

Media reported violence and female labor supply

Zahra Siddique

Discussion Paper 20 / 732

19 August 2020



School of Economics
University of Bristol
Priory Road Complex
Bristol BS8 1TU
United Kingdom

Media reported violence and female labor supply*

Zahra Siddique [†]

August 19, 2020

Abstract

This paper explores how safety concerns together with cultural norms associated with female purity have an impact on behavior such as female labor supply in a developing country context. In particular, I examine the effect of media reports of local physical and sexual assaults on urban women's labor force participation in India. This is done by combining nationally representative cross-sectional microeconomic surveys on labor force participation carried out between 2009 and 2012 with a novel geographically referenced data source on media reports of assaults. I find that a one standard deviation increase in lagged media reports per 1000 people of local sexual assaults reduces the probability that a woman is employed outside her home by 0.67 percentage points (or 5.5% of the sample average). I find evidence that this is a short lived effect, with female labor supply increasing to catch-up following an initial decline. The negative effect of media reported violence on female labor supply persists after controlling for the underlying level of violence against women reported to the police in a district or after controlling for exogenous gender specific labor demand shocks. I find these effects to be strongest among young women between the ages of 18 and 25. These effects are robust to changes in the estimation sample and empirical specification, as well as to placebo checks.

JEL Classification: J16, J22

Keywords: Economics of Gender, Labor Supply

*I have benefited from comments by Arnab Basu, Erlend Berg, Nancy Chau, Melanie Khamis, Kanika Mahajan, Christine Valente, Ferdinand Vieider and Kate Vyborny as well as numerous conference and seminar participants. All errors are my own

[†]Department of Economics, University of Bristol, Priory Road Complex, Bristol BS8 1TU, UK. E-mail: zahra.siddique@bristol.ac.uk. Phone: +44 (0) 118 378 7864.

1 Introduction

Gender gaps or differences in labor market outcomes across men and women exist worldwide, but these differences are particularly stark in much of the developing world (Jayachandran 2015). For instance, labor markets in developing countries are likely to be characterised by large gender gaps in labor force participation. Low female labor force participation, in turn, is associated with reduced incentives to invest in female human capital, reduced female bargaining power within households, and worse child health outcomes (Heath & Jayachandran 2005). Female labor supply within developing countries might also be particularly sensitive to sudden incidents of violence against women which are publicised in the media since such reporting can make the extreme consequences of being a violence victim salient. This paper investigates and quantifies these relationships using data from urban India.

An important feature of the Indian labor market today is the presence of large gender gaps in labor force participation (Fletcher et al. 2018). Despite high economic growth, increases in education attainment and declining fertility over the last thirty years, female labor force participation rates have stagnated in urban households (Klasen & Pieters 2015) and fallen in rural households (Afridi et al. 2018). This is in contrast to developed countries which have seen large gains by women in labor force participation over time (Blundell & MaCurdy 1999). For instance, in 2011-2012 just 20% of working age Indian women (age 15-65) in urban households and 27% in rural households were in the labor force, in comparison with 81% of working age Indian men in urban households and 84% in rural households.¹ Labor force participation rates in 2012 for women in the US were 68%, in the UK were 71%, in Sweden were 78% and in Germany were 72%.² India's female labor force participation rates also compare unfavourably with other developing countries. For instance, using Demographic and Health Survey data from 63 developing countries between 1986 and 2006, Bhalotra & Umana-Aponte (2010) note that employment among women age 20-49 was 64% in Africa, 43% in Asia and 50% in Latin America.

This paper investigates labor force participation decisions of urban women age 18-50 in India to understand how these women might play a more active role within the labor market. In particular, it examines a potential determinant of women's labor force participation decisions which is currently under-explored in the literature on female labor supply. Recent increases in media reports of violence against women in India, particularly of sexual assaults and rapes, may have an unintended negative effect of

¹Based on own calculations from the Employment and Unemployment schedule of the 68th round of the Indian National Sample Survey. Labor force status is defined using activity status over the previous year; a person is in the labor force if they are self employed, an unpaid family worker, a regular salaried employee, a casual worker or unemployed.

²Data extracted from <http://stats.oecd.org>.

detering women from going out for work in a society where the stigma costs of sexual assaults are high by making the extreme consequences of being a violence victim salient. This paper quantifies the effect of such reporting on whether women seek employment outside their homes. It makes a contribution to the literature on the distortive effects of fear as well as contributing to the literature on female labor supply determinants in a developing country context, and a smaller literature examining the causes and consequences of violence against women.

[Becker & Rubinstein \(2011\)](#) provide a framework to illustrate how small probability events (such as becoming a victim of violence) can have a large affect on people’s behavior. They incorporate into expected utility theory “situations in which the extreme consequences associated with the consumption of risky goods (or engaging in outside work) and the extent these turn into a salient phenomenon affect persons’ mental state, generate fear, and by that affect peoples’ utility and well-being.” People can make investments to overcome fear provided the benefits from consumption of risky goods or activity are sufficiently large. Those for whom the benefits are not sufficiently large substitute out of the risky activity so that the affect on their behavior appears to overstate the objective probability of being harmed by terror. [Becker & Rubinstein \(2011\)](#) test their theoretical predictions by examining how fear of terrorist attacks in the US and Israel has an impact on usage of goods and services subject to these attacks.

Recent research indicates that fear and safety concerns of women outside the home are likely to play an important role in whether they seek outside employment in India. [Muralidharan & Prakash \(2017\)](#) find that providing girls in the Indian state of Bihar with a bicycle improved education enrollment by making it quicker and safer for girls to travel to school. [Borker \(2018\)](#) finds that women are willing to choose a college in the bottom half of the quality distribution over a college in the top half at the University of Delhi for a travel route they perceive to be one standard deviation safer. [Chakraborty et al. \(2018\)](#) use cross-sectional data from the 2005 wave of the India Human Development Survey (IHDS) to find that in urban neighborhoods where the self reported level of sexual harassment against women is high, women are far less likely to seek outside employment.³

Fear of public spaces following media coverage of sexual assaults can create stress and anxiety, deterring women from going out for work. According to a survey carried out in Delhi in 2012 following the rape and subsequent death of a Delhi woman on a moving bus which was widely reported in the media, nearly

³Note several distinctions of this paper from [Chakraborty et al. \(2018\)](#); first, the regressor is not self-reported sexual harassment but media reports of assaults. Hence, while [Chakraborty et al. \(2018\)](#) find a general association of sexual harassment with female labor supply, this paper finds a more specific effect of sexual assault media reports on female labor supply. Second, the measure of sexual assault media reports used in this paper is less prone to measurement error in comparison with self-reported sexual harassment. Third, by using a panel this paper exploits variation over time as well as space in the empirical analysis.

73% respondents said that women face sexual violence or harassment in their neighborhoods, and more than half stated that these spaces are unsafe at all times. Almost 20% of the respondents stated they were fearful when going out alone in the daytime and an additional 10% percent stated they would not venture out alone at all. These fractions were 63% and 21% when respondents expressed safety concerns for going out after it was dark (UN & ICRW 2013).

Much of the literature on female labor supply in developing countries has examined the role of individual characteristics (such as age, education, race/ethnicity) or family attributes (such as spouse variables, number of children) in labor force participation decisions. For instance Klasen & Pieters (2015) and Afridi et al. (2018) investigate the role of these variables in participation decisions of married women in urban and rural India. A consistent finding across studies using Indian labor market data is that women from Muslim and high-caste Hindu households have lower labor force participation rates compared to women from low-caste Hindu households. This is attributed to a higher value placed on women's purity (involving limited interaction with men outside the family) by these social groups compared to low-caste Hindu households (Field et al. 2010, Klasen & Pieters 2015, Jayachandran 2015).

A small but increasing literature in economics looks at the causes and consequences of violence against women. Whilst improvements in women's economic position relative to men have been shown to reduce violence against them within developed countries (Aizer 2010, Anderberg et al. 2016), such improvements *increase* violence against them in developing country settings due to a backlash effect (Krishnan et al. 2010, Guarnieri & Rainer 2018, Bhalotra et al. forthcoming). This paper looks at the flip side of this relationship or whether violence against women (specifically that publicised in the media) has an impact on women's labor force participation. Therefore it makes a contribution to research examining the *consequences* of violence against women. Apart from the direct costs of such violence on victims as well as the harmful affects of such violence on health outcomes of children born to victims (Aizer 2011, Currie et al. 2018, Rawlings & Siddique forthcoming), this paper makes the case for an additional negative consequence: reduced involvement of women in the labor market following media publicised sexual violence due to a fear channel. This is the first study to examine and quantify such a channel.

This paper uses data from the 2009-10 and 2011-12 rounds of the Indian National Sample Surveys (NSS), and combines these data with a novel geographically referenced data source on media reports of physical and sexual assaults that occur in each respondent's local area. Whilst this data source (the Global Database on Events, Language and Tone, or GDELT) has been used in existing research in political science and economics, its use in research on violence against women remains unexplored. By combining

these different data sets, I am able to quantify the effect of media reported violence per capita in one's local area or district on labor supply decisions of urban women. To eliminate potential bias from local area specific unobservables as well as state specific time (quarter-year) shocks, I include local area and state-time fixed effects in the estimations. I find that a one standard deviation increase in lagged media reports per 1000 people of sexual assaults in the local area of a woman reduces the probability that she is employed outside her home by 0.67 percentage points (or 5.5% of the sample average). I do not find statistically significant effects of media reported violence on the labor supply of working age urban men. I find evidence that the effect of lagged media reports of sexual assaults on female labor supply is a short lived effect, with female labor supply increasing to catch-up following an initial decline. The negative effect of media reported sexual violence on female labor supply persists even after controlling for the level of crimes against women reported to the police in a district or after controlling for exogenous gender specific labor demand shocks. I find these effects to be strongest among young women between the ages of 18 and 25. Finally, I find that these effects are robust to changes in the estimation sample and empirical specification, as well as to placebo checks.

The rest of the paper is organised as follows: section 2 describes the data sets used as well as features of the estimation sample; section 3 lays out the framework and estimation methods employed in the study; section 4 discusses the estimation results; section 5 provides robustness checks and section 6 concludes.

2 Data and estimation samples

2.1 Data on the labor market

Data on the Indian labor market is taken from the Employment and Unemployment schedules of the Indian National Sample Survey (NSS). In most of the empirical analysis carried out in this paper, I use the two most recent 'thick' rounds of the NSS⁴: round 66 which was fielded from July 2009 to June 2010, and round 68 which was fielded from July 2011 to June 2012. I also make use of two earlier NSS thick rounds when carrying out a placebo experiment reported in section 5: round 62 which was fielded from July 2005 to June 2006, and round 64 which was fielded from July 2007 to June 2008. The unit of time in the empirical analysis is a quarter (also referred to as a sub-round). These data include a wealth of individual and household variables, and are the most widely used source of information on the Indian labor market.

⁴The 'thick' NSS Rounds use a large sample size, as opposed to 'thin' NSS Rounds which make use of smaller samples.

The dependent variable of interest in the empirical analysis is whether or not a female respondent works outside her home. I construct a labor supply measure L based on questions asked on daily activity status over the past week, specifically questions on whether the respondent spent any time in regular salaried or casual employment. Either of regular or casual employment is likely to involve work outside the home for women, so L is an indicator variable taking the value one if the respondent spent a non zero fraction of the past seven days in either regular salaried or casual employment. The NSS also queries respondents about the industry in which they worked in the past week based on weekly activity status which is used in later empirical analysis. It also asks respondents their usual principal activity status over the past year with an accompanying question on location of the workplace.

Possible confounders which are controlled for in the empirical analysis include own age and household social group, where social group is defined by caste and religious affiliation. I include indicator variables for whether a respondent belongs to the historically disadvantaged scheduled caste (SC) or scheduled tribe (ST) groups; whether a respondent is non-SCST and belongs to the Muslim religion; whether a respondent is non-SCST and belongs to the Other religion⁵; and whether a respondent is non-SCST and belongs to the low-caste Hindu group referred to as Other Backward Caste (OBC). The omitted category (unless stated otherwise) is Other Hindu, which consists primarily of high-caste Hindus.⁶

The NSS also includes a household level measure of per capita consumption over the past thirty days which is used to examine heterogeneity in further empirical analysis.

2.2 Data on media reported violence

Data on media reports of different kinds of assaults is extracted from the Global Database of Events, Language and Tone or GDELT. This is a very large, open source database which collects information on political events in the area of verbal and physical mediation and conflict based on an automated textual analysis of newswires.⁷ GDELT includes over a quarter-billion event records in over 300 categories across the globe, from 1979 to the present. Events in the database are sourced from digitalised newspapers, news agencies and web based news aggregators such as GoogleNews. Data is extracted from these sources using an open source coding algorithm TABARI (Text Analysis by Augmented Replacement Instructions) that searches through news articles for actions carried out by one actor on another as detailed in CAMEO (Conflict and Mediation Coding System), a widely used coding system in political science. CAMEO

⁵Other religion includes Christian, Sikh, Jain, Buddhist, Zoroastrian or Other.

⁶These dummy variables are constructed from two questions, the first asking respondents the social group (SC, ST, OBC, Other) of the household and the second asking respondents the religion of the household.

