Participatory Theater Empowers Women: Evidence from West Bengal, India*

Saraswata Chaudhuri† Karla Hoff‡ Jyotsna Jalan§ Sattwik Santra¶

December 15, 2025

Abstract

We evaluate an intervention by *Jana Sanskriti* (*JS*), a non-governmental organization in India, that uses community-based participatory theater to reduce violence against women. Using a survey of married women aged 18-49 in 92 villages, we find, on average, *JS* reduced the proportion that experienced physical abuse by 9 percentage points (pp) from a baseline of 33%. We also find *JS* increased several other measures of women's empowerment. Since the *JS*-intervention was not randomized, we estimate these causal effects by adjusting for potential confounders that *JS* informed us. We find that strong unadjusted confounders are needed to overturn these causal effects.

JEL Classification: C52, J12, J16, K42, L82, Z13

Keywords: domestic violence, gender, narratives, patriarchy, sensitivity analysis

*The World Bank Research Support Budget and the Centre for Training and Research in Public Finance and Policy, Kolkata provided financial support. We benefitted from comments by I. Andrews, S. Banuri, K. Basu, S. Bhattacharya, A. Brown, S. Dercon, V. Girard, W. Kets, M. Kolesar, M. Masten, S. Mitra, D. Mookherjee, V. Orozco, J. Ricart-Huguet, A. Sen, M. Shayo, R. Somanathan, P. Todd, R. Yarrow, V. Zinde-Walsh, and various conference and seminar participants. This work would not have been possible without the support of Jana Sanskriti, our field investigators, and J. Panda, who supervised the survey.

†Department of Economics, McGill University & Cireq. (saraswata.chaudhuri@mcgill.ca)

‡Department of Economics, Columbia University. (kh2830@columbia.edu)

§Centre for Studies in Social Sciences. (jjalan@gmail.com)

¶Centre for Training and Research in Public Finance & Policy. (sattwik@cssscal.org)

1 Introduction

Nearly one in three women worldwide – across all regions, ages, and socioeconomic groups – has experienced intimate partner violence at some point in their life (WHO, 2025).[1] Such violence imposes enormous costs on survivors, their families, communities, and entire economies. These costs are both economic (measurable in money) and non-economic (human and social harms that are harder to quantify) (Duvvury et al. (2013), Koenen et al. (2003), Egeland (1993)). Despite widespread recognition of these harms, domestic violence remains prevalent and, in many societies, is widely accepted. In low-income and lower-middle-income countries, respectively, a median of 58% and 33% of women believe a husband is justified in hitting or beating his wife under certain circumstances (Supplemental Appendix A, Table 6). These attitudes reflect deep-seated social norms that can perpetuate cycles of abuse.

Under international pressure, many developing countries enacted laws against domestic violence for the first time during the 1990's and 2000's. However, where domestic violence is socially accepted, the power of the law to reduce violence against women is limited (Htun and Jensenius, 2020). Conditional cash transfers can reduce domestic violence (Haushofer and Shapiro, 2016), but they are costly and their long-term effects – beyond ten months after intervention – are uncertain (Roy et al., 2019). Interventions like mandating a minimum age for marriage and increasing the minimum years of required schooling for girls have also been ineffective in raising the women's agency in the household (Erten and Keskin, 2018).

The early success of *edutainment* – entertaining narratives designed to induce social change – raised hopes that this could be a potential tool to reduce domestic violence. But *edutainment* designed for the purpose of making intimate partner unacceptable has not lived up to expectations. Only two evaluations of its impact have been done, and both found that it reduced by little or not at all the acceptability of such violence against women (Banerjee et al. (2019a); Cooper et al. (2020)). *Edutainment*, like entertainment, is better suited to influencing independent behaviors than dependent behaviors (behaviors that depend on

[1] We will use the terms intimate partner violence and domestic violence interchangeably.

societal approvals) (Bicchieri, 2017). Success stories of *edutainment* are consumption of iron-fortified salt (Banerjee et al., 2018), avoiding unsafe sex (Banerjee et al., 2019b), reporting corruption (Blair et al. (2019); Cooper et al. (2020)), doing well in school (Riley, 2024) – all examples of independent behaviors. It is, however, unlikely to sustain change in interdependent behaviors like domestic violence where beating is valorized, unless it triggers community-wide discussions and renegotiation of social norms.[2]

Our paper shows that community-based participatory theater can break down attitudes that legitimize domestic violence, and empower women. To our knowledge, our paper is the first large-scale quantitative impact-evaluation of community-based participatory theater.[3]

Participatory theater (also known as Forum Theater because it provides a forum for dialogue), first developed in Brazil as part of the Theater of the Oppressed, uses drama to discuss and challenge social norms harming marginalized groups.[4] It empowers community members without a voice to co-create and actively engage in performances, shaping how their stories are understood and promoting collective reflection and change. Instead of offering a solution to a problem, a play in participatory theater confronts the audience with a problem for them to solve. The plays are intentionally incomplete. The aim is not to give advice (e.g., don't beat your wife) but rather to induce participants to think about social change

[2]An Ethiopian woman observed "[i]t is sometimes necessary for husbands to beat their wives when they commit mistakes to correct them...it is also a sign of strong manhood" (Narayan et al., 2000). An expression of men in North India is "What kind of a man is he who does not hit?" (Chowdhry, 2015). In parts of India, a husband who does not beat/hit a "recalcitrant" wife risks a loss of honor, with adverse consequences on his family's social standing, economic opportunities, and his children's marriage prospects (Derne, 1994).

[3]Asante and Zakaria (2021) find that participatory theater reduced open defecation in a fishing community in Ghana. Thambu et al. (2020) find that participatory theatre when used as a pedagogy in Malaysian secondary schools, can foster moral reasoning, collaborative thinking, and active engagement among students. Both interventions were assessed through small-scale qualitative methods, utilizing direct observation, participant interviews, and focus group discussions to capture contextual data on participant experiences.

[4]Augusto Boal developed his popular-theatre ideas in Brazil during the 1950s–60s. After his exile, he first staged what became "Forum Theatre" in 1973 in Peru, during a Freire-inspired literacy project, when an audience member stepped in to replace an actor and propose an alternative outcome. He then formalized and disseminated the method through his mid-1970s publications and workshops in Europe.

by "playing with" the story.

The plays are provocative. Plays typically end with an accusation or a question directed at the audience to prompt the ensuing discussion. In the repeated enactments that follow an uninterrupted performance, members of the audience are encouraged to go on stage, play the role of a character, and through that role find ways to avert the oppression that occurs in the original story. Members of the audience, including wives who in their everyday lives might have little voice in their household, have the experience of averting oppression in the fictional drama. Participatory theater induces men and women, little by little, to think critically about their socially prescribed roles and try out rebellious actions onstage.

Our paper evaluates the effects of exposing villages to performances by the Jana Sanskriti Center for Theatre of the Oppressed (JS) over a period of about 16 years, aiming to understand how sustained engagement with such theater can foster empowerment and social transformation. JS is a grassroots organization founded in 1985 in rural West Bengal by Sanjoy Ganguly. JS is now one of the world's largest and longest-running community-based participatory theater groups (Yarrow, 2021). Like street theater, JS performs in public spaces without charging fees; shows are widely advertised and well attended. During our 2014 survey, JS had a core team and 13 village-based troupes in South 24 Parganas, each performing regularly in 12–19 villages. The actors are community members who have been trained to initially resist rebellious actions and respond in ways that reflect village norms.

A JS-performance begins with music, dance, and games to build rapport, followed by a 20-minute uninterrupted drama. Then the *Joker* (Theater facilitator) invites the audience to intervene. Actors replay the drama, and any audience member may stop the action, replace a character onstage, and attempt to change the story's outcome. These participants, called *spect-actor*s, both act in the play and propose strategies for change (Ganguly, 2010). Actors question and refine each *spect-actor*'s ideas as the *Joker* guides the discussion. The interactive segment, often lasting 2–3 hours, includes probing questions from the *Joker* and multiple audience interventions. Each play is usually performed on 3 days within 1–3 months

in the same village (Yarrow, 2017). The first performance is mostly entertainment; on later days, the social message deepens, and residents begin to rethink previously accepted norms. *Spect-actor*s craft alternative endings that prevent injustice, prompting the community to reflect on its own practices. Plays are adapted between performances to reflect local developments, making subsequent shows more dynamic and engaging. As events during the performances become common knowledge, they can gradually influence community norms.

The treatment effect we would ideally assess is the difference in the potential outcomes, e.g., level of physical abuse of wives by husbands in a village, with and without exposure of the village to *JS* performances. It is, of course, not possible to observe both potential outcomes. We focus, as is standard, on identifying and estimating the causal parameters that are the common aggregate/average measures of the treatment effect.

Once *JS* had started performing in a village, it continued to perform there regularly. That is, selection of the villages exposed to *JS* was a one-time decision. However, *JS* did not randomly select the villages where it performs. *JS* informed us that they sought to perform in relatively densely populated villages in South 24 Parganas that were geographically distant from towns, and not well connected to the outside world. It is possible that such villages had more conservative beliefs about gender roles and that, in turn, could influence the potential outcomes. That is, these selection variables (pretreatment characteristics) could be potential confounders. Hence, we adjust/control for them to circumvent the potential selection bias in identifying the causal parameters under a selection on observables assumption.

