

Reassessing China's Rural Reforms: The View from Outer Space ^{*}

Joel Ferguson Oliver Kim[†]

UC Berkeley UC Berkeley

This Version: October 31, 2023

[\[Link to Latest Version\]](#)

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 agricultural production, 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 causal evidence that areas that adopted the HRS sooner experienced faster 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, Lai Wei, Mingqi Zeng, and Silin Zhang for excellent research assistance. 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 for generous support. 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 Development Lunch, UC Berkeley Trade Lunch, UC Berkeley Macro Lunch, and UC Berkeley Economic History Workshop for valuable comments and feedback. All errors are our own.

[†]Job Market Paper. 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 growth and 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 was responsible, 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, giving it even greater political-economic significance.

But how much can we trust the conventional wisdom about the HRS? Prior studies have relied on official Chinese data, largely without the benefit of modern causal inference techniques. Compared to the scale of the Chinese miracle, modern research on Deng Xiaoping's reforms is surprisingly rare, in large part because Chinese economic statistics are notoriously unreliable. Foreign researchers, Chinese economists, and even top Chinese leaders have all observed that Chinese economic statistics can reflect political priorities more than the ground truth.¹ One of the great disasters of the Communist period, the Great Leap Forward famine, was partly caused by party cadres inflating agricultural production to placate their superiors (Yang 2013). More recently, after the Covid-19 pandemic, the Chinese government has stopped publishing a

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

wide range of basic economic statistics, highlighting the continued political sensitivity towards inconvenient facts.

This paper expands our understanding of the causes of the Chinese miracle by using data from a novel source—historical satellite imagery. This paper’s first major contribution is the creation of new satellite-based measures of historical agricultural production. Applying machine learning methods from remote sensing and environmental science to satellite data, we train a random forest to predict yield from a set of countries with similar crops and conditions to China. We then take the trained model to Chinese remote sensing data to form a new, highly disaggregated dataset of grain yields, independent of the official Chinese provincial data. We then verify that these satellite-based measures can accurately predict yield through a range of validation exercises, including detecting the effects of severe weather shocks.

This paper’s second major contribution is to combine these satellite-based measures of yield with treatment data on the staggered rollout of the HRS across China to give us credible causal estimates of the HRS’s effects. Since the HRS was first permitted only in remote and famine-stricken areas, a causal design needs to take into account selection of treated areas. To address this, we employ two separate empirical designs, which take advantage of the unique granularity of our satellite data in both space and time. First, we apply a novel *staggered difference-in-discontinuities* strategy. Like in a classic regression discontinuity, we identify the causal effects of the HRS right at the boundary between provinces that adopted the HRS and provinces that did not. Moreover, by observing the same province borders over time, we can use fixed effects to control for any unchanging differences. Under the weak assumption that areas just on either side of the border would have followed parallel trends in the absence of treatment, we can then causally identify the effects of the HRS. Second, we use separate treatment data on the county-level rollout of reform compiled by Almond et al. (2019) from historical gazetteers, and find similar null effects on predicted yields using a staggered differences-in-differences design.

Our central finding is that the Household Responsibility System had a negligible, near-zero effect on grain yields. Our difference-in-discontinuities estimates are statistically precise—we can reject effect sizes on grain yields as small as 0.2 tons per hectare, even five years after the onset of reform—and robust to a wide range of estimation approaches. We do not find evidence of larger effects in provinces with a higher share of work teams adopting HRS, and furthermore find no

evidence that the lack of discontinuities are caused by confounding treatments or spillovers into neighboring areas. Similarly, our county-level adoption estimates—using an entirely separate source of semi-official treatment data—find no evidence that counties adopted the HRS sooner experienced faster yield growth.

Note that this paper estimates the *causal* effect of the Household Responsibility System on yields. These results are entirely consistent with a *general* rise in agricultural yields (which we plan to measure in future work), just that the HRS was not a cause. We present preliminary evidence that another reform—a 1979 rise in state procurement prices, which brought them closer to free market levels—may instead have been the main factor behind rising yields. Another reasonable possibility, suggested by qualitative studies of rural China, is that there was rampant grain hiding from the state before reform, causing the official level of grain production to be understated (Chan and Unger 1982; Oi 1991). Grain hiding may have gradually lessened as the state loosened its grip on agriculture, creating the appearance of an acceleration in yields, with little change on the ground. But, for the moment, these hypotheses remain speculative—the present paper must remain silent on issues of aggregates.

Nonetheless, this paper challenges the near-universal view of the Chinese miracle that agricultural decollectivization boosted yields. This result has implications far beyond agriculture. One of the main drivers of Chinese macroeconomic growth has been the enormous movement of labor from rural to urban areas. 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 takeoff (Young 2003; Brandt et al. 2008). Moreover, for human welfare, 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.

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 et al. 1989). In

particular, it builds on Lin (1992), which estimates a production function over a panel of official provincial data and finds that the HRS accounted for half of overall agricultural output growth from 1978-84. It also builds on the subsequent work of Almond et al. (2019), which assembles county-level data on the rollout of the HRS to identify a positive effect on grain output per capita. (We will use the Almond et al. (2019) treatment data as an alternative identification strategy.) 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 paper 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 et al. 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). 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 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 et al. 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 et al. 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 hard 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 our machine learning model to predict yields from satellite imagery. [Section 4](#) explains why we need to use satellites to overcome reporting biases and outlines our empirical strategy. [Section 5](#) presents our main empirical results, and discusses their implications. [Section 6](#) concludes, and presents directions for future work.

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. More detail on the specific institutional structures of rural reform will be made available in an online History Appendix.

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 their net income among households based on their size ([Eisenman \(2018\)](#), p. 193). If the team produced a surplus, a system of

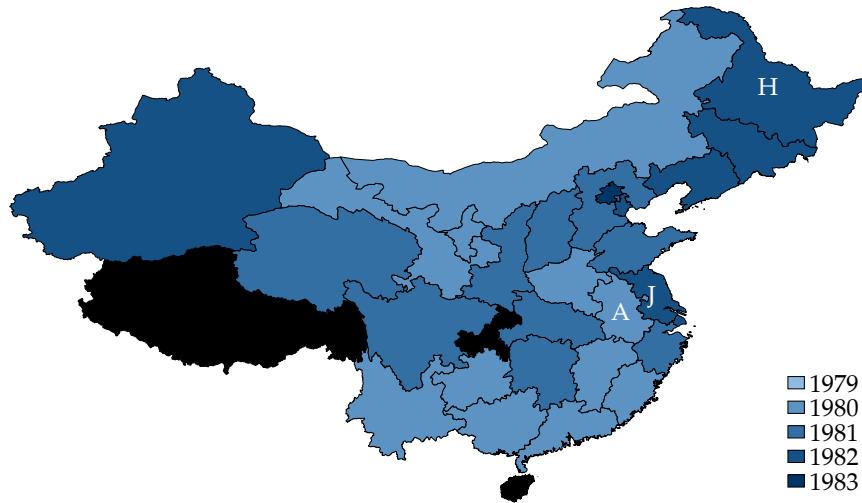
work points gave households a cash bonus, usually based on their observed labor inputs. But because supervision is costly, standard neoclassical agricultural models predict that there are strong incentives to shirk on a collective farm, overwhelming any of the potential efficiency gains from economies 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, in response to a severe drought, work teams in Anhui Province 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 household 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.

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. Figure 1 shows the date when over 50% of each province’s work teams reported the Household Responsibility System (HRS). Anhui was the clear national

Figure 1: Adoption Dates of Household Responsibility



This figure shows the dates at which over 50% of the 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. Provincial HRS adoption data is from Lin (1992).

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 (Chöng (2000), p. 94). When peasants began implementing household responsibility in 1978, Wan Li allowed the experiment to continue. Later, he protected 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 province with heavily mechanized agriculture. 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 (Chöng (2000)). Heilongjiang only caught up after the HRS was made national policy, 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 hardliners in Beijing soon openly clashed over the reform question.² At first, the hardliners held sway—November 1978's Third Plenum Regulations explicitly banned *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" (Chöng (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 stylized facts will inform our identification strategy: we will use the staggered

²Contrary to popular belief, Deng Xiaoping was not involved in the rural reform process until reforms were well underway. He made his first public comments on *baochan daohu* only on May 31, 1980, long after decollectivization had begun in Anhui.

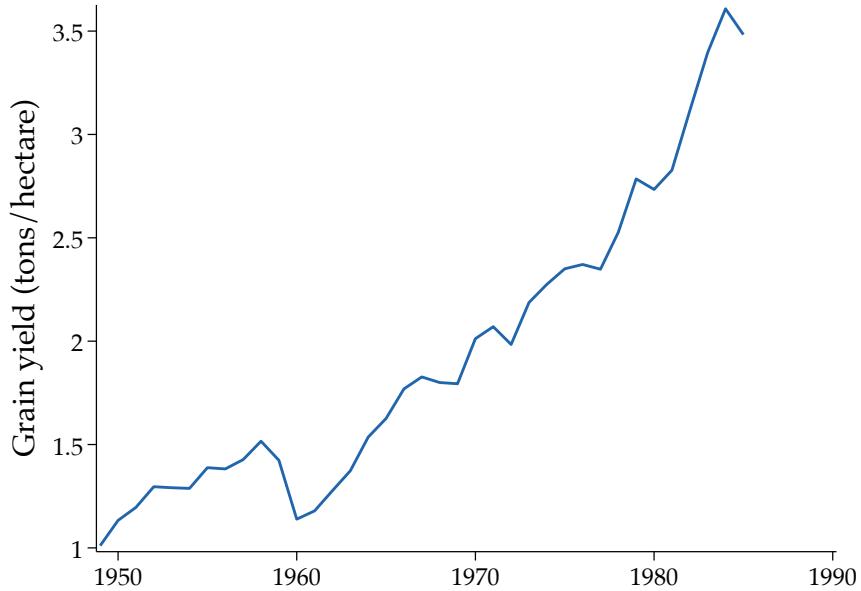
rollout of the HRS across provinces as a source of variation, while taking into account that areas that reformed first were negatively selected.

The Productivity Effects of Reform [Figure 2](#) shows how official grain yields rose 43% between 1978 and 1984, a 6% annual rate—a doubling of the 3% rate from 1949 to 1978. Grain output is the sum of wheat, rice, barley, and other cereal output (by weight), and the most commonly reported output statistic in official Chinese agricultural data. Spurred by this apparent success, the spread of decollectivization was gradually sanctioned by a growing number of provinces, until the newly christened Household Responsibility System (HRS) became national policy in 1982.

The scholarly consensus, summarized in [Table A5](#), has largely supported the official view that the HRS had a large, positive effect on agricultural growth. The outcome measures used by prior papers vary from agricultural TFP to grain per capita, but all show a sizable contribution of the HRS reform to agricultural growth. However, all have had to rely on central or local government data sources, which has led some to express skepticism about the effects of the HRS. A CIA report from 1983 attributed yield growth instead to “good weather, increased use of fertilizer, and other related reforms, such as a more rational state pricing system” (CIA [1983](#)). Looking even at the official national statistics, Bramall ([2004](#)) notes that agricultural output growth accelerated in 1976-80, when land reform hadn’t even been implemented by most work teams until 1982. He also notes that provinces that decollectivized earlier (Anhui, Sichuan) did not seem to grow faster than provinces that decollectivized later (Heilongjiang, Jiangsu).

The “more rational state pricing reform” mentioned in the CIA report refers to another plausible contributor to rising yields: the 1979 procurement price rises. In 1979, under Hua Guofeng, the state raised the average procurement price for quota grain by 20% and the bonus for above-quota grain from 30% to 50% ([Sicular 1988](#)). 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. McMillan et al. ([1989](#)), Lin ([1992](#)), and Bramall ([2004](#)) all attribute some portion of agricultural productivity growth to the price reform, but the former two sources (based on official aggregate statistics) emphasize the HRS as the main driver. Disentangling the effects of the HRS from the near-contemporaneous price reform will require careful

Figure 2: National grain yields



This figure shows national grain yields (output in kilograms / hectares of land) for China, from the State Statistical Bureau (SSB).

causal identification. We turn to constructing the satellite-based data needed for this analysis in the following section.

3 Measuring Yields with Satellites

3.1 Data Sources

This paper uses imagery from the Advanced Very High Resolution Radiometer (AVHRR) to measure agricultural yields. AVHRR, which was carried by National Oceanic and Atmospheric Administration (NOAA) satellites, collected imagery at red and near-infrared bands at a 4km resolution, twice daily, from late 1978 to 2013. To train our agricultural models, we pair this satellite data with precipitation and temperature data from the European Union's Copernicus ERA5 dataset, which is available at the 0.25 degree level at the hourly level from 1940 onwards.

We aggregate our satellite and weather data up both over space and time. To make computation feasible, we bin observations into a grid of 0.05-degree cells, which are roughly 5km squares around the latitude of Anhui Province (31 degrees N). We use the Global Land Surface Satellite-

Global Land Cover (GLASS-GLC) dataset, also derived from AVHRR, to mask out pixels that are not agricultural land (Liu et al. 2020).³ To observe crops over the growing cycle, and to close gaps in coverage caused by clouds, we aggregate AVHRR observations up to the weekly level. This gives us a dataset with 27,858,445 grid-cell observations across all of China, with coverage from October 1978 to December 1990. Full technical details are available in an online appendix.

