Land Tenure and Missing Women: Evidence from North India

Apoorva Lal*

Current version: November 27, 2021

While a large literature documents a robust positive relationship between property rights and economic growth, the effects of institutions on demographic and socio-political outcomes are less well understood. This paper studies this relationship by using variation in property rights systems induced by historically contingent assignment of districts in the North Indian state of Uttar Pradesh to study their effects on modern-day demographic outcomes. We find that villages where property rights were granted to the cultivators have more skewed sex-ratios, lower female literacy, and lower female labor force participation rate than villages where property rights were granted to the landlord. These findings suggest that, in contrast to the robust positive relationship between institutions and economic outcomes, economic institutions may, depending on existing social norms, ameliorate or exacerbate social inequalities.

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. 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 labor market, an excessively high savings rate, property value inflation, and increases in violent crime, and trafficking. In India, women's outcomes are systematically worse in a variety of measures, including literacy, labor force participation, and education attainment, and there has recently been a disturbing increase in violence against women (Jayachandran 2015). The broad consensus on the causes is "A combination of cultural preferences, government decree, and modern medical technology" (Denyer and Gowen 2018). To address the likely causes of these problems, it is imperative that we first understand the multiple causes of the underlying demographic disparity.

This paper explores the institutional roots of the demographic imbalance in North India using a spatial regression discontinuity design. We compare village-level outcomes for women (sex-ratio, female literacy, and labor force participation) on either side of a historical administrative boundary that generated variation in land-tenure institutions[1], and find that these outcomes are worse in areas where individual farmers, as opposed to absent landlords, were granted property rights in North India. Biased sex ratios have persistent effects on conservative views on gender roles (Grosjean and Khattar 2019) and therefore can be expected to perpetuate gender inequality.

A potential mechanism for the patterns we document is the old-age support norm in patrilineal and patrilocal societies like India (wherein sons support and house parents in

---

1. land-tenure systems in the Indian subcontinent are typically classified under the broad categories of *Raiyatwari* := property rights granted to the cultivator, *Mahalwari* := property rights granted to the village council, or *Zamindari* := property rights granted to the local landlord. In the present paper, we compare historically Zamindari villages with historically Mahalwari ones

their old age, while daughters usually move to their husband's hometown following marriage (Jayachandran 2015)). More concretely, it implies the following explanation: 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 labor force participation and literacy among women in *Mahalwari* villages also suggests that the increased bargaining power for men granted by land rights leads to lower investments in female children, which in turn results in lower female labor participation. This mechanism is consistent with household bargaining models wherein changes in factor endowments (land given to the man) increases their bargaining power within the household (Wang 2014; Brulé 2018). This may lead to one agent seeking to curtail the other's 'outside option', which, in this case, is remunerated labor.

To credibly identify the effects of land-tenure institutions on demographic outcomes, we use variation in historical land-tenure institutions produced by the gradual British conquest of India. In particular, we 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, 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 thought of as a natural experiment in close proximity to the border *within the same state*, thereby holding most institutions con-

stant. Villages in near district borders were assigned into different land tenure systems that granted property rights to the village/farmer (*Mahalwari*) or the local landlord (*Zamindari*) [2]. Furthermore, since land taxes were abolished in India following independence in 1947, observed effects of these land tenure systems should be purely institutional overhang and persistence, rather than being confounded by contemporary factors.

We contribute to the extensive literature that examines the persistent effects of historical institutions on contemporary development outcomes. (Banerjee and Iyer 2005; Dell 2010; Hornbeck 2010; Acemoglu, Johnson, and Robinson 2001; Sokoloff and Engerman 2000; Haber, Maurer, and Razo 2003). We also contribute to the literature showing that property rights institutions may exacerbate existing socio-political inequalities by altering bargaining power and influencing fertility decisions (Brulé 2018; Bhalotra, Brulé, and Roy 2018; Kulkarni 2017). The paper is closely related to Banerjee and Iyer (2005) (henceforth BI) and use localised institutional heterogeneity, land-tenure systems in colonial India, to study institutional persistence in determining current economic outcomes and 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 [3].

This paper also contributes to an extensive literature that tackles the 'missing women' puzzle, started by Amartya Sen in a famous article that claimed that more than 100 million women are missing (Sen 1990, 1992) with the simple observation that men outnumber women by a margin too large to be consistent with biological explanations. Anderson and Ray (2010) and Calvi (2020) decompose the gap by age and disease and argue that the

---

2. While there are a variety of land-tenure systems in the subcontinent, they are typically classified under the broad categories of *Raiyatwari* := property rights granted to the cultivator, *Mahalwari* := property rights granted to the village council, or *Zamindari*:= property rights granted to the local landlord. Of the two systems under consideration in Uttar Pradesh, Mahalwari can be thought of as a hybrid between Zamindari and Raiyawari.

3. Iversen, Palmer-Jones, and 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 our analysis for this reason), as *Zamindari* districts. Omitting these areas result in BI's findings to shrink in substantive and statistical significance.

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 (as a consequence of highly asymmetric bargaining power within the household) during illnesses are likely to be a major factor in the observed gap as well. This suggests that intra-household inequality, differential neglect, and asymmetry in bargaining power is a likely factor driving the observed demographic gap.

This paper also contributes to an interdisciplinary literature that attempts to link institutions and agricultural practices to modern economic and social outcomes for women through household bargaining. This paper's findings demonstrate the persistence of these household dynamics by studying long-run effects of a shock in the household distribution of assets. 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 recent tenancy reforms in China and West Bengal, India respectively. Like this paper, they find that tenancy reform exacerbated son preference by increasing the sex ratio. Bhalotra, Brulé, and Roy (2018) find that equal inheritance rights for women exacerbate son preference in India, which suggests that changes in the costs of having daughters (through expectations of inheritance, combined with patrilocal marriage norms) are highly consequential. Our findings also relate to Qian (2008), who finds strong evidence linking the sex ratio in China to labor-market conditions induced by variation in tea prices, and more generally to the strong link theorised and studied in Alesina, Giuliano, and Nunn (2013), who argue that traditional agricultural practices influenced the gender division of labor and evolution of gender norms, and find empirical evidence suggesting that descendants of groups that practised plough agriculture have more unequal gender norms. A major implication of this literature is that intra-household bargaining is likely to be a major factor driving the observed demographic imbalance between male and female children in the long run (Chakraborty and Kim 2010; Lundberg and Pollak 1993), which is consistent with the findings of this paper.

The political implications of missing women are likely highly consequential. There is a striking gender gap in political participation in India (Prillaman 2019), wherein women vote at lower rates at all levels of elections and participate in local democracy at alarmingly low rates. Both large-scale cross-country (Iversen and Rosenbluth 2006) and experimental (Chattopadhyay and Duflo 2004) studies document significant and highly consequential differences between men and women's policy preferences. Beaman et al. (2009) show that exposure to women in leadership positions increases women's political participation and weakens stereotypes about gender roles in the public and private spheres. A lopsided gender ratio in mahalwari villages likely exacerbates the problem of lack of role models, thereby resulting in low levels of political participation, which in turn exacerbates the non-representativeness of women's preferences in public policy (Pitkin 1967; Burrell 1996). Traditional conservative norms about the role of women in society are likely to persist in environments where norms regarding property rights favour male children, thereby lowering political participation (Gottlieb and Robinson 2016). Worse labor market opportunities for women is also linked with their espousing more conservative values, thereby perpetuating gender inequality (Blaydes and Linzer 2008). The findings of this paper of institutional and demographic persistence suggest that property-rights institutions put these two otherwise similar regions on different trajectories with regard to land rights and norms and practices regarding the role and position of women in society (sex-ratios, bargaining power, education and labor force participation, and political participation).

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 led to worse sex ratios, female literacy, and labor force participation. The rest of the paper is organized 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 and 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 north Indian state under consideration in this paper, Uttar Pradesh is home to approximately 204 million people, which would make it the 5th largest country in the world by population (IWS) at the time of measurement (2001). 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 (Metcalfe and 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). BI document that, for all of India, areas conquered at later dates were less likely to have a landlord system [4], 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.

---

4. 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.

The institutional variation in Awadh was an unforeseen consequence of the EIC's involvement in palace intrigue. After 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, he signed a treaty that resulted in the cession of "half the territories"[5], the loose selection of which was made on grounds of defending the EIC's territory against the frontier in the Northwest from the Afghans and Sikhs [6] (Fisher 1993). The annexation is illustrated in the juxtaposed maps of the region in Figure 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).

---

5. 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)

6. "The Nabob ceded to the Company the territory of Rohilcund, the Dooab, and Gurruckpoor, the two 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)
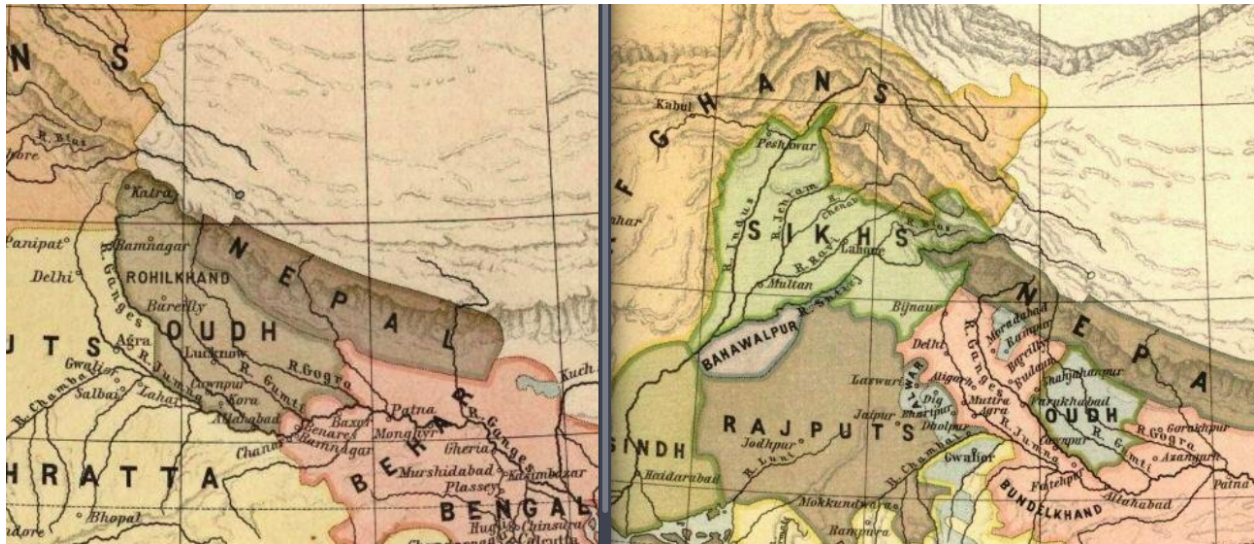
Figure 1: Inset of Region from 'India in 1795'(L) and 'India in 1805' (R) maps showing the partial annexation of Awadh in 1801, Joppen (1914)

The British annexed the remainder of the principality of Awadh in 1856 under the auspices of 'mismanagement' (Fisher 1998; Barry 2019). This is thought to have been one of the key factors precipitating the sepoy mutiny in 1857 (also considered the first war of Independence in India)(Kochanek and Hardgrave 2007). 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 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 exorbitant demands.

