Land Tenure and Missing Women: Evidence from North India *

Apoorva Lal
Stanford University

This paper examines the effects of colonial era land tenure institutions on modern day demographic outcomes in villages in North India. I exploit the staggered annexation of the kingdom of Awadh in North India by the British East India company in 1803 and 1856 and as a source of exogenous variation in land-tenure (as classified in Banerjee and Iyer 2005) to evaluate the effects of different property-rights systems on modern-day outcomes using a spatial regression discontinuity design on the 2001 village level census. I find that villages where property rights were granted to the cultivators (*mahalwari* villages) have more skewed sex-ratios, lower female literacy, and lower female labour force participation rate than villages where property rights were granted to the landlord (*zamindari* villages). I hypothesise that the likely mechanism is that farmers who have property to pass on to future generations have a stronger preference for male children. These property rights may also grant the men more bargaining power in the household, thereby entrenching intra-household inequalities and manifesting in worse educational and labour market outcomes for women.

*I thank Anand Swamy, Thorsten Rogall, Josh Gottlieb, Thomas Lemieux, Siwan Anderson, Avi Acharya, Saad Gulzar, Saumitra Jha, Vicki Fouka, Steve Haber, David Laitin, Alexander Lee, Dan Thompson, Jayash Paudel, Toby Nowacki, Jared Rubin, and seminar participants at UBC, UCLA, S for valuable comments.
Current version: June 3, 2019; **Corresponding author**: apoorval@stanford.edu

1 Introduction

A stark demographic imbalance exists in the two largest countries in the world: men outnumber women by 70 million in China and India (Denyer and Gowen 2018). The political, economic, and social implications of this are vast and are beginning to surface in both countries in a variety of economic and social channels: a distorted labour market, an excessively high savings rate, property value inflation, and increases in violent crime, trafficking, and prostitution. In India, women's outcomes are systematically worse in a variety of measures, including literacy, labour force participation, and education attainment, and there has recently been a disturbing increase in violence against women. To address the likely causes of these problems, it is imperative that we first understand the likely causes of the underlying demographic disparity. The broad consensus on the causes is articulated by (Denyer and Gowen 2018) as "A combination of cultural preferences, government decree, and modern medical technology". This paper explores the institutional roots of the demographic imbalance in North India, and argues that property rights play a pivotal role in the 'missing women' problem.

To this end, I use exogenous variation in land-tenure institutions produced by the gradual British conquest/annexation of India. In particular, I exploit the variation in land-tenure introduced in villages in the state of Uttar Pradesh by the staggered conquest of the erstwhile Awadh principality, previously a satellite state of the Mughal Empire, between 1805 and 1860. Half the districts in the principality were incorporated into the North-western provinces in 1801 as part of a treaty effectively imposed by the East India Company (EIC) upon the new *nawab* (king) of Awadh, Sa'adat Ali Khan, whom the EIC installed following a period of court intrigue over succession. The partial annexation was justified under auspices of security, with little regard to the economic output of the districts, since the British were yet to survey the region) (Fisher 1993; Fisher 1998). The remainder of the Awadh principality, including the capital Lucknow, was annexed half a century later, and this led to the implementation of a different land-tenure system in these districts. This can be used as natural experiment in close proximity to the border: villages in close proximity to the district borders were as-if-randomly assigned into different land tenure system that granted property rights to the village/farmer (*Mahalwari*) or the local landlord (*Zamindari*) [1]. Furthermore, since land taxes were abolished in India following independence in 1947, observed effects of these land

[1]The systems are defined as *Raiyatwari* := property rights granted to the cultivator, *Mahalwari* := property rights granted to the village council, or *Zamindari*:= property rights granted to the local landlord

tenure systems should be purely institutional overhang and persistence, rather than being confounded by contemporary factors.

This natural experiment lends itself to a spatial regression discontinuity design (assuming unobserved confounders are smooth at the discontinuity) to examine the effects of the *Zamindari* system on village demographics, in particular measures of sex-ratio, female labour force participation, and female literacy. I find that these measures are systematically worse for women in *Mahalwari* villages compared to *Zamindari* villages. The likely mechanism is that farmers in *Mahalwari* villages (through a common-ownership arrangement) had property rights to their land, and therefore preferred male progeny to hand-down their property to, while their contemporaries in *Zamindari* villages did not, since they were de-facto serfs cultivating land they had very low probability of owning. Lower levels of labour force participation and literacy among women in *Mahalwari* villages also suggests that the increased bargaining power for for men granted by land rights leads to lower female labour participation and literacy, which is consistent with canonical household bargaining models wherein changes in factor endowments (land given to the man) may lead to one agent seeking to curtail the other's 'outside option', which, in this case, is remunerated labour.

Our findings add to the burgeoning literature documenting the persistent effects of historical institutions on development outcomes, and are one of the first demostrating the link with demographics. Studies have documented persistent effects of property rights institutions in a wide variety of contexts contexts: Banerjee and Iyer (2005), Dell (2010). Banerjee and Iyer (2005) (henceforth BI) and Dell (2010) use highly localised institutional heterogeneity: land-tenure systems in colonial India for the former and forced mining labour institutions in the latter, to study institutional persistence in determining current economic outcomes. BI find that districts where the landlord was given property rights lagged in a variety of development outcomes, especially those pertaining to agricultural investment following the Green Revolution in the 1960s. Iversen, Palmer-Jones, and K. Sen (2013) question these findings by pointing out that BI's results hinge on coding the central provinces, comprising of much of modern day Madhya Pradesh, parts of Maharashtra, and Orissa (which are omitted from the analysis for this reason), as *Zamindari* districts. Omitting these districts from the analysis results in most of BI's primary results to shrink and lose statistical significance.

A separate body of literature tackles the 'missing women' puzzle, coined by Amartya Sen in a famous article that claimed that more than 100 million women are missing (A. Sen 1990; A. Sen 1992) (in the sense that men outnumber women by margin too large to be consistent with biological explanations). Notable papers in this literature, such as Anderson and Ray (2010) have decomposed the gap by age and disease and argue that the missing women problem is not simply a reflection of a gap at birth (which would suggest that sex-selective abortion is primarily responsible for the gap), but instead that the gap increases further along the age pyramid, which suggests that familial neglect during illnesses are likely to be a major factor in the observed gap as well. This suggests that intra-household inequality and asymmetry in bargaining power is a likely factor driving the gap.

A small but emerging literature, which this paper contributes to, attempts to link institutions and agricultural practices to modern economic and social outcomes for women. Most related to the topic at hand: two very recent papers, Almond, Li, and Zhang (2019) and Bhalotra et al. (2019), study the effects of tenancy reform in China and West Bengal, India respectively. Like this paper, they find that tenancy reform exacerbated son preference by increasing the sex ratio. Alesina, Giuliano, and Nunn (2013) theorise that traditional agricultural practices influenced the gender division of labour and evolution of gender norms, and find empirical evidence suggesting that descendents of societies that practiced plough agriculture have more unequal gender norms. Qian (2008) bridges the economic development and history literature and the 'missing women' literature, to the extent that it seeks to explain variation in the sex ratio in China using quasi-experimental variation in labour-market conditions. This suggests that intra-household bargaining is likely to be a major factor driving the observed demographic imbalance between male and female children. Chakraborty and Kim (2010) document variation in the sex ratio in historical data and draw on from intra-household bargaining models (Lundberg and Pollak 1993) to suggest that 'internal threat points' - determined by control over resources within the household - are a key determinant of household decisions and outcomes. Labour market participation is both an input (since it affects womens' bargaining power because it is an 'outside option' and is remunerated) and an equilibrium outcome in this framework, and a change in the distribution of resources (endowed by property rights) to the male head of the household may therefore plausibly alter it.

In summary, this paper tests the proposition that the intra-household distribution of resources affects the demographic imbalance between male and female children using a historical quasi-experiment, and finds that granting smallholders property rights it led to worse sex ratios, female literacy, and labour force participation. The rest of the paper is organised as follows: section 2 outlines the historical background, section 3 outlines a conceptual framework, section 4 describes the data, section 5 explains the identification strategy, section 6 summarises the results, and section 7 concludes.

2 Historical / Institutional Context

The Gangetic plains of North India are one of the most densely populated parts of the world thanks to the alluvial soil along the Ganges and its tributaries. The two north Indian states under consideration in this paper, Uttar Pradesh and Bihar, have a combined population of approx 227 million [2] , which independently would make them the 5th largest country in the world by population (IWS (n.d.)). The region's vast population and territory, fertile agricultural soil, and relative proximity to the old capital of Delhi has made it a prized possession for those who seek to rule India, and thus it has featured heavily in conflicts over territory throughout Indian history (B. Metcalfe and T. R. Metcalfe 2006). The transitional period between the Mughal Empire and the British colonial period was no exception.

Following a victory over the *Nawab* of Bengal and his French allies in the battle of Plassey in 1757, The EIC began its gradual conquest of India, and obtained revenue collection rights in Bihar and Bengal (Fisher 1998). The revenue system implemented in Bengal and Bihar is known as the Permanent (*Zamindari*) settlement, where the landlords' revenue dues to the government were fixed in perpetuity and the landlords were free to set revenue terms for the peasants, thereby granting the landlords de-facto property rights and petty chiefdom. The principality of Awadh had been an ally of the EIC since 1764 and contributed vast amounts of tribute and soldiers to its cause. However, by the turn of the century, the EIC controlled Bengal in the East and South India, and viewed Awadh as a weakness on the frontier, and began to plan its annexation (Fisher 1993).

Following the death of the incumbent ruler in 1797, the EIC first supported the putative heir to the throne, then deposed him a few months later and installed Sa'adat Ali Khan, the deceased ruler's brother, who had lived in exile under the company's protection for many years. The company initially demanded that he sign a treaty transferring all the territory of Awadh to the EIC in exchange for a generous pension, but following protests, signed a treaty that resulted in the cession of "half the territories"[3], the loose selection of which was made on grounds of defending the EIC's territory against the frontier in the the Northwest from the Afghans and Sikhs [4] (Fisher 1993). The annexation is illustrated in

[2]The exact population, based on the 2001 census that is analysed in this paper, is 227,045,470 distributed across 125,784 villages, which makes for an average village-level population of 1805

[3] Barnett (1980, p. 236) unequivocally states "There exists no adequate explanation in available records of Wellesley's acceptance in 1801 of half of Awadh rather than the entire state". Barnett (1980) goes on to argue that the available evidence suggests that the peculiar decision was made to eliminate the subsidiary alliance, which required the presence of a British garrison in Awadh territory, which the Awadh king was supposed to pay for but frequently shirked on, resulting in the accumulation of arrears. A full annexation was also undesirable because it would have obliged the British to defend the Nawab in perpetuity. Hence the decision to annex half-the-territory (with little to no interest or knowledge of which particular parts)

[4] "The Nabob ceded to the Company the territory of Rohilcund, the Dooab, and Gurruckpoor, the two

the juxtaposed maps of the region in fig 1 in 1795 (L) and 1805 (R). Awadh then remained under a state of 'indirect rule' wherein a representative of the East India company was present, but the company had next to no other institutional or bureaucratic presence (Fisher 1998). In the region of Awadh incorporated into the United Provinces in 1801, the land tenure arrangements were a continuation of the extant community/village based system in the Northwest, called the *Mahalwari*. In the *Mahalwari* system, the 'village bodies that jointly owned the village were responsible for land revenue' (BI, p. 1194). These villages were also known as '*bhaichara*' (brotherhood) communities and were praised by reform-minded British bureaucrats like Metcalfe and Holt Mackenzie. In this region, the land tenure revenue settlements were "avowedly short term, providing for two three-year and one four-year settlements" (Stokes 1983).

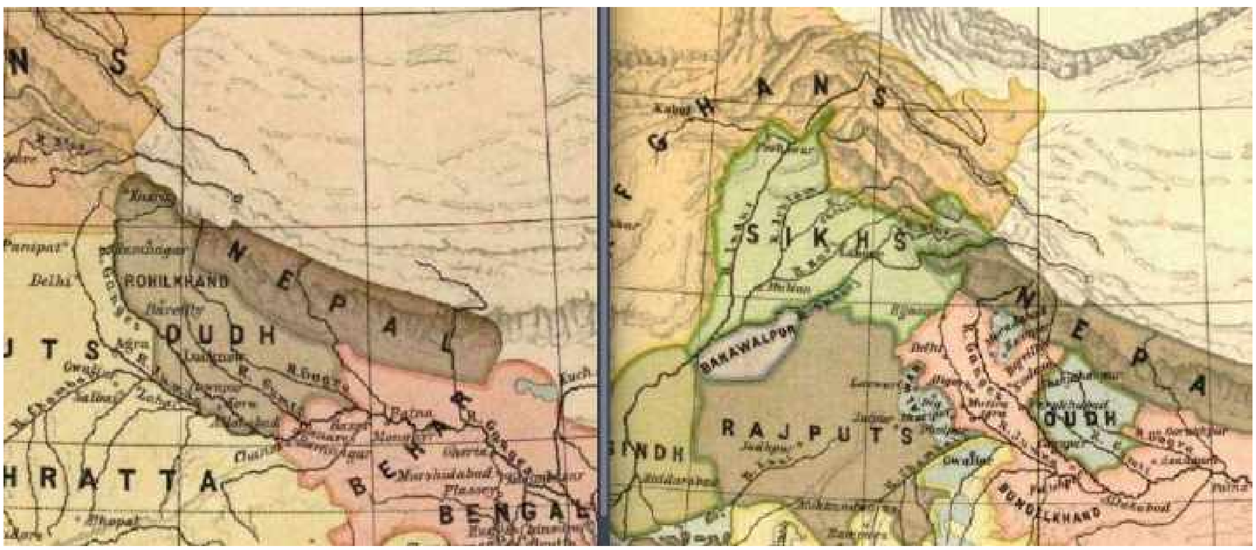


Figure 1: Inset of Region from 'India in 1795'(L) and 'India in 1805' (R) maps showing the partial annexation of Awadh in 1801, Joppen (1914)

The British annexed the remainder of the principality of Awadh in 1856 under the auspices of 'mismanagement' (Fisher 1998). This thought to have been one of the key factors precipitating in the sepoy mutiny in 1857 (also considered the first war of Independence in India). Following the mutiny, the British decided that they needed support from the landed aristocracy in the region and thereby implemented a system of tax revenue that closely resembled the *Zamindari* system (Stokes 1983). Landlords, called *talukdars* in the region, were

former being his frontier provinces ... and the latter bordering upon the company; and, engaged, further, to introduce a better system of management into the territories that which remained in his hands. ... The advantage of acquiring the means of placing upon this weak point additional numbers of the British troops, and thereby increasing its strength, and the general security of the provinces.", Arthur Wellesley (major-general and brother of Richard Wellesley, the Governor-General - the head of the British administration of India), *Memorandum on Marquess Wellesley's Government of India*, reproduced in Fisher (1993) given property rights, but their dues were not fixed in perpetuity like in Bihar and Bengal. The vast majority of cultivators in this region were effectively at-will share-croppers (Stokes 1983) who paid a share of their output to the 'malik' (typically a landlord who was a de-facto petty chief and had titles to land in several villages), and could be expelled from their land for failing to fulfil frequently exorbitantly demands.

BI document that, for all of India, areas conquered at later dates were less likely to have a landlord system [5], but the opposite is true in the regions under consideration in this paper. *Mahalwari* was instituted in the NWP districts that were annexed in 1801 as part of the 'Ceded and Conquered' North-West Provinces, while *Zamindari* was instituted in the remainder of Awadh that was annexed over half a century later.

In summary, the region ended up with a mix of *Zamindari* and *Mahalwari* districts: *Zamindari* was implemented in Bihar and Awadh, while *Mahalwari* was implemented in the rest of Uttar Pradesh (as illustrated in fig 2)[6]. The border between the districts of Awadh incorporated into the ceded territories in 1805 and those annexed with the remainder of Awadh in 1856 was largely arbitrary and based on the complicated politics of the indirect rule system, rather than underlying economic and geographical characteristics on the ground. This allows me to exploit the border and the resultant institutional heterogeneity as a natural experiment. This criterion is less likely to be fulfilled by the border between Uttar Pradesh and Bihar (which were part of different empires in the pre-colonial Period and had very different trajectories post-Independence), but the primary identification strategy is estimated on the entire sample and reported in the appendix as a limited but useful external validity check of the Awadh boundary estimates.

[5] a fact that they use to justify year of conquest as an instrument for land tenure system. However, the exclusion restriction may be violated if the date of conquest affected other institutions that determine contemporary growth, which is plausible

[6] A Labelled map of the Awadh Districts is in the appendix (fig 13)

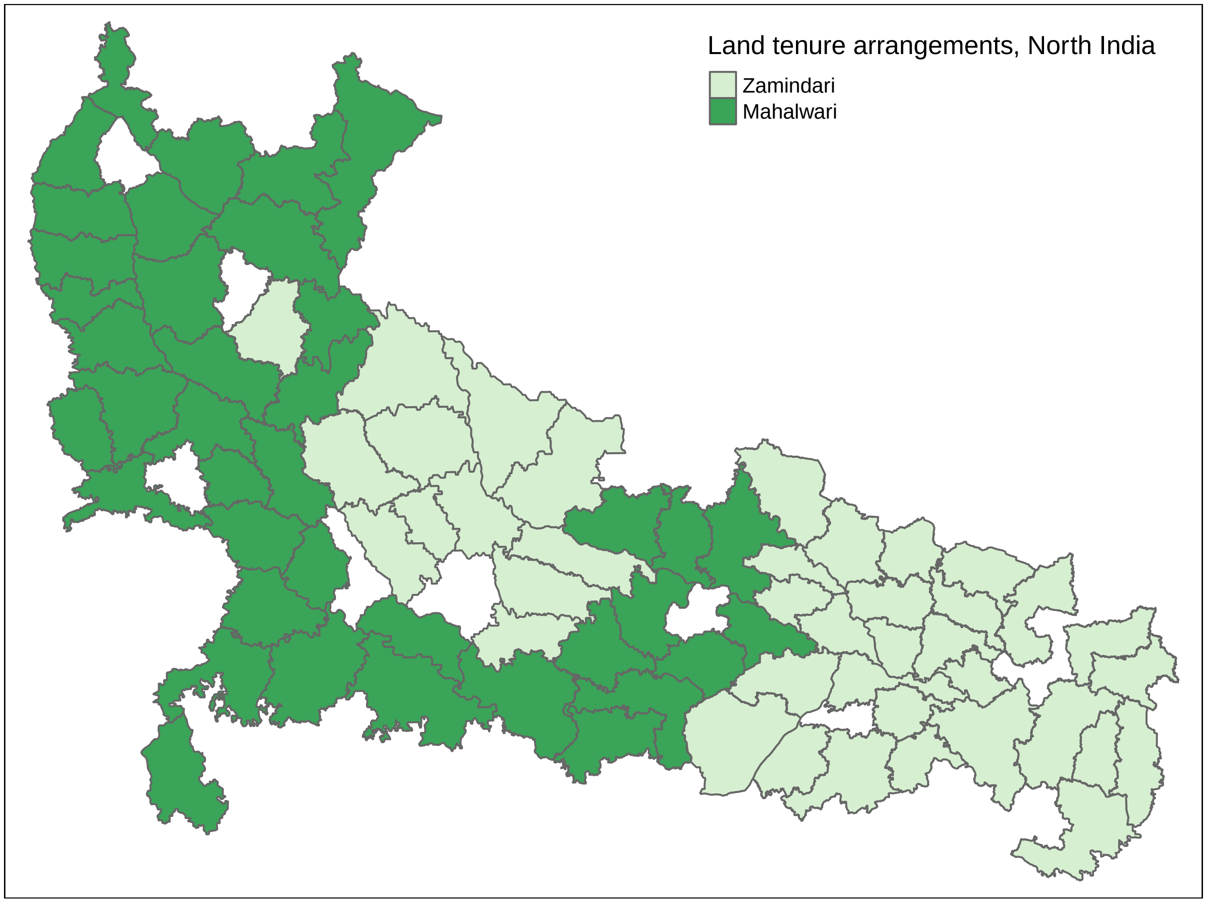


Figure 2: Land Tenure arrangements in North India, based on classifications from Banerjee and Iyer (2005)

3 Conceptual Framework

Development is a multi-faceted process that can be thought of as a function of a variety of processes, including property rights and gender equality, among other things. Applying the theory of the second best (Lipsey and Lancaster 1956) reasoning to this context, one may claim that when the ex-ante equilibrium values are far from the optimum, simply moving one of the inputs towards the social-welfare-maximising value (granting property rights) may in fact result in lower values of another input (gender equity), and potentially even lower overall social welfare [7]. This is entirely plausible in the context of India in the colonial period, where many markets were missing, and it is almost certain that social welfare was not maximised, even with the existing technology and institutions.

