Farming on the Margin: The Green Revolution and Farm Sizes in India

Holt Dwyer *

October 23, 2024

Abstract

The agricultural sector's evolution from numerous smallholdings to larger-scale farms is widely seen as part of the structural transformation process. However, since the 1960s average farm sizes in India and many other developing countries have markedly declined despite ongoing economic development. Can rapid increases in agricultural productivity account for this trend? I investigate the impact of the Green Revolution on farm sizes in India, employing a shift-share-style instrument for adoption of improved crop varieties (HYVs) to account for potential endogeneity. I find that HYV adoption significantly increased the numbers of very small farms (but not large farms) at the district level without affecting total farmland, thereby lowering average size. I explain this result with simple model of untitled farmland in which rising agricultural productivity increases subdivision of small farms without impacting larger holdings, and show that a calibrated Chen (2017) model captures this intuition and closely matches my empirical findings in partial equilibrium. Simulation with this calibrated model suggests that the general equilibrium effects of the Green Revolution work in the opposite direction: by forestalling increases in the price of food, in aggregate the Green Revolution prevented even greater proliferation of small farms.

*holtdwyer@ucsd.edu. Many thanks to those who gave thoughtful feedback on this paper, including Prashant Bharadwaj, Sam Bazzi, Gordon McCord, Tom Vogl, Julie Cullen, Gaurav Khanna, Craig McIntosh, Munseob Lee, and Phil Hoxie. All remaining errors are the author's.

1 Introduction

> Movement of labor from agriculture to nonagriculture and the associated increase in farm size through structural transformation are at the core of economic development.
>
> Deininger et al., "Structural Transformation of the Agricultural Sector In Low- and Middle-Income Economies" (2022)

Farm sizes differ markedly across countries, and with a clear pattern. Average farm size in countries in the top quintile of GDP per capita is 34 times as large as in in the bottom quintile (Adamopoulos and Restuccia, 2014), and this relationship is not merely cross-sectional: farm sizes also rose dramatically over time in the few developed countries for which long-run data are available (Eastwood et al., 2010). As a result of these patterns, consolidation of farmland into larger holdings is widely seen as part of the process of structural transformation.

Yet development is not in fact coinciding with increasing farm sizes in much of the world. While farms continue to grow larger in high-income countries, farm sizes declined on average between 1960 and 2000 in most low- and lower-middle-income countries, particularly in South Asia (Lowder et al., 2016). During this same period, there was rapid growth in agricultural productivity in developing countries, which converged towards developed country levels of productivity (Martin and Mitra, 2001). Why are developing countries such as India not following the path trodden by developed countries and evolving larger-scale farms as their economies develop? Is high agricultural productivity growth a plausible cause? What impact does an agricultural productivity shock have on the shape of the farm size distribution?

Understanding the changing shape of the farm size distribution, not merely the change in average farm size, matters for at least two reasons. First, farm profitability varies significantly and non-monotonically with farm size in India (Foster and Rosenzweig, 2022) for fundamental technological reasons (the interaction of hiring costs and scale economies in labor-saving machinery). Thus, the shape of the farm size distribution impacts the aggregate productivity of the largest sector in the economy in a way that is not captured by changes in average size. Second, in the presence of land frictions, increases in the value of farmland cause retention of labor in agriculture that is not necessarily efficient (Chen, 2017) and that delays the structural transformation process. Viewed through the lens of such a model, the farm size distribution's response to an agricultural productivity shock can help us understand which households are on the margin of leaving or remaining in agriculture, which is important for understanding how both agricultural and non-agricultural productivity change as labor shifts between sectors (Lagakos and Waugh, 2013). Changes in the shape of the farm size distribution are thus a potentially important channel through which agricultural productivity shocks impact structural transformation.

Assessing the impact of an agricultural productivity increase on the farm size distribution is difficult due to the complex interrelationship between farm size, agricultural productivity, and technology adoption. Yields and productivity each vary with farm size (Foster and Rosenzweig (2022), Aragón et al. (2022), Deininger et al. (2016), Barrett et al. (2010), among many others) and farm size may impact adoption of new agricultural technologies due to large farmers' greater ability to bear the risk of new technologies and pay fixed adoption costs (Just and Zilberman (1983), Feder and O'Mara (1981), Foster and Rosenzweig (1996)). Thus, investigating the impact of agricultural productivity increases on farm sizes requires both controlling for the initial farm size distribution and identifying a source of variation in productivity gains that is not subject to reverse causation from the size distribution itself.

In this paper, I examine historical data on the farm size distribution in India (home to a quarter of the world's farms) to assess the impact of an increase in agricultural productivity on the distribution of farm sizes. I leverage variation driven by the Green Revolution, a process of rapidly rising agricultural productivity in developing countries that began with the release of new high-yielding crop varieties (HYVs) from international agricultural research centers in the late 1960s. HYV varieties of rice, wheat, and maize were the first to be released because these crops had been the subject of extensive study in developed countries (Evenson and Gollin, 2003); HYVs for other crops were researched and released over subsequent decades. The increased agricultural productivity brought about by these crops had significant impacts on developing economies, increasing returns to education (Foster and Rosenzweig, 1996), reducing infant mortality (Bharadwaj et al. (2020), von der Goltz et al. (2020)), reducing population growth and increasing per-capita income (McArthur and McCord (2017), Gollin et al. (2021)).

As previously noted by Foster and Rosenzweig (1996) and others, the onset of the Green Revolution provides useful variation for studying the effects of agricultural productivity growth on developing economies for two main reasons. First, the technology (embodied in HYVs) was developed outside the developing countries themselves, reducing concerns about reverse causality in the innovation process. Second, regions were differentially exposed to the new technology on the basis of their suitability for the particular crops that happened to be released first, thus generating cross-sectional variation that is plausibly uncorrelated with omitted drivers of changes in the size distribution. Remaining threats to identification include the potential for unobserved drivers of background changes in the farm distribution (for example,)

In order to investigate the impacts of agricultural productivity on farm size, I construct an instrument for HYV adoption that combines pre-existing crop-wise potential for yield increases with exogenous variation across crops in the timing of HYV development. This instrument is similar to the proxy for exposure to the Green Revolution used in Gollin et al. (2021) and the instruments for agricultural productivity increases used in Moscona (2019), Nunn and Qian (2011), and Bustos et al. (2016). I then investigate the impacts of HYV adoption using a panel specification that controls for the impacts of the starting farm size distribution, unobserved state-level policies, and trends driven by pre-Green Revolution conditions including expected population growth.

The resulting estimates suggest that HYV adoption from 1970 to 1990 increased the relative total number of farms at the district level by roughly a third, with no impact on total farmland, thereby significantly reducing average farm size in more exposed districts. Results are qualitatively similar (although less precise) even when state-year fixed effects or controls for starting conditions are omitted. These impacts are concentrated in the left tail of the distribution: the number of extremely small farms between 0 and 1 hectare increased by roughly 20% of the entire initial number of farms and their share of farmland increased significantly. Meanwhile, there were essentially no impacts on the number of large farms and only small and insignificant effects on their share of farmland.

Why are these effects limited to the left tail of the distribution? How do these district-level estimates relate to the size distribution at the national level? I present a simple model in which farmland evolution is determined by inheritance decisions rather than sales, as was largely the case in this context, and show that the impact of a productivity increase in this model matches the pattern of impacts that I estimate. I then calibrate a Chen (2017) model with untitled farmland to 1970s India and show that the impact of a plausibly sized agricultural productivity shock in partial equilibrium closely matches my district-level estimates. Feeding this same shock into a general equilibrium version of the model allows me to simulate the effects of the Green Revolution on the overall size distribution. This exercise suggests that the general-equilibrium effects operate contrary to the partial-equilibrium effects, consistent with prior theory (Matsuyama, 1992). The simulation shows that by forestalling a demographically driven increase in the price of food, the Green Revolution substantially reduced the total number of farms in India (by roughly 44% of their 1970 total), slowing but not halting the process of subdivision.

Simple representative agent models with freely transacted farmland suggest that neutral or land-augmenting productivity increases will retain additional households in agriculture, thereby decreasing the representative farm size (Eastwood et al., 2010). However, there appears to be no prior empirical evidence supporting this prediction or estimating the possible magnitude of this effect. Given that farmland is not freely transacted in much of the developing world, and that general equilibrium price effects (not present in these models) may counteract direct productivity effects, both the direction and magnitude of changes in farm sizes are a priori unclear. Moreover, more recent models featuring non-degenerate farm size distributions (Adamopoulos and Restuccia (2014), Chen (2017)) have differing predictions about how the shape of the distribution should change; these have also not yet been empirically assessed. I provide empirical evidence on the effects of an important agricultural productivity shock and discuss how these effects compare to the predictions of existing models.

This paper is related to two main strands of literature. The first is the empirical literature on the impacts of the Green Revolution (Evenson and Gollin (2003), McArthur and McCord (2017), Moscona (2019), Bharadwaj et al. (2020), von der Goltz et al. (2020), Gollin et al. (2021)), which leverages differences in pre-existing suitability to study the structural transformation effects of agricultural productivity shocks that affect some areas more intensely than others. I provide the first causal estimates of the Green Revolution's impact on farm sizes, pointing out a potentially important secondary effect of agricultural productivity increases in the presence of real-world land frictions. The second is the applied theory literature on the determinants of the farm size distribution (Adamopoulos and Restuccia (2014), Adamopoulos and Restuccia (2020), Chen (2017)), which model the relationship between the farmer productivity and farm size distributions and use these models to investigate the importance of land and labor misallocation driven by real-world frictions. I argue that my empirical estimates are best explained by a model in which land frictions prevent reallocation of land across households, leaving inheritance decisions as the only margin for adjustment. This has the implication that increases in agricultural productivity may worsen the sector-distorting effect of untitled land noted in Foster and Rosenzweig (2007), Chen (2017), and Fernando (2022). This paper also relates indirectly to the large literature on the size-productivity relationship in agriculture (Foster and Rosenzweig (2022), Barrett et al. (2010), Adamopoulos and Restuccia (2014), Deininger et al. (2016), Aragón et al. (2022), and many others) by providing empirical evidence on the way in which the farm size distribution, which plays a role in determining aggregate agricultural productivity, is itself affected by agricultural productivity shocks.

2 Data

Data on both the farm size distribution and HYV adoption at the district level over time come from the ICRISAT's District Level Dataset (DLD), which consolidates data on Indian districts from multiple sources to the level of districts as they existed in approximately 1966.[1] The main outcome variables are the share of total farms and share of total farmland in each of five size categories used in the Agricultural Census: "marginal" (<1 ha), "small", (1-2 ha), "semimedium" (2-4 ha), "medium" (4-10 ha), and "large" (> 10 ha), reported every five years from 1970 to 1990. Roughly speaking, a typical farm household can manage a farm of 1-2 ha without hiring outside labor (Yamauchi, 2021); thus "marginal" and "small" farms can be thought of as farms that can be managed entirely with household labor. Starting demographic variables included in the controls are from the Indian Population Census via Vanneman and Barnes (2000). Further details on dataset construction can be found in Appendix D.

The variable of interest, *HYVShare*, is the gross share of cultivated area used to grow HYV crop varieties; this means that the estimates incorporate both expansion of HYV cropped area in one season area and intensification of HYV cropping into additional seasons.[2] Many states stopped reporting the share of land under high-yielding varieties

[1] This dataset is also sometimes referred to as the Village Dynamics in South Asia (VDSA) Meso Database (`https://vdsa.icrisat.org/vdsa-database.aspx`).

[2] In practice, the amount of cultivated area in a district is effectively constant over this time period; I divide by the 95th percentile of cultivated area in the district so that the denominator is constant.

around 1990, so I focus on the period from 1970[3] (India's first Agricultural Census) to 1990 for which the district panel is closer to balanced.

