Towns and Rural Land Inequality in India

Manaswini Rao, Juan Eberhard, Prashant Bharadwaj*

September 30, 2022

Abstract

Using the universe of land records from a large state in India, we document three empirical facts on rural land holding inequality at the village-level: 1) inequality is higher close to urban areas and decreases with distance, 2) this is due to fewer medium-sized farms (i.e. more small and large farms near urban areas), and 3) the distance to urban area-land holding inequality relationship depends on the size of the urban area - larger the urban area, greater the inequality close to such towns. A simple model where individual farmers face financial frictions, a U-shaped agriculture production function linking land size and farm productivity, and a significant urban opportunity cost of farm production, explains these patterns. While medium-sized farmers exit agriculture and large farmers consolidate, financial and land market frictions are key factors behind the preponderance of small farms even near towns.

JEL Codes: Economic Development (O1), Urban, Rural Regional, and Transportation Analysis (O18)

*Author affiliations: Rao and Bharadwaj - UC San Diego, Eberhad - Universidad Andres Bello, Facultad de Economia y Negocios. Contact: manaswini.rao@gmail.com, juan.eberhard@gmail.com, prbhardwaj@ucsd.edu. We thank seminar participants at Barcelona GSE Summer Forum 2022 Applied Micro Workshop, Columbia, Johns Hopkins U, NEUDC 2021, NYU, NU Singapore, Pacdev 2022, UConn, University of Kent, UPitt, ISI-Delhi, UC San Diego, UC Irvine, UEA London 2022, U Wisconsin-Madison. Data acknowledgements: Karthik Muralidharan and Siddharth George.

1 Introduction

Why do individuals in rural areas of developing countries persist in small scale farming? Why are small farmers unable to expand, and why do they not migrate to urban areas that offer a higher wage? These questions have been the focus of research in Development Economics for decades and are intimately connected with questions regarding the existence of poverty traps, market frictions, and structural transformation.[1] A key empirical fact that has been proposed in the literature that is consistent with models of poverty traps is that there are many more small farms in developing countries than in developed countries (Rigg, Salamanca, and Thompson 2016). Large scale farming is more efficient than small scale farming (especially since large farms can take advantage of mechanization); hence, the existence of many small farms is often taken as *prima facie* evidence of the existence of poverty traps, which might arise due to various market frictions and frictions related to rural-urban migration in developing countries (Foster and Rosenzweig 2017; Lagakos 2020). One of the primary motivating reasons behind this research is that if poverty traps are the main reason why people persist in states of under development (like small scale farming), then certain types of policy interventions - often dubbed as "big push" interventions - can enable structural transformations in the economy that raise welfare and place countries on a path of positive growth.[2]

In this paper we aim to provide new insight by examining empirical data on land holding inequality. While land inequality in rural areas might be of interest in and of themselves to answer questions regarding who owns land and the role of land concentration or land redistribution policies, we first focus on describing some of the basic spatial patterns of rural land inequality in relation to urban areas. This itself is a novel exercise since land records for the universe of land holders is generally not available in rural areas of developing countries and any survey-related datasets are generally not representative of the population-level land-holding distribution. We describe three features of spatial patterns in land inequality in rural India and argue that these patterns, when viewed through the lens of our simple model, shed light on the existence of preponderance of small farms, the importance of financial frictions, and its association with urbanization. We want to be upfront that this is not a paper about

[1] This research includes some of the seminal theoretical work in this space such as Rosenstein-Rodan (1961), Murphy, Shleifer, and Vishny (1989), Banerjee and Newman (1993), and Banerjee and Newman (1994). The theoretical work has lead to a large empirical literature attempting to find evidence for poverty traps (see Kraay and McKenzie 2014 for an excellent review, Gollin, Lagakos, and Waugh 2014 who argue about importance of agricultural productivity differences and income inequality, and Balboni, Bandiera, Burgess, Ghatak, and Heil 2020 as a recent example of empirical work on poverty traps).

[2] There are certainly several critics of this theory as well, for example, see Easterly (2006).

whether urbanization *causes* these land holding patterns to emerge or vice-versa; instead, the paper first describes and then seeks to understand through a model the set of market conditions and frictions that could generate such patterns. We also want to be clear that there could be other models that might generate our observed patterns. Our goal in putting forth this model was to focus on the role of financial frictions which has received tremendous attention in the literature.

We compute land inequality as a Gini coefficient at the village level. An important feature of Indian agriculture, especially in the state we examine, is that a vast majority of households are small, land owning, farmers (approximately 80%); therefore, the Gini makes sense as an intuitive measure of how land is held at the village level. Using the Gini as the basis for measuring land holding inequality, we highlight three empirical patterns: first, we note that land inequality is higher near towns and the Gini coefficient is lower in villages further away from towns. Second, higher Gini coefficient near towns is driven by fewer mid-sized farmers (with 5-8 acre plots). In other words, villages near towns have relatively more small *and* large farms (compared to middle sized farms). Third, the Gini-distance relationship is steeper for larger towns compared to smaller towns.

These village-level spatial patterns are consistent with a model where agricultural productivity is a U-shaped function (as in Cornia 1985; Barrett 1996, and more recently revisited by Gaurav and Mishra 2015 and Foster and Rosenzweig 2017) in farm size[3], farmers face financial constraints that decreases with the amount of land owned, and an urban wage[4] premium that dissipates with distance from the city (due to migration-type costs that increase with distance). With these key ingredients of the model, we are able to simulate a data generating process that fit the empirical patterns discussed above. Hence, according to this model, the process of structural transformation (one of the features of this is the act of moving out of agriculture - see Syrquin 1988) is possible for mid-sized farmers, close to towns, but not for small farmers close or far away from towns. This generates the observed distribution of higher land inequality near towns. Our model provides a lens as to why, even in the presence of high wages in towns and cities and other urban advantages, small farmers who live nearby are unlikely to move or expand their farming activities. Financial frictions, therefore, play a crucial role in linking the spatial patterns we observe to the ideas

[3] Large empirical literature had documented negative farm-size relationship with agricultural productivity, which is now being updated to non-monotonic or U-shaped relationship considering the presence of measurement error and heterogeneity in plots and production shocks (see Gollin and Udry 2021).

[4] We model the urban premium as wage but this can potentially be replaced by any other advantages offered by urban areas such as access to markets, better institutional quality near urban areas facilitating sale of land, etc.

of sub-scale agriculture.

Our paper is related to many different strands of the literature. First, it adds to the recent empirical literature on structural transformation and spatial frictions (Heise and Porzio 2021), including rural-urban migration (Young 2013; Bryan, Chowdhury, and Mobarak 2014; Bazzi, Gaduh, Rothenberg, and Wong 2016), by bringing attention to selective exit from agriculture based on asset (land) ownership rather than skills. Our patterns support a model where financial constraints bind for small farmers (as also highlighted by Bryan, Chowdhury, and Mobarak 2014) but earning relatively better returns for mid-sized farmers (as also modeled by Lagakos, Mobarak, and Waugh 2021 requiring low relative productivity in rural areas for migration). The patterns we demonstrate is supported not just by variations in the land inequality Gini index but also by the relative shares of mid-sized farmers with respect to small and large farmers.

Second, our paper models the relationship between land inequality and urban influence independent of the initial conditions, acknowledging that that these initial conditions could matter. In this regard, our paper extends the work by Bardhan, Luca, Mookherjee, and Pino (2014), where the authors examine the effect of land tenancy reform (leading to land redistribution) in West Bengal and subsequent demographic transitions generating land inequality through land subdivisions from inheritences. However, we do not find evidence of inequality driven by greater sub-divisions of particularly mid-sized farms close to towns. Further, the universality of land ownership data in our setting allows us to examine the change in land holding distributions over a span of 7 years, documenting increasing inequality closer to towns over time. That is, the observed patterns are not likely to in steady state.

Third, it complements the vast literature on urban development (such as in China documented by Kahn, Sun, Wu, and Zheng 2018, and in Africa documented by Henderson and Kriticos 2018). Saiz (2010) and Hsieh and Moretti (2019) note that land restrictions generate scarcity in urban housing supply that drive up prices and wages, affecting labor allocation across space and subsequently, economic growth. Our work sheds light on the role of urban opportunity on land distribution in surrounding rural areas, opening a new area for additional research in determining the causal relationships in both directions in light of Glaeser and Henderson (2017) and Henderson and Turner (2020) - whether the ability to consolidate land in rural areas leads to higher urban development in such areas, and whether higher wages in urban areas motivate increasing consolidation in surrounding rural areas.

Finally and importantly, we add to development economics literature that has focused on land ownership patterns from the point of view of tenure security and incentives for investments to improve agricultural productivity (Banerjee, Gertler, and Ghatak 2002; Burchardi, Gulesci, Lerva, and Sulaiman 2019; Adamopoulos and Restuccia 2020) and its relationship to migration (de Janvry, Emerick, Gonzalez-Navarro, and Sadoulet 2015), we document novel correlations between rural land inequality and urbanization. Our research has direct relevance to understanding the existence of poverty traps in developing countries. The most recent work on this suggests an important role for poverty traps in explaining why the poor stay poor (Balboni, Bandiera, Burgess, Ghatak, and Heil 2020).

As more countries move towards digitizing land records, such spatial patterns in land inequality will be important to document. While some countries may exhibit the patterns we describe above, others might not. Understanding these links between land inequality, financial constraints, proximity to towns, and economic growth will be key for thinking about the implications of land regulations and policies. Land regulations such as land ceiling policies often lower land inequality but in the process, create more small farmers. However, these policies are often done *without* regard to the potential for a U-shaped farm size productivity relationship and the presence of financial frictions. These two forces might lead to long run poverty traps after such land redistribution style policies. That said, our paper also suggests that land consolidation policies in the presence of financial frictions might be a heavy-handed way of bringing about structural change. The key take-away is that the preponderance of small farms we document are due to multiple market failures (in our case, for example, it would be financial frictions *and* labor/land market frictions that drive the U-shaped productivity relationship) and therefore solving any one of these failures through interventions might not lead to intended positive effects.

The rest of the paper is as follows. Section 2 describes the data sources, Section 3 documents the key empirical patterns, Section 4 presents a simple model describing the economic processes generating the observed correlations, Section 5 examines the implication of the model in data, Sections 6 and 7 discuss alternative theories and welfare implications, and Section 8 concludes.

2 Data

We exploit multiple data sources - administrative, satellite imagery, and survey-based - from India to empirically identify some key facts on the relationship between rural land inequality and proximity to urban centers. The basic unit of analysis in our data is a village, which is the smallest unit of administration in India. We construct the land inequality measures and exploit other outcomes at the village-level using disaggregated unit-level data, which we discuss in detail in the subsections below. Aggregating data at the village-level still provides thousands of observations at varying distances from a substantial number of towns.

2.1 Land Records

One of the central datasets is the land holding data spanning the universe of farmers in one large state in India. The main fields in this dataset that we use for our analysis include land ownership details - unique ids with basic demographic detail, area owned, and village identifiers. These are based on recent field-level updates to land records to capture the most accurate status of land ownership since they were also used for income transfers under both state and national-level farmer income support programs.

We classify a farmer as a small farmer if they own less than 5 acres of land, as a mid-sized farmer if they own between 5-8 acres of land, and as a large farmer for those owning more than 8 acres of land. This is similar to the categorization followed under the Agricultural Census by the Ministry of Agriculture and Farmer Welfare, Government of India.[5]

2.2 Census Data

In addition to the above mentioned administrative data on the universe of farmers, we also employ both the population census village-level aggregates and socio-economic and caste census micro-data for the particular state. The idea behind using two different sources of *population-level* data is to: (a) verify whether the patterns replicate in either datasets, and (b) address any measurement error issues based on whether land holding is recorded at the farmer-level or the household-level. The latter includes the set of landless households, therefore, the Gini coefficients measures inequality in access to land as a productive asset as well.