More concretely, property rights may affect the sex-ratio through a simple mechanism in that they endow the man with property to pass on to descendants, which likely increases son-preference given the marriage institution in North India. One way for this to occur is if a rational household (almost certainly headed by a man in this context) can increase future welfare by endowing their progeny a factor of production (agricultural land), thereby increasing their future income stream when they grow old. Since women in North India get married into other villages and emigrate from their village of birth (*exogamy*), the likelihood of this future income stream being realised is greatly increased if the household has a son. If the household can 'tweak' (through differential neglect, female infanticide, or sex-selective abortion, though the latter was unavailable until the late 20th century) the 50:50 ratio of male to female progeny, it will choose to do so in order to increase its future utility. This implies that heritable property results in son-preference in a completely rational framework. A simple 2-period model in the appendix (A.1) formalises this argument, and illustrates that $\frac{\partial z}{\partial \theta} > 0$ (where z is the extent of deviation from the 50:50 sex-ratio that would naturally occur, which is controlled by the family, and θ is a parameter for property rights), so one would expect the granting of property rights in a system with female exogamy to yield more unbalanced sex ratios. Thus, a key channel through which property rights are likely to contribute to increasing the sex ratio is through the realisation of future income by male progeny and not female ones. This happens because of the institution of exogamous marriage in the region.

Granting property rights to men may also lead to a decrease in female literacy and labour force participation under canonical household bargaining models. Worse labour market and

[7] Define A SWF: $W = f(x_1, x_2)$, where x_1 is a measure of property rights, and x_2 is a measure of gender equity. W is maximised at $W^* = f(x_1^*, x_2^*)$. Because of a variety of market failures, the ex-ante equilibrium may be $\tilde{S} = f(\tilde{x_1}, \tilde{x_2})$. The Lipsey-Lancaster argument in this context is that moving x_1 in the direction of the optimal (from $\tilde{x_1}$ towards x_1^*) may result in values of x_2 s.t. $x_2 < \tilde{x_2} < x_2^*$, and $S < \tilde{S} < S^*$.

educational outcomes for women being accompanied with higher income and productive capacity is consistent with a household bargaining framework in the spirit of Lundberg and Pollak (1993), as articulated (albiet verbally) by (Chakraborty and Kim 2010). Literacy (which is a form of human capital, albeit a minimal one) increases the woman's 'outside option' from participating in the labour market, and increases her bargaining power in household decisions, and thus it is in the interest of the man to curtail it. Since property rights grant the man with control over the single most important factor of production in an agricultural household, he can use this to limit the outside option for the woman, thereby resulting in lower literacy and labour force participation. Even though the woman working would increase the overall household budget, it would decrease the man's bargaining power, and the man therefore trades off a larger household budget for higher bargaining power. This may be compounded by (or feed into) conservative social mores: women not 'having to work' enhances a family's social standing in society in many developing countries. That female labour force participation as a whole has been shrinking in a time of very high economic growth in India suggests that this effect dominates in India (Economist 2018).

4 Data

The primary datasets analysed in this paper are the 2001 village-level geocoded Indian census and shapefiles [8], obtained from ML Infomap, and district-level land-tenure classifications from Banerjee and Iyer (2005). BI classify 166 districts in colonial India by land-tenure systems, 83 of which are in Bihar and Uttar Pradesh. Spatially merging these districts to the village level shapefiles and subsetting to within 2 decimal degrees of the relevant borders yields an analysis sample of 119,223 villages [9], of which 47,237 are in the state of Uttar Pradesh.

This paper's spatial regression discontinuity identification strategy relies on contiguous districts with different land-tenure systems, which significantly limits the number of regions that can be studied. This data restriction is exacerbated by the fact that Banerjee and Iyer (2005)'s classification of colonial era districts only covers approximately 200 of India's 600 or so districts, so the spatial contiguity requirement limits potential regions to north India. So, I focus on a region of India that is unambiguously classified as one of the two systems. In particular, the paper focuses on Uttar Pradesh and Bihar, the largest and 3rd largest states in India respectively, where the British implemented a mix of land-tenure systems because of

[8] The relevant provinces are Bihar and Uttar Pradesh. The latter is split into two separate datasets. Infomap (n.d.[a]), Infomap (n.d.[c]), Infomap (n.d.[b])

[9] I define a village as a geographical unit classified as 'village' by the census with a population ≥ 100

complexities arising from initial (quasi-) independence of the principality of Awadh, and its subsequent annexation following the sepoy mutiny of 1857. Iversen, Palmer-Jones, and K. Sen (2013) critique BI on their classification of the erstwhile Central Provinces as *Zamindari* areas (modern day provinces of Madhya Pradesh and Chhatisgarh), so I omit the southern *Bundelkhand* region of Uttar Pradesh, which is close to the central Provinces. This yields an analysis sample of the province of Uttar Pradesh, where, the identification strategy is bolstered by the natural experiment produced by the staggered annexation of the principality of Awadh described in section 2.

Since the primary dataset used in the analysis is the population census, some measures that would have been desirable are unavailable. The most notable is the absence of population counts disaggregated by age, which would allow me to document the extent to which the imbalanced sex ratio is stable across the age distribution or not. This would also serve as an indirect test of the Anderson and Ray (2010) finding that a large part of the missing women imbalance arises well after infancy. Unfortunately, the census data is not disaggregated by age and therefore does not permit this potentially interesting analysis.

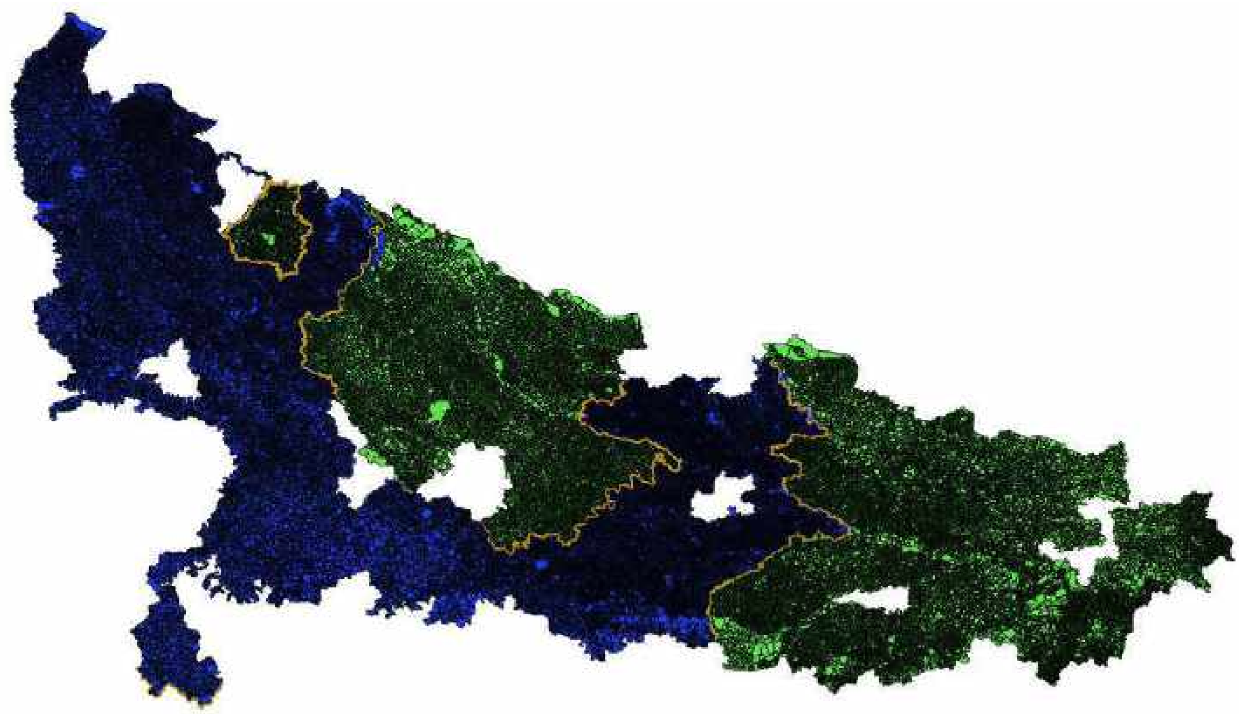


Figure 3: 120,000 villages in North India, with Z-M borders highlighted in yellow

4.1 Summary Statistics

Summary statistics for the relevant variables for the Awadh analysis sample are presented in table 1. Summary statistics for the different buffer sub-samples and the entire sample are included in the section A.3, and balance tables for the corresponding samples are reported in A.4. Boxplots of potential confounders (as a means of balance check) are included in A.8.

Table 1: Descriptive Statistics: All Villages, Awadh

Statistic	N	Mean	St. Dev.	Min	Max
Non-Landlord Status	47,237	0.497	0.500	0	1
Household Size	47,237	6.400	1.070	1.570	140.000
Population	47,237	1,335.000	1,293.000	100	25,433
Share Scheduled-Caste	47,237	0.248	0.189	0.000	1.000
Literacy	47,237	0.409	0.122	0.000	1.000
Female Literacy	47,237	0.285	0.124	0.000	0.962
Sex Ratio	47,237	1.110	0.137	0.540	11.200
Labour Force Participation Rate	47,237	0.345	0.101	0.000	1.000
Female Labour Force Participation Rate	47,237	0.184	0.175	0.000	0.904
Night-time Luminosity (DMSP 2001)	47,237	1.900	3.180	0.000	56.700

The average sex-ratio in the region(1.11) is well above what one would naturally expect (1 ± 0.02), and as illustrated in fig 4, the average is slightly higher in *Mahalwari* villages compared to *Zamindari* ones. This is especially noticeable in the bottom panel in fig 4, which plots the share of villages with each type of land-tenure institution conditional on having a particular value of the sex-ratio. Mahalwari villages are far likelier to have higher sex ratios (especially on the upper end of the spectrum, beyond the overall mean of 1.1).

However, in order to make a causal claim about the effects of land-tenure institutions on demographics, we need it to be the case that the 'treatment' (land-tenure) is uncorrelated with other factors that may drive modern day demographics [10]. Now, there is no reason to believe that the land tenure system in different districts of North India is exogenous; district level policies may drive differences in the sex-ratio that have nothing to do with the property rights institutions under consideration, and this may bias results. Thus, a naive comparison of various development outcomes between *Zamindari* and *Raiyatwari* districts is likely to be biased. This necessitates the analysis of a more localised variation in institutions where the exogeneity condition is more plausible.

[10] Econometrically, this is simply the exogeneity condition($\mathbb{E}(X\epsilon) = 0$) for OLS. Put differently, since we know there is omitted variables bias (OVB), the only way that the 'short regression' coefficient is unbiased is if either (1) the omitted variable is uncorrelated with Y or (2) the omitted variable is uncorrelated with X. Neither assumption is likely to be true *for the entire sample of villages*.

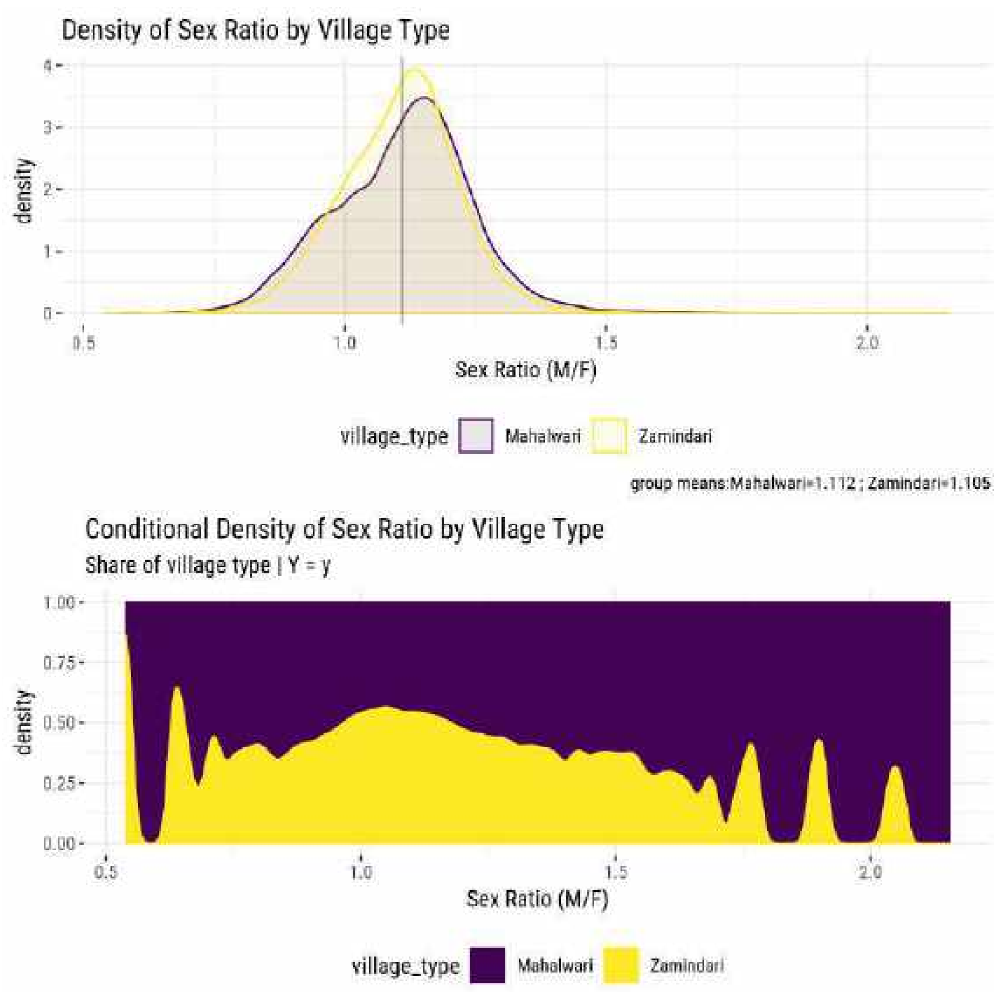


Figure 4: Density of Sex Ratio by Village Type : Unconditional (conventional density) in top panel, Conditional (i.e. share of mahalwari and zamindari villages for each value of the sex ratio on the X axis) in bottom panel

4.2 Border Identification

Fortunately, given the availability of high-resolution spatial data for the village-level census, one can alleviate this problem by zooming in and comparing villages that were on different sides of district (and land-tenure system) borders (highlighted in yellow in Figure 3), where the land-tenure system can plausibly be exogenous (since geography, climate etc. are likely to be the same in villages within 50 kilometres of each other on either side of the border). This is especially plausible given that the border between the districts that comprised of the boundary between Awadh ('Oudh' in fig 1) and the rest of North India was largely exogenous and was part of the 'half-the-territory' partitioning in the treaty in 1801 (Fisher 1998).

Given the high population density (and correspondingly high disaggregation of administrative units) in the region, one can zoom in very close to the border and still have enough sample size to estimate the effects of non-landlord systems. I omit the boundary between the North-Western Provinces and the Central Provinces (modern day Uttar Pradesh and Madhya Pradesh - on the south- western end of UP in the map below), in accordance with Iversen, Palmer-Jones, and K. Sen (2013)'s claim that CP's land tenure systems were highly heterogeneous and BI's findings were sensitive to how these were coded.

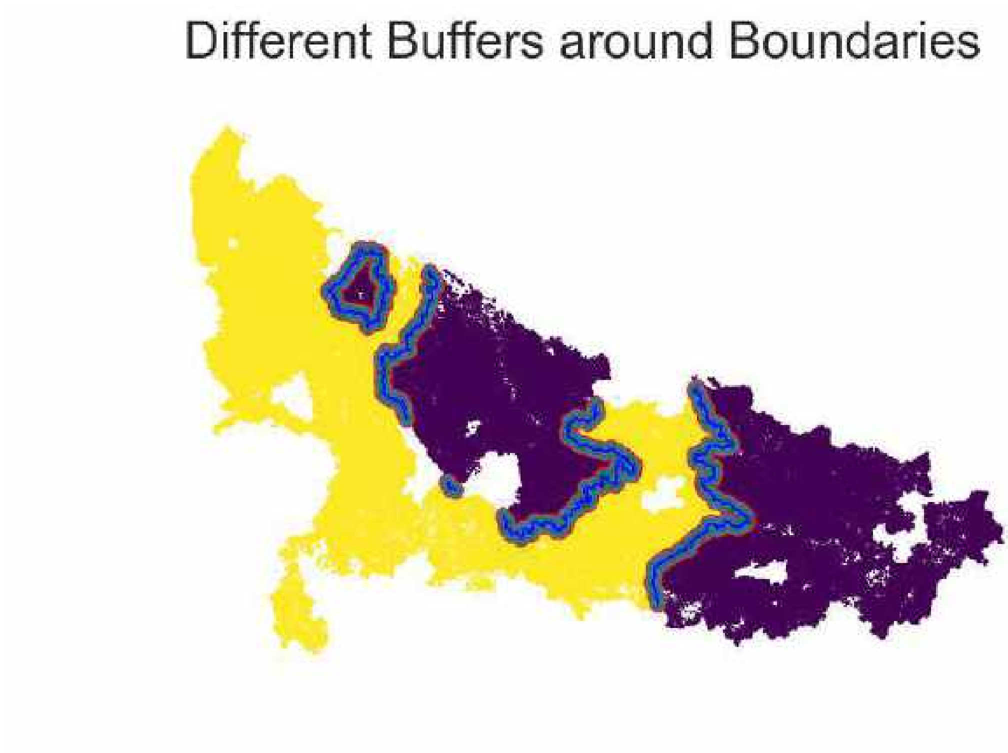


Figure 5: Buffer villages: 10(red), 5(green), and 1 km (blue)

I subset the analysis sample to within 0.1 decimal degrees (≑ 10 km), 0.05 decimal degrees (≃ 5 km), and 0.01 decimal degrees(≃1 km) of the border between *Zamindari* and non-*Zamindari* districts [11], and estimate regressions on these buffer samples separately, both for the basic OLS analysis outlined in 5.1, and the 2D Regression discontinuity 5.3. The bandwidth (distance from the border) is chosen by a data driven selection procedure for the non- parametric regression discontinuity design with Euclidian distance as the running variable 5.2.