3 Methodology

I identify the effects of HYV adoption on the farm size distribution using panel IV regressions of the following form. For outcome variable $Y_{l,d,t}$ (a share of land, share of farms, or normalized number of farms in category l, district d at time t), I estimate the following specification:

First stage:

$$HYVShare_{d,t} = \alpha'_{l,d} + \gamma'_{l,s(d),t} + \beta'_l \cdot PYI_{d,t} + \sum_{l \in L} \gamma'_l Y_{l,d,1970} \cdot t + X_{d,t}\nu'_l \epsilon'_{d,t} \qquad (1)$$

Second stage:

$$Y_{l,d,t} = \alpha_{l,d} + \gamma_{l,s(d),t} + \beta_l \cdot HYVShare_{d,t} + \sum_{l \in L} \gamma_l Y_{l,d,1970} \cdot t + X_{d,t}\nu_l + \epsilon_{d,t} \qquad (2)$$

The fixed effects α_d control for the level effect of time-invariant characteristics of district d (such as soil quality, elevation, ruggedness, location, or long-run climate). The state-time fixed effects $\gamma_{s(d),t}$ control for time trends as well as any unobserved state-level policies. Thus all identification in this specification is based on differences in relative changes from baseline across districts in the same state in the same year. The key coefficient β_l captures the impact of growing entirely HYV crops for a full cropping season on the farm size distribution. In discussing the magnitudes of the estimated effects of HYV adoption in the text, I rescale these coefficients by the actual average increase in the HYV share over this time period (roughly 0.3). Standard errors are clustered by district.

In the counterfactual in which the Green Revolution never happened, there might

[3] Although the first HYV crops were introduced to India in 1965-66, adoption was a slow process taking place over decades. It is thus reasonable to treat 1970 as approximately the start of the Green Revolution, as in Foster and Rosenzweig (1996), particularly given that farm sizes are unlikely to react instantaneously.

still have been differential changes in the distribution of farmland across districts driven by differences in the initial size distribution. For example, one might theorize that a change in returns to scale driven by expanded rental markets for machinery is responsible for the decline in the land share of "large" farms. If this were so, one might expect areas with initially high shares of land in "large" farms to experience different size dynamics than those with initially few "large" farms, even in the absence of the Green Revolution. To combat this concern, I control for linear trends interacted with the initial size distribution.[4] Thus any detected effects of the Green Revolution are not simply driven by correlation between HYV adoption and trends driven by the initial farm size distribution.

Other factors beyond year fixed effects and the initial farm size distribution may affect the distribution of farmland. $X_{d,t}$ contains additional controls that interact initial values of potentially relevant geographical and demographic variables observed in 1961 (prior to the Green Revolution) with linear time trends. These starting variables are the urban population share, share of labor in agriculture, literate share of the adult population, the ratio of rural minors to adults (a proxy for predetermined rural population growth), Hindu population share (Hindus and Muslims have different inheritance practices which might affect the evolution of farm sizes) and the rural population density. I also include an indicator for whether a district is on the coast (increasing access to trade might plausibly have pulled labor out of agriculture in these areas). These are denoted as "Addl Controls" in tables (their coefficients are not reported for space reasons). I also control for rainfall in the district in year t.

3.1 Instrumental Variable

It is reasonable to think that part of the correlation between HYV adoption and the farm size distribution is driven by differential adoption of HYVs by large versus small farms, at least in the early years of the Green Revolution (Feder et al. (1985), Foster and Rosenzweig (1996)). To avoid interpreting these endogenous correlations as the impact of HYVs, I instrument for HYV adoption with an instrument I refer to as Predicted Yield

[4]When the outcome is land shares, I use initial land shares; when it is a number share or normalized count, I use initial numbers in each category. A full set of starting shares (which sum to 1 for each district-year) interacted with time trends is collinear with a linear trend, which itself is collinear with the state-year fixed effects (which sum across all states to year indicators). Thus in practice the starting marginal share interaction is omitted from the controls to prevent multicollinearity.

Increase (PYI). PYI is the value of a "back-of-the-envelope" calculation of how much total agricultural yields would be expected to increase in a particular district from adopting HYV crops.[5] The intuitive argument for validity is that areas with greater pre-existing suitability for the the particular crops (rice, wheat, and maize; later also sorghum and pearl millet) for which HYV varieties became available should adopt HYVs faster for reasons unrelated to future shocks to the farm size distribution. As Gollin et al. (2021) put it, "[c]ountries that were heavily dependent on rice or wheat agriculture received an earlier—and potentially stronger—boost from the Green Revolution than those that relied on root crops or on other cereal grains, such as barley, sorghum, and millet." This same argument applies to districts within a large country such as India.

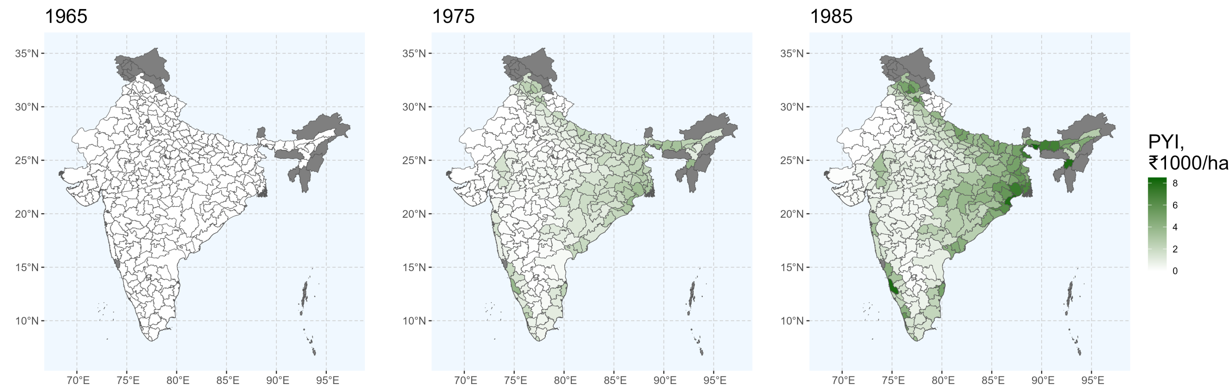
The instrument is constructed by interacting "exposure" to each of the five crops for which HYV areas are recorded[6] with a time trend giving the number of years since the HYV version of that crop was released (an exogenous driver of adoption), and then summing across crops. "Exposure" for a particular crop captures how much overall yields in the district would increase if farmers switched to using the HYV version of that crop while maintaining their existing crop mix. It is calculated by multiplying the initial area share of the crop by the corresponding location-crop-specific potential yield gains from switching from "low" to "high" inputs, according to the FAO's Global Agro-Ecological Zones database (IIASA, 2023). The GAEZ estimates combine characteristics of the soil, terrain, and climate to estimate the maximum possible yields of individual crops under different input conditions for gridded locations around the globe. Since they reflect theoretical models of crop growth rather than estimates from actual yields, they can be used as proxy for potential yield increases for particular crops that is not influenced by the endogenous choice of HYV adoption or the distribution of farmland (as is done in Bustos et al. (2016), Nunn and Qian (2011), Costinot et al. (2016), and Moscona (2019)).

The GAEZ potential yield estimates for each crop can be constructed making different assumptions about the varieties and techniques being used. Under "low input" conditions "[p]roduction is based on the use of *traditional cultivars* [...], labor intensive techniques, and no application of nutrients, no use of chemicals for pest and disease control and minimum conservation measures" (emphasis mine). Under high inputs, "[p]roduction is based on *improved or high yielding varieties*, is fully mechanized with low

[5]This is similar to the instrument used in Moscona (2019).

[6]Wheat, rice, maize, sorghum, and pearl millet.

Figure 1: Instrument Values by Year



Potential Yield Increase, defined in Equation 3

labor intensity and uses optimum applications of nutrients and chemical pest, disease and weed control." In other words, the gap between GAEZ low and high potential yields is a proxy for the difference in potential yields prior to and after the process of agricultural intensification brought about by the Green Revolution.

HYV crops are not adopted overnight; the back-of-the-envelope calculation makes the assumption that the HYV versions of crops are phased in at a constant rate from the date that each one becomes available.[7] Thus differences in PYI across districts are by construction initially small (since few years have elapsed since the new varieties were introduced), but gaps between areas more and less suitable for HYV adoption grow over time. This is illustrated in Figure 1.

After interacting the district-crop exposure measures with their corresponding crop-specific time trends, I sum these predicted yield increases together, aggregating different crops by weighting by their 1993 prices in thousands of rupees (thus the units of exposure are thousands of ₹$_{1993}$ per hectare). The instrument is defined as:

$$PYI_{d,t} = \sum_{c \in w,\, r,\, ma,\, s,\, mi} \underbrace{S_{c,d,1966} \cdot (Y_{H,c,d} - Y_{L,c,d}) \cdot P_{c,1993}}_{''Exposure''} \cdot \underbrace{\frac{\max(t - R_c, 0)}{50}}_{''Shift''} \tag{3}$$

Here $S_{c,d,1966}$ is the initial share of cropland in district d planted with crop c, $Y_{H,c,d}$

[7]The rate itself is unimportant as it simply rescales the instrument.

is the "high-input" yield for crop c in district d (thus the dark green term corresponds to the GAEZ forecast of yield increases), R_c is the year of the initial release for Green Revolution varieties of crop c (taken from Gollin et al. (2021)), and $P_{c,1993}$ is the 1993 price of that crop (from Bhalla and Singh (2001)).

To obtain unbiased estimates from an IV regression, the instrument (PYI) must be both relevant and valid. Relevance is testable; first-stage regressions show reasonably strong predictive power for actual variation in HYV shares. The exclusion restriction requires assuming that the hypothetical increase in total monetary yields from switching from low to high inputs (PYI) is uncorrelated with omitted factors that drive relative changes in the farm size distribution within states over time, and which are not absorbed by the included controls. This is plausible since, as argued by Evenson and Gollin (2003), Gollin et al. (2021), HYV development was exogenous from the perspective of developing countries, and there is thus no reason to believe that areas with higher and lower HYV suitability would have evolved differently had HYVs not become available.[8] This assumption would fail if, for example, differences in inheritance customs across districts within a state were highly correlated with potential yield increases from adopting Green Revolution technologies even conditional on controls. Were that the case, the effect of subdivision through inheritance could be mistakenly attributed to the Green Revolution in these specifications.

This instrument is akin to a "shift-share" or "Bartik" instrument in that it interacts a time-invariant location-industry "share" or "exposure" (the potential money-equivalent yield increases $S_{c,d,1966} \cdot (Y_{H,c,d} - Y_{L,c,d})P_{c,1993}$) with corresponding crop-time "shifts" (years since release of the HYV version of that crop). There are two differences from the typical Bartik instrument, which interacts initial industry shares of locations with industry-wide growth rates to generate the instrument. First, the "industries" are crops and the "exposure" includes not only the initial crop share but also the location-specific impact of the new version of that crop on total crop yields. Using location-specific predicted yields gives this measure of exposure more predictive power than assuming that new varieties have the same impact in all locations with the same initial shares, while leveraging variation that is unlikely to be correlated with omitted drivers of the farm size distribution. Second, the "shifts" are approximated using the time elapsed since the release date of new varieties. This ensures that the instrument inherits its time variation from exoge-

[8] I show in Appendix B that there is no evidence of relative price, yield, or area trends in HYV crops in India in the decade previous to the initial releases.

nous shocks (the release of new crop varieties) rather than the endogenous timing of adoption decisions which might suffer from reverse causality.

In settings where the number of "industries" is small, shift-share instruments depend for their validity on the assumption that the "share" component of the instrument is exogenous (Goldsmith-Pinkham et al., 2020). This assumption would fail if, for example, differences in inheritance customs across districts within a state were highly correlated with predicted yield increases from adopting Green Revolution technologies even conditional on controls. Were that the case, the effect of subdivision through inheritance could be mistakenly attributed to the Green Revolution in these specifications. Some readers may be concerned that the exposures used to generate this instrument are correlated with some driver of the farm size distribution that is not yet controlled for. In particular, readers may be concerned that although the FAO GAEZ yield increases are plausibly exogenous, the initial crop shares may be correlated with unobserved drivers of the farm size distribution that are not included in the controls. Although the lack of pre-period data prevents me from running a placebo test on earlier data,[9] I show in Appendix B that the crops for which HYV crops were released first (wheat, rice, and maize) were not experiencing relative changes in yields, prices, or cropped area in Indian districts prior to the Green Revolution, consistent with evidence in international data from Gollin et al. (2021). Thus there is no clear reason to think that there were different pre-existing trends in agriculture in "HYV crop" areas that would have driven relative changes in the farm size distribution in the absence of the Green Revolution.

Assuming that the identifying assumptions are correct, resulting estimates will capture a LATE which will give relatively greater weight to areas where HYV shares reacted more strongly to PYI. For the benefit of readers concerned about instrument strength, I also report weak-instrument-robust Anderson-Rubin p-values, as implemented by Finlay et al. (2016).