We obtained the data for these pretreatment characteristics of the villages from Census of India (1991). We obtained the data on the outcome variables of interest by conducting a survey in 2014 in the rural South 24 Parganas. We randomly selected 31 out of the total 125 villages in which *JS* had performed ("treatment villages") and 61 villages in which *JS* had never performed ("control villages"). See Figure 2 in Supplemental Appendix A. We used the voters list for the 2011 legislative assembly election in West Bengal to randomly select 3481 households from these villages. In the sampled households, we administered a survey

to the wife (between 18-49 years) and to her husband. The wives and their husbands were asked privately, by female and male interviewers respectively, the same questions on issues concerning women's autonomy. Only the wives, and not their husbands, were asked questions concerning actual incidence of domestic abuse, to reduce the risk of repercussions on the wives for participating in a survey that covered potentially illegal actions of domestic abuse by the husband. These latter questions were asked only toward the end of the interviews.[5]

Our estimates suggest that on average, this sustained intervention of about 16 years of exposure of a village to *JS*'s performances reduced the proportion of village households where wives reported being physically abused by their husbands by about 9 percentage points (pp) from an estimated baseline mean of 33% (mean potential outcome without treatment). This reduction is about 80% of the standard deviation of the observed physical abuse. The average effect on the treated villages was a little larger. Based on the wive's and husbands' responses we find similar effectiveness of the exposure of a village to *JS*'s performances in reducing not only other forms (e.g., sexual) of domestic abuse but also various indicators of patriarchal norms like limited or no participation in making decisions related to the visiting wife's family or relatives, number of children to have, use of contraceptives, among others.[6] It also increased their willingness to report to someone domestic violence that they witnessed.

Point-identification of these causal parameters is based on our selection on observables assumption that all the confounders between the treatment and the potential outcomes are adjusted for. This assumption cannot be tested statistically. However, this assumption is

[5]Other features of our survey: To avoid priming effects, we obtained agreement from *JS* not to perform any plays in a given village within the four-month period before the interviews were held in a village. To avoid demand effects, we never represented ourselves as evaluating *JS*, and the survey did not mention *JS* until the final section of questions. To avoid contagion, we completed a survey in each village in a single day.

[6]We obtain very similar estimates based on the estimators like (1) regression adjustment; (2) inverse probability weighting; (3) combination of regression and inverse probability weighting; (4) matching; and also machine learning methods such as (5) double-debiased machine learning estimators based on covariate selections using lasso; and (6) augmented inverse probability weighting or targeted maximum likelihood estimation where predictions are done by generalized random forest, that are commonly used in empirical analysis. We report all estimates, but focus our discussion on results for the regression adjustment estimator.

violated if there persists confounders unadjusted/uncontrolled for in our analysis. Therefore, we use sensitivity tools proposed by Masten and Poirier (2018, 2020) and Chernozhukov et al. (2024) to evaluate how robust our results are to violations of the selection on observables assumption. We report the magnitude of association that would be needed between possible unadjusted/uncontrolled confounders with the treatment and with the potential outcomes to cause enough bias in our causal estimates to overturn their signs. The reader ultimately decides the plausibility of such strength of association. However, given what we know about *JS*'s treatment assignment mechanism, we argue by using "benchmarking" exercises (Imbens, 2003) that such strength of association with the unadjusted/uncontrolled confounder – often referred to as "unobserved confounder" henceforth – would have to be unrealistically strong to overturn (change the sign of) our causal results for many of the outcomes.

Outline: Section 2 discusses mechanisms by wthich community-based participatory theater may influence behavior and norms. Section 3 explains our sampling and econometric strategies. Sections 4 and 5 present estimates of the impact of *JS* and a summary of their robustness to possible unobserved confounders. Section 6 concludes. Tables and figures are at the end of the paper. Supplemental Appendices A, B, C, D and E contain, respectively, background tables and figures, an extended version of the robustness and sensitivity analyses, synopses of 4 JS plays, our survey design, and our survey questionnaire. Further estimation and sensitivity results are similar to those we report here and are available from the authors.

2 Mechanisms in Forum Theater to foster social change

Community-based participatory theater (Forum Theater) fosters social change by influencing social norms and identities through three main mechanisms discussed below. Supplemental Appendix C presents synopses of some *JS* plays for reference to our discussion in this section.

2.1 Identity Shift through collective reframing and role playing:

During the performance, *spect-actor*s assume the roles of characters and experiment with alternative ways of behaving. From their actions and words and the reactions of the trained actors and the audience, a collectively created new ending to the story emerges. The collective realization that a better ending is possible is the first step towards reframing oppressive social norms.

Economists have also recognized recently that narratives provide mental frames through which people organize information and interpret reality. By reframing situations, new stories can have a large impact on behavior and aggregate outcomes. Benabou et al. (2018) show how new stories that raise self-image concerns influence the social norms that individuals follow. Stories influence macroeconomic fluctuations (Shiller, 2019), work effort (Akerlof and Rayo, 2020), health choices (Banerjee et al. (2018); Banerjee et al. (2019a)), fertility rates (La Ferrara et al., 2012), the ability to perform (Hoff and Stiglitz (2010); Riley (2024)) and racial hate and discrimination (Ang (2023); Esposito et al. (2023)). Akerlof (2020) characterizes the neglect of stories in mainstream economics as a "sin of omission". The distinctive contribution of participatory theater is to *give members of a community the opportunity to collectively change stories that justify their subjugation.* In doing so, it makes possible collective story editing and redistributes to oppressed groups an important power that is normally held only by socially dominant groups — the power of cultural representation.

2.2 Transmitting information and eliciting empathy

Ideology can act as a cognitive shield and block the absorption of information, thus preserving the false beliefs on which the ideology is based (Kahan et al., 2017). Exposure and experience can pierce such ideological blinders. In a field experiment in Israel, Jha and Shayo (2019) show that a few weeks of actively trading Israeli and Palestinian financial assets changed the participants' beliefs about the sociotropic costs of war. This shifted their votes in the 2015 Israeli election toward parties more supportive of the peace process. Engagement in

the stories of *JS* leads some in the audience to see for the first time the suffering of victims of domestic violence and feel empathy for them. Through narratives whose broad outlines would be familiar to most people in the audience, they can see themselves from a third-person perspective, as if they were spectators of their own lives. They can be 'transported' into the story and they may, at least for the moment, see things in an entirely new way. After watching a play on domestic violence, one member of the audience encountered the leading actress at a bus stop and pledged to her, "*Didi* (sister), I will not beat my wife again. I beat her quite often. When you were crying after being beaten by your husband in the play, I remembered my wife. She cries exactly like that when I beat her." (Ganguly (2010), p. 30).

2.3 Performing agency

After a behavior has become routinized and automatic, a higher level of engagement than passive viewing or listening is necessary to induce behavior change: individuals must not only think differently, they must also practice acting differently (Bourdieu, 1977). Forum theater gives people a relatively safe space in which to discuss and analyze their responses and those of others and to intervene. The *Joker* is required not to comment on the proceedings but to ask questions to advance the discussion and to ensure that all voices are heard (Mohan, 2010). Individuals rehearse social change onstage, which Yarrow (2012) describes as "performing agency". Evidence of intense parasocial engagement is that a member of the audience once threw a brick at an actor who was about to sell a child into slavery (Mills, 2009).

The audience learns immediately some of the effects of new strategies from the responses of the actors. *Spect-actors* may get social validation from their interactions with the actors and the applause of the audience. This may bolster the courage to adopt new behaviors. It is especially important for women to have the experience of "performing agency" or seeing their peers do it, since in rural India women are trained from childhood not to argue or be assertive. To see other women being assertive and confronting marital abuse on stage would broaden the range of actions that they, too, might believe are possible for them.

3 Sampling and Econometric Strategies

3.1 Sampling strategy for our survey

From the late-1990s to the present, *JS* has regularly performed in about 125 villages in the Sundarbans delta in the South 24 Parganas district of West Bengal (see Figure 2 in Supplemental Appendix A). These villages constitute the population of our treatment villages. We drew a random sample of 31 treatment villages from this population.

A fundamental issue in any impact evaluation is the choice of the control units. One should choose the control villages to be as similar as possible to the treatment villages in terms of pretreatment characteristics such as population, remoteness, communications facilities, etc. that (as we were informed by *JS*) influenced *JS*'s selection of villages for its performances. The objective of choosing the controls is to make the treatment assignment "as if random" once these pretreatment characteristics are adjusted for.

Selecting as controls the nearby villages within the same three administrative blocks as the treatment villages is not ideal for an intervention like the *JS* performances for which spillover effects are likely. Such proximity may make it difficult to isolate non-exposed from exposed villages and enforce compliance, as residents of control villages could easily travel to treatment villages to attend performances. To minimize spillovers and ensure non-exposure, we instead selected control villages from the three neighbouring administrative blocks since travel between blocks is limited due to the presence of intervening water bodies. We selected 61 control villages from these three neighboring administrative blocks. There are no contiguous control and treatment villages in our sample (Figure 2, Supplemental Appendix A). About 90% of the respondents in the treatment villages had heard of *JS*. No respondent in the control villages had heard of *JS*. This is strong evidence that substantive spillovers from exposure to *JS* did not occur. To the extent that minor spillovers may have occurred, our estimated impacts of village exposure to *JS* are likely biased towards zero.

Because the treatment villages were more populous (in 1991) than the average village

in the Sundarbans delta, we drew control villages with probabilities proportional to their 1991 population (see Supplemental Appendix D for our sampling strategy). Consequently, on average, the control villages turned out to be more densely populated, less remote, and better connected by communication facilities – the three characteristics that JS reported considering when selecting performance locations – than the treatment villages (Table 1).

Our analysis for impact evaluation focuses on the combined population of the treatment and control villages which is representative of the relatively densely populated (1991) villages in the Sundarbans delta of South 24 Parganas. When the parameter of interest is the impact on the treated, the target population is the 125 villages where JS has regularly performed.

Between March 2014 and March 2016, we conducted a household survey to collect data on domestic abuse and its social acceptance. The survey sample includes 1,635 married couples from 31 treatment villages, and 1,846 married couples from 61 control villages. The average age of female (male) respondents was 31 (37) in the treatment villages and 30 (36) in the control villages. Table 1 reports descriptive statistics for the outcomes obtained as village-level averages from our survey responses, and covariates (from Census of India (1991)) that likely influenced JS's village selection. Supplemental Appendix E presents the parts of our survey questionnaire on which the analysis in the paper draws.

