What Started the Chinese Miracle? The View from Outer Space*

Joel Ferguson
University of Wisconsin-Madison

Oliver Kim†
Coefficient Giving

This Version: March 27, 2026

[Link to Latest Version]

Abstract

We study one of the central reforms in China's economic miracle, the Household Responsibility System (HRS), which decollectivized agriculture starting in 1978. The HRS is commonly seen as having significantly boosted agricultural productivity—but this conclusion rests on unreliable official data. We use historical satellite imagery to generate new measurements of grain yield, independent of official Chinese statistics. Using two separate empirical designs that exploit the staggered rollout of the HRS across provinces and counties, we find no evidence that adopting the HRS sooner caused faster grain yield growth. We provide suggestive evidence that a procurement price increase in 1979 was a more important driver of yield growth. These results challenge our conventional understanding of decollectivization, land reform, and the origins of the Chinese miracle.

*We thank Edward Miguel, Benjamin Faber, Jón Steinsson, Barry Eichengreen, Brad DeLong, Andres Rodriguez-Clare, Cecile Gaubert, Marco Gonzalez-Navarro, Gerard Roland, Frederico Finan, Joshua Blumenstock, Solomon Hsiang, Kirill Borusyak, Davis Kedrosky, Advik Shreekumar, J. Landin Smith, and Karthik Tadepalli for helpful comments, Justin Yifu Lin, Douglas Almond, Björn Brey, and Matthias Hertweck for generously sharing their data, and Yiyang Chen, Zachary Shi, Tony W., Mingqi Zeng, and Silin Zhang for excellent research assistance. We also thank ChatGPT for making this project feasible. For generous financial support, we thank UC Berkeley's Center for Effective Global Action (CEGA), UC Berkeley's Clausen Center, UC Berkeley Haas's Fisher Center for Real Estate & Urban Economics, and the Structural Transformation and Economic Growth (STEG) Programme. Oliver Kim acknowledges support from the NSF Graduate Research Fellowship under Grant No. 1752814 and an Emergent Ventures Fellowship from the Mercatus Center at George Mason University. We thank participants at the UC Berkeley's Development Seminar, Economic History Seminar, Development Lunch, Trade Lunch, Macro Lunch, and Economic History Workshop for valuable comments and feedback. All errors are our own.

†Contact: oliverwkim@gmail.com

1 Introduction

China's rise has lifted hundreds of millions out of poverty and reshaped the global economy, making it a central model for developing countries worldwide. According to the World Bank, China has been responsible for 75% of global extreme poverty reduction and 23% of global GDP growth since 1980 (World Bank 2022). Understanding the causes of Chinese economic growth is thus a central question for development economics.

The conventional wisdom is that the Chinese economic miracle was caused by the market-oriented reforms of Deng Xiaoping. The first and perhaps the most important of these was the post-1978 Household Responsibility System (HRS), which dismantled Mao's collective farms and gave households market incentives to produce—in a literal sense, ending communism in agriculture. From 1978 to 1984, official grain yields surged by 43%, growing from 2.8 tons per hectare to 3.6 tons per hectare. The academic consensus is that the HRS played an important role, with a widely cited estimate by Lin (1992) stating that decollectivization contributed to 49% of all agricultural output growth from 1978-84. For a developing country where agriculture was still central—around 30% of value-added and 60% of employment in 1980—such a boost to productivity would have made the HRS a major factor in China's growth takeoff. Moreover, the perceived success of the HRS led to the expansion of other reforms that liberalized the Chinese economy, making it a cornerstone of political legitimacy for the post-Mao Communist Party.

But how much can we trust the conventional wisdom about the HRS? Compared to the scale of the Chinese miracle, modern research on Deng Xiaoping's reforms is surprisingly rare. Prior work has been limited by two data constraints. First, statistics below the provincial level from this period are scarce, making precise causal identification difficult. A particular concern is that the HRS was first adopted in poorer and drought-stricken areas, so naive comparisons of yield growth could reflect mean reversion rather than effects of the reform. Second, Chinese economic statistics are notoriously unreliable—foreign researchers, Chinese economists, and even top Chinese leaders have all observed that they can reflect political priorities more than the ground truth.[1] Notably, the Great Chinese Famine, which occurred two decades before the HRS and killed upwards of 30 million people, was exacerbated by local officials systematically over-reporting grain production (J. Yang 2013).

This paper tackles these challenges with data from a novel source: historical satellite imagery. Using images from the Landsat Multispectral Scanner (MSS) and the Advanced Very High Resolution Radiometer (AVHRR), we measure plant biomass on the Earth's surface with the Normalized Difference Vegetation Index (NDVI), a widely used proxy for agricultural yields. We verify that NDVI is an accurate predictor of yield in a wide array of geographic and historical contexts, and is precise enough to detect effect sizes similar in magnitude to those attributed to the HRS. The result is a dataset that is not only granular enough for precise identification—at

1. See Rawski (1976), Holz (2003), and Nakamura, Steinsson, and Liu (2016) for discussions of Chinese macroeconomic statistics. In leaked diplomatic cables, former Premier Li Keqiang once described China's GDP statistics as "man-made" and therefore unreliable (Rabinovitch 2010).

resolutions ranging from 60 meters (Landsat MSS) to 8 kilometers (AVHRR)—but also independent of the official Chinese statistical apparatus. To our knowledge, this is the first application of historical satellite imagery in economic history.

We exploit the spatial granularity of satellite data with two complementary identification strategies. A *difference-in-discontinuities* design exploits the staggered rollout of the HRS across provinces: by comparing satellite-measured vegetation on either side of provincial boundaries before and after HRS adoption, we absorb any fixed differences at the border and identify the HRS effect from changes in the discontinuity over time. A county-level staggered difference-in-differences design, using unofficial treatment data compiled by Almond, Li, and Zhang (2019) from historical gazetteers, provides a test across the interior. These two strategies draw on independent treatment data sources and cover complementary geographies—border areas and county interiors, respectively—strengthening the credibility of the combined evidence.

Our first and main finding is that the Household Responsibility System had a negligible effect on grain yields. The difference-in-discontinuities estimates are precisely zero—converting NDVI to yield units using the upper bound of the estimated NDVI–yield relationship, the 95% confidence interval rules out yield effects larger than 0.5%, even three years after the onset of reform. The county-level difference-in-differences confirms the result: counties did not experience faster yield growth after adopting the HRS. The null finding is consistent across both designs and is robust to a wide array of alternative estimators and specifications. An important nuance for interpreting these results is that grain yields measure land productivity, which is visible from space, while labor productivity is not. Our findings cannot rule out the HRS having raised labor productivity, thereby freeing households to allocate their time more flexibly. But grain yields were the central claim of the HRS narrative, and it is the land-productivity channel through which prior work attributed massive output gains.

If the HRS did not increase yields, what did? Our second finding is that another reform—Hua Guofeng's 1979 increase in state procurement prices—may have been a more important driver of China's agricultural takeoff. In 1979, the state raised the average price for quota grain by 20% and the bonus for above-quota grain from 30% to 50% (Sicular 1988b). Using a shift-share research design that interacts cross-sectional variation in crop suitability with the national procurement price shock, we find that grid cells more exposed to the price increase experienced significantly larger gains in NDVI. Controlling for HRS adoption leaves the price effect unchanged, while the HRS effect remains insignificant. This finding also helps resolve a basic timing puzzle: Chinese yields had already risen by almost 20% by 1980, when fewer than 20% of work teams had adopted the HRS. The procurement price reform, by contrast, was implemented nationally in 1979—aligning with the timing of the yield surge. Politically, the price reform is associated with Hua Guofeng, Mao's chosen successor, rather than with Deng Xiaoping, complicating the standard narrative that credits Deng's market reforms for China's agricultural takeoff.

By freeing households from the collective, the HRS may not have made farming more productive per hectare, but may have released labor for more productive non-agricultural work. Our

third finding is that the HRS increased urbanization and non-agricultural activity. Using Landsat imagery, we construct satellite-based measures of built-up settlement area to track peri-urban growth, and find that counties that adopted the HRS earlier saw an expansion of built-up land. These patterns are consistent with the explosive growth of the Township and Village Enterprises (TVEs)—the most dynamic sector of 1980s China—which drew labor out of agriculture and into rural industry.

These findings challenge a near-universal view among economists about the origins of the Chinese miracle. Growth accounting exercises have concluded that agricultural productivity growth was needed to release labor from rural areas, making agriculture, not manufacturing, the key ingredient in China's GDP takeoff (Young 2003; Brandt, Hsieh, and Zhu 2008). Ravallion and Chen (2007) find that growth in the rural sector was responsible for 75–80% of the fall in the national poverty rate from 1981–2001, the "bulk" of which they attribute to the HRS. If China has indeed lifted 800 million people out of extreme poverty since 1980—75% of the global total (World Bank 2022)—then our understanding of the drivers behind the vast majority of global poverty reduction over the past 40 years will have to be substantially revised. Our results shift the locus of explanation: rather than making land more productive, the HRS may have freed labor to pursue more productive work outside of agriculture.

Related Literature This paper revisits an older literature which established that the HRS had a major effect on agricultural productivity growth in China (Lin 1988; McMillan, Whalley, and Zhu 1989). In particular, it builds on Lin (1992), which, using a panel of official provincial data, finds that the HRS accounted for half of overall agricultural output growth from 1978-84. It also builds on the subsequent work of Almond, Li, and Zhang (2019), which finds a positive effect on grain output per capita using a county-level panel. We will use both the official province-level HRS treatment data from Lin (1992) and the unofficial county-level HRS data Almond, Li, and Zhang (2019) in our identification strategies. Our findings revise our understanding of one of the central reforms of China's liberalizing period, and should prompt greater skepticism about the economic statistics underlying our conception of the Chinese miracle.

This study also contributes to our broader understanding of the role of agriculture in development. There is a long-running debate in the literature around the effects of land reform and farm size on agricultural productivity: an influential view is that smallholder farms can be more efficient (Vollrath 2007; Kagin, Taylor, and Yúnez-Naude 2016), while an opposing view is that large farms benefit from scale economies that make them more efficient (Foster and Rosenzweig 2017; Adamopoulos and Restuccia 2020). Our paper joins a wave of new evidence reexamining Asia's postwar land reforms, which have found evidence of mixed success (Kim and Wang 2024; Kitamura 2022). China's transition from large-scale collectives to small-scale household farms—the largest land reform of its kind in history—has been enormously influential in that debate. This paper's null finding may temper some of the optimism around the universal efficiency of smallholder farming.

Finally, this paper contributes to the growing, fruitful intersection between remote sensing research and economics. Nightlights observed by satellites have become a well-accepted measure for economic activity, particularly when there are political incentives to inflate the official economic statistics (Henderson, Storeygard, and Weil 2012; Hodler and Raschky 2014; Martínez 2022). More recent research has used daytime satellite imagery to measure contemporary outcomes like agricultural output and poverty (Jean et al. 2016; Yeh et al. 2020; Huang, Hsiang, and Gonzalez-Navarro 2021), but to our knowledge this is their first application in economic history. Prior research has tended to overlook older satellites, like the Advanced Very High Resolution Radiometer (AVHRR) used in this paper, due perhaps to their lower resolution and how difficult they are to process. This paper shows that these earlier measurements, while imperfect, contain a wealth of useful economic information; the techniques developed in this paper can be applied to a wide range of historical settings—anywhere in the world, from 1978 onwards. Given the central place of agriculture in historical development, satellite-based measurement could open up vast new areas of economic research in settings like postcolonial sub-Saharan Africa and the Soviet Union, where statistics on agricultural incomes and output are either unreliable or nonexistent.

This paper is organized as follows. Section 2 introduces the historical context. Section 3 describes our data sources and presents supporting evidence that satellite imagery is a strong predictor of agricultural yields. Section 4 presents our main empirical results. Section 5 examines the role of procurement price reform. Section 6 discusses urbanization and structural transformation. Section 7 concludes.

2 Historical Context

Mao Zedong died in 1976, creating the political opening to reform the collectivized system of agriculture that had prevailed in China since 1953. Crucially for this study, decollectivization did not occur evenly across the country. While it was spurred by changes to national leadership, and often pushed for by peasants from the grassroots level, it was most of all constrained and shaped by provincial leaders, giving rise to the natural experiment at the heart of this paper.

The View from the Peasantry By the mid-1970s in China, every twenty to thirty agricultural households were organized into work teams, which owned the land collectively and shared both work responsibilities and output (Eisenman 2018). Every two to two dozen work teams were organized into brigades, which were largely responsible for rural industry; and every ten to twenty brigades were sorted into communes, which owned all the agricultural capital, and served as the centers for public administration and services in rural China (Kelliher (1992), p. 9).

The conventional view is that the collective farming system gave households little incentive to produce more. Teams had to sell a mandatory procurement quota in grain to the state, for which they received a fixed price. This procurement price was lower than the price that would have prevailed in a free market, such that the grain procurement effectively acted as a tax on

farmers (Kelliher 1992). After deducting costs, work teams divided most of their net income among households based on their size (Eisenman (2018), p. 193). If the team produced a surplus, a system of work points gave households a cash bonus, usually based on their observed labor inputs. But because supervision is costly, standard neoclassical agricultural models predict that there are strong incentives to shirk on a collective farm, overwhelming any of the potential efficiency gains from economies of scale (Lin 1988). Or, as one Hubei farmer succinctly put it: "you've got Brother Zhang and Brother Li—if one works more and the other works less, it all comes out about the same" (Kelliher (1992), p. 96).

In 1978, Anhui Province was struck by a severe drought. As an emergency measure to prevent starvation, work teams in Anhui began openly experimenting with "household responsibility". Rather than plant the winter wheat crop collectively, each working adult was allocated 0.1 hectares of land and three yuan for production costs (Kelliher (1992), p. 61). In exchange, they would turn over 100 kilograms of wheat to the collective at harvest. Any surplus could be kept by the household, to be used or marketed as they pleased. In effect, the old collective was broken. Soon, with the support of provincial leaders like Anhui's Wan Li and Sichuan's Zhao Ziyang, similar responsibility systems began to be implemented in other provinces. Though there were a wide range of different variations, the two most common were *baochan daohu*, which rewarded households for over-quota production while retaining cropping, management, and investment decisions with the collective, and *baogan daohu*, which gave households close to full responsibility over production and output. The latter was more popular among farmers because of the greater freedom it gave them. But both household responsibility systems were united in giving households a claim on any surplus output after fulfilling a procurement quota to the state.

In 1982, both *baogan daohu* and *baochan daohu* were finally legalized nationally, creating the formal Household Responsibility System. *Baogan daohu* (which literally means contracting directly with households) became the favored system in most provinces, with work teams shedding their control over households' agricultural decisions. Households were now directly responsible to the state for a mandatory procurement quota, but the residual claimants on any remaining surplus. The teams, now "villages" again after reform, kept ownership of the land, but the law guaranteed security of tenure for at least 15 years (Brandt, Turner, and Rozelle 2004). The tractors and agricultural capital of the communes were redistributed (Eisenman (2018), p. 258), and, accompanied by reforms to the procurement price system, rural markets gradually reopened (Skinner 1985). Rural China was now firmly on the path to transition out of socialism.

What factors shaped the adoption of the HRS across China? Responsibility systems were not, in fact, new to the Communist period: during the famine of the Great Leap Forward, farmers often broke from the collectives, and set up responsibility systems themselves in desperation, but these experiments were invariably crushed by the Party. Mao's death altered the national political environment and made deviations from orthodoxy possible—but reform needed sponsors in provincial leadership in order to survive.

The View from the Provinces Decollectivization proceeded unevenly across provinces, resisted in some areas, pioneered in others. B.1a shows the date when over 50% of each province's work teams reported implementing the Household Responsibility System (HRS). Anhui was the clear national reform leader starting in 1978, with most of its neighbors still beginning to adopt reform in 1981. In 1982, the HRS became national policy, and Anhui's neighbors caught up, reporting close to full adoption by 1983 (at least in their official statistics).

Why did provinces vary so much in their rates of reform? Historians have emphasized the importance of provincial leaders in setting the pace of decollectivization:

> Provincial leaders generally played a pivotal role in the entire rural-reform process. Although the impetus for change came from below and the issue was only settled with a series of central-level decisions in 1980-81, innovative provincial leaders encouraged and protected the survival and spread of responsibility systems within their respective local areas... (Teiwes and Sun (2016), p.75)

The paradigm was Anhui's First Party Secretary Wan Li, who had been appointed in 1977 for orthogonal political reasons—to reduce the military's influence in provincial politics (Chung (2000), p. 94). When peasants began implementing household responsibility in 1978, Wan Li allowed the experiment to continue. Later, he protected it from mid-level cadres who wanted to reverse the reforms, and encouraged further experimentation throughout his province (Teiwes and Sun (2016), p. 101). When faced with similar grassroots pushes for household responsibility, Sichuan's Zhao Ziyang and Guizhou's Ma Li permitted and even encouraged greater reform.

By contrast, more conservative provincial leaders could hamstring the progress of reform. In the north, Heilongjiang Province was a notable laggard. Heilongjiang's leaders resisted small-scale household farming, thinking it an inappropriate system for a wheat-growing region where agriculture was heavily mechanized. In June 1981, only 0.7% of work teams had adopted the HRS, showing the importance of provincial authorities in containing the spread of reform (Chung (2000)). Heilongjiang only caught up after the HRS was made national policy—and even then only under immense pressure from the center (Weber 2021)—and the decollectivization process was belatedly rushed in the early 1980s.

The View from the Top While changes to China's national leadership (namely, Mao's demise) created the political conditions needed for reform, the resulting power vacuum made provincial leaders crucial in determining the pace and spread of decollectivization.

Mao's immediate successor was Hua Guofeng, a relatively junior official he had handpicked from Hunan Province. Hua's authority was much weaker than Mao's, and reformers and hard-liners in Beijing soon openly clashed over the reform question. Contrary to popular belief, Deng Xiaoping was not involved in the rural reform process until reforms were well underway.[2] At first, the hardliners held sway—November 1978's Third Plenum Regulations explicitly banned

2. Deng made his first public comments on *baochan daohu* only on May 31, 1980, long after decollectivization had begun in Anhui.

baochan daohu (Teiwes and Sun (2016), p. 66). But, in response to lobbying from figures like Wan Li, a document from September 1979's Fourth Plenum carved out three exceptions for households engaged in sideline occupations, remote areas, and single, isolated households. A September 1980 Party notice expanded the exceptions again, this time to "poor and backward areas" and "production units heavily dependent on state subsidies". Finally, January 1982's Central Document No. 1 officially enshrined *baochan daohu* and *baogan daohu* as "the production responsibility systems of the socialist economy" (Chung (2000), p. 58).

This conflict at the top had two effects on the reform process. First, because of the confused and contradictory directives coming from Beijing, provincial leaders took the lead in implementing reform. Second, because of the resistance from hardliners, HRS was first permitted only in poorer, famine-stricken, and economically marginal areas, inducing *negative* selection in early adopters. These facts will inform our identification strategy: we will use the staggered rollout of the HRS across provinces as a source of variation, while taking into account that areas that reformed first were negatively selected.

The Productivity Effects of Reform Figure 1 summarizes the central timing challenge of identifying the causes of China's agricultural growth—productivity grew rapidly during a period of multiple simultaneous reforms.

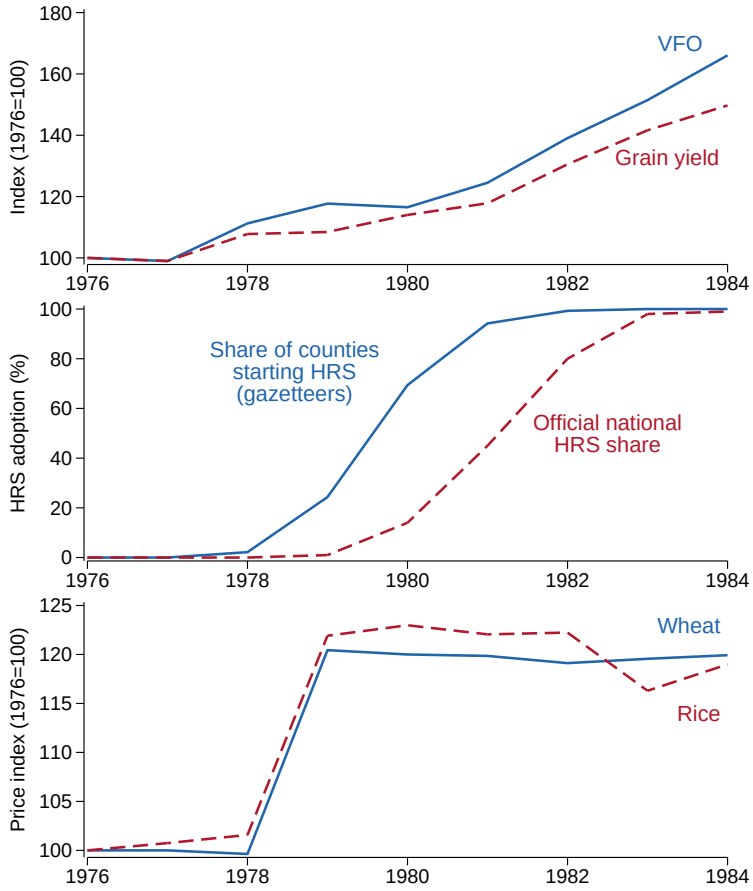
The top panel shows how, from 1976 to 1984, the value of farm output (VFO) increased by around 60% and grain yields rose by over 40%. Almost all of this growth was driven by grains—in particular, wheat, rice, and corn, which accounted for 75% of sown area and 85% of total output (H. Yang 1999). Rice yields grew by 1.4 tons per hectare, or 35%, and wheat yields grew by 1.1 tons per hectare, or 60%.

The middle panel tracks the HRS's spread using two series. The red line shows the nationwide share of work teams operating under the HRS from official statistics, as compiled by Lin (1992). The blue line shows the share of counties reporting the start of HRS adoption in unofficial county gazetteers (Almond, Li, and Zhang 2019).[3] The gazetteer series leads the official statistics, in part because it captures the *onset* of adoption (when a few villages in a county first contracted land) rather than the share of teams that have fully transitioned, and in part because gazetteers, compiled locally for historical purposes, may be subject to less political pressure than official reporting. Even so, in the official series fewer than 20% of work teams had adopted the HRS through 1980, while yields had already risen by almost 20%—well before the HRS reached the majority of the countryside, a tension emphasized by Bramall (2004).

The bottom panel shows the sharp rise in rice and wheat procurement prices beginning in 1979. That year, under Hua Guofeng, the state raised the average procurement price for farmers' required grain quota by 20% and the bonus for above-quota grain from 30% to 50% (Sicular 1988a). (In the Chinese context, grain refers to rice, wheat, corn, soybean, potatoes, millet,

3. Almond, Li, and Zhang (2019) define the start of land reform as "the year when collectively owned land was first contracted to individual households in a few villages for each county." Note also that this is an incomplete sample, with $N = 914$, roughly half of China's rural counties.

Figure 1: National Timing of Yield Growth, HRS Adoption, and Procurement Prices



This figure aligns three national time series on a common year axis (1976–84). Panel A plots the value of farm output (VFO) and grain-yield indices (1976=100) using Lin (1992) and State Statistical Bureau grain statistics. Panel B plots HRS adoption as the share of counties beginning HRS reforms based on county gazetteers (Almond, Li, and Zhang 2019) and the official national HRS share from Lin (1992). Panel C plots national wheat and rice procurement prices (1976=100) from digitized State Council price schedules.

sorghum, and some miscellaneous crops (H. Yang 1999).) Quota and above-quota prices for oil crops and cotton were raised, as was the quota price for sugar. These brought prices closer to the levels that would prevail on a free market.

The conventional academic view attributes the lion's share of China's agricultural takeoff to the HRS. (Table B.7 summarizes this literature.) Applying a structural model to national aggregate data, McMillan, Whalley, and Zhu (1989) attribute 78% of TFP growth from 1978 to 1984 to the HRS and 22% to the procurement price reform. Similarly, using a province-level panel regression, a widely cited study by Lin (1992) estimates that the HRS accounts for almost half of growth in farm output. More recently, using the above gazetteer adoption data to conduct a county-level event study, Almond, Li, and Zhang (2019) find large positive effects of HRS adoption on grain output per capita.

These estimates, however, share important limitations. McMillan, Whalley, and Zhu (1989)

and Lin (1992) rely on official data and predate modern causal identification methods, making it difficult to separate the contribution of the HRS from concurrent reforms like the 1979 price increases. Almond, Li, and Zhang (2019) exploit county-level variation in HRS adoption timing, but use grain output per capita—rather than the more common measure of yield—as their outcome variable, where the denominator is total county population instead of the agricultural labor force. Our paper applies modern causal inference techniques to satellite-based proxies for yield, to generate new estimates of the HRS's effect.

3 The View from Space: Measuring Yields with Satellites

3.1 Landsat MSS and AVHRR

This paper uses imagery from two satellite sensors to measure agricultural yields: the Landsat Multispectral Scanner (MSS) and the Advanced Very High Resolution Radiometer (AVHRR). The two satellites have complementary advantages and disadvantages.

Landsat MSS was launched into orbit with NASA's Landsat 1 satellite in 1972 and operated until the suspension of Landsat 5 in 2013. MSS captured images with high spatial resolution (around 60 meters per pixel) at the cost of a relatively low temporal resolution (points were revisited every 16-18 days, and not everywhere was captured). It captured data in four spectral bands—green ($0.5 - 0.6\mu$m), red ($0.6 - 0.7\mu$m), high-frequency near-infrared ($0.7 - 0.8\mu$m), and low-frequency near-infrared ($0.8 - 1.1\mu$m)—which allows for the calculation of both NDVI and GCVI. (Figure A.1 maps MSS's coverage across China from 1976 to 1984.[4])

AVHRR data is available from a series of National Oceanic and Atmospheric Administration (NOAA) satellites from late 1978 to 2013. In contrast to MSS, AVHRR had low spatial resolution (around 4 kilometers per pixel) but extremely high temporal resolution, with images captured twice a day. The earliest generations of AVHRR only had two spectral bands relevant for agriculture—red ($0.58 - 0.68\mu$m) and near-infrared ($0.725 - 1.00\mu$m)—which allows only for the calculation of NDVI. However, the high temporal resolution makes AVHRR data well-suited to measuring agricultural activity, as high-frequency NDVI allows us to measure changing biomass over the growing season.

One additional complication with AVHRR is that NOAA's widely used Climate Data Record (CDR) data only provides cleaned, processed, and calibrated AVHRR data from imagery captured after June 24, 1981. This is a problem for the present study, since the rollout of the HRS was well underway by this point. In order to capture the earliest stages of the Household Responsibility System, we processed additional data from the earlier TIROS-N and NOAA-6 satellites, which carried older AVHRR sensors but had been excluded from the Climate Data Record due to their lower data quality. (Figure A.5, panel (a), shows a timeline of these sensors.) Adding these older sensors allows us to extend our dataset back to late 1979, though level differences

4. Among Chinese territories at the time of the reform, we exclude Tibet due to the lack of HRS treatment data.

in NDVI measurements across sensors prohibit direct comparisons across certain time periods (A.5b). Further technical details are available in the appendix.

3.2 Measuring Yields with Satellites

Measuring agricultural yields with satellites relies on a simple biological observation: plants use light from the visible part of the electromagnetic spectrum to photosynthesize, while reflecting back higher-frequency light (Taiz et al. 2022). Viewed from space, a healthy plant will thus reflect more near-infrared (NIR) light relative to red or green light than a stressed plant. This insight motivates one of the most common measures of crop cover in environmental science, the Normalized Difference Vegetation Index (NDVI):

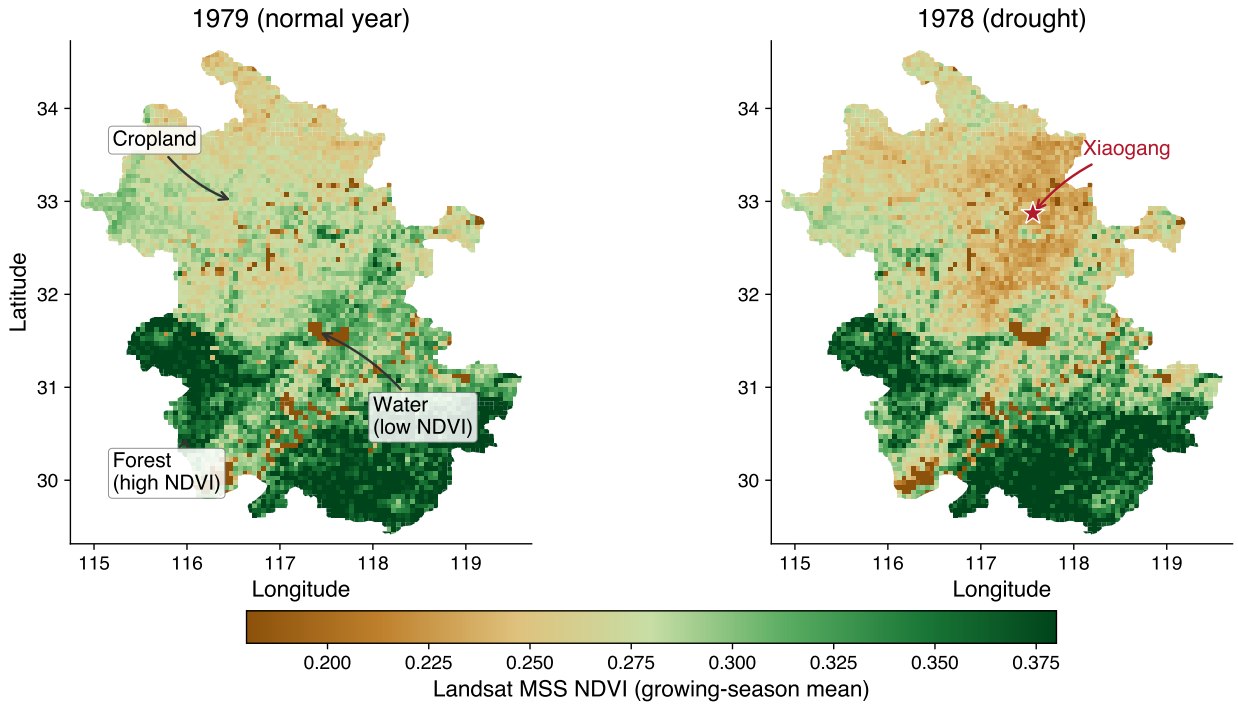
$$NDVI = \frac{NIR - Red}{NIR + Red}. \tag{1}$$

For our regression analyses, we aggregate NDVI observations in both space and time. Spatially, we compute the mean of pixel-level NDVI within fixed geographic units—0.05×0.05 grid cells (approximately 5×5 km) for border-based comparisons, and administrative counties for panel analyses. Temporally, we take the maximum across satellite images within a year as a proxy for yields, as the growing season peak of NDVI is strongly correlated with crop biomass and harvested output (David B Lobell et al. 2019).

Figure 2 shows how NDVI captures both cross-sectional and intertemporal variation in vegetation. The left panel shows mean growing-season NDVI in 1979, an ordinary year. Forests and wilderness areas, like the southern Dabie Mountains, have high NDVI and are colored in dark green. Water features and urban areas, like the U-shaped Chao Lake and the densely populated banks of the Yangtze River (which cuts through Anhui's lower third), have low NDVI and appear as brown. The NDVI of cropland and agricultural areas is between these extremes, and are colored a middle shade of green. The right panel shows NDVI in 1978, the year when the HRS was first adopted in Anhui. 1978 saw unusually low rainfall and poor yields, which is reflected in lower values of NDVI in brown. Northern Anhui and the Huai River basin were particularly hard-hit—Xiaogang village, where the HRS was reportedly first implemented as an emergency measure, is marked with a red star.

These figures demonstrate that satellite-based NDVI is capable of capturing significant variation in vegetation. However, capturing a drought visible from space is a lower bar than detecting the finer yield differences needed to identify a policy effect. In the extreme, one concerning possibility is that the null results we find are simply the product of measurement error or statistical noise. This section shows systematic evidence that this is not the case. We run a battery of validation tests to demonstrate that our satellite NDVI data (a) behaves like crop data in time series, (b) correlates strongly with official yields from administrative data, and (c) responds sensitively enough to observe small changes in yields in real-world regression settings.

Figure 2: Landsat MSS NDVI over Anhui Province: Normal Year vs. Drought



This figure shows mean growing-season Landsat MSS NDVI for Anhui Province in 1979 (a normal year, left) and 1978 (a drought year, right). NDVI from earlier Landsat sensors (Landsat-1 and -2) is cross-calibrated to Landsat-3 using the 1978–1981 overlap period. The left panel shows the expected pattern of land cover: forest in the Dabie Mountains (high NDVI), cropland on the Huai River plain, and Chao Lake (low NDVI). In 1978, the Huai River basin dried up, visible as a dramatic shift from green to brown tones across northern Anhui. A star marks Xiaogang village (Fengyang County), where households began experimenting with the Household Responsibility System in response to the drought.

Does NDVI behave like crop data in time series? A key test for NDVI in satellite imagery is that its descriptive statistics should behave like crop data throughout the agricultural calendar. In agricultural regions, biomass rises during the growing season of the principal crop, peaks just before the harvest, and bottoms out during the period when the land is left idle. Figure A.2 shows the profiles of NDVI throughout the year for each Chinese province, with Landsat MSS in blue and AVHRR in red, overlaid with planting and harvest windows from USDA crop calendar data. The NDVI curves for AVHRR are lower in level and flatter than for Landsat, since AVHRR's coarser resolution captures more non-agricultural land, systematically lowering NDVI and reducing the amplitude of seasonal changes. However, both Landsat and AVHRR NDVI show the expected seasonal pattern, rising during the planting season, peaking just before the harvest, and flattening out in the winter. (The annual maximum of NDVI, which typically occurs just before harvest, is often a suitable proxy for harvest yield.) Places where we observe two peaks (for instance, Henan and Jiangsu) are known for double-cropping, where winter wheat is harvested in May or June and a second crop (either maize or rice) is harvested in August or September.

