

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

Joel Ferguson Oliver Kim[†]

UC Berkeley UC Berkeley

Job Market Paper

This Version: November 27, 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 grain yield, independent of official Chinese statistics. Using two separate empirical designs that exploit the staggered rollout of the HRS across provinces and counties, we find no causal evidence that areas that adopted the HRS sooner experienced faster grain yield growth. These results challenge our conventional understanding of decollectivization, land reform, and the origins of the Chinese miracle.

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

[†]Contact: oliverwkim@gmail.com

1 Introduction

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

But how much can we trust the conventional wisdom about the HRS? Prior studies have often 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, because Chinese economic statistics are either unavailable or unreliable. In order to make precise *causal* claims about the HRS, we need disaggregated data, but sub-provincial statistics on agricultural productivity from the reform period are rare. Moreover, there are underlying concerns about the reliability of Chinese economic statistics—foreign researchers, Chinese economists, and even top Chinese leaders have all observed that they can reflect political priorities more than the ground truth.¹

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

This paper tackles these challenges and 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 in China. Using images from the Advanced Very High Resolution Radiometer (AVHRR), we can directly measure the amount of vegetation on the ground using the Normalized Difference Vegetation Index (NDVI), a common satellite-based measure which we show contains a strong signal of the reported yield (output over area). Then, to turn NDVI into interpretable yields, we apply machine learning methods from remote sensing and environmental science and 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 data. We then verify that these satellite-based measures can accurately predict yield through a range of validation exercises.

This paper’s second major contribution is to combine these satellite-based measures with treatment data on the staggered rollout of the HRS across China to give us the first credible causal estimates of the HRS’s effects on grain yields. 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 *difference-in-discontinuities* strategy that exploits the staggered rollout of the HRS at the province level. 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, semi-official 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 effects on yields as small as 5%, even three 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. Though our main focus is the *causal* effect of the Household Responsibility System on yields, as an additional exercise, we also use our random forest to predict aggregate grain yields across provinces in China. Our model predicts that aggregate yields did indeed increase from 1978 to 1984. The absence of a detectable causal effect for the HRS and the continued finding of yield growth suggests 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 China’s agricultural takeoff.

It is important to be precise about what we can and cannot say with satellite imagery. Yields—output over harvested area—are visible from space, but labor productivity is not. This paper’s findings are entirely consistent with the HRS increasing labor productivity in grain, freeing up time for industry and other crops. Nonetheless, increasing grain productivity—in by far the most important Chinese crops, rice and wheat—was a central part of the narrative around the HRS (Li et al. 2016). This paper challenges this near-universal view, a result with 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 GDP takeoff (Young 2003; Brandt et al. 2008).

The consequences for human welfare are equally large. 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 both the official province-level HRS treatment data from Lin (1992) and the unofficial county-level HRS data Almond et al. (2019) in our identification strategies. Our findings revise our understanding of one of the central reforms of China’s liberalizing period, and should prompt greater skepticism about the economic statistics underlying our conception of the Chinese miracle.

This 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 universal efficiency of smallholder farming.

Finally, this paper contributes to the growing, fruitful intersection between remote sensing research and economics. Nightlights observed by satellites have become a well-accepted measure for economic activity, particularly when there are political incentives to inflate the official economic statistics (Henderson 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](#) outlines our empirical strategy. [Section 5](#) presents our main empirical results, and discusses their implications. [Section 6](#) concludes, placing our results in historical context, 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 most of their net income among households based on their size (Eisenman (2018), p. 193). If the team produced a surplus, a system of work points gave households a cash bonus, usually based on their observed labor inputs. But because supervision is costly, standard neoclassical agricultural models predict that there are strong incentives to shirk on a collective farm, overwhelming any of the potential efficiency gains from economies of scale (Lin 1988). Or, as one Hubei farmer succinctly put it: “you’ve got Brother Zhang and Brother Li—if one works more and the other works less, it all comes out about the same” (Kelliher (1992), p. 96).

In 1978, Anhui Province was struck by a severe drought. As an emergency measure to prevent starvation, work teams in Anhui began openly experimenting with “household responsibility”. Rather than plant the winter wheat crop collectively, each working adult was allocated 0.1 hectares of land and three yuan for production costs (Kelliher (1992), p. 61). In exchange, they would turn over 100 kilograms of wheat to the collective at harvest. Any surplus could be kept by the household, to be used or marketed as they pleased. In effect, the old collective was broken. Soon, with the support of provincial leaders like Anhui’s Wan Li and Sichuan’s Zhao Ziyang, similar 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.

In 1982, both *baogan daohu* and *baochan daohu* were finally legalized nationally, formalizing these provincial reforms—and creating the Household Responsibility System. *Baogan daohu*—literally, contracting directly with households—became the favored system in most provinces, with work teams shedding their control over households’ agricultural decisions. Households were now directly responsible to the state for a mandatory procurement quota, but the residual claimants on any surplus that remained. The teams, now “villages” again after reform, kept

ownership of the land, but the law guaranteed security of tenure for at least 15 years (Brandt et al. 2004). The tractors and agricultural capital of the communes were redistributed (Eisenman (2018), p. 258), and, accompanied by reforms to the procurement price system, rural markets gradually reopened (Skinner 1985). Rural China was now firmly on the path to transition out of socialism.

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

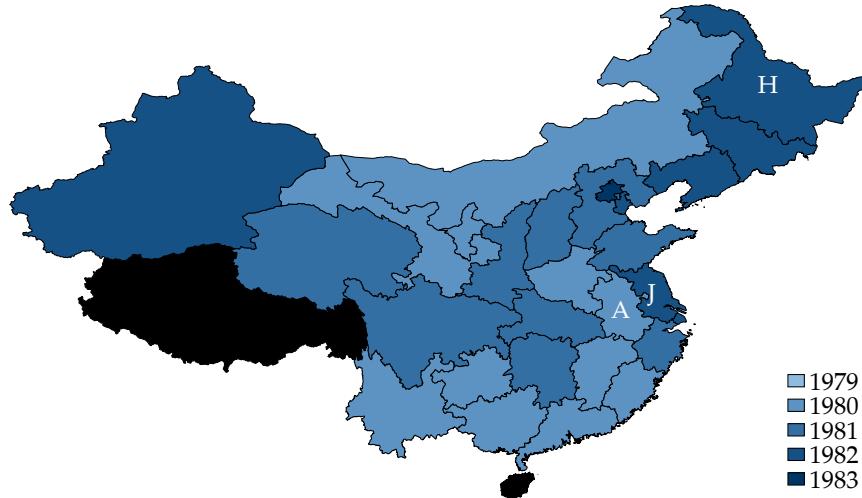
The View from the Provinces Decollectivization proceeded unevenly across provinces, resisted in some areas, pioneered in others. [Figure 1](#) shows the date when over 50% of each province's work teams reported implementing the Household Responsibility System (HRS). Anhui was the clear national reform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

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

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

The paradigm was Anhui's First Party Secretary Wan Li, who had been appointed in 1977 for orthogonal political reasons—to reduce the military's influence in provincial politics (Chöng (2000), p. 94). When peasants began implementing household responsibility in 1978, Wan Li allowed the experiment to continue. Later, he protected it from mid-level cadres who wanted to

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).

reverse the reforms, and encouraged further experimentation throughout his province (Teiwes and Sun (2016), p. 101). When faced with similar grassroots pushes for household responsibility, Sichuan’s Zhao Ziyang and Guizhou’s Ma Li permitted and even encouraged greater reform.

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

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

Mao’s immediate successor was Hua Guofeng, a relatively junior official he had handpicked from Hunan Province. Hua’s authority was much weaker than Mao’s, and reformers and hard-

liners in Beijing soon openly clashed over the reform question. Contrary to popular belief, Deng Xiaoping was not involved in the rural reform process until reforms were well underway.² At first, the hardliners held sway—November 1978’s Third Plenum Regulations explicitly banned *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).