4 Results

To examine the effects of HYV adoption on the farm size distribution, I run separate IVs for each of five farm size categories reported in the Agricultural Census data. I

[9] The 1970 Agricultural Census was India's first agricultural census.

show in Appendix A that total farmland is essentially fixed over time and unaffected by HYV adoption; I thus focus on the impacts on the numbers of farms and share of (unchanged) total farmland. Section 4.1 investigates impacts on the total number of farms in each category (normalized by the starting 1970 total in the district). Section 4.2 reports the effects of HYV adoption on the share of farms in each category (normalized by the contemporaneous rather than 1970 total). Section 4.3 reports the effects of HYV adoption on the share of farmland held by farms in each size category (rather than the share of farms by number in each category).

4.1 Results: Normalized Farm Numbers

In Table 1 I report the results of the main specification separately for five outcome variables: the number of farms in each of the five available size categories, normalized by the total number of farms in the district in 1970. I present the main IV results first, with corresponding OLS and reduced-form results shown in subsequent rows.

The estimated impact of HYV adoption on the number of farms in the marginal size category is large: it suggests that switching to HYV crops for a full season would increase the number of farms in the "marginal" size category by roughly 60% of the starting total number of farms in the district. Multiplying this point estimate by the actual adoption of HYVs over this time period still suggests that HYV adoption led to an increase in the number of marginal-sized farms that was roughly 20% of the initial total number of farms. There are also effects on the next smallest size categories, though the effects are only about a quarter as large. Effects on the two largest size cateories, "medium" and "large" farms, are more precisely estimated and essentially zero. Thus the bulk of the increase in the total number of farms found in Column 6 is attributable to the increase in the number of marginal farms.

4.2 Results: Farm Number Shares

In this section I examine effects on the share of farms in each size category (the shares in each category sum to 1 in each time period). The results suggest that HYV adoption significantly increases the share of farms that are in the "marginal" size category, with

Table 1: Normalized Farm Numbers: Results

	(1)	(2)	(3)	(4)	(5)	(6)
	[0-1)	[1-2)	[2-4)	[4-10)	[10+]	Total
HYV Share (IV)	0.605**	0.153	0.138*	0.054	-0.009	0.932**
	(0.300)	(0.098)	(0.071)	(0.039)	(0.030)	(0.365)
HYV Share (OLS)	0.071*	0.007	0.012	0.009	-0.005	0.088*
	(0.040)	(0.016)	(0.011)	(0.008)	(0.005)	(0.049)
Pot. Yield. Incr. (RF)	1.160**	0.259	0.234**	0.083	-0.014	1.708***
	(0.498)	(0.161)	(0.109)	(0.066)	(0.052)	(0.559)
N	1,125	1,127	1,127	1,114	1,116	1,127
First-stage F	14.28	14.39	14.43	13.78	13.94	14.43
M-P Eff. F	13.88	13.94	13.96	13.45	13.55	13.96
A-R p-value	0.06	0.08	0.03	0.18	0.75	0.01
Addl. Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All specifications include district and state-year fixed effects, linear trends interacted with starting district ag. worker share of pop., literate share, urban share, rural pop. density, rural minor/adult ratio, coastal indicator, and starting number shares, and annual rainfall. IV regressions instrument HYV Share with Potential Yield Increase. Row 3 displays the reduced form relationship between the instrument (in units of ₹1000) and normalized farm numbers. Standard errors (clustered by district) in parentheses. Key: * p<0.1, ** p<0.05, *** p<0.01

Table 2: Share of Farms: Results

	(1)	(2)	(3)	(4)	(5)
	[0-1)	[1-2)	[2-4)	[4-10)	[10+]
HYV Share (IV)	0.216**	-0.064	-0.060	-0.058	-0.034
	(0.086)	(0.047)	(0.037)	(0.040)	(0.022)
HYV Share (OLS)	0.026**	-0.010	-0.010	-0.002	-0.005
	(0.013)	(0.008)	(0.007)	(0.006)	(0.004)
Pot. Yield Incr. (RF)	0.405***	-0.126	-0.114*	-0.107	-0.058
	(0.126)	(0.079)	(0.062)	(0.066)	(0.036)
N	1,128	1,128	1,128	1,128	1,128
First-stage F	14.49	14.49	14.49	14.49	14.49
M-P Eff. F	14.03	14.03	14.03	14.03	14.03
A-R p-value	0.01	0.20	0.10	0.11	0.11
Addl. Controls	Yes	Yes	Yes	Yes	Yes

Notes: All specifications include district and state-year fixed effects, linear trends interacted with starting district ag. worker share of pop., literate share, urban share, rural pop. density, rural minor/adult ratio, coastal indicator, and starting number shares, and annual rainfall. IV regressions instrument HYV Share with Potential Yield Increase. Row 3 displays the reduced form relationship between the instrument (in units of ₹1000) and farm number shares. Standard errors (clustered by district) in parentheses. Key: * p<0.1, ** p<0.05, *** p<0.01

only small effects on the shares of farms in the larger size categories (which must decline to offset the increase in the marginal share). This is again a large effect on the distribution: given that the average district in the sample had approximately 40% of its farms in the "marginal" size category in 1970, these point estimates, when multiplied by the actual change in HYV share ($\approx$0.3) experienced by a typical district would suggest that there was a roughly 7 percentage point increase in the share of "marginal" farms (18% of their initial share) as the result of HYV adoption between 1970 and 1990. Of course, the IV captures a LATE that may differ significantly from the average treatment effect; the exercise merely demonstrates that the effects are potentially economically significant.

4.3 Results: Farmland Shares

In Table 3 I report the results of the main specification separately for five outcome variables: the shares of farmland in each of the five farm size categories. This is the same as in Section 4.2, except that the outcome variable is now a share of the number of farms, rather than the share of farmland, and the controls use starting land shares rather than starting number shares.

Table 3: Share of Farmland: Results

	(1)	(2)	(3)	(4)	(5)
	[0-1)	[1-2)	[2-4)	[4-10)	[10+]
HYV Share (IV)	0.126***	-0.011	-0.017	-0.035	-0.064
	(0.048)	(0.039)	(0.030)	(0.050)	(0.051)
HYV Share (OLS)	0.019**	-0.007	-0.000	0.007	-0.018*
	(0.008)	(0.007)	(0.008)	(0.009)	(0.011)
Pot. Yield Incr. (RF)	0.293***	-0.028	-0.040	-0.088	-0.137
	(0.093)	(0.081)	(0.063)	(0.104)	(0.106)
N	930	930	930	930	930
First-stage F	20.28	20.28	20.28	20.28	20.28
M-P Eff. F	18.45	18.45	18.45	18.45	18.45
A-R p-value	0.00	0.79	0.57	0.46	0.20
Addl. Controls	Yes	Yes	Yes	Yes	Yes

Notes: All specifications include district and state-year fixed effects, linear trends interacted with starting district ag. worker share of pop., literate share, urban share, rural pop. density, rural minor/adult ratio, coastal indicator, and starting land shares, and annual rainfall. IV regressions instrument HYV Share with Potential Yield Increase. Row 3 displays the reduced form relationship between the instrument (in units of ₹1000) and farmland share. Standard errors (clustered by district) in parentheses. Key: * $p<0.1$, ** $p<0.05$, *** $p<0.01$

The estimated effect on the share of farmland in "marginal" farms is statistically significant and meaningful; it suggests that full adoption of HYVs for one season would increase the share of farmland held in marginal-size farms by roughly 13 percentage points, more than doubling the starting share in the sample (10%). Even given that the increase in the HYV share in sample districts was only about 30 percentage points (0.3) on average over the period 1970-1990, this still suggests that a back-of-the-envelope calculation (ignoring treatment effect heterogeneity) would attribute a roughly 4 p.p. increase in the share of farmland held in marginal farms to the Green Revolution. The share of farmland in other size categories must necessarily decline as the marginal share increases, but the effects are smaller and hence not statistically significant. The estimates suggest that the area under small farms grew meaningfully at the expense of small amounts of farmland in larger size categories, although the estimates for each of the larger categories are not precise enough to rule out zero change in the given size bin.

Together, these results suggest that the onset of the Green Revolution increased the total number of farms in India by increasing the number of small farms, particularly those in the smallest category, "marginal" farms. As a result, the total number of farms increased, with marginal farms making up an increasing share of the total. The share of farmland in that category also increased, a change made possible by small declines across the other size categories.

Why is the entirety of the change in the farm size distribution concentrated at the lower end? In the following section I describe a simple model of farmland inheritance and show that the impact of an increase in agricultural productivity is to increase the number of very small farms.

5 Model

In this section, I build a highly parsimonious model of changes in the farm size distribution in India. It seeks to explain the following stylized facts from the earlier empirical sections:

1. HYV adoption raises the total number of farms.
2. This increase is driven by the smallest farms.
3. There is no impact on the number of farms in the largest categories.

The key insight of the model is that when the farmland distribution changes through inheritance (not sale) the only way a productivity shock affects the evolution of farm sizes is by changing decisions about whether or not to subdivide existing farms. By making farming more competitive with non-agricultural employment, an increase in agricultural productivity lowers the floor of the farm size distribution. This makes possible the further subdivision of small farms into yet smaller holdings that previously had insufficient scale to compete with the outside option. Before describing the model, it is important to lay the groundwork for its main assumptions by describing the institutions governing farmland in India in the 1970s.

5.1 Context

There are two key pieces of institutional knowledge necessary for thinking about the evolution of farmland in this context.

First, land transactions outside the household were very rare in India in this time period. Between the 1999 and 2008 rounds of the nationally representative REDS sample, less than 3% of farmers bought or sold land (Foster and Rosenzweig, 2011); in the ICRISAT VLS from 2009-2014 only 0.74% of all plot observations involved a purchase of land (Foster and Rosenzweig, 2017). Due to lack of clear legal title to land, very high transaction costs (Mearns (1999) notes that "[i]t is widely acknowledged that the costs involved in carrying out land transactions in India are enormous"), and potentially household-specific land productivity (see Rosenzweig and Wolpin (1985)), land transfers outside the household were uncommon in this setting. As noted in Chen (2017), this situation is widespread in developing countries and distorts the occupational choices of the agents unless the distribution of farmland is precisely the optimal one given their skill levels (which does not seem to be the case, see Chen et al. (2023) or Aragón et al. (2022)). Thus in practice the division of farmland is determined almost entirely by inheritance: as Foster and Rosenzweig (2002) note, in the nationwide rural ARIS/REDS panel between 1971 and 1982 "weighted by the size of the changes in landholdings, partition accounted for over 90% of the decline in average household landholdings. Clearly household division, rather than market transactions, plays a dominant role in the evolution of landholdings over time."

Second, the norms and laws governing inheritance among Hindus, who constitute a large majority of farmers in India, stipulated that farmland was to be equally divided among heirs, though with possible renegotiation among them. Under the Hindu Succession Act of 1956, when a Hindu head of household dies, his coparceners (typically the male heirs) can either keep a farm as joint undivided property or force an equal division of the land; high negotiation costs may limit households to those two options. This legislation formalized the traditional rules governing inheritance de facto. Ballabh and Walker (1986) describe the typical process of land inheritance in the ICRISAT villages as follows: "When the joint family splits or the father dies, the sons negotiate on the subdivision of the property. [...] If the sons cannot achieve unanimity on a fair division, the extreme form of land subdivision occurs where equal shares of each plot are allocated to each male heir." This tradition remains in effect today; in a recent survey in Gujarat, 97% of respondents report that the inheritance rule can be described as equal shares to all sons (or, more rarely, equal shares to all siblings), with 93% saying that this equal division is not influenced by the education or job prospects of the children; this norm of equal division is a highly accurate predictor of actual land inheritance (Fernando, 2022). (Foster and Rosenzweig (2002) also use these inheritance rules in their model of agricultural household division.)

This institutional context guides the model, which will feature evolution in landholdings driven by changes in inheritance rather than optimal sales.

5.2 Setup

There are two periods in the model, and two sectors, agriculture and non-agriculture. In the initial period there is a continuum of farming households with heterogeneous endowments of land holdings l. Households face a sectoral choice: they can work in agriculture, earning lifetime income $v(l)$, or dispose of their land and work in non-agriculture, earning $w(l)$, which includes any income from disposing of their farmland. Thus the decision to remain in or leave agriculture is determined by the ratio $v(l)/w(l)$, which is assumed to be weakly increasing: giving a household more farmland does not make farming less attractive relative to non-agriculture.[10] This situation is likely in contexts lacking well-functioning land markets; for example, consider a model where $w(l) = w(0) + kv(l)$ with $k < 1$: land can be sold, but only at a discount on its value to the seller. This modeling assumption is also found in Foster and Rosenzweig (2007): "migrants give up claims on agricultural profits in the event that they leave" their rural villages, since empirically "increases in landholding decrease the probability of exiting the village".