(Figure A.3 maps each province's dominant cropping system and the Qinling–Huaihe line that roughly divides wheat-dominant northern agriculture from rice-dominant southern agriculture.)

Another crucial time series test is how NDVI behaves *between* years. Figure A.4 shows the trend of the national average of Landsat MSS NDVI during the 1970s reform period.[5] The left panel shows mean NDVI, residualized against cell fixed-effects and month-of-year effects (to control for the differential availability of satellite imagery).[6] The right panel shows maximum NDVI for a given cell in a year. In both panels, MSS NDVI across China shows a clear positive trend from 1976 to 1984. (Appendix Figure A.9 compares, province by province, Landsat MSS NDVI to grain yields, though some provincial series are noisy due to small image samples.) Overall, observations from NDVI are consistent with a narrative that agricultural yields did indeed increase in aggregate across China throughout the late 1970s and early 1980s.

Is NDVI a good predictor of yields? We can directly test NDVI's ability to predict yield by comparing it to administrative data. We estimate the following specification

$$z(y)_{i,t} = \beta\, z(NDVI)_{i,t} + \alpha_i + \tau_t + \varepsilon_{i,t}$$

where for location i and year t, $z(y)_{i,t}$ is standardized grain yield from official statistics, $z(NDVI)_{i,t}$ is standardized annual maximum NDVI, α_i is a location fixed effect, and τ_t is a year effect. Both variables are expressed in standard-deviation units, so β is interpreted as the SD change in yield associated with a 1 SD increase in NDVI.

Table 1 shows regressions of grain yield from administrative data on annual maximum NDVI across several Asian countries—China, South Korea, Japan, and India—for both Landsat MSS (Panel A) and AVHRR (Panel B). In the pooled sample, a 1 SD increase in maximum NDVI is associated with a 0.081 SD increase in annual yield with AVHRR and a 0.037 SD increase with Landsat MSS, both highly significant. AVHRR NDVI is a significant positive predictor in every country sample, while Landsat MSS is significant for the pooled sample, China counties, and India, though imprecise for Japan and South Korea, largely due to the restrictively small sample sizes. In the appendix, Figure A.7 and Figure A.8 show the corresponding binscatters of yield on AVHRR and Landsat MSS NDVI. The quantile means are approximately linear, suggesting that a linear model is appropriate for when we convert NDVI into yield estimates.

These results accord with a large research literature in remote sensing and environmental science where NDVI has become a standard proxy for agricultural yields. Appendix Table B.7 summarizes the literature on NDVI for yield estimation, which dates back to Tucker (1979). Moreover, it has recently become widely used in development and agricultural economics, particularly in settings where administrative data is either unreliable or missing (Asher and Novosad 2020; Wuepper, Oluoch, and Hadi 2025; Cisse et al. 2026).

5. Recall that it is not possible to construct a similarly long panel with AVHRR, due to discrepancies between AVHRR sensors.

6. We use a province-level block boostrap with 100 draws to simulate the varying coverage of Landsat MSS, result-

Table 1: Predicting Yields with NDVI

Panel A: Landsat MSS					
	(1)	(2)	(3)	(4)	(5)
	Pooled	China (counties)	Korea	Japan	India
Max NDVI (MSS)	0.0368***	0.0820***	0.0117	0.0183	0.0600***
	(0.0110)	(0.0148)	(0.0423)	(0.0275)	(0.0118)
Observations	5623	1262	354	348	2783
R^2	0.800	0.949	0.637	0.888	0.820
Panel B: AVHRR					
Max NDVI (AVHRR)	0.0809***	0.0853***	0.0574*	0.131***	0.113***
	(0.00623)	(0.00864)	(0.0257)	(0.0338)	(0.00860)
Observations	16595	3360	2324	305	7603
R^2	0.808	0.958	0.635	0.902	0.857

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table reports estimates from $y_{i,t} = \beta \, \text{Max NDVI}_{i,t} + \alpha_i + \tau_t + \varepsilon_{i,t}$, where $y_{i,t}$ is administrative grain yield and Max NDVI is the annual peak NDVI in location i and year t. Panel A uses Landsat MSS and Panel B uses AVHRR; each panel reports pooled estimates and separate samples for Korea, Japan, India, and China (counties). Both outcome and NDVI are standardized (z-scores), so coefficients are interpreted in standard-deviation units. All specifications include location and year fixed effects; standard errors are in parentheses.

Are satellites sensitive enough? Given our finding of small or statistically insignificant effects of the HRS on yields, a natural concern is that our null results are the product of measurement error. NDVI may be a good predictor of administrative yields on average, but perhaps it is too noisy to detect the moderate-sized yield changes attributed to the HRS. We address this concern by showing that satellite-measured NDVI is sensitive enough to detect yield changes in regression settings comparable to those we use to evaluate the HRS.

First, since one of our main identification strategies is a regression discontinuity design, we show that our satellite data can detect discontinuities in yield from space[7]. In Asia, the India-Pakistan border in the Punjab is a well-known discontinuity, where since Partition yields have been systematically higher on the Indian side of the border. (Figure A.6 shows that this NDVI difference is still visible from space in the present day.) We estimate the following cross-sectional specification:

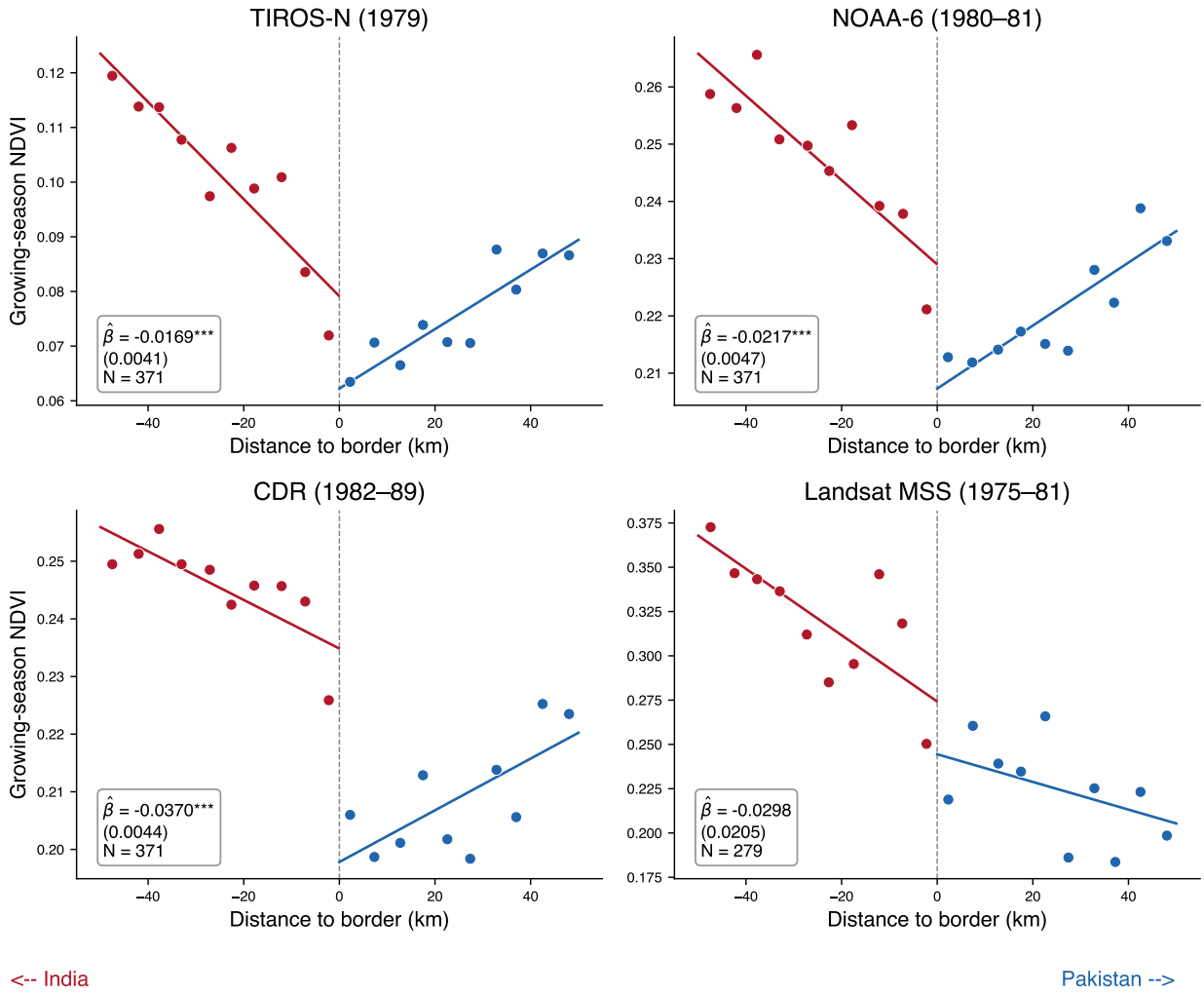
$$NDVI_i = \alpha + \beta \, India_i + \gamma \, R_i + \delta \, (India_i \times R_i) + \varepsilon_i \tag{2}$$

where $NDVI_i$ is the maximum NDVI for cell i, $India_i = 1$ if cell i is on the Indian side of the border, R_i is the distance to the border, and the interaction $India_i \times R_i$ allows the slope to differ

ing in asymmetric confidence intervals.

7. There is existing evidence, e.g. Wuepper et al. (2023), that more recently captured satellite imagery can detect even relatively small discontinuities in yields at administrative boundaries.

Figure 3: AVHRR-MSS Border-RD Validation: India–Pakistan Punjab



This figure applies the same border-based regression-discontinuity validation approach to the India–Pakistan Punjab border using AVHRR NDVI. The horizontal axis is signed distance to the border and the vertical reference marks the border cutoff; binned means and fitted trends are shown on each side. The plot checks whether the satellite outcome and RD design recover a clear discontinuity in a setting with a known policy and productivity divide.

on each side. The coefficient β captures the discontinuous jump in NDVI at the border—a positive estimate indicates higher agricultural productivity on the Indian side of the Punjab.

Figure 3 reports the results across all four satellite sensors. In the three AVHRR sensors, the discontinuity is highly significant ($\hat{\beta} = -0.017$ to -0.037, $p < 0.001$); in Landsat MSS, the estimate is similar in magnitude ($\hat{\beta} = -0.030$) though less precise due to sparser coverage.[8] Crucially, the AVHRR confidence intervals are tight enough to reject discontinuities smaller than 2–3 percentage points of NDVI, well below the yield gains attributed to the HRS.

8. The regression discontinuity slope is negative on the Pakistani side in Landsat MSS but positive for the AVHRR sensors. However, Table A.1 shows that this coefficient is not statistically significant.

Second, we show that NDVI responds to weather shocks—sources of variation that are clearly exogenous and strong enough to affect agricultural yields. Figure 4 shows that an event-study design—similar to the design we use to evaluate the HRS—has enough power to detect weather-driven yield changes. For each event time $k \in \{-3, \ldots, +3\}$, we estimate a separate local projection regression[9] (Jordà 2005):

$$z(NDVI)_{i,t} = \beta_k \, z(Rain)_{i,t-k} + \alpha_i + \tau_t + \varepsilon_{i,t} \tag{3}$$

where $z(Rain)_{i,t-k}$ is standardized growing-season rainfall k years before observation t, and α_i and τ_t are county and year fixed effects. The coefficient β_k at $k = 0$ captures the contemporaneous effect of rainfall on NDVI; the pre-event coefficients ($k < 0$) serve as placebo checks, since future weather should not predict current NDVI.

In both panels, the pre-event coefficients are close to zero and statistically insignificant, confirming that the design passes the placebo check. At $k = 0$, rainfall significantly reduces peak NDVI in both sensors: by 0.08 standard deviations with AVHRR ($p < 0.01$) and 0.16 standard deviations with Landsat MSS ($p < 0.05$), consistent with waterlogging and reduced insolation during China's monsoon growing season.[10] The magnitudes of these weather effects are comparable to the claimed effects of the HRS, demonstrating that both satellite sensors and the event-study design have sufficient power to detect yield shocks of this size.

3.3 Treatment Data

We measure the adoption of the HRS across China using two sources, mapped in Appendix Figure B.1.

First, we use official data on the share of teams in each province who had adopted the HRS by the end of the year from Lin (1992).[11] The data for 1981 and 1982 were originally sourced from the Research Center for Rural Development of the State Council, and the data for 1983 and 1984 were pulled from the China Agricultural Yearbook. To transform this share into a binary outcome for our regression discontinuity design, we code a province as having been "treated" with the HRS if the share of work teams in the province passes a certain threshold. B.1a plots when provinces' share of work teams adopting the HRS exceeds 50%.

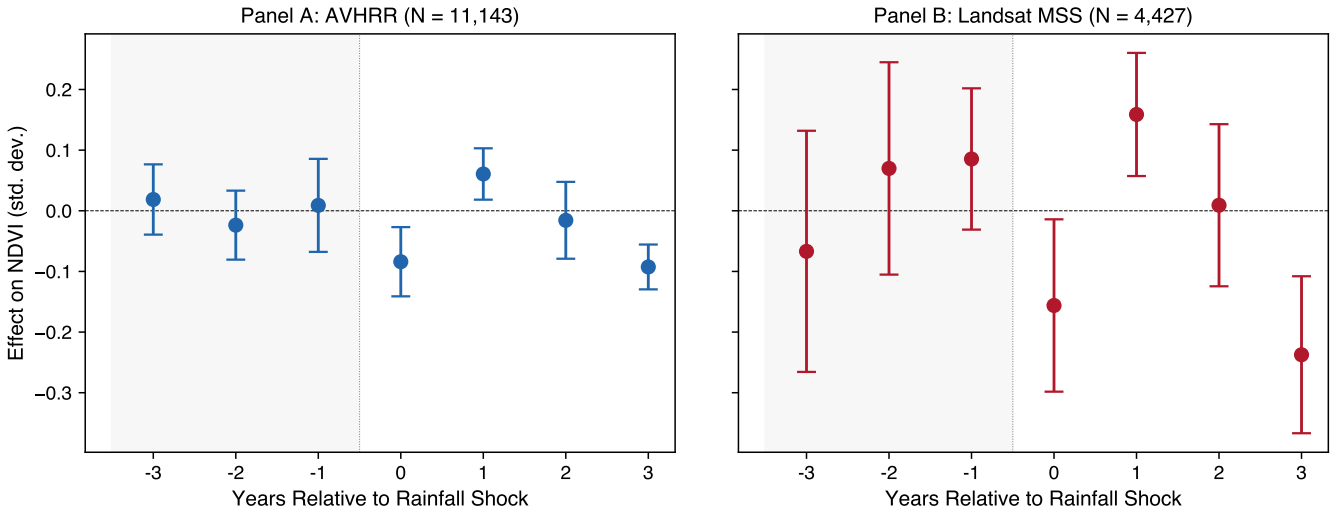
Second, we use county-level rollout data of the HRS collected from gazetteers by Almond, Li, and Zhang (2019). Gazetteers are unofficial records compiled locally largely for historical purposes, and have become an increasingly popular alternative in China studies to official data (Looney 2008). In Almond, Li, and Zhang (2019)'s binary definition, a county becomes treated

9. Unlike the HRS, rainfall shocks are recurrent, generating overlapping treatment periods. Because of this, the staggered difference-in-differences estimators used in Section 4 are not compatible with studying their effects.

10. The larger MSS point estimate may appear surprising given that MSS is a weaker predictor of administrative yields than AVHRR (Table 1). However, Landsat MSS's higher spatial resolution (80m vs. AVHRR's 8km) may allow it to capture localized weather responses more precisely at the county level, even as its sparser temporal coverage makes it a noisier predictor of yields overall.

11. We thank Justin Yifu Lin for generously sharing this data.

Figure 4: Rainfall Event Study: Effect of Rainfall Shocks on Satellite NDVI



This figure reports event-study coefficients from separate regressions of standardized county-level max NDVI on standardized growing-season rainfall at event time $k \in \{-3, \ldots, +3\}$, with county and year fixed effects. Panel A uses AVHRR NDVI (1982–1990); Panel B uses Landsat MSS NDVI (1978–1990, excluding pre-Landsat 3 years and urban-misclassified counties). The shaded region marks pre-event periods ($k < 0$), which serve as placebo checks. Standard errors are clustered by province; error bars show 95% confidence intervals.

with the HRS when "collectively owned land was first contracted to individual households in a few villages".

These treatment data series do not entirely agree. Figure B.7 compares the official provincial-level HRS treatment dates in Lin (1992) to the county gazetteer reporting of the beginning of HRS adoption, averaged up to the province level, from Almond, Li, and Zhang (2019). In every province, adoption of the HRS appears earlier in the gazetteers than in the provincial data. This discrepancy likely reflects definitional differences—Lin (1992) reports a continuous share of work teams who have adopted the HRS, while Almond, Li, and Zhang (2019) reports a binary outcome for when the first villages in a county *begin* adopting the HRS. But it also may reflect differing reporting incentives. Early in the reform period, household farming was technically illegal and the target of suppression by conservative officials (Chung 2000). Later, once the HRS became national policy in 1982, laggard provinces became incentivized to report rapid catch-up.

Rather than take a final stance over which data source is the absolute truth, we choose to use *both* sources, applying them to two separate identification strategies. Moreover, by using the same treatment variation as two major papers which support the current consensus, we ensure that any null results cannot be attributed to differences in treatment measurement.

4 Empirical Results

4.1 Empirical Strategy

Identifying the effects of the Household Responsibility System presents several challenges. The first problem is negative selection. Provinces and counties that were poorer or hard-hit were more likely to adopt reform earlier—famously, the first decollectivization experiments started in Anhui in 1978 as a response to drought. Appendix Table B.3 confirms this statistically: counties that adopted the HRS by 1979 or 1980 had significantly lower baseline grain yields and slower pre-reform yield growth[12]. Naive statistical comparisons of early and late adopters could thus confound the HRS's treatment effect with mean reversion. A second problem is simultaneous treatment: the spread of the HRS in the late 1970s overlapped with the 1979 procurement price reform. The price increases for a given crop were similar nationally, but areas varied in their exposure based on their crop mix. In Section 5, we will separately assess the procurement reform using a shift-share measure that leverages variation in crop suitability.

To address these concerns, we employ two complementary empirical approaches: a difference-in-discontinuities design and an event study design. Each leverages a distinct source of treatment variation: the difference-in-discontinuities approach uses the official Lin (1992) provincial adoption data, while the event study uses the unofficial county gazetteers from Almond, Li, and Zhang (2019). Moreover, each design complements some of the potential weaknesses of the other—together, they help present a coherent narrative of the HRS.

Our differences-in-discontinuities design (subsection 4.2) exploits staggered HRS adoption across provincial borders, using 5km grid cells as the unit of observation. We compare the gap in NDVI between provinces that adopted the HRS earlier and provinces that adopted it later, just across the border. Cell fixed effects absorb time-invariant geographic differences, and the fine spatial resolution both ensures that cells on either side of a border share similar growing conditions and provides sufficient statistical power to detect small effects. Moreover, by comparing cells just on either side of the same boundary, this design directly addresses negative selection, as villages just on either side of the provincial border likely experience very similar trends in unobserved determinants of agricultural productivity. We do not, however, require that there are no permanent differences in productivity across the border. The identifying assumption is that the discontinuity in untreated potential outcomes at the border,

$$\Delta_t(0) \equiv \lim_{r \downarrow 0} \mathbb{E}[Y_{it}(0) \mid R_i = r] - \lim_{r \uparrow 0} \mathbb{E}[Y_{it}(0) \mid R_i = r], \tag{4}$$

is stable over time: $\Delta_t(0) = \Delta_{t'}(0)$ for all t, t'. The border may have a fixed discontinuity reflecting persistent differences in provincial institutions or infrastructure, but this (expected) discontinuity is assumed to not change over time except because of the HRS. This is analogous to the parallel

12. We see qualitatively similar patterns for provinces, but these are not statistically significant, likely due to the small sample.

trends assumption in a typical difference-in-differences design, though it only need hold at the border. This assumption is generally weaker than the continuity assumption needed for a cross-sectional geographic RD ($\Delta_t(0) = 0$), and also weaker than a standard difference-in-differences parallel trends assumption, which would require parallel trends in untreated potential outcomes for all units on either side of the border rather than exclusively at the border.

Our event study design (subsection 4.3) applies recently-developed staggered event study estimators to the county gazetteer treatment data from Almond, Li, and Zhang (2019). The key assumption is that, absent the HRS, county-level NDVI in a given year would have evolved similarly to NDVI in counties that had not adopted HRS by that year. This parallel trends assumption is somewhat more stringent than in the border discontinuity design, but can be partially empirically supported by estimating pre-trends. Moreover, this county-level design addresses the concern that effects observed at the border may be merely local, i.e., that the local average treatment effect is a poor approximation of the average treatment effect. It is also less susceptible to spillover concerns: if news of the HRS spread across provincial borders, our border RD estimates would be biased toward zero, but county-level treatment variation is scattered across both the interior and border regions of provinces (see B.1b). One may be concerned that the treatment statistics we used before—the official provincial data on the share of work teams adopting the HRS—may themselves be manipulated. While still subject to some political reporting pressures, the gazetteer data used in our staggered event studies are viewed by historians as more likely to be critical of the central government (Looney 2008), and thus are an important check for the Lin (1992) treatment data.

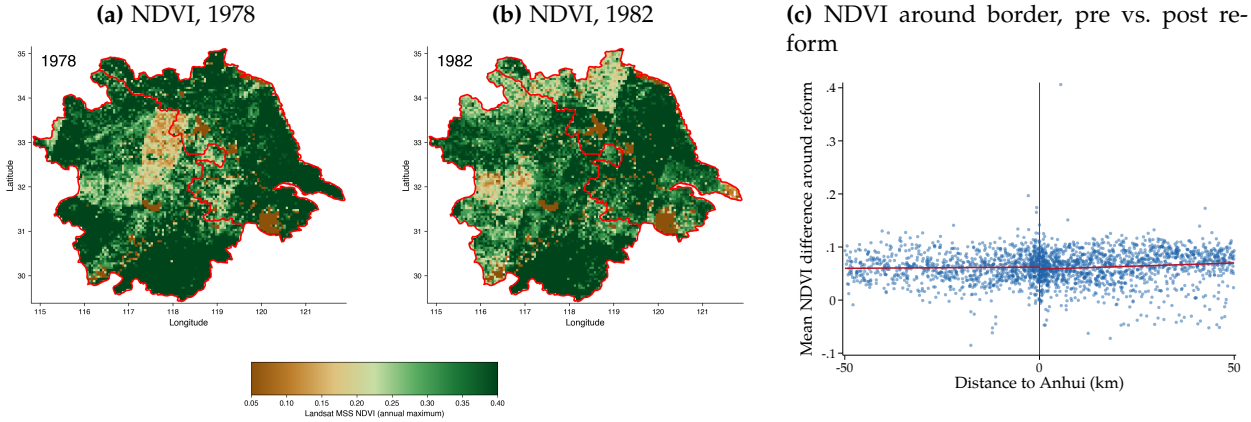
4.2 Differences-in-Discontinuities

Anhui and Jiangsu Case Study A natural first test for the effects of the Household Responsibility System is the provincial border between Anhui and Jiangsu. Starting in 1978, led by reformers like Wan Li, Anhui was a national pioneer in implementing the HRS, while Jiangsu was a conservative holdout. In 1980, Jiangsu's hardline provincial leader, Xu Jiatun, launched a campaign to stamp out the HRS, with the militant slogan "Resolutely Close Up the West Gate [to Anhui]" (Zhang 2019). Wang (1989) notes the lengths taken by Jiangsu's leaders to limit the spread of the HRS:

> In the villages, road intersections, hills, and ends of fields on the border between Jiangsu and Anhui, large slogans were painted, and numerous villages used high-decibel loudspeakers to launch fierce "attacks" on Anhui...

In the official data, by the end of 1981, only 7% of work teams in Jiangsu had adopted the HRS, compared to 85% in Anhui (Lin 1992). Unofficial sources report a similarly large gap—by one estimate, only 12% of Jiangsu work teams had adopted *baogan daohu* by December 1981, compared to 90% in Anhui in February 1982 just three months later (Teiwes and Sun 2016).

Figure 5: Regression Discontinuity around the Anhui-Jiangsu Border



This figure shows maps of Landsat MSS NDVI over Anhui and Jiangsu, before (left, 1978) and in 1982 (middle), the beginning of the Household Responsibility System in Anhui. Both maps use the same green-brown color scale and a shared NDVI legend. The right panel shows cell-level changes in NDVI between pre- and post-reform periods against distance to the border.

In theory, then, if the HRS had an effect on yields, we should expect to see a positive jump in measured agricultural productivity on the Anhui side of the border around 1978. Figure 5 explores this regression discontinuity around the Anhui-Jiangsu border, using Landsat MSS[13]. On the left, 5a shows NDVI within Anhui (left) and Jiangsu (right), with their provincial borders marked in red. There is no clear discontinuity in NDVI. This also reflects the geography of the Anhui-Jiangsu border, which does not follow any major geographic barriers and allows for similar weather conditions on either side. Similarly, in 5b, several years after reform, there is still no obvious discontinuity. On the right, 5c shows cell-level changes in NDVI between the pre-1978 and post-1978 periods (y-axis), scattered against the distance to Anhui (x-axis), with the left of the red vertical zero line representing Anhui. The scatter is almost entirely flat, with no significant gap between the lines of best fit. This suggests that agricultural productivity, as measured by NDVI, did not change differently between the two sides of the border.

We can formalize this graphical intuition by estimating the following specification:

$$y_{i,t} = \beta \left(Anhui_i \times Post_t \right) + f(R_i, t) + \alpha_i + \delta_t + \varepsilon_{i,t} \tag{5}$$

where $y_{i,t}$ is maximum yearly NDVI in Landsat MSS for cell i in year t, $Anhui_i = 1$ if cell i is in Anhui Province, $Post_t = 1$ after Anhui reportedly began adopting the HRS in 1978, $f(R_i, t)$ represents local linear functions of distance to the border (with slopes that may vary by side and, in some specifications, by year), α_i are cell fixed effects, and δ_t are year fixed effects. The coefficient β captures the change in the border discontinuity after decollectivization—that is, whether NDVI grew differentially on the Anhui side post-reform, after controlling for any fixed

13. AVHRR data is not available for most of 1978.

Table 2: Anhui-Jiangsu Regression Discontinuity Estimates

	(1) NDVI	(2) NDVI	(3) NDVI
Anhui x Post-reform	-0.0232***	-0.0113***	-0.00350
	(0.00446)	(0.00299)	(0.00646)
Observations	283116	283116	283116
R^2	0.651	0.652	0.652
Distance Slopes	No	Yes	Yes
Time-varying Slopes	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table reports difference-in-discontinuity estimates of the effect of HRS adoption on maximum yearly NDVI at the Anhui-Jiangsu provincial border, estimated using Equation 5. Anhui is coded as treated: by the end of 1981, 85% of work teams in Anhui had adopted the HRS, compared to 7% in Jiangsu (Lin 1992). The running variable is distance to the provincial border, estimated with local linear fits on each side. 95% confidence intervals are reported in brackets.

differences at the border. We use Landsat MSS for these regressions because of its availability pre-1978.

Table 2 reports estimates of β from Equation 5. Column 1 is the simplest specification with no distance slopes, where $f(R_i, t) = 0$. The coefficient is -0.023 ($p < 0.001$), indicating that NDVI actually *fell* on the Anhui side of the border relative to Jiangsu after the reform. Column 2 adds local linear slopes in distance to the border on each side, or $f(R_i, t) = R_i$. The estimate shrinks to -0.011 but remains highly significant. Finally, Column 3 allows these slopes to flexibly vary over time, where $f(R_i, t) = R_{i,t}$. The estimate attenuates further to -0.004 and is no longer statistically significant. Across all three specifications, there is no evidence that the HRS increased agricultural productivity at this border—the point estimates are uniformly negative and close to zero. Had the HRS generated even a modest improvement in yields, we would expect a positive discontinuity on the Anhui side; instead, the data are consistent with no effect at all. Figure A.12 in the appendix shows the corresponding year-by-year Anhui-minus-Jiangsu coefficients relative to 1977 and likewise shows no discrete positive break after 1978. Taken together, there is little *prima facie* evidence of a divergence in yields between one of the earliest pioneers of the HRS and one of the most stalwart holdouts.

National Difference-in-Discontinuities We can extend the logic of this Anhui-Jiangsu comparison to *all* provincial boundaries in China. This extension forms the basis for our first main empirical design: a difference-in-discontinuities approach that exploits the staggered rollout of HRS adoption across China's provinces.

To extend the specification given by equation (5) to a setting with multiple provincial pairs and treatment timings we follow the Dube et al. (2023) local projections approach, where treatment effects are estimated with separate regressions for each horizon h. Formally, for cell i in

Table 3: National Difference-in-Discontinuity Estimates

Panel A: Binary (> 75%) treatment						
	Landsat MSS			AVHRR		
HRS share > 75%	-0.00977	-0.00800	0.0288*	-0.00146	-0.00341	-0.00117
	(0.0125)	(0.0125)	(0.0129)	(0.00200)	(0.00211)	(0.00147)
Observations	3,061,126	3,061,126	3,061,126	1,439,311	1,439,311	1,439,311
Distance Slopes	No	Yes	Yes	No	Yes	Yes
Time-varying Slopes	No	No	Yes	No	No	Yes
Panel B: Continuous treatment						
	Landsat MSS			AVHRR		
HRS share	0.0358	0.0375	0.122**	0.00105	-0.00133	-0.00133
	(0.0495)	(0.0487)	(0.0437)	(0.00373)	(0.00409)	(0.00320)
Observations	3,061,126	3,061,126	3,061,126	1,439,311	1,439,311	1,439,311
Distance Slopes	No	Yes	Yes	No	Yes	Yes
Time-varying Slopes	No	No	Yes	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table reports difference-in-discontinuity estimates of the effect of HRS adoption on maximum yearly NDVI across all available provincial borders, estimated using Equation 6. Panel A reports the binary treatment specification (province treated when HRS adoption exceeds 75%), and Panel B reports the continuous treatment specification (HRS adoption share). In each panel, columns 1–3 use Landsat MSS outcomes and columns 4–6 use AVHRR outcomes. The running variable is distance to the provincial border, estimated with local linear fits on each side.

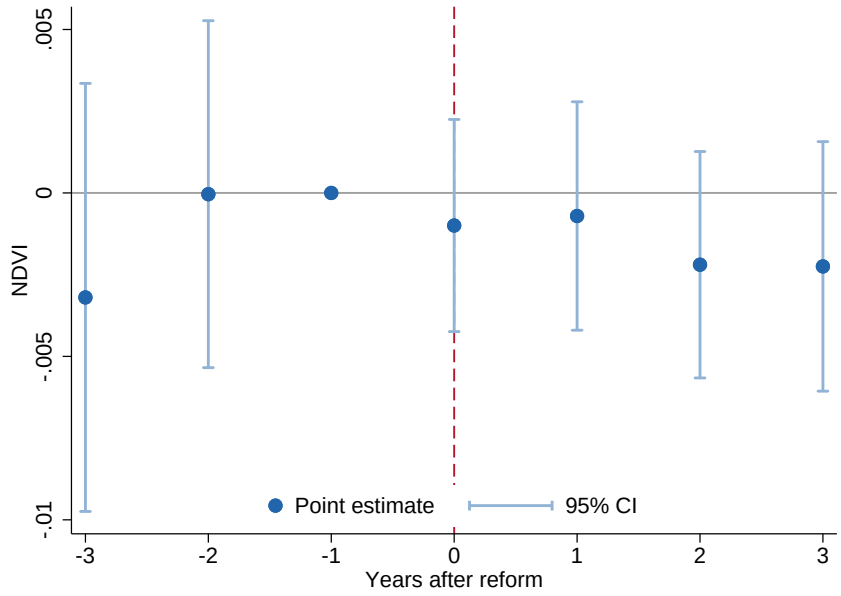
border group b in year t, at horizon h, we estimate

$$y_{i,b,t+h} - y_{i,b,t-1} = \beta_h \Delta D_{i,b,t} + \gamma_{b,t+h}(R_i \times B_{b,t}) + \delta_{b,t+h}(R_i \times B_{b,t} \times \Delta D_{i,b,t}) + G_{b,t} + e^h_{i,b,t} \quad (6)$$

where $D_i = 1$ if over 75% of the work teams in that cell's province have adopted the HRS, R_i is distance to the border, and $B_{b,t}$ is an indicator for being in border group b at time t. We include cell-level fixed effects (implicitly, by differencing out $y_{i,t-1}$), and border group-by-year fixed effects $G_{b,t}$. To avoid overfitting issues, we follow the recommendation of Gelman and Imbens (2019) and use local linear functions of distance for the RD, allowing for different slopes on the running variable across each border for each year.

We select 75% as the treatment threshold because it produces clean pre-trends and a sharp, persistent first stage. Figure B.4 in the appendix illustrates the logic: for each candidate threshold (25%, 50%, 75%, and 90%), we plot the event-study coefficients from regressing the *continuous* HRS adoption share on the binary treatment indicator. These first-stage event studies serve two purposes: they show how large a gap in HRS implementation intensity the binary variable opens between early- and late-adopting provinces, and they track whether pre-trends are flat and whether the gap persists after treatment. At the 75% cutoff, the pre-treatment coefficients are close to zero, the treatment effect on continuous HRS share is immediate and large, and it

Figure 6: Differences-in-discontinuities effect of HRS adoption on maximum NDVI



This figure shows the event study of maximum yearly NDVI following the treatment of provincial decollectivization, when 75% of the province's work teams report having adopted the HRS, estimated using Equation 6. The bars show the 95% confidence intervals around the point estimates, where the standard errors are clustered at the border-group level.

persists through the post-period. Figure B.5 confirms that the reduced-form null result on NDVI is robust to alternative thresholds of 25% and 50%.

