DISCUSSION PAPER SERIES

DP17189

MISSING WOMEN IN COLONIAL INDIA

Bishnupriya Gupta, James Fenske and Cora Neumann

ECONOMIC HISTORY

ISSN 0265-8003

MISSING WOMEN IN COLONIAL INDIA

Bishnupriya Gupta, James Fenske and Cora Neumann

Discussion Paper DP17189
Published 07 April 2022
Submitted 06 April 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

MISSING WOMEN IN COLONIAL INDIA

Abstract

We construct novel data on female population shares by age, district, and religion in South Asia from 1881 to 1931. Sex ratios skew male in Northern India and are more balanced in Southern and Eastern India, including Burma. Male-biased sex ratios emerge most visibly after age 10, and this is not specific to any one region, religion, or time period. Sikhs have the most male-biased sex ratios, followed by Hindus, Muslims, and Jains. The female share correlates across religious groups within districts. Evidence that sex ratios correlate with suitability for wheat and rice is weaker than suggested by the existing literature.

JEL Classification: J16, N35

Keywords: N/A

Bishnupriya Gupta - b.gupta@warwick.ac.uk
University of Warwick and CEPR

James Fenske - j.fenske@warwick.ac.uk
University of Warwick

Cora Neumann - c.neumann@warwick.ac.uk
University of Warwick

Acknowledgements
We are grateful to Sonia Bhalotra, Sahar Parsa, Debraj Ray, and seminar audiences at the Economic History Society Meeting, Humboldt University, Trinity College Dublin, the University of Nottingham, the University of Oxford, the University of Warwick, and the Warwick-SUFE Workshop 2021 for their comments, and to the British Academy/Leverhulme Small Research Grants scheme for funding.

MISSING WOMEN IN COLONIAL INDIA

JAMES FENSKE†, BISHNUPRIYA GUPTA†, AND CORA NEUMANN †

ABSTRACT. We construct novel data on female population shares by age, district, and religion in South Asia from 1881 to 1931. Sex ratios skew male in Northern India and are more balanced in Southern and Eastern India, including Burma. Male-biased sex ratios emerge most visibly after age 10, and this is not specific to any one region, religion, or time period. Sikhs have the most male-biased sex ratios, followed by Hindus, Muslims, and Jains. The female share correlates across religious groups within districts. Evidence that sex ratios correlate with suitability for wheat and rice is weaker than suggested by the existing literature.

1. INTRODUCTION

In 1990, Amartya Sen contrasted the low ratios of women to men in South Asia, West Asia, and China with those in Europe and North America and estimated that 100 million women were "missing" (Sen, 1990). Sen's concern with sex ratios was not a new one. In the report of the Indian census of 1881, officials noted with concern the difference in the sex ratio in the population with the ratio that was found in most European countries. Females were in deficit in India compared to Europe. The census enumerators noted that, "[i]t was in the North of India the greatest divergence had been observed; while in the South especially, and less so in the East this variation became less and less conspicuous" (Report of Census of 1881: p.51). They attributed the high mortality among female children to lack of care (p.52). The literature has estimated the proportion of missing women for several Asian countries (Gupta, 2005), by region for India (Agnihotri, Palmer-Jones and Parikh, 2002) and by age groups (Anderson and Ray, 2010). Evidence from recent decades suggests large increases in the ratio of males to females at birth, particularly in certain regions (Censuses 1991-2011). To the extent that women are missing due to decisions regarding the allocation of resources they receive, to the extent that these

†UNIVERSITY OF WARWICK
†UNIVERSITY OF WARWICK
†UNIVERSITY OF WARWICK
E-mail addresses: J.Fenske@warwick.ac.uk, B.Gupta@warwick.ac.uk, C.Neumann@warwick.ac.uk.
Date: April 5, 2022.
We are grateful to Sonia Bhalotra, Sahar Parsa, Debraj Ray, and seminar audiences at the Economic History Society Meetings, Humboldt University, Trinity College Dublin, the University of Nottingham, the University of Oxford, the University of Warwick, and the Warwick-SUFE Workshop 2021 for their comments, and to the British Academy/Leverhulme Small Research Grants scheme for funding.

decisions change with levels of economic development, and to the extent that these decisions themselves affect economic development, the problem of missing women is an economic problem (Doepke, Tertilt and Voena, 2012; Duflo, 2012; Jayachandran, 2015). Economists and economic historians have, then, contributed to the literature on missing women (e.g. Beltrán Tapia and Gallego-Martínez (2020); Bhalotra, Brulé and Roy (2018); Das Gupta (1987); Gupta (2014)).

The literature offers several explanations for both levels and increases in the male bias in the sex ratio, in South Asia and elsewhere. One explanation is a preference for sons that existed historically and has become accentuated with a decline in fertility and the availability of pre-natal sex selective technology (Basu, 1999; Das Gupta and Mari Bhat, 1997; Gupta, 2005; Jayachandran, 2017; Jayachandran and Pande, 2017; Pande, 2003). The male bias in the sex ratio can also arise from lower bargaining power in post-reproductive ages and women's disadvantage in intra-household resource allocation (Calvi, 2020). The literature has explained the preference for sons and lower social status of women within the family in terms of several factors. These explanations include the relative productivity of women in certain types of agriculture (Alesina, Giuliano and Nunn, 2018; Bardhan, 1974; Boserup, 1970; Carranza, 2014; Miller, 1981; Qian, 2008; Rosenzweig and Schultz, 1982), women's importance in industrial activity (Xue, 2018), and the roles of kinship norms and marital residence patterns (Dyson and Moore, 1983).

The validity of these explanations depends on both *where* women go missing and *when* women go missing, both over the life cycle and over time. Anderson and Ray (2010) compute the numbers of missing women by age cohort and show that 11% of the deficit of women in India in 2000 was perinatal. More than two-thirds of missing women went missing over the age of 15, and a large number went missing over the age of 70. Therefore, neither sex selection at birth nor mistreatment of young girls can be a complete explanation of India's current male-biased sex ratios.

Historical data allows us to look for explanations of missing women based on variables that are historically persistent. These variables could include regional norms, religious norms, or the component of women's participation in economic activity that is determined by geographical factors. If historical sex ratios that vary by region, religion, and geography correlate with sex ratios today, then it is evidence consistent with explanations of missing women based on historically persistent causes.

In this paper, we introduce two new data sets based on the censuses of colonial India and document both where and when women go missing. The censuses cover the years 1881, 1891, 1901, 1911, 1921, and 1931, and include the territories of present-day

Bangladesh, Burma, India, and Pakistan. The data sets report the percentage of the population that is female in each district in each census year, by religion, and for four broad age categories: 0-10, 10-20, 20-40, and over 40. The data sets report these percentages at the level of districts and also for a set of aggregated “parent” districts that are constant over time, despite the district splits and reorganizations that occurred between 1881 and 1931. To our knowledge, ours are the largest and most granular data that have yet been assembled on sex ratios by age, religion, and location for colonial South Asia.

We use these novel data sets to understand where and when women went missing in the population a century ago. We describe variation in sex ratios by *region*, *religion*, *geography*, and *age*. We begin with the question of where women go missing by *region*. Broadly, the literature has classified India into four broad regions: the North, which encompasses regions such as the Punjab and Uttar Pradesh; the East, which includes regions such as Bengal, Assam, and also Burma, which was under British rule until 1937; the West, which includes Gujarat, Maharashtra, and the territory of the former Bombay Presidency; and the South, which includes the territory of the former British-ruled Madras Presidency, and the princely states of Hyderabad, Mysore, Travancore, and Cochin.

In our data, we show that sex ratios in colonial India as a whole are male-biased, and that this is particularly the case in the Northern region. The South and East are more balanced in their sex ratios. This is consistent with the data from recent censuses. Late colonial sex ratios correlate strongly with sex ratios in the present, not only across broad regions, but even looking within states of modern India. Evidence of this persistence is strongest among children and adolescents. In our analysis, we thus confirm the regional divide recognized in past and recent studies and show persistence at a granular level.

Next, we discuss sex ratios by *religion*. The focus of the literature in the present day has been on son preference among Hindus. In our data, we show that sex ratios are most male-biased for Sikhs and still skew male, albeit less so, among Hindus, Muslims, and Jains. The share of the population that is female correlates across religious groups within a district. For example, if the Hindu cohort under age 10 in a district is more male, the Muslim cohort in the same district is more likely to skew male. This inter-religious correlation does not show trends over time, but is stronger for older cohorts than for younger cohorts within any given census wave. Indeed, we find positive inter-religious correlations across districts using the 2011 Indian census. In colonial India sex ratios show similarities across religions within regions and over the life cycle. This differs from the evidence on son preference from recent decades, and suggests that the difference between Hindu and Muslim sex ratios in the under 5 age group in the present is a result of greater use of sex selective abortions in Hindu communities (Bhalotra, Clots-Figueras and Iyer, 2021). The technology of sex selective abortion was not available in the colonial period.

We then turn to differences by *geography*. Past work has sought to explain the North-South divide in terms of factors such as female labor participation and the greater involvement of women in rice cultivation rather than wheat cultivation (Bardhan, 1974; Boserup, 1970; Miller, 1982). Spatial variation in the economic value of women has been attributed in the literature to women's participation in agricultural activity, such as hoeing, weeding, and planting. In rice-growing regions, women's labour force participation is high and their economic value makes son preference less important. In the drier areas where millet is grown and in the wheat-growing regions, women's participation in economic activity is low. In these regions, preference for sons prevails. Therefore the broad regional divide between North, South, and East may reflect a divide between rice and wheat cultivation. The literature provides additional reasons to anticipate that geography will correlate with sex ratios. In the industrial towns, the sex ratios were often male-biased due to migration of men working in industrial jobs. Docklands were yet another preserve for male employment; women were rarely employed in these areas. Men migrated, leaving their families in the village (Chandavarkar, 2003; Ojha, 2014). Therefore, coastal areas with ports may be expected to have a deficit of women in the population.

In our data, we correlate the female share of the population with several observable geographic variables. Basic correlations show that, in various age ranges, the share of the population that is female is greater in Eastern districts and in districts suited to the growing of cotton, dryland rice, and wetland rice, while it is lower at higher latitudes, and in districts suitable for the growing of wheat. However, once we perform regressions that account for the correlations between these variables, and include fixed effects that focus our comparisons within colonial provinces, these patterns are less pronounced. In these more demanding specifications, latitude is a negative predictor of the share female at younger ages. Eastern districts have more females in older age ranges, though this is partly a proxy for the presence of Buddhists in Eastern districts. Districts whose geography makes them suitable for malaria have more male-biased sex compositions in several age categories. The evidence that wheat suitability predicts more male populations is at times weak, and wetland rice suitability is rarely a significant predictor of more female populations. Suitability for dryland rice, by contrast, is a robust predictor of more female populations in later census years. In sum, we provide systematic evidence of how geography may predict women's status in society, as reflected in the sex ratio. The rice-wheat distinction that is discussed widely in the literature is not statistically robust, a conclusion that is also apparent using data on actual crop cultivation.

Finally, we examine patterns by *age*. We compute age-specific sex ratios, which capture when women go missing over the life cycle. Recent literature finds a large number of missing girls at birth in some regions in present-day India. It is more difficult to obtain data on sex ratios at birth for districts in colonial India. Evidence from the colonial censuses points to a male deficit before age 1 that becomes a female deficit by age 10 (Bhaskar and Gupta, 2007). Two factors have been cited to explain this pattern. First, male mortality at birth is higher in underdeveloped economies, and so life expectancy at birth was higher for females than for males up to 1921. Second, the technology of fetal sex selection has only become available from the 1970s. Male-biased sex ratios in colonial India were, then, outcomes of neglect of girls and of greater mortality among adult women. In our data, we show that sex ratios, while already slightly male-biased in the 0-10 age cohort, become much more strikingly male in the 10-20 cohort before becoming more balanced again in the 20-40 age range. The dip in the share of females in the age group 10-20 is similar in all regions, and suggests that child marriage and maternal mortality may be a factor everywhere. It is also consistent with under-reporting and misreporting of girls in this age cohort over all four regions. This age profile is apparent in each census wave between 1881 and 1931, though it is more modest in 1931 than in earlier years. Despite the regional differences in sex ratio, we find that variation in the sex ratio by age shows similar patterns across all regions.

1.1. **Contribution.** We contribute to a literature on gender in economics (Behrman et al., 1999; BenYishay et al., 2020; Bordalo et al., 2019; Croson and Gneezy, 2009; De Moor and Van Zanden, 2010; Giuliano, 2014; Goldin, 1995; Luke and Munshi, 2011; Niederle and Vesterlund, 2007), in particular to a branch of this literature that has focused on South Asia (Ashraf et al., 2020; Boserup, 1970; Carranza, 2014; Dyson and Moore, 1983; Lowes, 2020; Miller, 1982). We contribute to the literature on the origins of gender norms, which includes work on the prevalence of plough-dependent agriculture (Alesina, Giuliano and Nunn, 2013) and cotton spinning (Xue, 2018), among other factors. Because of data limitations, many of these papers have been compelled to “compress” history, ignoring the evolution of the outcomes they study between the distant past and the present day (Austin, 2008). By introducing new data that covers several points in time dating back to 1881, we are able to evaluate the degree to which one gender outcome – the sex ratio – has been persistent over more than a century within South Asia.

Economists and economic historians have also shown that historical events can lead to persistent differences in outcomes by gender. This is part of the larger literature on persistence in economics and economic history (Voth, 2020). These studies have identified long-run impacts of Christian missions (Cagé and Rueda, 2020; Nunn, 2014), colonial law (Anderson, 2018), historic sex ratios (Grosjean and Khattar, 2019), labor mobilization during the Second World War (Acemoglu, Autor and Lyle, 2004; Goldin and Olivetti, 2013; Goldin, 1991), male mortality during the First World War (Boehnke and Gay, 2020), and Africa's slave trades (Teso, 2019). They have shown that inherited culture affects female labor force participation and can be slow to change (Fernandez, 2007; Fernández, 2013; Fernández, Fogli and Olivetti, 2004; Fogli and Veldkamp, 2011). By providing panel data covering six points in time between 1881 and 1931, we make available a source that later researchers will be able to use to identify exogenous shocks to historic sex ratios, and to evaluate whether the impacts of these shocks have exhibited historical persistence.

We also contribute to the literature on women's relatively poor health outcomes and disproportionate mortality over the life cycle in developing countries, including within South Asia (Anderson and Ray, 2010; Bhaskar and Gupta, 2007; Sudha and Rajan, 1999). Research on this theme in South Asia has found, for example, that girls receive fewer prenatal investments (Bharadwaj and Lakdawala, 2013), that girls are breastfed less (Jayachandran and Kuziemko, 2011), that a preference for eldest sons constrains investments in daughters' health (Jayachandran and Pande, 2017), and that survival of girls is more sensitive to rainfall shocks (Rose, 1999). We introduce historic data on sex ratios at a much finer geographic level than has been available previously for South Asia. This permits us to quantitatively reexamine several hypotheses about the determinants of male-biased sex ratios in the region that have been tested only in more recent data, or supported only by coarse regional comparisons. Our findings suggest two revisions to the literature on gender norms in India. First, we find that regional differences are as salient as differences across religions within regions. Second, differences in son preference arising from rice and wheat agriculture are less pronounced than is generally claimed in the literature.

The remainder of this paper is as follows. In section 2, we explain our underlying sources of census data and describe the new data sets we produce. In section 3, we describe regional patterns in sex ratios and how they changed over time. In section 4, we discuss how sex ratios varied by religion across colonial India. In section 5, we explore the geographic correlates of sex ratios. In section 6, we explore age-specific sex ratios to understand where women go missing over the life cycle. Section 7 concludes.

2. DATA

2.1. **Overview.** In this project, we introduce two data sets based on the colonial Indian censuses of 1881, 1891, 1901, 1911, 1921, and 1931. The first data set reports variables for the changing set of districts that are covered by the census in each year – the "contemporaneous" districts. The second combines together districts that split or were reorganized over time into a consistent set of "parent" districts that are, with a few exceptions, constant over this time period.

For each of these two data sets, we report at the district level the percentage of the population that is female in a given census year, including by subcategories defined by four broad age bins – 0 to 10, 10 to 20, 20 to 40, and over 40 – and six religious categories – Buddhist, Hindu, Jain, Muslim, Sikh, and Tribal. The ultimate source of these data are the district-level tables labeled "Age, Sex and Civil Condition," within the Imperial Tables of each separate census wave. Rather than working with the ratio of females to males in a given age bin, we use the percentage of the population that is female in our analyses. We use percentages, rather than the sex ratio, because of the tendency of sex ratios to take extreme values, especially among small populations, creating skewed distributions. Readers interested in converting percentages to sex ratios can simply transform them as $SexRatio = \frac{PercentFemale}{1-PercentFemale}$.

2.2. **Sex ratios as an measure of missing women.** While the sex ratio in a given district, age bin, and religion depends on relative female mortality, and hence is an indicator of the degree to which women are missing, labor migration in some situations can also lead to sex imbalance when the work force in an industry has a strong sex imbalance. Workers in cotton mills in Bombay, for example, were disproportionately male (Morris, 1965, p. 65). Migration, including migration for marriage, is, however, unlikely to have a large influence on sex ratios in our sample. In 1881, at least 97% persons in India lived in the provinces in which they were born. In 1901, 91.3% of Indians in India lived in their district of birth (Collins, 1999, p. 252). Throughout, then, we will treat the percentage female as a noisy measure of relative mortality. Following the recent public health literature, we take sex ratios at the earliest ages as those that are least likely to be affected by other factors (Guilmoto et al., 2018).