2.2.1 Population Census

From the 2011 population census, we use the village-level primary census abstract documenting village population and amenities, including distance from the nearest town in km. We merge this with the town-level census by the nearest town, to obtain the corresponding town population. Note that this nearest town could be located in another district or another state,

[5]The Agricultural Census has a finer classification as follows: (a) marginal farmers as those with below 1 hectare/2.47 acres of land; (b) small farmers between 1-2 hectare or 2.47-5 acres; (c) semi-medium farmer as those with 2-4 hectare or 5-10 acres; (d) medium farmer as those with 4-10 hectare or 10-25 acre; and (e) large farmer as those with over 10 hectare or 25 acres of land. Since the number of farmers with over 10 acres is very small in our data, we chose to use our coarser classification which captures the land size cut-off between small and medium farmers.

altogether. Among the many variables present in the village census abstract and amenities modules, we mainly focus on the following: village-level population; number of households; total geographic, cultivated, irrigated, and non-agricultural areas; primary crop produced; road connectivity via district, state, or national highways; presence of a bank branch; and the name and distance to the nearest town.

2.2.2 Socio-Economic and Caste Census (SECC)

The Government of India administered a comprehensive socio-economic and caste census during the 2011 population census, and collected detailed information on education, occupation, income, and assets (including land) at the household-level.[6] We use this census micro-data for the corresponding state to verify the construction of the land inequality measure generated from the farmer-level land records data as well as to verify the key empirical patterns on the relationship between rural land inequality and proximity to urban areas. Importantly, this dataset captures all households within a village including the landless. Other variables that we examine in our analysis include the extent of farm mechanization by household.

2.3 Satellite Data

We use FAO GAEZ satellite data products to examine agricultural outcomes. First, we check whether the distance pattern we observe is due to any sorting by productivity of land for specific crops (mainly wetland rice and cotton) by examining if there are distance correlations in the potential yields for such crops. Second, we use the yield achievement ratio (yield and production gap) thematic product to examine whether the distance pattern in land holding inequality also has a distance pattern in better agricultural outcomes. The yield achievement ratio is our preferred productivity measure as it reflects the difference between realized yield and "predicted" yield based on soil and climatic suitability of a crop. This likely implies that the productivity is driven by farmers' production choices.

We download these products as raster images from FAO GAEZ v4 data portal. We then overlap these raster images with the population census administrative boundary shapefiles of villages for the state we study, to obtain the mean values of potential yields and yield achievement ratio, respectively, across pixels contained within each village polygon.

[6]Some of this data is even reported at the individual-level but land is mainly reported at the household-level

2.4 Household Panel Survey: IHDS

An important and complementary data source is the pan-India India Human Development Survey (IHDS) datasets, conducted in 2005 (Desai, Vanneman, and National Council of Applied Economic Research 2005) and again in 2012 (Desai, Vanneman, and National Council of Applied Economic Research 2012). This dataset consists of survey data of a representative sample of households across India, using a multi-stage random sampling design: villages (for rural sample) were first randomly selected to form the primary sampling unit (PSU), and within the sampled villages, the surveys were administered to a random sample of households. A critical feature of this dataset is that the sample of households can be tracked across the two waves. We use this dataset to examine life cycle changes at a household-level over time - attrition from the first survey round to the next, share of children, household splits, and land inheritance. In the absence of good quality, disaggregated data on migration, we interpret survey attrition to provide a plausible measure of migration.

2.5 Merging Datasets

We fuzzy merge land records data with census 2011 village amenities data using district, sub-district, and village names using the Hungarian Method, minimizing the string distance between the triple (village name, sub-district name, district name) in the two datasets, following similar procedure as in Norris, Pecenco, and Weaver (2021). The algorithm yields 75% match rate, which we further improve by manually matching the unmatched villages. We match village-level data with the characteristics of its nearest town - mainly population - by merging on the town census id.

2.6 Constructing Land Inequality Measure

Using the universe of land records, we construct measures of land inequality as a Gini index at the village-level. This is as fine grained as one could go given a village is also the basic unit of administration in India. We construct this empirically as:

$$Gini_v = 2 * \frac{1}{N * 100} \sum_{i \in v} (p_i - l_i)$$

where p_i is the percentile rank of farmer i in village v based on the size of their land parcel, and l_i is the cumulative share of land held by all farmers ordered by their percentile rank below i.

2.6.1 Robustness of the constructed measure

We address potential issues with the land records dataset as follows. First, to account for land registered under different household members, we also generate a household-level measure of farmland ownership using the SECC dataset by aggregating land ownership within a household. This, complementary survey-based dataset, also helps account for landless individuals and households in the same village while computing the overall village-level Gini index.

Second, it is possible that farmers owning fewer acres may rent land for cultivation from larger farmers. However, tenant farming is typically absent in states that had *ryotwari* system of land revenue during British India (Banerjee and Iyer 2005), such as in our context. On the other hand, this is important in erstwhile *zamindari* areas, where subsequent tenancy reforms led to improved tenure security for tenant farmers (for e.g., Operation Barga of West Bengal discussed in Banerjee, Gertler, and Ghatak 2002). While we don't observe lease details in the land records data, we are mainly interested in the distribution of land ownership as an asset to examine our research question, that could have important ramifications on the state of economic development through wealth inequality.

2.7 Summary Statistics

2.7.1 Land Inequality

A summary of the land records data as in Figure A1 shows that there are too many small and marginal farmers, who own less than their share of land within these villages. An overwhelming majority - over 85 percent - of rural land owner-cultivators are small farmers, with average land holding of about 2 acres. Mid-sized farmers, with average land holding of 6 acres, constitute about 10 percent of farmers whereas only 5 percent of all farmers are large farmers with average land holding of 12 acres. Despite being a small share, large farmers own close to a fifth of all land. On the other hand, mid-sized farmers constitute less than 20 percent of all farmers in a majority of the villages and in over 10 percent of the villages, there are no mid-sized farmers.

An absence of land concentration or more equal distribution should imply that each group of farmers own the same share of land as their share among all farmers. Indeed, the ratio between farm size at the 99th percentile and that at the 25th percentile shows a wide variation in our data (see Panel A Figure A2). Similarly, land inequality - measured as a village-level Gini index as described above - shows substantial variation (Panel B Figure A2). Using farmer-level data often underestimates inequality since the data does not include landless households. On the other hand, survey data incorporates self-reported data on landholding size by all households within a village, suggesting much higher rates of inequality (Panel B Figure A2).

3 Key Empirical Patterns

We document three key empirical patterns relating rural land inequality and proximity to towns, which we detail below. First, land concentration is higher in villages close to towns, which stabilizes after a certain distance (roughly 35-40 km from towns). Second, villages close to towns have fewer mid-sized farmers relative to either small or large farmers. Finally, the relationship between distance to town and land concentration becomes steeper as town size increases.

3.1 Negative Correlation Between Rural Land Inequality and Distance to Nearest Town

We find that landholding inequality is strongly and negatively correlated with the distance of the village to its nearest urban center. Specifically, we run the following regression specification as shown in Equation 1 to document the distance-inequality correlation in a non-parametric form:

$$Gini_{v(t)m} = \delta_t + \delta_m + \sum_{j \neq 35-40 \text{ km}} \beta_j D_{jv(t)m} + \epsilon_{v(t)m} \tag{1}$$

where $D_{jv(t)m}$ is a dummy that takes value 1 when village $v(t)$ in the vicinity of town t is within a specific 5 km distance bin. The leave-out group is villages that are in the $35-40$ km distance bin from the nearest town and the farthest bin includes all distances greater than 75 km. A distance of 0-5 km implies that the village is outside the town limits but is within the "peri-urban" area. Since the villages are administered separately from the perspective of revenue and are separate electoral enitities, they are by definition outside the city limits. Note that our choice of the leave-out group is just for the purpose of illustration and we could replace this with other distance bins as the reference group. δ_t and δ_m are nearest town and sub-district fixed effects, respectively. Standard errors are clustered by the nearest town.

Panel A Figure 1 and Table 2 show the distance correlation between rural land inequality and proximity to an urban area. The correlation is strongest in villages within 0-5 kms from a town, which dissipates with distance. The correlation stabilizes after 35-40 km, although villages over 70 kms away from any urban area is relatively more equal than those that are at 35-40 km distance. Table 2 presents the parametric estimates of this relationship modeled using OLS with (Panel A) and without fixed effects (Panel B). The Gini Index is transformed into standard deviation units relative to the "baseline" group of villages that are 35-40 km away. The coefficients imply a close to 0.1 σ increase in land inequality for every 10 km decrease in distance to town.

The observed patterns exist even (strongly) as a binary correlation without the nearest town and sub-district fixed effect or clustering standard error by the nearest town (Panel B Table 2). Given the population-level data and no causal parameter to estimate, we needn't cluster the standard errors per se (Abadie, Athey, Imbens, and Wooldridge 2017). Our intention behind adding these fixed effect is to account for as many time-invariant unobserved variables that could be correlated with both location of a village vis-a-vis an urban area and the landholding distribution within a village. Insofar as even correlations could be under or over-estimated due to omitted variables, we try to account for them in our specification as fixed effects. Further, in order to be more conservative with our inference, we choose to cluster standard errors by the nearest town to account for any spatial correlation between the villages in our data.

3.2 Plot-Size Correlation: Fewer Mid-Sized Farmers Near Towns

A second key fact that we observe is that there are fewer mid-sized farmers or households (those owning 5-8 acres of land) closer to towns relative to small and large farmers. We estimate Equation 1 but using the ratio of the number of mid-sized farmers to the number of small and large farmers, respectively as the dependent variable. Both specifications yield a negative correlation between the share of mid-sized farmers relative to the share of small and large farmers, respectively, and village's distance from its nearest town (Figure 1 and Table 2). Parametric estimates imply that every 10 km increase in distance from town is associated with a 0.65 percentage point increase in the share of mid-sized farmers as a ratio of the share of small farmers and nearly 16 percentage points increase as a ratio of the share of large farmers in a village.

3.3 Distance Correlation Steeper Near Larger Towns

A third fact that we observe is that the downward sloping distance gradient is stronger when the nearest town is a large town with respect to its population (for example, in the top quintile) compared to the distance gradient when the nearest town is a small town (first quintile). We modify the distance specification above to include an interaction term along with the distance bins as shown in Equation 2.

$$
\begin{aligned}
Gini_{v(t)m} &= \delta_t + \delta_m + \sum_{j \neq 35-40 \text{ km}} \gamma_j D_{jv(t)m} \text{ x Large Town}_{v(t)m} + \sum_{j \neq 35-40 \text{ km}} \beta_j D_{jv(t)m} \\
&+ \epsilon_{v(t)m} \qquad (2)
\end{aligned}
$$

The interaction variable, Large $\text{Town}_{v(t)m}$, is a dummy variable that takes a value 1 when the nearest town is in the top quintile of town population distribution. We restrict the sample to villages close to either the first or fifth quintile towns by town population. The coefficients on the interaction terms decreases with increase in distance from the nearest town (as depicted in Figure 1 or Table 2). We also examine this interaction using distance and town population as continuous variables (as shown in Table 2). The interaction is negative and meaningful. Specifically, the land inequality-distance to town correlation becomes more steep and negative around towns with population that is a hundred thousand more than other towns. This implies inequality worsens by 7-10% in villages close to large towns compared to those surrounding small towns (Columns 4 and 5 Table 2).

3.4 Robustness of Observed Empirical Patterns

These observations are robust to a variety of tests. First, the same patterns are observed using a completely different dataset - SECC, one that is administered as a primary survey by the central government's statistical office, and therefore, is a different data generating process than the state government's revenue records. This, census-based dataset, also helps account for land parcels registered under different household members to calculate land inequality at the farm household-level. Further, the Gini index measured using SECC data also captures landless households within a village (See Figure A3).

Second, the empirical patterns remain unaffected after dropping villages surrounding large metropolitan cities, as shown in Figure A4. This ensures that the observed patterns are not driven by outlier urban areas but one that is observed even in the vicinity of medium-sized towns.