We estimate this equation over our cell-level observations over all provinces in China from 1978 to 1990. Our baseline specification includes observations that are up to 50 kilometers away from the border. Our main object of interest is β_h, the change in the border discontinuity in the outcome y between a province not yet treated with land reform and one that experienced land reform h years ago. This β_h is the pooled effect over *all* provincial border pairs in China. In our results, we report β_h for each time horizon h, as an event study.

Figure 6 plots our estimates of the border discontinuity in maximum yearly NDVI at each year-horizon h after decollectivization, β_h from Equation 6. We do not find evidence that decollectivization increased NDVI at the border. Effect sizes are near 0 for all time horizons, with the tight standard error bounds—which we compute clustered by each of the 81 border groups—able to rule out positive effects as small as 0.01 NDVI units. Even 3 years after the onset of reform, the effects of decollectivization are not statistically different from 0 at the 95% confidence level, and the upper end of the confidence interval is smaller than 0.01 NDVI units.

Converting NDVI to Yield The estimates in Figure 6 are denominated in raw NDVI units. We convert them to implied yield effects using the NDVI–yield relationship from Table 1: in the China county sample, a one standard deviation increase in AVHRR NDVI is associated with

a 0.085 SD increase in grain production. Given that the standard deviation of AVHRR NDVI is 0.098 and mean county grain production is approximately 12,100 tons, any raw NDVI effect Δ_{NDVI} implies a percentage change in yield of

$$\frac{\Delta y}{\bar{y}} = \hat{\beta}_{\text{std}} \times \frac{\Delta_{\text{NDVI}}}{\sigma_{\text{NDVI}}} \times \frac{\sigma_y}{\bar{y}} = 0.085 \times \frac{1}{0.098} \times \frac{10{,}600}{12{,}100} \times \Delta_{\text{NDVI}} \approx 0.76 \times \Delta_{\text{NDVI}}.$$

Across all time horizons in Figure 6, the point estimates are near zero, and the upper bound of the 95% confidence interval never exceeds 0.005 NDVI units—implying a yield effect of at most 0.4% of mean county grain production. To obtain an absolute upper bound that also accounts for uncertainty in the NDVI–yield relationship itself, we replace $\hat{\beta}_{\text{std}}$ with the upper end of its 95% confidence interval ($0.085 + 1.96 \times 0.009 \approx 0.102$), giving a conversion factor of approximately $0.91 \times \Delta_{\text{NDVI}}$ and an implied upper bound of 0.5%. Even this most conservative estimate is two orders of magnitude smaller than the effects reported in the existing literature: Lin (1992) estimates a 47% increase in the gross value of farm output, McMillan, Whalley, and Zhu (1989) attributes 78% of TFP growth to the HRS, and Almond, Li, and Zhang (2019) estimates a 15% increase in grain per capita. Our upper bound is 30–200 times smaller, confirming that the null result is economically meaningful and not merely a consequence of imprecise measurement.

Robustness Checks We subject our national-level spatial discontinuity results to a battery of robustness checks. The results are similar across specifications.

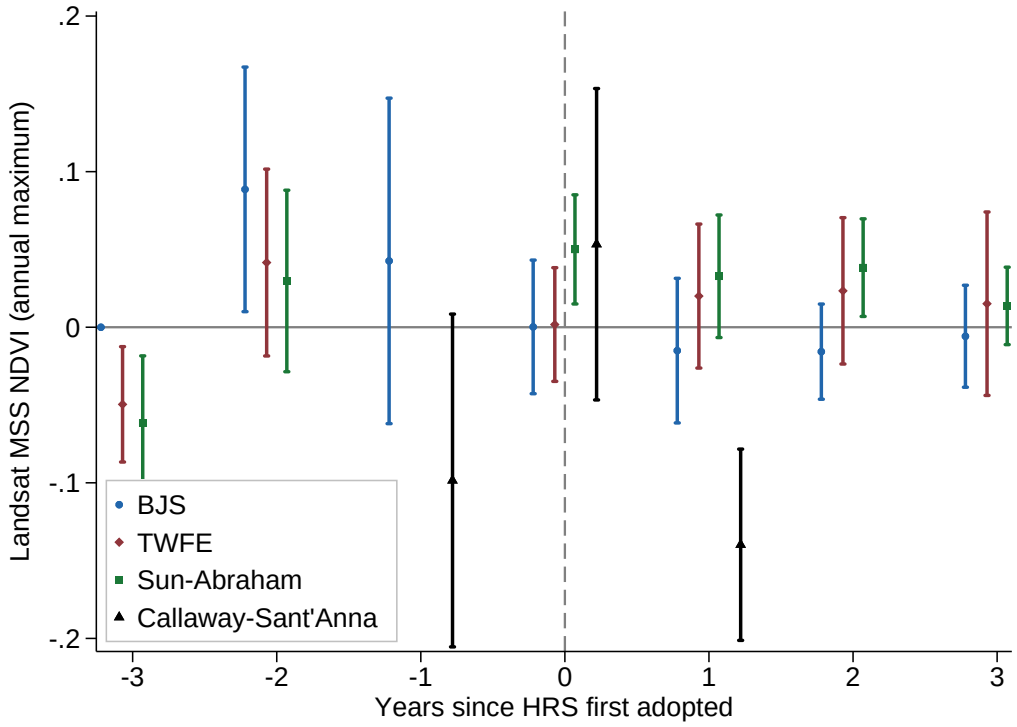
First, to see if our results are potentially being driven by outlier provinces, Figure A.11 estimates our border RD specification, leaving out one province from the sample at a time. The coefficients remain stable and near-zero across omissions.

Second, we test if our results are sensitive to the RD distance bandwidth parameter. Table B.4 shows that the main estimates are stable when we narrow the bandwidth from 50km to 25km. Moreover, to show that the results are not being driven by artifacts immediately along provincial boundaries, we impose 5km and 10km "donut holes", dropping the grid cell observations that are closest to the border. The results are similar.

Third, we relax the binary-treatment specification, Table B.6 uses quartiles of HRS adoption intensity, and Figure B.4 shows that the first-stage border discontinuity remains visible across alternative treatment cutoffs (25%, 50%, 75%, 90%). Figure B.5 confirms that the reduced-form null result on NDVI is robust to re-estimating with 25% and 50% treatment cutoffs. We also report appendix checks using alternative treatment series (Figure B.7 and Table B.9) and find similar results.

Finally, we test if spillovers are driving our null result. Figure B.6 illustrates the intuition underlying our test graphically; if implementation of the HRS spilled over across province borders (for instance, through word of mouth), we might expect NDVI on the late-adopter side to be highest immediately at the border, then decay towards the interior. Spillovers of this nature, with adoption of HRS on one side increasing yields and NDVI immediately across the border, would

Figure 7: Effect of County-Level HRS Adoption on Landsat NDVI



This figure shows the event study of annual maximum Landsat MSS NDVI following county-level HRS adoption from Almond, Li, and Zhang (2019) gazetteers, estimated using four different estimators: Borusyak, Jaravel, and Spiess (2023) imputation (BJS), two-way fixed effects (TWFE), Sun and Abraham (2021) interaction-weighted, and Callaway and Sant'Anna (2021). The x-axis shows years relative to when collectively owned land was first contracted to individual households. All specifications include county and year fixed effects. Standard errors are clustered at the county level.

attenuate our estimates relative to the true effect. Table B.5 tests spatial spillovers by allowing treatment effects to vary with distance on the early- and late-adopter sides of each border. We do not find evidence that effects are larger near the border on the late-adopter side.

4.3 County Rollout

We next consider the effect of HRS on productivity at the county level, using the Almond, Li, and Zhang (2019) binary variable for when villages in a county start adopting the HRS. This county rollout design complements the differences-in-discontinuities design, addressing the concern that effects at provincial borders may be merely local and unrepresentative of the broader effects of the HRS. Moreover, unofficial gazetteer data is not as vulnerable to concerns about data reliability or bias that apply to official provincial treatment data.

For county i at time t, we assume potential outcomes follow a two-way fixed specification:

$$y_{i,t} = \alpha_i + \delta_t + \sum_{h \in H} \beta_{i,t,h} 1\{t - E_i = h\} \tag{7}$$

where E_i is the year of HRS adoption in county i, $H \subseteq \mathbb{Z}_{\geq 0}$ a subset of non-negative integers denotes periods after adoption in which there is an effect of HRS adoption, and $y_{i,t}$ is county-level annual maximum Landsat MSS NDVI[14]. The untreated potential outcomes, given by $y_{i,t}(0) = \alpha_i + \delta_t$, are assumed to satisfy the standard parallel trends assumption: that absent HRS adoption NDVI would have evolved in treated counties similar to in untreated counties. For a given $h \in H$, the observation-specific treatment effect is $\tau_{i,t} = \beta_{i,t,h}$.

In models such as this—with unrestricted treatment effect heterogeneity across units, time, and horizon—the standard TWFE estimator is known to be inconsistent for the ATT and indeed does not necessarily correspond to any estimand that is a convex weighted average of treatment effects (De Chaisemartin and dHaultfoeuille 2020). In light of this, we use a number of recently-developed estimators that do consistently estimate causal estimands in addition to the traditional TWFE estimator.

fig:county_result_did<empty citation> presents results from four estimators: the Borusyak, Jaravel, and Spiess (2023) imputation estimator (BJS), two-way fixed effects (TWFE), the Sun and Abraham (2021) interaction-weighted estimator, and Callaway and Sant'Anna (2021). For all analyses we cluster standard errors at the county level. Before treatment, we generally do not observe evidence of statistically significant pre-trends in NDVI, supportive of the parallel trends assumption. In the few period - estimator pairs that do exhibit statistically significant pre-trends the magnitude of the estimates is small.

Turning to the post-treatment period, we find similar results to the differences-in-discontinuities design: we find no consistent evidence of meaningful increases in agricultural output after the adoption of the HRS. The majority of point estimates are not statistically different from zero at conventional confidence levels and the few point estimates that are statistically significant are very small in magnitude or negative. Broadly, the four estimators all point to the same conclusion: the HRS had a negligible effect on agricultural output up to three years after initial adoption. These results suggest that differences between core and border regions of provinces are not likely drivers of our difference-in-discontinuities findings. In other words, both research designs and treatment definitions we employ generate consistent findings - that the effect of the HRS on agricultural output per unit area was small and statistically indistinguishable from 0.

Converting to yield units using the MSS NDVI–yield relationship from Table 1 (and, conservatively, the upper end of its 95% confidence interval), the BJS upper bounds imply yield effects of at most 5% of mean county grain production—still far below the 15–47% effects reported in the prior literature. Appendix Figure B.9 shows that the null result is robust to alternative vege-

14. We use Landsat MSS because consistency between sensors allows us to form a consistent panel. AVHRR, by contrast, employed three different sensors over this period, which harms temporal comparability.

tation measures, including harvest-month MSS NDVI, AVHRR NDVI with a cropland mask, and z-scored MSS NDVI. Appendix Figure B.10 confirms that growing degree days do not change discontinuously at county-level HRS adoption, ruling out coincident weather shocks as a confound.

4.4 Interpreting Our Findings

Our findings clash with a longstanding academic consensus that the Household Responsibility System was the primary driver of China's agricultural takeoff.

Lin (1992) Table B.1 replicates the province-level results of Lin (1992). Column 1 corresponds to Lin (1992)'s main two-way fixed effect result, where the log value of farm output is the outcome variable. The coefficient on HRS share exactly matches table 5, column 6 from Lin (1992). Column 2 performs the same regression, using log grain output as the outcome variable, and finds a similar strong and significant effect of the HRS on growth. Columns 3 and 4 estimate the same regressions as columns 1 and 2, respectively, but substitute in the provincial averages of county HRS adoption from Almond, Li, and Zhang (2019) for the official provincial-level data from Lin (1992). (Note that the sample size is roughly halved.) However, in these regressions the gazetteer-reported HRS share is not predictive of growth either in the value of farm output or overall grain production.

Table B.2 repeats this exercise using log province-level AVHRR NDVI as the outcome variable, providing a direct test of whether satellite-measured vegetation responds to the HRS. Because AVHRR data begin in 1978, the sample is restricted to 223 province-year observations. Using Lin (1992)'s official HRS shares, the coefficient is positive and significant ($\hat{\beta} = 0.323$, $p < 0.01$), suggesting that provinces with greater reported HRS adoption experienced higher satellite greenness.[15]

These results also serve as a validation of the satellite data: province-level AVHRR NDVI is responsive to the same variation that drives Lin (1992)'s original findings, confirming that the satellite measures capture meaningful agricultural signals at a high level of aggregation. At the same time, the fact that the NDVI effect disappears with the gazetteer-based HRS measure—just as the VFO and grain effects do—suggests that the problem lies not with the outcome data but with the identification strategy. The official provincial HRS shares may be endogenous to agricultural productivity growth, generating a spurious correlation that appears across all outcome measures.

Almond, Li, and Zhang (2019): Yield vs. Labor Productivity We apply modern staggered event-study estimators to the county-level gazetteer data from Almond, Li, and Zhang (2019). One limitation of this data is that it does not contain an estimate of land under cultivation, meaning that we cannot directly calculate yield. Appendix Figure B.8 reports event studies for

15. We find similar null results when we substitute in the Almond, Li, and Zhang (2019) gazetteer treatment data.

Almond, Li, and Zhang (2019)'s gazetteer outcomes using three estimators: Borusyak, Jaravel, and Spiess (2023) imputation (BJS), TWFE, and Callaway-Sant'Anna. The grain-per-capita series rises after reform in the BJS specification, but the effect on log county grain output is imprecise and not robust across estimators.[16]

Why do our findings differ? One important distinction is that satellites measure land productivity—grain output per hectare—but not labor productivity. Our findings are entirely consistent with the HRS having raised labor productivity by freeing households to allocate their time more flexibly. Indeed, in Section 6 we find evidence that HRS adoption increased built-up settlement area, consistent with households redirecting labor toward non-agricultural activities and the explosive growth of Township and Village Enterprises. However, grain yields were the central claim of the HRS literature: Lin (1992) estimated a 47% increase in the gross value of farm output, McMillan, Whalley, and Zhu (1989) attributed 78% of TFP growth to the HRS, and Almond, Li, and Zhang (2019) estimated a 15% increase in grain per capita. Our null finding challenges these claims on their own terms.

Qualitative Evidence Our null result on grain yields is consistent with a revisionist literature that emphasizes the continuities between the pre-reform collectives and the post-reform household system (Eisenman 2018). If the collectives were less dysfunctional than the standard narrative suggests—for instance, if work teams had already developed effective monitoring and incentive mechanisms (Eisenman 2018)—then dismantling them would produce smaller efficiency gains than predicted by models of pure shirking (Lin 1988). More broadly, recent historical scholarship with access to party archives has found little policy daylight between Hua Guofeng and Deng Xiaoping on rural reform, suggesting that the political narrative crediting Deng's market-oriented reforms may have overstated the discontinuity of the HRS (Teiwes and Sun 2016; Torigian 2022). We discuss these revisionist perspectives further in Section 7.

5 Procurement Price Reform

If the HRS did not cause yields to grow, what did? The obvious candidate—and the only other major national-level agricultural policy change during the same period—is the 1979 rise in procurement prices, where the state raised the average price for quota grain by 20% and the bonus for above-quota grain from 30% to 50% (Sicular 1988a). (Figure C.2 shows the evolution of these procurement prices by crop.) Since there was limited variation in the price increases by province, their effects are not captured in the border discontinuity design.

5.1 Theoretical Framework

We present a simple theoretical framework to motivate how an increase in procurement prices can increase yields, with yield responses a function of the local suitability of crops.

16. Almond, Li, and Zhang (2019) did not report confidence intervals.

Individuals living in region r, $i \in N_r$, have preferences over leisure, measured as $1 - l_i$ (the share of time not spent working), and income $l_i w_r$. Income is the product of the share of time spent working during period t and the current regional wage, $w_{r,t}$, which is taken to be exogenous by the worker. Preferences are given by

$$U_i(l_{i,t}; w_{r,t}) = \sigma^{-1}(w_{r,t} l_{i,t} - \psi l_{i,t}^{\theta})^{1-\sigma} \tag{8}$$

where ψ determines the scale of disutility from working, and θ determines the wage elasticity of labor supply. Utility maximization implies that individual labor supply is given by $(w_{r,t}/\theta\psi)^{1/(\theta-1)}$, and total labor supply in region r at time t as $L_{r,t} = N_r(w_{r,t}/\theta\psi)^{1/(\theta-1)}$. Importantly, individual and regional labor supply are both monotonically increasing functions of wage.

Crop output is produced on a discrete set of plots using a Cobb-Douglas production function combining plot$\times$crop-specific productivity, $A_{p,k}$, and total labor employed on the plot, $L_{p,t}$:

$$Q_{p,t,k} = A_{p,k}^{1-\gamma} L_{p,t}^{\gamma} \tag{9}$$

Note that the productivity term comes from the inelastic supply of plots to agricultural production. The only production decisions that need to be made at the plot-level are what crop to allocate the plot to and how much labor to employ. Profit maximization implies that the marginal productivity of labor is equalized across crops:

$$\gamma P_{r,t,k} A_{p,k}^{1-\gamma} L_{p,t}^{\gamma-1} = w_{r,t} \tag{10}$$

where $P_{r,t,k}$ is the price of crop k in region r at time t. Despite prices being exogenous in this setting, work teams still had private incentives to optimize the allocation of effective labor that is actually employed in production across plots. Solving this expression for plot-level labor employment, we get

$$L_{p,t} = \left(\frac{\gamma P_{r,t,k} A_{p,k}^{1-\gamma}}{w_{r,t}} \right)^{1/(1-\gamma)} \tag{11}$$

We next turn to the allocation of plots to crops. Since wages and labor productivity are constant across all crops on a given plot, the managers' optimization problem over crops simplifies to choosing k to maximize $A_k P_{r,t,k}^{1/(1-\gamma)}$. Define $k_p^* := \arg\max_k \{A_k P_{r,t,k}^{1/(1-\gamma)}\}$ to be this plot-specific optimal crop choice.

To close the model, we use a labor market clearing condition. The sum of labor demand

across plots must be equal to total regional labor supply

$$N_r \left(\frac{w_{r,t}}{\theta \psi} \right)^{1/(\theta-1)} = \left(\frac{\gamma}{w_{r,t}} \right)^{1/(1-\gamma)} \underbrace{\sum_{p \in r} A_{p,k_p^*} P_{r,t,k_p^*}^{1/(1-\gamma)}}_{\Phi_{r,t}} \tag{12}$$

We can solve for $w_{r,t}$:

$$w_{r,t}^{1/(\theta-1)+1/(1-\gamma)} = \frac{\gamma^{1/(1-\gamma)} (\theta \psi)^{1/(\theta-1)} \Phi_{r,t}}{N_r} \tag{13}$$

which implies

$$\ln w_{r,t} = \frac{(\theta-1)(1-\gamma)}{\theta-\gamma} \left(\ln c_0 + \ln \Phi_{r,t} - \ln N_r \right) \tag{14}$$

where $c_0 = \gamma^{1/(1-\gamma)} (\theta \psi)^{1/(\theta-1)}$. Note that the effect of a price change affects wages through the term $\Phi_{r,t}$, with greater (productivity weighted) prices increasing the wage. Because k_p^* maximizes $A_{p,k_p} P_{r,t,k}^{1/(1-\gamma)}$, a price increase can only raise wages and thus labor supply. This is the mechanism through which price reform affects agricultural output in the model.

We can obtain an expression for plot-level output by substituting equation (11) into equation (9).

$$Q_{p,t,k_p^*} = A_{p,k_p^*}^{1-\gamma} \left(\frac{\gamma P_{r,t,k_p^*} A_{p,k_p^*}^{1-\gamma}}{w_{r,t}} \right)^{\gamma/(1-\gamma)} = \gamma^{\gamma/(1-\gamma)} w_{r,t}^{-\gamma/(1-\gamma)} A_{p,k_p^*} P_{r,t,k_p^*}^{\gamma/(1-\gamma)} \tag{15}$$

Summing plot-level output within a region and dividing by the number (area) of plots, we obtain region-level yield

$$\mathcal{Y}_{r,t} = |p \in r|^{-1} \sum_{p \in r} Q_{p,t,k_p^*} = |p \in r|^{-1} \gamma^{\gamma/(1-\gamma)} w_{r,t}^{-\gamma/(1-\gamma)} \underbrace{\sum_{p \in r} A_{p,k_p^*} P_{r,t,k_p^*}^{\gamma/(1-\gamma)}}_{\Psi_{r,t}} \tag{16}$$

Taking logs and substituting in our expression for $w_{r,t}$ obtained from the labor market clearing condition, we get:

$$\ln \mathcal{Y}_r = \kappa_r + \frac{\gamma(\theta-1)}{\theta-\gamma} \ln \Phi_{r,t} + \ln \Psi_{r,t} \tag{17}$$

where κ_r collects constants and terms that only vary at the region level. The terms $\Phi_{r,t}$ and $\Psi_{r,t}$ together determine the effect of a change in prices on yield.

5.2 Empirical Results

Informed by this model, we construct shift-share-style variables that interact cross-sectional variation in crop suitability with time-series variation in government procurement prices. For each grid cell c in province $r(c)$ and year t, the relevant productivity-weighted price changes are:

$$\chi_{c,t} = \sum_{k \in \mathcal{K}} \Delta \ln P_{r(c),t,k}^{(1-\gamma)^{-1}} \times S_{c,k} \tag{18}$$

$$\Lambda_{c,t} = \sum_{c \in \mathcal{C}} \Delta \ln P_{r(c),t,k}^{\gamma(1-\gamma)^{-1}} \times S_{c,k} \tag{19}$$

where $\mathcal{K} = \{\text{wheat, rice, sorghum, millet, maize, soybeans, cotton}\}$ denotes the set of crops for which procurement price data are available, $\Delta \ln P_{r(c),t,k}$ is the log change in the government procurement price of crop k in province $r(c)$ between years $t-1$ and t, and $S_{c,k}$ is the GAEZ agro-ecological suitability index for crop k at location c, measured in potential yield (kg/ha) under low-input, irrigated conditions. The variables $\chi_{c,t}$, $\Lambda_{c,t}$ can be interpreted as instruments for $\Phi_{r,t}$ and $\Psi_{r,t}$, respectively, where we have substituted the GAEZ suitability index for (unobservable) land productivity at the endogenously chosen optimal crop mix.

Figure C.1 maps the spatial distribution of growing suitability across China for six major food crops (wheat, rice, sorghum, millet, maize, and soybeans), using the FAO's Global Agro-Ecological Zone (GAEZ) data. Growing suitability is measured using the potential yield (in kilograms per hectare), estimated based on climate, soil, and terrain conditions.

Figure C.2 shows how these six crops' procurement prices changed over time. All six crops saw significant 20% price increases between 1978 and 1979. This rise in prices was largely national: there was limited variation between provinces in their prices for the key staples of wheat and rice, with somewhat greater variation for sorghum, millet, maize, and soybeans.

The identifying variation comes from the interaction of two sources: (i) the differential timing and magnitude of procurement price increases across crops, and (ii) the spatial variation in agronomic suitability across grid cells. Cells more suited to crops that experienced larger price increases receive a larger positive shock. Figure 8 maps the resulting shift-share treatment at the grid-cell level, showing that the central agricultural provinces experienced the greatest exposure to the 1979 price reform.

We estimate the following panel regression:

$$\text{NDVI}_{c,t} = \beta_1 \chi_{c,t} + \beta_2 \Lambda_{c,t} + \alpha_c + \gamma_t + \varepsilon_{c,t} \tag{20}$$

where $\text{NDVI}_{c,t}$ is the mean Landsat MSS NDVI for grid cell c in year t, α_c are grid cell fixed effects that absorb time-invariant characteristics of each location (including the level of crop suitability), γ_t are year fixed effects that absorb common shocks across all cells, and $\varepsilon_{c,t}$ is the error term.

In equation (17), an exogenous change in prices affects yields through the terms $\Phi_{r,t}$, $\Psi_{r,t}$. The analogous terms in the specification given by equation (20) are $\chi_{c,t}$ and $\Lambda_{c,t}$, with the total

effect of the price reform given by the sum of the coefficients $\beta_1 + \beta_2$. Because we use NDVI as the outcome, rather than yields, β_2 reflects the scale difference between these two measures. Similarly, the error term $\varepsilon_{c,t}$ reflects the imperfect correlation between yields and NDVI.

Table 4 reports the results, with standard errors clustered at the province level. The baseline estimate of the total effect of price changes is positive and statistically significant (column 1), and remains so when using maximum NDVI (column 2) or replacing year fixed effects with province $\times$ year fixed effects (column 3). Grid cells with greater exposure to the 1979 procurement price increases, as measured by crop suitability, experienced larger increases in NDVI, consistent with the price reform driving agricultural productivity growth. Figure C.5 reports a placebo version of this exercise in the appendix.

A natural concern is that our measure of price reform exposure may capture the broader HRS institutional reform rather than the price channel specifically, since both reforms occurred in the same period. To address this, column 4 of Table 4 controls directly for HRS adoption, measured as the continuous share of a province's households that had adopted the HRS. The estimated total effect of the price reform is essentially unchanged after including the HRS control, confirming that the procurement price channel is distinct from the institutional reform. The HRS coefficient itself is positive but statistically insignificant. Column 5 controls for a revenue-weighted procurement price index at the province-year level, constructed using 1978 base-period revenue shares of rice and wheat; the estimated total effect is again essentially unchanged. Column 6 uses log NDVI as the dependent variable, yielding a semi-elasticity interpretation.

How large are these effects in yield terms? The HRS difference-in-discontinuities point estimates from Section 4 are essentially zero: in the preferred AVHRR specification, the point estimate is -0.001 NDVI units, implying a yield effect of approximately -0.1%, and the 95% confidence interval rules out effects larger than 0.5%. The procurement price effect, by contrast, is clearly positive: in our baseline specification (column 1 of Table 4), the combined effect of a one-standard-deviation increase in price exposure is $+0.020$ Landsat MSS NDVI units ($p < 0.01$). Applying the MSS NDVI–yield relationship from Table 1 ($\hat{\beta}_{\text{std}} = 0.037$, roughly half the magnitude of the AVHRR relationship), this implies a yield increase on the order of 0.7% per standard deviation of price exposure.

This is modest—but it is a lower bound on the true effect for two reasons. First, our shift-share design identifies only the *differential* response of locations with different crop mixes to the price reform; the common component of the 1979 price increase, which raised procurement prices for all crops nationally, is absorbed by year fixed effects. The total national effect of the price reform is therefore not identified by this design and is likely substantially larger. Second, we use time-invariant agro-ecological suitability as the exposure measure rather than observed pre-reform crop allocations, which are unavailable at the grid-cell level; to the extent that actual cropping patterns deviated from GAEZ suitability, classical measurement error will attenuate our estimates.

These results suggest that the procurement price reform was a significant driver of the agri-

Table 4: Bartik Procurement Price Shock and NDVI

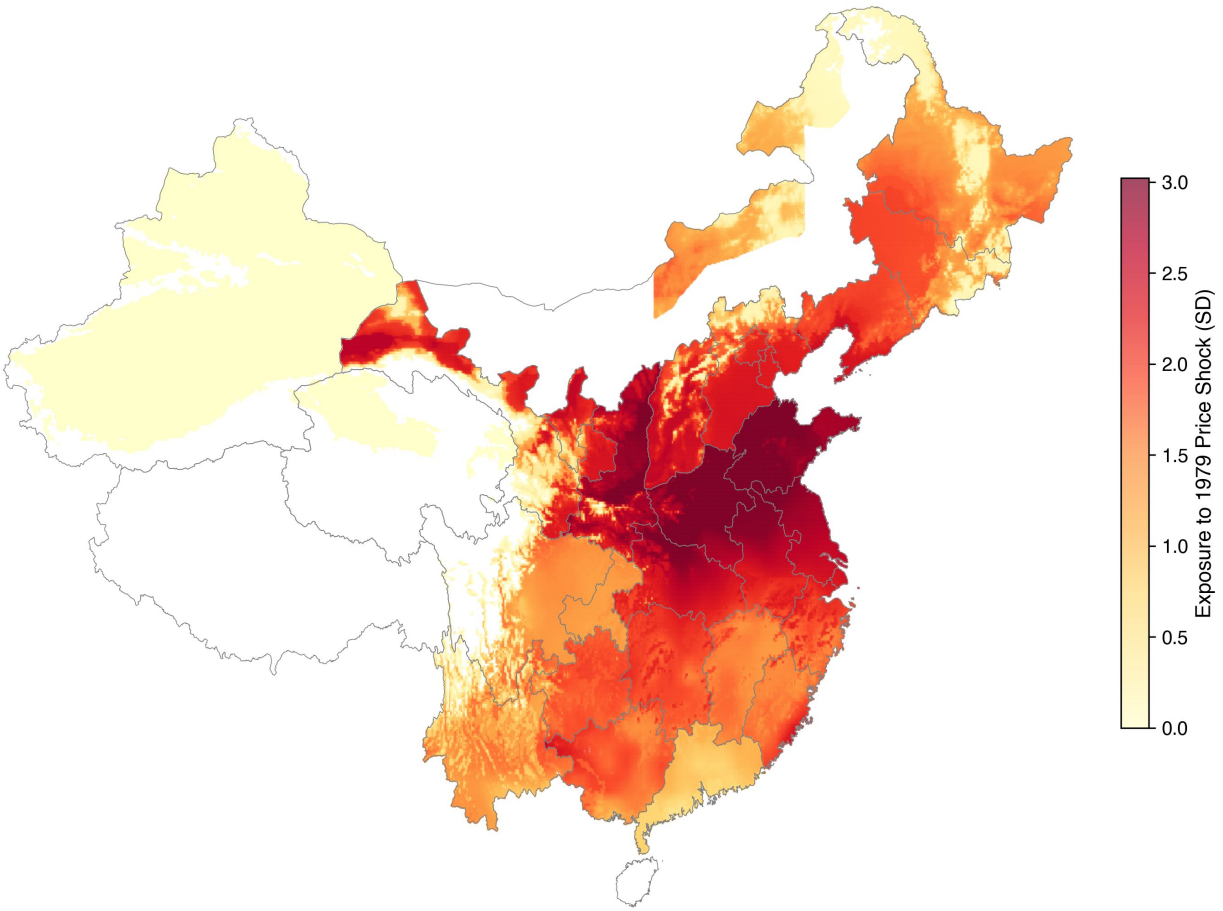
	(1) NDVI mean	(2) NDVI max	(3) NDVI mean	(4) NDVI mean	(5) NDVI mean	(6) Log NDVI
χ (SD) $\times$ Post	-0.180908***	-0.134081*	0.023293	-0.175536**	-0.211721***	-0.904790**
	(0.069687)	(0.071022)	(0.076813)	(0.069318)	(0.072831)	(0.376249)
Λ (SD) $\times$ Post	0.200747***	0.169274**	0.002346	0.195214***	0.231901***	0.972139**
	(0.070679)	(0.073518)	(0.082599)	(0.070278)	(0.073598)	(0.378403)
HRS (continuous)				0.011778		
				(0.012890)		
Log(price index)					0.549494**	
					(0.276246)	
$\hat{\beta}_1 + \hat{\beta}_2$	0.019838***	0.035193***	0.025639**	0.019678***	0.020180***	0.067349**
	(0.005493)	(0.012390)	(0.009767)	(0.005388)	(0.005792)	(0.029588)
N	2,013,306	2,013,306	2,013,306	2,013,306	1,909,478	1,877,195
R^2 (within)	-0.0236	0.0512	-0.0435	-0.0358	-0.4773	-0.0009
Cell FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	No	Yes	Yes	Yes
Prov $\times$ Year FE	No	No	Yes	No	No	No
HRS control	No	No	No	Continuous	No	No
Price index control	No	No	No	No	Yes	No
Cluster	Province	Province	Province	Province	Province	Province

This table reports estimates of the effect of the 1979 procurement price reform on Landsat MSS NDVI, estimated using Equation 20. The dependent variable is the difference-in-difference of NDVI (post minus pre-1979) at the grid-cell level. The independent variable is grid-cell exposure to the 1979 price shock, as defined in Equation 19. Column 1 is the baseline using mean NDVI. Column 2 uses maximum NDVI. Column 3 replaces year FE with province $\times$ year FE. Column 4 controls for HRS adoption (continuous share). Column 5 controls for a revenue-weighted procurement price index, constructed using 1978 base-period revenue shares of rice and wheat. Column 6 uses log NDVI. All specifications include cell fixed effects. Standard errors clustered at the province level are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

cultural productivity growth in China during the late 1970s and early 1980s. The shift-share estimates are robust to controlling for HRS adoption, to alternative fixed effect structures, and to placebo tests—and the HRS coefficient is itself insignificant when both channels are included in the same regression. This pattern is consistent with price reform, not decollectivization, lifting a binding constraint on agricultural output during this period.

Moreover, the timing of the yield surge reinforces this interpretation: as Figure 1 illustrates, national grain yields had already risen by nearly 20% by 1980, when fewer than 20% of work teams had adopted the HRS. The procurement price reform, by contrast, was implemented nationally in 1979 and would have affected planting decisions immediately—suggesting that much of the aggregate yield growth commonly attributed to the HRS may instead reflect the response to higher prices.

Figure 8: Exposure to 1979 Procurement Price Shock



This figure maps the suitability-weighted exposure to the 1979 procurement price reform at the grid-cell level. Exposure is defined as $\sum_{k \in \mathcal{K}} \Delta \ln P_{r(c),1979,k} \times S_{c,k}$, where $\Delta \ln P_{r(c),1979,k}$ is the log change in the provincial procurement price of crop k between 1978 and 1979, and $S_{c,k}$ is the GAEZ suitability index for crop k at location c. Darker shading indicates higher exposure. The central agricultural provinces show the greatest exposure, reflecting both high crop suitability and large price increases.