⁷Data available from <https://www.gdelproject.org>.

dictionaries include lists of approximately 15,000 actions (or verb phrases) and 60,000 political actors (Leetaru & Schrodt 2013).⁸

The database includes information on the type of event which was reported in the media, the day and location of the event, as well as the number of articles in which the event was reported.⁹ Of the different event categories, I extracted data on physical assault events or **physical assaults, not specified below**, which are described as **attack physical well-being of individuals without the use of weaponry, not otherwise specified**. This event category consists primarily of beatings reported in the media. An example provided by the CAMEO codebook¹⁰ is a news article which includes the sentence “Israeli soldiers routinely *beat up* Palestinian detainees on the occupied West Bank with the knowledge of senior officers, a court martial was told today.”¹¹ To these incidents I also add events involving **torture** (example: Security forces in Guinea have *tortured* scores of Sierra Leonean and Liberian refugees, whom authorities blame for a border conflict, Human Rights Watch (HWR) said Thursday) and **kill by physical assault** (example: A Palestinian prisoner *died* as a result of *torture* while in Israeli police custody, according to a report by a pathologist sent to Israel by Physicians for Human Rights). **Sexual assault** events are a separate event category, and are described as **sexually abuse, assault sexual integrity of individuals**. This event category consists of rapes and other sexual assaults reported in the media (example: U.S. border patrol agents *sexually abused* illegal Mexican immigrants with impunity, a human rights organization charged on Saturday).

Data on physical and sexual assault events is matched to individual districts using the 2011 Census administrative boundaries.¹² Events and the number of articles in which they are reported (or media reports) are then aggregated at the district and quarter-year level of aggregation, converted to per capita measures using the 2011 Census district population, and merged with individual level NSS data.

2.3 Data on crimes against women

To examine how media reporting of sexual assaults compares with underlying violence against women I use data on annual crimes by district which are reported to the police and available in publications by the National Crime Records Bureau (NCRB) at the Ministry of Home Affairs in India. I construct an

⁸These dictionaries as well as CAMEO codebooks may be accessed from the Computational Event Data System Project web site (<http://eventdata.parusanalytics.com/data.html>).

⁹For location the latitude and longitude of the landmark-centroid are provided.

¹⁰<http://data.gdeltproject.org/documentation/CAMEO.Manual.1.1b3.pdf>, page 81.

¹¹In this and the following examples, underlined text is used to indicate actors while italic font indicates the action carried out by one actor on another.

¹²I only keep incidents that have been identified by the database at the level of a city or landmark outside the US, which is 91.86% of all events over the respective time frame.

aggregate measure of crimes against women *CAW* that combines the following Indian Penal Code (or IPC) crime categories: rapes, assaults on women with intent to outrage her modesty and insults to the modesty of women. Since crimes against women are only available at the annual frequency, I use this measure from the previous year to when a woman is interviewed (2009 to 2012), hence using the measure from the years 2008 to 2011 for each district. The annual measures are further divided by four (to make these comparable to the quarterly media reported violence measures), converted to per capita measures of crimes against women using the 2011 Census district population and merged with the individual level NSS data at the district and year level of aggregation.

2.4 Estimation sample and summary statistics

The estimation sample consists of women from urban households age 18-50. These women are likely to be active in the labor market, and to be aware of media reporting on assaults taking place in their local area.

The dependent variable of interest is employment outside the home. Figure A1 shows the time allocation of men and women age 18-50 in the past week. I consider a woman to be employed outside her home if she spends a non-zero fraction of time in the past seven days in either regular salaried or casual employment according to daily activity status. From Figure A1, most of the time in the previous week is spent by women outside of the labor force, and there is little change in this pattern over time. This provides a striking contrast with men who spend a far higher fraction of time in regular salaried or casual employment, as well as in self employment.

Figure A2 shows the fraction of women and men age 18-50 from urban households who indicate their workplace is not within or adjacent to their dwelling by type of work based on principal activity status over the past year. For women, this fraction is highest among regular salaried workers and among casual workers, indicating that the assumption of regular salaried and casually employed women being employed outside the home is reasonable. For men the fraction indicating their workplace is not within or adjacent to their dwelling is high among regular salaried and casual workers, as well as among the self employed and unpaid family workers at close to or greater than 70%.

Table 1 gives descriptive statistics of variables used in the empirical analysis. The average age of women, at 33 years, is very similar over time (or across rounds). Women from disadvantaged caste groups such as OBC and SCST form 29% and 22% of the estimation sample. Women from high-caste groups form 29% of the estimation sample. Women from Muslim households form 16% and from Other religion

form just 4.5% of the estimation sample. Figure A3 gives the labor force participation rates for women by social group. Labor force participation rates are highest among low-caste women (OBC and SCST) and among women who belong to the Other religion category. Labor force participation rates are lowest among Muslim women, followed by high-caste Hindu women.

Descriptive statistics for the merged data on media reports of assaults and on crimes against women are given in panels B and C of Table 1. Crimes against women reported to the police are far higher than sexual assaults that appear in the media. Physical assaults are reported in the media more frequently than sexual assaults, on average. While media reports of both physical and sexual assaults show a slightly increasing trend over this relatively short time period (Figure 1), there is far more variation in media reported assaults across districts as indicated by the large standard errors.

The distribution of media reports of physical and sexual assaults when aggregated to Indian states is given in Figure 2. As with the distribution of media reported assault incidents (Figure A4), media reports of physical and sexual assaults are higher in Maharashtra and Uttar Pradesh compared with other Indian states. A scatterplot of media reports of sexual assaults across the current and previous time period by district is given in Figure A5, sub-figure (a), revealing some outliers. A few districts (those in Delhi, Mumbai and Jalpaiguri) report very high media reports of sexual assaults in some quarters; as a robustness check I drop these districts from the estimation sample and find this does not change the main results (see Table 5). Overall, there is quite a lot of variation in media reports of sexual assaults from one period to the next (as shown by a fairly wide scatter away from the 45 degree line of equality in Figure A5, sub-figure (b)) which I exploit in my empirical analysis.

A comparison of annual reported crimes against women and media reports of sexual assaults by district is given in Figure A6. These scatterplots give combinations of annual crimes against women and media reports of sexual assaults by district for each year from 2009 to 2012. Almost all points on the scatterplots lie above the 45 degree line of equality, indicating that reported crimes against women almost always exceed media reports of sexual assaults or that (unsurprisingly) most crimes against women which are reported to the police are not reported in media outlets. Figure A6 also shows that the distribution of media reports of sexual assaults is right-skewed with a high fraction of zeroes.¹³

¹³Given the right skewed distribution of media reported assaults, the empirical analysis uses per capita measures rather than the number of media reports of assaults. As a robustness check I also use the inverse hyperbolic sine transformation of media reports of physical and sexual assaults and find this does not change the results (see Table A1).

3 Framework and estimation methods

3.1 Baseline regressions

I examine the relationship between media reported violence and labor supply by estimating variations of the following reduced form labor force participation equation which incorporates lagged media reports of assaults in one’s own local area as additional regressors:

$$L_{idst} = \beta_0 + \beta_1 X_{idst} + \beta_2 PA_{ds,t-1} + \beta_3 SA_{ds,t-1} + \gamma_d + \gamma_{s \times t} + u_{idst} \quad (1)$$

The dependent variable L_{idst} is an indicator variable taking the value one if woman i from an urban household who resides in district d and state s while interviewing in NSS sub-round t spends a positive fraction over the past seven days in either regular or casual employment based on daily activity status.

A parsimonious set of covariates are used as controls X_{idst} in equation (1) which are plausibly exogenous, or uncorrelated with the error term u_{idst} . These include a quadratic in own age, and a set of household social group (caste and religion) indicator variables.

The regressors of particular interest in equation (1) are $PA_{ds,t-1}$ and $SA_{ds,t-1}$. $PA_{ds,t-1}$ is the number of media reports per 1000 people of physical assaults taking place in woman i ’s own district and state in the three months preceding NSS sub-round t in which labor force participation is elicited. Similarly, $SA_{ds,t-1}$ is the number of media reports per 1000 people of local sexual assaults in the previous time period. The coefficients on these regressors (β_2 and β_3) capture the effect of local violence (in the form of local physical and sexual assault media reports per capita) on the probability that a woman works outside her home. By including lagged physical and sexual assault media reports separately, it is possible to examine the differential effect of these two kinds of media reported violence on female labor supply. The coefficient on $S_{ds,t-1}$ (or β_3) quantifies the effect of an additional local sexual assault media report per 1000 people in the previous time period on labor force participation today, while holding the number of local physical assault media reports per 1000 people in the previous period constant.

To rule out potential bias from unobserved heterogeneity which is local area (or district) specific, and state-time (quarter-year) specific, I also include district fixed effects (γ_d), and state times NSS sub-round fixed effects ($\gamma_{s \times t}$) in the estimations. These sources of unobserved heterogeneity can be important in this context. For instance, district fixed effects allow me to rule out potential bias due to region specific cultural factors which may be correlated with both female labor supply and media reported assaults.¹⁴

¹⁴See [Dyson & Moore \(1983\)](#) for a discussion of the variation in gender inequality across north and south India, with a

Another source of bias could be from state and time (quarter-year) specific macroeconomic shocks that influence labor supply and are likely correlated with the regressors; to rule these out state-time fixed effects are also included in all regressions. Finally, standard errors are adjusted for clustering at the district level.

It is possible that omitted variable bias has not been completely eliminated in these specifications; for instance, there might be important district specific time-varying gender norms that are correlated with labor supply and lagged media reports of assaults. However, results from placebo checks in section 5 indicate that this is unlikely to be the case.

3.2 Inclusion of additional lags of media reported assaults

To further examine how the effects of media reported violence change over time, I estimate the following regressions:

$$L_{idst} = \rho_0 + \rho_1 X_{idst} + \sum_{l=-4}^{-1} \rho_{2,l} PA_{ds,t+l} + \sum_{l=-4}^{-1} \rho_{3,l} SA_{ds,t+l} + \gamma_d + \gamma_{s \times t} + \zeta_{idst} \quad (2)$$

These regressions include up to four lags of PA and SA as additional regressors; in all other respects equation (2) is the same as equation (1).

3.3 Inclusion of crimes against women

To investigate how the effects of media reported violence compare with crimes against women reported to the police in one's local area, I estimate the following regressions:

$$L_{idst} = \pi_0 + \pi_1 X_{idst} + \pi_2 PA_{ds,t-1} + \pi_3 SA_{ds,t-1} + \pi_4 CAW_{dst} + \gamma_d + \gamma_{s \times t} + \varsigma_{idst} \quad (3)$$

These regressions also include CAW_{dst} ; as previously described (sub-section 2.3), this variable is constructed by aggregating the number of rapes, assaults and insults reported to the police in district d and state s per 1000 people over the past year (further divided by four to make this a quarterly measure). Estimation of these regressions allows a check on whether increased media reporting of assaults has an impact on female labor force participation (π_2 and π_3), while holding the past level of police reported crimes against women in the district constant. In all other respects equation (3) is the same as equation (1).

stronger patrilineal system in the north in combination with stronger cultural norms associated with female purity.

3.4 Inclusion of exogenous labor demand shocks

Rather than increased sexual assault media reports leading to reduced female labor force participation, an alternative could be related to unobserved changes in the demand for labor having an impact on both labor supply decisions and the incidence of sexual assaults. For instance, it may be that a negative labor demand shock leads to a decrease in female labor force participation, which further results in a decrease in women’s bargaining power within the household and in society. This, in turn, could be associated with an increase in sexual assaults leading us to observe a negative relationship between sexual assault media reports and female labor supply. To rule this out, I explicitly control for plausibly exogenous labor demand shocks by construction of gender specific industry weighted measures of employment, following the approach by [Bartik \(1991\)](#), [Autor & Duggan \(2003\)](#) and [Aizer \(2010\)](#). These measures use cross-state variation in gender specific industrial composition together with national changes in log employment by industry to construct predicted log employment for each state and quarter by gender. Specifically, these measures are constructed as

$$\ln(\widehat{Emp}_{st}^g) = \sum_{ind} \kappa_{s,ind}^g \ln(Emp_{-st,ind}) \quad (4)$$

where ind are the two digit NIC2004 industry codes, $g \in \{m, f\}$ indexes gender, s state and t NSS sub-round.¹⁵ $\kappa_{s,ind}^g$ is the proportion of male (female) workers in industry ind in state s using data from the first sub-round of NSS round 66. This proportion indicates whether the industry is a male (female) employment intensive industry; the proportion is fixed over time so that changes in log employment do not reflect selective sorting across industries. $\ln(Emp_{-st,ind})$ is the log of the fraction nationally employed (among individuals of working age) in industry ind and NSS sub-round t , excluding state s . Removing the focal state s from measurement of the national average allows removal of changes in employment that may be caused by changes in the underlying characteristics of workers within the state itself.