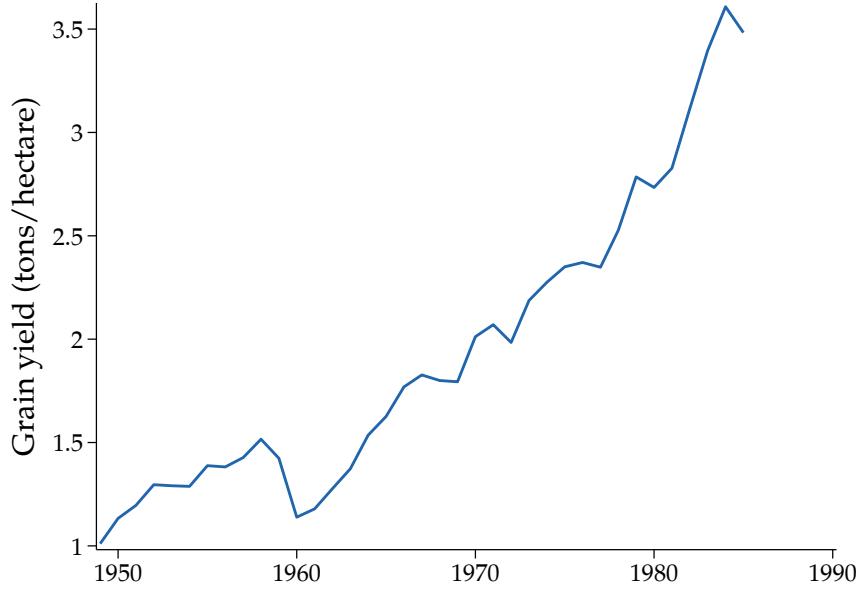
This conflict at the top had two effects on the reform process. First, because of the confused and contradictory directives coming from Beijing, provincial leaders took the lead in implementing reform. Second, because of the resistance from hardliners, HRS was first permitted only in poorer, famine-stricken,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

The Productivity Effects of Reform Figure 2 shows how official grain yields rose 43% in the six years between 1978 and 1984, a 6% annual rate—a doubling of the reported 3% growth rate from 1949 to 1978. Over just six years, reported rice yields grew by 1.4 tons, or 35%, while wheat yields grew by 1.1 tons, or 60%. 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 A4, has largely supported the official view that the HRS was responsible for this burst of 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. However, all prior econometric analyses have had to rely on central or local government

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

Figure 2: Reported grain yields in China



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

data sources. Other sources have expressed skepticism. 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. Moreover, 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: a major reform to the procurement price system in 1979. That year, 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 causal identification. We turn to constructing the satellite-based data needed for this analysis in the following section.

3 The View from Space: Measuring Yields with Satellites

3.1 Training Sample: Agricultural Data

This paper aims to produce estimates of agricultural yields without using any official Chinese data. We will present results using both raw satellite data, which we will show contains a strong signal of agricultural productivity, and using a machine learning model, which will transform these raw data into interpretable yields. To train this model, we assemble a pooled sample of agricultural data from several neighboring countries: Japan, Korea, India, and Pakistan. [Table A5](#) summarizes these sources in greater detail. We chose these countries to mimic the climatic and agricultural conditions of China's two main grain crops, rice and wheat.

Rice is the main crop in China, accounting for 50-55% of its official cereal output alone in the 1970s and 1980s. 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, South Korean county-level yields from 1981-2013, and Indian district-level yields from 1981-2005. 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.

Wheat is primarily grown in northern China, accounting for around 15-20% of official cereal production in the 1970s-80s. 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 was winter wheat, while in 2022 the USDA estimated it was 95% of China's total wheat output. To predict wheat in China, we include Indian district-level yields from 1981-2005 and Pakistani

district-level yields from Punjab province from 1990-2013.³ India and Pakistan have a similar winter wheat (or *rabi* wheat) crop to China, albeit with an earlier March-May harvest.

Though we do not employ it in our predictions or regressions, in the text we also use Chinese official rice and wheat yield data from the State Statistical Bureau as a point of comparison.

3.2 Satellite Data

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. We aggregate our satellite 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). This gives us a total of 345,608 grid cells in China, and 3,870,470 observations from 1978 to 1990.⁴

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

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

Dating back to the 1970s, NDVI has become one of the central measures in a vast literature predicting agricultural outcomes using satellite data.⁵

Figure 3 shows the evolution 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

³We 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.

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

⁵For examples, see Tucker (1979), Hamar et al. (1996), and Bognár et al. (2022). See Asher and Novosad (2020) for a recent application in development economics, studying the effect of rural roads in India.

Figure 3: Normalized Difference Vegetation Index (NDVI) over the year

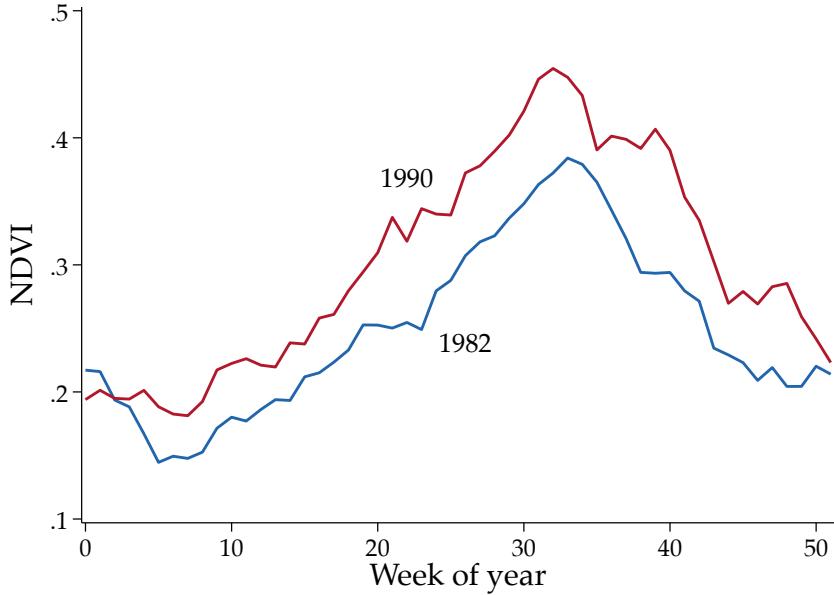


Figure 3 shows the evolution of NDVI over 1982 (blue) and 1990 (red) for an unweighted average of all Chinese provinces, using the NOAA’s AVHRR satellite data. Weekly observations are smoothed using a 4-week moving average.

and peaking in the fall—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.

Figure 4 shows scatters of annual maximum NDVI against reported grain yields for our sample of Asian countries. To ensure we are capturing time series variation in agricultural productivity, and not just the fixed cross-sectional differences between places, we de-mean the data, residualizing both NDVI and yield with respect to administrative unit fixed effects. In all cases, NDVI alone is predictive of yields, with a highly significant ($p < 0.001$) and positive coefficient