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)[7]. The border between the districts of Awadh

---

7. A Labelled map of the Awadh Districts is in the appendix (fig A1)

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 other land-tenure discontinuities in India, which are typically along state borders, which amount to a severe compound treatment problem since states often have very different histories and many policies change at the state border. For example, 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). We therefore restrict ourselves to estimating the treatment effect at the *within*-Uttar Pradesh boundary of late-period Awadh.
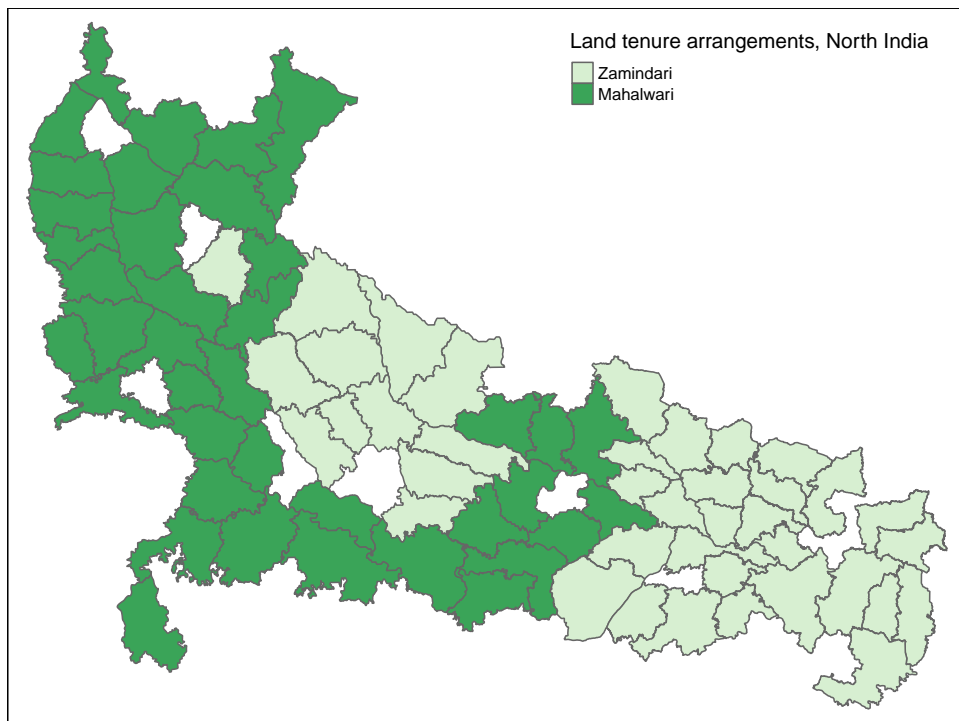
Figure 2: Colonial Era land tenure arrangements in North India, based on classifications from Banerjee and Iyer (2005), who in turn adapted it from Baden-Powell (1892)

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 result in lower values of another input (gender equity), and potentially even lower overall social welfare [8]. 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. In fact, in most places in north India, women are essentially cut off from their families following marriage, which means that inheritance passed on to them is effectively a transfer to a different household. This creates a very strong incentives for son preference (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 to increase its future utility. This implies that heritable property results

---

8. Define A social welfare function: $W = f(x_1, x_2)$, 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 alone 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 $S < \tilde{S} < S^*$.

in son-preference in a completely rational framework. A simple 2-period model in appendix C 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 to men 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 labor force participation under canonical household bargaining models. Worse labor market and 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 (albeit 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 labor 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 labor 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 labor 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 dataset analysed in this paper s the 2001 village-level geocoded Indian census [9] (Infomap 2001) merged with 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 shape-files and restricting the data to within 2 decimal degrees[10] of the relevant borders yields an analysis sample of 119,223 villages across the states of Uttar Pradesh and Bihar [11]. The primary results of the paper are based on the sample within Uttar Pradesh, which includes 47,237 villages.

This paper's spatial regression discontinuity identification strategy relies on contiguous districts with different land-tenure systems holding other institutions (such as state governments) constant as far as possible. 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, we focus on a region of India that is unambiguously classified as one of the two systems. Our focus on Uttar Pradesh is motivated by the mix of land-tenure systems present in the state thanks to complexities arising from the initial (quasi-) independence of the principality of Awadh, and its subsequent annexation following the sepoy mutiny of 1857. To our knowledge, the within-UP boundary between Awadh and non-Awadh districts is the only such boundary in contemporary India.

Iversen, Palmer-Jones, and 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

---

9. The relevant state is Uttar Pradesh. Results that include the boundary with Bihar, the adjacent state, are reported in the appendix

10. which is approximately 200 km. This is a substantively chosen cutoff for plausible smoothness of covariates. The estimation strategy uses a data-driven bandwidth selection procedure, which looks at narrower windows (Calonico, Cattaneo, and Titiunik 2015)

11. I define a village as a geographical unit classified as 'village' by the census with a population ≥ 100

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 for analysis that would elucidate mechanisms are unavailable. The most notable is the absence of population counts disaggregated by age, which would allow us to test the differential neglect hypothesis and document the extent to which the imbalanced sex ratio is stable across the age distribution. 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. Future research on this question may be able to use geo-located household-level census datasets that are as yet unavailable.

4.1   Summary Statistics

Summary statistics for the relevant variables for the Awadh analysis sample are presented in table 1 [12]. The average sex-ratio in the region(1.11) is well above what one would naturally expect (1.05), and as illustrated in fig 3, the average is slightly higher in *Mahalwari* villages compared to *Zamindari* ones. This is especially noticeable in the bottom panel in fig 3, 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 [13]. There is no reason

---

12. Balance tables for the corresponding samples are reported in B.1. The unit of observation throughout the analysis is a village.

13. This is simply the exogeneity condition($\mathbb{E}(X\epsilon) = 0$) for OLS. Put differently, since we know there is
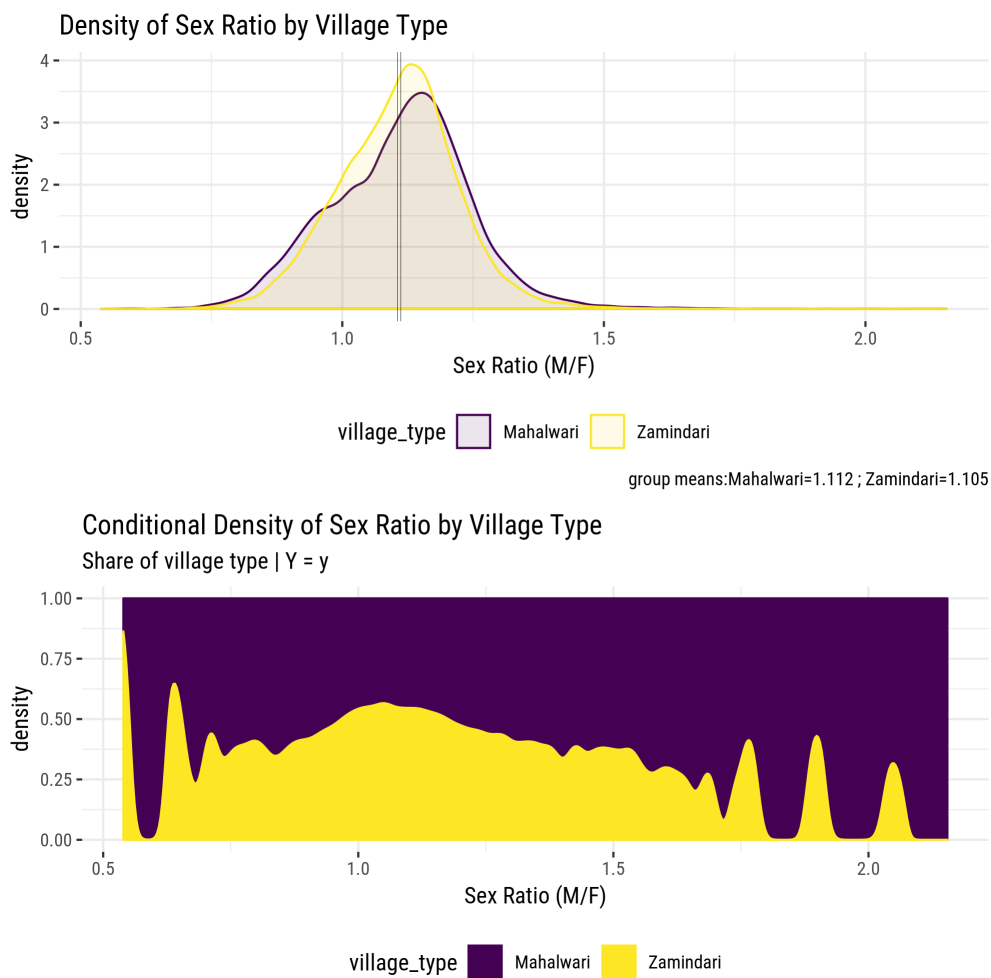
Figure 3: Density of Sex Ratio by Village Type : Unconditional (kernel density) in top panel, Conditional (i.e. share of Mahalwari and Zamindari villages for each value of the sex ratio on the X axis) in bottom panel. More extreme values of sex-ratio ($\leq .8, \geq 1.2$) are much more likely to be present in Mahalwari villages

Table 1: Descriptive Statistics: All Villages, Awadh

Statistic	N	Min	Max	Mean	St. Dev.
Non-Landlord Status	47,237	1	2	1.500	0.500
Household Size	47,237	1.570	140.000	6.400	1.070
Population	47,237	100	25,433	1,335.000	1,293.000
Share Scheduled-Caste	47,237	0.000	1.000	0.248	0.189
Literacy	47,237	0.000	1.000	0.409	0.122
Female Literacy	47,237	0.000	0.962	0.285	0.124
Sex Ratio	47,237	0.540	11.200	1.110	0.137
Labour Force Participation Rate	47,237	0.000	1.000	0.345	0.101
Female Labour Force Participation Rate	47,237	0.000	0.904	0.184	0.175
Share of Labourers: Marginal	47,237	0.000	1.000	0.263	0.209
Share of Labourers: Agricultural	47,237	0.000	1.000	0.120	0.130
Share of Labourers: Cultivators	47,237	0.000	0.841	0.166	0.080
Night-time Luminosity (DMSP 2001)	47,237	0.000	56.700	1.900	3.180

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 in ways 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.

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 [14], where the land-tenure system can plausibly be exogenous (since geography, climate are very similar in villages within 50 kilometres of each other on either side of the border. We test formally for this using rainfall

---

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

14. (highlighted in yellow in Figure 3)

and FAO crop suitability data and find no statistically detectable difference at the border). This is especially plausible given that the border between the districts that comprised of the boundary between Awadh ('Oudh' in figure 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 Mahalwari systems. We omit the boundary between the North-western Provinces and the Central Provinces (modern-day Uttar Pradesh and Madhya Pradesh - on the southwestern end of UP in the map below), in accordance with Iversen, Palmer-Jones, and Sen (2013)'s claim that CP's land tenure systems were highly heterogeneous and BI's findings were sensitive to how these were coded. All across-state boundaries are eliminated from the analysis, so the boundaries that separate treatment and control regions are all within the state of Uttar Pradesh, which alleviates concerns regarding major institutional variation at the cutoff.

We subset the analysis sample to within 1 decimal degree ($\approx$ 100 km) of the border between *Zamindari* and non-*Zamindari* districts[15]. The buffers are constructed using the shapefiles using the original WGS-84 EPSG-4326 projection. [16]. The bandwidth (distance from the border) is chosen by a data-driven selection procedure for the regression discontinuity design with Euclidian distance as the running variable. We report OLS regression coefficients for a very large number of bandwidths and specifications in appendix D.1.

