV estimates | | |
|---|---|---|---|---|---|---|
| Dep. Var | Grain Output | Machinery Power | Pesticide & Fertilizer Usage | Grain Output | Machinery Power | Pesticide & Fertilizer Usage |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Urban Expansion | -0.006*** | -0.008*** | 0.480*** | -0.023*** | -0.017*** | 1.130*** |
| | (0.001) | (0.001) | (0.061) | (0.005) | (0.003) | (0.220) |
| Control Variables | Yes | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted R-squared | 0.951 | 0.924 | 0.945 | | | |
| KP-F statistics | | | | 15.038 | 14.635 | 15.006 |
| Observations | 53537 | 53219 | 53077 | 53537 | 53219 | 53077 |

*Notes*: This table presents the results on the effects of urban expansion on agricultural production and input adjustment, with columns (1) to (3) presenting the OLS estimates and columns (4) to (6) the IV estimates. The endogenous variable is the urban built-up area, instrumented by the interaction between initial county GDP and national road investment. The observation is at the county-year level. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the county level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

[27] Due to the lack of direct statistical data on pesticide and fertilizer use at the county level, we use the proportion of total grain production at the county level to total grain production at the city level to disaggregate city-level pesticide and fertilizer use for the calculation. The city-level pesticide and fertilizer usage are derived from the City Statistical Yearbooks.

[28] To ensure that our results are not driven by the mechanical substitution between agricultural land and urban land, in Appendix Table A10, we examine the robustness of our estimates by directly controlling for the agricultural land area. The results remain significant.

# 7. Conclusion

This study provides new insights into the unintended consequences of urban expansion and farmland protection policies on agricultural land productivity. By integrating a theoretical model with empirical evidence, we demonstrate that urban expansion not only reduces the quantity of agricultural land but also degrades its quality. This process is driven by the selective expropriation of high-quality farmland for urban development, which forces agricultural production onto less productive land with greater terrain ruggedness and lower suitability for modern farming practices.

A key contribution of this study is the examination of the Agricultural Land Protection Policy (ALPP) and its effects on land quality. While the ALPP was designed to reduce farmland loss by imposing strict quantity-based conservation targets, our findings suggest that such policies, when implemented without strong quality considerations, may lead to unintended distortions. Specifically, local political leaders facing strong promotion incentives may comply with the policy in a way that preserves the required amount of farmland but shifts agricultural activity onto lower-quality land. This strategic response aligns with our theoretical predictions and highlights the risks of rigid land conservation policies that fail to consider variations in land productivity.

The broader implications of our findings extend beyond China. Many developing economies face similar tensions between urban expansion and agricultural sustainability, particularly in contexts where decentralized governance structures create incentives for local leaders to prioritize economic growth. Our results suggest that land protection policies must go beyond simple quantity-based restrictions and incorporate mechanisms to safeguard high-quality agricultural land. Potential policy solutions include differentiated farmland protection strategies that account for land fertility and productivity, financial incentives for local governments to preserve high-quality land, and more flexible land-use regulations that promote sustainable urban expansion.

Additionally, our findings indicate that urban expansion has broader consequences for agricultural modernization. The loss of high-quality farmland reduces the adoption of advanced agricultural technologies, such as mechanization, while increasing reliance on fertilizers to compensate for declining land productivity. These shifts could have long-term implications for food security and environmental sustainability, underscoring the need for a more integrated approach to land-use planning.

In conclusion, this study highlights the complex trade-offs between urban expansion, land protection policies, and agricultural productivity. While urbanization is a key driver of economic growth, its impact on agricultural land quality must be carefully managed. Future research should explore the role of alternative land-use policies, the long-term effects of farmland degradation, and the potential for technological innovations to mitigate productivity losses in agriculture. By addressing these challenges, policymakers can better balance economic development with the sustainability of agricultural production.

**CRediT authorship contribution statement**

**Hai Hong:** Conceptualization, Methodology, Software, Data curation, Visualization, Writing- Original draft preparation, Validation. **Yi Shi**: Conceptualization, Methodology, Software, Writing- Original draft preparation. **Zenghui Li**: Conceptualization, Methodology, Software, Writing- Original draft preparation. **Kevin Chen**: Conceptualization, Supervision, Writing- Reviewing and Editing. **Yi Shi** serves as a co-corresponding author.

**Acknowledge**

We are thankful for the support from Zhejiang University-International Food Policy Research Institute Center for International Development Study.

**Funding sources**

This work was supported by the National Social Science Fund of China (No. 21&ZD091, No. 23ZDA035); the Humanities and Social Science Fund of the Ministry of Education and China (22JJD790076); and the Agricultural Science and Technology (10-IAED-05-2024; 10-IAED-RY-01-2024).