3.2 Empirical framework

Let $D_i \in \{0, 1\}$ be the treatment indicator, i.e., $D_i = 1$ if JS has been performing in village i since 1998, and $D_i = 0$ if JS has never performed in village i. Until the time of our survey in 2014, JS never stopped performing in a village once it had started performing there. To fix ideas, we continue the exposition with one of the outcomes – physical abuse. This variable is defined as the proportion of households in a village in which a woman reports that her husband physically abuses her.

Let $Y_i(1)$ and $Y_i(0)$ denote the potential "physical abuse" in village i with and without treatment. In our context, it makes sense to think about potential outcomes for both

treatment and control villages given their geographical proximity and similarity in social, economic and demographic conditions. We assume no spillovers: i.e., the potential outcomes in a village do not depend on another village's exposure/nonexposure to the *JS* plays.

We observe $Y_i \equiv Y_i(D_i) = (1 - D_i)Y_i(0) + D_iY_i(1) = Y_i(0) + D_i(Y_i(1) - Y_i(0))$ for each village i. Since we can never observe $Y_i(1) - Y_i(0)$, i.e., the effect for any village i, the causal parameters of interest are the standard aggregate measures of the effect of *JS*: the average effect, i.e., ATE $= E[Y_i(1) - Y_i(0)]$, and the average effect on the treated, i.e., ATET $= E[Y_i(1) - Y_i(0)|D_i = 1]$, and on the untreated, i.e., ATUT $= E[Y_i(1) - Y_i(0)|D_i = 0]$.

These aggregate causal parameters are not necessarily equal since the *JS* intervention was not a randomized control trial. One might focus on ATE since it can better inform policies for broader implementation of Forum Theater to reduce domestic abuse and shift patriarchal norms. Furthermore, unlike in studies where ATET is of substantive interest since their treatment allocations are based on pretreatment outcomes or the expected future outcomes, the choice of treatment villages in our study was based on *JS*'s idiosyncratic criteria – densely populated, remote, not well-connected villages – and hence it possibly lacks broader appeal. Nevertheless, identification of ATET (ATUT) requires weaker assumptions. Therefore, our discussion considers both ATE and ATET. (We also report ATUT for completeness.)

Alternatively, our reported findings can be interpreted as the effects observed solely within our specific sample (Imbens, 2004). This sample-specific effect remains highly relevant for public policy evaluation. Under this interpretation, our reported standard errors are conservative, i.e., the causal effects are in fact statistically more significant than reported.[7]

3.3 Identification and Estimation

The central empirical challenge is that *JS* did not randomly choose the villages where it performs. They sought to perform in relatively densely populated villages that were geo-

[7]ATE/ATET/ATUT is often denoted by PATE/PATET/PATUT where "P" signifies population, and the sample analog by SATE/SATET/SATUT where "S" signifies sample. We do not distinguish the notation with "P"/"S"; our results can be interpreted either way.

graphically distant from towns, and not well connected to the outside world. These pre-treatment characteristics could be confounders between $Y_i(1), Y_i(0)$ and the intervention D_i. If "all" the confounders are observed, then denoting them by W_i, we can point-identify ATE/ATET/ATUT under overlap if we assume selection on observables (conditional independence/ignorability/unconfoundedness; see, e.g., Imbens (2004), Angrist and Pischke (2009)):

$$Y_i(0) \text{ and } Y_i(1) \text{ are independent of } D_i, \text{ conditional on } W_i \text{ for all } i. \tag{1}$$

That is, conditioning on W_i makes the treatment "as if random". While, in general, it is a strong assumption to maintain that we observe all W_i such that (1) holds, it may be less so in our case since JS informed us of village characteristics that affected their treatment assignment. Nevertheless, finding the correct set of confounders is always difficult in practice; see, e.g., VanderWeele and Shpitser (2011), Cinelli et al. (2024). Assuming that there exists a minimal set of observed variables conditioning on which makes the treatment as if random, we follow VanderWeele and Shpitser (2011)'s "disjunctive cause criterion" whereby we take W_i to be the set of pre-treatment variables that we believe cause either the treatment or the outcome or both.[8] Accordingly, we use data from the Census of India (1991) to constitute our W_i of the following variables (numbered using "W"). W1: (log) population density of the village; W2: (log) distance to the nearest town; W3: access to communication facilities (e.g., post, telegraph and telephone facility); W4: fraction of girls among kids under 6 years; and W5: literacy growth rate between 1981 and 1991. W1-W3 were suggested by JS. W4-W5 cannot also be ruled out as plausible confounders based on comments that we have received.[9]

To streamline the presentation, we focus on the results based on W1-W5, i.e., the more

[8]This strategy is, in principle, similar to Belloni et al. (2014)'s for covariate selection. However, their rationale is different. They want to guard against the pitfalls of post-model-selection inference in high dimensional settings, which is not applicable in our paper that works with low dimensional models except for a subset of supporting results where we select covariates using Lasso from pretreatment variables and their interactions up to third order.

[9]W4 is a measure of the extent to which a society values females. W5 is an indicator of the extent to which a community is committed to increasing its capacity to modernize.

conservative specification for confounders.[10] (Estimated effects based on only W1-W3 are almost the same as those based on W1-W5; see Table 7, Supplemental Appendix B.1.)

Since it is perhaps not controversial under our maintained framework of (1) to move between the potential outcomes and the directed acyclic graphs (DAGs) (see Imbens (2020), Section 3.5), it may help to visually represent using DAGs the causal mechanism in (1) and its potential violation that we have in mind. We posit that the JS plays help to change the unobserved gender norms and that, in turn, reduces abuse of women and empowers them.

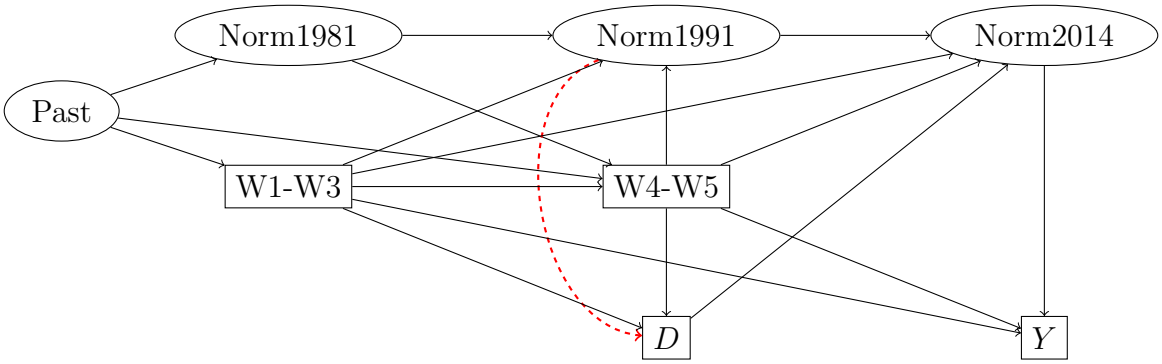


Figure 1: An example of the causal mechanism that is compatible with our framework. Norm1981, Norm1991, Norm2014 are unobserved gender norms in 1981, 1991 and 2014 respectively. Past denotes some unspecified and unobserved past. All unobserved (observed) variables have an ellipse (rectangle) around them. The node W1-W3 represents the three observed covariates that JS informed us about. The node W4-W5 represents the two other observed covariates. Excluding the arrows W1-W3 $\rightarrow Y$ and W4-W5 $\rightarrow Y$, or including an arrow Norm1981 $\rightarrow$ W1-W3 (which seems unlikely), does not alter the discussion. The red-dashed arrow represents a plausible violation of our key identifying assumption (1).

While Figure 1 is not strictly speaking a DAG since it does not specify the arrows, or lack thereof, among the elements inside the W1-W3 and W4-W5 nodes that jointly denote all the observed covariates W, it helps to consider it as a DAG to fix ideas.[11] The selection

[10]These W_i's are pretreatment variables affecting the selection of treatment villages and hence are not "bad" controls in the sense of Angrist and Pischke (2009) (Section 3.2.3). Furthermore, controlling for these pretreatment variables is unlikely to cause the so-called M-bias since it is difficult to imagine a plausible causal path between the treatment and the outcomes via the pretreatment variables (W_i) as "colliders" between two independent unobserved variables on the path (Cinelli et al., 2024) (more on M-bias in Footnote 13).

[11]We do not know enough to assume causal directions among the elements inside these nodes. Those arrows will not change the discussion unless they form a cycle, which is unlikely.

on observables, i.e., the conditional independence assumption in (1), holds if there is no red-dashed arrow from the unobserved gender norm in 1991 to the treatment.[12,13] On the other hand, the red-dashed arrow opens a back-door path between the treatment and the observed outcome: $D \leftarrow \text{Norm1991} \rightarrow \text{Norm2014} \rightarrow Y$. This path cannot be blocked by conditioning on W1-W3 and W4-W5, and this causes a violation of (1). While our conversations with *JS* suggest that the presence of the red-dashed arrow is unlikely, it is prudent to be skeptical.

Thus, the question remains whether conditioning on these pretreatment variables W_i's is sufficient for the independence of D_i with $Y_i(0)$ and $Y_i(1)$ in (1). This is our key assumption. It cannot be tested statistically. Hence we take the following approach in our presentation.

First, we will present in Section 4 our conclusions about the impact of *JS* by point-identifying and estimating the causal parameters of interest under a selection on observables assumption, i.e., the conditional independence condition in (1). There does not seem to be an overlap-failure in our data (Table 1) and, hence, these estimated causal effects should not suffer from extrapolation-bias as long as (1) holds. Moreover, these estimated effects are stable across different choices of commonly used estimators, and they are also precise

[12]Pretending that Figure 1 is a DAG, this can be checked by using the back-door criterion (Pearl, 2000) as follows. First, note that all back-door paths from D to Y (paths with an arrow pointing at D) pass through the node W1-W3 and/or W4-W5. Second, the node W4-W5 is either a "fork" or a "chain" on all the back-door paths from D to Y not through the node W1-W3. Third, W1-W3 is always either a "fork" or a "chain" on any back-door path from D to Y through the node W1-W3. Fourth, although there are two back-door paths from D to Y where the node W4-W5 is a "collider", conditioning on it still does not unblock these two back-door paths since the node W1-W3 lies on these paths as a "fork" and a "chain" respectively and hence conditioning on the node W1-W3 blocks the respective back-door paths again. Therefore, conditioning jointly on W1-W3 and W4-W5 means conditioning on a "fork" and/or "chain" and hence it blocks all the back-door paths from D to Y. Now it follows by the "do-calculus" that conditioning on W1-W3 and W4-W5 gives the causal effect of D on the outcome. This conclusion does not change if we remove the node W4-W5.