---

15. These degree-km conversions are approximate and are correct for 23 degrees North of the equator, which is the region of the globe for North India

16. Since we are considering very small bandwidths, the choice of Euclidian, as opposed to Haversine/Great-circle, distance is inconsequential for estimated distances. This is because for small distances, the earth can be approximated by a flat plane, therefore justifying the choice of Euclidian distance. When larger distances are involved, however, the choice of projection matters a great deal.
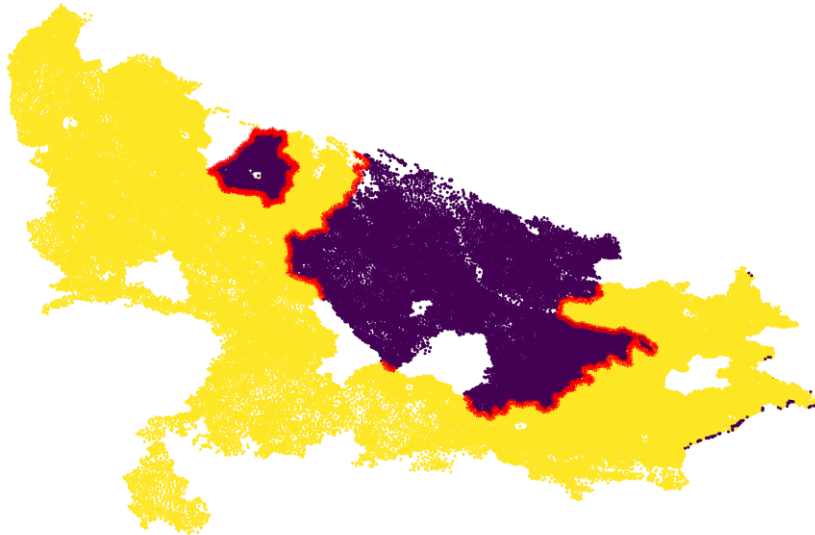
Boundary around Oudh

Figure 4: Identifying Boundaries around Awadh

5 Empirical Strategy

We describe the research design for the analysis below and report the results in the corresponding section in 6.

5.1 Spatial Regression Discontinuity Design

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 estimated using flexible local-linear regressions to approximate conditional expectation functions 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

19

is the difference between the CEFs at c. [17] The intuition behind this result is that if we believe that potential confounders for 'treated' and 'control' units are smooth functions at the cutoff c, we can treat them as a nuisance parameter and isolate the treatment effect by estimating the expectation functions conditional on the running variable flexibly on either side of the threshold.

To this end, we use the robust local-linear regression-based regression-discontinuity design estimator and confidence intervals proposed in Calonico, Cattaneo, and Titiunik (2014) (henceforth CCT 2014) [18].we 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 Mahalwari villages, and use this distance as a running variable with a cutoff at $\text{Distance} = 0$. 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 on a bandwidth h [19], ϕ_b are line-segment fixed effects, which ensure that villages on either side of the same border are compared [20] . A triangular kernel is used to weight the observations closest to the cutoff heavily and decreasing in either direction

---

17.
$$\hat{\tau}_{SRD} = \lim_{x\downarrow c}\mathbb{E}[Y_i|X_i = x] - \lim_{x\uparrow c}\mathbb{E}[Y_i|X_i = x]$$

18. estimated using the accompanying R package RDRobust (Calonico, Cattaneo, and Titiunik 2015)

19. estimated using local polynomial regressions selected by the Mean Squared Error (MSE) expansion of the sharp RD estimator as proposed in CCT (2014). 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)

20. This requires an ad-hoc partitioning of the border into a finite number of segments, since in theory a line of finite length can be subset into an infinite number of line segments (Keele and Titiunik 2015). We construct 20 and find that the substantive conclusions remain the same for a variety of choices for the line segment count

on the interval $[c-h, c+h]$. Using alternative kernels and estimating with and without bias-correction yields substantively and statistically similar results.

We also use two alternative methods of estimating spatial regression discontinuity designs: a 'local-randomization' approach where a linear specification is fit on either side of the window with progressively narrower bandwidth, and a parametric specification that includes separate cubic polynomials for longitude and latitude (first popularised by Dell (2010)) and report them in appendix sections D.1 and D.3 respectively. These also yield substantively similar conclusions with sometimes wider standard errors.

6   Results

6.1   Distance Regression Discontinuity Results

Tables 2, 3, and 4 report local-linear and robust regression discontinuity estimates, robust standard errors, and the number of observations used in the estimation on either side of the cutoff for sex ratio, female literacy, and female labor force participation respectively. The SRD coefficient is calculated as the difference between the local polynomial regression functions (linear in columns 1 and 3, and quadratic in 2 and 4), with line-segment fixed effects as controls (in columns 3 and 4). The CCT bandwidths are different on the two sides of the threshold and are reported in the table alongside the corresponding number of observations. In our preferred specifications (which include linear CEF and line-segment fixed effects - column 3 in all three tables below), Mahalwari villages have 0.0151 more men per woman (in more interpretable terms: 1.51 additional men for every 100 women relative to Zamindari regions, which is 1.5% higher than the control mean, 1.1). The sign and magnitudes of the effects are comparable to those found in the OLS estimation (reported in D.1) and are more precisely estimated. The effects on female literacy and labor force participation are marginally significant upon the inclusion of segment fixed effects (t-stat $\approx 1.8 - 1.9$ in column 3 in tables 3 and 4). While results reported in tables 2, 3, and

are estimated using local-linear regressions, results for quadratic CEFs are visually and quantitatively similar and presented in figure .

	1	2	3	4
Local-Linear	-	-	-	-
Coef	0.0291	0.0265	0.0151	0.0155
SE	0.004	0.0043	0.0032	0.0036
t-stat	7.3544	6.2238	4.7501	4.3519
CI	(0.0214,0.0369)	(0.0182,0.0349)	(0.0089,0.0214)	(0.0085,0.0225)
Robust	-	-	-	-
Coef	0.0302	0.0273	0.0156	0.0167
SE	0.0045	0.0046	0.0036	0.0038
t-stat	6.724	5.8823	4.2735	4.373
CI	(0.0214,0.039)	(0.0182,0.0363)	(0.0084,0.0227)	(0.0092,0.0241)
	—	—	—	—
Polynomial Order	1	2	1	2
Bandwidth (Degrees)	(-0.1859,0.1859)	(-0.3525,0.3525)	(-0.1766,0.1766)	(-0.297,0.297)
# Obs	(8080,11873)	(12258,16546)	(7795,11426)	(11013,15429)
Segment FEs			Y	Y

Table 2: Sex-Ratio RD estimates - Awadh

	1	2	3	4
Local-Linear	-	-	-	-
Coef	-0.0174	-0.0151	-0.0069	-0.0054
SE	0.0037	0.0038	0.0036	0.0041
t-stat	-4.7071	-3.9594	-1.9337	-1.3403
CI	(-0.0247,-0.0102)	(-0.0226,-0.0076)	(-0.014,0.0001)	(-0.0134,0.0025)
Robust	-	-	-	-
Coef	-0.0159	-0.0146	-0.0055	-0.0043
SE	0.0041	0.0042	0.0039	0.0043
t-stat	-3.8838	-3.4897	-1.4155	-0.9935
CI	(-0.0239,-0.0079)	(-0.0229,-0.0064)	(-0.0132,0.0021)	(-0.0128,0.0042)
	—	—	—	—
Polynomial Order	1	2	1	2
Bandwidth (Degrees)	(-0.1447,0.1447)	(-0.3109,0.3109)	(-0.1299,0.1299)	(-0.2189,0.2189)
# Obs	(6761,9887)	(11320,15727)	(6228,9061)	(9035,13212)
Segment FEs			Y	Y

Table 3: Female Literacy RD estimates - Awadh

6.2 Robustness Checks

6.2.1 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)

Sex Ratio (M/F)
Awadh

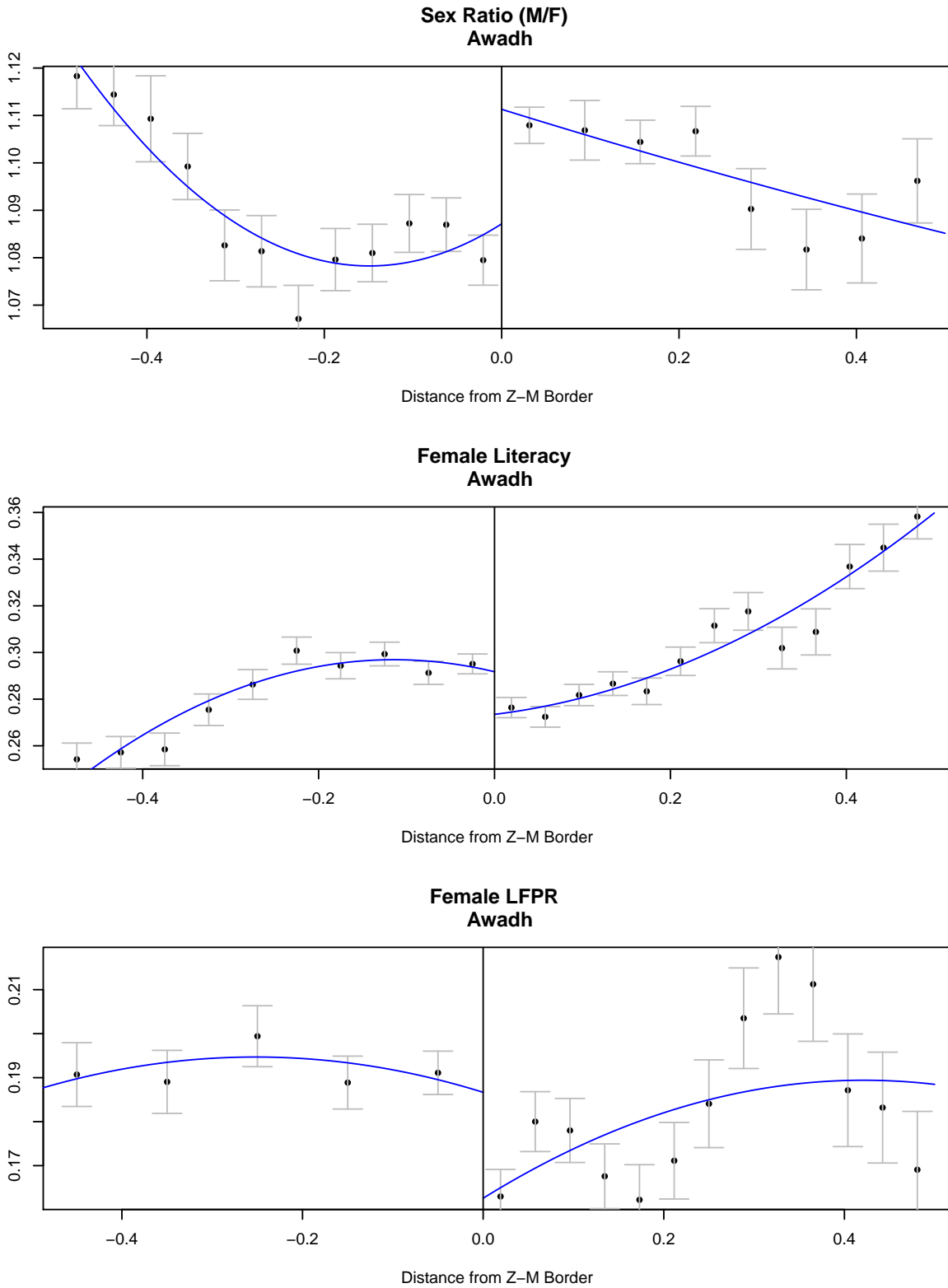**Female Literacy**
Awadh

Female LFPR
Awadh

Figure 5: 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

	1	2	3	4
Local-Linear	-	-	-	-
Coef	-0.0242	-0.0212	-0.0078	-0.0076
SE	0.0053	0.0055	0.0042	0.0049
t-stat	-4.5933	-3.8224	-1.8587	-1.5337
CI	(-0.0345,-0.0139)	(-0.032,-0.0103)	(-0.016,0.0004)	(-0.0172,0.0021)
Robust	-	-	-	-
Coef	-0.0259	-0.0219	-0.0076	-0.0081
SE	0.006	0.0061	0.0049	0.0054
t-stat	-4.3318	-3.5946	-1.5512	-1.4956
CI	(-0.0377,-0.0142)	(-0.0339,-0.01)	(-0.0171,0.002)	(-0.0187,0.0025)
	——	——	——	——
Polynomial Order	1	2	1	2
Bandwidth (Degrees)	(-0.1617,0.1617)	(-0.3295,0.3295)	(-0.2126,0.2126)	(-0.3249,0.3249)
# Obs	(7330,10741)	(11746,16128)	(8861,13011)	(11652,16041)
Segment FEs			Y	Y

Table 4: Female LFPR RD estimates - Awadh

because most plausible development outcomes that may be in the error term are post-treatment – i.e. they are 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. Furthermore, given the research design, potential confounders can be expected to be smooth at the discontinuity, since the estimation sample is a narrow band along the border.