Third, the distance correlation is evident in all-India data as well. We use the IHDS all-India sample survey village module data from 2005 to compute the measure of land inequality as well as obtain distance to the nearest town variable. The village module includes kinship (jati)-group level composition by village population and village land ownership in percentage terms. We compute land inequality as the ratio between maximum per capita land ownership and minimum per capital land ownership across the different jati groups in a village. While this is not the same as the Gini index, it captures a measure of inequality in terms of the composition of total village land ownership. Column 1 Table A1 presents this relationship. As noted, distance to nearest town is negatively correlated with land inequality in rural areas.[7]

4 Model

We build a parsimonious model that can describe the observed empirical patterns. Our model posits that urbanization encourages mid-sized farmers to exit cultivation for better wage opportunities in urban areas relative to their returns to agricultural production with their existing land endowment.[8] At the same time, they are not as credit constrained as small farmers to finance their transaction and migration costs. Therefore, our model starts with farmer-level optimization problem that involves maximizing a value function that incorporates the trade-off between returns from investing in agriculture given their current land endowment and their ability to borrow to expand their agricultural land, and net returns from migrating to the nearest urban area, which depend on the characteristics of that town (i.e. urban wages, which in turn, is a function of town population) and the cost of migration (which depends on distance to the town). Next, we aggregate individual farmer decisions over the distribution of farmers within a village to show how aggregating these individual decisions affects village-level land inequality Gini index. It is important to note that the goal of the model is to map individual farmer's constraints to the empirical patterns observed at village level by distance to its nearest town.

We make two simple assumptions: First, agricultural productivity follows a U-shape function with respect to plot-size as in Foster and Rosenzweig (2017). This implies higher productivity among small and large farmers but lower productivity among mid-sized farmers.

[7]Since the IHDS sample is not representative of the village-level land size distribution (specifically, this dataset over-represents larger land-owning households relative to the population-level distribution), we cannot test for the plot-size correlation. Further, this dataset also does not have identifying information on the nearest town to be able to obtain the town population to test the town size correlation.

[8]The model only assumes that higher urban wages may impact the present value of land-holding. This leads to mid-sized farmers having the following two choices: (a) migrate to the nearest town, or (b) become a landless farm laborer due to higher farm labor wages. Thus, the optimal response for mid-sized farmers would be to sell their initial land endowment as continuing to cultivate their own farm is not optimal. We discuss this in greater detail in this section.

Second, financial constraint is a function of plot-size, with a minimum plot-size requirement in order to borrow.[9] The presence of financial constraint implies that a farmer cannot borrow indefinitely to increase their land endowment to achieve economies of scale.

4.1 Farmer-level optimization

Consider a representative farmer with land L with returns $A(L)$, and current debt D with interest rate r, who lives for infinite periods. At any point in time, the farmer faces the decision of continuing with agriculture by expanding their land endowment or permanently migrating to the nearest town. First, we consider their optimization problem with respect to their land-holding as follows:

$$\begin{aligned} V(L, D, P_L) &= \max_{c,N} u(c) + \beta V(L', D') \\ c + rD + T &\leq A(L) \\ L' &= L + N \\ D' &= D + P_L N \\ N &\leq \phi L \end{aligned} \tag{3}$$

where T is the land transaction fixed cost, P_L is the current land price, N is the extent of new land purchased, c is current consumption, and $0 < \phi \leq 1$ is a measure of financial friction (such as collateral requirements). This set up implies that the farmer's value function incorporates current period consumption and the discounted present value of future value from cultivating land that incorporates a potentially larger land endowment (L') and a larger debt level (D') to finance the expansion in their land endowment.

The farmer compares returns from their agricultural enterprise with the returns from migrating to the nearest town after selling their land, which earns them $P_L L - D$ (after paying off their current debt). We denote $M(w, d)$ as the value of permanently migrating to the nearest town (and starting to work in the next period) that is at distance d with urban wage w, or equivalently

$$M(w, d) = \sum_{t=1} \beta^t u(w, d) \tag{4}$$

Therefore, the full problem for a farmer with land L and debt D facing current land

[9] This restriction is introduced using transaction costs for selling and buying land.

prices P_L, who lives near a town at distance d with current urban wage w is

$$F(L, D, P_L, w, d) = \max\{V(L, D, P_L), u(P_L L - D) + M(w, d)\} \tag{5}$$

This set up also implies that the dynamics of the distribution of land-holdings would depend on both the initial land-holdings and the initial debt distribution. In our simulations, we create heterogeneity across villages using random initial distribution of land (that is right-skewed) and debt, and we study how this joint distribution evolves over time.[10]

Figure A5 shows the intuition of this model. The top panels show how the shape of the value function changes when financial frictions are introduced. The top-left panel shows that even though the agricultural production follows a U-shaped function, the value function is not U-shaped but is rather monotonic. This is due to the result that, under no financial frictions, all farmers have optimal land-holdings, therefore, no one has any mid-sized land that corresponds to the trough of the agricultural U-shaped production function. Introducing financial frictions generates non-monotonicity in the value function as shown in the top-right panel. This non-monotonicity is a key property of our model. The bottom-left panel shows that the value function is not a simple mirror of the U-shaped production function, where the trough corresponds to about 8.45 acres of land. How severe are the financial frictions would affect the shape of the value function, and thus, how similar it is to the U-shaped agricultural production function. Finally, the bottom-right panel shows how the farmer compares the agricultural value function that incorporates the U-shaped production function with the value function associated with migration to the nearest town that depends on urban wages and distance to the nearest town. This implies that there exists a lower and upper threshold of land-holding - L^* and $L^{**} > L^*$, such that

$$u(P_L L - D) + M(w, d) = V(L^*, D) \tag{6}$$

$$u(P_L L - D) + M(w, d) = V(L^{**}, D) \tag{7}$$

The above condition provides a range of land-sizes for which migration provides better returns than remaining in agriculture. This corresponds to mid-sized farmers who find selling their land and migrating to the nearby town as more attractive. Moreover, the attractiveness of towns increases as their population increases and consequently, wages (larger towns have agglomeration economies that increases the returns to labor and other factors of production).

[10] In this set up, the distribution of land is changing while converging to the long-run steady state.

These changes in towns' pull are represented by vertical movements of the value function of migration $M(w, d)$.

Finally, we assume that the farmer takes current urban wage as given even though migration from other nearby villages would likely affect urban wages in equilibrium. We believe that this is a plausible assumption since in order to forecast changes in urban wages due to migration of other farmers from other villages in the vicinity, a farmer would need information on the land and debt distribution in every village. Given this large information requirement, it is unlikely that a farmer has this level of detailed information. Relaxing this assumption would mean that the farmer would account for the expected value of future urban wage rather than consider a deterministic current wage. This would give rise to predictions that are qualitatively similar to those from our model above.

4.2 Simulating Village-Level Land Inequality

In order to use the model to study the empirical patterns described earlier, we must aggregate the optimal solution for each farmer at the village level subject to a fixed supply of land within a village, and to examine if the Gini coefficient generated using the optimal decisions of all the farmers in every village presents the empirical patterns discussed. We do this through simulations.

We start with 1500 villages with 500 farmers in each. Every farmer is endowed with some land and debt initially, with the village-level distributions following random normal distribution with a right skew. The villages are scattered around 75 towns with random normal distribution of population (with parameters $\mu = 100,000$ and $\sigma = 30,000$) and distance (with parameters $\mu = 25$ km and $\sigma = 15$ km). This ensures that the initial conditions imply no correlation between Gini coefficient at the village-level and the distance between the village and its nearest town.

We assume standard functional forms for utility (log), returns to migration ($w = (s/1000)^{0.75} - d^{0.5}$ where s is the nearest town population and d is the distance between the village and the town) farmers. We consider the relative price of land within a village and financial constraints to be exogenous parameters.

Farmers optimize their value function discussed in the previous section, based on which, either they either increase their land holdings, or maintain their initial land endowment, or sell their land in order to migrate to the nearest town. These decisions would change the land and debt distribution in each village, and consequently affect the Gini coefficient at the village-level as well as the town's size.

An important caveat is that our model is based on partial equilibrium conditions arising from the farmer optimization problem and we do not explicitly model general equilibrium, steady-state conditions. There are two key reasons for this. First, modeling general equilibrium in this context is computationally complex requiring distributions of land and debt for each village, town size distribution, and urban wages in equilibrium, in addition to endogenizing the relative price of land and financial constraint parameters. Second and more importantly, the general equilibrium response requires a long time horizon, similar to the macro-economic processes of economic development. Therefore, what we are able to show is the path to equilibrium (which we corroborate with evidence from data that we discuss later).

A Note on Debt vs. Savings One could argue that farmers can engage in land transactions through savings, instead of debt. While this is certainly possible, it is less probable because: (a) the empirical support for this is weak (very rarely do small farmers accumulate savings to purchase large tracts of land to become a large landowner) and, (b) the same constraints on the credit market could also apply to savings (for example, see Ashraf, Karlan, and Yin 2006; Brune, Giné, Goldberg, and Yang 2016; Breza and Chandrasekhar 2019). Therefore, we only incorporate debt in our model.

Model Simulations Figure 2 shows the distribution of land inequality Gini, its correlation with the distance to the nearest town, and the relative distance gradient between top quintile and bottom quintile towns using simulated data following the data generating process as per our model. Panel A shows the distance correlation subsequent to farmer-level optimization given their initial endowments and debt levels. Panel B shows the town-size correlation and Panel C shows the distance correlation based on the heterogeneity of binding financial frictions.

These pattern are similar to what we observe in the data, suggesting that the model likely explains an important phenomenon of selective migration or exit from farming sub-optimal land area by farming households. That is, the observations could be driven by either or both of the following: (a) mid-sized farmers and farming households sell their agricultural land and migrate to the nearest town, and/or (b) downsize or increase their landholding so that they either become small farmers or large farmers themselves to avoid farming a sub-optimal farm size.

Financial constraint binds for small farmers who are unable to buy land from the mid-sized farmers to achieve scale agriculture nor do they migrate since their returns from agriculture dominate the net returns from migration. On the other hand, large farmers are able to consolidate and increase the scale of their farm by purchasing land from the mid-sized farmers. Mid-sized farmers can also choose to expand by purchasing land from those wanting to sell. Note from Section 2.7 that only a small fraction of farmers are mid-sized, consistent with the thin land markets documented by researchers (Rajagopalan 2020). The small number of mid-sized farms who face this trade-off is also consistent with why we observe such low permanent migration in India (Munshi and Rosenzweig 2016; Morten 2019).

We test these implications from the model in our data and rule out a few alternative explanations, suggesting that factor market frictions could be an important reason for the preponderance of small farms.

5 Testing Model Implications

5.1 Role of Financial Constraints

The main role of financial constraint is that it prevents small farmers from acquiring more land to reach economies of scale and optimal productivity at a larger plot size. One metric to determine presence of financial constraints is whether or not a village has a bank branch (including regular commercial/retail banks, cooperative banks, and agricultural credit societies) that increases access to credit even among small farming households. This is plausible since agriculture sector loans, particularly targeting small and marginal farmers, are considered priority sector as per national policy (RBI 2021) and the respective state governments generally underwrite such loans, i.e., make budgetary provisions to compensate the banks in the event of non-payment (Phadnis and Gupta 2015). Access to finance can enable small farmers to acquire more farmland, mid-sized farmers to sell their land, and large farmers to further consolidate their land. This should lower inequality as both mid-sized and small farmers exit and become large, respectively, generating a more homogeneous land holding distribution among the remaining farmers.

An empirical test of this hypothesis from our model would be to examine how inequality evolves over time in the presence of financial constraints (given the dynamic nature of the farmer optimization problem). Since we only have cross-sectional data from a specific data source, we test the hypothesis using variation in the nearest town population to capture the evolution of rural land inequality with increasing urbanization. Further, we leverage variation in village-level access to formal financial institutions (i.e. commercial bank, cooperative bank, or agricultural credit societies) to empirically model the presence or absence of financial constraints. Specifically, we estimate Equation 2 separately for villages with banks and those without banks.

Panel A Figure 3 juxtaposes the town-size and distance correlation with and without financial constraints, showing a clear distance-town size interaction gradient in the presence of financial constraints and no such gradient in its absence. Columns 1 and 4 of Table 3 document the regression coefficients equivalent to this graph. We note that the distance-town size interaction term is negative and significant (both statistically and in magnitude) in the presence of financial friction (limited access to formal financial institutions in a village) relative to the coefficient in the absence of financial frictions. This suggests that rural land inequality is correlated with the urbanization process only when there is limited access to formal financial institutions. On the other hand, the marginally significant distance coefficient in Column 4 indicates rural land inequality at baseline but one that does not worsen with increasing urbanization in the presence of financial institutions.