To train our models and validate our satellite-based data, we use ground truth data at the province level from the *Agricultural Statistics of the People's Republic of China, 1949-1990*, compiled by the State Statistical Bureau. We pair this with the original provincial level panel collected by Lin (1992). To further test and validate our satellite-based yields, we also compile ground truth data from nearby countries: prefecture-level yield data from Japan's Ministry of Agriculture, Forestry, and Fisheries and Indian district-level yield data from the International Crop Research Institute for the Semi-Arid Tropics (ICRISAT).

3.2 The Normalized Difference Vegetation Index (NDVI)

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 the higher-frequency light that can damage their growth (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):

$$NDVI = \frac{NIR - Red}{NIR + Red}. \quad (1)$$

Dating back to the 1970s, NDVI has become one of the central measures in a vast literature measuring agricultural outcomes using satellite data.⁴ When combined with weather data like precipitation and temperature, models using NDVI can predict yield with a high degree of accuracy.

[Figure 3](#) presents two key features of NDVI in our setting. First, [Figure 3a](#) shows the evolution

³The GLASS dataset is only available for 1982 onwards. For observations before 1982, we use the 1982 cropland mask. Results after 1982 are robust to not using the cropland mask at all.

⁴For examples, see Tucker (1979), Hamar et al. (1996), and Bognár et al. (2022).

Figure 3: Features of the Normalized Difference Vegetation Index (NDVI)

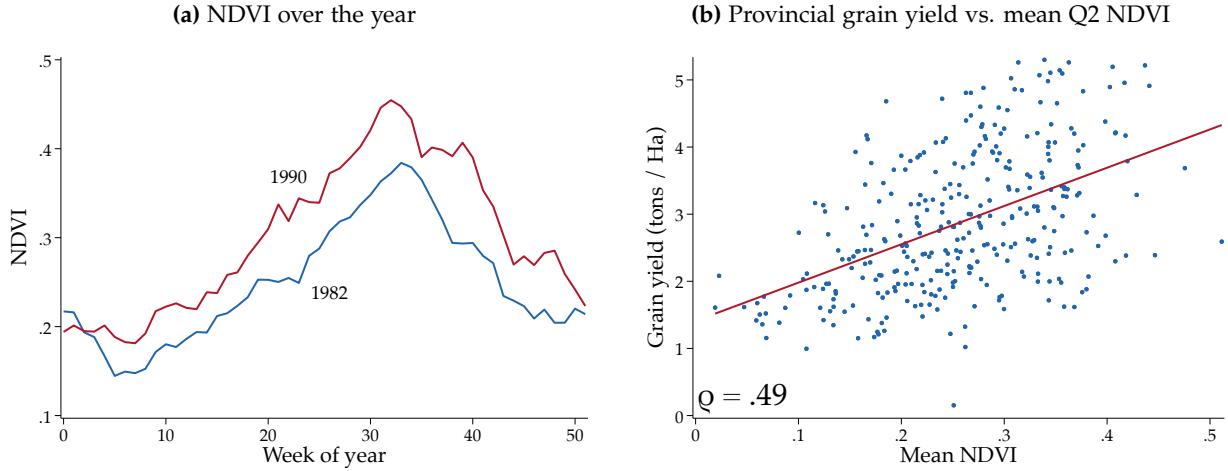


Figure 3a shows the evolution of NDVI over 1982 (blue) and 1990 (red) for an unweighted average of all Chinese provinces, using the NOAA's AVHRR satellite data. Weekly observations are smoothed using a 4-week moving average. Figure 3b shows the relationship between provincial grain yield and mean quarter 2 NDVI, with the correlation coefficient $\rho = 0.49$. The line of best fit is in red. Q2 NDVI is calculated from Landsat MSS measurements, with the mean computed over the whole of each province. Provincial grain yield is from the State Statistical Bureau (SSB).

of NDVI from the AVHRR over the course of 1982 and 1990 in China. As one might expect, NDVI shows a strong seasonal pattern, rising steadily in the spring and peaking in the fall. This peak corresponds to right before the main harvest, when there is the maximum amount of biomass on the ground. The main staples in China, wheat and rice, are mostly harvested in Q2 and Q3: winter wheat is harvested from May to June, spring wheat is harvested from August to September, while single-cropped rice is harvested from August through October (USDA 2023). We can exploit the temporal richness of the AVHRR data and include multiple observations throughout the year in our models, allowing us to predict the yields of these major grains, while ignoring other crops that peak at different times. We also note that the overall level of NDVI is significantly higher in 1990 than in 1982, suggesting that agricultural productivity did indeed increase over that time period.

Second, Figure 3b plots provinces' Q2 mean NDVIs against their official grain yields for 1976-86. Even in a simple univariate model, mean NDVI clearly predicts grain yields, with a correlation coefficient $\rho = 0.49$. As a quick rule of thumb, a 0.1 increase in Q2 NDVI predicts an additional 0.57 tons/Ha in grain yield, as shown by the red line of best fit. Of course, there remains unexplained variation in yields: other variables like rainfall and temperature clearly

matter for prediction, and, conditional on these factors, the relationship between yield and NDVI is unlikely to always be simply linear. But NDVI clearly has strong predictive content, which we can exploit in more sophisticated models.

3.3 Yield Prediction

Random Forests We can now turn to a more flexible machine learning models of yield prediction, random forest regression, which has been deployed successfully in a number of other studies to predict agricultural yields using remotely sensed data (Jeong et al. 2016; Cao et al. 2020; Han et al. 2020; Marques Ramos et al. 2020).

Random forests work by combining the predictions of a large number, or forest, of regression trees. These individual regression trees predict an outcome by progressively splitting the data into smaller subsamples according to the values of their covariates (or “features”), then assigning the outcome mean for that subsample. For instance, if the outcome is “hours of leisure” and a feature is “age”, a reasonable split might be “ > 65 years old”. Each individual tree is estimated on a different bootstrapped sample of the original data, using only a random subset of the available features to determine each split, which is chosen to minimize prediction error. The predictions of these trees are then averaged to form the random forest’s overall prediction. We use the “honest trees” variant of Wager and Athey (2018), where disjoint subsamples are used to fit the trees and estimate the prediction. This approach produces predictions that are asymptotically Gaussian and unbiased.

Random forests have two properties that make them well-suited to our agricultural prediction setting. First, the individual regression trees are effectively non-parametric matching estimators, which make them good at capturing non-linear relationships between variables like temperature, rainfall, and NDVI. Second, because of their “ensemble approach” of combining a large number of weaker individual predictors, random forests are less prone to overfitting than other methods, and can achieve strong predictive power out-of-sample with minimal tuning of their hyperparameters (Athey and Imbens 2019). We do not use Chinese data in training to avoid re-introducing bias into our model, making these out-of-sample properties particularly important.

Training Sample We train our model on a pooled sample of agricultural data from several countries: Japan, Korea, India, and the United States.⁵ We chose these countries to mimic the climatic growing and cultural conditions of China's two main grain crops, wheat and rice. To make our main outcome variable, we compute grain yields from the sum of wheat and rice output in tons, divided by the sum of wheat and rice cultivated area.

Wheat is primarily grown in northern China. The overwhelming majority of wheat in China is winter wheat, which is sown at the end of September and harvested in early or mid-June the next year (USDA 2023). During the early reform period, Myers (1978) estimates that 87% of total wheat sown area is winter wheat, while as of 2022 the USDA estimates it is 95% of China's total wheat output. To predict wheat in China, we include Indian prefecture-level yields from 1981-2005 ($N = 7511$) and US county-level observations from 1981-1993 ($N = 932$).⁶ India has a similar winter wheat (or *rabi* wheat) crop to China, albeit with an earlier March-May harvest, while the US Midwest, besides being a major winter wheat producer, has a remarkably similar climate to the North China Plain.

Rice is the main crop in southern China. Single-cropped rice, which is more common farther north, is planted in April through June and harvested in August through October. For double-cropped rice, which is most common farther south, the first, early crop is planted in March through May and harvested in July, while the second, late crop is planted in July through August and harvested in October through November. To predict rice, we include Japanese prefecture-level yields from 1981-2013 ($N = 994$), South Korean county-level yields from 1981-2013 ($N = 7557$), and Indian prefecture-level yields from 1981-2005 ($N = 7511$). All three countries have a similar rice seasons to China, centered around the summer monsoon, with some variation based on latitude. Single-cropping is predominant in Japan and Korea, where labor is relatively expensive, but double-cropping occurs in the south. We are working on adding Taiwan to the sample, which is climatically very similar to Southern China, and double-cropping is prevalent.

Features As features, we include the mean NDVI, precipitation, temperature for each month over the previous year. To flexibly capture the shape of NDVI over the seasons, we also include

⁵We are also working on adding Taiwan to the sample.

⁶We thank Björn Brey and Matthias Hertweck for generously sharing their data and shapefiles, which match the ICRISAT yield data with a harmonized set of Indian districts.

Figure 4: Predicting grain yields with a random forest, trained on pooled foreign sample

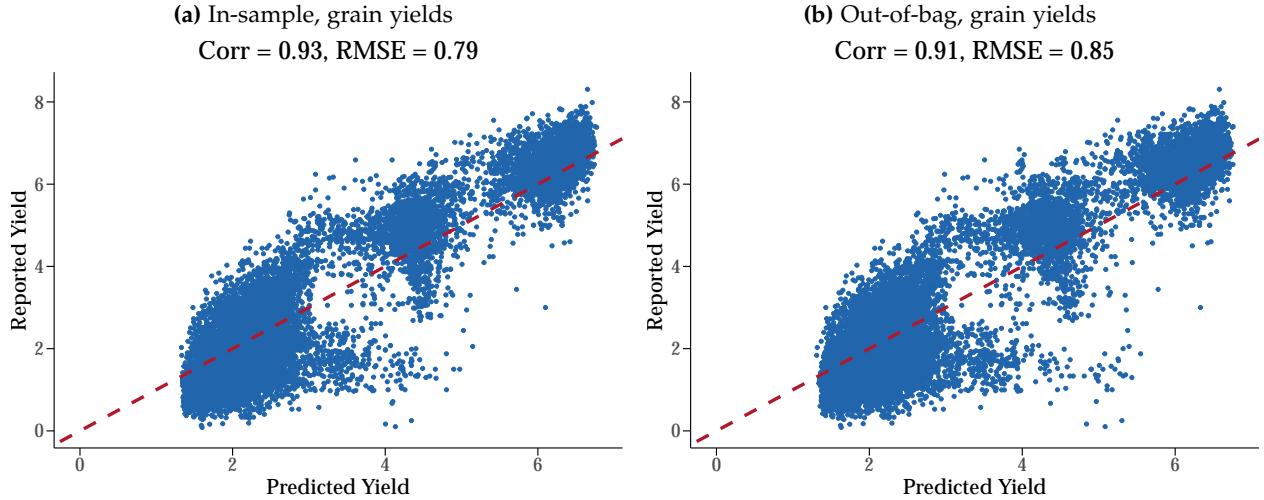


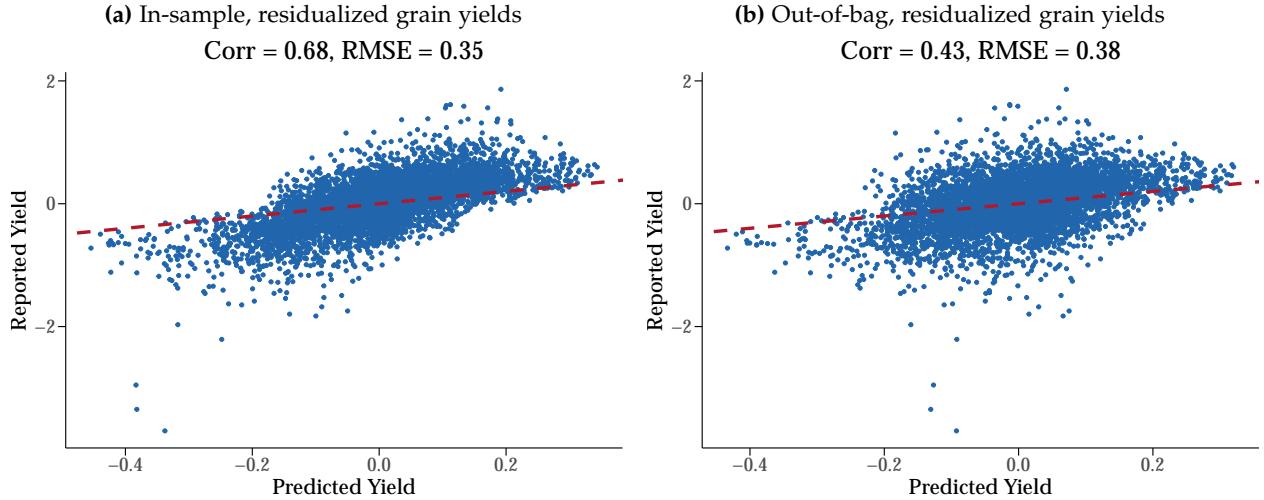
Figure 4a plots the predicted grain yield against the reported grain yield for the pooled foreign sample, using a random forest model with monthly NDVI, precipitation, and average temperature as well as FAO-GAEZ potential yields as predictors. Figure 4b plots the out-of-sample predictions of provincial grain yield from 1981-1990, using the same random forest model estimated with foreign data. In both subfigures, the red 45 degree line shows the benchmark of perfect correspondence between predicted and actual yield.

the mean, minimum, maximum, and standard deviation of NDVI over the course over the year. We also include agro-climatic potential yields of wheat and rice under irrigation conditions from the UN FAO's Global Agro-Ecological Zones (GAEZ) dataset, which computes the theoretical yield from a model using crop experiments.