Despite this, we use the limited number of controls available at the village level: 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) to test the robustness of RD results. We report results for specifications with controls in table 5. The first column includes all villages in the sample and therefore the result is likely biased (because the argument about balance in controls around the narrow band around the border no longer applies) because of omitted variables bias. Columns 2-4, however, rely on similar specifications to those in **??** and are reasonably credible (because they subset to a narrow band around the border). The estimated coefficients for sex ratio are significantly larger upon the inclusion of controls,

which suggests that the covariance between Mahalwari and the controls is nonzero.

Table 5: 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

6.2.2 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. To this end, we report the density of the observations by Euclidian distance to the border in fig 6. Sorting is 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. Because the unit of observation in the analysis is a village, we face the additional complication that there are mechanically more villages on the mahalwari side because they tend to be smaller in area than zamindari villages (because more smallholders own their land), which shows up in the mass of villages close to the border on the zamindari side in 6. Because households aren't entitled to land on the other side of the border, there is little reason to suspect manipulation of the running variable (McCrary 2008). Nevertheless, the heuristic check of the 'donut test', to test whether agents sort across to the favourable side when close to the cutoff, leaving a large chasm in the density plot on one of the two sides of the cutoff, appears to pass in fig 6.

6.2.3 labor 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 labor market conditions for women vary discontinuously across the border. Neither Male labor force participation (which is a useful placebo because potential labor market imperfections should affect both genders equally) nor share of village engaged in agriculture (figs 7 8) appears to vary discontinuously across the border, which suggests that labor market opportunities are not responsible for the observed effects on the sex ratio, female literacy, and female labor force participation.

7 Conclusion

This paper finds that villages property rights were granted to cultivators have systematically worse outcomes for women on a variety of measures: sheer numbers (the sex ratio), literacy, and labor force participation. The observed effects are consistent with a mechanism whereby the presence of heritable land in the household exacerbates cultur-
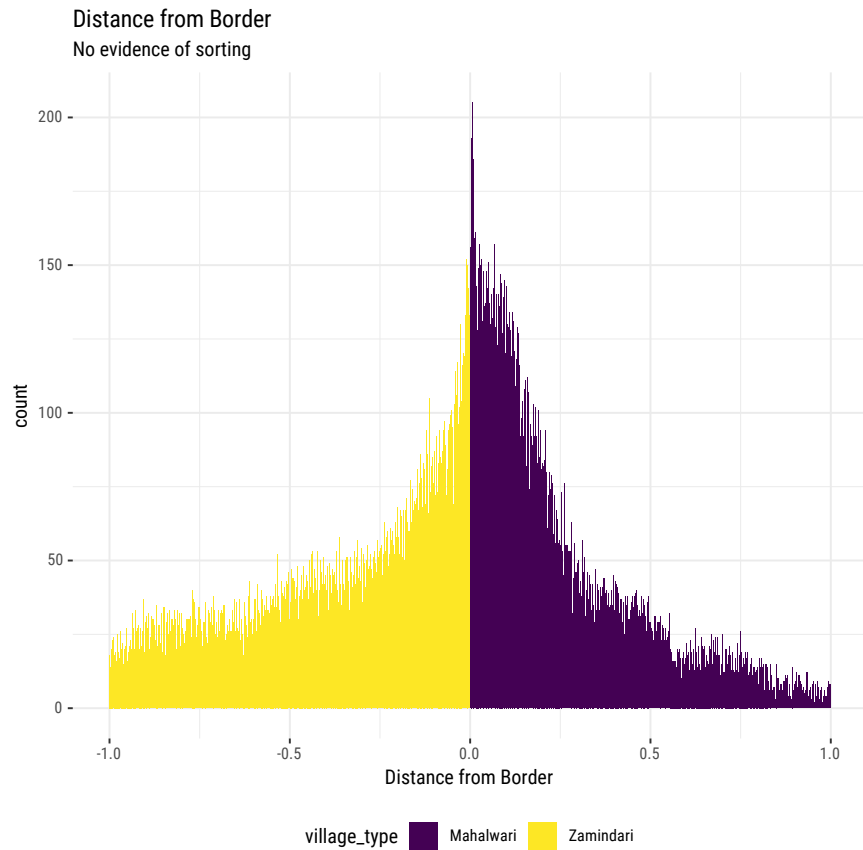
Figure 6: Histogram of Distance to Border for all villages. Sorting is close-to-impossible across the threshold. The higher number of villages on the Mahalwari side is simply the result of there being smaller villages on that side of the border
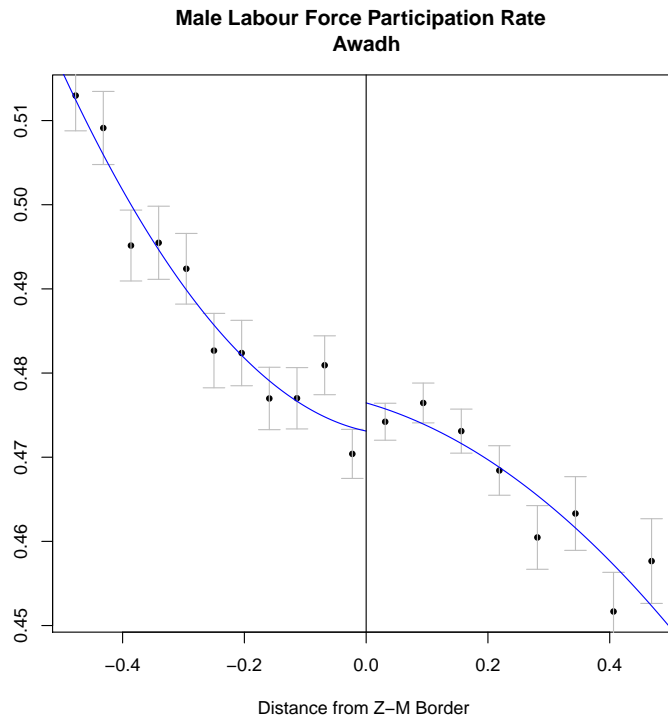
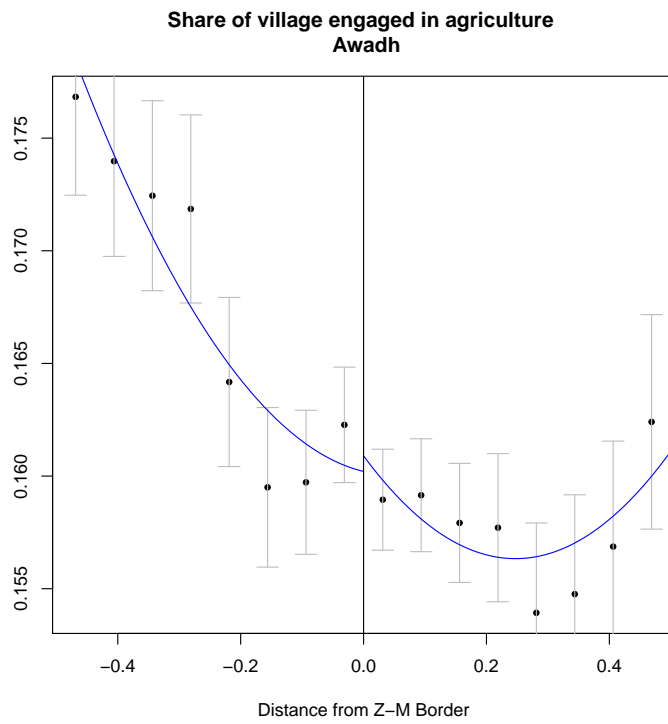**Male Labour Force Participation Rate
Awadh**

Distance from Z–M Border

Figure 7: Male LFPR

**Share of village engaged in agriculture
Awadh**

Distance from Z–M Border

Figure 8: Agricultural Share of labor

ally widespread son preference, and increased bargaining power for men results in worse female labor force and 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 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 labor 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 we 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 the effects of institutions are mediated by existing socio-economic relations. 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. Findings from economic history such as this one are useful because they dissuade silver-bullet thinking, especially one that has been championed vociferously as property rights.

Bibliography

ACEMOGLU, Daron, Simon JOHNSON, and James A. ROBINSON. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American economic review* 91 (5): 1369–1401.

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): 469–530.

ALMOND, Douglas, Hongbin LI, and Shuang ZHANG. 2019. "Land Reform and Sex Selection in China." *The journal of political economy,* 000–000. https://doi.org/10.1086/701030.

ANDERSON, Siwan, and Debraj RAY. 2010. "Missing Women: Age and Disease." *The Review of Economic Studies* 77 (4): 1262–1300.

BADEN-POWELL, Baden Henry. 1892. *The land-systems of British India.* Vol. 2. Clarendon Press.

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): 1190–1213.

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

BARRY, Ellen. 2019. "The Jungle Prince of Delhi." *The New York Times* (November). https://www.nytimes.com/2019/11/22/world/asia/the-jungle-prince-of-delhi.html.

BEAMAN, Lori, Raghabendra CHATTOPADHYAY, Esther DUFLO, Rohini PANDE, and Petia TOPALOVA. 2009. "Powerful women: does exposure reduce bias?" *The Quarterly journal of economics* 124 (4): 1497–1540.

BHALOTRA, Sonia, Rachel BRULÉ, and Sanchari ROY. 2018. "Women's inheritance rights reform and the preference for sons in India." *Journal of development economics* (August). http://www.sciencedirect.com/science/article/pii/S0304387818300294.

BHALOTRA, Sonia, Abhishek CHAKRAVARTY, Dilip MOOKHERJEE, and Francisco J PINO. 2019. "Property Rights and Gender Bias: Evidence from Land Reform in West Bengal." *American economic journal. Applied economics* 11, no. 2 (April): 205–237. https://www.aeaweb.org/articles?id=10.1257/app.20160262.

BLAYDES, Lisa, and Drew A LINZER. 2008. "The political economy of women's support for fundamentalist Islam." *World Politics* 60 (4): 576–609.

BRULÉ, Rachel. 2018. "Reform, Representation & Resistance: The Politics of Property Rights' Enforcement." *Journal of Politics.*

BURRELL, Barbara C. 1996. *A Woman's Place is in the House: Campaigning for Congress in the Feminist Era.* University of Michigan Press.

CALONICO, Sebastian, Matias D. CATTANEO, and Rocio TITIUNIK. 2014. "Robust Nonparametric Confidence Intervals for Regression-discontinuity Designs." *Econometrica* 82 (6): 2295–2326.

———. 2015. "Rdrobust: An r Package for Robust Nonparametric Inference in Regression-Discontinuity Designs." *R Journal* 7 (1): 38–51.

CALVI, Rossella. 2020. "Why are older women missing in India? The age profile of bargaining power and poverty." *Journal of Political Economy.*

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): 989–1012.

CHATTOPADHYAY, Raghabendra, and Esther DUFLO. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72 (5): 1409–1443.

CONLEY, Timothy G. 1999. "GMM estimation with cross sectional dependence." *Journal of econometrics* 92 (1): 1–45.

DELL, Melissa. 2010. "The Persistent Effects of Peru's Mining Mita." *Econometrica* 78 (6): 1863–1903.

DENYER, Simon, and Annie GOWEN. 2018. ""There's Too Many Men": What Happens When Women Are Outnumbered on a Massive Scale" [in en]. *Washington Post.*

ECONOMIST, The. 2018. "Why India Needs Women to Work." *The Economist* (July). 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): 104–117.

