Gender Bias and Male Backlash as Drivers of Crime Against Women: Evidence from India

Debasis Bandyopadhyay*, James Jones†, Asha Sundaram‡

Draft September 2018

Abstract

We explore the relationship between the gender gap in earning potential and crime against women in India. We exploit survey data from India between 2004 and 2011 to construct measures of earning potential for men and women and combine them with administrative records on both domestic violence and rapes and indecent assaults in Indian districts. We employ an instrumental variable estimation strategy that exploits trade shocks to tackle the endogeneity of earning potential. We provide evidence of a backlash effect, where a lower gender gap is associated more rapes and indecent assaults, particularly in Indian states with high gender bias. As seen in developed counties, a lower gender gap is associated with low domestic violence, consistent with better outside options for women affording them greater bargaining power in the home. However, this results only holds for states with low gender bias. There is evidence of backlash in the home in states with high gender bias. Our study highlights that gender equity may exacerbate crime against women, particularly in the presence of gender biased institutions or culture.

Keywords: Crime Against Women, Backlash, Gender Bias, Gender gaps, India

JEL Codes: O12, J16

*University of Auckland: debasis@auckland.ac.nz

†University of Auckland: james.jones@auckland.ac.nz

‡University of Auckland. Corresponding Author: a.sundaram@auckland.ac.nz, Contact: 12 Grafton Road, Auckland, New Zealand.

1 Introduction

The World Bank identifies violence against women and girls as a global pandemic that is not just devastating for victims, but has significant economic costs[1]. Recognizing its importance, a large literature in economics and other social sciences has investigated drivers of crime against women. Recent studies challenge the conventional wisdom that crime against women will duly fall as economic development brings greater opportunities for women [2]. We contribute to this debate by exploring the relationship between the gap in earning potential between genders (henceforth, gender gap) and crime against women, including rapes, indecent assaults and domestic violence. Additionally, we ask if this relationship depends on the presence of gender bias. Our study provides evidence of a backlash effect, where a lower gender gap is associated with more rapes and indecent assaults against women, particularly in a gender biased environment.

Much of the literature on crime against women focuses on domestic violence. Income shocks, such as exchange rate shocks or rainfall shocks in agrarian societies have been identified as important drivers of domestic violence against women (Sekhri & Storeygard (2014); Cools *et al.* (2015); Munyo & Rossi (2015); Abiona & Koppensteiner (2016)). Female empowerment, captured by the relative income or employment of women, has been identified as another key contributor (Angelucci (2008); Aizer (2010); Eswaran & Malhotra (2011); Hidrobo *et al.* (2016); Munyo & Rossi (2015); Anderberg *et al.* (2016)). The literature on domestic violence finds four main channels via which female empowerment, or lower gender gaps in employment and wages may affect domestic violence. Firstly, in a household bargaining model, a lower gender wage gap increases the woman's bargaining power and leads to lower domestic violence by improving her outside option (Aizer (2010); Abiona & Koppensteiner (2016); Munyo & Rossi (2015)). Anderberg *et al.* (2016) find that greater unemployment of men is associated with lower levels of domestic violence due to the reduced position of men in the household. These studies imply that improved development and equality may reduce crime against women. Secondly, in an exposure context, Chin (2011) finds that greater employment of women (and hence a lower employment gap) is associated with women being out of the house and a lower exposure to domestic violence. Hence, better relative employment opportunities for women are associated with lower domestic violence.

A third channel affecting the relationship between gender gaps and domestic violence is evolutionary backlash. Eswaran & Malhotra (2011) argue that domestic violence stems from paternity uncertainty in our evolutionary past, where males feel jealous when their partners are exposed to other males, which is likely to be the case with better employment opportunities for women. They posit that violence may be used by men to keep women in the home and away from other men. In their framework, a lower gender employment gap

[1] 'Violence Against Women and Girls', Brief, The World Bank, Washington DC, April 4 2018, accessed on 13 September at http://www.worldbank.org/en/topic/socialdevelopment/brief/violence-against-women-and-girls. The brief estimates the economic cost of violence against women in Latin America as 3.7% of GDP.

[2] See Brysk & Mehta (2017) and references therein.

resulting from greater employment opportunities for women would lead to a backlash from men and more domestic violence.

The final channel is the interaction between gender gaps, social norms and gender-biased institutions. For example, a lower gender wage gap may be associated with greater marital conflict when this violates established gender identity norms. Bertrand *et al.* (2015) find that when the wife earns more than the husband, marriage satisfaction is lower and divorce rates are higher. In contrast, Amaral (2014) find that institutional improvement in female inheritance leads to greater bargaining power for women within a marriage and decreases domestic violence. Alesina *et al.* (2016) look across countries in Africa and find that historical social norms which rendered women more valuable are associated with less family violence today. Amaral & Bhalotra (2017) posit that violence against women can be expressed as a function of sex ratios. Using district level data, they estimate that the elasticity of violence with respect to the surplus of men aged 20-24 is unity. They suggest that one channel for increased number of males leading to increased domestic violence is the cultural bias generated by a male dominated sex ratio at birth.

Economic studies focusing on crime against women that are not domestic violence are scarce. One of the few papers to do so, Amaral *et al.* (2015), looks at India's National Rural Employment Guarantee Scheme and studies the impact of increased employment for women on crimes committed against them. The study finds evidence that increased female labor force participation is associated with increased gender based violence, consistent with a societal exposure effect. In this context, more women outside the home and in employment leads to more exposure to violence. The study does not look at the impact of female wages or at male employment and wages on crime against women.

An alternative to the exposure channel for crime against women is the backlash hypothesis. As men and women become more equal in employment and wages, traditional male dominance is usurped. A minority of men backlash against this, attempting to regain power using the criminal power channel. This would suggest that a narrowing of gender gaps may be accompanied by an increase in rapes and indecent assaults against women. While this channel has been explored in more detail in other social sciences[3], it has been scarcely studied in the economics literature .

In this paper, we propose to fill this gap in the literature. We study the relationship between gender gaps and crime against women both in the domestic context and outside in India. We use data on reported crime against women across Indian districts, including rapes, assaults, harassment and domestic violence and construct district level gender gaps in earning potential using survey data. Our measure of earning potential for each gender in an Indian district is the gender-specific employment weighted average across industries of wages earned by that gender nationally. This is arguably exogeneous to shocks to labor supply from crime. Further, we tackle the potential concerns that wages are endogenously determined and recorded with

[3]for example see Martin *et al.* (2006), Whaley (2001) and the references therein.

measurement error using an instrumental variable (IV) estimation strategy that exploits trade-induced shocks to labor demand that affect the earning potential of each gender differently. Finally, we exploit the wide variation in culture across Indian states to delve into the role of gender bias in moderating the relationship between gender gaps and crime against women. Our primary measure of gender bias is the inverse of the percentage of women in the state that report involvement in major household decisions.

Our results show that a lower gender gap in earning potential is associated with higher levels of rapes and indecent assaults committed against women. A one percent decrease in the gender gap is associated with an increase in rapes and indecent assaults of 1.8 percent. Decomposing the effect of the gap into effects coming from changes in male and female earning potential, we find that lower male and higher female earning potential are associated with an increase in rapes and indecent assaults. These relationships are consistent with a backlash from men, the dominant social class (Zaluar (1995)). To our knowledge, ours is the first study to present evidence consistent with a male backlash effect on crime against women outside of the domestic violence context. We find that a narrower gender gap in employment is associated with an increase in rapes and indecent assaults. This is driven by higher levels of female employment, consistent with both a backlash effect, or a greater exposure to unsafe spaces. We find that the earning potential and employment effects are exacerbated in areas with high gender bias.

A potential concern for many studies on crime, including ours, is that our data are on reported crimes. Any effect of gender gaps on crime against women may be driven by an increase in reporting of crime, rather than an increase in instances of crime. We tackle this in several ways. Firstly, the total number of elected female representatives in a district is included as a control variable. This has been suggested to affect the level of reporting for crime against women (Iyer *et al.* (2012)), and is a proxy for institutional support for reporting . Secondly, we exclude reports of harassment from our main analysis and focus on the more extreme crimes of rape and indecent assault. Instances of rapes, indecent assaults and harassment are likely to stem from similar root causes. Although all of these crimes will suffer from under reporting, harassment is likely to be the most sensitive and rape and assault the least sensitive to it, given that evidence might be easier to obtain in the case of the latter. In fact, when we repeat our main regressions by including harrassment, we find evidence that higher female earning potential is associated with a much larger increase in reported crime. We do not observe a large change in the effect of male earning potential, suggesting that the male effect captures backlash. Finally, the fact that the negative relationship between male earning potential and crime against women is magnified in gender biased areas where institutions are potentially more favorable to men is also evidence against a potential alternate hypothesis to backlash - that lower power for men may be associated with less intimidation against reporting.

For domestic violence, a one percent decrease in the gender gap, is associated with a 1.9 percent decrease in domestic violence, consistent with the bargaining effect seen in developed countries (Aizer (2010)). This is much larger in areas where gender bias is low. Additionally, there is evidence of the opposite relationship in areas of high gender bias. In gender biased areas, a decrease in the gap is associated with an increase in domestic violence. Separating out male and female earning potential, we find that male earning potential is negatively associated with domestic violence. This demonstrates a potential backlash effect from men in high gender bias areas. Overall the domestic violence results suggest that, in line with developed countries, an increase in bargaining power can reduce violence against women in the home, but only in areas of lower gender bias[4].

Our study contributes to the literature on the drivers of crime against women, both in and outside the home. We provide evidence for a backlash effect in rapes and indecent assault that is robust to reverse causation, potential reporting bias and a host of robustness checks that employ alternate measures of gender gaps and gender bias. We show that this backlash effect is exacerbated in regions with higher gender bias. We hence highlight that the process of development, often associated with narrowing gender gaps as women increase labor force participation and command higher wages, may increase crime against them, particularly when institutions or culture are gender-biased. We find domestic violence is dominated by the bargaining effect as seen in developed countries but only in areas of lower gender bias, with a potential backlash effect in areas of high gender bias. Finally, we also contribute to the growing literature on gender gaps and their implications for growth and economic performance (Klasen *et al.* (2018)).

2 Empirical Analysis

2.1 Empirical specification

We present alternate hypotheses. If a decrease in the earning potential of men relative to women leads to a backlash effect, a narrow gender gap will be associated with a higher level of crime against women. Next, we hypothesize that these effects are stronger in areas with more gender bias. To empirically examine these hypotheses, we first estimate the following specification:

$$lnCAW_{i,t} = \alpha + \beta_1 lnG_{i,t} + \boldsymbol{\beta}_2 \boldsymbol{X}_{i,t} + \mu_i + \tau_t + \epsilon_{i,t}. \tag{1}$$

Here, $CAW_{i,t}$ is indecent assaults and rapes in district i at time t, $G_{i,t}$ refers to the gender gap, $\boldsymbol{X}_{i,t}$ includes a set of control variables at the district level. μ_i and τ_t are district and year fixed effects respectively and $\epsilon_{i,t}$ is the idiosyncratic error term. In the fully specified model, control variables include log total crime,

[4]Note that reporting bias in this case would work against the bargaining power channel. Our result hence underestimates the bargaining power effect in the presence of reporting bias.

log mean per capita household expenditure to capture the level of development, log working-age population of men and women, inequality (measured as mean per capita expenditure of the household at the seventy-fifth percentile relative to the household at the twenty-fifth percentile), controls for district composition including percentage of urban working-age population, percentage of employment in manufacturing, percentage of employment in agriculture, controls for education levels including the percentage of working-age individuals with a high-school education or above, the education gap (ratio of men to women of working-age with a high-school education or above) and the number of female elected representatives to the state legislature to control for political representation of women.

Since total crime is controlled for, the resulting analysis highlights the relationship between key independent variables and crime against women relative to overall crime. Note that total crime accounts for law and order in the district. District fixed effects account for time-invariant, unobserved shocks at the district level that determine both gender gaps and crime against women. Year effects control for year-specific shocks to crime and the gender gaps. Standard errors are clustered at the district level. The coefficient β_1 on the gender employment or earning potential gap is expected to be negative with backlash - a smaller gap between men and women is associated with more crime against women. Further, we expect any backlash effect to be strongest in areas of high gender bias, where men might perceive greater losses as power structures re-balance.

The primary source for our measure of gender bias is survey information on female participation in household decision making, with alternative measures being the percentage of females with access to Essential Services and Opportunities (ESO) sourced from McKinsey and a variation on the Missing Women variable as in (Anderson & Ray (2010, 2012)). For each gender bias variable, the states are ordered from best to worst, then grouped into low, medium and high levels of relative gender bias[5]. From this, we estimate separately:

$$lnCrime_{i,t} = \alpha_B + \beta_{1,B} lnG_{i,t} + \boldsymbol{\beta}_{2,B} \boldsymbol{X}_{i,t} + \mu_i + \tau_t + \epsilon_{i,t} \quad i \in I_B \quad B = \{low, med, hig\}. \tag{2}$$

Here, i is a district in set I_B, the set of all districts in states that have gender bias level B. The hypothesis is that $|\beta_{1,low}| < |\beta_{1,med}| < |\beta_{1,high}|$ since a lower gender gap is associated with greater crime in areas with higher gender bias.

Next, we study domestic violence by estimating equation (1) and equation (2) with domestic violence as the dependent variable. In the case of domestic violence, β_1 is negative if a lower gender gap is associated with backlash from men. However, β_1 is positive if a lower gender gaps are associated with better outside options for women and hence greater bargaining power, lowering violence against them. The mechanism between gender bias and domestic violence is less clear. On the one hand, men may be more controlling and react with more violence to smaller gaps in areas of high gender bias, so the backlash effect may dominate.

[5] The "low" gender bias states are not unbiased, they are just less biased than the rest of India. Detailed descriptions of variables are presented in the Data Appendix.

On the other hand, if women have a lower position in society, an improved outside option may dramatically improve their relative position and the bargaining effect could prevail in areas of high bias.

2.2 Earning potential

Our key independent variable of interest captures the gap in earning potential between men and women in Indian districts. We first calculate earning potential for groups of individuals: male or female, further broken down into high qualified, (Hq), and low qualified, (Lq), male or female[6]. The higher (lower) the potential earnings for a group, the more (less) economically empowered individuals are. Changes in power structures between groups are likely to influence criminal power dynamics. For example, previously dominant individuals may attempt to regain lost power by acting against the newly enfranchised, particularly when group identity is strong, as is likely in gender biased areas.