5.3 Period 1

As a result of these assumptions, there is a minimal farm size l^* where $v(l^*)/w(l^*) = 1$ which is the floor of the initial farm size distribution: $F_0(l^*) = 0$. This is illustrated as

[10] I also make the technical assumptions that $\lim_{l\to\infty} v(l)/w(l) > 1$ so that there is some farm size above which it is optimal to remain in agriculture, and $v(0)/w(0) < 1$, so that there are some non-farming households in the initial period.

the intersection between the blue $v(l)/w(l)$ curve and the horizontal line at 1 on the left side of Fig. 2.

At the end of the Period 1, the initial household head retires and is replaced by a measure $\alpha > 1$ of heirs. I treat the ratio α as fixed for two reasons. First, the population of adult heirs available in Period 2 (approximately 1990) is largely predetermined by fertility decisions made eighteen years before (1972 or earlier); for the Green Revolution to significantly impact this ratio, large relative fertility impacts would have needed to emerge almost immediately (1966-72). Second, I show in Appendix F that there do not appear to be strong effects of HYV adoption on rural fertility or population growth over this time period.

5.4 Period 2

At the start of Period 2, each initial household's heirs face a sectoral choice, which they make collectively. They can each accept an equal $1/\alpha$ share of the family farm, each earning lifetime income $v(l^0/\alpha)$, or they can leave the farm undivided in the hands of a measure-one subset of the heirs and send the remaining $\alpha - 1$ measure of heirs into non-agriculture.

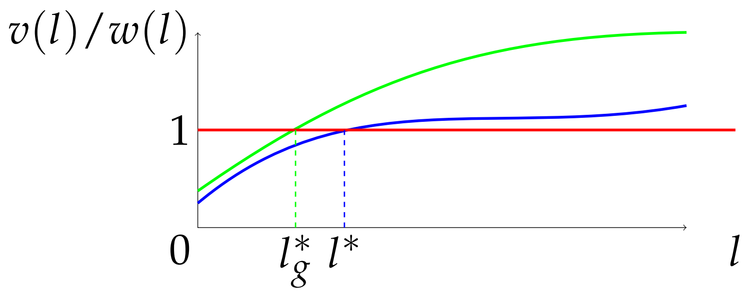
The heirs of sufficiently large farms clearly prefer to remain in agriculture: for high enough l^0, $v(l^0/\alpha)/w(l^0/\alpha) > 1$. But because proposed farms are smaller than existing farms, there is also a measure of farms in the proposal distribution that are below the size threshold at which farmers exit agriculture. These heirs are in households where $l^0 \geq l^*$ but $l^0/\alpha < l^*$; they are at the lowest end of the farm size distribution. Rather than each accepting lifetime income $v(l^0/\alpha) < w(l^0/\alpha)$ for the household as a whole, the heirs make a Pareto-improving agreement to leave the family farm undivided in the hands of a measure-one subset of heirs (who earn $v(l^0) > v(l^0/\alpha)$) and send the remaining heirs into non-agriculture (where they earn $w(l^0/\alpha) > v(l^0/\alpha)$).

5.5 Green Revolution

How does the Green Revolution affect this process? Suppose that at the end of the initial period, before the proposal distribution is made, the introduction of HYVs increases

$v(l)$ by increasing the profitability of farming without affecting $w(l)$. As a result of the growing relative attractiveness of agriculture, there is a new, lower floor for the farm size distribution l_g^*. The impact of the Green Revolution is the difference in the Period 2 farm size distribution with vs. without this lowering of the farm size floor.

Figure 2: Increasing Relative Productivity Lowers Farm Size Floor



As a result of this lower floor, some households which would previously have maintained the family farm undivided now find it optimal for all heirs, not only a subset, to remain in agriculture, and subdivide the family farm. Farms above αl^* are large enough that their heirs will be above the farm size floor even without the Green Revolution; thus subdivision at the upper end of the distribution is not affected. However, farms between αl_g^* and αl^*, which were previously too small to subdivide, are now productive enough that all heirs find it optimal to remain in agriculture; they divide into α times as many farms arrayed between l_g^* and l^*, below what was previously the floor of the size distribution.

Thus, the Green Revolution increases the total number of farms entirely through an increase at the bottom end of the distribution, without affecting the evolution of large farms, as in my empirical estimates.

5.6 Background Farmland Evolution

In the above model, changes in the farm size distribution occur entirely through subdivision of farmland in inheritance, which leads to farmland traveling down the size ladder as it is subdivided among heirs. Are the actual time dynamics of the farm size distribution consistent with this hypothesized process of division? Since total farm area is essentially constant (as shown in Appendix A), all new farms result from either the division of larger farms or the consolidation of more numerous smaller holdings. If these rates of division and consolidation are approximately constant, it should be possible to model next period's farm size distribution as a function of the present farm size distribution. The resulting matrix of coefficients will be akin to the Markov transition probabilities of a square meter of farmland currently in category i transitioning into category j through division or consolidation. Such a relationship is not causal, but represents a statistical pattern that a realistic model should match. Here I present the results of a set of these descriptive OLS regressions, which follow the specification

$$Y_{l,d,t} = \alpha_{s(d)} + \gamma_{l,t} + \sum_{l \in L} \gamma_l Y_{l,d,t-5} + X_{d,1961}\nu_l + \epsilon_{d,t} \tag{4}$$

where $Y_{l,d,t}$ is the number of farms in category l as a share of the total number of farms in that district in 1970. District fixed effects cannot be included (they would cause short-panel Nickell (1981) bias), but I control for state and year fixed effects and include the same controls as in the main specification (without time trends, as this regression models a single time period). The variables of interest are the coefficients on the five-year lags of the farm size distribution. In the table below, rows give the coefficients on the lagged number of farms in each category, while columns give the category being predicted. I omit stars from this table to focus on the magnitudes of the coefficients.

The matrix of coefficients has diagonal elements close to 1, indicating that the number of farms in a given category is, perhaps unsurprisingly, highly persistent over a given five-year period. There are also economically meaningful coefficients just below the diagonal: the MSE-minimizing forecast of farms in a given size category is higher when there are more farms in the next higher farm size category in the previous period. This statistical relationship predicts most of the variation in the farm size distribution, especially for the largest size cateories. Meanwhile elements far from the diagonal are generally of negligible magnitudes: having a high number of large-sized farms does not predict the creation of numerous small farms, nor the reverse.

Together, this evidence points to a farmland evolution process in which farmland mainly travels stepwise down the size ladder, with some fraction of farms in size category k subdividing by a factor of roughly 2 to create two additional farms in category $k-1$ in the next Agricultural Census round and the rest remaining undivided, as in my

Table 4: Share of Farms: Autoregressive OLS Model

	[0-1]/Tot70	[1-2]/Tot70	[2-4]/Tot70	[4-10]/Tot70	[10+]/Tot70
5 Yr Lag:[0-1]/Tot70	1.06	0.01	-0.03	-0.01	-0.01
	(0.03)	(0.01)	(0.01)	(0.01)	(0.00)
5 Yr Lag:[1-2]/Tot70	0.11	0.75	-0.05	-0.09	-0.01
	(0.10)	(0.06)	(0.05)	(0.03)	(0.02)
5 Yr Lag:[2-4]/Tot70	-0.10	0.46	1.00	0.11	0.03
	(0.12)	(0.08)	(0.08)	(0.05)	(0.04)
5 Yr Lag:[4-10]/Tot70	0.00	-0.01	0.16	0.84	-0.14
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
5 Yr Lag:[10+]/Tot70	-0.04	-0.01	-0.06	0.23	1.09
	(0.03)	(0.03)	(0.03)	(0.06)	(0.04)
N	896	898	898	885	887
R^2	0.88	0.84	0.92	0.95	0.97
Addl Controls	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions include state-year fixed effects, starting district ag. worker share of pop., literate share, urban share, rural pop. density, rural minor/adult ratio, and coastal indicator, and annual rainfall. Standard errors (clustered by district) in parentheses. Significance stars omitted.

inheritance model.

6 Modeling Exercise

Do these empirical estimates of the impact of HYV adoption agree with the predictions of existing models? What were the aggregate (general equilibrium) effects of the Green Revolution on the farm size distribution? To answer those questions, I calibrate the model of Chen (2017) to 1970s India and run two exercises in the calibrated model. First, I show that the partial equilibrium implications of the calibrated model align well with my empirical estimates: an agricultural TFP increase sized to match my estimated effects on the total number of farms has impacts that are similarly concentrated at the bottom end of the distribution, and is of a plausible magnitude. Second, I close the model and feed in the same shock to agricultural productivity. In this closed economy model, relative price effects outweigh the effects on agricultural productivity on farm profitability, significantly reducing the overall number of farms.

6.1 Model Overview

I give a brief overview of the model; for a full description see Chen (2017). The model is a one-period heterogeneous-agent model (impacts are therefore comparative statics). The economy consists of two sectors: agriculture, where production is performed by individual households, and non-agriculture, which consists of a representative firm. The production functions in these sectors are:

Agricultural production (household):

$$\pi(z_a, l) = \max_k p \cdot A\kappa z_a \left[\omega k^\eta + (1-\omega)(z_a l)^\eta\right]^{\frac{\gamma}{\eta}} - rk$$

Non-agricultural production (representative firm):

$$\pi(K, L) = AK^\alpha N^{1-\alpha} - rK - wN, \quad N = \int_{s(i)=n} z_n^i di$$

Households have Stone-Geary preferences with a subsistence requirement for agricultural output c_a that generates non-homothetic demand for food at low income levels:

Household utility:

$$u(c) = \phi \ln(c_a - \bar{c}) + (1 - \phi) \ln(c_n)$$

Households are heterogeneous in their agricultural and non-agricultural skill and their endowment of untitled land. Agricultural and non-agricultural productivities z_a and z_n are distributed Weibull(ζ_a) and Weibull(ζ_n), with their joint distribution governed by a Frank copula with parameter ρ. Untitled land is distributed

$$\ln l^i = \beta_0 + \beta_1 \ln z_a^i + \epsilon^i$$

Note that the relationship between agricultural skill and farm size is not necessarily optimal and includes noise.

Households face a sectoral choice between working in agriculture or non-agriculture; they supply one unit of labor inelastically to their chosen sector. In order to match the large differences in labor productivity between agriculture and non-agriculture, the model features a non-monetary wedge ξ that prevents workers from leaving agriculture unless their gain in income is sufficiently high.

Households remain in agriculture if:

$$\pi(z_a, l) > (1 - \xi) \cdot w \cdot z_n$$

Land belonging to households who choose to leave agriculture is rebated to the remaining farmers proportionally to their existing holdings; in equilibrium all occupational decisions are rational given the equilibrium landholdings.

In partial equilibrium, capital and occupations are chosen optimally and all farmland is distributed to farmers given the price of agricultural output p and the rental rate of capital r (non-agricultural output is the numeraire). In general equilibrium, p and r adjust to clear markets for agricultural and non-agricultural output and capital.