5 Empirical Strategy

The three estimation strategies of used in the paper are described in the following subsections, and results for each strategy are presented in the corresponding sub-section in section 6.

5.1 OLS on Boundary Samples

I estimate regressions of the form:

$$Y_{ijb} = \alpha + \beta \text{Non-Landlord}_j + \gamma \text{Non-Landlord}_j \times s_{ij}^p + \phi_b + \epsilon_i \tag{1}$$

$$\text{s.t. } |s_{ij}| \leq k \quad \forall k = \{0.5, 0.4995, \ldots, 0.0001\} \tag{2}$$

where Y_{ij} is an outcome (sex ratio [12], female labour force participation [13], female literacy [14]) for village i in district j, Non- Landlord$_j$ is a dummy variable for whether district j is classified as a non- landlord district, and ϕ_b is a vector of line-segment fixed effects (which are constructed by dividing the border between the Zamindari and Mahalwari districts into 10 segments and generating dummies, which ensures that the estimation compares villages on either side of the same part of the boundary), and s_{ij} is distance to the border with a p^{th} order of the spline; 0 means no spline.

[11] These degree to km conversions are approximate and are correct for 23 degrees North of the equator, which is the region of the globe for North India. The buffers are constructed using the shapefiles using the original WGS-84 EPSG-4326 projection. Mercator projection, which would permit constructing buffers in metres, were unstable because the region spans multiple Mercator regions

[12] defined as $\frac{\text{Male Population}_i}{\text{Female Population}_i}$ for village j with closest boundary b- see appendix for detailed variable list

[13] defined as $\frac{\text{Female Workers}_i}{\text{Female Population}_i}$ for village j

[14] defined as $\frac{\text{Female Literates}_i}{\text{Female Population}_i}$ for village j

5.2 Euclidian Distance Non-Parametric Regression Discontinuity

The sharp RD estimate is evaluated as the difference between the CEFs of the 'treated' and 'control' groups evaluated at the cutoff (Imbens and Lemieux 2008). In the current applied econometrics literature, the treatment effect is is estimated using flexible CEFs estimated using local polynomial regressions on either side of the cutoff, consistent with Hahn, Todd, and Van der Klaauw (2001), which shows that if the average potential outcomes are continuous functions of the score at c, the sharp-RD treatment effects is the difference between the CEFs at c.

$$\hat{\tau}_{SRD} = lim_{x \downarrow c} \mathbb{E}[Y_i | X_i = x] - lim_{x \uparrow c} \mathbb{E}[Y_i | X_i = x]$$

To this end, I use the robust non-parametric regression-discontinuity design estimator and confidence intervals proposed in Calonico, Cattaneo, and Titiunik (2014) (henceforth CCT 2014) [15]. I first construct Euclidian distance Distance_{ij} to the nearest point on the Zamindari-Raiyatwari district boundary, where negative values denote landlord villages and positive values denote non-landlord villages, and use this distance as a running variable with a cutoff at $Distance = 0$. The estimating equation is the following regression

$$Y_{ijb} = \alpha + \gamma \mathbb{1}_{[D \leq 0]} f(\text{Distance}_{ij}) + \psi \mathbb{1}_{[D>0]} g(\text{Distance}_{ij}) + \phi_b + \epsilon_i$$

where $\mathbb{1}_{D>0}$ is the indicator function for non-Zamindari villages, f and g are smooth functions of distance (different degree polynomials are estimated and reported) on a bandwidth h estimated using local polynomial regressions selected by the Mean Squared Error (MSE) expansion of the sharp RD estimator as proposed in CCT 2014 [16], ϕ_b are line-segment fixed effects. A triangular kernel is used to weight the observations closest to the cutoff heavily and decreasing in either direction on the interval $[c - h, c + h]$.

[15] estimated using the accompanying R package RDRobust (**calonicoRdRobustPackageRobust2015**)

[16] the MSE- optimal bandwidth choice is

$$h_{MSE} = \left(\frac{V}{2(p+1)B^2} \right)^{1/(2p+3)} n^{-1/(2p+3)}$$

where B and V represent the bias and variance of the RD point estimator $\hat{\tau}$, and p is the degree of the polynomial (Cattaneo, Idrobo, and Titiunik 2018)

5.3 Semi-Parametric Spatial Regression Discontinuity

I estimate semi-parametric regressions with smooth polynomial functions of 2 running variables (longitude and latitude) per (Dell 2010). Keele and Titiunik (2015) contend that Dell (2010)'s approach is limited by the fact that the longitude and latitude are the same for all households contained in the same cluster (district), even though the relevant unit of observation for all estimating equations is at the household level. Since I have geographical coordinates for the centroids for each village (rather than simply the district that contains the village, which would be the analogue to the Dell, 2010 case), the semi-parametric regression is a sub-optimal estimation technique for the available data and is reported as a check for the non-parametric estimates, which are more flexibly estimated and make full use of the data.

$$Y_{ijb} = \alpha + \beta \text{Non-Landlord}_j + f(\text{Location}_{ij}) + \phi_b + \epsilon_{ijb}$$

6 Results

6.1 OLS on boundary samples

The regressions are estimated using the framework outlined in 5.1 close to the boundaries of the Awadh principality (highlighted in red in fig 6), which compares villages within the state of Uttar Pradesh, and heteroskedasticity-robust (HC2) standard errors are reported. In the 10 and 5 km buffers, I find a statistically and economically significant increase in sex-ratio (i.e. more men than women) of approximately 0.015 (1.38 % on an already high base level of 1.09 in the control group) in *Mahalwari* villages. Similarly, female literacy is 1.2 percentage points (or 4.14% on a base of 0.29) lower, and female labour force participation is 1 percentage point (or 5.6%) lower. These magnitudes suggest large negative effects of *Mahalwari* land-tenure arrangements on the sex ratio and women's outcomes (conditional on the 'treatment' being as-if-random in close proximity of the border).

While one might be concerned that estimating the same regressions on the entire sample will be driven by villages on the boundary between Bihar and Uttar Pradesh, which are different states and have had different political parties in power (and consequently different political and economic institutions), these results serve as an approximate check of the external validity of the main estimates and are presented in the appendix A.10. The direction of the effect on sex-ratio, female literacy, and female labour force participation is identical, and the magnitudes are slightly smaller but statistically significant for the larger sample.

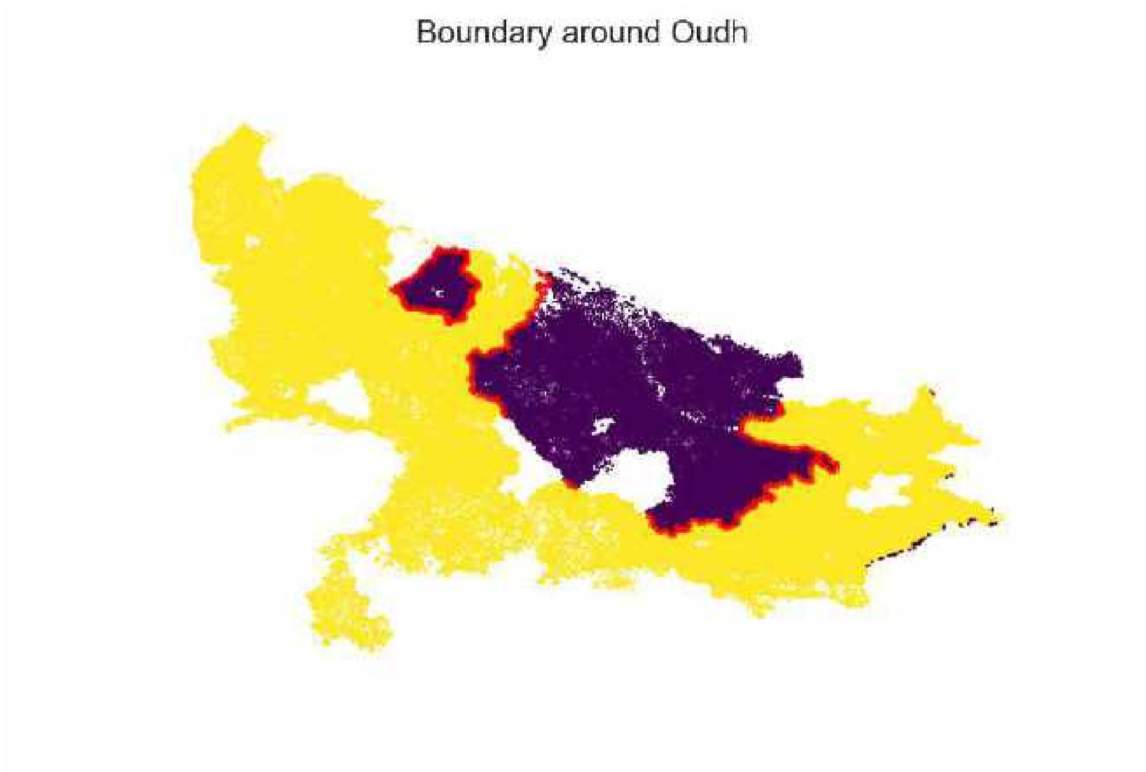


Figure 6: Identifying Boundares around Awadh

Table 2: Sex Ratio (M/F) - Awadh

	Dependent variable:		
	Sex Ratio		
	All	10 km	5 km
	(1)	(2)	(3)
Non-landlord	−0.004***	0.015***	0.014***
	(0.001)	(0.002)	(0.002)
Line-segment fixed effects	X	X	X
Control Mean	1.11	1.09	1.09
Observations	47,237	17,931	9,052
R^2	0.243	0.243	0.395

Table 3: Literacy - Awadh

	Dependent variable:		
	Female Literacy		
	All	10 km	5 km
	(1)	(2)	(3)
Non-landlord	0.050***	−0.012***	−0.010***
	(0.001)	(0.002)	(0.002)
Line-segment fixed effects	X	X	X
Control Mean	0.262	0.29	0.289
Observations	47,237	17,931	9,052
R^2	0.135	0.159	0.150

Table 4: Female Labour Force Participation - Awadh

	Dependent variable:		
	Female Labour Force Participation		
	All	10 km	5 km
	(1)	(2)	(3)
Non-landlord	0.013***	−0.010***	−0.010***
	(0.002)	(0.002)	(0.003)
Line-segment fixed effects	X	X	X
Control Mean	0.186	0.179	0.174
Observations	47,237	17,931	9,052
R^2	0.156	0.220	0.240

6.1.1 OLS Coefficient Stability for Bandwidth choices

To rule out the possibility that arbitrary choices of buffer are responsible for the observed effect, I estimate 100 separate regressions for each outcome to test the stability of the coefficient for different bandwidths j (drawn from a decreasing sequence from 0.5 to 0, with a step of 0.005). The results for this coefficient stability exercise are plotted in 7). Each point in these graphs is the estimated $\hat{\beta}_j$ for non-landlord villages estimated on the sample of villages s.t. distance to border $s_i \leq j \ \forall i, j = \{0.5, 0.4995, \ldots, 0.000001\}$, with its corresponding (robust) standard error in blue. Moving to the right on the X axis illustrates the estimated $\hat{\beta}$ on progressively narrower buffers; a relatively stable pattern indicates that high-leverage outlier villages are not driving the estimated effect (since they are almost certainly dropped as the cutoff gets narrower). The estimate switching signs with narrower cutoffs (such as for female literacy) suggests that a simple comparison of means was inadequate for that outcome, especially since the effect remains negative and stable for all subsequent thresholds. The corresponding plot for the entire sample is reported in the appendix (fig 15).

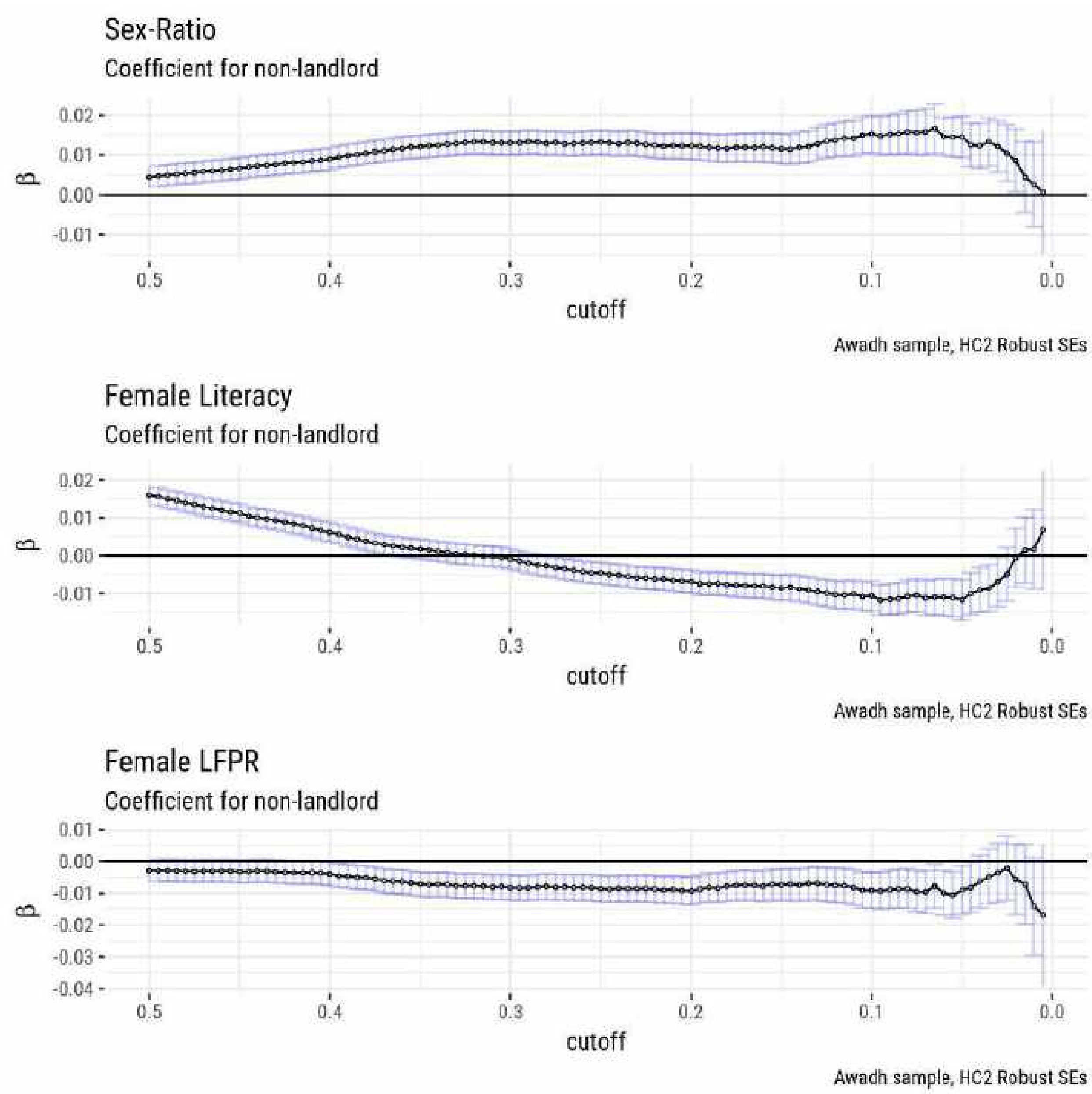


Figure 7: Non-landlord Coefficients with varying cutoff (in decimal degrees): The sign-reversal for female literacy may be driven by potential imbalance when the bandwidth is wide. The estimates are closer to the preferred specification and sample on the right of the figure (i.e. bandwidth of 10 km and 5 km). Corresponding figures with the linear and quadratic splines are in the appendix sec A.5

6.2 Distance Regression Discontinuity

Tables 5, 6, and 7 report sharp regression discontinuity estimates, standard errors, and the number of observations used in the estimation on either side of the cutoff. The SRD coefficient (row 1 of the tables below) is calculated as the difference between the local polynomial regression functions (linear in columns 1 and 3, and quadratic in 2 and 4) using the estimation framework proposed by (Calonico, Cattaneo, and Titiunik 2014) and described in section 5.2, with line-segment fixed effects as controls (in column 3 and 4). The CCT bandwidths are different on the two sides of the threshold and are reported in rows 7 and 9 (with the corresponding number of observations on each side of the threshold in rows 8 and 10).

The sign and magnitudes of the effects are comparable to those found in the OLS estimation and are more precisely estimated (as suggested by the robust SE in row 2 and z-score in row 3); however, the effect on female labour force participation is statistically indistinguishable from zero. Results for quadratic CEFs are presented in fig 8 [17]. Corresponding results for the entire sample are in the appendix (section A.10.2).