Validation Figure 4 shows the predictive performance of the random forest model trained on the pooled data. On the left, Figure 4a shows the in-sample performance, plotting actual yields from Japan, Korea, India, and the USA against the model's predicted yields. The model's in-sample predictions are quite tight around the actual yields, with a correlation coefficient of close to 1. The predictions are also clustered largely symmetrically around the dotted 45-degree line, suggesting that there is no systematic bias in the predictions.

One natural concern is that the model is overfitting towards the training observations. A common test of overfitting in the random forest literature is the *out-of-bag* error. Recall that each tree is trained on a bootstrapped sample of the original data. We can compare left-out observations to predictions of trees not trained on that data, to get a metric similar to a cross-validation error. Moreover, when estimating the trees, we cluster the bootstrap samples by administrative units,

Figure 5: Predicting grain yield residuals with a random forest, trained on pooled foreign sample



[Figure 4a](#) plots the predicted residualized grain yield against the reported grain yield for the pooled foreign sample, using a random forest model with monthly NDVI, precipitation, and average temperature as well as FAO-GAEZ potential yields as predictors. [Figure 4b](#) plots the out-of-sample predictions of provincial grain yield from 1981-1990, using the same random forest model estimated with foreign data. In both subfigures, the red 45 degree line shows the benchmark of perfect correspondence between predicted and actual yield.

drawing bootstrap samples at the unit (not observation) level, simulating the process of observing new provinces or prefectures out-of-sample. On the right, [Figure 4b](#) shows the out-of-bag performance. The correlation coefficient between actual and predicted yields remains quite high, around 1, while the root mean squared errors (RMSE) expand slightly.

Another reasonable concern is that random forests can predict cross-sectional variation in yields, but cannot capture the critical intertemporal variation we need to estimate the causal effects of the HRS. To test this, we create regress yields on unit-level fixed effects and keep the residuals. We then train the random forest to predict these residualized yields, purged of any cross-sectional variation. [Figure 5a](#) shows the in-sample predictive performance, while [Figure 5b](#) shows the out-of-bag performance. Overall predictive performance weakens, with a correlation coefficient of 0.68 in-sample and 0.43 out-of-bag, but the random forest is still clearly able to predict intertemporal variation in yields.

A final, critical test of random forests is if we can detect the effect of real-world events on its predicted yields in a regression setting. (If not, finding a statistical null effect for the HRS might simply be because the random forest is predicting a vector of noise.) A natural candidate is extreme weather events—in particular, a negative shock to temperature during the winter wheat

Table 1: Effect of Weather Shocks on Model-Predicted Yields

	Full RF	RF (no weather)
Weather shock	−0.057* (0.019)	−0.256 (0.153)
R2	0.933	0.606
Num.Obs.	2 766 247	2 766 247

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

harvest months that is 1SD below its historical mean. (Winter wheat is particularly vulnerable to cold right before it is harvested.) We can test this with our cell-level predicted Chinese data. For cell i in year t , we estimate

$$\ln y_{i,t} = \alpha + \beta Shock_{i,t} + \delta_i + \varepsilon_{i,t}$$

where $Shock_{i,t} = 1$ if the weather shock occurs in year t and δ_i is a unit-level fixed effect. We cluster standard errors at the province-by year level. Table 1 plots these estimates using two different outcome variables: predicted yields from a random forest trained on our pooled sample and a full set of features, and predicted yields from a random forest trained on our pooled sample but only the NDVI features (i.e., excluding all weather variables) Both show a clear yield response to weather shocks in China, indicating that our satellite-based measures are able to detect the effects of time-varying shocks to agricultural productivity in China.

4 Empirical Strategy

4.1 Reporting Bias in Official Yields

To highlight the usefulness of satellite data in measuring the true level of Chinese agricultural activity, we can begin with a relatively simple statistical exercise: a replication of the influential core results of Lin (1992). We replicate the paper’s main regression, estimating the following

log-transformed Cobb-Douglas production function:

$$\begin{aligned}\ln(Y_{it}) = & \alpha_1 + \alpha_2 \ln(\text{Land}_{it}) + \alpha_3 \ln(\text{Labor}_{it}) + \alpha_4 \ln(\text{Capital}_{it}) + \alpha_5 \ln(\text{Fertilizer}_{it}) + \alpha_6 \text{HRS}_{it} \\ & + \alpha_7 \text{MP}_{t-1} + \alpha_8 \text{GP}_t + \alpha_9 \text{NGCA}_{it} + \alpha_{10} \text{MCI}_{it} + \alpha_{11} T_t + \sum_{j=12}^{39} \alpha_j D_j + \varepsilon_{it}\end{aligned}\quad (2)$$

where for province i in year t , Y_{it} is agricultural output (value of farm output in Lin (1992), grain yield or mean NDVI in our study), $\text{HRS}_{i,t}$ is the share of work teams in a province that have adopted household farming, $\text{MP}_{i,t-1}$ is the lagged index of market prices divided by input prices, GP_t is the index of above-quota prices divided by manufactured input prices, NGCA_{it} is the share of total sown area dedicated to nongrain crops, MCI_t is a multiple cropping index, T_t is a linear time trend, and D_j is a province-level dummy variable.

Table 2 shows the estimates of Equation (2) for a panel of provinces from 1974-87.⁷ Column 1 shows a replication of the original result from Lin (1992), using the log-value of farm output as the outcome. The estimated coefficients are identical to the original paper. In particular, the coefficient on HRS (0.220) is positive and highly significant—multiplying it by 0.99, the increase in share of HRS teams nationwide from 1978-84, and dividing it by total growth of 42 percentage points yields the paper’s headline result that the HRS accounted for half of output growth. The coefficient on HRS remains highly significant even while including year effects instead of a time trend to make the specification a more traditional two way fixed effects regression (column 2),⁸ or by switching the outcome variable to grain yields instead of the log-value of farm of output, which removes the influence of prices from the outcome (column 3). However, when we use mean NDVI in Q2 as the outcome instead of an official government statistic (column 4), we find that the coefficient on HRS is in fact negative, and no longer statistically different from 0.⁹ It is unlikely that this is driven by measurement error or lack of predictive power in NDVI—recall from Figure 3b that the same measure of provincial NDVI is highly predictive of grain yield on its own.

⁷Lacking data on the HRS for 1980, Lin (1992) drops the observations for that year, leaving a total of $N = 476$.

⁸In Figure A18, we also run the event study using the Callaway and Sant’Anna (2021) estimator to account for staggered timing. We find similarly small effects of HRS adoption on NDVI, but also find small pre-trends, highlighting concerns about selection.

⁹Due to limitations of the satellite data, some of the observations are dropped when using NDVI as an outcome, particularly for 1974, resulting in $N = 300$. The estimated results for columns 1-3 are qualitatively similar when restricting the sample to this subset. Results available upon request.

Table 2: Lin (1992) Replication

	(1) VFO	(2) Grain yield	(3) Q2 NDVI	(4) Pred yield
HRS share	0.177*** (0.0459)	0.221*** (0.0535)	-0.0502 (0.0985)	0.0415 (0.0976)
ln(Land)	0.470*** (0.110)	-0.502* (0.205)	0.0655 (0.377)	-1.954 (1.134)
ln(Labor)	0.125*** (0.0275)	0.0493 (0.0513)	-0.0548 (0.100)	0.160 (0.0815)
ln(Capital)	0.173** (0.0593)	0.0354 (0.108)	-0.0359 (0.229)	0.243 (0.251)
ln(Fertilizer)	0.149*** (0.0240)	0.165*** (0.0438)	0.206* (0.0872)	-0.269 (0.143)
Nongrain crops percentage	0.762** (0.235)	0.290 (0.440)	-0.143 (0.903)	0.220 (0.670)
Multiple cropping index	0.187* (0.0934)	-0.115 (0.175)	-0.158 (0.381)	-0.643 (0.359)
Year		-0.00211 (0.0136)	0.0130 (0.0284)	-0.000409 (0.0447)
Market / input price (t-1)		-0.00491 (0.127)	-0.147 (0.256)	-0.164 (0.158)
Govt / input price t		-0.0553 (0.0430)	-0.0355 (0.0853)	0.0821 (0.173)
Constant	3.284*** (0.596)	7.649 (27.02)	-27.13 (56.40)	14.04 (89.21)
Observations	364	364	300	168
R ²	0.865	0.536	0.104	0.187

Standard errors in parentheses

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

Rather, our null finding suggests that the difference in our results from those of Lin (1992) are driven by a *positive* correlation between the progress of the HRS and reporting errors in the official grain yield. Consider the following toy example. We can model the official yield for province i in year t as:

$$\tilde{y}_{i,t} = \lambda y_{i,t} + \varepsilon_{i,t}$$

where $y_{i,t}$ is the true yield, $\tilde{y}_{i,t}$ is the reported yield, λ is a bias factor, and $\varepsilon_{i,t}$ is a mean-zero error term. Suppose the true relationship between yield and the HRS is

$$y_{i,t} = \beta_0 + \beta_1 HRS_{i,t} + \eta_{i,t}$$

but since we only have official yield data what we actually estimate is

$$\tilde{y}_{i,t} = \tilde{\beta}_0 + \tilde{\beta}_1 HRS_{i,t} + \tilde{\eta}_{i,t}. \quad (3)$$

Then the estimated coefficient will be the sum of the true coefficient β and a classic omitted variable bias term:

$$\tilde{\beta}_1 = \lambda \beta_1 + \frac{\text{Cov}(HRS_{i,t}, \varepsilon_{i,t})}{\mathbb{V}(HRS_{i,t})}.$$

If either $\text{Cov}(HRS_{i,t}, \varepsilon_{i,t}) \neq 0$ or $\lambda \neq 1$, then $\beta_1 \neq \tilde{\beta}_1$ and our estimate of β_1 will be biased. We can check if the first of these conditions holds by bringing in the satellite data, which can be modeled as

$$\bar{y}_{i,t} = \kappa y_{i,t} + \theta_{i,t}$$

where κ captures the linear relation between satellite-based NDVI and yield, and $\theta_{i,t}$ is an error term. Unlike above, we can safely assume that $\text{Cov}(\theta_{i,t}, HRS_{i,t}) = 0$ for satellite imagery. Plugging in NDVI as the outcome in [Equation \(3\)](#), we find that the satellite-based regression coefficient $\bar{\beta}_1 = \kappa \beta$. The satellite-based regression results in column 4 suggest that $\bar{\beta}_1 \approx 0$. If we rule out the

degenerate case where $\kappa = 0$, then it follows that the true relationship between NDVI and HRS is $\beta_1 = 0$. Since we found a positive $\tilde{\beta}_1$, it must then be that $\text{Cov}(HRS_{i,t}, \varepsilon_{i,t}) > 0$. This positive correlation is consistent with a story where provinces sought to present good economic news about the rollout of reform and exaggerated their grain yields as the HRS spread. NDVI from satellites allows us to overcome these reporting biases and recover the true effect of the HRS.

4.2 Causally Identifying the Effect of the HRS

Besides showing the potential pitfalls of correlated measurement error, the previous subsection introduced the startling possibility that the true effect of the HRS on yields was zero. Of course, this estimate was imprecise—a 95% confidence interval includes fairly large contributions to overall growth—and a panel regression approach may not reveal a true causal effect, the reasons for which we will outline below. We can now proceed to more modern causally identification approaches, using the finer variation made available by satellite imagery to precisely estimate the effects of the HRS.

Our identification strategy exploits the staggered timing of HRS adoption across China’s provinces. Our main treatment variable is from the provincial-level data from Lin (1992). We code a province as having been “treated” with the HRS if the share of work teams in the province who have adopted HRS exceeds 50%.¹⁰ Figure 1 maps when provinces crossed this threshold in adopting the HRS. 11 provinces crossed the 50% threshold in 1981; 9 more did in 1982; 7 did in 1983; and the last province did in 1984.

The major empirical concern with identifying the causal effect of the HRS is selection: as noted in Section 2, reform was originally targeted at areas that were poor, remote, or at risk of famine—recall that the first decollectivization experiments began in Anhui in 1978, in response to drought. In other words, negative selection was baked into the very design of the reform: provinces that adopted earlier were more likely to be worse-off.

We address this concern by using a *staggered difference-in-discontinuities* design, which extends the differences-in-discontinuities design first formally described by Butts (2021) to a setting with staggered timing. Intuitively, differences-in-discontinuities itself extends the logic of a border regression discontinuity (RD) design to a multi-period setting. In a conventional border RD,

¹⁰Specifications with other cutoffs are available in the appendix. The results are similar.

assuming that all relevant factors other than treatment vary smoothly across the border, any discontinuities in outcomes observed at the border can be causally attributed to the treatment. Weather events (like Anhui’s drought in 1978), climate, and other natural characteristics relevant for agriculture are likely to be continuous in space, and satisfy this assumption across province borders. However, provincial boundaries also reflect economic and political differences that are unlikely to satisfy the classic RD assumption for causal identification.