Land Consolidation Another key implication of our model is that large farmers, who face no financial constraints, could buy the agricultural land sold by mid-sized farmers to increase scale of production. Therefore, farm sizes at the top of the landholding distribution should be large relative to sizes at lower percentiles of the distribution close to towns compared to farther away. Indeed, we note this in Panel B Figure 3 where the ratio of farm size at the 99th percentile relative to farm size at the 25th percentile shows a negative town-size-distance gradient. The figure suggests that farms at the 99th percentile of size distribution is nearly 20 times as large as farms at the 25th percentile of the distribution within a village. Further, this gradient is more discernible among villages without formal financial institutions (figure on the right, Panel B).

Columns 2 and 5 Table 3 present the regression coefficients from an equivalent regression specification. Similar to the Gini index, we find that the relative farm size at the top of distribution is negatively associated with distance in the presence of urbanization and financial constraints. While we also note a negative gradient in the presence of formal financial institutions in a village, the figure in Panel B Figure 3 suggests some non-monotonicity. Importantly, we do not find large differences in farm sizes in the immediate peri-urban areas when access to finance is not a constraint.

Composition of Farmers by Plot-Size In order to examine whether the increasing inequality is driven by fewer mid-sized farmers, we test whether the share of mid-sized farmers as a percentage of total farmers in a village is positively associated with an increase in distance from town in the presence of urbanization and financial frictions. Panel C Figure 3 and Columns 3 and 6 of Table 3 document this. Both the figure and table depict clearly that the composition of farmers (share of mid-sized farmers) shows no distance correlation in the absence of financial frictions. The correlation is only evident in the presence of frictions.

5.2 Fewer Mid-Sized Farmers Near Towns Over Time

Central to all the model implications is that mid-sized farmers exit sub-optimal agriculture. They can do this by downsizing or upsizing, or through complete sectoral reallocation by migrating to urban areas to earn wages from the non-agricultural sector. Testing this is hard mainly due to data challenges: (a) panel data on the entirety of land holding distribution is hard to come by, (b) data on migration at the household or individual-level is not available; even aggregate data on migration by landholding size distribution is unavailable.

We overcome these challenges by exploiting all-India household panel data from IHDS to examine: (a) whether there is differential attrition in the subsequent survey rounds, and (b) compute a transition matrix of landholding between the two survey rounds, both based on the households' initial landholding size. Analysis of attrition investigates whether the entire household is missing in the second round, which is a likely (although not a conclusive) indicator of permanent migration. The transition matrix allows us to examine whether mid-sized farmers downsize or upsize over time.

Column 1 Table 4 examines attrition by households' initial landholding size. The sample includes all rural land-holding households from IHDS wave 1 from 2005, with the dependent variable as an indicators for whether the household was surveyed during wave 2 in 2012. We include village fixed effect to account for time invariant village-level unobservables or shocks that occur between 2005 and 2012, common to all households across the land holding distribution. The coefficients indicate that mid-sized farming households are 1.6 percentage points more likely to be missing relative to small or large farming households in 2012 in villages in the peri urban areas and this relative likelihood of their absence decreases with distance. Columns 2 and 3 of Table 4 examines attrition within subsamples based on the presence of formal financial institutions in the village. As seen in the previous subsection, we note that all the correlation is mainly being driven by villages without financial institutions, i.e., those where small farmers are more likely to face financial constraints.

Table 5 presents the landholding transition matrix by initial landholding distribution between the two survey rounds. Over 92% of the initial small farm households (< 5 acres) continue to remain small in 2012 and only a small fraction increase their landholding. Among mid-sized farms in 2005, 42% become small and about 28% become large in 2012 and only 30% remain mid-sized. Among large farm households, the majority remain large although a substantial fraction downsize. Importantly, a greater share among these erstwhile large and mid-sized farms become small farms compared to being mid-sized in 2012. Perhaps an important reason for the downsizing we observe could be due to lifecycle events such as inheritance, but we show later that such events do not systematically alter the landholding distributions to generate the empirical patterns on rural land inequality and urbanization that we observe.

5.3 Plausible Welfare Implications

While welfare analysis of land inequality is outside the scope of this paper, we find increased mechanization and village-level agricultural productivity (measured as average yield achievement ratio defined by FAO as the ratio of actual yields to potential yields) closer to towns (see Panel A Figure 4).

There are likely multiple channels behind these observed welfare effects. First, the exit of mid-sized farmers may leave behind more productive farms (i.e. small and large), which increases overall productivity at the village level. Second, land consolidation by large farmers, for which we find suggestive evidence in the form for increasing ratio of land sizes across the distribution closer to towns, could lead to increased mechanization that support increasing returns to scale.

Finally, this observed correlation is also consistent with the plausible urban influence through access to markets and higher agricultural wages that increases overall income and consumption levels in the surrounding villages (we find negative distance correlation with agricultural wages in Columns 2-3 Table A1).

5.4 Not a steady state

The model suggests a dynamic process in the evolution of rural land inequality over time as a function of distance to the nearest town and urbanization. However, we model only partial equilibrium conditions, focusing on farmer-level optimization over their choice of landholding size, which could lead to increasing land inequality if there are missing mid-sized farms. We do not model steady state, which requires town population, urban wages, and prices to converge to their steady state values. We find support for this choice empirically, where we find suggestive evidence of worsening land inequality in villages close to towns over time. Using the fact that our two sources of landholding data are separated by 7 years, we investigate changes in land inequality by comparing Gini index constructed using SECC data by only including landed households (in order to make it comparable to the administrative data) with the subsequent farmer-level administrative data. We note that the Gini index continues to increase in villages close to town, particularly in those surrounding larger urban areas (see Panel B Figure 4).

6 Discussion

Policy recommendations to address wealth inequality, particularly land, is hard because it requires reducing frictions across multiple markets (land, labor, and capital). Given the history of land redistribution and land ceiling legislations in the past, further land redistribution may be politically infeasible and economically unclear as the best policy response.[11] An important fact to bear in mind is that farms get fragmented over time due to lifecycle events such as inheritance, resulting in more small farms that can't take advantage of economies of scale. On the other hand, land consolidation policies without taking into account the constraints facing small farmers in switching to better opportunities (such as migrating to urban areas to earn better wages) may not be welfare enhancing.

In this paper, we bring to fore some key empirical facts about rural land inequality that is shaped by who remains in agriculture. This choice is based on the relative attractiveness of farming enterprise with respect to urban wage income and whether or not binding financial constraints keep households tied to sub-scale agriculture. It is certainly plausible that more than one specific mechanism could be at play. We tested a variety of alternate explanations, which are unable to explain all the observed patterns. We present the results from examining these explanations below.

6.1 Other (Complementary) Explanations

First, and most importantly, can the observed patterns be explained by family-size-choice based on landholding size since land is mainly acquired through inheritance (Desai, Vanneman, and National Council of Applied Economic Research 2012)?[12] For example, if mid-sized farmers choose to have more children in order to have more family labor to replace costly hired labor near towns, then the observed patterns could be explained by this. However, we

[11] There are very few 100 acre farms in the state. There are no longer large farms and the current "large farms" are only 12 acres on average

[12] Close to 90% of the land owning rural households in IHDS surveys report acquiring land through inheritance.

find no support for this claim. Table A2 show no differential patterns for mid-sized farm households by distance to the nearest town in: (a) share of household members under 15 years of age in 2005 (many of whom would reach adulthood by 2012), (b) whether households splits (that typically is associated with subdivision of land), (c) changes in family size, and (d) the likelihood of land acquisition through inheritance.

Second, we examine whether proximity to towns influencing crop-choice could affect choice of farm-size if urban crops (such as vegetables and fruits) make either small or very large farms more optimal (and this isn't inconsistent with the U-shaped agricultural productivity assumption). With respect to the composition of farm sizes, recall that we observe higher land consolidation (relative size of farm at the 99th percentile compared to that at the 25th percentile) as well as a larger share of small farms near towns. The dominant crops in the region we study are paddy and cotton, including on small farms. We note no differential pattern in crop-choice by distance to town in the presence of large towns (see Figure A6 and Column 4 Table A1).[13] This suggests that crop-choice is plausibly driven by the large share of small, subsistence farmers (rather than crop-choice driving the choice of farm size).

Third, mid-sized farmers may be differentially skilled relative to small farmers by distance to town, and thus more valued in urban areas. This explanation is based on the role of skill-based rural to urban migration in generating income inequality following rural-urban wage gap as discussed in Young (2013). Though this explanation is not inconsistent with our model, which suggests that the mid-sized farm households exit sub-scale agriculture and plausibly invest in human capital, we do not find strong distance correlation in differential skilling by household farm size. Specifically, we do not find that mid-sized farm households are: (a) relatively more educated (years of schooling) relative to small or large farm households with distance to the nearest town (Column 1 Table A3), (b) any more likely to have salaried employment that varies with distance (Column 2 Table A3), (c) any more likely to have a higher share of farm workers among family members that varies with distance (Column 3 Table A3), and (d) have higher or lower total income (as a proxy for unobserved skills) relative to small or large farmers that varies with distance to the nearest town (Column 4 Table A3). Further, we examine whether the patterns correlate with the presence of secondary school in a village and find no differential correlation between access to higher levels of schooling and rural land inequality (see Panel B Figure A6, where both types of villages - those with and those without secondary schools - display similar distance correlation).

[13] Vegetables and perishables, while important from urban consumption point of view, are not among the main three crops reported in the census data for our context and therefore, unlikely to influence the choice of farm-size.

Finally, having a good road - defined as national, state, or district highways or even rural roads as examined by Asher and Novosad (2020) - through a village could bring about changes in the local village economy and reduce migration costs.Therefore, placement of roads could also drive the observed distance correlations between rural land inequality and urbanization. While there is some evidence that inequality is worse when rural roads are of better quality (by lowering the cost of migration in our model), we observe similar inequality in the peri urban villages irrespective of the quality of road (Panel C Figure A6). Again, this explanation does not contradict our model but rather, complements by providing potential variation in the cost of migration, which can be modeled as a function of road quality-adjusted distance to town.

6.2 A Note on Endogeneity of Distance to Town

We do not make any causal claims based on the observed empirical patterns. We acknowledge that towns could endogenously locate close to villages with historically greater land consolidation and higher agricultural productivity with strong persistence over time. However, a few considerations: (a) we examine the patterns using most recent land holding data available, and (b) the persistence in land holding patterns and inequality vis-a-vis the timing of urban settlement and growth is likely to be broken by a series of land reforms implemented in the 1960s and 1970s. These reforms provided tenure to tenant farmers through redistribution of land and imposed ceilings on land holding sizes.

While our context and timing of data could address concerns of simultaneity bias, our coefficients could be over or under-estimated due to omitted variables bias, mainly arising from omitting variables that are correlated with distance to towns. These could be geographic and environmental features (such as soil suitability and aquifers), or non-agricultural opportunities that may drive both the location of towns as well as the extent of consolidation affecting land inequality in villages. We do not find any evidence in support of such explanations. Table A4 shows a lack of any meaningful distance correlations with soil suitability for major crops, access to surface water resources (rivers, springs, canals, or lakes/ponds), access to ground water resources (by presence of tubewells, open, or covered wells), and village-level non-agricultural land use and growth in non-agricultural employment.

7 Conclusion

This paper is one of the first attempts at documenting patterns in land inequality in rural India using the universe of land records in a large state as well as validating some of the key implications using all India representative household sample panel data. We find a very clear spatial pattern in rural land inequality based on the distance between the rural area and its nearest town.

These patterns are consistent with a model where mid-sized farmers find it more attractive to exit agriculture near urban areas. While small and large plot owners do not face this trade-off, small farmers face financial constraints to acquire more land to exploit economies of scale in agricultural production enjoyed by large farmers. Our model, along with the empirical observations, point to the role of urbanization, non-linearities in farm productivity by farm size, and financial constraints in creating asset size traps and incomplete structural transformation where only a specific group is able to take advantage of urban opportunities, leaving many to rely on subsistence agriculture.

References

Abadie, A., S. Athey, G. Imbens, and J. Wooldridge (2017):.

Adamopoulos, T., and D. Restuccia (2020): "Land Reform and Productivity: A Quantitative Analysis with Micro Data," *American Economic Journal: Macroeconomics*, 12(3), 1–39.

Asher, S., and P. Novosad (2020): "Rural Roads and Local Economic Development," *American Economic Review*, 110(3), 797–823.