6 Urbanization and Structural Transformation

6.1 The Rise of Township and Village Enterprises

Did the Household Responsibility System have effects beyond agriculture? One possibility is the potential connection with rural industrialization.

During the reform period, China became known for an unusual pattern of rural industrialization, where millions of small factories appeared not in cities but in villages and rural towns, transforming the countryside into a major site of industrial production. In particular, Township and Village Enterprises (TVEs)—industrial and commercial enterprises, which often sprang from the commune and brigade enterprises of the collectivized period—drove much of this rural industrial growth.[17] During the 1980s, the TVEs were the most dynamic sector of the Chinese economy, with employment growing from 28 million in 1978 to 135 million in 1996, and their share of GDP quadrupling from 6% in 1978 to 26% in 1996 (Naughton 2007). Remarkably, TVEs emerged unexpectedly without the direction of the central Chinese state—in the words of Deng Xiaoping, they were "a new force that just came into being spontaneously" (Vogel (2013), p. 445).

Another important mediating factor during this period was the severe migration restrictions imposed by the hukou household registration system. Under the hukou, rural households were tied to their place of registration and largely barred from migrating to urban areas. The system was only partly relaxed beginning in 1984, and even then significant barriers to migration remained. This suggests that the labor freed up by agricultural reform may have had few outlets beyond local rural industry.

Taken together, the combination of land reform and the restrictions on migration may have inadvertently caused industrialization to cluster in the countryside. A large body of scholarship has drawn a connection between China's agricultural reform and the growth of rural industrialization (Oi 2002; Huang 2008). However, to our knowledge, the linkage between the HRS and the TVEs has not been conclusively shown using modern econometric evidence.

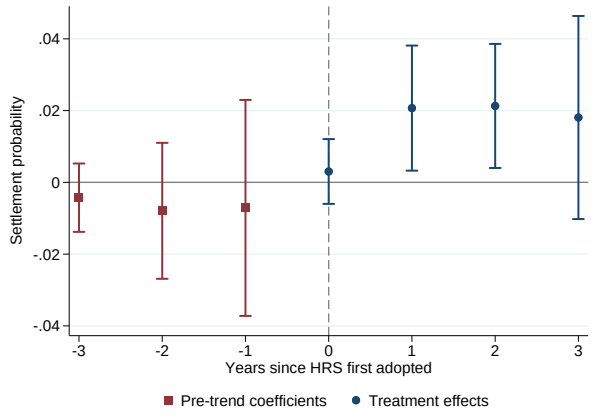
6.2 Measuring Settlement

To study structural transformation beyond agriculture, we track changes in built-up surface around villages using satellite imagery. We train a random forest classifier on Landsat MSS spectral indices to predict settlement probability—the likelihood that a pixel contains at least 5% built-up area, using the World Settlement Footprint (WSF) as a source of training data. This lower threshold captures not just dense urban cores but also the diffuse built-up expansion characteristic of TVE-driven rural industrialization. The appendix reports the core diagnostics for the prediction pipeline: Figure D.1 summarizes hyperparameter tuning; Figure D.2 compares predicted and reference maps; Figure D.3 reports calibration performance; and Figure D.4 shows which features are most predictive out of sample.

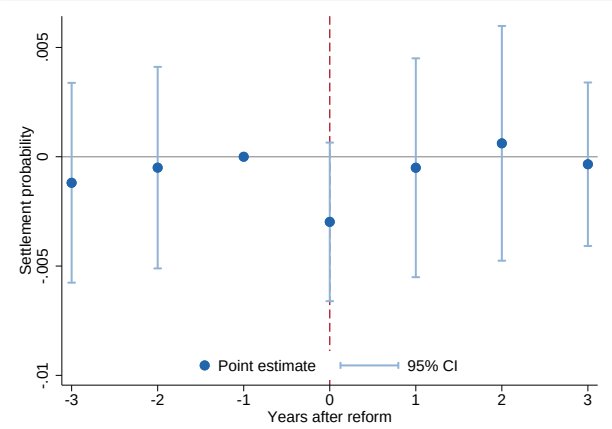
17. Scholars debate the extent to which TVEs were firms run and owned by local governments or private enterprises.

Figure 9: HRS Effects on Settlement Probability

(a) County-level DiD (Borusyak, Jaravel, and Spiess (2023))



(b) Differences-in-discontinuities (Dube et al. (2023))



This figure reports dynamic estimates of the effect of HRS adoption on predicted settlement probability. Settlement probability is the random-forest-predicted likelihood that a Landsat MSS pixel contains at least 5% built-up area. Panel (a) uses the Borusyak, Jaravel, and Spiess (2023) imputation estimator with county-level treatment timing from Almond, Li, and Zhang (2019). Panel (b) uses the Dube et al. (2023) local projections differences-in-discontinuities estimator at provincial borders (50km bandwidth), with border-group-by-year fixed effects and local linear distance controls. Error bars show 95% confidence intervals.

We validate that satellite-measured settlement tracks economic activity by merging with village-level income data from the Contemporary Chinese Village Gazetteer (CCVG), a dataset compiled from unofficial local records. Figure E.1 and Table E.1 show that villages with more surrounding GHSL built-up area have higher log income per capita, with a clear positive gradient in the binscatter and positive coefficients across 500m, 1km, and 5km buffers. Figure E.2 and Table E.2 report the same exercise using WSF urban fraction, with consistent results across buffers and with and without fixed effects.

Figure 9 reports dynamic estimates of the effect of HRS adoption on settlement probability using two complementary designs. Panel (a) uses county-level treatment timing from Almond, Li, and Zhang (2019) and the Borusyak, Jaravel, and Spiess (2023) imputation estimator. Pre-trends are flat, and settlement probability increases by approximately 2 percentage points in the first two years following adoption, before fading by year 3. This is consistent with modest growth in built-up surface following decollectivization—potentially reflecting the early expansion of rural industry and settlement that accompanied the TVE boom. Panel (b) applies the Dube et al. (2023) differences-in-discontinuities design—the same border-based approach used for our main NDVI results—to settlement probability at provincial borders. Here the effect is a precise null: all coefficients are near zero, with confidence intervals that rule out effects larger than 1 percentage point.

The contrast between the two panels has a number of potential interpretations. One is that

the differing results reflect different estimands: the ATT for the county-level result and the LATE for the provincial-level. Another is that the differences in treatment definition are meaningful, and that county-level first adoption captures the onset of HRS adoption better than the provincial share of work teams that had adopted HRS rising above 50%. It may also be the case that the county-level result reflects broad trends correlated with reform timing rather than a causal effect of decollectivization on built-up surface. Overall, we take our findings as suggestive evidence that the adoption of HRS induced an increase in built-up area, consistent with a reallocation of labor away from agricultural production.

7 Conclusion

Prima facie, one would expect that a policy change credited with fundamentally transforming Chinese agriculture would be visible with satellites—particularly when other, more minute changes (like weather shocks) are also visible. We find consistent evidence, across two different identification strategies, that the causal effect of the Household Responsibility System on grain yields is statistically indistinguishable from 0. The standard errors on these estimates are precise; we are able to reject even relatively minor changes in agricultural yield at the 95% confidence level.

We should reiterate what we can precisely say with satellite data. This paper has focused entirely on grain yields, which are plausibly measurable from space. Labor productivity is not. This paper's findings are entirely consistent with the HRS increasing labor productivity, freeing up time for non-agricultural work—indeed, we find evidence suggesting that the HRS induced a small amount of growth in built-up land area, supportive of this hypothesis. Nonetheless, in people-rich, land-poor China, where rice and wheat were overwhelmingly the main caloric sources, grain yields were a central preoccupation of leadership—the procurement targets that starved millions to death during the Great Leap Forward were set in terms of unrealistic grain yields (Liu and Zhou 2022). Finding no evidence of an effect of the HRS on grain yields is in tension with the academic and popular consensus, which holds that the HRS was *the* major driver of agricultural productivity growth from the late 1970s to early 1980s.

Prior treatments of China's rural reform period, from Lin (1992) to Swinnen and Rozelle (2006), have all considered the procurement price changes a significant contributor to rising agricultural yields, but based on prior panel regression evidence, they have been de-emphasized in favor of the HRS as the central cause. This paper's (lack of) causal evidence may shift the balance between these two explanations. At the microeconomic level, models of collective agriculture tend to emphasize free rider problems and monitoring costs, which overwhelm any of the productivity benefits of larger scale (Lin 1988). However, even under collective agriculture, there are marginal incentives to produce more, particularly if these prices are brought closer in line with free-market levels. This paper's results touch on fundamental questions about what drives the efficiency differences between capitalist and socialist systems—is it who owns the claims on output, or is it getting prices right?

These results also interact with our understanding of the history of China's reform period at the highest levels of power. Hua Guofeng, Mao's direct successor, has long been seen as a dogmatic Maoist opposed to reform, a view best summarized by his "Two Whatevers" slogan—that he would uphold whatever decisions Mao made, and whatever instructions he gave. In this light, the Reform and Opening period has commonly been seen as the result of a victory of a reform coalition led by Deng Xiaoping over a conservative faction led by Hua.

A recent wave of revisionist history, with new access to party archives and close contacts with former Party leaders, has begun to revise these long-held views of Hua Guofeng, and the common narrative of Reform and Opening (Teiwes and Sun 2016; Eisenman 2018; Torigian 2022). Close study of party conference materials by Teiwes and Sun (2016) reveals that there was little policy daylight between Hua and Deng—if anything, Hua was more involved in rural policy reforms at an early stage. It was under Hua that the experiment with Special Economic Zones began (Vogel (2013), p. 185). It was Hua who issued the first, cautious statement of support of household responsibility by a top leader, when in 1979 he authorized in the special cases of "isolated households in mountainous and remote areas" (Teiwes and Sun (2016), p. 70). And it was under Hua that procurement prices were raised in 1979.

This last point may partly explain the emphasis on the now-famous Household Responsibility System, which is closely associated with Deng, relative to the less-known procurement price reform. As early as 1983, Deng's quotes concerning household responsibility were already being rewritten for his *Selected Works* to edit out his caution and make his support seem stronger than it was at the time (Teiwes and Sun (2016), p. 157). By contrast, during Hua's self-criticism in fall 1980, as he was gradually being stripped of authority by Deng and his allies, he made a special point of claiming "sole responsibility" for the price reform of 1979 (Teiwes and Sun (2016), p. 65). Our conventional narratives of the Household Responsibility System may reflect history being written by the winners.

On a broader note, satellite imagery from sources like AVHRR present an enormous opportunity for historical research. This paper demonstrates how remote sensing data can be applied, with relatively low computational cost, to produce agricultural output data almost anywhere in the world, from 1978 on. While substantial measurement error concerns remain, we document here that with careful calibration these data can be used to detect the effects of weather shocks and cross-country differences in yields over small spatial scales. The relatively high level of disaggregation provided by AVHRR also allows for the application of careful, modern causal identification to historical questions. This paper's new causal evidence, challenging long-held views of one of the major reforms of the Chinese miracle, suggests that other surprising possibilities may exist in historical satellite imagery, waiting to be discovered.

References

Adamopoulos, Tasso, and Diego Restuccia. 2020. "Land Reform and Productivity: A Quantitative Analysis with Micro Data." *American Economic Journal: Macroeconomics* 12, no. 3 (July): 1–39. ISSN: 1945-7707, 1945-7715, accessed May 8, 2021. https://doi.org/10.1257/mac.20150222.

Almond, Douglas, Hongbin Li, and Shuang Zhang. 2019. "Land Reform and Sex Selection in China." *Journal of Political Economy* 127, no. 2 (April): 560–585. ISSN: 0022-3808, accessed April 26, 2022. https://doi.org/10.1086/701030.

Asher, Sam, and Paul Novosad. 2020. "Rural Roads and Local Economic Development." *American Economic Review* 110, no. 3 (March): 797–823. ISSN: 0002-8282, accessed November 18, 2023. https://doi.org/10.1257/aer.20180268.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2023. *Revisiting Event Study Designs: Robust and Efficient Estimation*, arXiv:2108.12419, September. Accessed November 12, 2023. arXiv: 2108.12419 [econ].

Bramall, Chris. 2004. "Chinese Land Reform in Long-Run Perspective and in the Wider East Asian Context." *Journal of Agrarian Change* 4 (1-2): 107–141. ISSN: 1471-0366, accessed October 2, 2023. https://doi.org/10.1111/j.1471-0366.2004.00074.x.

Brandt, Loren, Chant-tai Hsieh, and Xiaodong Zhu. 2008. "Growth and Structural Transformation in China." In *China's Great Economic Transformation*, 1st ed., edited by Loren Brandt and Thomas G. Rawski, 683–728. Cambridge University Press, April. ISBN: 978-0-521-88557-7 978-0-521-71290-3 978-0-511-75423-4, accessed June 20, 2023. https://doi.org/10.1017/CBO9780511754234.018.

Brandt, Loren, Matthew A. Turner, and Scott Rozelle. 2004. "Local Government Behavior and Property Right Formation in Rural China." *Journal of Institutional and Theoretical Economics* 160 (4): 627. ISSN: 0932-4569, accessed November 1, 2023. https://doi.org/10.1628/093245604276032.

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics*, Themed Issue: Treatment Effect 1, 225, no. 2 (December): 200–230. ISSN: 0304-4076, accessed June 15, 2023. https://doi.org/10.1016/j.jeconom.2020.12.001.

Chung, Jae-ho. 2000. *Central Control and Local Discretion in China: Leadership and Implementation during Post-Mao Decollectivization*. Studies on Contemporary China. Oxford ; New York: Oxford University Press. ISBN: 978-0-19-829777-2.

Cisse, Abdoulaye, Alain de Janvry, Joel Ferguson, Marco Gonzalez-Navarro, Samba Mbaye, Elisabeth Sadoulet, and Mame Mor Anta Syll. 2026. "Irrigation Infrastructure and Satellite-Measured Land Cultivation Impacts: Evidence from the Senegal River Valley." *Journal of Development Economics* 179 (February): 103615. ISSN: 0304-3878, accessed March 7, 2026. https://doi.org/10.1016/j.jdeveco.2025.103615.

De Chaisemartin, Clément, and Xavier dHaultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American economic review* 110 (9): 2964–2996.

Dube, Arindrajit, Daniele Girardi, Òscar Jordà, and Alan M. Taylor. 2023. "A Local Projections Approach to Difference-in-Differences Event Studies." *NBER Working Paper No. 31184*, accessed May 21, 2023.

Eisenman, Joshua. 2018. *Red China's Green Revolution: Technological Innovation, Institutional Change, and Economic Development under the Commune.* New York: Columbia University press. ISBN: 978-0-231-18666-7 978-0-231-18667-4.

Foster, Andrew, and Mark Rosenzweig. 2017. *Are There Too Many Farms in the World? Labor-Market Transaction Costs, Machine Capacities and Optimal Farm Size.* Technical report w23909. Cambridge, MA: National Bureau of Economic Research, October. Accessed May 5, 2021. https://doi.org/10.3386/w23909.

Gelman, Andrew, and Guido Imbens. 2019. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics* 37, no. 3 (July): 447–456. ISSN: 0735-0015, 1537-2707, accessed June 11, 2023. https://doi.org/10.1080/07350015.2017.1366909.

Gibson, John. 2020. "Aggregate and Distributional Impacts of China's Household Responsibility System." *Australian Journal of Agricultural and Resource Economics* 64 (1): 14–29. ISSN: 1467-8489, accessed August 7, 2023. https://doi.org/10.1111/1467-8489.12329.

Henderson, J. Vernon, Adam Storeygard, and David N Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102, no. 2 (April): 994–1028. ISSN: 0002-8282, accessed May 9, 2021. https://doi.org/10.1257/aer.102.2.994.

Hodler, Roland, and Paul A. Raschky. 2014. "Regional Favoritism." *The Quarterly Journal of Economics* 129 (2): 995–1033. ISSN: 0033-5533, accessed June 17, 2023. JSTOR: 26372565.

Holz, Carsten A. 2003. ""Fast, Clear and Accurate": How Reliable Are Chinese Output and Economic Growth Statistics?" *The China Quarterly* 173 (March): 122–163. ISSN: 0305-7410, 1468-2648, accessed February 23, 2022. https://doi.org/10.1017/S0009443903000081.

Huang, Jingfeng, Xiuzhen Wang, Xinxing Li, Hanqin Tian, and Zhuokun Pan. 2013. "Remotely Sensed Rice Yield Prediction Using Multi-Temporal NDVI Data Derived from NOAA's-AVHRR." Edited by Wengui Yan. *PLoS ONE* 8, no. 8 (August): e70816. ISSN: 1932-6203, accessed November 18, 2023. https://doi.org/10.1371/journal.pone.0070816.

Huang, Luna Yue, Solomon M. Hsiang, and Marco Gonzalez-Navarro. 2021. *Using Satellite Imagery and Deep Learning to Evaluate the Impact of Anti-Poverty Programs.* Working Paper, Working Paper Series 29105. National Bureau of Economic Research, July. Accessed April 26, 2022. https://doi.org/10.3386/w29105.

Huang, Yasheng. 2008. *Capitalism with Chinese Characteristics: Entrepreneurship and the State.* Cambridge New York: Cambridge University Press. ISBN: 978-0-511-75421-0 978-0-511-43004-6 978-0-511-42966-8.

Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty." *Science* 353, no. 6301 (August): 790–794. ISSN: 0036-8075, 1095-9203, accessed August 5, 2021. https://doi.org/10.1126/science.aaf7894.

Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review* 95, no. 1 (March): 161–182. ISSN: 0002-8282, accessed March 23, 2026. https://doi.org/10.1257/0002828053828518.

Kagin, Justin, J. Edward Taylor, and Antonio Yúnez-Naude. 2016. "Inverse Productivity or Inverse Efficiency? Evidence from Mexico." *The Journal of Development Studies* 52, no. 3 (March): 396–411. ISSN: 0022-0388, accessed August 19, 2023. https://doi.org/10.1080/00220388.2015.1041515.

Kelliher, Daniel Roy. 1992. *Peasant Power in China: The Era of Rural Reform, 1979-1989.* Yale Agrarian Studies. New Haven: Yale University Press. ISBN: 978-0-300-05465-1.

Kim, Oliver, and Jen-Kuan Wang. 2024. *Land Reform in Taiwan, 1950-1961: Effects on Agriculture and Structural Change.* Accessed March 18, 2026. https://doi.org/10.2139/ssrn.4951831.

Kitamura, Shuhei. 2022. *Tillers of Prosperity: Land Ownership, Reallocation, and Structural Transformation,* May. Accessed March 18, 2026. https://doi.org/10.31219/osf.io/wh5qx.

Lin, Justin Yifu. 1988. "The Household Responsibility System in China's Agricultural Reform: A Theoretical and Empirical Study." *Economic Development and Cultural Change* 36, no. S3 (April): S199–S224. ISSN: 0013-0079, accessed June 21, 2023. https://doi.org/10.1086/edcc.36.s3.1566543.

———. 1992. "Rural Reforms and Agricultural Growth in China." *The American Economic Review* 82 (1): 34–51. ISSN: 0002-8282, accessed April 18, 2022. JSTOR: 2117601.

Liu, Chang, and Li-An Zhou. 2022. "Radical Target Setting and China's Great Famine." *The Journal of Law, Economics, and Organization* 38, no. 1 (March): 120–160. ISSN: 8756-6222, accessed November 22, 2023. https://doi.org/10.1093/jleo/ewab025.

Lobell, David B, George Azzari, Marshall Burke, Sydney Gourlay, Zhenong Jin, Talip Kilic, and Siobhan Murray. 2019. "Eyes in the sky, boots on the ground: Assessing satellite-and ground-based approaches to crop yield measurement and analysis." *American Journal of Agricultural Economics,* aaz051.

Lobell, David B., David Thau, Christopher Seifert, Eric Engle, and Bertis Little. 2015. "A Scalable Satellite-Based Crop Yield Mapper." *Remote Sensing of Environment* 164 (July): 324–333. ISSN: 0034-4257, accessed August 21, 2023. https://doi.org/10.1016/j.rse.2015.04.021.

Looney, Kristen. 2008. *'Village Gazetteers, a New Source in the China Field.'* SSRN Scholarly Paper 3027667. Rochester, NY: Social Science Research Network, July. Accessed April 26, 2022.

Martínez, Luis R. 2022. "How Much Should We Trust the Dictator's GDP Growth Estimates?" *Journal of Political Economy* 130, no. 10 (October): 2731–2769. ISSN: 0022-3808, accessed June 17, 2023. https://doi.org/10.1086/720458.

McMillan, John, John Whalley, and Lijing Zhu. 1989. "The Impact of China's Economic Reforms on Agricultural Productivity Growth." *Journal of Political Economy* 97 (4): 781–807. ISSN: 0022-3808, accessed June 8, 2023. JSTOR: 1832191.

Nakamura, Emi, Jón Steinsson, and Miao Liu. 2016. "Are Chinese Growth and Inflation Too Smooth? Evidence from Engel Curves." *American Economic Journal: Macroeconomics* 8, no. 3 (July): 113–144. ISSN: 1945-7707, 1945-7715, accessed February 23, 2022. https://doi.org/10.1257/mac.20150074.

Naughton, Barry. 2007. *The Chinese Economy: Transitions and Growth.* Cambridge, Mass: MIT Press. ISBN: 978-0-262-64064-0 978-0-262-14095-9.

Oi, Jean Chun. 2002. *Rural China Takes off: Institutional Foundations of Economic Reform.* Repr. Berkeley: Univ. of California Press. ISBN: 978-0-520-21727-0.

Rabinovitch, Simon. 2010. "China's GDP Is "Man-Made," Unreliable: Top Leader." *Reuters* (December). Accessed September 11, 2023.

Ravallion, Martin, and Shaohua Chen. 2007. "China's (Uneven) Progress against Poverty." *Journal of Development Economics* 82, no. 1 (January): 1–42. ISSN: 03043878, accessed July 7, 2023. https://doi.org/10.1016/j.jdeveco.2005.07.003.

Rawski, Thomas G. 1976. "On the Reliability of Chinese Economic Data." *The Journal of Development Studies* 12, no. 4 (July): 438–441. ISSN: 0022-0388, 1743-9140, accessed April 17, 2022. https://doi.org/10.1080/00220387608421599.

Sicular, Terry. 1988a. "Agricultural Planning and Pricing in the Post-Mao Period." *The China Quarterly*, no. 116, 671–705. ISSN: 0305-7410, accessed June 11, 2023. JSTOR: 654756.

———. 1988b. "Grain Pricing: A Key Link in Chinese Economic Policy." *Modern China* 14, no. 4 (October): 451–486. ISSN: 0097-7004, 1552-6836, accessed October 3, 2023. https://doi.org/10.1177/009770048801400404.

Skinner, G. William. 1985. "Rural Marketing in China: Repression and Revival." *The China Quarterly*, no. 103, 393–413. ISSN: 0305-7410, accessed May 3, 2023. JSTOR: 653964.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*, Themed Issue: Treatment Effect 1, 225, no. 2 (December): 175–199. ISSN: 0304-4076, accessed November 12, 2023. https://doi.org/10.1016/j.jeconom.2020.09.006.

Swinnen, Johan F. M., and Scott Rozelle. 2006. *From Marx and Mao to the Market: The Economics and Politics of Agricultural Transition.* Oxford New York: Oxford University Press. ISBN: 978-0-19-928891-5.

Taiz, Lincoln, I. M. Møller, Angus S. Murphy, and Eduardo Zeiger, eds. 2022. *Plant Physiology and Development*. Seventh edition. New York, NY: Sinauer Associates : Oxford University Press. ISBN: 978-0-19-757724-0 978-0-19-761423-5.

Teiwes, Frederick C., and Warren Sun. 2016. *Paradoxes of Post-Mao Rural Reform: Initial Steps toward a New Chinese Countryside, 1976-1981.* London ; New York, NY: Routledge, Taylor & Francis Group. ISBN: 978-1-138-85658-5.

Torigian, Joseph. 2022. *Prestige, Manipulation, and Coercion: Elite Power Struggles in the Soviet Union and China after Stalin and Mao.* New Haven London: Yale University Press. ISBN: 978-0-300-25423-5.

Tucker, Compton J. 1979. "Red and Photographic Infrared Linear Combinations for Monitoring Vegetation." *Remote Sensing of Environment* 8, no. 2 (May): 127–150. ISSN: 0034-4257, accessed September 9, 2023. https://doi.org/10.1016/0034-4257(79)90013-0.

Vogel, Ezra F. 2013. *Deng Xiaoping and the Transformation of China.* First Harvard University Press paperback edition. Cambridge, Massachusetts London, England: Belknap Press of Harvard University Press. ISBN: 978-0-674-72586-7.

Vollrath, Dietrich. 2007. "Land Distribution and International Agricultural Productivity." *American Journal of Agricultural Economics* 89 (1): 202–216. ISSN: 1467-8276, accessed August 19, 2023. https://doi.org/10.1111/j.1467-8276.2007.00973.x.

Wang, Lixin. 1989. *Mao's Legacy in Anhui: Rural Reform 1978-80.* Technical report JPRS-CAR-89-079. Springfield, VA: U.S. Department of Commerce, National Technical Information Service, July.

Weber, Isabella. 2021. *How China Escaped Shock Therapy: The Market Reform Debate.* Routledge Studies on the Chinese Economy. Abingdon, Oxon ; New York, N.Y: Routledge. ISBN: 978-1-138-59219-3 978-1-032-00849-3.

World Bank. 2022. *Four Decades of Poverty Reduction in China: Drivers, Insights for the World, and the Way Ahead.* The World Bank, August. ISBN: 978-1-4648-1877-6 978-1-4648-1878-3, accessed October 14, 2022. https://doi.org/10.1596/978-1-4648-1877-6.

Wuepper, David, Wyclife Agumba Oluoch, and Hadi Hadi. 2025. "Satellite Data in Agricultural and Environmental Economics: Theory and Practice." *Agricultural Economics* 56 (3): 493–511. ISSN: 1574-0862, accessed March 6, 2026. https://doi.org/10.1111/agec.70006.

Wuepper, David, Haoyu Wang, Wolfram Schlenker, Meha Jain, and Robert Finger. 2023. *Institutions and global crop yields.* Technical report. National Bureau of Economic Research.

Yang, Hong. 1999. "Growth in China's Grain Production 1978–1997: A Disaggregate Analysis." *World Development* 27, no. 12 (December): 2137–2154. ISSN: 0305750X, accessed November 9, 2025. https://doi.org/10.1016/S0305-750X(99)00107-2.

Yang, Jisheng. 2013. *Tombstone: The Great Chinese Famine, 1958 - 1962.* 1., American paperback ed. New York: Farrar, Strauss and Giroux. ISBN: 978-0-374-53399-1.

Yeh, Christopher, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. 2020. "Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-Being in Africa." *Nature Communications* 11, no. 1 (December): 2583. ISSN: 2041-1723, accessed August 5, 2021. https://doi.org/10.1038/s41467-020-16185-w.

You, Jiaxuan, Xiaocheng Li, Melvin Low, David Lobell, and Stefano Ermon. 2017. "Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data." *Proceedings of the AAAI Conference on Artificial Intelligence* 31, no. 1 (February). ISSN: 2374-3468, 2159-5399, accessed November 21, 2023. https://doi.org/10.1609/aaai.v31i1.11172.

Young, Alwyn. 2003. "Gold into Base Metals: Productivity Growth in the People's Republic of China during the Reform Period." *Journal of Political Economy* 111, no. 6 (December): 1220–1261. ISSN: 0022-3808, 1537-534X, accessed June 20, 2023. https://doi.org/10.1086/378532.

Zhang, Qi. 2019. *Revolutionary Legacy, Power Structure, and Grassroots Capitalism under the Red Flag in China.* Cambridge: Cambridge University Press. ISBN: 978-1-108-47492-4 978-1-108-67824-7.

Appendix

A Satellite Validation **45**

B The Household Responsibility System **64**

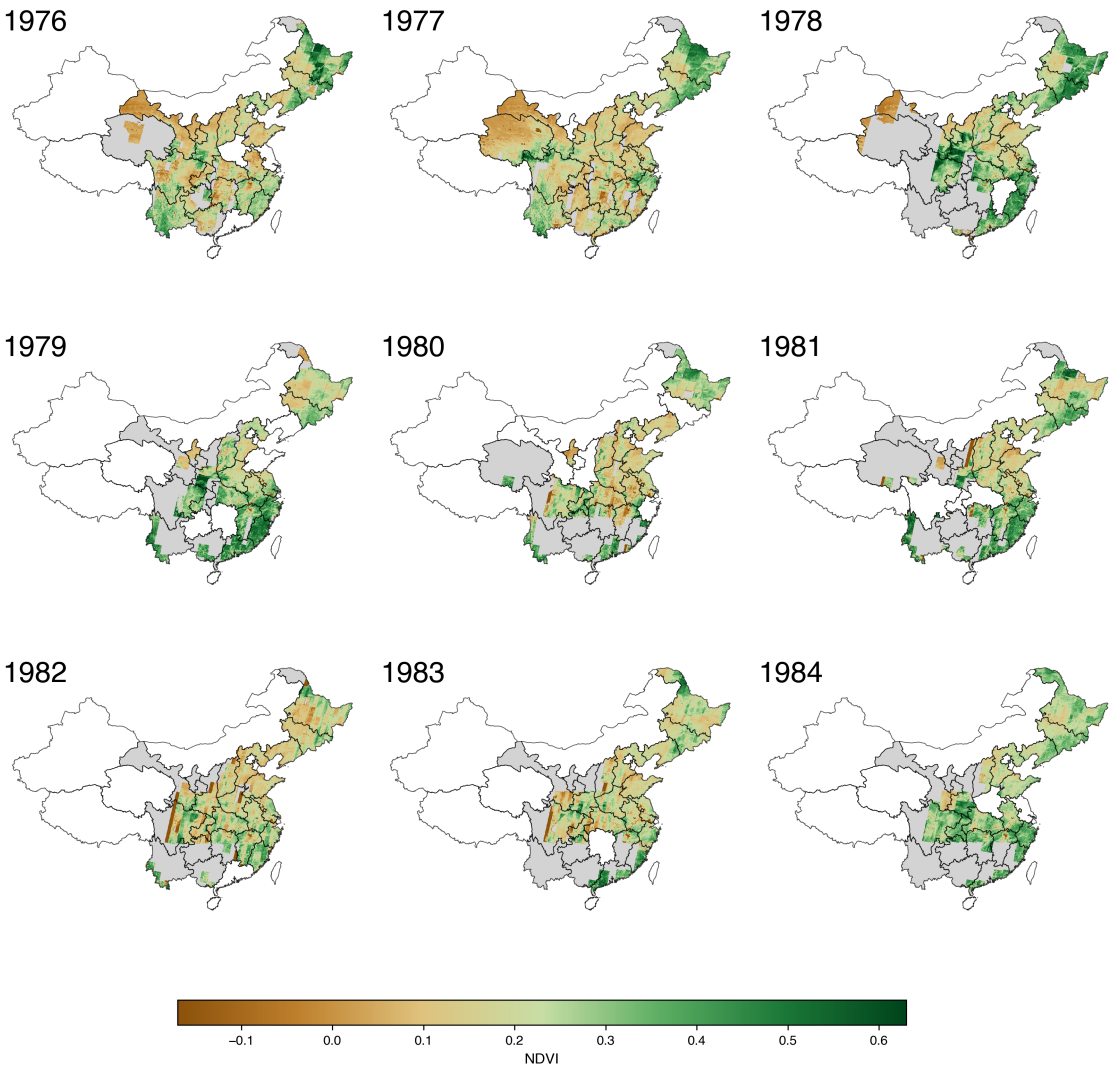
C Procurement Reform **81**

D Urban Prediction **86**

E Non-agricultural Growth **90**

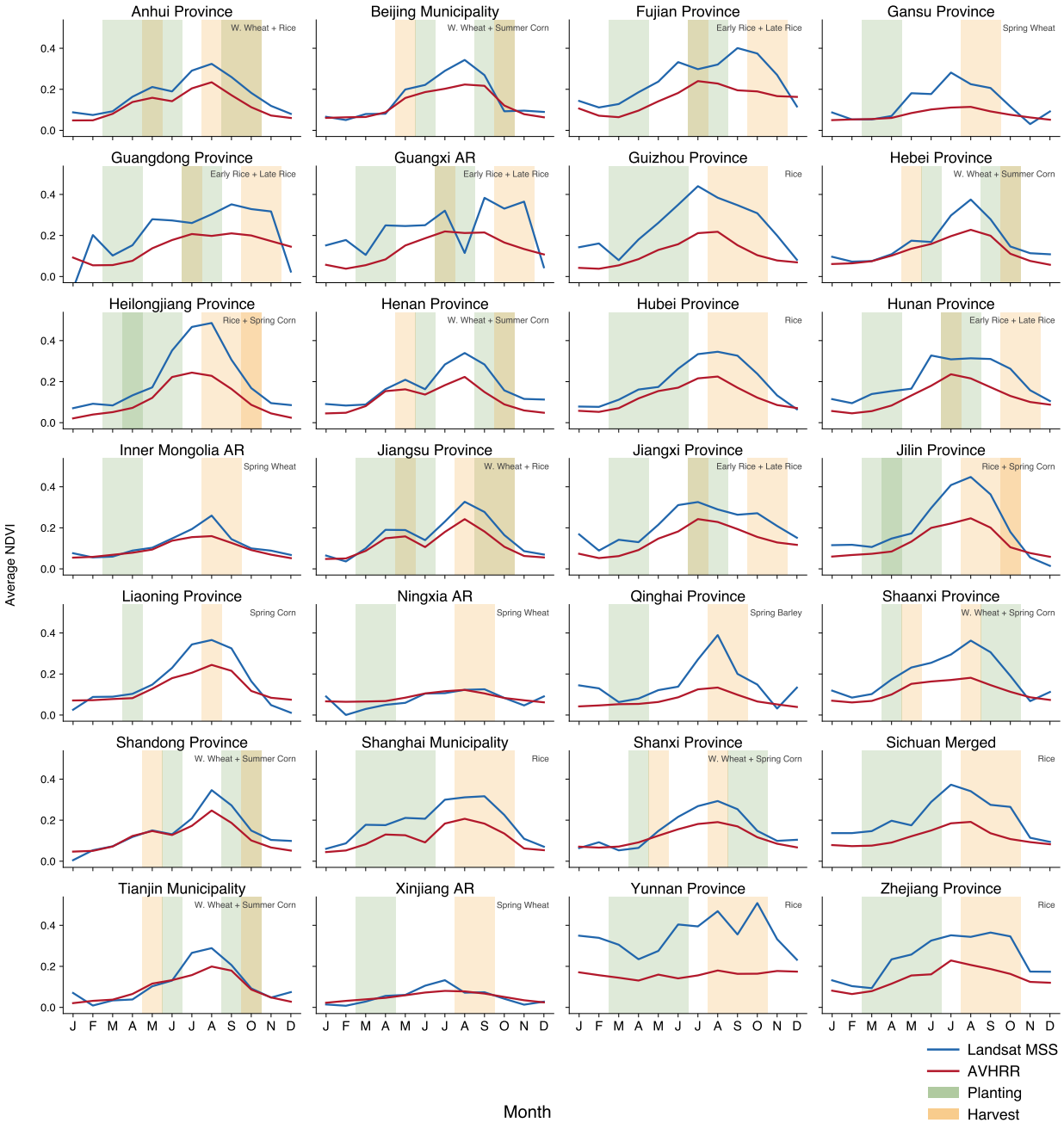
A Satellite Validation

Figure A.1: Landsat MSS NDVI Coverage Across China, 1976–1984



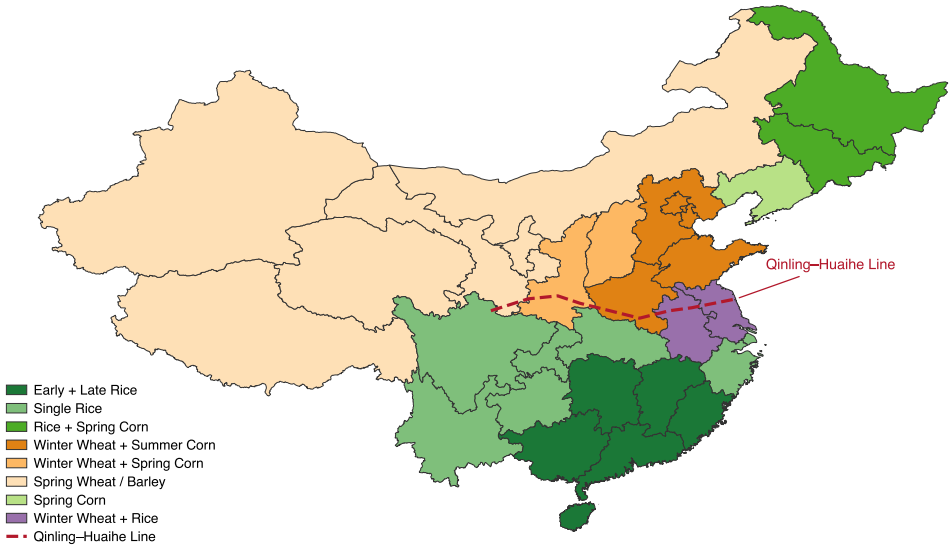
This figure shows annual Landsat MSS NDVI coverage over China from 1976 to 1984. All panels use the same green-brown color palette and a common NDVI scale with one shared legend.