It is similarly unlikely that, except in a small number of identifiable cases of major cities, mines, and penal colonies, urbanization and structural change drive the patterns we observe in sex ratios. Urbanization was slow in India; while 8.7% of the population lived in towns above 5,000 persons in 1872, by 1931 this had risen only to 11.1% (Visaria and Visaria, 1983, p. 519). Similarly, the dominance of agriculture changed little. In the 1881 Census Report (p. 350), approximately 77% of workers were in agriculture.[1] The

[1]To arrive at this figure, we subtract 48,794,195 persons with unspecified or unstated occupations (a large portion of which are children) from 115,417,956 persons to reach an estimate of 66,623,761 workers. We divide 51,089,021 agriculturalists by this number.

1931 Census Report (p. 276) noted that 67% of workers worked in the "exploitation of animals and vegetation" – the category most closely corresponding to agriculture.

2.3. **Background: the colonial censuses.** The first attempt by the British to systematically count the population of India was undertaken between 1867 and 1872, but this census left out large portions of the country and was imprecise due to different regions being enumerated in different years.[2] Following this inadequate first attempt, British authorities created a system to accurately count the population once every ten years, starting in 1881. Over time, more territories were censused.

To improve the process, in 1901 the country was divided into blocks and a large workforce was recruited, which was split into three levels: Charge Superintendents, Supervisors, and Enumerators. Enumerators were responsible for counting people in one block, or roughly 30-50 houses. They then reported their tables to the Supervisors, who in turn combined these tables to send them to their Charge Superintendents, who then combined these tables and sent them to the Census Commissioner. The number of people involved in collecting the information was large. For example, for the 1901 Census, 1.3 million enumerators, 122,000 supervisors and 9,800 superintendents were employed. In order to count people by different categories, e.g. by religion, age, or marital status, in the first two census waves each person had been ticked into all their categories and then all the ticks were added up to calculate the population counts. Starting in 1901, however, each counted person's categories were recorded on a slip, and then slips were counted together by category. The slips were colour coded for religion and shapes and symbols were used to record sex and civil condition. This system reduced the workforce required as well as the frequency of making errors. During January and February of 1901 information from households was collected by the enumerator in charge of the block. On March 1, this was checked and updated and on March 2, the totals were calculated at the level of the block. Due to the time it took to combine all tables and count the different sub-groups, the whole process took roughly six months.

2.4. **Datasets.** Our analysis takes place at three geographical levels: contemporaneous districts, parent districts, and regions. We take the contemporaneous districts from each census wave, as they are the level at which population counts were reported. The majority of our descriptive and analytical work is done at the contemporaneous district level. Due to the changes in districts over time, we use parent districts for comparisons across years. These parent districts take into account splits and merges of districts over the years. To this end, we create two data sets.

[2]We base this account on the Census Reports from 1881 to 1931.

Our first data set covers the census years between 1881 and 1931 and takes as the unit of analysis the set of districts that exists in any given census year, reflecting both changes in the set of districts that were censused and the set that existed. Census coverage of districts becomes more complete over time. Much of Northern Burma, for example, is not covered in 1881 and 1891. Similarly, the entirety of the Central India Agency is reported as a single unit in the 1881 census, while the separate parts of the agency that have their own entries in later years differ across years. Coverage of the population of the Andaman and Nicobar Islands prior to 1921 includes only the penal colony of Port Blair.

Adding to these differences in coverage, districts were sometimes split (as with the division of Khandesh in to East and West Khandesh after 1906) and other times reorganized (as with the districts of Berar after 1905). While this shifting composition of districts complicates any use of the data in a panel setting, it preserves the granularity of the underlying information and reduces the scope for measurement error due to aggregation of districts together. In this data set, we have 313 districts in 1881, 327 in 1891, 366 in 1901, 380 in 1911, 408 in 1921, and 506 in 1931.

Our second data set also covers the census years between 1881 and 1931, but takes as its unit of analysis a consistent set of parent districts that aggregate together districts that were split and reorganized over time. In the above example, West and East Khandesh are merged together in all years to form the aggregate district of "Khandesh." Where districts were created by combining portions of several districts, this compels us to form synthetic units, such as the aggregation that we label "Hazara, Peshawar, Kohat, Bannu and Dera Ismail Khan."

These data are not, however, a balanced panel. First, there are a number of districts that are reported so inconsistently across census years in the original data that we have not attempted to form parent districts with these. Second, some districts are not censused in earlier years (for example, in northern Burma), and so are missing in these years. This data set contains 305 districts.

To show more general trends and to account for the geographical divide in percentage female that has been identified in the literature, we use four regions also used by Gupta (2021) – East, North, South, and West. The districts are assigned to these regions through their provinces, which vary across census years. In the East, these are the Andaman and Nicobar Islands, Assam, Behar, Bengal, Burma, Orissa, and, where they are separately reported, Calcutta and Mayurbhanj State. In the North, these are Ajmer-Merwara, Baluchistan, Delhi, Jammu and Kashmir, the North West Frontier Province, Punjab (including the Native States), the Rajputana Agency, and the United Provinces. In the South, these are Cochin, Coorg, Hyderabad, the Madras Presidency, Mysore, and

Travancore. In the West, these are Baroda, Berar, Bombay, the Central India Agency, the Central Provinces, Gwalior, and the Western India States Agency.[3] We provide maps of these regions in the Online Appendix.

2.5. **Harmonizing the data.** There are some limitations of the original data that must be accounted for in order for us to draw comparisons between districts and over time. In particular, there are differences both across census waves and across provinces within any given census wave in how religions and age categories are reported.

While the original data include 17 distinct religious groups, we focus our analysis on the six main religious groups in Colonial India: Buddhist, Hindu, Jain, Muslim, Sikh, and Tribal (i.e. Animist). Because the Christians and several other religious groups are only recorded in a few districts, we have focused on these six groups in order to draw meaningful comparisons for the whole country. While Hindu and Muslim populations are recorded in virtually every district and age group, population counts for the other four religious groups are missing in some districts because either there were no inhabitants of that religious group in the district or the number was small and thus they were grouped together with other religions to form an "Other" category. Nevertheless, the six studied religions are present in a majority of districts and can thus be examined comparatively. There is a regional component to the distribution of religious groups; for example, Buddhists are mainly present in the East and North, while Jains are mainly present in the North and West. Hindus and Muslims together made up 89.8% of the population in 1931, while adding Tribal populations brings this to 92.2. The reporting of the number of Christians by district is not systematic and often included in the category "other." Therefore we do not report the share of women in this religious group, which was 1 percent of the population in 1901.[4]

Across years and between provinces, population counts were reported in different age bins. For example, in 1911 the United Provinces, Travancore, and Burma reported population counts in 5-year bins for early ages (0-5, 5-10, 10-15, 15-20) and in 20-year bins for later ages (20-40, 40-60), while Coorg, Cochin, and the Central India Agency reported population counts in 5-year bins until age 70. In other years and provinces, for example in 1881 across the whole country, children are counted in a wider age bin of 0-9 or 0-10. In order to make our analysis comparable across provinces and years, we

[3]Some of these provinces are often separately considered as a "Central" region in the literature. However, many districts of these regions, as given by colonial borders, would be part of present-day Gujarat or Maharashtra.

[4]Calculated from p. 572-3 of the Census of India, 1931, Vol. I. – India. Part II. – Imperial Tables. The 1931 Census Report of India (p. 200) provides all-India sex ratios for both Christians and Parsis. The Christian population had 1,002 girls per boy in the 0-1 age bin and 952 females per 1,000 males of all ages. These are very similar to the Hindu figures of 1017 and 953, respectively. For Parsis, the sex ratios were lower, at 981 and 940 in the 0-1 and all ages categories.

chose age bins that can be reported consistently across provinces and years while still allowing for meaningful interpretation. These age bins are 0-10, 10-20, 20-40, and over 40. The coarseness of the early census years and for certain provinces in later censuses compels us to use a 0-10 age category, rather than computing under-5 sex ratios.

2.6. **Control variables.** To our data, we have added a small number of additional control variables. We use our data to compute shares of the total district population for our six religious categories, as well as for a residual "Other" category. One limitation of the census data is that, when a religion is uncommon in a province (for example, Sikhs in Burma), the category is not separately reported in the original census data, and is instead included in the "Other" category. This compels us to code the population share of that religion as zero in these instances, which will induce some minor measurement error in these religious shares.

We have also added a number of district geographic characteristics that we have computed using geographic information systems (GIS) data. These include, for example, measures of the suitability of a district for the cultivation of certain crops.

In order to compute geographic covariates for the districts in our data, we begin by creating a one-to-many correspondence between the historic districts mapped for each census and the third-level administrative units (e.g. thesils) for which GADM, the Database of Global Administrative Areas, provides shapefiles.[5] We join these shapefiles to several sources of geographic data. For continuous variables such as latitude, we compute an area-weighted average of the variable over the polygons that together constitute the historic district. When the underlying data is in raster format, we average over raster cells within a sub-district. Sub-districts too small to contain a raster cell when the data has a coarse underlying resolution are assigned the value of the nearest raster cell. We will follow this procedure for all raster data. For dummy variables such as presence of a coastline, we take the maximum over this set of sub-district polygons. That is, if any part of a district has a coastline, then the district as a whole has a coastline.

We select covariates based on their common use in the economics literature, on their importance to Indian agriculture, and on their relevance to broad hypothesis in the literature about the determinants of women's roles in South Asian economies. We provide complete maps of these geographic variables in the Online Appendix. Because districts and their boundaries change over time, we plot these maps for each census year.

[5]We base these correspondences off several scanned maps. For present-day India, we make extensive use of Chandramouli, Singh and Sethi (2011). We also employ the map included in the 1931 census, which Fenske and Kala (2021) have previously converted to shapefile format, maps from the Imperial Gazetteer of India made available by the Digital South Asia Library, and maps available in the census reports of the specific provinces in several census years.

In our data, *Coastal* is an indicator equal to 1 if the district has a coastline. We compute this variable using the polygon map of oceans provided by Natural Earth Data. While this variable has not featured heavily in the literature on sex ratios, it will correlate with past exposure to trade and predict the presence of port cities (Jha, 2013). *Latitude* is the latitude of a district's centroid, which we compute ourselves in ArcMap. *Longitude* is the longitude of a district's centroid, analogously defined. While neither of these variables will be of interest in and of themselves, it has become conventional to control for them in order to reduce the influence of omitted variables that vary continuously over space (e.g. (Michalopoulos and Papaioannou, 2013)). These will also proxy, broadly, for more Northern and more Eastern areas of India.

Malaria Suitability is the Kiszewski et al. (2004) index of the stability of malaria transmission, and is originally available as raster data. While this has not featured prominently in the literature on sex ratios, it is both a variable of interest and conventional control variable in several studies in development (Alsan, 2015; Cervellati, Esposito and Sunde, 2021). Diseases such as malaria increase the infant mortality rate generally, and excess male mortality may be greatest when infant mortality is otherwise high (Alkema et al., 2014; Drevenstedt et al., 2008).

Major River is a dummy for whether any part of the district is intersected by a river in the polyline shapefile of "Major Rivers of the World" provided by the World Bank.[6] The rivers that intersect South Asia in this source are the Indus, Ganges, Brahmaputra, Salween, and Irrawaddy – the latter two of which are in Burma. This dummy will be a correlate of several possible sources of relative female survival, including land suitability for agriculture, the importance of irrigation, commerce, and population density. *Ruggedness* is the Nunn and Puga (2012) index of terrain ruggedness. The underlying raster data consist of the average deviation in elevation between a given grid cell and its 8 immediate neighbors. Within South Asia, larger values typically correspond to mountainous terrain in Himalayan regions. While this has not featured prominently in the literature on sex ratios, it is well known that the Himalayan regions of South and Southeast Asia that score highly on this index are distinctive along several cultural and institutional dimensions (Scott, 2009). Polyandry, for example, is more common in these areas (Korn, 2000).

Cotton Suitability is the district's potential output of cotton under rain-fed agriculture and with low inputs. The underlying data here are from the Food and Agriculture Organization of the United Nations' Global Agro-Ecological Zones (FAO-GAEZ) project.

[6]`https://datacatalog.worldbank.org/dataset/major-rivers-world/resource/c8d8958a-51e4-4e84-88a6-75e5abc26ea4`

These are initially available as raster data, and measure exogenous potential crop output based on factors beyond human control. This is measured in kilograms dry weight per hectare. While the FAO computations of potential output are based on modern-day remote sensing data, they have become widely used in economic history (e.g. Dimico, Isopi and Olsson (2017); Dupraz (2019)).

Cotton is chosen here as one of India's main export cash crops; it may have increased rural incomes, but may also have increased food insecurity (Harnetty, 1971; Satya, 1997) We compute several analogous suitability measures from the FAO data: *Dryland Rice Suitability* is the district's potential output of dryland rice, *Tea Suitability* its potential output of tea, *Wetland Rice Suitability* is the district's potential output of wetland rice, and *Wheat Suitability* captures a district's potential output of wheat. Rice and wheat are included because they have featured prominently in the literature on sex ratios. This literature argues that, while women play a major role in paddy rice production, wheat is intensive in male labor, which may affect the economic value of men and women as discussed in the introduction. Tea, by contrast, is a crop that was largely cultivated by migrant workers who were roughly half female; this was not the case in other crops, and was driven by female participation in plucking of leaves (Gupta and Swamy, 2017).

The existing evidence on colonial sex ratios is collected at the province level, and the regional categories reported in past studies use aggregate, province-level data. Our regional data is constructed using district-level evidence. Further, our data set includes information by age cohort, and so allows us to track changes over time in the same age cohort and across age groups in a census year.

3. Regional Variations in Sex Ratios in Colonial India

3.1. **Literature.** From the early censuses, enumerators noted the North-South difference in sex ratios, and considered the East to be somewhere in between these two regions. Bardhan (1974), using data from the 1950s and 1960s, noted that female life expectancy was higher than male life expectancy in the Eastern states of Assam, West Bengal, Bihar, and Orissa, while male life expectancy was higher in the Northern states of Punjab and Rajasthan. In Southern States such as Tamil Nadu and Kerala, survival of girls was more likely. Bardhan attributes this pattern to a number of variables, including regional differences in the age of marriage, female literacy, maternal medical care, and neglect of girls.

The literature has explored factors behind the North-South division, framing it around the Boserup (1970) thesis of women's labour force participation. Bardhan (1974) speculated that women worked in rice cultivation in the South and East, and had little involvement in agricultural lands that are suitable for wheat and millet in the North and

North-West; therefore they were more valued in the South and the East. Rosenzweig and Schultz (1982), similarly, find a positive correlation between female employment and girls' survival. Kishor (1993) finds a negative correlation between area cropped to rice and under-5 female mortality in census data from the 1980s.