Earning potential for a group k in i is calculated as the weighted average national wage for the group across industries, where the weights are industry employment shares of the group in the district:

$$ln\overline{EP}_{i,t,k} = \frac{\sum_{j=1}^{J_i} \left(Emp_{i,04,k,j} * ln\overline{Wage}_{t,k,j}\right)}{\sum_{j=1}^{J_i} Emp_{i,04,k,j}}, \tag{3}$$

where, $ln\overline{Wage}_{t,k,j}$is the India average wage at time t for an individual in group k in industry j. We use the total weekly earnings and the total number of days worked in the week to derive the average daily wage for each person[7]. k is the person's group, either $male$ or $female$ or, in the detailed analysis: Hq $male$, Hq $female$, Lq $male$ and Lq $female$. j is the two digit NIC 1998 industry code for the industry of principal employment for the individual. $Emp_{i,04,k,j}$ is the total number of people in district i of type k employed in industry j in 2004, the first period for our data.[8]

The relative gap in earning potential in district i is defined as follows:

$$ln\overline{EP}_{i,t,gap} = ln\overline{EP}_{i,t,male} - ln\overline{EP}_{i,t,female}. \tag{4}$$

Similarly, Hq and Lq gaps are defined using the earning potential for Hq or Lq men and women. Given that earning potential in i is derived from industry national average wages, it could be argued that this variable is independent of the level of crime in i. However, if industry employment is geographically concentrated, the industry wage might be endogeneous to crime. Furthermore, as discussed in the Data Appendix, wage

[6] An individual is classified as Hq if they have a high-school education or above, otherwise they are classified as Lq

[7] Each day of part time work is counted as 0.5 of a day. Full detail on variable creation is provided in the Data Appendix.

[8] Any individual not reporting a wage, along with their associated survey weightings, are excluded when calculating the average wages for individuals of type k. If average wage information is missing for an industry, those industries are excluded from the employment weightings in i and excluded from $\sum_{j=1}^{J_i} Emp_{i,04,k,j}$. If the denominator is zero, we define $ln\overline{EP}_{i,t,k}$ as zero. For example, 38 districts out of 565 (6%) report zero employed Hq women in 2004, so earning potential for Hq women in these districts is set to zero. If Hq women are not employable in a district, then the measure of earning potential for Hq women is definitionally zero.

information may be susceptible to measurement error. To tackle these concerns, we use an instrumental variable approach that uses trade shocks to instrument for industry wages as detailed in the next section.

2.3 Identification

We instrument for the national industry wage using import tariffs on the final good produced by the industry and on inputs it employs in production. Adapting the specification in Topalova (2010), we define the exposure to tariffs (ETT) for group k in district i at time t as the weighted average of all the industry tariffs at time t, weighted by the first period (2004) industry employment composition of group k in district i :

$$ETT_{i,t,k} = \frac{\sum_{j=1}^{J_i}\left(Emp_{i,04,k,j} * \overline{Tar}_{t,j}\right)}{\sum_{j=1}^{J_i} Emp_{i,04,k,j}} \quad j \in J_{i,04}$$

Similarly, exposure to input tariffs (EIT), is defined as the weighted average of all industry input tariffs:

$$EIT_{i,t,k} = \frac{\sum_{j=1}^{J_i}\left(Emp_{i,04,k,j} * \overline{inpTar}_{t,j}\right)}{\sum_{j=1}^{J_i} Emp_{i,04,k,j}} \quad j \in J_{i,04}$$

In the above $\overline{Tar}_{t,j}$ is the simple average of the tariff on all goods produced in industry j at time t[9]. j is the NIC 1998 2 digit industry classification. $\overline{inpTar}_{t,j}$ is the weighted average of the tariff on each good used as inputs in industry j, weighted by the fraction of total inputs to j represented by each good. $Emp_{i,04,k,j}$ is the total number of people in district i of type k employed in industry j in 2004. For non-trade industries we set $\overline{Tar}_{t,j}$ and $\overline{inpTar}_{t,j}$ to zero. $\sum_{j=1}^{J_i} Emp_{i,04,k,j}$ is the total number of people of type k employed in i in 2004.[10] Employment in non traded industries is included when deriving total employment in i so that the impact of any tariffs or input tariffs are also scaled by the relative size of the tradeable sector in a district.

The resulting exposure variables provide the potential exposure to tariffs or input tariffs at time t for each type of worker in district i. They hence capture the potential level of protection from competition groups in a district experience. This is likely to influence wages and employment for the group in a district. Besides, since tariffs are set for India as a whole by the central (federal) government, we argue that ETT and EIT for a group are exogenous to the level of crime against women, providing useful instruments in our investigation.

[9] Full detail available in the Data Appendix

[10] If this total is zero, then there are no workers of type k to be exposed to tariff effects and definitionally our exposure measures are zero for this also. Hence we set $ETT_{i,t,k} = 0$ and $EIT_{i,t,k} = 0$ if $\sum_{j=1}^{J_i} Emp_{i,04,k,j} = 0$.

3 Data

Full detail on all data is included in the Data Appendix. Data on gender gaps and control variables are sourced from the Employment and Unemployment surveys of the National Sample Survey Organization (NSSO), India. Five rounds of data are used for the years 2004-05, 2005-06, 2007-08, 2009-10 and 2011-12. The surveys collect information on household per capita expenditure, the district, state and rural or urban area the household is located in, and individual variables like age, gender, level of education, employment status in the principal activity (we consider both paid and unpaid employment in both the formal and informal sectors), industry of employment and the daily cash wage (adjusted for working for half the day) during the week. From the survey information, we construct district level variables using survey weights.

Table 1: Average Indian district-level crime per 10,000 working age population

year	Total Crime	CAW	DV
2004	32.95	1.08	0.84
2005	30.64	1.08	0.90
2007	31.74	1.11	1.06
2009	31.99	1.10	1.18
2011	32.03	1.15	1.27

Note: CAW represents: Rape and Indecent Assault. DV represents domestic violence. Total crime remains fairly constant over time, however CAW and DV show an upward trend.

Data on crime are at the district level, sourced from the National Crime Records Bureau (NCRB), India. We consider Rapes and Assaults on the Modesty of Women (Indecent Assaults) under crimes against women. A variation of this variable that is more sensitive to reporting bias also includes Insult to the Modesty of Women (Harassment). Cruelty by Husband or his Relatives is used as the measure of Domestic Violence. Data on total crimes at the district level registered under the Indian Penal Code are also sourced from the NCRB. We collect data on the number of female representatives in the district elected to the state legislature from the Election Commission's Election Results[11].

Data on import tariffs come from product level data from the TRAINS database (downloaded from the WITS World Bank database). Information on the inputs used by each industry come from the input-output transactions table (IOTT 1994). State variables describing potential gender bias are derived from three sources. The primary variable uses survey information on the participation of women in household decision making from the third National Family Health Survey (NHFS) from 2004-2005. A secondary measure uses demographic information from the NSSO survey to constructs a variation of the Missing Women measure described in Anderson & Ray (2012). The final bias variable uses an index of access for women to essential services and opportunities produced by the McKinsey Global Institute in 2015[12].

[11]Data accessed from http://eci.nic.in/eci_main1/ElectionStatistics.aspx on 13 November 2017

[12]Data accessed from https://www.mckinsey.com/featured-insights/employment-and-growth/the-power-of-parity-advancing-womens-

Table 1 shows total crime, crime against women and domestic violence over the years in our sample. Total crime per 10,000 working age population is roughly stationary over time. However, crime against women displays an upward trend. Domestic violence shows a noticeable upward trend over time. Table 2 shows the evolution of gender gaps in employment and earning potential over time. All gaps are positive showing a bias towards men, which is consistent over time.

Figure 1 shows female lack of decision making, the primary proxy for gender bias, across Indian states. Dark colours signify less participation of women in household decisions, which may indicate implicit gender bias in a state. We see significant variation in gender bias across India, with the southern and eastern states displaying lower levels of gender bias relative to northern states. We exploit this variation to identify the moderating role of gender bias in determining the relationship between gender gaps and crime against women.

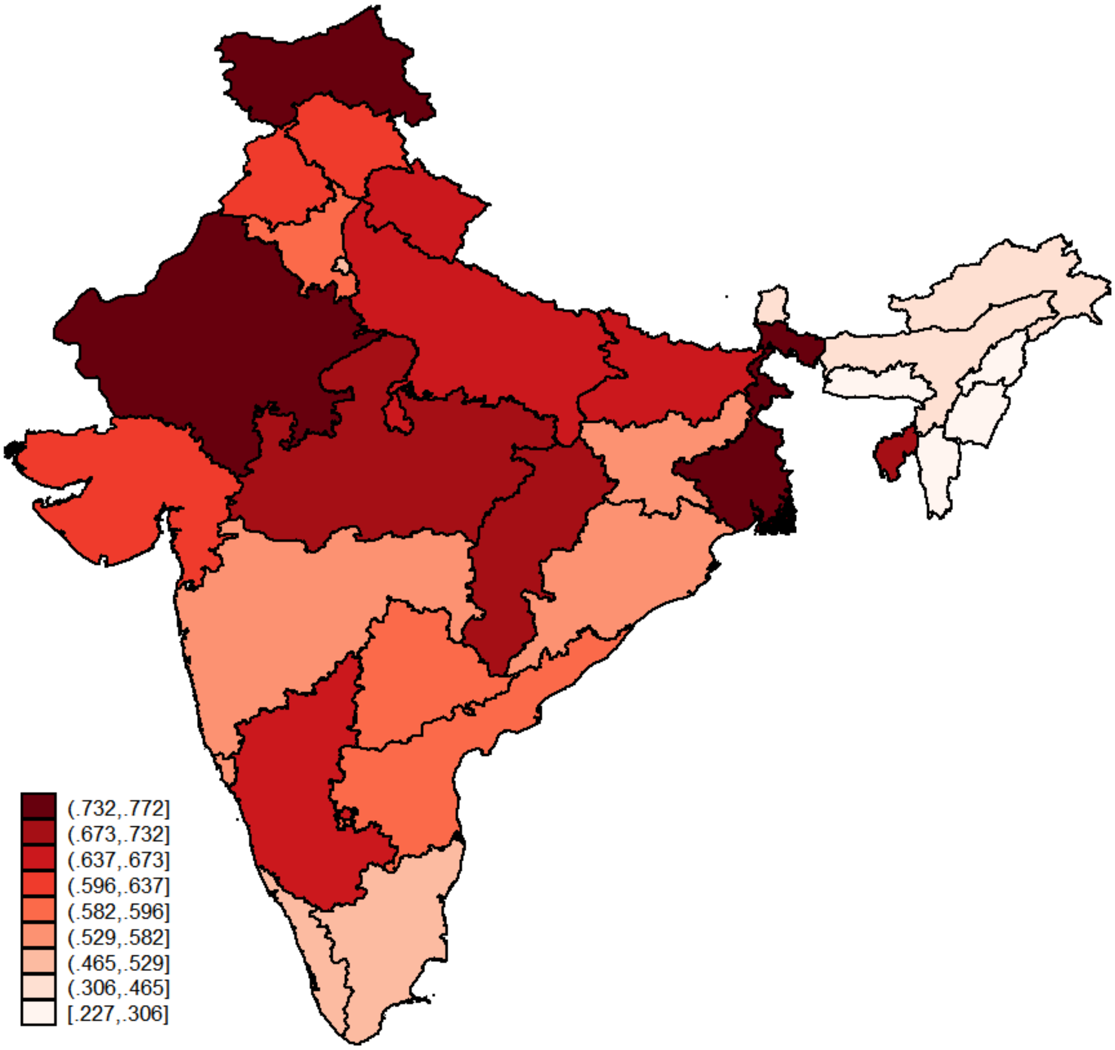
Table 2: Evolution of gender gaps in employment and EP over time

Year	ln(Mean EP gap)	ln(Emp. gap)
2004	0.494	1.187
2005	0.439	1.334
2007	0.463	1.310
2009	0.400	1.468
2011	0.374	1.584

Note: ln(Mean EP gap) is the log of the ratio of average male earning potential over average female earning potential. ln(Emp gap) is the log of the ratio of total male employment over total female employment. Both variables as presented here are India wide averages of the district level variables for each year. By construction, parity between genders is 0. A positive (negative) gap is in favour of men (women).

equality-in-india on 13 November 2017

Figure 1: Average lack of female participation in decision making by State in 2005



Note: This variable is the percentage of women who report they are not allowed to participate in one or more of the following four household decisions: Decisions regarding: the woman's own health care, making major household purchases, making purchases for daily household needs and visits to the woman's family or relatives. Darker colours represents fewer women with decision making abilities and could indicate implicit gender bias in that state. The India average for women who report they can participate in all four decisions is only 37%. Compared to developed countries, all of India is relatively gender biased according to this measure. The survey question only required participation in the decision making process. Alternative versions of the questionnaire, which asked how many women could make a decision on their own, or had the final say, displayed very low responses in the affirmative (International Institute for Population Sciences (2007)).

4 Results

4.1 Rapes and Indecent Assaults

4.1.1 Earning Potential: Baseline

Table 3 presents results from the estimation of equation (1), where the dependent variable is rapes and indecent assaults against women and the independent variable of interest, the gender gap in earning potential, is defined as in equation (4). The gender gap in total employment is included to account for potential exposure effects. In columns (1) through (3), other control variables are introduced sequentially. Across all three columns, a smaller gender gap in employment is significantly associated with an increase in crime against women, consistent with a greater exposure of women to unsafe spaces accompanying increased relative employment. The coefficient on the gap in earning potential is insignificant in all OLS regressions. Total crime shows statistically significant positive associations with crime against women. In contrast to the findings of Iyer *et al.* (2012), who looked at older data for India, we find higher female representation is significantly associated with lower crime against women.

Columns (4) and (5), show results from the instrumental variables estimation strategy. In column (4), the gender gap in earning potential is instrumented using district exposure to input and output tariffs for $Hq\,men$, $Hq\,women$, $Lq\,men$, $Lq\,women$. In this regression, the employment gap enters as a control variable only and is not instrumented. In column (5), we instrument both the gap in earning potential and the employment gap. Although not the main variable of interest, the employment gap is likely to be endogenously determined if female employment choices are influenced by the level of crime against women in a district. Similarly, if improved employment opportunities for women is associated with evolutionary backlash or an increased exposure to unsafe working spaces, a smaller employment gap will be associated with more crime against women.

The first stage regressions are presented in Table 12 in Appendix I, and show that our instruments are strong. Seven of the eight instruments are strongly correlated with the gap in earning potential as seen in column (1). The first stage regressions when instrumenting the decomposed gap, presented in the same table in columns (3a) and (3b), show that these correlations are driven by the expected gender associations. Exposure to tariffs, as weighted by Hq and Lq male employment, is positively correlated with male earning potential. The reverse is the case for input tariffs, where lower tariffs may mean cheaper factor inputs and a higher return for labour. The same relationships hold for female earning potential, with the exception of tariff exposure for Hq females, which is uncorrelated with any variable and included only for completeness. The first stage when both employment and earning potential gaps are instrumented together is shown in columns (2a) and (2b). For the employment gap, a high exposure to input tariffs as weighted by employment of Lq women and Hq men, is correlated with a smaller employment gap between men and women, with the opposite effect for exposure to tariffs. We present first stage statistics in all our tables, all of which show we can comfortably reject the null hypothesis of weak instruments.

