Unlanded: Distribution of Land in Bihar

Lucas Shen* Gaurav Sood†

January 16, 2026

Abstract

Land is the primary store of wealth in rural India. We use newly digitized administrative land records from Bihar, India's poorest and among its most rural states, to document the distribution of land ownership. More than forty years after rural land reforms that were geared toward redistributing land more equitably, we find that 60% of the people own 18.5% of the land. The land is owned primarily by men—76.5% of the landowners are men. Muslims comprise 7.7% of matched ownership accounts versus 17.7% of Bihar's population. Scheduled Castes and Extremely Backward Classes are also substantially underrepresented. Conditional on being a land owner, women's holdings look broadly similar to men's, while non-Muslim and upper-caste owners are more prevalent in the upper tail.

Keywords: Land concentration, Caste inequality, Measurement, India
JEL codes: Q15, J15, J16, D31, C81

*Email: lucas@lucasshen.com, National University of Singapore
†Email: gsood07@gmail.com

1 Introduction

Land remains the central store of rural wealth and political power in much of India. Who owns land—and how much—shapes access to credit, the ability to weather economic shocks, and bargaining power in local labor markets. These linkages are especially salient in Bihar, where caste, class, and landholding have been deeply intertwined since the colonial zamindari system.

Despite land's importance, basic facts about *ownership* are not readily available at fine levels of geography or for important social groups. Much of the empirical literature relies on the Agricultural Census and related official statistics, which are organized around the concept of an *operational holding*—the land operated as a single technical unit, regardless of ownership and inclusive of leased-in/leased-out cultivation arrangements. Operational holdings are indispensable for understanding production, but they do not map cleanly to de jure ownership, and the gap between operation and ownership can be large precisely where tenancy and informality are pervasive (Rawal 2008). Household surveys can measure ownership more directly, but they are typically not designed to estimate land concentration (and its correlates) at the level of districts or smaller administrative units, nor do they always support detailed group-wise comparisons at scale.

This paper complements existing survey- and census-based evidence by using administrative microdata from digitized land records. We assemble plot-level Records of Rights (RoR) from the Government of Bihar's land records portal and use these records to describe the distribution of recorded land ownership in Bihar. Because these data enumerate individual plots and the associated account holders, they enable high-resolution descriptive facts about concentration in ownership and how ownership differs across groups. Our empirical focus is deliberately descriptive: we document how land is distributed across owners in Bihar and how that distribution varies by gender, religion, and broad caste group.

Our analysis contributes to long-standing debates about inequality in land rights. A substantial body of work emphasizes how gender norms and legal practice jointly restrict women's effective land ownership and control (Agarwal 1988, 1994; Rao 2008). Recent work using digitized land records documents the continuing gender gap in recorded ownership and highlights the promise (and limitations) of administrative data for monitoring progress (Jain et al. 2023). Parallel research underscores the social patterning of land inequality, including disparities across caste and related social categories (Bakshi 2008). For Bihar specifically, evidence of economic inequality at socially meaningful levels of

stratification motivates careful measurement of assets and entitlements (Joshi et al. 2018). Our contribution is to provide a transparent, reproducible description of recorded land ownership in Bihar at a scale that is difficult to achieve with conventional surveys.

2 Background and Motivation

Land is the principal productive asset in rural India, constituting over 85 percent of household wealth (Government of India 2021). It shapes household livelihoods (Ellis 2000), access to credit (Narayanan and Chakraborty 2019), and intergenerational wealth transmission (Agarwal 1994). The relationship between land and social stratification is not merely correlational but constitutive: caste hierarchy was historically expressed through, and reinforced by, differential access to land, with upper castes dominating ownership while lower castes were relegated to landless labor or marginal tenancy (Chakravarti 2001; Bakshi 2008). Gender compounds these patterns. The Hindu Succession (Amendment) Act of 2005 granted daughters equal inheritance rights, yet customary norms governing inheritance, marital residence, and administrative practice continue to constrain women's access (Rao 2008): women constitute only 14 percent of rural landowners and hold 11 percent of agricultural land (Agarwal et al. 2021). Understanding who owns how much land and how ownership is distributed across caste, gender, and religion is therefore central to diagnosing inequality and designing policy responses.

Given the depth of caste and gender stratification in land ownership, a natural question is whether policy can reshape these distributions. Cross-state panel evidence suggests that legislated land reforms are associated with poverty reduction (Besley and Burgess 2000), and studies of specific reform episodes find meaningful effects: tenancy reform in West Bengal strengthened tenant rights and raised productivity through improved bargaining power and tenure security (Banerjee et al. 2002). Yet implementation has often been partial and politically constrained, with substantial gaps between legislative intent and on-the-ground outcomes (Bandyopadhyay 1986).

Bihar presents an especially stark case. Despite being the first Indian state to abolish zamindari (1950), land reform implementation was systematically undermined by legislative loopholes, delayed enforcement, and political capture by landed interests (Jannuzi 1974; Frankel 2005). The Bihar Land Reforms Commission (2008) found that rural landlessness actually *increased* from 67 percent in 1993–94 to 75 percent by 2001. Village studies document extreme concentration: in two North Bihar villages surveyed by the Project on Agrarian Relations in India (PARI), the top 5 percent of households owned

72–74 percent of all land, while the bottom 50 percent owned nothing (Kumar 2022). In Nayanagar village, Bhumihar households, who make up 25 percent of the population, controlled 97.5 percent of land; Scheduled Caste households, comprising 34 percent of households, owned just 0.2 percent (Kumar 2022). These patterns reflect the tight overlap between caste and class in Bihar's agrarian structure, where social hierarchy is inscribed in the distribution of productive assets (Chakravarti 2001; Prasad 2021). Research further shows that in Bihar, where stratification operates at the level of jati, economic inequality varies substantially even within broad administrative categories (Joshi et al. 2018).

But despite decades of research on land inequality, systematic granular evidence on ownership distributions by social group remains sparse—a gap rooted in fundamental limitations of existing data sources. The quinquennial Agricultural Census measures *operational holdings*—land cultivated as a single technical unit, regardless of title—rather than ownership per se. Because operational holdings incorporate tenancy, sharecropping, and management arrangements, they can diverge substantially from ownership distributions; Rawal (2008) demonstrates that conflating these concepts leads to systematic misinterpretation of land concentration. Moreover, the Agricultural Census excludes landless households entirely and does not record religion.

Household surveys such as the NSS and IHDS collect self-reported ownership alongside demographic characteristics, enabling analysis by caste and gender at aggregate levels (Bakshi 2008), but sample sizes preclude the distributional comparisons—the 90th percentile of SC owners versus UC owners, for instance—needed to characterize inequality within and across groups at finer geographic scales. The Socio-Economic and Caste Census (SECC) of 2011–12 collected both caste status and landholding at the household level, but the government released these variables separately rather than as cross-tabulations; the detailed OBC and caste-specific data were never made public, with officials citing data quality concerns. As a result, no existing public data source provides distributions of land ownership by detailed caste category at scale.

India's digitization of land records under the Digital India Land Records Modernization Programme (DILRMP)[1] creates new research possibilities. Unlike surveys, administrative Records of Rights approach universe coverage and contain demographic fields—including caste (jati) entries and owner names—that permit disaggregation by social group. Jain et al. (2023) pioneered this approach, analyzing 16,000 digital land record copies from 12 states to document gender bias in recorded ownership; they found that even when women

[1]Originally launched as the National Land Records Modernization Programme (NLRMP) in 2008, the program was revamped as DILRMP in 2016 under the Department of Land Resources, Ministry of Rural Development. See https://dolr.gov.in/.

hold titles, they face discrimination through limited access to single titles, smaller plot sizes, and inferior land quality. Such administrative data can reveal patterns that surveys cannot capture—not just average differences but the entire distribution of holdings within and across groups.