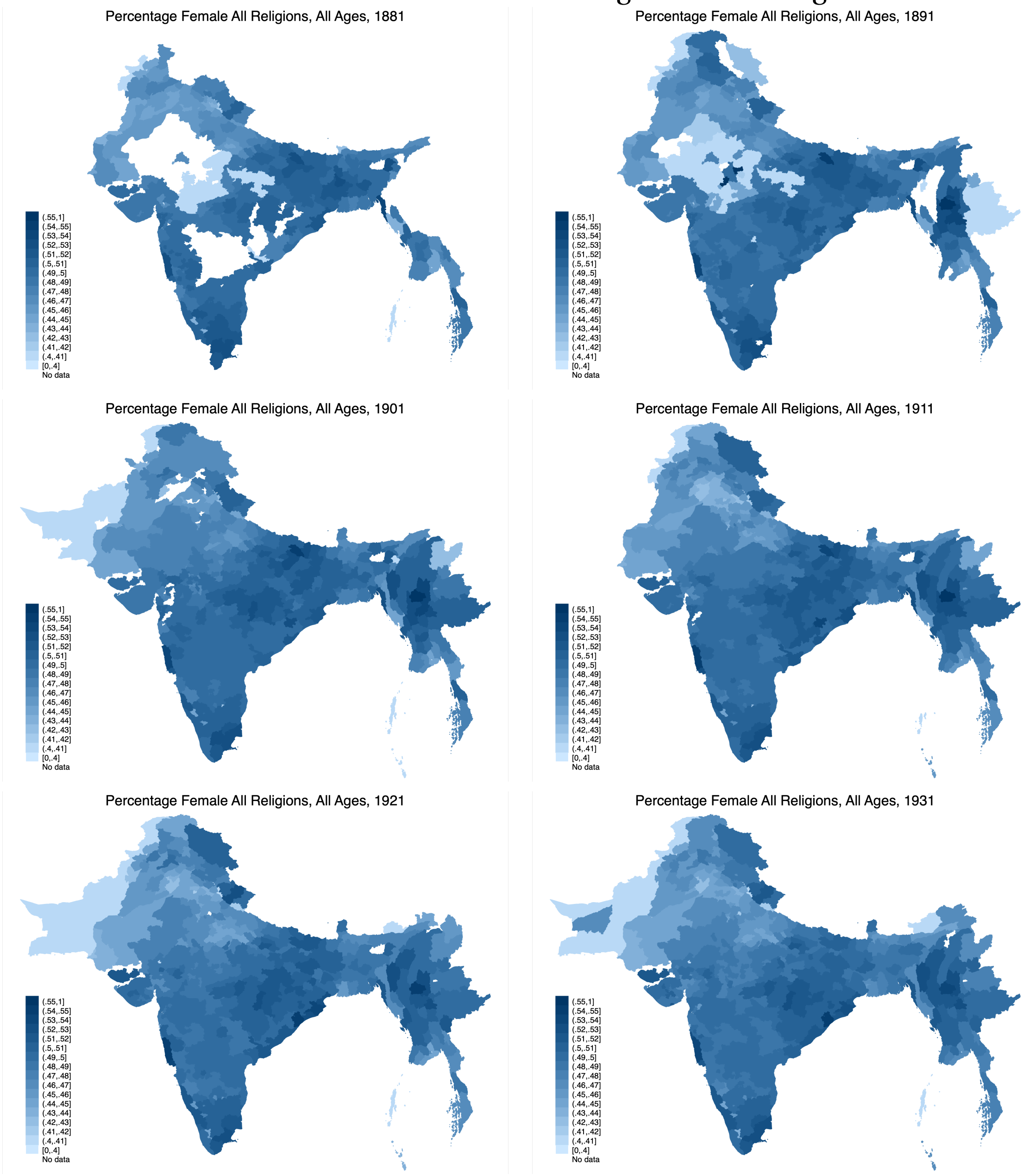
Other scholars have instead stressed kinship and marriage patterns. Dyson and Moore (1983) argue that differences in kinship and marriage conventions between the North and the South explain the lower status of women and greater son preference in Northern India, where women marry outside their place of birth, and therefore have less value to the parental family. Kishor (1993) cites differences in kinship structure and female labor force participation as important determinants of child mortality in India, but shows they cannot fully explain the North-South gap. Clark (2000), using data from the early 1990s, finds a similar pattern of more male-biased sex ratios in the North, particularly in families with few children, or among those who are disadvantaged. Miller (1981, 1982) used census data from 1961 to show a North-South divide in women's labour force participation, restrictive social norms for women, and male-biased sex ratios. She stresses the relative prevalence of dowry, rather than reciprocal marriage payments, in the North.

3.2. **Patterns.** In the analysis that follows, we show patterns for the four regions defined in section 2 – East, North, South, and West. In Table 1, we show the percentage of the population that is female in each of these regions in each census year, alongside the total population (male and female) within that region and year and the number of districts from which these totals are computed. Because we aggregate these numbers from the data on districts, and not all districts are covered in the Age, Sex, and Civil Condition sections of the census, there will be differences from comparable totals in the census reports. While the total sex ratio for all of India is approximately 49% in each year, it is clear that there are regional differences. In the South, the percent female is above 50% in all years, while the percentage female in the North never exceeds 48%. Like the South, the East has a less male-biased sex ratio in all years. Sex ratios are clearly male-biased in the West, albeit less so than in the North.

This table only provides one snapshot of the regional variation available in our data. In the Online Appendix, we provide a complete breakdown of the percentage of the population that is female by census year, region, religion, and age bin.

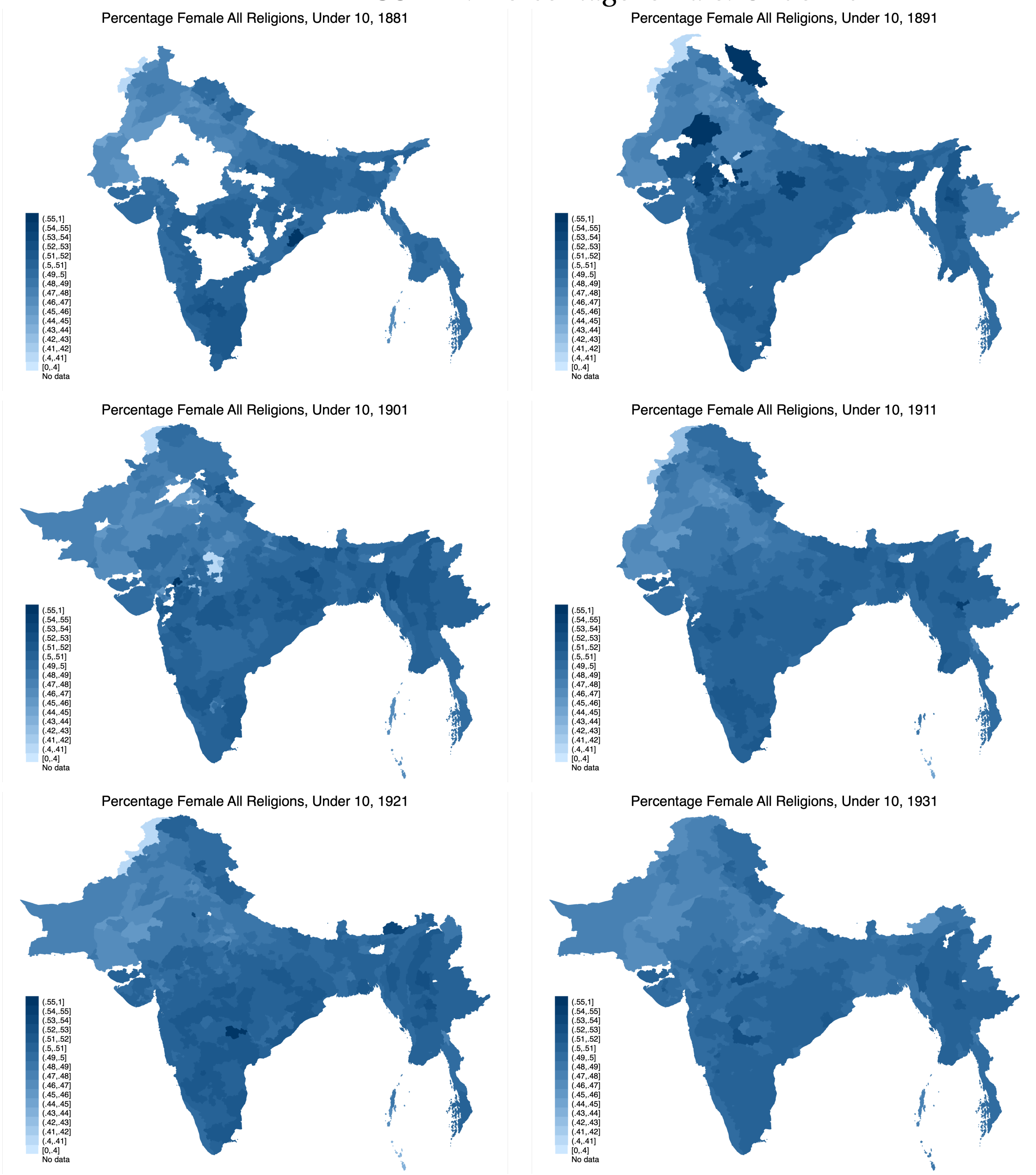
Because these broad regional tables can mask sub-regional differences, in Figure 1 we present a map of the percentage of the population that is female in each district in each census year. These numbers capture the population across all religions. District boundaries here are defined by our correspondence of the census districts with the sub-district

FIGURE 1. Percentage female: All ages



polygons made available by GADM. All maps we show will be based on this correspondence. An immediate geographic separation is apparent. North of a line that roughly connects Surat on the western coast of India to Uttarakhand in the Himalayas, the percentage of the population that is female is markedly lower than elsewhere. Higher female percentages of the population are visible particularly in southern peninsular India and in contemporary Burma. While the stability of this geographic pattern is obscured

FIGURE 2. Percentage female: Under 10



by the fact that Baluchistan, much of Burma, and many of princely states are not censused in 1881 (Hyderabad, Rajasthan, and the Orissa Tributary states all stand out for their absence), this general geographic pattern is visible across census waves.

In Figure 2, we present similar maps of the population aged 0-10. Here, the regional patterns are roughly similar. The general gradient separating the northwest from the rest of South Asia is again clear. Some of the more extreme outliers now have more balanced sex ratios, reflecting the prevalence of male migrants of working age in some of these frontier areas. These maps are only a small sample of what we can show using our data. In the Online Appendix, we provide a complete set of maps of the percentage female by religion, age bin, and census year.

3.3. **Persistence.** To what extent do the spatial patterns in colonial sex ratios persist into the present? To answer this question, we turn to the 2011 Census of India,[7] and compute the percentage female for each district that existed in 2011, for age bins that mimic those we can construct for the colonial period: 0-9, 10-19, 20-39, and 40+. We show maps of the percentage female in each of these age bins across districts in Figure 3. In the 0-9 and 10-19 age bins, the familiar pattern of male bias appears in the north west, particularly in Punjab and Haryana, as well as in inland Maharashtra.

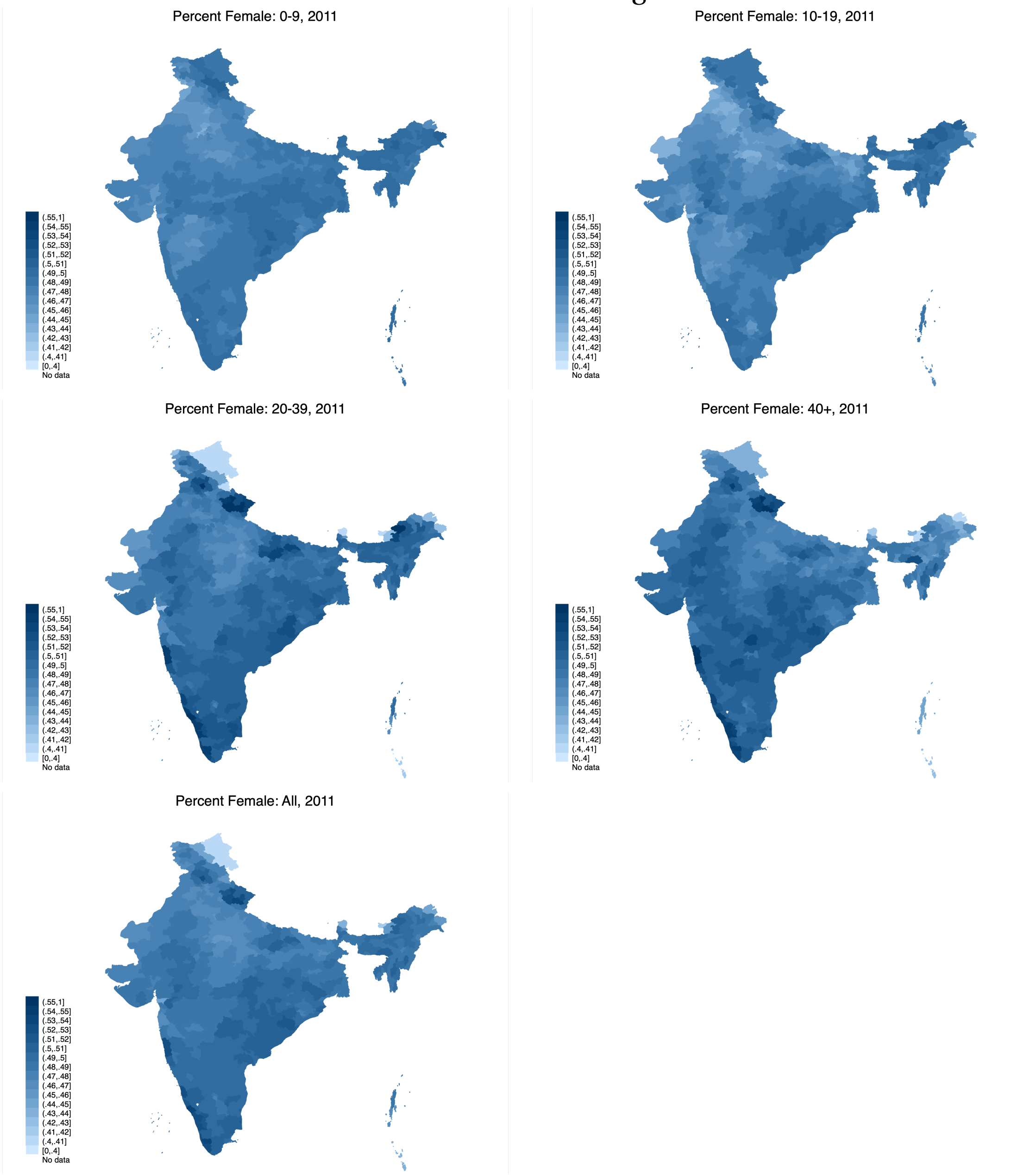
In Table 2, we measure the correlation between past and present sex ratios. We regress the percentage female in a district in 2011 on its percentage female in 1931, by age bin. We report specifications both with and without fixed effects for the states that existed in 2011. To merge districts from 2011 to districts from 1931, we begin by matching each 2011 district to all its sub-districts in the shapefile provided by GADM. We then use our correspondence between the GADM sub-districts and the districts mapped in Chandramouli, Singh and Sethi (2011) to identify the 1931 district from which any 2011 district has been carved. If a 2011 district has a unique match in 1931 (for example, Deoria was entirely within Gorakhpur in 1931) we assign it the percentage female of that 1931 match. If the merge is not unique (for example, Firozabad was created from parts of both Agra and Mainpuri) we create an area-weighted average of the percentages female from all such matches.

Three conclusions are apparent from this table. First, colonial sex ratios predict current sex ratios, even at the district level. In the columns without fixed effects, the r-squared values range from 0.116 in the 40+ age bin to 0.293 in the 10-19 age bin. These are equivalent (taking the square root) to correlation coefficients between .341 and .541. Second, the coefficients and r-squared values are greater at younger ages; persistence in the sex ratio is most pronounced among children and adolescents. Third, the correlations between the colonial and present-day percentages female survive controlling for state fixed effects. It is not only the regions of colonial India that had male-biased sex ratios that continue to have them today – persistence is apparent across districts *within* a given state.

In Figure 4, we show the scatterplots of the raw data underlying these correlations. Two conclusions emerge from these figures. The first is that outliers, rather than driving the correlations in Table 2, undermine them. The points that lie far from the central cloud of points are often frontier districts in regions such as Arunachal Pradesh that

[7] `https://censusindia.gov.in/2011census/C-15.html`

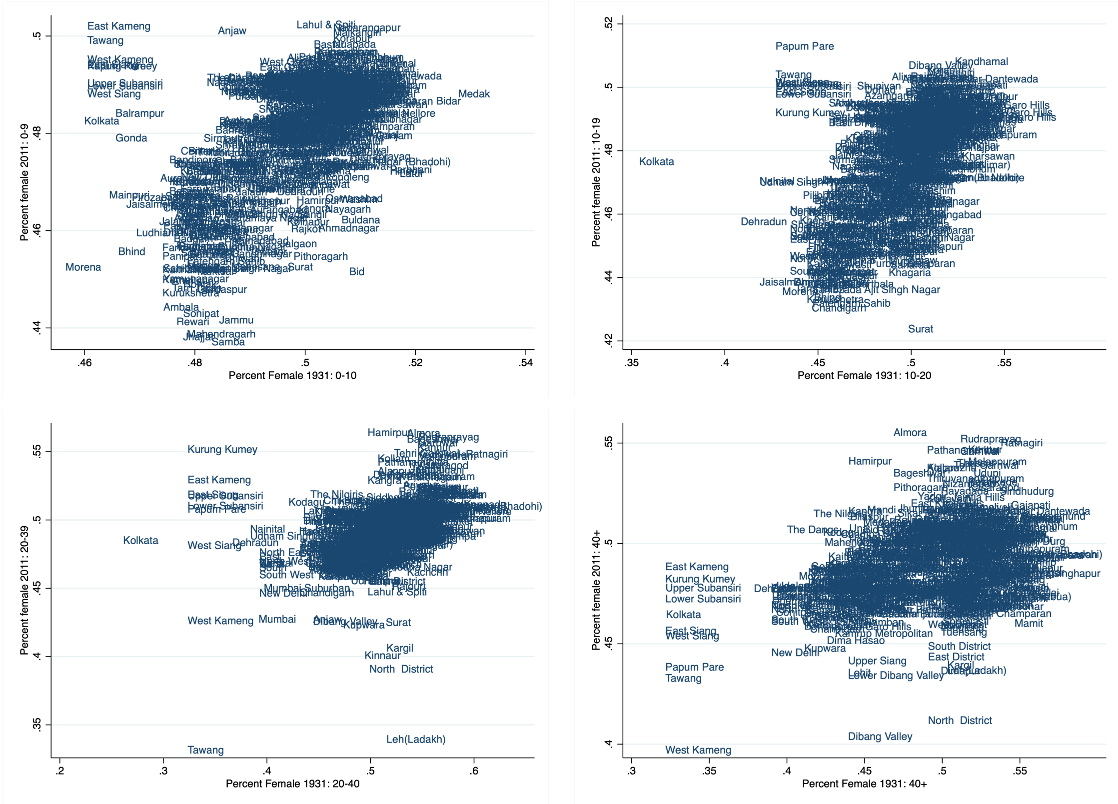
FIGURE 3. Percentage female: 2011



have lower-quality data and male-biased sex ratios in either the colonial or modern periods. Or large cities such as Kolkata in which adult populations were once overwhelmingly male but have since become more balanced. The second conclusion is that it is in the 0-10 and 10-20 age bins that past and present sex ratios are most clearly correlated and the relationship flattens at later ages.