**Declaration of generative AI and AI-assisted technologies in the writing process**

During the preparation of this work, the authors used ChatGPT to improve the readability of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

**Declaration of interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

# Reference


Abdulai, A., Owusu, V., Goetz, R., 2011. Land tenure differences and investment in land improvement measures: Theoretical and empirical analyses. Journal of Development Economics 96, 66–78. https://doi.org/10.1016/j.jdeveco.2010.08.002

Adamopoulos, T., 2025. Spatial Integration and Agricultural Productivity: Quantifying the Impact of New Roads. American Economic Journal: Macroeconomics 17, 343–378. https://doi.org/10.1257/mac.20200149

Alesina, A., Tabellini, G., 2008. Bureaucrats or politicians? Part II: Multiple policy tasks. Journal of Public Economics 92, 426–447. https://doi.org/10.1016/j.jpubeco.2007.06.004

Alesina, A., Tabellini, G., 2007. Bureaucrats or Politicians? Part I: A Single Policy Task. American Economic Review 97, 169–179. https://doi.org/10.1257/aer.97.1.169

Amundson, R., Berhe, A.A., Hopmans, J.W., Olson, C., Sztein, A.E., Sparks, D.L., 2015. Soil and human security in the 21st century. Science 348, 1261071. https://doi.org/10.1126/science.1261071

Angrist, J.D., Pischke, J.-S., 2008. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press. https://doi.org/10.2307/j.ctvcm4j72

Axbard, S., Deng, Z., 2024. Informed Enforcement: Lessons from Pollution Monitoring in China. American Economic Journal: Applied Economics 16, 213–252. https://doi.org/10.1257/app.20210386

Baker, G.P., 1992. Incentive Contracts and Performance Measurement. Journal of Political Economy 100, 598–614. https://doi.org/10.1086/261831

Besley, T., 1995. Property Rights and Investment Incentives: Theory and Evidence from Ghana. Journal of Political Economy 103, 903–937. https://doi.org/10.1086/262008

Borrelli, P., Robinson, D.A., Fleischer, L.R., Lugato, E., Ballabio, C., Alewell, C., Meusburger, K., Modugno, S., Schütt, B., Ferro, V., Bagarello, V., Oost, K.V., Montanarella, L., Panagos, P., 2017. An assessment of the global impact of 21st century land use change on soil erosion. Nat Commun 8, 2013. https://doi.org/10.1038/s41467-017-02142-7

Borusyak, K., Hull, P., Jaravel, X., 2025. A Practical Guide to Shift-Share Instruments. Journal of Economic Perspectives 39, 181–204. https://doi.org/10.1257/jep.20231370

Borusyak, K., Hull, P., Jaravel, X., 2022. Quasi-Experimental Shift-Share Research Designs. The Review of Economic Studies 89, 181–213. https://doi.org/10.1093/restud/rdab030

Bridges, E.M., Oldeman, L.R., 1999. Global Assessment of Human-Induced Soil Degradation. Arid Soil Research and Rehabilitation 13, 319–325. https://doi.org/10.1080/089030699263212

Brueckner, J.K., 2000. Urban Sprawl: Diagnosis and Remedies. International Regional Science Review 23, 160–171. https://doi.org/10.1177/016001700761012710

Brueckner, J.K., Largey, A.G., 2008. Social interaction and urban sprawl. Journal of Urban Economics 64, 18–34. https://doi.org/10.1016/j.jue.2007.08.002

Burchfield, M., Overman, H.G., Puga, D., Turner, M.A., 2006. Causes of Sprawl: A Portrait from Space. The Quarterly Journal of Economics 121, 587–633. https://doi.org/10.1162/qjec.2006.121.2.587

Cai, H., Wang, Z., Zhang, Q., 2017. To build above the limit? Implementation of land use regulations in urban China. Journal of Urban Economics 98, 223–233. https://doi.org/10.1016/j.jue.2016.03.003

Callaway, B., Goodman-Bacon, A., Sant'Anna, P., 2024a. Difference-in-Differences with a Continuous Treatment. SSRN Journal. https://doi.org/10.2139/ssrn.4716682

Callaway, B., Goodman-Bacon, A., Sant'Anna, P., 2024b. Event-Studies with a Continuous Treatment. SSRN Journal. https://doi.org/10.2139/ssrn.4716683

Cameron, A.C., Gelbach, J.B., Miller, D.L., 2008. Bootstrap-Based Improvements for Inference with Clustered Errors. Review of Economics and Statistics 90, 414–427. https://doi.org/10.1162/rest.90.3.414

Chan, J., 2022. Farming output, concentration, and market access: Evidence from the 19th-century American railroad expansion. Journal of Development Economics 157, 102878. https://doi.org/10.1016/j.jdeveco.2022.102878

Chatterjee, S., Lebesmuehlbacher, T., Narayanan, A., 2025. Road infrastructure and skill premium: Evidence from Indian manufacturing. Journal of Development Economics 174, 103408. https://doi.org/10.1016/j.jdeveco.2024.103408