3 Data

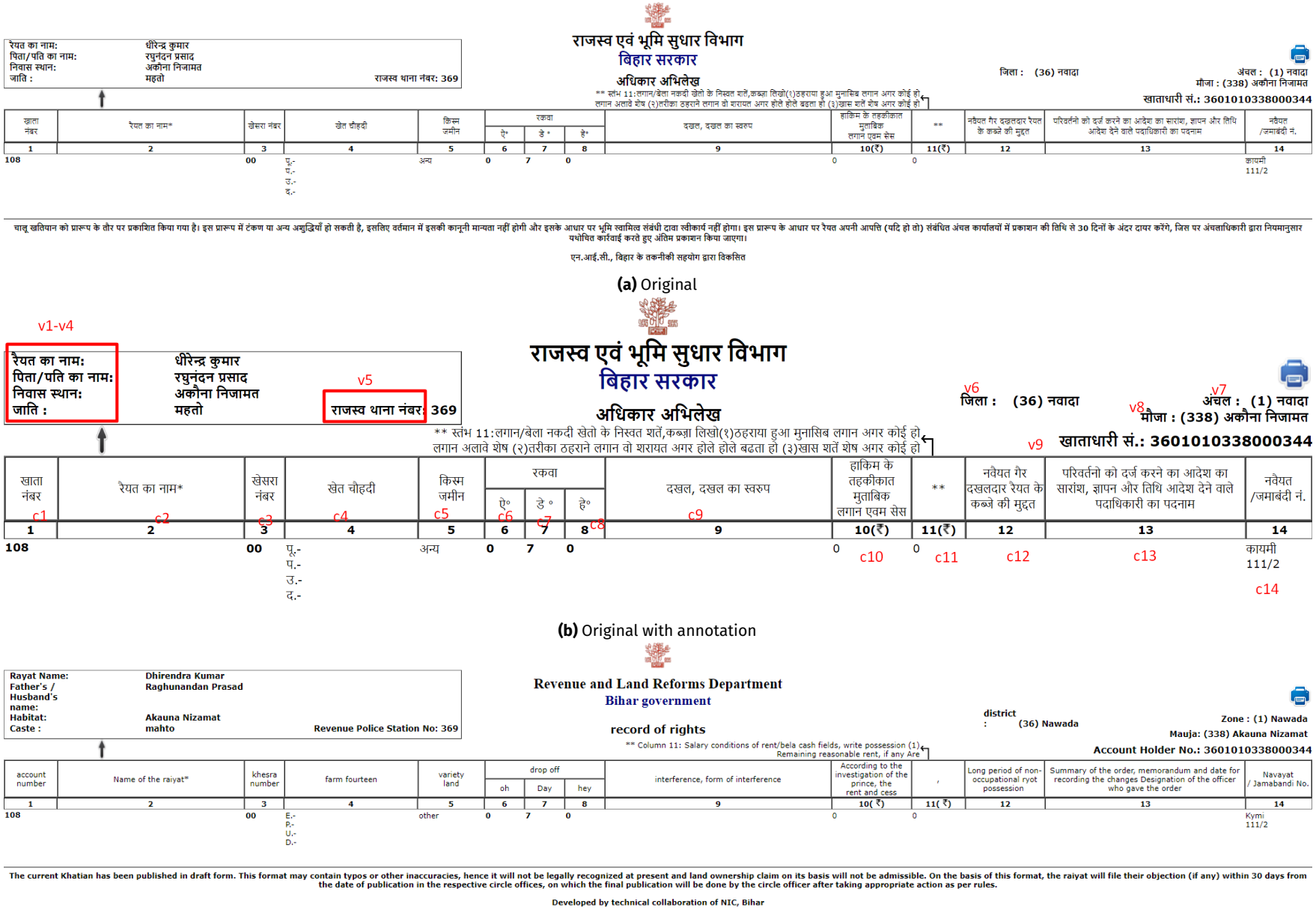
The foundational data for this study are Bihar's digitized Records of Rights (RoR), accessed through the Government of Bihar land records portal (Government of Bihar 2025). Digitization in India has been pursued through successive national programs and is currently organized under DILRMP (Department of Land Resources, Ministry of Rural Development, Government of India 2021). We scraped these records in 2022; the data are archived at (Sood 2022). In total, we obtained approximately 41.87 million plot records and 12.12 million ownership accounts.

Each plot record contains the account number, owner name (ryot/raiyat), father's name, residence, district, and jati entry. The jati field records caste subgroup for Hindus and religion for non-Hindus. Table 1 summarizes the available fields; Figure 1 shows an example record.

Some plots in the raw data have zero or negative recorded area, likely reflecting data-entry errors or placeholder entries. We drop these observations, bringing our analysis sample to approximately 38.58 million plots and 11.90 million accounts. We also trim the top 1 percent of the area distribution in figures, as extreme outliers exist in the raw register (e.g., the maximum recorded area exceeds eight million acres); these outliers do not affect percentiles below the 99th. Appendix A provides additional details on data cleaning.

The two primary metrics of recorded land ownership we examine are total area (in acres) and the number of plots. We compute area of the plot, we use the fields 6–7 Table 1), and calculate $Acre + \frac{Decimal}{100}$, consistent with the portal's acre/decimal reporting. The hectare field is used only for reference and we ignore it. (This is based on conversations with land administration officials.) We treat each RoR entry as a plot and use the account holder number as the unit of recorded ownership when describing distributions across owners. Figure 2 visualizes recorded land accounts across the Bihar districts.

We infer owner gender using the `naampy` package (Laohaprapanon et al. 2022), which provides a deep learning classifier trained on Indian Electoral Rolls data. The original names are recorded in Hindi; we transliterate to English using `indicate` (Chintalapati



(a) Original

(b) Original with annotation

(c) Google translated

Figure 1. Example of Bihar land record. Screenshot of originating source on `http://land.bihar.gov.in`. Google translation of Hindi to English comes with errors. See Sood (2022). Please see Table 1 for a description of selected annotated fields.

and Sood 2022a) when required for downstream inference (Appendix A).

Caste information comes directly from the jati field in the land records. We clean and standardize these entries through a multi-step process: transliterating 3,435 unique Hindi entries to 2,468 English forms, consolidating variant spellings into 129 standardized caste names (e.g., mapping both "brahmana" and "Brahmin" to Brahmin). We then map the caste names to caste codes (e.g., BC-1/007) before finally mapping them into five broad administrative categories: General or Upper Caste (UC), Backward Caste 2 (BC2), Backward Caste 1 or Economically Backward Caste (EBC), Scheduled Caste (SC), and Scheduled Tribe (ST). Using this approach, we successfully match 26.2 million land records and 7 million accounts to one of the five caste groups. For non-Hindus, the jati field encodes religion rather than caste; we use this to distinguish Muslim from non-Muslim owners in religion-based analyses.

Table 1. *Fields available in Bihar land records*

	Field	Description
v1	name of ryot	Name of the plot owner or landholder (ryot/raiyat) - the person in whose name the land is registered
v2	name of father	Name of the owner's father or husband
v3	residence	Address or location where the owner resides
v4	jati	Caste and subcaste of the owner
v5	revenue unit	Revenue Police Station number - administrative jurisdiction for land record management
v6	district	District name and code where the land is located
v7	zone/anchal	Subdivision or zone level (anchal), an administrative unit between district and village
v8	mouza	Village or revenue village name (mouza) where the land parcel is situated
v9	account no.	Owner's unique account number used to identify and track all land holdings in the revenue system
c6	plot area (decimal)	Land area measured in decimal units
c7	plot area (acre)	Land area measured in acres
c8	plot area (hectare)	Land area measured in hectares

Note: *See Figure 1 for an example of a land record and the information available.*

For descriptive comparisons to population baselines, we use the 2023 Bihar Caste-Based Survey (Government of Bihar 2023), which provides population shares by caste category for Bihar. District boundary data are from the SHRUG platform (Asher et al. 2021).