To explicitly examine whether controlling for exogenous labor demand shocks has an impact on the relationship between media reported violence and female labor supply, I estimate the following regressions:

¹⁵The current weekly activity status is used to find the industry in which an individual respondent to the NSS worked in the last week. Four digit industry codes identifying the industry in which a worker is employed changed from the NIC2004 classification in round 66 to the NIC2008 classification in round 68. I use a matching provided by http://mospi.nic.in/sites/default/files/main_menu/national_industrial_classification/nic_2008_17apr09.pdf to convert the four digit NIC2008 industry codes to four digit NIC2004 industry codes. A few four digit NIC2008 industry codes could not be matched to NIC2004 codes so these observations were dropped. All industry codes were then converted to two digit NIC2004 codes for subsequent empirical analysis.

$$L_{idst} = \delta_0 + \delta_1 X_{idst} + \delta_2 PA_{ds,t-1} + \delta_3 SA_{ds,t-1} + \delta_4 \ln(\widehat{Emp}_{st}^m) + \delta_5 \ln(\widehat{Emp}_{st}^f) + \gamma_d + \gamma_t + \varphi_{idst} \quad (5)$$

These regressions include the male and female specific measures of exogenous labor demand shocks given by (4). Since variation in these measures comes from across states and NSS sub-round, these regressions include NSS sub-round fixed effects γ_t rather than state times NSS sub-round fixed effects $\gamma_{s \times t}$ for identification; in all other respects equation (5) is the same as equation (1).

3.5 Heterogeneity

Given that purity concerns may have differing importance for different groups of women and that some groups might have different incentives than others to invest in overcoming fear, it may be that the effect of sexual assault media reports differs across groups. To investigate these kind of heterogeneities, I estimate equation (1) on different sub-samples of working age women. The advantage of this approach is that it allows for the effect of all regressors (and not just $SA_{ds,t-1}$) on female labor supply to be different across sub-samples. I investigate heterogeneity in the effect of $SA_{ds,t-1}$ on female labor supply across several dimensions: the first is across urban women who belong to different age groups, the second across urban women from households with differing levels of income (based on quartiles of the NSS round specific household per capita consumption distribution), the third across urban women from different social groups (defined by caste and religious affiliation) and the fourth across working age women from rural vs urban households.

4 Estimation results and discussion

4.1 Baseline regressions

The results from estimating equation (1) are reported in Table 2: columns (I)-(III) use a sample of women age 18-50 from urban households while columns (IV)-(V) use the corresponding sample of urban men. The results reported in column (I) include controls for a quadratic in age and a set of indicators for social group. With only this parsimonious set of controls included, $SA_{ds,t-1}$ has a positive statistically insignificant effect on labor force participation and there is no effect of $PA_{ds,t-1}$. Inclusion of district fixed effects in column (II) changes the results dramatically; there is still no effect of $PA_{ds,t-1}$ on labor force

participation but $SA_{ds,t-1}$ now has a negative statistically significant effect. This shows that district-specific unobservables are likely to be correlated with labor force participation L and with $SA_{ds,t-1}$, so that controlling for district fixed effects and exploiting within-district variation reveals a negative effect of $SA_{ds,t-1}$ on female labor supply. Further addition of (state \times NSS sub-round) fixed effects in column (III) results in a smaller change in the coefficients of interest. In the preferred specification reported in column (III), $PA_{ds,t-1}$ does not have a statistically significant effect on female labor force participation while $SA_{ds,t-1}$ has a statistically significant negative effect. A one standard deviation increase in $SA_{ds,t-1}$ ($= 0.03$) reduces the probability that a woman works outside her home by 0.67 percentage points using the sample of urban women age 18-50; this effect is 5.5% of the sample average of labor supply. In terms of effect size, the coefficient estimate on $SA_{ds,t-1}$ is likely to be an underestimate since it does not include the effect of media reports of sexual assaults taking place in neighboring districts which might also have a negative impact on female labor force participation decisions.

The effect of age and social group affiliation on female labor supply is consistent with existing research. Estimates reported in columns (I)-(III) of Table 2 indicate a quadratic effect of age on female labor supply, with labor force participation first increasing (with a positive coefficient on Age) and then decreasing (with a negative coefficient on Age^2) as age increases. Labor force participation is lowest among Muslim women, who are 4.6 percentage points less likely to be working outside the home compared to high-caste Hindu women (the omitted category). The lowest caste group of SCST has higher labor force participation in comparison with high-caste Hindu women. Women belonging to the SCST social group are 6.9 percentage points more likely to be working outside the home compared to high-caste Hindu women.

Since the distribution of media reported violence is right skewed with a high fraction of zeroes, I also estimate equation (1) using the inverse hyperbolic sine transformation of media reports per 1000 people of local assaults. The results are reported in Table A1; the sign and significance of the coefficients remains unchanged compared with results reported in columns (I)-(III) of Table 2.

The results in columns (I)-(III) of Table 2 also provide an interesting contrast with those from estimating equation (1) on a sample of urban men who are also 18-50 years old, as reported in columns (IV)-(V) of Table 2. The coefficient on $PA_{ds,t-1}$ for urban men is positive but statistically insignificant while the coefficient on $SA_{ds,t-1}$ is quite close to zero (column (IV) of Table 2). I also estimate equation (1) for an additional labor supply measure L^{ALT} on the sample of urban men; estimation results are reported in column (V) of Table 2. L^{ALT} is an indicator variable taking the value one if men are in regular employment, casual employment, self-employment, or unpaid family work based on daily activity

status. For this alternative measure of employment, the coefficient on $PA_{ds,t-1}$ is now close to zero; the coefficient on $SA_{ds,t-1}$ is also fairly close to zero. Given that a high fraction of men who are self employed or unpaid family workers also work outside the home (Figure A2), the measure L^{ALT} is more likely to capture employment outside the home for men than L . While there is no consistent and strong relationship between $SA_{ds,t-1}$ and male labor supply, a somewhat positive effect of $PA_{ds,t-1}$ on whether or not men are working in regular or casual employment suggests that higher media reporting of local physical violence might be associated with urban men switching from self employment and unpaid family work to regular and casual employment.

4.2 Inclusion of additional lags of media reported assaults

Results after inclusion of additional lags of media reports of assaults (as given by equation (2)) are reported in column (I) of Table 3, and reveal an interesting pattern. Lags of local physical assault media reports per 1000 people have no impact on female labor supply. While the coefficient on $SA_{ds,t-1}$ continues to be negative and statistically significant, the coefficient on $SA_{ds,t-2}$ is large and positive but misses statistical significance at the 5% level. The coefficient on $SA_{ds,t-3}$ is also positive but smaller than the coefficient on $SA_{ds,t-2}$ while that on $SA_{ds,t-4}$ is close to zero and statistically insignificant. This provides suggestive evidence that women might immediately decrease labor force participation following increased media reporting of sexual violence but try to catch up later; in other words these effects might not persist in the longer term. I also estimate equation (2) on sub-samples of working age women after splitting the sample by the quartile of the per capita consumption distribution of the household that a woman belongs to (see Appendix Table A2). These results provide suggestive evidence that the pattern of effects is coming from women who belong to the highest income households, for whom the coefficient on $SA_{ds,t-1}$ is large and negative while the coefficient on $SA_{ds,t-2}$ is substantively large, positive and statistically significant. This provides additional support for a catch-up, whereby women from the highest income households realize that they could have overreacted to the media reported violence.

4.3 Inclusion of crimes against women

Results after inclusion of crimes against women in the baseline regressions (as given by equation (3)) are reported in column (II) of Table 3. Coefficients on $SA_{ds,t-1}$ remain negative and statistically significant. This provides suggestive evidence that the negative effect of $SA_{ds,t-1}$ on female labor supply is a behavioral effect that persists even when comparing women across districts with the same level of police reported

crimes against women (such as rapes, assaults and insults) or underlying violence; in other words, the effect I find seems likely due to fear and not changes in the underlying probability of being assaulted. However, it is important to note that if the labor supply response is driven by actual crimes and not reporting but crimes are measured with error (due to possible under-reporting), then the coefficient on reporting would likely still be positive.

The effect of police reported crimes against women per 1000 people on female labor supply is also substantively large and negative, but not statistically significant due to the large standard error on these coefficients. This indicates that the cumulative effects of sexual assault media reports and (somewhat imperfectly measured) underlying violence against women on female labor supply might well be quite large. This is also consistent with the large effect sizes reported in [Chakraborty et al. \(2018\)](#).

4.4 Inclusion of exogenous labor demand shocks

Results from estimation of equation (4) are reported in columns (III)-(IV) of Table 3; column (III) uses contemporaneous measures of exogenous labor demand while column (IV) uses one quarter lags of these measures. $SA_{ds,t-1}$ continues to have a negative and statistically significant effect on female labor supply in both cases, while there is no effect of $PA_{ds,t-1}$ on female labor supply. Contemporaneous exogenous increases in employment within female dominated industries increase female labor supply while such increases within male dominated industries decrease it, but these effects are not statistically significant. One period lags in these measures also have the same signed effects but (unsurprisingly) the size of the coefficients is now very close to zero. This set of estimation results allows me to rule out an alternative story in which we observe a spurious relationship between sexual assault media reports and female labor supply due to changes in labor demand.

4.5 Heterogeneity

To examine the mechanisms behind the relationship between lagged sexual assault media reports and female labor supply, I estimate and report results for equation (1) across different sub-samples of women. This allows me to explore whether and how this effect is stronger among some groups of women compared to others.

Purity concerns are likely to be more important for younger than older women. Sexual assault victims also tend to be younger women; in 2017 approximately 30% of police registered rape victims in India were women younger than 18 and approximately 50% were between 18 and 30 years old (based on data made

available by the National Crime Records Bureau, accessed via <https://data.gov.in/>). To examine whether the effect of lagged sexual assault media reports is larger among younger women, I estimate equation (1) across different sub-samples of working age urban women, where I split the sample based on the woman’s age (between 18 to 25, between 26 to 33, between 34 to 41 or between 42 to 50 years old). The results are reported in Panel A of Table 4 and show the effect to be strongest among the youngest women (age 18 to 25) (column (II), Panel A of Table 4). In a completely interacted regression using the complete sample I find that the effect for young women (age 18 to 25) is statistically significantly different in comparison to older women (who are either 26 to 33 or 42 to 50 years old).¹⁶

Media reports of local sexual assaults are likely to generate feelings of anxiety and fear among women. It is possible that fear leads women to magnify the subjective probability that they might become a victim, despite little to no change in the objective probability of this happening. Such fear could explain why women become less likely to work outside their homes in response to higher media reports of local sexual assaults in the previous time period. It is also possible that some groups of women have the economic incentives to overcome this fear and work outside their homes; one such group may be women from poor households. At the same time women from the highest income households are least likely to have the economic incentives to overcome this fear. I examine this kind of heterogeneity by estimating equation (1) across different sub-samples of working age urban women, where I split the sample based on the per capita consumption of the household that women belong to. I separately examine women who belong to households with per capita consumption less than the 25th percentile of the NSS round specific household per capita consumption distribution, women who belong to households with per capita consumption between the 25th percentile and median of the distribution, women who belong to households with per capita consumption between the median and 75th percentile of the distribution and women who belong to households with per capita consumption higher than the 75th percentile of the distribution (see Figure A7 for the distribution of household consumption per capita among urban households across NSS rounds). The results are reported in Panel B of Table 4. Consistent with a fear channel, I find suggestive evidence that the negative effect of media reports of local sexual assaults on female labor supply is lowest among women from the poorest households who have the economic incentives to invest in overcoming their fear (column (II), Panel B of Table 4), and strongest among women from the highest income households for whom these incentives are likely to be absent (column (V), Panel B of Table 4). However, in a completely interacted regression using the complete sample I find that the effect for women from the wealthiest

¹⁶Results using a completely interacted regression specification on the complete sample are available on request.

households is not statistically significantly different in comparison to the other groups of women (where the groups are based on household per capita consumption).

Two groups of women who have the lowest labor force participation rates are Muslim and high-caste Hindu women. I examine whether these two groups of women also have a relationship between media reported violence and labor supply that differs from other women. I estimate equation (1) separately for Muslim and high-caste Hindu women (as well as women from low-caste Hindu groups, OBC and SCST). The results are reported in Panel C of Table 4. Column (I) in Panel C reports the estimation results for Muslim women and column (II) in Panel C for high-caste Hindu women; I find suggestive evidence that the negative effect of media reports of local sexual assaults on female labor supply are strongest among high-caste Hindu women. In contrast, this negative effect is weaker among Muslim women. It is also negative but statistically insignificant among SCST women. Among OBC women it is surprisingly positive (but statistically insignificant) while the effect of media reports of local physical assaults is large and negative. If the sample of OBC women is further restricted to younger women (age 18-25) the effects become very similar to those among other low-caste SCST women. The pattern of results for older OBC women from urban households might arise if these women are less affected by purity concerns but deterred from going out to work by increased media reporting of local physical violence. However, in a completely interacted regression using the complete sample I find that the effect for high-caste Hindu women is not statistically significantly different in comparison to the other social groups (apart from the OBC group).