Table 5: Sex-Ratio RD estimates - Awadh

	Parameter	(1)	(2)	(3)	(4)
1	Coeff	0.0302	0.0273	0.0156	0.0167
2	SE	0.0045	0.0046	0.0036	0.0038
3	z	6.724	5.8823	4.2735	4.373
4	P Val	0	0	0	0
5	CI Lower	0.0214	0.0182	0.0084	0.0092
6	CI Upper	0.039	0.0363	0.0227	0.0241
7	Left Bandwidth	0.1859	0.3525	0.1766	0.297
8	N (left of c)	8080	12258	7795	11013
9	Right Bandwidth	0.1859	0.3525	0.1766	0.297
10	N (right of c)	11873	16546	11426	15429
11	Polynomial Degree	1	2	1	2
12	Line-segment Fixed Effects			X	X

[17] higher order polynomial figures are included in the appendix (section A.11)

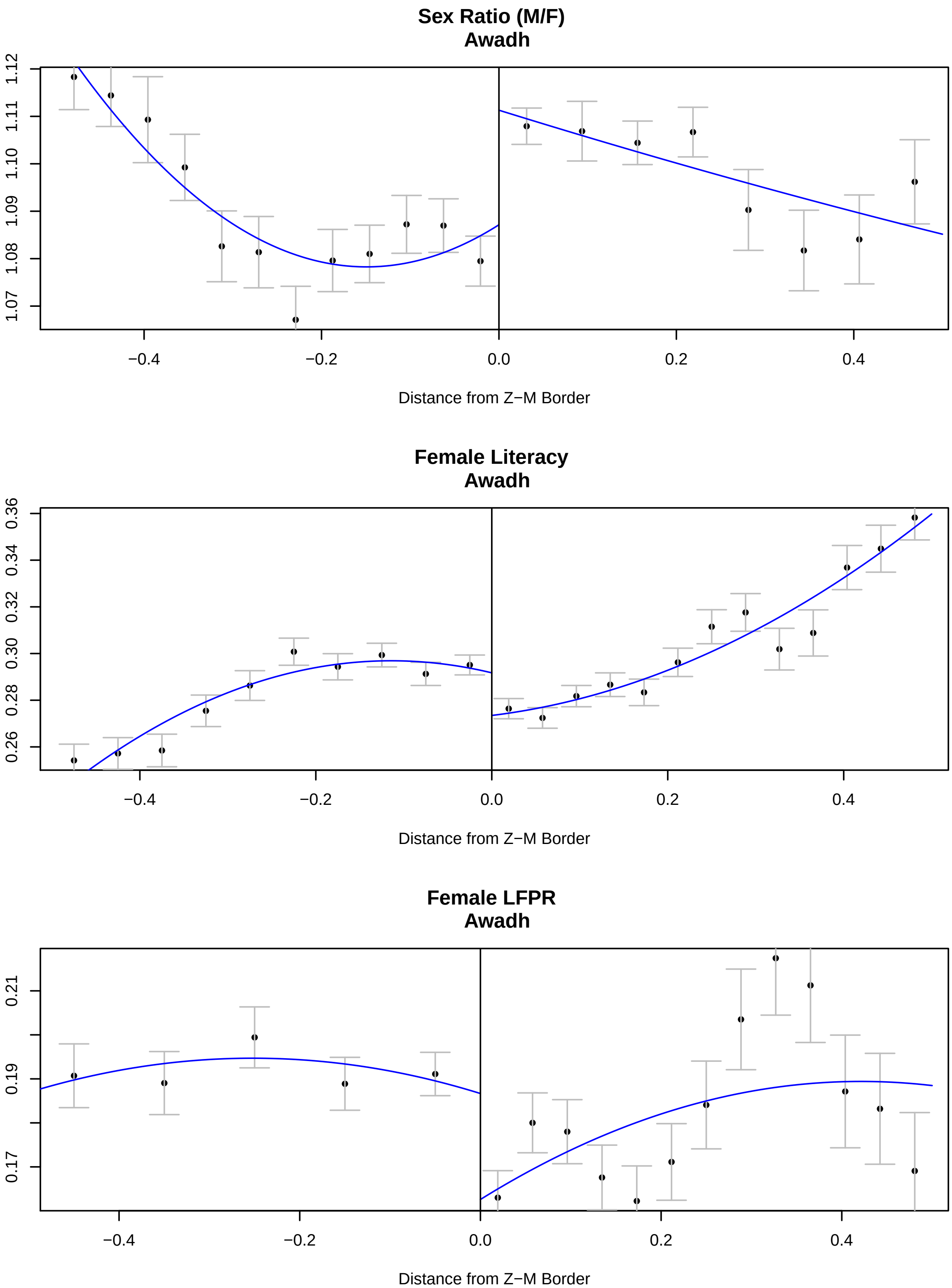


Figure 8: RD estimates with Euclidian distance as running variable. The MSE optimal bandwidths are [-0.35, 0.30] for Sex Ratio, [-0.31, 0.22] for Female Literacy, and [-0.33, 0.32] for LFPR

Table 6: Female Literacy RD estimates - Awadh

	Parameter	(1)	(2)	(3)	(4)
1	Coeff	-0.0159	-0.0146	-0.0055	-0.0043
2	SE	0.0041	0.0042	0.0039	0.0043
3	z	-3.8838	-3.4897	-1.4155	-0.9935
4	P Val	0.0001	0.0005	0.1569	0.3204
5	CI Lower	-0.0239	-0.0229	-0.0132	-0.0128
6	CI Upper	-0.0079	-0.0064	0.0021	0.0042
7	Left Bandwidth	0.1447	0.3109	0.1299	0.2189
8	N (left of c)	6761	11320	6228	9035
9	Right Bandwidth	0.1447	0.3109	0.1299	0.2189
10	N (right of c)	9887	15727	9061	13212
11	Polynomial Degree	1	2	1	2
12	Line-segment Fixed Effects			X	X

Table 7: Female LFPR RD estimates - Awadh

	Parameter	(1)	(2)	(3)	(4)
1	Coeff	-0.0259	-0.0219	-0.0076	-0.0081
2	SE	0.006	0.0061	0.0049	0.0054
3	z	-4.3318	-3.5946	-1.5512	-1.4956
4	P Val	0	0.0003	0.1208	0.1347
5	CI Lower	-0.0377	-0.0339	-0.0171	-0.0187
6	CI Upper	-0.0142	-0.01	0.002	0.0025
7	Left Bandwidth	0.1617	0.3295	0.2126	0.3249
8	N (left of c)	7330	11746	8861	11652
9	Right Bandwidth	0.1617	0.3295	0.2126	0.3249
10	N (right of c)	10741	16128	13011	16041
11	Polynomial Degree	1	2	1	2
12	Line-segment Fixed Effects			X	X

6.3 Semiparametric Spatial Regression Discontinuity

The functions are estimated using in the framework outlined in section 5.3. Once again, the sign and magnitudes of the effects are comparable to those found in the OLS and local polynomial RD estimation; however, the effect on female labour force participation is statistically indistinguishable from zero for some choices of bandwidth. Corresponding results for the entire sample are in the appendix (section A.10.3).

Table 8: Awadh - Female Literacy - Semiparametric Spatial RD estimates

	Dependent variable:			
	Female Literacy			
	10 km	5 km	10 km	5 km
	(1)	(2)	(3)	(4)
Non-landlord	−0.015***	−0.012***	−0.009***	−0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
Polynomial Degree	2	2	3	3
LSFE	X	X	X	X
Observations	17,931	9,052	17,931	9,052
R^2	0.204	0.204	0.224	0.220
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table 9: Awadh - Female LFPR - Semiparametric Spatial RD estimates

	Dependent variable:			
	Female LFPR			
	10 km	5 km	10 km	5 km
	(1)	(2)	(3)	(4)
Non-landlord	−0.008***	−0.009***	−0.003	−0.007**
	(0.002)	(0.003)	(0.002)	(0.003)
Polynomial Degree	2	2	3	3
LSFE	X	X	X	X
Observations	17,931	9,052	17,931	9,052
R^2	0.249	0.260	0.254	0.265
Adjusted R^2	0.249	0.260	0.253	0.264
Note:		*p<0.1; **p<0.05; ***p<0.01		

6.3.1 Semiparametric RD Coefficient Stability for Bandwidth choices

Using the methodology outlined in section 6.1.1, I estimate 100 different regressions with the semiparametric estimating equation in section 5.3 to rule out the possibility that particular choices of bandwidth yield the observed effect. The coefficient remains large, stable, and statistically significant as one zooms in (illustrated by moving from left to right on the graph). The observed difference in female labour force participation rate is small and statistically insignificant from this estimation strategy.

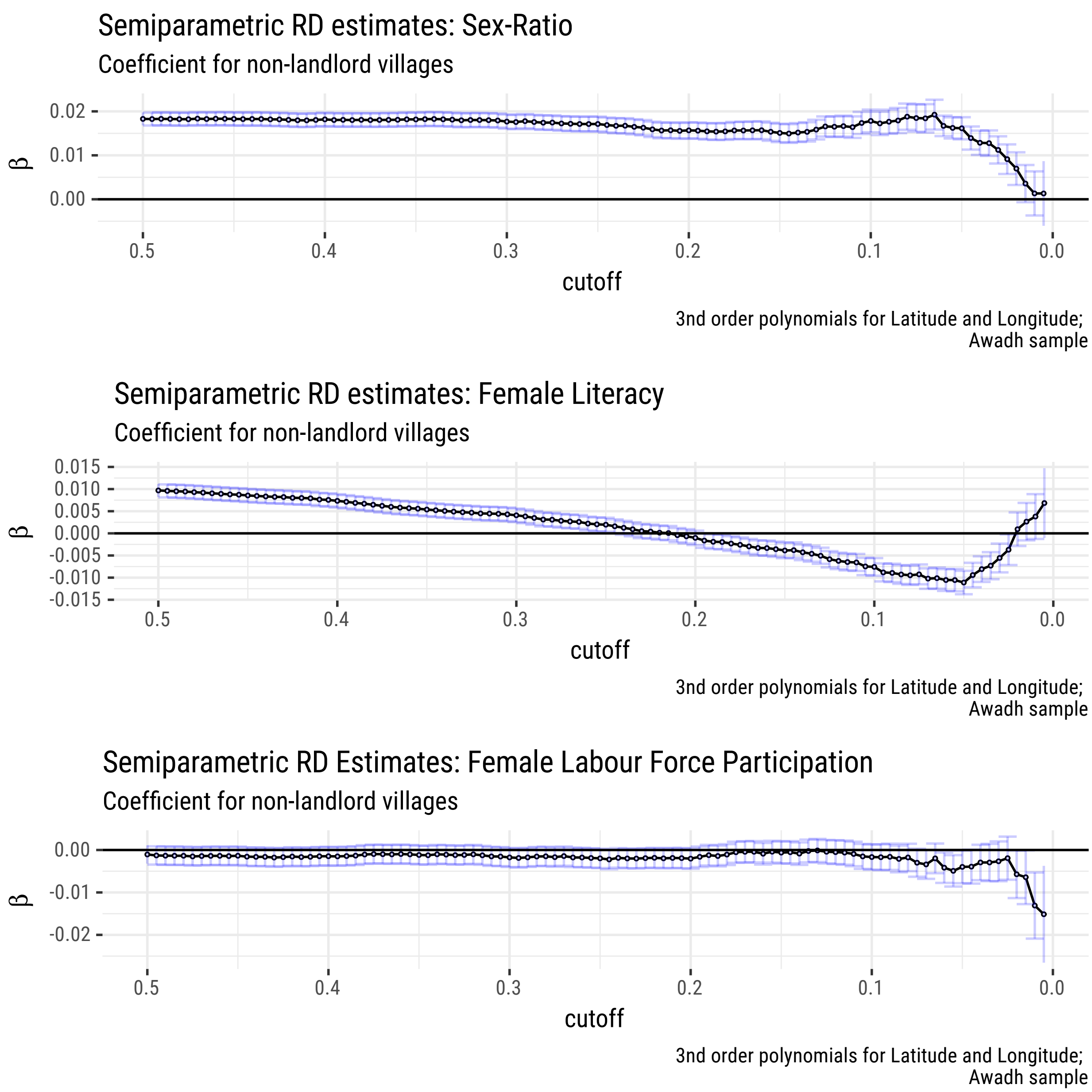


Figure 9: Semi-parametric RD Non-landlord Coefficients: Awadh

6.4 Robustness Checks

6.4.1 Controls in OLS

The primary estimates in this paper do not include controls (beyond line-segment fixed effects, which are meant to ensure that geographically proximate villages are compared) because most plausible development outcomes that may be in the error term are themselves likely to be outcomes of the notional 'experiment' (that of allocating different land tenure arrangements) because of the long gap between the measurement of Y (2001 census) and X (which were determined by 1860), and thus will bias the estimated coefficient for land-tenure. Available variables at the village level are also limited: population size, caste composition (Chakraborty and Kim (2010) document more balanced sex ratios among 'lower' caste groups), luminosity (as a proxy of income, aggregated to the village level from DMSP rasters, NOAA (n.d.)).

Despite this, the magnitude of this difference is likely useful information. The estimated coefficients for sex ratio are significantly larger upon the inclusion of controls, which suggests that the covariance between non-landlord and the controls is nonzero. Corresponding regressions for other outcomes for Awadh, as well as estimates for the entire sample are included in the appendix (sec A.9)Awadh. The two types of villages show similar trends for luminosity for the 21 year DMSP archive (fig A.7), which suggests that public good provision is comparable across the two types of villages when zooming close to the border.

6.4.2 Testing for sorting across the border

An obvious check for the validity of the regression discontinuity design is to look at whether there is sorting across the discontinuity. This is largely implausible in this particular context, since simply moving from a *Zamindari* village to a *Mahalwari* one does not entitle a farmer to land; he (or his forbears) must have been present there during the distribution of titles in the early Nineteenth century. Nevertheless, the unit of observation in this context makes this difficult to detect (since households may be sorting but entire villages may still be present). This does not appear to be the case based on village counts; indeed, the opposite seems to be true (i.e. the density of villages is very high close to the discontinuity).

Table 10: Sex Ratio (M/F) - Awadh

	Dependent variable:			
	Sex Ratio			
	All	10 km	5 km	1 km
	(1)	(2)	(3)	(4)
Non-landlord	−0.005***	0.032***	0.033***	0.021***
	(0.001)	(0.003)	(0.003)	(0.006)
Share Scheduled-Caste	0.005	−0.085***	−0.088***	−0.107***
	(0.003)	(0.008)	(0.008)	(0.018)
log Population	−0.005***	−0.001	0.0001	0.0003
	(0.001)	(0.002)	(0.002)	(0.004)
Longitude	−0.044***			
	(0.001)			
Latitude	0.028***			
	(0.001)			
Village Mean Luminosity (2001)	−0.0004	0.001	−0.001***	−0.003***
	(0.0004)	(0.001)	(0.0004)	(0.001)
Constant	3.940***	1.120***	1.110***	1.130***
	(0.055)	(0.010)	(0.014)	(0.029)
Observations	47,237	17,931	9,052	2,068
R^2	0.262	0.018	0.031	0.031
Adjusted R^2	0.262	0.018	0.031	0.029
Note:				*p<0.1; **p<0.05; ***p<0.01

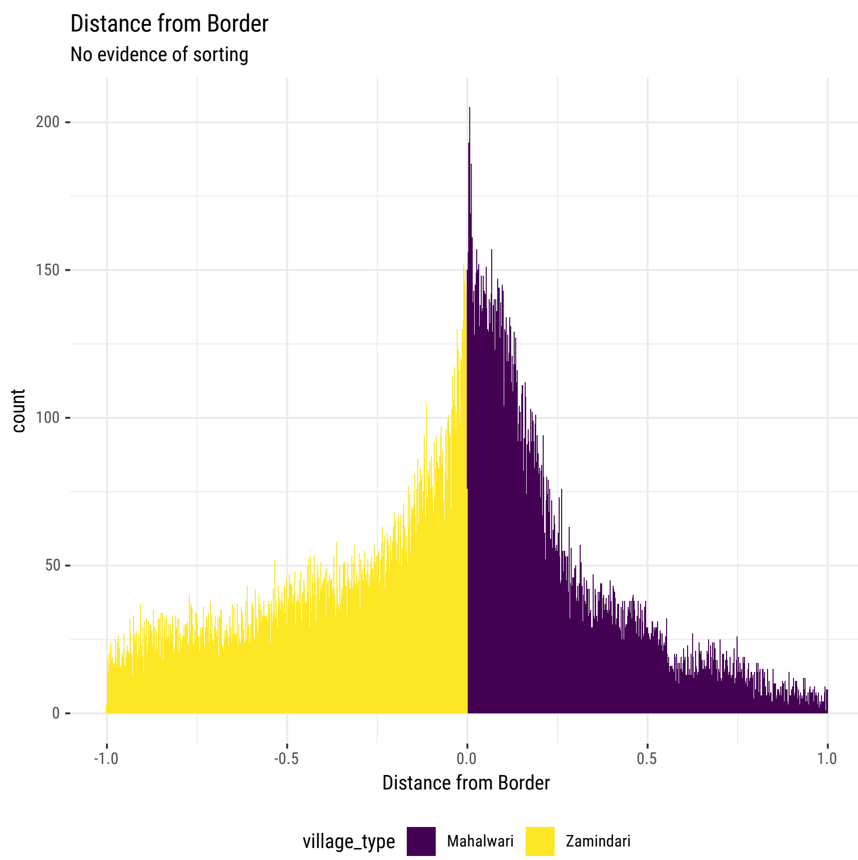


Figure 10: Histogram of Distance to Border for all villages

6.4.3 Continuity in other potential factors

6.4.4 Labour market conditions

It is also possible to assess the validity of the RD by checking if other plausible (observable) factors that may determine the sex ratio and labour market conditions for women vary discontinuously across the border. Neither Male labour force participation (which is a useful placebo because in theory, potential labour market imperfections should affect both genders equally) nor share of village engaged in agriculture (figs 11 12) appear to vary discontinuously across the border, which suggests that labour market opportunities are not responsible for the observed effects on the sex ratio, female literacy, and female labour force participation.

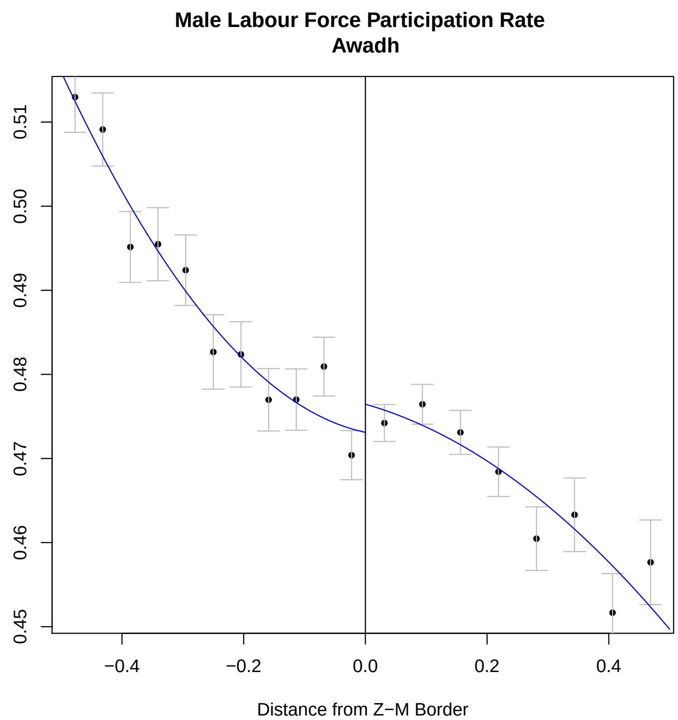


Figure 11: Male LFPR

7 Conclusion

This paper finds that villages where *Mahalwari* was implemented have systematically worse outcomes for women on a variety of measures: sheer numbers (the sex ratio), literacy, and labour force participation. The observed effects are consistent with a mechanism whereby the presence of heritable land in the household exacerbates culturally widespread son preference, and increased bargaining power for men results in worse female labour force and

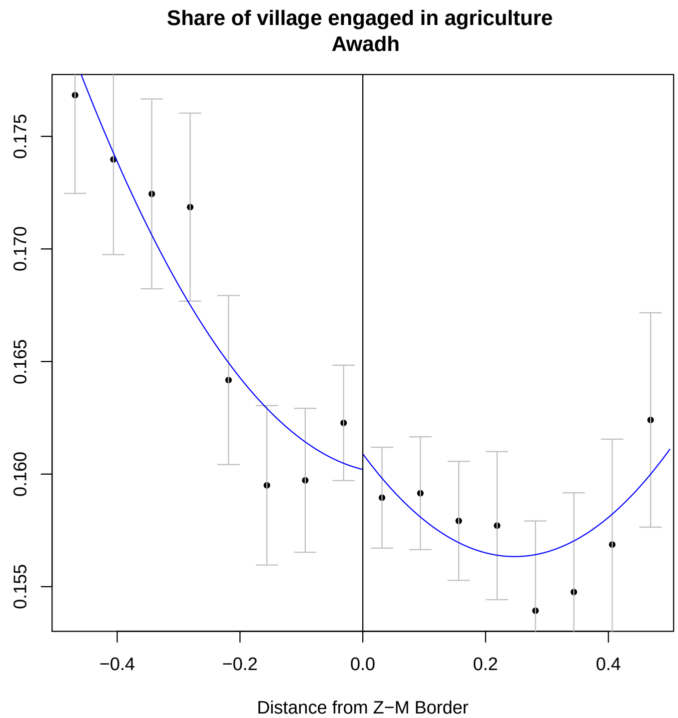


Figure 12: Agricultural Share of Labour

literacy outcomes. This paper documents the adverse side-effects of institutions that have generally been considered beneficial for development, in particular agricultural investment (as documented in Banerjee and Iyer (2005), though questioned by Iversen, Palmer-Jones, and K. Sen (2013)). While this may be true, the findings of this paper suggest that these institutions also entrench intra-household inequalities and hierarchies.