Chaurey, R., Le, D.T., 2022. Infrastructure maintenance and rural economic activity: Evidence from India. Journal of Public Economics 214, 104725. https://doi.org/10.1016/j.jpubeco.2022.104725

Chen, T., Kung, J.K., 2019. Busting the "Princelings": The Campaign Against Corruption in China's Primary Land Market*. The Quarterly Journal of Economics 134, 185–226. https://doi.org/10.1093/qje/qjy027

Chen, T., Kung, J.K.-S., 2016. Do land revenue windfalls create a political resource curse? Evidence from China. Journal of Development Economics 123, 86–106. https://doi.org/10.1016/j.jdeveco.2016.08.005

Chen, Y.J., Li, P., Lu, Y., 2018. Career concerns and multitasking local bureaucrats: Evidence of a target-based performance evaluation system in China. Journal of Development Economics 133, 84–101. https://doi.org/10.1016/j.jdeveco.2018.02.001

Cheng, M., Du, J., Ye, C., Zhang, Q., 2022. Your misfortune is also mine: Land expropriation, property rights insecurity, and household behaviors in rural China. Journal of Comparative Economics 50, 1068–1086. https://doi.org/10.1016/j.jce.2022.06.002

Conley, T.G., Hansen, C.B., Rossi, P.E., 2012. Plausibly Exogenous. Review of Economics and Statistics 94, 260–272. https://doi.org/10.1162/REST_a_00139

Dall Schmidt, T., Jensen, P.S., Naz, A., 2018. Agricultural productivity and economic development: the contribution of clover to structural transformation in Denmark. J Econ Growth 23, 387–426. https://doi.org/10.1007/s10887-018-9159-1

Deininger, K., Jin, S., 2006. Tenure security and land-related investment: Evidence from Ethiopia. European Economic Review 50, 1245–1277. https://doi.org/10.1016/j.euroecorev.2005.02.001

Deng, X., Huang, J., Rozelle, S., Uchida, E., 2008. Growth, population and industrialization, and urban land expansion of China. Journal of Urban Economics 63, 96–115. https://doi.org/10.1016/j.jue.2006.12.006

Dewatripont, M., Jewitt, I., Tirole, J., 2000. Multitask agency problems: Focus and task clustering. European Economic Review 44, 869–877. https://doi.org/10.1016/S0014-2921(00)00059-3

Dewatripont, M., Jewitt, I., Tirole, J., 1999. The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies. Rev Econ Studies 66, 199–217. https://doi.org/10.1111/1467-937X.00085

Dubois, P., 2002. Moral hazard, land fertility and sharecropping in a rural area of the Philippines. Journal of Development Economics 68, 35–64. https://doi.org/10.1016/S0304-3878(02)00005-6

Duranton, G., Puga, D., 2015. Urban Land Use, in: Handbook of Regional and Urban Economics. Elsevier, pp. 467–560. https://doi.org/10.1016/B978-0-444-59517-1.00008-8

Fallah, B.N., Partridge, M.D., Olfert, M.R., 2011. Urban sprawl and productivity: Evidence from US metropolitan areas. Papers in Regional Science 90, 451–473. https://doi.org/10.1111/j.1435-5957.2010.00330.x

Fan, H., Liu, G., Wang, H., Zhao, X., 2023. Land use supervision and environmental pollution: multitasking bureaucrats and spillovers across regulations. The Journal of Law, Economics, and Organization ewad028. https://doi.org/10.1093/jleo/ewad028

Fretz, S., Parchet, R., Robert-Nicoud, F., 2022. Highways, Market Access and Spatial Sorting. The Economic Journal 132, 1011–1036. https://doi.org/10.1093/ej/ueab070

Glaeser, E.L., Kahn, M.E., 2004. Chapter 56 - Sprawl and Urban Growth, in: Henderson, J.V., Thisse, J.-F. (Eds.), Handbook of Regional and Urban Economics, Cities and Geography. Elsevier, pp. 2481–2527. https://doi.org/10.1016/S1574-0080(04)80013-0

Goldsmith-Pinkham, P., Sorkin, I., Swift, H., 2020. Bartik Instruments: What, When, Why, and How. American Economic Review 110, 2586–2624. https://doi.org/10.1257/aer.20181047

Han, L., Kung, J.K.-S., 2015. Fiscal incentives and policy choices of local governments: Evidence from China. Journal of Development Economics 116, 89–104. https://doi.org/10.1016/j.jdeveco.2015.04.003

Harari M., 2020. Cities in Bad Shape: Urban Geometry in India. Am. Econ. Rev. 110, 2377–2421. https://doi.org/10.1257/aer.20171673

He, C., Liu, Z., Lu, W., Xu, M., 2022. Dataset of urban built-up area in China (1992-2020) V1.0. National Tibetan Plateau Data Center. https://doi.org/10.11888/HumanNat.tpdc.272851