I also estimate equation (1) separately on the estimation sample of rural women to investigate heterogeneity across working age rural and urban women; the results are reported in column (II) of Table A3. The coefficient on $SA_{ds,t-1}$ is still negative but much smaller in size and no longer statistically significant while the coefficient on $PA_{ds,t-1}$ becomes surprisingly large, positive and statistically significant. However, if the estimation sample is further restricted to younger women in rural households (as given in columns (III) and (IV) of Table A3), effects of media reported violence become closer to those documented for urban women. The pattern of results for older women from rural households might arise if these women substitute for male members who might work less when media reports of local physical violence increase or increased physical violence might lead to lost income which in turn compels older women to work more.

5 Robustness checks

In this section I report robustness checks to examine whether the negative relationship between lagged sexual assault media reports per 1000 people and female labor supply persists after alterations of the estimation sample, empirical specification and in a placebo experiment. I find the negative relationship to be robust to alterations across these dimensions, and to these additional checks.

Figure A5 indicated a very large number of sexual assault reports over this time period come from a few districts, including those in the city of Delhi. As a robustness check, I drop these districts from the estimation sample and re-estimate equation (1). The results are reported in column (I) of Table 5. The negative relationship between $SA_{ds,t-1}$ and female labor supply persists in the smaller sample.

As another robustness check, I drop women who are currently studying in an educational institution from the estimation sample and re-estimate equation (1). The results are reported in column (II) of Table 5. The relationship between $SA_{ds,t-1}$ and labor supply is very similar to before.

As a further check I estimate the following equation:

$$L_{idst} = \theta_0 + \theta_1 X_{idst} + \sum_{l=-1}^{+1} \theta_{2,l} PA_{ds,t+l} + \sum_{l=-1}^{+1} \theta_{3,l} SA_{ds,t+l} + \gamma_d + \gamma_{s \times r} + \omega_{idst} \quad (6)$$

This empirical specification adds contemporaneous media reports per 1000 people of local physical and sexual assaults as well as one period leads of these measures to the set of regressors in equation (1). Inclusion of one period leads provides a useful placebo test. The estimation results for equation (6) are reported in column (III) of Table 5. Reassuringly, the coefficients on the one period leads or $PA_{ds,t+1}$ and $SA_{ds,t+1}$ are statistically insignificant. The coefficients on contemporaneous media reports of local physical and sexual assaults or $PA_{ds,t}$ and $SA_{ds,t}$ are also statistically insignificant, indicating that the effect of media reported violence on labor supply takes place with a one period (or quarter) lag. $SA_{ds,t-1}$ continues to have a negative relationship with female labor supply.

I also estimate a further specification in which additional lags of media reported assaults are included together with contemporaneous and one period leads:

$$L_{idst} = \lambda_0 + \lambda_1 X_{idst} + \sum_{l=-4}^{+1} \lambda_{2,l} PA_{ds,t+l} + \sum_{l=-4}^{+1} \lambda_{3,l} SA_{ds,t+l} + \gamma_d + \gamma_{s \times r} + \vartheta_{idst} \quad (7)$$

Estimation results are reported in column (IV) of Table 5. As before, the coefficients on one period leads and contemporaneous media reports per 1000 people of physical and sexual assaults are statistically

insignificant. $SA_{ds,t-1}$ continues to have a negative relationship with female labor supply. These results can also be seen visually in Figure 3, which plots the coefficients on leads, lags and contemporaneous media reports per 1000 people of local sexual assaults from estimation of equation (7). This Figure shows that the coefficient on the one period lag is negative and statistically significant, while the two period lag is positive but misses statistical significance.

As a final check, I carry out a placebo experiment in which I estimate equation (1) using the estimation sample of working age urban women (age 18-50) from two earlier NSS rounds (rounds 62 and 64) which surveyed respondents between 2005 and 2008. However, I assign PA and SA (or placebo treatments) to women in this estimation sample from rounds 66 and 68 so that these are actually the 16 period or 4 year leads of these variables. Estimation results are reported in column (V) of Table 5 and show a statistically insignificant effect of the placebo treatments on female labor supply.

6 Conclusion

I find that the labor force participation of urban women in India is reduced following increases in lagged media reports per 1000 people of sexual assaults in one's local area. These effects come primarily from young women, are short-lived, persist despite ruling out several sources of unobserved heterogeneity, and are robust to a number of checks. Promising avenues for future research include exploring whether similar effects exist in other developing country settings and contexts. Also useful to explore further would be the use of high frequency data on female employment to further understand the relationship between media reported violence and female labor supply.

The results reported in this paper highlight the importance of addressing safety concerns of women in India, particularly younger women. One set of longer term interventions involve strengthening of a policing and legal framework that protects women from sexual assaults. Another shorter term intervention could involve provision of special transport facilities for women. [Aguilar et al. \(forthcoming\)](#) examine a program that reserves subway cars for women in Mexico City and find that this reduces sexual harassment towards women while [Kondylis et al. \(2018\)](#) find that riding in women reserved safe spaces reduces harassment against women in Rio de Janeiro. However, both studies also find negative effects; in Mexico City there is an increase in non-sexual aggression incidents among men while in Rio de Janeiro there is stigmatization of women who ride female only cars. This indicates that policy prescriptions to address women's safety concerns need some care and thought to be implemented most effectively.

Tables

TABLE 1
Descriptive statistics

	Round 66	Round 68	Total
Panel A: Demographic and labor market characteristics			
Employed in salaried or casual work last week	0.1192	0.1217	0.1204
Age	32.6008	32.7259	32.6632
Hindu Other	0.2966	0.2806	0.2886
SCST	0.2184	0.2254	0.2219
Hindu OBC	0.2853	0.2880	0.2867
Muslim	0.1534	0.1616	0.1575
Other Religion	0.0454	0.0444	0.0449
N	46216	45907	92123
	Round 66	Round 68	Total
Panel B: Media reported violence			
Physical assaults per 1000 people at $(t - 1)$	0.0009 (0.0030)	0.0010 (0.0110)	0.0010 (0.0080)
Physical assault reports per 1000 people at $(t - 1)$	0.0061 (0.0221)	0.0068 (0.0894)	0.0065 (0.0650)
Sexual assaults per 1000 people at $(t - 1)$	0.0006 (0.0031)	0.0006 (0.0046)	0.0006 (0.0039)
Sexual assault reports per 1000 people at $(t - 1)$	0.0042 (0.0226)	0.0039 (0.0333)	0.0041 (0.0284)
N	46216	45907	92123
	Round 66	Round 68	Total
Panel C: Crimes against women			
Crimes against women per 1000 people	0.0178 (0.0154)	0.0184 (0.0176)	0.0181 (0.0166)
N	45467	44742	90209

Notes: Each cell gives the average value of a variable in the sub-sample indicated in the column head. Standard deviations are given in parentheses.

Source: Data on demographic and labor market characteristics (Panel A) is from rounds 66 (2009-10) and 68 (2011-12) of the Employment and Unemployment schedules, Indian National Sample Survey (NSS). Estimation sample is restricted to women from urban households who are 18-50 years of age. Data on media reported incidents and reports of assaults (Panel B) is extracted from the Global Database of Events, Language, and Tone (GDELT) which is then merged with individual level NSS data at the district and year-quarter level of aggregation. Data on crimes against women (Panel C) is extracted from publications by the National Crime Records Bureau (NCRB) at the Ministry of Home Affairs in India which is then merged with individual level NSS data at the district and year level of aggregation.

TABLE 2
Media reported violence and labor supply

<i>Sample:</i>	Women			Men	
<i>Dependent variable:</i>	<i>L</i>	<i>L</i>	<i>L</i>	<i>L</i>	<i>L^{ALT}</i>
	(I)	(II)	(III)	(IV)	(V)
Physical assault reports per 1000 people at $(t - 1)$	-0.0541 (0.0585)	-0.0269 (0.0650)	-0.0568 (0.0849)	0.1932 (0.1091)	0.0046 (0.0684)
Sexual assault reports per 1000 people at $(t - 1)$	0.2763 (0.2022)	-0.1969*** (0.0697)	-0.2313** (0.0910)	-0.0069 (0.1325)	0.0767 (0.0802)
SCST	0.0599*** (0.0064)	0.0688*** (0.0051)	0.0689*** (0.0052)	0.1535*** (0.0079)	0.0458*** (0.0043)
Hindu OBC	0.0144** (0.0060)	0.0064 (0.0041)	0.0072 (0.0041)	0.0345*** (0.0060)	0.0387*** (0.0039)
Muslim	-0.0435*** (0.0051)	-0.0465*** (0.0059)	-0.0458*** (0.0059)	-0.0182* (0.0082)	0.0813*** (0.0051)
Other Religion	0.0594*** (0.0116)	0.0380*** (0.0090)	0.0363*** (0.0089)	-0.0460*** (0.0119)	-0.0052 (0.0074)
Age	0.0180*** (0.0012)	0.0182*** (0.0011)	0.0182*** (0.0011)	0.0598*** (0.0020)	0.1189*** (0.0013)
Age ²	-0.0002*** (0.0000)	-0.0002*** (0.0000)	-0.0002*** (0.0000)	-0.0008*** (0.0000)	-0.0015*** (0.0000)
Controls	Yes	Yes	Yes	Yes	Yes
District FE	No	Yes	Yes	Yes	Yes
State \times NSS sub-round FE	No	No	Yes	Yes	Yes
N	92123	92123	92123	94692	94692

Notes: Each column reports results from a separate regression. Results reported in columns (I)-(III) are from estimating equation (1) for the sample of urban women while results reported in columns (V)-(VI) are from estimating equation (1) for the sample of urban men. The dependent variable in columns (I)-(IV) is L , which takes the value one if a person spends non-zero time in regular or casual employment in the last week. The dependent variable in column (V) is L^{ALT} , which takes the value one if a person spends non-zero time in regular, casual, self-employment or unpaid family work in the last week. Standard errors are clustered at the district level and reported in parentheses; * p-value < 0.05, ** p-value < 0.025, *** p-value < 0.01.

Source: Data from rounds 66 (2009-10) and 68 (2011-12) of the Employment and Unemployment schedules, Indian National Sample Survey (NSS). The estimation sample in columns (I)-(III) is restricted to women from urban households who are 18-50 years of age while the estimation sample in columns (IV)-(V) is restricted to men from urban households who are 18-50 years of age. Data on reports of physical and sexual assaults is extracted from the Global Database of Events, Language, and Tone (GDELTA).

TABLE 3
Additional controls

	(I)	(II)	(III)	(IV)
Physical assault reports per 1000 people at $(t - 1)$	-0.1368 (0.1315)	-0.0779 (0.0860)	-0.0063 (0.0622)	-0.0170 (0.0705)
Sexual assault reports per 1000 people at $(t - 1)$	-0.2653** (0.1136)	-0.2300** (0.0924)	-0.2153*** (0.0692)	-0.1658*** (0.0639)
Physical assault reports per 1000 people at $(t - 2)$	-0.0507 (0.1264)			
Sexual assault reports per 1000 people at $(t - 2)$	0.2097 (0.1243)			
Physical assault reports per 1000 people at $(t - 3)$	0.0400 (0.0585)			
Sexual assault reports per 1000 people at $(t - 3)$	0.1738 (0.1732)			
Physical assault reports per 1000 people at $(t - 4)$	0.0269 (0.0239)			
Sexual assault reports per 1000 people at $(t - 4)$	0.0154 (0.1076)			
Crimes against women per 1000 people		-0.3664 (0.2921)		
Industry weighted log of female employment at (t)			0.1062 (0.0774)	
Industry weighted log of male employment at (t)			-0.2225 (0.1203)	
Industry weighted log of female employment at $(t - 1)$				0.0665 (0.0949)
Industry weighted log of male employment at $(t - 1)$				-0.1672 (0.1527)
Controls	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
NSS sub-round FE	No	No	Yes	Yes
State \times NSS sub-round FE	Yes	Yes	No	No
N	92123	90209	92123	69038

Notes: Each column reports results from a separate regression. Results reported in column (I) are from estimating equation (2), in column (II) from estimating equation (3), in column (III) from estimating equation (5) and in column (IV) from estimating equation (5) using lagged measures of gender specific and industry weighted log employment. All regressions are estimated using the sample of urban women. The dependent variable is L , which takes the value one if a person spends non-zero time in regular or casual employment in the last week. Standard errors are clustered at the district level and reported in parentheses; * p-value < 0.05, ** p-value < 0.025, *** p-value < 0.01.

Source: Data from rounds 66 (2009-10) and 68 (2011-12) of the Employment and Unemployment schedules, Indian National Sample Survey (NSS). Estimation sample is restricted to women from urban households who are 18-50 years of age. It is further restricted to women for whom data on crimes against women could be matched at the district and NSS round level of aggregation in column (II). Data on reports of physical and sexual assaults is extracted from the Global Database of Events, Language, and Tone (GDELT). Data on annual crimes against women by district is from publications by the National Crime Records Bureau (NCRB) at the Ministry of Home Affairs in India.