In Figure 5, we extend the analysis from Table 2. For each census year between 1961 and 2011, we regress the percentage female in a district on its percentage female in 1931.

FIGURE 4. Percent female: 2011 versus 1931



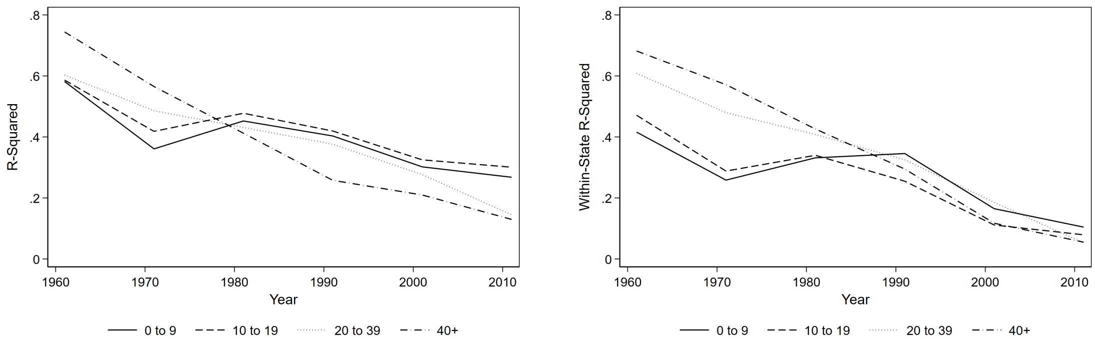
For the years 1961-1991, we use data from Vanneman and Barnes (2000), while for 2001 and 2011 we obtain data directly from the census. We use the same procedure as we did for Table 2 to merge post-colonial districts to those that existed in 1931. In Figure 5 we show the R^2 from a bivariate regression of the percentage female in a given year on the 1931 percentage female. We also run this same regression with state fixed effects, and show the R^2 net of these fixed effects. It is clear from both figures that the importance of colonial sex ratios diminishes over time; while correlation between the 1931 percentage female and the 2011 percentage female that constitutes the result in Table 2 is statistically significant, it is much weaker than the correlations for earlier years.[8]

In sum, our district-level data have allowed us to document the stability of spatial patterns in the sex ratio over more than a century both across and within regions, and to identify colonial-era variation in the sex ratio at a more granular geographic level than has been done previously.

[8] In the Online Appendix, we perform a similar exercise, reporting correlation coefficients for the percentage female across census waves, separately by age bin.

FIGURE 5. R^2 due to 1931 percentage female over time

Without State Fixed Effects / With State Fixed Effects

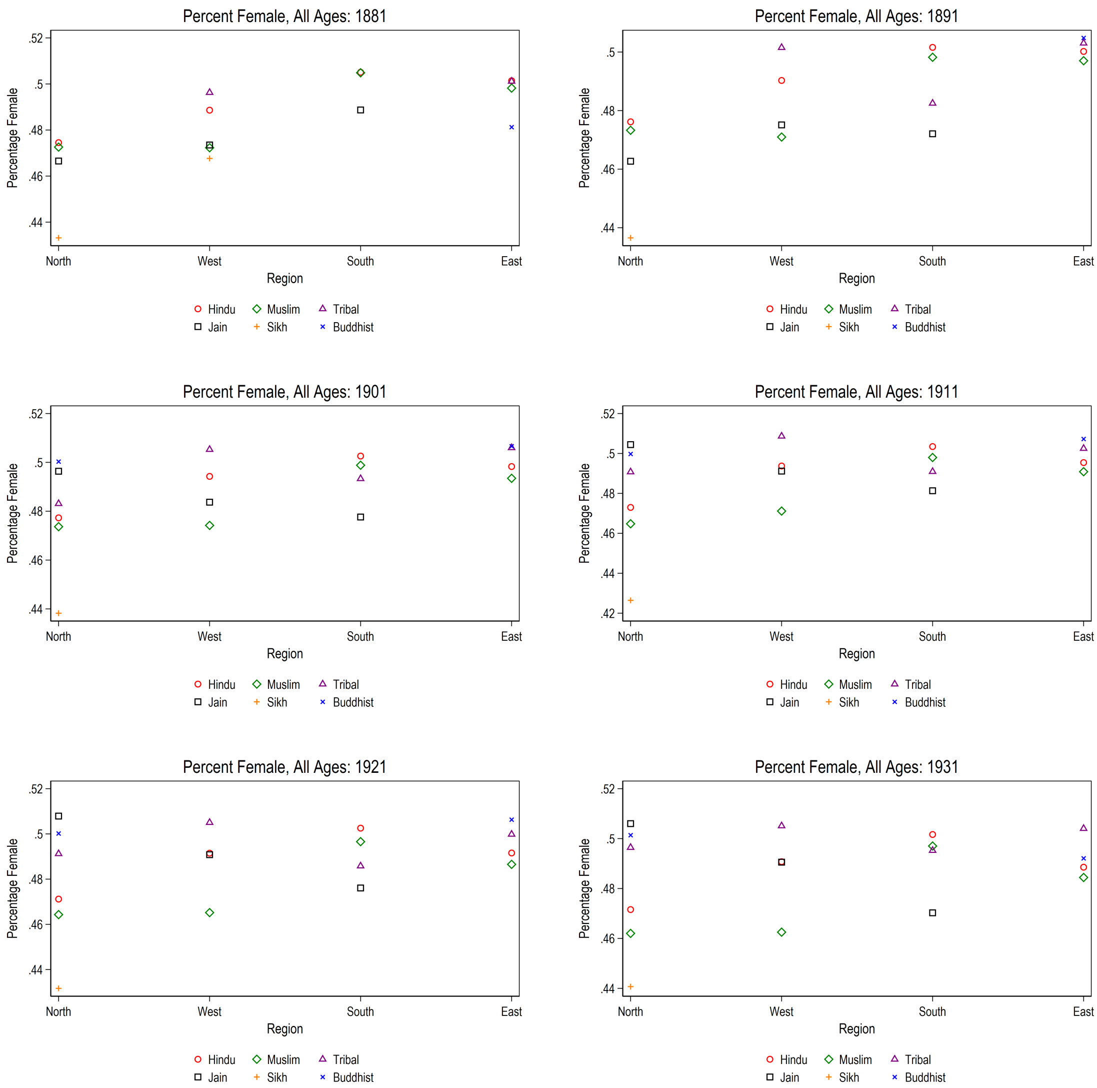


4. RELIGION AND SEX RATIOS IN COLONIAL INDIA

4.1. **Literature.** Religious norms, especially the importance of sons and male relatives in performing the last rites, have been cited as an explanation of son preference in Hindu communities (Jayachandran, 2015; Visaria, 2015). Visaria (2015) suggests that there were no general patterns across religions in India. The female deficit was higher among Muslims compared to Hindus in North India, but not in the East (p 499). Male-biased sex ratios were more prevalent in the higher castes (Chakraborty and Kim, 2010). The census reports documented meticulously the sex ratio by age cohort by region and religion, but rarely commented on Hindu-Muslim differences in sex ratios. Province-level data from the census of 1901, do not show systematic differences in sex ratio across religions, although they show differences across provinces. The sex ratio for Muslims was less biased against females than the sex ratio for Hindus and Sikhs in the North. In Bengal, a province with a large Muslim population, the share of females in the population declined faster among Muslims after age 30 (Census Report of 1901, pp. 238–249).

Recent evidence on sex ratios shows a different picture. Son preference is stronger among Hindus compared to Muslims (Borooah et al., 2009; Borooah and Iyer, 2005; Jayachandran, 2015). While Buddhist sex ratios are less male-biased than Hindu sex ratios in the present, they are more male-biased than those of Muslims and have been growing increasingly unbalanced (Narayan, 2018). Muslim women show lower son preference in surveys (Visaria, 2015). The sex ratio at birth is more balanced among Muslims and the Muslim population have better child survival rates compared to Hindus (Barooah and Iyer, 2005). Bhalotra, Clots-Figueras and Iyer (2021) find that abortion is not widely practised among Muslims and therefore in recent data, Muslims have more equal sex ratios at birth compared to Hindus.

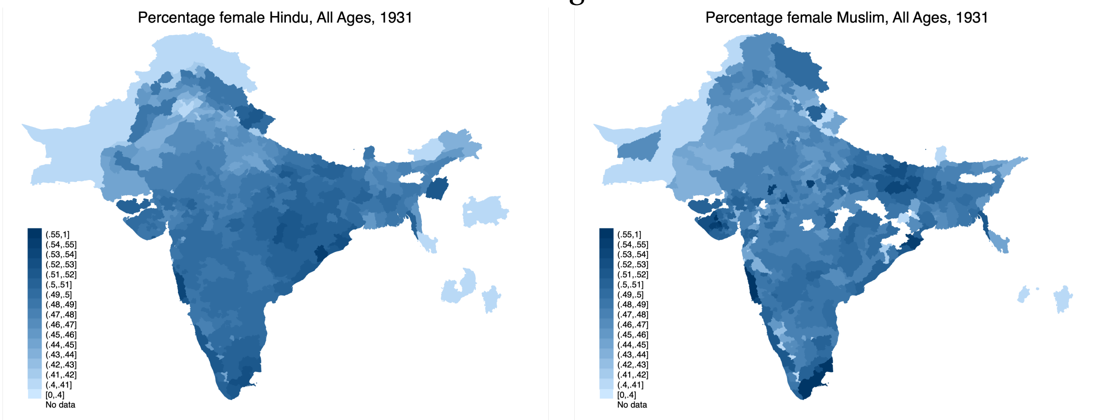
FIGURE 6. Percentage female by region and religion



4.2. **Patterns.** Our data shows that, while there are differences in sex ratios across religions, these differences vary by region. While Sikhs have the most male-biased sex ratios and Buddhists have the most balanced sex ratios, these groups are regionally concentrated. Consider Figure 6, in which we show the percentage of the population that is female by religion and region.[9] While it is clear that there are differences in the share female by religion, the share female for any one religious group often differs markedly by region. In many cases, a group that is more predominantly female than another religious group in one region is instead less female in another region. Within regions, we

[9]We exclude religion-region combinations with fewer than 10,000 persons in these figures. A complete breakdown of the percentage female by religion and region is contained in the Online Appendix.

FIGURE 7. Percentage female: Hindus and Muslims



will show below that religions often exhibited correlations in sex ratio across districts. This is especially true of Hindus and Muslims, the two largest groups. Figure 7 shows the sex ratio by religion for Hindus and Muslims for 1931.[10] Each of these maps show regional variation similar to what we find for the overall population.

In our data, we can compute the sex ratios by religion at the level of districts. In Table 3, we assess the degree to which female percentages in the population are correlated across religious communities in the same district. In particular, using the sample of districts in 1901, we compute correlation coefficients for the percentage female across religions. The top panel (A) reports the correlations for all ages, while the bottom panel (B) reports the correlations for the 0-10 age bin. Note that the number of observations will vary considerably across comparisons. Where a religion is a small minority in any given province, data on the sex composition of that minority are not reported by age, and so we cannot include districts from that province in the comparison. As an extreme example: for Buddhists (found predominantly in Burma) and Sikhs (found predominantly in the Punjab), we have only four observations with which to compute a correlation for the under 10 age bin.

Despite this limitation, a clear patten emerges: in districts where the population composition is male-biased for one religion, the population composition tends to be male-biased for others. The correlation in sex ratio between Hindus and Muslims is high. It is weaker between Hindus and other religious groups. In many cases, this correlation is statistically significant at conventional levels. Consider Hindus and Muslims, India's two largest religious groups. In the under-10 age bin, the correlation in percentage female is just shy of 0.3, and is significant at the 5% level. The Muslim percentage female similarly correlates significantly with the Buddhist percentage female, and the Tribal

[10]The maps of other census years and for other religions are shown in the Online Appendix.

percentage female correlates with the Hindu and Jain percentages female. This broad pattern holds for the "all ages" bin as well.

Because Table 3 represents only a sub-sample of the possible correlations across religions that we can report, we provide several additional tables in the Online Appendix. In these tables, we report the correlations between the percentages female of different religions across districts for each age bin and for each census wave in our data.

The inter-religious correlations across population sex compositions are greater at later ages. Consider the Hindu-Muslim correlation in the percentage female across districts in 1901. This is just below 0.3 in the 0-10 age bin, 0.46 in the 10-20 age bin, 0.78 in the 20-40 age bin, and 0.76 in the over 40 age bin. In 1931, the equivalent correlations are again rising, from 0.13 in the youngest bin to 0.51 in the oldest bin. In 1881, the increase is from 0.22 to 0.74. The Hindu-Sikh and Muslim-Sikh correlations are often not significantly positive in the 0-10 age bin but significant at older ages.

4.3. **Persistence.** In Table 4, we consider whether these interreligious correlations across districts are a feature only of the colonial period. Using data from the 2011 Census of India,[11] we report correlation coefficients analogous to those in Table 3 for the most recent Indian census. The top panel (A) reports these correlations for all ages, while the bottom panel (B) reports correlations for the 0-9 age group. Here too we find significant correlations between the percentages of the population that is female coming from different religious groups in the same district. For Hindus and Muslims, the correlation is slightly above 0.40 for all ages, and just below 0.33 for the 0-10 age bin. Several other correlations are significant and positive at the 5% level, including the Hindu-Buddhist, Hindu-Sikh, Muslim-Christian, Muslim-Sikh, and Sikh-Jain correlations.

In both the historic and modern periods, the fact that sex ratios are correlated for several religious pairs even in the under 10 age bin suggests that the patterns we find are not artifacts of urbanization or of regions in which opportunities for gendered labor led to sex-selective migration. While we only report correlations for the 0-9 and all ages cohorts in the text, we show in the Online Appendix that the inter-religious correlations across districts in 2011 are similar for other age bins as well.

In sum, excepting the very male-skewed sex ratios we find among Sikhs, our results suggest that differences across regions are shared across religions. The common age profile in the sex ratio across religions that we discuss in Section 6.2 and the correlation in sex ratios across religions over space has, we believe, gone largely overlooked in the literature. The striking similarity of Hindu and Muslim sex ratios in the colonial period differs from the present day. In the 2011 census, the percent female in the 0-9 age bin was 47.8% for Hindus and 48.5% for Muslims. This is consistent with the interpretation

[11] `https://censusindia.gov.in/2011census/C-15.html`

in Bhalotra, Clots-Figueras and Iyer (2021) that the Muslim religious aversion to abortion helps explain their less male-biased sex ratios in the present; in the colonial era, sex-selective abortion was not available.

5. Geography and Sex Ratios in Colonial India

5.1. **Literature.** The literature on India has largely focused on possible links between agricultural patterns and the sex ratio. In Section 3, we have discussed the literature that attributes the North-South gradient in sex ratios in part to the relative prevalence of wheat and rice cultivation. While the existing literature on Indian demography and history has focused on agricultural practices as predictors of the sex ratio, economists have tended to focus more narrowly on whether geographic determinants of agricultural practices predict the sex ratio. Measures of geographic suitability for various types of agriculture have the advantage of being largely exogenous to human action, and so allow the researcher to avoid confounding effects of the possible co-determination of agricultural practices and gender outcomes. Prior empirical work has shown that, across countries and across districts in the present day, areas more suited to plough agriculture have more male populations (Alesina, Giuliano and Nunn, 2018; Carranza, 2014). The rice-wheat distinction highlighted by the literature on India does not map cleanly into the use of the plough; Alesina, Giuliano and Nunn (2013) identify both crops as "plough-positive" and, in the 1921 Agricultural Statistics of India, the share of cultivated area planted to either crop correlates positively with the number of ploughs per capita.[12] To our knowledge, other geographic variables that we include have been comparatively neglected in the literature on India – these include ruggedness, presence of a coastline, and suitability for cash crops like tea and cotton.

5.2. **Patterns.**

5.2.1. *Correlations.* In Table 5, we show pairwise correlations between our key geographic variables and the percentage female across districts in the 1901 sample. We do this separately for each of the age bins in our data. Several, but not all, of these correlations are significant at the 5% level. For example, the share female in the 10-20 bin is greater in coastal districts, and the share female is greater in cotton-suitable districts in all age bins. Greater suitability for dryland and wetland rice predict a greater female share across all age bins. Wheat suitability, by contrast, predicts a more male population in all age bins.

[12]By contrast, the share planted to cotton or tea correlates negatively with this measure of plough prevalence. In the 1910 Agricultural Statistics of India, the share of cultivated land that is irrigated correlates negatively across districts with the share of cultivated land planted to rice and positively with the share planted to wheat. The correlation of cotton or tea cultivation with irrigation is insignificant in these data.

Across all age bins, latitude correlates negatively with the share female, reflecting the greater prevalence of male-biased sex ratios in Northern India. In all but the over 40 bin, longitude correlates positively with the female share, reflecting the greater share of girls in eastern India and Burma. Correlations of the share female with malaria vary in sign across age bins, while ruggedness is largely a negative predictor of the share female.