Figure 3: Calibrated Parameters

Name	Value	Justification
A	1	Normalized
κ	1	Normalized
η	0.24	Capital-land elasticity of 1.32 (Adamopoulos & Restuccia 2014)
γ	0.54	Labor share in ag. 0.46 (Gollin et al. 2014)
α	0.33	$\frac{1}{3}$ capital share in non-ag
ϕ	0.005	Long-run ag. employment share 0.05%
L	1.59	Aggregate land endowment per household, India 1970
ρ	2.24	Spearman correlation 0.35 between z_a, z_n (Lagakos & Waugh 2013)

Figure 4: Estimated Parameters

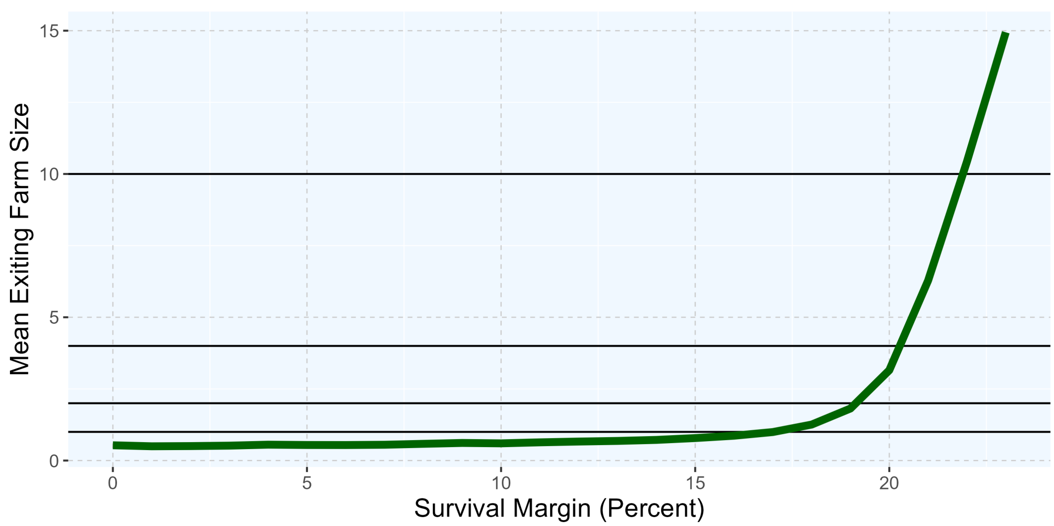
Name	Model	Target	Target Description	Source
ζ_n	0.355	0.339	Non-ag hhold Gini	Bhalla (2003) (urban households)
ζ_a	1.087	1.108	Var. of log farm output	ARIS 1970
$\bar{c}$	0.681	0.683	Ag. emp. share	1970 Pop. Census via IDD
ω	0.268	0.270	Land share	Chen (2017)
K	1.198	1.190	Capital-Output Ratio	Calculated from PWT v. 6.3
σ_e	1.280	1.280	Var. of log farm size	Estimated from Ag Census
β_1	-0.04	-0.06	Corr. of log skill, farm size	VDSA 1975
ξ	0.413	0.417	Ag. Share of GDP	Ministry of Agriculture

6.2 Calibration

The model features 16 parameters. 8 are calibrated based on prior estimates in the literature (Table 3); the remaining 8 are jointly estimated by finding parameters that make the closed economy match eight corresponding data moments. I follow the same procedure as Chen (2017), except that I target data moments corresponding to 1970s India rather than present-day Malawi (I also update the total land endowment, a calibrated parameter). The calibrated model matches most of the targeted moments quite closely; see Table 4.

The calibrated Chen model shares the implication of the simple model that the response of farms to changes in agricultural productivity is driven almost entirely by extremely small farms. In Figure 5 I summarize the distribution of a measure of closeness to exit: how much would agricultural productivity κ have to decline to push a given farm household to make the opposite sectoral choice? As can be seen, for all but extremely large changes in agricultural productivity, the mean farm size of the exiting farms is below 1 hectare.

Figure 5: Mean Exiting Farm Size by Output Shock



6.3 Simulation

The calibrated model captures the main implication of my simple model, that changes in agricultural productivity crowd in additional small farms with little effect further up the distribution. But are the empirical estimates of the Green Revolution's impacts plausible in the calibrated model, in terms of magnitude and shape? To answer that question requires performing a comparison in the model that mirrors the comparison made by the estimator: how does the farm size distribution in a typical district in 1990 differ from the one an initially identical district that experienced a smaller agricultural productivity increase?

This comparison is a partial equilibrium comparison performed in a world where the Green Revolution did happen; a representative district in 1990 differs from a representative district in 1970 as a result of these changes, which also affect the equilibrium prices of agricultural output and for capital (taken as given at the district level). I thus update four parameters to mimic changes to the Indian economy between 1970 and 1990.

I increase the number of households by the predicted increase in the adult population based on the 1971 Population Census and the total supply of capital by carrying forward the average 1957-1966 real GDP growth rate and investment share of GDP in the Penn World Tables. I also increase non-agricultural TFP A and agricultural productivity $A\kappa$ by extrapolating the sectoral TFP growth rates for the 1980s in India KLEMS to the full time period. I re-solve the model with this updated parameterization to find the impact of these changes on the national economy.[11]

I then examine the comparative statics between a representative district in 1990 (a small open economy that is a microcosm of the larger closed economy) and the same district with a lower level of agricultural productivity (taking 1990 equilibrium prices as given). What size gap in agricultural productivity corresponds to the difference in the total number of farms estimated in Column 6 of Table 1? When exposed to a productivity increase sized to match the estimated impact on the total, does the shape correspond to the estimates in the preceding columns?

Generating an impact on the total number of farms that matches Column 6 of Table 1 requires agricultural productivity in the treated and untreated district to differ by a factor of 1.43 (or productivity growth to differ by 1.8 percentage points per year); this is 87% of the agricultural TFP growth rate observed in India KLEMS in the 1980s. This is likely an over-estimate of the share of agricultural TFP growth over this period that can be ascribed to the Green Revolution; however the IV estimates capture a LATE which may differ from the impact on a perfectly representative district. The necessary increase in agricultural TFP is at least of a plausible magnitude.

Does the shape of the impacts match those in Table 1? In Table 5 I show that the overall shape of the impacts matches my empirical estimates: the vast majority of the impact is driven by additional farms in the 0-1 hectare range, with small impacts on the next two categories and negligible changes at the top of the distribution.

The empirical estimates and the comparative statics from this experiment both repre-

[11] Approximating the Indian economy as closed is reasonable in this time period. As noted by the FAO (FAO, 2000), "Until 1991, India followed an inward-looking development strategy with a trade regime characterized by quantitative restrictions, licensing and high tariffs. As a result, domestic markets were virtually insulated from changes in world market prices."

Table 5: Empirical vs. Model Results: Share of 1970 Farm Total

Category	Estimate	Experiment
Marginal	0.18	0.25
Small	0.05	0.04
Semi-Medium	0.04	0.01
Medium	0.02	-0.01
Large	0.00	-0.01
Total	0.28	0.28

sent partial equilibrium differences; districts within the same state face the same prices of capital and agricultural output and any differences between them driven by the Green Revolution do not include the impact of price changes required to clear markets at the national level.

To estimate the aggregate impact of the Green Revolution on the farm size distribution at the national level, I run the same experiment run at the district level at the national level, comparing economies which resemble the 1990 calibrated economy but differ in agricultural productivity by a factor of 1.43. This exercise suggests that the aggregate effects of the Green Revolution on the farm size distribution are larger in magnitude and in the opposite direction of the partial equilibrium effects: the Green Revolution in aggregate reduced the total number of farms by 44% of their initial total, with 91% of this impact concentrated in farms of less than one hectare. While the Green Revolution made agriculture relatively profitable in the areas it most heavily impacted, thus retaining extremely small farms, at the aggregate level it put downward pressure on food prices that had strong countervailing effects. Thus in aggregate the Green Revolution reduced the subdivision of farmland that would otherwise have occurred as the result of increased population pressure on the food supply in an environment where farmland was in approximately fixed supply.

7 Conclusions

Using an instrumental variables design in a district panel, I find that HYV adoption significantly increases the relative number of very small farms and the share of total farm land that they encompass, with much smaller impacts on the rest of the farm size distribution. I demonstrate in a simple model that in a context where inheritance is the primary mechanism of farmland evolution, agricultural productivity shocks impact the small but not the large end of the farm size distribution. I then calibrate an existing model of the distribution of untitled farmland to 1970s India and show that the calibrated model has the same qualitative predictions for the impacts of agricultural productivity on the farm size distribution at the district level. Using the calibrated model to map between district-level estimates and aggregate effects, I find that the increase in agricultural productivity that best replicates my district-level estimates causes impacts of the opposite sign at the aggregate level: relative to the counterfactual, the Green Revolution lowered the total number of farms in India by roughly 44% of its 1970 total, significantly slowing the decline in average farm sizes.

References

Adamopoulos, T. and D. Restuccia (2014): "The Size Distribution of Farms and International Productivity Differences," *The American Economic Review*, 104, 1667–1697, publisher: American Economic Association.

——— (2020): "Land Reform and Productivity: A Quantitative Analysis with Micro Data," *American Economic Journal: Macroeconomics*, 12, 1–39.

Aragón, F. M., D. Restuccia, and J. P. Rud (2022): "Are small farms really more productive than large farms?" *Food Policy*, 106, 102168.

Ballabh, V. and T. Walker (1986): "Land Fragmentation, Subdivision, and Consolidation in India's Dry Semi-Arid Tropics (SAT)," .

Bandyopadhyay, D. (1986): "Land Reforms in India: An Analysis," *Economic and Political Weekly*, 21, A50–A56, publisher: Economic and Political Weekly.

Barrett, C. B., M. F. Bellemare, and J. Y. Hou (2010): "Reconsidering Conventional Explanations of the Inverse Productivity–Size Relationship," *World Development*, 38, 88–97.

Bhalla, G. and G. Singh (2001): *Indian Agriculture: Four Decades of Development*, Sage Publications.

Bharadwaj, P., J. Fenske, N. Kala, and R. A. Mirza (2020): "The Green revolution and infant mortality in India," *Journal of Health Economics*, 71, 102314.

Brulé, R. (2010): "Changes in India's Property Rights Regime and the Implications for Improved Gender Parity: Evidence from Regression Discontinuity Analysis of Panel Data for 17 Indian States," .

Bustos, P., B. Caprettini, and J. Ponticelli (2016): "Agricultural Productivity and Structural Transformation: Evidence from Brazil," *American Economic Review*, 106, 1320–1365.

Chen, C. (2017): "Untitled Land, Occupational Choice, and Agricultural Productivity," *AMERICAN ECONOMIC JOURNAL-MACROECONOMICS*, 9, 91–121, num Pages: 31 Place: Nashville Publisher: Amer Economic Assoc Web of Science ID: WOS:000411828400003.

Chen, C., D. Restuccia, and R. Santaeulàlia-Llopis (2023): "Land Misallocation and Productivity," *American Economic Journal: Macroeconomics*, 15, 441–465.

Costinot, A., D. Donaldson, and C. Smith (2016): "Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from 1.7 Million Fields around the World," *Journal of Political Economy*.

Deininger, K., S. Jin, Y. Liu, and S. K. Singh (2016): "Can Labor Market Imperfections

Explain Changes in the Inverse Farm Size–Productivity Relationship?" Working Paper.

Eastwood, R., M. Lipton, and A. Newell (2010): "Chapter 65 Farm Size," in *Handbook of Agricultural Economics*, Elsevier, vol. 4, 3323–3397.

Evenson, R. E. and D. Gollin (2003): "Assessing the Impact of the Green Revolution, 1960 to 2000," *Science*, 300, 758–762, publisher: American Association for the Advancement of Science.

FAO (2000): "AGRICULTURE, TRADE AND FOOD: COUNTRY CASE STUDIES - Volume II," .

Feder, G., R. E. Just, and D. Zilberman (1985): "Adoption of Agricultural Innovations in Developing Countries: A Survey," *Economic Development and Cultural Change*, 33, 255–298, publisher: University of Chicago Press.

Feder, G. and G. T. O'Mara (1981): "Farm Size and the Diffusion of Green Revolution Technology," *Economic Development and Cultural Change*, 30, 59–76, publisher: University of Chicago Press.

Fenske, J., B. Gupta, and C. Neumann (2022): "Missing Women in Colonial India," .

Fernando, A. N. (2022): "Shackled to the Soil? Inherited Land, Birth Order, and Labor Mobility," *Journal of Human Resources*, 57, 491–524.

Finlay, K., L. Magnusson, and M. E. Schaffer (2016): "WEAKIV: Stata module to perform weak-instrument-robust tests and confidence intervals for instrumental-variable (IV) estimation of linear, probit and tobit models," *Statistical Software Components*, publisher: Boston College Department of Economics.

Foster, A. D. and M. R. Rosenzweig (1996): "Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution," *The American Economic Review*, 86, 931–953, publisher: American Economic Association.

——— (2002): "Household Division and Rural Economic Growth," *The Review of Economic Studies*, 69, 839–869, publisher: [Oxford University Press, Review of Economic Studies, Ltd.].

——— (2007): "Chapter 47 Economic Development and the Decline of Agricultural Employment," in *Handbook of Development Economics*, ed. by T. P. Schultz and J. A. Strauss, Elsevier, vol. 4, 3051–3083.