Holmstrom, B., Milgrom, P., 1991. Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. The Journal of Law, Economics, and Organization 7, 24–52. https://doi.org/10.1093/jleo/7.special_issue.24

Hong, H., Huang, Y., 2025. Order! the border: Multitasking, air pollution regulation and local government responses. Journal of Environmental Economics and Management 131, 103135. https://doi.org/10.1016/j.jeem.2025.103135

Hou, L., Liu, P., Tian, X., 2023. Grassland tenure reform and grassland quality in China. American J Agri Economics 105, 1388–1404. https://doi.org/10.1111/ajae.12357

Huang, W., Luo, M., Ta, Y., Wang, B., 2024. Land expropriation, household behaviors, and health outcomes: Evidence from China. Journal of Development Economics 171, 103358. https://doi.org/10.1016/j.jdeveco.2024.103358

Huntington, H., Shenoy, A., 2021. Does insecure land tenure deter investment? Evidence from a randomized controlled trial. Journal of Development Economics 150, 102632. https://doi.org/10.1016/j.jdeveco.2021.102632

Irwin, E.G., Bockstael, N.E., 2007. The evolution of urban sprawl: Evidence of spatial heterogeneity and increasing land fragmentation. Proc. Natl. Acad. Sci. U.S.A. 104, 20672–20677. https://doi.org/10.1073/pnas.0705527105

Kung, J.K., Chen, S., 2011. The Tragedy of the Nomenklatura: Career Incentives and Political Radicalism during China's Great Leap Famine. The American Political Science Review 105, 27–

45.
Levinsohn, J., Petrin, A., 2003. Estimating Production Functions Using Inputs to Control for Unobservables. Rev Econ Studies 70, 317–341. https://doi.org/10.1111/1467-937X.00246
Li, H., Zhou, L.-A., 2005. Political turnover and economic performance: the incentive role of personnel control in China. Journal of Public Economics 89, 1743–1762. https://doi.org/10.1016/j.jpubeco.2004.06.009
Li, H., Zhu, J., 2024. Property rights and land quality. American J Agri Economics 106, 1619–1647. https://doi.org/10.1111/ajae.12440
Li, X., Liu, C., Weng, X., Zhou, L.-A., 2019. Target Setting in Tournaments: Theory and Evidence from China. The Economic Journal 129, 2888–2915. https://doi.org/10.1093/ej/uez018
Lichtenberg, E., Ding, C., 2009. Local officials as land developers: Urban spatial expansion in China. Journal of Urban Economics 66, 57–64. https://doi.org/10.1016/j.jue.2009.03.002
Litina, A., 2016. Natural land productivity, cooperation and comparative development. J Econ Growth 21, 351–408. https://doi.org/10.1007/s10887-016-9134-7
Ma, S., Mu, R., 2020. Forced off the farm? Farmers' labor allocation response to land requisition in China. World Development 132, 104980. https://doi.org/10.1016/j.worlddev.2020.104980
Marcus, M., Sant'Anna, P.H.C., 2021. The Role of Parallel Trends in Event Study Settings: An Application to Environmental Economics. Journal of the Association of Environmental and Resource Economists 8, 235–275. https://doi.org/10.1086/711509
Nechyba, T.J., Walsh, R.P., 2004. Urban Sprawl. Journal of Economic Perspectives 18, 177–200. https://doi.org/10.1257/0895330042632681
Nickell, S., 1981. Biases in Dynamic Models with Fixed Effects. Econometrica 49, 1417. https://doi.org/10.2307/1911408
Nunn, N., Puga, D., 2012. Ruggedness: The Blessing of Bad Geography in Africa. Review of Economics and Statistics 94, 20–36. https://doi.org/10.1162/REST_a_00161
Oster, E., 2019. Unobservable Selection and Coefficient Stability: Theory and Evidence. Journal of Business & Economic Statistics 37, 187–204. https://doi.org/10.1080/07350015.2016.1227711
Oueslati, W., Salanié, J., Wu, J., 2019. Urbanization and agricultural productivity: some lessons from European cities. Journal of Economic Geography 19, 225–249. https://doi.org/10.1093/jeg/lby001
Petty, W.S., 1662. A Treatise of Taxes and Contributions.
Ricardo, D., 1817. On the Principles of Political Economy, and Taxation, Cambridge Library Collection - British and Irish History, 19th Century. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781107589421
Roth, J., 2022. Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends. American Economic Review: Insights 4, 305–322. https://doi.org/10.1257/aeri.20210236
Sha, W., 2023. The political impacts of land expropriation in China. Journal of Development Economics 160, 102985. https://doi.org/10.1016/j.jdeveco.2022.102985
Song, Y., Zenou, Y., 2006. Property tax and urban sprawl: Theory and implications for US cities. Journal of Urban Economics 60, 519–534. https://doi.org/10.1016/j.jue.2006.05.001
Sotelo, S., 2020. Domestic Trade Frictions and Agriculture. Journal of Political Economy 128, 2690–2738. https://doi.org/10.1086/706859
Stock, J., Yogo, M., 2002. Testing for Weak Instruments in Linear IV Regression (No. t0284).