GOTTLIEB, Jessica, and Amanda Lea ROBINSON. 2016. "The effects of matrilineality on gender differences in political behavior across Africa." *British Journal of Political Science.*

GROSJEAN, Pauline, and Rose KHATTAR. 2019. "It's raining men! Hallelujah? The long-run consequences of male-biased sex ratios." *The Review of Economic Studies* 86 (2): 723–754.

HABER, Stephen, Noel MAURER, and Armando RAZO. 2003. *The politics of property rights: political instability, credible commitments, and economic growth in Mexico, 1876-1929.* Cambridge University Press.

HAHN, Jinyong, Petra TODD, and Wilbert VAN DER KLAAUW. 2001. "Identification and Estimation of Treatment Effects with a Regression-discontinuity Design." *Econometrica* 69 (1): 201–209.

HORNBECK, Richard. 2010. "Barbed Wire: Property Rights and Agricultural Development." *The Quarterly Journal of Economics* 125 (2): 767–810.

HUNTER, John D. 2007. "Matplotlib: A 2D Graphics Environment." *Computing in science & engineering* 9 (3): 90–95.

IMBENS, Guido W., and Thomas LEMIEUX. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of econometrics* 142 (2): 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. 2001. *Uttar Pradesh East, India: Village Socio-Demographic and Economic Census Data, 2001* [in en]. https://purl.stanford.edu/jj024vd6848.

IVERSEN, Torben, and Frances ROSENBLUTH. 2006. "The Political Economy of Gender: Explaining Cross-National Variation in the Gender Division of Labor and the Gender Voting Gap." *American journal of political science* 50, no. 1 (January): 1–19. http://doi.wiley.com/10.1111/j.1540-5907.2006.00166.x.

IVERSEN, Vegard, Richard PALMER-JONES, and Kunal SEN. 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, no. 12 (December): 1631–1646. ISSN: 0022-0388. https://doi.org/10.1080/00220388.2013.807502.

IWS. *Ten Countries with the Highest Population in the World.* https://www.internetworldstats.com/stats8.h

JAYACHANDRAN, Seema. 2015. "The Roots of Gender Inequality in Developing Countries." *Annual review of economics* 7, no. 1 (August): 63–88. https://doi.org/10.1146/annurev-economics-080614-115404.

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): 127–155.

KOCHANEK, Stanley A, and Robert L HARDGRAVE. 2007. *India: Government and politics in a developing nation.* Cengage Learning.

KULKARNI, Parashar. 2017. "The British Academy Brian Barry Prize Essay: Can Religious Norms Undermine Effective Property Rights?: Evidence from Inheritance Rights of Widows in Colonial India." *British Journal of Political Science* 47 (3): 479499. https://doi.org/10.1017/S0007123416000363.

Lipsey, Richard G., and Kelvin Lancaster. 1956. "The General Theory of Second Best." *The review of economic studies* 24 (1): 11–32.

Lundberg, Shelly, and Robert A. Pollak. 1993. "Separate Spheres Bargaining and the Marriage Market." *Journal of political Economy* 101 (6): 988–1010.

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of econometrics* 142 (2): 698–714.

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

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

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

Pitkin, Hanna F. 1967. *The concept of representation.* Vol. 75. Univ of California Press.

Prillaman, Soledad Artiz. 2019. "Strength in Numbers: How Women's Groups Close India's Political Gender Gap." *Working Paper.*

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): 1251–1285.

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

———. 1992. "Missing Women." *BMJ: British Medical Journal* 304 (6827): 587.

Sokoloff, Kenneth L., and Stanley L. Engerman. 2000. "Institutions, Factor Endowments, and Paths of Development in the New World." *Journal of Economic perspectives* 14 (3): 217–232.

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

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): 22–30.

Wang, Shing-Yi. 2014. "Property rights and intra-household bargaining." *Journal of development economics* 107 (March): 192–201. http://www.sciencedirect.com/science/article/pii/S0304387813001752.

Wickham, Hadley. 2010. "Ggplot2: Elegant Graphics for Data Analysis." *J Stat Softw* 35 (1): 65–88.

Wickham, Hadley, Romain Francois, Lionel Henry, and K. Müller. 2015. "Dplyr: A Grammar of Data Manipulation." *R package version 0.4* 3.

Part

Appendix

Table of Contents

A Labelled Map - Awadh

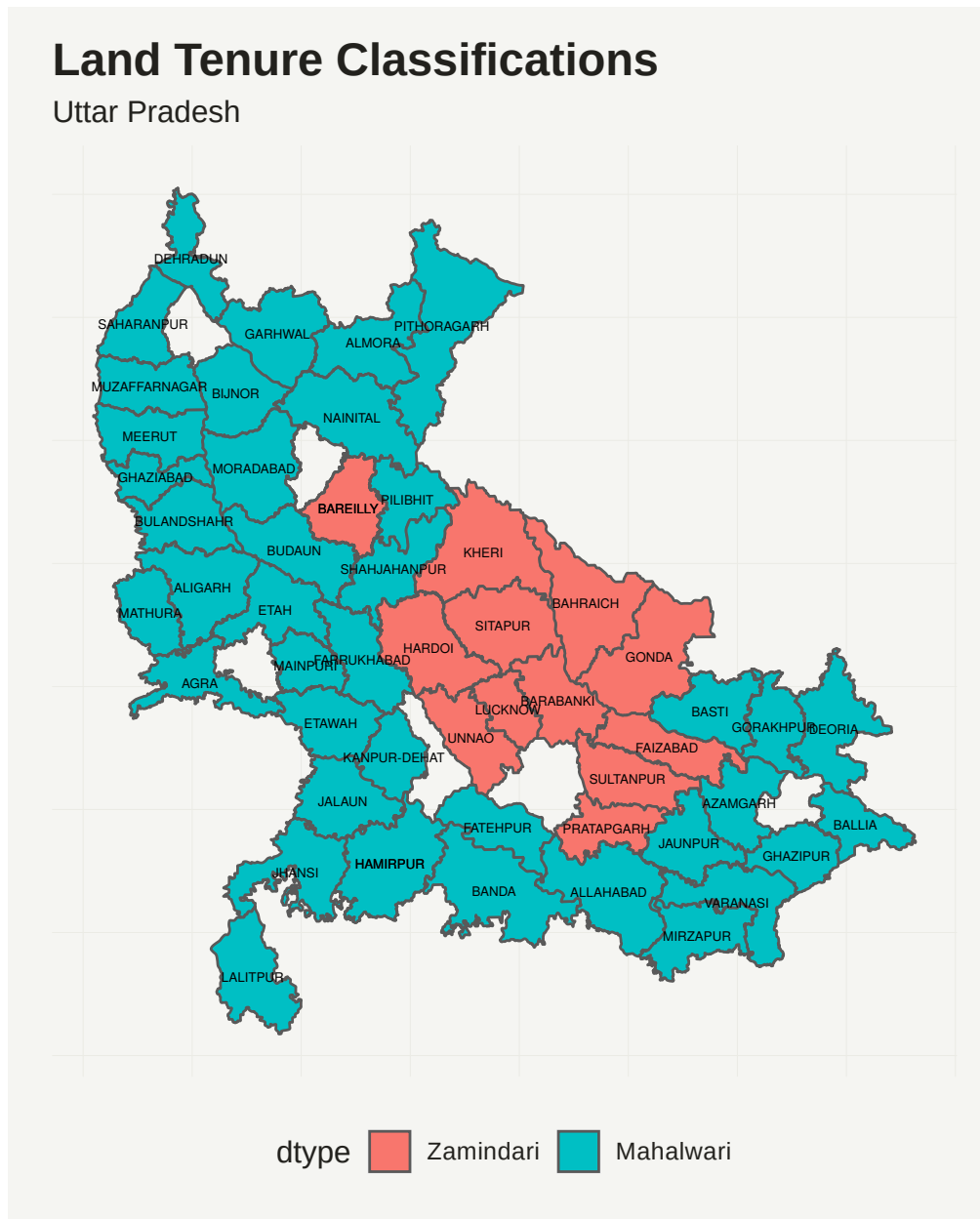## Land Tenure Classifications
Uttar Pradesh

Figure A1: Labelled Map of UP Districts with Land Tenure Classifications

B Summary Statistics and Balance Tables

B.1 Balance Tables

Here,we 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 A1: 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 A2: 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 A3: 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

C 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 [21]. If the household has no property rights, θ is zero.

The household's maximisation problem is:

$$\max_{c_1, c_2, s} U(c_1, c_2) \text{ s.t.} \tag{1}$$

$$c_1 + s = y_1 \tag{2}$$

$$c_2 = \theta s + y_2 \tag{3}$$

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:

$$\max_{c_1, c_2, s} U(c_1, c_2) \text{ s.t.} \tag{4}$$

$$c_1 + \frac{c_2}{\theta} = y_1 + \frac{y_2}{\theta} \tag{5}$$

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:

---

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

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

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.

C.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

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

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:

$$FOC(c_1) \rightarrow u'(c_1) = \lambda$$

$$FOC(c_2) \rightarrow u'(c_2) = \frac{\lambda}{(0.5 + z)\theta}$$

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

$$FOC(\mu) \rightarrow z = 0.1$$

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

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).

D Additional Estimates

D.1 Least Squares on Boundary Samples

As an interpretable benchmark case, we first estimate ordinary least squares regressions on a sub-sample of villages close to the border between historically Zamindari and Raiyatwari regions (the 'control' and 'treatment' in our sample respectively) for which confounders (geography, other demographic characteristics etc.) can be plausibly believed to be similar on both sides of the border. We run regressions of the form:

$$Y_{ijb} = \alpha + \beta \text{Mahalwari}_j + \gamma \text{Mahalwari}_j \times s_{ij}^p + \phi_b + \epsilon_i \tag{6}$$

$$\text{s.t. } |s_{ij}| \leq k \quad \forall k = \{0.5, 0.4995, \ldots, 0.0001\} \tag{7}$$

where Y_{ij} is an outcome (sex ratio [22], female labor force participation [23], female literacy [24]) for village i in district j, Mahalwari$_j$ is a dummy variable for whether district j is classified as a Mahalwari district (i.e. a non-landlord system), 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 equal-length 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 pth order of the spline; 0 means no spline. Heteroskedasticity-Robust standard errors are presented throughout [25], and results with Conley standard errors are presented in the appendix [sec D.2] and are substantively similar.

The following regressions are estimated using the framework outlined in D.1 close to the boundaries of the Awadh principality[26], which compares villages within the state of Uttar Pradesh, and robust (HC2) standard errors are reported. 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 transportability validity of the main estimates and are presented in the appendix D.6. The direction of the effect on sex-ratio, female literacy, and female labor force participation is identical, and the magnitudes are slightly smaller but statistically significant for the larger sample.

To rule out the possibility that arbitrary choices of buffer are responsible for the observed effect,we 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 A2). Each point in these graphs is the estimated $\hat{\beta}_j$ for Mahalwari villages estimated on the

---

22. defined as $\frac{\text{Male Population}_i}{\text{Female Population}_i}$ for village i. See appendix for detailed variable list

23. defined as $\frac{\text{Female Workers}_i}{\text{Female Population}_i}$ for village i

24. defined as $\frac{\text{Female Literates}_i}{\text{Female Population}_i}$ for village i

25. the analysis sample has an insufficient number of districts for clustering at the district level to be viable, even though the treatment is assigned at the district level

26. (highlighted in red in fig 4)

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. Figures with linear and quadratic spline included in the D.5.