As with any research design that relies on a regression discontinuity, these findings are only plausibly causal near the cutoff (which, in this case, is the border between the Awadh principality's districts in 1801 AD), and therefore can only be extrapolated to other contexts when buttressed by plausible mechanisms and theory. The paper suggests some generalisable mechanisms for the observed effect, but these cannot be tested using available data. The ideal (hypothetical) experiment to study the effects of land-tenure systems would block-randomise land-tenure systems and collect individual and household level data on labour supply, fertility, and consumption rather than relying on village level aggregates. Furthermore, since the discontinuity in question is a geographical and administrative one, additional work must be done to verify that the boundary in question is truly exogenous (which I attempt to do with historical sources in section 2). In addition to this, since the boundaries in question are administrative, it is possible that a 'compound treatment' problem (Keele and Titiunik 2015) may arise such that the observed difference is not merely the result of land-tenure institutions but of district level policies adopted subsequently. While this is unlikely given that district level governance in India is not sizeable and most policies are adopted at the state level, it is a limitation worth keeping in mind.

The policy implications of this paper are not, however, that serfdom-like agricultural institutions that existed in *Zamindari* villages are preferable to giving property rights to the cultivator or village bodies; overall welfare was likely worse in *Zamindari* villages. This paper's findings simply suggest that property rights may serve to cement or exacerbate existing inequalities as they may exist in a given environment regardless of their (likely positive) first-order effects on investment and consumption. Therefore property rights may not be the panacea for development that some thinkers, like De Soto (2002), may believe it to be. Findings from economic history such as this one are useful because they dissuade from silver-bullet thinking, especially one that has been championed vociferously as property rights.

Bibliography

Alesina, Alberto, Paola Giuliano, and Nathan Nunn (2013). "On the Origins of Gender Roles: Women and the Plough". *The Quarterly Journal of Economics* 128.2, pp. 469–530.

Almond, Douglas, Hongbin Li, and Shuang Zhang (2019). "Land reform and sex selection in China". *Journal of Political Economy* 127.2, pp. 000–000.

Anderson, Siwan and Debraj Ray (2010). "Missing Women: Age and Disease". *The Review of Economic Studies* 77.4, pp. 1262–1300.

Banerjee, Abhijit and Lakshmi Iyer (2005). "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India". *American economic review* 95.4, pp. 1190–1213.

Barnett, Richard B. (1980). *North India Between Empires: Awadh, the Mughals, and the British, 1720-1801.* 31. Univ of California Press.

Bhalotra, Sonia et al. (Apr. 2019). "Property Rights and Gender Bias: Evidence from Land Reform in West Bengal". *American Economic Journal: Applied Economics* 11.2, pp. 205–37. doi: 10.1257/app.20160262. url: http://www.aeaweb.org/articles?id=10.1257/app.20160262.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik (2014). "Robust Nonparametric Confidence Intervals for Regression-discontinuity Designs". *Econometrica* 82.6, pp. 2295–2326.

Cattaneo, Matias D., Nicolás Idrobo, and Rocio Titiunik (2018). *A Practical Introduction to Regression Discontinuity Designs: Volume I.* Cambridge University Press, forthcoming.

Chakraborty, Tanika and Sukkoo Kim (2010). "Kinship Institutions and Sex Ratios in India". *Demography* 47.4, pp. 989–1012.

De Soto, Hernando (2002). *The Other Path: The Economic Answer to Terrorism.* Basic Books New York. isbn: 0-465-01610-3.

Dell, Melissa (2010). "The Persistent Effects of Peru's Mining Mita". *Econometrica* 78.6, pp. 1863–1903.

Denyer, Simon and Annie Gowen (2018). ""There's Too Many Men": What Happens When Women Are Outnumbered on a Massive Scale". en. *Washington Post.*

Economist, The (July 2018). "Why India Needs Women to Work". *The Economist.* issn: 0013-0613.

Fisher, Michael Herbert (1993). *The Politics of the British Annexation of India, 1757-1857.* Oxford University Press, USA.

— (1998). *Indirect Rule in India: Residents and the Residency System 1764-1857.* Oxford University Press, USA. isbn: 0-19-564419-0.

Gaure, Simen (2013). "Lfe: Linear Group Fixed Effects". *The R Journal* 5.2, pp. 104–117.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw (2001). "Identification and Estimation of Treatment Effects with a Regression-discontinuity Design". *Econometrica* 69.1, pp. 201–209.

Hunter, John D. (2007). "Matplotlib: A 2D Graphics Environment". *Computing in science & engineering* 9.3, pp. 90–95.

Imbens, Guido W. and Thomas Lemieux (2008). "Regression Discontinuity Designs: A Guide to Practice". *Journal of econometrics* 142.2, pp. 615–635.

Imbens, Guido W. and Donald B. Rubin (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press. isbn: 0-521-88588-4.

Infomap, ML (n.d.[a]). *Bihar, India: Village Socio-Demographic and Economic Census Data, 2001*. en. https://purl.stanford.edu/js637zp2537.

— (n.d.[b]). *Uttar Pradesh East, India: Village Socio-Demographic and Economic Census Data, 2001*. en. https://purl.stanford.edu/jj024vd6848.

— (n.d.[c]). *Uttar Pradesh West, India: Village Socio-Demographic and Economic Census Data, 2001*. en. https://purl.stanford.edu/db699ws4965.

Iversen, Vegard, Richard Palmer-Jones, and Kunal Sen (Dec. 2013). "On the Colonial Origins of Agricultural Development in India: A Re-Examination of Banerjee and Iyer, 'History, Institutions and Economic Performance'". *The Journal of Development Studies* 49.12, pp. 1631–1646. issn: 0022-0388. doi: 10.1080/00220388.2013.807502.

IWS (n.d.). *Ten Countries with the Highest Population in the World*. https://www.internetworldstats.com/s

Joppen, Charles (1914). *Historical Atlas of India: For the Use of High Schools, Colleges, and Private Students*. Longmans, Green.

Jordahl, K. (2014). "GeoPandas: Python Tools for Geographic Data". *URL: https://github.com/geopandas/geopandas*.

Keele, Luke J. and Rocio Titiunik (2015). "Geographic Boundaries as Regression Discontinuities". *Political Analysis* 23.1, pp. 127–155.

Lipsey, Richard G. and Kelvin Lancaster (1956). "The General Theory of Second Best". *The review of economic studies* 24.1, pp. 11–32.

Lundberg, Shelly and Robert A. Pollak (1993). "Separate Spheres Bargaining and the Marriage Market". *Journal of political Economy* 101.6, pp. 988–1010.

McKinney, Wes (2011). "Pandas: A Foundational Python Library for Data Analysis and Statistics". *Python for High Performance and Scientific Computing*, pp. 1–9.

Metcalfe, Barbara and Thomas R. Metcalfe (2006). *A Concise History of Modern India*. Cambridge/New York: Cambridge University Press.

NOAA (n.d.). *Nighttime Lights Annual Composites V4 | Dataset Overview | National Centers for Environmental Information (NCEI)*. https://www.ngdc.noaa.gov/metaview/page?xml=NOAA/

Qian, Nancy (2008). "Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance". *The Quarterly Journal of Economics* 123.3, pp. 1251–1285.

Sen, Amartya (1990). "More than 100 Million Women Are Missing". *New York Review of Books*, pp. 61–66.

— (1992). "Missing Women." *BMJ: British Medical Journal* 304.6827, p. 587.

Stokes, Eric (1983). "Agrarian Relations: Northern and Central India". *The Cambridge Economic History of India* 2, pp. 1750–1970.

Walt, Stéfan van der, S. Chris Colbert, and Gael Varoquaux (2011). "The NumPy Array: A Structure for Efficient Numerical Computation". *Computing in Science & Engineering* 13.2, pp. 22–30.

Wickham, Hadley (2010). "Ggplot2: Elegant Graphics for Data Analysis". *J Stat Softw* 35.1, pp. 65–88.

WICKHAM, Hadley et al. (2015). "Dplyr: A Grammar of Data Manipulation". *R package version 0.4* 3.

A Appendix

A.1 Household Model with Exogamy and Endogenous Sex-Selection

Households live for 2 periods, and receive endowments of income (y_1, y_2) in the first and second period respectively. They decide how much to consume (c_1) and how much to save (s) for the next generation in the first period in order to consume c_2 in the second period. The household discounts future utility by a discount factor β. The household's utility is time-separable, so the household's preferences are given by

$$U(c_1, c_2) = u(c_1) + \beta u(c_2)$$

Credit markets are non-existent, so the household cannot borrow to finance consumption, or save through any channel outside property. The share of saving that passes into the next period is controlled by an exogenous parameter θ, which denotes property rights [18]. If the household has no property rights, θ is zero.

The household's maximisation problem is:

$$\max_{c_1, c_2, s} U(c_1, c_2) \text{ s.t.} \tag{3}$$

$$c_1 + s = y_1 \tag{4}$$

$$c_2 = \theta s + y_2 \tag{5}$$

Since there is no utility for positive savings at the end of the two periods, the constraints are binding. Substituting s from (1.2) into (1.3) yields a lifetime budget constraint :

$$c_1 + \frac{c_2}{\theta} = y_1 + \frac{y_2}{\theta}$$

This yields a simplified optimisation problem of the form:

$$\max_{c_1, c_2, s} U(c_1, c_2) \text{ s.t.} \tag{6}$$

$$c_1 + \frac{c_2}{\theta} = y_1 + \frac{y_2}{\theta} \tag{7}$$

The Lagrangian can be written as

$$L(c_1, c_2, \lambda) = u(c_1) + \beta u(c_2) + \lambda \left[y_1 + \frac{y_2}{\theta} - c_1 - \frac{c_2}{\theta} \right]$$

The solution to this problem is:

[18] This is identical to the conventional 2 period model where borrowing is permitted, except $\theta = (1 + r)$ where r is the interest rate.

$$
\begin{aligned}
FOC(c_1) \rightarrow & u'(c_1) = \lambda \\
FOC(c_2) \rightarrow & u'(c_2) = \frac{\lambda}{\theta} \\
FOC(\lambda) \rightarrow & c_1 + \frac{c_2}{\theta} = y_1 + \frac{y_2}{\theta}
\end{aligned}
$$

Putting together the first 2 FOCs yields the following Euler equation:

$$
\frac{u'(c_1)}{\beta u'(c_2)} = \theta
$$

Under a no-property-rights regime ($\theta = 0$), so the problem reduces to 2 separate optimisation problems for each time period (which is equivalent to perfect impatience where future utility is discounted fully i.e. $\beta = 0$). So, the optimal consumption for the household is to consume the entire income in each period (i.e. $c_t^* = y_t$).

However, if the household has property rights ($\theta > 0$), then it will save for the next generation ($s^* = y_1 - c_1^* > 0$). Therefore, $\frac{\partial s}{\partial \theta} > 0$, i.e. savings are increasing in property rights.

A.1.1 Exogamy and Sex-selection

Now, let us introduce gender of the progeny and exogamy into the model. Define progeny $\zeta = \{m, f\}$, (male and female) where $p(\zeta = m) = 0.5 + z$, where z is a household's choice parameter and represents the fact that the household can increase the likelihood of a male child through sex-selective abortion / differential neglect / female infanticide. Let this be bounded under 0.1 (or some other arbitrary number ψ below 0.5, because otherwise the solution only yields male children, which is biologically unsustainable and inconsistent with reality).

Exogamy: Furthermore, let it be the case that female progeny are married into another village and do not contribute to the household's budget, so saving for them is ineffective (i.e. s does not enter into the 2nd period budget if $\zeta = f$). So, the θs component in the 2nd period budget only gets realised if the household's child is male. This effectively means that the expectation of the saving component of the budget is $(0.5 + z)\theta s + 0 \times (1 - (0.5 + z)) = (0.5 + z)\theta s$.

The household's problem is now

$$
\begin{aligned}
& \max_{c_1, c_2, s, z} U(c_1, c_2) \text{ s.t.} \\
& c_1 + s = y_1 \\
& c_2 = \theta(0.5 + z)s + y_2 \\
& z \leq 0.1
\end{aligned}
$$

The lifetime budget constraint is

$$c_1 + \frac{c_2}{\theta(0.5+z)} = y_1 + \frac{y_2}{\theta(0.5+z)}$$

and the Lagrangian is:

$$L(c_1, c_2, \lambda, \mu) = u(c_1) + \beta u(c_2) + \lambda \left[y_1 + \frac{y_2}{(0.5+z)\theta} - c_1 - \frac{c_2}{(0.5+z)\theta} \right] + \mu[z - 0.1]$$

The solution to this problem is:

$$\begin{aligned}
FOC(c_1) &\rightarrow u'(c_1) = \lambda \\
FOC(c_2) &\rightarrow u'(c_2) = \frac{\lambda}{(0.5+z)\theta} \\
FOC(\lambda) &\rightarrow c_1 + \frac{c_2}{(0.5+z)\theta} = y_1 + \frac{y_2}{(0.5+z)\theta} \\
FOC(\mu) &\rightarrow z = 0.1
\end{aligned}$$

$$\frac{u'(c_1)}{\beta u'(c_2)} = (0.5+z)\theta$$

Under a no-property-rights regime ($\theta = 0$), the last constraint is slack ($\mu = 0$), so $z^* = 0$, and therefore there is no son preference and the solution is the same as the no-property-rights solution above (since s is not included in the 2nd period budget regardless of the gender of the progeny, and households consume whatever income they have in each period). However, under a property rights system ($\theta > 0$, the household will prefer sons because that increases 2nd period consumption, so $z^* = 0.1$ (or whatever else we set the upper bound on son-preference to be). In summary, $\frac{\partial z}{\partial \theta} > 0$. This simple model illustrates that in a model with exogamy and endogenous sex-selection, households will prefer sons if they have property rights (which endow them with more consumption in the future).

A.2 Labelled Map - Awadh

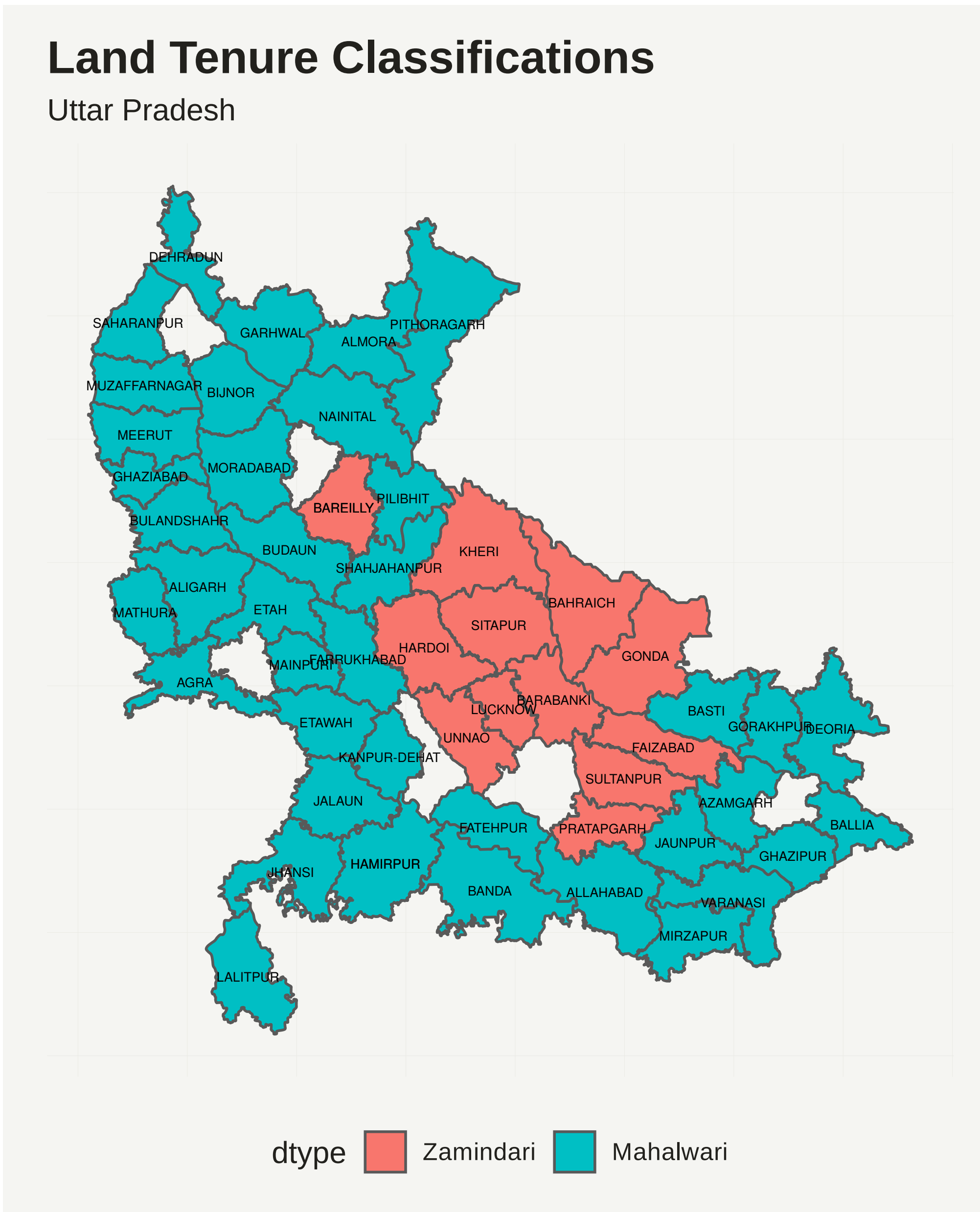


Figure 13: Labelled Map of UP Districts with Land Tenure Classifications

A.3 Summary Statistics - Full Sample, 10, 5 KM buffers

Table 11: Descriptive Statistics: All Villages

Statistic	N	Mean	St. Dev.	Min	Max
Non-Landlord Status	119,223	0.512	0.500	0	1
Household Size	119,223	6.500	1.080	1.570	140.000
Population	119,223	1,574.000	1,828.000	100	99,506
Share Scheduled-Caste	119,223	0.226	0.193	0.000	1.000
Literacy	119,223	0.402	0.130	0.000	1.000
Female Literacy	119,223	0.276	0.129	0.000	1.000
Sex Ratio	119,223	1.100	0.130	0.516	14.500
Labour Force Participation Rate	119,223	0.345	0.103	0.000	1.000
Female Labour Force Participation Rate	119,223	0.200	0.173	0.000	1.000
Night-time Luminosity (DMSP 2001)	119,223	2.000	3.770	0.000	63.000