National Bureau of Economic Research, Cambridge, MA. https://doi.org/10.3386/t0284

Suri, T., 2011. Selection and Comparative Advantage in Technology Adoption. Econometrica 79, 159–209. https://doi.org/10.3982/ECTA7749

Tan, Y., Wang, Z., Zhang, Q., 2020. Land-use regulation and the intensive margin of housing supply. Journal of Urban Economics 115, 103199. https://doi.org/10.1016/j.jue.2019.103199

Tian, W., Wang, Z., Zhang, Q., 2024. Land allocation and industrial agglomeration: Evidence from the 2007 reform in China. Journal of Development Economics 171, 103351. https://doi.org/10.1016/j.jdeveco.2024.103351

Wang, Z., Zhang, Q., Zhou, L.-A., 2020. Career Incentives of City Leaders and Urban Spatial Expansion in China. The Review of Economics and Statistics 102, 897–911. https://doi.org/10.1162/rest_a_00862

Xi, T., Yao, Y., Zhang, M., 2018. Capability and opportunism: Evidence from city officials in China. Journal of Comparative Economics 46, 1046–1061. https://doi.org/10.1016/j.jce.2018.04.001

Xu, C., 2011. The Fundamental Institutions of China's Reforms and Development. Journal of Economic Literature 49, 1076–1151. https://doi.org/10.1257/jel.49.4.1076

Yang, J., Huang, X., 2021. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. Earth Syst. Sci. Data 13, 3907–3925. https://doi.org/10.5194/essd-13-3907-2021

Yao, Y., Zhang, M., 2015. Subnational leaders and economic growth: evidence from Chinese cities. J Econ Growth 20, 405–436. https://doi.org/10.1007/s10887-015-9116-1

Zhao, X., Xie, Y., 2022. The effect of land expropriation on local political trust in China. Land Use Policy 114, 105966. https://doi.org/10.1016/j.landusepol.2021.105966

# Online Appendix (Not for Publication)

## Appendix A: Additional Results

### Additional Figures

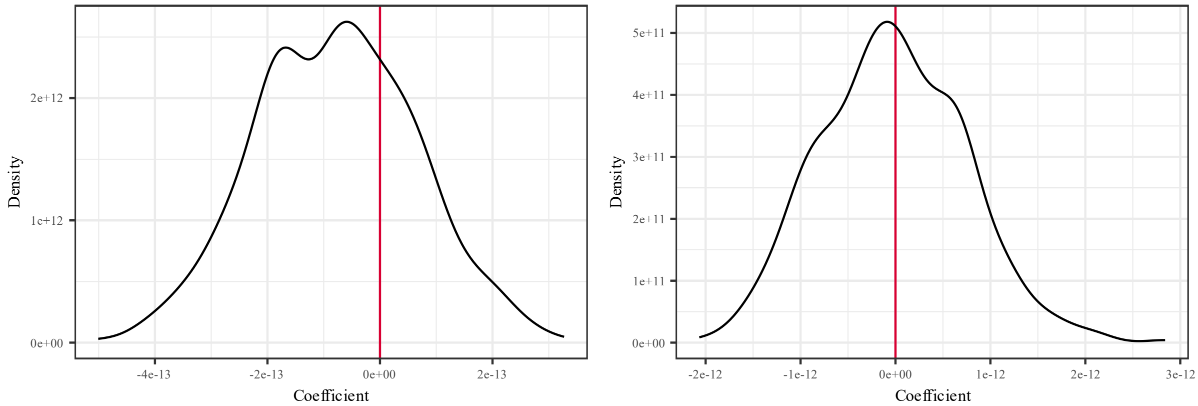



**Figure A1 Placebo Test by Randomly Assigning Initial GDP or Road Investment**

*Notes*: This figure displays the distribution of the placebo coefficients estimated through the random assignment of initial GDP (left panel) and road investment (right panel). All placebo regressions include county fixed effects and year fixed effects, as well as baseline control variables. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects.