TABLE 4
Heterogeneity

Panel A: heterogeneity by age					
<i>Sample:</i>	Age 18 – 50 (I)	Age 18 – 25 (II)	Age 26 – 33 (III)	Age 34 – 41 (IV)	Age 42 – 50 (V)
Physical assault reports per 1000 people at $(t - 1)$	-0.0568 (0.0849)	0.0861 (0.1291)	-0.1761 (0.1328)	-0.0705 (0.1224)	-0.1366 (0.1866)
Sexual assault reports per 1000 people at $(t - 1)$	-0.2313** (0.0910)	-0.5545*** (0.1561)	-0.1533 (0.1547)	-0.0631 (0.1633)	-0.0954 (0.2048)
Controls	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes
State \times NSS sub-round FE	Yes	Yes	Yes	Yes	Yes
N	92123	26008	23847	22470	19798
Panel B: heterogeneity by HH consumption quartile					
<i>Sample:</i>	all (I)	HH cons < 25 perc. (II)	HH cons 25 – 50 perc. (III)	HH cons 50 – 75 perc. (IV)	HH cons > 75 perc. (V)
Physical assault reports per 1000 people at $(t - 1)$	-0.0568 (0.0849)	0.0468 (0.4115)	0.0016 (0.3387)	-0.1871 (0.1223)	-0.0258 (0.0850)
Sexual assault reports per 1000 people at $(t - 1)$	-0.2313** (0.0910)	-0.0718 (0.4495)	-0.2333 (0.1997)	-0.1277 (0.1601)	-0.3447*** (0.0827)
Controls	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes
State \times NSS sub-round FE	Yes	Yes	Yes	Yes	Yes
N	92123	15380	22422	25317	29004
Panel C: heterogeneity by social group					
<i>Sample:</i>	Muslim (I)	high-caste Hindu (II)	OBC Age 18 – 50 (III)	OBC Age 18 – 25 (IV)	SCST (V)
Physical assault reports per 1000 people at $(t - 1)$	-0.1670 (0.2615)	0.3043 (0.1996)	-0.4418*** (0.1242)	-0.3411 (0.3670)	-0.3347 (0.2491)
Sexual assault reports per 1000 people at $(t - 1)$	-0.0898 (0.1786)	-0.4670*** (0.1203)	0.3171 (0.1872)	-0.3248 (0.2640)	-0.3321 (0.2806)
Controls	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes
State \times NSS sub-round FE	Yes	Yes	Yes	Yes	Yes
N	14506	26587	26408	7150	20443

Notes: Each column in a Panel reports results from a separate regression. Panel A splits the sample into four groups (columns (II)-(V)) based on a woman's age and reports results from estimating equation (1) for each group. Panel B splits the sample into four groups (columns (II)-(V)) based on the per capita consumption level of the household that a woman belongs to and reports results from estimating equation (1) for each group. Panel C splits the sample into four social groups (Muslim, high-caste Hindu, OBC and SCST) and reports results from estimating equation (1) for each group; the OBC group is further split by age (with results reported in columns (III)-(IV) of Panel B). The dependent variable is L , which takes the value one if a person spends non-zero time in regular or casual employment in the last week. Standard errors are clustered at the district level and reported in parentheses; * p-value < 0.05, ** p-value < 0.025, *** p-value < 0.01.

Source: Data from rounds 66 (2009-10) and 68 (2011-12) of the Employment and Unemployment schedules, Indian National Sample Survey (NSS). Estimation sample is restricted to women from urban households who are 18-50 years of age, unless specified otherwise. Data on reports of physical and sexual assaults is extracted from the Global Database of Events, Language, and Tone (GDELT).

TABLE 5
Robustness checks

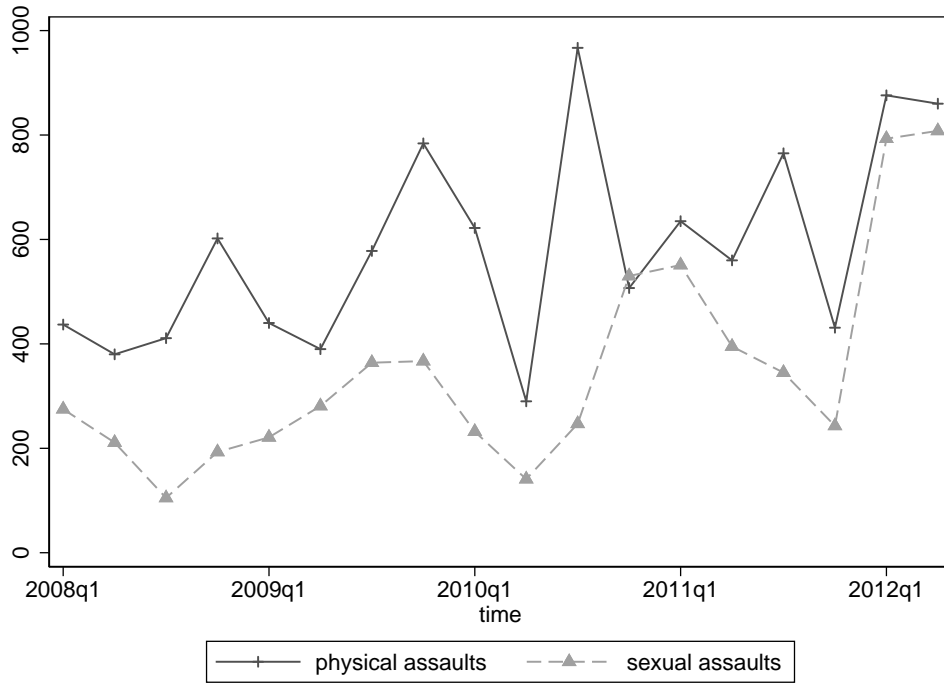
	excluding outliers (I)	excluding students (II)	alternative specifications (III)	alternative specifications (IV)	placebo exp (V)
Physical assault reports per 1000 people at $(t + 1)$			0.0086 (0.0297)	0.0326 (0.0385)	
Sexual assault reports per 1000 people at $(t + 1)$			-0.0239 (0.1274)	0.0458 (0.1316)	
Physical assault reports per 1000 people at (t)			0.0162 (0.1247)	-0.0514 (0.1434)	
Sexual assault reports per 1000 people at (t)			-0.1019 (0.0709)	-0.0856 (0.0875)	
Physical assault reports per 1000 people at $(t - 1)$	-0.0356 (0.0901)	-0.0734 (0.0855)	-0.0724 (0.0820)	-0.1614 (0.1223)	0.0037 (0.0679)
Sexual assault reports per 1000 people at $(t - 1)$	-0.2168* (0.0976)	-0.2407** (0.0962)	-0.2196* (0.1028)	-0.2587* (0.1163)	0.0155 (0.0903)
Physical assault reports per 1000 people at $(t - 2)$				-0.0805 (0.1087)	
Sexual assault reports per 1000 people at $(t - 2)$				0.2391 (0.1370)	
Physical assault reports per 1000 people at $(t - 3)$				0.0221 (0.0604)	
Sexual assault reports per 1000 people at $(t - 3)$				0.1846 (0.1971)	
Physical assault reports per 1000 people at $(t - 4)$				0.0315 (0.0234)	
Sexual assault reports per 1000 people at $(t - 4)$				-0.0510 (0.1398)	
Controls	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes
State \times NSS sub-round FE	Yes	Yes	Yes	Yes	Yes
N	90374	83731	92123	92123	100980

Notes: Each column reports results from a separate regression. Results reported in columns (I) and (II) are from estimating equation (1) on different samples of urban women. Results reported in column (III) are from estimating equation (6) and results reported in column (IV) from estimating equation (7). Results reported in column (V) are from a placebo experiment described in Section 5. The dependent variable is L , which takes the value one if a person spends non-zero time in regular or casual employment in the last week. Standard errors are clustered at the district level and reported in parentheses; * p-value < 0.05, ** p-value < 0.025, *** p-value < 0.01.

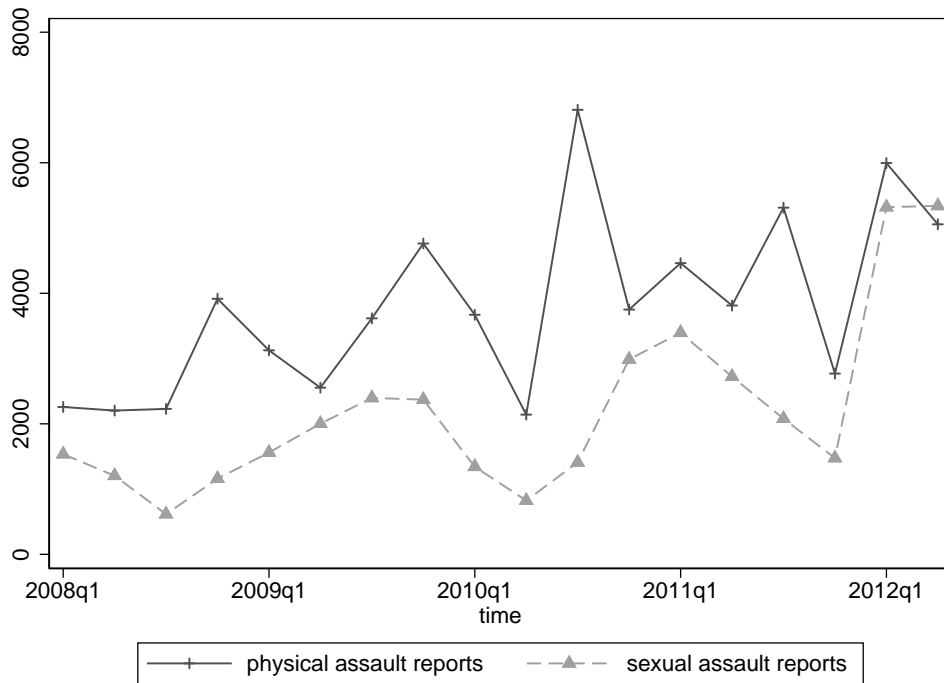
Source: Data from rounds 66 (2009-10) and 68 (2011-12) of the Employment and Unemployment schedules, Indian National Sample Survey (NSS) in columns (I)-(IV) and from rounds 62 (2005-06) and 64 (2007-08) in column (V). Estimation sample restricted to women from urban households who are 18-50 years of age, and 1) excludes women from Delhi, Mumbai (Maharashtra) and Jalpaiguri (West Bengal) in column (I), 2) excludes women currently enrolled in education in column (II). Data on reports of physical and sexual assaults is extracted from the Global Database of Events, Language, and Tone (GDELT).