Figure 4: Annual Maximum Normalized Difference Vegetation Index (NDVI) vs. Yield

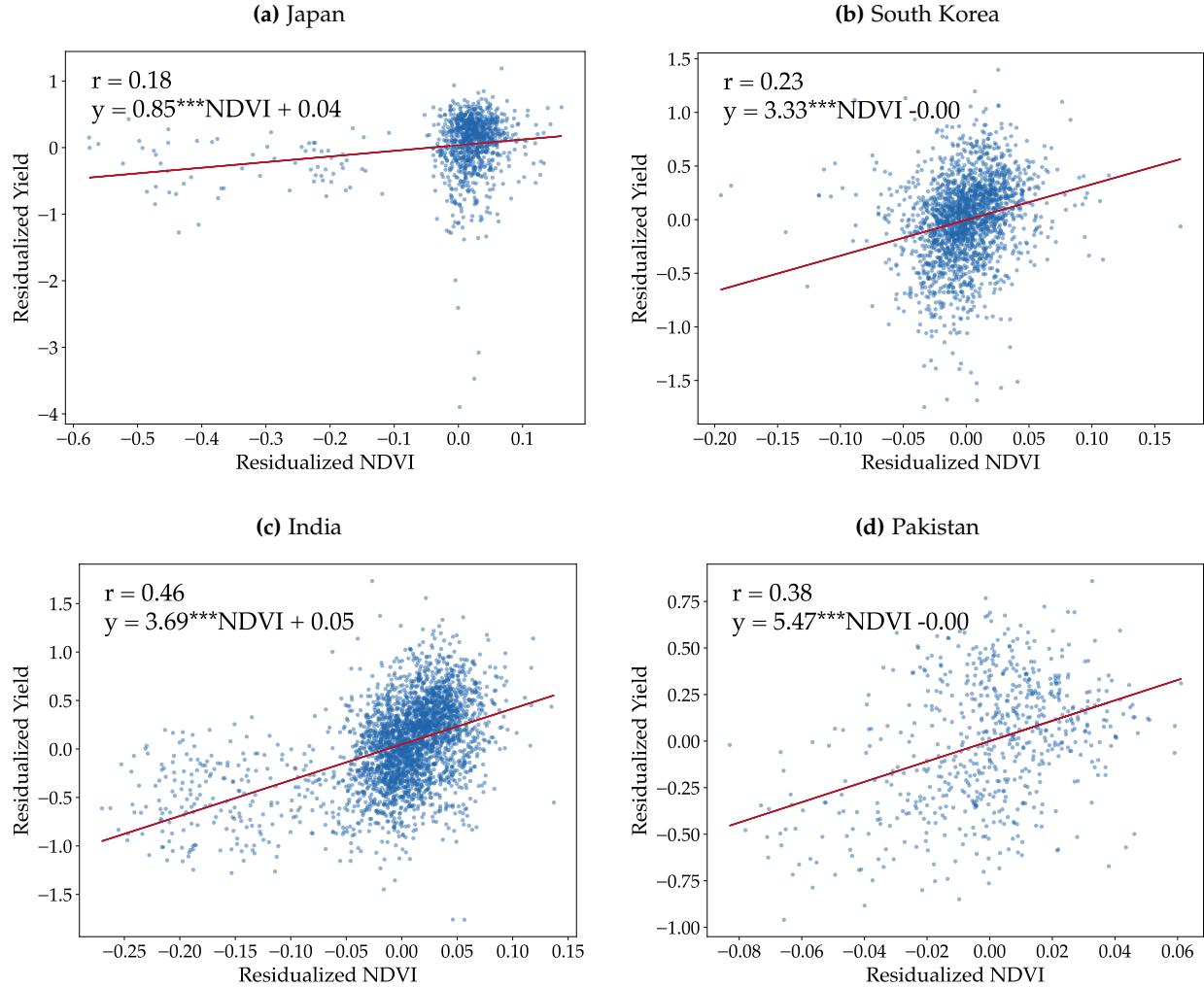


Figure 4 plots the annual maximum NDVI residual against the rice yield residual in Japan (top-left), South Korea (top-right), and the wheat yield residual in India (bottom-left) and Pakistan (bottom-right). Yield and NDVI are residualized with respect to a unit fixed effect. Correlation coefficients r and regression equations shown on the plots. $* \leq 0.05, ** \leq .01, *** \leq .001$.

across all settings. The coefficients differ across samples, ranging from less than 1 for Japanese rice yields to more than 5 for Pakistani wheat yields, highlighting that these relationships are context-dependent. And of course, there remains unexplained variation in yields—while the regression coefficients are significant, the correlation coefficients range from around 0.2-0.4. The scatters suggest that the relationships between yield and NDVI is unlikely to be simply linear, and richer sets of covariates may be needed for precise prediction. In particular, beyond the simple yearly maximum, we can build more sophisticated models that exploit variation in NDVI

within the year to better predict yields on the ground.

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 large number of studies to predict agricultural yields using remotely sensed data.⁶

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 “rice yield” and a feature is “average temperature during the growing season”, a simple one-split tree (given the dangers of high heat) might split at “ $> 30^{\circ}\text{C}$ ”. When predicting, the model assigns yields exceeding $> 30^{\circ}\text{C}$ the average from training data above this split. More sophisticated trees could split on further covariates, for instance “rainfall $>$ historical mean”. 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.

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 NDVI and yield. 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.

Features NDVI from sensors like AVHRR can be noisy because of idiosyncrasies like clouds, aerosols, and variations in satellite positioning. To smooth out this variation and to capture the

⁶See van Klompenburg et al. (2020) and Bali and Singla (2022) for systematic reviews of the machine learning and crop prediction literature. Recent examples in China include Cao et al. (2020) and Han et al. (2020).

lower-frequency patterns of crop growth, for each AVHRR pixel we fit harmonic regressions in NDVI with two pairs of terms:

$$f(t) = a_0 + (a_1 \cos t + b_1 \sin t) + (a_2 \cos 2t + b_2 \sin 2t)$$

where t is the days of the year from the start of planting, scaled into radians.⁷ We then use the fitted harmonic regressions to predict NDVI at 16 evenly spaced points in the harvest year (e.g., 0 days from the start of planting, 24 days from the start of planting, etc.), up to nine months out of the year.⁸ We use international crop calendars from the US Department of Agriculture to set the beginning and end of each crop’s annual cycle. This approach allows us to focus on the growing cycle of a specific staple crop, interpolating even when observations are occasionally missing.

Next, for each unit of analysis (e.g., an Indian district or a Japanese prefecture) at each of the 16 points in the year, we construct a histogram of the fitted NDVI values of the pixels inside that unit, counting the number of pixels that fall into each bin. (This is an adaptation to random forests of the deep learning feature reduction approach of You et al. (2017).) We use a total of 10 evenly spaced bins up to a maximum NDVI value of 0.3, above which plant cover is more likely to be forest than crops. We then use the frequencies of these histograms as features, rather than the raw NDVI values, giving us a total of $16 \times 10 = 160$ NDVI-based features for each yearly observation of an administrative unit. This helps capture the variation in vegetation *within* an administrative unit and *across* time, in a scale-invariant way.

Outcomes We predict grain yields (tons of output over area in hectares) for two major crops: rice and wheat. When training the random forest, in the training sample, we set the grain yield to be equal to the crop with the highest reported acreage. As we are ultimately interested in changes in yields, not their levels, as our main outcome, we train our model to predict log yield residuals, from regressing log-yields on their administrative unit fixed effects—or equivalently, de-meaning log yields by their historical average.

⁷Recall from high school (though we certainly didn’t) that the linear combination of sine and cosine is equal to a single sine function with phase shift and scaled amplitude: $a \cos x + b \sin x = c \cos(x + \gamma)$, where $c = \text{sgn}(a)\sqrt{a^2 + b^2}$ and $\gamma = \arctan(-b/a)$. Harmonic regression is frequently used in remote sensing to smooth noisy NDVI series, see for example Wang et al. (2020).

⁸Nine months covers the growing season for almost all crops, and avoids capturing rising NDVI from next year’s planting.

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

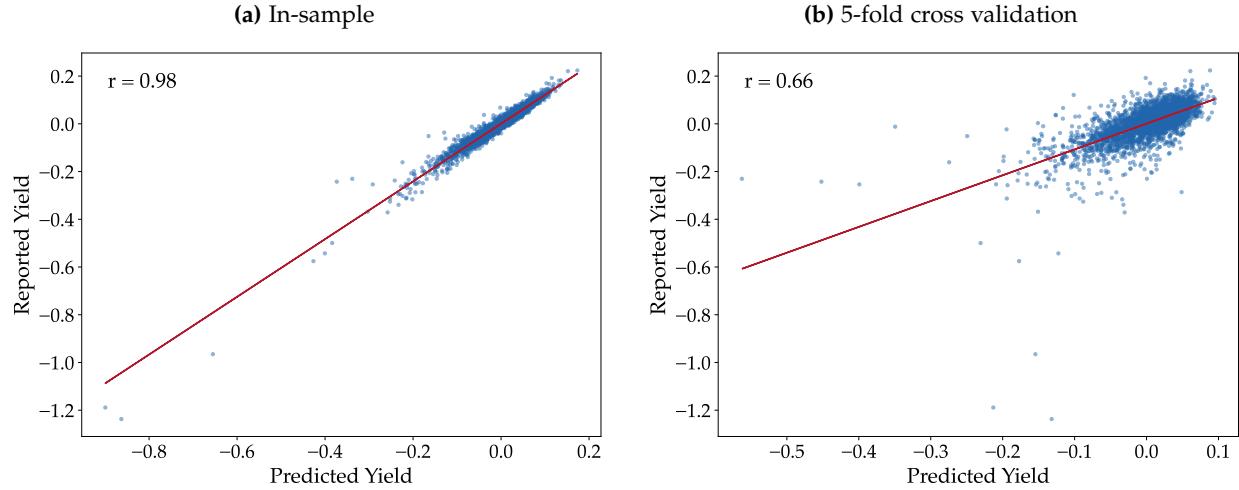


Figure 5a plots the predicted residualized grain yield against the reported grain yield for the pooled foreign sample, using a random forest model with transformed NDVI histograms as features. Figure 5b plots the out-of-sample predictions of provincial grain yield using the same random forest model trained using five-fold cross-validation. The line of best fit is in red.

Validation Figure 5 shows the predictive performance of the random forest model trained on the pooled data. On the left, Figure 5a shows the in-sample performance, plotting reported yields from the pooled training sample 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. Of course, a natural concern is that the model is overfitting towards the training set. On the right, Figure 5b plots the result of 5-fold cross-validation, where each predicted log-yield residual (x-axis) is estimated on four parts of the data and compared to the reported yields from the left-out fifth (y-axis). This approach provides a robust estimation of the model’s predictive performance by ensuring that the predictions are made on data not used during training. The predictive performance naturally decreases from the in-sample exercise, but the correlation coefficient remains relatively high at over 0.6, comparing favorably to other yield prediction machine learning models in more modern contexts (Cao et al. 2020; Bognár et al. 2022).