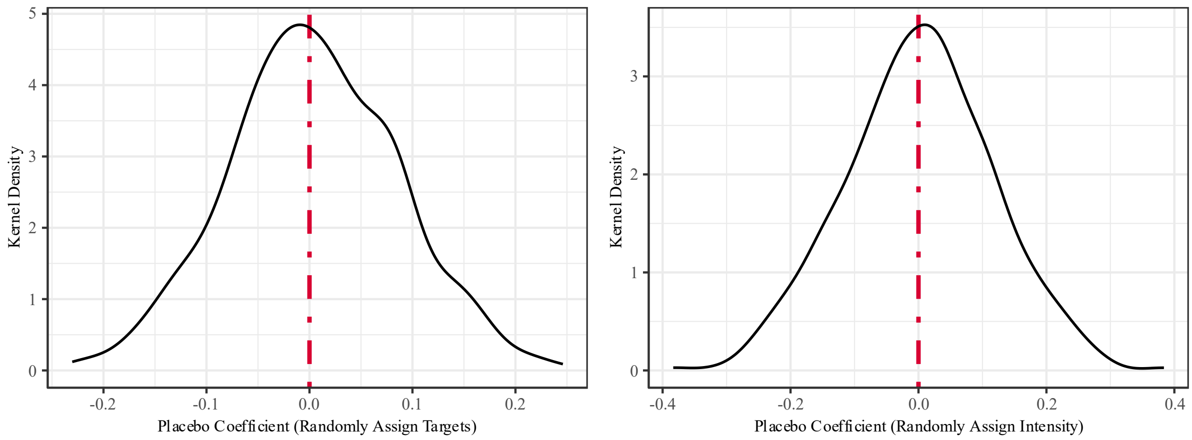



**Figure A2 Placebo Test by Randomly Assigning Protection Target/Treatment Intensity**

*Notes*: This figure displays the distribution of the placebo coefficients estimated through the random assignment of protection target (left panel) and treatment intensity (right panel). All placebo regressions include county fixed effects and year fixed effects, as well as baseline control variables. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects.

## Additional Tables

**Table A1 The Effects of Urban Expansion on Other Land**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Dep. Var. | Grassland | | Forest | | Wetland | | Other Land Quality | |
| Urban Expansion | 0.190 | -2.757* | -0.312* | -0.462 | -0.122 | -0.060 | 0.004 | -0.023 |
| | (0.268) | (1.501) | (0.160) | (0.978) | (0.084) | (0.199) | (0.015) | (0.060) |
| Observations | 53,560 | 53,560 | 53,560 | 53,560 | 53,560 | 53,560 | 53,560 | 53,560 |
| County FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Method | OLS | IV | OLS | IV | OLS | IV | OLS | IV |
| KP-F Statistics | - | 9.009 | - | 9.009 | - | 9.009 | - | 9.009 |

*Notes*: This table presents the results on the effects of urban expansion on the quantity of other land covers, including grassland, forest, and wetland, and the effects on terrain ruggedness of other land (excluding urban and agricultural land). Odd columns present the OLS estimates, and even columns present the IV estimates. The endogenous variable is the urban built-up area, instrumented by the interaction between initial county GDP and national road investment. The observation is at the county-year level. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the county level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

**Table A2 Robustness of Using Alternative IV**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. Var. | Land Quantity | | Land Quality | |
| $1(HighGDP)_i \times 1(Post2013)_t$ | -15.380*** | | 1.588*** | |
| | (4.914) | | (0.567) | |
| Urban Expansion | | -1.663*** | | 0.172*** |
| | | (0.534) | | (0.063) |
| | | | | |
| Specification | Reduced Form | IV | Reduced Form | IV |
| Observations | 53,560 | 53,560 | 53,560 | 53,560 |
| County FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes |
| KP-F Statistics | - | 185.4 | - | 185.4 |

*Notes*: This table presents the robustness results on the effects of urban expansion on the quantity and quality of agricultural land, using the alternative IV. Odd columns present the reduced-form estimates, and even columns present the IV estimates. The observation is at the county-year level. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the county level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

**Table A3 Control Instruments in OLS Specification**

| | (1) | (2) |
|---|---|---|
| Dep. Var. | Land Quantity | Land Quality |
| Urban Expansion | -1.123*** | 0.034** |
| | (0.167) | (0.015) |
| $GDP_i \times Infras_{t-1}$ | -0.009 | 0.004 |
| | (0.018) | (0.003) |
| | | |
| Observations | 53,560 | 53,560 |
| County FE | Yes | Yes |
| Year FE | Yes | Yes |
| Control Variables | Yes | Yes |
| Adjusted R-squared | 0.982 | 0.975 |

*Notes*: This table examines the exclusion restriction assumption by directly controlling for IV in the OLS specification. The first column presents the effects on agricultural land quantity, and the second column presents the effects on land quality. The observation is at the county-year level. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the county level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

**Table A4 Examining the Effects of the Instrument on Population Changes**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. Var. | Rural Population | Total Population | Agricultural Employment | Industrial Employment |
| $GDP_i \times Infras_{t-1}$ | -0.000 | 0.003 | 3.503 | 3.396 |
| | (0.001) | (0.003) | (11.699) | (14.569) |
| Observations | 53,560 | 53,560 | 53,560 | 53,560 |
| County FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Control Variables | Yes | Yes | Yes | Yes |
| Adjusted R-squared | 0.985 | 0.969 | 0.979 | 0.925 |

*Notes*: This table examines the exclusion restriction assumption by regressing a set of population and employment indicators on the instrument. The observation is at the county-year level. The dependent variables are the county's rural population in column (1), total population in column (2), agricultural employment in column (3), and industrial employment in column (4). Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the county level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