——— (2011): "Are Indian Farms too Small? Mechanization, Agency Cost, and Farm Efficiency." .

——— (2017): "Are There Too Many Farms in the World? Labor-Market Transaction Costs, Machine Capacities and Optimal Farm Size (Economic Growth Center)," 80.

——— (2022): "Are There Too Many Farms in the World? Labor Market Transaction Costs, Machine Capacities, and Optimal Farm Size," *Journal of Political Economy*, 130, 636–680, publisher: The University of Chicago Press.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020): "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 110, 2586–2624.

Gollin, D., C. W. Hansen, and A. M. Wingender (2021): "Two Blades of Grass: The Impact of the Green Revolution," *Journal of Political Economy*, 129, 2344–2384, publisher: The University of Chicago Press.

IIASA, F. a. (2023): "Global Agro Ecological Zones," .

Just, R. E. and D. Zilberman (1983): "Stochastic Structure, Farm Size and Technology Adoption in Developing Agriculture," *Oxford Economic Papers*, 35, 307–328, publisher: Oxford University Press.

Lagakos, D. and M. E. Waugh (2013): "Selection, Agriculture, and Cross-Country Productivity Differences," *American Economic Review*, 103, 948–980.

Lowder, S. K., J. Skoet, and T. Raney (2016): "The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide," *World Development*, 87, 16–29.

Martin, W. and D. Mitra (2001): "Productivity Growth and Convergence in Agriculture versus Manufacturing," *Economic Development and Cultural Change*, 49, 403–422.

Matsuyama, K. (1992): "Agricultural productivity, comparative advantage, and economic growth," *Journal of Economic Theory*, 58, 317–334.

McArthur, J. W. and G. C. McCord (2017): "Fertilizing growth: Agricultural inputs and their effects in economic development," *Journal of Development Economics*, 127, 133–152.

Mearns, R. (1999): *Access to Land in Rural India*, Policy Research Working Papers, The World Bank.

Moscona, J. (2019): "Agricultural Development and Structural Change, Within and Across Countries," Working Paper.

Nickell, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417–1426, publisher: [Wiley, Econometric Society].

Nunn, N. and N. Qian (2011): "The Potato's Contribution to Population and Urbanization: Evidence From A Historical Experiment*," *The Quarterly Journal of Economics*, 126, 593–650.

Rosenzweig, M. R. and K. I. Wolpin (1985): "Specific Experience, Household Structure, and Intergenerational Transfers: Farm Family Land and Labor Arrangements in Developing Countries," *The Quarterly Journal of Economics*, 100, 961–987, publisher: Oxford University Press.

Sanghi, A., K. K. Kumar, and J. W. J. McKinsey (1998): "India Agriculture and Climate

Data Set," .

Sircar, A. K. and S. Pal (2014): "What is preventing Women From Inheriting Land? A Study of the Implementation of the Hindu Succession (Amendment) Act 2005 in Three States in India," Tech. rep., Landesa Rural Development Institutet.

Tandel, V., A. Dutta, S. Gandhi, and A. Narayanan (2023): "Women's right to property and the child quantity-quality trade-off: evidence from India," *Journal of Population Economics*.

Vanneman, R. and D. Barnes (2000): "Indian District Data, 1961-1991: machine-readable data file and codebook." .

von der Goltz, J., A. Dar, R. Fishman, N. D. Mueller, P. Barnwal, and G. C. McCord (2020): "Health Impacts of the Green Revolution: Evidence from 600,000 births across the Developing World," *Journal of Health Economics*, 74, 102373.

Willmott, C. J. and K. Matsuura (2018): "University of Delaware Terrestrial Precipitation v. 5.01," .

Yamauchi, F. (2021): "CHANGING FARM SIZE AND AGRICULTURAL PRODUCTIVITY IN," Tech. rep., International Food Policy Research Institute.

A Appendix A: Total Land, First Stages, OLS Versions, Robustness

A.1 Effects of HYV Adoption on Total Farmed Land

Table 6: Total Ag Census Farmland (1970-1990)

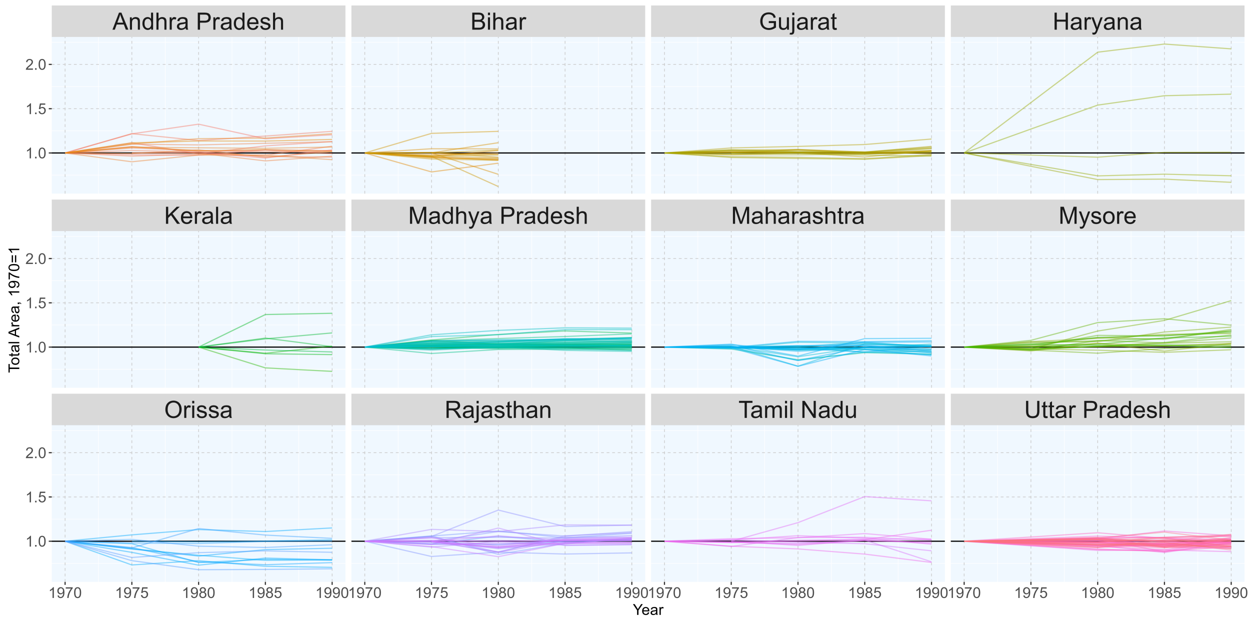
	(1) Log Tot. Land	(2) Log Tot. Land	(3) Log Tot. Land	(4) Log Tot. Land
Year	0.001	0.004		
	(0.000)	(0.003)		
HYV Share		-0.228	0.139	0.049
		(0.188)	(0.174)	(0.145)
N	1,049	1,028	1,028	930
First-stage F		20.45	22.41	20.03
M-P Eff. F		11.22	15.92	20.03
Add'l Controls	No	No	No	Yes

S.E.s (clustered by district) in parentheses.

* p<0.1, ** p<0.05, *** p<0.01

In Column 1, I report the results of a simple OLS regression of the log of total farmland, regressed on a district fixed effect (so that variation is within-district over time), rainfall, and the year. The point estimate on Year is effectively zero, suggesting no average change in log farmland over this time period. In the second column, I report the results of estimating the same equation but adding in HYVShare as an explanatory variable and instrumenting for it with PYI. The estimate on HYVShare is far from statistically significant despite a fairly strong first stage. In Column 3 I replace the single Year variable with state-year fixed effects, to see whether an effect of HYVShare is being masked by state-level trends in total farmland that counteract the effect of HYVs. Estimates of the impact of HYVs are once again extremely imprecise despite reasonably strong instrumentation. Column 4 repeats the specification of Column 3 but with the Additional Controls described above. As there is no clear evidence that HYV adoption changes overall agricultural land, or that overall farmland is changing in important ways over this time period, I focus on its effects on the division of farmland.

Figure 6: Total Farmland Essentially Constant



A.2 First Stages

Here I present the first stages of the IV regressions presented in Section 4.

Table 7: Farm Land Share Spec.: First Stage

	HYV Share	
PYI, ₹1000	2.197***	(0.488)
N	930	
First-stage F	20.28	
M-P Eff. F	18.45	
Addl. Controls	Yes	
State-Year FEs	Yes	

S.E.s (clustered by district) in parentheses.

* p<0.1, ** p<0.05, *** p<0.01

Table 8: Farm Number Share Spec.: First Stage

	HYV Share	
PYI, ₹1000	1.827***	(0.483)
N	1,125	
First-stage F	14.28	
M-P Eff. F	12.77	
Addl. Controls	Yes	
State-Year FEs	Yes	

S.E.s (clustered by district) in parentheses.

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

A.3 OLS Versions

Here I show the OLS equivalents of the main results.

A.3.1 Farmland Shares: OLS

	10+ Ha.	0-1 Ha.	1-2 Ha.	2-4 Ha.	4-10 Ha.	10+ Ha.
HYV Share (IV)	-0.064	0.019**	-0.007	-0.000	0.007	-0.018*
	(0.051)	(0.008)	(0.007)	(0.008)	(0.009)	(0.011)
State-Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Addl. Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	930	930	930	930	930	930

A.3.2 Farm Number Shares: OLS

	0-1 Ha.	1-2 Ha.	2-4 Ha.	4-10 Ha.	10+ Ha.
HYV Share	0.026**	-0.010	-0.010	-0.002	-0.005
	(0.013)	(0.008)	(0.007)	(0.006)	(0.004)
State-Year FEs	Yes	Yes	Yes	Yes	Yes
Addl. Controls	Yes	Yes	Yes	Yes	Yes
N	1,128	1,128	1,128	1,128	1,128

A.3.3 Normalized Farm Numbers: OLS

	[0-1]/Tot70	[1-2]/Tot70	[2-4]/Tot70	[4-10]/Tot70	[10+]/Tot70	Tot/Tot70
HYV Share	0.071*	0.007	0.012	0.009	-0.005	0.088*
	(0.040)	(0.016)	(0.011)	(0.008)	(0.005)	(0.049)
State-Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Addl. Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	1,125	1,127	1,127	1,114	1,116	1,127

A.4 Robustness to Dropping Additional Controls

The main specifications feature a vector of "Additional Controls" discussed in Section 3. Here I show the results of specifications without those additional controls. The instrumentation is somewhat weaker, leading to less precise estimates, but the overall pattern of estimated effects is largely unchanged.

A.4.1 Farmland Share: Without Additional Controls

	0-1 Ha.	1-2 Ha.	2-4 Ha.	4-10 Ha.	10+ Ha.
HYV Share	0.118***	-0.043	-0.026	-0.018	-0.030
	(0.046)	(0.050)	(0.032)	(0.051)	(0.053)
N	931	931	931	931	931
First-stage F	13.35	13.35	13.35	13.35	13.35
M-P Eff. F	11.13	11.13	11.13	11.13	11.13
A-R p-value	0.01	0.40	0.41	0.72	0.57
Addl Controls	Yes	Yes	Yes	Yes	Yes

S.E.s (clustered by district and state-year) in parentheses.

* p<0.1, ** p<0.05, *** p<0.01

A.4.2 Farm Number Shares: Without Add'l Controls

	0-1 Ha.	1-2 Ha.	2-4 Ha.	4-10 Ha.	10+ Ha.	LnTotNum
HYV Share	0.193*	-0.064	-0.049	-0.055	-0.024	0.416
	(0.106)	(0.063)	(0.040)	(0.041)	(0.020)	(0.348)
N	1,138	1,138	1,138	1,138	1,138	1,138
First-stage F	11.39	11.39	11.39	11.39	11.39	11.39
M-P Eff. F	11.76	11.76	11.76	11.76	11.76	11.76
A-R p-value	0.06	0.32	0.22	0.15	0.20	0.23
Addl Controls	Yes	Yes	Yes	Yes	Yes	Yes

S.E.s (clustered by district and state-year) in parentheses.

* p<0.1, ** p<0.05, *** p<0.01

A.4.3 Normalized Farm Numbers: Without Add'l Controls

	[0-1]/Tot70	[1-2]/Tot70	[2-4]/Tot70	[4-10]/Tot70	[10+]/Tot70	Tot/Tot70
HYV Share	0.419	0.150	0.149*	0.035	-0.009	0.746
	(0.465)	(0.106)	(0.083)	(0.041)	(0.026)	(0.552)
N	1,135	1,137	1,137	1,124	1,126	1,137
First-stage F	11.18	11.33	11.35	10.20	10.41	11.35
M-P Eff. F	11.59	11.69	11.71	10.56	10.72	11.71
A-R p-value	0.37	0.12	0.04	0.40	0.71	0.17
Addl Controls	Yes	Yes	Yes	Yes	Yes	Yes

S.E.s (clustered by district and state-year) in parentheses.

