The Paradox of Seclusion: Regional Differences in Female Employment and Wages in Urban India*

Leila Gautham +

December 2025

Abstract

Why are urban gender wage gaps lower in northern than in southern states of India, despite greater gender equality (in non-wage dimensions) in the south? I show that this is due to greater suppression of women's low-wage employment in the north, resulting in stronger positive selection: selection-corrected gaps that impute wages for the non-employed based on observed and unobserved characteristics are similar for both north and south. I suggest that stronger social norms in the north that stigmatize women's wage work produce lower employment rates, particularly among less-educated, low-wage women who do not have access to white-collar jobs. These patterns of participation introduce significant selection biases in the measurement of gender wage gaps and help explain why urban gender wage gaps are lower in the north.

Keywords: gender wage inequality; female employment; selection; gender norms.

JEL Classification: J16, J21, J22, J31.

* *Acknowledgements.* I am grateful to Nancy Folbre, Ihsaan Bassier, Deepankar Basu, Joya Misra, Suman Seth, and two anonymous referees for comments. I also thank the discussant and audience for the presentation of an earlier version of this paper at the 2021 Eastern Economic Association conference in New York. The author reports that there are no competing interests to declare. The programs and code used to create the results in this paper are available at the following repository: "Gautham, Leila (2025), "Replication package for The Paradox of Seclusion: Regional Differences in Female Employment and Wages in Urban India," Mendeley Data, V1, doi: 10.17632/4b9mgh3zf2.1 and should enable a full replication of main tables and figures.

+ Department of Economics, University of Leeds. Address: GM.20, Maurice Keyworth, Leeds LS2 9JT. Email: l.gautham@leeds.ac.uk.

1. Introduction

This paper investigates the puzzle of why urban gender wage gaps in the southern states of India are larger than those in the north, despite the consensus that conservative gender norms are stronger in the north, and gender inequality in non-wage dimensions is higher. For example, northern states do substantially worse in terms of gender parities in educational attainment, women's self-reported indices of bargaining power or autonomy, sex ratios, female child mortality, fertility rates, and age at marriage (Dyson & Moore, 1983; Malhotra, Vanneman, & Kishor, 1995; Chakraborty & Kim, 2010; Rammohan & Vu, 2018).[i] And yet, gender gaps in median wages are far higher in the south than in the north, at a difference of 12 percentage points.[ii] Discrimination in the wage-setting process, differences in human capital investments, and unequal unpaid work burdens are some inter-related factors that might contribute to gender wage inequality (Blau & Kahn, 2006). Therefore, we would expect settings characterized by greater (non-wage) gender inequality to have larger gender wage gaps. This is at odds with observed patterns of regional gender wage gaps in India.

In this paper, I estimate selection-corrected gender wage gaps separately for the north and the south. As wages are observed only for employed individuals, and as female employment rates are low, correcting for selection bias is essential for measuring women's relative wages (Neal, 2004; Blau & Kahn, 2006; Olivetti & Petrongolo, 2008). If we have positive selection (i.e., a positive association between employment and potential wages), women's observed wages would be greater than their potential wages, on average. As male employment rates are typically high, there is less scope for selection to affect male wages (as borne out in my empirical analysis), and the observed gender wage gap would then be smaller than the gender gap in potential wages. The magnitude of this discrepancy would be positively related to the degree of female non-participation in employment. Across Indian states, urban gender wage gaps are lower in states that have low rates of urban female wage employment (Figure 1).[iii]

I focus on urban areas as earlier work on regional gender differences has been restricted to rural, agricultural work. Urban residents are 31% of India's population, numbering at 377.1 million individuals (MOSPI, 2016). North-south differences in agricultural gender wage gaps in India were first noted by Ester Boserup and attributed to the greater supply of female agricultural wage labour in the south (Boserup, 1970). Mahajan & Ramaswami (2017) empirically test and provide evidence in favor of this hypothesis for rural workers. However, a north-south difference in the urban gender wage gap also exists and is larger in magnitude than the difference in the rural gap.[iv] Compositional differences in human capital characteristics induced by low and variable employment rates would be more important in explaining differences in the gender wage gaps for non-manual, relatively skilled work, whereas low female-male labour substitutability might primarily apply to physical labour (Bhalotra & Fernández Sierra, 2018). The reasons underlying the north-south difference are therefore likely to differ across rural and urban contexts: more than two-thirds of rural wage workers are occupationally classified as manual labourers, compared to only a fifth of urban wage workers, making it important to study urban labour markets separately.[v]

I show that lower gender wage gaps in the north are due to stronger positive selection of women into wage work in the north compared to the south (i.e., women's employment and potential wages are more strongly positively associated with each other in the north). My key result is that correcting for selection eliminates nearly all of the north-south difference in gender wage gaps. Selection-correction does not change male median wages, but lowers female median wages, and to a greater degree in the north. My preferred method of correcting for selection involves imputing the position of the non-employed in relation to the median wage on the basis of observable characteristics (Olivetti & Petrongolo, 2008; Blau et al., 2024). I find similar results when I utilize robustness checks that allow for selection based on unobserved characteristics.

But why do northern states have stronger positive female selection into wage work in the first place? To understand regional differences in female employment, I present a household choice model of how gender norms may influence female selection into wage work. A key feature of this model is the absence of a social stigma associated with women's white-collar work (Goldin, 1995; Klasen & Pieters, 2015). A higher stigma lowers wage employment at all levels but has the least effect on highly educated women, given their access to white-collar jobs. The model demonstrates how a higher stigma strengthens positive selection of women into wage work, resulting in lower *observed* gender wage gaps. I provide suggestive evidence that norms against women's work are stronger in the north. Consistent with the model's predictions, female employment rates as a whole are lower in the north, and the north-south employment differential is the largest for less-educated (i.e., low-wage) women. This helps explain stronger positive female selection into wage employment in the north relative to the south, and why observed gender wage gaps are more likely to understate true gender wage inequality in the north.

This paper contributes to the literature on gender wage inequality and selection in labour markets (Neal, 2004; Blau & Kahn, 2006; Olivetti & Petrongolo, 2008; Mulligan & Rubinstein, 2008): my paper is among the first to systematically examine the effect of selection on wages in India, a context where low and variable female participation rates introduce significant bias in gender wage gap measurement. Few studies of gender wage inequality account for selection biases in developing countries (Seneviratne, 2020), and only one other paper—Lee and Wie (2017)—studies India (in conjunction with China). However, they rely solely on the Heckman (1979) two-step correction method, using marital status and children as excluded instruments. The identifying assumption that these instruments affect employment probability but not wages is implausible, given recent evidence that children have negative effects on women's wages (Kleven et al., 2019).

In contrast, I examine median wages and apply a selection-on-observables strategy that relies only on correctly imputing wage positions with respect to the median. I subject this analysis to a battery of robustness checks, including estimation strategies that allow for selection on unobservables (these include the Heckman correction, but I use instruments—e.g., co-residence with parents-in-law—that are plausibly exogeneous to women's wages). Further, they focus only on adjusted gaps (i.e., gender differences in wage offers adjusted for human capital covariates), whereas I examine the total gender wage gap, which includes gender wage inequality arising from human capital characteristics, consistent with my focus on understanding overall differences in gender wage inequality between the north and the south. Importantly, I also discuss the *mechanisms* underlying selection: I argue that gender norms stigmatizing women's manual labour induce stronger positive female selection into wage work.

I also contribute to a growing literature on the effects of gender norms on economic outcomes (Fernandez, 2007; Jayachandran, 2020), by linking micro-level dynamics involving social stigmas against women's work and female employment with macro-level outcomes such as the gender wage gap. Recent work has focused on low female labour force participation rates in India (Afridi et al., 2018; Klasen & Pieters, 2015; Sarkar et al., 2019). However, comparatively little attention has been paid to regional differences in female participation rates and the implications of low and variable participation rates for the measurement and interpretation of gender wage gaps. Finally, this paper also extends the demographic literature on the north-south divide in patriarchal institutions and gender inequality in India (Dyson & Moore, 1983), reconciling the paradox of lower urban gender wage gaps in settings with greater gender inequality.

The next section summarizes the data used. Sections 3 and 4 describe my empirical strategy and results from the estimation of selection-corrected wage gaps, showing that stronger positive selection of women into wage work in the north results in an understatement of gender wage inequality, relative to the south. Section 5 offers a discussion of the mechanisms underlying differential female selection into wage work, suggesting that stronger gender norms in the north suppress the participation of low-wage women. Section 6 concludes.

2. Data

My primary dataset for measuring employment and wages is the nationally representative Employment-Unemployment Schedule (EUS) of the Indian National Sample Survey (NSS). I pool four rounds of the NSS-EUS for the years 2004–5, 2007–8, 2009–10 and 2011–12. I terminate my analysis in 2012, after which the NSS-EUS was replaced with the Periodic Labour Force Survey (PLFS). Given differences in survey design, I do not pool the NSS-EUS and PLFS. However, I reproduce my main results using PLFS 2017-2019 data (Appendix Table S.1), showing that time trends do not undermine the key results (details on PLFS data, including differences with the NSS-EUS, are described in Appendix A.2).

I restrict my sample to individuals residing in urban areas,[vi] and to the 25–54 age group; excluding workers under 25 helps eliminates individuals who opt out of employment to complete their education, while excluding those over 54 abstracts from issues of partial or full retirement, consistent with practices in the prior literature (Klasen & Pieters, 2015; Blau et al., 2024). Wage employment rates are roughly stable over this portion of the lifecycle (Appendix Figure S.1). My wage measure is the log of real daily wages, obtained by dividing total weekly earnings by total days worked in that week.[vii] Wages are expressed as 2011–12 rupees using the state-level Consumer Price Index for Industrial Workers to deflate earnings in urban areas. Participation in wage work is defined as a dummy variable that equals 1 when an individual has engaged in wage work in the previous week (consistent with the weekly recall period for wages) and has non-missing wages. I follow Dyson and Moore's (1983) classification of Indian states as "north" and "south." The north includes the states of Gujarat, Rajasthan, Uttar Pradesh, Madhya Pradesh, Punjab, and Haryana; south includes Kerala, Tamil Nadu, Andhra Pradesh, Karnataka, and Maharashtra.[viii]

Educational attainment consists of the following categories: no schooling, less than primary school, primary school, middle school, secondary school, higher secondary, and diploma, college or above. Years of potential experience are constructed as age minus years of education minus 6. Caste is among the principal categories for exclusion and differentiation in Indian society, captured here in four broad administrative categories: scheduled castes (SC), scheduled tribes (ST), other backward classes (OBC) and "others," a residual category that roughly contains dominant/privileged castes. The NSS does not directly indicate parent-child or spousal relationships within the household, but I use the "relationship to household head" variable to match children, parents, and spouses to each other (Sobek & Kennedy, 2010; Gautham, 2022). All estimates are weighted using NSS-EUS sample weights, normalized to unity for each year. See Appendix A for further details on data and variable construction.

Sample means for all variables are shown in Table 2, separately by gender, region, and wage participation status. Compositional patterns are consistent with stronger positive selection of women into wage work in the north. In the south, employed and non-employed women are similar in terms of average years of education (7.2 vs. 7.4) and shares with secondary or higher education (42 vs. 41 percent). In the north, employed women are markedly better educated than non-employed women (7.6 vs. 6.5 years; 49 vs. 37 percent with a secondary education or higher), while men's schooling varies little by employment status in either region. Among the employed, the female-to-male wage ratio is higher in the north than in the south (0.79 vs. 0.69), a pattern that anticipates the finding that observed gaps understate true gender wage inequality to a greater extent in the north.

In the subsequent analysis, I report the urban gender wage gap as the raw (unadjusted) difference in log wages between women and men. This choice reflects my interest in overall gender wage inequality that combines gender differences in human capital and in returns to these characteristics. "Adjusted" gender wage gaps (i.e., gender wage gaps adjusted for human capital characteristics), however, show similar patterns to unadjusted gaps (Table S.3).

3. Selection-corrected median gender wage gaps

The metric of interest is the gender gap in potential wages, which may differ from the observed gender wage gap that is based only on those currently employed. Recovering the gender gap in potential wages (i.e., correcting for selection) requires us to impute the potential wage distribution of the non-employed. Let w denote log wages and $F(w|g)$ the cumulative distribution of log potential wages by gender $g \in \{m, f\}$, with m denoting men and f women. If $s \in \{0, 1\}$ is an indicator for participation in wage work, then

$$F(w|g) = F(w|g, s = 1)P(s = 1| g) + F(w| g, s = 0)[1 - P(s = 1 | g)] \qquad (1)$$

where $F(w|g, s = 0)$ is not directly observed. If it is different from $F(w|g, s = 1)$, the gender wage gap computed from the observed wage distribution will be a biased estimate for the gap in potential wages.