Ashraf, N., D. Karlan, and W. Yin (2006): "Tying Odysseus to the Mast: Evidence From a Commitment Savings Product in the Philippines*," *The Quarterly Journal of Economics*, 121(2), 635–672.

Balboni, C., O. Bandiera, R. Burgess, M. Ghatak, and A. Heil (2020): "Why do people stay poor?," .

Banerjee, A., P. Gertler, and M. Ghatak (2002): "Empowerment and Efficiency: Tenancy Reform in West Bengal," *Journal of Political Economy*, 110(2), 239–280.

Banerjee, A., and L. Iyer (2005): "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India," *American Economic Review*, 95(4), 1190–1213.

Banerjee, A. V., and A. F. Newman (1993): "Occupational Choice and the Process of Development," *Journal of Political Economy*, 101(2), 274–298.

Banerjee, A. V., and A. F. Newman (1994): "Poverty, incentives, and development," *The American Economic Review*, 84(2), 211–215.

Bardhan, P., M. Luca, D. Mookherjee, and F. Pino (2014): "Evolution of land distribution in West Bengal 1967–2004: Role of land reform and demographic changes," *Journal of Development Economics*, 110, 171–190.

Barrett, C. B. (1996): "On price risk and the inverse farm size-productivity relationship," *Journal of Development Economics*, 51(2), 193–215.

Bazzi, S., A. Gaduh, A. D. Rothenberg, and M. Wong (2016): "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia," *American Economic Review*, 106(9), 2658–98.

Breza, E., and A. G. Chandrasekhar (2019): "Social Networks, Reputation, and Commitment: Evidence From a Savings Monitors Experiment," *Econometrica*, 87(1), 175–216.

Brune, L., X. Giné, J. Goldberg, and D. Yang (2016): "Facilitating Savings for Agriculture: Field Experimental Evidence from Malawi," *Economic Development and Cultural Change*, 64(2), 187–220, Publisher: University of Chicago Press.

Bryan, G., S. Chowdhury, and A. M. Mobarak (2014): "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, 82(5), 1671–1748.

Burchardi, K. B., S. Gulesci, B. Lerva, and M. Sulaiman (2019): "Moral Hazard: Experimental Evidence from Tenancy Contracts*," *The Quarterly Journal of Economics*, 134(1), 281–347.

Cornia, G. A. (1985): "Farm size, land yields and the agricultural production function: An analysis for fifteen developing countries," *World development*, 13(4), 513–534.

de Janvry, A., K. Emerick, M. Gonzalez-Navarro, and E. Sadoulet (2015): "Delinking Land Rights from Land Use: Certification and Migration in Mexico," *American Economic Review*, 105(10), 3125–49.

Desai, S., R. Vanneman, and N. D. National Council of Applied Economic Research (2005): "India Human Development Survey (IHDS)," .

——— (2012): "India Human Development Survey (IHDS)," .

Easterly, W. (2006): "Reliving the 1950s: the big push, poverty traps, and takeoffs in economic development," *Journal of Economic Growth*, 11(4), 289–318.

Foster, A. D., and M. R. Rosenzweig (2017): "Are there too many farms in the world? Labor-market transaction costs, machine capacities and optimal farm size," Discussion paper, National Bureau of Economic Research.

Gaurav, S., and S. Mishra (2015): "Farm size and returns to cultivation in India: Revisiting an old debate," *Oxford Development Studies*, 43(2), 165–193.

Glaeser, E., and J. V. Henderson (2017): "Urban economics for the developing World: An introduction," *Journal of Urban Economics*, 98, 1–5.

Gollin, D. (2018): "Farm size and productivity: lessons from recent literature," p. 40.

Gollin, D., D. Lagakos, and M. E. Waugh (2014): "Agricultural Productivity Differences across Countries," *American Economic Review*, 104(5), 165–170.

Gollin, D., and C. Udry (2021): "Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture," *Journal of Political Economy*, 129(1), 1–80, Publisher: The University of Chicago Press.

Heise, S., and T. Porzio (2021): "The Aggregate and Distributional Effects of Spatial Frictions," *Working Paper*.

Henderson, J. V., and S. Kriticos (2018): "The Development of the African System of Cities," *Annual Review of Economics*, 10(1), 287–314.

Henderson, J. V., and M. A. Turner (2020): "Urbanization in the Developing World: Too Early or Too Slow?," *Journal of Economic Perspectives*, 34(3), 150–73.

Hsieh, C.-T., and E. Moretti (2019): "Housing Constraints and Spatial Misallocation," *American Economic Journal: Macroeconomics*, 11(2), 1–39.

Kahn, M. E., W. Sun, J. Wu, and S. Zheng (2018): "The Revealed Preference of the Chinese Communist Party Leadership: Investing in Local Economic Development versus Rewarding Social Connections," NBER Working Papers 24457, National Bureau of Economic Research, Inc.

Kraay, A., and D. McKenzie (2014): "Do poverty traps exist? Assessing the evidence," *Journal of Economic Perspectives*, 28(3), 127–48.

Lagakos, D. (2020): "Urban-rural gaps in the developing world: Does internal migration offer opportunities?," *Journal of Economic perspectives*, 34(3), 174–92.

Lagakos, D., A. M. Mobarak, and M. E. Waugh (2021): "The Welfare Effects of Encouraging Rural-Urban Migration," *Working Paper*.

Lobell, D. B. (2013): "The use of satellite data for crop yield gap analysis," *Field Crops Research*, 143, 56–64.

Morten, M. (2019): "Temporary Migration and Endogenous Risk Sharing in Village India," *Journal of Political Economy*, 127(1), 1–46, Publisher: The University of Chicago Press.

Munshi, K., and M. Rosenzweig (2016): "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap," *American Economic Review*, 106(1), 46–98.

Murphy, K. M., A. Shleifer, and R. W. Vishny (1989): "Industrialization and the big push," *Journal of political economy*, 97(5), 1003–1026.

Norris, S., M. Pecenco, and J. Weaver (2021): "The Effects of Parental and Sibling Incarceration: Evidence from Ohio," *American Economic Review*.

Phadnis, A., and A. Gupta (2015): "The Politics of Farm Loan Waivers," *Economic and Political Weekly*, 54(23), 7–8.

Rajagopalan, S. (2020): "Opinion | Ease the sale of agricultural land to help our farmers," Section: Opinion.

RBI, R. B. o. I. (2021): "Reserve Bank of India - Master Directions," .

Rigg, J., A. Salamanca, and E. C. Thompson (2016): "The puzzle of East and Southeast Asia's persistent smallholder," *Journal of Rural Studies*, 43, 118–133.

Rosenstein-Rodan, P. N. (1961): "Notes on the theory of the 'big push'," in *Economic Development for Latin America*, pp. 57–81. Springer.

Saiz, A. (2010): "The Geographic Determinants of Housing Supply*," *The Quarterly Journal of Economics*, 125(3), 1253–1296.

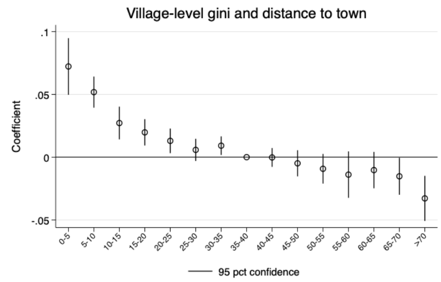
Syrquin, M. (1988): "Patterns of structural change," *Handbook of development economics*, 1, 203–273.

Young, A. (2013): "Inequality, the Urban-Rural Gap, and Migration*," *The Quarterly Journal of Economics*, 128(4), 1727–1785.

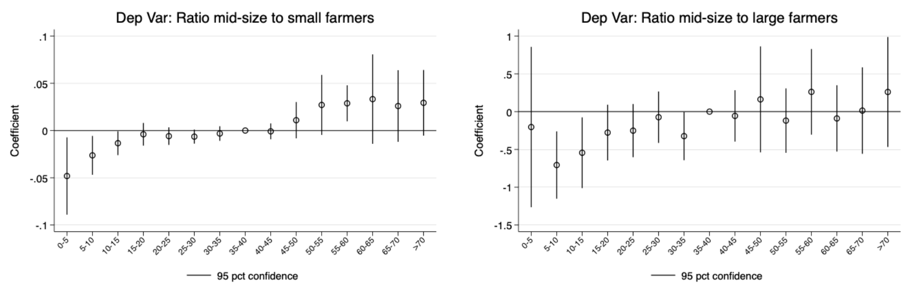
Figures

Figure 1: Key Facts

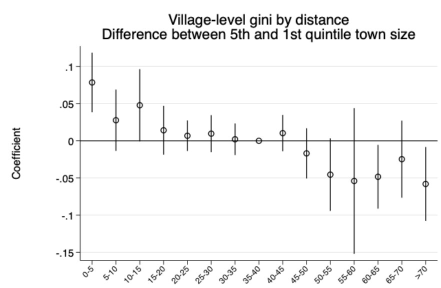
Panel A: Distance Correlation



Panel B: Farm-Size Correlation



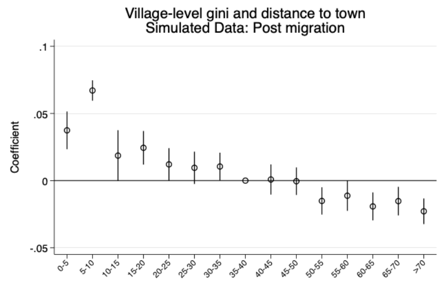
Panel C: Town-Size Correlation



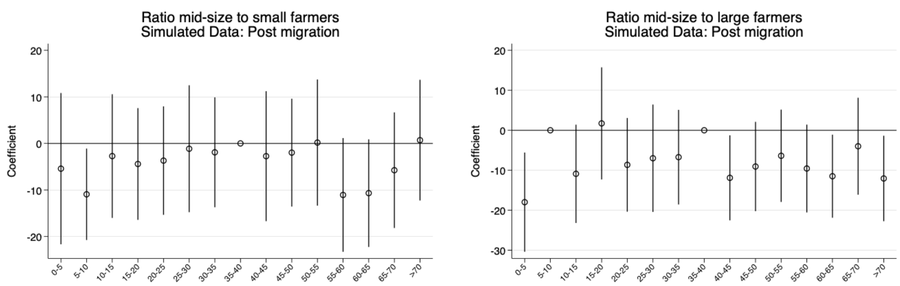
Notes: The panels above document empirical facts about rural landholding inequality based on the distance of the rural area (i.e., a village) and closest town. The x-axes record 5 km distance bins, where the bin 35-40 km as the leave-out group and the y-axes correspond to regression coefficients on the specific distance bins when either gini coefficient is the dependent variable or the ratio of the number of mid to small or the number mid to large farmers based on their landholding sizes. Small and marginal farmers are those with less than 5 acres of land. Mid-sized farmers are those with 5-8 acres of land. Large farmers are those with more than 8 acres of land. The last figure plots the coefficients of the interaction terms between distance bins and the size of the nearest town (based on population quintiles). The regression specifications control for the nearest town and sub-district fixed effects and cluster standard errors by the nearest town.

Figure 2: Model Simulations

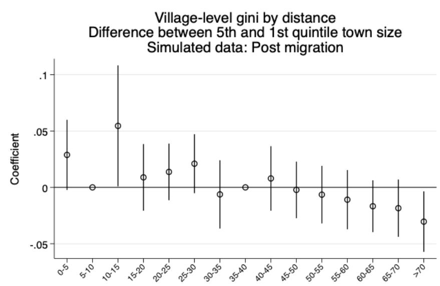
Panel A: Simulated Distance Correlation



Panel B: Simulated Farm-Size Correlation



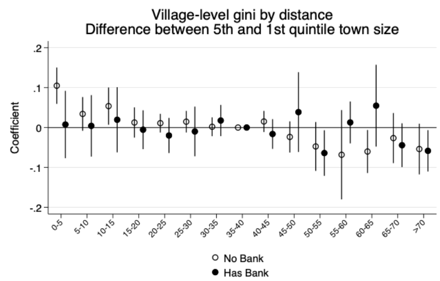
Panel C: Simulated Town-Size Correlation



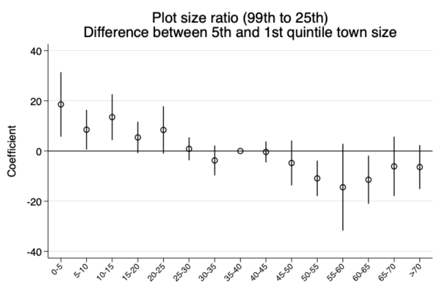
Notes: Above graphs are based on simulated data, following data generating process as per our model. Simulations include 1500 villages, distributed at random distances from 75 towns of varying population sizes. The distance and population distributions are drawn from random normal distributions with $\mu_d = 25, \sigma_d = 15$ and $\mu_s = 100,000, \sigma_s = 30,000$, respectively. Each village has 500 farmers, each with initial land endowment and debt-levels drawn from skewed distributions. We assume log utility functional form, fixed relative price of (agricultural) land within a village, and a net urban wage following $w = (s/1000)^{0.75} - d^{0.5}$. The above graphs show the resulting Gini correlations after one round of farmer-level optimization.