4 Ownership of Land

We describe the distribution of recorded land ownership in Bihar using two complementary measures: the number of plots linked to an ownership account and the total

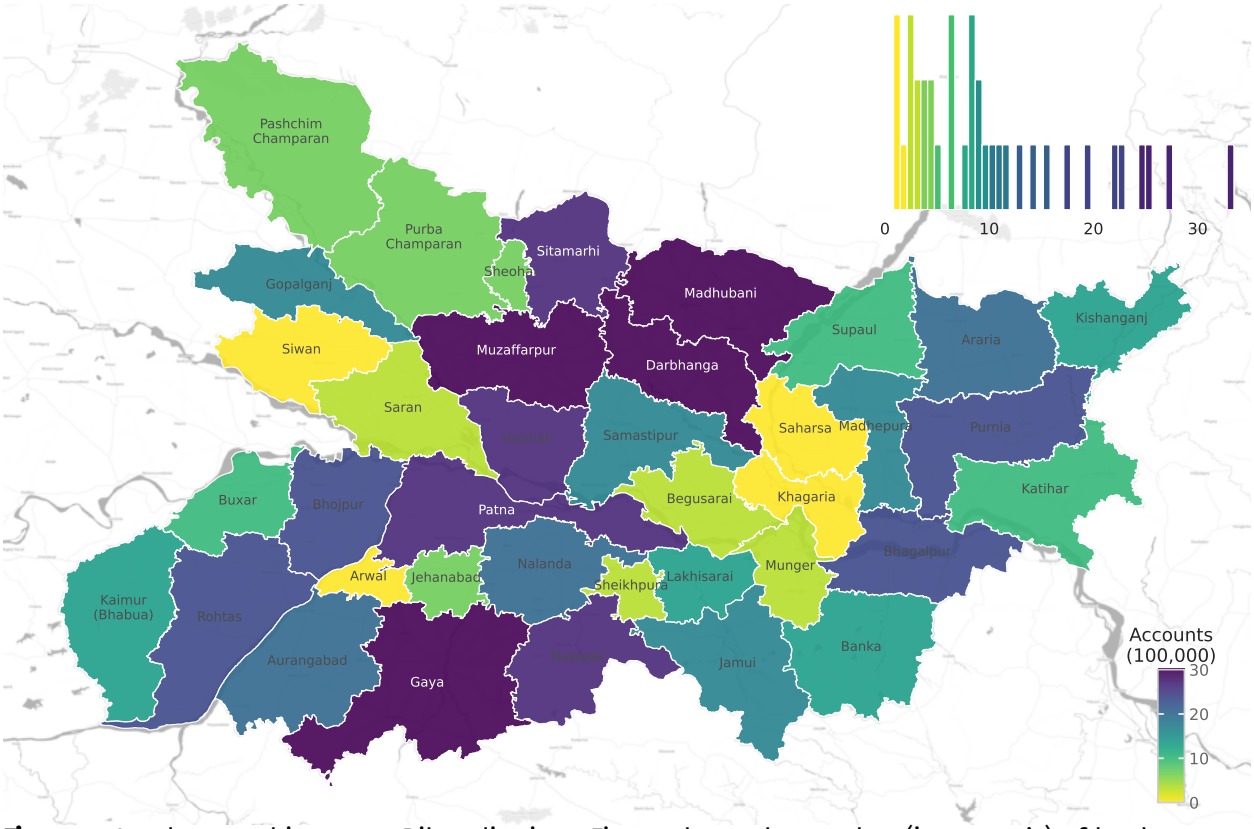


Figure 2. Land ownership across Bihar districts. Figure shows the number (in 10,000's) of land account holders in each of the 38 districts.

recorded area (in acres) summed over those plots.[2] We first characterize the overall distribution, then examine differences across gender, religion, and caste.

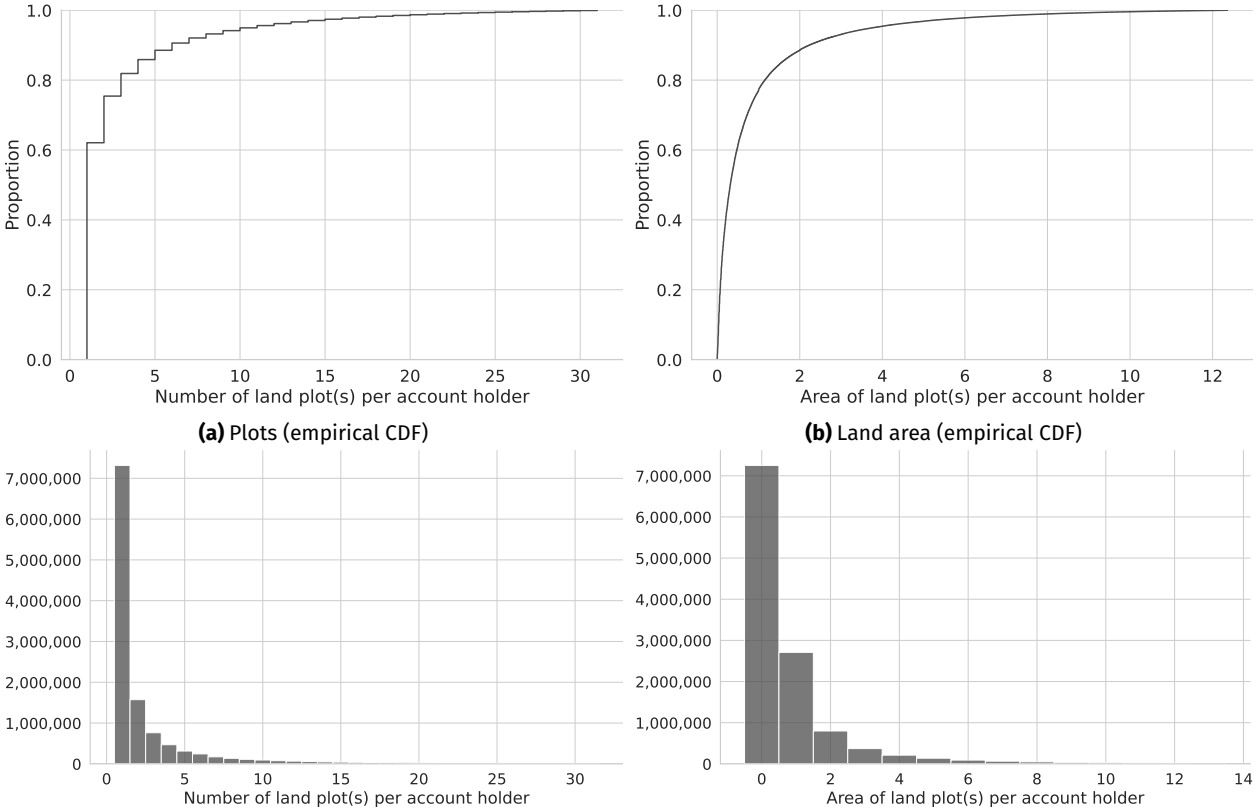
Figure 3 displays the distributions (trimmed at the 99th percentile); Tables A1 and A2 report percentiles across the full range. Both distributions are sharply right-skewed. The median account holds one plot and 0.32 acres. At the 90th percentile, an account holds seven plots totaling 2.42 acres; at the 99th percentile, 31 plots totaling 12.36 acres. Relative to the median, the 90th percentile holds roughly 7 times more land and the 99th percentile roughly 40 times more.

The concentration of plots is similarly pronounced. The 60th percentile account still holds just one plot, meaning the bottom 60 percent of accounts—approximately 7.2 million—collectively hold only 18.5 percent of all plots.[3] The 70th percentile holds two plots, the 80th three, and the 90th seven. At the 99th percentile, accounts hold 31 plots. This pattern—a large mass of single-plot holders alongside a small number of accounts

[2] An "account" is the administrative unit in the land register and need not map one-to-one to a unique individual or household.

[3] This calculation concerns the share of plots, not area. Converting to area shares requires information on plot sizes, which vary across the distribution.

with many plots—is the baseline against which we assess group-level differences below.



(a) Plots (empirical CDF)

(b) Land area (empirical CDF)

(c) Plots (histogram)

(d) Land area (histogram)

Figure 3. Distribution of land in Bihar. From ∼38.65 million land records and ∼11.9 million account holders. Column 1 is the number of plots per account holder. Column 2 is the total area of plot(s) per account holder. Row 1 shows the empirical cumulative distributions. Row 2 shows the histograms. Area unit is in acres. All subfigures have the 99th percentile trimmed for visual articulation. See Table A1 and Table A2 for the percentile values at the right tail, including the 99th percentile.

5 Land Ownership by Group

Comparisons across groups are clearer when we separate two margins: (i) *the extensive margin*—what percentage own the land compared to baseline—and (ii) *the intensive margin*—how many plots and how much area recorded owners hold. We benchmark representation on the extensive margin against population base rates from Bihar's caste-based survey (Government of Bihar 2023) and, for gender, adult population shares from the 2011 Census of India.