A differences-in-discontinuities design allows us to relax this assumption. By observing the same geographic points repeatedly over time, we can use fixed effects to control for any time-invariant discontinuities along the border. The identifying assumption then becomes that cells just on either side of the border follow parallel trends in their potential outcomes, akin to the parallel trends assumption in a differences-in-differences design. In other words, we can control for any fixed differences (natural, political, or otherwise) at the border, alleviating concerns about selection, and we can identify the effects of the HRS from changes in the border discontinuity over time—assuming that no other policies change at the same time.

Another concern, highlighted in the recent differences-in-differences literature, is that applying traditional two-way fixed effect estimators in settings with staggered event timing can result in negative own-treatment weights. This problem arises when previously treated units are used as a control group for just-treated units, since previously treated units may be experiencing lagged or heterogeneous treatment effects—a problem that is likely to apply here, as more work teams in the province adopt the HRS and the effect on productivity phases in. To address this, we follow the Dube et al. (2023) local projections approach, where treatment effects are estimated with separate regressions for each horizon h , and the control group is restricted to “clean” never-treated units. Formally, for cell i in border group b in year t , at horizon h , we estimate

$$y_{i,b,t+h} - y_{i,b,t-1} = \beta_h \Delta D_{i,b,t} + \gamma_{b,t+h} (R_i \times B_{b,t}) + \delta_{b,t+h} (R_i \times B_{b,t} \times \Delta D_{i,b,t}) + B_{b,t} + e_{i,b,t}^h \quad (4)$$

where $D_i = 1$ if over 50% of the work teams in that cell’s province have adopted the HRS, $B_b = 1$ if cell is closest to border $b \in \{\text{Anhui-Jiangsu, Guangdong-Guangxi...}\}$, and R_i is distance to the border. We include cell-level fixed effects (implicitly, by differencing out $y_{i,t-1}$), and border-by-year fixed effects $B_{b,t}$. Following Gelman and Imbens (2019), to avoid overfitting, we use local

linear functions of distance for the RD, allowing for different slopes on the running variable across each border for each year. We cluster standard errors by border b . To prevent negative weighting, for each time period t we restrict the sample to just-treated units and “clean control” (or never-treated) units:

$$\begin{cases} \text{newly treated: } & \Delta D_{i,t} = 1 \\ \text{or clean control: } & D_{i,t+h} = 0 \end{cases}$$

We estimate this equation over our cell-level observations over all provinces in China from 1974Q1 to 1986Q4. Our baseline specification includes observations that are up to 50 kilometers away from the border. Our results are robust to different distance thresholds and orders of polynomials; see the appendix for more details.

Our chief 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 will report β_h for each time horizon h , as an event study.

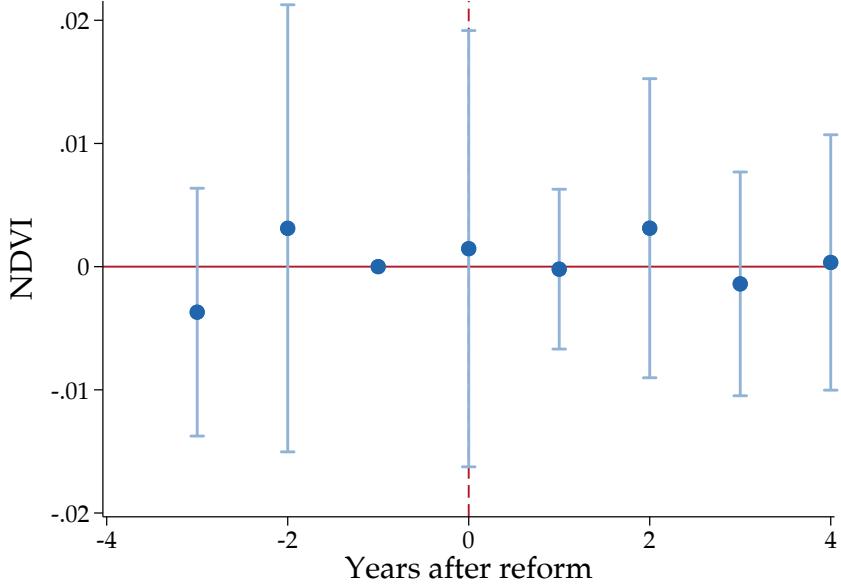
5 Empirical Results

5.1 Differences-in-Discontinuities

[Figure 6](#) plots our estimates of the border discontinuity at each year-horizon h after decollectivization, β_h from [Equation 4](#). Panel A shows the results with Q2 NDVI as an outcome variable, and panel B shows Q3 NDVI. We do not find evidence that decollectivization increased NDVI at the border. Effect sizes are near 0 for all time horizons in both Q2 and Q3, with the tight standard error bounds able to rule out positive effects as small as 0.1 NDVI units. Even 5 years after the onset of reform, the effects of decollectivization are barely statistically different from 0 at the 95% confidence level, with a point estimate of less than 0.05 NDVI units. These results are robust to varying the distance threshold ([Figure A14](#)), and using a more traditional two-way fixed effects estimator ([Figure A13](#)).

The obvious next question is how these NDVI differences translate into agricultural yields.

Figure 6: Effect of HRS adoption on Peak NDVI, national sample

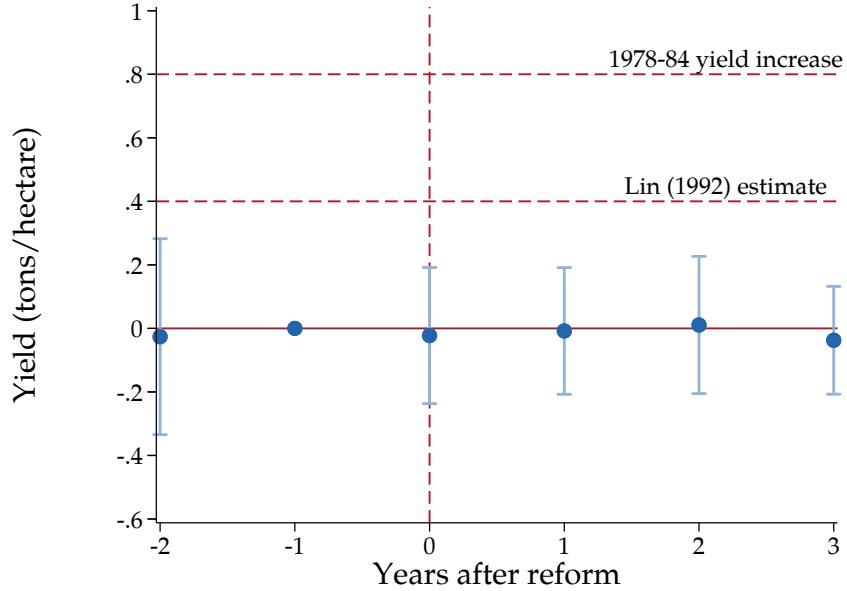


This figure shows the event study of peak yearly NDVI following the treatment of provincial decollectivization, estimated using [Equation 4](#). The bars show the 95% confidence intervals around the point estimates.

We can convert NDVI into yields using the random forest model estimated on the pooled Japanese and Indian data in [Section 3.3](#). For each cell, we estimate the yield using the observed NDVI, precipitation, and temperature over the previous year. We then re-estimate [Equation \(4\)](#) using the model-generated yields as an outcome variable. To take into account the fact that yields are themselves estimated data, we use a two-stage bootstrap procedure when estimating the standard errors. In the first stage, we draw a bootstrap sample of the rural observatory data, clustering by quarter and village, then estimate the yield-NDVI relationship using the random forest. In the second stage, we draw a bootstrap sample of the satellite data, which we plug into the estimated yield-NDVI relationship from the previous stage to produce model-estimated yields. We then estimate the regression discontinuity, with estimated yields as the outcome variable. We repeat this for 100 bootstrap samples, using the sample standard deviation of the regression discontinuity coefficient estimates to form the standard error.

[Figure 7](#) plots the effects of the rollout of the HRS on estimated yields. The relative pattern of point estimates remains similar to the raw NDVI estimates in [Figure 6](#), while the standard errors expand due to the additional uncertainty of estimating the relationship between NDVI

Figure 7: Effect of HRS adoption on estimated yield, national sample



This figure shows the event study of yield following the treatment of provincial decollectivization, estimated using [Equation 4](#). The bars show the 95% confidence intervals around the point estimates. The horizontal dashed line shows the national yield increase from 1978-84 recorded in the official data.

and yield. The point estimates on yield remain mostly indistinguishable from 0. Nonetheless, we retain enough statistical precision that we can reject small effects of the HRS on grain yields—we can rule out an effect of 0.2 tons per hectare, even five years after the onset of reform.

Our estimates so far have relied on a binary coding of HRS treatment. How should we think about the *continuous* effect of an increase in HRS share—did provinces with higher HRS shares experience faster yield growth? To gauge the effects of a continuous change in the provincial HRS share (the dose-response), we can modify [Equation \(4\)](#) by interacting the binary treatment $\Delta D_{i,b,t}$ with $HRS_{i,b,t}$, the share of provincial work teams who have adopted HRS,

$$\begin{aligned}
 y_{i,b,t+h} - y_{i,b,t-1} = & \alpha_h \Delta D_{i,b,t} + \beta_h \Delta D_{i,b,t} \times HRS_{i,b,t} \\
 & + \gamma_{b,t+h} (R_i \times B_{b,t}) + \delta_{b,t+h} (R_i \times B_{b,t} \times \Delta D_{i,b,t}) \\
 & + B_{b,t} + e_{i,b,t}^h.
 \end{aligned} \tag{5}$$

[Table 3](#) summarizes the results of estimating [Equation \(5\)](#). We do not find evidence that the

Table 3: Continuous Effects of the HRS on Yield

	Horizon (h years after reform)			
	0	1	2	3
<i>Linear</i>				
α_h	1.16 (0.78)	1.05 (0.67)	1.15 (0.94)	-0.00 (0.94)
β_h	-1.33 (0.93)	-1.29 (0.78)	-1.45 (1.11)	-0.01 (1.07)
<i>Binned</i>				
$70\% \leq HRS < 80\%$	0.23 (0.23)	0.20 (0.16)	0.13 (0.29)	0.15 (0.29)
$80\% \leq HRS < 90\%$	0.09 (0.27)	-0.01 (0.20)	0.12 (0.33)	0.04 (0.26)
$90\% \leq HRS < 100\%$	-0.02 (0.22)	0.04 (0.14)	0.03 (0.27)	0.01 (0.22)

This table shows the regression estimates for [Equation \(5\)](#) for different horizons h . The top row shows the results for a linear interaction with the share of provincial work teams who have adopted the HRS. The bottom row shows an interaction with indicators for 10% bins of households who have adopted the HRS. Standard errors are calculated using the two-stage bootstrap procedure.

effect size is increasing in HRS. In the linear interaction specification, the constant effect of α_h is positive, but the point estimates decrease as the share of HRS increases ($\beta_h < 0$). We can also estimate a non-parametric version of the same regression, where instead of a linear interaction with HRS we include indicators for binned 10 percentage point values of HRS share. Similar to the linear specification with $\beta_h < 0$, we find that the estimates get smaller for bins with larger HRS shares. The bootstrapped standard errors are large enough that we cannot rule out that any of these estimates are different from 0. Thus, we do not find evidence that the HRS had a positive heterogeneous effect on yield, where provinces with larger HRS shares experienced higher yield growth.

How should we judge the magnitude of these effects? The official increase in national grain yields from 1978 to 1984, 0.8 tons, is plotted as the horizontal dashed line. McMillan et al. ([1989](#)) finds that the HRS was responsible for 78% of the 1978-84 increase in agricultural TFP, while Lin ([1992](#)) concludes that the HRS's contribution was closer to 90%. By contrast, we find that our 95% confidence intervals overlap with 0, and rule out a 25% contribution of the HRS to overall grain yields. While TFP is a different quantity from yield (land productivity), these results are quantitatively far too small to be consistent with the view that the HRS had a large and

transformative effect on agricultural productivity.