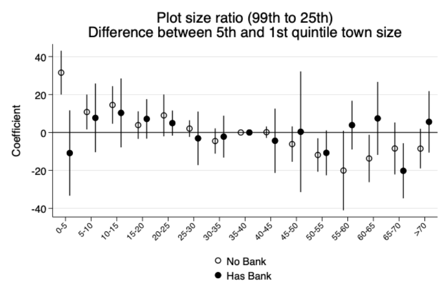
Figure 3: Model Implications

Panel A: Distance Correlation, Urbanization, Access to Finance

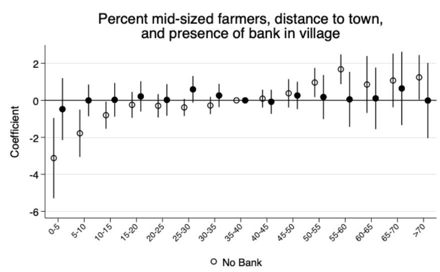


Panel B: Land Consolidation by Large Farmers





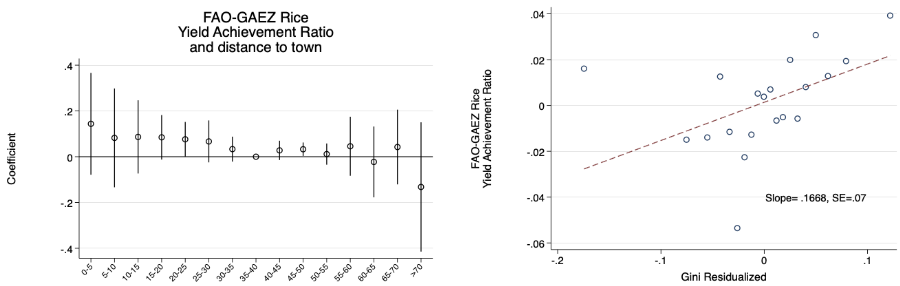
Panel C: Share of mid-sized farmers



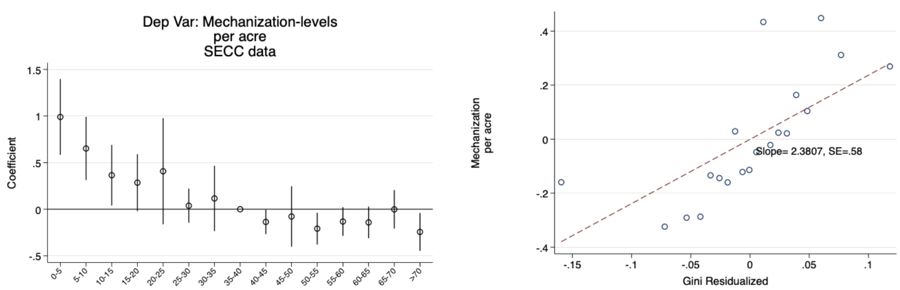
Notes: Panel A plots the interaction term between distance bin and town size with village-level gini as the dependent variable across two specifications - one where villages have a bank or formal financial institution for credit access in hollow circle and another where villages do not have any banks or formal financial institutions in solid black circles. Panel B, the figure of left plots the interaction term between distance bin and town size with the ratio of plot sizes at 99th and 25th percentile as the dependent variable. Panel B right figure examines the ratio of plot sizes at 99th and 25th percentile by town size and distance interaction. Panel C is also constructed similarly but with the percent of mid-sized farmers as the dependent variable. All specifications include the nearest town and sub-district fixed effects and cluster standard errors by the nearest town.

Figure 4: Welfare Implications

Panel A: Productivity



Panel B: Factor Intensity



Notes: Panel A presents the distance correlation in yield achievement ratio for wetland paddy using FAO GAEZ data (left) and its correlation with rural land inequality (right). Panel B presents the distance correlation in mechanization per unit area (capital intensity) using SECC data (left) and its correlation with rural land inequality (right). All specifications include the nearest town and sub-district fixed effects and cluster standard errors by the nearest town.

Tables

Table 1: Data Sources

Variable	Source	Obs	Year	Scope
Farmer-level Gini	Land Records	12,843	2017	One State (Universe)
Household-level Gini	SECC	9,984	2012	One State (Universe)
Dist. to Town (km)	Census	10,686	2011	One State
Town Size	Census	49	2011	One State
Village Bank	Census	10,686	2011	One State
Village Road	Census	10,686	2011	One State
Village Sec. School	Census	10,686	2011	One State
Village Water Src	Census	10,686	2011	One State
Agri Outcomes	FAO GAEZ	NA	2010	Raster Image (All India)
HH Panel	IHDS	21919	2005, 2012	All India Sample
Village Module	IHDS	15627	2005, 2012	All India Sample

Notes: Between 2011 census and 2017, the state created new villages from existing villages. The data on towns pertain to the nearest towns from each of the village in the dataset. The sampling frame in the India Human Development Survey (IHDS) is entire country, from which villages were randomly selected, stratified by state, to form the primary sampling units (PSU). From each PSU, a sample of households were randomly selected. The original sample consists of 41554 households. Please refer to IHDS documentation for further details. The final observations used for analysis in this paper is restricted based on household-level panel among landed households in 2005.

Table 2: Key Facts: Table

	(1) Gini	(2) Mid-Small Ratio	(3) Mid-Large Ratio	(4) Gini (All)	(5) Gini 1st vs. 5th Quintile
Panel A: Main Specification					
Dist (10 km)	-0.0973***	0.00651*	0.159***	-0.0957***	-0.0256
	(0.0265)	(0.00351)	(0.0473)	(0.0284)	(0.0423)
Dist (10 km) x Town Pop				-0.000314	-0.00187**
				(0.000463)	(0.000752)
Observations	9917	9914	9262	9917	3924
Mandal Fixed Effect	Y	Y	Y	Y	Y
Town Fixed Effect	Y	Y	Y	Y	Y
Panel B: Without FE					
Dist (10 km)	-0.0583***	0.00401***	0.0434**	-0.0603***	-0.0374***
	(0.00457)	(0.000542)	(0.0218)	(0.00501)	(0.00676)
Town Pop (100,000)				0.0202***	0.0205***
				(0.00207)	(0.00211)
Dist x Town Pop				-0.00170***	-0.00231***
				(0.000409)	(0.000417)
Observations	9918	9915	9264	9918	3928
Sub-District Fixed Effect	N	N	N	N	N
Town Fixed Effect	N	N	N	N	N

Standard errors in parentheses
* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Panel A of the above table corresponds to the regressions behind Figure 1. Panel B represents simple bi-variate/interaction specification without any fixed effect. Gini Index is transformed into standard deviation units relative to villages in the 35-40 km distance bin. Distance to the nearest town is measured in multiples of 10 km and the town population is in multiples of 100,000.

Table 3: Model Implications: Table

	(1) Gini No Bank	(2) Ratio 99-25 No Bank	(3) Perc. Mid-Sized No Bank	(4) Gini Bank	(5) Ratio 99-25 Bank	(6) Perc. Mid-Sized Bank
Dist x Town Pop	-0.00201**	-0.0766***	0.00795***	-0.000741	-0.0666***	-0.000224
	(0.000817)	(0.0120)	(0.00178)	(0.000476)	(0.0137)	(0.00395)
Dist (10 km)	-0.0121	-0.235	-0.0609	-0.0608*	-1.314*	-0.177
	(0.0459)	(0.633)	(0.110)	(0.0319)	(0.756)	(0.193)
Observations	3212	3212	3212	679	679	679
Sub-District Fixed Effect	Y	Y	Y	Y	Y	Y
Town Fixed Effect	Y	Y	Y	Y	Y	Y

Standard errors in parentheses
* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Columns 1-3 represent estimates from Equation 2 using the subset of villages without formal financial institutions. Columns 4-6 represent estimates from Equation 2 using the subset of villages with formal financial institutions. The dependent variable in Columns 1 and 4 - Gini Index - is transformed into standard deviation units relative to villages in the 35-40 km distance bin. The dependent variable in Columns 2 and 5 is the ratio of farm size between farmers at 99th relative to 25th percentile landholding distribution, and the dependent variable in Columns 3 and 6 is the percentage of mid-sized farmers in a village. Distance to the nearest town is measured in multiples of 10 km and the town population is in multiples of 100,000. Standard errors are clustered by the nearest town.

Table 4: Selective Absence of Mid-Sized Farm HH in Household Panel Survey

	(1) Absent (2012) All	(2) Absent (2012) Bank	(3) Absent (2012) No Bank
Mid-Sized x Dist (km)	-0.00118**	-0.000255	-0.00142***
	(0.000467)	(0.00115)	(0.000509)
Mid-Sized Farm (2005)	0.0160*	-0.00822	0.0230**
	(0.00912)	(0.0216)	(0.0100)
Observations	18159	4221	13938
Village Fixed Effect	X	X	X

Standard errors in parentheses
* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: The dependent variable is a dummy variable if a household from IHDS-1 (2005) is not found/surveyed in IHDS-2 (2012). Columns 2 and 3 are conditioned on whether or not a village has a bank/formal financial institution. The leave-out group is households that are either small (< 5 acres) or large (> 8 acres) in 2005. We use Eicker-Huber-White robust standard errors to account for any heterogeneity since villages are randomly sampled and the main coefficient of interest is the distance interaction term.

Table 5: Transition Matrix of Land Ownership Changes

Plot Size (2005)	<= 5 Acre (2012)	5-8 Acre (2012)	>8 Acre (2012)
< 5	0.925	0.0393	0.0361
5-8	0.424	0.302	0.275
8-10	0.265	0.209	0.525
10-20	0.205	0.136	0.659
>20	0.171	0.0642	0.765

Notes: Plot sizes measured in acres. ≤ 5 includes landless households. 5-8 includes households with 8 acres.

Appendix

A1 Additional Data Sources

Other measures of village-level agricultural yields include NDVI product made available through BHUVAN NOEDA (by India Space Research Organization or ISRO) that uses Oceansat-2 Ocean Color Monitor (OCM2) sensor data. Lobell (2013) highlights the importance of using satellite data products such as NDVI and other vegetation indices for measuring yields when traditional field-based estimates are not available or are inaccurate. The satellite Oceansat-2 collects images every 2 weeks, covering a wide 1420 km area at a high radiometric resolution mapping to a 1 km x 1 km spatial resolution. The NDVI product generated by Oceansat-2 is highly correlated to that generated using MODIS sensor data, a more commonly used satellite data source. Using the NDVI raster images from the main growing season for each year from 2010 onward, we overlay the 2011 census village shapefiles for our state and compute the average NDVI value contained within each village boundary. The correlations look similar to the results from FAO GAEZ yield achievement ratio if we use NDVI based measure instead.

A2 Model Appendix

A2.1 Empirical Support for Model Assumptions

The two key assumptions in our model are: (a) non-monotonicity of agricultural production as a function of land size (the "U-shaped" yield function) and (b) presence of financial frictions also a function of existing land size that limits expansion of land through debt/borrowing.

A2.1.1 Land-Size and Agricultural Productivity

This assumption is largely based on recent advances in the literature on agricultural productivity and economic development such as Foster and Rosenzweig (2017) with the larger literature summarized by Gollin (2018). While a replication exercise is beyond the scope of this paper, we provide support for this assumption in our data.

In order to show this, we use village-level average NDVI as a proxy for agricultural productivity in the village and examine whether this follows a U-shape relationship based on the percentage of small, mid-sized, and large farmers within a village, in the absence of plot-level data on yields.