Table 2. *Benchmarking representation in the land register against population base rates*

Group	Population share	Share among recorded land owners	Representation ratio
Gender			
Women	47.7%	21.7%	0.45
Men	52.3%	78.3%	1.50
Religion			
Muslim	17.70%	6.4%	0.36
Non-Muslim	82.30%	83.6%	1.14
Caste			
Unreserved / General (UC)	15.52%	36.0%	2.32
Backward Class (BC)	27.12%	38.6%	1.42
Extremely Backward Class (EBC)	36.01%	17.6%	0.49
Scheduled Castes (SC)	19.65%	6.8%	0.35
Scheduled Tribes (ST)	1.68%	1.0%	0.60

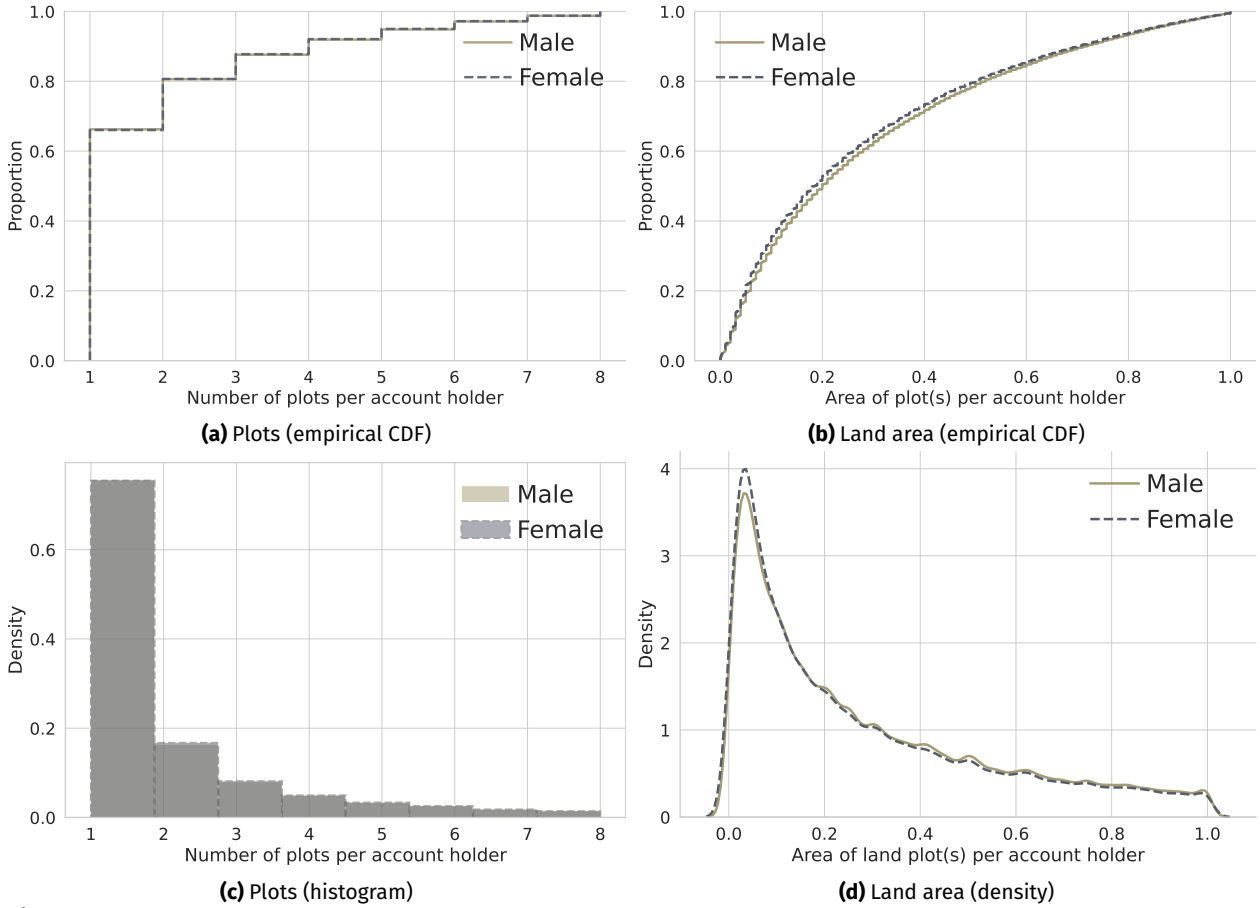
Note: *Population shares for caste and religion come from Bihar's caste-based survey released in October 2023 (Government of Bihar 2023). Because the land register often records religion rather than caste for non-Hindus, caste owner shares in this table are computed within the Hindu subset with standardized caste categories; the resulting caste representation ratios should be interpreted as descriptive benchmarks rather than exact landownership rates. Gender population shares (age 20+) are from the 2011 Census of India. "Share among land recorded owners" uses the land register subsamples for which the corresponding attribute can be inferred or standardized. The representation ratio is the landowner share divided by the baseline population share; it is* not *an absolute landownership rate because the land register does not cover landless households. Figure A2 shows an alternative visualization.*

5.1 Inequality by gender

We infer gender for approximately 11.9 million ownership accounts linked to 38.6 million plot records. Women constitute 21.7% of matched accounts and 19.6% of matched plots—less than half their 47.7% share of the adult population (Table 2).

On the intensive margin, however, the distributions among recorded owners are similar by gender (Figure 4). The median woman and man each hold one plot; at the 80th percentile, both hold three plots; at the 99th percentile, women hold 28 plots versus 32 for men (Table A3). For land area, the median woman holds 0.28 acres compared to 0.33 for men. At the 90th percentile, women hold 2.13 acres versus 2.50 for men; at the 99th, 11.25 versus 12.66 acres (Table A4).

Gender gaps in recorded land ownership are thus driven primarily by the extensive margin—who appears in the register—rather than by differences in holdings among those recorded.



(a) Plots (empirical CDF)

(b) Land area (empirical CDF)

(c) Plots (histogram)

(d) Land area (density)

Figure 4. Distribution of land in Bihar by gender. From ∼22.9 million land records and ∼6.78 million account holders. Column 1 is the number of plots per account holder. Column 2 is the total area of plot(s) per account holder. Row 1 shows the empirical cumulative distributions. Row 2 shows the histograms. Area unit is in acres. The density plots are trimmed for visual articulation. See Table A3 and Table A4 for the percentile values at the right tail, including the 99th percentile.