5.2 Threats to Identification

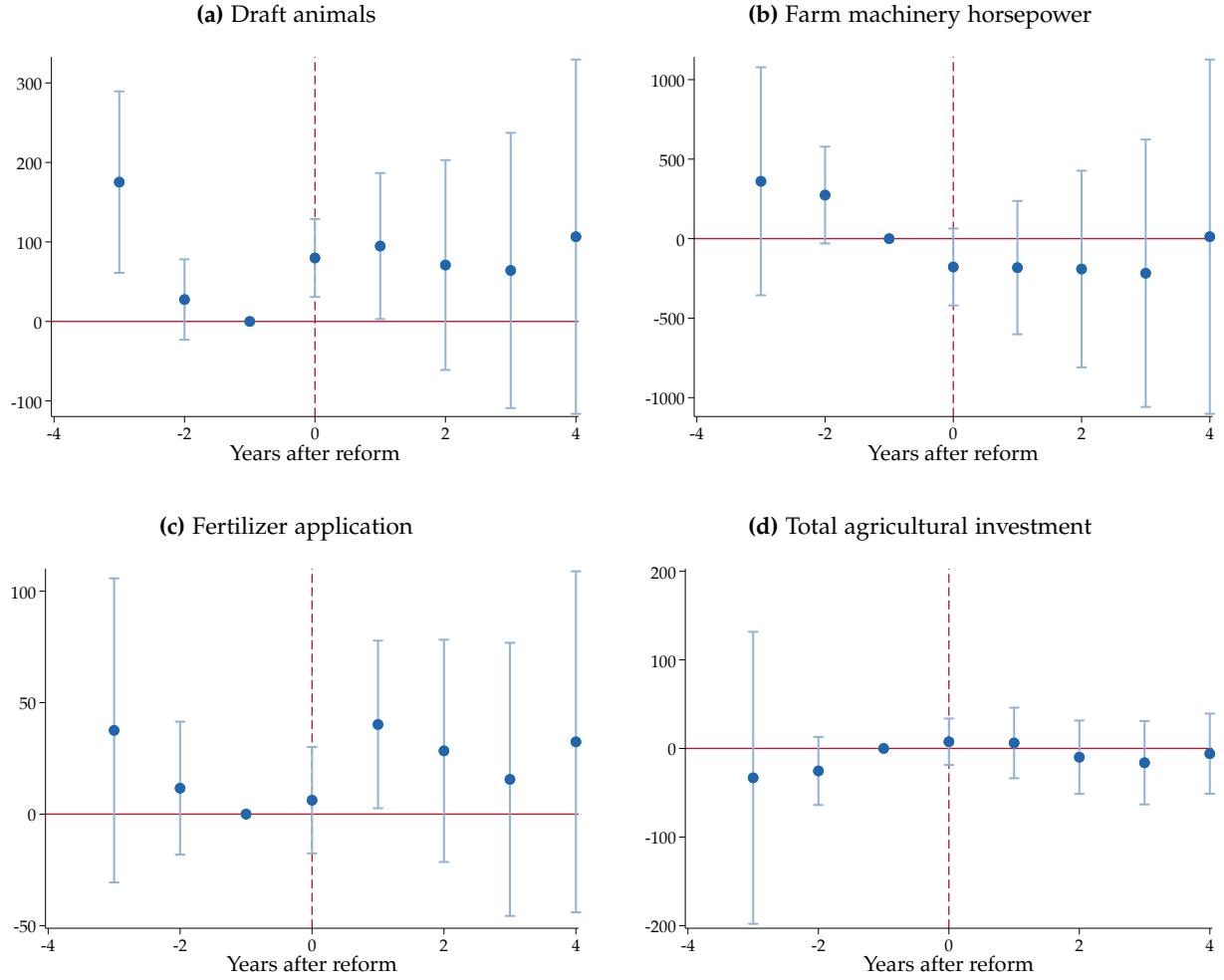
Given these surprising null results, and the literature's strong consensus in support of a strong effect of the HRS, it is natural to ask what factors other than the HRS might be driving the patterns we observe. The main threats to causal identification in our staggered-difference-in-discontinuities framework are confounding treatment, spillovers, and core vs. border effects.

Confounding Treatment A critical assumption for our identification strategy is that no other policy changes affected agriculture around decollectivization that could cause differences along provincial borders. Other than the HRS, the major agricultural policy change during this study period was the 1979 rise in procurement prices. However, these price rises did not differ substantially by province.

We do not observe any other national policy changes affecting agriculture during this time period. To check this statistically, [Figure 8](#) plots event studies of key agricultural inputs at the provincial level—number of draft animals, total horsepower of farm machinery, total fertilizer application, and total agricultural investment—from the State Statistical Bureau, around when 50% of work teams adopted the HRS. We do not observe any changes in these variables around the time of treatment. Moreover, because our main empirical finding is a null effect, if one's prior is that the HRS was a boon for productivity, the main threat to identification would be the unlikely coincidence of *negative* relative productivity shocks in treated areas, happening simultaneously with the HRS and cancelling out its positive effects. While noisy, the event study plots show that it is unlikely that this kind of negative compound treatment is driving our results.

Spillovers Our causal estimates are derived from discontinuities measured at province borders. A common concern in this kind of geographic design is that the treatment of reform “spilled over” into neighboring provinces. If farmers were able to migrate across provinces and bring decollectivizing reforms with them, or if land reform spread through word-of-mouth across the border, then yields may have also risen in “control” provinces, biasing the estimated effect of land reform towards 0.

Figure 8: Agricultural Inputs around Decollectivization



This figure shows the event study of province-level agricultural inputs—draft animals, the total horsepower of farm machinery, total fertilizer application, and total agricultural investment—around the treatment event of provincial decollectivization. Input data from the State Statistical bureau (SSB).

Two institutional features in this setting make spillovers unlikely. First, China's *hukou* (household registration) system effectively banned migration during this time period. Each person was categorized as "agricultural" or "industrial" based on their place of registration (typically, their birthplace), and had to seek an official transfer to move (Cheng and Selden 1994). Official transfers were rare. Even in the 1980s, the share of the population living in a location different from their *de jure* residence was only 0.6% (Chan 2009). The *hukou* system's migration restrictions were only relaxed in 1984, when migrants seeking work in small towns were allowed to move.

Second, a strong system of ideological control sought to prevent the spread of the HRS. Until

1982, the legal status of the HRS was unclear, and more conservative provinces tried to contain its spread across their boundaries—in one notable example, slogans were even broadcast across the Anhui-Jiangsu border “denouncing Anhui’s revival of capitalism” (Teiwes and Sun (2016), p. 142). Given the high capacity of the Communist Chinese state, the support of cadres and higher-level officials were likely necessary for the HRS to systematically spread; indeed, research on the spread of the HRS has emphasized the role of provincial governments in controlling when farmers could switch over to the HRS (Bai and Kung 2014).

Naturally, we do not have any data to observe any illicit spread of the HRS, so we cannot directly rule out the possibility of spillovers. However, we can test for the *effects* of spillovers under the assumption that, if they exist, they would decay in distance away from a treated area—like if the reform spread across province borders through word-of-mouth. Using a similar functional form as before, we can estimate

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

where the focus is now on γ_{t+h} , the slope of the outcome in a not-yet-treated province neighboring a treated province as one approaches the border. By not interacting distance with border-by-year effects $B_{b,t}$ as in Equation (4), we pool slopes across all untreated provinces and all years into a single coefficient. If $\gamma_{t+h} < 0$, then yields increase on average as distance to a treated province decreases, suggesting that treatment is spilling over into its untreated neighbor. We can also observe the evolution of these spillovers over time by estimating Equation (6) for each horizon h .

Table 4 shows our estimates of γ_{t+h} from Equation (6). We cannot reject that the slope γ_{t+h} is different from 0, at all time horizons. The point estimates are negative at the 1 and 3-year horizons, though the standard errors are large and the point estimates are small. We can also observe the effect of distance nonparametrically, by substituting the linear distance term in Equation (6) with indicator variables for each 10km distance bin—i.e., if a cell’s centroid is 0-10km away from the border, 10-20km away, etc. With this more flexible approach, we still find no evidence that yields in untreated provinces are increasing as one approaches the border with a treated province. These results collectively suggest that spillovers are not driving our null results.

Table 4: Spillovers into untreated provinces, in distance to the border

	Horizon (h years after reform)			
	0	1	2	3
<u>Linear</u>				
γ_{t+h} (in 1000 km)	0.03 (0.07)	-0.08 (0.14)	0.02 (0.19)	-0.12 (0.16)
<u>Binned</u>				
$0\text{km} \leq R_i < 10\text{km}$	-0.00 (0.00)	0.00 (0.04)	0.00 (0.06)	0.01 (0.02)
$10\text{km} \leq R_i < 20\text{km}$	-0.00 (0.00)	0.00 (0.04)	0.00 (0.06)	0.01 (0.01)
$20\text{km} \leq R_i < 30\text{km}$	-0.00 (0.00)	0.00 (0.04)	0.00 (0.06)	0.00 (0.01)
$30\text{km} \leq R_i < 40\text{km}$	-0.00 (0.00)	0.00 (0.04)	0.00 (0.06)	0.00 (0.01)

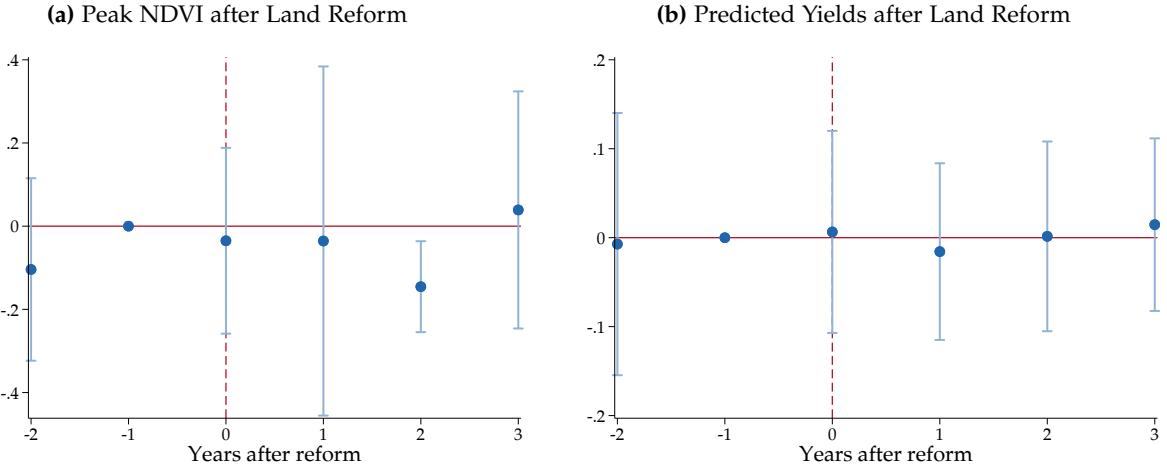
This table shows the spillover estimates of [Equation \(6\)](#) on estimated yields in untreated provinces for different yearly horizons h . The first row shows the linear specification in absolute distance to the border (in 1000s of kilometers, to make the effect visible). The second set of rows shows the binned specification, where distance is grouped into 10 kilometer bins from the border with a treated province. 40-50km is the omitted category. Standard errors are computed using the two-stage bootstrap procedure described in [Section 5.1](#).

5.3 Alternative Strategy: County Rollout

We can test the robustness of our null finding by using an alternative identification strategy. [Almond et al. \(2019\)](#) collect county-level data of the rollout of the HRS from county gazetteers, creating a separate source of treatment variation from the central government records used in [Lin \(1992\)](#). In their definition, a county becomes treated when “collectively owned land was first contracted to individual households in a few villages”. [Figure A12](#) maps out when the counties in the [Almond et al. \(2019\)](#) dataset become treated. We combine this county-level rollout treatment with mean NDVIs and predicted yields of counties as our outcome variable.

We view this as county rollout design as complementary to the differences-in-discontinuities design, answering two major concerns with the previous approach. First, 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. By contrast, the gazetteers used by [Almond et al. \(2019\)](#) are semi-official sources compiled locally, largely for historical purposes. While still subject to some political reporting pressures, they are viewed by historians as more likely to be critical of the central government ([Looney 2008](#)), and thus are an important check for

Figure 9: Effect of HRS on NDVI and Yields, National County Rollout



This figure shows the event study of NDVI (left panel) and yield (right) following the treatment of the onset of county decollectivization, estimated using Equation (7). The bars show 95% confidence intervals.

the Lin (1992) treatment data.

Second, a central concern with the differences-in-discontinuities design is that the effects at provincial borders are merely local. Perhaps the effects of land reform on yields were highest at the centers of provinces, while the periphery were unaffected. Or, as discussed in Section 5.2, maybe news of the HRS was more likely to spread over province border regions, biasing the estimated effects towards 0. Measuring county-level outcomes can help ease concerns that the borders of treated areas differ systematically from their “core” regions. Figure A12 shows that counties in the Almond et al. (2019) data are scattered throughout the border and core regions of provinces and, moreover, by looking at averages over a whole county as an outcome, we incorporate information about both the border and core of the counties.

For county i at time t at year horizon h after treatment, we estimate

$$y_{i,t+h} - y_{i,t-1} = \beta \Delta D_{i,t} + \delta_t^h + \varepsilon_{i,t}^h \quad (7)$$

where y_i is either county-level mean NDVI or predicted grain yield, $D_{i,t} = 1$ in the year when “collectively owned land was first contracted to individual households in a few villages”, and δ_t^h is a year effect. We cluster standard errors at the county level, allowing for intertemporal correlation. As before, to prevent negative weighting, for each time period t we restrict the

sample to just-treated units ($\Delta D_{i,t} = 1$) and “clean control” (or never-treated) units ($D_{i,t+h} = 0$).

[Figure 9](#) plots these county-level results, with yearly peak NDVI (left) and predicted yields (right) as outcomes. Before treatment, we do not observe evidence of statistically significant pre-trends in yields or NDVI, suggesting that we can assume parallel trends between treated and control counties. Turning to the post-treatment period, we find similar results to the differences-in-discontinuities design: we cannot reject that the effect of land reform on yields is different from 0 at the 95% confidence level for all time periods, including looking up to 3 years after the onset of reform. We can also reject, at the 95% confidence level, that effects are as low as 0.1 NDVI units for most periods (the exception is $h = 2$ years after reform). Differences between core and border regions are thus not likely driving our main results. In other words, the effects of county-level adoption of HRS appear to be similar to those from province-level discontinuities—small and statistically indistinguishable from 0.