I infer wages for the non-employed based on observable characteristics using probabilistic imputation (Neal, 2004; Olivetti & Petrongolo, 2008; Blau et al., 2024). This method circumvents the strong identifying assumptions imposed by structural methods (such as

Heckman, 1979), does not assume positive selection unlike some techniques to tighten bounds (Blundell et al., 2007), and produces estimates that are representative of the broader population, unlike the identification at infinity approach (Chamberlain, 1986; Mulligan & Rubinstein, 2008).[ix] Assuming selection on observables presumes that all factors jointly influencing employment and wages are captured by observed covariates. However, as unmeasured variables may affect both participation and pay. Still, it remains a standard approach in wage-gap studies (e.g., Neal, 2004; Olivetti & Petrongolo, 2008; Blau et al., 2024), especially when the chosen covariates are carefully selected to represent key influences. If crucial variables remain unobserved, however, estimates could be biased. To mitigate these concerns, I apply three alternative strategies that allow for selection to occur on unobservable characteristics: panel imputation that draws on adjacent rounds to gain information on the non-employed; Heckman correction which explicitly models participation decisions; and an identification at infinity approach that estimates wage gaps on a sample with a high probability of employment. These additional methods help validate the robustness of my findings against potential violations of the selection-on-observables assumption.

3.1. Probabilistic imputation

I focus on median gaps (defined as the difference in the median log female and male potential log wages), motivated by the fact that it only requires correct imputation of wage positions of the non-employed relative to the median, rather than the precise value of those wages. Following Olivetti and Petrongolo (2008), I first estimate the probability that an employed individual has a wage below their gender- and region-specific median wage using a probit model (by gender and region):

$$Pr(BM_i = 1|\boldsymbol{X}_i) = \Phi(\boldsymbol{\gamma}'\boldsymbol{X}_i) = \hat{P}_i \qquad (2)$$

where $BM_i = 1$ if individual i earns less than their gender- and region-specific median, and 0 otherwise, $\boldsymbol{X}_i$ is a vector of controls including education (7 categories), years of potential experience, caste, marital status, own children under 5 and under 14, subregion, and year (regression results presented in Table S.4), and Φ is the cumulative distribution function of the standardized normal distribution.

Second, the coefficients $\boldsymbol{\gamma}$ from estimation (2) are then used to predict probabilities of having a potential wage below the gender- and region-specific median for the non-employed. These predicted probabilities $\hat{P}_i$ are then used as sampling weights for the non-employed. I create an "imputed sample" with two copies for each non-employed observation: one with a below-median wage (assigned a weight $\hat{P}_i$) and the other with an above-median wage (assigned a weight $1 - \hat{P}_i$).[x] Finally, I estimate median gender wage gaps from the full sample (employed individuals and the imputed sample for the non-employed). To arrive at a correct reference median wage (by gender and region), I iterate over all three steps until the median wage for each gender-region group converges to the reference median wage.[xi] Standard errors are obtained by bootstrapping (500 replications) over the entire procedure.

3.2. Results

The observed gender gap in median log wages is at 48 log points (38 percent) in the north and at 69 log points (50 percent) in the south (Panel A of Table 3). Male wages are higher in the south compared to the north, reflecting both higher levels of education (Table 2) and higher levels of economic output and growth in southern states. This is consistent with findings from rural labour markets: male wages are higher in the south compared to the north, while female wages in both regions are similar (Mahajan & Ramaswami, 2017). While regional differences in male wages deserve greater scrutiny, this paper focuses on gender wage differences: more specifically, why does a region (i.e. the south) with more progressive outcomes for women (including lower gender gaps in education) have higher gender wage inequality compared to the north?

Probabilistic imputation of the non-employed to below- or above-median reduces median female wages (i.e., indicating positive selection for women) in both the north and south but has minimal effects on male wages (Panel B). This matches our expectation that selection effects are stronger for women than for men. The reduction in female wages is greater for northern women (i.e., stronger positive selection). Therefore, selection-corrected gender wage gaps in the north and south converge to 72 and 74 log points, respectively (difference not statistically significant). In other words, the 21 log-point north–south difference in observed gender gaps is almost entirely a selection artifact, and correcting for selection largely eliminates the north–south difference in the gender wage gap.

Does selection also help explain differences in *rural* gender wage gaps? Prior research attributes north-south differences in the rural wage gap to relatively greater female labour supply in the south (Boserup, 1970, Mahajan & Ramaswami, 2017). Given that rural work consists primarily of manual labour, there might be less scope for selection on human capital characteristics to play a role in influencing wages. Consistent with this, rural gaps and north-south differences in rural gaps remain largely unchanged after applying selection correction

(Table S.5). Conversely, this paper does not attempt to estimate the impact of labour supply on the wage structure (i.e., to model general equilibrium effects). Instead, it takes the wage structure as given and estimates the impact of selection into wage work. Differences in labour supply might further increase gender wage gaps in the north relative to the south, but existing research suggests that the limited female-male substitutability assumption underlying the impact of female labour supply on women's relative wages has limited applicability for urban labour markets (Bhalotra & Fernández Sierra, 2018).

4. Robustness checks

Across all three strategies that allow for selection on unobservables (panel imputation, Heckman correction, and identification at infinity), the core finding persists: once selection is corrected, regional differences in gender wage gaps effectively disappear. Where differences remain, they are small and statistically indistinguishable from zero.

4.1. Panel imputation

I use panel data (from the Indian Human Development Survey or IHDS, described in full detail in Appendix A.2) from 2004–5 and 2011–12 to impute wages for the non-employed based on wages in the adjacent round. For example, I augment the 2011–12 wage samples with the 2011–12 non-employed who participated in wage work in the 2004–5 round and impute their position with respect to the median in the 2011–12 wage distribution based on their position in the 2004–5 wage distribution.

Median gender gaps in log hourly wages estimated by region from the IHDS are similar to those in the NSS-EUS but not identical. This reflects differences in time period coverage (the IHDS does not contain data for 2007–8 or 2009–10) and the wage measure used (hourly versus daily wages). Problematically, however, the IHDS is not representative at the state level, and the following estimates are subject to that caveat. There is a 16-log point difference between gender gaps in the north and south (Table 4.A). Given that only a small fraction of the non-employed have wages that can be recovered from adjacent rounds, the change in the median gender wage gap from the expanded sample is not large. However, it does alter the north-south difference in the gender wage gap in the expected direction: the gender wage gap in the north rises by 9 log points, while the gender gap in the south rises by 6 log points (Table 4.B). The difference between the two correspondingly shrinks by 2 log points.

Panel C uses the probabilistic imputation method described earlier (using the same set of observed covariates as with the NSS: education, potential experience, caste, marital status, children, region, and year; sample means in Table S.6) to infer wage positions relative to the median for the remaining non-employed (while the panel imputation allows for selection on unobservables, this hybrid approach leverages observable characteristics for the remaining non-employed). This eliminates nearly all of the remaining difference in the gender wage gap between the north and the south.

4.2. Heckman correction

Second, returning to the NSS-EUS, I apply the conventional Heckman (1979) method of structurally modelling selection into wage work. I estimate the first-step equation:

$$Pr(s_i = 1|\boldsymbol{Z}_i) = \Phi(\boldsymbol{\pi}'\boldsymbol{Z}_i) \qquad (3)$$

where $s_i = 1$ if individual i participates in wage work and 0 otherwise, and the covariate vector $\boldsymbol{Z}_i$ includes all wage covariates $\boldsymbol{X}_i$ (education, experience, caste, marital status, children, subregion, year) as well as a set of excluded instruments: spousal employment and the education of the household head (to proxy for non-wage income and household resource constraints); and the presence of a father-in-law or mother-in-law in the household (to capture stricter adherence to social norms). The relevance of the latter instrument is supported by causal research showing that co-residence with parents-in-law is a significant shifter of married women's labour supply (Dhanaraj & Mahambare, 2019). First-step estimation results are presented in Table S.7.

As there is arguably less of a selection problem for male employment and given the practical difficulty of finding convincing instruments for men, I apply the Heckman correction only to female wages. As the second step, I estimate the wage equation:

$$w_i = \boldsymbol{\beta}'^{\boldsymbol{X}_i} + \theta\lambda_i(\boldsymbol{\pi}'\boldsymbol{Z}_i) + u_i \qquad (4)$$

where $\lambda_i(\boldsymbol{\pi}'\boldsymbol{Z}_i)$ is the inverse Mills ratio obtained from step 1. I use the coefficients $\boldsymbol{\beta}$ to predict wages for women not in wage work (wage regressions shown in Table S.7).

Observed mean gender gaps in the north and the south are 43 log points (35 percent) and 60 log points (45 percent) respectively (Table 5.A). Including the Heckman selection-correction term to extrapolate wages for women not in wage work (while continuing to use observed average male wages), reduces average female log wages, and increases gender gaps to 110 log points in the north and 103 log points in the south (Table 5.B): the north-south difference reverses sign (but is not statistically significant). Thus, even when explicitly modelling selection on unobservables, the smaller gender wage gap in the north appears to be the result of female selection into wage work.

Identification hinges on the instruments satisfying the exclusion restriction: specifically, co-residence with parents-in-law, spousal employment, and education of the head must not directly affect women's wages. Two of my excluded instruments (spousal employment and education) might affect a woman's wages directly (e.g., through network effects). I re-estimate the Heckman model using a pared-down instrument set that retains only parents-in-law co-residence variables (Table 5.C); the resulting gaps are qualitatively unchanged (though standard errors are larger), reinforcing the baseline findings.

4.3. Identification at infinity

Finally, I apply the identification at infinity method by finding a segment of the population for which the probability of employment approaches unity and estimating the gender wage gap within this group (Chamberlain, 1986; Mulligan & Rubinstein, 2008). Specifically, I predict the probability of employment, through a probit of participation on the same set of observed characteristics as before, separately by gender. I then retain only those observations with a predicted probability greater than 0.8 and compute median gender gaps within this group. The choice of 0.8 follows the value employed in the extant literature (Mulligan & Rubinstein, 2008; Blau et al., 2024), and balances the need for validity (i.e., keeping a sample with high attachment to wage work, with the predicted probability of employment approaching unity), with retaining a sufficiently large sample.[xii] With this restriction, gaps shrink to 23 and 27 log points in the north and south (Table 5, panel D), respectively, given that women with high attachment to wage work also have better observed characteristics, on balance, than the rest of the female population (see Table S.8 for probit regression and Table S.9 for characteristics of the identification at infinity sample). Here, as well, the north-south difference in the gender wage gap converges, pointing to the same conclusion: correcting for selection eliminates substantive regional differences. However, that the gaps themselves cannot be taken as representative of the labour market as a whole as they apply to workers with a high likelihood of engaging in wage work (i.e., workers with high levels of human capital).

5. Participation differentials and stigmas against working women

My empirical results demonstrate stronger positive selection of women into wage work in the north: i.e., relatively fewer low-wage women enter wage work in the north, resulting in lower observed gender wage gaps. But what explains such patterns of participation? Goldin (1995) argues that, in developing countries, social norms against women's work outside the home apply primarily to women engaging in low-wage manual jobs, rather than "respectable" white-collar jobs. Stronger norms are therefore more likely to suppress the employment of less-educated, low-wage women, producing stronger positive female selection into wage work. In this section, I provide suggestive (not causal) evidence in favor of this argument: first, I illustrate how social norms influence female selection into wage employment, with implications for gender wage gaps, using a stylised household choice model.[xiii] I then document north-south differences in gender norms, while also examining other potential drivers of female wage employment and showing that these are unlikely to explain north-south differences in female employment.

5.1. Household choice model of female wage participation

Consider a household of two adults: a woman and her spouse.[xiv] Although the model foregrounds supply-side determinants of wage participation, it is important to note that outcomes are jointly shaped by region-specific labour demand conditions (a channel I return to later). Household preferences are defined over consumption of a market good c, the value ϕ of the non-market good that is produced if the woman does not engage in wage work (e.g., housework or childcare that is foregone if $s = 1$), and a cost δ imposed by norms that stigmatize women's wage work outside the home (i.e., δ applies if $s = 1$):

$$u = u\,(c\,, \phi(1 - s)\,, \delta s\,) \qquad (5)$$

with utility u increasing in market and non-market consumption, but decreasing in the stigma: i.e., $u_1 > 0$, $u_2 > 0$, and $u_3 < 0$ (u_j being the derivative of u with respect to its j^{th} argument). For tractability, I assume the functional form

$$u = \ln(c) + \phi(1 - s) - \delta s \qquad (6)$$

which assumes that market consumption has diminishing marginal utility as represented through the log functional form, whereas preferences concerning non-market production and the stigma are modelled additively to facilitate transparent interpretations of tradeoffs. By comparing u when $s = 0$ to $s = 1$, and applying the budget constraint $c = w_m + sw_f$ where w_m and w_f are male and female wages, we can obtain the following expression for whether the woman will participate in wage work (steps in Appendix B):

$$s = 1\left[ln\left(1 + \frac{w_f}{w_m}\right) - \phi - \delta > 0\right] \qquad (7)$$

Intuitively, the relative gain in consumption from female wages must outweigh lost non-market production as well as the stigma from working. As we expect, the probability of female participation increases with potential female wages (i.e., positive selection), declines with male wages (higher male wages raise baseline consumption, reducing the marginal gain from female participation), and decreases with the value of the household good.