Figures

Figure 1
Media reporting of assaults over time



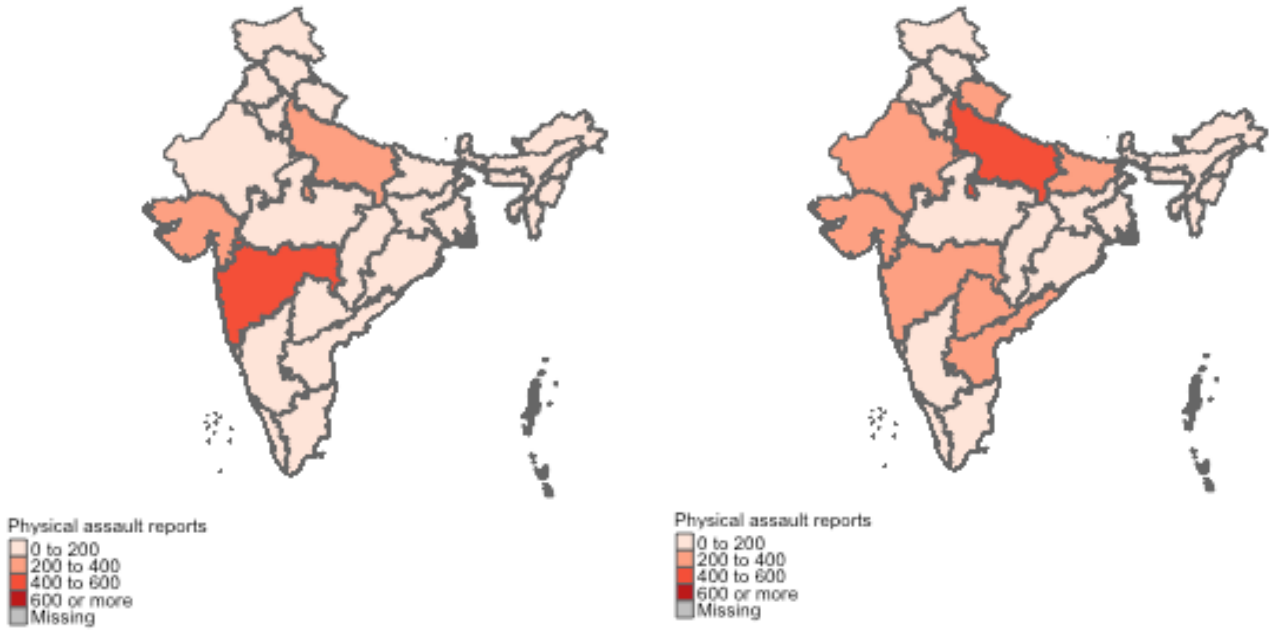
(a) Physical and sexual assaults



(b) Reports of physical and sexual assaults

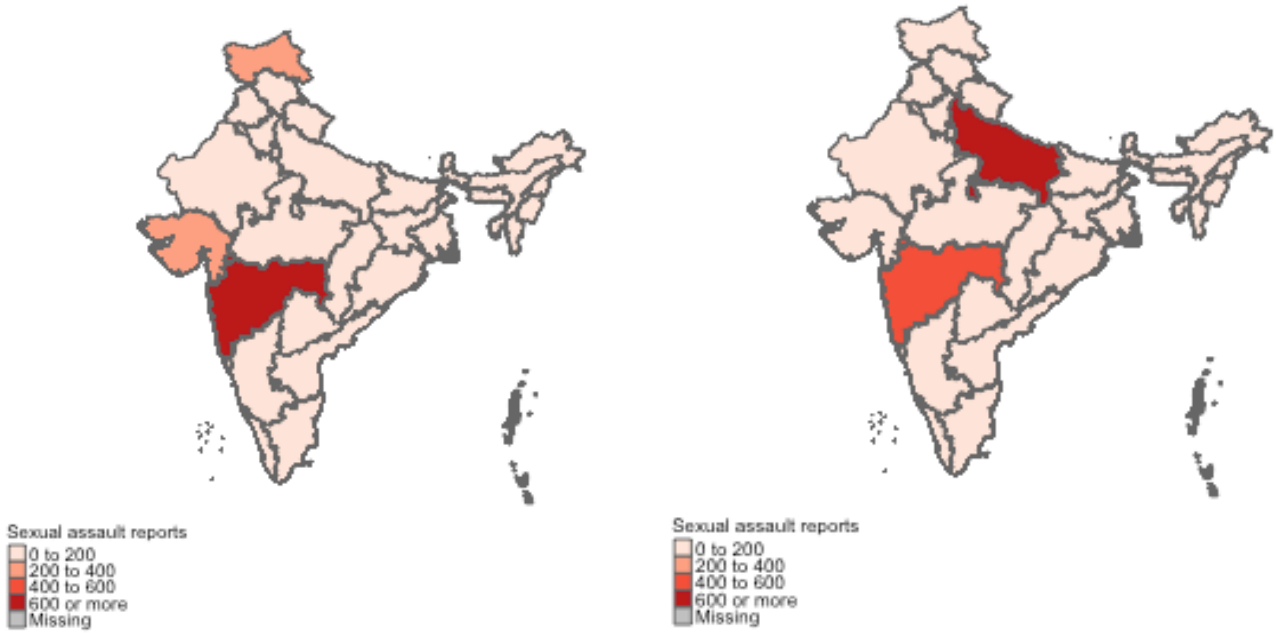
Source: Data on incidents and reports of assaults is extracted from the Global Database of Events, Language and Tone (GDELT).

Figure 2
Media reported assault reports across states



(a) Physical assault reports, April 2009 to March 2010

(b) Physical assault reports, April 2011 to March 2012

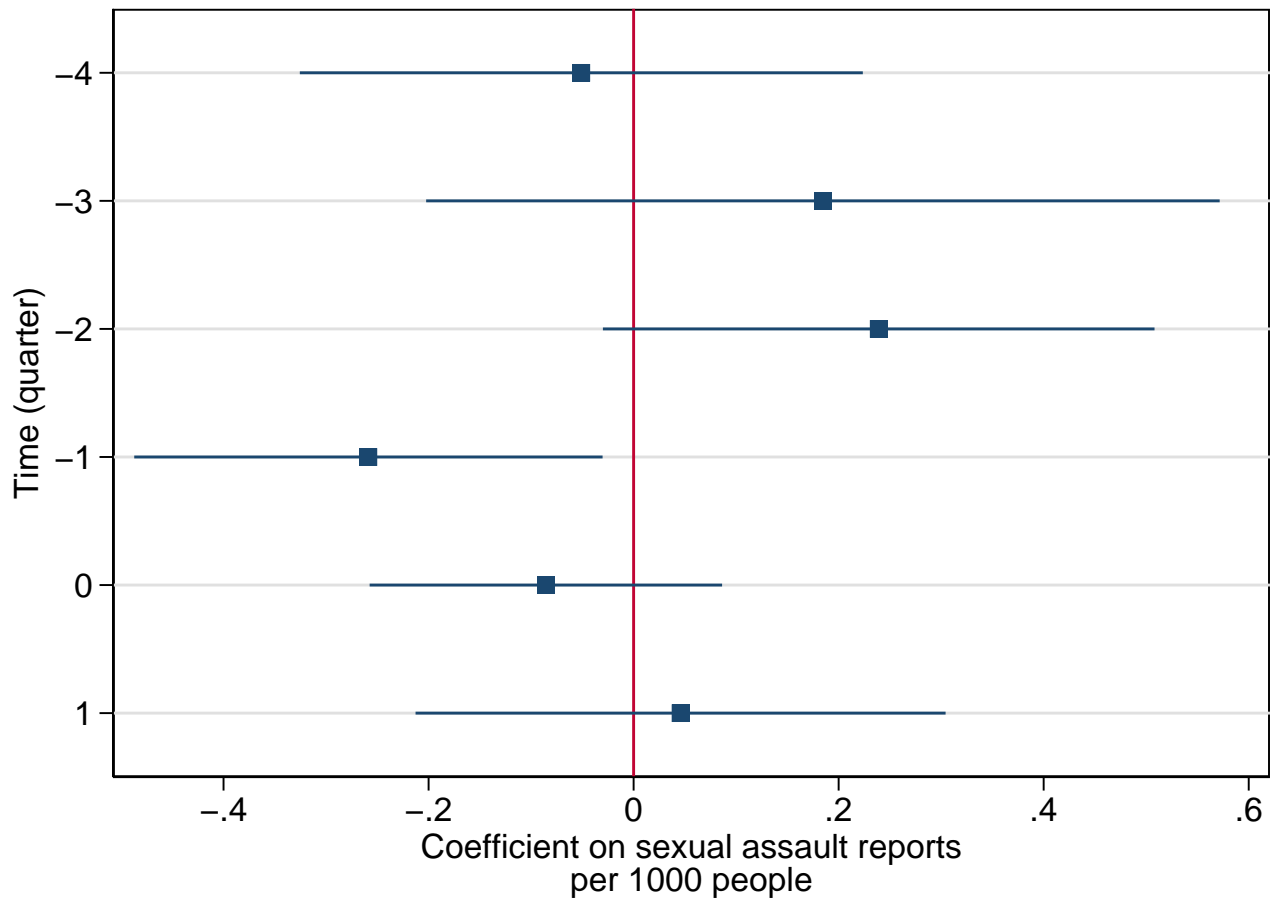


(c) Sexual assault reports, April 2009 to March 2010

(d) Sexual assault reports, April 2011 to March 2012

Source: Data on assault reports is extracted from the Global Database of Events, Language and Tone (GDELT).

Figure 3
Coefficient Plot from estimation of equation (7)



Source: Data from rounds 66 (2009-10) and 68 (2011-12) of the Employment and Unemployment schedules, Indian National Sample Survey (NSS). Estimation sample is restricted to women from urban households who are 18-50 years of age. Data on reports of assaults is extracted from the Global Database of Events, Language, and Tone (GDELT).

References

- Afridi, F., Dinkelman, T. & Mahajan, K. (2018), 'Why are fewer married women working in rural india? a decomposition analysis over two decades', *The Journal of Population Economics* **31**, 783–818.
- Aguilar, A., Gutierrez, E. & Villagran, P. S. (forthcoming), 'Benefits and unintended consequences of gender segregation in public transportation: Evidence from mexico city's subway system', *Economic Development and Cultural Change* .
- Aizer, A. (2010), 'The gender wage gap and domestic violence', *American Economic Review* **100**(4), 1847–59.
- Aizer, A. (2011), 'Poverty, violence and health', *Journal of Human Resources* **46**(3), 518–538.
- Anderberg, D., Rainer, H., Wadsworth, J. & Wilson, T. (2016), 'Unemployment and domestic violence: Theory and evidence', *Economic Journal* **126**, 1947–1979.
- Autor, D. & Duggan, M. (2003), 'The rise in the disability rolls and the decline in unemployment', *Quarterly Journal of Economics* **118**(1), 157–205.
- Bartik, T. (1991), *Who Benefits from State and Local Development Policies?*, Kalamazoo, MI: W.E. Upjohn Institute.
- Becker, G. & Rubinstein, A. (2011), Fear and the response to terrorism: An economic analysis. Technical report, Center for Economic Performance, LSE.
- Bhalotra, S., Kambhampati, U., Rawlings, S. & Siddique, Z. (forthcoming), 'Intimate partner violence: The influence of job opportunities for men and women', *The World Bank Economic Review* .
- Bhalotra, S. & Umana-Aponte, M. (2010), The dynamics of women's labour supply in developing countries. IZA Discussion Paper 4879.
- Blundell, R. & MaCurdy, T. (1999), Labor supply: A review of alternative approaches, in A. O. & D. Card, eds, 'Handbook of Labor Economics', Vol. 3, Elsevier, pp. 1559–1695.
- Borker, G. (2018), Safety first: Perceived risk of street harassment and educational choices of women. Unpublished Manuscript.
- Chakraborty, T., Mukherjee, A., Rachapalli, S. & Saha, S. (2018), 'Stigma of sexual violence and women's decision to work', *World Development* **103**, 226–238.
- Currie, J., Mueller-Smith, M. & Rossin-Slater, M. (2018), Violence while in utero: The impact of assaults during pregnancy on birth outcomes, Working Paper 24802, National Bureau of Economic Research. URL: <http://www.nber.org/papers/w24802>
- Dyson, T. & Moore, M. (1983), 'On kinship structure, female autonomy, and demographic behavior in india', *Population and Development Review* **9**(1), 35–60.
- Field, E., Jayachandran, S. & Pande, R. (2010), 'Do traditional institutions constrain female entrepreneurship? a field experiment on business training in india', *The American Economic Review* **100**(2), 125–129.
- Fletcher, E., Moore, C. & Pande, R. (2018), Women and work in india: Descriptive evidence and a review of potential policies. Unpublished Manuscript.
- Guarnieri, E. & Rainer, H. (2018), Female empowerment and male backlash. Unpublished Manuscript.

- Heath, R. & Jayachandran, S. (2005), The causes and consequences of increased female education and labor force participation in developing countries, *in* A. S., L. Argys & S. Hoffman, eds, ‘Oxford Handbook of Women and the Economy’, Oxford University Press.
- Jayachandran, S. (2015), ‘The roots of gender inequality in developing countries’, *Annual Review of Economics* **7**, 63–88.
- Klasen, S. & Pieters, J. (2015), ‘What explains the stagnation of female labor force participation in urban india?’, *The World Bank Economic Review* **29**(3), 449–478.
- Kondylis, F., Legovini, A., Vyborny, K., Zwager, A. & Andrade, L. (2018), Demand for safe spaces: Avoiding harassment and stigma. Unpublished Manuscript.
- Krishnan, S., Rocca, C. H., Hubbard, A. E., Subbiah, K., Edmeades, J. & Padian, N. S. (2010), ‘Do changes in spousal employment status lead to domestic violence? insights from a prospective study in bangalore, india’, *Social Science and Medicine* **70**, 136–143.
- Leetaru, K. & Schrodtt, P. A. (2013), ‘Gdelt: Global data on events, location, and tone, 1979-2012’, *International Studies Association Annual Conference* .
- Muralidharan, K. & Prakash, N. (2017), ‘Cycling to school: Increasing secondary school enrollment for girls in india’, *American Economic Journal: Applied Economics* **9**(3), 321–350.
- Rawlings, S. & Siddique, Z. (forthcoming), ‘Domestic violence and child mortality in the developing world’, *Oxford Bulletin of Economics and Statistics* .
- UN, W. & ICRW (2013), Unsafe: An epidemic of sexual violence in delhi’s public spaces: Baseline findings from the safe cities delhi programme. [https://www.icrw.org/wp-content/uploads/2016/10/Baseline-Research-of-Safe-Cities-programme-\(1\)\[smallpdf.com\].pdf](https://www.icrw.org/wp-content/uploads/2016/10/Baseline-Research-of-Safe-Cities-programme-(1)[smallpdf.com].pdf).

Appendix Tables and Figures

TABLE A1

Inverse hyperbolic sine transformation for media reports of assaults

	(I)	(II)	(III)
arcsinh(Physical assault reports per 1000 people at $(t - 1)$)	0.0751 (0.1064)	-0.0299 (0.0692)	-0.0572 (0.0898)
arcsinh(Sexual assault reports per 1000 people at $(t - 1)$)	0.1544 (0.1622)	-0.1992*** (0.0705)	-0.2343** (0.0924)
SCST	0.0599*** (0.0063)	0.0688*** (0.0051)	0.0689*** (0.0052)
Hindu OBC	0.0143** (0.0060)	0.0064 (0.0041)	0.0072 (0.0041)
Muslim	-0.0438*** (0.0051)	-0.0465*** (0.0059)	-0.0458*** (0.0059)
Other Religion	0.0595*** (0.0116)	0.0380*** (0.0090)	0.0363*** (0.0089)
Age	0.0180*** (0.0012)	0.0182*** (0.0011)	0.0182*** (0.0011)
Age ²	-0.0002*** (0.0000)	-0.0002*** (0.0000)	-0.0002*** (0.0000)
Controls	Yes	Yes	Yes
District FE	No	Yes	Yes
State \times NSS sub-round FE	No	No	Yes
N	92123	92123	92123

Notes: Each column reports results from a separate regression. Results reported in columns (I)-(III) are from estimating equation (1) for the sample of urban women, where the inverse hyperbolic sine transformation has been applied to lagged media reports of physical and sexual assaults. The dependent variable is L , which takes the value one if a person spends non-zero time in regular or casual employment in the last week. Standard errors are clustered at the district level and reported in parentheses; * p-value < 0.05, ** p-value < 0.025, *** p-value < 0.01.