Table 12: Descriptive Statistics: Villages within 10 km of border

Statistic	N	Mean	St. Dev.	Min	Max
Non-Landlord Status	25,300	0.576	0.494	0	1
Household Size	25,300	6.750	0.912	2.730	17.000
Population	25,300	1,301.000	1,424.000	100	28,836
Share Scheduled-Caste	25,300	0.209	0.183	0.000	1.000
Literacy	25,300	0.401	0.118	0.003	0.878
Female Literacy	25,300	0.276	0.120	0.000	0.827
Sex Ratio	25,300	1.090	0.165	0.586	11.200
Labour Force Participation Rate	25,300	0.325	0.099	0.010	0.956
Female Labour Force Participation Rate	25,300	0.169	0.167	0.000	0.929
Night-time Luminosity (DMSP 2001)	25,300	1.640	2.810	0.000	48.000

Table 13: Descriptive Statistics: Villages within 5 km of border

Statistic	N	Mean	St. Dev.	Min	Max
Non-Landlord Status	12,155	0.562	0.496	0	1
Household Size	12,155	6.750	0.913	2.730	17.000
Population	12,155	1,293.000	1,432.000	100	28,836
Share Scheduled-Caste	12,155	0.210	0.186	0.000	1.000
Literacy	12,155	0.399	0.119	0.003	0.776
Female Literacy	12,155	0.274	0.122	0.000	0.738
Sex Ratio	12,155	1.090	0.148	0.586	5.480
Labour Force Participation Rate	12,155	0.323	0.097	0.038	0.956
Female Labour Force Participation Rate	12,155	0.164	0.165	0.000	0.929
Night-time Luminosity (DMSP 2001)	12,155	1.500	2.800	0.000	37.200

A.4 Balance Tables

Here, I report treatment and control means and standard errors for potential confounders, the difference in means, as well as standardised difference in means, defined by Imbens and Rubin (2015) as

$$\frac{\overline{X}_t - \overline{X}_c}{\sqrt{(\hat{\sigma}_t^2 + \hat{\sigma}_c^2)/2}}$$

where $\bar{X}_t$ and $\hat{\sigma}_t$ are treatment mean and standard deviations respectively. The normalisation is desirable because a t-test of equality of means mechanically yields low p-values of equal means in large samples (like the ones under consideration), even though the underlying covariate balance may be reasonable. When normalised, the rule of thumb is that a difference in means of less than 0.5 implies that the overlap condition is less likely to have been violated. By this heuristic, the zamindari and non-zamindari samples are well balanced along observables.

Table 14: Balance Table - Full Sample

	Control Mean	Control SE	Treatment Mean	Treatment SE	Difference in Means	Normalised Difference
Household Size	6.22	0.01	6.57	0.01	0.35	0.10
Total Population	1430.39	8.19	1239.11	8.19	-191.28	-0.05
Share Scheduled Caste	0.27	0.00	0.23	0.00	-0.04	-0.07
Share Literate	0.38	0.00	0.43	0.00	0.05	0.15
Labour Force Participation Rate	0.35	0.00	0.34	0.00	-0.01	-0.05
Village Night Lights (2001) Mean	2.02	0.02	1.74	0.02	-0.28	-0.03

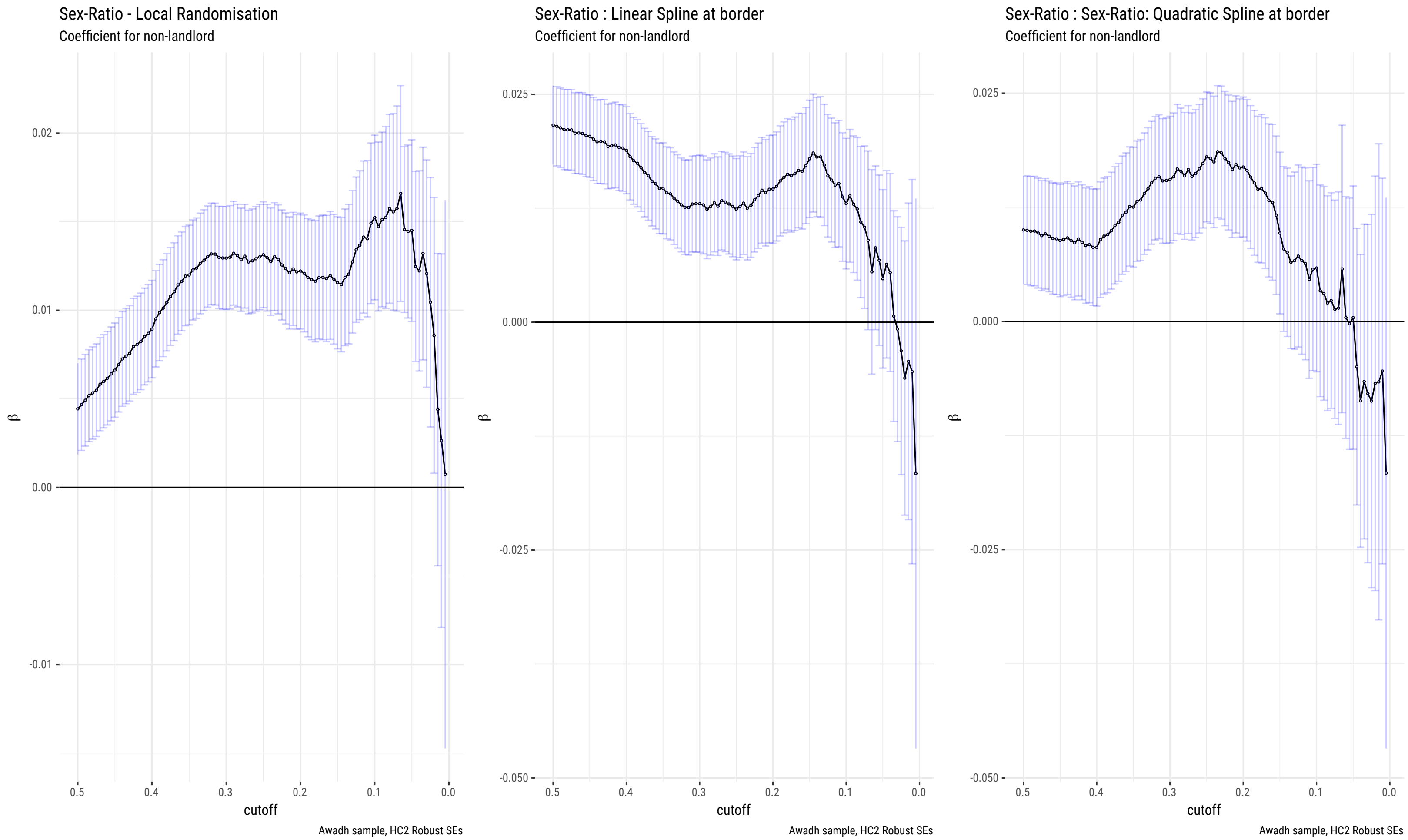
Table 15: Balance Table - 10 km buffer

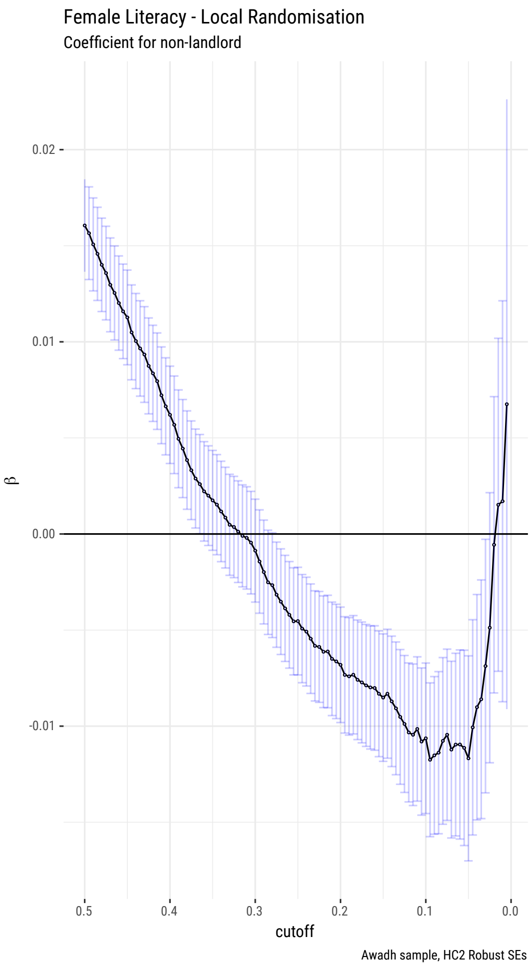
	Control Mean	Control SE	Treatment Mean	Treatment SE	Difference in Means	Normalised Difference
Household Size	6.63	0.01	6.62	0.01	-0.01	-0.00
Total Population	1239.32	10.09	1136.10	10.09	-103.22	-0.04
Share Scheduled Caste	0.23	0.00	0.21	0.00	-0.02	-0.04
Share Literate	0.41	0.00	0.39	0.00	-0.02	-0.06
Labour Force Participation Rate	0.33	0.00	0.33	0.00	-0.01	-0.03
Village Night Lights (2001) Mean	1.96	0.03	1.62	0.03	-0.33	-0.04

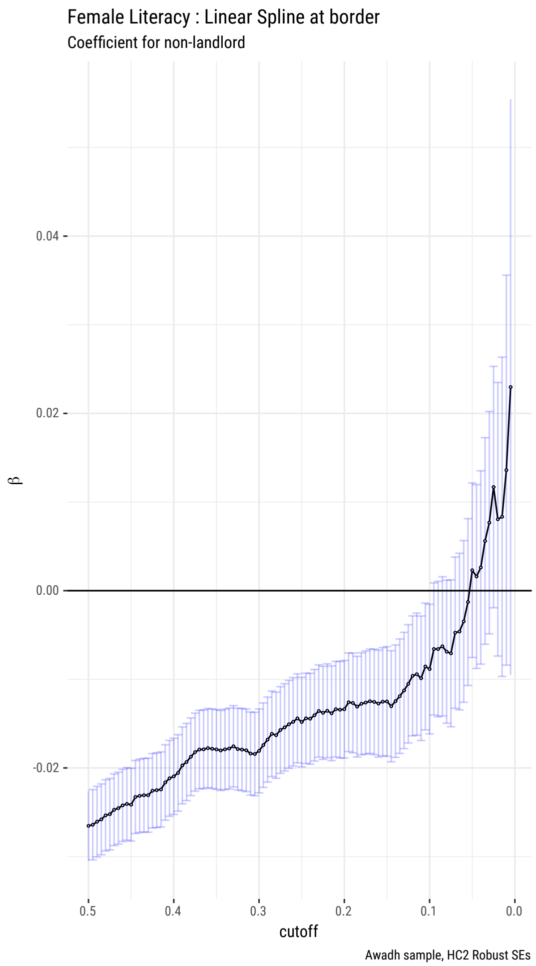
Table 16: Balance Table - 5 km buffer

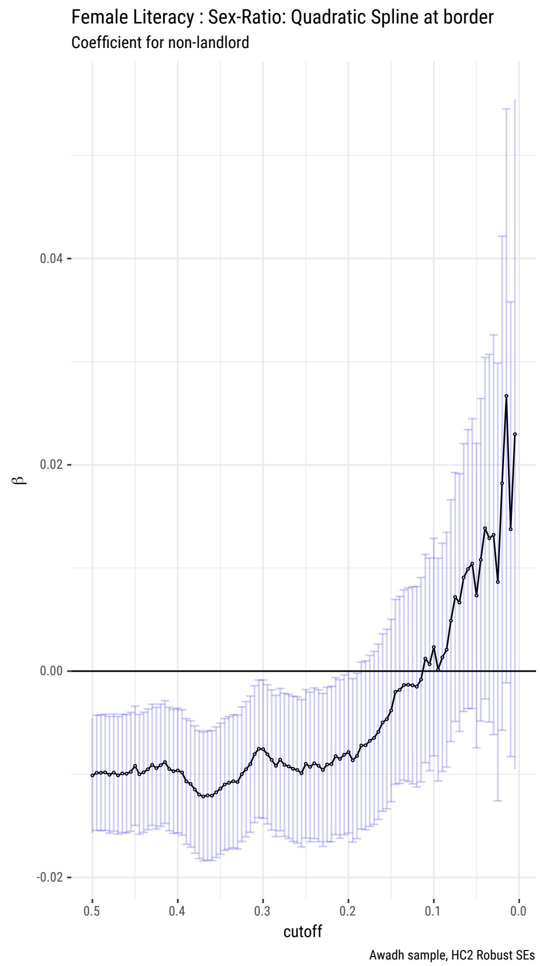
	Control Mean	Control SE	Treatment Mean	Treatment SE	Difference in Means	Normalised Difference
Household Size	6.64	0.01	6.63	0.01	-0.02	-0.01
Total Population	1224.61	14.28	1145.56	14.28	-79.05	-0.03
Share Scheduled Caste	0.23	0.00	0.21	0.00	-0.02	-0.03
Share Literate	0.41	0.00	0.39	0.00	-0.02	-0.06
Labour Force Participation Rate	0.33	0.00	0.32	0.00	-0.01	-0.03
Village Night Lights (2001) Mean	1.79	0.04	1.50	0.04	-0.29	-0.04

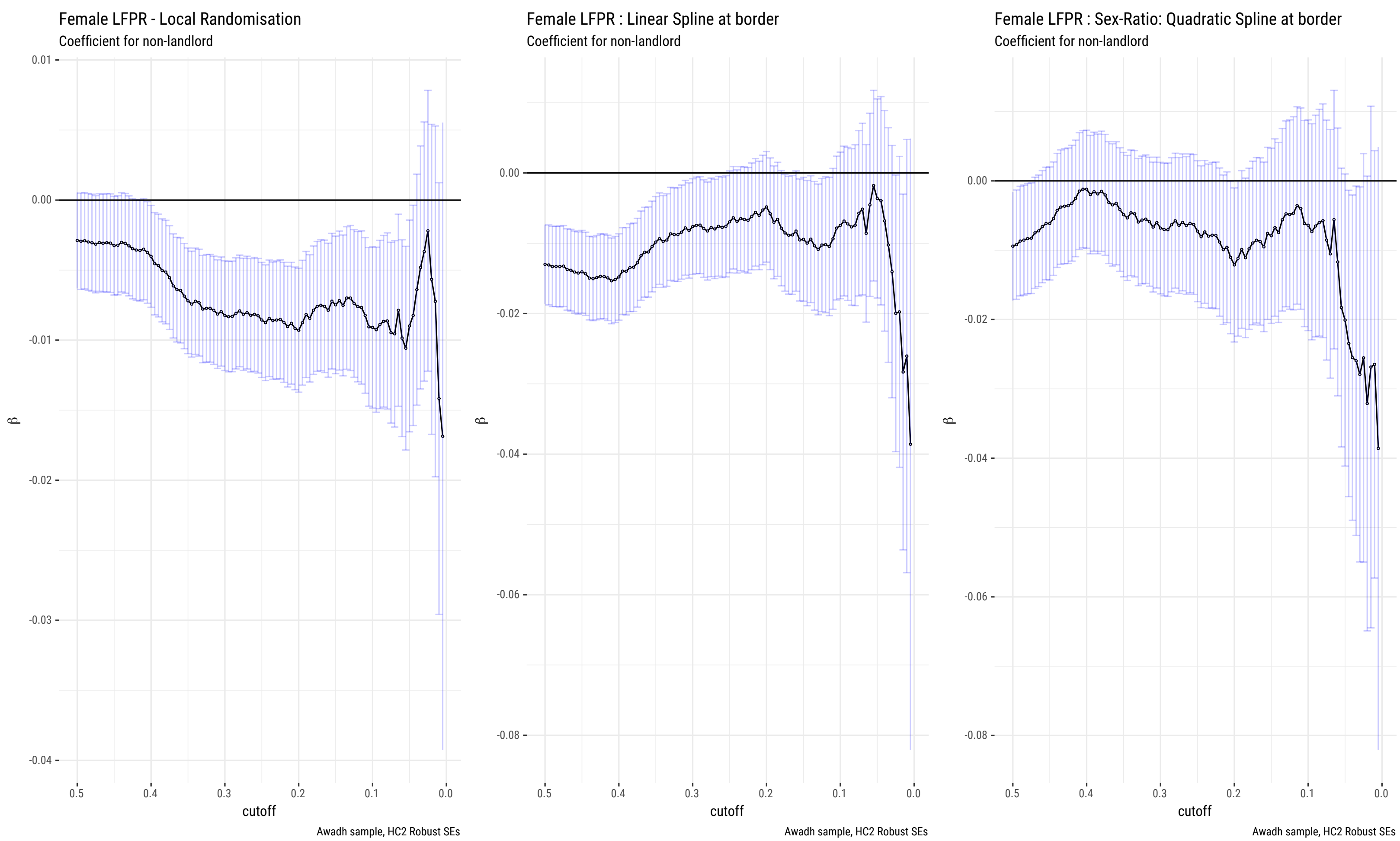
A.5 Coefficient Plots with linear and quadratic splines




Female Literacy - Local Randomisation
Coefficient for non-landlord
β
0.02
0.01
0.00
-0.01
0.5
0.4
0.3
0.2
0.1
0.0
cutoff
Awadh sample, HC2 Robust SEs


Female Literacy : Linear Spline at border
Coefficient for non-landlord
β
0.04
0.02
0.00
-0.02
0.5
0.4
0.3
0.2
0.1
0.0
cutoff
Awadh sample, HC2 Robust SEs


Female Literacy : Sex-Ratio: Quadratic Spline at border
Coefficient for non-landlord
β
0.04
0.02
0.00
-0.02
0.5
0.4
0.3
0.2
0.1
0.0
cutoff
Awadh sample, HC2 Robust SEs