Table 3: Gender gaps regressed on Rapes and Indecent Assaults

	ln(Rapes, Indecent Assaults)				
VARIABLES	OLS (1)	OLS (2)	OLS (3)	IV (4)	IV (5)
ln(Mean EP gap)	0.080	0.082	0.067	-1.793***	-2.215***
	[0.112]	[0.109]	[0.108]	[0.490]	[0.759]
ln(Emp gap)	-0.047***	-0.050***	-0.049***	-0.052***	-0.492***
	[0.012]	[0.012]	[0.012]	[0.012]	[0.129]
ln(Total Crimes)	0.741***	0.741***	0.749***	0.771***	0.779***
	[0.056]	[0.056]	[0.056]	[0.049]	[0.054]
ln(MPCE)	-0.087*	-0.097**	-0.096**	-0.080	-0.064
	[0.047]	[0.049]	[0.049]	[0.049]	[0.077]
ln(Male Working population)	-0.031	-0.034	-0.067	-0.110	0.551**
	[0.068]	[0.069]	[0.068]	[0.075]	[0.215]
ln(Female Working population)	0.095	0.092	0.127*	0.052	-0.571***
	[0.065]	[0.069]	[0.071]	[0.080]	[0.218]
ln(Inequality)		0.022	0.022	0.013	-0.115
		[0.046]	[0.045]	[0.043]	[0.072]
urbanisation %age		0.002	0.004	0.006	0.003
		[0.015]	[0.014]	[0.015]	[0.019]
%age emp.in Agriculture		-0.133*	-0.128*	-0.109	-1.012***
		[0.077]	[0.077]	[0.086]	[0.281]
%age emp. in Manufacturing		0.182	0.191	0.221	-0.911**
		[0.153]	[0.151]	[0.154]	[0.390]
%age with Hq education		0.002	0.008	0.013	-0.044
		[0.020]	[0.021]	[0.022]	[0.037]
ln(Hq gap)			0.021	0.024	0.001
			[0.016]	[0.016]	[0.025]
Elected female representatives			-0.045***	-0.046***	-0.037*
			[0.015]	[0.013]	[0.019]
Year and District FE	Yes	Yes	Yes	Yes	Yes
Observations	2,810	2,810	2,810	2,810	2,810
R-squared	0.251	0.254	0.259	0.204	-0.512
Number of panel	562	562	562	562	562
First-stage Statistics					
Sanderson-Windmeijer F-stat:	-	-	-	31.66	3.17
Sanderson-Windmeijer p value:	-	-	-	0.0000	0.0025
Kleibergen-Paap rk Wald F-stat:	-	-	-	31.66	2.74
Stock Yogo LIML 10% maximal IV bias critical value:				3.97	
Stock Yogo LIML 15% maximal IV bias critical value:				2.73	

Note: Data are for the years 2004-05, 2005-06, 2007-08, 2009-10 and 2011-12. EP gap is defined as in equation (4). Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1. All variables are district level totals or averages for each time period: ln(Rapes, Indecent Assaults) is the log of the total number of reports of rapes and indecent assaults. ln(Mean EP gap) is the log of the ratio of average male earning potential over average female earning potential. ln(Emp gap) is the log of the ratio of total male employment over total female employment. ln(Total Crimes) is the log of the total number of crimes. ln(MPCE) is the log of mean per capita household expenditure. ln(Male Working population) is the log of the total number of males aged 15-64. ln(Female Working population) is the log of the total number of females aged 15-64. ln(Inequality) is the log of the ratio of MPCE of the household at the seventy-fifth percentile relative to the household at the twenty-fifth percentile. urbanisation %age is the the percentage of households that are located in urban sectors. %age emp. in Agriculture is the percentage of individuals employed in farming, forestry and fishing as a proportion of total working population. %age emp. in Manufacturing s the percentage of individuals employed in manufacturing as a proportion of total working population. %age with Hq education is the percentage of individuals with a high school education or above as a proportion of total working population. ln(Hq gap) is the log of the ratio of total number of high school educated males over total high school educated females. Elected female representatives is the total number of sitting female elected representatives to the state legislature. To guard against a weak instruments problem IV regressions use continuously updated GMM estimators (CUE) .

From column (4), the coefficient on the earning potential gap is now negative and strongly significant. This suggests that after countering for any potential endogeneity in earning potential, a smaller gap is associated with higher crime against women, consistent with the backlash hypothesis. From column (5), the coefficient on the gender employment gap is an order of magnitude larger after instrumentation. Large coefficients in each IV regression are consistent with both measurement error in gender gaps attenuating coefficients toward zero or unobserved district-specific shocks being associated with greater crime against women and gender gaps simultaneously. Overall, the evidence is consistent with backlash - better relative opportunities for women relative to men are associated with greater crime against women.

4.1.2 Relationship Between Earning Potential and Gender Bias

Table 4 present the IV regression for the gender gap in earning potential when the sample is separated by districts in states with relatively low, medium and high levels of gender bias[13]. Column (1) presents the baseline results for the whole sample, reproduced from Table 3 for comparison. Columns (2-4) show the regressions for gender biased regions as determined by female participation in household decision making, our primary measure of gender bias. Column (2) displays the sub-sample regression for relatively low gender bias, with columns (3) and (4) showing medium and high gender bias respectively. The coefficients on the gap in earning potential in each regression have been presented on separate lines for ease of viewing. Columns (5-7) and (8-10) show the regressions when the sample is split according to alternative measures of gender bias as robustness checks. Columns (5) and (8) represent low gender bias, (6) and (9) medium, and (7) and (10) high bias. Columns (5-7) display bias as determined by the percentage of females in a state who have access to Essential Services and Opportunities. Columns (8-10) shows bias based on a variation of the Missing Women variable in Anderson & Ray (2012).

Across all measures of gender bias, the results are consistent. The negative correlation between rapes and indecent assaults and the gender gap in earning potential is driven by areas of high gender bias. In such regions, the magnitude of the effect is larger than in the overall sample. In low gender bias regions, the sign on the coefficient changes direction, although it also becomes insignificant. The negative relationship between the gender gap in employment and rapes and indecent assaults is only significant in regions of medium and high gender bias, suggesting that the exposure effect may not hold in regions of low gender bias. In other words, more women out in employment relative to men does not necessarily mean more crime against women unless there is gender bias.

[13] Instruments for the employment gap are less strong than for earning potential. Also, IV results obtained from instrumenting for the employment gap are almost always in line with OLS results. Hence, in subsequent analyses, we only instrument for earning potential and use employment as a control.

Table 4: IV regressions for gap in Earning Potential, regressed on Rapes and Indecent Assaults, separated by Gender Bias level

	ln(Rapes, Indecent Assaults)									
Measure of Gender bias:	Base[14]	Inverse Decision %age			Missing Women			Inverse ESO		
VARIABLES	IV (1)	IV (2)	IV (3)	IV (4)	IV (5)	IV (6)	IV (7)	IV (8)	IV (9)	IV (10)
ln(Mean EP gap)	-1.793***									
	[0.490]									
ln(Mean EP gap): Low GB		1.493			1.107			-0.784		
		[0.997]			[0.895]			[0.943]		
ln(Mean EP gap): Med GB			-0.617			-2.929***			-0.872	
			[0.616]			[0.849]			[0.691]	
ln(Mean EP gap): High GB				-2.964***			-2.299***			-2.270***
				[0.713]			[0.781]			[0.717]
ln(Emp. Gap)	-0.052***	0.051	-0.094***	-0.050***	-0.037	-0.045**	-0.063***	0.038	-0.111***	-0.042***
	[0.012]	[0.056]	[0.025]	[0.016]	[0.026]	[0.023]	[0.018]	[0.046]	[0.031]	[0.014]
ln(Total Crimes)	0.771***	0.690***	0.922***	0.814***	0.932***	0.821***	0.598***	0.783***	0.546***	0.824***
	[0.049]	[0.099]	[0.081]	[0.071]	[0.085]	[0.062]	[0.089]	[0.068]	[0.111]	[0.078]
ln(MPCE)	-0.080	-0.181*	-0.046	-0.071	-0.035	-0.233**	-0.111	-0.165*	-0.039	-0.101
	[0.049]	[0.105]	[0.072]	[0.081]	[0.070]	[0.091]	[0.086]	[0.087]	[0.074]	[0.078]
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District and Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,810	745	915	1,150	915	875	1,020	905	650	1,255
R-squared	0.204	0.247	0.357	0.031	0.341	0.126	0.092	0.331	0.201	0.206
Number of panel	562	149	183	230	183	175	204	181	130	251
First-stage Statistics										
Sanderson-Windmeijer F-stat:	31.66	8.525	20.00	14.01	9.751	9.311	13.98	17.61	7.810	22.62
Sanderson-Windmeijer p value:	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Kleibergen-Paap rk Wald F-stat:	31.66	8.525	20.00	14.01	9.751	9.311	13.98	17.61	7.810	22.62
Stock Yogo LIML 10% maximal IV bias critical value:							3.97			

Note: Data are for the years 2004-05, 2005-06, 2007-08, 2009-10 and 2011-12. Gender gaps are defined as in equation (4). Standard errors are clustered at the district level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. All non-gender bias variables are district level totals or averages for each time period: ln(Rapes, Indecent Assaults) is the log of the total number of reports of rapes and indecent assaults. ln(Mean EP gap) is the log of the ratio of average male earning potential over average female earning potential. ln(Emp gap) is the log of the ratio of total male employment over total female employment. ln(Total Crimes) is the log of the total number of crimes. ln(MPCE) is the log of mean per capita household expenditure. Gender bias variables define methods to group districts in states of varying degrees of gender bias. Inverse Decision %age defines gender bias by the percentage of women who report they are not allowed to participate in at least one of four household decisions. Missing Women is the ratio of men to women outside of the fertile age range in a district, divided by the expected ratio for men to women outside the fertile age range, with the expected ratio based on India and state averages. Inverse ESO is the inverse of the Mckinsey Institute index for the percentage of women in a state with access to essential services and opportunities. The subsample regressions for regions of low gender bias are presented in columns (2), (5) and (8). Medium gender bias is represented in columns (3), (6) and (9). High gender bias is shown in columns (4), (7) and (10). To guard against a weak instruments problem IV regressions use continuously updated GMM estimators (CUE) .

[14] Reproduced from Table 3

Results in Table 3 also allow us to argue against an alternate explanation to backlash. If earning potential does influence institutions, then a large gender gap in favour of men may favor institutions where women are intimidated against reporting crime. For example, if reports are dismissed or blame reversed to victims because males make up the majority of contributors to community groups or tax payers in a region. Following this, a narrower gender gap in earning potential may be associated with high levels of crime against women not due to backlash, but due to reporting effects. For example a narrower gap may mean institutions are less biased and take reports of crime against women more seriously. However, in regions of high gender bias, institutions are fixed in favour of men due to historical precedence. The strong negative correlation between the size of gap and crime seen in the subsample regressions with high gender bias, shown in columns (4), (7) and (10) hence highlights a plausible backlash effect. This is also consistent with the idea that men may backlash more in areas of high gender bias as they have more societal power to lose when women gain in equality.

4.1.3 Earning Potential: Robustness Checks

For robustness, we create several other versions of the earning potential variable defined in equation 3 by replacing $ln\overline{Wage}_{t,j,k}$ with other variations of industry average wages. Firstly, we create CPI adjusted earning potential using the industry average wages when each individual wage has been adjusted by a state level urban and rural CPI modifier[15]. Secondly, in addition to reported cash wages, we include payments received "in kind" (goods or services received in lieu of wages and other income such as rent, returns on assets). Finally, "excluding own district" earning potential creates industry average wages for district i using information from all individuals not in that district. Given that there are 565 districts, this variable is very similar to the main variable of interest[16]. Table 5 displays the results when the regression in column (4) of Table 3 is repeated using the alternative measures of earning potential. The baseline results are repeated in the first column for comparison. Each measure is instrumented using the tariff and input tariff variables.

The results in Table 5 show a consistent and significant negative association between the earning potential gender gaps and rapes and indecent assaults against women. The coefficients on the gaps are broadly similar in magnitude, with the exception of total earning potential, which includes wages and payments in kind. This coefficient is much larger in magnitude, and still very significant. Total earnings are recorded with a lot more noise than standard wages, and hence earning potential derived from wages alone is used in our main specification.

[15] Consumer Price Index (CPI) data comes from the State Level Consumer Price Index (Rural/Urban) for 2011, published by the Central Statistics Office of India. Data accessed from https://data.gov.in/resources/state-level-consumer-price-index-ruralurban-2011 on 13 November 2017

[16] Full detail available in the Data Appendix

Table 5: Robustness Checks for Gender gap in Earning Potential regressed on Rapes and Indecent Assaults:

	ln(Rapes, Indecent Assaults)			
VARIABLES	Base: IV (1)[17]	IV (2)	IV (3)	IV (4)
ln(Mean EP gap)	-1.793***			
	[0.490]			
ln(Mean EP gap, CPI adj.)		-1.814***		
		[0.494]		
ln(Mean Total EP gap)			-3.910***	
			[0.923]	
ln(Mean EP$_{-i}$ gap)				-1.793***
				[0.489]
ln(Emp. gap)	-0.052***	-0.052***	-0.056***	-0.052***
	[0.012]	[0.012]	[0.014]	[0.012]
ln(Total Crimes)	0.771***	0.770***	0.782***	0.771***
	[0.049]	[0.049]	[0.051]	[0.049]
ln(MPCE)	-0.080	-0.080	-0.054	-0.080
	[0.049]	[0.049]	[0.052]	[0.049]
Other Controls	Yes	Yes	Yes	Yes
District and Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	2,810	2,810	2,810	2,810
R-squared	0.204	0.203	0.027	0.204
Number of panel	562	562	562	562
First-stage Statistics				
Sanderson-Windmeijer F-stat:	31.66	31.00	13.89	31.26
Sanderson-Windmeijer p value:	0.0000	0.0000	0.0000	0.0000
Kleibergen-Paap rk Wald F-stat:	31.66	31.00	13.89	31.26
Stock Yogo LIML 10% maximal IV bias critical value:		3.97		

Note: Data are for the years 2004-05, 2005-06, 2007-08, 2009-10 and 2011-12. EP gap is defined as in equation (4). Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1. All variables are district level totals or averages for each time period: ln(Rapes, Indecent Assaults) is the log of the total number of reports of rapes and indecent assaults. ln(Mean EP gap) is the log of the ratio of average male earning potential over average female earning potential. ln(Mean EP gap, CPI adj) ratio of average male earning potential over average female earning potential where wages have been adjusted by a state level urban and rural CPI modifier. ln(Mean Total EP gap) is the log of the ratio of average male total earning potential over average female total earning potential, where the total includes wages and payments made in kind. ln(Mean EP$_{-i}$ gap) is the log of the ratio of average male earning potential over average female earning potential where earning potential in district i is derived from data from $-i$, i.e. all districts excluding i. ln(Emp gap) is the log of the ratio of total male employment over total female employment. ln(Total Crimes) is the log of the total number of crimes. ln(MPCE) is the log of mean per capita household expenditure. To guard against a weak instruments problem IV regressions use continuously updated GMM estimators (CUE) .