[13]After the discussion in Footnote 12, now we can clarify the statement about M-bias from Footnote 10 using Figure 1. Even if we pretend that the unobserved nodes Past and Norm1981 are independent (which is unlikely), the back-door path on which W4-W5 is a collider between these two nodes, i.e., the path $D \leftarrow$ W1-W3 $\leftarrow$ Past $\rightarrow$ W4-W5 $\leftarrow$ Norm1981 $\rightarrow$ Norm1991 $\rightarrow$ Norm2014 $\rightarrow$ Y also contains the node W1-W3 as a "chain". Hence, conditioning on W1-W3 will block this path and prevent M-bias.

— both zero and non-zero effects are estimated precisely. This suggests that problems of limited overlap are not a pressing concern in our analysis (Khan and Tamer, 2010).[14] Thus, our estimated causal effects are reliable provided that our key assumption (1) is true.

Then, in Section 5, we will analyze the sensitivity of our conclusions (from Section 4) to departures from the conditional independence assumption in (1) using the methods proposed by Masten and Poirier (2018, 2020) and Chernozhukov et al. (2024). Such departures signify the insufficiency of the covariates W_i in generating the required conditional independence between the potential outcomes and the treatment in (1) such that the treatment is "as if random". Consider Figure 1 for concreteness and note that everything else fixed, a larger departure signifies greater strength of the red-dashed arrow. The causal results in Section 4 turn out to be robust/insensitive in most cases and require implausibly strong departures from the conditional independence assumption (1) to be overturned (vanish or change sign).

4 Empirical results: Causal effects

An overarching goal of *JS* is to promote open discussion within a community of the oppressive consequences of patriarchal norms and thereby reduce adherence to them and change the beliefs in which they are embedded. We test the hypothesis whether Forum Theatre can shift the focus of attention in spousal violence from the manhood of the assailant to his cruelty to his wife and make domestic violence socially unacceptable. We do this by estimating the impact of exposure to *JS* performances on various indicators of dominance – spousal abuse, legitimacy of domestic violence, participation in community actions to prevent domestic violence and knowledge about legal recourse in case of incidents of domestic violence.

[14]The pvalue of Imai and Ratkovic (2014)'s over-identification test of the null that the covariates are balanced post-adjustment is .95. Also, superimposing the kernel density estimates of the estimated propensity scores in the treatment villages and control villages visually suggests good overlap for estimation of ATET. "Moving the goalpost" by trimming propensity score (Crump et al., 2009) or by focusing on the "overlap population" (Li et al., 2018) has almost no impact on our estimates (Table 8, Supplemental Appendix B.2).

We discuss the regression adjustment (RA) estimates of the ATE and ATET and their respective baselines ($E[Y_i(0)]$ and $E[Y_i(0)|D_i = 1]$) for all these outcomes. These results (and ATUT and more) using RA are reported in Tables 2 and 3. Table 7 in Supplemental Appendix B.1 reports ATE and ATET estimated using alternative methods based on adjustment of all the covariates W1-W5 and the subset W1-W3.[15] The estimates in Table 7 are very stable across methods of estimation and the two nested choices of covariate sets.

4.1 Spousal abuse

Acts of domestic violence as reported by the wives:

The definitions of spousal abuse and the detailed instructions to the field investigators on how to ask the related questions are in Supplemental Appendix E.[16] We estimate the ATE of a village's exposure to *JS* to be a reduction in the proportion of physically abused married women by 9.4 pp (pvalue < 0.01) from the baseline ($E[Y_i(0)]$) of 32.6% (Table 2, col. 1). Exposure to *JS* reduced the proportion of sexually abused married women by close to half — 8.2 pp (pvalue < 0.01) from the baseline ($E[Y_i(0)]$) of 17.7% (Table 2, col. 2). When considering *any* form of spousal abuse — physical, sexual or emotional — exposure to *JS* increased the proportion of abuse-free married women by 15.9 pp (pvalue < 0.01) from the baseline ($E[Y_i(0)]$) of 45.5% (Table 2, col. 3). ATETs for physical abuse and abuse-free relationship are larger and more significant than ATEs. ATET on sexual abuse is smaller than ATE and not significant at the 5% level. This is likely because the baseline ($E[Y_i(0)|D = 1] = 7.6\%$) is already low.

[15] These alternative methods are inverse probability weighting (IPW); combination of IPW and RA viz. IPWRA and augmented IPW (AIPW); matching based on the observed covariates and propensity scores respectively; and estimation using machine learning methods viz. AIPW based on covariate selection using lasso from up to 3rd order interactions among the elements of W_i, and AIPW and targeted maximum likelihood based on prediction using covariates with generalized random forest. We consider $W_i = W1_i - W5_i$ and $W_i = W1_i - W3_i$.

[16] Rough estimates from the NFHS 2015-16 (2017), the round closest to the time of our survey, suggest that 34.3% of married women of age 15-49 experienced physical abuse and 41.7% experienced physical, sexual or emotional abuse in rural South 24 Parganas. These statistics are not too different from those, i.e., 28.7% and 48.7%, in our sample.

***Legitimacy of domestic violence*:**

Field investigators asked women and their husbands whether a husband was justified to hit or beat his wife under any of the following circumstances: she goes out without telling him, neglects the children and household work, argues with him, refuses to have sex with him, disrespects her in-laws, is suspected of having an illicit relationship, and has not borne a son.[17] There is no meaningful effect of *JS* on this outcome using the wives' response (Table 2, col. 4). This is not surprising because the estimated baseline $E[Y_i(0)]$ of only 9% is low. The effect of *JS* is only a decrease of .7 pp. Although not small in relative terms (8% decrease from the baseline, and about 12% of the standard deviation of the observed Y_i), is not meaningful in absolute terms. By contrast, *JS* greatly reduced the proportion of husbands who said that wife-beating was justified in at least one of the aforementioned circumstances. The proportion fell by 16.3 pp (pvalue $< .01$) from the baseline of 21.7% (Table 3, col. 1). The observations for the ATETs are similar to those for the ATEs. Interestingly, here the estimated $E[Y_i(1)]$ for the husbands is lower than that for their wives (5.4% versus 8.3%), although the baseline $E[Y_i(0)]$ is not.

***Awareness that domestic violence is against the law*:**

India prohibited domestic violence for the first time in 2005 through a comprehensive law – the Protection of Women from Domestic Violence Act (PDVA). The new law gives a female victim the right to file a police report and claim monetary compensation for medical treatment and loss of earnings, and also to claim maintenance from her husband for herself and her children. The law bars the husband and in-laws from evicting her from the marital home. However, laws designed to protect women from domestic violence are of no use if women are not aware of them. To raise awareness among women of the PDVA law, many *JS* plays refer to women's rights under the law. The core *JS* team instructs the satellite teams to mention them whenever relevant in the interactive sessions with the audience.

[17]Rough estimates from NFHS 2015-16 (2017) suggest that 7.2% women and 30.3% men in rural South 24 Parganas believe that this is justified. By contrast, 8.1% of the wives and 17.8% of husbands justified wife-beating under these circumstances in our sampled villages.

A village's exposure to *JS* reduced the proportions of wives and husbands who were not aware of the legal protections of the wife against domestic violence under the PDVA by 12.6 and 10.6 pp (pvalue $< .01$) from their respective baselines of 53.7% and 55.9% respectively (Table 2, col. 5 and Table 3, col. 2). The ATETs for wives and husbands are noticeably higher, viz. 16.3 and 14.9 pp respectively. Interestingly, ignorance of the PDVA in villages with and without exposure to *JS* is smaller among wives than husbands.

4.2 The wife's voice in her household

***Empowerment of women*:**

Field investigators asked each female respondent and, separately, her husband if she had not participated with him in making decisions in *any* of the 7 domains: education of the children, family health care, major household purchases, her visits to her relatives, the children's marriages, the number of children to bear, and the use of contraception.[18] Based on the wives' responses, a village's exposure to *JS* reduced this non-participation by almost half — a reduction of 8.4 pp (pvalue $< .01$) from a baseline of 17.5% (Table 2, col. 6). The husbands' responses to the same question about his wife's non-participation in joint decision making is a reduction of only 3 pp (pvalue $< .01$) from the baseline of 5.8% (Table 3, col. 3). The ATETs based on the wife's and the husband's responses are somewhat smaller – 5.2 and 2.7 percentage point reductions (pvalue $< .05$), respectively, – but the impact remains substantial relative to the respective baselines ($E[Y_i(0)|D_i = 1]$) of 16.2 % and 4.8%. The great dissimilarity in the baselines in the wives' and husbands' responses suggests that many husbands do not notice the exclusion of their wives from decision-making. Regardless, the *JS* intervention seems to reduce the "voicelessness" of the wives.

[18]This the final outcome in our discussion for which we could obtain from NFHS 2015-16 (2017) rough estimates for the rural South 24 Parganas. However, these estimates — about 1% for both wives and husbands — differ much from our sample where 15.7% and 4.8% wives and husbands respectively report that wives had not participated in joint decision making.

4.3 Engagement in community activities to alleviate violence against women

***Willingness to report acts of domestic violence witnessed as a bystander*:**

Our survey asked the married women and their husbands if they would report to anyone acts of domestic violence that they witnessed as a bystander. A village's exposure to *JS* made almost universal the willingness of individuals to report such incidents to relatives or institutions such as a police station or the village council. *JS* increased the willingness of women to reporting such incidents by 10.3 pp (pvalue $< .01$) from an already high baseline of 83.5% (Table 2, col. 7). For men, the increase was 5.1 pp (pvalue < 0.01) from an even higher baseline of 92.2% (Table 3, col. 4). The ATETs are larger increases from slightly lower baselines. The near universal willingness of married individuals, when their village is exposed to JS, to report acts of domestic violence witnessed by them supports the hypothesis that *JS* changed village norms: wife beating no longer seemed to villagers to be a normal behavior.