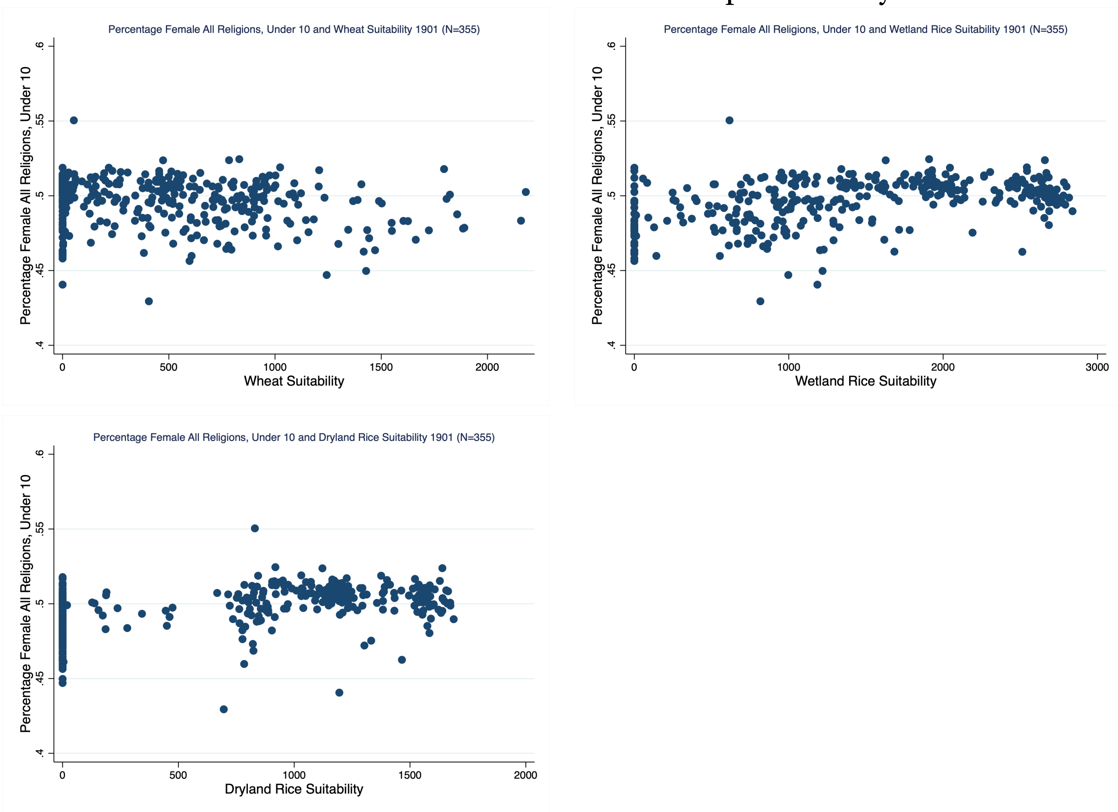
In Figure 8, we demonstrate that these patterns are apparent in the raw data. We present scatterplots of the under-10 sex ratio against three of our key geographic variables in the sample of districts from 1901. These variables are dryland rice suitability, wetland rice suitability, and wheat suitability. Although the correlations are not steep, the positive correlation of the share female with suitability for both types of rice and the negative correlation with suitability for wheat are visible, and clearly not driven by outliers. Because these are only a subset of the possible correlations, we report in the Online Appendix the full set of scatterplots between the percentage female in each age group and census year with each of our geographic variables. Note that, both here and in the Online Appendix, we omit outlier points with female populations outside the range 40-60% from these figures.

While we only present correlations for 1901 in the text, in the Online Appendix we present analogous tables of pairwise correlations for the other years in our data. While not all correlations have the same signs and significance in all years, there are general patterns that emerge. Dryland rice and wetland rice suitability are both positive predictors of the share female across several age bins and census years, while wheat suitability is a negative predictor across many age bins and census years. Because of the limited degree of urbanization and structural change between 1881 and 1931, we do not expect these suitability measures to lose predictive power over time due to declining importance of agriculture.

Latitude correlates negatively, and longitude positively, with the share female in several age bins and census years. Coastal districts have a greater share of younger women in many years. Cotton suitability is an inconsistent predictor of the share female across census years, while tea only positively predicts the presence of younger women from 1901 onwards.

Many of the theories that link the sex ratio to agricultural conditions do so by linking agricultural practices to the economic role of women. We use Table 6 to demonstrate that there is a strong correlation between female labor force participation and the share of the population that is female. We use data from Fenske, Gupta and Yuan (2020), who estimate the female labor force participation rate by dividing the number of women reported as working in the occupational section of the census by the total female population in a district. These data cover the directly-ruled districts of India for the census

FIGURE 8. Correlations with Crop Suitability: Under 10



years between 1901 and 1931. For each census year, we compute the correlation between the population that is female and this measure of women's role in the labor market. In each census year, there is a positive and significant correlation between the two variables.

Similarly, we use Table 6 to show that the share of the population that is female correlates significantly with three other variables of interest – literacy, child marriage, and urbanization. We compute literacy as the share of the total population that is literate using data from Chaudhary and Fenske (2020), child marriage as the percentage of girls under 15 who are married using data from the censuses, and urbanization as the percentage of the population living in cities above 5,000 persons using data from Fenske, Kala and Wei (2021). We do not have literacy or urbanization for all districts and or years. While greater literacy correlates negatively with the share of the population that is female, greater rates of child marriage predict a greater prevalence of women in the population. Cities are disproportionately male.

5.2.2. *Regressions.* Because these pairwise correlations do not account for the likely correlations between the geographic variables we consider, we use Table 7 to report corresponding regression results. For 1901, we regress the percentage of the population that is female on the geographic controls reported in Table 5. We do this separately for each of our major age bins. For ease of interpretation, we standardize the percentage female to have a normal distribution with mean zero and standard deviation one, and we do the same for all continuous geographic controls. Indicators (coastal, river) are unchanged. So: coefficients can be interpreted as standardized betas – how a one standard deviation change in a geographic control predicts changes in the outcome variable, again in standard deviations.

While some of the correlations in Table 5 retain their signs and significance, this is not universally the case. Latitude negatively predicts the share female in all but the oldest age bin, but this is significant only for the youngest bin. Longitude predicts a higher share female in the two youngest age bins. That is: the East-West gradient in South Asia emerges as a predictor of sex ratios as robust as the North-South gradient when controlling for other observable geographic characteristics of districts.

The correlation between the coastal dummy and the percentage female in the 10-20 age bin is no longer significant. Correlations with ruggedness are no longer significant. Malaria suitability now becomes a negative predictor of the female share in several age bins. Cotton suitability correlates with a greater share female from adolescence onwards. Wheat suitability is no longer a significant predictor of sex ratios. Dryland rice suitability, but not wetland rice suitability, significantly predicts a greater share female in all but one age bin.

In the Online Appendix, we show the robustness of these correlations to controlling for the religious composition of the population. That is, treating Hindus as the omitted category, we control for the share of the population that is Buddhist, Jain, Muslim, Sikh, Tribal, and Other. Recall that, because not all religions are separately enumerated in all districts, we are compelled to code their populations as zero when they are not reported, and so their inclusion in the Other category will introduce measurement error. However, as this only occurs where a religious group is a small share of the population, this error should be slight.

Controlling for religion in this way, the most major change is that the coefficients on longitude fall in sign and significance in several columns; longitude may be a proxy for the large Buddhist populations of the East. Other changes are more minor. Latitude still correlates negatively with the under-10 percent female in 1901. Malaria suitability remains a significant predictor, but only at ages 20 and above. The negative correlation with coastal remains in the Over 40 group but becomes marginally insignificant for all ages. Cotton suitability remains a positive predictor in the 10-20 and 20-40 age bins. The predictive capacity of dryland rice suitability remains.

The conditional correlations reported in Table 7 are based off comparisons of all districts in the data, and so may be confounded by unobserved determinants of sex ratios that vary across broad regions of South Asia. In Table 8, then, we add province fixed effects to our regressions. That is, the regression coefficients reported in this table now capture comparisons across districts *within* the provinces of colonial India, rather than comparisons across provinces. Because fixed effects can exacerbate attenuation bias due to measurement error, we do expect some coefficients to become insignificant due to this loss of precision.

Despite this caveat, there are meaningful patterns that emerge in Table 8. First, latitude continues to predict lower under-10 female shares, even conditional on province fixed effects and other geographic observables. The correlation between longitude and the share female, however, weakens, becoming significant only at the 10% level in one column and losing significance in the others; the East-West gradient in sex ratios is a gradient that exists across broad regions, more than within them. Malaria suitability continues to correlate negatively with the share female. Dryland rice suitability remains a positive predictor in all but one age bin.

While suitability measures have become widely used in economics due to their exogeneity, they may not reflect actual cultivation practices. Using data from the 1921 Agricultural Statistics of India, we replace our suitability measures of rice, wheat, cotton, and tea with actual shares of cultivated land planted to these crops. We report these results in the Online Appendix. Our general conclusion that the importance of rice and wheat has been overstated in the literature remains. Rice acreage correlates negatively with the under-10 share female and positively with the 10-20 share female; neither correlation remains significant with province fixed effects or controlling for religion. In only one specification is the partial correlation between the share female and wheat cultivation significant.

For space, we have only reported regression results with geographic covariates for 1901. In the Online Appendix, we report equivalent results for all census years. Across years, greater latitudes predict lower shares female, particularly at younger ages.[13] Longitude only becomes a regular predictor of greater female shares from 1901 onwards. From 1911 onwards, this becomes significant at the 5% level even with province fixed effects.

To summarize, these correlations and regression results confirm that Northern India is different; the share female at younger ages correlates negatively with latitude. Eastern India has higher shares female, particularly in the later years of our sample. Malaria predicts fewer women in 1901 and after. Cotton suitability is a positive, but inconsistent,

[13]Removing Punjab from the data (not reported) does not lead to large qualitative changes in this pattern.

predictor of the share of the population that is female. Wheat suitability, by contrast, is a sometimes negative but usually inconsistent predictor of the female share. Wetland rice suitability is a poor predictor of the female share. Dryland rice suitability is often correlated positively with the share of the population that is female.

Existing claims in the literature have often been made on the basis of broad regional comparisons and, in the colonial period, have often been made without data on either the cultivation of or the suitability for the key crops of rice and wheat, and without adjusting for possible confounders. Once these are adjusted for, the wheat-rice dichotomy may be overstated, particularly for wetland rice. By contrast, the relevance of other factors, including cotton and malaria, has been relatively overlooked.

6. Age and Sex Ratios in Colonial India

6.1. **Age and Sex Ratios.**

6.1.1. *Literature.* In this section, we address the question "when do when go missing"? We describe how sex ratios evolve across age bins, by region and by religion. Dyson (2018) provides estimates that life expectancy in the territory of today's India rose from 26.3 to 29.6 over the period 1891 to 1931, while for women the comparable increase was from 27.2 to 30.1.[14] Unlike the evidence from recent data, which finds the sex ratio to be unnaturally male biased at birth in some regions, the evidence from colonial censuses shows a male deficit in the age group 0-1. This became a female deficit by ages 5-10 (Bhaskar and Gupta, 2007). Life expectancy at birth moved in favour of males only from 1921, as mortality declined faster for males than females (Mayer, 1999; Visaria and Visaria, 1983). Although the regional differences in overall sex ratios have been documented (Visaria and Visaria, 1983), we know little about regional variations in sex ratios by age cohorts and about how these variations change over time. To the best of our knowledge, our is the first paper to document changes in sex ratios by age group for all of colonial India and thereby throw light on the question of missing women over the life cycle.

Bardhan (1974) cites maternal mortality as one explanation of regional differences in female survival rates in India in the 1950s and 1960s; in states like Uttar Pradesh and Rajasthan, female marriage ages were lower than the national average, and fertility was higher. This would lead us to expect women to go missing in the 10-20 age bin. Evidence on maternal mortality rates in colonial India are relatively sparse. Most women gave birth at home, particularly in rural India and were cared for by family members or traditional midwives (Muraleedharan, 1994). The high risk of dying at childbirth in the Madras Presidency was described by an Indian commentator in 1866 in dramatic

[14]See Tables 7.1 and 8.1 in Dyson (2018).

language: "mothers would eat a specially prepared death meal before going into labour, in the presence of her family and friends" (Lang, 2005). Lang points out that British officials laid the blame on ill-trained traditional midwives, and training midwives in modern methods became a goal. The Countess of Dufferin Fund was created in 1885 to provide maternity services and, in 1914, the colonial government accepted this as a policy. However, the use of these services remained limited (Jat et al., 2015).

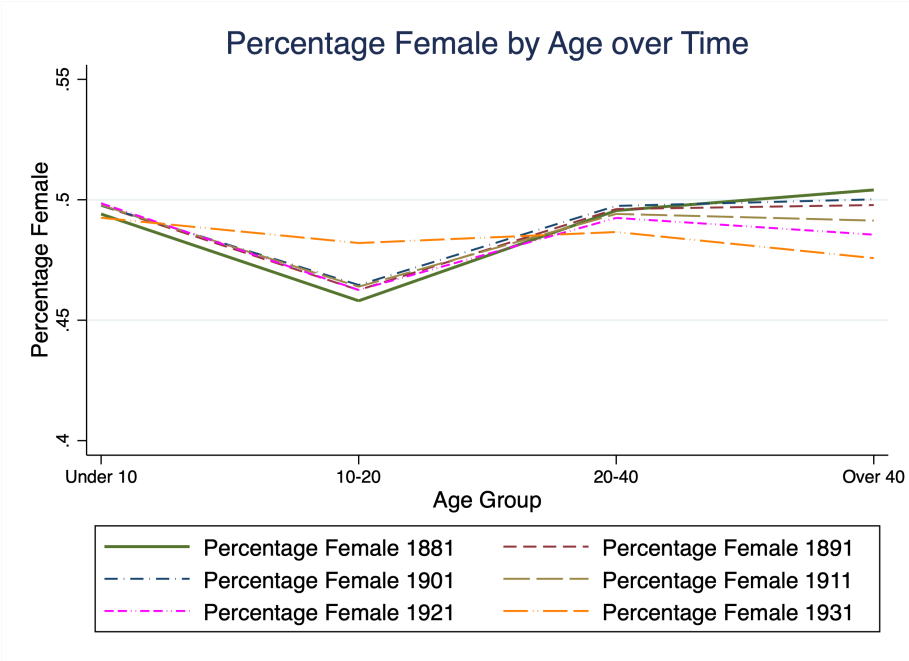
An investigation of 16 municipalities in the Madras Presidency in the early 1930s revealed a maternal mortality rate of 15.4 per 1,000, more than three times the rate in England and Wales at that time. The overall female death rate between ages 15 and 40 was 14.11 per 1,000, versus 12.37 for men (Nair, 2011, p. 244). An investigation of more than 7,000 hospital confinements for childbirth in Madras, Madura, Trichinopoly and Coimbatore in 1927 and 1928 found similar rates of 17.89 maternal deaths per 1,000 births (Muraleedharan, 1994). This was comparable to 16 per 1,000 in Bombay hospitals between 1925 and 1929 (Balfour and Talpade, 1930). Most births were at home, without access to the medical interventions available in hospitals. We would therefore expect the average maternal mortality to be higher than what was reported in the urban hospitals. The evidence from the hospital found the age of the mother to be an important factor in maternal mortality as well as neonatal mortality. For every 1000 births, mortality was 51 in Madras Presidency if the mother's age was under 15 and 31-33 for mothers aged 15-25. Comparable figures for the United States in 1921 as cited in the report were 20 and 6.8, respectively. The study concluded that maternal mortality was highest in the case of first birth for mothers under the age of 15. Poor nutrition of the mother contributed to high maternal and infant mortality.

6.1.2. *Patterns.* In Figure 9, we show the fraction of the population that is female for each of our four age bins, in each of our six census years. Note that we aggregate these numbers from our district-level data, so they will differ slightly from comparable totals in the census reports, as will the numbers by age and religion below.

Two immediate patterns are apparent. The first is that the female deficit in the population emerges most strikingly after age 10. While even in the 0-10 cohort, the percentage of the population that is female is universally less than 50%, in the 10-20 cohort it falls below 47% in all census years except for 1931. In the cohort aged 20-40, the sex ratio reverted back towards the 50% mark, while remaining male biased. In the early census waves, the percentage female for the cohorts aged over 40 is slightly less male biased than that for the cohort aged 20-40. In the years 1911 and later, sex ratios again become more male biased in this over 40 cohort.

Two broad patterns can account for these trends in colonial censuses by age: differences in mortality across sexes by age, and differential misreporting of age by gender.

FIGURE 9. Percentage female by age and census year



The deficit of females in the population under 10 was discussed in various censuses and the enumerators concluded that, while infanticide of girls was practised in isolated communities in Punjab, Sind, Rajasthan and Cutch, this was not of a magnitude that could impact the overall sex ratio and the mostly likely reason for the female deficit was a lack of care for girls and, therefore, higher female mortality in this age group (Report of Census of 1901, p116). The female deficit occurring in adolescence is likely to result in part from maternal mortality in a context where the median age of marriage in 1911 was 12.9 years (Bhat and Halli, 1999, p. 137). The reduction in the female deficit in the working age population can be attributed in part to the demographic evidence from famines in colonial India that male mortality was higher than female mortality in the working-age population (Dyson, 1991; Lardinois, 1985). In part it was due to reporting biases.