[17] Reproduced from table 3

Table 6: Decomposition of the gap in Earning Potential, regressed on Rapes, and Indecent Assaults

	ln(Rapes, Indecent Assaults)			
VARIABLES	OLS (1)	IV (2)	IV (3)	IV (4)
ln(Mean Male EP)	-0.756**	-1.766***		
	[0.358]	[0.525]		
ln(Mean Female EP)	-0.199	1.742***		
	[0.126]	[0.584]		
ln(Mean Lq Male EP)			-1.618***	
			[0.568]	
ln(Mean Lq Female EP)			1.704**	
			[0.804]	
ln(Mean Hq Male EP)				-0.531**
				[0.245]
ln(Mean Hq Female EP)				-0.059
				[0.116]
ln(Male Emp.)	0.072	0.037		
	[0.114]	[0.106]		
ln(Female Emp.)	0.047***	0.051***		
	[0.012]	[0.012]		
ln(Male Lq Emp.)			0.083	
			[0.077]	
ln(Female Lq Emp.)			0.038***	
			[0.012]	
ln(Male Hq Emp.)				-0.023
				[0.031]
ln(Female Hq Emp.)				-0.005
				[0.004]
ln(Total Crimes)	0.745***	0.769***	0.764***	0.754***
	[0.056]	[0.050]	[0.050]	[0.048]
ln(MPCE)	-0.107**	-0.079	-0.061	-0.098**
	[0.048]	[0.049]	[0.051]	[0.048]
Other Controls	Yes	Yes	Yes	Yes
District and Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	2,810	2,810	2,810	2,810
R-squared	0.263	0.208	0.182	0.245
Number of panel	562	562	562	562
First-stage Statistics				
Sanderson-Windmeijer F-stat:	-	24.35	7.98	16.45
Sanderson-Windmeijer p value:	-	0.0000	0.0000	0.0000
Kleibergen-Paap rk Wald F-stat:	-	20.41	6.98	13.72
Stock Yogo LIML 10% maximal IV bias:		3.97	3.78	3.78

Note: Data are for the years 2004-05, 2005-06, 2007-08, 2009-10 and 2011-12. Gender gaps are defined as in equation (4). Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1. All variables are district level totals or averages for each time period: ln(Rapes, Indecent Assaults) is the log of the total number of reports of rapes and indecent assaults. ln(Mean Male EP) is the log of the average male earning potential. ln(Mean Female EP) is the log of the average female earning potential. ln(Mean Hq Male EP) is the log of the average male earning potential for males with a high-school level of education or above. ln(Mean Hq Female EP) is the log of the average female earning potential for females with a high-school level of education or above. ln(Mean Lq Male EP) is the log of the average male earning potential for males with less than a high-school level of education. ln(Mean Lq Female EP)is the log of the average female earning potential for females with less than a high-school level of education. ln(Male Emp.) is the log of the total number of employed males. ln(Female Emp.) is the log of the total number of employed females. ln(Male Hq Emp.) is the log of the total number of employed males with a high-school level of education or above. ln(Female Hq Emp.) is the log of the total number of employed females with a high-school level of education or above. ln(Male Lq Emp.) is the log of total number of employed males with less than a high-school level of education. ln(Female Lq Emp.) is the log of total number of employed females with less than a high-school level of education. ln(Total Crimes) is the log of the total number of crimes. ln(MPCE) is the log of mean per capita household expenditure. To guard against a weak instruments problem IV regressions use continuously updated GMM estimators (CUE) .

4.1.4 Earning Potential: Decomposing male and female effects

Table 6 shows the results when the gender gap in earning potential is broken down first by male and female earning potential, and then by high qualified and low qualified males and females. For completeness, the employment gap is decomposed to match. Column (1) shows the OLS regression with male and female earning potential. We now observe a significant and negative relationship between male earning potential and reports of rapes and indecent assaults, consistent with the backlash hypothesis. After instrumenting with the trade instruments in column (2), the coefficient on male earning potential gains magnitude and significance. The coefficient on female earning potential becomes positive and highly significant. The female effect is consistent with either backlash from men when females have more societal power, or more reporting by females of existing crimes for the same reason. From the employment result in column (2), we see the negative relationship between the employment gap and crime comes from female employment. This is consistent with the exposure effect, where more females in employment results in more crimes committed against them. The interaction between male and female earning potential with gender bias is shown in Table 14 in Appendix I. The results confirm the backlash effects seen in Table 4.

Further decomposition of earning potential by Hq and Lq individuals in columns (3) and (4) respectively show that the main backlash and exposure effects are driven by individuals with a "less than high-school" level of education. There is still evidence of a negative association between Hq male earning potential and rapes and indecent assaults, consistent with backlash, although it is slightly less significant and of smaller magnitude. The coefficient on Hq female earning potential is insignificant. This could be due to conflicting interactions. On the one hand, a high earning potential for Hq females may lead to more reports of crime due to an increased backlash from men or an increase in reporting when Hq women have greater societal power. On the other hand, Hq women also have more to lose when making a report of a rape or indecent assault. They may have a highly valued job or a position that is hard to come by and may risk losing it by making a criminal report. This is particularly true if the perpetrator is a colleague or supervisor. Lq women may face less pressure to under-report.

4.1.5 Earning Potential and Harassment

Table 7 shows the results when reports of harassment are included along with reports of rapes and indecent assaults. Column (1) shows the results using the gender gap in earning potential and column (2) displays the decomposition of the gap into male and female effects. Harassment is much more sensitive than rapes and indecent assaults to issues of under reporting, hence it is noteworthy that the coefficient on female earning potential, shown in column (2), is much larger when including harassment than when compared to rapes and assaults alone, as shown in Table 6. Harassment is recorded with a large amount of noise in India, so we qualitatively interpret these results as indicating that increased female earning potential is associated with increased levels of reporting. By the same token, since the coefficient on the gap in earning potential, shown in column (1), is similar in magnitude to the baseline, and the negative correlation with male earning potential, shown in column (2), is much less inflated by the inclusion of harassment, we have evidence that the male effect is likely not influenced by reporting factors. Hence in our baseline specification, the negative coefficient on male earning potential cannot be explained as simply a reporting effect. Instead we have further evidence of male backlash when the previously dominant social class faces a lower earning potential. The first stage statistics suggest that we achieve good instrumentation.

Table 7: Male and female Earning Potential, regressed on Rapes, Indecent Assaults, and Harassment

VARIABLES	ln(Rapes, Indecent Assaults, Harassment) IV (1)	IV (2)
ln(Mean EP gap)	-2.057***	
	[0.521]	
ln(Mean Male EP)		-2.771***
		[0.825]
ln(Mean Female EP)		6.680***
		[1.051]
ln(Emp. gap)	-0.048***	
	[0.013]	
ln(Male Emp.)		-0.002
		[0.130]
ln(Female Emp.)		0.061***
		[0.019]
ln(Total Crimes)	0.683***	0.736***
	[0.047]	[0.053]
ln(MPCE)	-0.039	0.076
	[0.051]	[0.066]
Other Controls	Yes	Yes
District and Year Fixed Effects	Yes	Yes
Observations	2,810	2,810
R-squared	0.091	-0.561
Number of panel	562	562
First-stage Statistics		
Sanderson-Windmeijer F-stat:	31.66	24.35
Sanderson-Windmeijer p value:	0.0000	0.0000
Kleibergen-Paap rk Wald F-stat:	31.66	20.41
Stock Yogo LIML 10% maximal IV bias:	3.97	3.78

Note: Data are for the years 2004-05, 2005-06, 2007-08, 2009-10 and 2011-12. Gender gaps are defined as in equation (4). Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1. All variables are district level totals or averages for each time period: ln(Rapes, Indecent Assaults, Harassment) is the log of the total number of reports of rapes, indecent assaults and harassment. ln(Mean EP gap) is the log of the ratio of average male earning potential over average female earning potential. ln(Mean Male EP) is the log of the average male earning potential. ln(Mean Female EP) is the log of the average female earning potential. ln(Emp gap) is the log of the ratio of total male employment over total female employment. ln(Male Emp.) is the log of the total number of employed males. ln(Female Emp.) is the log of the total number of employed females. ln(Total Crimes) is the log of the total number of crimes. ln(MPCE) is the log of mean per capita household expenditure. To guard against a weak instruments problem IV regressions use continuously updated GMM estimators (CUE).

4.2 Domestic Violence

4.2.1 Earning Potential Baseline

Table 8 presents results from the estimation of equation (1), where the dependent variable is reports of domestic violence (cruelty by the husband or his relatives). In columns (1) through (3), other control variables are introduced sequentially. Across all three columns, a larger gender gap in earning potential is associated with more domestic violence, however, these results are not statistically significant. In addition, the magnitude of this effect is not economically significant. Total crime shows statistically significant positive associations with domestic violence, in line with policing or law and order effects. Increased urbanisation percentage shows a negative association with domestic violence, which could be due to close proximity with neighbours reigning in the actions of some urban males. Female representation, and the gender gap in employment, are insignificant for domestic violence.

To account for endogeneity, columns (4) and (5), show results from the instrumental variables estimation strategy. In column (4), the gender gap in earning potential is instrumented using the eight trade variables, and both earning potential and employment gaps are instrumented in column (5). From column (4), a one percent decrease in the gender employment gap is associated with a 1.9% decrease in domestic violence and the coefficient is now statistically significant at the 1% level. This is consistent with relative female empowerment increasing bargaining power and hence reducing the levels of domestic violence. The employment gap remains insignificant after instrumentation.

4.2.2 Relationship Between Earning Potential and Gender Bias

Table 9 presents the IV regression for the gender gap in earning potential on domestic violence when the sample is separated by districts in states with relatively low, medium and high levels of gender bias. Across all measures of gender bias, we observe a positive, significant and large coefficient on the gap in earning potential in regions of low gender bias, as seen in columns (2), (5) and (8). These coefficients are much larger than the baseline result, repeated in column (1), suggesting that the bargaining effect may be stronger in relatively low gender biased areas[18]. From the primary gender bias variable, shown in columns (2-4) we observe a reversal of the sign on the earning potential gap as the degree of gender bias increases. There is now a negative association between the gap in earning potential and domestic violence in areas of high gender bias. This may indicate a potential backlash effect overriding the bargaining effect when institutions or culture are more gender biased.

In the gap regressions in Table 9, the other gender bias variables only provide minimal corroboration for a potential backlash effect. However, further evidence is presented in Table 10, which decomposes the earning potential gap into male and female effects. The baseline in column (1) displays results consistent with the bargaining hypothesis. An increase in female earning potential is associated with a decrease in reports of domestic violence, with the opposite case for male earning potential. However, while the female effect is highly significant, the male effect is only significant at the ten percent level. Looking across columns (2-4), the reason for this becomes apparent. In areas of low gender bias, shown in column (2), male earning potential is positively correlated at the one percent level with domestic violence, and in regions of high gender bias,

[18] As discussed in the Data Appendix, all of India is potentially gender biased, with our variables just capturing differing degrees in this bias. The "low" gender biased areas are likely to be more biased than developed countries where the bargaining effect is well documents (Aizer (2010)).

shown in column (4), it is negatively correlated to the same significance. These results are corroborated with the other measures of gender bias, although only ESO shows a similarly significant relationship. This suggest that there may be a backlash effect, where low male earning potential leads to an increase in instances of domestic violence, that is stronger than the bargaining effect in areas of high gender bias.

In Table 11, earning potential is further decomposed into Hq and Lq male and female effects. The baseline results from Table 10 are repeated in column (1). The bargaining effects hold across varying levels of qualification, however the magnitude of the effect is larger amongst low qualified individuals. Overall we provide evidence for the existence of a bargaining effect documented for the United States by Aizer (2010) also in the case of developing countries like India. However, this effect is only present in less gender bias regions, which are more similar to the US in terms of gender bias. In areas of high gender bias, we find that the bargaining effect breaks down, and a potential backlash effect dominates.

Table 8: Gender Gaps Regressed on Domestic Violence

	ln(Domestic Violence)				
VARIABLES	OLS (1)	OLS (2)	OLS (3)	IV (4)	IV (5)
ln(Mean EP gap)	0.195	0.179	0.179	1.861***	1.678**
	[0.277]	[0.268]	[0.267]	[0.659]	[0.670]
ln(Emp gap)	0.012	0.010	0.010	0.012	-0.064
	[0.019]	[0.019]	[0.020]	[0.018]	[0.115]
ln(Total Crimes)	0.856***	0.861***	0.861***	0.836***	0.839***
	[0.069]	[0.069]	[0.068]	[0.061]	[0.061]
ln(MPCE)	0.106	0.089	0.091	0.141*	0.141*
	[0.073]	[0.073]	[0.073]	[0.072]	[0.072]
ln(Male Working population)	0.072	0.043	0.047	0.032	0.143
	[0.089]	[0.089]	[0.095]	[0.097]	[0.194]
ln(Female Working population)	0.042	0.139	0.135	0.125	0.027
	[0.090]	[0.094]	[0.100]	[0.105]	[0.189]
ln(Inequality)		-0.044	-0.045	-0.079	-0.099
		[0.064]	[0.065]	[0.060]	[0.066]
urbanisation %age		-0.067***	-0.067***	-0.059***	-0.060***
		[0.018]	[0.018]	[0.019]	[0.019]
%age emp.in Agriculture		-0.120	-0.118	-0.156	-0.323
		[0.136]	[0.136]	[0.127]	[0.281]
%age emp. in Manufacturing		-0.124	-0.123	-0.107	-0.310
		[0.225]	[0.225]	[0.237]	[0.379]
%age with Hq education		0.024	0.023	0.023	0.012
		[0.030]	[0.031]	[0.033]	[0.037]
ln(Hq gap)			-0.005	-0.004	-0.007
			[0.016]	[0.018]	[0.019]
Elected female representatives			-0.009	-0.017	-0.017
			[0.025]	[0.020]	[0.020]
Year and District FE	Yes	Yes	Yes	Yes	Yes
Observations	2,810	2,810	2,810	2,810	2,810
R-squared	0.235	0.241	0.241	0.221	0.215
Number of panel	562	562	562	562	562
First-stage Statistics					
Sanderson-Windmeijer F-stat:	-	-	-	31.66	3.17
Sanderson-Windmeijer p value:	-	-	-	0.0000	0.0025
Kleibergen-Paap rk Wald F-stat:	-	-	-	31.66	2.74
Stock Yogo LIML 10% maximal IV bias critical value:				3.97	
Stock Yogo LIML 15% maximal IV bias critical value:				2.73	

Note: Data are for the years 2004-05, 2005-06, 2007-08, 2009-10 and 2011-12. EP gap is defined as in equation (4). Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1. All variables are district level totals or averages for each time period: ln(Domestic Violence) is the log of the total number of reports of domestic violence. ln(Mean EP gap) is the log of the ratio of average male earning potential over average female earning potential. ln(Emp gap) is the log of the ratio of total male employment over total female employment. ln(Total Crimes) is the log of the total number of crimes. ln(MPCE) is the log of mean per capita household expenditure. ln(Male Working population) is the log of the total number of males aged 15-64. ln(Female Working population) is the log of the total number of females aged 15-64. ln(Inequality) is the log of the ratio of MPCE of the household at the seventy-fifth percentile relative to the household at the twenty-fifth percentile. urbanisation %age is the the percentage of households that are located in urban sectors. %age emp. in Agriculture is the percentage of individuals employed in farming, forestry and fishing as a proportion of total working population. %age emp. in Manufacturing s the percentage of individuals employed in manufacturing as a proportion of total working population. %age with Hq education is the percentage of individuals with a high school education or above as a proportion of total working population. ln(Hq gap) is the log of the ratio of total number of high school educated male over total high school educated females. Elected female representatives is the total number of sitting female elected representatives to the state legislature. To guard against a weak instruments problem IV regressions use continuously updated GMM estimators (CUE).