Panel A Figure A7 shows that the data supports non-monotonicity where larger percentage of mid-sized farmers reduce overall village-level productivity compared to villages with fewer mid-sized farmers relative to small or large farmers.

A2.1.2 Financial Constraints, ϕ

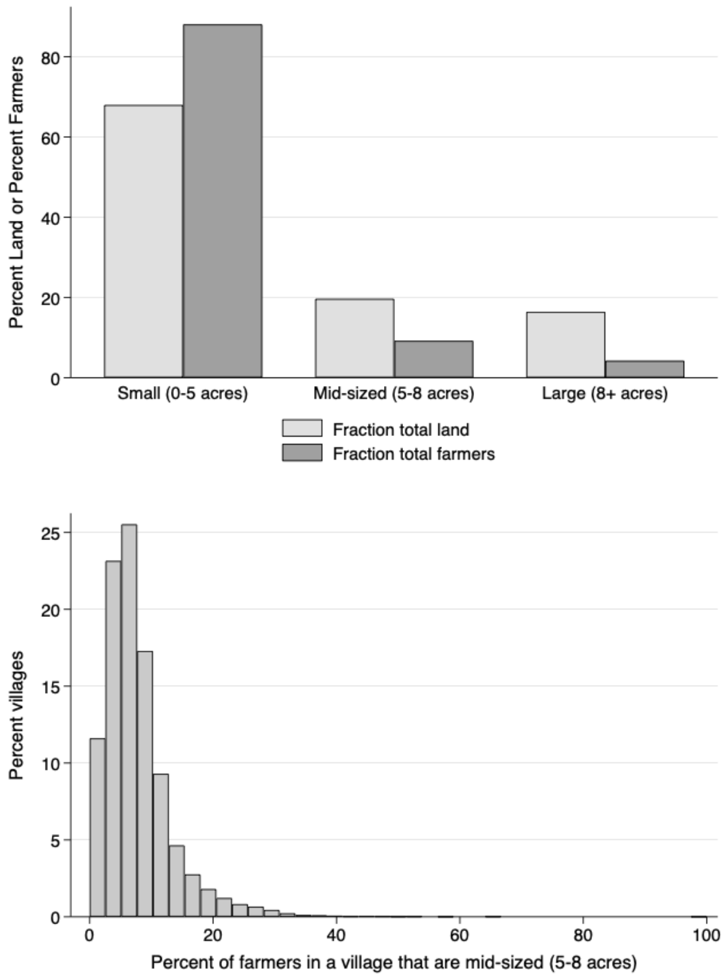
In order to find support for the assumption of financial frictions, we examine land holding changes using household-level panel data from IHDS (rounds 2005 and 2012). We find that households exhibit positive, negative, and no change in the amount of land owned between the two rounds. Among households that report acquiring additional land, we find that the median share of new land acquired relative to initial land holding (which we denote as ϕ in our model) is 0.69 and close 60% of the households report a change less than the land their owned in the prior period (see Panel B Figure A7).

Further, we find that more mid-sized farmers sell land compared to small farmers or large farmers. The transition matrix Table 1 shows that over 92% of small farmers remain small whereas only 30% mid-sized farmers remain mid-sized between the two survey rounds. Similar to small farmers, a greater share of large farmers continue to remain large (over 76% of farmers with over 20 acres of land continue to remain large).

Appendix Figures

Figure A1: Land size distribution

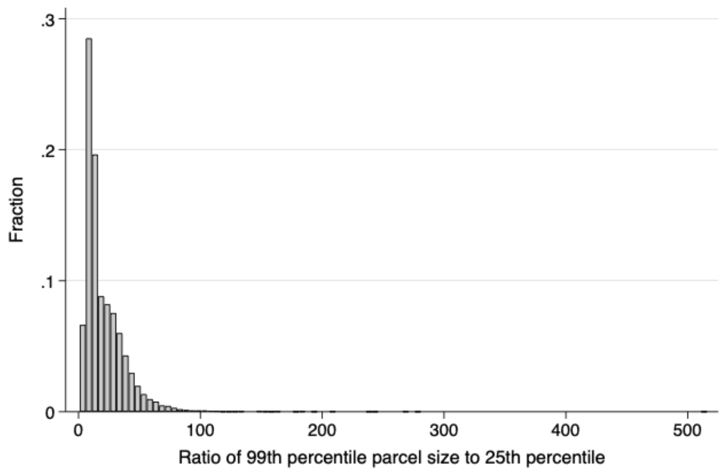
Panel A



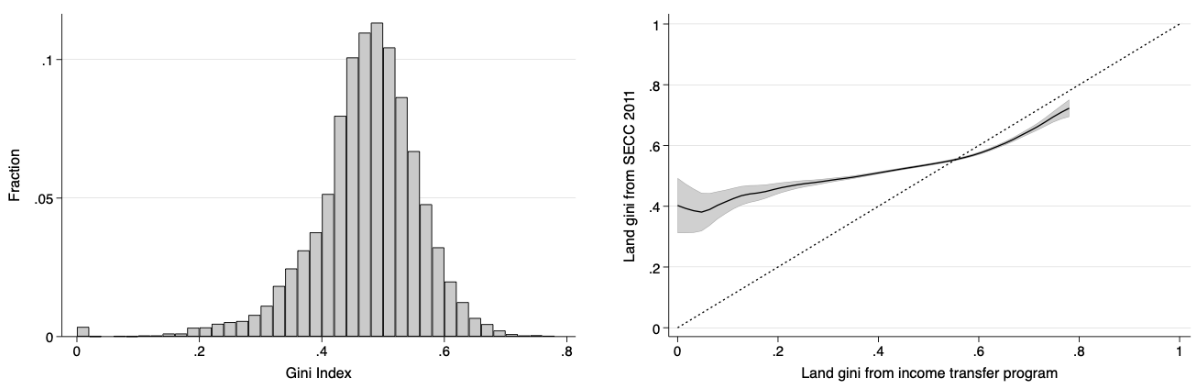
Notes: The figures present the summary of farmer size groups in the land records data.

Figure A2: Land inequality and land concentration

Panel A: Land holding size at 99th percentile relative to 25th percentile

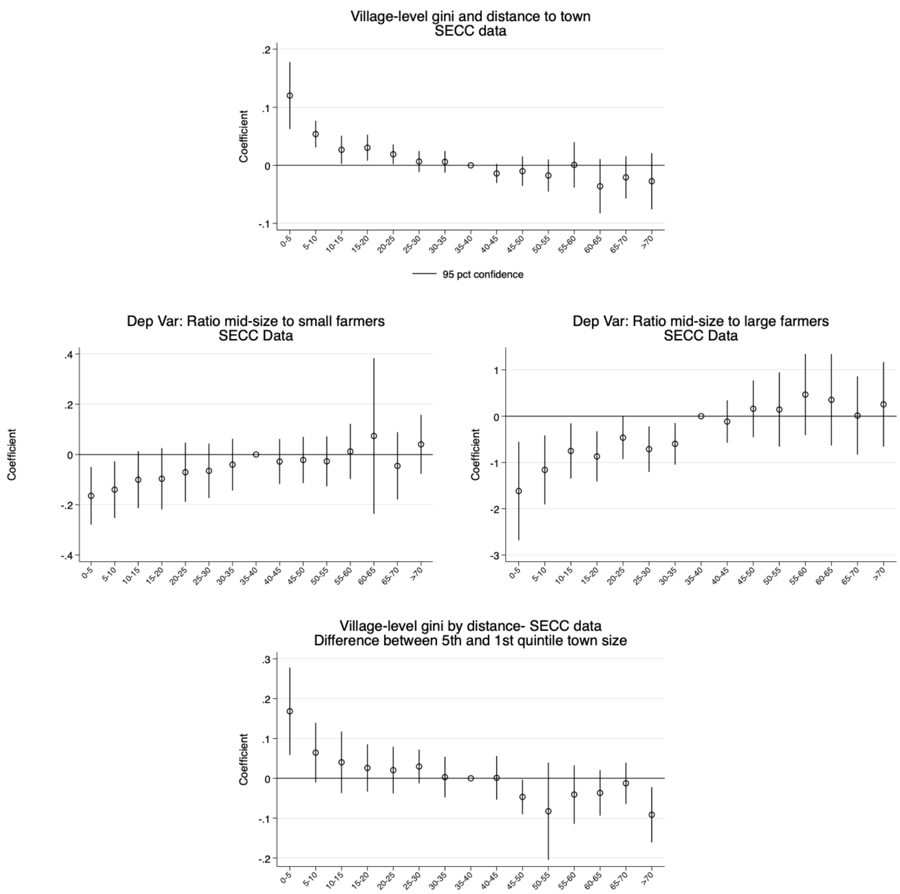


Panel B: Village-level land holding gini index



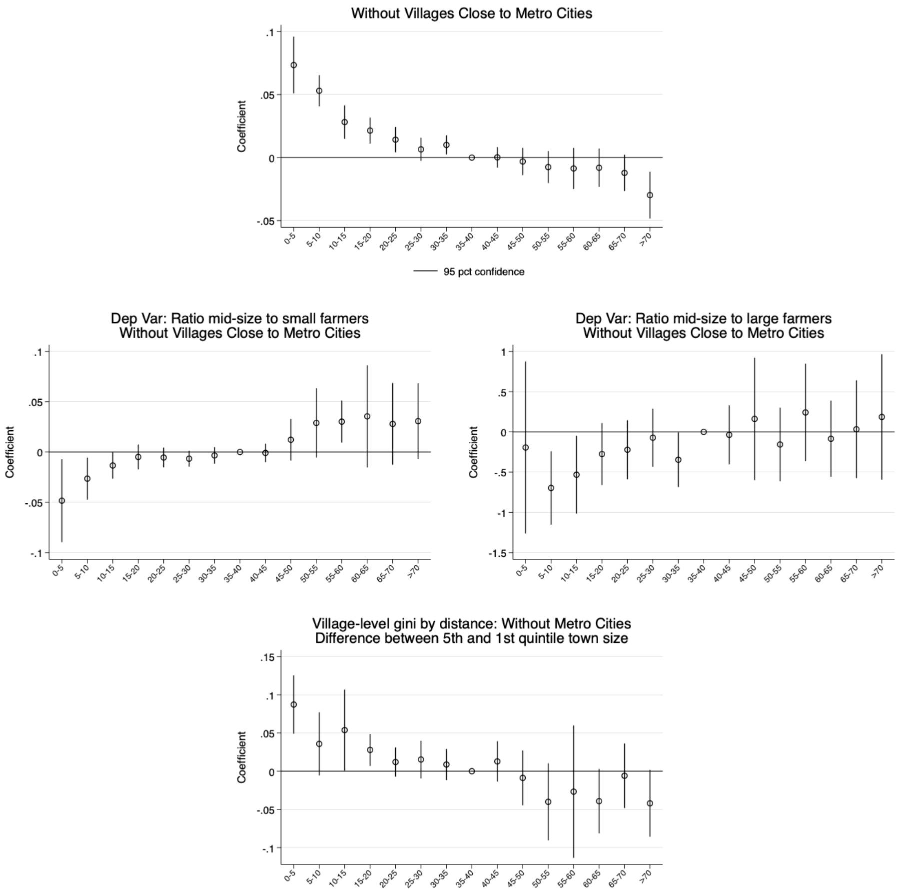
Notes: Panel A presents the distribution of the ratio between 99th percentile land holding size and 25th percentile. Panel B presents the distribution of village-level Gini index using farmer-level data (land records used for income transfer) and the correlation between Gini constructed using survey data (self reported landholding at the household level).

Figure A3: Robustness: SECC data on land ownership by household



Notes: The figures above recreate Figure 1 using household-level census data from the Socio-Economic and Caste Census (SECC) for the study state.