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

A.5 Robustness to Dropping State-Year Fixed Effects

Agriculture is a state subject in the Indian system of federalism, and states pursue a variety of agricultural policies that might conceivably impact the farm size distribution and be correlated (negatively or positively) with PYI. There is also likely regional variation in the growth in demand for non-agricultural labor beyond what can be controlled for using starting demographics. Since these state policies and regional trends are difficult to comprehensively measure and control for, the preferred specification includes state-year fixed effects so that the estimator is driven by differences in changes between districts in the same state. For the interested reader, I show results from specifications that employ district and year fixed effects rather than district and state-year fixed effects.

The effects on farmland shares are still highly significant and of larger magnitudes than with the state-year fixed effects. However, without state-year fixed effects there are no estimated effects on the farm number shares, a finding that is somewhat in tension with the clear effects on farmland shares. Estimated effects on the number of farms as a share of the starting total are of similar magnitudes to the preferred specification, but lower instrument strength (and hence precision) in this specification renders them no longer statistically significant (with the exception of the effect on the total number of farms, which remains significant at the 10% level).

A.5.1 Farmland Share: No State-Year FEs

	0-1 Ha.	1-2 Ha.	2-4 Ha.	4-10 Ha.	10+ Ha.
HYV Share	0.238***	0.003	0.026	-0.102	-0.165**
	(0.086)	(0.046)	(0.035)	(0.067)	(0.077)
First-stage F	12.50	12.50	12.50	12.50	12.50
M-P Eff. F	12.26	12.26	12.26	12.26	12.26
A-R p-value	0.00	0.96	0.46	0.07	0.01
Addl. Controls	Yes	Yes	Yes	Yes	Yes
N	930	930	930	930	930

A.5.2 Farm Number Shares: No State-Year FEs

	0-1 Ha.	1-2 Ha.	2-4 Ha.	4-10 Ha.	10+ Ha.	LnTotNum
HYV Share	0.150	-0.076*	-0.036	-0.028	-0.009	0.507
	(0.093)	(0.045)	(0.039)	(0.042)	(0.019)	(0.341)
First-stage F	9.05	9.05	9.05	9.05	9.05	9.05
M-P Eff. F	8.66	8.66	8.66	8.66	8.66	8.66
A-R p-value	0.04	0.09	0.32	0.45	0.61	0.06
Addl. Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	1,128	1,128	1,128	1,128	1,128	1,128

A.5.3 Normalized Farm Numbers: No State-Year FEs

	[0-1]/Tot70	[1-2]/Tot70	[2-4]/Tot70	[4-10]/Tot70	[10+]/Tot70	Tot/Tot70
HYV Share	0.621	0.141	0.122*	0.049	0.007	0.906*
	(0.384)	(0.120)	(0.070)	(0.042)	(0.025)	(0.518)
First-stage F	9.05	8.96	9.06	7.82	7.84	9.06
M-P Eff. F	8.65	8.58	8.67	7.67	7.70	8.67
A-R p-value	0.04	0.17	0.03	0.27	0.80	0.02
Addl. Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	1,125	1,127	1,127	1,114	1,116	1,127

B Appendix B: Potential Price/Yield/Area Pre-Trends

In this section I first provide evidence that there were no significant relative yield or price trends for Green Revolution crops prior to the Green Revolution. I hope to persuade readers that although there is no way to demonstrate a lack of pre-trends in the farm size distribution driven by initial crop shares (because sufficiently detailed data before 1970 do not exist), there is no clear reason to believe that the share of land planted with Green Revolution crops was related to increasing or decreasing profitability of agriculture prior to the Green Revolution. Next, I show that the main results remain qualitatively the same even when using an instrument that does not use initial shares (i.e. $S_{c,d,1966} = 1 \; \forall c, d$ in Equation 3) and that the general conclusions remain the same (although some estimates lose significance) when using this instrument and additionally controlling directly for $S_{c,d,1966}$ (interacted with linear time trends, in the linear specification).

B.1 No Pre-Period Yield or Price Trends

In this section, I demonstrate that there is no evidence to support the idea that crops for which HYV varieties were developed during the Green Revolution were on differential yield or price trends prior to the Green Revolution.

B.1.1 Relative Yield Trends

I first present results from Gollin et al. (2021) relevant to this discussion. In that paper, the authors study the impact of exposure to the Green Revolution, constructing their measure of exposure to the Green Revolution by interacting pre-Green Revolution farmland shares of particular crops with modeled yield trends that are crop- but not crop-location-specific. Since the yield increases they study vary only at the crop-time level (not the crop-location-time level), and time fixed effects would absorb all variation in their exposure measure if shares were equal, all variation in their measure of exposure (and hence their identification) stems from the initial crop distribution (the shares rather than the shifts). The authors acknowledge that the variation comes from these initial shares, explaining that "[c]ountries that were heavily dependent on rice or wheat agriculture received an earlier—and potentially stronger—boost from the Green Revolution than those that relied on root crops or on other cereal grains, such as barley, sorghum, and millet." They then attempt to show that there are no pre-trends in yields by estimating the impact of the Green Revolution on crop yields using an event-study design. Their estimating equation (run on international yield data) is:

$$\ln yield_{c,s,t} = \mu_{c,s} + \mu_{s,t} + \sum_{j} \beta_{c,j} \cdot \mathbf{1}(t = R_c + j) + \delta \ln area_{c,s,t} + \epsilon_{c,s,t} \tag{5}$$

Here c is crop, s country, and R_c the year in which a ; area is included as a control to combat the downward bias in estimated yields that would otherwise result from crops being grown on increasingly marginal land as acreage under that crop expands. The result of this estimation is shown below:

Fig. 2 from Gollin, Hansen, Wingender (2022)

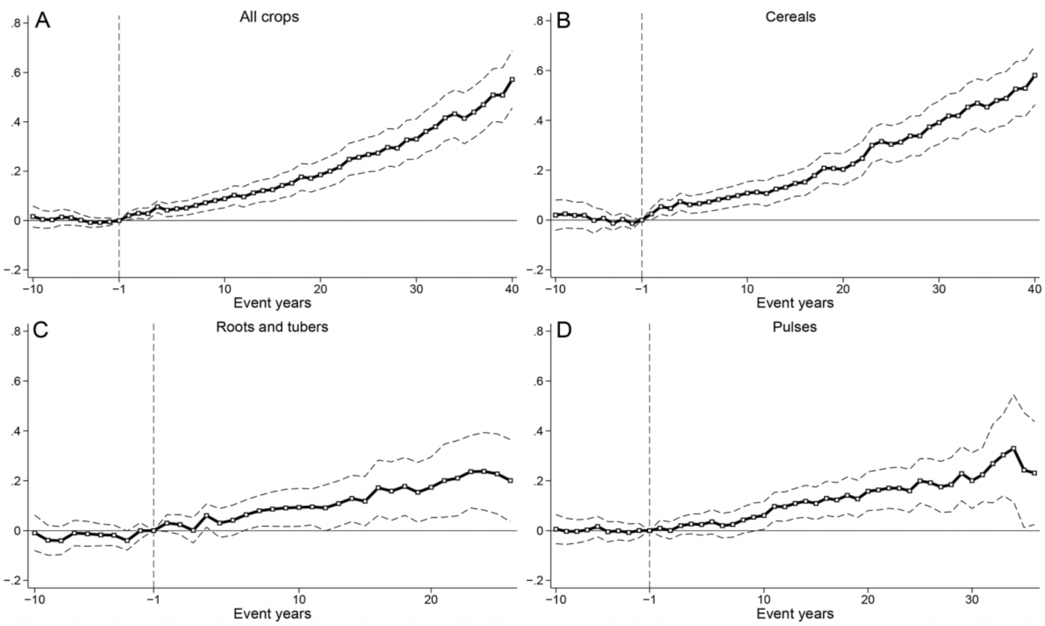


Fig. 2.—Baseline crop-level event-study estimates based on equation (10). The dependent variable is ln yields. The sample period is 1945–2010. In *A*, we show estimates under the assumption that the treatment effect is of the same magnitude across crops. In *B–D*, we report effects by crop type (cereals, roots and tubers, pulses), all from the same regression. Both regressions (i.e., *A* and *B–D*) include controls for ln harvested area, country-by-year fixed effects, and crop-by-country fixed effects. The omitted comparison event-year is the year before an HYV release (vertical line). The dashed curves indicate the 95% confidence bands. Standard errors are clustered at the crop-by-country level.

As can be seen, there are no significant pre-trends in yields of Green Revolution crops prior to the release date of their first Green Revolution high-yielding varieties, either compared to crops overall or compared to crops in their same family (cereals, roots and tubers, or pulses).

B.1.2 Relative Price, Yield, Area Trends in India

Next, I investigate the possibility of relative price trends in 1956-64 (prior to the Green Revolution) using data from Sanghi et al. (1998) sourced from DES wholesale price reports. I run the following specification of prices for crop c in district d at time t, defining a "HYV crop" as one of the three (rice, wheat, and maize) that would have HYV varieties released in 1965-66. I include crop-district fixed effects (absorbing initial differences in yields/prices/areas in 1957) and district-time fixed effects (so that identification is based on relative changes in yields/prices/areas between soon-to-be-HYV crops and other crops within the same district over time). I linearly interpolate through missing values and then limit to crops for which the full 1957-1964 time series is available so that the regression is not biased by changing crop composition of the "HYV" and "non-HYV" categories over time.

$$\ln y_{c,d,t} = \alpha_{c,d} + \gamma_{d,t} + \beta \cdot HYV \times t + \epsilon_{c,d,t} \tag{6}$$

	(1) Log Yield	(2) Log Area	(3) Log Price
Wheat/Rice/Maize=1 × Year	-0.000	-0.000	-0.000
	(-0.65)	(-1.10)	(-0.93)
N	27,162	27,135	29,826
District × Year FEs	Yes	Yes	Yes
District × Crop FEs	Yes	Yes	Yes
Clustering	District	District	District

t-stat in parens. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

As can be seen, there is no evidence of statistically significant relative trends in yields, cropped areas, or prices prior to the Green Revolution. This supports the plausibility of the assumption that the initial shares are exogenous with respect to unobserved drivers of the farm size distribution.

C Appendix C: Arguments Against Land Redistribution and Inheritance Reform as Causes

Are state-level policies an omitted variable driving these results? In my main specifications, I control for state-year fixed effects, so any state-level policy that affects all districts equally does not bias my results. However, readers may be concerned that unobserved variation in two types of policy (land reform and inheritance reform) across districts within a state are driving my results. I argue that neither one is likely to significantly bias my regressions because the de facto impact of these de jure reforms was extremely small.

C.1 Land Reforms

In early post-independence years, there were significant changes to the control of farmland in India. The first and most sweeping change was the "abolition of intermediaries" in the 1950s. Under British rule large areas of land had been granted to zamindaris (tax collectors) in certain regions, and the farmers on that land converted to tenants under a tax farming scheme; zamindari abolition reversed this process, granting ownership rights to permanent tenants. This process was complete by the early 1960s. Subsequently, some variety of "land reform" legislation was passed by most states in India. These reforms came in multiple forms, the most common being a) rules increasing the share of output that sharecropping tenants were to receive, b) "land ceilings" which appropriated from large landowners all land above a given limit and redistributed it to landless households, and c) "land consolidation", in which land was reallocated across landholders in a village with the goal of creating more geographically contiguous landholdings (thought to increase the efficiency of mechanization and irrigation) out of small dispersed parcels, *without* changing the total amount of land held by each farmer. Most of this legislation was also passed before the start of my panel; the only states in my sample with land reform legislation passed during the study period are Haryana (1972), Andhra Pradesh (1973), and Rajasthan (1973) (see Mearns (1999)).