How does a stronger stigma affect participation? If we follow Goldin's (1995) argument, a stronger stigma would have only a minimal impact on women with access to white-collar jobs, and a larger impact for women in manual labour. We can model female participation by education levels, taking education as an easily observable proxy for access to white-collar jobs. With Klasen and Pieters' (2015) assertion that women need at least a secondary education to gain access to white-collar jobs that are not subject to social stigma, we might model the stigma as being greater for women with education less than the threshold (e.g., women with less than a secondary education). Taking the stigma as a continuous, decreasing function of education does not change the substantive result. A stronger stigma would reduce participation overall, but to a greater extent for less-educated, low-wage women who cannot access white-collar jobs, therefore strengthening positive female selection into wage work, and resulting in a smaller observed gender wage gap.

5.2. Gender norms and north-south participation differentials

An array of studies suggest that patriarchal norms are stronger in the north of India compared to the south (Singh et al., 2022; Dyson & Moore, 1983). Evidence on the perceived stigma of women working outside the home can be gleaned from reported views on women's employment. Further, the extent to which women report that they can move freely outside the home, or the control that their spouses exert over their movement, might also serve as proxies for such norms. Urban individuals in the south are more likely to disagree that a woman earning more than her husband is likely to cause problems, or that children suffer if their mother works for pay (Table 6.A; details on data used in Appendix A). They are also less likely to disagree that having a job is the best way for a woman to be an independent person.

We also observe greater restrictions on women's travel in the north. A higher fraction of women in the south report that they do not need permission from their spouse or a senior family member to visit a grocery shop, or to travel a short distance by train or bus, and a greater fraction report having travelled in the past five years to (another) metro city (Table 6.B). Women in the south are also more likely to report that they can leave town alone or go shopping alone. Another manifestation of social norms against women's presence outside the home is reflected in the control that their spouses exert over their movement—women in the north are more likely to report that their husbands insist on knowing where they are, or do not permit them to meet friends, or do not trust them with money, or are jealous if they talk with other men (Table 6.C). Finally, women in the south are more likely to live closer to the natal families, are slightly more likely to have their own mother present in the household, and are considerably less likely to be living with their parents-in-law (Table 6.D). Research from India finds that support from natal kin enhances women's autonomy, while co-residence with parents-in-law encourages a closer observation of patriarchal norms (Dyson & Moore, 1983; Dhanaraj & Mahambare, 2019).

What implications do stronger patriarchal norms against working women have for women's participation in wage work and, consequently, for gender wage inequality? As the model outlined earlier suggests, the north—with a stronger stigma—would see lower female participation in wage work overall, relative to the south. However, this gap would be the largest among less-educated women who cannot access white-collar work. So, stronger gender norms in the north increase positive female selection into wage work. Taking secondary education as the benchmark qualification needed for urban women to access white-collar job (Klasen & Pieters, 2015), we see that participation rates in wage work among both less and highly educated women are similar in the south, but are about 50 percent higher among highly-educated women compared to less-educated in the north (Figure 2.a). Also, consistent with the norms applying to women but not men, we do not see participation differentials across education groups among men (Figure S.2). As borne out in the wage analysis, stronger positive selection of women into wage work in the north reduces the observed wage gap. Thus, paradoxically, a context with less progressive gender norms would see a smaller gender wage gap.

5.3. Other drivers of participation differentials

Demonstrating a causal relationship between gender norms and patterns of female participation is outside the scope of this paper. Other factors influencing participation differentials include household resources, unpaid work responsibilities, sectoral demand for female labour, and employer discrimination, and might instead drive observed north-south differences. I consider each and provide descriptive evidence inconsistent with these factors.

Household resources and care constraints

A lack of household resources (i.e., income outside of the woman's own earnings) might "push" women into wage work out of necessity in overcoming financial constraints. Conversely, unpaid work responsibilities (such as domestic chores or caring for children) might raise the opportunity cost of women's work outside the home and discourage wage work. Therefore, if less-educated women in the north are more likely to belong to less financially-constrained households, or have greater unpaid work constraints, this might explain stronger positive female selection into wage work in the north.

Looking at spousal education and wages as a proxy for resource constraints suggests that this is not the case. Average spousal years of education and wages are lower in the north compared to the south (and therefore unlikely to explain women's lower participation). More educated spouses might have more liberal views on women's employment. However, female wage participation is lower in the north than in the south, across all groups of spousal education and wages (Figure S.3).

To verify more rigorously that north-south differences in female participation are not an artifact of household resources or care constraints, I net out these influences from participation rates. Specifically, I regress participation on age, caste, and year fixed effects, and proxies for resource constraints (spousal education) and unpaid work requirements (own children under five, presence of own mother, or mother- or father-in-law). Replacing spouse's education, as a proxy for resource constraints, with the education of the household head or household earnings yields similar results. The residuals from this regression are free of variation driven by these characteristics: i.e., they represent participation rates having controlled for proxies for household resources or care constraints. Figure 2.b plots these residuals of female participation in wage work. North-south differences in these residuals across highly-educated and less-educated women remain similar: both types of women continue to participate at roughly similar rates in the south, while participation rates among highly-educated women relative to less-educated women are 60 percent higher in the north.

Sectoral demand

North-south participation differentials could also be influenced by the structure of demand for female labour. In particular, if the availability of white-collar jobs is greater in the north, this could elicit greater participation among highly educated women (relative to less educated women) and help explain stronger positive selection into wage work in the north. Simple descriptive analysis does not support this hypothesis. To the extent that male wage labour exogenous, male sectoral shares in wage employment by region are informative: the share of white-collar jobs (defined as administrators, managers, professionals, and clerks) in male wage employment are similar across regions—indeed, slightly higher in the south than in the north (33 versus 30 percent) (Figure S.4). Furthermore, the share of women in total white-collar employment is exactly the same (at about 25 percent) in both the north and south (Table S.11).

These figures suggest that the number of white-collar jobs (relative to the size of the workforce) is not higher in the north.

The availability of white-collar jobs, as measured by the number of such jobs divided by the number of individuals who have at least a secondary school education (and would therefore qualify for such jobs) also does not show the north to have higher white-collar jobs available (Table S.11). This holds across genders, and when college education is taken as the necessary benchmark for being qualified to hold a white-collar job. This similarity in the availability of white-collar jobs (and therefore the demand for female white-collar workers) across regions is not surprising: the bulk of white-collar employment is concentrated in non-tradeable, public sector services such as public administration, education, and health. The higher share of white-collar jobs in women's employment in the north appears to be due to a relative under-supply of female labour for non-white-collar jobs.

Employer discrimination

A final, demand-related, driver of regional differences in female wage participation could be that social disapproval of women's manual wage work may manifest both as norms that are internalized within households (which is the channel emphasized here), and as employers discriminating against the employment of women in particular types of work. Employers in the north may simply be less likely to employ women in non-white-collar jobs. Distinguishing between these two channels is beyond the scope of this paper. However, assigning a dominant role to the second channel would suggest that employers discriminate only on the basis of employment, and not on wages. As observed wages for northern women (relative to men) are higher than for women in the south, selection effects would have to be powerful enough to outweigh (and reverse) the negative effects of wage discrimination on wages. Disentangling the respective roles of gender discrimination in the labour market and internalized norms—possibly through an examination of participation amongst north-south migrants—is a promising avenue for future research.

6. Conclusion

When women's participation in paid employment is variable, gender gaps in wages do not accurately reflect true gender inequality in wages. The analysis presented here drives this point home. I show that constructing selection-corrected gender wage gaps by inferring wages for the non-employed based on observable characteristics eliminates the difference in urban gender gaps between the north and the south. Selection-corrected wages for men are similar to observed wages; however, selection-corrected wages for women are lower than observed wages, and to a degree for women in the north. We see stronger positive female selection into wage work in the north.

Female participation in wage work in urban India is low, and may depend on household income, job opportunities, the opportunity cost in terms of forgone home production, and the social stigma associated with particular kinds of work. Women with low levels of education may have access only to public, manual work that carries a particularly high social stigma. I provide descriptive evidence for higher levels of social stigma against working women in the northern states and demonstrate how this could contribute to lower rates of female participation in wage work, particularly among less educated women. Paradoxically, therefore, stronger patriarchal norms that stigmatize women's participation in wage work also result in lower observed gender wage gaps.

This insight, while derived through a comparative analysis within the Indian context, may apply to other developing countries where social norms against women working outside the home are widespread, and female participation rates are low. For example, women's relative participation rates in wage employment are much lower in countries with stronger gender norms against women's work (Figure 3.a; see also Jayachandran, 2020). Strikingly, however, such norms show no relationship with gender wage gaps (Figure 3.b). While this associational pattern is not the focus of investigation of this paper, it suggests that the paradox of low gender wage gaps in contexts with strong patriarchal norms or high gender inequality in non-wage dimensions is not one that is restricted to the Indian context.

As this research demonstrates, observed gender wage gaps might be a misleading metric for policy progress towards gender inequality. Policies that remove barriers against the participation of low-skilled women, for instance, might worsen observed gender wage gaps. Correcting for female selection into wage work is important to measure progress towards gender wage equality.

[i] Table 1 illustrates some of these urban gender disparities using Census data from 2011 (latest available): educational outcomes for urban women are higher in the south than in the north, both in absolute terms, and relative to men. Sex ratios (the number of women per 1000 men) are considerably higher in the south.

[ii] See Table 1, using data from National Sample Survey 2004–12 data for urban workers between the ages of 25 and 54. Section 2 offers a more detailed discussion of the data used, and the north-south definition.
[iii] A similar pattern holds if we replace female employment rates with female-male employment ratios.
[iv] The rural female-male wage ratio is 0.66 in the north, and 0.56 in the south: a 10-percentage point difference, contrasted to the 12-percentage point urban difference.
[v] Rural estimates of selection-corrected gaps are also provided for comparison, although they are not the primary focus of the paper.
[vi] The NSS follows the Census definition of urban area as all places with a municipality, corporation, or notified town area committee, or those places that have a minimum population of 5,000, at least 75 percent of male working population in non-agricultural work, and a population density of at least 1,000/square mile.
[vii] The NSS-EUS does not have information on hours worked. However, based on information from the PLFS 2017-2019, weekly hours worked are similar for employed women in the north and the south. Results that utilize hourly rather than daily wages show similar conclusions (see Table S.1 and Table 4).
[viii] My key results are robust to the exclusion of Maharashtra or Kerala from the southern grouping, as well as the inclusion of additional states (Chandigarh, Delhi, and Bihar) to the northern group (Table S.2). According to the 2011 Census, the total population in the north and the south, so defined, is 455 million and 364 million, respectively (together constituting about 68% of the total Indian population); however, urbanization rates are higher in the south (42%) compared to the north (28%), and urban population sizes in both regions are roughly comparable at 127 million and 154 million for the north and south, respectively.
[ix] Limitations of structural methods include reliance on an excluded variable that affects employment but not wages and strong assumptions regarding functional form. With respect to worse-case bounds: the positive selection restriction is not sufficient to obtain a meaningful range. For instance, we might construct worst-case bounds on median potential wages M by estimating the lower bound on median wages assuming that $F(M|g, s = 0) = 1$ and upper bound with $F(M|g, s = 0) = 0$. For women, rates of non-participation in wage work are 91 and 82 percent in the north and south, respectively, and as they are greater than 50 percent, we cannot recover bounds for median female wages. The identification at infinity approach involves estimating the wage gap in a segment of the population for whom the probability of employment approaches one—this sample is likely to be unrepresentative of the general population.
[x] In practice, when I assign someone a wage below (above) the median, I assign a wage of -5 (13), this value being well below (above) the minimum (maximum) observed log wage for all gender-region wage distributions. As I use NSS-EUS sampling weights throughout, I multiply the weights $\hat{P}_i$ with the sampling weights provided by the survey.
[xi] Convergence is defined as less than a 0.1 rupee difference between reference median wage and the resulting selection-corrected median wage.
[xii] In Table S.10, I employ different values for the thresholds, and show that my key conclusions are preserved across thresholds.
[xiii] I adopt a unitary household framework primarily for analytical clarity and tractability, as my primary interest is in aggregate household-level outcomes rather than intra-household distribution (Lundberg and Pollak 1996); similar outcomes concerning the former (i.e., impact of stigma on women's labour force participation) are obtained in a cooperative or non-cooperative bargaining framework.
[xiv] Of urban women in the 25–54 age group, about 88% are currently married, 8% are widowed, divorced, or separated, and the remaining 4% have never been married (pooled NSS-EUS 2004–12).

References

Afridi, F., Dinkelman, T., & Mahajan. K. (2018). Why are fewer married women joining the work force in rural India? A decomposition analysis over two decades. *Journal of Population Economics*, 31(3), 783–818.

Bhalotra, S. & Fernández Sierra, M. (2018). The distribution of the gender wage gap. IZA Discussion Paper.