The majority of female respondents preferred reporting an incident of domestic violence that they witnessed to a formal institutions. Less than 45% expressed a willingness to report the incident to a member of either their paternal or maternal family or to other villagers. the husbands of the female respondents, however, showed no any preference for reporting to an individual over an institution.

***Willingness to participate in community drive to destroy illegal liquor shops*:**

Gender violence is aggravated by alcohol and substance abuse (WHO, 2014). According to the NFHS 2019-21 (2023), 85% of women in West Bengal suffer domestic violence if their husband often gets drunk. The social and economic burdens on families of excessive alcohol consumption are a central theme in some *JS* plays. We test whether *JS* encouraged married men and women to participate in community drives to destroy illegal liquor shops. Our survey asked, "If there was an illegal liquor shop in your community, would you participate

in demolishing it?" We estimate the effect on the proportion of married women who are willing to participate in the demolition activities to be an increase by 17.8 pp (pvalue < 0.01) from the baseline of 46.4% (Table 2, col. 8). It increased the proportion of husbands who are willing to participate by 20.5 pp (pvalue $< .01$) from the baseline of 71.2% (Table 3, col. 5). This means that nearly all the husbands are willing to participate in community drives to make their villages free of illegal liquor shops. The ATET based on the husband's response is similar to the ATE. On the other hand, the ATET based on the wife's response is much larger than the ATE, viz. 22.4 pp from a similar baseline ($E[Y_i(0)|D_i = 1]$) of 45.5%.

4.4 Placebo/negative Outcomes

To assess the selection on observables assumption in (1), which underlies our adjustment for JS's non-random treatment allocation, it is useful to consider outcomes that JS should not influence. Ideally, these would be pretreatment values of the main outcomes, but such data are unavailable. (If they were available, we would have included them in W_i in (1)). Instead, we consider a posttreatment household asset — ownership of a color TV — as a placebo/negative outcome, since JS performances are not expected to influence it.

A naive regression of ownership of a color TV on JS exposure yields a coefficient of –.072 (pvalue .13). However, once we adjust for the covariates, as suggested by our selection on observables assumption (1), — in the same way that we did for the main outcomes — the estimates and their significance become much smaller. The ATE estimates based on, e.g., RA and IPW, are -.035 and -.045, with associated pvalues of .49 and .39, respectively. ATET estimates based on RA and IPW are even less significant with pvalues .52 and .89, respectively. Results based on AIPW and machine learning estimators are similar. (The results are also similar for ownership of a refrigerator.) Provided that one believes that JS should not affect ownership of a color TV, it is reassuring that the adjustment for the covariates used in our analysis seems to correct for the difference between the treatment and control villages and thereby avoid spurious inference of labeling a non-effect as an effect.

4.5 Why/how did JS reduce physical abuse?

Our results suggest that about 16 years of exposure to JS performances helped various indicators of women's empowerment, including, perhaps most importantly, the reduction in physical abuse of wives. We conjectured that prolonged exposure to JS performances helped change gender norms in the treated villages and thereby reduced physical abuse of wives. An alternative mechanism is also suggested by our results discussed in Section 4.1, viz. JS raised awareness of the PDVA law and that, in turn, could have reduced physical abuse of wives. That is, knowledge of PDVA is a mediator through which JS could have had an indirect effect on physical abuse. This can be tested statistically under suitable assumptions — e.g., sequential ignorability of treatment assignment and mediator as in Imai et al. (2010) and suitable overlap — using STATA. More precisely, denoting the potential outcome and mediator as composite counterfactuals $Y_i(d, m)$ and $M_i(d)$ for $d = 0, 1$ and $m \in [0, 1]$, and relating it to our notation by noting that $Y_i(d) \equiv Y_i(d, M_i(d))$ for $d = 0, 1$, we can consider the decomposition of ATE (Decomposition 1 in STATA's mediation analysis module) as:

$$\begin{aligned} ATE &= E[Y_i(1) - Y_i(0)] \equiv E[Y_i(1, M_i(1)) - Y_i(0, M_i(0))] \\ &= (E[Y_i(1, M_i(1)) - Y_i(1, M_i(0))]) + (E[Y_i(1, M_i(0)) - Y_i(0, M_i(0))]) \\ &= \text{Average Natural Indirect Effect} + \text{Average Natural Direct Effect}. \end{aligned}$$

We can then estimate the Average Natural Indirect Effect (ANIE) and quantify the mediation. A weakness of our survey that makes such mediation analysis less credible is that our survey does not clarify the temporal ordering of knowledge of PDVA (i.e., mediator M) and physical abuse (i.e. outcome Y). Nevertheless, for the sake of the discussion, let us assume the temporal ordering that knowledge of PDVA preceded physical abuse. Our ATE estimate in Section 4.1 was more than 9% reduction in physical abuse (it is close to -11% if covariates are not interacted with treatment, as in done in STATA's mediation module). Relative to such large ATE estimates, we find our ANIE estimates here to be very negligible. In partic-

ular, considering the four measures of M — proportion of households where both wife and husband did not know about PDVA, where wife did not know but husband knew, where wife knew but the husband did not, and where both wife and husband knew — our estimated ANIEs are, respectively, -.56% (pvalue 81%), .65% (pvalue 40%), .02% (pvalue 87%) and 1.73% (pvalue 33%). In other words, we find little evidence of the alternative mechanism, i.e., of a considerable indirect effect of JS on reduction in physical abuse through the increase in knowledge (awareness/fear) of the PDVA law, to contradict our conjecture that sustained exposure to JS changed the gender norms on physical abuse in the treated villages.

5 Sensitivity analysis of estimated causal effects

The causal effects in Section 4 were identified by our selection on observables assumption in (1) that conditioning on the observed covariates W_i is sufficient to make the treatment D_i independent of the potential outcomes $Y_i(0)$ and $Y_i(1)$. (1) cannot be validated statistically. However, we can analyze the sensitivity of our causal results to violations of (1). We do this following two distinct routes due to Masten and Poirier (2018) and Chernozhukov et al. (2024).[19] We report the strength of unobserved confounders needed to overturn our causal conclusions. It is ultimately up to the readers to decide if they believe that such strength of the unobserved confounder, as deduced from our sensitivity analysis, is likely.

Nevertheless, using standard benchmarking exercises, we find that an unobserved confounder generally needs to be nearly as strong as, or sometimes stronger than, JS's stated

[19]There are several seminal contributions to sensitivity analysis in economics; see, among others, Imbens (2003), Altonji et al. (2005) and Oster (2019). We apply the sensitivity analysis due to Masten and Poirier (2018)) and Chernozhukov et al. (2024) for three reasons. First, as noted by their authors, these methods are valid under weak assumptions. Second, these methods work with bounded sensitivity parameters. Third, like the methods in the aforementioned papers, these methods were also developed by and pitched for economists and, hence, are more familiar to our readers. Various other methods of sensitivity analyses from the biostatistics, economics, epidemiology and statistics literature, when applied in our context, also suggest that the selection on observable assumption in (1) has to be violated to an implausibly large degree to overturn our causal results.

selection criterion to be able to overturn our conclusions. Such extreme requirements for the strength of unobserved confounders suggest that our causal results are robust/insensitive to reasonable violations of our selection on observables condition. It is useful to think of the red-dashed arrow in Figure 1 when considering the strength of the unobserved confounder.

For brevity, we only present the summary of the sensitivity results here. A self-contained description of the sensitivity analysis with precisely-stated assumptions, definitions, results, and a visual representation using various sensitivity plots is presented in Supplemental Appendix B. This appendix could be useful because the details of the analyses in this section based on the relatively new sensitivity-tools may not be widely familiar in applied research.

5.1 Conditional c-dependence and breakdown points

Following Masten and Poirier (2018), D_i is conditionally "c-dependent" on $Y_i(j)$ for $j = 0, 1$ given W_i, if conditioning on $Y_i(j)$ in addition to W_i does not change the probability of receiving the treatment by more than $c \in [0, 1]$. $c = 0$ implies the conditional independence in assumption (1). Larger values of c imply larger violations of (1). Under the assumptions that D_i is conditionally c-dependent on $Y_i(j)$ for $j = 0, 1$ given W_i, Masten and Poirier (2018) provide the upper and lower bounds on ATE and ATET that depend on c, with $c = 0$ signifying point-identification. To fix ideas, consider a case where ATE < 0. Masten and Poirier (2020) define the breakdown point c^* as the smallest c-dependence needed to overturn the result that ATE < 0. To assess whether such a value c^* is reasonable, Masten and Poirier (2020) recommend using the observed covariates to obtain the plausible c-dependence benchmarks that are identifiable from the data. In particular, the k-th benchmark, c_k is taken as the maximum difference between the probability of receiving the treatment conditional on two sets — the set of all covariates and the set without the k-th covariate $W_{i,k}$.

Table 4 presents the breakdown point c^* for each outcome studied in our paper, using the "tesensitivity" package in STATA. It also reports the elements of the observed confounders in W_i, with c_k's larger than c^*. Our causal conclusions will not survive if the inclusion of $Y_i(j)$

in the conditioning set already consisting of W_i should have similar impact on the propensity score as the inclusion of $W_{i,k}$ in the conditioning set already consisting of the other covariates. For most outcomes, including the key outcome of physical abuse, this requires $Y_i(j)$ to have an impact similar to that of the village's population density i in 1991, or its distance from the nearest town in 1991, which were by JS's account key variables in their selection of the treatment villages. In other words, the unobserved confounder needs to be extremely strong to overturn our causal conclusions.