As we prepare to turn to our main regressions, a critical test of our random forest model will be 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 with our model yields might simply be because the random forest is predicting a vector of noise.) A natural candidate is weather shocks—in particular, temperature shocks during the growing season. For simplicity, we use the main single-

Table 1: Detecting weather shocks in Chinese grid cells

| | (1) | (2) |
|-------------------------------------|---------------------|------------------------|
| >1SD temp. shock in growing months | 0.0519* (0.0237) | |
| <-1SD temp. shock in growing months | | -0.0650*** (0.0136) |
| Observations | 1733239 | 1733239 |
| R^2 | 0.475 | 0.476 |

Table 1 shows the effects of positive and negative temperature shocks on predicted log-yields for Chinese grid cells.

crop rice growing season, which is April through October. We merge in monthly minimum temperature data from TerraClimate for each grid cell, then calculate the cell-level historical mean and standard deviation. We then compute two different shocks: a positive shock, where the growing season temperature is 1SD higher than the historical mean; and a negative shock, where it is 1SD lower. Then, with our gridded Chinese data, for cell i in year t , we estimate

$$y_{i,t} = \alpha + \beta Shock_{i,t} + \delta_i + \varepsilon_{i,t} \quad (2)$$

where $y_{i,t}$ is our predicted log-yield residuals from the random forest, $Shock_{i,t} = 1$ if the weather shock occurs in year t and δ_i is a unit-level fixed effect. We cluster standard errors $\varepsilon_{i,t}$ at the province level.

Table 1 shows the estimated effects of these weather shocks on our model-predicted log yield residuals. We find that a positive >1 SD weather shock increases measured grain yields, by around 5%, while a negative <-1 SD weather shock decreases measured grain yields, by around 6%. Both estimates are statistically significant at the $p < 0.05$ level. Note that we did not use any weather variables as features when training our model—the model is able to detect the effects of these weather shocks using features that are entirely transformations of NDVI observed by satellites. The magnitude and sign of these estimates accord with the agronomic and environmental science literatures in China. For positive temperature shocks, Chen et al. (2016) find that an increase in minimum temperature during the vegetative stage *increased* the yield of paddy rice, while Huang et al. (2013) find that a 1-degree increase to growing-season temperatures increased grain yield by 7% (for scale, the population SD is about 0.8 degrees in our dataset). Similarly, for

negative temperature shocks, Liu et al. (2013) find that rice yields in Heilongjiang fell about 0.4 tons/ha for each 1 degree fall in temperature (roughly a 10% yield shock in our setting), while Li et al. (2022) find that cold stress during the critical heading to flowering stages was negatively associated with rice yield in Yunnan, the Northeastern Plain, and south of the Yangtze.

These validation exercises confirm that our model is sensitive enough to detect relatively small (around 5%) shocks to yields in a regression setting. We can now turn to identifying the causal effects of the HRS.

4 Empirical Strategy

Our main 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%.⁹ 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. This critical feature is not captured in earlier provincial panel regression-based estimates of the HRS’s effect, like Lin (1992), which lacks a causal design. This raises the concern, for instance, that the HRS is merely capturing catch-up in places that were fundamentally poorer.

We address this concern by using a *difference-in-discontinuities* design, which extends the logic of a border regression discontinuity (RD) design to a multi-period setting. In a conventional border RD, 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

⁹Specifications with other cutoffs are available in the appendix. The results are similar.

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.

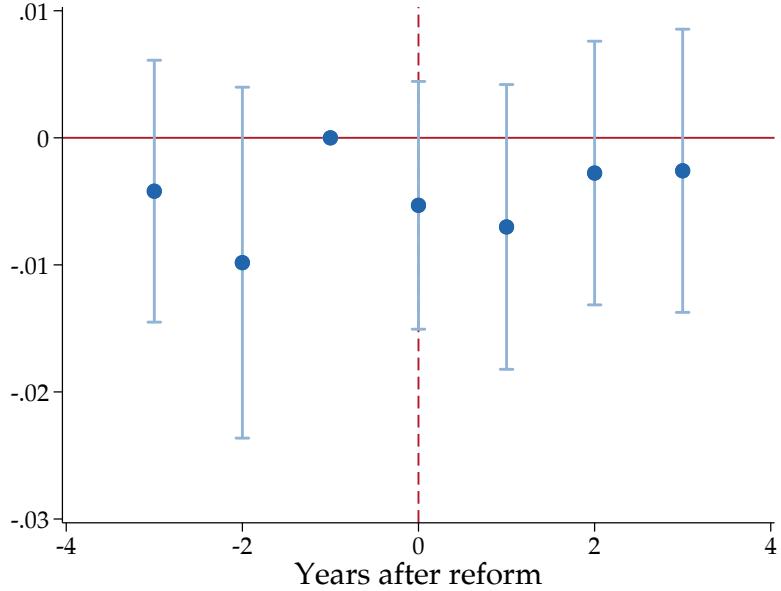
We follow the Dube et al. (2023) local projections approach, where treatment effects are estimated with separate regressions for each horizon h . Formally, for cell i in border group b in year t , at horizon h , we estimate

$$y_{i,b,t+h} - y_{i,b,t-1} = \beta_h \Delta D_{i,b,t} + \gamma_{b,t+h} (R_i \times B_{b,t}) + \delta_{b,t+h} (R_i \times B_{b,t} \times \Delta D_{i,b,t}) + B_{b,t} + e_{i,b,t}^h \quad (3)$$

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 estimate this equation over our cell-level observations over all provinces in China from 1978 to 1990. Our baseline specification includes observations that are up to 50 kilometers away from the border. Our 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.

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



This figure shows the event study of maximum yearly NDVI following the treatment of provincial decollectivization, when 50% of the province’s work teams report having adopted the HRS, estimated using [Equation 3](#). The bars show the 95% confidence intervals around the point estimates, where the standard errors are clustered at the border-group level.

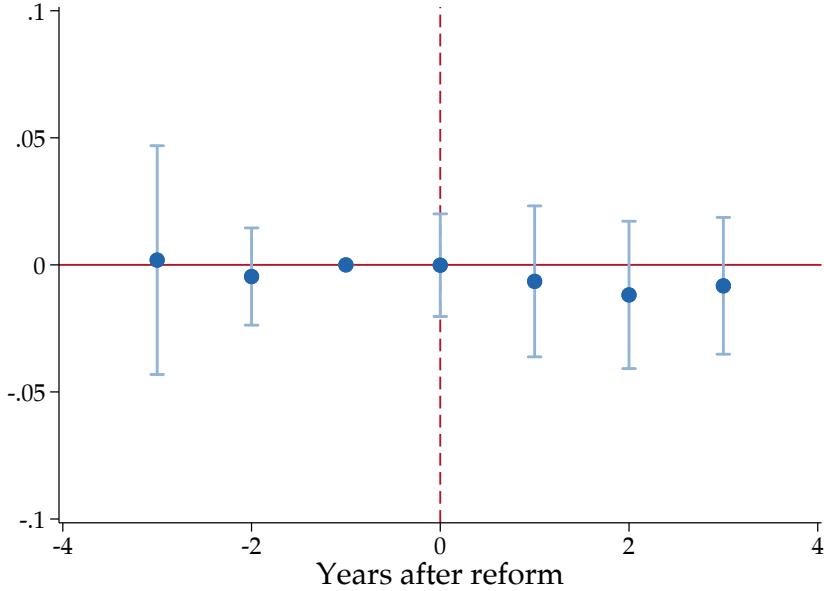
5 Empirical Results

5.1 Differences-in-Discontinuities

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

The obvious next question is how these maximum NDVI differences translate into agricultural yields. Recall from [Section 3.3](#) that maximum NDVI is highly predictive of harvest yields even in a simple univariate regression. A quick back-of-the-envelope calculation, based on the slopes in the scatter plots in [Figure 4](#), suggests that, even with the largest of slope coefficients (over 5), our

Figure 7: Differences-in-discontinuities effect of HRS adoption on estimated log yield



This figure shows the event study of model-estimated log grain yield following the treatment of provincial decollectivization, estimated using [Equation 3](#). The bars show the 95% confidence intervals around the point estimates, formed using a two-stage bootstrap procedure described in the text.

NDVI point estimates would translate into yield changes smaller than 0.05 tons per hectare—a tiny fraction of the reported 0.8 ton grain yield increase from 1978-84. For more precise estimates, we can convert satellite measurements into yields by using the random forest model trained on the pooled sample in [Section 3.3](#). For each cell, we estimate the yield using the remote sensing data, then re-estimate [Equation \(3\)](#) using the model-generated yields as an outcome variable.

[Figure 7](#) plots the effects of the provincial rollout of the HRS on our model-estimated log yields. The point estimates on yield remain indistinguishable from 0. Moreover, we retain enough statistical precision that we can reject small effects of the HRS on grain yields—we can rule out an effect of 0.05 log-points, or around 5%, even three years after the onset of reform. (A word on inference and the calculation of these standard errors in a moment.) These results are robust to variations on our core specification—we can vary the distance threshold by 10km increments and allow for a 5km discontinuity “donut hole”, where we drop grid cells whose centroid is 5km or less from the border ([Figure A12](#)). Moreover, the magnitudes are consistent with the simpler maximum NDVI results, produced using raw satellite data without the intermediary of a model.

We compute standard errors using a two-stage bootstrap procedure—because our yields are themselves estimated data, conventional clustered standard errors may understate the confidence intervals. In the first stage, we draw a bootstrap sample of the training data, clustering by administrative unit, 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 500 bootstrap samples, using the sample standard deviation of the regression discontinuity coefficient estimates to form the standard error.