Female LFPR - Local Randomisation
Coefficient for non-landlord
β
cutoff
Awadh sample, HC2 Robust SEs
0.01
0.00
-0.01
-0.02
-0.03
-0.04
0.5
0.4
0.3
0.2
0.1
0.0
Female LFPR : Linear Spline at border
Coefficient for non-landlord
β
cutoff
Awadh sample, HC2 Robust SEs
0.00
-0.02
-0.04
-0.06
-0.08
Female LFPR : Sex-Ratio: Quadratic Spline at border
Coefficient for non-landlord
β
cutoff
Awadh sample, HC2 Robust SEs

A.6 List of Variables in Geocoded Census

RURAL POPULATION

R_TOT_NM_HH	Total rural number of household
R_TOT_POP	Total rural population
R_M_POP	Rural male population
R_F_POP	Rural female population
R_TOT_L6	Total rural pop below 6 years
R_M_L6	Rural male pop below 6 years
R_F_L6	Rural female pop below 6 years
R_TOT_SC	Total rural scheduled caste
R_M_SC	Rural male scheduled caste
R_F_SC	Rural female scheduled caste
R_TOT_ST	Total rural scheduled tribe
R_M_ST	Rural male scheduled tribe
R_F_ST	Rural female scheduled tribe

RURAL LITERACY

R_TOT_LIT	Total rural literates
R_M_LIT	Rural male literates
R_F_LIT	Rural female literates
R_TOT_ILLT	Total rural Illiterates
R_M_ILLT	Rural male Illiterates
R_F_ILLT	Rural female Illiterates

RURAL WORKERS

R_TOT_W	Total rural workers
R_M_W	Rural male workers
R_F_W	Rural female workers
R_TOT_MNW	Total rural main workers
R_M_MNW	Rural male main workers
R_F_MNW	Rural female main workers
R_TOT_CULT	Total rural Cultivators
R_M_CULT	Rural male Cultivators
R_F_CULT	Rural female Cultivators
R_TOT_AGLB	Total rural agricultural labourers
R_M_AGLB	Rural male agricultural labourers
R_F_AGLB	Rural female agricultural labourers
R_TOT_MFHH	Total rural Household industry workers
R_M_MFHH	Rural male Household industry workers
R_F_MFHH	Rural female Household industry workers
R_TOT_OTH_W	Total rural other workers
R_M_OTH_W	Rural male other workers
R_F_OTH_W	Rural female other workers
R_TOT_MRW	Total rural Marginal workers other workers
R_M_MRW	Rural male Marginal workers other workers
R_F_MRW	Rural female Marginal workers other workers
R_T_MRG_CUL	Total rural Marginal workers as cultivators
R_M_MRG_CUL	Rural male Marginal workers as cultivators
R_F_MRG_CUL	Rural female Marginal workers as cultivators
R_T_MRG_AGL	Total rural Marginal workers as agricultural labourers
R_M_MRG_AGL	Rural male Marginal workers as agricultural labourers
R_F_MRG_AGL	Rural female Marginal workers as agricultural labourers
R_T_MRG_HH	Total rural Marginal workers household industry workers
R_M_MRG_HH	Rural male Marginal workers household industry workers
R_F_MRG_HH	Rural female Marginal workers household industry workers
R_T_MRG_OTH	Total rural Marginal workers as other workers
R_M_MRG_OTH	Rural male Marginal workers as other workers
R_F_MRG_OTH	Rural female Marginal workers as other workers
R_TOT_NNW	Total rural Non-workers
R_M_NNW	Rural male Non-workers
R_F_NNW	Rural female Non-workers

A.7 Trends in Luminosity

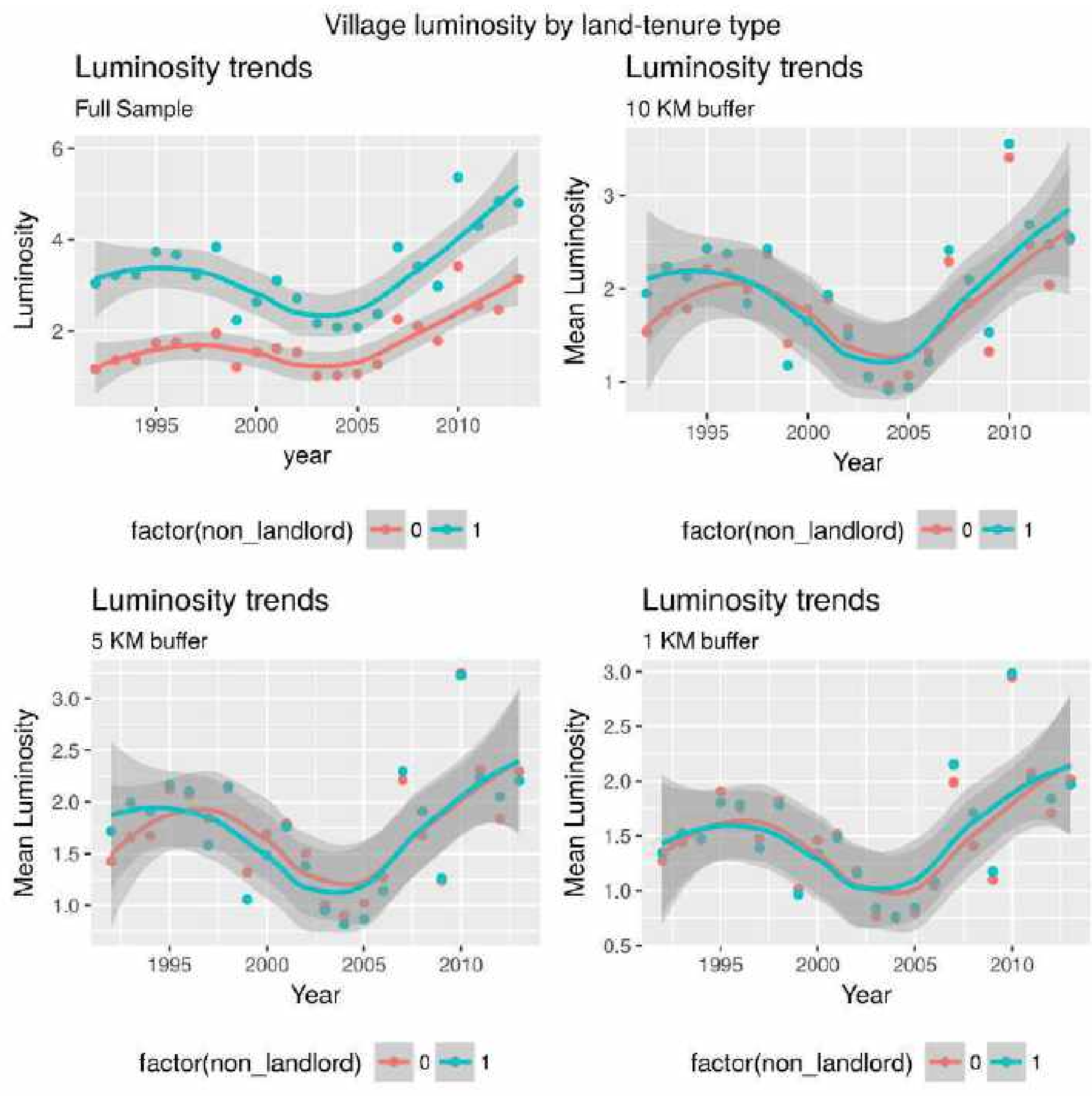
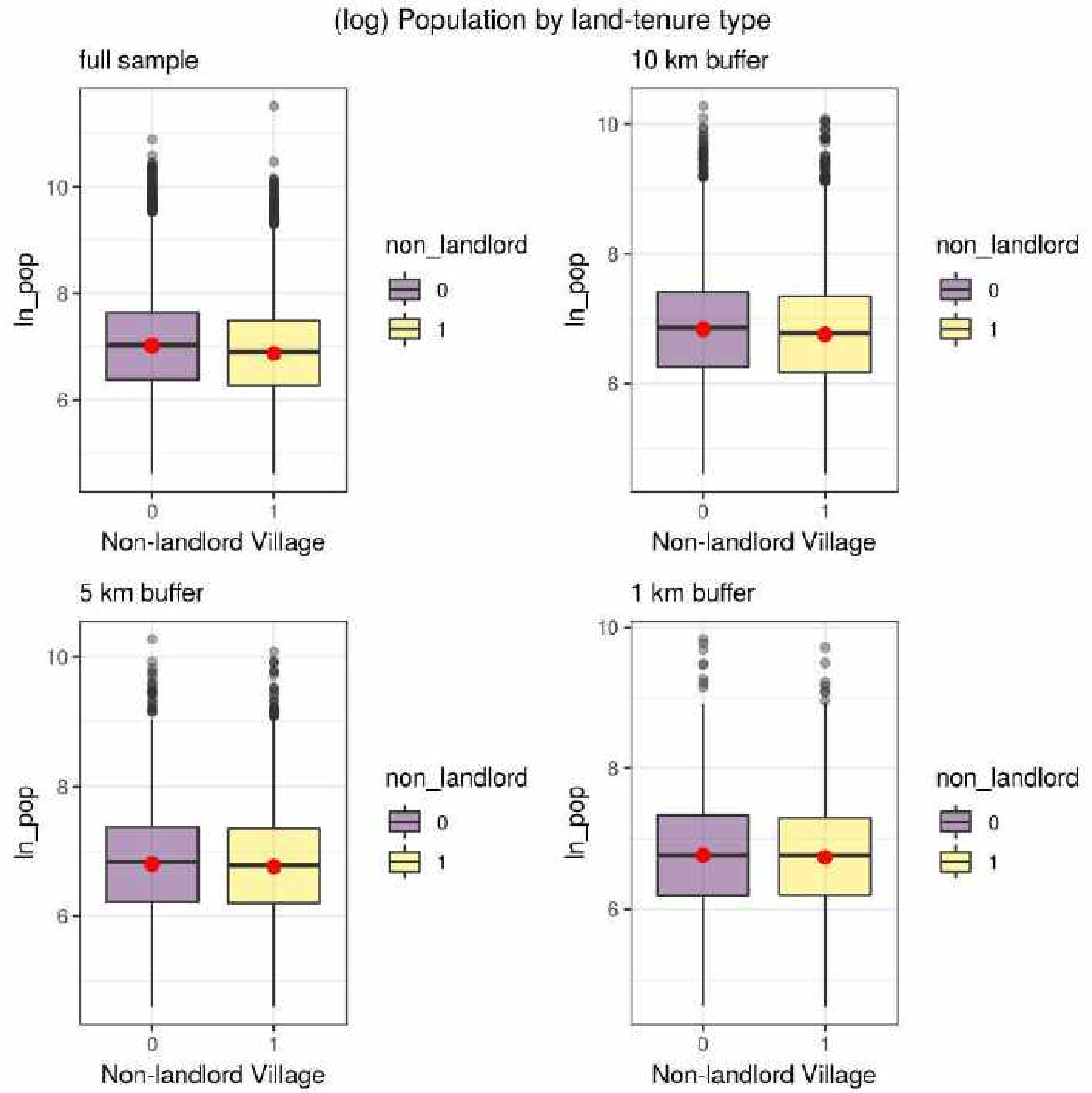
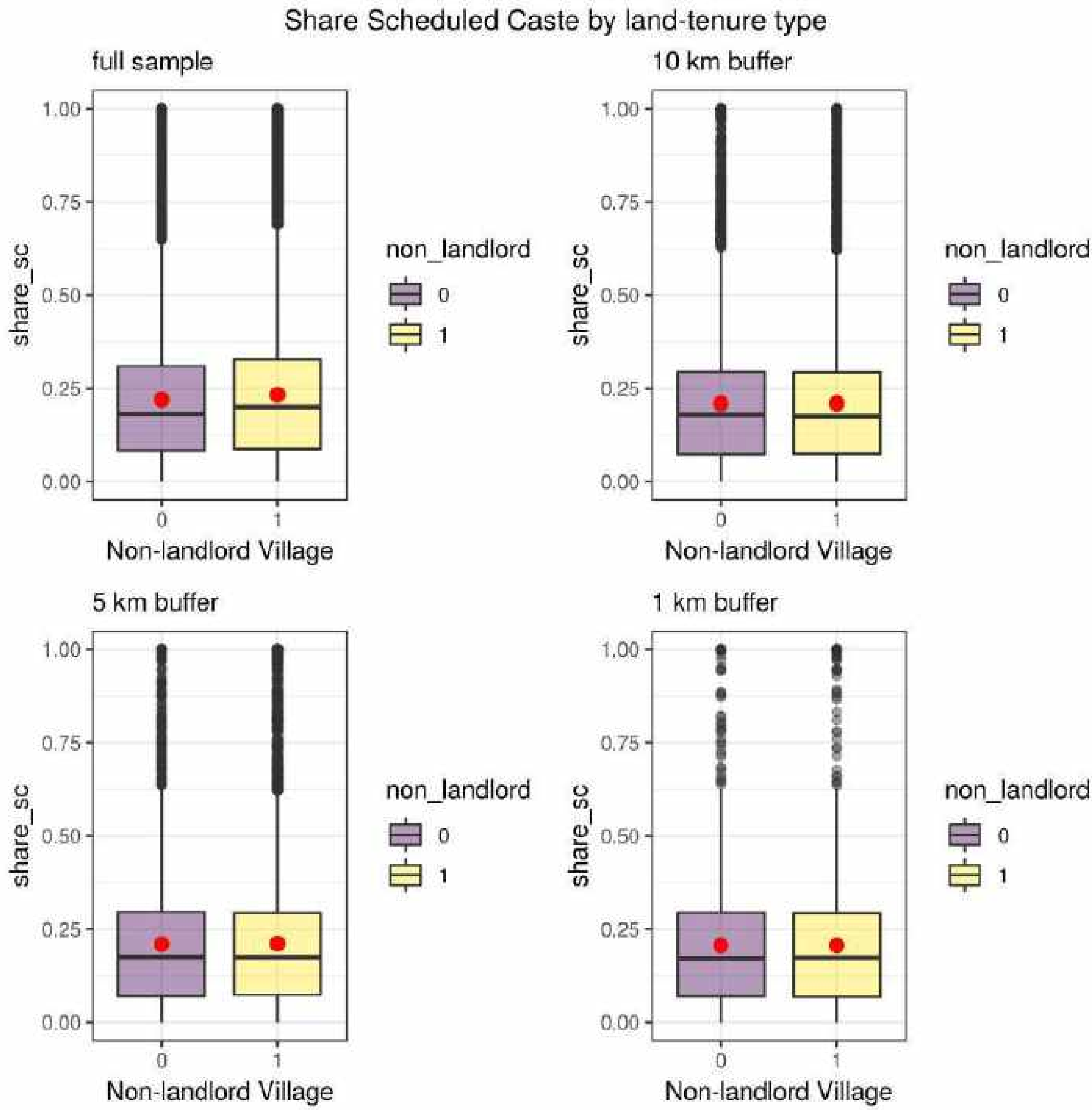
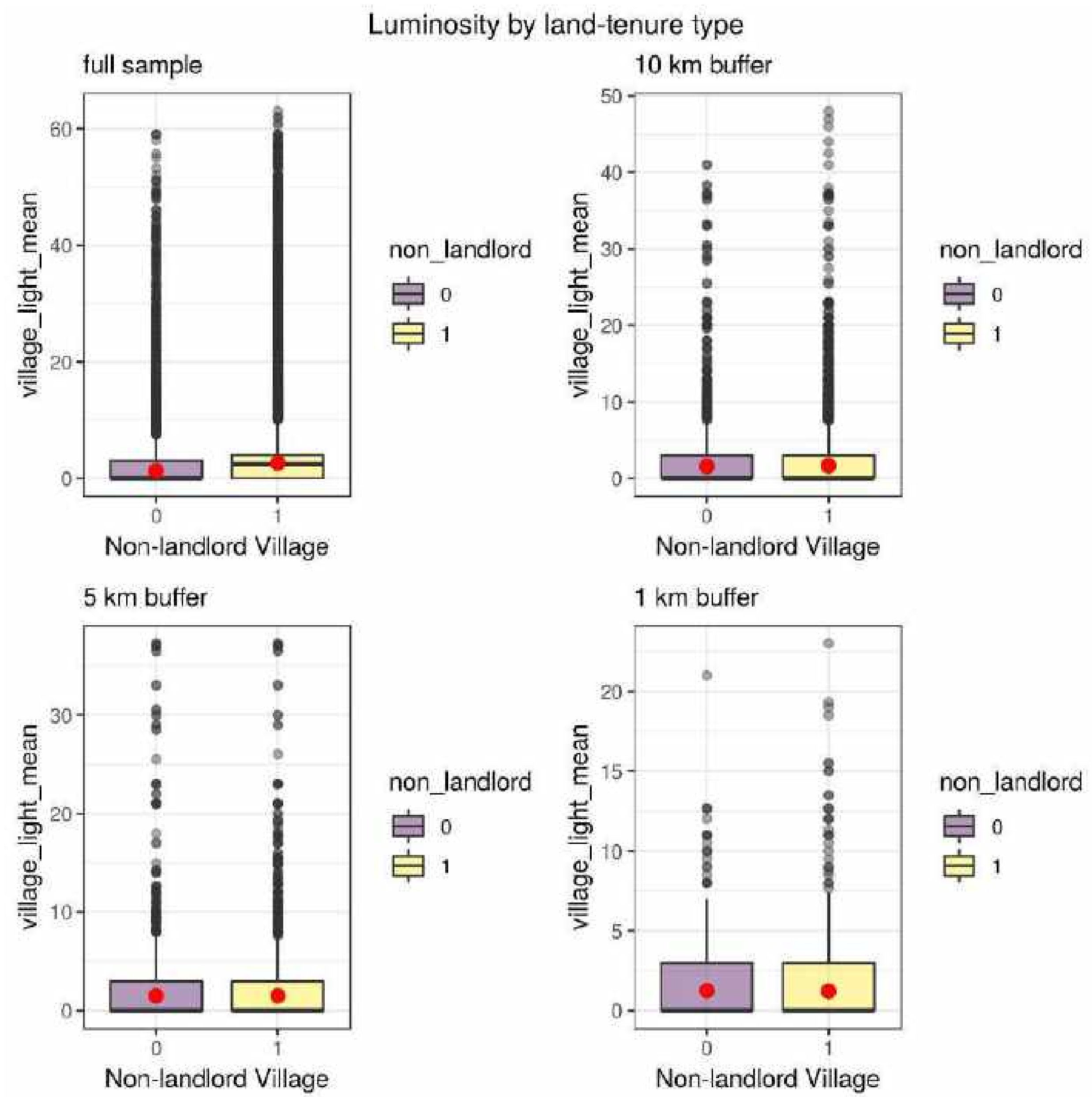


Figure 14: Trends in village luminosity - Awadh

A.8 Boxplots of potential confounders by treatment status




Share Scheduled Caste by land-tenure type
full sample
10 km buffer
5 km buffer
1 km buffer
share_sc
1.00
0.75
0.50
0.25
0.00
0
1
Non-landlord Village
non_landlord
0
1


Luminosity by land-tenure type
full sample
10 km buffer
5 km buffer
1 km buffer
village_light_mean
Non-landlord Village
non_landlord
0
1