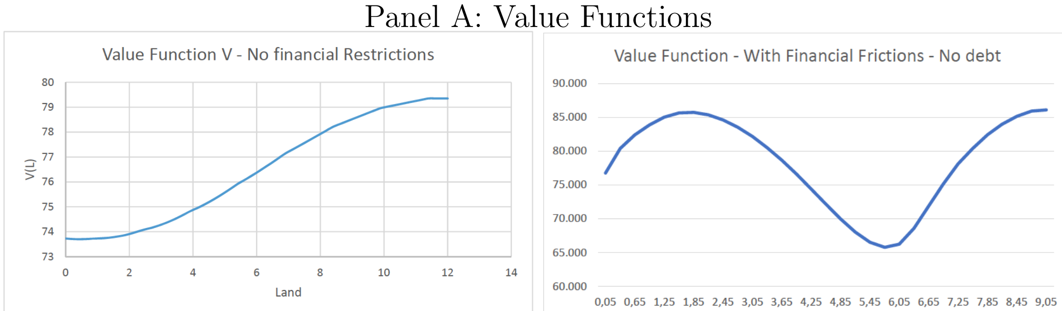
Figure A4: Robustness: Without Metro



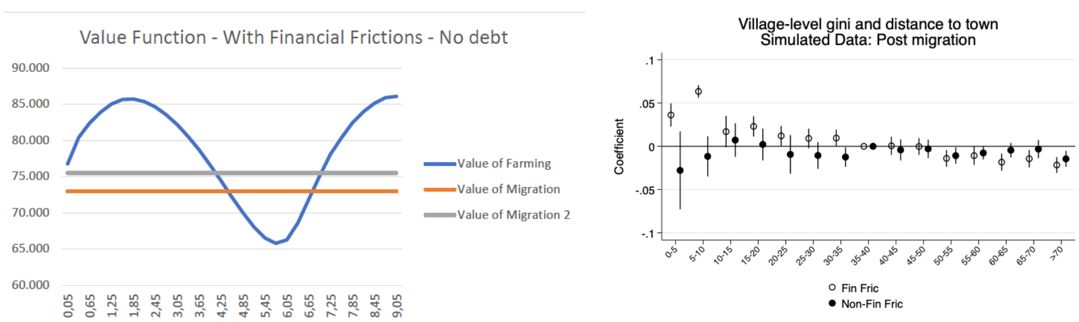
Notes: The figures above recreate Figure 1 after dropping villages in the vicinity of metro cities.

Figure A5: Model Intuition

Panel A: Value Functions



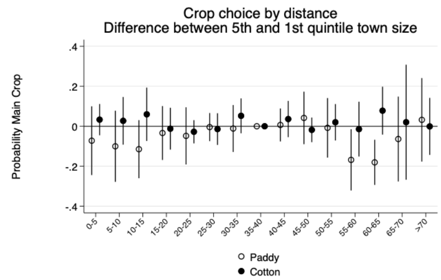
Panel B: Financial Frictions, Urban Opportunity, and Rural Land Inequality



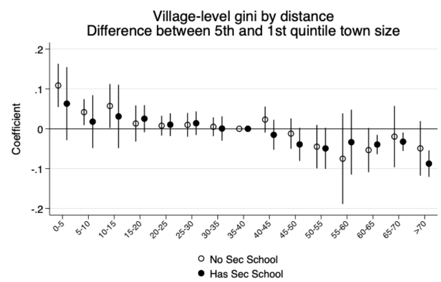
Notes: We use quadratic agricultural production function and log-linear utility function. Without financial constraints, the value function is monotonic, generating a unique threshold value of land, below which all land-owners would sell and migrate (Panel A, left). However, binding financial constraints generates non-monotonicity in the value function (Panel A, right). This creates an upper and lower bound of land holding size, and only those with land in this range would find it profitable to migrate. This range increases with an increase in returns to migration, i.e. the urban wage net of migration costs (Panel B, left). Simulated data as per the model with 2 periods of optimization generates the distance correlation pattern we observe empirically (Panel B, right).

Figure A6: Other Explanations

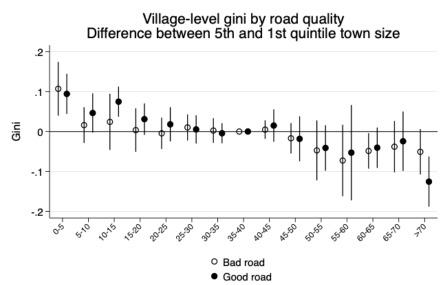
Panel A: Crop-choice



Panel B: Access to Skilling



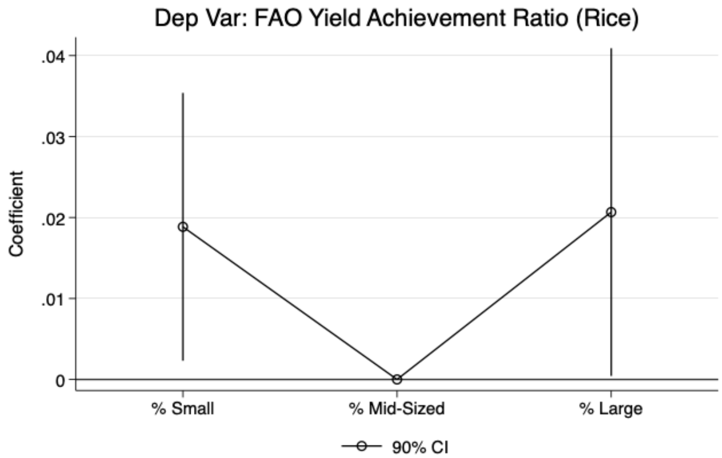
Panel C: Migration Costs



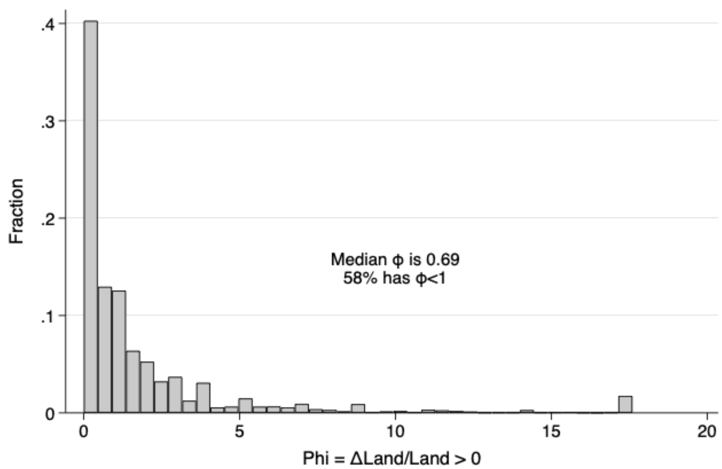
Notes: The above figure examines whether crop-choice (Panel A) and village-level amenities - importantly, access to secondary schooling (Panel B), and good quality roads (Panel C) help explain the observed town-size correlations.

Figure A7: Supporting Model's Assumptions

Panel A: Land-Size and Agricultural Productivity

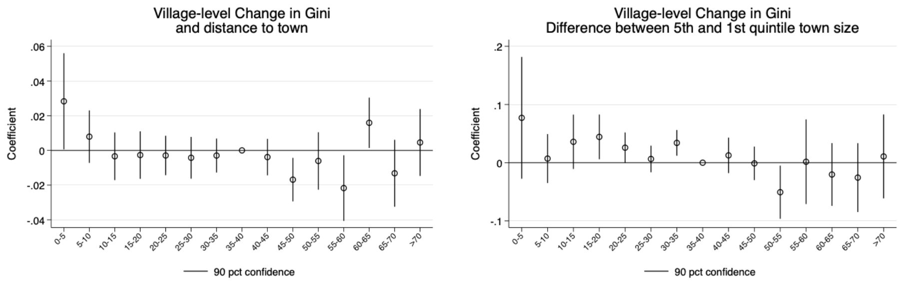


Panel B: Financial Constraints ($\phi < 1$)



Notes: Panel A is generated using village-level average FAO yield achievement ratio data for rice as the dependent variable and percentage of small and large farmers as the explanatory variables (with percentage of mid-sized farmers as the leave-out group). Panel B plots the distribution of ϕ using IHDS data, calculated as the percentage change in land holding between the two survey rounds in 2005 and 2012 when new land was acquired.

Figure A8: Path to Equilibrium: Not a Steady State



Notes: The graphs above are similar to Figure 1 except that the dependent variable is change in gini over time. The first time-period corresponds to 2011 when SECC data was collected. The second time-period corresponds to 2017-18 when land records were updated for the implementation of farmer income support program. We exclude all landless households from SECC to make it comparable to the land records data from later time-period.

Appendix Tables

Table A1: Key Patterns: Using All India Household Panel Data

	(1) Land Inequality	(2) Ploughing Wage	(3) Harvest Wage	(4) Urban Crop
Dist Town (km)	-0.235***	-0.182***	-0.0785***	0.000276
	(0.0251)	(0.0153)	(0.0126)	(0.000227)
Observations	21919	14066	14141	15627
District Fixed Effect	X	X	X	X

Standard errors in parentheses
* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Data in this table are from IHDS Wave 1 (2005). Dist (km) is the distance in kilometers to the nearest town. Column 1 uses the entire rural sample from IHDS-1 (2005) across India. The dependent variable in Column 1 is the ratio between the maximum and minimum land per capita by jati groups in a village. Columns 2-4 are village-level male agricultural wages for specific tasks and classification of main crops grown as an urban crop (fresh fruits and vegetables). Since village identity is not disclosed, we can only use distance to nearest town reported in the survey. Subsequently, we are unable to estimate the town-size interaction since the nearest town identifier is not available in this dataset. We include fixed effect at the district-level following the sampling strategy to account for any spatial correlation.

Table A2: Testing the Family Size Explanation

	(1) Share Children (2005)	(2) HH Split (2012)	(3) Change HH Size (2012)	(4) Land Inherited (2012)
Mid-Sized x Dist (km)	-0.000977	0.000756	-0.0000222	-0.0000348
	(0.000608)	(0.000909)	(0.0000509)	(0.000909)
Mid-Sized Farm (2005)	-0.000633	-0.000819	0.000471	0.00757
	(0.0111)	(0.0180)	(0.00140)	(0.0166)
Observations	13170	13170	13170	10286
Village Fixed Effect	X	X	X	X

Standard errors in parentheses
* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Data in this table are from the IHDS Panel. Dist (km) is the distance in kilometers to the nearest town. Farm size classifications are as per the household landholding sizes in Wave 1 of the IHDS survey (in 2005). Dependent variables except the share of children, are from 2012 survey round (Wave 2). All specifications include village fixed effect, so the coefficients are with respect to small or large farm households within a village as the reference group.

Table A3: Testing the Differential Skilling Explanation

	(1) Schooling Years	(2) Share Salaried	(3) Share Family Farm Labor	(4) Total HH Income
Mid-Sized x Dist (km)	-0.00457	-0.0000240	0.000579	203.7
	(0.0106)	(0.000216)	(0.000599)	(175.2)
Mid-Sized Farm (2005)	0.536***	-0.00735*	0.0216*	-2887.7
	(0.196)	(0.00396)	(0.0115)	(3266.2)
Observations	15573	15573	15573	15572
Village Fixed Effect	X	X	X	X

Standard errors in parentheses
* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Data in this table are from the IHDS Wave 1. Dist (km) is the distance in kilometers to the nearest town. Farm size classifications are as per the household landholding sizes in Wave 1 of the IHDS survey (in 2005). Column 1 reports the correlations using maximum years of schooling among the household members as the dependent variable. Columns 2 and 3 report correlation using share of household members engaged in salaried and own farm labor, respectively, as the dependent variables. Column 4 examines total household income's correlation with farm size and distance interaction. All specifications include village fixed effect, so the coefficients are with respect to small or large farm households within a village as the reference group.

Table A4: Potential Omitted Variables

	(1) FAO Rice Suitability	(2) FAO Cotton Suitability	(3) Surface Water Availability	(4) Ground Water Availability	(5) Non-Agri Vill Area (Percent)	(6) Percent Change Non-Agri Employment
Dist (10 km)	3.044	-0.0467	0.000640	-0.00758	-0.293	12.87
	(2.019)	(0.0482)	(0.00665)	(0.00557)	(0.272)	(7.878)
Observations	10686	10686	10686	10686	10668	7148
Sub-District Fixed Effect	X	X	X	X	X	X
Town Fixed Effect	X	X	X	X	X	X

Standard errors in parentheses
* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: The dependent variables are village-level characteristics, amenities, and non-agricultural opportunities, respectively, from data pertaining to the study state as described in Table 1. As before, distance to the nearest town is expressed in terms of multiples of 10 km. The specifications include sub-district and town fixed effect and standard errors are clustered by the nearest town.