The scale of the effects I find appears to be too large to be explained by the apparently limited direct effect of land reform legislation in India. The implementation of tenancy reforms was in general "weak, non-existent or counterproductive" Mearns (1999), tending if anything to increase the consolidation of operated holdings by making landowners unwilling to lease out land. There is general consensus that state land reform efforts have overall had small direct effects Mearns (1999). Bandyopadhyay (1986) finds that by 1986, less than 2% of cultivated area had been officially declared surplus (less had been distributed). In short, "arable land subject to ceiling-surplus or state redistribution is insignificant by comparison with potential land transfers through inheritance or through the market" Mearns (1999). Some authors believe that consolidations in Haryana, Punjab, and western UP were effective (Ballabh and Walker, 1986) (although without direct evidence), but these programs do not change the total land ownership of the household and thus have no obvious direct effect on the distribution of total farm sizes. Moreover, all these pieces of legislation place limits on ownership rather than operations; it is possible for the ownership distribution to change without a corresponding change in the division of land into farms as defined by management.

C.2 Changes in Inheritance Laws

One key mechanism driving the subdivision of farmland is the separation of joint family property. Under the Hindu Succession Act of 1956, which legally formalized prevailing cultural practices, upon the death of a Hindu head of household joint family property (inherited from a paternal line ancestor) becomes joint property of "coparceners", who can own it jointly or petition to split it into separate holdings. In the absence of a will this division applies to all property, whether joint family or personal. This is likely the situation in the majority of inheritances; even in the 21st century field work suggests that 20% or less of inheritances are governed by a formal will (Brulé, 2010). Field work by Ballabh and Walker (1986) suggests that this proposed equal division of land among coparceners is indeed how land was typically divided in rural India in the 1970s and 1980s. Under the 1956 law, only male heirs were deemed to be "coparceners." Expansion of this right to women might be expected to accelerate the subdivision of farmland if it in fact leads to greater operational division of previously jointly managed joint holdings.

However, there are strong reasons to suspect that this is not a major factor in the evolution of farm sizes in the period of study (1970-1990). Only three states, Kerala, Andhra Pradesh, and Tamil Nadu, allowed women to be coparceners prior to 1990.[12] In the latter two states the reforms gave coparcenary rights only to *unmarried* daughters (Tandel et al., 2023) in the last few years of the sample (1986 and 1989). Since the median age of marriage for women was below 18 at this time, women who were eligible for coparcenary rights by remaining unmarried until their father's death were likely rare. Moreover, even when women have a de jure right to inherited property, prevailing custom in rural areas often made this right difficult to act on in practice (Mearns, 1999). Even after the 2005 Supreme Court ruling that gave coparcenary rights to women everywhere (regardless of married status), field work in northern India suggests that less than 1% of women even ask for a portion of land inheritance (Brulé, 2010), while a (non-representative) survey in

[12] In fact Kerala abolished the concept of joint family property entirely.

AP, MP, and Bihar as late as 2011 found that the amount women expected to inherit was on average roughly 10% of the equal division by then stipulated by the law (Sircar and Pal, 2014). Thus it is unlikely that changes in inheritance laws prior to 1990 are driving these results.

D Appendix D: Data Appendix

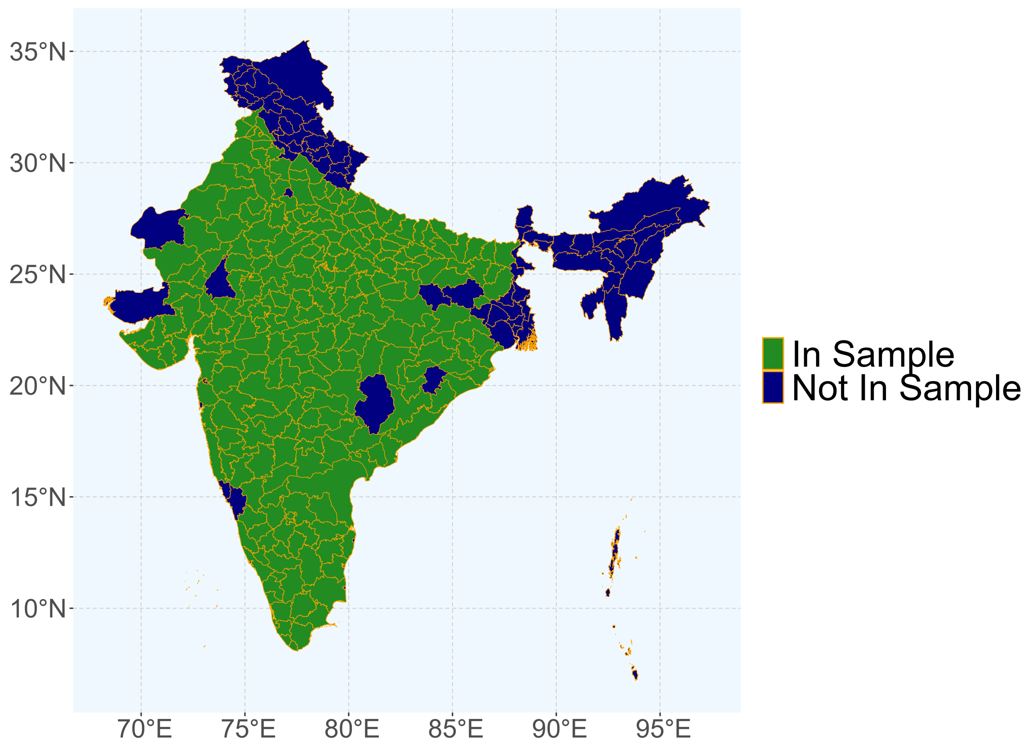
Data on both the farm size distribution and HYV adoption at the district level over time come from the ICRISAT's District Level Dataset (DLD), which consolidates data on Indian districts from multiple sources to the level of districts as they existed in approximately 1966.[13] Data cover (up to) 313 Indian districts from 1966 to 2016. The dataset covers the major agricultural states of India (see Fig. 7). It is missing data for Jammu & Kashmir, the northeastern states, districts which consist entirely of major cities and thus have no agricultural data (Delhi, Kolkata, etc.), and areas which were not governed by India in 1966 (Goa, Puducherry, etc.). I also exclude West Bengal, where district-level data on HYV rice (the main local HYV crop) simply rescale the overall state-level adoption data, rendering within-state variation uninformative, as well as eight extremely rugged districts where farmland is less than 25% of district area.

Data on the number and size of farms in the Agricultural Census are collected in Phase I of the Agricultural Census from the village patwari, who maintains village-level records of land managed by each household.[14] The concept of a farm measured in the Agricultural Census is the "Operational Holding", which is defined as "all land which is used wholly or partly for agricultural production and is operated as one technical unit by one person alone or with others without regard to title, legal form, size or location". This distinction between holding and ownership is important, as is the method of collection. Because legal rights to land were often outdated or unclear in this setting, statistics on de jure land ownership may give a misleading picture of de facto farm sizes. Similarly, legal ceilings on landholdings (despite being seldom enforced) could motivate farmers to under-report the true size of their holdings; data maintained and submitted by the

[13] This dataset is also sometimes referred to as the Village Dynamics in South Asia (VDSA) Meso Database (https://vdsa.icrisat.org/vdsa-database.aspx).

[14] Except in Punjab and the states (West Bengal, Kerala, Odisha, and the Northeast States) that were part of the "Permanent Settlement" during the British colonial period, where the census is conducted as a survey.

Figure 7: Farmland Distribution Sample



village patwari may suffer less from this self-interested misreporting.

Data for other variables comes from several sources. Occasional missing values of the key variables (farm sizes and HYV crop area) in the DLD are patched from the India Agricultural and Climate Dataset (Sanghi et al., 1998) or from the underlying published Agricultural Census reports. Uttar Pradesh is missing HYV shares for 1990; I impute these with 1989 values. Potential crop yields (used in creating the instrument for HYV adoption) are from the Global Agro-Ecological Database (IIASA (2023), discussed in Section 3.1). Demographic variables not present in the DLD are from Vanneman and Barnes (2000), which collects Indian Population Census data from the 1961-1991 rounds. Ruggedness is derived from the CGIAR SRTM 90m DEM. Rainfall is from the University of Delaware Terrestrial Precipitation dataset (Willmott and Matsuura, 2018).

India's districts have proliferated and changed boundaries (and names) over time. The most common occurrence is a split dividing a single district into multiple districts; in rarer cases a new district is formed by merging areas descended from 1961 district "A" with areas descended from 1961 district "B". In the first case I consolidate data for the child districts up to the level of the larger parent district as it existed in 1961 via a crosswalk developed by Fenske et al. (2022); in the latter case I create a "super-district" that combines the 1961 districts "A" and "B". This results in a slightly smaller number of observations but does not require making any assumptions about how population or other variables were distributed within 1961 districts. I refer to these consolidated "super-districts" as "districts" in the following discussion.

E Appendix E

In the main regressions, I control for district fixed effects (which control for all time-invariant characteristics of districts) and for the initial 1970 farm size distribution (so that results are not driven simply by the starting farmland configuration). Nevertheless, readers may be concerned that some unobserved characteristics of districts not absorbed by these controls were causing farm sizes to develop differently in areas with different levels of PYI even before the Green Revolution. If so, there should presumably have been meaningful differences in the farmland distribution in areas with high versus low (future) levels of PYI even at the beginning of the sample. Here I show the results of two cross-sectional regressions that address this concern. I regress the land and number shares in the different farm size categories on the 1990 values of PYI[15], and include state fixed effects (so that comparisons are within states, as in my main regressions) and the standard Additional Controls. I normalize PYI to have a standard deviation of 1, so the coefficient corresponds to a 1 s.d. increase in PYI in 1990.

[15] Recall that PYI is initially zero for all districts and grows over time.

Table 9: Relationship of 1970 land shares with 1990 PYI

	0-1 Ha.	1-2 Ha.	2-4 Ha.	4-10 Ha.	10+ Ha.
1 S.D. 1990 PYI	0.009**	0.014***	0.010*	-0.005	-0.028***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.011)
N	205	205	205	205	205
State FEs	Yes	Yes	Yes	Yes	Yes
Addl Controls	Yes	Yes	Yes	Yes	Yes

S.E.s (clustered by district) in parentheses.

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 10: Relationship of 1970 number shares with 1990 PYI

	0-1 Ha.	1-2 Ha.	2-4 Ha.	4-10 Ha.	10+ Ha.
1 S.D. 1990 PYI	0.031***	0.004	-0.007**	-0.018***	-0.010
	(0.010)	(0.004)	(0.003)	(0.006)	(0.006)
N	237	237	237	237	237
State FEs	Yes	Yes	Yes	Yes	Yes
Addl Controls	Yes	Yes	Yes	Yes	Yes

S.E.s (clustered by district) in parentheses.

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

As can be seen, there is a relationship between 1990 PYI and the 1970 farm size distribution, but it is small in magnitude relative to the effects of the Green Revolution, with a 1 s.d. increase in PYI moving the shares in each category by no more than about 3 p.p.. Thus the relationship between future PYI and the initial farm size distribution, while statistically strong, is economically small in magnitude and seems unlikely to explain the effects found in the main regressions.

F Appendix F

Here I investigate whether the effects on the farm size distribution are driven by effects on population growth or fertility. I examine the effects of HYV adoption on three outcomes, using the same specification as in Section 4.1. In Column 1, I examine the effect of HYV adoption on a proxy for fertility, defined as the ratio of children age 0-4 to females age 15-34. Effects are fairly small and far from significant. In Column 2, I examine the effects of HYV adoption on the log of the rural population of the district. The point estimate is positive but quite small, and cannot be statistically distinguished from zero. In Column 3 I examine the impact of HYV adoption on the ratio of children to adults in rural areas in the district. The point estimate is small and again not statistically significant. Together, these results suggest that changes in fertility or population growth are not a major mechanism linking HYV adoption to changes in the number of farms.

Table 11: Effects of HYV Adoption on Fertility and Rural Population

	FertProxy	LnRrPop	RrChild:Adult
HYV Share	0.047	0.019	-0.079
	(0.127)	(0.132)	(0.109)
First-stage F	11.90	12.80	12.80
M-P Eff. F	12.21	13.08	13.08
A-R p-value	0.72	0.88	0.45
Addl. Controls	Yes	Yes	Yes
N	679	700	700

Column 1 shows the impact of HYV share on a proxy for rural fertility rates defined as children age 0-4 per female age 15-34 in the district. Column 2 shows the impact of HYV Share on log rural population. Column 3 shows the impact on the ratio of children to adults in rural areas. The specification is the same as in Section 4.1, but includes only the years 1970, 1980, and 1990. Census population figures are taken from the following year's Population Census.