Table 9: IV regressions for EP gaps, regressed on Domestic Violence, Separated by Gender Bias level

	ln(Domestic Violence)									
Measure of Gender bias:	Base[19]	Inverse Decision %age			Missing Women			Inverse ESO		
VARIABLES	IV (1)	IV (2)	IV (3)	IV (4)	IV (5)	IV (6)	IV (7)	IV (8)	IV (9)	IV (10)
ln(Mean EP gap)	1.861***									
	[0.659]									
ln(Mean EP gap): Low GB		6.864***			5.801***			4.912***		
		[1.655]			[2.080]			[1.204]		
ln(Mean EP gap): Med GB			4.789***			-1.197			1.797*	
			[1.326]			[1.095]			[1.084]	
ln(Mean EP gap): High GB				-2.033**			0.965			-0.906
				[0.822]			[0.785]			[0.872]
ln(Emp. Gap)	0.012	-0.002	0.018	0.017	0.056	0.001	0.003	-0.022	-0.033	0.018
	[0.018]	[0.082]	[0.045]	[0.021]	[0.052]	[0.039]	[0.018]	[0.045]	[0.047]	[0.021]
ln(Total Crimes)	0.836***	0.476***	0.885***	1.110***	0.869***	1.032***	0.674***	0.659***	0.697***	1.057***
	[0.061]	[0.094]	[0.113]	[0.083]	[0.118]	[0.080]	[0.105]	[0.077]	[0.154]	[0.099]
ln(MPCE)	0.141*	0.080	0.216*	0.051	0.108	-0.049	0.256**	0.074	0.152	0.129
	[0.072]	[0.126]	[0.128]	[0.112]	[0.128]	[0.126]	[0.104]	[0.097]	[0.153]	[0.112]
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District and Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,810	745	915	1,150	915	875	1,020	905	650	1,255
R-squared	0.221	0.034	0.150	0.320	0.198	0.298	0.238	0.249	0.148	0.253
Number of panel	562	149	183	230	183	175	204	181	130	251
First-stage Statistics										
Sanderson-Windmeijer F-stat:	31.66	8.525	20.00	14.01	9.751	9.311	13.98	17.61	7.810	22.62
Sanderson-Windmeijer p value:	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Kleibergen-Paap rk Wald F-stat:	31.66	8.525	20.00	14.01	9.751	9.311	13.98	17.61	7.810	22.62
Stock Yogo LIML 10% maximal IV bias critical value:							3.97			

Note: Data are for the years 2004-05, 2005-06, 2007-08, 2009-10 and 2011-12. Gender gaps are defined as in equation (4). Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1. All non-gender bias variables are district level totals or averages for each time period: ln(Domestic Violence) is the log of the total number of reports of domestic violence. ln(Mean EP gap) is the log of the ratio of average male earning potential over average female earning potential. ln(Emp gap) is the log of the ratio of total male employment over total female employment. ln(Total Crimes) is the log of the total number of crimes. ln(MPCE) is the log of mean per capita household expenditure. Gender bias variables define methods to group districts in states of varying degrees of gender bias. Inverse Decision %age defines gender bias by the percentage of women who report they are not allowed to participate in at least one of four household decisions. Missing Women is the ratio of men to women outside of the fertile age range in a district, divided by the expected ratio for men to women outside the fertile age range, with the expected ratio based on India and state averages. Inverse ESO is the inverse of the Mckinsey Institute index for the percentage of women in a state with access to essential services and opportunities. The subsample regressions for regions of low gender bias are presented in columns (2), (5) and (8). Medium gender bias is represented in columns (3), (6) and (9). High gender bias is shown in columns (4), (7) and (10). To guard against a weak instruments problem IV regressions use continuously updated GMM estimators (CUE) .

[19] Repeated from table 11

Table 10: IV regressions for male and female EP, regressed on Domestic Violence, Separated by Gender Bias level

	ln(Domestic Violence)									
Measure of Gender bias:	Base	Inverse Decision %age			Missing Women			Inverse ESO		
VARIABLES	IV (1)	IV (2)	IV (3)	IV (4)	IV (5)	IV (6)	IV (7)	IV (8)	IV (9)	IV (10)
ln(Mean Male EP)	1.334*									
	[0.721]									
ln(Mean Female EP)	-2.448***									
	[0.842]									
ln(Mean Male EP): Low GB		6.356***			5.430***			4.862***		
		[1.779]			[2.026]			[1.248]		
ln(Mean Female EP): Low GB		-6.757***			-4.440*			-4.747***		
		[1.677]			[2.288]			[1.265]		
ln(Mean Male EP): Med GB			4.823***			-0.767			1.696	
			[1.370]			[1.155]			[1.442]	
ln(Mean Female EP): Med GB			-4.329***			-2.414			-1.857	
			[1.514]			[1.603]			[1.297]	
ln(Mean Male EP): High GB				-2.809***			0.236			-2.238**
				[0.925]			[1.057]			[1.024]
ln(Mean Female EP): High GB				0.807			-1.134			-0.558
				[0.981]			[0.808]			[1.024]
ln(Male Emp.)	-0.121	-0.386*	0.261	0.136	-0.039	0.297	-0.464**	-0.056	-0.169	-0.084
	[0.141]	[0.220]	[0.310]	[0.209]	[0.275]	[0.254]	[0.189]	[0.183]	[0.359]	[0.228]
ln(Female Emp.)	-0.017	0.003	-0.008	-0.025	-0.040	-0.029	-0.002	0.027	0.031	-0.031
	[0.019]	[0.081]	[0.047]	[0.022]	[0.053]	[0.042]	[0.018]	[0.046]	[0.052]	[0.023]
ln(Total Crimes)	0.816***	0.483***	0.893***	1.055***	0.883***	0.904***	0.663***	0.659***	0.699***	1.026***
	[0.062]	[0.094]	[0.114]	[0.089]	[0.118]	[0.091]	[0.101]	[0.078]	[0.152]	[0.101]
ln(MPCE)	0.110	0.057	0.226*	0.036	0.105	-0.077	0.219**	0.080	0.148	0.142
	[0.075]	[0.131]	[0.131]	[0.109]	[0.129]	[0.131]	[0.105]	[0.097]	[0.176]	[0.111]
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District and Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,810	745	915	1,150	915	875	1,020	905	650	1,255
R-squared	0.211	0.053	0.158	0.355	0.218	0.253	0.242	0.253	0.145	0.258
Number of panel	562	149	183	230	183	175	204	181	130	251
First-stage Statistics										
Sanderson-Windmeijer F-stat:	24.35	8.699	17.71	12.27	8.596	7.803	14.37	15.88	9.862	15.84
Sanderson-Windmeijer p value:	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Kleibergen-Paap rk Wald F-stat:	20.41	7.545	15.24	10.75	7.484	6.644	12.57	13.49	7.765	13.66
Stock Yogo LIML 10% maximal IV bias critical value:							3.78			

Note: Data are for the years 2004-05, 2005-06, 2007-08, 2009-10 and 2011-12. Gender gaps are defined as in equations (4). Standard errors are clustered at the district level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. All variables are district level totals or averages for each time period: ln(Domestic Violence) is the log of the total number of reports of domestic violence. ln(Mean Male EP) is the log of the average male earning potential. ln(Mean Female EP) is the log of the average female earning potential. ln(Male Emp.) is the log of the total number of employed males. ln(Female Emp.) is the log of the total number of employed females. ln(Total Crimes) is the log of the total number of crimes. ln(MPCE) is the log of mean per capita household expenditure. Gender bias variables define methods to group districts in states of varying degrees of gender bias. Inverse Decision %age defines gender bias by the percentage of women who report they are not allowed to participate in at least one of four household decisions. Missing Women is the ratio of men to women outside of the fertile age range in a district, divided by the expected ratio for men to women outside the fertile age range, with the expected ratio based on India and state averages. Inverse ESO is the inverse of the Mckinsey Institute index for the percentage of women in a state with access to essential services and opportunities. The subsample regressions for regions of low gender bias are presented in columns (2), (5) and (8). Medium gender bias is represented in columns (3), (6) and (9). High gender bias is shown in columns (4), (7) and (10). To guard against a weak instruments problem IV regressions use continuously updated GMM estimators (CUE) .

Table 11: Male and female wage and employment regressed on Domestic Violence, for gender gap, male and female separately and for Hq vs Lq individuals

	ln(Domestic Violence)		
VARIABLES	IV (1) [20]	IV(2)	IV(3)
ln(Mean Male EP)	1.334*		
	[0.721]		
ln(Mean Female EP)	-2.448***		
	[0.842]		
ln(Mean Male Lq EP)		2.079**	
		[0.989]	
ln(Mean Female Lq EP)		-3.939**	
		[1.565]	
ln(Mean Male Hq EP)			1.032**
			[0.406]
ln(Mean Female Hq EP)			-0.827***
			[0.229]
ln(Male Emp.)	-0.121		
	[0.141]		
ln(Female Emp.)	-0.017		
	[0.019]		
ln(Male Lq Emp.)		-0.136	
		[0.118]	
ln(Female Lq Emp.)		0.001	
		[0.020]	
ln(Male Hq Emp.)			-0.023
			[0.046]
ln(Female Hq Emp.)			-0.007
			[0.007]
ln(Total Crimes)	0.816***	0.815***	0.858***
	[0.062]	[0.065]	[0.065]
ln(MPCE)	0.110	0.069	0.052
	[0.075]	[0.081]	[0.081]
Other Controls	Yes	Yes	Yes
District and Year Fixed Effects	Yes	Yes	Yes
Observations; Number of panel	2,810; 562	2,810; 562	2,810; 562
R-squared	0.211	0.043	0.107
First-stage Statistics			
Sanderson-Windmeijer F-stat:	24.35	7.978	16.45
Sanderson-Windmeijer p value:	0.0000	0.0000	0.0000
Kleibergen-Paap rk Wald F-stat:	20.41	6.981	13.72
Stock Yogo LIML 10% maximal IV bias:		3.78	

Note: Data are for the years 2004-05, 2005-06, 2007-08, 2009-10 and 2011-12. Gender gaps are defined as in equations (4). Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1. All variables are district level totals or averages for each time period: ln(Domestic Violence) is the log of the total number of reports of domestic violence. ln(Mean Male EP) is the log of the average male earning potential. ln(Mean Female EP) is the log of the average female earning potential. ln(Mean Hq Male EP) is the log of the average male earning potential for males with a high-school level of education or above. ln(Mean Hq Female EP) is the log of the average female earning potential for females with a high-school level of education or above. ln(Mean Lq Male EP) is the log of the average male earning potential for males with less than a high-school level of education. ln(Mean Lq Female EP)is the log of the average female earning potential for females with less than a high-school level of education. ln(Male Emp.) is the log of the total number of employed males. ln(Female Emp.) is the log of the total number of employed females. ln(Male Hq Emp.) is the log of the total number of employed males with a high-school level of education or above. ln(Female Hq Emp.) is the log of the total number of employed females with a high-school level of education or above. ln(Male Lq Emp.) is the log of total number of employed males with less than a high-school level of education. ln(Female Lq Emp.) is the log of total number of employed females with less than a high-school level of education. ln(Total Crimes) is the log of the total number of crimes. ln(MPCE) is the log of mean per capita household expenditure. To guard against a weak instruments problem IV regressions use continuously updated GMM estimators (CUE) .

5 Conclusion

We study the relationship between gender gaps and crime against women. We find that a smaller gender gap in earning potential is associated with increased crime against women, consistent with backlash. To our knowledge, ours is the first study in the economics literature to find evidence for backlash leading to more crime against women outside of the context of domestic violence. The backlash effect is particularly exacerbated in areas exhibiting gender bias. For domestic violence, we find evidence for empowerment leading to lower violence as bargaining power for women increases in areas with low gender bias. This is in line with evidence from advanced countries like the United States. There is evidence that poorer earning potential for men in states with high gender bias is associated with more domestic violence. Our study underscores the role of gender biased institutions or culture in exacerbating crime against women as a result of backlash to better opportunities for women relative to men. In the home, female empowerment can be associated with lower violence against women by affording them greater bargaining power.

[20] Baseline Repeated from Table 10

References

Abiona, Olukorede, & Koppensteiner, Martin Foureaux. 2016. *The Impact of Household Shocks on Domestic Violence: Evidence from Tanzania*. Tech. rept. University of Leicester.

Aizer, Anna. 2010. The gender wage gap and domestic violence. *American Economic Review*, **100**(4), 1847–1859.

Alesina, Alberto, Brioschiz, Benedetta, & La Ferrara, Eliana. 2016. *Violence Against Women: A Cross-cultural Analysis for Africa*. Tech. rept. Harvard University.

Amaral, Sofia. 2014. *Do Improved Property Rights Decrease Violence Against Women in India?* Tech. rept. University of Essex.

Amaral, Sofia, & Bhalotra, Sonia. 2017. *Population Sex Ratios and Violence Against Women: The Long-Run Effects of Sex Selection in India*. Tech. rept. University of Essex.

Amaral, Sofia, Bandyopadhyay, Siddhartha, & Sensarma, Rudra. 2015. Employment programmes for the poor and female empowerment: The effect of NREGS on gender-based violence in India. *Journal of Interdisciplinary Economics*, **27**(2), 199–218.

Anderberg, Dan, Rainer, Helmut, Wadsworth, Jonathan, & Wilson, Tanya. 2016. Unemployment and Domestic Violence: Theory and Evidence. *Economic Journal*, **126**(597), 1947–1979.

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

Anderson, Siwan, & Ray, Debraj. 2012. The Age Distribution of Missing Women in India. *Economic & Political weekly*, **47**(47-48), 87–95.

Angelucci, Manuela. 2008. Love on the rocks: Domestic violence and alcohol abuse in rural Mexico. *B.E. Journal of Economic Analysis and Policy*, **8**(1), 43.

Bertrand, Marianne, Kamenica, Emir, & Pan, Jessica. 2015. Gender identity and relative income within households. *Quarterly Journal of Economics*, **130**(2), 571–614.

Brysk, Alison, & Mehta, Aashish. 2017. When development is not enough: Structural change, conflict and gendered insecurity. *Global Society*, **31**(4), 441–459.

Chin, Yoo Mi. 2011. Male backlash, bargaining, or exposure reduction?: Women's working status and physical spousal violence in India. *Journal of Population Economics*, **25**(1), 175–200.

Cools, Sara, Flatø, Martin, & Kotsadam, Andreas. 2015. *Weather Shocks and Violence Against Women in Sub-Saharan Africa*. Tech. rept.

Eswaran, Mukesh, & Malhotra, Nisha. 2011. Domestic violence and women's autonomy in developing countries: Theory and evidence. *Canadian Journal of Economics*, **44**(4), 1222–1263.