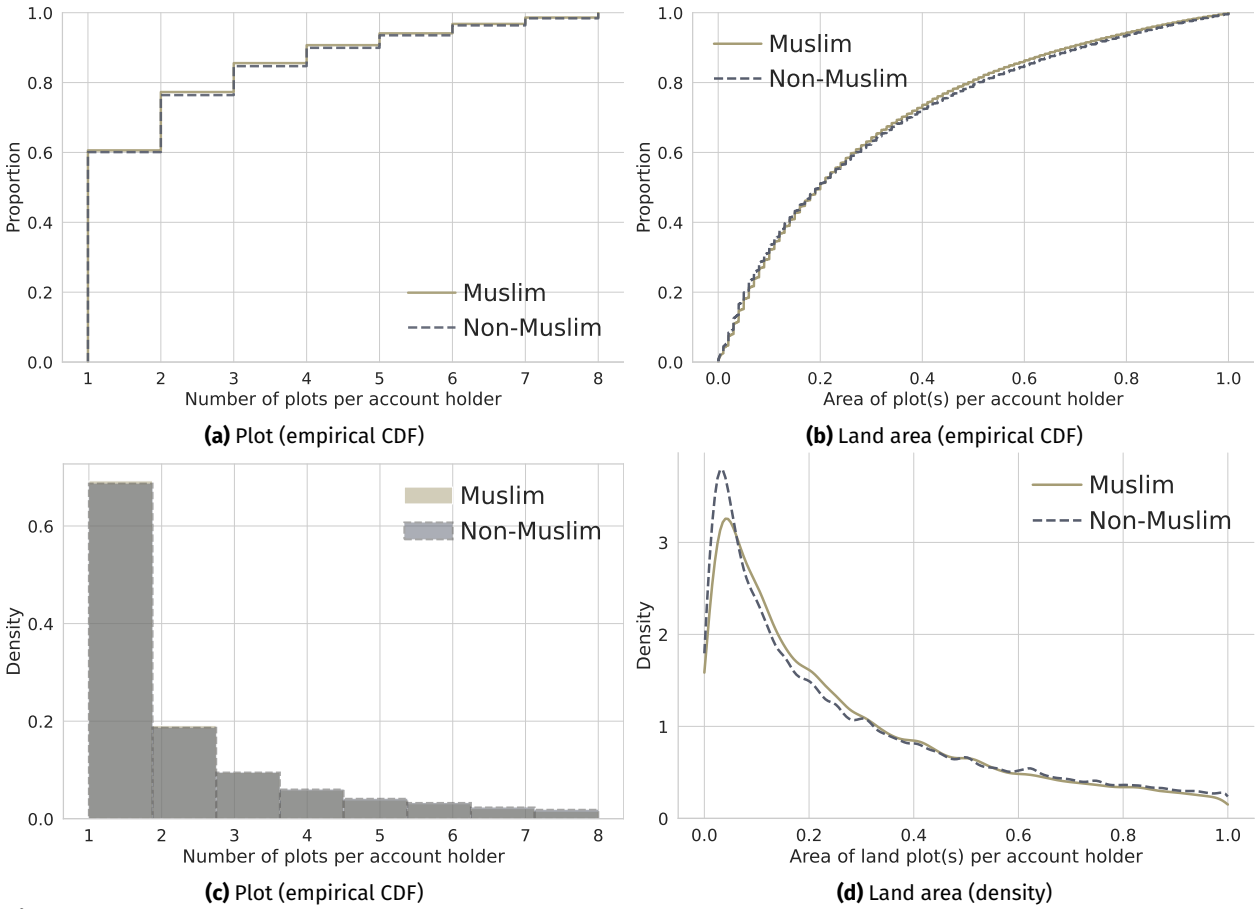
5.2 Inequality by religion

We classify religion for approximately 6.8 million ownership accounts linked to 25.3 million plot records. Muslims account for 6.4% of matched accounts, less than half their 17.7% population share (Table 2).

On the intensive margin, Muslim and non-Muslim owners are similar in the lower and middle portions of the distribution but diverge in the upper tail (Figure 5). The median owner in both groups holds one plot. At the 90th percentile, Muslim owners hold seven plots versus nine for non-Muslims; at the 99th, 27 versus 35 (Table A5). For land area, the median Muslim owner holds 0.29 acres compared to 0.34 for non-Muslims. At the 90th percentile, Muslims hold 2.0 acres versus 2.63; at the 99th, 8.96 versus 12.37 acres (Table A6).

Religious gaps thus operate on both margins: Muslims are underrepresented among

recorded owners and, conditional on ownership, hold less land—particularly in the upper tail of the distribution.[4]



(a) Plot (empirical CDF) **(b)** Land area (empirical CDF)

(c) Plot (empirical CDF) **(d)** Land area (density)

Figure 5. Distribution of land in Bihar by religion. From ∼25.3 million land records and ∼6.8 million account holders. Column 1 is the number of plots per account holder. Column 2 is the total area of plot(s) per account holder. Row 1 shows the empirical cumulative distributions. Row 2 shows the histograms. Area unit is in acres. The density plots are trimmed for visual articulation. See Table A5 and Table A6 for the percentile values at the right tail, including the 99th percentile.

5.3 Inequality by caste

We standardize caste for approximately 7.0 million Hindu ownership accounts linked to 26.2 million plot records. Because the land register often records religion rather than caste for non-Hindus, we restrict caste comparisons to Hindu accounts with standardized categories.

On the extensive margin, representation diverges sharply from population shares (Table 2). Upper Caste (UC) owners constitute 36.0% of matched accounts despite being

[4]We find similar distributions using a larger sample recovered from a neural model that infers religion by name (Appendix A).

15.5% of the population. Backward Class (BC) owners are similarly overrepresented (38.6% of owners versus 27.1% of population). By contrast, Extremely Backward Class (EBC) owners are underrepresented (17.6% versus 36.0%), and Scheduled Caste (SC) owners severely so (6.8% versus 19.7%).

On the intensive margin, distributions differ substantially across all percentiles (Figure 6). For plot counts, the median UC owner holds two plots while median SC, EBC, and BC owners hold one. At the 90th percentile, UC owners hold 11 plots, BC owners 9, EBC owners 6, and SC owners 4. At the 99th percentile, these values are 40, 36, 25, and 17, respectively (Table A7).

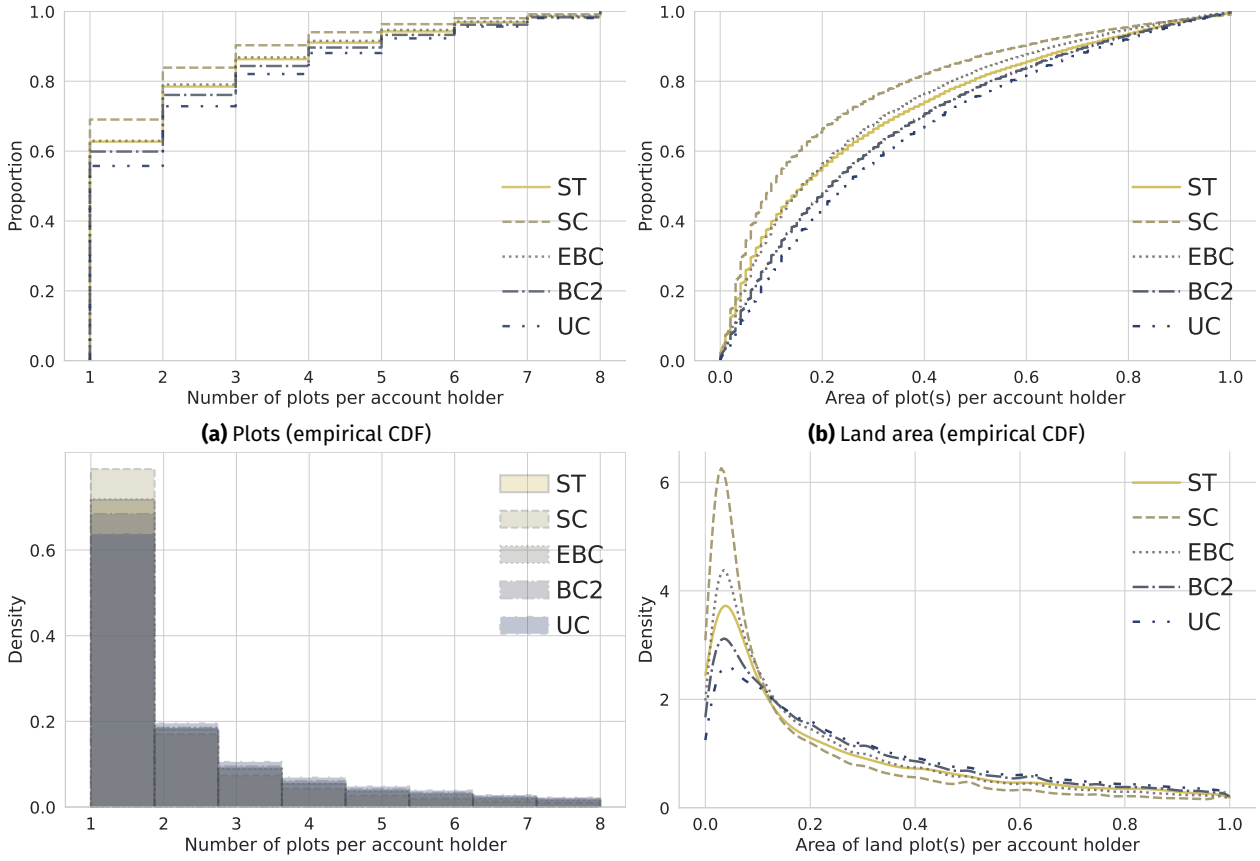
Differences are starker for land area (Table A8). At the median, UC owners hold 0.49 acres compared to 0.36 for BC, 0.23 for EBC, and 0.13 for SC. At the 90th percentile, UC owners hold 3.87 acres versus 1.05 for SC—a ratio of 3.7. At the 99th percentile, 16.46 versus 4.67 acres. These percentile-by-percentile comparisons show that caste differences are not confined to the extreme upper tail: the entire distribution of holdings is shifted rightward for UC and BC relative to EBC and SC.

Caste gaps thus operate on both margins and are the largest of the three dimensions we examine. UC households are overrepresented among owners by a factor of more than two, and conditional on ownership, hold substantially more land at every point in the distribution.