Sex-Ratio
Coefficient for non-landlord

Awadh sample, HC2 Robust SEs

Female Literacy
Coefficient for non-landlord

Awadh sample, HC2 Robust SEs

Female LFPR
Coefficient for non-landlord
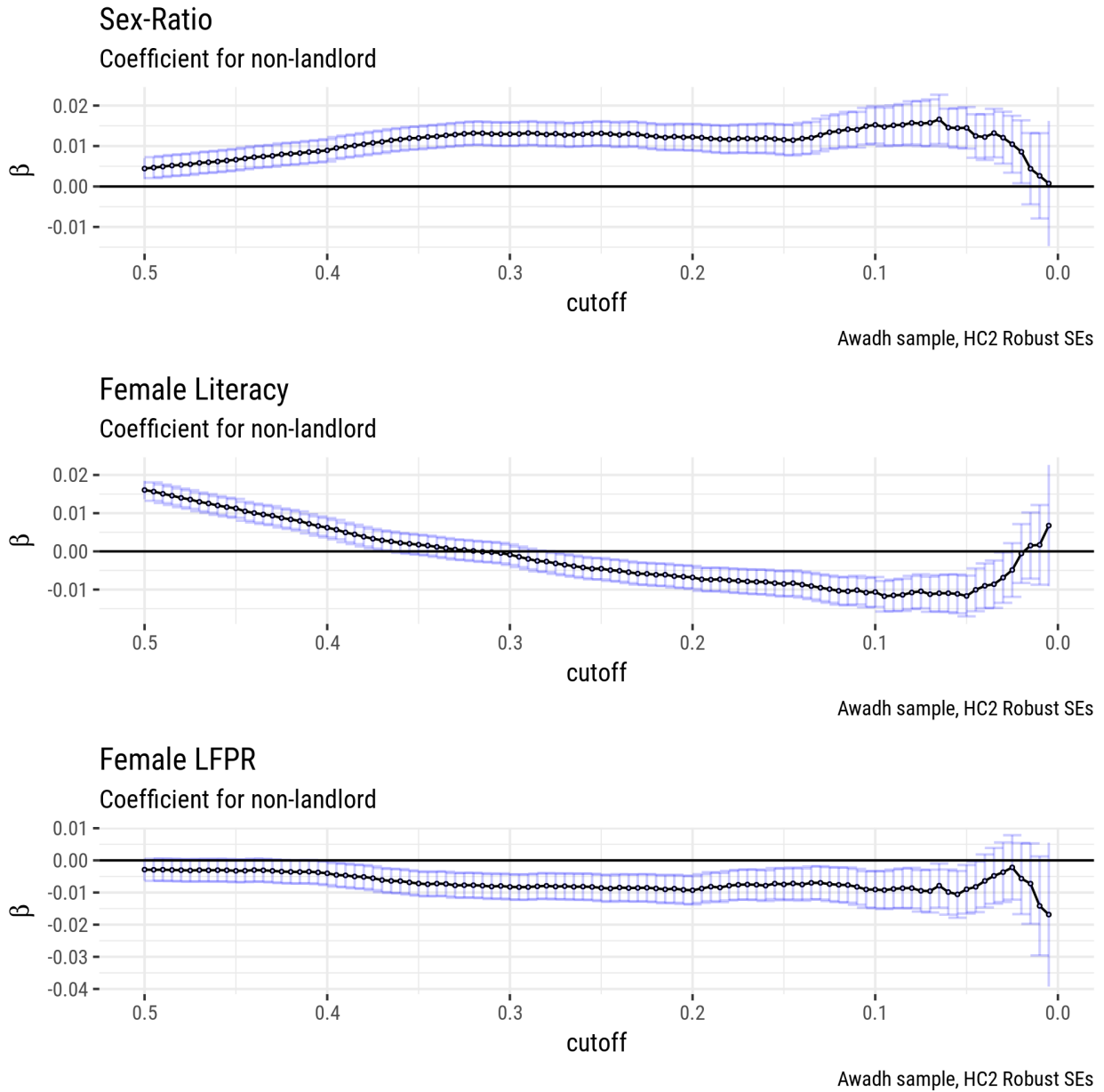
Awadh sample, HC2 Robust SEs

Figure A2: Mahalwari 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 reported in fig D.5

D.2 Conley Standard Errors

This section reports regression estimates with spatially autocorrelated error terms known as Conley standard errors (CONLEY 1999). We use fairly conservative bandwidths that allow for standard errors to be correlated over 200, 100, and 50 km respectively for the 3 estimation samples.

Table A4: Sex Ratio

	(10 KM)	(5 KM)	(1 KM)
Non-Landord	0.033***	0.035***	0.022***
	(0.004)	(0.003)	(0.007)
Constant	1.089***	1.090***	1.102***
	(0.060)	(0.057)	(0.043)
R^2	0.009	0.016	0.006
Conley Bandwidth	200 km	100 km	50 km

Table A5: Female Literacy

	(10 KM)	(5 KM)	(1 KM)
Non-Landord	-0.023	-0.022	-0.012*
	(0.017)	(0.014)	(0.007)
Constant	0.290***	0.289***	0.274***
	(0.038)	(0.036)	(0.028)
R^2	0.009	0.008	0.002
Conley Bandwidth	200 km	100 km	50 km

Table A6: Female LFPR

	(10 KM)	(5 KM)	(1 KM)
Non-Landord	-0.027***	-0.030***	-0.035***
	(0.010)	(0.011)	(0.007)
Constant	0.179***	0.174***	0.177***
	(0.050)	(0.046)	(0.036)
R^2	0.006	0.008	0.010
Conley Bandwidth	200 km	100 km	50 km

D.3   Parametric Spatial Regression Discontinuity

We estimate 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 we 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. The functional form f is a cubic polynomial for latitude and longitude as well as an interaction term between the two.

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

D.3.1   Parametric RD Coefficient Stability for Bandwidth choices

Using the methodology outlined in section D.1, we estimate 100 different regressions with the parametric estimating equation in section D.3 to rule out the possibility that particular choices of bandwidth yield the observed effect. We report results in A3 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 labor force participation rate is small and statistically insignificant from this estimation strategy.

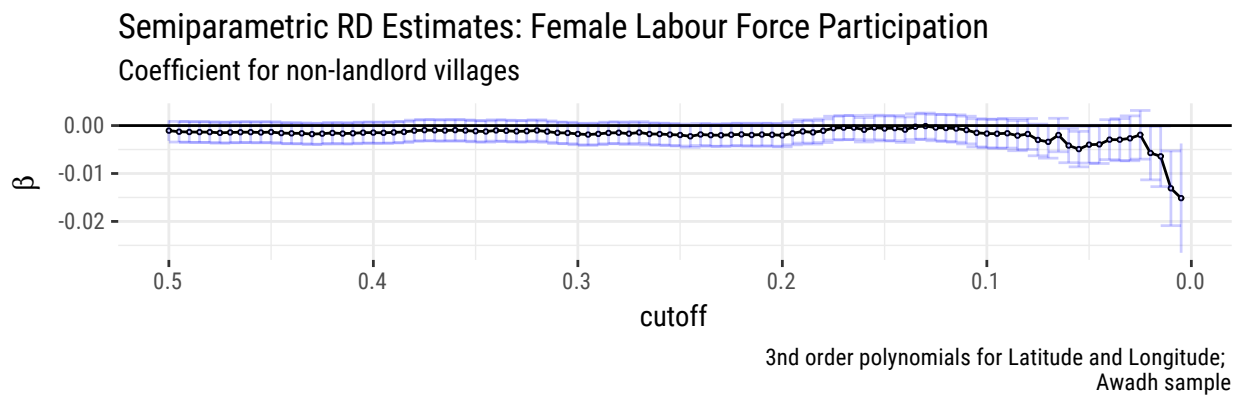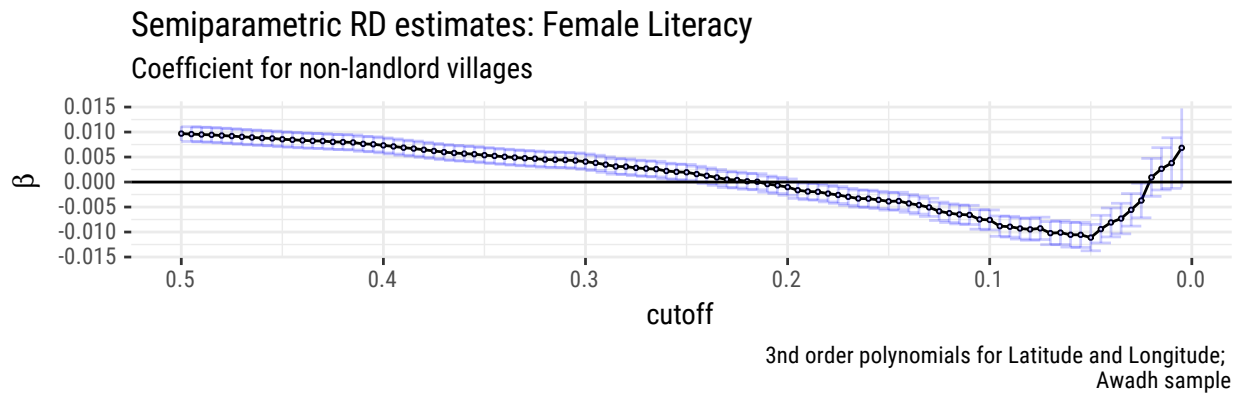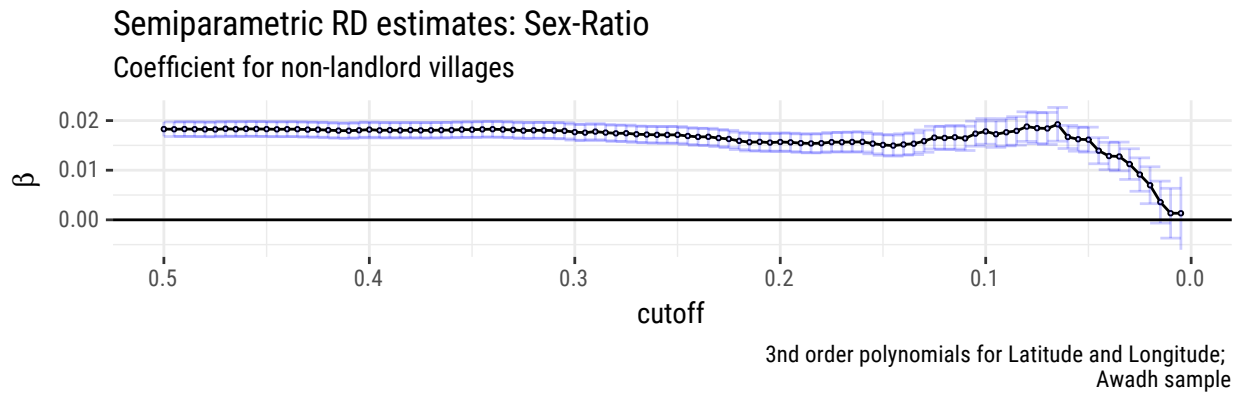## D.4   Linear and Quadratic Splines for OLS