The breakdown point for the ATE on experiencing no abuse is 0.81. Hence, overturning the result that JS increased the proportion of households with no abuse requires $Y_i(j)$ (in 2014) to have, roughly speaking, an impact on the propensity score of the JS treatment that is much larger than that of any observed determinants of the treatment that we consider based on JS's account.

There are four outcomes for which our results are much more vulnerable to violations of the selection on observables condition in (1). These are: (i) sexual abuse, (ii) the wife's belief that beating her is justified, (iii) the wife's belief that she does not participate in household decision-making, and (iv) the husband's belief that his wife does not participate in household decision-making. This fragility was expected a priori (even without any formal sensitivity analysis) because the estimated effects of JS on these outcomes is very small.[20] These are important outcomes by all accounts, and hence the failure of JS to have robust impact on these four outcomes is noteworthy. ATET is less sensitive than ATE in general, but not for the key outcomes of physical abuse and abuse-free relationship.

5.2 Long and short models: Analysis with Riesz representors

Sensitivity analyses proposed by Altonji et al. (2005), Oster (2019), Cinelli and Hazlett (2020), among others, are well-suited for empirical applications that model homogeneous

[20]The baselines for these outcomes were relatively low. Hence there was little room for improvement. All else equal, it is easier to overturn a small effect than a larger effect.

treatment effects, i.e., the so-called partially linear models. Since we allow for heterogeneous effects, i.e., D_i appears by itself and also interacted with W_i in the regression adjustment, we instead follow Chernozhukov et al. (2024)'s approach to what they call the nonparametric model for ATE/ATET to obtain our second set of results on the sensitivity analysis of (1).

Suppose that the selection on observables assumption in (1), i.e., the so-called "short model" where every confounder is observed, is false. Let U_i be the unobserved confounder for village i. We will analyze how badly false the short model has to be for our causal conclusions to be overturned. For concreteness, consider Figure 1 and think of U_i as the gender norms in 1991 (Norm1991), the short model as the DAG, i.e., nonparametric structural model, without the red-dashed arrow, and the so-called "long model" as the one with the red-dashed arrow. ATE (and, in general, ATET) is not point-identified if the short model (1) is false.

Under the assumption of conditional independence in the long model, i.e. $Y_i(0)$ and $Y_i(1)$ are independent of D_i conditional on W_i and U_i, Chernozhukov et al. (2024) obtained upper and lower bounds on the partially identified ATE and ATET as functions of two unidentifiable sensitivity parameters $\eta^2_{(Y_i \sim U_i | D_i, W_i)}$ and $1 - \nu^2_{(D_i \sim U_i | W_i)}$. To fix ideas, focus on ATE. (The case for ATET is similar except that $\nu^2_{(D_i \sim U_i | W_i)}$ is different because ATET has a different Riesz representor.) $\eta^2_{(Y_i \sim U_i | D_i, W_i)} \in [0, 1]$ measures the additional gain in explanatory power that U_i provides for Y_i, beyond what is already explained by D_i and W_i. $1 - \nu^2_{(D_i \sim U_i | W_i)} \in [0, 1]$ measures the relative gain in the average precision of the treatment D_i model due to U_i. $\eta^2_{(Y_i \sim U_i | D_i, W_i)} = 0$ and $1 - \nu^2_{(D_i \sim U_i | W_i)} = 0$ if "U_i is irrelevant" in the sense that the short model (1) is true. Then ATE is point-identified; otherwise it is partially identified $[\tau_L, \tau_U]$.

When $\widehat{\tau} < 0$ (i.e., our estimated ATE < 0), then the sensitivity analysis studies the plausibility of the values of the sensitivity parameters $\left(\eta^2_{(Y_i \sim U_i | D_i, W_i)}, 1 - \nu^2_{(D_i \sim U_i | W_i)}\right)$ that make $\widehat{\tau}_U \geq 0$. In particular, in Table 5, we report the estimated Robustness Value, which is the value $r_0 \in [0, 1]$ such that $\widehat{\tau}_U(r_0, r_0) = 0$ when $\widehat{\tau} < 0$ (alternatively, $\widehat{\tau}_L(r_0, r_0) = 0$ when $\widehat{\tau} > 0$). To account for the estimation error for both τ and the identifiable part of τ_U, we do a similar analysis with the upper 95% confidence bound of τ_U when $\widehat{\tau}_U < 0$.

The counterpart of the Robustness Value in this context is denoted by Robustness Value (.05) and is defined as $r_0 \in [0,1]$ such that the upper 95% confidence bound is 0 when $\eta^2_{(Y_i \sim U_i | D_i, W_i)} = 1 - \nu^2_{(D_i \sim U_i | W_i)} = r_0$.[21] When $\widehat{\tau}_L > 0$, we look instead at the lower 95% confidence bound of τ_L. We also do a similar analysis for ATET.

All these results are reported in Table 5 based on not only the RA estimates as in the rest of the paper, but also the IPW estimates because the Riesz representor, which is fundamental to the construction in Chernozhukov et al. (2024), is directly associated with IPW estimation. The results based on the RA and IPW estimates (and also other estimates) are very similar and convey the same message about the sensitivity of our results to (1) — our causal conclusions survive even under large but equal value of the sensitivity parameters.[22]

An extended and more flexible version of this sensitivity analysis, including contour plots and benchmarks based on observed covariates, similar to Imbens (2003), Cinelli and Hazlett (2020) and Chernozhukov et al. (2024) is presented in Supplemental Appendix B.3.2. We summarize the main observations here.

The ATE/ATET on the four outcomes – (i) sexual abuse, (ii) the wife's belief that beating her is justified, (iii) the wife's belief that she does not participate in household decision-making, and (iv) the husband's belief that his wife does not participate in household decision-making – that were fragile in Section 5.1 continue to be fragile. Additionally, we observe that the RA estimate for ATE and IPW estimate for ATET on the reduction in the proportion of husbands who believe that there is no law against domestic violence can be

[21] While Chernozhukov et al. (2024) provide standard errors based on the semiparametrically efficient estimators, we use nonparametric percentile bootstrap to obtain the confidence bounds. We justify this as follows. We wish to report this sensitivity analysis with the RA and IPW estimators (result for other estimators are also available from us), since the results while similar are not identical. This means that the standard error based on the efficiency bound formula may be less appropriate with our small sample. On the other hand, formulae for standard errors based on parametric treatment of these estimators are tedious and different for these estimators. We use nonparametric bootstrap to bypass all these.

[22] Results under "extreme scenarios" as in Cinelli and Hazlett (2020) (e.g., setting $\eta^2_{(Y_i \sim U_i | D_i, W_i)} = 1$ and studying sensitivity to $1 - \nu^2_{(D_i \sim U_i | W_i)}$), but under Chernozhukov et al. (2024)'s setup, are available from us.

overturned if the unobserved confounder is as influential as the remoteness of the village, the key factor in JS's selection. On the other hand, the other effects are quite robust. The unobserved confounder needs to be more influential than each of the observed covariates to overturn the effect on the decrease in physical abuse, one of the key outcomes; the IPW estimate of ATET is a little more fragile and can be overturned if the unobserved confounder is as influential as the remoteness of the village. These are unlikely given what we learned from JS about their selection criteria. The effect on the other outcomes cannot be overturned by an unobserved confounder that is as strong as the observed benchmarks.

The statistical conclusions on the causal effects and their sensitivity analysis for most of these highly correlated outcomes do not change even with the use of conservative Bonferroni critical values to account for multiple testing by controlling for the family-wise error rate.

6 Conclusion

In 2014, we conducted a household survey – the first of its kind – to obtain outcome data from randomly chosen treatment and control villages in the South 24 Parganas district of West Bengal, India. JS had been performing its community-based participatory theater in the area since 1998. Our estimates based on the survey data show that within less than 16 years of exposure to JS's performances significantly reduced the acceptability and prevalence of domestic violence. The intervention also promoted a sense of collective responsibility among couples, leading to an increased willingness among both husbands and wives to report domestic violence they witnessed and to join efforts to demolish illegal liquor shops.

The results suggest that long-term community engagement with participatory theater shifts entrenched social norms and behaviors surrounding gender-based violence. We obtained these results under the identifying assumption that JS selected the villages where it performed based on observable characteristics, all of which we controlled for in our estimations. Our sensitivity analysis suggests that most of our results are robust except to very

strong unobserved confounders between JS's selection of villages and the villages' potential response to JS's performances. This however is unlikely given JS's description of their village selection methodology. To our knowledge, our results are the first large-scale and rigorous evidence of the impact of community-based participatory theater.

Collectively held beliefs and social norms have powerful effects coordinating individuals. There are many distinct sets of beliefs that can be widely held in a society. Differences may arise as a result of differences in history (e.g., Cohen (2001); Hoff and Stiglitz (2010); Rustagi (2024)), and there are no automatic processes that move a society from a Pareto inefficient outcome to an outcome that makes everyone better off. Community-based participatory theater gives communities a platform for collectively finding ways to reduce oppressive practices and adopting norms that will sustain them. The plays cause many spectators to feel empathy for victims of oppression. The plays give people the ability to imagine less oppressive relationships and explore ways to bring them about. This is not something that can be done by an individual in isolation—for example, through watching an *edutainment* video alone or with anonymous others. It requires a change in attitudes by a sufficiently large proportion of a community and common knowledge of the changes. These theaters create a relatively safe forum in which individuals discuss social norms in the context of a narrative of oppression in which everyone has the same facts, no individual stands accused, and the *Joker* ensures that everyone is free to speak and to act onstage. It is literally a rehearsal of social change.

As Rovelli (2023) points out, not just facts but also the ability to reorganize concepts are necessary to change longstanding beliefs. Narratives help people organize their understanding of relationships and play a crucial role in anchoring social attitudes (Bruner, 1991). Community-based participatory theater presents stories that capture problems that almost everyone in a community has either experienced, seen, or heard about. Through active participation to reach different (more empowering) conclusions to the story of the drama, exposure to these theatres by JS may help individuals reorganize concepts and change their beliefs. It would be interesting to see interventions by participatory theaters in the context

of norms and social meanings that underpin, e.g., the tolerance of corruption, corporal punishment of children, and child marriage. Analysis of networks (and other mediations), dosage (intensity and duration) and dynamics in the context of impact evaluation of participatory theater in some of these cases by means of multi-period randomized control trials focusing on scalability and cost effectiveness of the intervention is the subject of our future research.