**Table A5 Including Additional Controls**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. Var. | Land Quantity | Land Quality | Land Quantity | Land Quality |
| Urban Expansion | -0.957*** | 0.277*** | -0.746** | 0.194*** |
| | (0.367) | (0.069) | (0.356) | (0.056) |
| Observations | 53,500 | 53,500 | 52,560 | 52,560 |
| County FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes |
| KP-F Statistics | 15.38 | 15.38 | 13.82 | 13.82 |

*Notes*: This table examines the exclusion restriction assumption by controlling for variables through which the instrument could affect agricultural land quality/quantity. The odd columns present the effects on agricultural land quantity, and the even columns present the effects on land quality. In columns (1) and (2), we control for time-varying agricultural output and agricultural firm entry, whereas in columns (3) and (4), we additionally control for historical agricultural productivity, interacted with full sets of year dummies. The observation is at the county-year level. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the county level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

**Table A6 Controlling for Additional Time Trends/Fixed Effects**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. Var. | Land Quantity | Land Quality | Land Quantity | Land Quality |
| Urban Expansion | -1.542** | 0.181** | -1.542** | 0.235* |
| | (0.724) | (0.090) | (0.743) | (0.137) |
| Observations | 53,560 | 53,560 | 53,560 | 53,560 |
| County FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | - | - |
| Controls | Yes | Yes | Yes | Yes |
| Province-Year Trends | Yes | Yes | - | - |
| Province-Year FE | - | - | Yes | Yes |
| KP-F Statistics | 10.47 | 7.721 | 10.16 | 8.170 |

*Notes*: This table examines the robustness of the baseline estimates by controlling for province-year trends or province-year fixed effects. The odd columns present the effects on agricultural land quantity, and the even columns present the effects on land quality. In columns (1) and (2), we control for province-specific year trends, whereas in columns (3) and (4), we control for province-year fixed effects. The observation is at the county-year level. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the county level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

**Table A7 Controlling for Lagged Dependent Variable and IV**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. Var. | Land Quantity | Land Quality | Land Quantity | Land Quality |
| Urban Expansion | -0.347* | 0.092** | -1.515*** | 0.193** |
| | (0.196) | (0.037) | (0.564) | (0.090) |
| Observations | 50,882 | 50,882 | 50,882 | 50,882 |
| County FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes |
| KP-F Statistics | 14.25 | 14.30 | 9.177 | 9.177 |

*Notes*: This table examines the robustness of the baseline estimates by controlling for lagged dependent variables and using the lagged instrument. The odd columns present the effects on agricultural land quantity, and the even columns present the effects on land quality. In columns (1) and (2), we control for the lagged term of the dependent variable, whereas in columns (3) and (4), we use both the interaction of the county's initial GDP and national-level road investment and its lagged term as instruments. The observation is at the county-year level. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the county level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

**Table A8 Adjusting the Clustering Level**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. Var. | Land Quantity | Land Quality | Land Quantity | Land Quality |
| Urban Expansion | -1.372* | 0.149* | -1.372* | 0.149* |
| | (0.720) | (0.081) | (0.782) | (0.083) |
| Observations | 53,560 | 53,560 | 53,560 | 53,560 |
| County FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes |
| KP-F Statistics | 11.34 | 11.34 | 9.009 | 9.009 |

*Notes*: This table examines the robustness of the baseline estimates by adjusting the clustering level. The odd columns present the effects on agricultural land quantity, and the even columns present the effects on land quality. In columns (1) and (2), we two-way cluster the standard errors at the county and province-year level, whereas in columns (3) and (4), we cluster the standard errors at the city level. The observation is at the county-year level. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

**Table A9 Summary of Robustness Checks**

| Dep. Var. Agricultural Land Quality | (1) Alternative Definition | (2) Two-way Cluster | (3) Wild Cluster Bootstrap | (4) City-specific Trend | (5) Group-Year Fixed Effects | (6) Confounding Policy |
|---|---|---|---|---|---|---|
| ***Panel A: OLS specification*** | | | | | | |
| *ALPPTarget* | -0.276*** | 18.067*** | 18.067*** | 12.665** | 23.255*** | 18.062*** |
| | (0.090) | (4.783) | (1.948) | (5.174) | (5.896) | (4.911) |
| Observations | 46,380 | 46,360 | 46,380 | 46,360 | 46,360 | 46,320 |
| Adjusted R-squared | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 |
| ***Panel A: DiD specification*** | | | | | | |
| *TargetIntensity*×*Post2010* | -0.165*** | 11.375*** | 11.375*** | 7.601** | 12.807*** | 11.371*** |
| | (0.058) | (3.002) | (1.130) | (3.601) | (2.719) | (3.077) |
| Observations | 46,380 | 46,360 | 46,380 | 46,360 | 46,360 | 46,320 |
| County FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Province by Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted R-squared | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 |