Hidrobo, Melissa, Peterman, Amber, & Heise, Lori. 2016. The effect of cash, vouchers, and food transfers on intimate partner violence: Evidence from a randomized experiment in Northern Ecuador. *American Economic Journal: Applied Economics*, **8**(3), 284–303.

International Institute for Population Sciences. 2007. National Family Health Survey (NFHS-3), 2005-06: India: Volume I. *International Journal of Health Care Quality Assurance*.

Iyer, Lakshmi, Mani, Anandi, Mishra, Prachi, & Topalova, Petia. 2012. The power of political voice: Women's political representation and crime in India. *American Economic Journal: Applied Economics*, **4**(4), 165–193.

Klasen, Stephan, Minasyan, Anna, Zenker, Juliane, & Vollmer, Sebastian. 2018. *Educational gender gaps and economic growth: A systematic review and meta-regression analysis*. Tech. rept. Courant Research Centre: Poverty, Equity and Growth.

Martin, Kimberly, Vieraitis, Lynne M., & Britto, Sarah. 2006. Gender equality and women's absolute status: A test of the feminist models of rape. *Violence Against Women*, **12**(4), 321–339.

Munyo, Ignacio, & Rossi, Martín A. 2015. *The Effects of Real Exchange Rate Fluctuations on the Gender Wage Gap and Domestic Violence in Uruguay Institutions for Development Sector Inter-American Development Bank*. Tech. rept. Inter-American Development Bank (IDB).

Sekhri, Sheetal, & Storeygard, Adam. 2014. Dowry deaths: Response to weather variability in India. *Journal of Development Economics*, **111**, 212–223.

Topalova, Petia. 2010. Factor immobility and regional impacts of trade liberalization: Evidence on poverty from India. *American Economic Journal: Applied Economics*, **2**(4), 1–41.

Whaley, R B. 2001. the Paradoxical Relationship Between Gender Inequality and Rape. *Gender & Society*, **15**(4), 531–555.

Zaluar, Alba. 1995. *The drug trade, crime and policies of repression in Brazil*. Tech. rept. 1.

Appendix I

Table 12: First Stage for IV Regressions when Instrumenting the gap in Earning Potential, the gap in Earning Potential and the gap in Employment, and when Instrumenting Male and Female Earning Potential.

	E.g Table 3 column (4)	E.g Table 3 column (5)		E.g Table 6 column (2)	
VARIABLES	ln(Mean EP gap) IV First-Stage (1)	ln(Mean EP gap) IV First-Stage (2a)	ln(Emp gap) IV First-Stage (2b)	ln(Mean EP male) IV First-Stage (3a)	ln(Mean EP fem) IV First-Stage (3b)
ETT for Lq males	0.0118***	0.0119***	-0.0287	0.0135***	0.0017
	[0.003]	[0.003]	[0.033]	[0.001]	[0.003]
ETT for Lq females	-0.0083***	-0.0084***	0.0419**	-0.0002	0.0081***
	[0.002]	[0.002]	[0.021]	[0.000]	[0.002]
ETT for Hq males	0.0048**	0.0047**	0.0539*	0.0048***	0.0000
	[0.002]	[0.002]	[0.029]	[0.001]	[0.002]
ETT for Hq females	-0.0003	-0.0003	-0.0136	-0.0002	0.0002
	[0.001]	[0.001]	[0.019]	[0.000]	[0.001]
EIT for Lq males	-0.0190***	-0.0191***	0.0262	-0.0202***	-0.0012
	[0.003]	[0.003]	[0.029]	[0.001]	[0.003]
EIT for Lq females	0.0153***	0.0155***	-0.0761***	-0.0004	-0.0157***
	[0.003]	[0.003]	[0.029]	[0.000]	[0.003]
EIT for Hq males	-0.0035***	-0.0032***	-0.0832***	-0.0029***	0.0005
	[0.001]	[0.001]	[0.028]	[0.000]	[0.001]
EIT for Hq females	0.0018***	0.0017***	0.0269	0.0003	-0.0015***
	[0.001]	[0.001]	[0.019]	[0.000]	[0.001]
ln(Emp. gap)	-0.0028				
	[0.002]				
ln(Emp. male)				-0.0035	0.0051
				[0.004]	[0.010]
ln(Emp. fem)				-0.0000	-0.0028*
				[0.000]	[0.002]
ln(Total Crimes)	0.0044	0.0043	0.0319	0.0032**	-0.0013
	[0.003]	[0.003]	[0.065]	[0.001]	[0.003]
ln(MPCE)	0.0051	0.0049	0.0644	0.0000	-0.0049
	[0.009]	[0.008]	[0.126]	[0.002]	[0.008]
Other Controls	Yes	Yes	Yes	Yes	Yes
Year and District FE	Yes	Yes	Yes	Yes	Yes
Observations	2,810	2,810	2,810	2,810	2,810
Number of panel	562	562	562	562	562
Sanderson-Windmeijer F-stat:	31.66	29.17	3.17	60.39	24.35
Sanderson-Windmeijer p value:	0.0000	0.0000	0.0025	0.0000	0.0000

Note: These first stage results are the same for all regression in this paper which instrument the same variable or set of variables. Data are for the years 2004-05, 2005-06, 2007-08, 2009-10 and 2011-12. Gender gaps are defined as in equations (4). Standard errors are clustered at the district level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. All variables are district level totals or averages for each time period: ln(Mean EP gap) is the log of the ratio of average male earning potential over average female earning potential. ln(Mean Male EP) is the log of the average male earning potential. ln(Mean Female EP) is the log of the average female earning potential. ln(Emp gap) is the log of the ratio of total male employment over total female employment. ln(Male Emp.) is the log of the total number of employed males. ln(Female Emp.) is the log of the total number of employed females. ln(Total Crimes) is the log of the total number of crimes. ln(MPCE) is the log of mean per capita household expenditure. ETT for Lq males is the district exposure to trade tariffs, as weighted by employment of low qualified males. ETT for Lq females is the district exposure to trade tariffs, as weighted by employment of low qualified females. ETT for Hq males is the district exposure to trade tariffs, as weighted by employment of high qualified males. ETT for Hq females is the district exposure to trade tariffs, as weighted by employment of high qualified females. EIT for Lq males is the district exposure to input trade tariffs, as weighted by employment of low qualified males. EIT for Lq females is the district exposure to input trade tariffs, as weighted by employment of low qualified females. EIT for Hq males is the district exposure to input trade tariffs, as weighted by employment of high qualified males. EIT for Hq females is the district exposure to input trade tariffs, as weighted by employment of high qualified females. Individuals are classified as High qualified (Hq) if they posses a high-school level of education or above. Otherwise they are are classified as Low qualified (Lq). To guard against a weak instruments problem IV regressions use continuously updated GMM estimators (CUE) .

APPENDIX I

Table 13: IV regressions for male and female Earning Potential, regressed on Rapes and Indecent Assaults, Separated by Gender Bias level

	ln(Rapes, Indecent Assaults)									
Measure of Gender bias:	Base	Inverse Decision %age			Missing Women			Inverse ESO		
VARIABLES	IV (1)	IV (2)	IV (3)	IV (4)	IV (5)	IV (6)	IV (7)	IV (8)	IV (9)	IV (10)
ln(Mean Male EP)	-1.766***									
	[0.525]									
ln(Mean Female EP)	1.742***									
	[0.584]									
ln(Mean Male EP): Low GB		1.933*			0.766			0.087		
		[1.148]			[0.925]			[1.032]		
ln(Mean Female EP): Low GB		-1.316			-1.711*			1.096		
		[1.025]			[1.013]			[0.995]		
ln(Mean Male EP): Med GB			0.452			-1.006			-0.933	
			[0.757]			[1.080]			[0.799]	
ln(Mean Female EP): Med GB			1.090			4.496***			0.717	
			[0.671]			[1.229]			[0.748]	
ln(Mean Male EP): High GB				-3.090***			-2.286***			-2.573***
				[0.879]			[0.855]			[0.775]
ln(Mean Female EP): High GB				2.890***			1.351*			1.895**
				[0.854]			[0.745]			[0.839]
ln(Male Emp.)	0.037	-0.079	0.260	0.146	-0.061	-0.196	0.167	-0.029	0.032	0.306*
	[0.106]	[0.207]	[0.159]	[0.171]	[0.160]	[0.191]	[0.176]	[0.170]	[0.194]	[0.159]
ln(Female Emp.)	0.051***	-0.050	0.108***	0.049***	0.026	0.041	0.058***	-0.013	0.108***	0.041***
	[0.012]	[0.055]	[0.025]	[0.015]	[0.028]	[0.027]	[0.017]	[0.048]	[0.032]	[0.014]
ln(Total Crimes)	0.769***	0.718***	0.950***	0.809***	0.912***	0.864***	0.586***	0.811***	0.547***	0.824***
	[0.050]	[0.103]	[0.082]	[0.073]	[0.087]	[0.069]	[0.088]	[0.071]	[0.112]	[0.077]
ln(MPCE)	-0.079	-0.166	-0.012	-0.079	-0.044	-0.197**	-0.113	-0.178**	-0.053	-0.090
	[0.049]	[0.107]	[0.075]	[0.082]	[0.070]	[0.096]	[0.085]	[0.088]	[0.080]	[0.078]
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District and Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,810	745	915	1,150	915	875	1,020	905	650	1,255
R-squared	0.208	0.248	0.351	0.047	0.335	-0.183	0.126	0.328	0.222	0.219
Number of panel	562	149	183	230	183	175	204	181	130	251
First-stage Statistics										
Sanderson-Windmeijer F-stat:	24.35	8.699	17.71	12.27	8.596	7.803	14.37	15.88	9.862	15.84
Sanderson-Windmeijer p value:	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Kleibergen-Paap rk Wald F-stat:	20.41	7.545	15.24	10.75	7.484	6.644	12.57	13.49	7.765	13.66
Stock Yogo LIML 10% maximal IV bias critical value:							3.78			

Note: Data are for the years 2004-05, 2005-06, 2007-08, 2009-10 and 2011-12. Gender gaps are defined as in equations (4). Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1. All non-gender bias variables are district level totals or averages for each time period: ln(Rapes, Indecent Assaults) is the log of the total number of reports of rapes and indecent assaults. ln(Mean EP gap) is the log of the ratio of average male earning potential over average female earning potential. ln(Emp gap) is the log of the ratio of total male employment over total female employment. ln(Total Crimes) is the log of the total number of crimes. ln(MPCE) is the log of mean per capita household expenditure. Gender bias variables define methods to group districts in states of varying degrees of gender bias. Inverse Decision %age defines gender bias by the percentage of women who report they are not allowed to participate in at least one of four household decisions. Missing Women is the ratio of men to women outside of the fertile age range in a district, divided by the expected ratio for men to women outside the fertile age range, with the expected ratio based on India and state averages. Inverse ESO is the inverse of the Mckinsey Institute index for the percentage of women in a state with access to essential services and opportunities. The subsample regressions for regions of low gender bias are presented in columns (2), (5) and (8). Medium gender bias is represented in columns (3), (6) and (9). High gender bias is shown in columns (4), (7) and (10). To guard against a weak instruments problem IV regressions use continuously updated GMM estimators (CUE) .

Appendix II: Data Appendix

Data sources

Industry, district, household, and individual level information from several large datasets for multiple time periods are integrated and aggregated in this paper. Data on gender gaps and control variables are sourced from the Employment and Unemployment surveys of the National Sample Survey Organization (NSSO), India. We use five rounds of data for the years 2004-05, 2005-06, 2007-08, 2009-10 and 2011-12. The surveys collect information on household per capita expenditure, the district, state and rural or urban area the household is located in, and individual variables like age, gender, level of education, employment status in the principal activity, industry of employment and daily wages. State and district administrative changes in effect in later rounds are concorded to the geopolitical boundaries as defined in NSS 61 for 2004-05, the first round of our data. Similarly, NIC industry classifications defined in 2004 and 2008 are concorded to NIC 1998. Each survey begins in July and ends in June the following year. We classify all information collected in a given round under the year in which the survey began. For example, NSSO round 61 collected data though 2004 to 2005. This is classified in this study as $t = 2004$.

Data on crime are at the district level, sourced from the National Crime Records Bureau (NCRB), India. We consider Rapes and Assaults on the Modesty of Women (Indecent Assaults) under crimes against women. A variation of this variable that is more sensitive to reporting bias also includes Insult to the Modesty of Women (Harassment). Cruelty by Husband or his Relatives is used as the measure of domestic violence. Data on total crimes at the district level registered under the Indian Penal Code are also sourced from the NCRB. Crime data is recorded as year end crime statistics for crime committed during the years of 2005, 2006, 2008, 2010 and 2012. Our interest is the effect of employment and wage gaps on crime hence crime data is concorded with a half year lead on the NSSO information. For example, crimes committed during January to December 2005 is allocated to $t = 2004$ in this analysis. This helps to correct for some of the potential reverse causation of crime on employment or wages, which is further accounted for with the IV analysis.

We collect data on the number of female representatives in the district elected to the state legislature from the Election Commission's Election Results for 2000-2010[21]. This is implanted as a snapshot of the number of sitting female politicians as of the year end for each year of 2003, 2004, 2006, 2008 and 2010. We are interested in controlling for any impact female representation has on gender gaps or crime. Hence this data is integrated with a 6 month lag on the NSSO data. For example, the total number of female representatives at year end in 2003 is allocated to $t = 2004$.

Data on import tariffs come from product level data from the WITS World Bank database. This is converted to overall tariffs for each NIC industry classification based on the simple average of tariffs on the

[21] Data accessed from http://eci.nic.in/eci_main1/ElectionStatistics.aspx on 13 November 2017

products being produced by each industry. Input tariffs are calculated using a weighted average of the tariffs on products used as inputs to each industry. Tariff information is recorded as a snapshot of product tariffs at year end for the years of 2003, 2004, 2006, 2008 and 2010. Given that we expect employment and wages to react sluggishly to changes in tariff levels, we integrate tariffs with a half year lag on the NSSO data. For example tariff data for the year ending in 2003 is allocated to $t = 2004$.

State level variables describing potential gender bias are derived from three sources. The primary variable uses survey information on the participation of women in household decision making from the third National Family Health Survey (NHFS) from 2004-2005. A secondary measure uses demographic information from the NSSO survey to constructs a variation of the Missing Women measure described in Anderson & Ray (2012). The final bias variable uses an index of access for women to essential services and opportunities in the workplace produced by the McKinsey Global Institute in 2015[22].

Consumer Price Index (CPI) data comes from the State Level Consumer Price Index (Rural/Urban) for 2011, published by the Central Statistics Office of India[23]. This dataset contains statewise urban and rural CPI modifiers for 2011 which are used to deflate the wage data.

Definition of types of individuals

Many of the following variables are defined for a person of group k. Primary classification is either $male$ or $female$ according to the gender information provided by an individual in the NSSO survey. We further categorise individuals into 2 subtypes based on their reported education level. An individual is $High\,Qualified$ (Hq) if they have a high-school education or above, and $Low\,Qualified$ (Lq) if not. The detailed classification hence defines 4 types of individual: $Hq\ male$, $Hq\ female$, $Lq\ male$ and $Lq\ female$. Individuals not reporting an education level are included in any variables using the gender classifications and excluded from any variables using the gender-skill classifications (~2,000 observations out of ~2,500,000).