Figure A.2: NDVI Seasonal Profiles by Province with Crop Calendar



This figure plots monthly average NDVI profiles by province for Landsat MSS (blue) and AVHRR (red). For each province, we first average county-level NDVI within each year-month, then average those year-month means across years to obtain one within-year seasonal profile. Shaded regions indicate planting (green) and harvest (orange) windows for each province's principal crop(s) based on USDA crop calendar data. Province subtitles report the dominant cropping system. NDVI peaks align closely with the period between planting and harvest, confirming that satellite NDVI tracks the agricultural growing season.

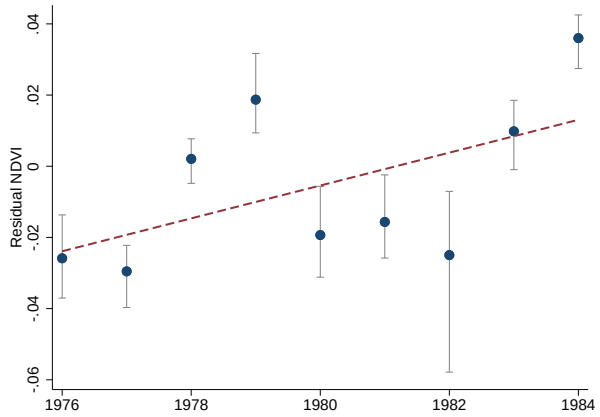
Figure A.3: Dominant Cropping Systems by Province



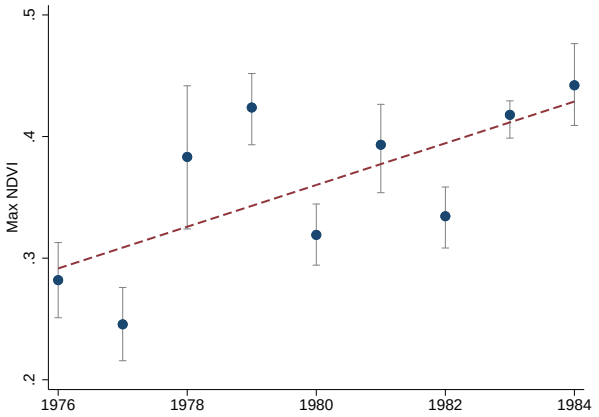
This map shows the dominant cropping system for each Chinese province based on USDA crop calendar data. Colors are grouped by crop family: greens for rice-based systems (darker for double-cropping), amber/gold tones for wheat-based systems (darker for winter wheat rotations), and purple for the wheat–rice transition zone. The dashed line marks the approximate Qinling–Huaihe divide, the traditional boundary between rice-dominant agriculture to the south and wheat-dominant agriculture to the north.

Figure A.4: China-wide Residual NDVI Time Series from Landsat MSS Cell Panel, 1976–1984

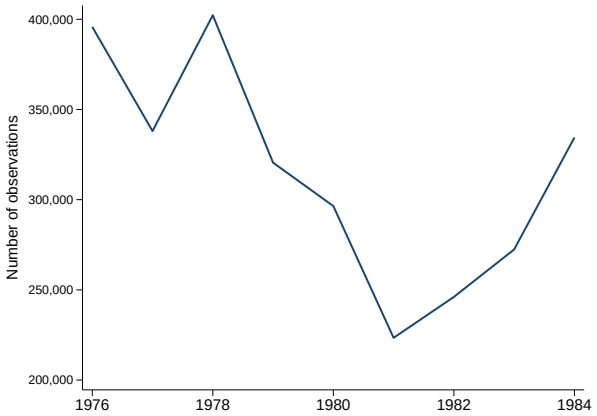
(a) Annual mean residual NDVI



(b) Annual maximum NDVI



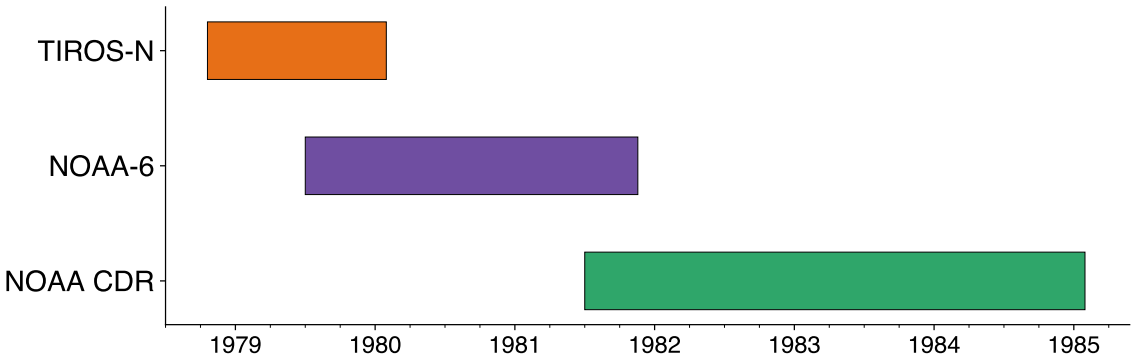
(c) Observation counts by year



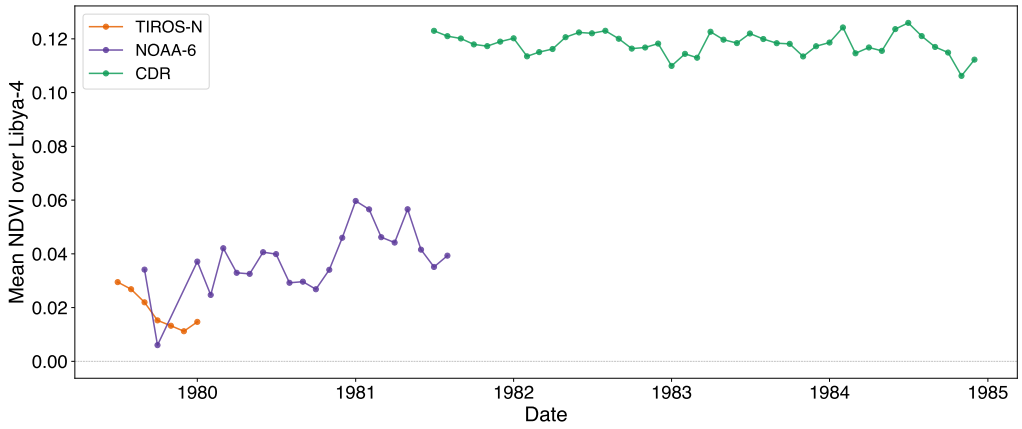
Panel (a) plots annual mean residual NDVI using the Landsat MSS cell-level panel. Residuals are obtained from a specification with cell fixed effects and month-of-year effects; points show annual means and the dashed line is the linear trend. Vertical bars show 95% confidence intervals from a province block bootstrap. Panel (b) reports the corresponding annual maximum NDVI time series. Panel (c) reports the number of cell-level observations used in each annual estimate.

Figure A.5: AVHRR Sensor Timeline and Calibration Checks

(a) AVHRR sensor timeline



(b) Libya-4 NDVI time series



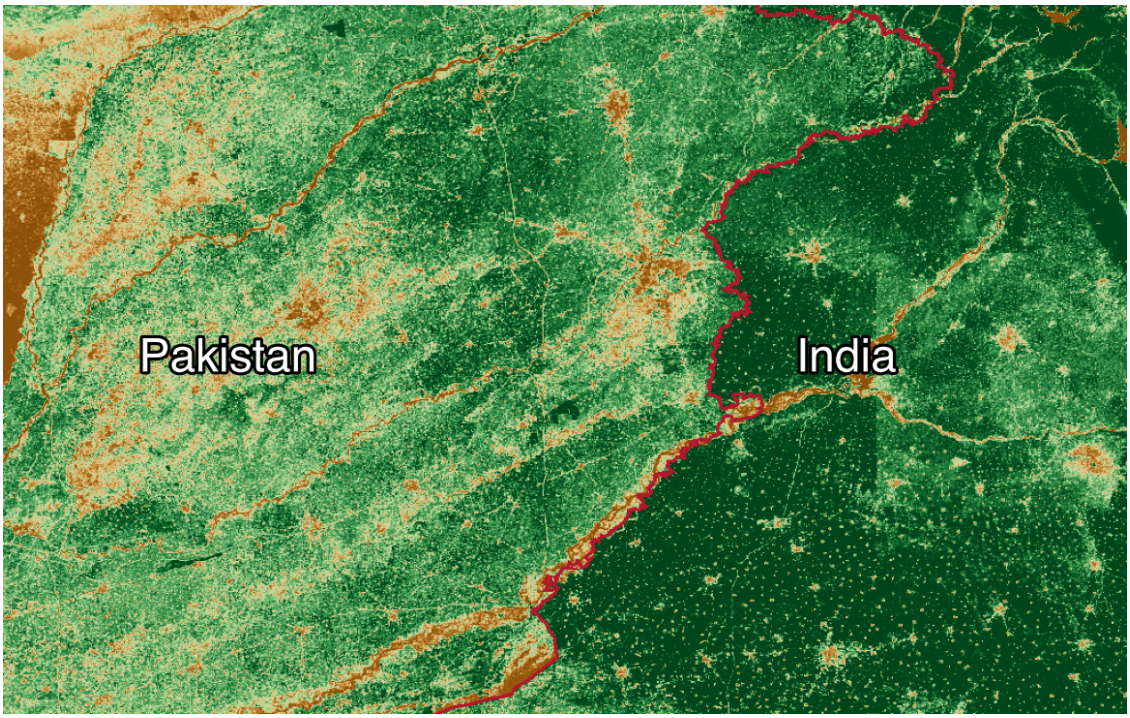
Panel (a) reports the sequence of AVHRR instruments and their periods of operation over the sample window, documenting sensor handoffs and motivating calibration and harmonization checks so that observed NDVI changes are less likely to reflect platform transitions rather than true vegetation dynamics. Panel (b) reports the AVHRR NDVI time series for the Libya-4 pseudo-invariant calibration site, used as a stability check for AVHRR reflectance over time.

Table A.1: India–Pakistan Punjab: Border Discontinuity and Within-Side Slopes

Sensor Era	Discontinuity $\hat{\beta}$	Pak. slope	India slope	N
TIROS-N (1979)	-0.0169*** (0.0041)	0.0005*** (0.0001)	-0.0009*** (0.0001)	371
NOAA-6 (198081)	-0.0217*** (0.0047)	0.0006*** (0.0001)	-0.0007*** (0.0001)	371
CDR (198289)	-0.0370*** (0.0044)	0.0004*** (0.0001)	-0.0004*** (0.0001)	371
Landsat MSS (197581)	-0.0298 (0.0205)	-0.0008 (0.0006)	-0.0019*** (0.0005)	279

This table reports the border discontinuity coefficient and within-side NDVI-distance slopes for the India–Pakistan Punjab border shown in Figure 3, estimated using Equation 2. The discontinuity column reports $\hat{\beta}$; the slope columns report the coefficient on distance (km) from separate within-side regressions. Robust (HC1) standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure A.6: Contemporary Reference Map: India–Pakistan Punjab Border



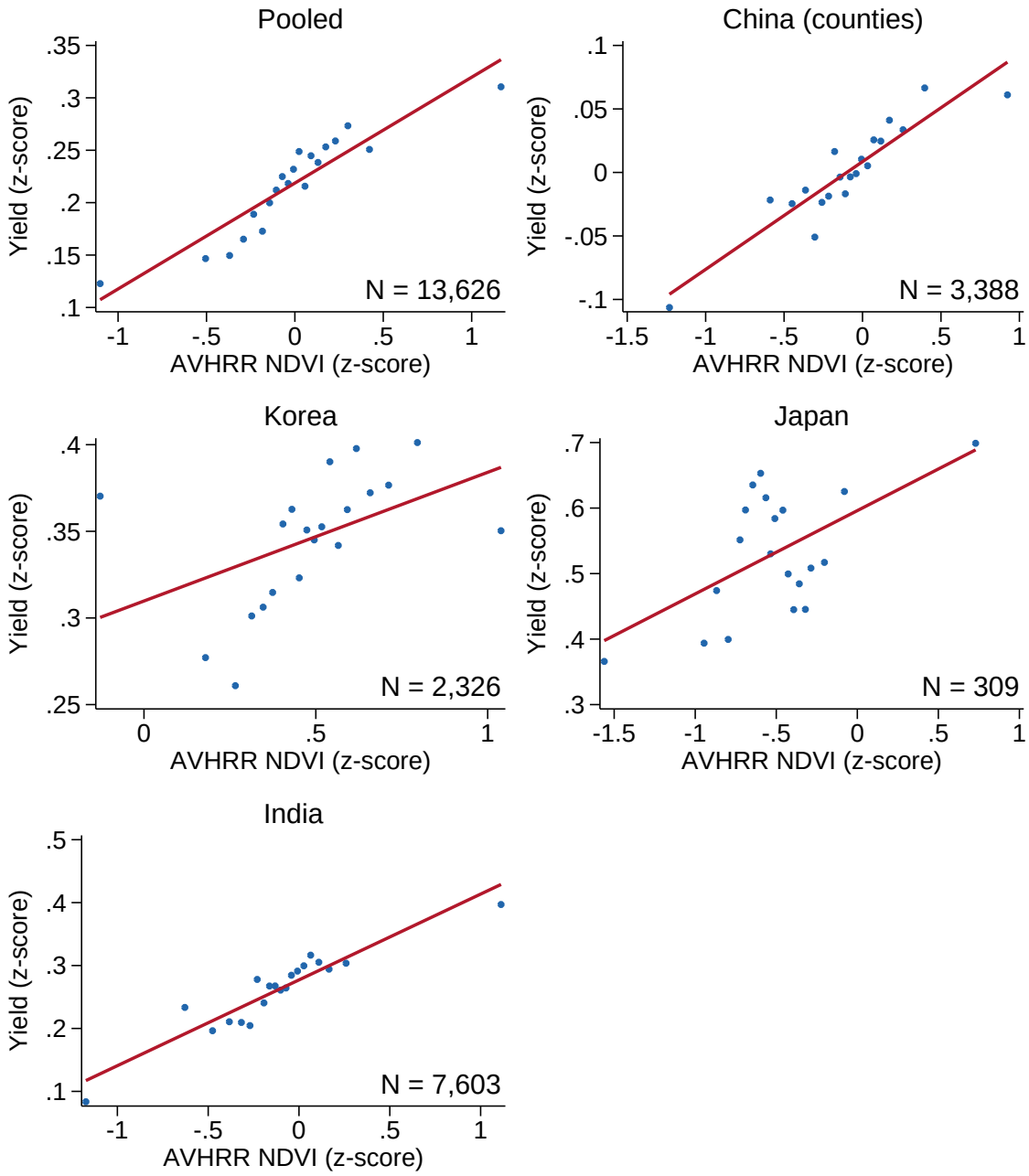
This figure provides the contemporary geographic reference map for the India–Pakistan Punjab validation border used in Figure 3. It documents the spatial context of the border segment used for the cross-border RD validation exercise.

Table A.2: Summary of NDVI-based yield prediction studies at administrative scales.

Study	Country	Scale	Crop	Sensor	R^2	Outcome
County / District level						
Doraiswamy et al. (2005)	USA	County	Corn	MODIS	0.84	Levels
Bolton and Friedl (2013)	USA	County	Corn	MODIS	0.78	Levels
Johnson (2014)	USA	CRD	Corn	MODIS	0.80	Levels
Sakamoto et al. (2014)	USA	County	Corn	MODIS	0.77	Levels
David B. Lobell et al. (2015)	USA	County	Corn	Landsat	0.70	Anomalies
You et al. (2017)	USA	County	Soybean	MODIS	0.72	Anomalies
Jain et al. (2016)	India	District	Wheat	SkySat	0.58	Levels
Burke and Lobell (2017)	Africa	District	Maize	MODIS	0.50	Anomalies
Funk and Budde (2009)	Zimbabwe	District	Maize	MODIS	0.70	Anomalies
Jin et al. (2019)	Africa	District	Maize	Sentinel-2	0.55	Levels
Prefecture / Oblast level						
Becker-Reshef et al. (2010)	Ukraine	Oblast	Wheat	MODIS	0.80	Levels
Schwalbert et al. (2020)	Brazil	Municipal.	Soybean	MODIS	0.73	Levels
Mkhabela et al. (2011)	Canada	CCS	Wheat	MODIS	0.72	Levels
Province / State / National level						
Huang et al. (2013)	China	Province	Rice	AVHRR	0.77	Levels
Lobell et al. (2007)	Global	National	Mixed	Multi	0.69	Levels

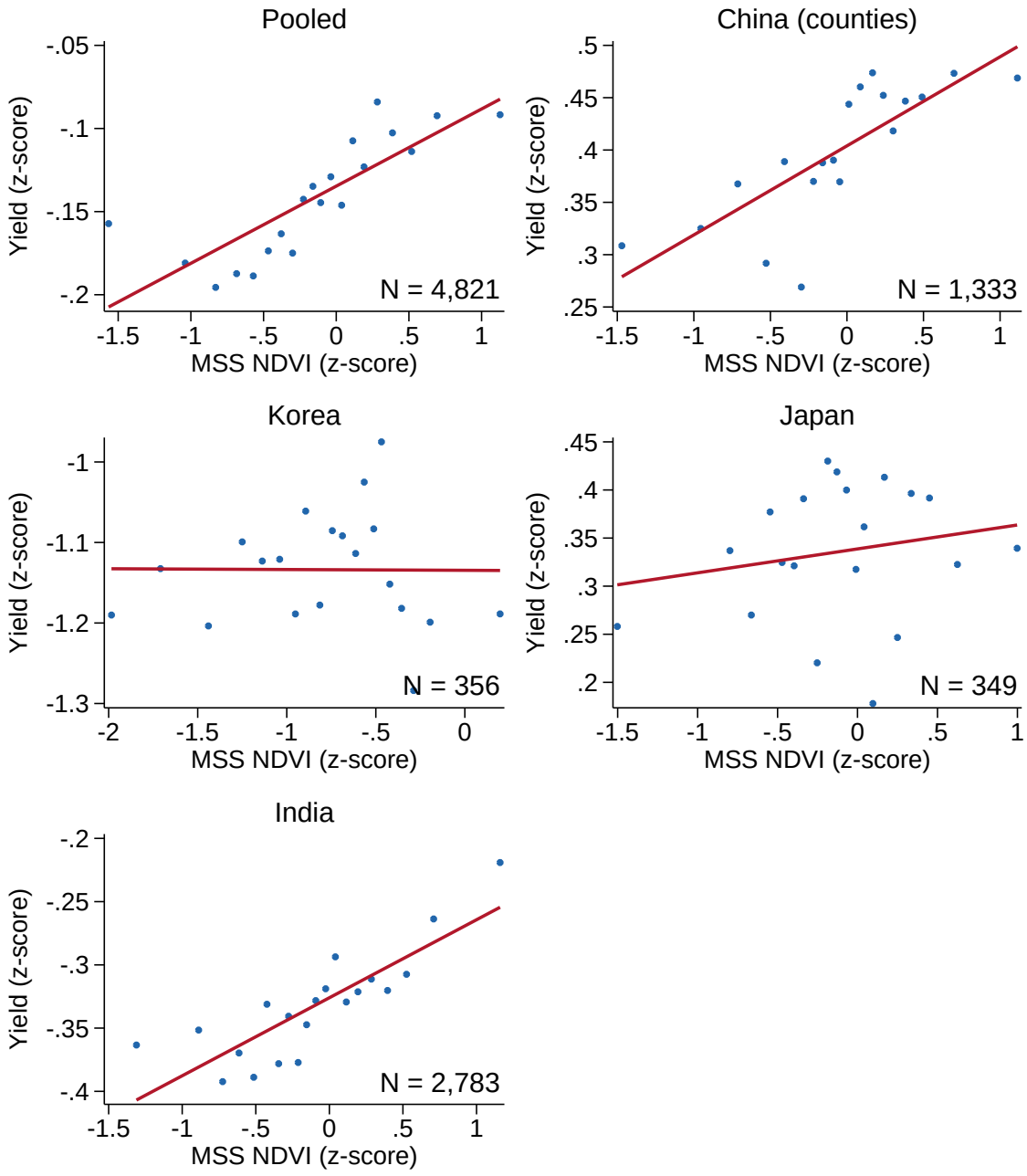
Notes: This table is adapted from `/Users/oliverwkim/Documents/Claude/NDVI_lit_review/ndvi_admin_scale.tex` and summarizes reported predictive fit (R^2) across administrative scales, crops, and sensors.

Figure A.7: AVHRR NDVI and Yield: Small-Multiple Binscatter Validation



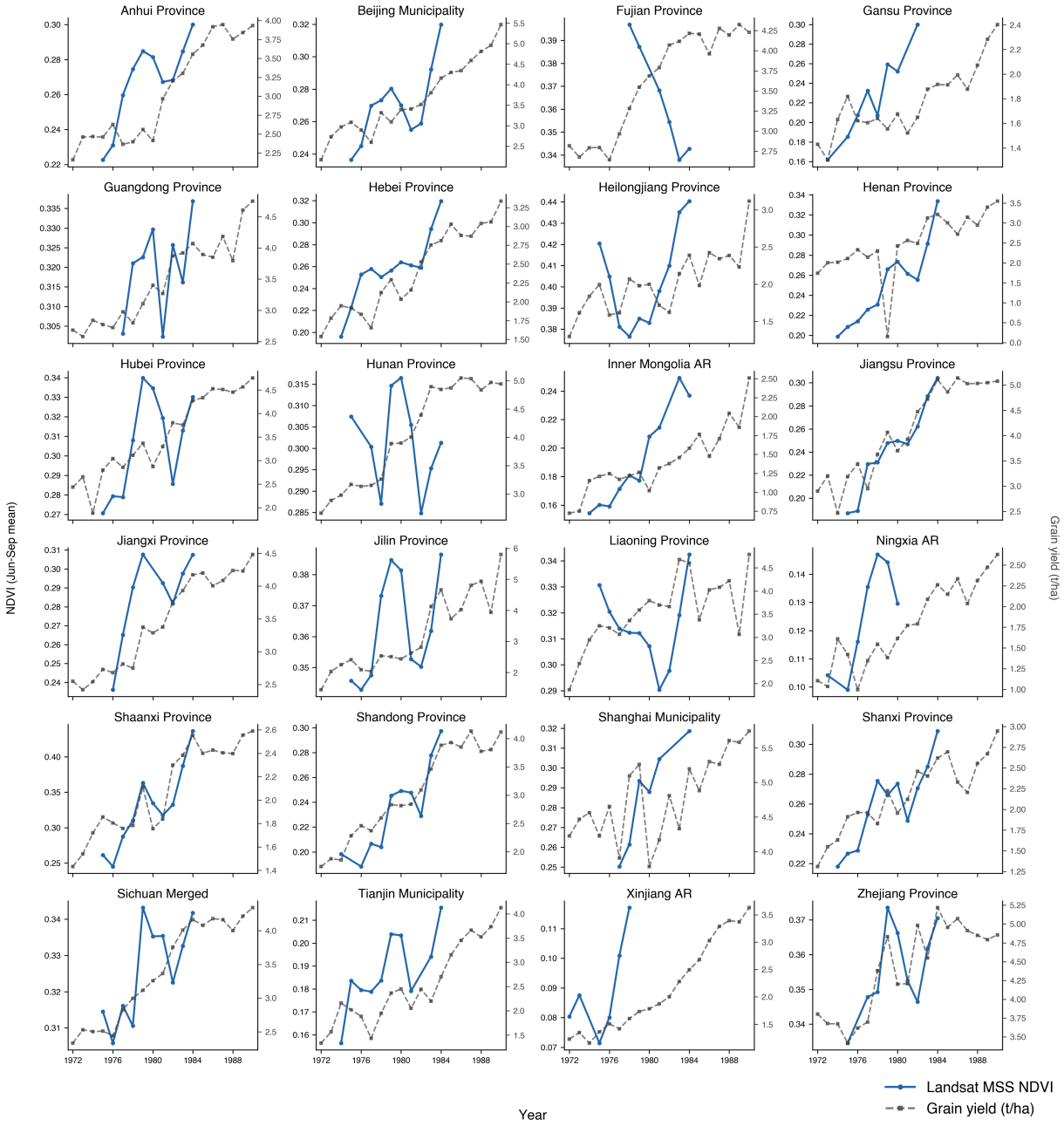
This figure reports binscatter diagnostics for the AVHRR NDVI–yield validation regressions. Each panel plots annual-maximum AVHRR NDVI against standardized administrative grain yield for the pooled sample and each country-specific validation sample. Within each panel, outcomes are residualized by region and year fixed effects and displayed as binned means with fitted linear trends.

Figure A.8: Landsat MSS NDVI and Yield: Small-Multiple Binscatter Validation



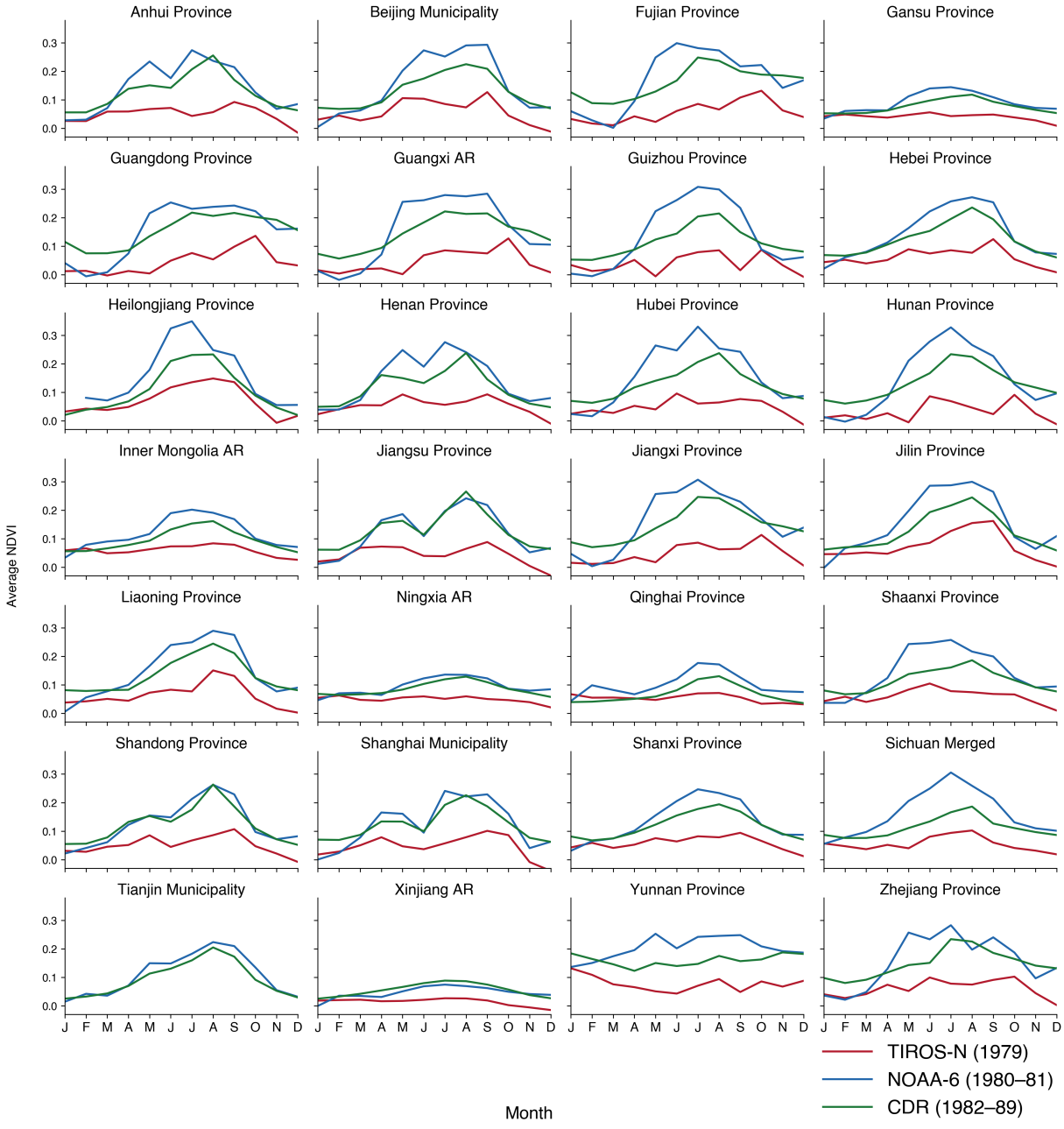
This figure reports binscatter diagnostics for the Landsat MSS NDVI–yield validation regressions. Each panel plots annual-maximum MSS NDVI against standardized administrative grain yield for the pooled sample and each country-specific validation sample. Within each panel, outcomes are residualized by region and year fixed effects and displayed as binned means with fitted linear trends.

Figure A.9: Province-Level Landsat MSS NDVI and Grain Yield Time Series



This figure overlays province-level annual Landsat MSS NDVI (blue, left axis) and official grain yield (gray dashed, right axis, t/ha) in small multiples. NDVI is constructed as the province-year mean of June–September MSS observations, excluding province-years with fewer than 32 growing-season observations, then smoothed with a centered three-year moving average; provinces with fewer than four annual NDVI points are omitted. Grain yield is from the Statistical Yearbook series. The panels provide a within-province visual co-movement check over 1972–1990.