Source: Data from rounds 66 (2009-10) and 68 (2011-12) of the Employment and Unemployment schedules, Indian National Sample Survey (NSS). Estimation sample is restricted to women from urban households who are 18-50 years of age. Data on reports of physical and sexual assaults is extracted from the Global Database of Events, Language, and Tone (GDELT).

TABLE A2

Heterogeneity by per capita household consumption quartile with additional lags

<i>Sample:</i>	all	HH cons < 25 perc.	HH cons 25 – 50 perc.	HH cons 50 – 75 perc.	HH cons > 75 perc.
	(I)	(II)	(III)	(IV)	(V)
Sexual assault reports per 1000 people at $(t - 1)$	-0.2653** (0.1136)	-0.0208 (0.4564)	-0.2494 (0.2332)	-0.0975 (0.2102)	-0.5102*** (0.1390)
Sexual assault reports per 1000 people at $(t - 2)$	0.2097 (0.1243)	-0.6414 (0.6184)	-0.1541 (0.4256)	0.0240 (0.2581)	0.6057*** (0.1436)
Sexual assault reports per 1000 people at $(t - 3)$	0.1738 (0.1732)	1.0688 (0.6631)	0.2472 (0.2656)	0.1324 (0.2277)	-0.2076 (0.2179)
Sexual assault reports per 1000 people at $(t - 4)$	0.0154 (0.1076)	-0.7123 (0.4756)	0.1675 (0.2804)	-0.0791 (0.2986)	-0.2106 (0.1617)
Controls	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes
State \times NSS sub-round FE	Yes	Yes	Yes	Yes	Yes
N	92123	15380	22422	25317	29004

Notes: Each column reports results from a separate regression. Results are from estimating equation (2) after splitting the sample into four groups (columns (II)-(V)) based on the per capita consumption level of the household that a woman belongs to and estimating the equation separately for each group. The dependent variable is L , which takes the value one if a person spends non-zero time in regular or casual employment in the last week. Standard errors are clustered at the district level and reported in parentheses; * p-value < 0.05, ** p-value < 0.025, *** p-value < 0.01.

Source: Data from rounds 66 (2009-10) and 68 (2011-12) of the Employment and Unemployment schedules, Indian National Sample Survey (NSS). Estimation sample restricted to women from urban households who are 18-50 years of age. Data on reports of physical and sexual assaults is extracted from the Global Database of Events, Language, and Tone (GDELT).

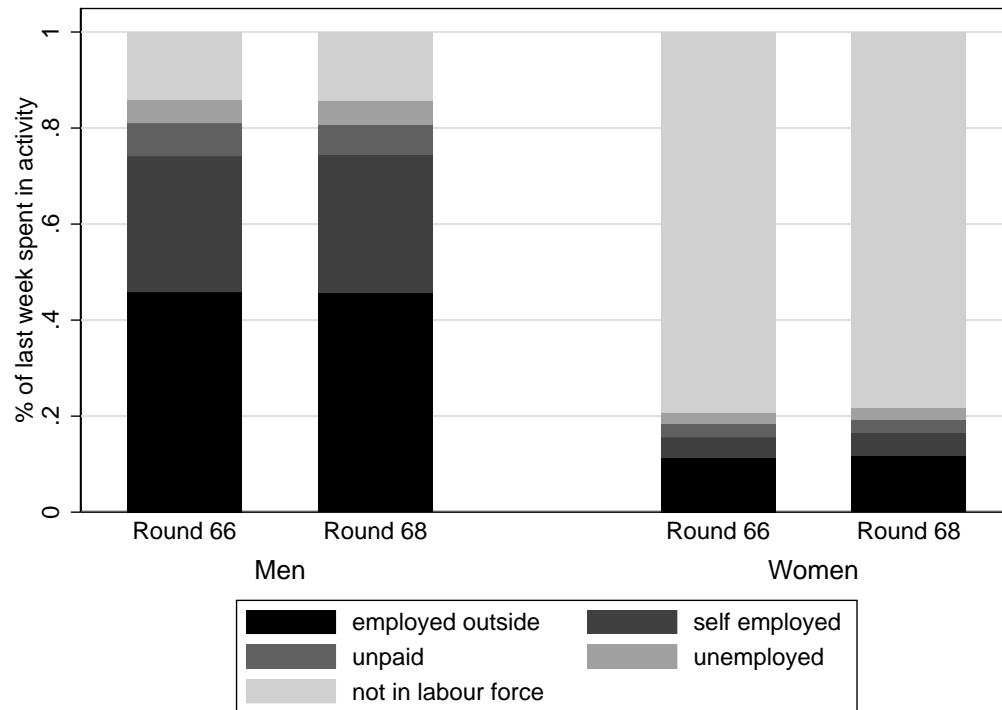
TABLE A3
Rural women

<i>Sample:</i>	urban Age 18 – 50 (I)	rural Age 18 – 50 (II)	rural Age 18 – 25 (III)	rural Age 18 – 22 (IV)
Physical assault reports per 1000 people at $(t - 1)$	-0.0568 (0.0849)	0.3556*** (0.0999)	0.2760 (0.1947)	0.2470 (0.1902)
Sexual assault reports per 1000 people at $(t - 1)$	-0.2313** (0.0910)	-0.0410 (0.1650)	-0.1914 (0.2201)	-0.3193 (0.1939)
Controls	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
State \times NSS sub-round FE	Yes	Yes	Yes	Yes
N	92123	139597	41243	26275

Notes: Each column reports results from a separate regression. Results in column (I) are from estimating equation (1) on the sample of urban women and in column (II)-(IV) from estimating equation (1) on the sample of rural women. Results in columns (II)-(IV) use estimation samples after imposing different age restrictions. The dependent variable is L , which takes the value one if a person spends non-zero time in regular or casual employment in the last week. Standard errors are clustered at the district level and reported in parentheses; * p-value < 0.05, ** p-value < 0.025, *** p-value < 0.01.

Source: Data from rounds 66 (2009-10) and 68 (2011-12) of the Employment and Unemployment schedules, Indian National Sample Survey (NSS). Estimation sample is restricted to women who are 18-50 years of age, and who belong to urban households in column (I) and rural households in columns (II)-(IV). Data on reports of physical and sexual assaults is extracted from the Global Database of Events, Language, and Tone (GDELT).

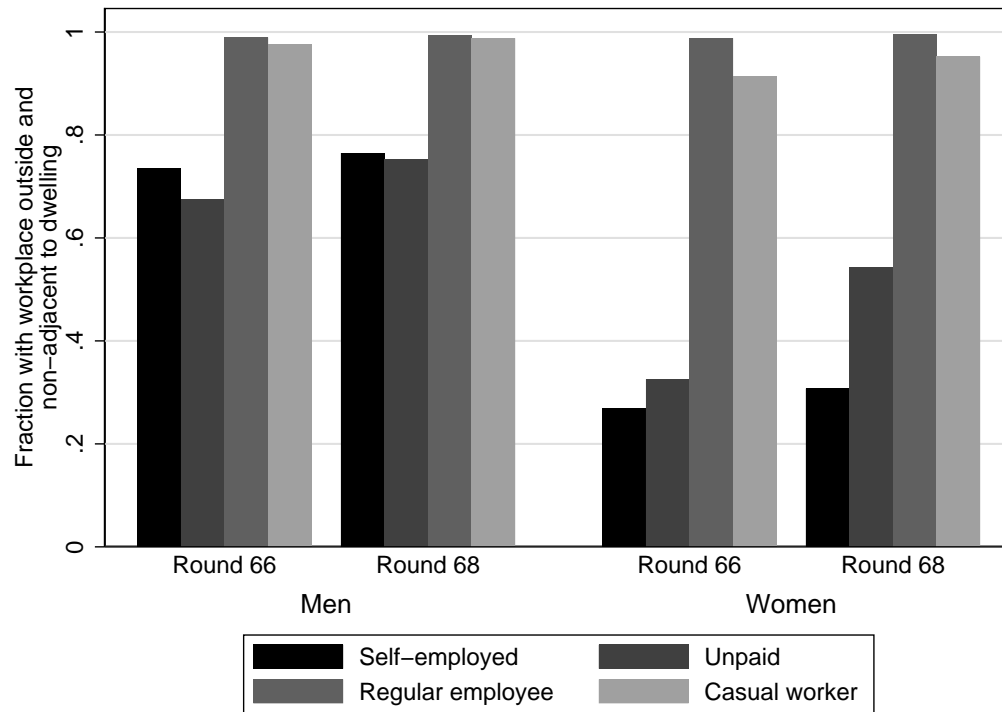
Figure A1
Time allocation in the past week



Source: Data from rounds 66 (2009-10) and 68 (2011-12) of the Employment and Unemployment schedules, Indian National Sample Survey (NSS). Sample is restricted to men and women from urban households who are 18-50 years of age.

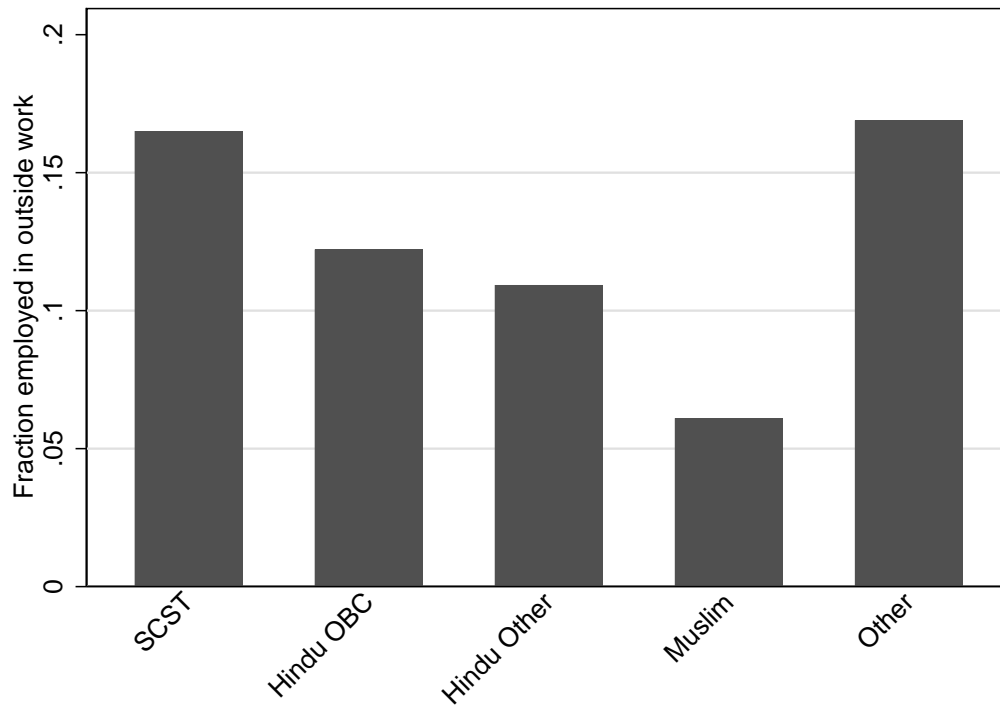
Figure A2

Location of workplace by type of work undertaken over the past year



Source: Data from rounds 66 (2009-10) and 68 (2011-12) of the Employment and Unemployment schedules, Indian National Sample Survey (NSS). Sample is restricted to men and women from urban households who are 18-50 years of age.