Blau, F. D., & Kahn, L. M. (2006). The U.S. gender pay gap in the 1990s: Slowing convergence. *Industrial and Labor Relations Review* 60 (1), 45–66.

Blau, F. D., Kahn, L. M., Boboshko, N., & Comey, M. (2024). The impact of selection into the labor force on the gender wage gap. *Journal of Labor Economics* 42 (4), 000–000.

Blundell, R., Gosling, A., Ichimura, H., & Meghir, C. (2007). Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica* 75 (2), 323–363.

Boserup, E. (1970). *Woman's role in economic development*. London: Allen and Unwin.

Chakraborty, T., & Kim, S. (2010). Kinship institutions and sex ratios in India. *Demography* 47 (4), 989–1012.

Chamberlain, G. (1986). Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics* 32 (2), 189–218.

Desai, S., & Vanneman, R. (2018). India Human Development Survey-II (IHDS-II), 2011-12.

Dhanaraj, S., & Mahambare, V. (2019). Family structure, education and women's employment in rural India. *World Development*, 115, 17-29.

Dyson, T., & Moore, M. (1983). On kinship structure, female autonomy, and demographic behavior in India. *Population and Development Review* 9 (1), 35–60.

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

Gautham, L. (2022). It takes a village: Childcare and women's paid employment in India. *Population and Development Review*, 48(3), 795-828.

Goldin C. (1995). The U-shaped female labor force function in economic development and economic history. In: *Investment in Women's Human Capital and Economic Development*. University of Chicago Press; 1995. pp. 61–90.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* 47 (1), 153–161.

IIPS (2017). National family health survey (NFHS-4), 2015-16: India. Technical report, International Institute for Population Sciences.

Jayachandran, S. (2020). Social norms as a barrier to women's employment in developing countries. NBER Working Paper, National Bureau of Economic Research.

Klasen, S., & Pieters, J. (2015). What explains the stagnation of female labor force participation in urban India? *World Bank Economic Review* 29 (3), 449–478.

Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4), 181-209.

Lee, J. W., & Wie, D. (2017). Wage structure and gender earnings differentials in China and India. *World Development* 97, 313–329.

Lundberg, S., & Pollak, R. A. (1996). Bargaining and distribution in marriage. *Journal of Economic Perspectives*, 10(4), 139–158.

Mahajan, K., & Ramaswami, B. (2017). Caste, female labor supply, and the gender wage gap in India: Boserup revisited. *Economic Development and Cultural Change* 65 (2), 339–378.

Malhotra, A., Vanneman, R., & Kishor, S. (1995). Fertility, dimensions of patriarchy, and development in India. *Population and Development Review*, 281–305.

Mulligan, C. B., & Rubinstein, Y. (2008). Selection, investment, and women's relative wages over time. *Quarterly Journal of Economics* 123 (3), 1061–1110.

Neal, D. (2004). The measured black-white wage gap among women is too small. *Journal of Political Economy* 112 (1).

Olivetti, C., & Petrongolo, B. (2008). Unequal pay or unequal employment? A cross-country analysis of gender gaps. *Journal of Labor Economics* 26 (4), 621–654.

Rammohan, A., & Vu, P. (2018). Gender inequality in education and kinship norms in India. *Feminist Economics* 24 (1), 142–167.

Sarkar, S., Sahoo, S., & Klasen, S. (2019). Employment transitions of women in India: A panel analysis. *World Development* 115, 291–309.

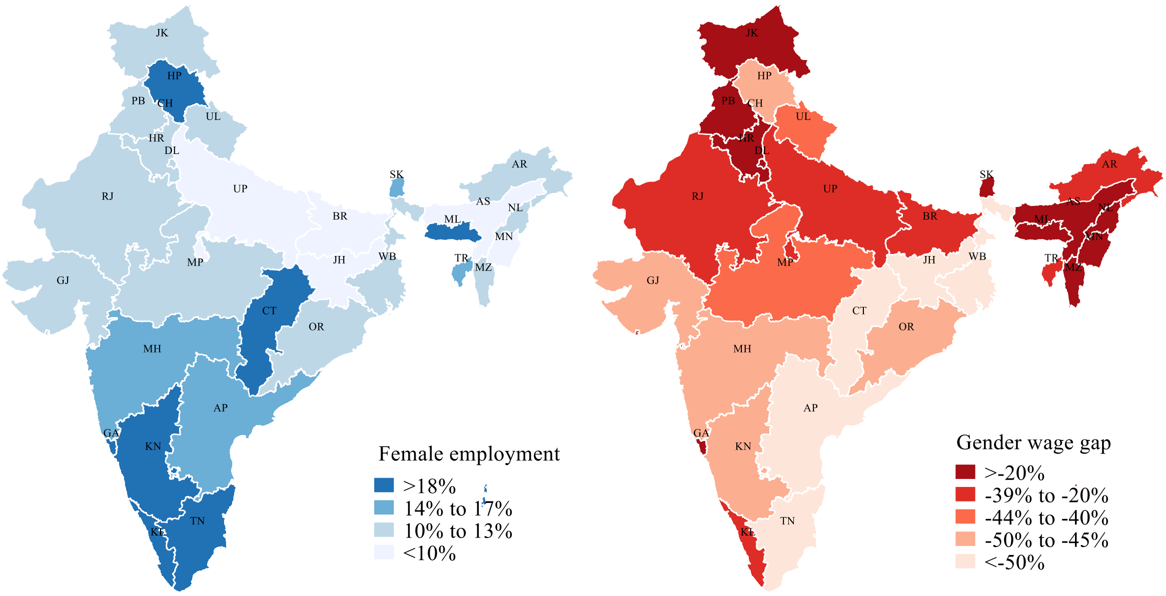
Seneviratne, P. (2020). Gender wage inequality during Sri Lanka's post-reform growth: A distributional analysis. *World Development*, 129, 104878.

Singh, A., Chokhandre, P., Singh, A. K., Barker, K. M., Kumar, K., McDougal, L., James, K., & Raj, A. (2022). Development of the India Patriarchy Index: Validation and testing of temporal and spatial patterning. *Social Indicators Research* 159 (1), 351–377.

Sobek, M., & Kennedy, S. (2010). The development of family interrelationship measures for international census data. Population Association of America.

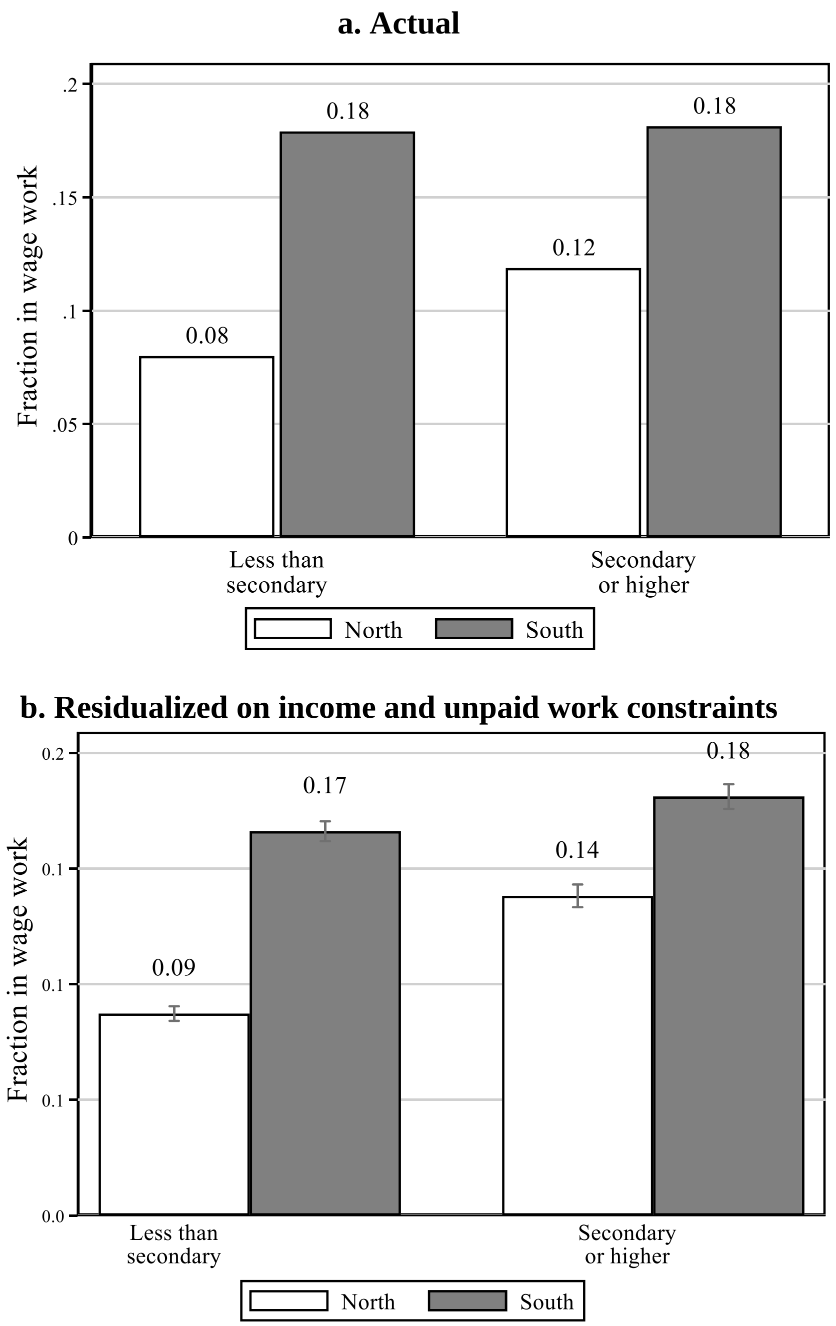
Figures

***Figure 1**. Urban female wage employment and gender wage gaps.*



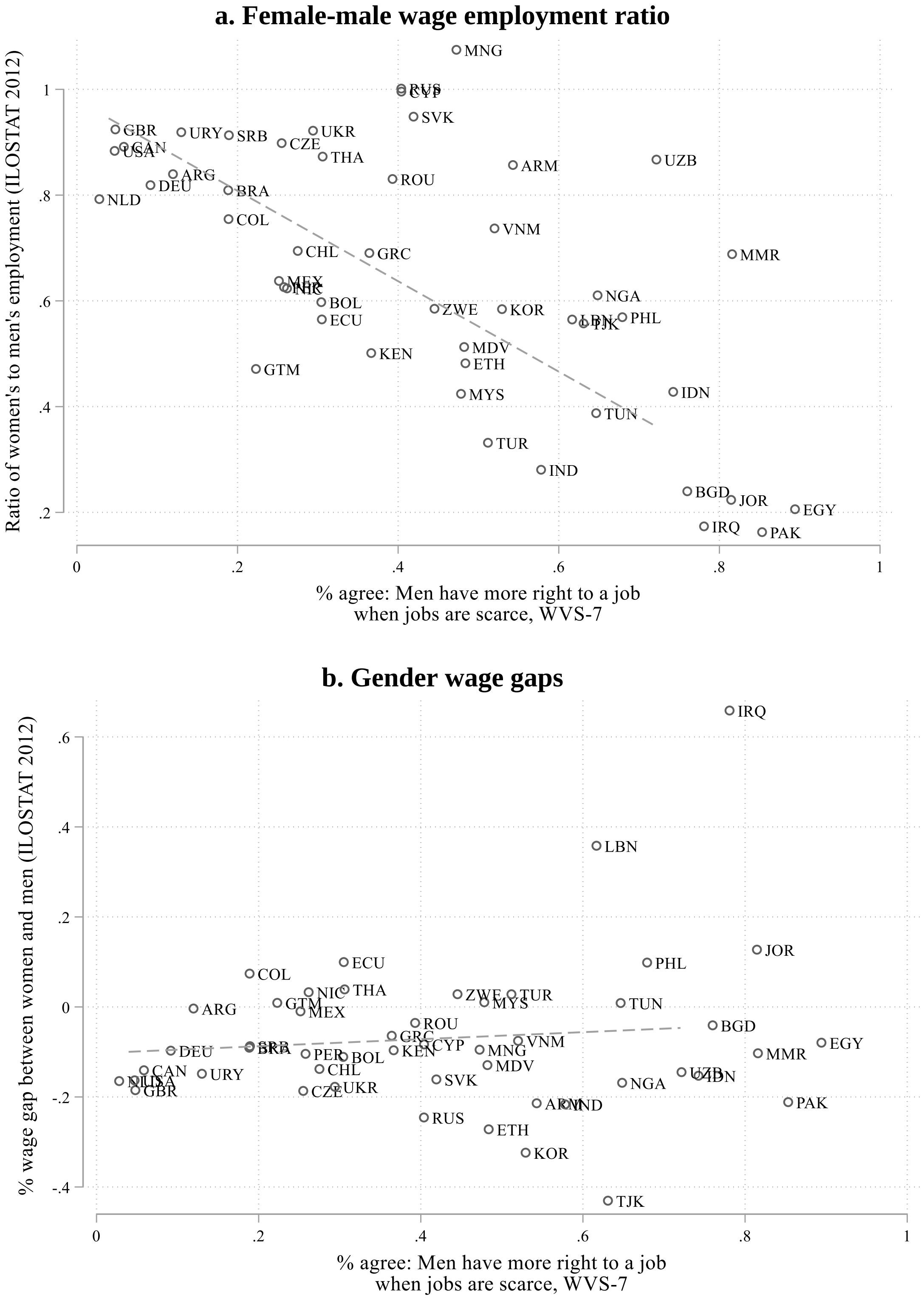
Source: NSS-EUS (2004–2012), urban individuals, ages 25–54. Female wage employment rates are the number of women in wage work as a percentage of the total number of women. Gender wage gap calculated as the percent difference between median female and male wages.