Dyson (1989, p. 165) notes that the colonial censuses had some “female age shifting into the reproductive span,” i.e. women misstating their age towards their prime reproductive years. There was under reporting of unmarried girls in the age-group 9-15 and of married women of 15-20 (Census of 1901, p. 115). This can account for some of the deficit at ages 15-19. Note that in-migration of men for employment can explain a deficit of women in adolescence in some districts, but not over the whole of India. The Census report of 1901 confirms that there were significant reporting issues with respect to adolescent females. Although we are not aware of systematic data at the level of districts for all provinces, data for age groups spanning five years is reported for selected provinces. These show that the female deficit rose sharply in the age group 10-20 and reversed in the age group 20-30. These changes reflect in part maternal mortality and in part misreporting. In comparison to more recent censuses, this pattern is quite different. The number of missing girls at birth has risen since the 1970s, with the availability of prenatal sex selection technology. Anderson and Ray (2010) and Anderson and Ray (2012) find a share of missing women in older age groups.

The second pattern that is apparent is that the changes in sex ratios across cohorts becomes dampened over time. In each successive census wave, the drop in the percentage female between the 0-10 and 10-20 cohorts is less pronounced. In 1931, the age profile of the share female is visibly flatter than in other census years. These changes over time likely reflect a weakening of the processes noted above – disproportionate female mortality in adolescence, and underreporting of female adolescents.[15]

It is unlikely that a reduction in child marriage in 1931 explains the flattening of the age profile in 1931. Dyson (2018) estimates that the mean age of marriage for women was 12.7 in both 1921 and 1931.[16] Roy and Tam (2021) have shown that the 1929 announcement of the Child Marriage Restraint Act *raised* rates of child marriage in anticipation of its coming into force six months later. Anecdotal evidence suggests this rush to marry girls was urgent enough that some high-caste girls married lower-caste men and men with disabilities (Gupta, 2014).

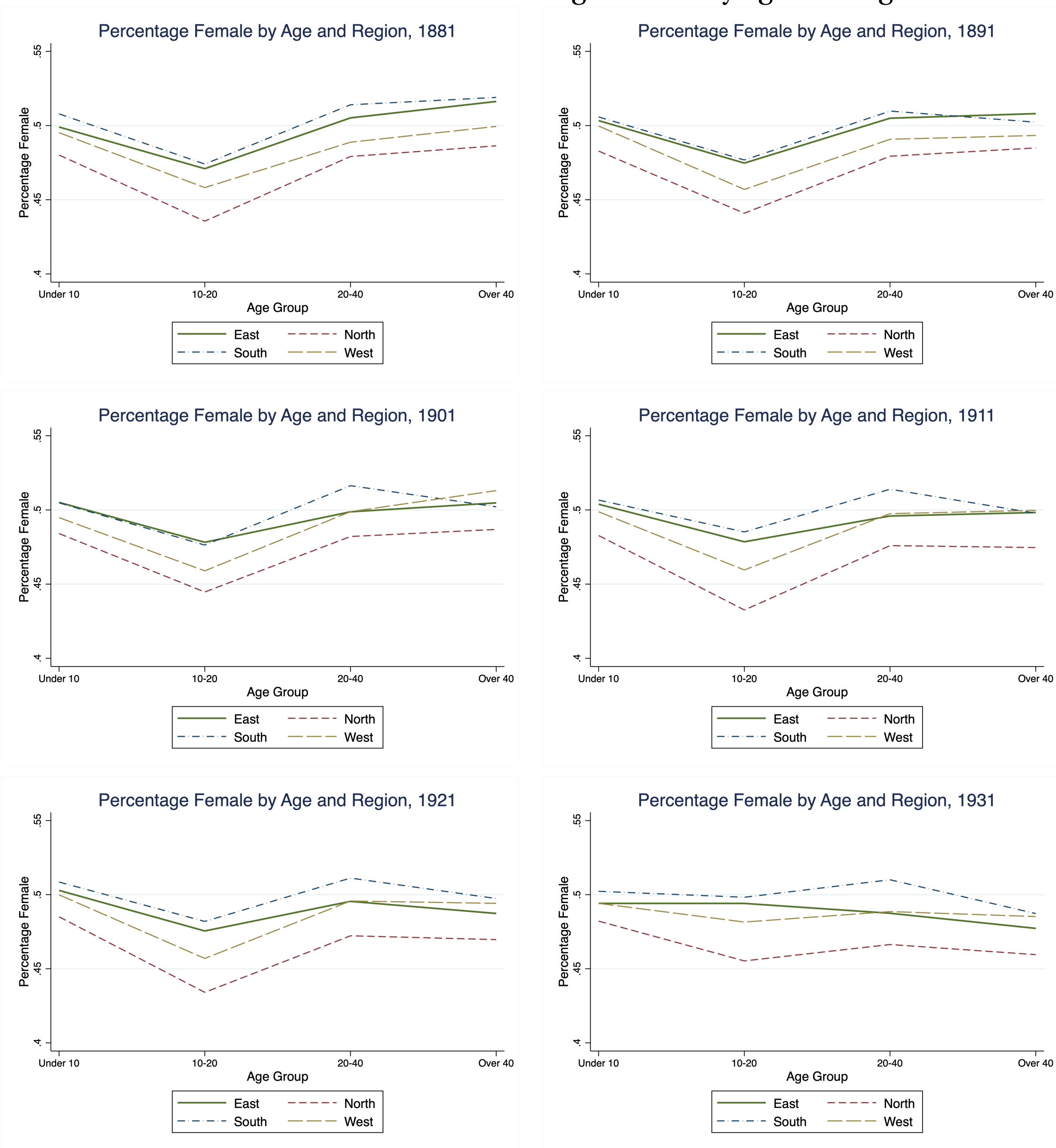
In Figure 10, we present figures analogous to Figure 9, but we separate our sample into the four broad regions we consider – East, North, South and West. Here, three broad patterns are apparent. First, the general trend of sex ratios becoming more male-biased in the 10-20 age bin is not confined to any one region of India. Before 1931, a reduction in the share female between the 0-10 and 10-20 cohorts is visible in all four regions. It is still present in all four in 1931, though it has become slight in the Eastern region.

Second, there are pronounced differences across regions. It is in the North that sex ratios are most male-biased, and in the South where they are least male-biased. Indeed, in some years both the 0-10 and 20-40 cohorts are greater than 50% female in the Southern region. The reduction in the share of the population that is female between the 20-40 and over 40 cohorts is apparent in several census years in both the North and South, but not in Western and Eastern India.

[15]If we restrict the 1931 sample to only those districts present in 1921 (not reported), the dampening is very similar and so is not driven by a change in sample composition. The major changes in composition are the Central India Agency and Western India States Agency being divided into more districts in 1931 than in 1921.

[16]See table 8.1 in Dyson (2018).

FIGURE 10. Percentage female by age and region



Third, the dampening of the differences across cohorts in 1931 is not driven by only one region alone. The reduction in the share female between the 0-10 and 10-20 cohorts, as well as the subsequent increase in the share female in the 20-40 cohort are less pronounced in all four of the major regions.

6.1.3. *Persistence.* A striking finding we have noted is that the low female population share in the 10-20 age bin is common across regions; this, we believe, has largely gone overlooked. This is consistent with a role for child marriage and maternal mortality in explaining the age-specific sex ratio even in regions where the overall sex ratio is relatively more balanced. The North shows a large share of missing women over all ages. The pattern in missing women over the life cycle in colonial India shows some differences from that in the present, arising in part from changes in the availability of medical intervention in pregnancy. Key among these are prenatal sex selection and wider access to hospital birth. In recent years, a significant proportion of missing women are in older age cohorts (Calvi, 2020). Anderson and Ray (2010) and Anderson and Ray (2012), for example, find that relative mortality due to cardiovascular disease in older women is higher than in developed countries.

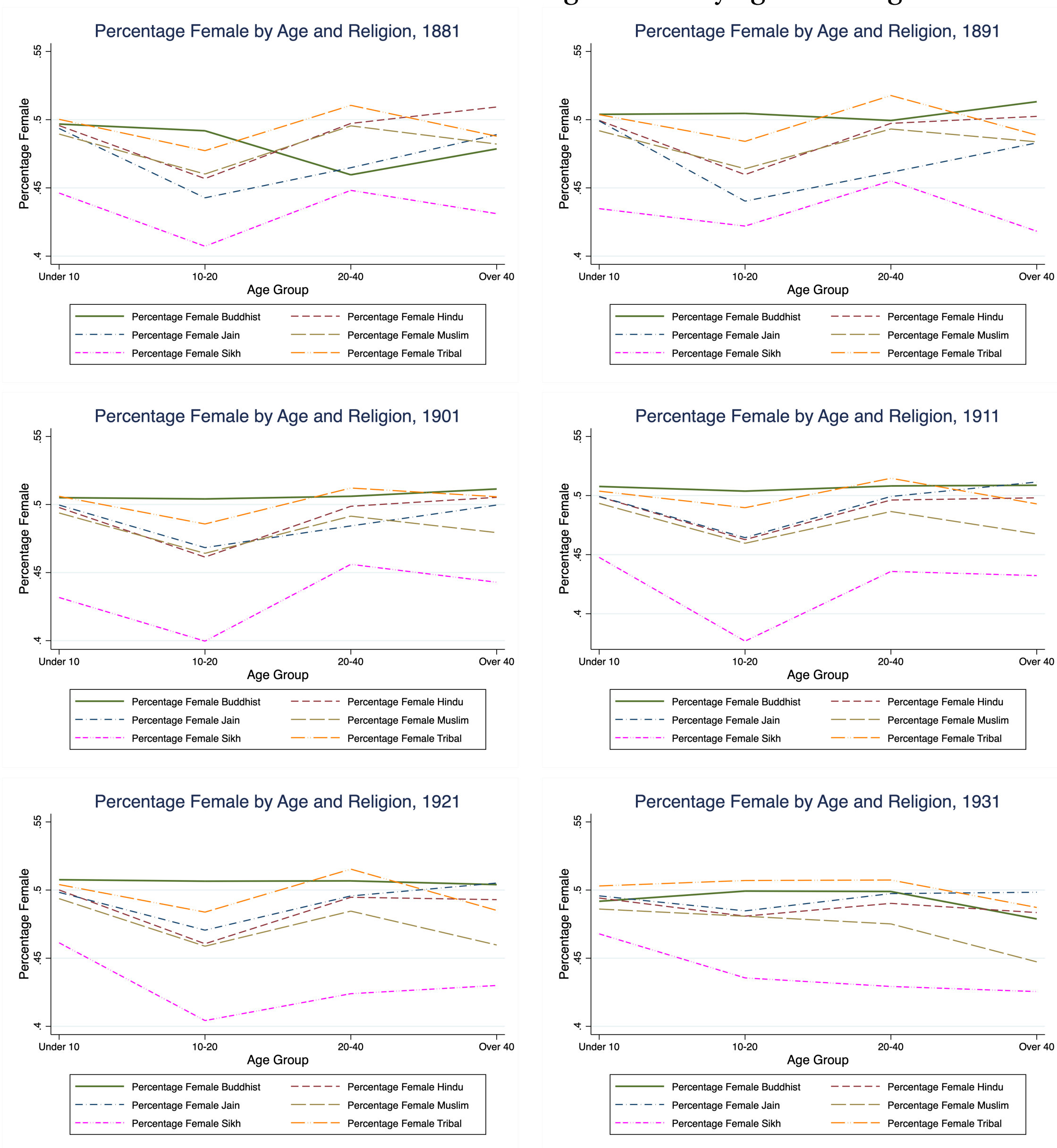
6.2. **Age and Sex Ratios by Religion.**

6.2.1. *Patterns.* In Figure 11, we present graphs similar to Figure 10, except that we now separate the sex ratio trends across age bins by religion, rather than by region. A number of key patterns stand out. The first is that there are clear differences in sex ratios across religions that are apparent throughout the life cycle. Sikhs have, in all census waves, exceptionally male populations in all age bins. In 1881 and 1911, for example, the percentage female is 45% or less in all age bins. Hindus, Muslims, and Jains all have male-biased sex ratios, while for Tribal populations and Buddhists, sex ratios are less male biased. Indeed, for Buddhists, the share female is often greater than 50%. This is in start contrast, then, to their male-biased sex ratios in the present.

Second, the general pattern of a lower share female in the 10-20 age bin than in either the 0-10 or 20-40 bin is not driven by any single religion. In several years, this characteristic dip in the 10-20 age bin is visible for Sikhs, Hindus, Muslims, and Jains. Third, the dampening of the differences in the share female across age bins in 1931 is driven by a similar dampening among Jains, Muslims and Hindus that is not visible for Sikhs. For Hindus, the reduced share of women in the over 40 age bin becomes more pronounced in 1931. For Sikhs, this becomes prevalent enough in 1931 that the share female in the over 40 population is even lower than in the adolescent age bin.

Third, although the dip in the share of females in the age group 10-20 is reversed in the age group 20-40 in all religious groups, the pattern differs across religions after age 40. The share of the female population declined for Sikhs, Muslims, and Tribals. While male-biased sex ratios are present in early stages of the life cycle for Muslims and Sikhs, for Tribals this appears to be the case only in later life. Although life expectancy at birth began to rise for males from 1921, male mortality was higher in the older age groups. Higher mortality for men from famines and from diseases such as tuberculosis may explain more favourable sex ratios for females after age 40. Mortality was high among adult males of working age in the nineteenth century famines (Dyson, 1991). Dyson attributes this to greater mobility of men and to the lower likelihood of pregnancy during

FIGURE 11. Percentage female by age and religion



famines, which reduced maternal mortality. In data from the 1970s, male mortality was higher than female mortality above the age of 35, due in part to greater prevalence of tuberculosis among men (Padmanabha, 1982).

7. CONCLUSION

In this paper, we have introduced two novel data sets covering sex ratios in colonial South Asia between 1881 and 1931. We have shown that the familiar pattern of more male-biased sex ratios in Northern India and more balanced sex ratios in Southern India and modern Burma go back more than a century, and are visible in granular, district-level data. Across districts, colonial sex ratios correlate strongly with present-day sex ratios, particularly at younger ages, and even comparing districts within the same modern state. Sex ratios generally became more male-biased in adolescence, a pattern that is visible across several regions, religions, and census waves. Sex ratios are most male biased among Sikhs, but the fraction female is less than 50% among Hindus, Muslims, and Jains. Female shares correlate across religious groups over space, a pattern that is also visible in modern data. While female shares correlate with the suitability for rice and wheat cultivation, many of these correlations become weaker when adjusting for confounders and looking within, rather than across, regions of South Asia.

The existing literature has provided many explanations of male-biased sex ratios in India, and our descriptive results provide an opportunity to reevaluate these theories in light of evidence from the colonial period. While many of these theories are rooted in a preference for sons, this preference is only a proximate cause of missing women and requires explanation. Some proposed causes, such as fertility decline, are unique to recent decades, and so cannot account for sex ratios in India's past. Other explanations are, however, relevant to both the past and present.

The persistence we find in sex ratios, not only across broad regions but within states, is indicative that slowly changing, "deep" determinants of sex ratios do indeed matter. This persistence is, however, strongest at younger ages and has been declining over time, suggesting that newer causes are needed to fully account for missing women at older ages in the present. Other patterns we find are also suggestive of changes in the causes of missing women over the twentieth century, including the dampening of the female deficit in adolescence, the relatively male-skewed sex ratios for Muslims in colonial India, and the relatively more equal sex ratios for Buddhists in the past.

We can broadly group explanations from the literature into four sets. The first set of explanations considers religion and social norms. The literature has stressed religious reasons for Hindu son preference, and the particularly skewed sex ratios among the upper castes. We do not find, however, that differences across regions are the same in all regions. While the male-biased sex ratios we find among Sikhs and Hindus are consistent with these explanations, it is clear that both Muslim and Buddhist sex ratios have diverged from those in the colonial era. One explanation is that attitudes towards the modern technology of abortion are critical in understanding modern outcomes, but not colonial patterns. That we find strong correlations over space across religions, at younger age bins, in both the past and present, suggests that either there are common determinants of the sex ratio across religions, that there are spillovers in social norms across religions, or both. That the deficit of women in adolescence is common to several religions suggests they share more causes of the sex ratio than is usually supposed.

The second set of explanations considers kinship and marital residence patterns, notably the greater prevalence of patrilocal residence and lower ages of marriage in the North. Although we do find this regional divide in colonial data, we show that sex ratios are much more balanced in the 0-10 age group than in the 10-20 age group, which suggests that mortality in adolescence is quantitatively more important than neglect of girls in the colonial data. Further, this pattern of missing women in adolescence is found in several regions, regardless of the prevailing type of kinship and marital residence. Maternal mortality, then, is a problem everywhere.

The third set of explanations considers the economic role of women. Much of this literature focuses on the cultivation of rice and wheat, and its consequences for female labor force participation, as a possible explanation for the North-South gradient in sex ratios. Related work stresses the importance of deep tillage. If women are valued primarily for their fertility rather than their economic contribution, their bargaining power may be lower in post-reproductive ages. Our evidence here has been mixed. We do find that sex ratios correlate with female labor force participation. We do not, however, find that either cropping patterns or their deeper determinants are robust predictors of colonial sex ratios. This is particularly true for wheat and for wetland rice.