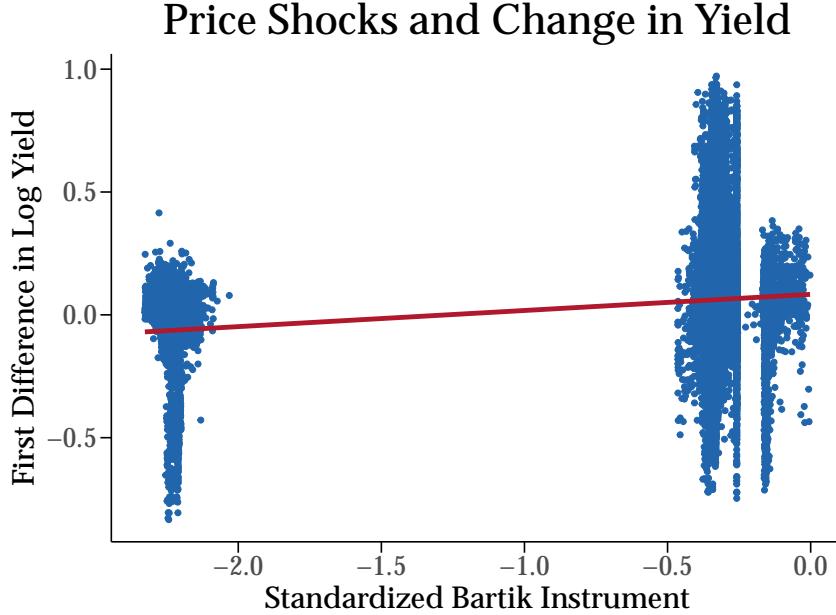
Why does this finding differ so much from Almond et al. (2019), which used the same treatment variation but found that gazetteer grain output per capita increased by 3.8 percent per year in a county? In [Figure A17](#) we redo the Almond et al. (2019) event study using the Dube et al. (2023) estimator, but with the original grain per capita data from that paper. We find no evidence for an effect of HRS adoption on county grain output per capita, suggesting that the discrepancy may be driven by negative weights from staggered timing.

5.4 Another Candidate: Procurement Prices

If the HRS did not cause yields to grow, what did? One obvious candidate 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 1988). Since these price rises did not differ by province, their effects are not captured in the border discontinuity design. However, the price rises differed between crops, and areas varied in their exposure to the price changes by their suitability for various crops, plausibly generating variation across China.

To test this hypothesis more formally, for each cell i in China, we can construct a “price shock” measure that interacts the price rises for crop c in year t with their agro-climatic potential yields

Figure 10: Correlation between Price Shocks and Yield Growth



from the UN FAO-GAEZ dataset:

$$Z_{i,t} = \sum_c (PotYield_{i,c} \times \Delta \ln p_{c,t}) \quad (8)$$

Figure 10 plots the relationship between these price shocks and the observed yield changes between 1980 and 1979. The correlation is positive, and the coefficient on price shocks is statistically significant at $p = 0.05$. At the moment, we do not have data on crop shares, so we cannot use these price shocks to directly instrument for the actual exposures to price changes. In this sense, this remains largely a reduced-form or correlational exercise. Nonetheless this suggestive, correlational evidence is consistent with price shocks contributing to yield growth.

6 Conclusion

We find consistent evidence, across a range of different identification strategies, that the causal effect of the Household Responsibility System on yield 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. This finding is deeply at odds 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.

A natural next question, given this evidence, is: did agricultural output increase at all? In future work, we are planning on developing satellite-based models to predict grain output *in aggregate*. Our answers at this stage are conjectural, but one striking possibility is that agricultural yields did not increase at all, or at least as much as was claimed. There is ample qualitative evidence to suggest that there was widespread grain hiding in poor and remote areas, in response to onerous state procurement (Chan and Unger 1982; Oi 1991). Rising reported yields may just have been a function of farmers starting to honestly report their yields as the state's agricultural regime relaxed its grip.

Extraordinary claims require extraordinary evidence. The Chinese economic miracle, which, like clockwork, posted steady rates of 10% annual growth, is one such example. In the wake of the Covid-19 pandemic, there have been growing number of pessimistic reassessments of China's future growth prospects, arguing for a prolonged slowdown. This paper's results suggest that that skepticism should also be extended to the recent past. Perhaps what we are witnessing in the world's second-largest economy is not a sharp deceleration, but a steady return to the historical norm.

References

- Adamopoulos, Tasso, and Diego Restuccia. 2020. "Land Reform and Productivity: A Quantitative Analysis with Micro Data". *American Economic Journal: Macroeconomics* 12, no. 3 (): 1–39.
- Almond, Douglas, Hongbin Li, and Shuang Zhang. 2019. "Land Reform and Sex Selection in China". *Journal of Political Economy* 127, no. 2 (): 560–585.
- Athey, Susan, and Guido W. Imbens. 2019. "Machine Learning Methods That Economists Should Know About". *Annual Review of Economics* 11 (1): 685–725.
- Bai, Ying, and James Kai-sing Kung. 2014. "The Shaping of an Institutional Choice: Weather Shocks, the Great Leap Famine, and Agricultural Decollectivization in China". *Explorations in Economic History* 54 (): 1–26.
- Bognár, Péter, et al. 2022. "Testing the Robust Yield Estimation Method for Winter Wheat, Corn, Rapeseed, and Sunflower with Different Vegetation Indices and Meteorological Data". *Remote Sensing* 14, no. 12 (): 2860.
- 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.
- Brandt, Loren, Chant-tai Hsieh, and Xiaodong Zhu. 2008. "Growth and Structural Transformation in China". In *China's Great Economic Transformation*, 1st ed., ed. by Loren Brandt and Thomas G. Rawski, 683–728. Cambridge University Press.
- Butts, Kyle. 2021. "Geographic Difference-in-Discontinuities".
- 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 (): 200–230.
- Cao, Juan, et al. 2020. "Identifying the Contributions of Multi-Source Data for Winter Wheat Yield Prediction in China". *Remote Sensing* 12, no. 5 (): 750.
- Chan, Anita, and Jonathan Unger. 1982. "Grey and Black: The Hidden Economy of Rural China". *Pacific Affairs* 55 (3): 452–471. JSTOR: [2757120](#).
- Chan, Kam Wing. 2009. "The Chinese Hukou System at 50". *Eurasian Geography and Economics* 50, no. 2 (): 197–221.
- Cheng, Tiejun, and Mark Selden. 1994. "The Origins and Social Consequences of China's Hukou System". *The China Quarterly* 139 (): 644–668.
- Chöng, Chae-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.
- CIA. 1983. *China: Reforming Agriculture with the Responsibility System*. Tech. rep. EA 83-10241.
- Dube, Arindrajit, et al. 2023. "A Local Projections Approach to Difference-in-Differences Event Studies". NBER Working Paper No. 31184.
- Eisenman, Joshua. 2018. *Red China's Green Revolution: Technological Innovation, Institutional Change, and Economic Development under the Commune*. New York: Columbia University press.
- Foster, Andrew, and Mark Rosenzweig. 2017. *Are There Too Many Farms in the World? Labor-Market Transaction Costs, Machine Capacities and Optimal Farm Size*. Tech. rep. w23909. Cambridge, MA: National Bureau of Economic Research.

- 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 (): 447–456.
- Gibson, John. 2020. "Aggregate and Distributional Impacts of China's Household Responsibility System". *Australian Journal of Agricultural and Resource Economics* 64 (1): 14–29.
- Hamar, D., et al. 1996. "Yield Estimation for Corn and Wheat in the Hungarian Great Plain Using Landsat MSS Data". *International Journal of Remote Sensing* 17, no. 9 (): 1689–1699.
- Han, Jichong, et al. 2020. "Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China". *Remote Sensing* 12, no. 2 (): 236.
- Henderson, J. Vernon, Adam Storeygard, and David N Weil. 2012. "Measuring Economic Growth from Outer Space". *American Economic Review* 102, no. 2 (): 994–1028.
- Hodler, Roland, and Paul A. Raschky. 2014. "Regional Favoritism". *The Quarterly Journal of Economics* 129 (2): 995–1033. JSTOR: [26372565](#).
- Holz, Carsten A. 2003. "'Fast, Clear and Accurate': How Reliable Are Chinese Output and Economic Growth Statistics?" *The China Quarterly* 173 (): 122–163.
- 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.
- Jean, Neal, et al. 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty". *Science* 353, no. 6301 (): 790–794.
- Jeong, Jig Han, et al. 2016. "Random Forests for Global and Regional Crop Yield Predictions". Ed. by Jose Luis Gonzalez-Andujar. *PLOS ONE* 11, no. 6 (): e0156571.
- 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 (): 396–411.
- Kelliher, Daniel Roy. 1992. *Peasant Power in China: The Era of Rural Reform, 1979-1989*. Yale Agrarian Studies. New Haven: Yale University Press.
- Lin, Justin Yifu. 1992. "Rural Reforms and Agricultural Growth in China". *The American Economic Review* 82 (1): 34–51. JSTOR: [2117601](#).
- . 1988. "The Household Responsibility System in China's Agricultural Reform: A Theoretical and Empirical Study". *Economic Development and Cultural Change* 36, no. S3 (): S199–S224.
- Liu, Han, et al. 2020. "Annual Dynamics of Global Land Cover and Its Long-Term Changes from 1982 to 2015". *Earth System Science Data* 12, no. 2 (): 1217–1243.
- Looney, Kristen. 2008. 'Village Gazetteers, a New Source in the China Field.' SSRN SCHOLARLY PAPER 3027667. Rochester, NY: Social Science Research Network.
- Marques Ramos, Ana Paula, et al. 2020. "A Random Forest Ranking Approach to Predict Yield in Maize with Uav-Based Vegetation Spectral Indices". *Computers and Electronics in Agriculture* 178 (): 105791.
- Martínez, Luis R. 2022. "How Much Should We Trust the Dictator's GDP Growth Estimates?" *Journal of Political Economy* 130, no. 10 (): 2731–2769.
- 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. JSTOR: [1832191](#).

- Myers, Ramon H. 1978. "Wheat in China—Past, Present and Future". *The China Quarterly*, no. 74: 297–333. JSTOR: [652694](#).
- 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 (): 113–144.
- Oi, Jean C. 1991. *State and Peasant in Contemporary China: The Political Economy of Village Government*. 1. paperback print. Berkeley: Univ. of California Press.
- Rabinovitch, Simon. 2010. "China's GDP Is "Man-Made," Unreliable: Top Leader". *Reuters* ().
- Ravallion, Martin, and Shaohua Chen. 2007. "China's (Uneven) Progress against Poverty". *Journal of Development Economics* 82, no. 1 (): 1–42.
- Rawski, Thomas G. 1976. "On the Reliability of Chinese Economic Data". *The Journal of Development Studies* 12, no. 4 (): 438–441.
- Sicular, Terry. 1988. "Agricultural Planning and Pricing in the Post-Mao Period". *The China Quarterly*, no. 116: 671–705. JSTOR: [654756](#).
- Taiz, Lincoln, et al., eds. 2022. *Plant Physiology and Development*. Seventh edition. New York, NY: Sinauer Associates : Oxford University Press.
- 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.
- Tucker, Compton J. 1979. "Red and Photographic Infrared Linear Combinations for Monitoring Vegetation". *Remote Sensing of Environment* 8, no. 2 (): 127–150.
- USDA. 2023. *China and East Asia - Crop Calendars*. https://ipad.fas.usda.gov/rssiws/al/crop_calendar/che.aspx.
- Vollrath, Dietrich. 2007. "Land Distribution and International Agricultural Productivity". *American Journal of Agricultural Economics* 89 (1): 202–216.
- Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests". *Journal of the American Statistical Association* 113, no. 523 (): 1228–1242.
- World Bank. 2022. *Four Decades of Poverty Reduction in China: Drivers, Insights for the World, and the Way Ahead*. The World Bank.
- Yang, Jisheng. 2013. *Tombstone: The Great Chinese Famine, 1958 - 1962*. 1., American paperback ed. New York: Farrar, Strauss and Giroux.
- Yeh, Christopher, et al. 2020. "Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-Being in Africa". *Nature Communications* 11, no. 1 (): 2583.
- 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 (): 1220–1261.