How should we assess the magnitude of these effects? The official increase in national grain yields from 1978 to 1984 was 0.8 tons per hectare, or 43% (0.357 log-points). Using a decomposition of aggregate time series, McMillan et al. (1989) finds that the HRS was responsible for 78% of the 1978-84 increase in agricultural TFP, while using a panel regression of provinces, 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 even a 5% effect on yields. While TFP is a different quantity from yield (land productivity), our differences-in-discontinuities estimates are quantitatively far too small to be consistent with the view that the HRS had a large and transformative effect on yields.

Staggered Timing One potential complication with our differences-in-discontinuities design is the now well-known result that two-way fixed estimators can be biased when there is staggered treatment timing (Borusyak et al. 2023; Goodman-Bacon 2021; Sun and Abraham 2021)—a concern that applies to the rollout of the HRS. However, even with staggered treatment, the local projections estimator will remain unbiased if treatment effects are homogeneous across cohorts. We argue that this assumption is reasonable in our setting, given our finding of a tight null effect. This null effect holds both in shorter time horizons and in longer ones, as we will show in the later county differences-in-differences design.¹⁰

As a robustness check in the appendix Table A6, we also apply the “clean control” local projections approach proposed by Dube et al. (2023), where we restrict comparisons to treated

¹⁰Moreover, given the lack of treatment dynamics, the assumption of static treatment effects is also reasonable.

Table 2: Continuous Effects of the HRS on Yield

| | Horizon (h years after reform) | | | |
|-------------------------|-----------------------------------|------------------|------------------|------------------|
| | 0 | 1 | 2 | 3 |
| <i>Binned</i> | | | | |
| $50\% \leq HRS < 60\%$ | -0.01 (0.02) | -0.01 (0.01) | -0.02 (0.01) | -0.02 (0.0) |
| $60\% \leq HRS < 70\%$ | -0.01 (0.01) | -0.06 (0.04) | -0.05 (0.05) | -0.03 (0.04) |
| $70\% \leq HRS < 80\%$ | 0.00 (0.01) | 0.01 (0.01) | -0.02 (0.02) | -0.01 (0.01) |
| $80\% \leq HRS < 90\%$ | 0.00 (0.01) | 0.02 (0.02) | 0.02 (0.03) | 0.02 (0.02) |
| $90\% \leq HRS < 100\%$ | 0.01 (0.04) | 0.03 (0.04) | -0.01 (0.04) | 0.02 (0.02) |

This table shows the regression estimates for Equation (4) 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.

provinces bordering “clean control” units that haven’t yet been treated. Restricting the sample in this way constrains us to looking only one year after reform (since $h = 1$ requires that 2 years have elapsed where one province has been treated and its neighbor has not). However, restricting to clean control comparisons does not change the results.

Continuous Treatment 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 (3) by interacting the treatment indicator $D_{i,b,t}$ with 10% bin indicators of $HRS_{i,b,t}$, the share of the province’s work teams who have adopted HRS—e.g., if the HRS share falls between 50%-60%, 60%-70%, etc. Formally,

$$y_{i,b,t+h} - y_{i,b,t-1} = \beta_{h,0} \Delta D_{i,b,t} \times \mathbb{1}\{50 < HRS < 60\}_{i,b,t} + \dots + \beta_{h,4} \Delta D_{i,b,t} \times \mathbb{1}\{90 < HRS < 100\}_{i,b,t} \\ + \gamma_{b,t+h} (R_i \times B_{b,t}) + \delta_{b,t+h} (R_i \times B_{b,t} \times \Delta D_{i,b,t}) + B_{b,t} + e_{i,b,t}^h. \quad (4)$$

Table 2 summarizes the results of estimating Equation (4). We do not find evidence that the effect size is increasing in HRS—the 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.

5.2 Threats to Identification

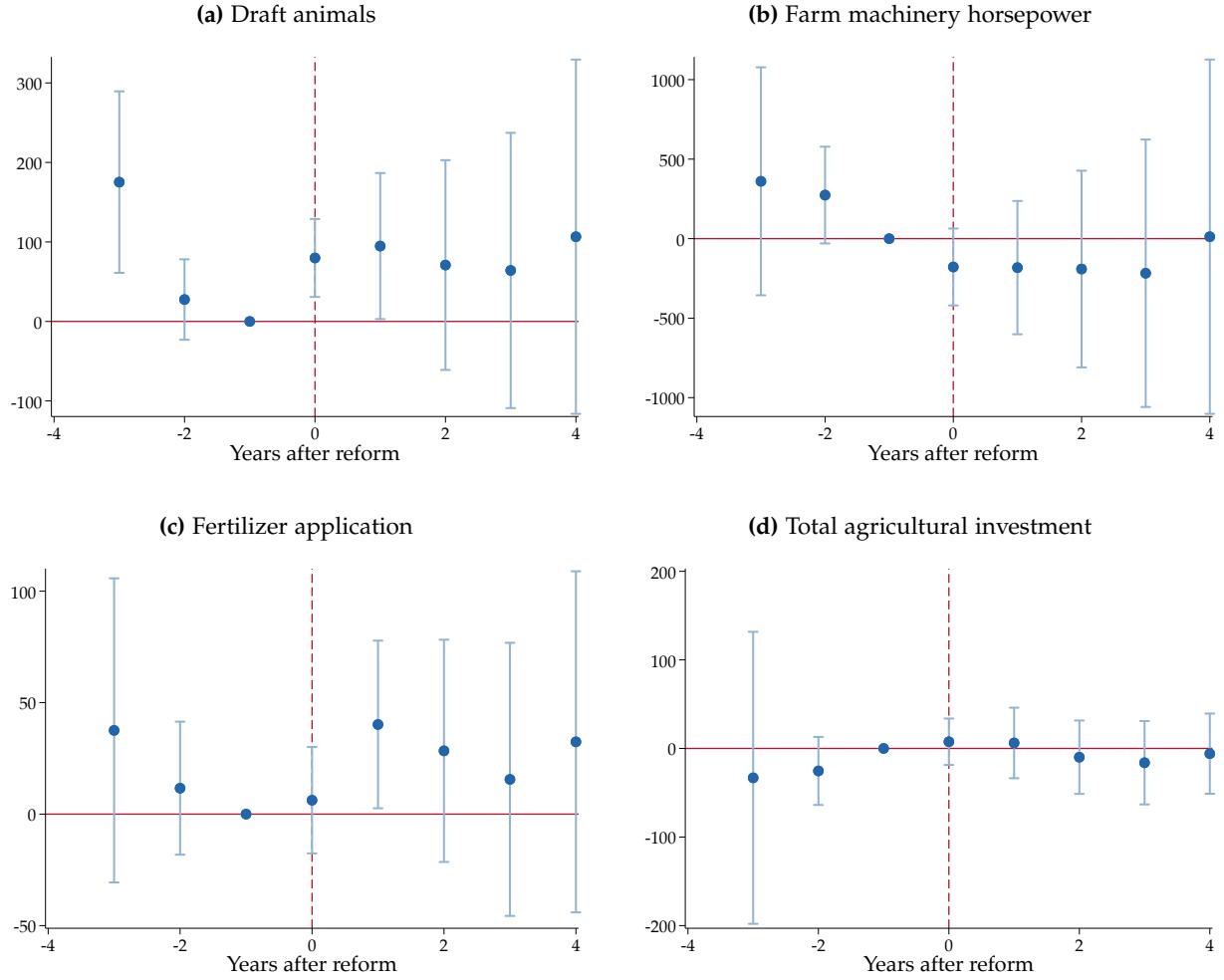
Given these surprising null results, and the literature's strong consensus in support of a strong effect of the HRS, it is natural to ask what factors other than the HRS might be driving the patterns we observe. The main threats to causal identification in our staggered-difference-in-discontinuities 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, and thus cannot influence the estimated border effects.

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

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).

border, then yields may have also risen in “control” provinces, biasing the estimated effect of land reform towards 0.

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

$$y_{i,b,t+h} - y_{i,b,t-1} = \beta_h \Delta D_{i,b,t} + \gamma_{t+h} R_i + \delta_{t+h} (R_i \times \Delta D_{i,b,t}) + B_{b,t} + e_{i,b,t}^h \quad (5)$$

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 (3), 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. (Figure A13 illustrates this logic.) We can also observe the evolution of these spillovers over time by estimating Equation (5) for each horizon h .

Table 3 shows our estimates of γ_{t+h} from Equation (5). 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 (5) 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

Table 3: 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 \(5\)](#) 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.