The fourth set of explanations is those that are absent from the literature. We have found a number of robust correlates of sex ratios, including suitability for malaria, suitability for cotton, and the presence of a major river, that have gone largely overlooked in conventional accounts. Existing explanations are, then, likely to be incomplete.

There are, however, a number of limitations to our approach. Because we use the censuses of colonial India as our underlying source, we are unable to push our estimates back in time further than the second half of the nineteenth century. There will be gaps in our coverage due to the gaps in the original sources; not all districts are covered in all years, not all religions are covered in all districts, and the inconsistencies in the age ranges reported in the census compel us to use more aggregate age bins in our analysis than would be ideal. Further, any deficiencies of the original data collection process, such as the misreporting of women by age, will affect our data. It is our hope that later historians and demographers will uncover additional sources that will fill the gaps in our data, and will develop techniques permitting them to produce plausible corrections for the errors in the original sources.

It is also our hope that these data will be of use to future researchers. Economists and economic historians have posited that the prevalence of missing women depends on a number of variables, including agricultural practices, weather shocks, price shocks, cultural change, and economic development (e.g. Alesina, Giuliano and Nunn (2013); Jayachandran (2017); Qian (2008); Rose (1999)). We have introduced new data that we expect will be used by later researchers to examine the external validity of these hypotheses, as well as novel hypotheses, in a broad historical setting.

References

Acemoglu, Daron, David H Autor and David Lyle. 2004. "Women, war, and wages: The effect of female labor supply on the wage structure at midcentury." *Journal of Political Economy* 112(3):497–551.

Agnihotri, Satish, Richard Palmer-Jones and Ashok Parikh. 2002. "Missing women in Indian districts: a quantitative analysis." *Structural Change and Economic Dynamics* 13(3):285–314.

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.

Alesina, Alberto, Paola Giuliano and Nathan Nunn. 2018. "Traditional agricultural practices and the sex ratio today." *PLOS One* 13(1):e0190510.

Alkema, Leontine, Fengqing Chao, Danzhen You, Jon Pedersen and Cheryl C Sawyer. 2014. "National, regional, and global sex ratios of infant, child, and under-5 mortality and identification of countries with outlying ratios: a systematic assessment." *The Lancet Global Health* 2(9):e521–e530.

Alsan, Marcella. 2015. "The effect of the tsetse fly on African development." *American Economic Review* 105(1):382–410.

Anderson, Siwan. 2018. "Legal origins and female HIV." *American Economic Review* 108(6):1407–39.

Anderson, Siwan and Debraj Ray. 2010. "Missing women: age and disease." *The Review of Economic Studies* 77(4):1262–1300.

Anderson, Siwan and Debraj Ray. 2012. "The age distribution of missing women in India." *Economic and Political Weekly* pp. 87–95.

Ashraf, Nava, Natalie Bau, Nathan Nunn and Alessandra Voena. 2020. "Bride price and female education." *Journal of Political Economy* 128(2):591–641.

Austin, Gareth. 2008. "The 'reversal of fortune' thesis and the compression of history: perspectives from African and comparative economic history." *Journal of International Development: The Journal of the Development Studies Association* 20(8):996–1027.

Balfour, Margaret I and Shakuntala K Talpade. 1930. "The maternity conditions of women mill-workers in India." *The Indian medical gazette* 65(5):241–249.

Bardhan, Pranab K. 1974. “On life and death questions.” *Economic and Political Weekly* 9(32/34):1293–1304.

Basu, Alaka Malwade. 1999. “Fertility decline and increasing gender imbalance in India, including a possible South Indian turnaround.” *Development and Change* 30(2):237–263.

Behrman, Jere R, Andrew D Foster, Mark R Rosenweig and Prem Vashishtha. 1999. “Women’s schooling, home teaching, and economic growth.” *Journal of Political Economy* 107(4):682–714.

Beltrán Tapia, Francisco J and Domingo Gallego-Martínez. 2020. “What explains the missing girls in nineteenth-century Spain?” *The Economic History Review* 73(1):59–77.

BenYishay, Ariel, Maria Jones, Florence Kondylis and Ahmed Mushfiq Mobarak. 2020. “Gender gaps in technology diffusion.” *Journal of Development Economics* 143:102380.

Bhalotra, Sonia, Irma Clots-Figueras and Lakshmi Iyer. 2021. “Religion and abortion: The role of politician identity.” *Journal of Development Economics* 153:102746.

Bhalotra, Sonia, Rachel Brulé and Sanchari Roy. 2018. “Women’s inheritance rights reform and the preference for sons in India.” *Journal of Development Economics* 146:102275.

Bharadwaj, Prashant and Leah K Lakdawala. 2013. “Discrimination begins in the womb: evidence of sex-selective prenatal investments.” *Journal of Human Resources* 48(1):71–113.

Bhaskar, V and Bishnupriya Gupta. 2007. “India’s missing girls: biology, customs, and economic development.” *Oxford Review of Economic Policy* 23(2):221–238.

Bhat, PN Mari and Shiva S Halli. 1999. “Demography of brideprice and dowry: Causes and consequences of the Indian marriage squeeze.” *Population Studies* 53(2):129–148.

Boehnke, Jörn and Victor Gay. 2020. “The Missing Men: World War I and Female Labor Force Participation.” *Forthcoming, Journal of Human Resources* .

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli and Andrei Shleifer. 2019. “Beliefs about gender.” *American Economic Review* 109(3):739–73.

Borooah, Vani K, Quy-Toan Do, Sriya Iyer and Shareen Joshi. 2009. “Missing women and India’s religious demography.” *World Bank Policy Research Working Paper* (5096).

Borooah, Vani K and Sriya Iyer. 2005. “Religion, literacy, and the female-to-male ratio.” *Economic and Political Weekly* 40(5):419–427.

Boserup, Ester. 1970. *Woman’s role in economic development.* Routledge.

Cagé, Julia and Valeria Rueda. 2020. “Sex and the mission: The conflicting effects of early Christian investments on the HIV epidemic in sub-Saharan Africa.” *Journal of*

Demographic Economics 86(3):213–257.

Calvi, Rossella. 2020. "Why are older women missing in India? The age profile of bargaining power and poverty." *Journal of Political Economy* 128(7):2453–2501.

Carranza, Eliana. 2014. "Soil endowments, female labor force participation, and the demographic deficit of women in India." *American Economic Journal: Applied Economics* 6(4):197–225.

Cervellati, Matteo, Elena Esposito and Uwe Sunde. 2021. "Epidemic shocks and civil violence: Evidence from malaria outbreaks in Africa." *The Review of Economics and Statistics* pp. 1–46.

Chakraborty, Tanika and Sukkoo Kim. 2010. "Kinship institutions and sex ratios in India." *Demography* 47(4):989–1012.

Chandavarkar, Rajnarayan. 2003. *The Origins of Industrial Capitalism in India: Business strategies and the working classes in Bombay, 1900-1940.* Cambridge University Press.

Chandramouli, C, A.P. Singh and R.C. Sethi. 2011. *Census of India 2011: Administrative Atlas of India.* Office of the Registrar General & Census Commissioner, India, Ministry of Home Affairs, Government of India.

Chaudhary, Latika and James Fenske. 2020. "Did railways affect literacy? Evidence from India." *CAGE working paper no. 529* .

Clark, Shelley. 2000. "Son preference and sex composition of children: Evidence from India." *Demography* 37(1):95–108.

Collins, William J. 1999. "Labor mobility, market integration, and wage convergence in late 19th century India." *Explorations in Economic History* 36(3):246–277.

Croson, Rachel and Uri Gneezy. 2009. "Gender differences in preferences." *Journal of Economic Literature* 47(2):448–74.

Das Gupta, Monica. 1987. "Selective discrimination against female children in rural Punjab, India." *Population and Development Review* 13(1):77–100.

Das Gupta, Monica and P.N. Mari Bhat. 1997. "Fertility decline and increased manifestation of sex bias in India." *Population Studies* 51(3):307–315.

De Moor, Tine and Jan Luiten Van Zanden. 2010. "Girl power: The European marriage pattern and labour markets in the North Sea region in the late medieval and early modern period." *The Economic History Review* 63(1):1–33.

Dimico, Arcangelo, Alessia Isopi and Ola Olsson. 2017. "Origins of the Sicilian Mafia: The market for lemons." *The Journal of Economic History* 77(4):1083–1115.

Doepke, Matthias, Michele Tertilt and Alessandra Voena. 2012. "The economics and politics of women's rights." *Annual Review of Economics* 4(1):339–372.

Drevenstedt, Greg L, Eileen M Crimmins, Sarinnapha Vasunilashorn and Caleb E Finch. 2008. "The rise and fall of excess male infant mortality." *Proceedings of the National*

Academy of Sciences 105(13):5016–5021.

Duflo, Esther. 2012. "Women empowerment and economic development." *Journal of Economic Literature* 50(4):1051–79.

Dupraz, Yannick. 2019. "French and British colonial legacies in education: Evidence from the partition of Cameroon." *The Journal of Economic History* 79(3):628–668.

Dyson, Tim. 1989. The Historical Demography of Berar, 1881-1980. In *India's Historical Demography*, ed. Tim Dyson. London: Curzon Press Ltd. chapter 6, pp. 150–196.

Dyson, Tim. 1991. "On the demography of South Asian famines: Part I." *Population Studies* pp. 5–25.

Dyson, Tim. 2018. *A population history of India: From the first modern people to the present day.* Oxford University Press.

Dyson, Tim and Mick Moore. 1983. "On kinship structure, female autonomy, and demographic behavior in India." *Population and Development Review* 9(1):35–60.

Fenske, James, Bishnupriya Gupta and Song Yuan. 2020. "Demographic Shocks and Women's Labor Market Participation: Evidence from the 1918 Influenza Pandemic in India." *CEPR Discussion Paper No. DP15077* .

Fenske, James and Namrata Kala. 2021. "Linguistic Distance and Market Integration in India." *The Journal of Economic History* 81(1):1–39.

Fenske, James, Namrata Kala and Jinlin Wei. 2021. "Railways and Cities in India." *CAGE working paper no. 559* .

Fernandez, Raquel. 2007. "Women, work, and culture." *Journal of the European Economic Association* 5(2-3):305–332.

Fernández, Raquel. 2013. "Cultural change as learning: The evolution of female labor force participation over a century." *American Economic Review* 103(1):472–500.

Fernández, Raquel, Alessandra Fogli and Claudia Olivetti. 2004. "Mothers and sons: Preference formation and female labor force dynamics." *The Quarterly Journal of Economics* 119(4):1249–1299.

Fogli, Alessandra and Laura Veldkamp. 2011. "Nature or nurture? Learning and the geography of female labor force participation." *Econometrica* 79(4):1103–1138.

Giuliano, Paola. 2014. "Female labor force participation: Persistence and evolution." *The New Palgrave Dictionary of Economics, Edited by Steven N. Durlauf and Lawrence E. Blume* .

Goldin, Claudia. 1995. "The U-shaped female labor force function." *"Economic Development and Economic History" in Investment in Women's Human Capital, TP Schultz (ed.) Chicago: The University of Chicago Press* .

Goldin, Claudia and Claudia Olivetti. 2013. "Shocking labor supply: A reassessment of the role of World War II on women's labor supply." *American Economic Review*

103(3):257–62.

Goldin, Claudia D. 1991. "The role of World War II in the rise of women's employment." *The American Economic Review* 81(4):741–756.

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.

Guilmoto, Christophe Z, Nandita Saikia, Vandana Tamrakar and Jayanta Kumar Bora. 2018. "Excess under-5 female mortality across India: a spatial analysis using 2011 census data." *The Lancet Global Health* 6(6):e650–e658.

Gupta, Bishnupriya. 2014. "Where have all the brides gone? Son preference and marriage in India over the twentieth century." *The Economic History Review* 67(1):1–24.

Gupta, Bishnupriya. 2021. *An Economic History of India: From the Mughal Empire to Modern Times.* Book Manuscript. Under contract with Cambridge University Press.

Gupta, Bishnupriya and Anand V Swamy. 2017. "Reputational consequences of labor coercion: evidence from Assam's tea plantations." *Journal of Development Economics* 127:431–439.

Gupta, Monica Das. 2005. "Explaining Asia's "missing women": a new look at the data." *Population and Development Review* 31(3):529–535.

Harnetty, Peter. 1971. "Cotton exports and Indian agriculture, 1861-1870." *The Economic History Review* 24(3):414–429.

Jat, Tej Ram, Prakash R Deo, Isabel Goicolea, Anna-Karin Hurtig and Miguel San Sebastian. 2015. "Socio-cultural and service delivery dimensions of maternal mortality in rural central India: a qualitative exploration using a human rights lens." *Global Health Action* 8(1):24976.

Jayachandran, Seema. 2015. "The roots of gender inequality in developing countries." *Annual Review of Economics* 7(1):63–88.

Jayachandran, Seema. 2017. "Fertility decline and missing women." *American Economic Journal: Applied Economics* 9(1):118–39.

Jayachandran, Seema and Ilyana Kuziemko. 2011. "Why do mothers breastfeed girls less than boys? Evidence and implications for child health in India." *The Quarterly Journal of Economics* 126(3):1485–1538.

Jayachandran, Seema and Rohini Pande. 2017. "Why are Indian children so short? The role of birth order and son preference." *American Economic Review* 107(9):2600–2629.

Jha, Saumitra. 2013. "Trade, institutions, and ethnic tolerance: Evidence from South Asia." *American Political Science Review* 107(4):806–832.

Kishor, Sunita. 1993. ""May God give sons to all": gender and child mortality in India." *American Sociological Review* 58(2):247–265.

Kiszewski, Anthony, Andrew Mellinger, Andrew Spielman, Pia Malaney, Sonia Ehrlich Sachs and Jeffrey Sachs. 2004. "A Global Index Representing the Stability of Malaria Transmission." *The American Journal of Tropical Medicine and Hygiene* 70(5):486–498.

Korn, Evelyn. 2000. "On the formation of family structures." *Public Choice* 105(3):357–372.

Lang, Sean. 2005. "Drop the demon dai: Maternal mortality and the state in colonial Madras, 1840–1875." *Social History of Medicine* 18(3):357–378.

Lardinois, Roland. 1985. "Famine, epidemics and mortality in South India: A Reappraisal of the Demographic Crisis of 1876-1878." *Economic and Political Weekly* pp. 454–465.

Lowes, Sara. 2020. "Matrilineal kinship and spousal cooperation: Evidence from the matrilineal belt." *Working Paper* .

Luke, Nancy and Kaivan Munshi. 2011. "Women as agents of change: Female income and mobility in India." *Journal of Development Economics* 94(1):1–17.

Mayer, Peter. 1999. "India's falling sex ratios." *Population and Development Review* 25(2):323–343.

Michalopoulos, Stelios and Elias Papaioannou. 2013. "Pre-colonial ethnic institutions and contemporary African development." *Econometrica* 81(1):113–152.

Miller, Barbara D. 1981. *The endangered sex: neglect of female children in rural north India.* Cornell University Press.

Miller, Barbara D. 1982. "Female labor participation and female seclusion in rural India: A regional view." *Economic Development and Cultural Change* 30(4):777–794.

Morris, Morris David. 1965. *The emergence of an industrial labor force in India.* University of California Press.

Muraleedharan, VR. 1994. "Diet, Disease and Death in Colonial South India." *Economic and Political Weekly* 29(1/2):55–63.

Nair, Rahul. 2011. "The construction of a 'population problem' in Colonial India 1919–1947." *The Journal of Imperial and Commonwealth History* 39(2):227–247.

Narayan, Swati. 2018. "Religion and Female–Male Ratios in India." *Indian Journal of Human Development* 12(3):441–452.

Niederle, Muriel and Lise Vesterlund. 2007. "Do women shy away from competition? Do men compete too much?" *The Quarterly Journal of Economics* 122(3):1067–1101.

Nunn, Nathan. 2014. Gender and missionary influence in colonial Africa. In *African development in historical perspective*, ed. Emmanuel Akyeampong, Robert Bates and James A Robinson. Cambridge, UK: Cambridge University Press pp. 489–512.

Nunn, Nathan and Diego Puga. 2012. "Ruggedness: The Blessing of Bad Geography in Africa." *The Review of Economics and Statistics* 94(1):20–36.

Ojha, Shubhankita. 2014. “Regulating Work: Decasualisation of Dock Labour in Colonial India.” *Social Scientist* 42(3/4):77–89.

Padmanabha, P. 1982. “Mortality in India: a note on trends and implications.” *Economic and Political Weekly* pp. 1285–1290.

Pande, Rohini P. 2003. “Selective gender differences in childhood nutrition and immunization in rural India: the role of siblings.” *Demography* 40(3):395–418.

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.

Rose, Elaina. 1999. “Consumption smoothing and excess female mortality in rural India.” *Review of Economics and Statistics* 81(1):41–49.

Rosenzweig, Mark R and T Paul Schultz. 1982. “Market opportunities, genetic endowments, and intrafamily resource distribution: Child survival in rural India.” *The American Economic Review* 72(4):803–815.