***Figure 2**. Urban female participation in wage work by secondary education.*



Source: Pooled NSS-EUS 2004–2012, urban married women, ages 25–54. Female wage participation residuals are obtained from regression of participation on age, caste, and year fixed effects, and proxies for non-wage income (spouse's education) and unpaid work requirements (child under 5, presence of own mother, or mother- or father-in-law). Mean female participation rate in each region are added back in to the residuals to maintain comparable levels. Ranges indicate 95% confidence intervals.

***Figure 3**. Cross-country gender norms, gender employment gaps, and gender wage gaps.*



Source: Information on gender norms obtained from World Value Survey Wave 6 (2010-2014): percentage of respondents per country who agree or strongly agree with the statement, "When jobs are scarce, men have more right to a job than women" (Similar results obtained for other variables on gender attitudes.) Information on gender employment and wage gaps from ILOSTAT 2012: percentage of 25-54 population in wage employment and percentage difference in average hourly earnings between women and men (if 2012 data not available, closest available year used). Linear fit weighted by 2012 population.

Tables

Table 1. *Urban gender disparities by region, 2004-12*

	North	South	Difference
Population means			
Sex ratios			
Among children (0-6)	877	932	
Among all	897	970	
Women's educational outcomes			
Years of education	6.66	7.49	
Fraction literate	0.71	0.82	
Fraction completed primary school	0.63	0.71	
Fraction completed secondary school	0.39	0.45	
Fraction completed higher secondary	0.27	0.32	
Female-male ratio in educational outcomes			
Years of education	0.80	0.85	
Fraction literate	0.83	0.90	
Fraction completed primary school	0.81	0.88	
Fraction completed secondary school	0.77	0.80	
Fraction completed higher secondary	0.78	0.79	
NSS-EUS sample means			
Wages and employment			
Female-male median wage ratio	0.62	0.50	0.12***
Female wage employment	0.09	0.18	-0.09***
Female-male participation ratio	0.19	0.30	-0.11***
Observations (NSS-EUS)	86,342	101,964	

*Source: State-population-weighted sex ratios and educational outcomes obtained from Census 2011 Tables: Sex ratio (females per 1000 males) of total population by residence and Education by sex for population aged 25-54. Labour market outcomes obtained from pooled rounds NSS-EUS 2004-12, ages 25-54 (N=188,306). North includes the states Gujarat, Rajasthan, Uttar Pradesh, Madhya Pradesh, Punjab, and Haryana; south includes Kerala, Tamil Nadu, Andhra Pradesh, Karnataka, and Maharashtra. * p<0.10, ** p<0.05, *** p<0.01.*

Table 2. *Sample means, pooled NSS-EUS 2004-12*

	North				*South*			
	Women		**Men**		**Women**		**Men**	
	E	NE	E	NE	E	NE	E	NE
Real daily wage (2011-12 rupees)	272	.	344	.	277	.	399	.
Education								
No school	0.34	0.33	0.16	0.16	0.30	0.19	0.09	0.09
Some school	0.04	0.05	0.06	0.06	0.08	0.07	0.06	0.06
Primary completed (Grade 4)	0.08	0.11	0.11	0.11	0.11	0.13	0.10	0.11
Middle completed (Grade 7)	0.06	0.13	0.15	0.16	0.10	0.19	0.18	0.19
Secondary completed (Grade 10)	0.07	0.12	0.14	0.16	0.07	0.17	0.17	0.20
Higher secondary completed (Grade 12)	0.06	0.09	0.10	0.12	0.05	0.09	0.09	0.11
Diploma, college or above	0.36	0.16	0.28	0.23	0.30	0.15	0.31	0.22
Education (years)	7.63	6.49	8.66	8.53	7.21	7.42	9.33	8.89
Potential experience (years)	24.0	24.4	22.0	22.6	23.8	23.7	21.5	22.9
Caste								
ST	0.05	0.02	0.03	0.01	0.03	0.02	0.03	0.01
SC	0.23	0.13	0.18	0.11	0.21	0.12	0.16	0.10
OBC	0.26	0.37	0.33	0.39	0.46	0.49	0.47	0.49
Other	0.45	0.48	0.46	0.49	0.30	0.37	0.34	0.40
Marital status								
Never married	0.08	0.02	0.09	0.12	0.10	0.03	0.14	0.14
Married	0.74	0.93	0.89	0.86	0.71	0.90	0.84	0.85
Divorced, separated, or widowed	0.18	0.05	0.02	0.02	0.19	0.06	0.01	0.01
Child under 5	0.25	0.33	0.35	0.38	0.18	0.25	0.32	0.31
Child under 14	0.63	0.68	0.65	0.70	0.55	0.58	0.63	0.63
Spouse characteristics								
Spouse’s years of education	8.36	8.56	6.66	6.64	7.74	9.03	7.90	7.64
Spouse in wage work	0.70	0.45	0.11	0.04	0.73	0.54	0.17	0.08
Spouse is self-employed	0.15	0.45	0.05	0.10	0.12	0.34	0.05	0.13
Father-in-law coresident	0.07	0.14	0.00	0.00	0.05	0.07	0.00	0.00
Mother-in-law coresident	0.08	0.14	0.00	0.00	0.07	0.09	0.01	0.01
Observations	4152	38387	19933	23870	9253	42187	29231	21293

Source: Pooled NSS-EUS 2004-12, urban individuals, ages 25-54. Notes. 'NE' and 'E' denote non-participation in wage work and participation in wage work, respectively. I use spousal and parental links to construct dummy variables for the spousal education, spousal employment, the presence of a father- or mother-in-law in the household, and the presence of own children under the age of 5 and under the age of 14. Years of education are imputed from education categories following Kingdon and Theopold (2008).

Table 3. *Median log daily wages, by gender and region.*

	North	South	North-south difference
A. Observed (employed)			
Women	4.83***	4.86***	-0.03
	(0.02)	(0.02)	(0.03)
Men	5.31***	5.55***	-0.24***
	(0.01)	(0.01)	(0.02)
Gender gap	-0.48***	-0.69***	0.21***
	(0.03)	(0.02)	(0.04)
B. Probabilistic imputation (employed and non-employed)			
Women	4.57***	4.79***	-0.21
	(0.42)	(0.13)	(0.36)
Men	5.30***	5.53***	-0.23***
	(0.02)	(0.01)	(0.04)
Gender gap	-0.72***	-0.74***	0.01
	(0.35)	(0.13)	(0.38)
Observations			
Panel A: Women	4152	9253	
Panel A: Men	19933	29231	
Panel B: Women	42539	51440	
Panel B: Men	43803	50524	

*Source: Pooled rounds NSS-EUS 2004-12; urban individuals, ages 25-54. Panel A reports median log daily wages for employed wage workers only, by gender and region. "Gender gap" is women minus men within region; "North–south difference" is north minus south for the corresponding statistic (so positive values indicate a higher statistic in the north). Panel B augments the employed sample by probabilistically imputing wages for the non-employed (see text for details); underlying probit regressions to predict below-median wage probability shown in Table S.3. Bootstrapped standard errors (500 replications). * p<0.10, ** p<0.05, *** p<0.01.*

Table 4. *Median log hourly wages, by gender and region (IHDS).*

	North	South	North-south difference
A. Observed			
Women	2.01***	2.20***	-0.19***
	(0.04)	(0.02)	(0.03)
Men	2.58***	2.93***	-0.35***
	(0.02)	(0.00)	(0.02)
Gender gap	-0.57***	-0.73***	0.16***
	(0.05)	(0.02)	(0.05)
B. Imputation from adjacent round			
Women	1.97***	2.20***	-0.24***
	(0.04)	(0.02)	(0.04)
Men	2.62***	3.00***	-0.37***
	(0.02)	(0.00)	(0.02)
Gender gap	-0.66***	-0.79***	0.14***
	(0.03)	(0.02)	(0.04)
C. Probabilistic imputation			
Women	1.98***	2.30***	-0.32
	(0.30)	(0.16)	(0.22)
Men	2.66***	3.00***	-0.23***
	(0.04)	(0.02)	(0.04)
Gender gap	-0.68***	-0.70***	0.02
	(0.30)	(0.15)	(0.21)
Observations			
Panel A: Women	1250	2263	
Panel A: Men	5103	6885	
Panel B: Women	1710	2893	
Panel B: Men	5838	7615	
Panel C: Women	7279	9888	
Panel C: Men	8489	9606	

*Source: IHDS 2004-5, 2011-12; urban individuals, ages 25-54. Panel A reports median log daily wages for employed wage workers only, by gender and region. Panel B includes those non-employed with observed wages in adjacent round. Panel C uses probabilistic imputation (see text for details) to impute missing wages for remaining non-employed. Bootstrapped standard errors (500 replications). * p<0.10, ** p<0.05, *** p<0.01.*

Table 5. *Mean log daily wages, by gender and region (selection on unobservables).*

	North	South	North-south difference
A. Mean, observed			
Women	5.04***	5.06***	-0.01
	(0.03)	(0.02)	(0.04)
Men	5.48***	5.65***	-0.18***
	(0.01)	(0.01)	(0.01)
Gender gap	-0.43***	-0.60***	0.17***
	(0.03)	(0.03)	(0.04)
B. Mean, Heckman correction			
Women	4.36***	4.62***	-0.26
	(0.37)	(0.21)	(0.26)
Gender gap	-1.10***	-1.03***	-0.07
	(0.38)	(0.21)	(0.26)
C. Mean, Heckman correction (only parents-in-law co-residence used)			
Women	3.95***	4.33***	-0.38
	(0.73)	(0.96)	(1.38)
Gender gap	-1.52**	-1.32	-0.20
	(0.63)	(0.96)	(1.38)
D. Mean, identification at infinity			
Women	5.03***	5.32***	-0.29***
	(0.06)	(0.05)	(0.09)
Men	5.26***	5.59***	-0.33***
	(0.02)	(0.02)	(0.02)
Gender gap	-0.23***	-0.27***	0.04
	(0.07)	(0.05)	(0.10)
Observations			
Panel A: Women	4152	9253	
Panel A: Men	19933	29231	
Panels B and C: Women	42539	51440	
Panels B and C: Men	43803	50524	
Panel D: Women	462	1162	
Panel D: Men	2099	4760	

*Source: Pooled rounds NSS-EUS 2004-12 individuals in urban areas, ages 25-54. Panel A reports mean log daily wages for employed wage workers only, by gender and region. See text for details on methods in Panels B-D. Panel B uses the full set of excluded instruments (spousal employment, education of the household head, and co-residence of parents-in-law) while panel C only uses co-residence with parents-in-law. Probit regressions for wage employment used in Heckman correction in Panel B and identification at infinity shown in Tables S.7 and S.8. Bootstrapped standard errors (500 replications). * $p<0.10$, ** $p<0.05$, *** $p<0.01$.*

Table 6. *Views on women's employment, movement, and spousal control, by region*

	North	South	Difference
A. Views on women's employment (WVS)			
Fraction that disagree or strongly disagree:			
If a woman earns more than her husband, it's almost certain to cause problems	0.29	0.44	-0.15***
Having a job is the best way for a woman to be an independent person	0.15	0.10	0.05
If a mother works for pay, children suffer	0.15	0.30	-0.15**
When jobs are scarce, men should have more right to a job than women	0.29	0.29	0.01
B. Women's travel (IHDS and NFHS)			
From IHDS: Do not need permission from spouse/senior family member to:			
Visit grocery shop	0.14	0.23	-0.09***
Travel short distance by train or bus	0.04	0.07	-0.03*
Visit relative or friend in the neighborhood	0.10	0.09	0.01
Travelled in the past five years to:			
(Another) metro city	0.29	0.45	-0.16***
(Another) village	0.90	0.88	0.02
From NFHS			
Can leave town alone	0.50	0.52	-0.02**
Can go shopping alone	0.65	0.71	-0.06***
C. Spousal control over movement (NFHS)			
Husband insists on knowing where she is	0.21	0.12	0.09***
Husband does not permit her to meet friends	0.22	0.16	0.06***
Husband does not trust her with money	0.25	0.11	0.14***
Husband jealous if talking with other men	0.26	0.23	0.03**
D. Natal kin and household structure (IHDS)			
Natal family close enough to visit in a day	0.54	0.63	-0.09***
Co-resident in the household:			
Mother	0.06	0.08	-0.02***
Mother-in-law	0.61	0.53	0.08***
Father-in-law	0.61	0.53	0.08***
Observations			
Panel A (WVS)	246	514	
Panel B and D (IHDS)	3941	4654	
Panel B and C (NFHS)	7,104	4,595	

*Source: World Values Survey (WVS-5) 2012, urban individuals. Indian Human Development Survey (IHDS) Eligible Woman module 2011-12, ever-married women, ages 15-49, in urban and India National Family Health Survey (NFHS) 2015-16, ever-married women, ages 15-49, in urban areas. See Appendix A for more details on dataset, sample, and variables for each survey. Difference column shows t-test for differences between two groups: * p<0.10, ** p<0.05, *** p<0.01.*

Appendix A: Datasets

A.1 National Sample Survey: Employment Unemployment Schedule (NSS-EUS)

I combine four rounds (2004–5, 2007–8, 2009–10, 2011–12) of the quinquennial NSS-EUS. Key variables are defined as follows:

Wage work. Includes both regular wage/salaried work (remunerated with a salary or wages on a regular basis) as well as casual wage work (wages received according to a contract that is daily/periodically renewed).