Figure A.10: Province-Level AVHRR Monthly NDVI by Satellite Era



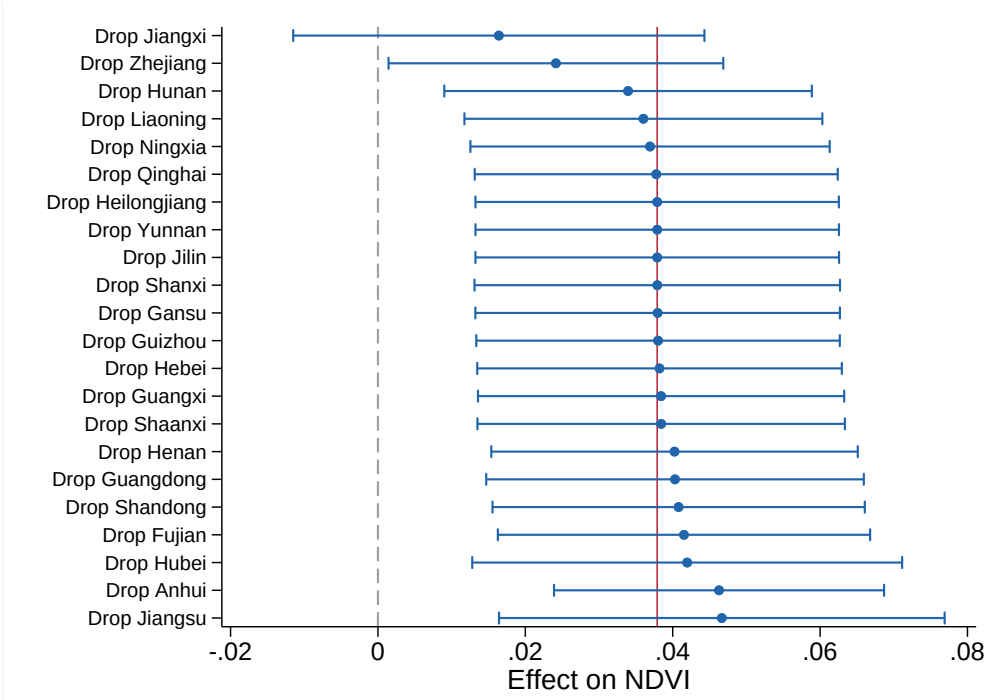
This figure reports province-level seasonal NDVI profiles from AVHRR, split by satellite era: TIROS-N (1979), NOAA-6 (1980–81), and AVHRR CDR (1982–89). For each province and month, NDVI is averaged across available cells and then across years within each era (using contemporaneous observations with year offset 0). All panels use a common y-axis scale to make cross-era level differences comparable, providing a visual sensor-consistency check in the monthly seasonal pattern.

Table A.3: Landsat MSS Observation Counts, Province by Year

	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	Total
Anhui Province	1871	25616	29605	38432	57932	84433	27393	0	41844	62097	84071	453294
Fujian Province	3257	2031	27617	6532	21626	19914	1280	26078	15199	11624	12775	147933
Gansu Province	0	0	0	0	0	5816	0	2038	0	0	4069	11923
Guangdong Province	0	3554	17451	28255	0	22138	5659	10856	5913	9911	6974	110711
Guangxi Zhuang Autonomous Region	11	5432	9455	10407	2006	5410	1553	2428	2129	2417	1896	43144
Guizhou Province	1220	3281	7351	9498	294	778	2432	151	3203	3173	4030	35411
Hebei Province	13026	0	0	0	0	0	0	0	0	0	0	13026
Henan Province	3567	28253	36705	38596	0	0	22423	0	34843	67138	31142	262667
Hubei Province	1660	28320	20056	27126	0	44353	0	33899	0	0	0	155414
Hunan Province	0	13656	15564	18001	26500	0	10731	13232	15175	17578	16695	147132
Jiangsu Province	1567	14281	38761	33169	44130	60563	27843	52556	33396	48305	69298	423869
Jiangxi Province	35	12086	22693	25405	27673	42522	12922	35469	27723	18623	22051	247202
Jilin Province	3949	0	0	0	0	0	0	0	0	0	0	3949
Liaoning Province	6446	0	59321	0	84649	0	68949	0	0	0	0	219365
Ningxia Hui Autonomous Region	848	14383	25084	26769	22603	6990	10185	106	199	89	58	107314
Qinghai Province	0	0	0	0	3193	0	1049	2911	0	0	0	7153
Shaanxi Province	4314	0	0	0	0	8325	24503	0	0	0	24870	62012
Shandong Province	7279	19084	46290	56321	81984	0	38496	0	38815	0	0	288269
Shanxi Province	3210	0	0	0	0	0	27388	0	0	0	0	30598
Yunnan Province	0	0	0	0	513	0	0	0	0	79	445	1037
Zhejiang Province	2397	6532	39978	19641	29362	19308	13670	43705	27663	31373	56074	289703
Total	54657	176509	395931	338152	402465	320550	296476	223429	246102	272407	334448	3061126

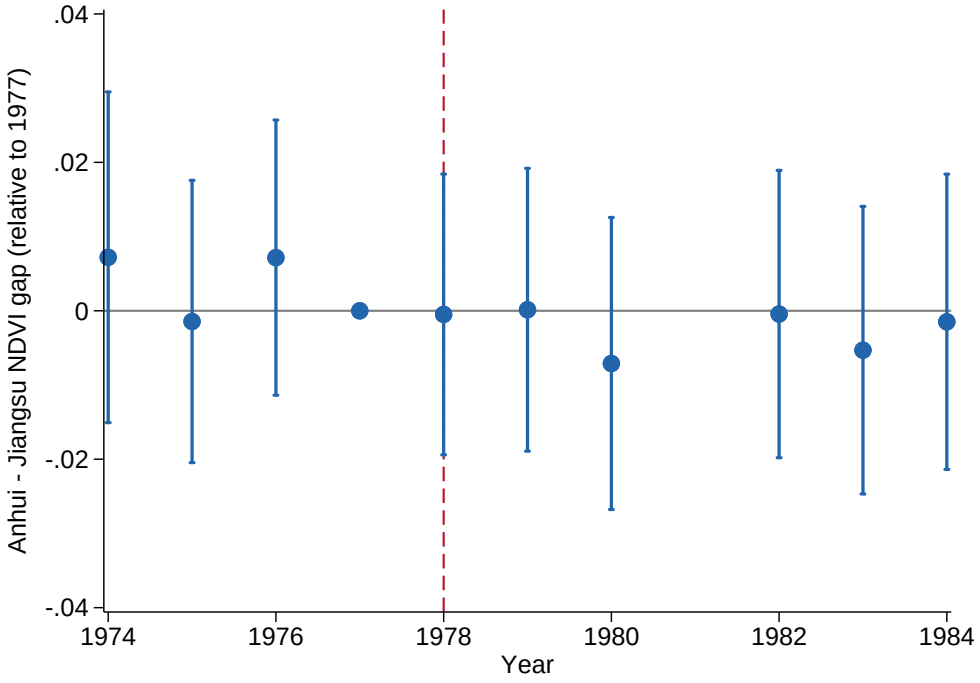
Notes: This table reports province-by-year counts of Landsat MSS observations used in the Landsat regression sample.

Figure A.11: Leave-One-Province-Out RD Estimates of HRS Effects on Landsat NDVI



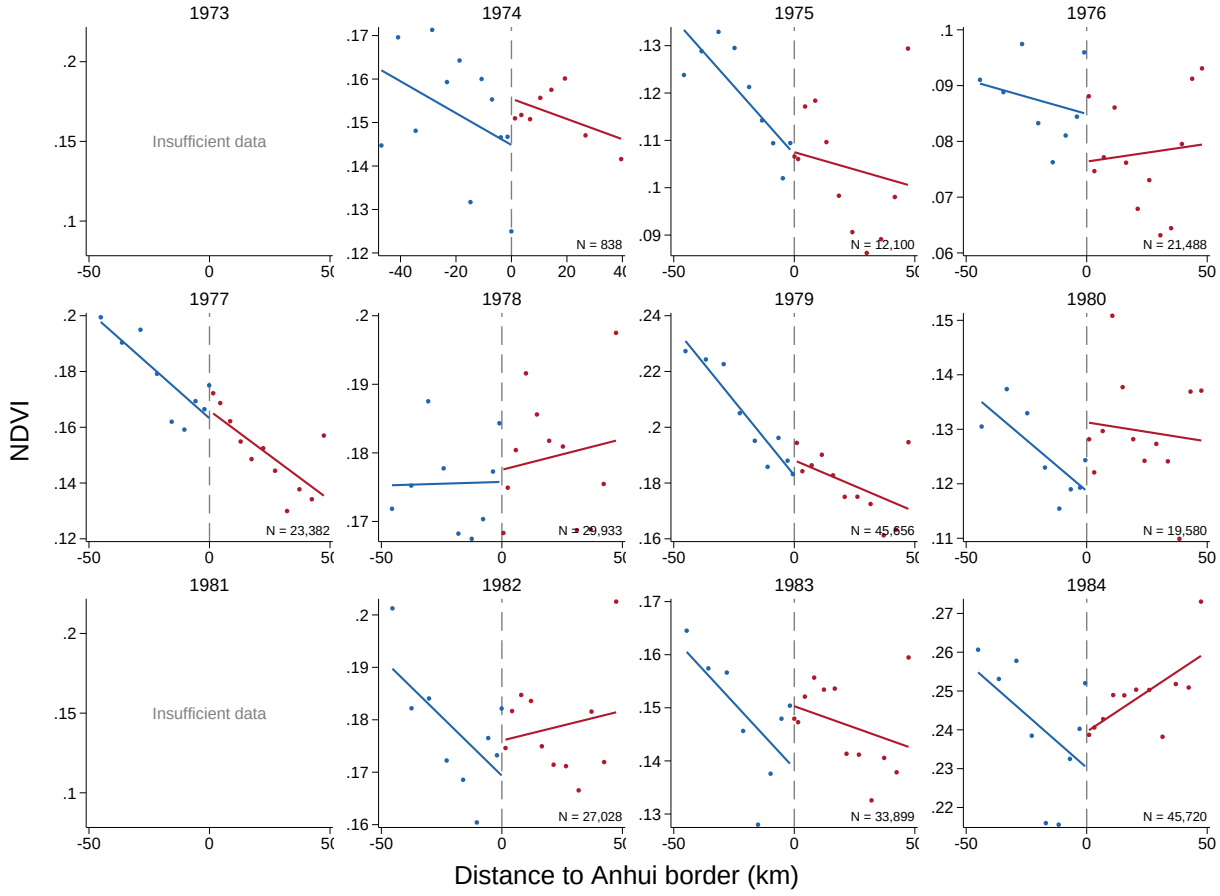
This figure reports a leave-one-province-out robustness exercise for the Landsat MSS border RD specification. Each row re-estimates the main RD effect after excluding one province from the sample (labeled "Drop [province]"). Points are coefficient estimates and horizontal bars are 95% confidence intervals. The estimates remain near zero across omissions, indicating that the baseline null effect is not driven by any single province.

Figure A.12: Anhui-Jiangsu Event-Study RD Estimates of HRS Effects on Landsat NDVI



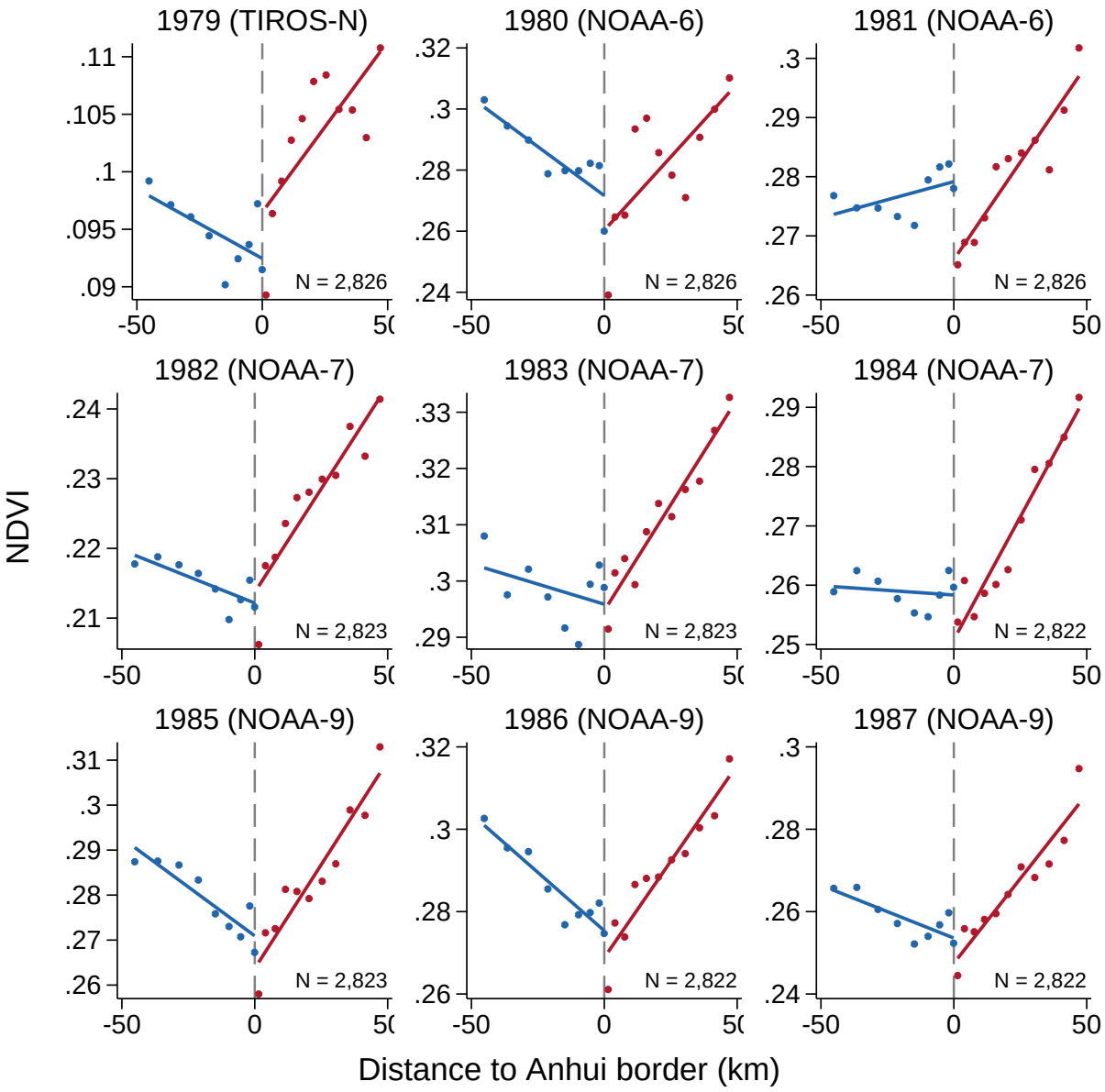
This figure reports year-specific difference-in-discontinuity coefficients at the Anhui-Jiangsu border, estimated from the same Landsat MSS specification as column 3 of Table 2, but replacing the post indicator with a full set of Anhui-by-year interactions and omitting 1977 as the reference year. The outcome is maximum yearly NDVI, the running variable is signed distance to the provincial border, and the specification includes cell fixed effects, year fixed effects, image-month fixed effects, and side-specific distance slopes that vary by year. Points are coefficient estimates and vertical bars are 95% confidence intervals. The dashed vertical line marks 1978, the onset of Anhui's reform period. There is no point for 1981 because the Anhui-Jiangsu border sample has no usable Landsat MSS observations in that year.

Figure A.13: Year-by-Year Landsat MSS RD Binscatters: Anhui–Jiangsu Border, 1973–1984



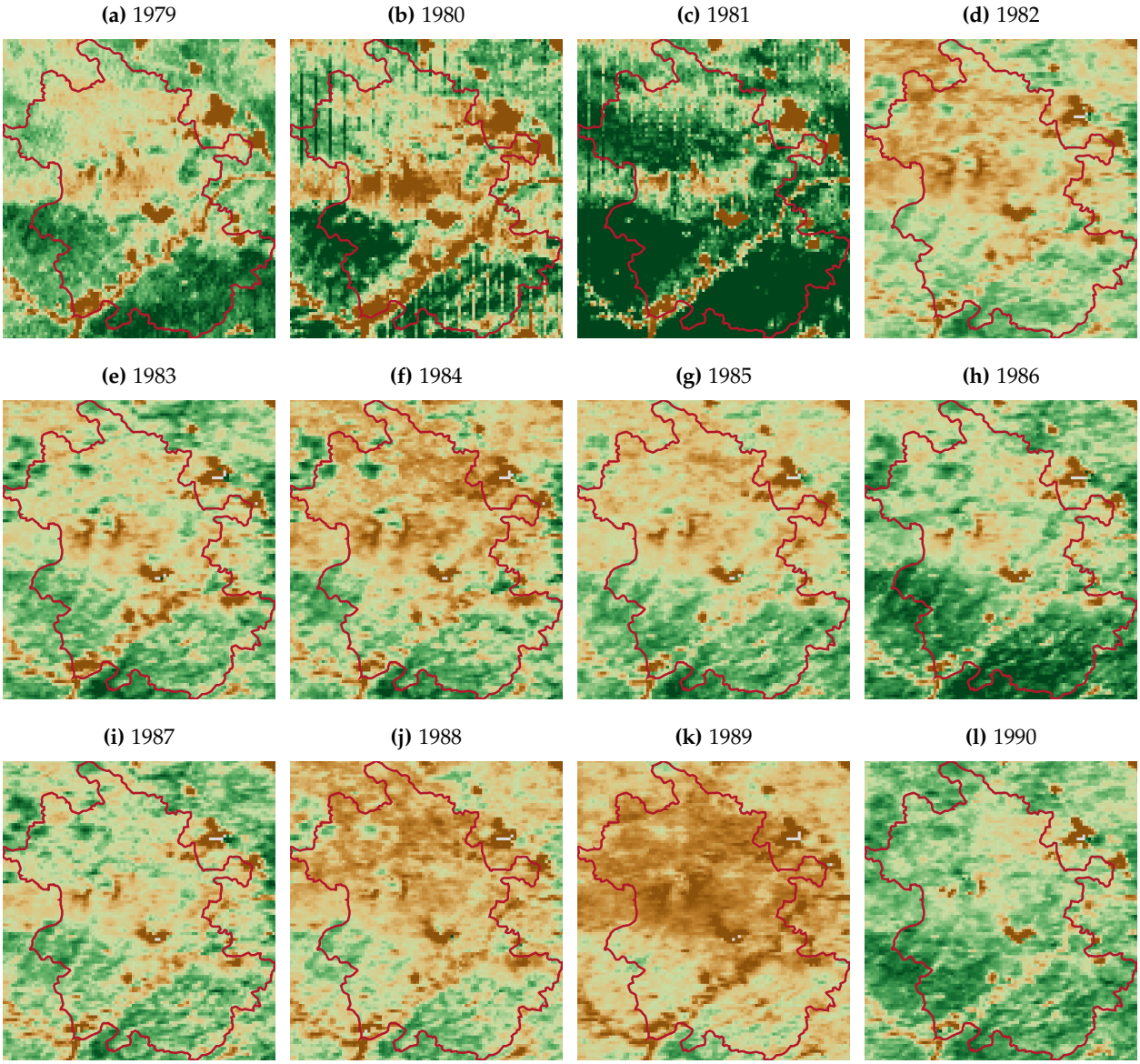
This figure shows year-by-year binned scatterplots of Landsat MSS NDVI against signed distance to the Anhui–Jiangsu provincial border for each year 1973–1984. Blue dots and fit lines denote the Jiangsu (non-reform) side; red dots and fit lines denote the Anhui (early-reform) side. Positive distance corresponds to cells inside Anhui. Each panel uses 20 equal-width distance bins with a linear fit on each side of the border. In pre-reform years (1973–1977), there is no systematic discontinuity at the border. After Anhui begins adopting the Household Responsibility System in 1978, a gap emerges with Anhui-side NDVI increasingly exceeding Jiangsu-side NDVI near the border. Some early years (1973, 1974, 1981) have limited spatial coverage due to sparse Landsat MSS acquisitions.

Figure A.14: Year-by-Year AVHRR RD Binscatters: Anhui–Jiangsu Border, 1979–1987



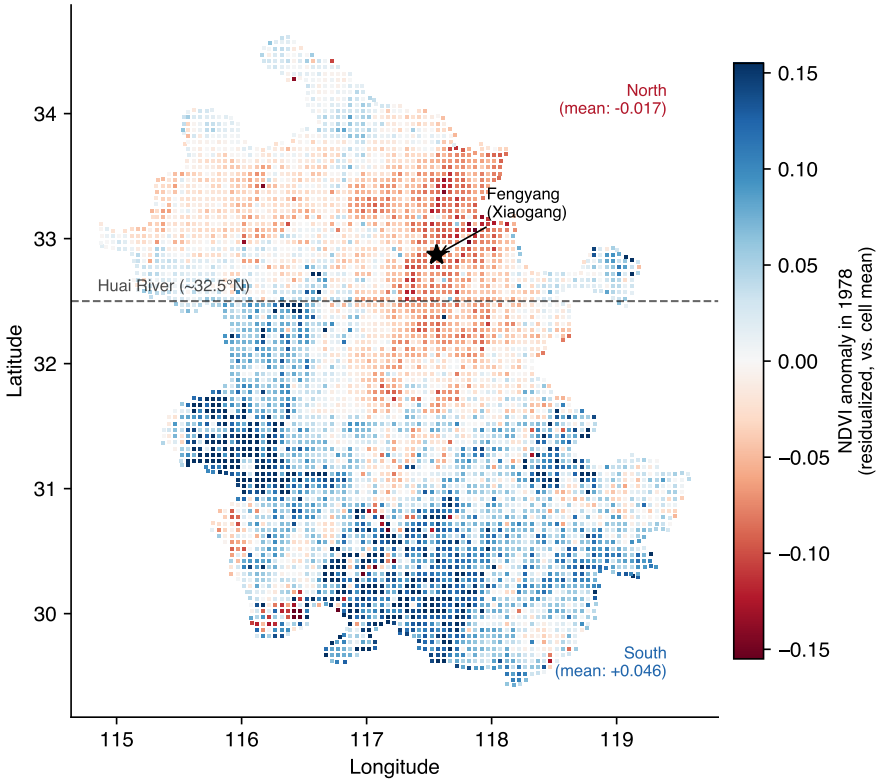
This figure shows year-by-year binned scatterplots of AVHRR NDVI against signed distance to the Anhui–Jiangsu provincial border for each year 1979–1987. Blue dots and fit lines denote the Jiangsu (non-reform) side; red dots and fit lines denote the Anhui (early-reform) side. Positive distance corresponds to cells inside Anhui. Each panel uses 20 equal-width distance bins with a linear fit on each side of the border. Panels are labeled by satellite platform: TIROS-N (1979), NOAA-6 (1980–1981), NOAA-7 (1982–1984), and NOAA-9 (1985–1987). In all years and across all four satellite platforms, Anhui-side NDVI exceeds Jiangsu-side NDVI near the border, with the discontinuity visible across sensor generations.

Figure A.15: Rice-Season AVHRR NDVI in Anhui Province, 1979–1990



This figure reports annual small-multiple maps of Anhui's AVHRR NDVI during the rice harvest season from 1979 to 1990. In each panel, the mapped value is the within-season maximum NDVI for each grid cell, with the province boundary overlaid. Green indicates higher vegetation levels and brown indicates lower levels. The sequence provides a year-by-year visual check of spatial NDVI patterns in one of the earliest HRS-adopting provinces.

Figure A.16: Spatial Pattern of the 1978 Anhui Drought in Landsat MSS NDVI

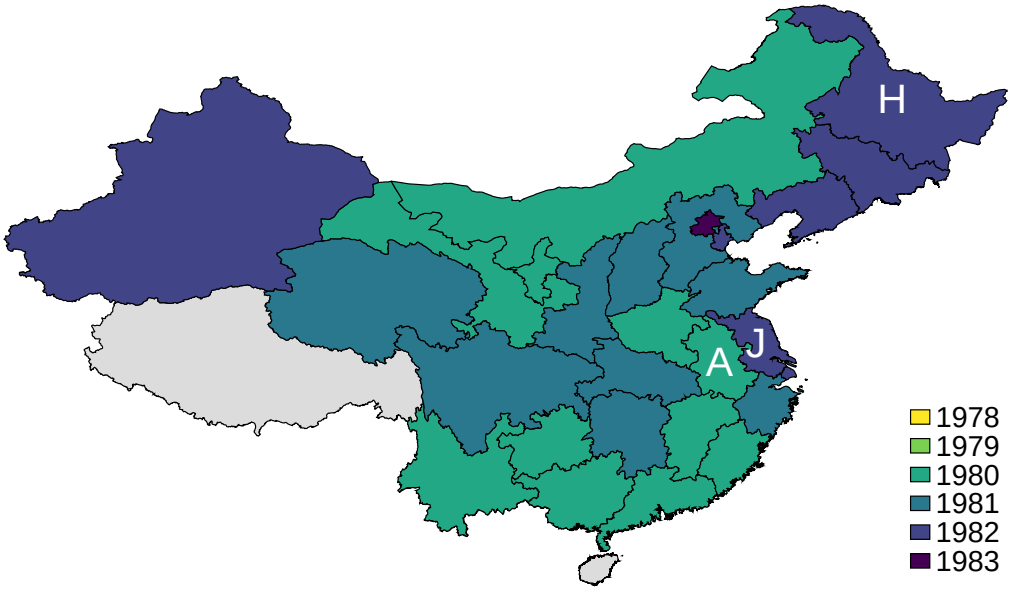


This figure maps cell-level 1978 Landsat MSS NDVI anomalies within Anhui. For each grid cell, annual mean NDVI is first residualized on the number of Landsat MSS images available for that cell-year, then the plotted anomaly is the 1978 residual minus the cell's average residual in all non-1978 years. The dashed horizontal line marks the Huai River (approximately 32.5°N), the star marks Fengyang County (Xiaogang), and the north/south annotations report the mean anomaly on either side of the river. More negative values indicate locations where 1978 NDVI fell further below the cell's baseline, showing that the drought shock was concentrated in northern Anhui.

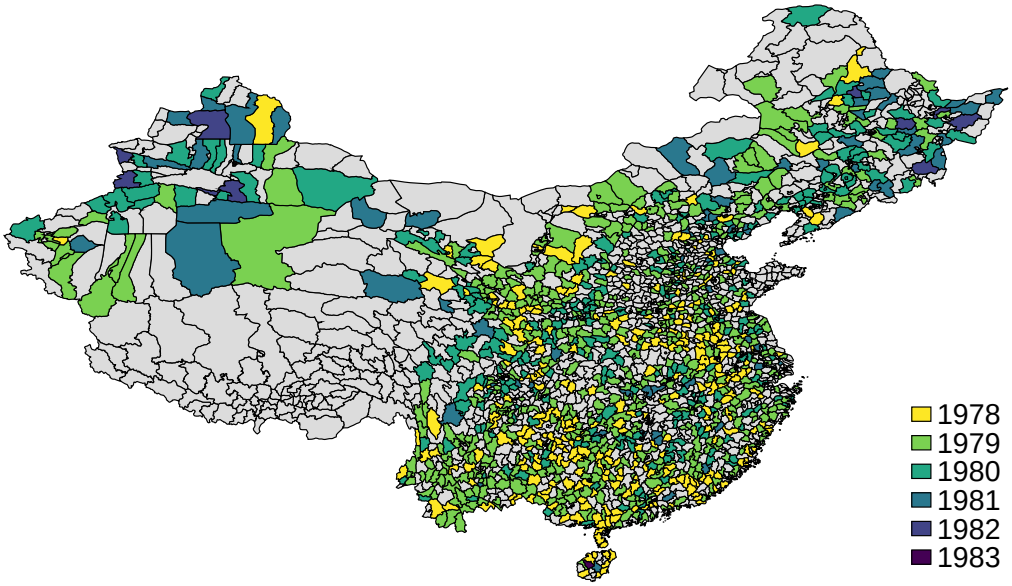
B The Household Responsibility System

Figure B.1: Household Responsibility System Treatment Maps

(a) Provincial treatment timing (Lin (1992))



(b) County treatment timing (Almond, Li, and Zhang (2019))



Panel (a) shows the dates at which over 50% of a province's work teams reported having adopted the Household Responsibility System (HRS). "A" marks Anhui Province, "J" marks Jiangsu Province, and "H" marks Heilongjiang Province. Panel (b) shows when Chinese counties (using 1990 boundaries) adopted the HRS using Almond, Li, and Zhang (2019)'s definition: when "collectively owned land was first contracted to individual households in a few villages".

Table B.1: Lin (1992) Replication

	(1)	(2)	(3)	(4)
	Lin (1992) HRS		Almond et al. (2019) HRS	
	VFO	log(Grain output)	VFO	log(Grain output)
HRS share	0.149**	0.172***	-0.0296	-0.0196
	(0.0500)	(0.0492)	(0.0587)	(0.0426)
ln(Land)	0.587***	-0.297**	0.552***	-0.118
	(0.0667)	(0.107)	(0.144)	(0.195)
ln(Labor)	0.151***	0.102*	0.156**	0.107*
	(0.0256)	(0.0424)	(0.0472)	(0.0535)
ln(Capital)	0.0960*	0.0493	0.0119	-0.241**
	(0.0396)	(0.0620)	(0.0693)	(0.0751)
ln(Fertilizer)	0.168***	0.140***	0.188***	0.170***
	(0.0216)	(0.0352)	(0.0264)	(0.0280)
Nongrain crops percentage	0.784***	-0.850*	0.637*	-1.597***
	(0.224)	(0.368)	(0.309)	(0.344)
Multiple cropping index	0.205*	0.273*	0.238*	0.310**
	(0.0832)	(0.138)	(0.111)	(0.119)
Year		-0.000289		0.0634***
		(0.0102)		(0.0125)
Market / input price (t-1)		0.0700		-0.361*
		(0.0858)		(0.167)
Govt / input price t		-0.0335		-0.148**
		(0.0347)		(0.0452)
Constant	2.624***	10.09	2.932***	-114.8***
	(0.341)	(20.08)	(0.735)	(24.67)
Observations	476	442	260	240
R^2	0.894	0.527	0.864	0.680

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Province-level panel regressions with province and year fixed effects, replicating Lin (1992) Table 5. VFO is log value of farm output in constant prices. Columns 1–2 use Lin's official provincial HRS adoption shares; columns 3–4 substitute county-level gazetteer HRS adoption from Almond et al. (2019), averaged to the province level. The smaller sample in columns 3–4 reflects provinces with available gazetteer data.

Table B.2: Lin (1992) Replication with Satellite NDVI

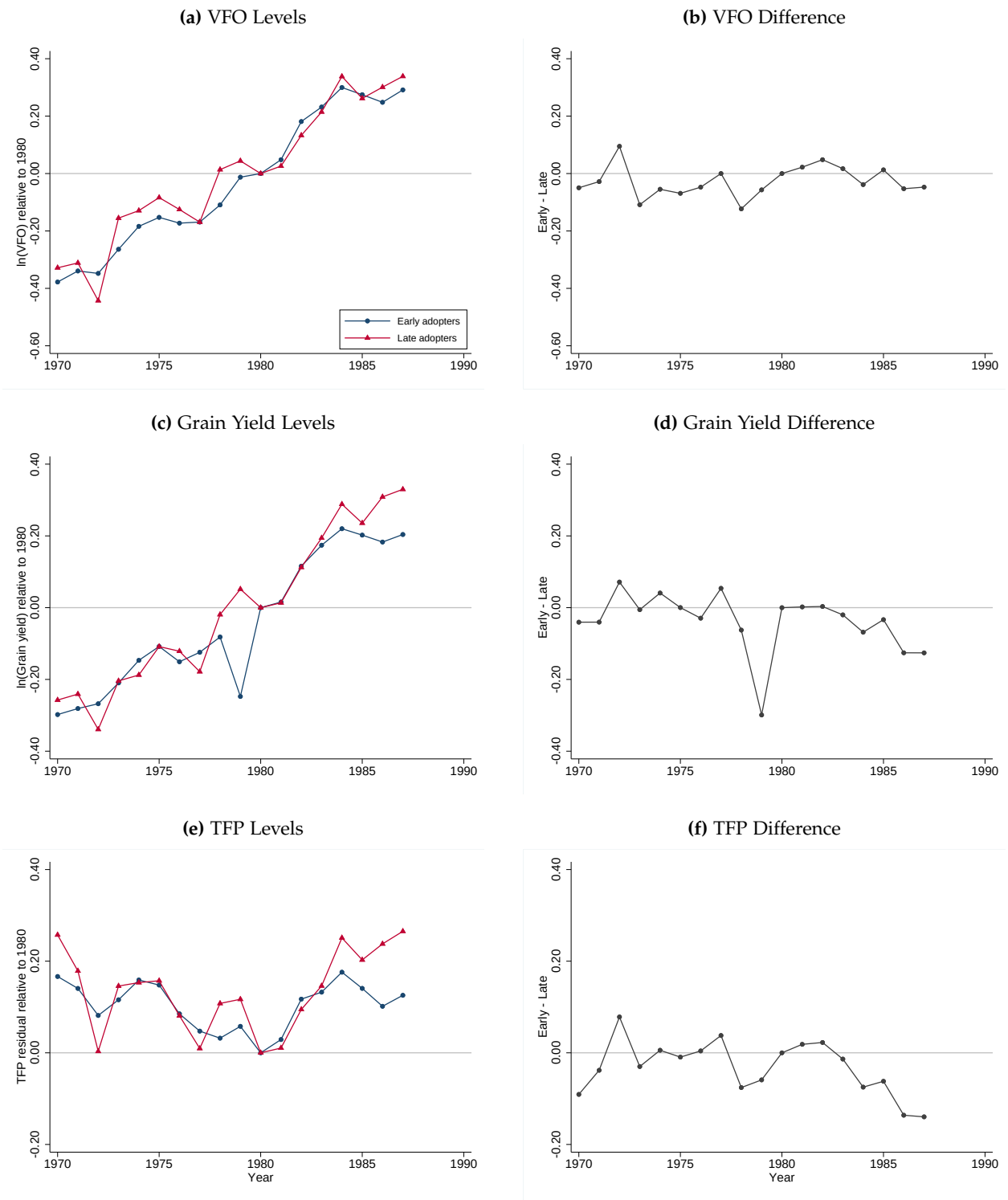
	Lin (1992) HRS			Almond et al. (2019) HRS		
	(1) VFO	(2) log(Grain)	(3) log(NDVI)	(4) VFO	(5) log(Grain)	(6) log(NDVI)
HRS share	0.149**	0.172***	0.323**	-0.0296	-0.0196	0.194
	(0.0500)	(0.0492)	(0.110)	(0.0587)	(0.0426)	(0.160)
ln(Land)	0.587***	-0.297**	3.563**	0.552***	-0.118	3.963**
	(0.0667)	(0.107)	(1.211)	(0.144)	(0.195)	(1.247)
ln(Labor)	0.151***	0.102*	-0.137	0.156**	0.107*	-0.164
	(0.0256)	(0.0424)	(0.0963)	(0.0472)	(0.0535)	(0.114)
ln(Capital)	0.0960*	0.0493	-0.171	0.0119	-0.241**	-0.334
	(0.0396)	(0.0620)	(0.274)	(0.0693)	(0.0751)	(0.301)
ln(Fertilizer)	0.168***	0.140***	0.142	0.188***	0.170***	0.258
	(0.0216)	(0.0352)	(0.139)	(0.0264)	(0.0280)	(0.151)
Nongrain crops percentage	0.784***	-0.850*	-2.643**	0.637*	-1.597***	-3.196***
	(0.224)	(0.368)	(0.859)	(0.309)	(0.344)	(0.894)
Multiple cropping index	0.205*	0.273*	0.123	0.238*	0.310**	-0.0632
	(0.0832)	(0.138)	(0.434)	(0.111)	(0.119)	(0.448)
Year		-0.000289	0.0730		0.0634***	0.122*
		(0.0102)	(0.0457)		(0.0125)	(0.0486)
Market / input price (t-1)		0.0700	-0.259		-0.361*	-0.388
		(0.0858)	(0.206)		(0.167)	(0.220)
Govt / input price t		-0.0335	0.0345		-0.148**	0.0498
		(0.0347)	(0.149)		(0.0452)	(0.199)
Constant	2.624***	10.09	-165.1	2.932***	-114.8***	-263.5**
	(0.341)	(20.08)	(92.18)	(0.735)	(24.67)	(97.15)
Observations	476	442	223	260	240	215
R^2	0.894	0.527	0.457	0.864	0.680	0.427

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Province-level panel regressions with province and year fixed effects, replicating Lin (1992) Table 5. VFO is log value of farm output in constant prices. NDVI is log province-level AVHRR NDVI (1978–1987). Columns 1–3 use Lin's official provincial HRS adoption shares; columns 4–6 substitute county-level gazetteer HRS adoption from Almond et al. (2019), averaged to the province level. The smaller sample in columns 3 and 6 reflects the availability of satellite data (1978–1987); the smaller sample in columns 4–6 reflects provinces with available gazetteer data.