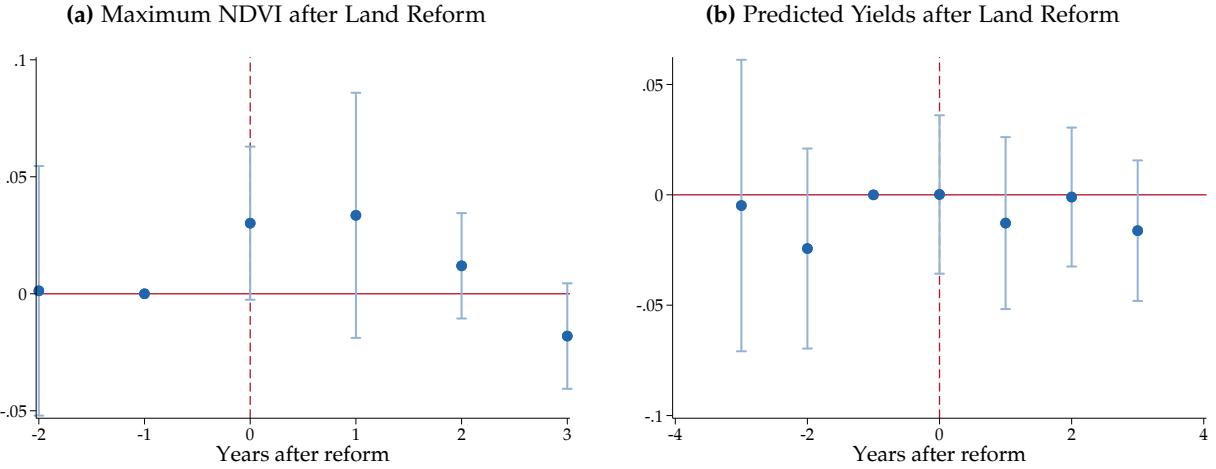
evidence that yields in untreated provinces are increasing as one approaches the border with a treated province. These results collectively suggest that spillovers are not driving our null results.

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 A11](#) 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.

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 (6). The bars show 95% confidence intervals.

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 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 ease concerns that the borders of treated areas differ systematically from their “core” regions. Figure A11 shows that counties in the Almond et al. (2019) data are scattered throughout the border and core regions of provinces. Moreover, by using 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

$$y_{i,t+h} - y_{i,t-1} = \beta \Delta D_{i,t} + \delta_t^h + \varepsilon_{i,t}^h \quad (6)$$

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. Following Dube et al. (2023), 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 annual maximum 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. 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? One possibility is there was a divergence between yield and labor productivity. Only yield, or land productivity, is visible from space; we cannot directly observe labor inputs. (The original Almond et al. (2019) data also lacked yearly data on acreage, preventing a calculation of yield.) If decollectivization increased labor productivity, freeing up labor for other uses, but did not increase the amount of output over each unit of land, this could reconcile these two findings. At present, we must remain silent on this possibility.

6 Wrapping Up: Aggregate Effects

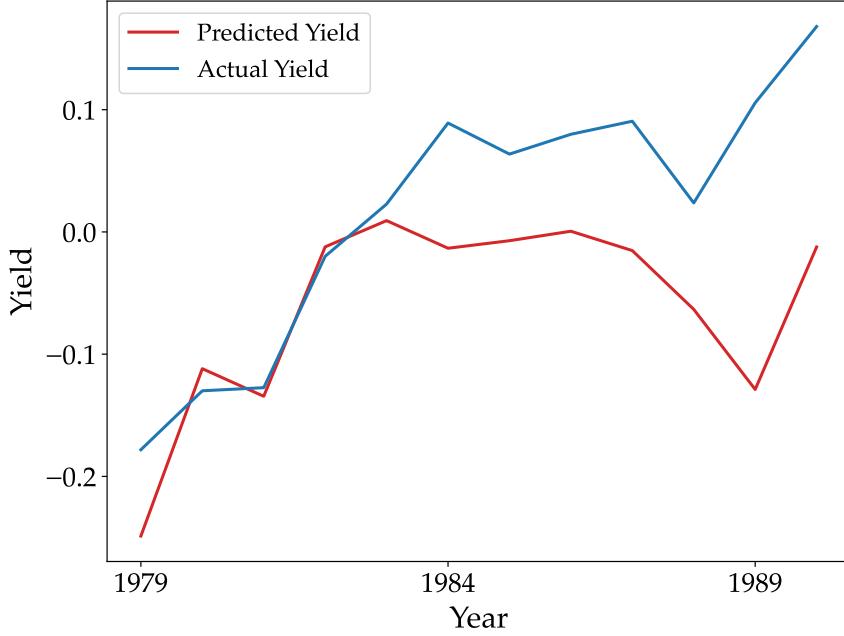
6.1 Predicting Aggregate Yields

The natural next question is—did yields increase at all?

Figure 10 plots our best estimate of aggregate rice yield for all provinces in China, using the main random forest model described in Section 3 to predict province-level NDVI features.¹¹

¹¹This reports the unweighted average across all provinces. We are working on a more sophisticated version that

Figure 10: Aggregate Yield Growth



This figure compares, at the Chinese province level, the mean model-predicted rice yields (blue) against the mean reported yields (red).

Our model predicts that aggregate yields did indeed increase from 1979-84—indeed, our predictions closely match the dynamics of the official yield series at a high frequency. However, the model predicts that yields stalled throughout the mid-late 1980s, while the reported yield series continued to increase.

It is important to interpret these aggregate results with some caution. Predicting province-level yields is a different exercise from predicting grid-cell level yields—particularly with Chinese provinces, which can be as large as European countries. Unlike with predicting responses to shocks, predictions of secular changes in yields are sensitive to the choice of countries in the training sample, which determine the domain of the outcome variable it can predict. Nonetheless, the majority of models trained on reasonable permutations of the training sample seem to show a yield increase throughout the 1980s. The model presented here represents our best current estimate—one that is best able to match the aggregate variation in Japan and South Korea (Figure A14), two relatively similar contexts, completely out-of-sample.

weights by aggregate output, which has to be estimated separately.

6.2 Conclusion

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

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

If the HRS did not cause yields to grow, what did? The obvious candidate—and the only major national-level agricultural policy change during the same period—is the 1979 rise in procurement prices, where the state raised the average price for quota grain by 20% and the bonus for above-quota grain from 30% to 50% (Sicular 1988). Since these price rises did not differ by province, their effects are not captured in the border discontinuity design. Unfortunately, we are somewhat limited in what we can formally prove about the procurement price reform—AVHRR began transmitting images only in late 1978, such that we lack a pre-period for the 1979 price change. In the appendix, Figure A15 plots the grid-level price shock (based on their underlying FAO-GAEZ agro-climatic potential yields, multiplied by the price change) against the observed

yield changes from 1979 to 1980. The correlation is positive, and the coefficient on price shocks is statistically significant at $p = 0.05$.

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

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

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

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

On a broader note, satellite imagery from sources like AVHRR present an enormous opportunity for historical research. This paper demonstrates how the latest advances in machine learning can be applied, with relatively low computational cost, to produce agricultural output data anywhere in the world, from 1978 on. The relatively high level of disaggregation provided by AVHRR also allows for the application of careful, modern causal identification to historical questions. In future work, we plan to continue to exploit this novel data source to uncover other drivers (or, in this case, non-drivers) of China's economic transformation, such as reforms to migration policy and the rise of the Township and Village Enterprises (TVEs). This paper's new causal evidence, challenging long-held views of one of the major reforms of the Chinese miracle, suggests that other surprising possibilities may exist in historical satellite imagery, waiting to be discovered.