6 Discussion

Using approximately 38.7 million plot records linked to 11.9 million account holders, we document a sharply right-skewed distribution of recorded land ownership in Bihar. The median account holder owns 0.34 acres, while the 99th percentile holds 12.9 acres. This pattern is consistent with an agrarian economy in which a large mass of smallholders coexists with a comparatively small number of large recorded owners.

Three considerations shape how these statistics should be read. First, Records of Rights measure de jure recorded claims rather than de facto control. Where tenancy, informal cultivation, or intra-household arrangements are common, ownership and operation can diverge; RoR-based inequality is thus best understood as inequality in recorded claims—economically meaningful for collateral, program eligibility, and bargaining power, but not identical to cultivated area. Second, digitization scales legacy measurement issues rather than resolving them. Our need to drop plots with zero or negative recorded area, and the non-trivial share of missing or ambiguous jati entries, are reminders that



(a) Plots (empirical CDF) **(b)** Land area (empirical CDF)

(c) Plots (histogram) **(d)** Land area (density)

Figure 6. Distribution of land in Bihar by caste. From ∼26.2 million land records and ∼7.0 million account holders. Column 1 is the number of Plots per account holder. Column 2 is the total area of Plot(s) per account holder. Row 1 shows the empirical cumulative distributions. Row 2 shows the histograms. Area unit is in acres. The density plots are trimmed for visual articulation. See Table A7 and Table A8 for the percentile values at the right tail, including the 99th percentile.

digitization changes the form of errors as much as it expands access. Third, the right tail warrants particular caution: large holdings may reflect genuine concentration, but could also reflect delayed mutation after inheritance, aggregation of family land under a single account, or heterogeneous recording practices across districts.

The gender pattern illustrates the value of distinguishing extensive from intensive margins. Women account for roughly one-quarter of account holders—about half their share of the adult population—yet conditional on ownership, the distributions of plots and area are broadly similar to men's through most of the distribution, with differences emerging primarily in the upper tail. This combination suggests a land rights regime in which women are less likely to appear on titles even when households possess land, consistent with evidence that legal reforms granting equal inheritance have not translated into equal recorded ownership due to persistent customary norms and administrative barriers (Agarwal 1994; Rao 2008; Jain et al. 2023).

For caste and religion, we find differences on both margins. Upper Caste owners are overrepresented relative to population (36 percent of owners versus 15.5 percent of population) and hold more land conditional on ownership at every percentile. Scheduled Caste owners are severely underrepresented (6.8 percent versus 19.7 percent) and hold less throughout the distribution. Muslims are similarly underrepresented (7.7 percent versus 17.7 percent), with smaller holdings concentrated in the upper tail. These patterns are consistent with historical accounts of caste-based land concentration in Bihar (Chakravarti 2001; Kumar 2022).

However, several compositional factors complicate causal interpretation of these group differences. Age structure varies across groups: Muslims in India have a younger age distribution due to higher fertility, meaning a larger share has not yet reached the life-cycle stage at which land is typically inherited or accumulated (Sowell 2019; Altonji and Blank 1999). Urbanization and occupational patterns also differ: Muslims have historically been more concentrated in urban areas and in trades such as weaving and commerce that do not involve agricultural landholding (Sachar et al. 2006), so lower recorded ownership of agricultural land may partly reflect residential and occupational sorting rather than exclusion from land per se. Geographic concentration adds further complexity, as Scheduled Tribes are located in specific districts with distinct tenure histories, and caste composition varies across Bihar's agro-climatic zones. These confounders do not invalidate the descriptive patterns—they are real features of the land register—but they caution against interpreting raw group comparisons as direct estimates of discrimination or dispossession. Disentangling compositional effects from ownership gaps that persist within age, location, and occupation cells would require individual-level data that the land register does not provide.

Two directions for future work follow naturally. The first is linkage: combining plot-level ownership with village- or household-level socioeconomic data would permit measurement of landlessness rates by group and analysis of how recorded ownership predicts welfare and program access. The second is granularity: moving beyond broad reservation categories toward jati-level estimates could reveal heterogeneity masked by administrative groupings (Joshi et al. 2018), though such analysis requires careful validation given that names, migration, and inter-district variation can induce systematic misclassification. More broadly, Bihar's digitized Records of Rights offer an unusually rich measurement opportunity. Using these data well requires treating land administration not as a neutral measurement device but as an institution whose incentives, frictions, and historical legacies shape what becomes data in the first place.

References

Bina Agarwal. Who sows? who reaps? women and land rights in india. *The Journal of Peasant Studies*, 15(4):531–581, 1988. doi:10.1080/03066158808438377.

Bina Agarwal. *A Field of One's Own: Gender and Land Rights in South Asia*. Cambridge South Asian Studies. Cambridge University Press, Cambridge, 1994.

Bina Agarwal, Pervesh Anthwal, and Malvika Mahesh. How many and which women own land in india? inter-gender and intra-gender gaps. *Journal of Development Studies*, 57 (11):1807–1829, 2021. doi:10.1080/00220388.2021.1887478.

Joseph G. Altonji and Rebecca M. Blank. Race and gender in the labor market. *Handbook of Labor Economics*, 3:3143–3259, 1999.

Sam Asher, Tobias Lunt, Ryu Matsuura, and Paul Novosad. Development research at high geographic resolution: An analysis of night-lights, firms, and poverty in india using the SHRUG open data platform. *The World Bank Economic Review*, 35(4):845–871, 2021. doi:10.1093/wber/lhab003.

Aparajita Bakshi. Social inequality in land ownership in india: A study with particular reference to west bengal. *Social Scientist*, 36(9/10):95–116, 2008.

D. Bandyopadhyay. Land reforms in india: An analysis. *Economic and Political Weekly*, 21 (25/26):A50–A56, 1986. URL https://www.jstor.org/stable/4375823.

Abhijit V. Banerjee, Paul J. Gertler, and Maitreesh Ghatak. Empowerment and efficiency: Tenancy reform in west bengal. *Journal of Political Economy*, 110(2):239–280, 2002. doi:10.1086/338744.

Timothy Besley and Robin Burgess. Land reform, poverty reduction, and growth: Evidence from india. *The Quarterly Journal of Economics*, 115(2):389–430, 2000. doi:10.1162/003355300554809.

Anand Chakravarti. Caste and agrarian class: A view from bihar. *Economic and Political Weekly*, 36(17):1449–1462, 2001.

Rajashekar Chintalapati and Gaurav Sood. Indicate: Transliterate indic languages to english, August 2022a. URL https://github.com/in-rolls/indicate. Accessed 2025-12-24.

Rajashekar Chintalapati and Gaurav Sood. pranaam: Predict religion based on name, May 2022b. URL https://github.com/appeler/pranaam. Accessed 2025-12-24.

Department of Land Resources, Ministry of Rural Development, Government of India. Guidelines of digital india land records modernization programme (DILRMP), 2021–2026. Technical report, Ministry of Rural Development, Government of India, 2021. URL https://landrevenue.rajasthan.gov.in/content/dam/landrevenue/settlementdepartment/pdffolder/Scheme%20and%20Section/DILRMP/Final%20Guidelines%20of%20DILRMP%202021-26.pdf. Accessed 2025-12-24.

Frank Ellis. *Rural Livelihoods and Diversity in Developing Countries*. Oxford University Press, 2000.

Francine R. Frankel. *India's Political Economy, 1947–2004*. Oxford University Press, 2nd edition, 2005.

Government of Bihar. Bihar caste-based survey report 2023, October 2023. URL https://state.bihar.gov.in/. Released October 2, 2023 by Vivek Kumar Singh, Additional Chief Secretary. Data collection January–August 2023. Total population: 13.07 crore. OBC 27.13%, EBC 36.01%, SC 19.65%, ST 1.68%, General 15.52%.

Government of Bihar. Bihar bhumi / land records portal (RoR access), 2025. URL http://land.bihar.gov.in/. Accessed 2025-12-24.

Government of India. All india debt and investment survey: Nss 77th round. Technical report, National Statistical Office, Ministry of Statistics and Programme Implementation, 2021.

Charu Jain, Disha Saxena, Somnath Sen, and Deepak Sanan. Women's land ownership in india: Evidence from digital land records. *Land Use Policy*, 133:106835, 2023. doi:10.1016/j.landusepol.2023.106835.