A.9 OLS estimates with village controls

A.9.1 Awadh

Table 17: Literacy - Awadh

	Dependent variable:			
	Female Literacy			
	All	10 km	5 km	1 km
	(1)	(2)	(3)	(4)
Non-landlord	0.024***	−0.022***	−0.021***	−0.011**
	(0.001)	(0.002)	(0.003)	(0.005)
Share Scheduled-Caste	−0.002	0.045***	0.038***	0.042***
	(0.003)	(0.005)	(0.007)	(0.013)
log Population	−0.007***	−0.010***	−0.010***	−0.013***
	(0.001)	(0.001)	(0.002)	(0.003)
Longitude	−0.022***			
	(0.001)			
Latitude	−0.062***			
	(0.001)			
Village Mean Luminosity (2001)	0.004***	0.003***	0.004***	0.005***
	(0.0002)	(0.0004)	(0.001)	(0.001)
Constant	3.770***	0.338***	0.344***	0.343***
	(0.058)	(0.009)	(0.012)	(0.025)
Observations	47,237	17,931	9,052	2,068
R^2	0.176	0.025	0.025	0.022
Adjusted R^2	0.176	0.025	0.024	0.020
Note:	*p<0.1; **p<0.05; ***p<0.01			

Table 18: Female Labour Force Participation - Awadh

	Dependent variable:			
	Female Labour Force Participation			
	All	10 km	5 km	1 km
	(1)	(2)	(3)	(4)
Non-landlord	-0.007^{***}	-0.025^{***}	-0.027^{***}	-0.033^{***}
	(0.002)	(0.003)	(0.004)	(0.008)
Share Scheduled-Caste	0.007	0.067^{***}	0.083^{***}	0.113^{***}
	(0.004)	(0.007)	(0.010)	(0.022)
log Population	-0.002^{**}	-0.011^{***}	-0.004^{*}	-0.005
	(0.001)	(0.002)	(0.002)	(0.005)
Longitude	0.024^{***}			
	(0.001)			
Latitude	-0.052^{***}			
	(0.001)			
Village Mean Luminosity (2001)	-0.003^{***}	0.003^{***}	0.005^{***}	0.009^{***}
	(0.0002)	(0.0004)	(0.001)	(0.002)
Constant	-0.341^{***}	0.231^{***}	0.175^{***}	0.175^{***}
	(0.088)	(0.011)	(0.015)	(0.032)
Observations	47,237	17,931	9,052	2,068
R^2	0.146	0.016	0.023	0.044
Adjusted R^2	0.146	0.016	0.023	0.042
Note:			*p<0.1; **p<0.05; ***p<0.01	

A.9.2 Full Sample

Table 19: Sex Ratio (M/F)

	Dependent variable:			
	Sex Ratio			
	All	10 km	5 km	1 km
	(1)	(2)	(3)	(4)
Non-landlord	−0.035***	0.026***	0.030***	0.022***
	(0.001)	(0.002)	(0.003)	(0.006)
Share Scheduled-Caste	−0.019***	−0.042***	−0.050***	−0.077***
	(0.002)	(0.006)	(0.007)	(0.016)
log Population	0.001**	−0.004***	−0.004**	−0.002
	(0.0005)	(0.001)	(0.002)	(0.003)
Longitude	−0.018***			
	(0.0002)			
Latitude	0.004***			
	(0.0004)			
Village Mean Luminosity (2001)	0.0002	0.003***	0.001	−0.002
	(0.0001)	(0.001)	(0.001)	(0.001)
Constant	2.480***	1.100***	1.110***	1.110***
	(0.025)	(0.008)	(0.011)	(0.025)
Observations	119,223	25,300	12,155	2,594
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table 20: Female Literacy

	Dependent variable:			
	Female Literacy			
	All	10 km	5 km	1 km
	(1)	(2)	(3)	(4)
Non-landlord	0.033***	−0.013***	−0.017***	−0.011**
	(0.001)	(0.002)	(0.002)	(0.005)
Share Scheduled-Caste	−0.028***	0.023***	0.019***	0.034***
	(0.002)	(0.004)	(0.006)	(0.013)
log Population	0.001**	−0.004***	−0.004***	−0.006**
	(0.0004)	(0.001)	(0.001)	(0.003)
Longitude	−0.013***			
	(0.0002)			
Latitude	−0.023***			
	(0.0005)			
Village Mean Luminosity (2001)	0.004***	0.004***	0.005***	0.005***
	(0.0001)	(0.0004)	(0.001)	(0.001)
Constant	1.940***	0.296***	0.302***	0.298***
	(0.029)	(0.007)	(0.010)	(0.021)
Observations	119,223	25,300	12,155	2,594

Note: *p<0.1; **p<0.05; ***p<0.01

Table 21: Female Labour Force Participation

	Dependent variable:			
	Female Labour Force Participation			
	All	10 km	5 km	1 km
	(1)	(2)	(3)	(4)
Non-landlord	0.016***	−0.012***	−0.015***	−0.024***
	(0.001)	(0.002)	(0.003)	(0.006)
Share Scheduled-Caste	0.058***	0.082***	0.094***	0.115***
	(0.003)	(0.006)	(0.009)	(0.020)
log Population	−0.007***	−0.010***	−0.006***	−0.008**
	(0.001)	(0.001)	(0.002)	(0.004)
Longitude	−0.003***			
	(0.0003)			
Latitude	−0.041***			
	(0.001)			
Village Mean Luminosity (2001)	−0.002***	0.002***	0.003***	0.008***
	(0.0001)	(0.0004)	(0.001)	(0.001)
Constant	1.530***	0.223***	0.190***	0.193***
	(0.039)	(0.009)	(0.012)	(0.026)
Observations	119,223	25,300	12,155	2,594
Note:			*p<0.1; **p<0.05; ***p<0.01	

A.10 Estimates on Full Sample

A.10.1 OLS

Table 22: Sex Ratio (M/F)

	Dependent variable:		
	Sex Ratio		
	All	10 km	5 km
	(1)	(2)	(3)
Non-landlord	-0.020^{***}	0.009^{***}	0.009^{***}
	(0.001)	(0.002)	(0.002)
Line-segment fixed effects	X	X	X
Control Mean	1.09	1.07	1.07
Observations	119,223	25,300	12,155
R^2	0.164	0.229	0.307

Table 23: Female Literacy

	Dependent variable:		
	Female Literacy		
	All	10 km	5 km
	(1)	(2)	(3)
Non-landlord	0.054^{***}	-0.004^{**}	-0.008^{***}
	(0.001)	(0.001)	(0.002)
Line-segment fixed effects	X	X	X
Control Mean	0.245	0.283	0.283
Observations	119,223	25,300	12,155
R^2	0.085	0.100	0.090

Table 24: Female Labour Force Participation

	Dependent variable:		
	Female Labour Force Participation		
	All	10 km	5 km
	(1)	(2)	(3)
Non-landlord	0.020***	0.003	0.002
	(0.001)	(0.002)	(0.003)
Line-segment fixed effects	X	X	X
Control Mean	0.203	0.176	0.173
Observations	119,223	25,300	12,155
R^2	0.082	0.157	0.183

A.10.2 Nonparametric Distance RD

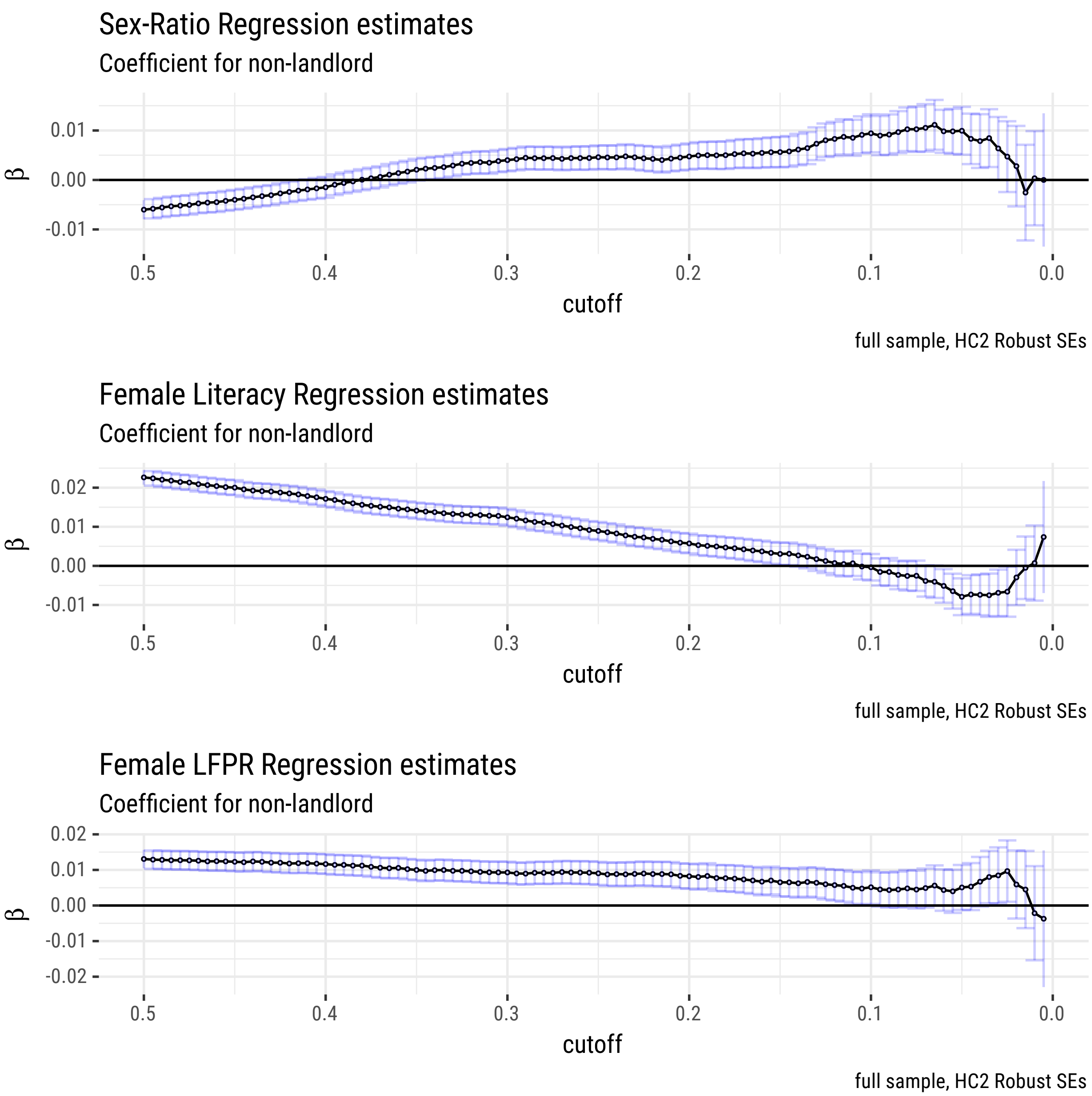


Figure 15: Non-landlord OLS Coefficients : Full Sample

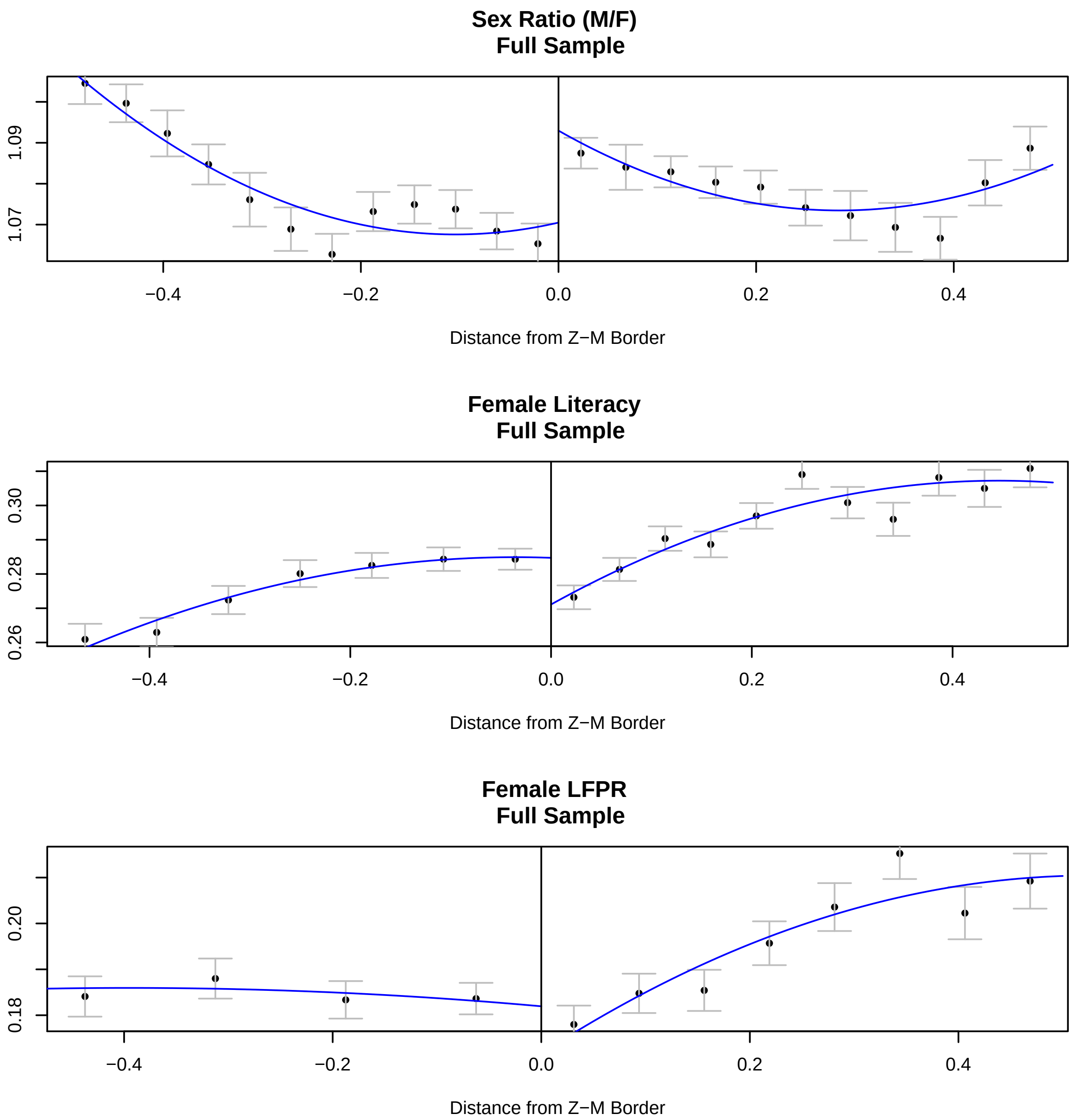


Figure 16: RD Estimates for entire sample

Table 25: Sex-Ratio RD estimates - Full Sample

	Parameter	(1)	(2)	(3)	(4)
1	Coeff	0.0222	0.0228	0.012	0.0076
2	SE	0.0035	0.0039	0.0025	0.003
3	z	6.3344	5.8282	4.8192	2.5191
4	P Val	0	0	0	0.0118
5	CI Lower	0.0153	0.0151	0.0071	0.0017
6	CI Upper	0.0291	0.0305	0.0169	0.0135
7	Left Bandwidth	0.282	0.3601	0.4429	0.4983
8	N (left of c)	17109	20498	23935	25996
9	Right Bandwidth	0.282	0.3601	0.4429	0.4983
10	N (right of c)	25421	30125	34464	37175
11	Polynomial Degree	1	2	1	2
12	Line-segment Fixed Effects			X	X

Table 26: Female Literacy RD estimates - Full Sample

	Parameter	(1)	(2)	(3)	(4)
1	Coeff	-0.0147	-0.0148	-0.0084	-0.0084
2	SE	0.0027	0.003	0.0025	0.0028
3	z	-5.5247	-4.969	-3.2999	-3.0058
4	P Val	0	0	0.001	0.0026
5	CI Lower	-0.0199	-0.0206	-0.0134	-0.0138
6	CI Upper	-0.0095	-0.009	-0.0034	-0.0029
7	Left Bandwidth	0.2933	0.4819	0.3088	0.5016
8	N (left of c)	17609	25384	18294	26096
9	Right Bandwidth	0.2933	0.4819	0.3088	0.5016
10	N (right of c)	26122	36383	27069	37346
11	Polynomial Degree	1	2	1	2
12	Line-segment Fixed Effects			X	X

A.10.3 Semiparametric 2D RD - Full Sample

Table 27: Female LFPR RD estimates - Full Sample

	Parameter	(1)	(2)	(3)	(4)
1	Coeff	-0.0089	-0.0075	0.0041	0.003
2	SE	0.0034	0.0042	0.0034	0.004
3	z	-2.6596	-1.7685	1.2114	0.7513
4	P Val	0.0078	0.077	0.2257	0.4525
5	CI Lower	-0.0155	-0.0158	-0.0025	-0.0049
6	CI Upper	-0.0023	0.0008	0.0107	0.0109
7	Left Bandwidth	0.3786	0.4638	0.3323	0.4081
8	N (left of c)	21264	24729	19355	22506
9	Right Bandwidth	0.3786	0.4638	0.3323	0.4081
10	N (right of c)	31185	35485	28532	32741
11	Polynomial Degree	1	2	1	2
12	Line-segment Fixed Effects			X	X

Table 28: Full Sample - Sex Ratio (M/F) - Semiparametric Spatial RD estimates

	Dependent variable:			
	Sex Ratio (M/F)			
	10 km	5 km	10 km	5 km
	(1)	(2)	(3)	(4)
Non-landlord	0.009***	0.010***	0.009***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
Polynomial Degree	2	2	3	3
LSFE	X	X	X	X
Observations	25,300	12,155	25,300	12,155
R^2	0.240	0.323	0.280	0.376
Note:	*p<0.1; **p<0.05; ***p<0.01			

Table 29: Female Literacy - Semiparametric Spatial RD estimates

	Dependent variable:			
	Female Literacy			
	10 km	5 km	10 km	5 km
	(1)	(2)	(3)	(4)
Non-landlord	-0.007^{***}	-0.009^{***}	-0.006^{***}	-0.007^{***}
	(0.001)	(0.002)	(0.001)	(0.002)
Polynomial Degree	2	2	3	3
LSFE	X	X	X	X
Observations	25,300	12,155	25,300	12,155
R^2	0.147	0.146	0.199	0.205
Note:	*p<0.1; **p<0.05; ***p<0.01			

Table 30: Female LFPR - Semiparametric Spatial RD estimates

	Dependent variable:			
	Female LFPR			
	10 km	5 km	10 km	5 km
	(1)	(2)	(3)	(4)
Non-landlord	0.001	0.001	−0.0001	0.0002
	(0.002)	(0.003)	(0.002)	(0.003)
Polynomial Degree	2	2	3	3
LSFE	X	X	X	X
Observations	25,300	12,155	25,300	12,155
R^2	0.206	0.220	0.206	0.220
Note:	*p<0.1; **p<0.05; ***p<0.01			

A.11 Functional Forms for RD

A.11.1 Cubic