References

- Adamopoulos, Tasso, and Diego Restuccia. 2020. "Land Reform and Productivity: A Quantitative Analysis with Micro Data". *American Economic Journal: Macroeconomics* 12, no. 3 (): 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.
- Asher, Sam, and Paul Novosad. 2020. "Rural Roads and Local Economic Development". *American Economic Review* 110, no. 3 (): 797–823.
- 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.
- Bali, Nishu, and Anshu Singla. 2022. "Emerging Trends in Machine Learning to Predict Crop Yield and Study Its Influential Factors: A Survey". *Archives of Computational Methods in Engineering* 29, no. 1 (): 95–112.
- 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.
- Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2023. *Revisiting Event Study Designs: Robust and Efficient Estimation*. arXiv: 2108.12419 [econ].
- Bramall, Chris. 2004. "Chinese Land Reform in Long-Run Perspective and in the Wider East Asian Context". *Journal of Agrarian Change* 4 (1-2): 107–141.
- 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.
- Brandt, Loren, Matthew A. Turner, and Scott Rozelle. 2004. "Local Government Behavior and Property Right Formation in Rural China". *Journal of Institutional and Theoretical Economics* 160 (4): 627.
- 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, Kam Wing. 2009. "The Chinese Hukou System at 50". *Eurasian Geography and Economics* 50, no. 2 (): 197–221.
- Chen, Shuai, Xiaoguang Chen, and Jintao Xu. 2016. "Assessing the Impacts of Temperature Variations on Rice Yield in China". *Climatic Change* 138, no. 1 (): 191–205.
- 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.
- Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing". *Journal of Econometrics*, Themed Issue: Treatment Effect 1, 225, no. 2 (): 254–277.
- 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.
- Huang, Min, et al. 2013. "Impact of Temperature Changes on Early-Rice Productivity in a Sub-tropical Environment of China". *Field Crops Research* 146 (): 10–15.
- Jean, Neal, et al. 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty". *Science* 353, no. 6301 (): 790–794.
- 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.
- Li, Xiaoyun, et al. 2016. "Patterns of Cereal Yield Growth across China from 1980 to 2010 and Their Implications for Food Production and Food Security". *PLoS ONE* 11, no. 7 (): e0159061.
- Li, Zhenwang, et al. 2022. "Long-Term Dynamic of Cold Stress during Heading and Flowering Stage and Its Effects on Rice Growth in China". *Atmosphere* 13, no. 1 (): 103.
- 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, Chang, and Li-An Zhou. 2022. "Radical Target Setting and China's Great Famine". *The Journal of Law, Economics, and Organization* 38, no. 1 (): 120–160.
- 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.
- Liu, Xiaofei, et al. 2013. "Impact of Chilling Injury and Global Warming on Rice Yield in Heilongjiang Province". *Journal of Geographical Sciences* 23, no. 1 (): 85–97.
- Looney, Kristen. 2008. 'Village Gazetteers, a New Source in the China Field.' SSRN SCHOLARLY PAPER 3027667. Rochester, NY: Social Science Research Network.
- 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.
- 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.
- Skinner, G. William. 1985. "Rural Marketing in China: Repression and Revival". *The China Quarterly*, no. 103: 393–413. JSTOR: 653964.
- Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects". *Journal of Econometrics*, Themed Issue: Treatment Effect 1, 225, no. 2 (): 175–199.
- Swinnen, Johan F. M., and Scott Rozelle. 2006. *From Marx and Mao to the Market: The Economics and Politics of Agricultural Transition*. Oxford New York: Oxford University Press.
- 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.
- Torigian, Joseph. 2022. *Prestige, Manipulation, and Coercion: Elite Power Struggles in the Soviet Union and China after Stalin and Mao*. New Haven London: Yale University Press.
- 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>.

- van Klompenburg, Thomas, Ayalew Kassahun, and Cagatay Catal. 2020. "Crop Yield Prediction Using Machine Learning: A Systematic Literature Review". *Computers and Electronics in Agriculture* 177 (): 105709.
- Vogel, Ezra F. 2013. *Deng Xiaoping and the Transformation of China*. First Harvard University Press paperback edition. Cambridge, Massachusetts London, England: Belknap Press of Harvard University Press.
- Vollrath, Dietrich. 2007. "Land Distribution and International Agricultural Productivity". *American Journal of Agricultural Economics* 89 (1): 202–216.
- Wang, Sherrie, et al. 2020. "Mapping Twenty Years of Corn and Soybean across the US Midwest Using the Landsat Archive". *Scientific Data* 7, no. 1 (): 307.
- Weber, Isabella. 2021. *How China Escaped Shock Therapy: The Market Reform Debate*. Routledge Studies on the Chinese Economy. Abingdon, Oxon ; New York, N.Y: Routledge.
- World Bank. 2022. *Four Decades of Poverty Reduction in China: Drivers, Insights for the World, and the Way Ahead*. The World Bank.
- 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.
- You, Jiaxuan, et al. 2017. "Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data". *Proceedings of the AAAI Conference on Artificial Intelligence* 31, no. 1 () .
- 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

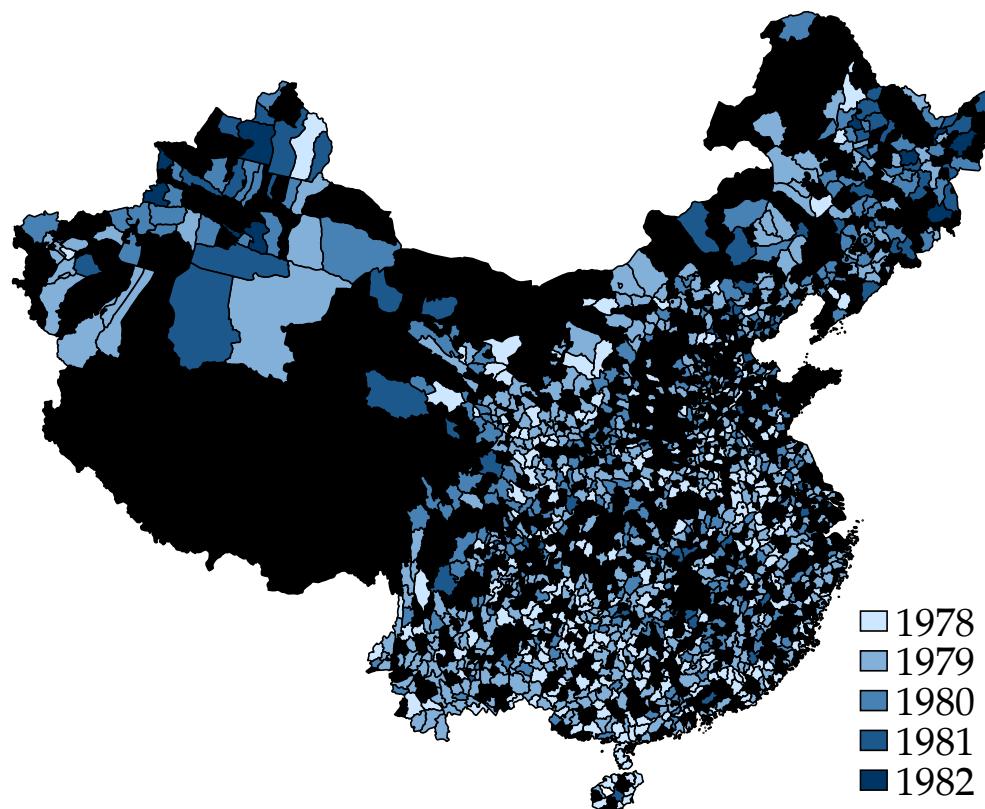
Appendix Figures

| | |
|--|----|
| A11 Maps of County Treatment Dates, Almond et al. (2019) | 41 |
| A12 Effect of Provincial HRS Adoption on Yield | 42 |
| A13 Spillovers | 43 |
| A14 Predicting aggregate rice yields in time series, out-of-sample | 43 |
| A15 Correlation between Price Shocks and Yield Growth | 44 |

Appendix Tables

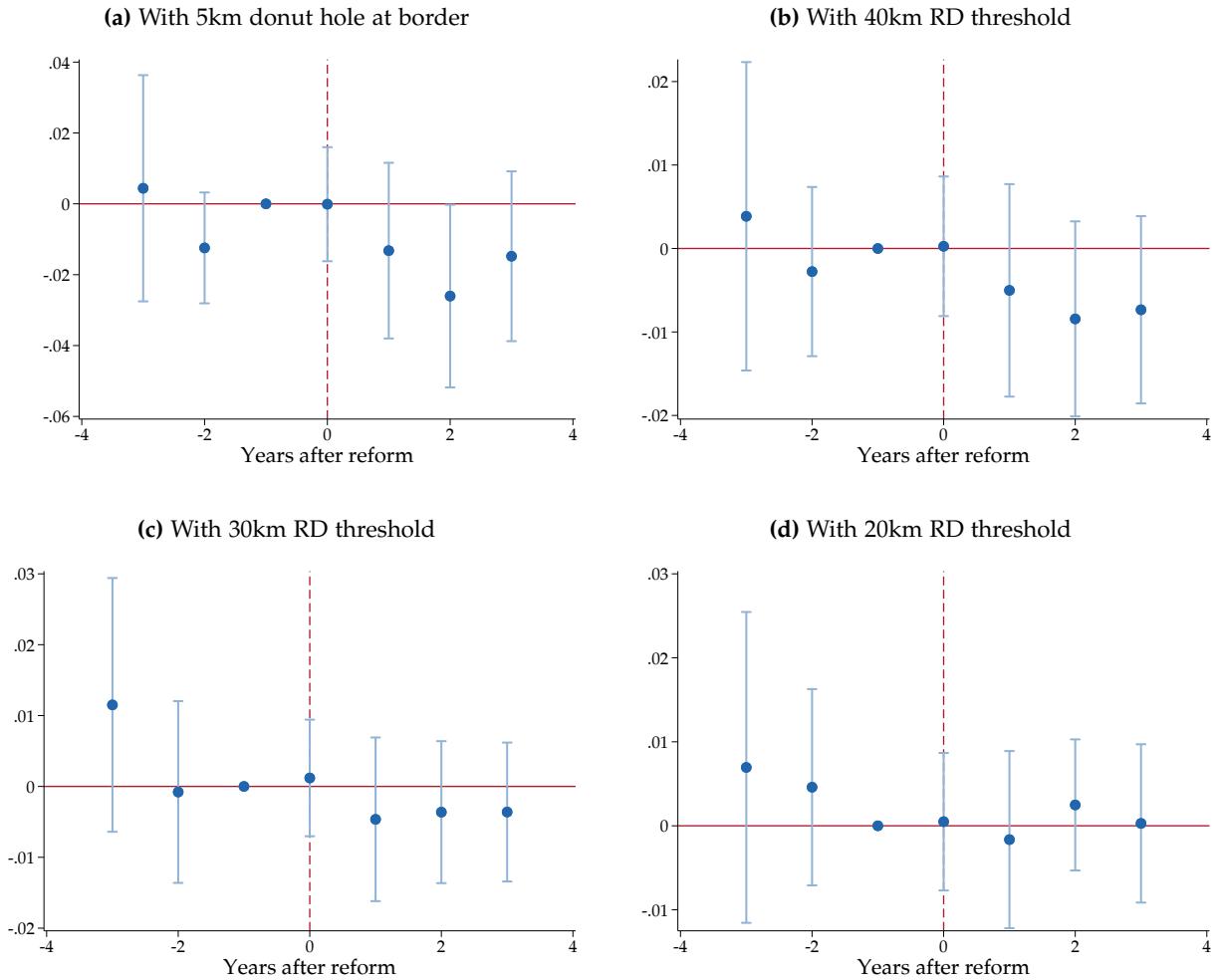
| | |
|---|----|
| A4 Literature Estimates of the Effects of the Household Responsibility System | 45 |
| A5 Training Data Sample | 45 |
| A6 Estimation w Clean Controls | 45 |