F. Tomasson Jannuzi. *Agrarian Crisis in India: The Case of Bihar*. University of Texas Press, Austin, 1974.

Shareen Joshi, Nishtha Kochhar, and Vijayendra Rao. Jati inequality in rural bihar. Policy Research Working Paper 8512, World Bank, July 2018. URL https://documents.worldbank.org/curated/en/216201531237182176/jati-inequality-in-rural-bihar.

Awanish Kumar. Land and caste relations in north bihar: Observations from two villages. *Review of Agrarian Studies*, 12(1):139–177, 2022. doi:10.25003/RAS.12.01.0007.

Suriyan Laohaprapanon, Gaurav Sood, and Rajashekar Chintalapati. naampy: Infer sociodemographic characteristics from indian names, July 2022. URL https://github.com/appeler/naampy. Accessed 2025-12-24.

Sudha Narayanan and Judhajit Chakraborty. Land as collateral in india. *Economic & Political Weekly*, 54(23):50–56, 2019.

Indulata Prasad. Caste-ing space: Mapping the dynamics of untouchability in rural bihar, india. *CASTE: A Global Journal on Social Exclusion*, 2(1):132–152, 2021. doi:10.26812/caste.v2i1.232.

Nitya Rao. *"Good Women Do Not Inherit Land": Politics of Land and Gender in India*. Social Science Press, New Delhi, 2008. Distributed by Orient Blackswan.

Vikas Rawal. Ownership holdings of land in rural india: Putting the record straight. *Economic and Political Weekly*, 43(10):43–47, 2008. URL https://www.jstor.org/stable/40277229.

Rajindar Sachar et al. Social, economic and educational status of the muslim community of india. Technical report, Prime Minister's High Level Committee, Cabinet Secretariat, Government of India, New Delhi, 2006.

Lucas Shen. Forestplot, 2022. URL https://github.com/LSYS/forestplot. MIT license.

Gaurav Sood. Bihar land records (2022), 2022. URL https://doi.org/10.7910/DVN/BI4KZS.

Thomas Sowell. *Discrimination and Disparities*. Basic Books, New York, revised and enlarged edition, 2019.

Appendix

A Data: Bihar Land records

13 account holders across 14 land records have negative entries for decimals. We start with the digitized Bihar land records of 41,871,025 rows of land records from 12,127,297 individual accounts. The key focus is on the land metadata: acres, decimals, and hectares (field #6 #7 #8 in Table 1). We ignore the hectares field as it is only used for reference internally by the department. Land area is computed as acres + decimal (divided by 100). No land records have negative values for acres. 14 records have negative values for decimals, which we drop. 64,472 records (0.15%) across 51,342 accounts (0.42%) have decimals above 99. We drop such observations. Finally, 3,216,631 records (7.7%) across 1,079,669 accounts (8.9%) have a computed total land area of 0 acres, which we drop. The final sample used for analyses therefore includes 38,589,908 land records from 11,904,901 accounts.

The land records also come with excessively large numbers at the right tail (e.g., more than 100,000 acres per land (not per account)). 91 land records (0.0002%) have total acres exceeding 100,000 acres (approximately the 99.99th percentile). These come from 81 account holders (0.0007%). We apply a visual (soft) censor at the 99th percentile in the distribution graphs for visual clarity.

Additionally, we analyze the land records by religion, gender, and caste. To supplement our coding, we infer religion using the full Hindi name (see Figure A1 for the most common last names). 1129 accounts (0.003%) have no recorded names. These drop out, leaving 3,197,303 unique names across 11,904,466 accounts. We clean Hindi names to remove data entry artifacts and formatting characters. Non-Devanagari characters were removed, including English letters, numbers, punctuation marks (e.g., degree symbols, brackets, braces, exclamation marks, question marks), special symbols (ampersands, plus signs, hyphens, pipes), Devanagari digits, and Unicode control characters (zero-width joiners, directional formatting marks), so that only Devanagari script characters and spaces are retained. Multiple consecutive spaces were collapsed to single spaces, and leading/trailing whitespace was trimmed. Names that became empty strings after cleaning were excluded, leaving 3,193,161 unique names. We then infer religion by passing all 3,193,161 cleaned, unique Hindi names through a neural model that tokenizes the name and matches learned patterns from 4M Bihar land-record names to obtain “Muslim” vs. “Non-Muslim” labels Chintalapati and Sood (2022b). In all, we successfully classified

religion for 9,978,187 (83.8%) accounts linked to 33,377,090 land records (86.5%). Through this, we find similar distributional patterns (please see https://github.com/in-rolls/land).

We infer gender using English (transliterated) first names. First, we transliterate the 3,193,161 cleaned, unique Hindi names to English. We drop the 115 Hindi names with failed transliteration and extract English first names via token order to obtain 390,653 unique English first names. We then infer gender using `nammpy` (Laohaprapanon et al. 2022), which indexes the 390,653 first names against the Indian Electoral Roll to compute the proportion of registered males and females with those first names. For first names not found in the rolls, it uses a character-level neural network trained on the electoral labels. We label a first name as "female" when either the electoral-roll female proportion or the model's predicted female probability is > 0.5, and "male" otherwise. In all, we successfully classified gender for 11,904,801 accounts ($\sim$100%) linked to 38,589,535 land records ($\sim$100%).

Last name	Religion	Count
Devi	Non-Muslim	172,507
Singh	Non-Muslim	121,827
Yadavas	Non-Muslim	113,381
Siha	Non-Muslim	68,161
Sah	Non-Muslim	66,308
Ray	Non-Muslim	64,250
Mahato	Non-Muslim	62,392
Mandal	Non-Muslim	60,880
Khatun	Muslim	38,940
Thakur	Non-Muslim	35,708
Chaudhuri	Non-Muslim	34,429
Jha	Non-Muslim	33,695
Vagairah	Non-Muslim	33,431
Das	Non-Muslim	32,973
Paswan	Non-Muslim	32,127
Rama	Non-Muslim	30,579
Prasad	Non-Muslim	30,400
Mishra	Non-Muslim	28,002
Miyan	Muslim	27,387
Sharma	Non-Muslim	27,071
Kumar	Non-Muslim	21,529
Pande	Non-Muslim	21,130
Manjhi	Non-Muslim	21,019
Pandit	Non-Muslim	20,008
Miya	Muslim	19,406
Khan	Muslim	19,144
Ali	Muslim	18,590
Bhagat	Non-Muslim	18,348
Tiwari	Non-Muslim	17,441
Gopa	Non-Muslim	17,021

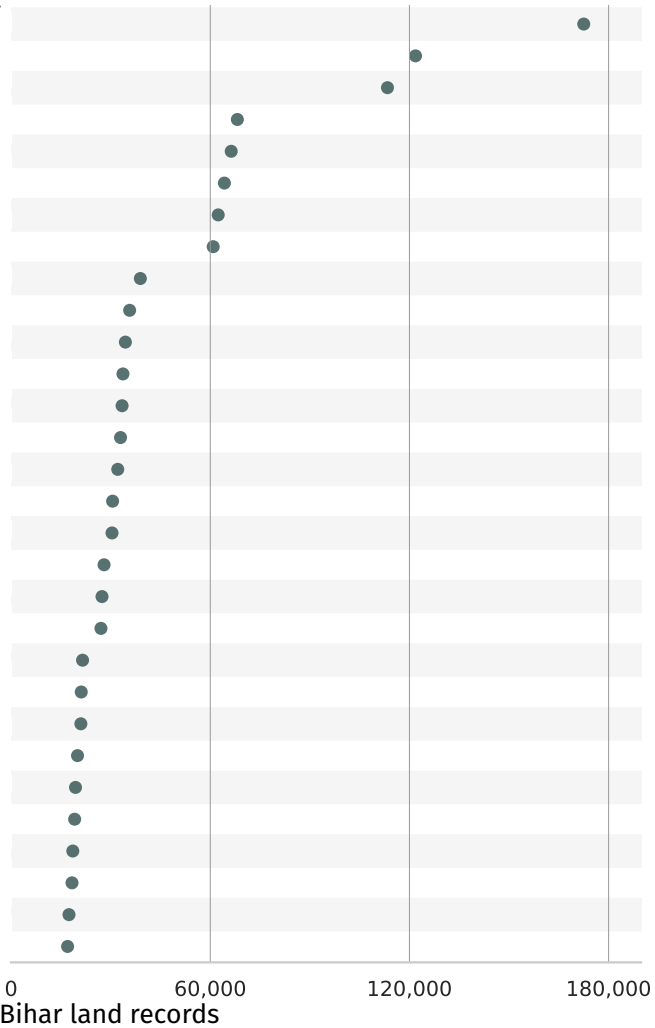


Figure A1. 30 most common last names in Bihar land records

Note: The figure reports the 30 most common (transliterated) last names found in the dataset Shen (2022). Religious classification is derived from the original Hindi names using *pranaam* (Chintalapati and Sood 2022b).