Semiparametric RD estimates: Sex-Ratio

Coefficient for non-landlord villages

3nd order polynomials for Latitude and Longitude;
Awadh sample

Semiparametric RD estimates: Female Literacy

Coefficient for non-landlord villages

3nd order polynomials for Latitude and Longitude;
Awadh sample

Semiparametric RD Estimates: Female Labour Force Participation

Coefficient for non-landlord villages

3nd order polynomials for Latitude and Longitude;
Awadh sample

Figure A3: Semi-parametric RD Mahalwari Coefficients: Awadh

D.5 Coefficient Plots with linear and quadratic splines

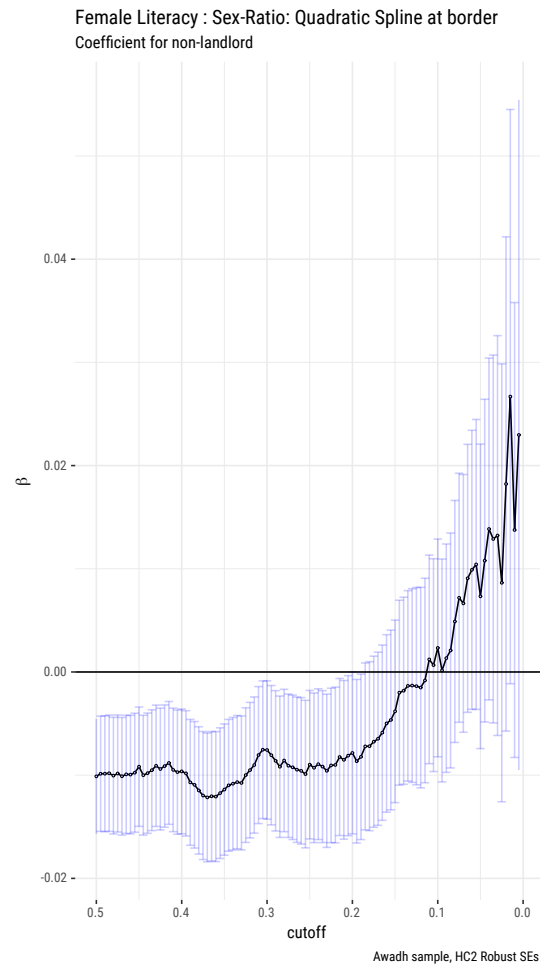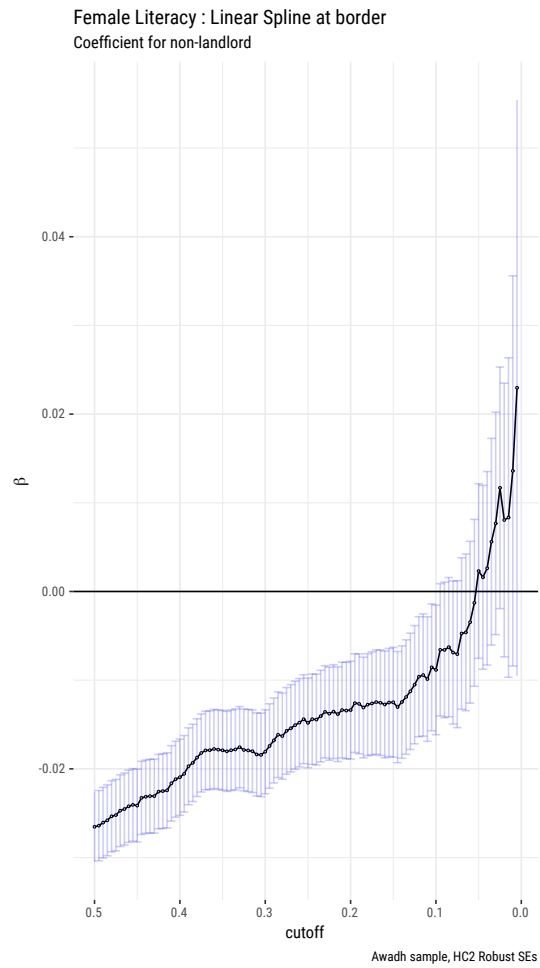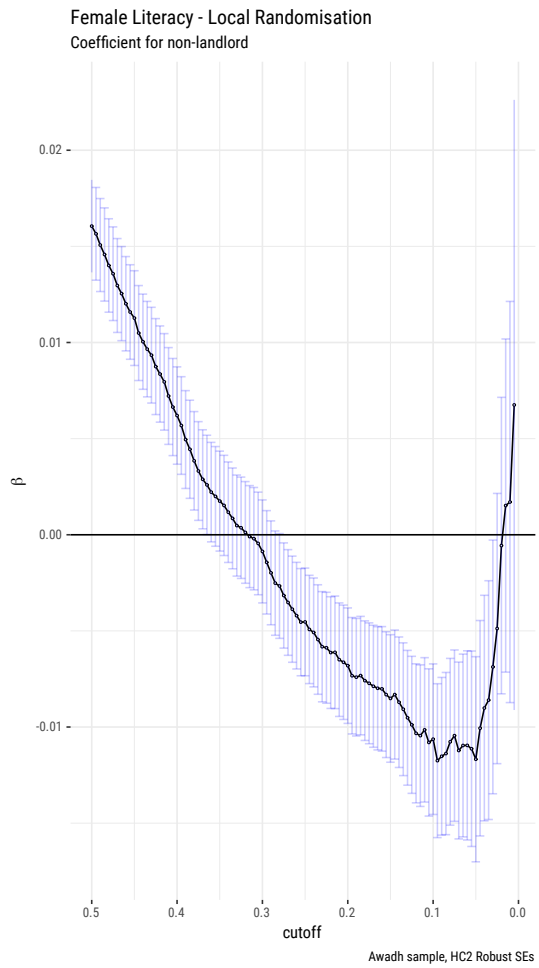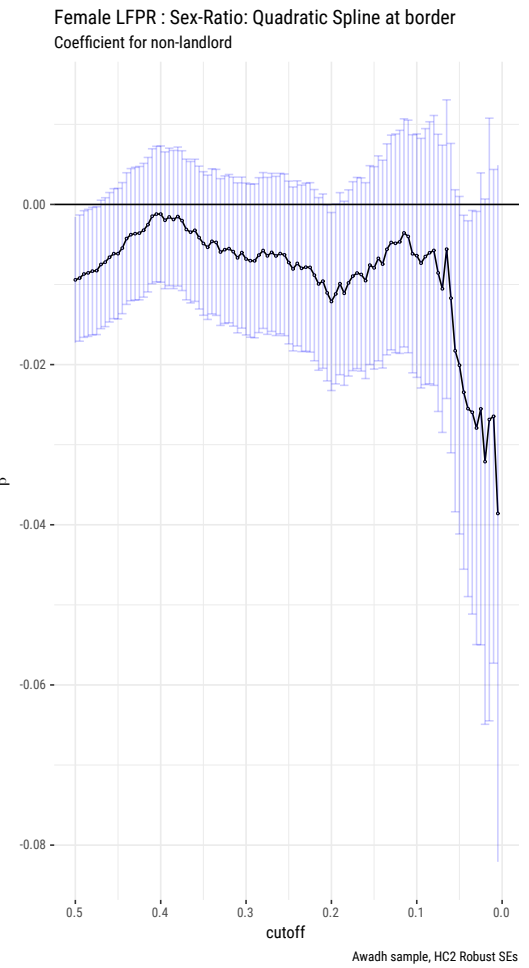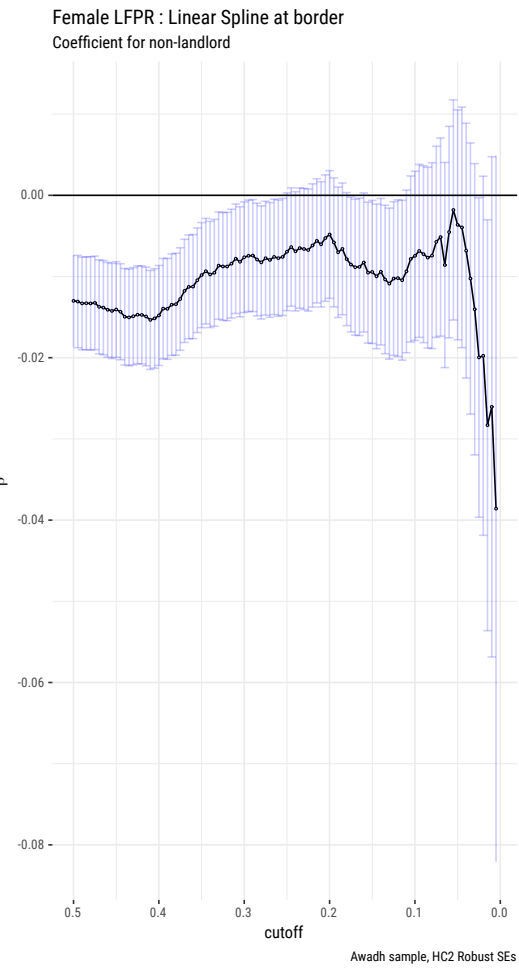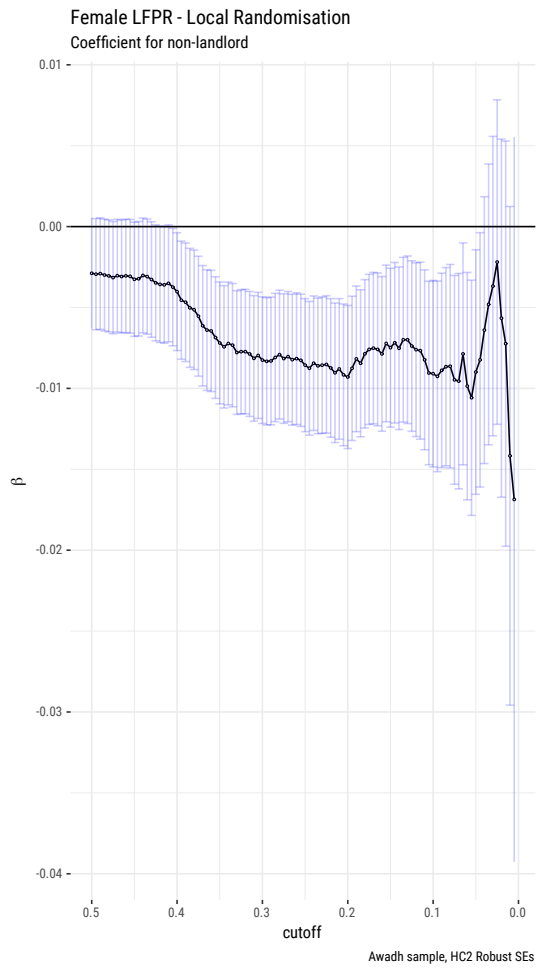Sex-Ratio - Local Randomisation
Coefficient for non-landlord
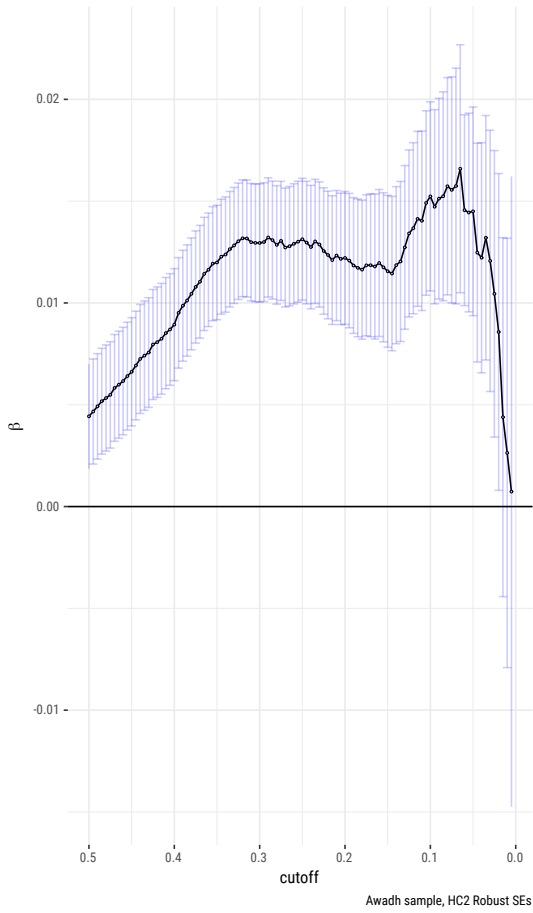
Sex-Ratio : Linear Spline at border
Coefficient for non-landlord
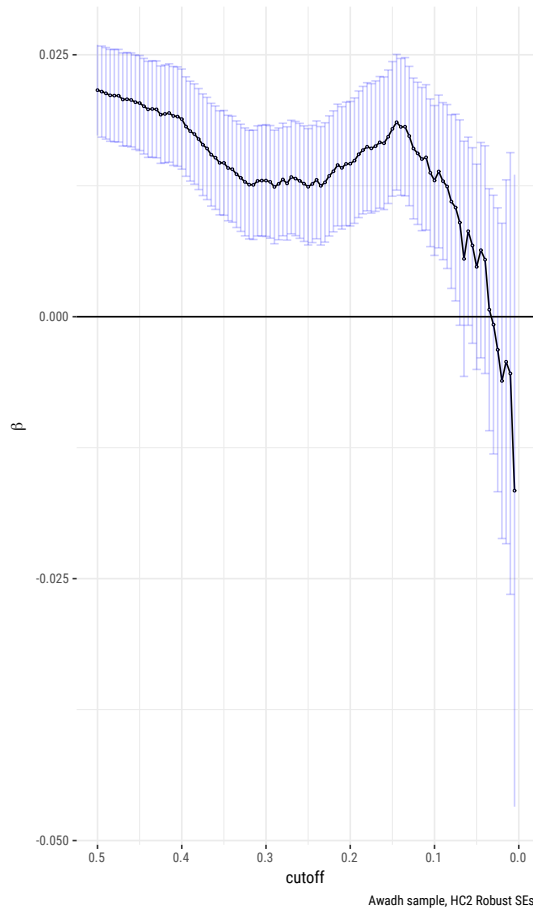
Sex-Ratio : Sex-Ratio: Quadratic Spline at border
Coefficient for non-landlord