Mean Per Capita Expenditure (MPCE)

MPCE is derived from the NSSO household level survey data. Monthly household consumer expenditure is defined in the survey as the following: Total household spending on non-durable goods for the 30 days prior to the survey, added to a monthly average from the previous 365 days of spending on durable goods and services. Non-durable goods include taxes, rent, transportation and utilities. Examples of durable goods include furniture, appliances, vehicles and long term services such as tuition fees and institutional medical expenses.

[22]Data accessed from https://www.mckinsey.com/featured-insights/employment-and-growth/the-power-of-parity-advancing-womens-equality-in-india on 13 November 2017

[23]Data accessed from https://data.gov.in/resources/state-level-consumer-price-index-ruralurban-2011 on 13 November 2017

To create per person expenditure, we divide the monthly consumer expenditure for each household by the number of individuals in that household. Similar to wages and other income data, there is a large right tail in the distribution of per person expenditure in a district. There are also a non-zero number of households reporting no per person expenditure. These mainly represent agrarian households who consume their own production and barter for other goods and services. To facilitate comparison and preserve those zeros we take the natural logarithm of one plus the per person expenditure. Finally, to aggregate to the district level, this variable is averaged across all households in a district using the household level weightings provided in the NSSO survey data. To summarise, MPCE in district i at time t as defined as:

$$lnMPCE_{i,t} = \sum_{h=1}^{H_{i,t}} ln\left(1+\frac{E_h}{N_h}\right) * \frac{W_h}{W_{H_{i,t}}} \qquad h \in H_{i,t},$$

where, h belongs to set $H_{i,t}$, the total number of households in district i at time t. E_h is the monthly household consumer expenditure in household h. N_h is the total number of individuals in household h. W_h is the household survey weighting for household h. $W_H = \sum_{h=1}^{H_{i,t}} W_h$, hence $\frac{W_h}{W_{H_{i,t}}}$ is the average household survey weighting for household h when aggregating to the district level. Using this weighting means that the average per person expenditure is as representative as possible at the district level, given the nature of the survey process. $ln\,(x)$ represents the natural logarithm transformation of x.

Households not reporting any expenditure are excluded from the MPCE calculations, along with their associated survey weightings (7 out of ~530,000 households). These households, and the individuals therein, are not necessarily excluded from other variable calculations.

Inequality

The level of inequality in a district is derived from the per person expenditure for each household in a district, as defined in 5. Inequality in i at time t is defined as the ratio of per person expenditure of the 75th percentile household, divided the per person expenditure of the 25th percentile household. Percentiles are calculated after taking account of the representative survey weighting. Hence: inequality is in district i at time t is defined as following:

$$lnIneq_{i,t} = ln\left(1+\frac{E_{h=p75}}{N_{h=p75}}\right) - ln\left(1+\frac{E_{h=p25}}{N_{h=p25}}\right) \qquad h \in H_{i,t},$$

where $h = p25\,(p75)$ represents the household in i at the 25th (75th) percentile after taking account of the survey weighting for each household in a district. E_h is the monthly household consumer expenditure in household h. N_h is the total number of individuals in household h. Households not reporting any expenditure are excluded from the Inequality calculations, along with their associated survey weightings.

Employment

Information on employment comes from the individual data in the NSSO surveys. Individuals reported their usual principal activity over the 7 days prior to the survey date, which is classified into one of 13 different categories. Examples of such categories include: attended domestic duties, attended educational institution, and did not work but looking for work. We categorise a person as being employed if they met one of the following 6 categories: worked in household enterprise (self-employed) as own account worker, worked in household enterprise (self-employed) as an employer, worked as helper in household enterprise (unpaid family worker), worked as regular salaried / wage employee, worked as casual wage labour in public works, or worked as casual wage labour in other types of work. The total employment of people of type k in district i at time t is given by:

$$lnEmp_{i,t,k} = ln\left(1 + \sum_{n=1}^{N_{i,t,k}} emp_n * W_n\right) \quad n \in N_{i,t,k},$$

where, n is in the set of $N_{i,t,k}$, which is the total number of individuals of type k in district i at time t. emp_n is a dummy variable that is 1 if person n is categorised as employed, and 0 otherwise. W_n is the survey weighting for person n. Some districts have zero employed people of a certain type, so we take the natural logarithm of one plus this variable to preserve these zeros. The relative gaps in employment in district i at time t are defined as follows:

$$lnEmp_{i,t,gap} = lnEmp_{i,t,male} - lnEmp_{i,t,female},$$

$$lnEmp_{i,t,Lq\,gap} = lnEmp_{i,t,Lq\,male} - lnEmp_{i,t,Lq\,female},$$

$$lnEmp_{i,t,Hq\,gap} = lnEmp_{i,t,Hq\,male} - lnEmp_{i,t,Hq\,female}.$$

Industry level employment for 2004, the first period of data, is used when deriving the wage and trade variables. The number of people of type k employed in industry j in district i in 2004 is given by:

$$Emp_{i,04,k,j} = \sum_{n=1}^{N_{i,04,k,j}} emp_n * W_n \quad n \in N_{i,04,k,j},$$

where, n is in the set of $N_{i,04,k,j}$, which is the total number of individuals of type k in district i in 2004 in j, where j is a NIC 1998 2 digit industry classification. emp_n is a dummy variable that is 1 if person n is categorised as employed, and 0 otherwise. (Definitionally, there is no industry information for individuals who are not classified as employed). W_n is the survey weighting for person n. Similarly, district level employment is defined as:

$$Emp_{i,04,k} = \sum_{n=1}^{N_{i,04,k}} emp_n * W_n \quad n \in N_{i,04,k},$$

where, n is in the set of $N_{i,04,k}$, which is the total number of individuals of type k in 2004 in district i. emp_n is a dummy variable that is 1 if person n is categorised as employed, and 0 otherwise. W_n is the survey weighting for person n.

Earning Potential

Industry Average wages

Wage data comes from the NSSO survey. Individuals reported their daily earnings each of the 7 days prior to the survey date and the intensity of work on that day (either full or part time). The NIC industry code for individual's principal employment activity is also recorded. When deriving the total number of days worked for an individual for the week, each full time day is counted as 1, and each day of part time work is counted at 0.5. We use the total weekly earnings and the total number of days worked to derive the average daily wage for each person. The India average wage at time t for an individual of type k in industry j is now defined as follows:

$$ln\overline{Wage}_{t,k,j} = \sum_{n=1}^{N_{t,k,j}} ln\left(1 + \overline{wage}_n\right) * \frac{W_n}{W_{N_{t,k,j}}} \quad n \in N_{t,k,j},$$

where, n belongs to the set of $N_{t,k,j}$ which is the total number of individuals at time t of type k in industry j. $\overline{wage}_n$ is the average cash wage for individual n from the previous 7 days. k is the person's type. j is the two digit NIC industry code for the primary industry of employment for the individual. All rounds of data are concorded to NIC 1998 industry codes. At the two digit level we have ˜60 industry categories. W_n is the individual weighting for person n. $W_{N_{t,k,j}} = \sum_{n=1}^{N_{t,k,j}} W_n$, hence $\frac{W_{n_{k,j}}}{W_{N_{k,j}}}$ is the individual average survey weighting.

Individuals not reporting a wage are excluded from the average wage calculations, even if they report employment. The associated survey weightings for such individuals are also excluded. Any individual of group $-k$, along with their associated survey weightings, are excluded when calculating the average wages for individuals of group k. There are also some individuals reporting zero wages (˜2,300 out of ˜387,000). For example, some of these individuals report working in family run business. To preserve these zeros we add one to each reported wage before performing the natural logarithm transformation.

For robustness, several other versions of this variable are created. Firstly, in addition to reported cash wages, we also include payments received "in kind", i.e goods or services received in lieu of wages and other income like rent, returns on assets etc. We define $\overline{wage\,tot}_n$as the average of the sum of cash and payments in kind received over the last 7 days, adjusted for part time work as appropriate. An individual reporting cash wage information but no "in kind" information is treated as reporting zero "in kind" payments, and vice versa. Any individual not reporting either wage is excluded.

State level urban/rural CPI adjustment is achieved by replacing $\overline{wage}_n$ with $\overline{wage\,CPI}_n$ where $\overline{wage\,CPI}_n = \overline{wage}_n * CPI$. CPI is a specific purchasing power adjustment based on an individual's state of residence and their urban or rural status. This CPI information is for the earliest time period available (2011)[24] and the same deflater is used for each round of survey data. Inflation over time, and district level CPI adjustments, are naturally captured in the time and district fixed effects in each regression. The purpose of the urban-rural state deflater is to account for CPI differences not captured in those fixed effects.

Further robustness checks are achieved by using "excluding own district" Indian average wages, defined as the following:

$$ln\overline{Wage}_{i,t,k,j} = \sum_{n=1}^{N_{-i,t,k,j}} ln\left(1 + \overline{wage}_n\right) * \frac{W_n}{W_{N_{-i,t,k,j}}} \qquad n \in N_{-i,t,k,j},$$

where i references the district. In the above, each India average industry wage is calculated for a district i based on information form all the individuals in $-i$, i.e. all individuals not in that district. Given that there are 565 districts this variable is very similar to the main variable of interest.

Earning Potential (EP)

Earning potential in i is now defined as the potential a person of type k could earn in district i, based on the India average earnings for their type, and the industrial employment composition of the district in which they reside:

$$ln\overline{EP}_{i,t,k} = \frac{\sum_{j=1}^{J_i}\left(Emp_{i,04,k,j} * ln\overline{Wage}_{t,k,j}\right)}{\sum_{j=1}^{J_i} Emp_{i,04,j,k}}$$

where, $Emp_{i,04,k,j}$ is the total number of people in district i of type k employed in industry j in 2004, the first period for our data. If there are no reported wages at time t for type k in industry$j$ (across the whole of India), then the number of type k people employed in j in district i (in 2004) are excluded when calculating the total employment of type k in i. This means that if average wage information is missing for an industry, those industries are excluded in i from the employment weightings and from $\sum_{j=1}^{J_i} \Upsilon_{t,k,j} * Emp_{i,04,k,j}$. Industries with wages of zero are included. (We will show wage information is much more sparse than employment information in the next few paragraphs). There are a few districts where total employment is zero for some k. For example 38 districts out of 565, (6%), report zero employed Hq women in 2004. In these circumstances

[24] Data accessed from https://data.gov.in/resources/state-level-consumer-price-index-ruralurban-2011 on 13 November 2017

we define the earning potential for Hq women as zero. If Hq women are not employable in a district, then the earning potential for Hq women is definitionally zero. Hence if $\sum_{j=1}^{J_i} \Upsilon_{t,k,j} * emp_{i,04,k,j} = 0$ we set $ln\overline{WD}_{i,t,k} = 0$.

It could be more useful to create state level earning potential using state level industry average wages instead of the all India averages. However, we are not able to do this as wage information is relatively under-reported in our dataset.To illustrate the relative lack of wage information, we can look at the 2004-2005 survey information. Of ~602,000 individual survey participants (prior to survey weighting), ~220,000 reported employment and ~86,000 reported a wage. It is not appropriate to replace the missing wage information with zeros. Some individuals may have chosen not to report a wage, others may be business owners who don't officially take a salary but make all purchases through the business. To ensure that the wage information is as representative as possible we minimise the number of bins into which these data points are categorised. These categories multiply quickly when using 60 industry classifications and 4 groups of individuals. Further subdivision by the 32 states spreads the data too thin, even if we use the NIC 1 digit codes which results in 10 industry classifications instead of 60. Too few individuals in a given category can amplify any sampling errors and result in unrepresentative information for that category after survey weighting multipliers are applied.

In the robustness checks, $ln\overline{Wage}_{t,j,k}$ is replaced by $ln\overline{Wage\,tot}_{t,j,k}$, $ln\overline{Wage\,CPI}_{t,j,k}$, $ln\widetilde{Wage}_{t,j,k}$ or $ln\overline{Wage}_{i,t,j,k}$ as defined in section 5. The relative gaps in earning potential in district i are defined as follows:

$$ln\overline{EP}_{i,t,gap} = ln\overline{EP}_{i,t,male} - ln\overline{EP}_{i,t,female},$$

$$ln\overline{EP}_{i,t,Lq\,gap} = ln\overline{EP}_{i,t,Lq\,male} - ln\overline{EP}_{i,t,Lqfemale},$$

$$ln\overline{EP}_{i,t,Hq\,gap} = ln\overline{EP}_{i,t,Hqmale} - ln\overline{EP}_{i,t,Hq\,female}.$$

Working population and Qualification level,

Information on the working population and skill level comes from the NSSO surveys where individuals reported their age and level of education. We categorise a person as being of working population if they are aged between 15 and 64. We categorise a person as being Hq if they posses a high school level of education or above. Otherwise they are categorised as Lq. For the following variables the notation k' is used to signify that only 2 types of person are considered: $male$ and $female$.. Any person not reporting an education level is excluded from the education calculation and not the working population variable (~2,000 observations out of ~2,500,000). The total number of working age population of people of type k' in district i at time t is given by:

$$lnWorkPop_{i,t,k'} = ln\left(1 + \sum_{n=1}^{N_{i,t,k}} wp_n * W_n\right) \qquad n \in N_{i,t,k'},$$

and the total number of working age people of type k' with high school education or above in district i at time t is given by:

$$lnHHC_{i,t,k'} = ln\left(1 + \sum_{n=1}^{N_{i,t,k}} hhc_n * wp_n * W_n\right) \qquad n \in N_{i,t,k'},$$

where, n is in the set of $N_{i,t,k'}$, which is the total number of individuals of type k' in district i at time t. wp_n is a dummy variable that is 1 if person n is categorised as of working population, and 0 otherwise. hhc_n is a dummy variable that is 1 if person n has a high school level of eduction or above, and 0 otherwise. W_n is the survey weighting for person n. There are no districts with zero working population or Hq individuals, however if there were, these zeros would need to be preserved. In keeping with other variables, we take the natural logarithm of one plus. Working population is used in the regressions as $male$ and $female$ separately. The qualification level is used as a gap, defined by:

$$lnHq_{i,t,gap} = lnHHC_{i,t,male} - lnHHC_{i,t,female}.$$

Further to this, the overall working age qualification level of district i at time t is defined as the total number of high school and above educated individuals of working age, divided by the total number of working age individuals:

$$lnHq\,Level_{i,t} = ln\left(1 + HHC_{i,t,male} + HHC_{i,t,female}\right) - ln\left(1 + WorkPop_{i,t,male} + WorkPop_{i,t,female}\right),$$

where $HHC_{i,t,k'}$ is the total number of high-school educated individuals of working age of type k'. $WorkPop_{i,t,k'}$ is the total number of individuals of working age of type k'.