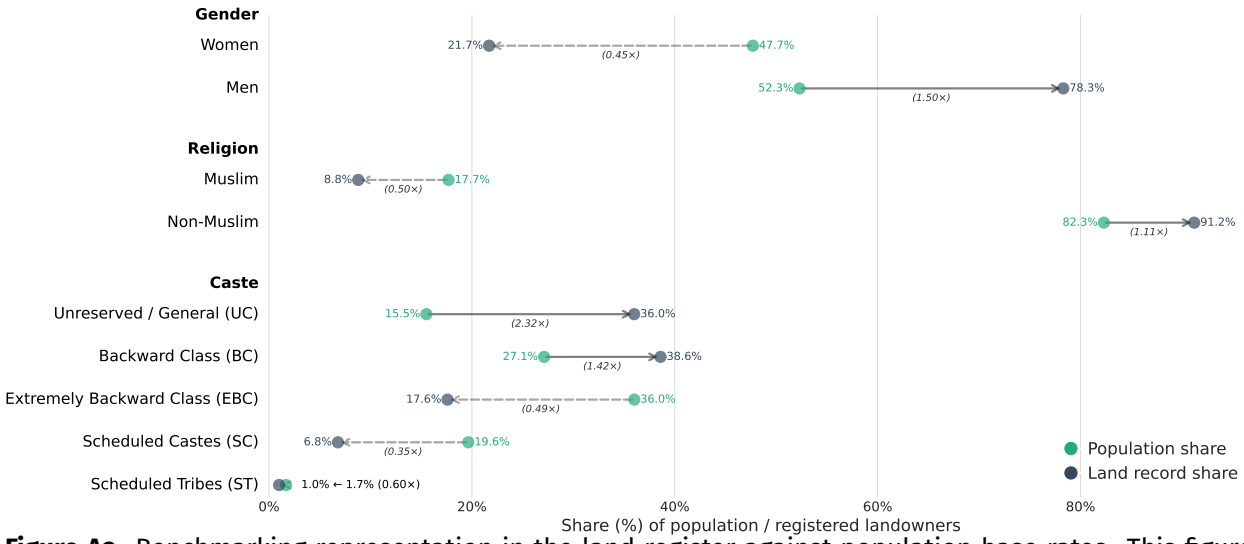


Figure A2. Benchmarking representation in the land register against population base rates. This figure visualizes the numbers reported in Table 2.

B Percentile tables

Table A1. *Number of plots owned per account holder, by percentiles*

Percentiles	Plots
0.00	1
0.10	1
0.20	1
0.30	1
0.40	1
0.50	1
0.60	1
0.70	2
0.80	3
0.90	7
0.95	12
0.96	14
0.97	17
0.98	22
0.99	31
1.00	3,095

Note: See Figure 3 for the distributions up to the 99th percentile (31 plots per account holder).

Table A2. *Area of plot per account holder, by percentiles*

Percentiles	Land area
0.00	0.00
0.10	0.03
0.20	0.07
0.30	0.13
0.40	0.21
0.50	0.32
0.60	0.49
0.70	0.75
0.80	1.21
0.90	2.42
0.95	4.31
0.96	5.09
0.97	6.21
0.98	8.10
0.99	12.36
1.00	8,888,890.90

Note: Units of area in acres. See Figure 3 for the distributions up to the 99th percentile.

B.1 By gender

Table A3. *Number of plots owned per account holder, by percentiles and gender*

	Plots	
Percentiles	**Women**	**Men**
0.00	1	1
0.10	1	1
0.20	1	1
0.30	1	1
0.40	1	1
0.50	1	1
0.60	1	1
0.70	2	2
0.80	3	3
0.90	7	7
0.95	12	13
0.96	14	15
0.97	17	18
0.98	21	24
0.99	31	34
1.00	1,565	3,098

Note: See Figure 4.

Table A4. *Area of plot per account holder, by percentiles and gender*

	plot area	
Percentiles	**Women**	**Men**
0.00	0.00	0.00
0.10	0.03	0.04
0.20	0.07	0.08
0.30	0.12	0.14
0.40	0.20	0.22
0.50	0.31	0.34
0.60	0.46	0.51
0.70	0.72	0.79
0.80	1.17	1.29
0.90	2.36	2.62
0.95	4.25	4.75
0.96	5.04	5.59
0.97	6.19	6.88
0.98	8.14	9.01
0.99	12.70	14.18
1.00	930,000.00	8,888,890.90

Note: Units of area in acres. See Figure 4.

B.2 By religion

Table A5. *Number of plots owned per account holder, by percentiles and religion*

	Plots	
Percentiles	**Muslim**	**Non-Muslim**
0.00	1	1
0.10	1	1
0.20	1	1
0.30	1	1
0.40	1	1
0.50	1	1
0.60	2	2
0.70	2	3
0.80	4	4
0.90	7	9
0.95	11	15
0.96	13	17
0.97	15	21
0.98	19	25
0.99	27	35
1.00	303	579

Note: See Figure 5.

Table A6. *Area of plot per account holder, by percentiles and religion*

	Plot area	
Percentiles	**Muslim**	**Non-Muslim**
0.00	0.00	0.00
0.10	0.04	0.04
0.20	0.08	0.08
0.30	0.13	0.14
0.40	0.20	0.22
0.50	0.29	0.34
0.60	0.42	0.51
0.70	0.64	0.79
0.80	1.02	1.30
0.90	2.00	2.63
0.95	3.52	4.61
0.96	4.16	5.39
0.97	5.08	6.55
0.98	6.30	8.43
0.99	8.96	12.37
1.00	21,203.07	541,016.60

Note: Units of area in acres. See Figure 4.

B.3 By caste

Table A7. *Number of plots owned per account holder, by percentiles and caste*

	Plots				
Percentiles	**UC**	**BC2**	**EBC**	**SC**	**ST**
0.00	1	1	1	1	1
0.10	1	1	1	1	1
0.20	1	1	1	1	1
0.30	1	1	1	1	1
0.40	1	1	1	1	1
0.50	2	1	1	1	1
0.60	2	2	2	1	2
0.70	3	3	2	2	2
0.80	5	5	3	2	4
0.90	11	9	6	4	7
0.95	18	16	10	7	11
0.96	21	18	12	8	13
0.97	24	22	14	10	15
0.98	30	27	18	12	19
0.99	40	36	25	17	26
1.00	569	579	347	205	155

Note: See Figure 6.

Table A8. *Area of plot per account holder, by percentiles and caste*

	plot area				
Percentiles	**UC**	**BC2**	**EBC**	**SC**	**ST**
0.00	0.00	0.00	0.00	0.00	0.00
0.10	0.06	0.04	0.03	0.02	0.03
0.20	0.12	0.09	0.06	0.03	0.05
0.30	0.21	0.15	0.10	0.05	0.10
0.40	0.32	0.24	0.15	0.08	0.18
0.50	0.49	0.36	0.23	0.13	0.29
0.60	0.73	0.55	0.35	0.21	0.46
0.70	1.13	0.83	0.53	0.34	0.73
0.80	1.92	1.34	0.85	0.57	1.17
0.90	3.87	2.62	1.61	1.05	2.22
0.95	6.62	4.44	2.75	1.75	3.82
0.96	7.66	5.15	3.20	2.00	4.36
0.97	9.13	6.20	3.89	2.40	5.27
0.98	11.50	7.96	5.02	3.10	6.80
0.99	16.46	11.70	7.65	4.67	9.34
1.00	80,950.00	322,255.87	52,312.26	541,016.60	758.20

Note: Units of area in acres. See Figure 6.