Appendix

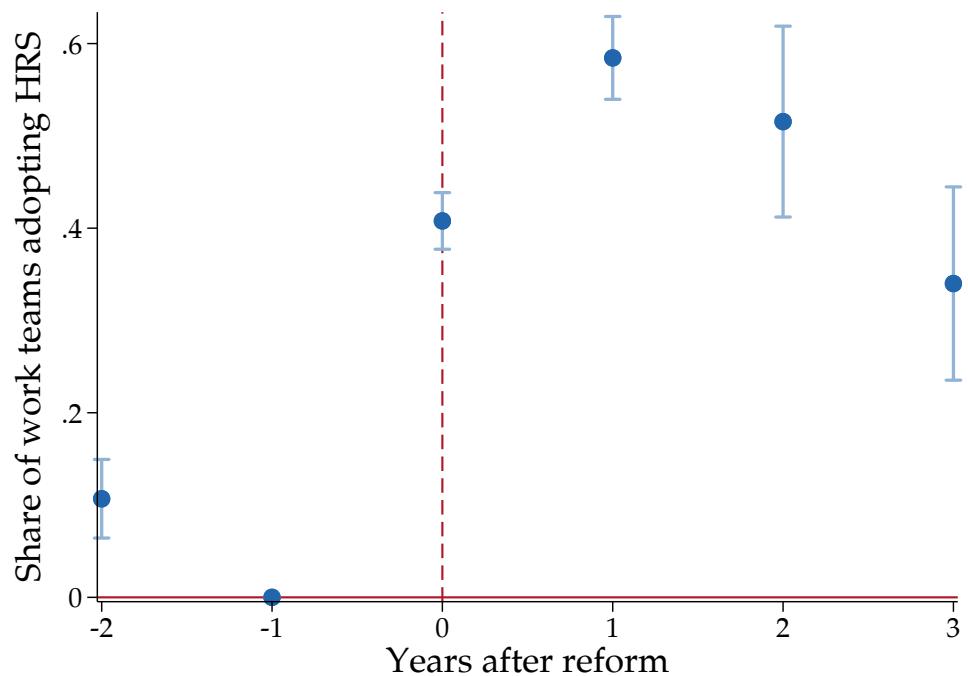
Appendix Figures

A11 Event study of HRS share to treatment indicator	40
A12 Maps of County Treatment Dates, Almond et al. (2019)	41
A13 Effect of HRS adoption on grain per capita, TWFE, provincial border discontinuity	42
A14 Effect of HRS adoption on estimated yield, national sample, different bandwidths .	43
A15 Effect of HRS adoption on estimated yield, national sample, different “donut holes”	44
A16 Effect of HRS adoption on NDVI, individual border effects	45
A17 Effect of HRS adoption on grain per capita, county-level panel	46
A18 Effect of HRS adoption on value of farm output, province-level panel	47

Appendix Tables

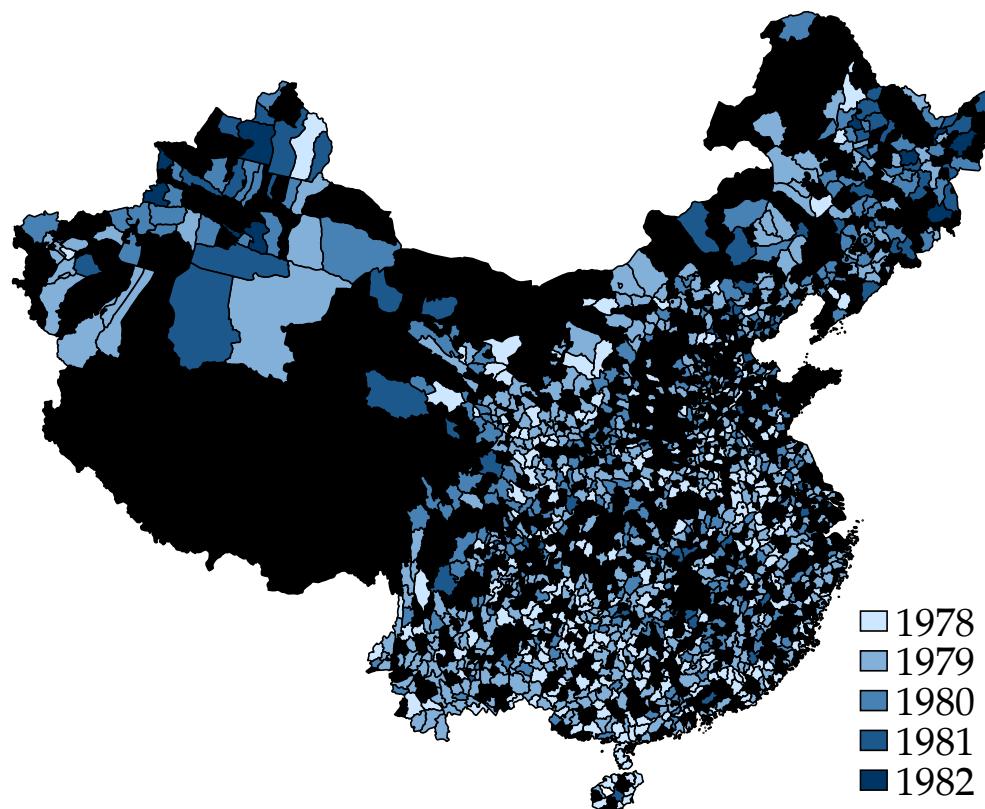
A5 Literature Estimates of the Effects of the Household Responsibility System	48
---	----

Figure A11: Event study of HRS share to treatment indicator



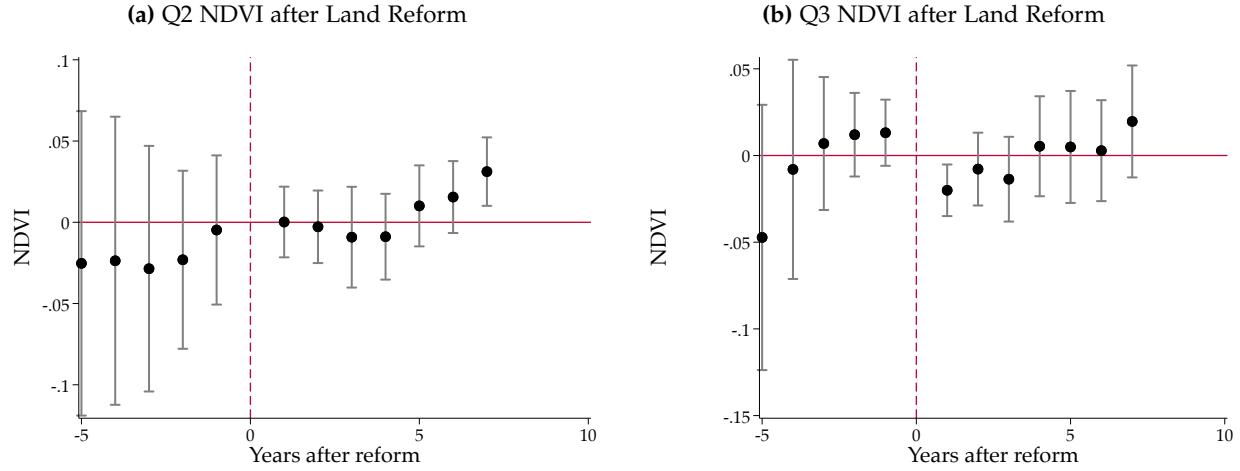
This figure shows how the continuous share of provincial work teams that have adopted the Household Responsibility System (HRS) responds to the event of the indicator variable where provincial HRS share > 50%.

Figure A12: Maps of County Treatment Dates, Almond et al. (2019)



This map shows when Chinese counties (using 1990 boundaries) adopted the Household Responsibility System (HRS) using Almond et al. (2019)'s definition: when "collectively owned land was first contracted to individual households in a few villages".

Figure A13: Effect of HRS adoption on grain per capita, TWFE, provincial border discontinuity

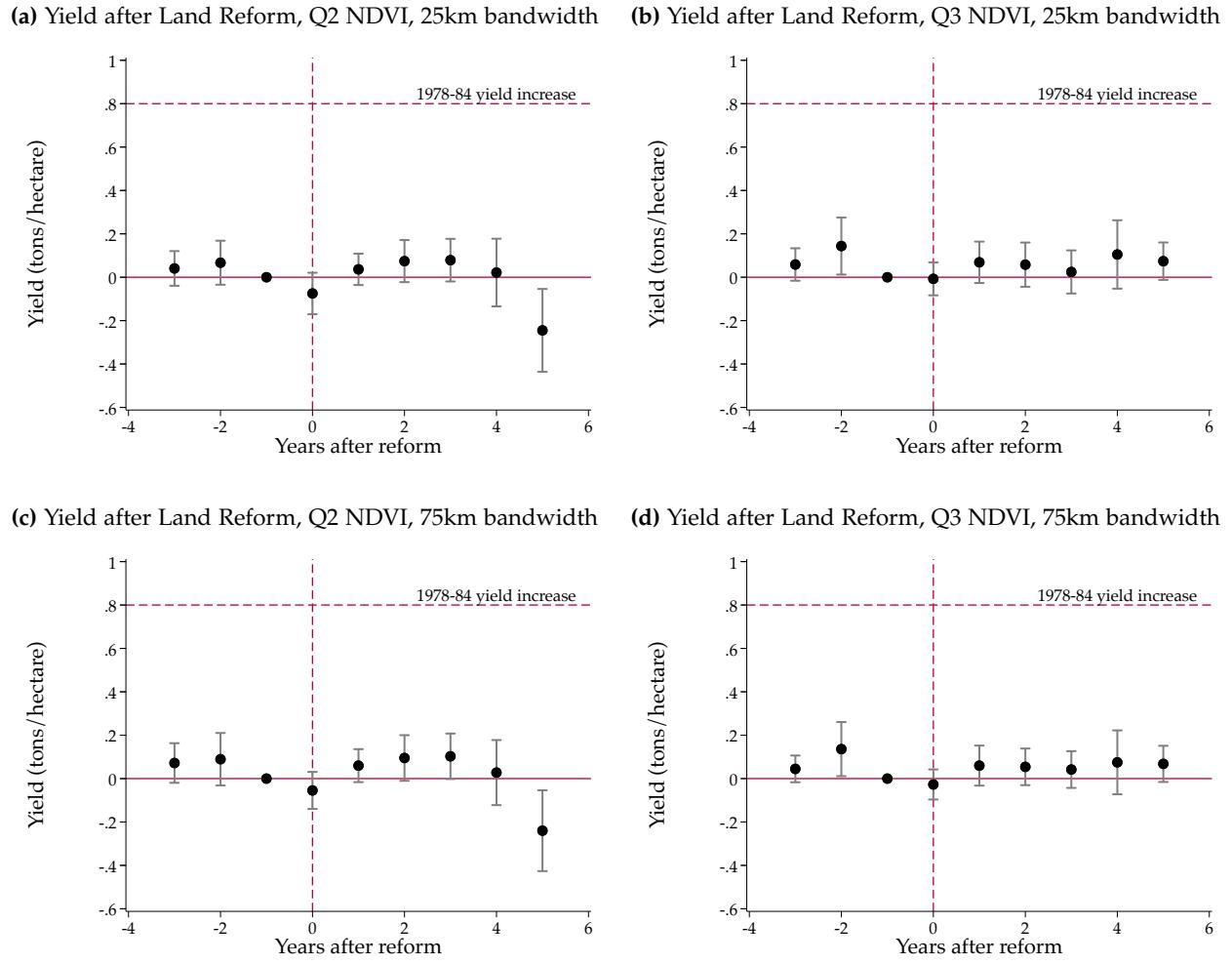


This figure shows the traditional two-way fixed effects (TWFE) estimator applied to the border regression discontinuity design for Q2 (panel (a)) and Q3 (panel (b)) NDVI. We estimate the following equation:

$$y_{i,b,t} = \sum_{h=-4}^6 H_h + \sum_{b=1}^M \gamma_{b,t+h} (R_i \times B_b \times T_t) + \sum_{b=1}^M \delta_{b,t+h} (R_i \times B_b \times T_t \times D_{i,b,t}) + \delta_i + T_t \times B_b + e_{i,b,t}^h.$$

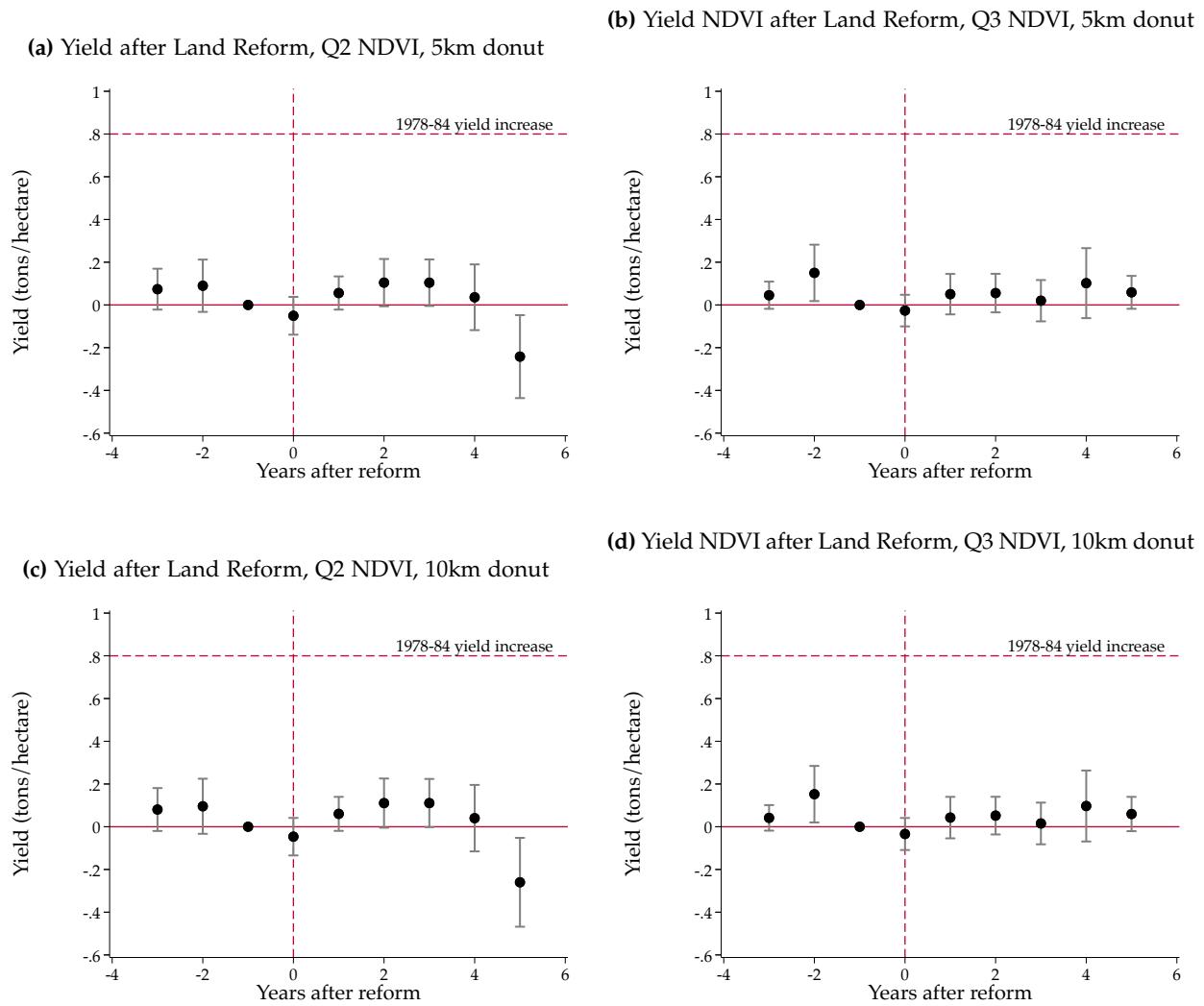
where H is an event-time indicator around treatment