Urbanisation Percentage

The urbanisation percentage of a district is derived from the NSSO household level survey data. Households are recorded as being either urban or rurally located. The urbanisation percentage is defined as the total number of individuals living in urban households divided by the total number of individuals in all households in the district:

$$lnUrban\%_{i,t} = ln\left(1 + \sum_{h=1}^{H_{i,t}} N_h * ub_h * W_h\right) - ln\left(1 + \sum_{h=1}^{H_{i,t}} N_h * W_h\right) \qquad h \in H_{i,t},$$

where, h belongs to set $H_{i,t}$, the total number of households in district i at time t. ub_h is a dummy variable that is 1 if household h is classified as urban and 0 otherwise. N_h is the total number of individuals in household h. W_h is the household survey weighting for household h. In keeping with other variable creation, the natural logarithm of 1 plus the variable is used to preserve any potential zeros.

Agricultural and Manufacturing percentage of a district

The percentages of people employed in agriculture and in manufacturing are derived from the NSSO individual level survey data. We categorise a person as being employed in Agriculture if they report a NIC98 industry code corresponding to Agriculture, Forestry or Fishing. We categorise a person as working in the manufacturing sector if they report one of 22 manufacturing NIC98 industry codes. The total number of employed people at time t in agriculture or manufacturing in district i is defined as follows:

$$lnEmp_{i,t,Agriculture} = ln\left(1 + \sum_{n=1}^{N_{i,t}} empA_n * W_n\right) \qquad n \in N_{i,t},$$

$$lnEmp_{i,t,manufacturing} = ln\left(1 + \sum_{n=1}^{N_{i,t}} empM_n * W_n\right) \qquad n \in N_{i,t},$$

where, n is in the set of $N_{i,t}$, which is the total number of individuals in district i at time t. $empA_n$ is a dummy variable that is 1 if person n is categorised as employed in agriculture, and 0 otherwise. $empM_n$ is the same for manufacturing. W_n is the survey weighting for person n. The relative percentages of employment in district i at time t are defined as follows:

$$lnEmp\%_{i,t,Agriculture} = lnEmp_{i,t,Agriculture} - ln\left(1 + Emp_{i,t,male} + Emp_{i,t,female}\right)$$

$$lnEmp\%_{i,t,manufacturing} = lnEmp_{i,t,manufacturing} - ln\left(1 + Emp_{i,t,male} + Emp_{i,t,female}\right),$$

where $Emp_{i,t,male}$ and $Emp_{i,t,female}$ are as defined in section 5.

Elected Female representatives

Data on female representation come from the Election Commission's Election Results. The number of female candidates elected to any Assembly Constituency (AC) in a district is recorded for each year between 2000 and 2011. Representatives sit for a term of five years, however AC elections are staggered so there are a number of AC elections every year in each district. We count the number of elected females currently sitting as of the end of the year preceding the survey date. i.e. the current number of elected representatives as of the end of 2003 is allocated to $t = 2004$. This variable is created by adding up the number of female representative elected in a district in each of the preceding five years:

$$ElecFem_{i,t} = \sum_{y=t-5}^{y=t-1} elecfem_{i,y},$$

where $elecfem_{i,y}$ is the number of female representatives elected in district i in year y. Note that the available Electoral Commission results only go as far back as the year 2000. Hence for $t = 2004$ we miss any incumbents who were elected in 1999 and are sitting their fifth year. Districts that do not report any elected females for a given year are considered to have elected zero female representatives in that year.

State level Gender Bias

Decision Making Percentage (DC%)

The primary measure of potential gender bias comes from the third National Family Health Survey for 2005-2006. They produce a variable for the percentage of women in a state who say that they participate, either alone or with their partner, in four household decisions. The decisions are regarding the following: the woman's own health care, making major household purchases, making purchases for daily household needs and visits to the woman's family or relatives. This variable best indicates how all of India is potentially gender biased, and how our gender bias variables may just be capturing the differing degrees of bias. For example, the India average for the percentage of women who report participating in the decision of when to see her own family is only 60%, with the maximum reaching just over 90%. This means that even in the best state, almost 10% of women have no say in seeing their own family. The India average for women reporting participation in all 4 decisions is only 37%. We take the inverse of this variable, given by:

$$invDC = \left(\frac{100 - DC}{100}\right).$$

Higher $invDC$ indicates less participation in decision making, and hence potentially more gender bias in a state

Missing Women

Following Anderson & Ray (2012) we consider a relative lack of women outside of the fertile age range (OFA) to be a potential indicator of gender bias in a state. To create a structural measure of gender bias, only the first period is considered, hence in the following $t = 2004$ and has been omitted for simplicity. We use individual age information reported in the NSSO survey data. The population of OFA *male* or *female* individuals in a state is given by:

$$P_{s,k'} = \sum_{n=1}^{N_{s,k'}} p_n * W_n \quad n \in N_{s,k'},$$

where, k' is either $male$ or $female$. n is in the set of $N_{s,k'}$, which is the total number of individuals of type k' in state s. p_n is a dummy variable that is 1 if person n is aged under 16 or older than 45, and 0 otherwise. W_n is the survey weighting for person n. We further define the expected OFA population of individuals of type k' for state s as the following:

$$\hat{P}_{s,k'} = \gamma_s * \sum_{s=1}^{S} P_{s,k'},$$

where $\sum_{s=1}^{S} P_{s,k'}$ is the total OFA population of type k' in all sates in India. γ_s is the fraction of India's total population attributable to state s. When working out the population fractions all ages are included. This creates an expected OFA population for the state based on how many OFA individuals of that type there are in India and the overall proportion of the total Indian population residing in that state. We now define the missing women ratio for a state s as the ratio of OFA men to OFA women, divided by the expected ratio of OFA men to OFA women:

$$MW_s = \frac{\frac{P_{s,male}}{P_{s,female}}}{\frac{\hat{P}_{s,male}}{\hat{P}_{s,female}}} = \frac{P_{s,male}}{P_{s,female}} * \frac{\hat{P}_{s,female}}{\hat{P}_{s,male}},$$

where, $\hat{P}_{s,k'}$ represents the expected OFA population of individuals of type k' outside the fertile age range for state s.

This variable effectively represents deviation from the state average ratio of men to women outside of the fertile age range. Although the baseline for this result is the India average, which itself may be biased when compared to other countries, it allows states to be ranked, from least to most extreme results. Table 14 shows the comparison between all the gender bias variables and the missing women variables presented in Anderson & Ray (2012). Broadly, our missing women variables are in line with the results in Anderson and Ray.

Essential Services and Opportunities (ESO)

An alternative to DC% and Missing Women for a potential measure of gender bias comes from the McKinsey Global Institute (MGI) report: the power of parity: advancing women's equality in India, from November 2015. Given the nature of societal gender bias, it is likely areas of high gender bias in 2015 were also of high gender bias in 2004-2011. Furthermore, the MGI variables are derived from different datasets, which acts as a robustness check on our measure of gender bias. We use their index for essential services and enablers of economic opportunity. This state level variable is made up of 5 components: The percentage of women with unmet needs for family planning, the maternal mortality rate, and the male/female gaps in education level, financial inclusion, and digital inclusion. This index is produced as a percentage, where higher ESO means more progressive scores in the 5 components, and hence could indicate less gender bias in a state. In keeping with our other gender bias variables, we take the inverse of this percentage:

$$invESO = \left(\frac{100 - ESO}{100}\right),$$

here, higher $invESO$ may indicated more gender bias in a state.

Gender bias comparison Tables

Table 14: Gender Bias comparison

State	MW	Measure of Gender Bias invESO%	invDC%	%age of MW[25]
GOA	0.8318	9	53	-
Chhattisgarh	0.8914	20	73	-
Kerala	0.9144	6	53	0.24
Uttaranchal	0.9185	13	64	-
Tamilnadu	0.9511	13	51	0.29
Andhra Pradesh	0.9677	17	60	0.39
Rajasthan	0.9705	24	77	0.42
Karnataka	0.9711	14	65	0.45
Orissa	0.9742	21	58	0.57
West Bengal	0.9749	13	76	0.53
Himachal Pradesh	0.9776	12	61	0.58
Meghalaya	0.9866	23	23	-
Gujarat	0.9929	17	63	0.4
Maharashtra	0.9955	13	55	0.79
Uttar Pradesh	1.0064	22	66	0.65
Jammu & Kashmir	1.0212	17	75	-
Madhya Pradesh	1.0226	21	71	0.93
Jharkhand	1.0274	21	58	-
Manipur	1.0316	11	31	-
Haryana	1.0328	18	58	1.23
Delhi	1.0358	11	48	
Tripura	1.0374	14	70	-
Punjab	1.0380	17	63	0.86
Mizoram	1.0668	11	30	-
Sikkim	1.0920	14	41	-
Assam	1.1029	16	39	0.67
Nagaland	1.1115	14	27	-
Arunachal Pradesh	1.1293	21	47	-
Bihar	1.1318	23	67	0.77

Key, based on conclusions from Anderson & Ray (2012)	
Red	Make up 37% of all missing women in India
Blue	Have lowest level of missing women

Note: Ordered by MW, low to high. Data are for MW and invDC% are for the years 2004-05. Data for invESO% is from 2015. All variables are State level totals or averages. invDC% is the Inverse Decision making percentage: the percentage of women who report they are not allowed to participate in at least one of four household decisions. MW is Missing Women, the ratio of men to women outside of the fertile age range in a district, divided by the expected ratio for men to women outside the fertile age range, with the expected ratio based on India and state averages.invESO% is Inverse ESO, the inverse of the Mckinsey Institute index for the percentage of women in a state with access to essential services and opportunities. %age of MW comes from Anderson & Ray (2012) and represents the total number of missing women as a percentage of total observed women in a state, - indicates a sate not considered in that paper.

[25] from Anderson & Ray (2012)

Crime

Data on crime are at the district level, sourced from the National Crime Records Bureau (NCRB), India. We consider Rapes, Assaults on the Modesty of Women (Indecent Assaults) as our main crime against women variable, denoted in the equation below as CAW. A version of this variable, denoted here as CAWR, that is potentially more sensitive to any reporting bias is achieved by including Insult to the Modesty of Women (Harassment) under crimes against women. We denote Cruelty by Husband or his Relatives as Domestic Violence, (DV). Data on total crimes at the district level registered under the Indian Penal Code are also sourced from the NCRB. Crime data is recorded as year end crime statistics for crime committed during that year. Our interest is the effect of employment and wage gaps on crime hence crime data is concorded with a half year lead on the NSSO information. For example, crimes committed during January to December 2005 is allocated to $t = 2004$ in this analysis. The dependent variables in this investigation are now defined as follows:

$$lnCAW_{i,t} = ln\left(1 + \sum Rapes_{i,t} + \sum IndecentAssaults_{i,t}\right),$$

$$lnCAWR_{i,t} = ln\left(1 + \sum Rapes_{i,t} + \sum IndecentAssaults_{i,t} + \sum Harassment_{i,t}\right),$$

$$lnDV_{i,t} = ln\left(1 + \sum DomesticViolence_{i,t}\right),$$

where $\sum Crime_{i.t}$ is the total reported $Crime$ in district i at time t. The natural logarithm of 1 plus the variable is used to preserve any districts that have zero reported crimes. Any districts for which crime data are missing are excluded from this report.

Exposure of a District to Trade Tariffs

The exposure of a district to trade tariffs (ETT) describes the potential level of protection from competition workers in a district experience. This may influence wages and employment in a district. Tariffs are set for India as a whole, hence ETT is exogenous to the level of crime in a district, providing useful instruments in our investigation. Data on product level tariffs comes from the WITS world Bank database. Information on the inputs and output to each industry come from the input-output transactions table (IOTT 1994).

We define the average tariff for industry j in time period t as the simple average of the tariffs on all goods produced by industry j:

$$\overline{Tar}_{t,j} = \frac{1}{G_{t,j}} * \sum_{g=1}^{G_{t,j}} tar_g \quad g \in G_{t,j},$$

where, g is in the set of $G_{t,j}$ which is all goods produced in industry j at time t. tar_g is the tariff on good g at time t[26]. For non-trade industries we set $\overline{Tar}_{t,j} = 0$. Following this, input tariffs are defined as the

[26] Technically, $\overline{Tar}_{t,j}$ is the simple average tariff on all HS (Harmonized System) 4-digit products in industry j. We first concord HS

weighted average of the tariff on each good used as an input to industry j, adjusted for the fraction of total inputs to j represented by each good:

$$\overline{inpTar}_{t,j} = \frac{\sum_{y=1}^{G_{t,j}} tar_g * input\,to\,j_g}{\sum_{g=1}^{G_{t,j}} input\,to\,j_g} \quad g \in G_{t,j},$$

where g is a good in the set of $G_{t,j}$, which is all goods used as inputs for industry j at time t. $input\,to\,j_g$ is the total inputs of g to industry j at time t. If no inputs to j are recorded for a given g, $input\,to\,g = 0$. Note: goods produced by industry j are included in the set $G_{t,j}$ as some products of j may be also used as inputs in j. $\sum_{g=1}^{G_{t,j}} input\,to\,j_g$ is total inputs to j. tar_g is the tariff on good g at time t. Non-trade industries are not included in $G_{t,j}$.

We now adapt the specification in Topalova (2010) and define ETT for people of type k in district i at time t as the weighted average of the all industry tariffs at time t, weighted by the first period industry employment composition for type k in district i :

$$ETT_{i,t,k} = \frac{\sum_{j=1}^{J_i} \left(Emp_{i,04,k,j} * \overline{Tar}_{t,j}\right)}{\sum_{j=1}^{J_i} Emp_{i,04,k,j}} \quad j \in J_{i,04},$$

Similarly, exposure to input tariffs (EIT), is defined as the weighted average of all industry input tariffs:

$$EIT_{i,t,k} = \frac{\sum_{j=1}^{J_i} \left(Emp_{i,04,k,j} * \overline{inpTar}_{t,j}\right)}{\sum_{j=1}^{J_i} Emp_{i,04,k,j}} \quad j \in J_{i,04}.$$

In the above, $Emp_{i,04,k,j}$ is the total number of people in district i of type k employed in industry j in 2004, as defined in section 5. For non-traded industries, there are no tariffs or input tariffs, so our exposure measures are definitionally zero[27]. $\sum_{j=1}^{J_i} Emp_{i,04,k,j}$ is the total number of people of type k employed in i in 2004. If this total is zero, then there are no workers of type k to be exposed to tariff effects and definitionally our exposure measures are zero for this also. Hence we set $ETT_{i,t,k} = 0$ and $EIT_{i,t,k} = 0$ if $\sum_{j=1}^{J_i} Emp_{i,04,k,j} = 0$. We include employment in non traded industries when deriving total employment in i. This means that the impact of any tariffs are also scaled by the relative size of the tradeable sector is in a district. The resulting exposure variables provide the potential impact of tariffs or input tariffs at time t for each type of worker in district i.

4-digit products to input-output sectors from the IOTT 1994. Each IOTT sector is then mapped to a 3-digit NIC 1998 sector.

[27]This is not strictly true, particularly for input tariffs. For example, if you are in the restaurant business you may use imported tables and chairs and be exposed to input tariffs in that regard.