Awadh sample, HC2 Robust SEs

Female Literacy - Local Randomisation
Coefficient for non-landlord

Female Literacy : Linear Spline at border
Coefficient for non-landlord

Female Literacy : Sex-Ratio: Quadratic Spline at border
Coefficient for non-landlord

Awadh sample, HC2 Robust SEs

49
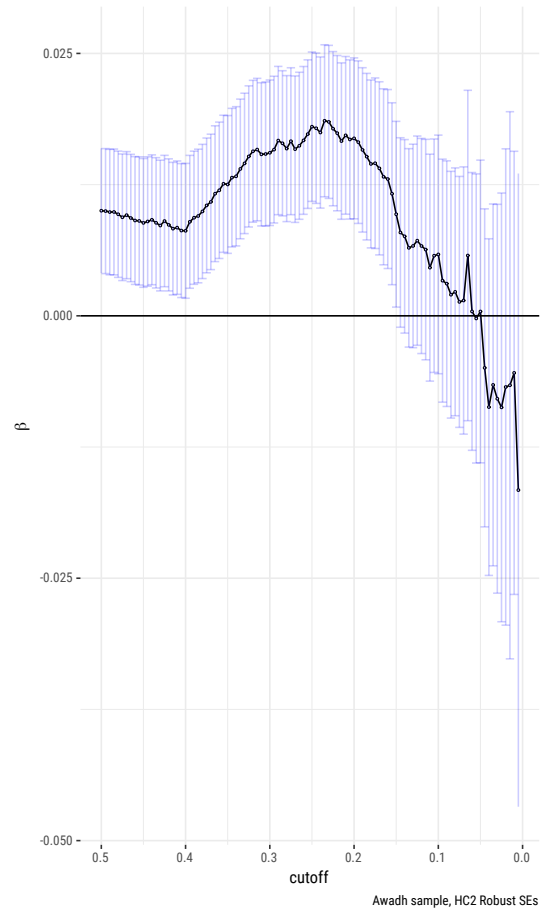
D.6 Estimates including villages in Bihar

D.6.1 OLS

Sex-Ratio Regression estimates
Coefficient for non-landlord

full sample, HC2 Robust SEs

Female Literacy Regression estimates
Coefficient for non-landlord

full sample, HC2 Robust SEs

Female LFPR Regression estimates
Coefficient for non-landlord

full sample, HC2 Robust SEs
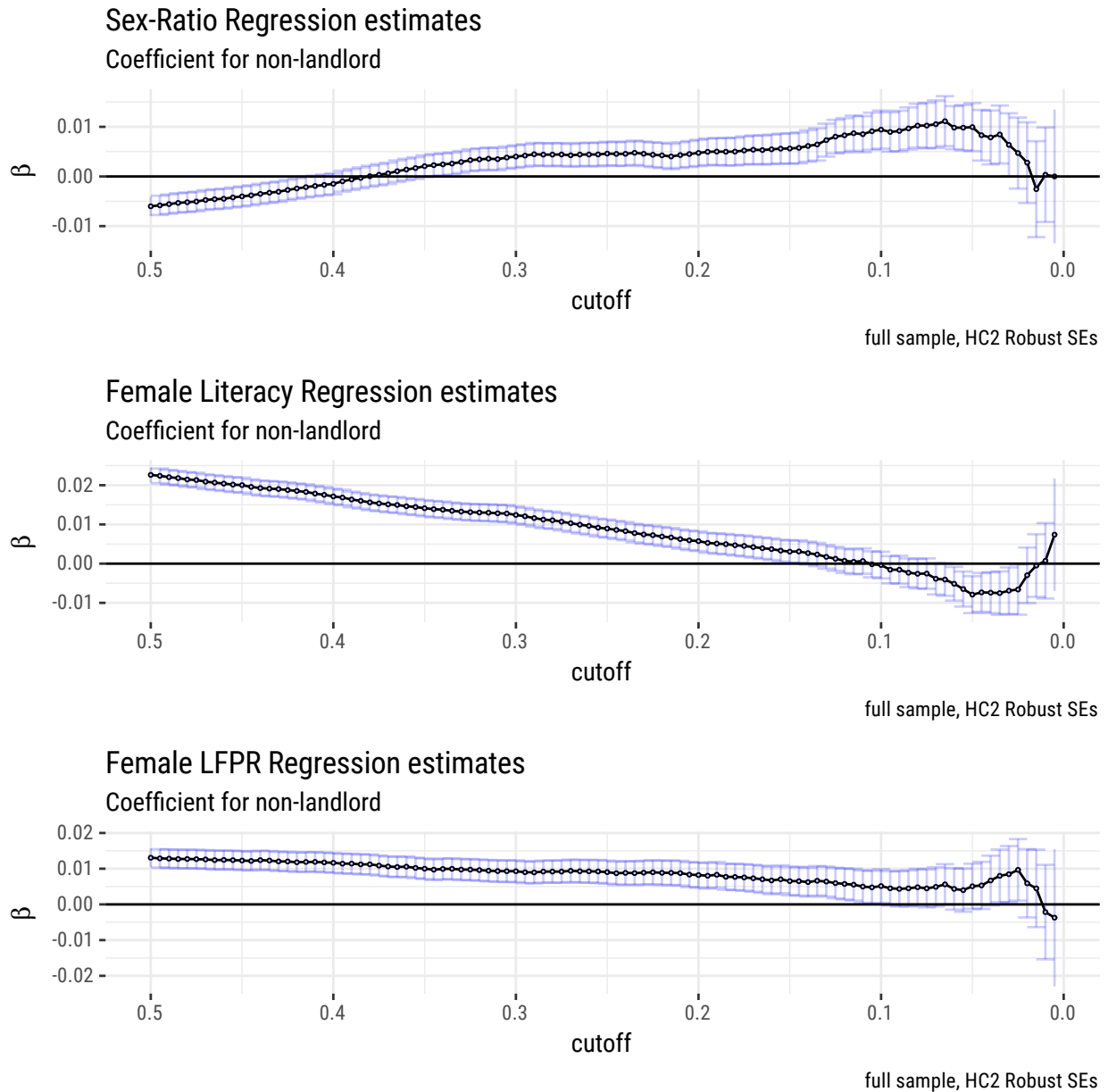
Figure A4: Mahalwari OLS Coefficients : UP and Bihar Sample

Sex Ratio (M/F)
Full Sample

Distance from Z–M Border

Female Literacy
Full Sample

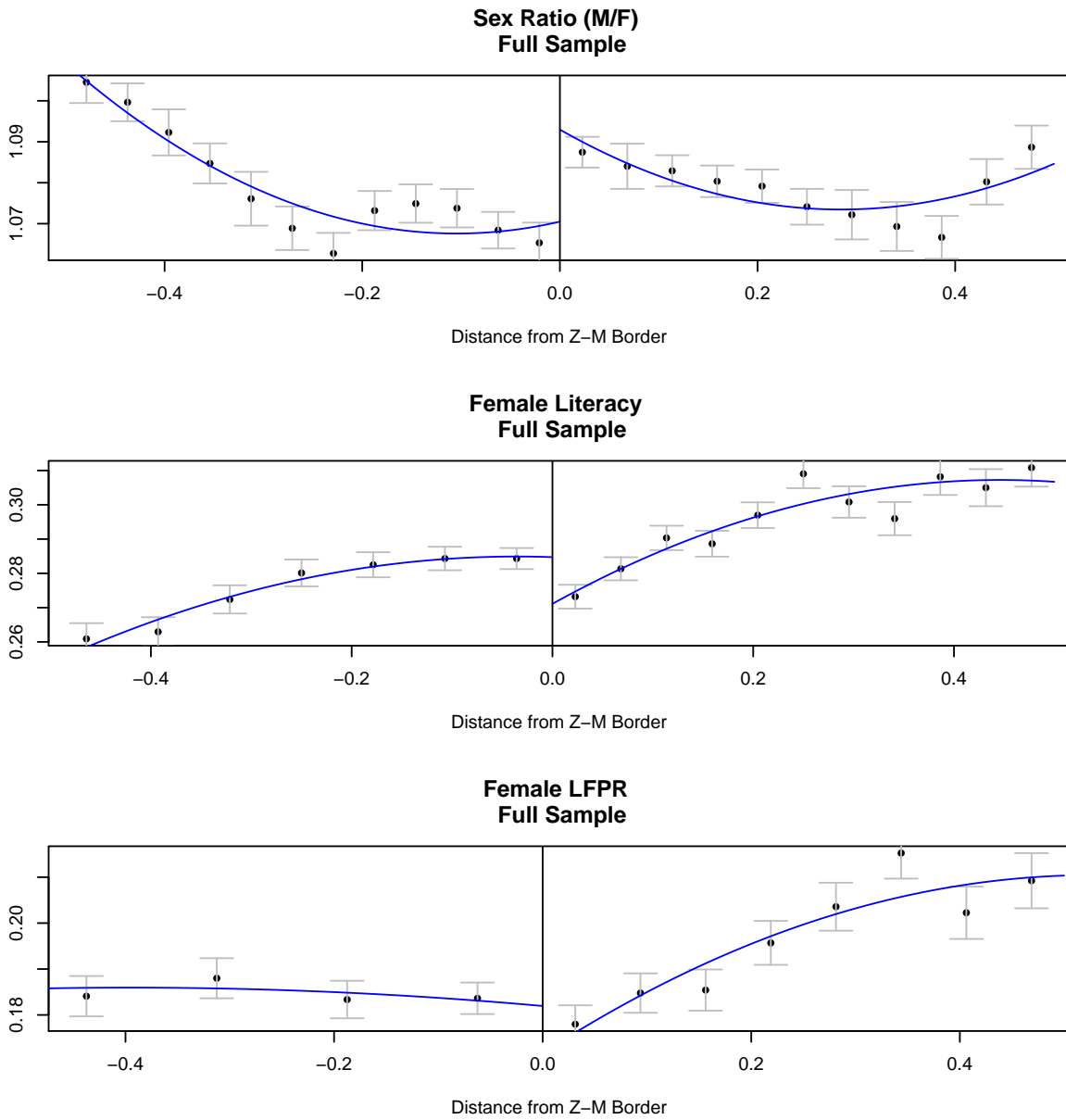Distance from Z–M Border

Female LFPR
Full Sample

Distance from Z–M Border

Figure A5: RD Estimates for UP and Bihar sample