7 Bibliography

Akerlof, G. A. (2020). Sins of omission and the practice of economics. *Journal of Economic Literature*, 58:405–418.

Akerlof, R. N. M. and Rayo, L. (2020). Stories at work. *AEA Papers and Proceedings*, 110:199–214.

Altonji, J. G., E., T., and Taber, C. R. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy*, 113: 151–184.

Ang, D. (2023). The birth of a nation: Media and racial hate. *American Economic Review*, 116:1424–1460.

Angrist, J. and Pischke, J. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Asante, E. and Zakaria, B. S. (2021). Participatory theatre in environmental health education: The woarabeba experience. *American Research Journal of Humanities and Social Sciences*, 12:121–128.

Banerjee, A., Barnhardt, S., and Duflo, E. (2018). Can iron-fortified salt control anemia? evidence from two experiments in rural bihar. *Journal of Development Economics*, 133:127–146.

Banerjee, A., La-Ferrara, E., and Orozco, V. (2019a). The entertaining way to behavioral change: Fighting hiv with mtv. NBER Working Paper No. 26096.

Banerjee, A., La-Ferrara, E., and Orozco, V. (2019b). Entertainment, education, and attitudes toward domestic violence. *American Economic Review*, 109:133–137.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies*, 81:608–650.

Benabou, R., Falk, A., and Tirole, J. (2018). Narratives, imperatives and moral persuasion. NBER Working Paper No. 24798.

Bicchieri, C. (2017). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press.

Blair, G., Littman, R., and Paluck, E. L. (2019). Motivating the adoption of new community-minded behaviors: An empirical test in nigeria. *Science Advances*, 5:eaau5175.

Bourdieu, P. (1977). *Outline of a Theory of Practice*. Cambridge University Press, New York.

Bruner, J. (1991). The Narrative Construction of Reality. *Critical Inquiry*, 18: 1–21.

Census of India (1991). Primary census abstracts.

Chernozhukov, C., Cinelli, C., Newey, W., Sharma, A., and Syrgkanis, V. (2024). Long Story Short: Omitted Variable Bias in Causal Machine Learning. Working Paper.

Chowdhry, P. (2015). Popular perceptions of masculinity in rural north indian oral traditions. *Asian Ethnography*, 74:5–36.

Cinelli, C., Forney, A., and Pearl, J. (2024). A Crash Course in Good and Bad Controls. *Sociological Methods and Research*, 53: 1071–1104.

Cinelli, C. and Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82:39–67.

Cohen, D. (2001). Cultural variation: Considerations and implications. *Psychological Bulletin*, 127: 451–471.

Cooper, J., Green, D., and Wilke, A. (2020). Reducing Violence against Women in Uganda through Video Dramas: A Survey Experiment to Illuminate Causal Mechanisms. *American Economic Review*, 110:615–619.

Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96: 187–199.

Derne, S. (1994). Hindu men talk about controlling women: Cultural ideas as a tool of the powerful. *Sociological Perspectives*, 37:203–227.

Duvvury, N., Callan, A., Carney, P., and Raghavendra, S. (2013). Intimate Partner Violence: Economic Costs and Implications for Growth and Development. Technical report, World Bank, Washington DC.

Egeland, B. (1993). A history of abuse is a major risk factor for abusing in the next generation. In Gelles, R. and Loseke, D., editors, *Current controversies on family violence*, pages 197–208. Sage Publications, Newbury Park, CA.

Erten, B. and Keskin, P. (2018). For Better or for Worse?: Education and the Prevalence of Domestic Violence in Turkey. *American Economic Journal: Applied Economics*, 10: 64–105.

Esposito, E., Rotesi, T., Saia, A., and Thoenig, M. (2023). Reconciliation Narratives: The Birth of a Nation after the US Civil War. *Ameican Economic Rebiew*, 113:1461–1504.

Ganguly, S. (2010). *Jana Sanskriti: Forum Theatre and Democracy in India*. Routledge, Oxon.

Haushofer, J. and Shapiro, J. (2016). The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya. *The Quarterly Journal of Economics*, 131:1973–2042.

Hoff, K. and Stiglitz, J. (2010). Equilibrium fictions: A cognitive approach to societal rigidity. *American Economic Review*, 100:141–146.

Htun, M. and Jensenius, F. R. (2020). Violence against women: Laws, norms, and challenges ahead. *Daedalus*, 149:144–159.

Imai, K., Keele, L., and Yamamoto, T. (2010). Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Statistical Science*, 25: 51–71.

Imai, K. and Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society, Series B*, 76:243–263.

Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review*, 93: 126–132.

Imbens, G. W. (2004). Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. *Review of Economics and Statistics*, 86:4–29.

Imbens, G. W. (2020). Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics. *Journal of Economic Literature*, 58: 1129–1179.

Jha, S. and Shayo, M. (2019). Valuing peace: The effects of financial market exposure on votes and political attitudes. *Econometrica*, 87:1561–1588.

Kahan, D. M., Peters, E., Dawson, E. C., and Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioral Public Policy*, 1:54–86.

Khan, S. and Tamer, E. (2010). Irregular Identification, Support Conditions, and Inverse Weight Estimation. *Econometrica*, 78:2021–2042.

Koenen, C. K., Moffitt, T. E., Caspi, A., Taylor, A., and Purcell, S. (2003). Domestic violence is associated with environmental suppression of IQ in young children. *Development and psychopathology*, 15: 297–311.

La Ferrara, E., Chong, A., and Duryea, S. (2012). Soap operas and fertility: Evidence from brazil. *American Economic Journal: Applied Economics*, 4:1–31.

Li, F., Morgan, K. L., and Zaslavsky, A. M. (2018). Balancing Covariates via Propensity Score Weighting. *Journal of the American Statistical Association*, 113: 390–400.

Masten, M. A. and Poirier, A. (2018). Identification of treatment effects under conditional partial independence. *Econometrica*, 86:317–351.

Masten, M. A. and Poirier, A. (2020). Inference on breakdown frontiers. *Quantitative Economics*, 11:41–111.

Masten, M. A., Poirier, A., and Zhang, L. (2024). Assessing Sensitivity to Unconfoundedness: Estimation and Inference. *Journal of Business and Economic Statistics*, 42: 1–13.

Mills, S. (2009). Theatre for transformation and empowerment: a case study of jana sanskriti theatre of the oppressed. *Development in Practice*, 19:550–559.

Mohan, D. (2010). *Scripting Power: Jana Sanskriti, On and Offstage*. Camp, Kolkata.

Narayan, D., Patel, R., Schafft, K., Rademacher, A., and Koch-Schulte, S. (2000). *Voices of the poor. Can anyone hear us?* Oxford University Press, Oxford.

NFHS 2015-16 (2017). National family health survey 2015-16.

NFHS 2019-21 (2023). National family health survey 2019-21.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37: 187–204.

Pearl, J. (2000). *Causality*. Cambridge University Press, Cambridge.

Riley, E. (2024). Role Models in Movies: The Impact of Queen of Katwe on Students' Educational Attainment. *Review of Economics and Statistics*, 106: 334–351.

Rovelli, C. (2023). *While Holes*. Riverhead Books.

Roy, S., Hidrobo, M., Hoddinott, J., , and Ahmed, A. (2019). Transfers, Behavior Change Communication, and Intimate Partner Violence. *The Review of Economics and Statistics*, 101: 865–877.

Rustagi, D. (2024). Historical Self-Governance and Norms of Cooperation. *Econometrica*, 92: 1473–1502.

Shiller, R. (2019). *Narrative Economics*. Princeton University Press, Princeton, NJ.

Thambu, N., Shanmugavelu, G., and Mahayudin, Z. (2020). The use of forum theatre as creative pedagogy in teaching and learning of moral education in secondary schools in malaysia. *EPRA International Journal of Multidisciplinary Research*, 6:2455–3662.

VanderWeele, T. J. and Shpitser, I. (2011). A new criterion for confounder selection. *Biometrics*, 67: 1406–1413.

WHO (2014). *Global Status Report on Alcohol and Health, 2014*. WHO Press, Geneva.

WHO (2025). *Violence against women prevalence estimates, 2023: global, regional and national prevalence estimates for intimate partner violence against women and non-partner sexual violence against women*.

Yarrow, R. (2012). Performing agency. *Studies in South Asian Film and Media*, 4:211–226.

Yarrow, R. (2017). From performers to spectactivists: Jana sanskriti's training for agency in and beyond theatre. *Indian Theatre Journal*, 1:29–37.

Yarrow, R. (2021). *Jana Sanskriti*. Routledge, London.