Figure B.2: Early vs. Late Adopters of the HRS



This figure compares agricultural outcomes for early and late adopters of the Household Responsibility System, using the provincial panel data from Lin (1992). The left column shows levels for early and late adopters; the right column shows the difference (early minus late). Early adopters are provinces where the share of work teams adopting the HRS was above the median level of adoption in 1981 (0.33); late adopters are the remaining provinces. TFP is calculated by residualizing log VFO with respect to log land, labor, capital, and fertilizer.

Table B.3: Negative Selection into HRS Adoption

Panel A: County-level adoption (Almond, Li, and Zhang 2019)

	(1) Adopt 1979	(2) Adopt 1980	(3) Adopt 1981	(4) Adopt 1982+	(5) Adopt year
Avg. yield 1974–76 (SD)	-0.039*	-0.043	0.041	0.045*	0.171***
	(0.018)	(0.025)	(0.022)	(0.018)	(0.047)
Yield growth 1974–78 (SD)	-0.037*	-0.051**	0.024	0.070***	0.213***
	(0.015)	(0.017)	(0.026)	(0.012)	(0.053)
Observations	406	406	406	406	406
R^2	0.013	0.014	0.010	0.098	0.071

Standard errors (robust) in parentheses.

Both regressors standardized to mean zero, unit variance.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Panel B: Province-level adoption (Lin 1992)

	(1) Adopt 1981	(2) Adopt 1982	(3) Adopt 1983+	(4) Adopt year
Avg. yield 1974–76 (SD)	-0.062	-0.022	0.084	0.179
	(0.069)	(0.093)	(0.103)	(0.153)
Yield growth 1974–78 (SD)	-0.166*	0.098	0.067	0.224
	(0.066)	(0.086)	(0.118)	(0.172)
Observations	26	26	26	26
R^2	0.140	0.043	0.060	0.112

Standard errors (robust) in parentheses. $N = 28$ provinces.

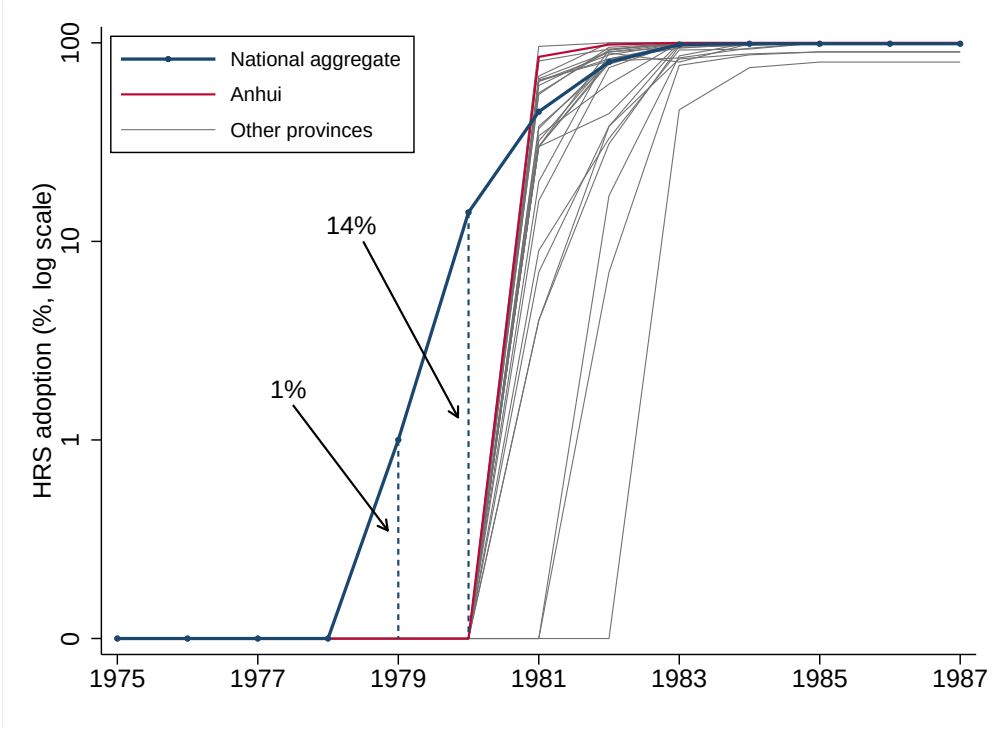
Both regressors standardized to mean zero, unit variance.

Adoption year = year province first exceeds 50% HRS share (Lin 1992).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Each column regresses an adoption indicator (or continuous adoption year) on two pre-reform predictors: average grain yield in 1974–1976 and grain yield growth from 1974 to 1978, both standardized to mean zero and unit variance. Panel A uses county-level gazetteer data from Almond, Li, and Zhang (2019) ($N = 406$ counties). Panel B uses province-level data from Lin (1992) ($N = 28$ provinces), where the adoption year is the first year the province exceeds 50% HRS share. Negative coefficients on early-adoption indicators indicate that worse-performing areas adopted first.

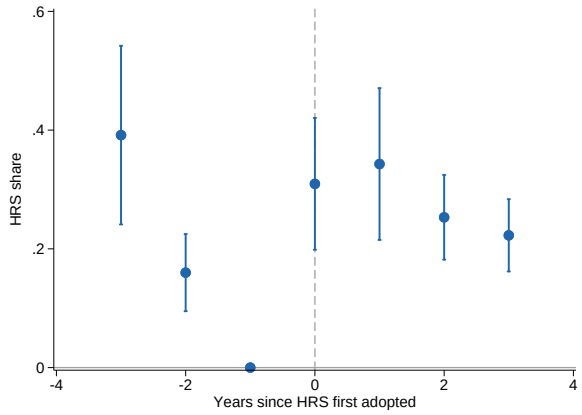
Figure B.3: HRS Adoption: National Aggregate vs. Provincial Data



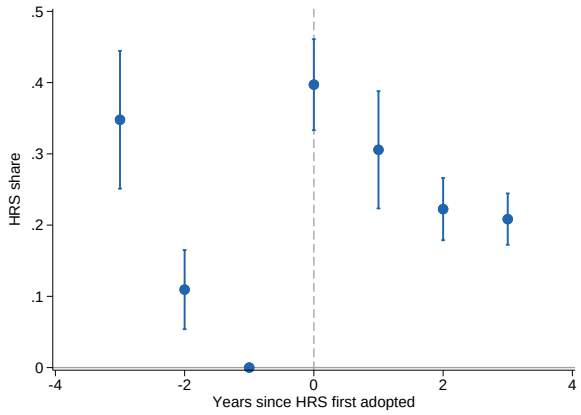
This figure plots the share of work teams that have adopted the HRS on a log scale, using data from Lin (1992). The national aggregate (dark blue) reports 1% adoption in 1979 and 14% in 1980, but all individual provinces (gray lines) report zero adoption until 1980 or later. Anhui Province (red) is highlighted as the earliest adopter. The discrepancy between the national aggregate and the provincial series suggests inconsistencies in the underlying adoption data.

Figure B.4: First-Stage Border Discontinuity in HRS Adoption Intensity Across Treatment Thresholds

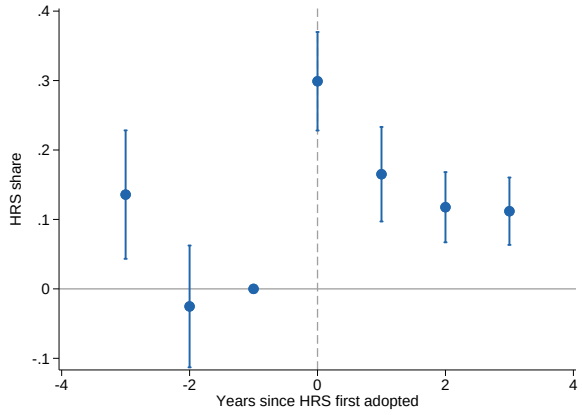
(a) Binary treatment cutoff: HRS share ≥ 0.25



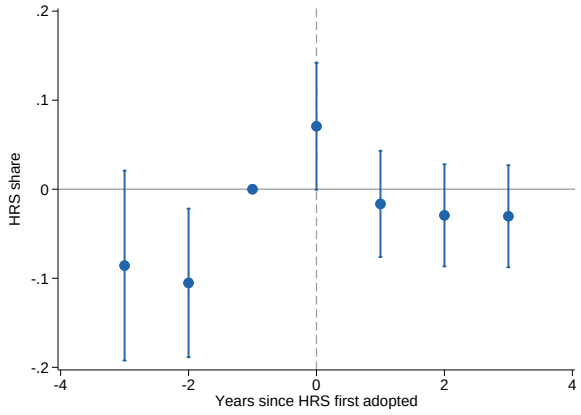
(b) Binary treatment cutoff: HRS share ≥ 0.50



(c) Binary treatment cutoff: HRS share ≥ 0.75



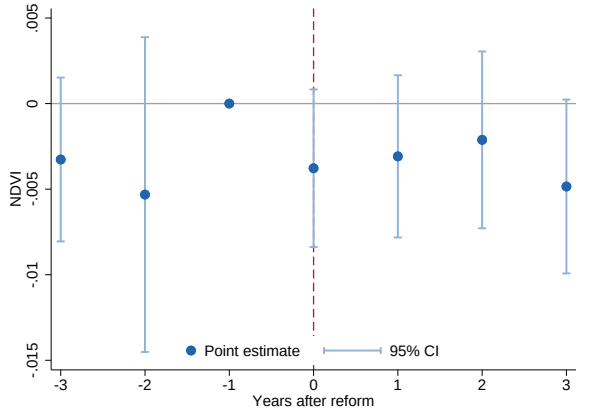
(d) Binary treatment cutoff: HRS share ≥ 0.90



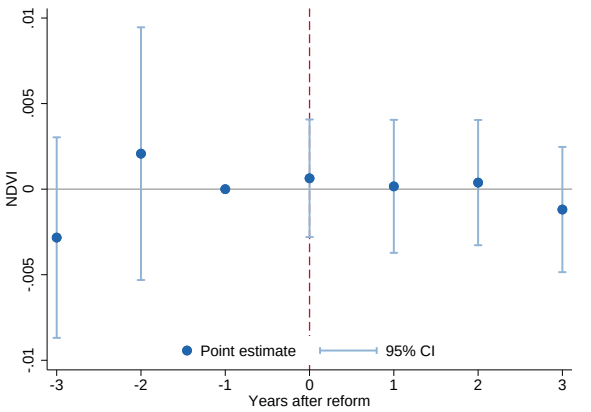
This figure reports the first-stage diagnostic for the border diff-in-discontinuities design using alternative binary treatment rules. In each panel, treatment status is defined by whether province-year HRS adoption share exceeds the listed cutoff, while the outcome is the underlying continuous HRS adoption share (0 to 1). The discontinuity estimate at the province border therefore measures how much reform intensity differs across the border under each treatment definition. Across the 25%, 50%, 75%, and 90% cutoffs, the border jump in HRS share remains visible, consistent with meaningful treatment-intensity separation at the design cutoff.

Figure B.5: Reduced-Form Border Discontinuity in NDVI Across Treatment Thresholds

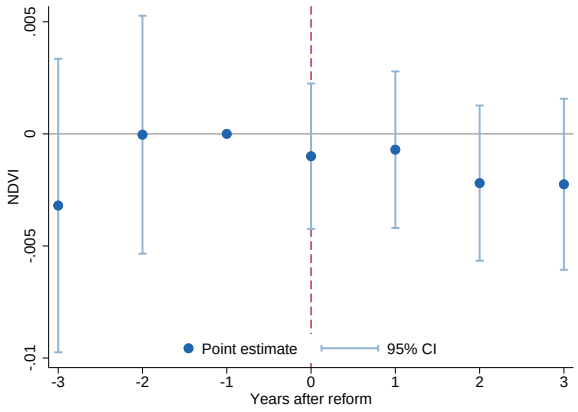
(a) Binary treatment cutoff: HRS share ≥ 0.25



(b) Binary treatment cutoff: HRS share ≥ 0.50

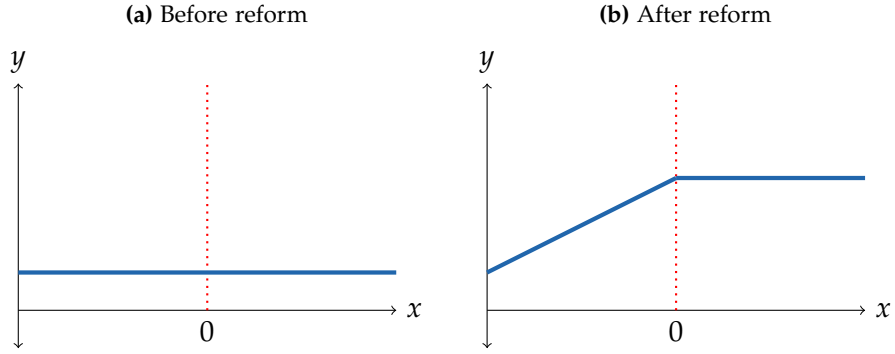


(c) Binary treatment cutoff: HRS share ≥ 0.75



This figure reports the reduced-form differences-in-discontinuities estimates of HRS adoption on maximum NDVI using alternative binary treatment rules. In each panel, treatment status is defined by whether province-year HRS adoption share exceeds the listed cutoff. Across the 25%, 50%, and 75% cutoffs, the estimated effect of decollectivization on NDVI remains near zero and statistically insignificant, confirming that the null result is not driven by the choice of treatment threshold.

Figure B.6: Spillovers



This conceptual figure illustrates the spillover pattern that could attenuate border RD estimates toward zero even when the true treatment effect is positive. The horizontal axis is signed distance to the provincial border ($x = 0$). Before reform (left), outcomes are flat across space. After reform (right), the treated side has a higher level, while outcomes on the untreated side rise as the border is approached, consistent with spillovers that decay with distance from the treated neighbor.

Table B.4: Robustness of Border RD Estimates to Bandwidth and Donut Hole

	Bandwidth				Donut	
Panel A: Landsat MSS						
HRS share > 50%	0.0108	0.00725	0.00886	0.00850	0.0148	0.0119
	(0.00849)	(0.00736)	(0.00603)	(0.00572)	(0.0109)	(0.0116)
Observations	2,014,094	1,755,998	1,467,244	1,304,534	1,544,287	1,301,348
Bandwidth	50km	40km	30km	25km	50km	50km
Donut	—	—	—	—	5km	10km
Panel B: AVHRR						
HRS share > 50%	-0.000909	-0.00105	-0.00109	-0.00115	-0.000823	-0.000603
	(0.00179)	(0.00158)	(0.00136)	(0.00125)	(0.00238)	(0.00274)
Observations	1,439,311	1,232,287	1,002,555	876,806	1,150,040	987,969
Bandwidth	50km	40km	30km	25km	50km	50km
Donut	—	—	—	—	5km	10km

Standard errors in parentheses

All specifications include time-varying distance slopes

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table reports robustness of the main border diff-in-discontinuities estimate to alternative bandwidth restrictions and donut holes. Columns 1–4 progressively narrow the bandwidth from 50km to 25km around the provincial border. Columns 5–6 exclude grid cells within 5km or 10km of the border (donut hole), retaining a 50km outer bandwidth. All specifications include cell fixed effects, border-group–year fixed effects, image-month fixed effects, and time-varying distance slopes. Standard errors are clustered by border group.

Table B.5: Border RD Slopes by Side: Testing for Spatial Spillovers

	Linear slopes		10km bin slopes	
	(1)	(2)	(3)	(4)
	MSS	AVHRR	MSS	AVHRR
HRS share $> 50\%$	0.00563	-0.000692	0.0139	0.00383
	(0.00781)	(0.00226)	(0.00958)	(0.00199)
HRS $\times$ Dist. (early adopter)	0.000418	0.000120		
	(0.000234)	(0.0000640)		
HRS $\times$ Dist. (late adopter)	0.0000485	0.0000673		
	(0.000319)	(0.0000711)		
HRS $\times$ 0–10km (early)			-0.00713	-0.00383
			(0.00814)	(0.00232)
HRS $\times$ 10–20km (early)			0.00516	-0.00363*
			(0.00588)	(0.00167)
HRS $\times$ 20–30km (early)			0.00592	-0.00135
			(0.00538)	(0.00117)
HRS $\times$ 30–40km (early)			0.00695	0.000197
			(0.00417)	(0.000928)
HRS $\times$ 0–10km (late)			-0.0201	-0.00498
			(0.0121)	(0.00280)
HRS $\times$ 10–20km (late)			-0.00951	-0.00281
			(0.0129)	(0.00247)
HRS $\times$ 20–30km (late)			-0.0108	-0.00324
			(0.0116)	(0.00216)
HRS $\times$ 30–40km (late)			-0.0113	-0.00218
			(0.0103)	(0.00210)
Observations	2,014,094	1,439,311	2,014,094	1,439,311

Standard errors in parentheses

Reference category: 40–50km bin. All at 50km bandwidth.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table tests whether the treatment effect varies with distance from the border, separately for early-adopting and late-adopting provinces within each border group. Columns 1–2 interact the treatment indicator with distance from the border (in km) on each side; a negative coefficient indicates that the treatment effect is stronger near the border. Columns 3–4 replace the linear distance interaction with 10km bin dummies (reference: 40–50km bin). "Early adopter" is the province in each border pair that first crosses the 50% HRS adoption threshold. If the HRS spills over across borders, we would expect the late-adopter side to show larger treatment effects near the border—i.e., negative coefficients on the late-adopter distance interactions. All specifications use a 50km bandwidth with cell, border-group–year, and image-month fixed effects. Standard errors are clustered by border group.

Table B.6: Dosage: HRS Adoption Intensity Quartiles

	Landsat MSS			AVHRR		
	(1)	(2)	(3)	(4)	(5)	(6)
HRS $\in (0, 25\%]$	-0.0786*	-0.0605	-0.0598	0.0130	0.0147	0.0164
	(0.0370)	(0.0372)	(0.0358)	(0.00917)	(0.00838)	(0.00924)
HRS $\in (25, 50\%]$	-0.00863	-0.0127	-0.0134	0.00744	0.00784	0.0101
	(0.00775)	(0.00870)	(0.00931)	(0.00842)	(0.00843)	(0.00849)
HRS $\in (50, 75\%]$	-0.00673	-0.00916	-0.00874	0.0106	0.00889	0.0112
	(0.0215)	(0.0218)	(0.0220)	(0.00902)	(0.00833)	(0.00822)
HRS $> 75\%$	0	0	0	0.00815	0.00619	0.0101
	(.)	(.)	(.)	(0.00850)	(0.00869)	(0.00835)
Observations	2,014,094	2,014,094	2,014,094	1,439,311	1,439,311	1,439,311
Distance Slopes	No	Yes	Yes	No	Yes	Yes
Time-varying Slopes	No	No	Yes	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

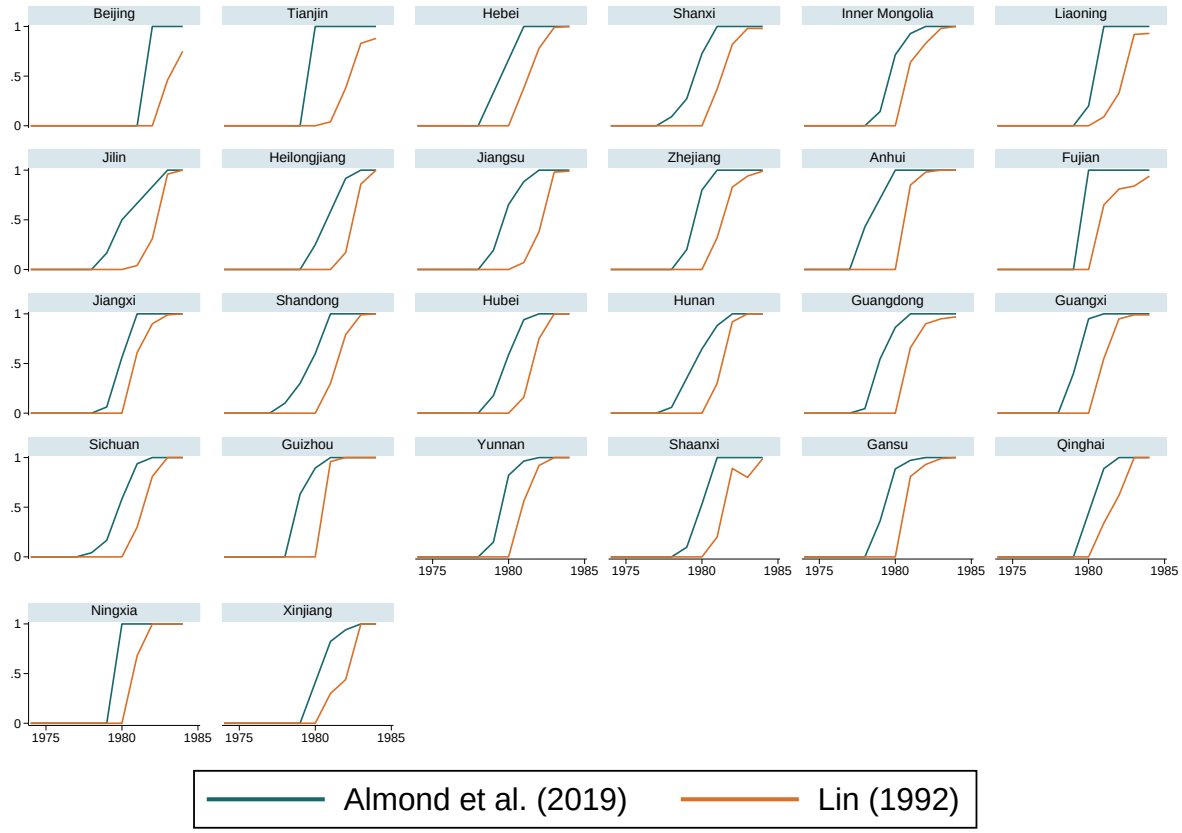
This table replaces the binary treatment indicator (HRS $> 50\%$) with quartile dummies for HRS adoption intensity. The reference category is HRS $= 0$ (no adoption). Quartile bins are: $(0, 25\%]$, $(25, 50\%]$, $(50, 75\%]$, and $> 75\%$. Columns 1–3 use Landsat MSS; columns 4–6 use AVHRR. All specifications include distance slopes as controls (not shown). All specifications use a 50km bandwidth with cell, border-group–year, and image-month fixed effects. Standard errors are clustered by border group.

Table B.7: Literature Estimates of the Effects of the Household Responsibility System

	Years	Outcome	HRS Effect (pct pts)	HRS share
McMillan, Whalley, and Zhu (1989)	1978-84	Agricultural TFP	78	78%
Lin (1992)	1978-84	Gross value of farm output	46.9	48.7%
Almond, Li, and Zhang (2019)	1974-84	Grain per capita	15.1	80%
Gibson (2020)	1979-84	Grain output	53	177%

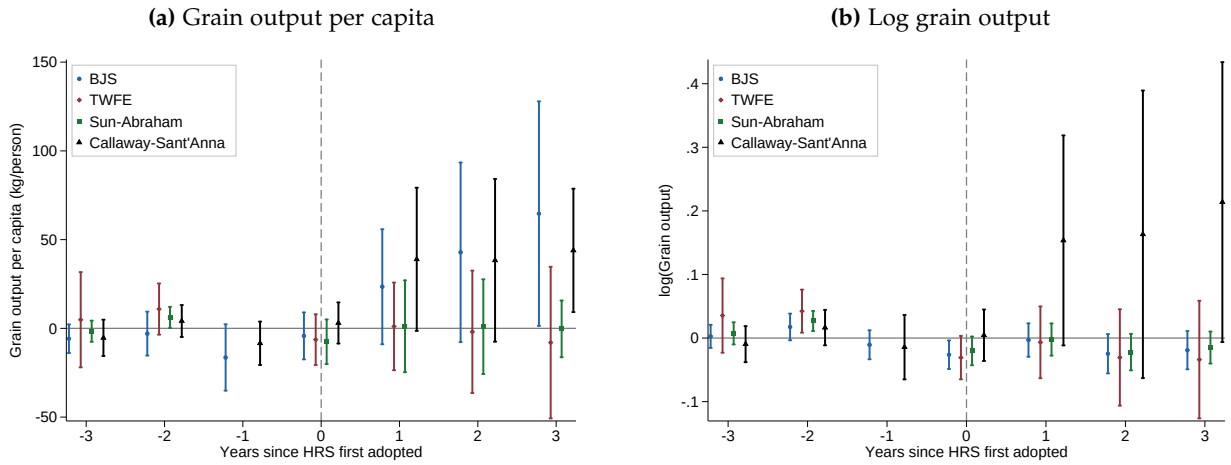
This table summarizes the literature's estimates of the HRS's effect on agricultural outcomes. The effect for Almond, Li, and Zhang (2019) is the percentage change between grain per capita observed 4 years after treatment (388.3 kg per capita) and grain yield at time of treatment (337.2 kg per capita), from their replication data. The overall growth for Almond, Li, and Zhang (2019) is the unweighted average county-level change in grain per capita from 1978 to 1984. We calculate overall grain output for Gibson (2020) using the official aggregate grain output increase from 1978-84 (30%).

Figure B.7: Comparing Treatment Variation in Almond, Li, and Zhang (2019) and Lin (1992)



This figure compares Household Responsibility System treatment dates in Lin (1992) and Almond, Li, and Zhang (2019) by province. The Almond, Li, and Zhang (2019) county-level data on when "collectively owned land was first contracted to individual households in a few villages" is collapsed to provincial means, and compared with the official provincial level share of work teams who have adopted HRS in Lin (1992).

Figure B.8: County Gazetteer Outcomes: BJS, TWFE, and Callaway-Sant'Anna



This figure reports county-level event studies for the Almond, Li, and Zhang (2019) gazetteer outcomes, with three estimators overlaid in each panel: Borusyak, Jaravel, and Spiess (2023) imputation (BJS), TWFE, and Callaway-Sant'Anna.

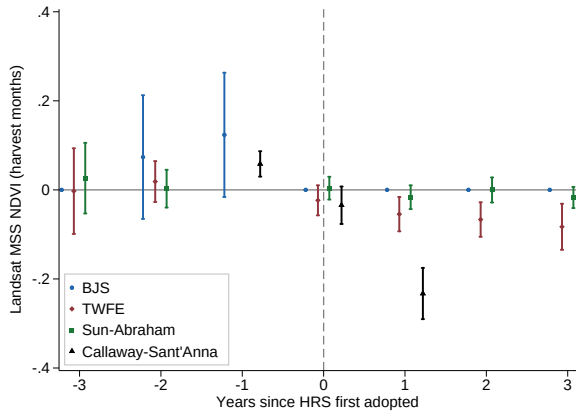
Table B.8: Pooled Pre- and Post-Treatment Effects

	Grain per capita		log(Grain output)	
	Pre	Post	Pre	Post
BJS	-4.405	31.678	0.010	-0.018
	(4.998)	(19.344)	(0.009)	(0.012)
TWFE	7.863	-3.802	0.039	-0.025
	(10.391)	(14.464)	(0.023)	(0.031)
Sun-Abraham	2.266	-1.428	0.017*	-0.015
	(2.620)	(9.334)	(0.007)	(0.010)
Callaway-Sant'Anna	-0.604	31.052*	0.003	0.134
	(2.962)	(14.912)	(0.008)	(0.074)

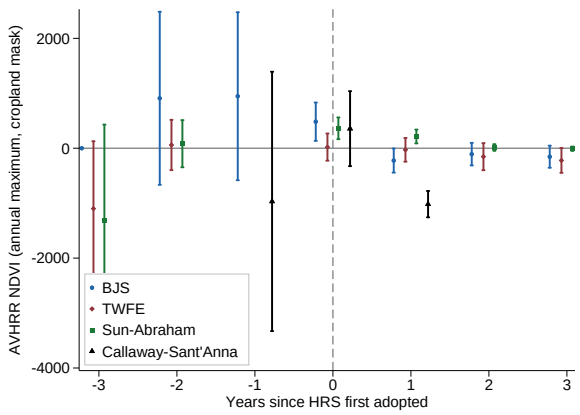
Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Pre: avg. of $h \in \{-3, -2\}$; Post: avg. of $h \in \{0, 1, 2, 3\}$.

Figure B.9: County-Level NDVI Event Study: Alternative Measures

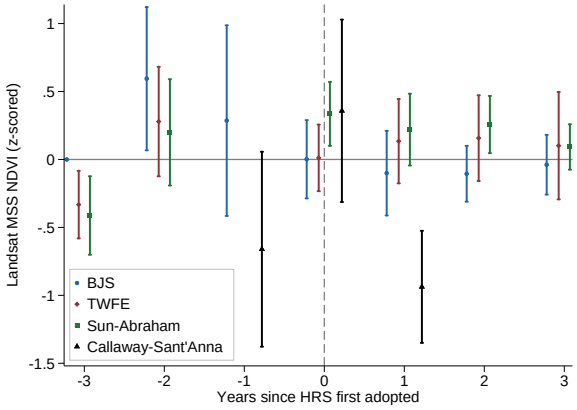
(a) Harvest-month MSS NDVI



(b) AVHRR NDVI (cropland mask)

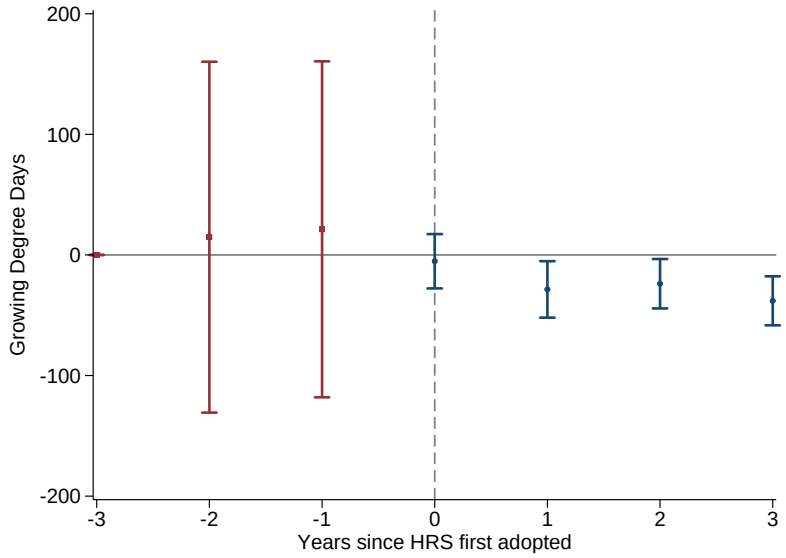


(c) Z-scored MSS NDVI



This figure repeats the county-level event study from Figure 7 using alternative vegetation index measures: (a) harvest-month (June–July) Landsat MSS NDVI, (b) AVHRR NDVI with a cropland mask, and (c) z-scored Landsat MSS NDVI. All panels show four estimators: BJS, TWFE, Sun-Abraham, and Callaway-Sant'Anna.

Figure B.10: County-Level GDD Placebo: Weather Does Not Jump at HRS Adoption



This figure shows the BJS imputation event study with growing degree days (GDD) as the outcome, as a placebo test. If weather does not change discontinuously at the time of county-level HRS adoption, any NDVI effect is not driven by coincident weather shocks. The absence of significant pre-trends or post-treatment effects confirms that weather is not confounding the NDVI event study.