Family inter-relationships. The "relationship to household head" variable is coded in the following way in the NSS-EUS for each household member: self (i.e., this individual is the household head); spouse of head; married child; spouse of married child; unmarried child; grandchild; father/mother/father-in-law/mother-in-law; brother/sister/brother-in-law/sister-in-law/other relatives; servants/employees/other non-relatives.

For example, suppose a family of four individuals: individual 1 is coded as "self" (i.e., they are the head), individual 2 is the "spouse of head," and individuals 3 and 4 are the "unmarried child of head." I then assume that the first two individuals are married to each other, and the third and fourth individuals are their children.

In an intergenerational household, provided there is just one married child, I can identify the spouse and children of that married child. When multiple married children of the household head live in the same household, I follow the IPUMS algorithm (Sobek & Kennedy 2010) in using gender, age, and person orderings to identify spousal and parental links: spouses are of opposite gender and are close in age (<10-year difference) and listed adjacent to each other in person numbering within the household, followed by their children. To exclude individuals for whom links are indeterminate, I restrict my sample to individuals who are either household heads, spouse of household heads, married children of household heads, and their spouses (they constitute 97% of the original sample). I experimented with excluding married children of heads when there is more than one married child, but this did not affect my results.

Subregion. Refers to NSS-defined regions, which group districts based on geographic and socioeconomic similarity. There are 23 northern and 20 southern NSS regions. I use "subregion" to distinguish from broader "north" and "south" state groupings, as district-level wage data are sparse, especially for women.

A.2 Periodic Labour Force Survey (PLFS)

The NSS-EUS was discontinued after its last round in 2011-2012 and replaced with the PLFS. I pool annual PLFS data from 2017, 2018, and 2019 (excluding data from 2020-2022 due to pandemic-related disruptions). Differences between the NSS-EUS and PLFS that undermine comparability include sampling stratification (based on consumption expenditure in the NSS but shifted to educational criteria in the PLFS). The PLFS also collects data on hours worked in the reference week, enabling the computation of an hourly wage variable by dividing earnings with hours worked (unlike for the NSS-EUS where. All other variables are constructed in the same manner as the NSS-EUS.

A.3 Indian Human Development Survey (IHDS)

I supplement the NSS (a cross-sectional dataset that cannot track individuals over time) with a smaller but also nationally representative panel household survey—the Indian Human Development Survey (IHDS) (2004–5 and 2011–12) (Desai and Vanneman 2018). Wage information for the non-employed is based on their wage positions in adjacent rounds. My dependent variable is real log hourly wages, calculated by dividing total wage earnings in the previous year by total hours worked over that year, and expressed as 2011–12 rupees using state-level deflators. Following conventional practice for the IHDS, I define employment in wage work as having worked at least 240 hours in the previous year, though including individuals with less than 240 hours does not alter my results. The samples of urban individuals between the ages of 25 to 54 who were engaged in wage work are 6,353 and 9,148 for the north and south, respectively. Including the non-employment with available past wage information increases these sample sizes to 7,548 and 10,508 (increasing coverage from 38 to 45 percent for the north, and from 47 to 54 percent for the south). Sample means for the currently employed, employed only in adjacent rounds, and the never-employed, are shown in Table S.6, separately by region and gender.

A.4 World Values Survey (WVS)

The WVS is a cross-national survey measuring individuals' beliefs/values on topics ranging from gender roles to political engagement. For cross-country analyses (Figure 2), I use Wave 6 (2010–2014; N=97,220 respondents, 66 countries), measuring social norms via agreement with the statement, "When jobs are scarce, men have more right to a job than women." India-specific analysis (Table 6) combines Waves 5 (2012) and 6 (2016), restricting to urban residents in northern (N=246) and southern states (N=514). In addition to the question on job scarcity, I include attitudinal questions on whether "a woman earning more than her husband causes problems," "having a job is best for women's independence," or "children suffer if a mother works." All WVS estimates utilize provided sampling weights.

A.5 National Family Health Survey (NFHS)

The NFHS is a nationally representative survey of Indian households, focusing primarily on health and family welfare (IIPS 2017). I draw on the NFHS 2015–2016, restricting the sample to ever-married women aged 15–49 in urban areas across northern and southern states (7,104 and 4,595 observations, respectively). I use questions on mobility with respect to shopping or going outside town ("Are you usually allowed to go to the following places alone?"), with binary dummy for affirmative response coded. Spousal control is coded as affirmative responses to the questions on her relationship with her husband: "he insists on knowing where she is at all times," "he does not permit her to meet friends," "he does not trust her with money," and "he is jealous if she talks with other men."

Appendix B: Model

Starting with the household preference function in equation (6):

$$u = \ln(c) + \phi(1 - s) - \delta s$$

we can compare u when $s = 0$ to $s = 1$, and apply the budget constraint $c = w_m + sw_f$ where w_m and w_f are male and female wages. Then,

$$u(s = 0) = \ln(w_m) + \phi$$

and

$$u(s = 1) = \ln\left(w_m + w_f\right) - \delta$$

(remember that $\delta < 0$ as the stigma imposes a cost).

Then, $s = 1$ only if $u(s = 1) > u(s = 0)$, i.e., if

$$\ln\left(w_m + w_f\right) - \delta > ln\ (w_m) + \phi$$

or

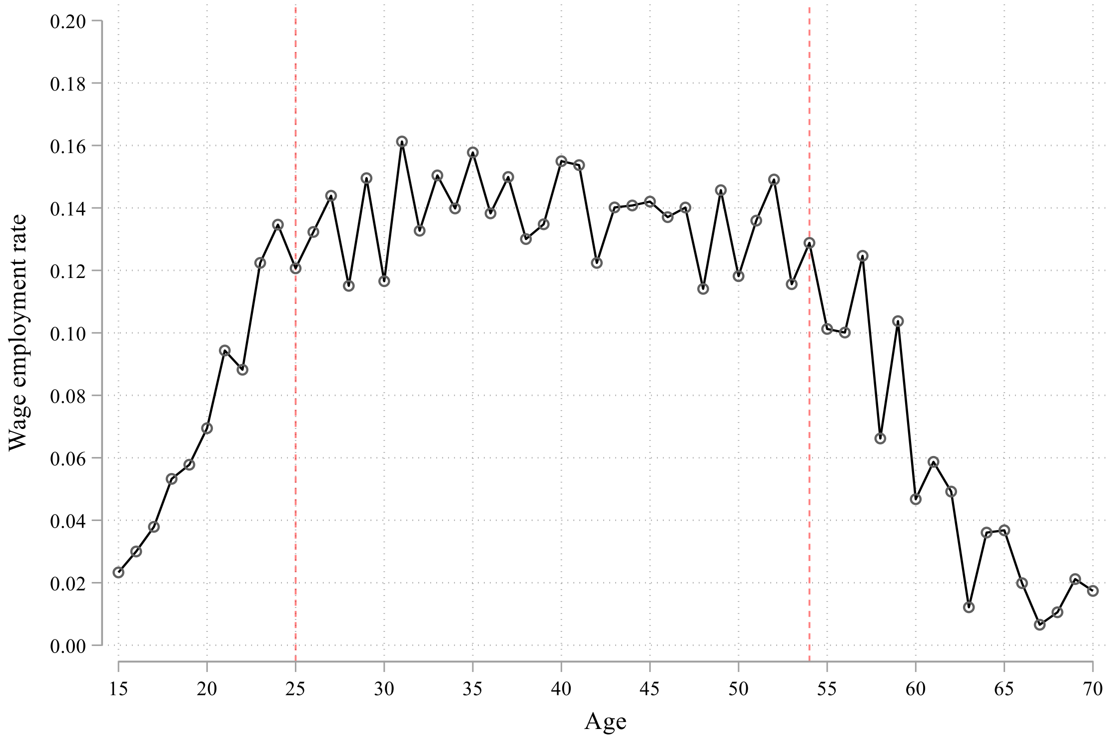
$$\ln\left(w_m + w_f\right) - \ln\ (w_m) - \delta - \phi > 0$$

or

$$ln\ \left(1 + \frac{w_f}{w_m}\right) - \delta - \phi > 0$$

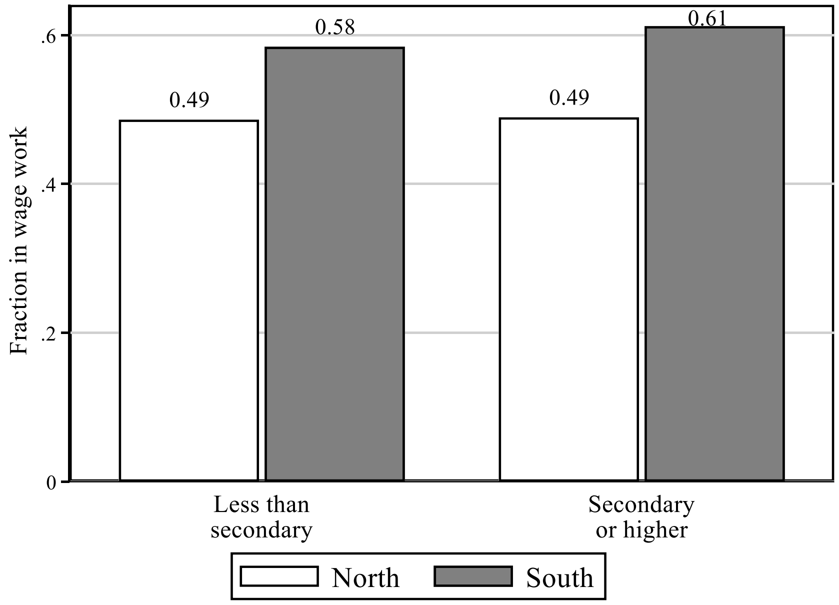
Supplementary Materials

Figure S.1*. Share of urban women in wage employment, by age.*



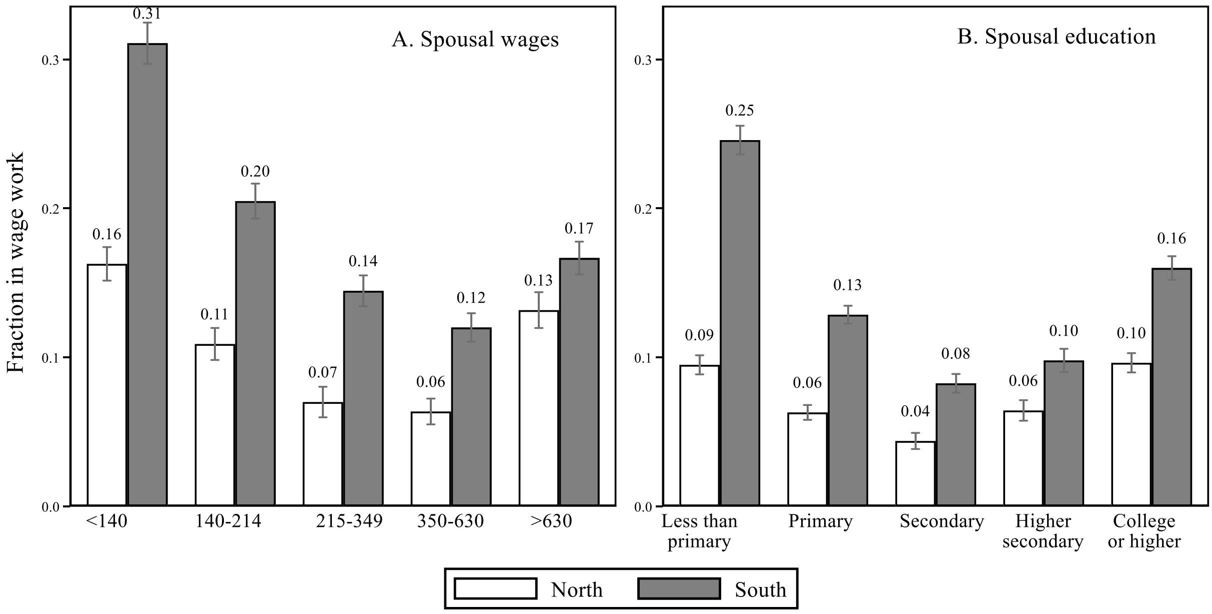
Source: NSS-EUS (2004-2012), urban women, ages 15-70. Female wage employment rates are the number of women in wage work as a percentage of the total number of women belonging to that age group.