Figure A11: Maps of County Treatment Dates, Almond et al. (2019)



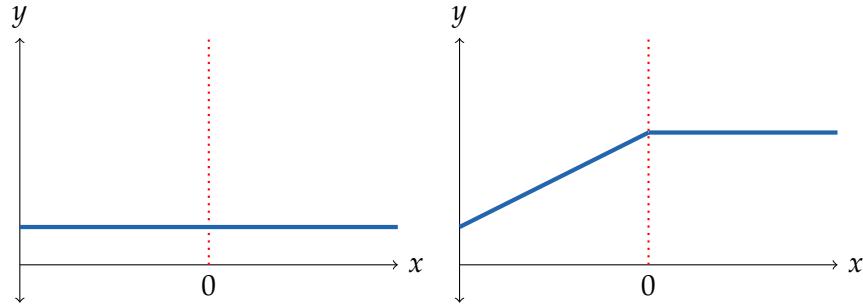
This map shows when Chinese counties (using 1990 boundaries) adopted the Household Responsibility System (HRS) using Almond et al. (2019)'s definition: when "collectively owned land was first contracted to individual households in a few villages".

Figure A12: Effect of Provincial HRS Adoption on Yield



This figure shows the effect of the provincial HRS rollout on variations on the main specification: dropping a 5km “donut hole” of the closest observations to the border, and changing the threshold of observations to include in the regression discontinuity by 10km increments.

Figure A13: Spillovers



This figure illustrates the pattern of yields we would expect before (left) and after (right) the reform that might cause a null finding, under the assumption that (1) the true effect is positive and (2) the spillovers decay with distance to the boundary.

Figure A14: Predicting aggregate rice yields in time series, out-of-sample

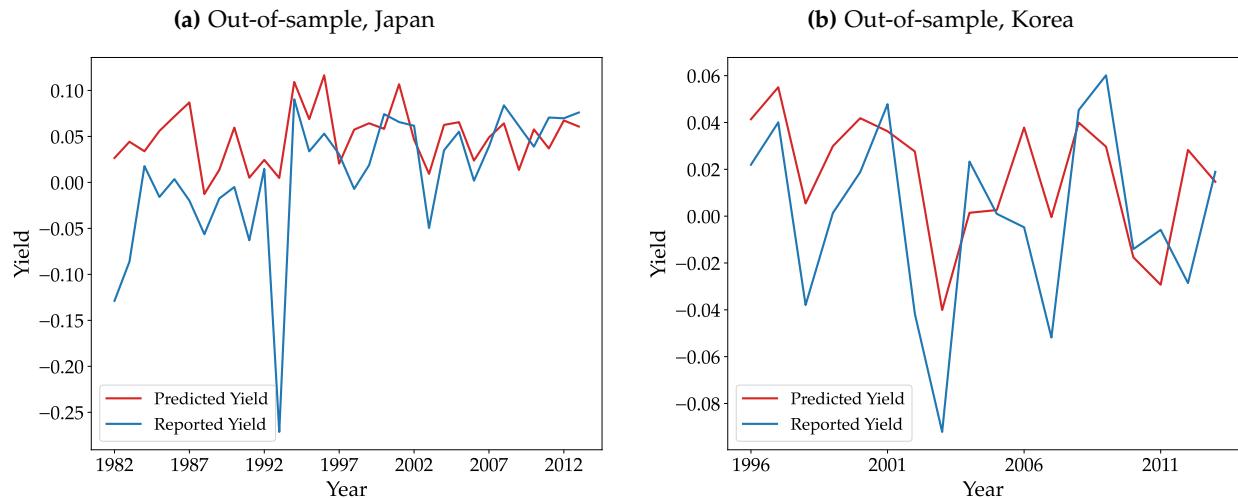
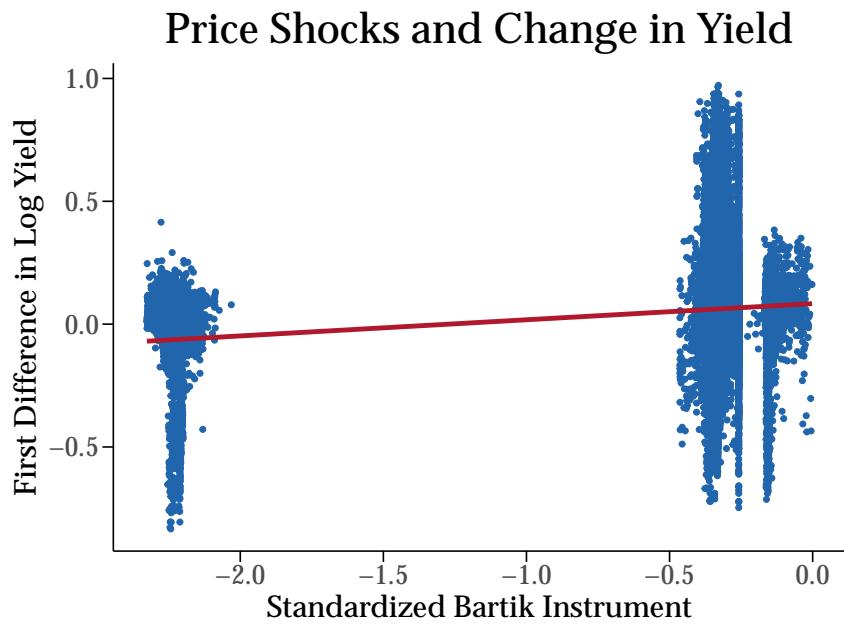


Figure A14 plots the aggregate predicted log rice yield residual against the reported log rice yield residual for Japan (left) and South Korea (right). The aggregate rice yield is the unweighted mean of prefecture- and county-level yields, respectively. Log rice yield residuals are formed by taking the log of the rice yield, and residualizing with respect to a unit fixed effect.

Figure A15: Correlation between Price Shocks and Yield Growth



This figure shows the correlation between yield growth and procurement price shocks at the grid cell level, where price shocks are defined as $Z_{i,t} = \sum_c (PotYield_{i,c} \times \Delta \ln p_{c,t})$.

Table A4: 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%).

Table A5: Training Data Sample

| Country | Years | Crops | Unit | N | Source |
|-------------|-----------|-------------|------------|------|---|
| Japan | 1982-2013 | Rice | Prefecture | 950 | Ministry of Agriculture, Forestry and Fisheries |
| South Korea | 1982-2013 | Rice | County | 2265 | KOSTAT |
| India | 1982-2005 | Rice, wheat | District | 3647 | ICRISAT; harmonized by Brey and Hertweck (2023) |
| Pakistan | 1990-2013 | Wheat | District | 638 | Punjab Crop Reporting Service (CRS) |

Table A6: Estimation w Clean Controls

| Main specification | Horizon (h years after reform) | | | | | | |
|-----------------------------------|-----------------------------------|----------|----------|----------|----------|----------|----------|
| | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
| | 0.002 | -0.005 | 0.000 | -0.000 | -0.006 | -0.012 | -0.008 |
| | (0.011) | (0.005) | (0.000) | (0.005) | (0.008) | (0.007) | (0.007) |
| Dube et al. (2023) Clean Controls | 0.002 | -0.005 | 0 | -0.000 | 0.012 | . | . |
| | (0.011) | (0.005) | (0) | (0.005) | (0.005) | . | . |