Roy, Sutanuka and Eddy HF Tam. 2021. “Impact of British Colonial Gender Reform on Early Female Marriages and Gender Gap in Education: Evidence from Child Marriage Abolition Act, 1929.” *Working Paper, University of Oxford* .

Satya, Laxman D. 1997. *Cotton and famine in Berar, 1850-1900.* Manohar Publishers and Distributors.

Scott, James C. 2009. *The art of not being governed.* Yale University Press.

Sen, Amartya. 1990. “More than 100 million women are missing.” *The New York review of books* 37(20):61–66.

Sudha, S and S Irudaya Rajan. 1999. “Female demographic disadvantage in India 1981–1991: Sex selective abortions and female infanticide.” *Development and change* 30(3):585–618.

Teso, Edoardo. 2019. “The long-term effect of demographic shocks on the evolution of gender roles: Evidence from the transatlantic slave trade.” *Journal of the European Economic Association* 17(2):497–534.

Vanneman, Reeve and Douglas Barnes. 2000. *Indian District Data, 1961-1991: machine-readable data file and codebook.* College Park, Maryland: Center on Population, Gender, and Social Inequality.

Visaria, Abhijit. 2015. Religion and son preference in India and Bangladesh: Three essays on comparing Hindus and Muslims on son preference and sex differentials in child health. PhD thesis University of Pennsylvania.

Visaria, Leela and Pravin Visaria. 1983. V - Population (1757–1947). In *The Cambridge Economic History of India. Volume 2: c.1757-c.1970,* ed. Dharma Kumar and Meghnad Desai. Cambridge, UK: Cambridge University Press pp. 463–532.

Voth, Hans-Joachim. 2020. "Persistence: myth and mystery." *Forthcoming, Bisin and Federico, Handbook of Historical Economics* .

Xue, Melanie Meng. 2018. "High-Value Work and the Rise of Women: The Cotton Revolution and Gender Equality in China." *Working Paper: New York University* .

Table 1. Percent Female by Region

Year	Region	Age Group	Religion	Total Population	Female Population	Percent Female	Districts
1881	All	All	All	218,513,488	107,036,368	0.490	313
1881	East	All	All	77,879,456	38,901,016	0.500	85
1881	North	All	All	68,014,832	32,153,816	0.473	122
1881	South	All	All	34,585,484	17,469,608	0.505	34
1881	West	All	All	38,033,720	18,511,924	0.487	72
1891	All	All	All	261527392	128323552	0.491	327
1891	East	All	All	87,442,432	43,659,020	0.499	99
1891	North	All	All	73,386,912	34,794,292	0.474	109
1891	South	All	All	55,436,664	27,776,536	0.501	55
1891	West	All	All	45,261,376	22,093,702	0.488	64
1901	All	All	All	288,797,920	141,924,672	0.491	366
1901	East	All	All	95183440	47371448	0.498	108
1901	North	All	All	83,109,440	39,568,864	0.476	129
1901	South	All	All	59,158,796	29,698,648	0.502	44
1901	West	All	All	51,346,248	25,285,704	0.492	85
1911	All	All	All	310,586,272	151,784,928	0.489	380
1911	East	All	All	103,952,560	51,486,344	0.495	109
1911	North	All	All	89319328	41892736	0.469	144
1911	South	All	All	65,072,468	32,707,190	0.503	64
1911	West	All	All	52,241,904	25,698,652	0.492	63
1921	All	All	All	313,544,544	152,601,360	0.487	408
1921	East	All	All	106,821,520	52,536,576	0.492	115
1921	North	All	All	87,600,008	40,971,624	0.468	139
1921	South	All	All	66,393,796	33,296,438	0.501	65
1921	West	All	All	52,729,204	25,796,720	0.489	89
1931	All	All	All	348,890,624	169,314,256	0.485	506
1931	East	All	All	118,339,632	57,857,312	0.489	114
1931	North	All	All	96,779,920	45,213,448	0.467	142
1931	South	All	All	74,184,472	37,161,520	0.501	66

Notes: These figures sum populations over all districts reporting civil condition by age in a given census year.

Table 2. Correlations Between Percentages Female in 1931 and 2011

	Percent Female 0-9		Percent Female 10-19	
Percent Female 1931: 0-10	0.565***	0.365***		
	(0.057)	(0.056)		
Percent Female 1931: 10-20			0.404***	0.234***
			(0.037)	(0.052)
Observations	625	625	625	625
R-squared	0.269	0.708	0.302	0.625
State FE	No	Yes	No	Yes
	Percent Female 20-39		Percent Female 40+	
Percent Female 1931: 20-40	0.215***	0.168***		
	(0.041)	(0.048)		
Percent Female 1931: 40+			0.166***	0.120***
			(0.018)	(0.028)
Observations	625	625	625	625
R-squared	0.146	0.410	0.131	0.640
State FE	No	Yes	No	Yes

Notes: Observations are at the 2011 district level. *** Significant at 1%. ** Significant at 5%. * Significant at 10%. Robust standard errors in parentheses.

Table 3. Correlations across Religions in 1901

Panel A. All Ages

	Perc. Fem. Buddhist	Perc. Fem. Hindu	Perc. Fem. Jain	Perc. Fem. Muslim	Perc. Fem. Sikh
Perc. Fem. Hindu	-0.1975				
P-Value	0.1693				
No. of Observations	50				
Perc. Fem. Jain	0.0400	0.8679			
P-Value	0.9745	0.0000			
No. of Observations	3	88			
Perc. Fem. Muslim	0.1732	0.7309	0.3554		
P-Value	0.2339	0.0000	0.0007		
No. of Observations	49	338	88		
Perc. Fem. Sikh	0.7016	0.2761	0.1144	0.2495	
P-Value	0.2984	0.0158	0.4821	0.0298	
No. of Observations	4	76	40	76	
Perc. Fem. Tribal	0.0182	0.5001	0.6942	0.4536	0.2955
P-Value	0.9102	0.0000	0.0000	0.0000	0.1516
No. of Observations	41	143	45	135	25

Panel B. Population aged 0-10

	Perc. Fem. Buddhist	Perc. Fem. Hindu	Perc. Fem. Jain	Perc. Fem. Muslim	Perc. Fem. Sikh
Perc. Fem. Hindu	0.0365				
P-Value	0.8031				
No. of Observations	49				
Perc. Fem. Jain	-1.0000	0.1824			
P-Value	1.0000	0.0948			
No. of Observations	2	85			
Perc. Fem. Muslim	0.6689	0.2872	0.1587		
P-Value	0.0000	0.0000	0.1468		
No. of Observations	48	338	85		
Perc. Fem. Sikh	0.9185	0.0938	-0.1602	0.2241	
P-Value	0.0815	0.4332	0.3654	0.0584	
No. of Observations	4	72	34	72	
Perc. Fem. Tribal	-0.1583	0.1944	0.5146	0.1286	0.1258
P-Value	0.3359	0.0205	0.0004	0.1386	0.5972
No. of Observations	39	142	43	134	20

Notes: Observations are at the contemporaneous district level. P values of 0.0000 are less than 0.0001.

Table 4. Correlations across Religions in 2011

Panel A. All Ages

	Pct. Female: All	Pct. Female: Buddhist	Pct. Female: Christian	Pct. Female: Hindu	Pct. Female: Jain	Pct. Female: Muslim
Pct. Female: Buddhist	0.1221					
P-Value	0.0020					
No. of Observations	640					
Pct. Female: Christian	0.3894	0.3664				
P-Value	0.0000	0.0000				
No. of Observations	640	640				
Pct. Female: Hindu	0.3945	0.2919	0.5727			
P-Value	0.0000	0.0000	0.0000			
No. of Observations	640	640	640			
Pct. Female: Jain	0.0659	0.1104	0.0923	0.1884		
P-Value	0.0961	0.0052	0.0197	0.0000		
No. of Observations	638	638	638	638		
Pct. Female: Muslim	0.3597	-0.0609	0.2080	0.4074	0.0550	
P-Value	0.0000	0.1240	0.0000	0.0000	0.1654	
No. of Observations	640	640	640	640	638	
Pct. Female: Sikh	0.1439	0.0924	0.3237	0.4913	0.1597	0.4880
P-Value	0.0003	0.0194	0.0000	0.0000	0.0001	0.0000
No. of Observations	640	640	640	640	638	640

Panel B. Population aged 0-9

	Pct. Female: All	Pct. Female: Buddhist	Pct. Female: Christian	Pct. Female: Hindu	Pct. Female: Jain	Pct. Female: Muslim
Pct. Female: Buddhist	0.0516					
P-Value	0.1940					
No. of Observations	636					
Pct. Female: Christian	0.3713	-0.0186				
P-Value	0.0000	0.6401				
No. of Observations	640	636				
Pct. Female: Hindu	0.8357	-0.0646	0.2853			
P-Value	0.0000	0.1034	0.0000			
No. of Observations	640	636	640			
Pct. Female: Jain	0.0038	0.0228	0.0659	0.0013		
P-Value	0.9235	0.5690	0.0982	0.9740		
No. of Observations	631	627	631	631		
Pct. Female: Muslim	0.4425	0.0411	0.2667	0.3293	0.1197	
P-Value	0.0000	0.3012	0.0000	0.0000	0.0026	
No. of Observations	640	636	640	640	631	
Pct. Female: Sikh	0.0903	0.0141	0.0945	0.1070	0.0863	0.0374
P-Value	0.0231	0.7230	0.0174	0.0071	0.0311	0.3481
No. of Observations	633	630	633	633	625	633

Notes: Observations are at the 2011 district level. P values of zero correspond to p<0.00005.

Table 5. Correlations of Percentage Female with Geographic Variables in 1901

	Under 10	10-20	20-40	Over 40	All Ages
Latitude	-0.4181	-0.3880	-0.1903	-0.1824	-0.2391
p-value	0.0000	0.0000	0.0003	0.0005	0.0000
No. of Observations	357	357	357	357	357
Longitude	0.3182	0.4567	0.0506	0.0073	0.1411
p-value	0.0000	0.0000	0.3406	0.8906	0.0076
No. of Observations	357	357	357	357	357
Costal	0.1022	0.1177	-0.0737	-0.1120	-0.0536
p-value	0.0537	0.0262	0.1649	0.0344	0.3122
No. of Observations	357	357	357	357	357
Ruggedness	-0.0538	-0.1028	-0.1924	-0.2348	-0.1946
p-value	0.3107	0.0522	0.0003	0.0000	0.0002
No. of Observations	357	357	357	357	357
Major River	0.0094	0.0679	0.0089	-0.0143	0.0347
p-value	0.8590	0.2007	0.8676	0.7871	0.5135
No. of Observations	357	357	357	357	357
Malaria Suitability	0.1330	0.2275	-0.1066	-0.1339	-0.0220
p-value	0.0119	0.0000	0.0441	0.0113	0.6789
No. of Observations	357	357	357	357	357
Cotton Suitability	0.1204	0.2109	0.2693	0.2619	0.2722
p-value	0.0229	0.0001	0.0000	0.0000	0.0000
No. of Observations	357	357	357	357	357
Wheat Suitability	-0.1941	-0.2053	-0.1228	-0.1817	-0.1547
p-value	0.0002	0.0001	0.0203	0.0006	0.0034
No. of Observations	357	357	357	357	357
Dryland Rice Suitability	0.4266	0.4682	0.1459	0.1699	0.2322
p-value	0.0000	0.0000	0.0057	0.0013	0.0000
No. of Observations	357	357	357	357	357
Wetland Rice Suitability	0.3758	0.4617	0.1070	0.0241	0.1794
p-value	0.0000	0.0000	0.0434	0.6499	0.0007
No. of Observations	357	357	357	357	357
Tea Suitability	0.1513	0.2547	-0.0581	-0.1645	0.0027
p-value	0.0042	0.0000	0.2735	0.0018	0.9595
No. of Observations	357	357	357	357	357

Notes: Observations are at the contemporaneous district level. P values of 0.0000 are less than 0.0001.

Table 6. Correlation of Percentage Female with Female Labour Force Participation, Literacy, Child Marriage, and Urbanization

	Percentage Female: All Ages					
	1881	1891	1901	1911	1921	1931
Female Labour Force Participation			0.4064	0.4896	0.4409	0.3129
p			0.0000	0.0000	0.0000	0.0000
N			191	198	198	197
Literacy	-0.3444	-0.4912	-0.4398	-0.4131	-0.4417	
p	0.0000	0.0000	0.0000	0.0000	0.0000	
N	188	188	195	202	200	
Child Marriage	0.2054	0.1493	0.1425	0.1290	0.1466	0.1670
p	0.0003	0.0074	0.0071	0.0127	0.0030	0.0002
N	309	321	356	372	408	506
Urbanization	-0.2137	-0.1952	-0.2032	-0.2901	-0.1952	-0.2648
p	0.0002	0.0004	0.0001	0.0000	0.0001	0.0000
N	290	320	357	380	408	506

Notes: Observations are at the contemporaneous district level. p=0.0000 reported if p<0.0005.

Table 7. Correlates of Percentage Female: Regression Results for 1901

	Under 10	10-20	20-40	Over 40	All Ages
Latitude	-0.346**	-0.0890	-0.000879	0.120	-0.00771
	(0.137)	(0.140)	(0.156)	(0.146)	(0.168)
Longitude	0.232**	0.388***	0.101	0.197	0.159
	(0.0965)	(0.0947)	(0.156)	(0.148)	(0.164)
Ruggedness	-0.0541	-0.00761	-0.0171	-0.0730	-0.0322
	(0.174)	(0.191)	(0.163)	(0.157)	(0.192)
Malaria Suitability	-0.152***	-0.183**	-0.278**	-0.336***	-0.245**
	(0.0520)	(0.0759)	(0.116)	(0.0928)	(0.114)
Major River	0.180	0.278**	0.358**	0.372**	0.395***
	(0.115)	(0.127)	(0.156)	(0.144)	(0.153)
Coastal	-0.129	0.0161	-0.207	-0.280*	-0.236
	(0.101)	(0.107)	(0.180)	(0.169)	(0.175)
Cotton Suitability	-0.00718	0.198***	0.301**	0.299**	0.299**
	(0.0688)	(0.0739)	(0.138)	(0.131)	(0.145)
Wheat Suitability	0.119	-0.148	-0.207	-0.221	-0.198
	(0.136)	(0.132)	(0.157)	(0.147)	(0.166)
Wetland Rice Suitability	-0.00676	-0.0830	0.00261	-0.158	-0.0587
	(0.116)	(0.122)	(0.118)	(0.106)	(0.120)
Dryland Rice Suitability	0.227***	0.259***	0.157	0.375***	0.238***
	(0.0722)	(0.0763)	(0.0973)	(0.0903)	(0.0877)
Tea Suitability	-0.0137	0.0946*	-0.0487	-0.135*	-0.0132
	(0.0487)	(0.0530)	(0.0936)	(0.0816)	(0.0890)
Observations	357	357	357	357	357
R-squared	0.264	0.357	0.176	0.249	0.201

Notes: Observations are at the contemporaneous district level. *** Significant at 1%. ** Significant at 5%. * Significant at 10%. Robust standard errors in parentheses. All regressions contain a constant that is not reported.

Table 8. Correlates of Percentage Female: Regression Results for 1901 with Province Fixed Effects.

	Under 10	10-20	20-40	Over 40	All Ages
Latitude	-0.583**	-0.316	-0.201	-0.165	-0.280
	(0.286)	(0.318)	(0.260)	(0.242)	(0.293)
Longitude	0.445*	0.318	0.334	0.187	0.327
	(0.251)	(0.259)	(0.236)	(0.233)	(0.248)
Ruggedness	-0.0750	-0.0770	-0.0805	-0.177	-0.139
	(0.237)	(0.271)	(0.208)	(0.193)	(0.247)
Malaria Suitability	-0.103*	-0.165*	-0.283***	-0.363***	-0.269***
	(0.0586)	(0.0899)	(0.107)	(0.0805)	(0.102)
Major River	0.241**	0.253*	0.356**	0.285**	0.344**
	(0.115)	(0.136)	(0.152)	(0.135)	(0.144)
Coastal	-0.107	0.0448	-0.0875	-0.0562	-0.0506
	(0.0921)	(0.0879)	(0.121)	(0.118)	(0.0967)
Cotton Suitability	-0.0245	0.156**	0.172*	0.0686	0.117
	(0.0771)	(0.0752)	(0.0908)	(0.0783)	(0.0833)
Wheat Suitability	0.261	-0.0490	-0.142	0.0429	-0.0230
	(0.225)	(0.248)	(0.217)	(0.194)	(0.238)
Wetland Rice Suitability	-0.127	-0.138	-0.0435	-0.199	-0.119
	(0.152)	(0.163)	(0.137)	(0.125)	(0.149)
Dryland Rice Suitability	0.159*	0.197**	0.112	0.356***	0.181*
	(0.0966)	(0.0984)	(0.120)	(0.113)	(0.108)
Tea Suitability	-0.00586	0.126*	-0.0406	-0.157**	-0.0171
	(0.0607)	(0.0683)	(0.0990)	(0.0781)	(0.0824)
Observations	357	357	357	357	357
R-squared	0.410	0.460	0.445	0.514	0.500

Notes: Observations are at the contemporaneous district level. *** Significant at 1%. ** Significant at 5%. * Significant at 10%. Robust standard errors in parentheses. All regressions contain a constant that is not reported.