Figure A14: Effect of HRS adoption on estimated yield, national sample, different bandwidths



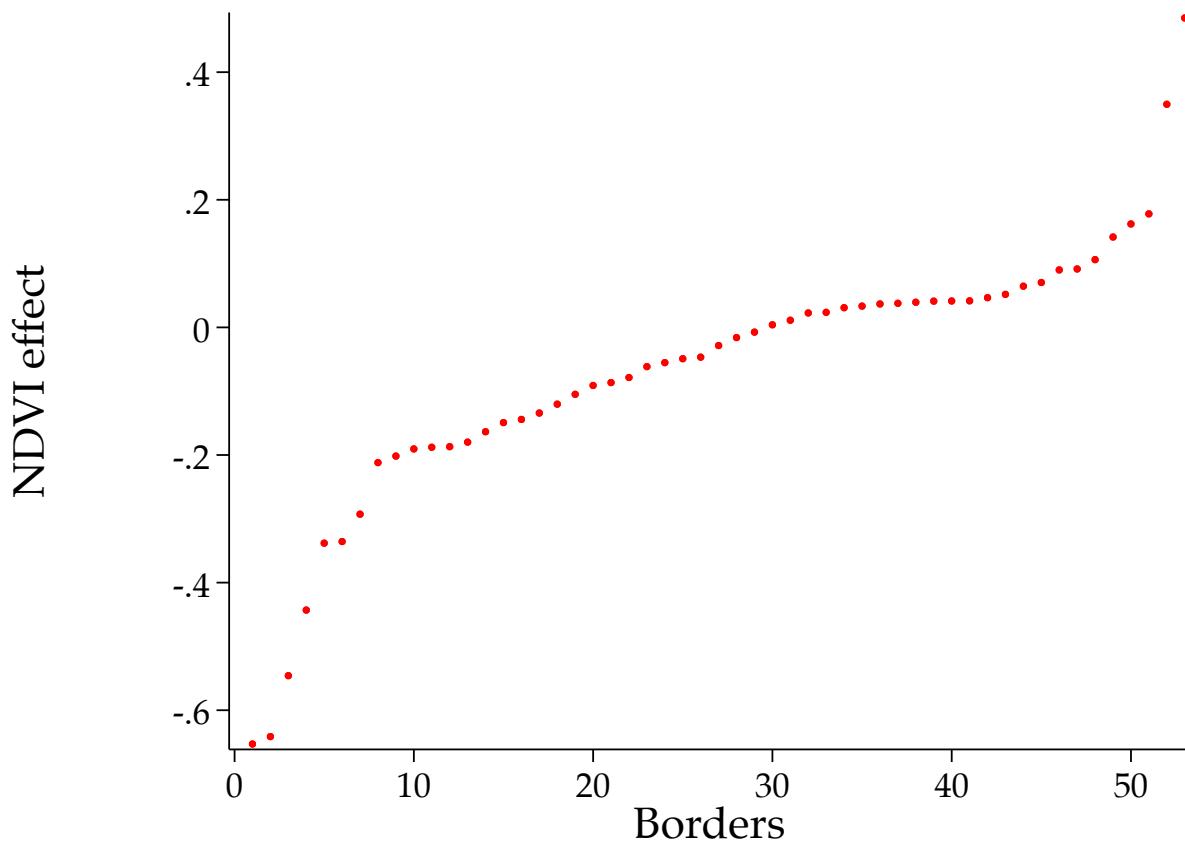
This figure shows the event study of yield following the treatment of provincial decollectivization, estimated using [Equation 4](#), using alternative distance bandwidths for the regression discontinuity—25km in the top row, 75km in the bottom row. The horizontal dashed line shows the national yield increase from 1978-84 recorded in the official data.

Figure A15: Effect of HRS adoption on estimated yield, national sample, different “donut holes”



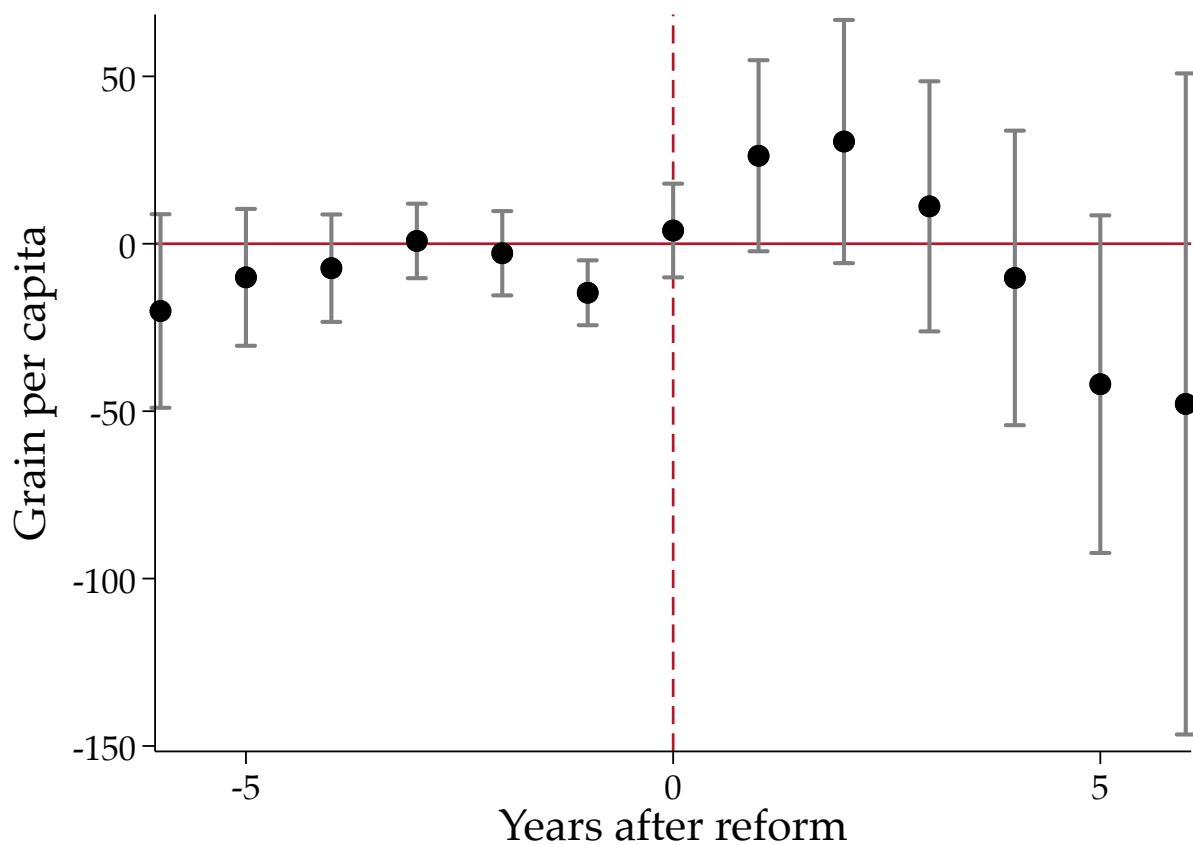
This figure shows the event study of yield following the treatment of provincial decollectivization, estimated using [Equation 4](#), dropping observations within a 5km “donut hole” immediately around the provincial border. The horizontal dashed line shows the national yield increase from 1978-84 recorded in the official data.

Figure A16: Effect of HRS adoption on NDVI, individual border effects



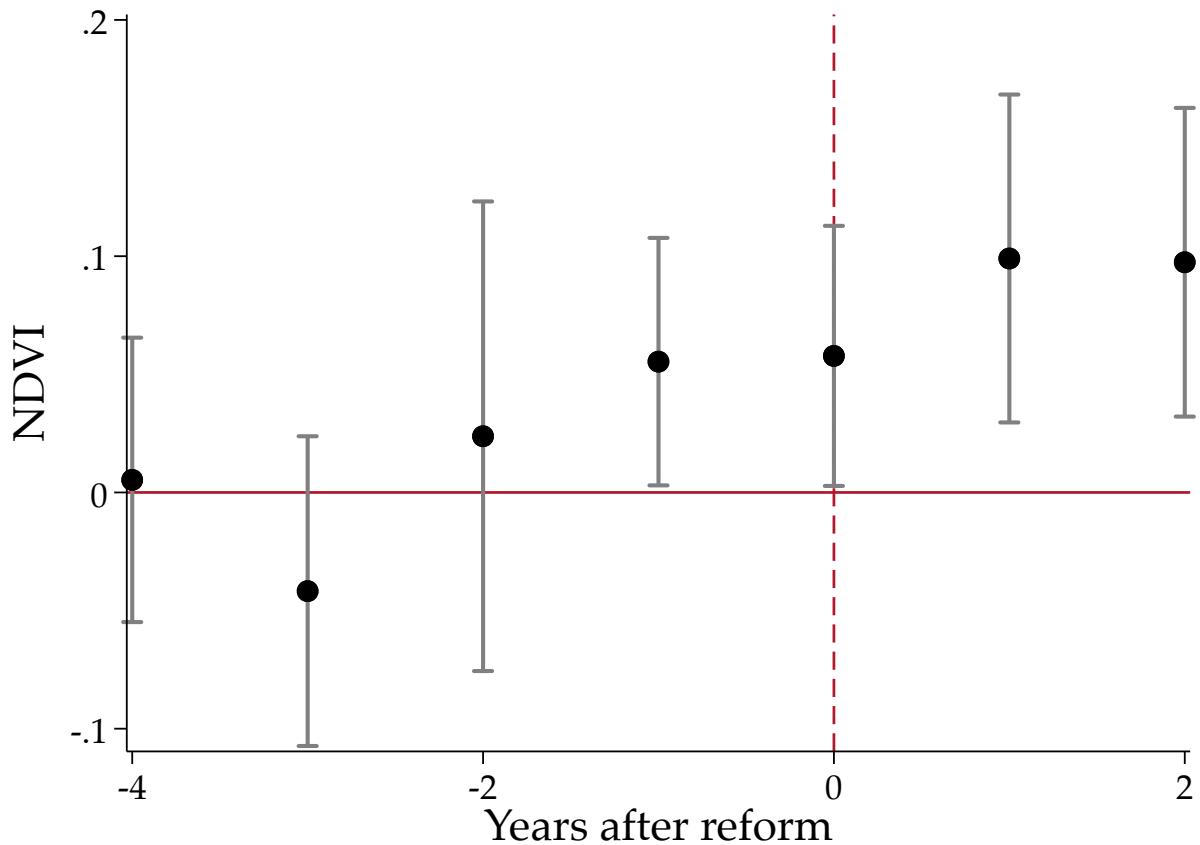
This figure shows the TWFE estimator applied to each border pair individually.

Figure A17: Effect of HRS adoption on grain per capita, county-level panel



This figure shows the Callaway and Sant'Anna (2021) estimator applied to the Almond et al. (2019) county rollout data.

Figure A18: Effect of HRS adoption on value of farm output, province-level panel



This figure shows the Callaway and Sant'Anna (2021) estimator applied to Lin (1992) provincial data. Treatment is defined as the first time when 50% of the work teams in the province report that they have adopted the HRS.

Table A5: Literature Estimates of the Effects of the Household Responsibility System

	Years	Outcome	HRS Effect (pct pts)	HRS con- tribution
McMillan et al. (1989)	1978-84	Agricultural TFP	78	78%
Lin (1992)	1978-84	Gross value of farm output	46.9	48.7%
Almond et al. (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 et al. (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 et al. (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%).