***Figure S.2**. Urban male participation in wage work by secondary education.*



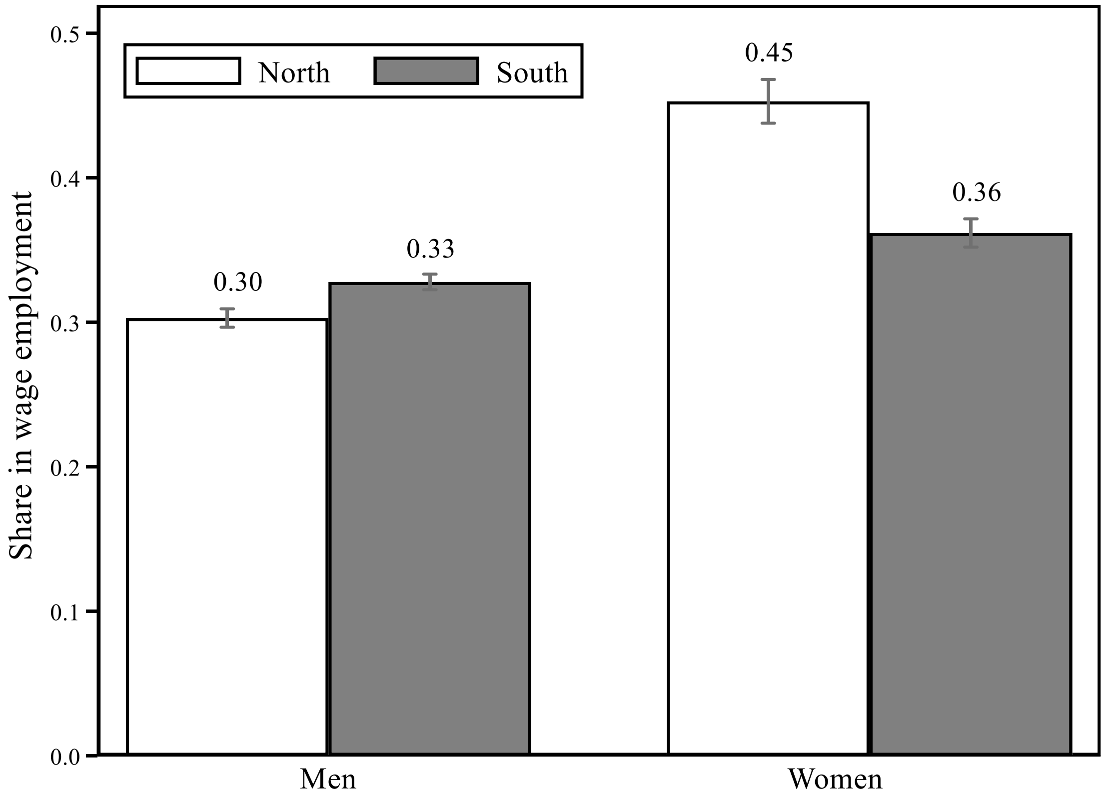
Source: Pooled NSS-EUS 2004–2012, urban men, ages 25–54.

Figure S.3.*Urban female wage participation by spousal characteristics.*



Source: Pooled NSS 2004-2012, urban married women, ages 25-54. Spousal daily wages are expressed in 2011-12 rupees and grouped into 5 equally sized groups. Average years of education are 6.6 and 7.4 in the north and south, respectively. Average spousal wages are 14 percent higher in the south compared to the north.

Figure S.4. *Share of white-collar jobs in total urban wage employment.*



Source: NSS-EUS (2004-2012), urban individuals, ages 25-54. White collar jobs defined as administrators, managers, professionals, and clerks.

Table S.1 *Median log hourly wages by gender and region, PLFS 2017-2019*

	North	South	Difference
A. Observed			
Women	5.70***	5.79***	-0.09***
	(0.02)	(0.03)	(0.03)
Men	5.99***	6.21***	-0.22***
	(0.01)	(0.02)	(0.02)
Gender gap	-0.29***	-0.42***	0.14***
	(0.02)	(0.03)	(0.03)
B. Probabilistic imputation			
Women	5.44***	5.72***	-0.28**
	(0.20)	(0.07)	(0.20)
Men	5.95***	6.21***	-0.26***
	(0.01)	(0.02)	(0.02)
Gender gap	-0.51***	-0.49***	-0.02
	(0.20)	(0.07)	(0.20)
Observations			
Panel A: Women	9818	20532	
Panel A: Men	46232	59653	
Panel B: Women	85731	90373	
Panel B: Men	53663	45392	

*Source: Pooled rounds PLFS 2017, 2018, and 2019. Sample restricted individuals in urban areas, ages 25-54. Computation of gender wage gaps and probabilistic imputation procedures exactly the same as in Table 3. Bootstrapped standard errors (500 replications) in parentheses. * p<0.10, ** p<0.05, *** p<0.01.*

Table S.2. *Median log daily wages, by gender and region (alternative region groupings).*

	North	South	North-south difference
A. Excluding Maharashtra			
A.1. Observed (employed)			
Women	4.83***	4.79***	0.04
	(0.03)	(0.01)	(0.03)
Men	5.31***	5.49***	-0.18***
	(0.02)	(0.01)	(0.02)
Gender gap	-0.48***	-0.70***	0.22***
	(0.04)	(0.01)	(0.04)
Observations			
Women	4,152	9,253	
Men	19,933	29,231	
A.2. Probabilistic imputation			
Women	4.57***	4.65***	-0.08
	(0.42)	(0.08)	(0.49)
Men	5.30***	5.44***	-0.15***
	(0.02)	(0.02)	(0.02)
Gender gap	-0.72***	-0.79***	0.07
	(0.35)	(0.08)	(0.49)
Observations			
Women	42,539	51,440	
Men	43,803	50,524	
B. Excluding Kerala			
B.1. Observed (employed)			
Women	4.83***	4.81***	0.02
	(0.03)	(0.02)	(0.04)
Men	5.31***	5.53***	-0.22***
	(0.02)	(0.01)	(0.02)
Gender gap	-0.48***	-0.71***	0.23***
	(0.04)	(0.02)	(0.04)
Observations			
Women	4,152	9,253	
Men	19,933	29,231	
B.2. Probabilistic imputation			
Women	4.57***	4.72***	-0.15
	(0.42)	(0.12)	(0.49)
Men	5.30***	5.52***	-0.22***
	(0.02)	(0.01)	(0.02)
Gender gap	-0.72***	-0.80***	0.07
	(0.35)	(0.12)	(0.49)
Observations			
Women	42,539	51,440	
Men	43,803	50,524	
C. Including additional states in the north			
C.1. Observed (employed)			
Women	4.96***	4.85***	0.11**
	(0.05)	(0.02)	(0.05)

Men	5.37***	5.55***	-0.18***
	(0.02)	(0.01)	(0.02)
Gender gap	-0.41***	-0.69***	0.29***
	(0.05)	(0.02)	(0.05)
Observations			
Women	4,152	9,253	
Men	19,933	29,231	
C.2. Probabilistic imputation			
Women	4.65**	4.79***	-0.14
	(0.41)	(0.13)	(0.43)
Men	5.40***	5.53***	-0.13***
	(0.03)	(0.01)	(0.01)
Gender gap	-0.75***	-0.74***	-0.01
	(0.32)	(0.13)	(0.43)
Observations			
Women	42,539	51,440	
Men	43,803	50,524	

*Source: Pooled rounds NSS-EUS 2004-12 individuals in urban areas, ages 25-54. In panel C, the states of Chandigarh, Delhi, and Bihar are included in the northern grouping. See text for details on probabilistic imputation. Bootstrapped standard errors (500 replications). * p<0.10, ** p<0.05, *** p<0.01.*

Table S.3. *Adjusted median log daily wages, by region.*

	North	South	North-south difference
Adjusted gender gap (employed)	-0.44***	-0.50***	0.06***
	(0.01)	(0.02)	(0.02)
Adjusted gender gap, probabilistic imputation	-0.55***	-0.55***	0.00
	(0.02)	(0.04)	(0.06)
Adjusted gap, Heckman correction	-0.92***	-0.85***	-0.37***
	(0.15)	(0.22)	(0.13)
Adjusted gap, identification at infinity	-0.43	0.06	-0.49
	(0.51)	(0.09)	(0.57)

Source: Pooled rounds NSS-EUS 2004-12 individuals in urban areas, ages 25-54. See text for details on implementation of probabilistic imputation, Heckman correction and identification at infinity methods. Sample and estimation are identical to Tables 3 and 5, except that adjusted gaps computed from regression of log wages on gender dummy and full set of covariates. Bootstrapped standard errors (500 replications). p<0.10, ** p<0.05, *** p<0.01.*

Table S.4. *Probit regressions for earning below gender- and region-specific median wage.*

	North		South	
	Women	Men	Women	Men
Education (reference: No schooling)				
Some school	-0.24***	-0.49***	-0.27***	-0.28***
	(0.11)	(0.11)	(0.10)	(0.08)
Primary completed	-0.17	-0.65***	-0.43***	-0.55***
	(0.14)	(0.08)	(0.10)	(0.07)
Middle completed	-0.50***	-0.89***	-0.69***	-0.84***
	(0.19)	(0.08)	(0.10)	(0.06)
Secondary completed	-0.78***	-1.39***	-1.26***	-1.30***
	(0.18)	(0.08)	(0.11)	(0.06)
Higher secondary completed	-1.22***	-1.81***	-1.79***	-1.80***
	(0.18)	(0.09)	(0.13)	(0.07)
Diploma, college or above	-2.10***	-2.70***	-2.74***	-2.71***
	(0.17)	(0.08)	(0.11)	(0.07)
Potential experience	-0.02**	-0.04***	-0.02***	-0.04***
	(0.01)	(0.00)	(0.00)	(0.00)
Caste (reference: Other)				
ST	-0.16	0.35***	0.14	0.18
	(0.17)	(0.10)	(0.13)	(0.12)
SC	0.07	0.22***	-0.06	0.16***
	(0.11)	(0.05)	(0.08)	(0.04)
OBC	0.27***	0.23***	0.06	0.13***
	(0.11)	(0.04)	(0.07)	(0.04)
Year dummies (reference: 2012)				
2004	0.58***	0.11**	0.58***	0.37***
	(0.11)	(0.05)	(0.07)	(0.04)
2007	0.43***	0.08	0.38***	0.12***
	(0.11)	(0.05)	(0.08)	(0.04)
2009	0.26**	0.16***	0.16**	0.11***
	(0.11)	(0.05)	(0.08)	(0.04)
Child under 5	-0.11	0.10**	0.06	0.00
	(0.10)	(0.05)	(0.08)	(0.04)
Child under 14	-0.00	-0.11**	-0.06	-0.08**
	(0.10)	(0.05)	(0.07)	(0.04)
Marital status (reference: currently married)				
Never married	0.41	-0.07	0.81***	0.12
	(0.52)	(0.14)	(0.22)	(0.10)
Widowed/divorced/separated	0.13	0.56**	0.58***	0.21
	(0.53)	(0.22)	(0.22)	(0.19)
Constant	0.42	0.44**	-2.05***	0.15
	(0.84)	(0.21)	(0.36)	(0.15)
Observations	4152	19,933	9,253	29,231

*Source: Pooled rounds NSS-EUS 2004-12 individuals in urban areas, ages 25-54. Each column shows gender- and region-specific probit regression for probability of employed individual earning a wage that is below their gender- and region-specific median wage. Dummy variables for subregions (NSS-regions) included but not shown. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.*

Table S.5. *Rural median log daily wages, by gender and region.*

	North	South	North-south difference
A. Observed (employed)			
Women	4.34***	4.35***	-0.01
	(0.02)	(0.00)	(0.02)
Men	4.76***	4.94***	-0.18***
	(0.00)	(0.01)	(0.01)
Gender gap	-0.42***	-0.59***	0.17***
	(0.02)	(0.02)	(0.03)
B. Probabilistic imputation (employed and non-employed)			
Women	4.37***	4.38***	-0.01
	(0.03)	(0.01)	(0.04)
Men	4.76***	4.92***	-0.16***
	(0.01)	(0.02)	(0.03)
Gender gap	-0.39***	-0.54***	0.15***
	(0.04)	(0.02)	(0.05)
Observations			
Panel A: Women	6,699	16,987	
Panel A: Men	24,268	28,827	
Panel B: Women	71,167	63,729	
Panel B: Men	70,196	59,583	

*Source: Pooled rounds NSS-EUS 2004-12 individuals in rural areas, ages 25-54. See text for details on probabilistic imputation. Bootstrapped standard errors (500 replications). * p<0.10, ** p<0.05, *** p<0.01.*

Table S.6. *Sample means, IHDS 2004-5 and 2011-12*

	Women			*Men*		
	E	E(A)	NE	E	E(A)	NE
North						
Hourly wage (2012 Rs.)	14.87	.	.	24.21	.	.
Age	38.00	38.20	37.45	38.06	38.27	37.49
Education years	5.94	5.79	7.72	8.92	8.97	9.75
Caste						
ST	0.02	0.02	0.01	0.02	0.02	0.01
SC	0.27	0.25	0.15	0.21	0.13	0.11
OBC	0.42	0.39	0.40	0.41	0.41	0.40
Married	0.73	0.83	0.91	0.87	0.81	0.76
Child under 5	0.05	0.06	0.07	0.08	0.06	0.05
South						
Hourly wage (2012 Rs.)	15.64	.	.	20.49	.	.
Age	38.10	38.21	37.21	37.49	36.69	37.34
Education years	5.75	4.38	6.65	8.21	7.90	9.70
Caste						
ST	0.05	0.06	0.02	0.04	0.01	0.01
SC	0.24	0.19	0.13	0.20	0.15	0.08
OBC	0.31	0.34	0.26	0.27	0.30	0.25
Married	0.77	0.88	0.92	0.89	0.87	0.84
Child under 5	0.07	0.11	0.10	0.10	0.11	0.09
Observations						
North	1250	460	6669	5103	735	2651
South	2263	630	6995	6885	730	1991

Source: IHDS 2004-5 and 2011-12, urban individuals, ages 25-54.Notes: 'E', 'E(A)', and 'NE' denote individuals currently in wage work, individuals in wage work in adjacent rounds, and those not in wage work in either round, respectively.