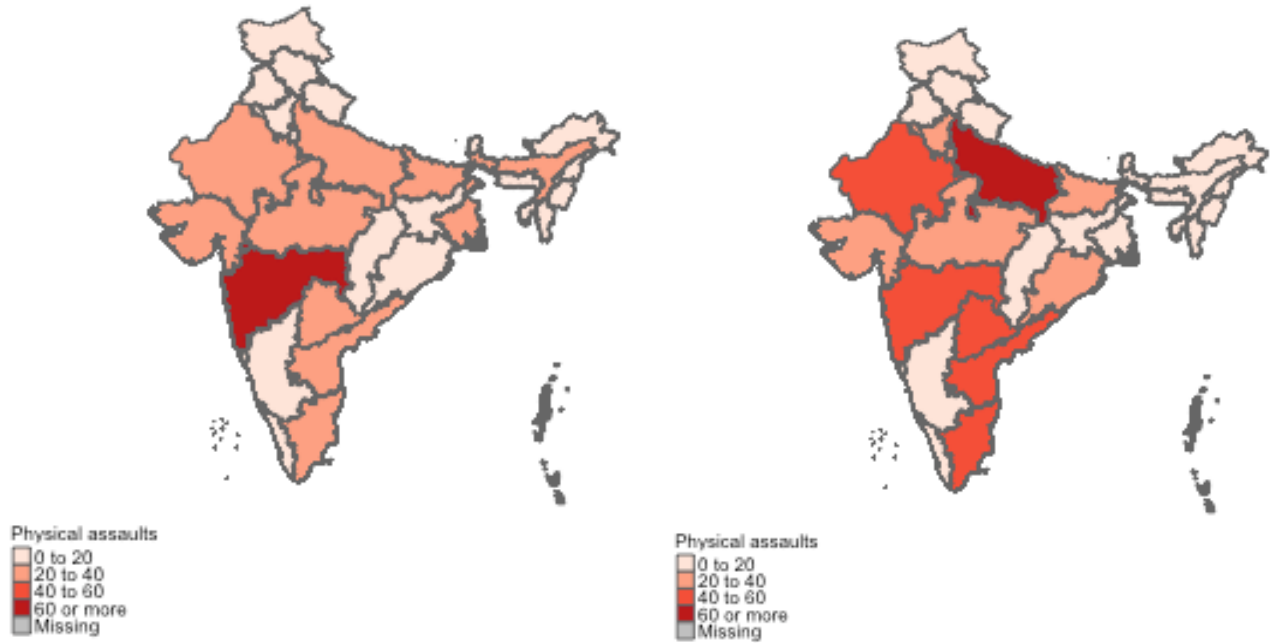
Figure A3
Labor supply of urban women by social group



Notes: SCST refers to women belonging to Scheduled Caste or Scheduled Tribe households. Hindu OBC refers to women belonging to Hindu Other Backward Caste households. Hindu Other refers to women belonging to non-SCST, non-OBC (primarily high-caste) Hindu households. Muslim refers to women belonging to Muslim households. Other refers to women belonging to Christian, Sikh, Jain, Buddhist, Zoroastrian or Other religion households.

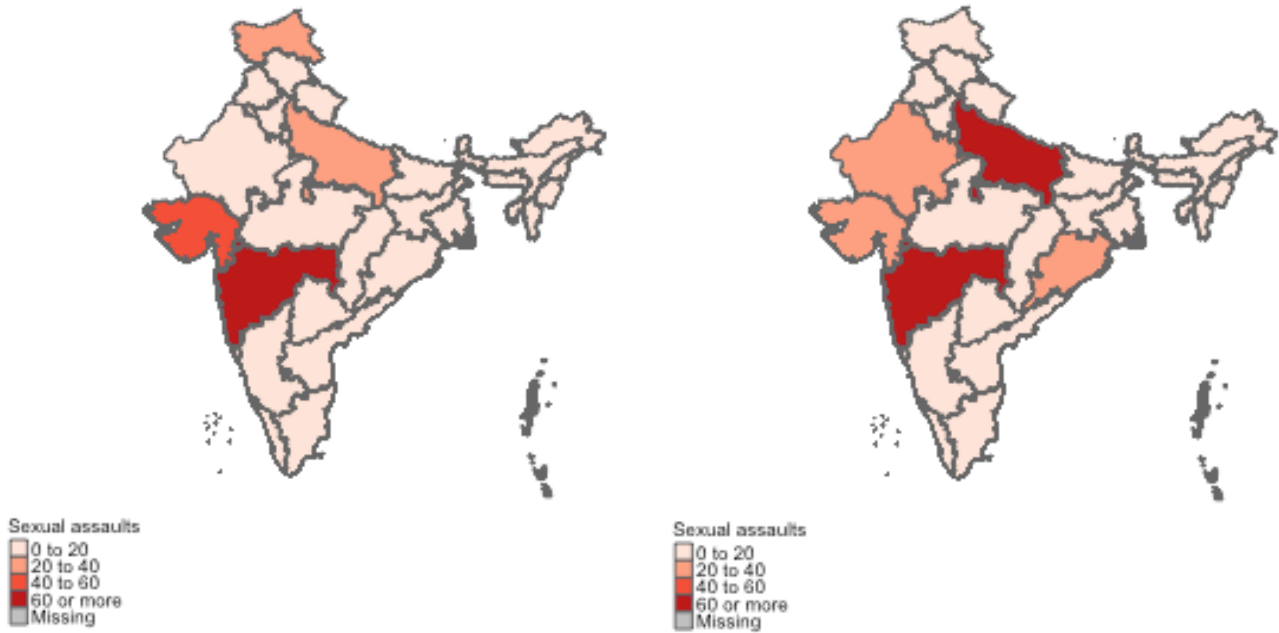
Source: Data from rounds 66 (2009-10) and 68 (2011-12) of the Employment and Unemployment schedules, Indian National Sample Survey (NSS). Sample is restricted to women from urban households who are 18-50 years of age.

Figure A4
Media reported assaults across states



(a) Physical assaults, April 2009 to March 2010

(b) Physical assaults, April 2011 to March 2012

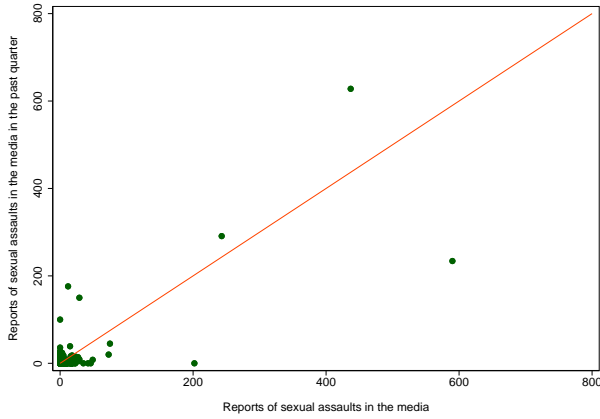


(c) Sexual assaults, April 2009 to March 2010

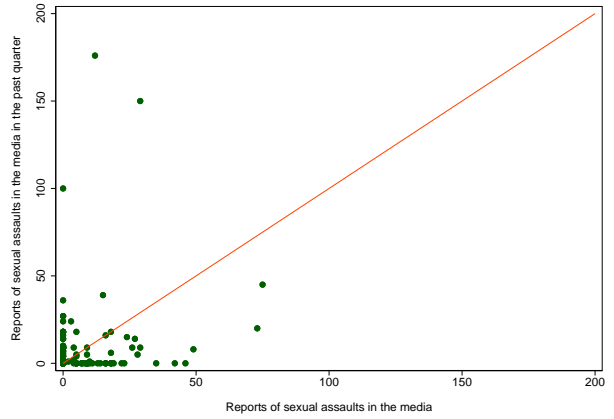
(d) Sexual assaults, April 2011 to March 2012

Source: Data on assaults is extracted from the Global Database of Events, Language and Tone (GDELT).

Figure A5
Variation in media reports of sexual assaults



(a) variation, including outliers



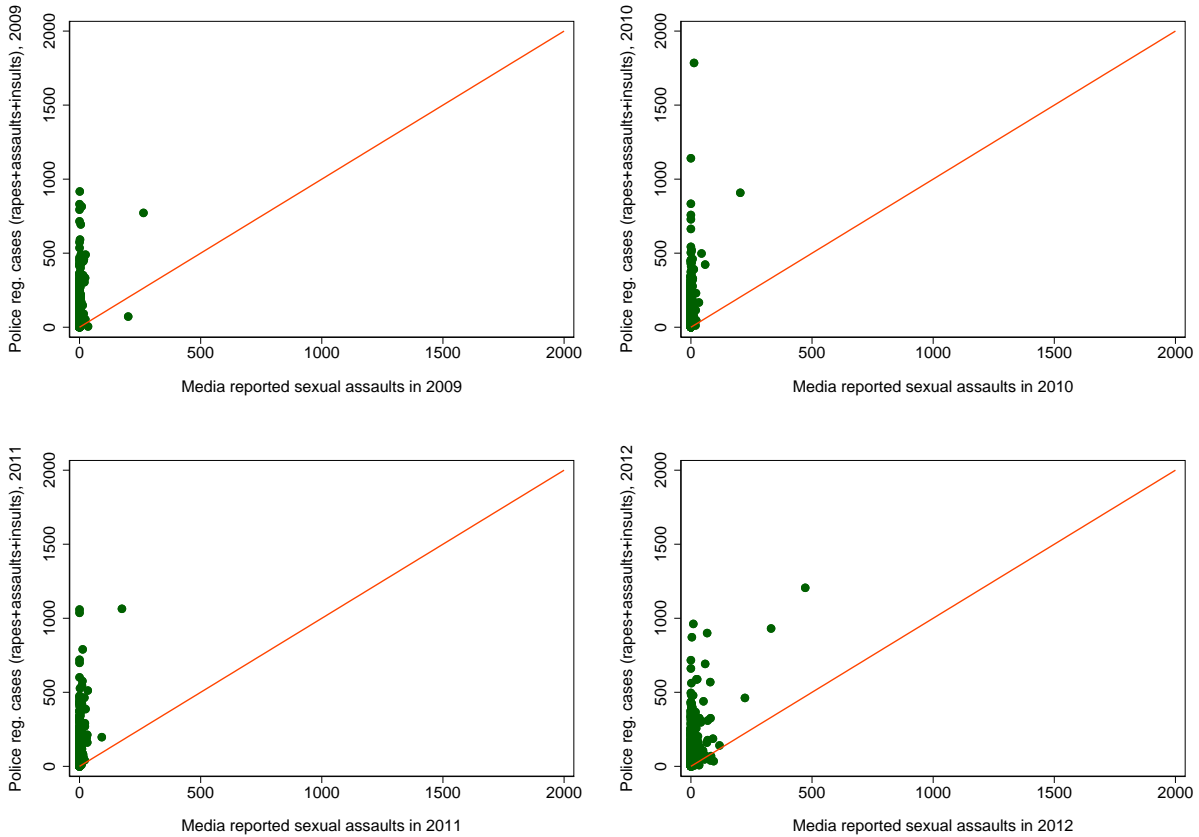
(b) variation, excluding outliers

Notes: Each point on the scatterplot gives the combination of media reports of sexual assaults in the current and previous quarter for the same district. Panel (b) excludes the districts of New Delhi, North Delhi, Mumbai (Maharashtra state) and Jalpaiguri (West Bengal state) for quarters where sexual assault media reports exceed 200. Superimposed is the 45 degree line of equality.

Source: Data on reports of assaults is extracted from the Global Database of Events, Language and Tone (GDELT), aggregated at the district and quarter level.

Figure A6

Police registered crimes against women and media reporting of sexual assaults

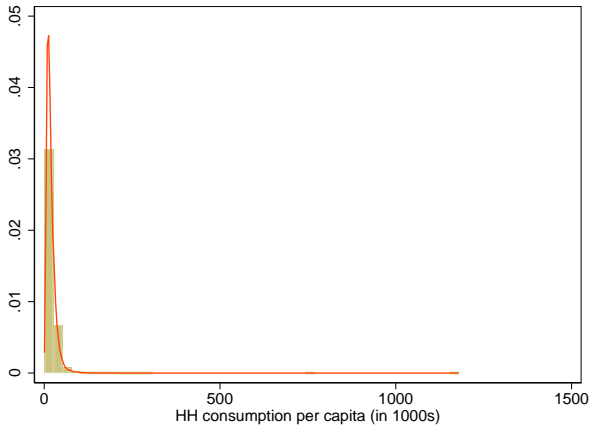


Notes: Each point on the scatterplot gives the combination of police registered crimes against women (the sum of Indian Penal Code or IPC categories of rape, assault on women with intent to outrage her modesty and insult to the modesty of women) and media reported sexual assaults for a district in a particular year. Superimposed is the 45 degree line of equality.

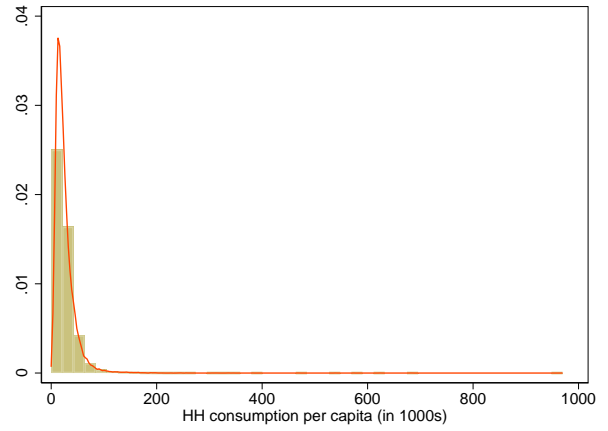
Source: Data on annual crimes against women by district is from publications by the National Crime Records Bureau (NCRB) at the Ministry of Home Affairs in India. Data on incidents and reports of assaults is extracted from the Global Database of Events, Language and Tone (GDELT), aggregated at the district and year level.

Figure A7

Distribution of Household Consumption per capita by NSS round



(a) NSS round 66



(b) NSS round 68

Notes: Household Consumption per capita is the value of household consumption per capita over the last 30 days in rupees.

Source: Data from Rounds 66 (2009-10) and 68 (2011-12) of the Employment and Unemployment schedules, Indian National Sample Survey. Sample restricted to urban households.