*Notes*: This table summarizes the robustness checks for the effects of the ALPP on the quality of agricultural land, with Panel A presenting the OLS specification (Equation (9)) and Panel B the DiD specification (Equation (10)). In column (1), we use an alternative measure of land quality. In column (2), we two-way cluster the standard error at the city and province-year level. In column (3), we cluster the standard error at the province level and exploit the Wild Cluster Bootstrap to estimate the corresponding standard error. In column (4), we control for city-specific year trends. In column (5), we absorb the group-specific year fixed effects for high-treatment and low-treatment groups. In column (6), we control for potential confounding policy. The observation is at the county-year level. County and Province by Year fixed effects are included for every specification. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the city level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

**Table A10 The Effects of Urban Expansion on Agricultural Production and Input Adjustment (Robustness)**

| Dep. Var | OLS estimates | | | IV estimates | | |
|---|---|---|---|---|---|---|
| | Grain Output | Machinery Power | Pesticide & Fertilizer Usage | Grain Output | Machinery Power | Pesticide & Fertilizer Usage |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Urban Expansion | -0.006*** | -0.008*** | 0.462*** | -0.023*** | -0.017*** | 1.112*** |
| | (0.001) | (0.001) | (0.060) | (0.005) | (0.003) | (0.220) |
| Control Variables | Yes | Yes | Yes | Yes | Yes | Yes |
| County FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted R-squared | 0.952 | 0.924 | 0.945 | | | |
| KP-F statistics | | | | 14.980 | 14.576 | 14.948 |
| Observations | 53537 | 53219 | 53077 | 53537 | 53219 | 53077 |

*Notes*: This table presents the robustness checks on the effects of urban expansion on agricultural production and input adjustment, by directly controlling for agricultural land area. Columns (1) to (3) present the OLS estimates and columns (4) to (6) the IV estimates. The endogenous variable is the urban built-up area, instrumented by the interaction between initial county GDP and national road investment. The observation is at the county-year level. Control variables include the agricultural GDP, rural income, industry structure, and income disparity, all measured at their 2000 values, the distance to the provincial capital, the distance to the coastline, and the distance to the nearest highway and railway in 2010. All controls are time-invariant and are interacted with time fixed effects. Standard error is clustered at the county level. * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.

# Appendix B: Additional Data Description

## National Taxation Survey Dataset (NTSD)

The National Taxation Survey Dataset (NTSD) is an administrative dataset compiled from tax records of firms operating in China. It is one of the most comprehensive and authoritative datasets available for studying firm behavior, taxation, and economic activities in China. The dataset covers a wide range of tax-related variables, offering a detailed view of firms' financial and operational performance, as well as their interactions with the tax system.

The NTSD is constructed based on tax filings and financial reports submitted by firms to the State Taxation Administration (STA). It includes data from multiple sources, such as Value-added tax (VAT) records that report firms' sales, purchases, and tax liabilities. Corporate income tax (CIT) filings provide insights into firms' revenue, profits, and taxable income. The dataset also includes other taxes and fees, such as business tax, excise tax, and social security contributions.

Besides detailed taxation information, the dataset also records comprehensive firm registration information, covering firm ownership structure, industry classification (based on China's national economic classification), location (province, city, or county), and registration year. What's more, the dataset also includes firms' financial performance (e.g., revenue, costs, profits, asset holdings, and liabilities), and a wide range of input information (e.g., wages, employment, capital goods investment, and intermediate goods inventory). This information allows us to calculate the TFP for each firm.

## Land Transaction Dataset

China's land transaction dataset is an administrative database that records transactions of state-owned land use rights across different cities and provinces. The dataset is primarily maintained by municipal land bureaus and real estate transaction centers, with some national-level integration conducted by the Ministry of Natural Resources and other government agencies. It is constructed based on publicly disclosed land transaction records from government land auction platforms, local land transaction centers, and official notices of land supply.

The dataset contains detailed transaction-level information on each land parcel that has been transferred through public bidding, auction, or negotiation. The core variables in the dataset include: 1. Transaction Information: The date of the land transfer, the mode of transfer (auction, tender, or negotiation), the transaction price (final winning bid or negotiated price), and the floor price per square meter. 2. Land Parcel Attributes: The land's location (including city, district, and sometimes specific geographic coordinates), total land area (measured in square meters), plot ratio, and land use type (residential, commercial, industrial, or mixed-use). Some datasets also include information on zoning regulations and land development conditions. 3. Buyer Information: The identity of the purchasing entity, which could be a real estate developer, an industrial firm, or other land users. The dataset sometimes distinguishes between different types of buyers, such as private enterprises, state-owned enterprises, or joint ventures. 4. Contractual Terms: The lease duration (typically 70 years for residential land, 50 years for commercial land, and 40 years for industrial land), specific land use restrictions, and any additional contractual obligations imposed on the buyer (e.g., requirements for infrastructure investment or affordable housing provisions).