Table S.7. *Employment and wage regressions for Heckman estimation (women).*

	Employment (probit)		**Wage prediction**	
	North	South	North	South
Education (reference: No schooling)				
Some school	-0.11	-0.15***	0.22***	0.13***
	(0.07)	(0.05)	(0.06)	(0.03)
Primary completed	-0.09	-0.23***	0.23***	0.16***
	(0.06)	(0.05)	(0.05)	(0.03)
Middle completed	-0.12	-0.47***	0.38***	0.24***
	(0.08)	(0.05)	(0.06)	(0.04)
Secondary completed	0.05	-0.45***	0.76***	0.73***
	(0.08)	(0.06)	(0.06)	(0.04)
Higher secondary completed	0.30***	-0.23***	1.25***	1.13***
	(0.09)	(0.07)	(0.06)	(0.05)
Diploma, college or above	1.10***	0.69***	1.95***	1.92***
	(0.08)	(0.07)	(0.06)	(0.03)
Potential experience	0.00*	-0.01***	0.02***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Caste (reference: Other)				
ST	0.56***	0.47***	0.18***	0.10**
	(0.08)	(0.06)	(0.07)	(0.05)
SC	0.55***	0.43***	0.11**	0.03
	(0.05)	(0.04)	(0.04)	(0.03)
OBC	0.11***	0.08***	-0.12***	-0.07***
	(0.04)	(0.03)	(0.03)	(0.02)
Year dummies (reference: 2012)				
2004	0.07	0.03	-0.29***	-0.28***
	(0.05)	(0.03)	(0.03)	(0.02)
2007	-0.02	-0.00	-0.23***	-0.14***
	(0.05)	(0.03)	(0.03)	(0.02)
2009	0.04	0.03	-0.08**	-0.06***
	(0.05)	(0.03)	(0.03)	(0.02)
Child under 5	-0.14***	-0.36***	0.02	-0.12***
	(0.05)	(0.04)	(0.03)	(0.03)
Child under 14	0.17***	0.11***	-0.08***	-0.00
	(0.05)	(0.03)	(0.03)	(0.02)
Marital status (reference: currently married)				
Never married	1.50***	1.11***	0.48***	0.44***
	(0.30)	(0.18)	(0.16)	(0.07)
Widowed/divorced/separated	0.68***	0.53***	0.19***	0.20***
	(0.07)	(0.05)	(0.06)	(0.03)
Excluded instruments				
Spousal education (reference: No schooling)				
Some school	-0.07	-0.10*		
	(0.07)	(0.05)		
Primary completed	-0.09	-0.23***		
	(0.06)	(0.04)		

Middle completed	-0.20***	-0.34***		
	(0.07)	(0.04)		
Secondary completed	-0.38***	-0.49***		
	(0.07)	(0.05)		
Higher secondary completed	-0.34***	-0.69***		
	(0.08)	(0.06)		
Diploma, college or above	-0.48***	-0.74***		
	(0.07)	(0.05)		
Spousal employment (reference: spouse not employed)				
Wage employment	-0.14***	-0.15***		
	(0.05)	(0.04)		
Self-employment	-0.70***	-0.69***		
	(0.05)	(0.04)		
Presence in the household of				
Father-in-law	-0.13	-0.07		
	(0.08)	(0.07)		
Mother-in-law	-0.24***	-0.16**		
	(0.08)	(0.06)		
Lambda (Inverse Mills ratio)			0.75***	0.67***
			(0.34)	(0.19)
Constant	-1.24***	-0.05	3.46***	3.67***
	(0.15)	(0.11)	(0.32)	(0.16)
Observations	42,539	51,440	4,152	9,253

*Source: Pooled rounds NSS-EUS 2004-12 individuals in urban areas, ages 25-54. Each column shows gender- and region-specific probit regression for probability of employment, with covariates including all covariates in wage regressions and excluded instruments (spousal education and employment, and presence of father- and mother-in-law in the household). Dummy variables for subregions (NSS-regions) included but not shown. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.*

Table S.8. *Probit regressions for identification at infinity estimation.*

	North		South	
	Women	Men	Women	Men
Education (reference: No schooling)				
Some school	-0.16**	-0.05	-0.24***	-0.11**
	(0.07)	(0.06)	(0.05)	(0.05)
Primary completed	-0.17***	-0.04	-0.41***	-0.19***
	(0.06)	(0.05)	(0.04)	(0.04)
Middle completed	-0.27***	-0.14***	-0.72***	-0.20***
	(0.07)	(0.04)	(0.04)	(0.04)
Secondary completed	-0.19***	-0.22***	-0.83***	-0.24***
	(0.07)	(0.05)	(0.05)	(0.04)
Higher secondary completed	0.02	-0.18***	-0.71***	-0.27***
	(0.08)	(0.05)	(0.06)	(0.05)
Diploma, college or above	0.74***	0.12**	0.14***	0.10**
	(0.07)	(0.05)	(0.05)	(0.04)
Potential experience	0.01**	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Caste (reference: Other)				
ST	0.65***	0.55***	0.55***	0.55***
	(0.08)	(0.07)	(0.06)	(0.06)
SC	0.62***	0.45***	0.53***	0.51***
	(0.05)	(0.04)	(0.04)	(0.03)
OBC	0.12***	0.02	0.12***	0.12***
	(0.04)	(0.03)	(0.03)	(0.02)
Year dummies (reference: 2012)				
2004	0.07	-0.10***	0.02	-0.10***
	(0.04)	(0.03)	(0.03)	(0.03)
2007	-0.02	-0.06*	0.01	-0.00
	(0.04)	(0.03)	(0.03)	(0.03)
2009	0.03	-0.01	0.02	0.05*
	(0.04)	(0.03)	(0.03)	(0.03)
Child under 5	-0.15***	-0.06**	-0.33***	-0.02
	(0.04)	(0.03)	(0.03)	(0.03)
Child under 14	0.16***	-0.12***	0.10***	0.01
	(0.04)	(0.03)	(0.03)	(0.02)
Marital status (reference: currently married)				
Never married	1.86***	0.05	1.38***	0.35***
	(0.30)	(0.11)	(0.17)	(0.07)
Widowed/divorced/separated	1.07***	-0.09	0.92***	0.10
	(0.05)	(0.08)	(0.04)	(0.09)
Constant	-1.72***	0.40***	-0.46***	0.66***
	(0.14)	(0.10)	(0.10)	(0.08)
Observations	42,539	43,803	51,440	50,524

*Source: Pooled rounds NSS-EUS 2004-12 individuals in urban areas, ages 25-54. Each column shows gender- and region-specific probit regression for probability of employment with covariates including all covariates in wage regressions. Dummy variables for subregions (NSS-regions) included but not shown. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.*

***Table S.9**. Sample means for identification at infinity sample*

	North		*South*	
	Women	Men	Women	Men
Real daily wage (2011-12 rupees)	246.30	259.77	352.21	388.49
Some school	0.02	0.04	0.04	0.03
Primary completed	0.04	0.11	0.06	0.08
Middle completed	0.02	0.14	0.10	0.16
Secondary completed	0.07	0.14	0.07	0.13
Higher secondary completed	0.07	0.11	0.07	0.08
Diploma, college or above	0.56	0.32	0.47	0.48
Potential experience	17.25	14.50	17.51	12.88
Caste (reference: Other)				
ST	0.02	0.05	0.02	0.05
SC	0.14	0.17	0.17	0.23
OBC	0.22	0.29	0.42	0.41
Child under 5	0.05	0.30	0.00	0.19
Child under 14	0.16	0.45	0.12	0.34
Marital status (reference: Married)				
Never married	0.63	0.69	0.68	0.74
Widowed/divorced/separated	0.18	0.02	0.21	0.01
Observations	462	2099	1162	4760

Source: Pooled NSS-EUS 2004-12, urban individuals, ages 25-54. Restricted to employed individuals with a predicted probability of employment greater than 0.8 (underlying probit regression given in Table S.8).

Table S.10. *Mean log daily wages, by gender and region (identification at infinity correction)*

	North	South	North-south difference
A. Threshold: 0.5			
Women	5.06***	5.17***	-0.11
	(0.06)	(0.04)	(0.08)
Men	5.57***	5.67***	-0.10***
	(0.02)	(0.01)	(0.02)
Gender gap	-0.52***	-0.50***	-0.01
	(0.06)	(0.04)	(0.08)
B. Threshold: 0.6			
Women	5.07***	5.25***	-0.18**
	(0.07)	(0.04)	(0.09)
Men	5.47***	5.77***	-0.29***
	(0.03)	(0.01)	(0.02)
Gender gap	-0.41***	-0.52***	0.11
	(0.08)	(0.04)	(0.10)
C. Threshold: 0.7			
Women	5.08***	5.35***	-0.27***
	(0.06)	(0.04)	(0.09)
Men	5.40***	5.72***	-0.31***
	(0.03)	(0.03)	(0.03)
Gender gap	-0.33***	-0.37***	0.04
	(0.07)	(0.04)	(0.09)
D. Threshold: 0.8			
Women	5.03***	5.32***	-0.29***
	(0.06)	(0.05)	(0.09)
Men	5.26***	5.59***	-0.33***
	(0.02)	(0.02)	(0.02)
Gender gap	-0.23***	-0.27***	0.04
	(0.07)	(0.05)	(0.10)
E. Threshold: 0.9			
Women	5.03***	5.22***	-0.19**
	(0.06)	(0.04)	(0.10)
Men	5.24***	5.47***	-0.23***
	(0.02)	(0.02)	(0.03)
Gender gap	-0.20***	-0.24***	0.04
	(0.06)	(0.05)	(0.10)
Observations			
Panel A: Women	664	2014	
Panel A: Men	11747	26457	
Panel B: Women	534	1638	
Panel B: Men	6520	17016	
Panel C: Women	491	1371	
Panel C: Men	3517	9871	
Panel D: Women	462	1162	
Panel D: Men	2099	4760	
Panel E: Women	462	1092	
Panel E: Men	1948	3901	

*Source: Pooled rounds NSS-EUS 2004-12 individuals in urban areas, ages 25-54. Identification at infinity computations same as in Table 5, panel D, except for the use of differing thresholds applied to predicted probability of employment. Bootstrapped standard errors (500 replications) in parentheses. * p<0.10, ** p<0.05, *** p<0.01.*

Table S.11. *Availability of white-collar jobs, by region.*

	All		*Men*		*Women*	
	North	South	North	South	North	South
(1) *N* (White-collar)	10348	10859	7775	8195	2300	2637
Gender shares in total	1.00	1.00	0.75	0.75	0.25	0.25
(2) *N* (At least secondary school completed)	48127	39854	27214	23070	20577	16756
(3) *N* (College or higher)	21305	15684	12026	9397	9132	6275
Ratio of white-collar jobs to number of highly-educated individuals						
(1)/(2)	0.22	0.27	0.29	0.36	0.11	0.16
(1)/(3)	0.49	0.69	0.65	0.87	0.25	0.42

Source: Pooled NSS-EUS 2004-12, urban individuals, ages 25-54. NSS sampling weights used to generate all estimates. White-collar jobs defined as administrators, managers, professionals, and clerks. Less than 7 percent of white-collar jobs are held by those who have less than a secondary school qualification, and about 63 percent of white-collar jobs are held by workers with a college education or higher.