Table B.9: Lin Replication: Lin vs Teiwes & Sun Treatment

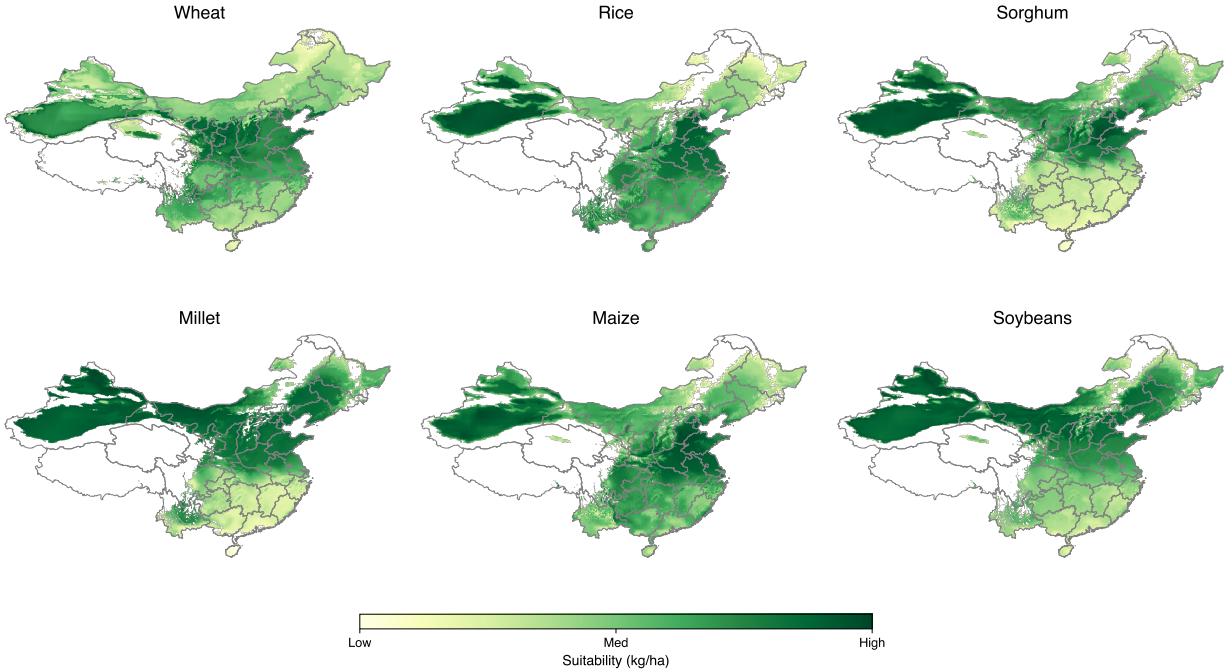
	(1) VFO (Lin)	(2) VFO (step)	(3) VFO (linear)	(4) Grain (Lin)	(5) Grain (step)	(6) Grain (linea
HRS share	0.149**			0.172***		
	(0.0500)			(0.0492)		
ln(Land)	0.587***	0.244*	0.251*	-0.297**	-0.720**	-0.734**
	(0.0667)	(0.0995)	(0.0994)	(0.107)	(0.239)	(0.243)
ln(Labor)	0.151***	0.247***	0.247***	0.102*	0.203***	0.222***
	(0.0256)	(0.0301)	(0.0301)	(0.0424)	(0.0606)	(0.0615)
ln(Capital)	0.0960*	0.155***	0.152***	0.0493	0.103	0.0959
	(0.0396)	(0.0389)	(0.0388)	(0.0620)	(0.0717)	(0.0728)
ln(Fertilizer)	0.168***	0.170***	0.171***	0.140***	0.156***	0.158***
	(0.0216)	(0.0200)	(0.0200)	(0.0352)	(0.0391)	(0.0397)
Nongrain crops percentage	0.784***	0.724***	0.700**	-0.850*	-0.756	-0.883*
	(0.224)	(0.213)	(0.213)	(0.368)	(0.409)	(0.415)
Multiple cropping index	0.205*	0.109	0.109	0.273*	0.183	0.144
	(0.0832)	(0.0760)	(0.0759)	(0.138)	(0.148)	(0.150)
Year				-0.000289	-0.00979	0.00578
				(0.0102)	(0.0110)	(0.0117)
HRS share (T&S stepwise)		-0.00262			0.166**	
		(0.0438)			(0.0505)	
HRS share (T&S linear)			-0.0370			0.0542
			(0.0479)			(0.0591)
Market / input price (t-1)				0.0700	0.0981	-0.0259
				(0.0858)	(0.0903)	(0.0952)
Govt / input price t				-0.0335	-0.0466	-0.0826*
				(0.0347)	(0.0354)	(0.0348)
Constant	2.624***	4.342***	4.312***	10.09	31.15	0.897
	(0.341)	(0.526)	(0.525)	(20.08)	(21.94)	(23.27)
Observations	476	408	408	442	391	391
R^2	0.894	0.912	0.912	0.527	0.541	0.529

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C Procurement Reform

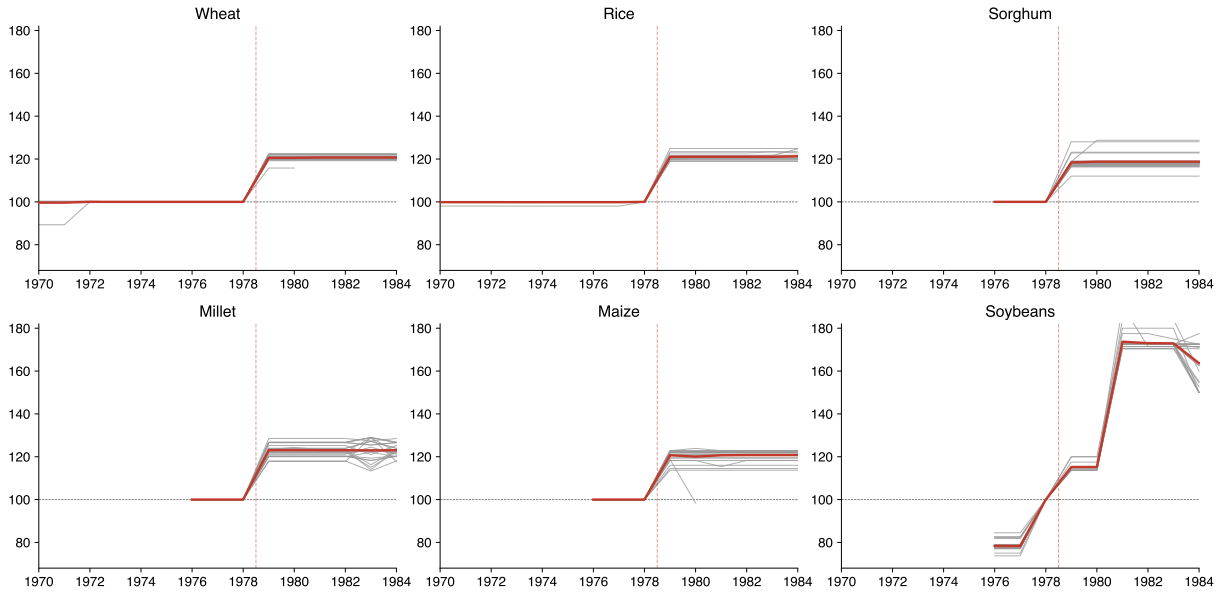
Figure C.1: Crop Suitability across China (GAEZ)



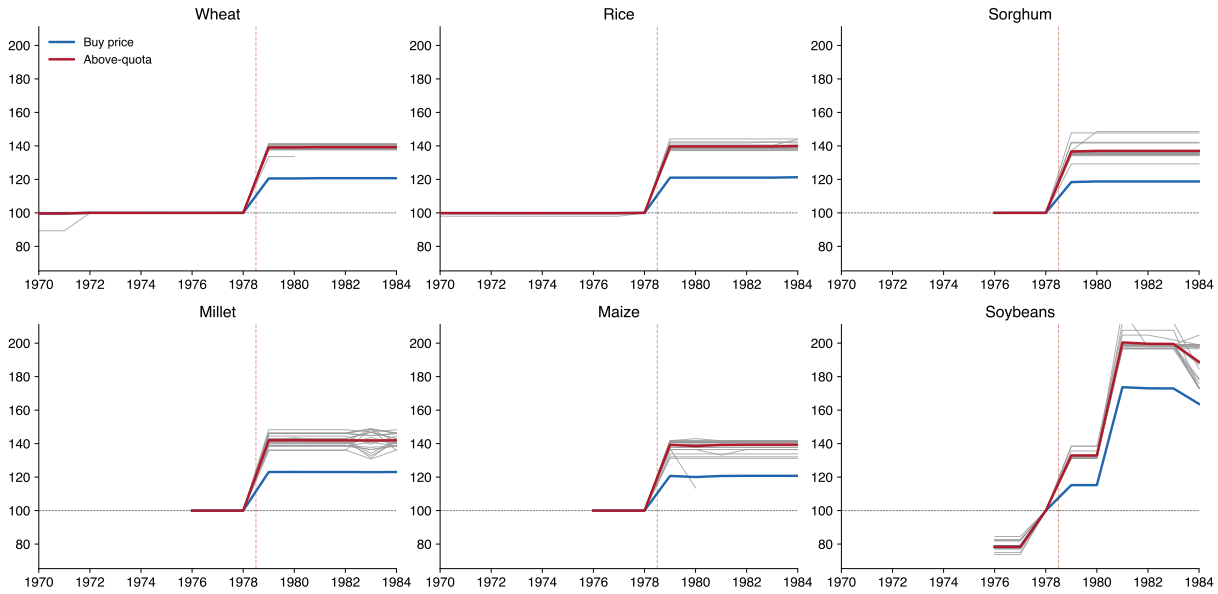
This figure shows the agro-ecological suitability of land across China for six major crops: wheat, rice, sorghum, millet, maize, and soybeans. Suitability is measured in potential yield (kg/ha) from the Global Agro-Ecological Zones (GAEZ) database, which estimates attainable yields based on climate, soil, and terrain conditions. Darker shading indicates higher suitability.

Figure C.2: Procurement Price Trends by Crop

(a) Standard procurement buy prices (1978=100)

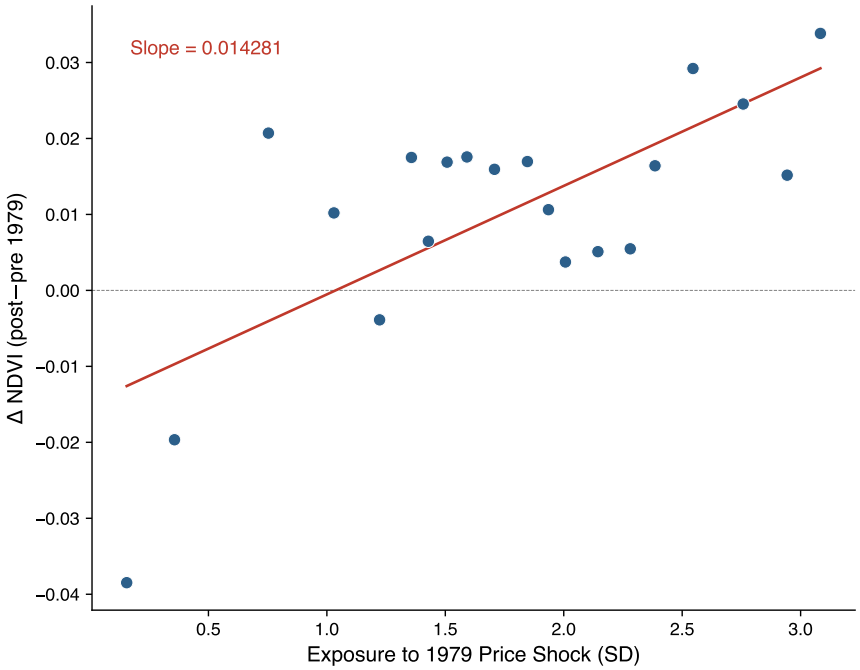


(b) Constructed above-quota series (130% pre-1979, 150% post-1979)



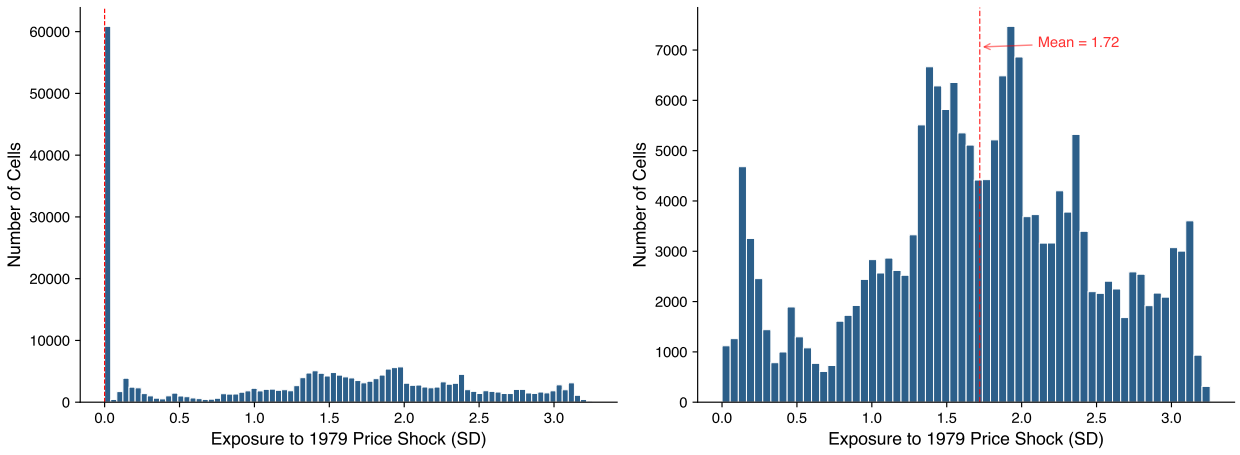
Panel (a) reports normalized procurement buy-price indices (1978=100) for six Bartik crops: wheat, rice, sorghum, millet, maize, and soybeans. Gray lines are province-specific series and the highlighted line is the cross-province mean. Panel (b) reports a constructed above-quota series that applies a 1.3 multiplier before 1979 and a 1.5 multiplier from 1979 onward to the same underlying crop prices, then re-normalizes to 1978=100; it overlays standard buy-price means and above-quota means for each crop. The vertical dashed marker at the 1978–1979 break indicates the timing of the procurement-price reform.

Figure C.3: Bartik Exposure and NDVI Growth (Binscatter)



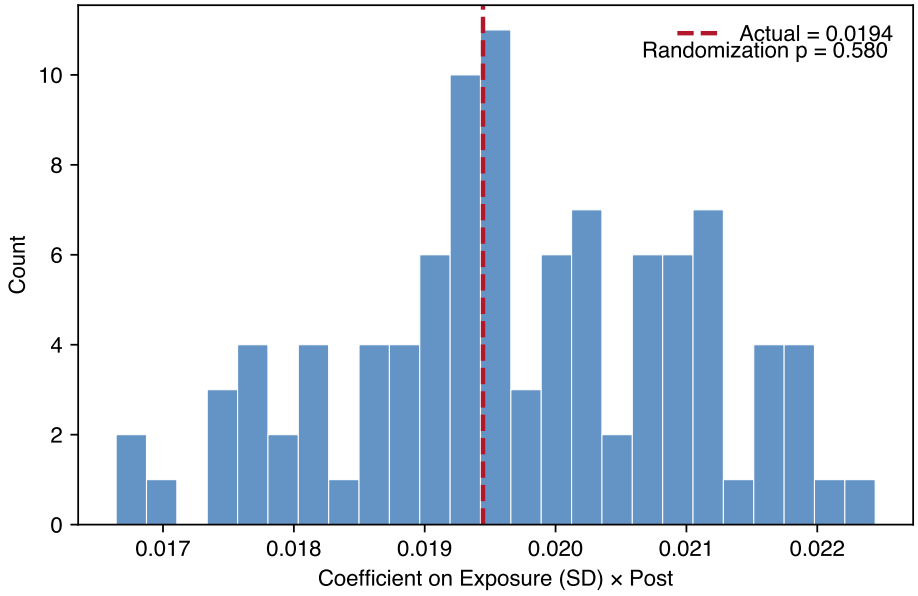
This figure plots binned means of post-1979 NDVI growth against grid-cell Bartik exposure to procurement price changes.

Figure C.4: Distribution of Grid-Cell Bartik Exposure



This figure shows the distribution of Bartik exposure across grid cells. Exposure is defined as the crop-suitability-weighted 1979 procurement price shock.

Figure C.5: Bartik Randomization-Inference Placebo (Permuted Crop Exposure Mapping)

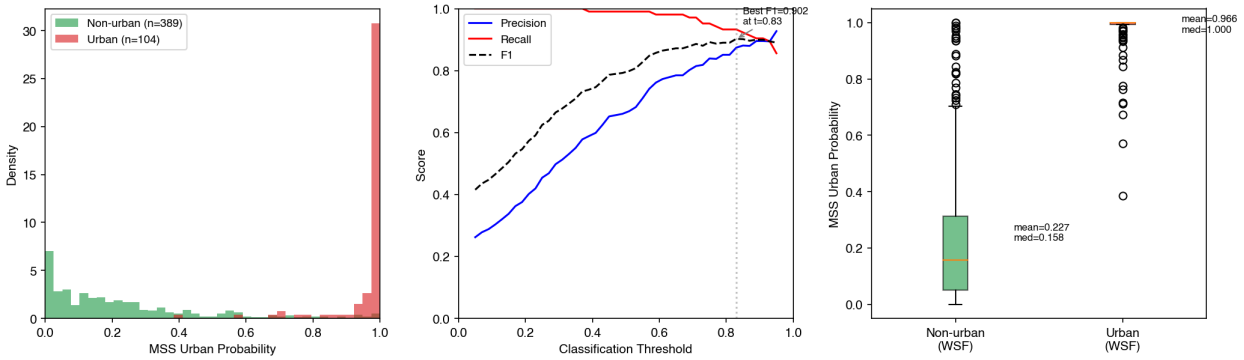


This figure reports a randomization-inference placebo for the Bartik procurement-price specification. In each placebo draw, we randomly permute which 1979 crop-specific procurement price shock is paired with each crop's GAEZ suitability layer, recompute exposure, and re-estimate the baseline static DiD with cell and year fixed effects. The histogram shows coefficients from 100 placebo draws; the dashed vertical line is the actual estimate from the true mapping, and the annotation reports the two-sided randomization p-value.

D Urban Prediction

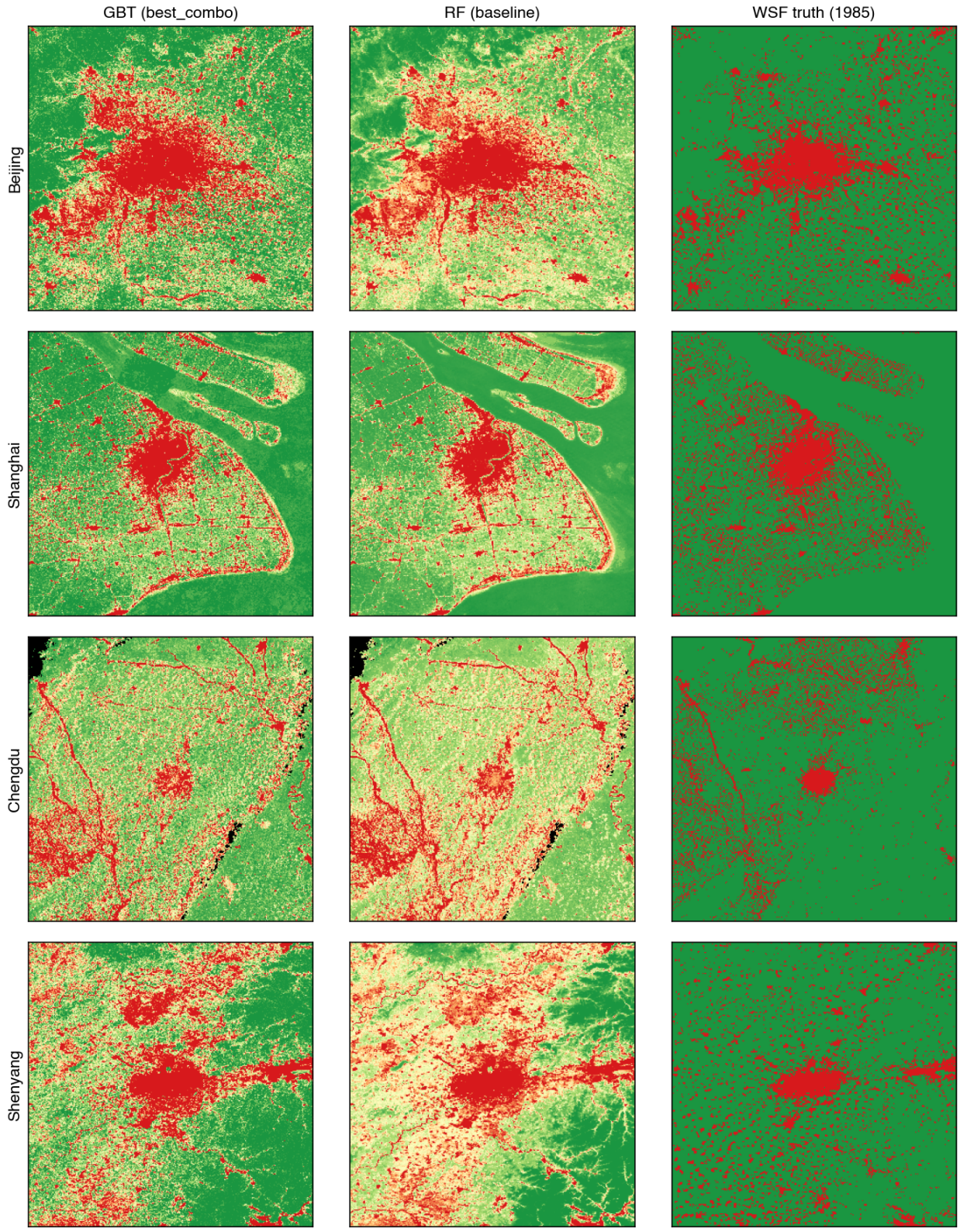
This section reports tuning and validation outputs for the urban-prediction pipeline.

Figure D.1: Urban Prediction Tuning: Best Hyperparameter Combination



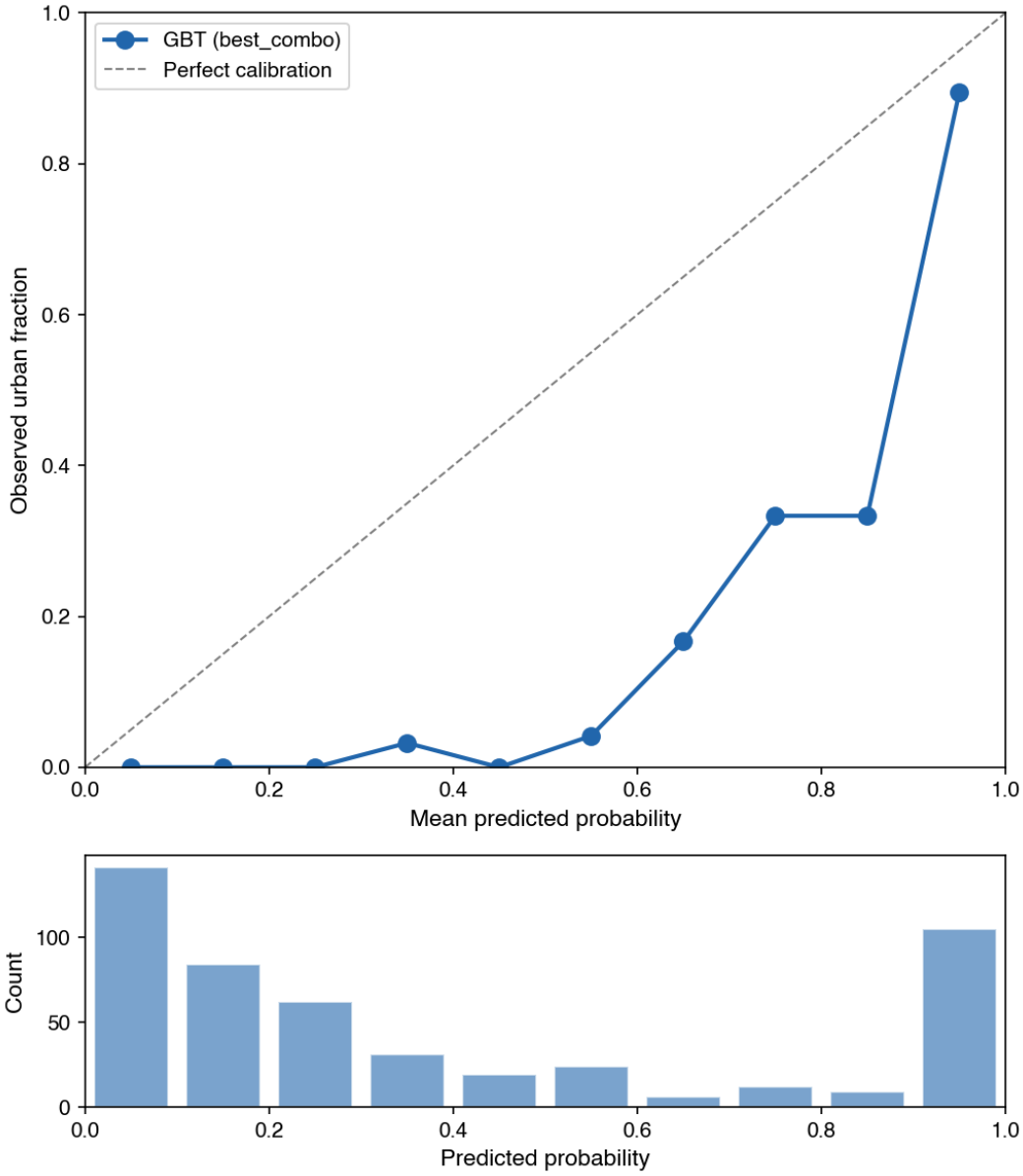
This figure reports the tuning output for the selected best-performing model configuration in the urban prediction exercise.

Figure D.2: Urban Prediction Validation: Comparison Maps



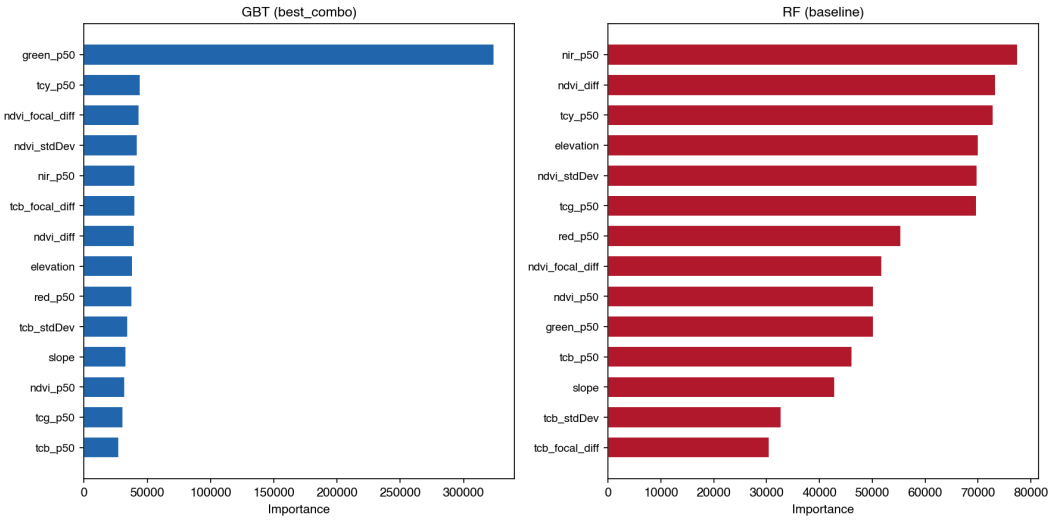
This figure compares predicted and reference urban-surface patterns across the validation maps used in the urban prediction pipeline.

Figure D.3: Urban Prediction Validation: Calibration



This figure reports calibration performance for the urban prediction model, comparing predicted values to validation outcomes.

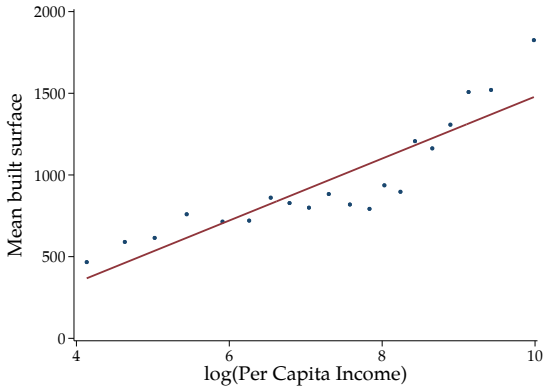
Figure D.4: Urban Prediction Validation: Feature Importance



This figure reports model-based feature-importance estimates from the urban prediction validation exercise.

E Non-agricultural Growth

Figure E.1: Built-up surface area predicts village income per capita



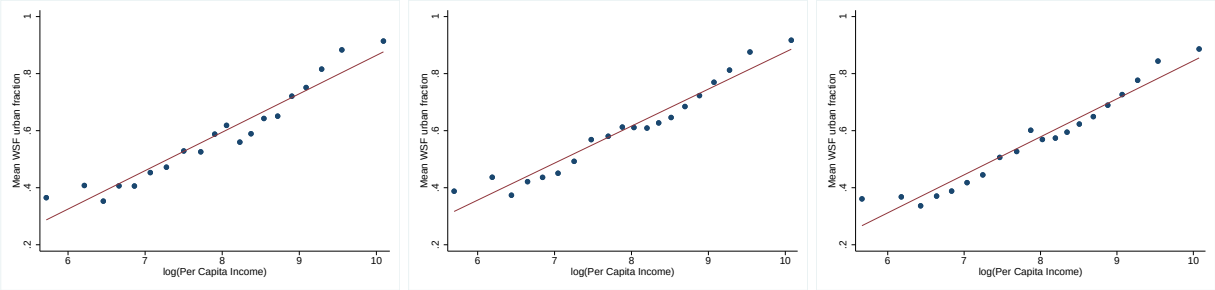
This figure plots binned means from a binscatter of mean GHSL built-up surface area in a 500m radius around each village center against log village income per capita (yuan) from the Contemporary Chinese Village Gazetteer.

Table E.1: Built-up surface area predicts village income per capita

	500m radius		1km radius		5km radius	
	(1)	(2)	(3)	(4)	(5)	(6)
log(Income per capita)	189.651***	118.132***	186.661***	123.931***	154.742***	84.111***
	(9.753)	(21.934)	(8.334)	(17.693)	(5.035)	(12.451)
Constant	-416.745***	97.626	-443.527***	7.942	-559.984***	-53.923
	(72.569)	(158.307)	(62.013)	(127.696)	(37.461)	(89.863)
Observations	4,792	4,468	4,792	4,468	4,792	4,468
R-squared	0.073	0.940	0.095	0.941	0.165	0.922
Location FE	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes

This table reports regressions of mean GHSL built-up surface area on log village income per capita from the Contemporary Chinese Village Gazetteer. Columns 1 and 2 use a 500m buffer around village coordinates; columns 3 and 4 use 1km; columns 5 and 6 use 5km. Columns 2, 4, and 6 include village and year fixed effects. Standard errors clustered by village are in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

Figure E.2: WSF urban fraction predicts village income per capita



This figure replicates the income-validation binscatter using mean WSF urban fraction as the outcome. The three panels use 500m, 1km, and 5km buffers around village coordinates and plot binned means against log village income per capita from the Contemporary Chinese Village Gazetteer.

Table E.2: WSF urban fraction predicts village income per capita

	500m radius		1km radius		5km radius	
	(1)	(2)	(3)	(4)	(5)	(6)
log(Income per capita)	0.135***	0.039***	0.130***	0.046***	0.133***	0.033***
	(0.004)	(0.010)	(0.003)	(0.008)	(0.003)	(0.006)
Constant	-0.483***	0.264***	-0.422***	0.232***	-0.488***	0.292***
	(0.031)	(0.081)	(0.026)	(0.064)	(0.022)	(0.050)
Observations	3,566	3,248	3,714	3,380	3,772	3,432
R-squared	0.253	0.886	0.295	0.905	0.387	0.933
Location FE	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes

This table reports regressions of mean WSF urban fraction on log village income per capita from the Contemporary Chinese Village Gazetteer. Columns 1 and 2 use a 500m buffer around village coordinates; columns 3 and 4 use 1km; columns 5 and 6 use 5km. Columns 2, 4, and 6 include village and year fixed effects. Standard errors clustered by village are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Data Appendix

Administrative Data Sources

Table .3: Administrative Data Sources Used in NDVI–Yield Validation

Sample	Source	Unit	Admin years	MSS years (N)	AVHRR years (N)	Coverage notes
South Korea	KOSTAT	*si*/*gun* (ADM2; *gu* dropped)	1965–2023	1972–1976 (408)	1996–2013 (2,324)	1977–1995 gap; no MSS-AVHRR overlap.
Japan	MAFF crop statistics	Prefecture (ADM1)	1959–2021	1982–1995 (180)	1979–2013 (305)	Overlap in 1982–1995.
India	ICRISAT (Brey-Hertweck harmonization)	District (ADM2)	1967–2020	1972–1994 (2,783)	1979–2005 (7,603)	Overlap in 1979–1991.
China (counties)	Almond-Li-Zhang gazetteers	County (ADM3)	1970–1990	1979–1984 (1,026)	1979–1984 (3,360)	Continuous overlap in 1979–1984.

Notes: Observation counts in parentheses are the regression samples reported in Table 1. "Admin years" reports the underlying administrative-series span used in the validation pipeline.