Table 1: **Descriptive statistics**

	Control Villages (sample size is 61)					*Treatment Villages (sample size is 31)*				
Variables	**Mean**	**Med**	**Std**	**Min**	**Max**	**Mean**	**Med**	**Std**	**Min**	**Max**
Village characteristics from 1991 (pretreatment) that we control for										
Log Population density/ha	2.425	2.406	.544	1.208	4.023	1.782	1.740	.443	.706	3.117
Log Distance (km) to nearest town	2.817	2.833	.691	1.386	3.912	3.782	4.094	.520	2.485	4.605
Access to communication facilities	.230	0	.424	0	1	.097	0	.301	0	1
Fraction of girls among kids 6 & under	.498	.490	.032	.406	.593	.491	.486	.031	.429	.574
Literacy growth rate (1981-1991)	.071	.066	.071	-0.160	.267	.084	.083	.090	-0.152	.285
Wife's responses: Outcomes										
Whether physically abused	.314	.294	.124	.062	.586	.236	.237	.079	.080	.458
Whether sexually abused	.185	.176	.087	0	.360	.135	.120	.071	.031	.306
Did not face any abuse	.480	.485	.121	.161	.742	.580	.594	.099	.367	.781
Physical abuse of wives is justified	.093	.086	.065	0	.296	.058	.057	.042	0	.156
Don't know of domestic violence laws	.534	.542	.116	.308	.840	.380	.379	.062	.288	.533
No joint decisions	.181	.185	.084	0	.417	.110	.089	.091	0	.500
Would report acts of domestic violence	.854	.871	.091	.581	1	.956	.960	.032	.879	1
Would demolish illegal liquor shops	.469	.462	.139	.083	.800	.679	.688	.115	.233	.837
Husband's responses: Outcomes										
Physical abuse of wives is justified	.230	.226	.120	.034	.481	.075	.052	.074	0	.268
Don't know of domestic violence laws	.555	.559	.113	.250	.913	.418	.407	.149	.172	.867
No joint decisions	.062	.048	.055	0	.200	.021	.016	.027	0	.111
Would report acts of domestic violence	.929	.938	.061	.760	1	.974	.978	.034	.867	1
Would demolish illegal liquor shops	.723	.720	.140	.250	.944	.908	.920	.065	.667	1

Source: Census of India, 1991 for the observed covariates, i.e., the pretreatment village characteristics that we control for (top panel), and Primary survey, 2014-2016 for the outcomes of interest (bottom panel).

Table 2: Regression Adjustment (RA) results based on Wife's responses

ESTIMAND	Physical abuse (1)	Sexual abuse (2)	Abuse-free relation-ship (3)	Physical abuse by husband is justified (4)	No knowledge of PDVA (5)	No joint decisions (6)	Report acts of domestic violence (7)	Demolish illegal liquor shops (8)
ATE E[Y(1) – Y(0)]	-.094 (0.03)	-.082 (0.02)	.159 (0.03)	-.007 (0.01)	-.126 (0.02)	-.084 (0.03)	.103 (0.02)	.178 (0.05)
ATT E[Y(1) – Y(0)|D=1]	-.115 (0.03)	-.026 (0.02)	.172 (0.03)	-.025 (0.02)	-.163 (0.03)	-.052 (0.02)	.158 (0.03)	.224 (0.04)
ATUT E[Y(1) – Y(0)|D=0]	-.083 (0.03)	-.110 (0.03)	.153 (0.04)	.003 (0.01)	-.108 (0.03)	-.100 (0.03)	.076 (0.02)	.154 (0.06)
Naïve regression E[Y(1)|D=1] - E[Y(0)|D=0]	-.078 (0.02)	-.050 (0.02)	.100 (0.02)	-.035 (0.01)	-.155 (0.02)	-.071 (0.02)	.102 (0.01)	.210 (0.03)
E[Y(0)]	.326 (0.02)	.177 (0.01)	.455 (0.02)	.090 (0.01)	.537 (0.02)	.175 (0.01)	.835 (0.01)	.464 (0.02)
E[Y(1)]	.232 (0.02)	.096 (0.02)	.615 (0.03)	.083 (0.01)	.411 (0.02)	.091 (0.02)	.939 (0.01)	.642 (0.04)
E[Y(0)|D=0]	.314 (0.02)	.185 (0.01)	.480 (0.02)	.093 (0.01)	.534 (0.01)	.181 (0.01)	.854 (0.01)	.469 (0.02)
E[Y(0)|D=1]	.351 (0.03)	.161 (0.02)	.408 (0.03)	.083 (0.01)	.542 (0.03)	.162 (0.02)	.799 (0.02)	.455 (0.03)
E[Y(1)|D=0]	.230 (0.02)	.076 (0.02)	.632 (0.04)	.096 (0.01)	.426 (0.03)	.081 (0.03)	.930 (0.01)	.623 (0.05)
E[Y(1)|D=1]	.236 (0.01)	.135 (0.01)	.580 (0.02)	.058 (0.01)	.380 (0.01)	.110 (0.02)	.956 (0.01)	.679 (0.02)

Standard errors in parentheses. No cell color indicates pvalue $\leq$ 1%, indicates pvalue between 1% and 5% and indicates pvalue $>$ 5%.

Table 3: Regression Adjustment (RA) results based on Husband's responses

ESTIMAND	Physical abuse by husband is justified (1)	No knowledge of PDVA (2)	No joint decisions (3)	Report acts of domestic violence (4)	Demolish illegal liquor shops (5)
ATE E[Y(1) – Y(0)]	-.163 (0.02)	-.106 (0.03)	-.030 (0.01)	.051 (0.01)	.205 (0.03)
ATT E[Y(1) – Y(0)|D=1]	-.117 (0.03)	-.149 (0.04)	-.027 (0.01)	.065 (0.01)	.217 (0.03)
ATUT E[Y(1) – Y(0)|D=0]	-.186 (0.02)	-.084 (0.05)	-.031 (0.01)	.044 (0.01)	.199 (0.03)
Naïve regression E[Y(1)|D=1] - E[Y(0)|D=0]	-.155 (0.02)	-.137 (0.03)	-.041 (0.01)	.045 (0.01)	.185 (0.02)
E[Y(0)]	.217 (0.02)	.559 (0.02)	.057 (0.01)	.922 (0.01)	.712 (0.02)
E[Y(1)]	.054 (0.01)	.453 (0.03)	.027 (0.01)	.973 (0.01)	.918 (0.02)
E[Y(0)|D=0]	.230 (0.02)	.555 (0.01)	.062 (0.01)	.929 (0.01)	.723 (0.02)
E[Y(0)|D=1]	.192 (0.03)	.567 (0.03)	.048 (0.01)	.909 (0.01)	.691 (0.03)
E[Y(1)|D=0]	.043 (0.02)	.471 (0.05)	.031 (0.01)	.973 (0.01)	.922 (0.03)
E[Y(1)|D=1]	.075 (0.01)	.418 (0.03)	.021 (0.00)	.974 (0.01)	.908 (0.01)

Standard errors in parentheses. No cell color indicates pvalue $\leq 1\%$, indicates pvalue between 1% and 5% and indicates pvalue $> 5\%$.

Table 4: Masten-Poirier's breakdown point analysis

	ATE		ATT	
Outcomes	**Breakdown point (c^*)**	**Confounders with $c_k \leq c^*$**	**Breakdown point (c^*)**	**Confounders with $c_k \leq c^*$**
Wife's responses				
Physical abuse	0.37	W3, W4, W5	0.18	W3, W4, W5
Sexual abuse	0.06	W3, W4, W5	0.11	W3, W4, W5
Abuse-free relationship	1.00	All of them	0.22	W3, W4, W5
Physical abuse by husband is justified	0.03	W3, W4	0.07	W3, W4, W5
No knowledge of PDVA	0.24	W3, W4, W5	0.20	W3, W4, W5
No joint decisions	0.02	W3, W4	0.19	W3, W4, W5
Report acts of domestic violence	1.00	All of them	0.53	All of them
Demolish illegal liquor shops	0.24	W3, W4, W5	0.32	W3, W4, W5
Husband's responses				
Physical abuse by husband is justified	0.24	W3, W4, W5	0.28	W3, W4, W5
No knowledge of PDVA	0.14	W3, W4, W5	0.35	W3, W4, W5
No joint decisions	0.08	W3, W4, W5	0.17	W3, W4, W5
Report acts of domestic violence	0.40	W3, W4, W5	0.25	W3, W4, W5
Demolish illegal liquor shops	0.29	W3, W4, W5	0.63	All of them

(c-)dependence of D_i on $Y_i(1)$ (and $Y_i(0)$) as "strong" as the observed confounders' c_k's will overturn causal effects. The labels for the observed confounders [and their c_k's] are:

- W1: (log) population density of the village [.419]
- W2: (log) distance to the nearest town [.440]
- W3: access to communication (post, telegraph and telephone facility) facilities [.012]
- W4: fraction of girls among kids under 6 years [.016]
- W5: literacy growth rate between 1981 and 1991 [.036]

Table 5: Chernozhukov et al (2024)'s Robustness Values and Robustness Values (.05) of the RA and IPW short estimates for ATE and ATET

Outcomes	RA (ATE)		RA (ATET)		IPW (ATE)		IPW (ATET)	
	Robustness Value	Robustness Value (0.05)	Robustness Value	Robustness Value (0.05)	Robustness Value	Robustness Value (0.05)	Robustness Value	Robustness Value (0.05)
Wife's responses								
Physical abuse	.30	.15	.35	.18	.27	.14	.34	.15
Sexual abuse	.31	.11	.12	.01	.26	.16	.10	.01
Abuse-free relationship	.44	.21	.47	.32	.38	.24	.47	.35
Physical abuse by husband is justified	.04	.01	.15	.01	.11	.01	.17	.06
No knowledge of PDVA	.37	.22	.45	.31	.42	.28	.44	.32
No joint decisions	.31	.06	.21	.05	.29	.17	.17	.01
Report acts of domestic violence	.42	.29	.56	.43	.47	.34	.60	.49
Demolish illegal liquor shops	.41	.17	.48	.32	.43	.27	.43	.29
Husband's responses								
Physical abuse by husband is justified	.45	.30	.35	.19	.42	.29	.24	.01
No knowledge of PDVA	.29	.11	.38	.22	.36	.25	.34	.17
No joint decisions	.22	.08	.20	.06	.26	.14	.25	.12
Report acts of domestic violence	.32	.17	.38	.23	.31	.19	.39	.19
Demolish illegal liquor shops	.47	.32	.49	.36	.44	.32	.40	.26

Supplemental Appendix:

Participatory Theater Empowers Women:

Evidence from West Bengal, India

S. Chaudhuri[23] K. Hoff[24] J. Jalan[25] S. Santra[26]

Table of Contents

[23]Department of Economics, McGill University & Cireq. (saraswata.chaudhuri@mcgill.ca)
[24]Department of Economics, Columbia University. (kh2830@columbia.edu)
[25]Centre for Studies in Social Sciences. (jjalan@cssscal.org)
[26]Centre for Training and Research in Public Finance and Policy. (sattwik@cssscal.org)

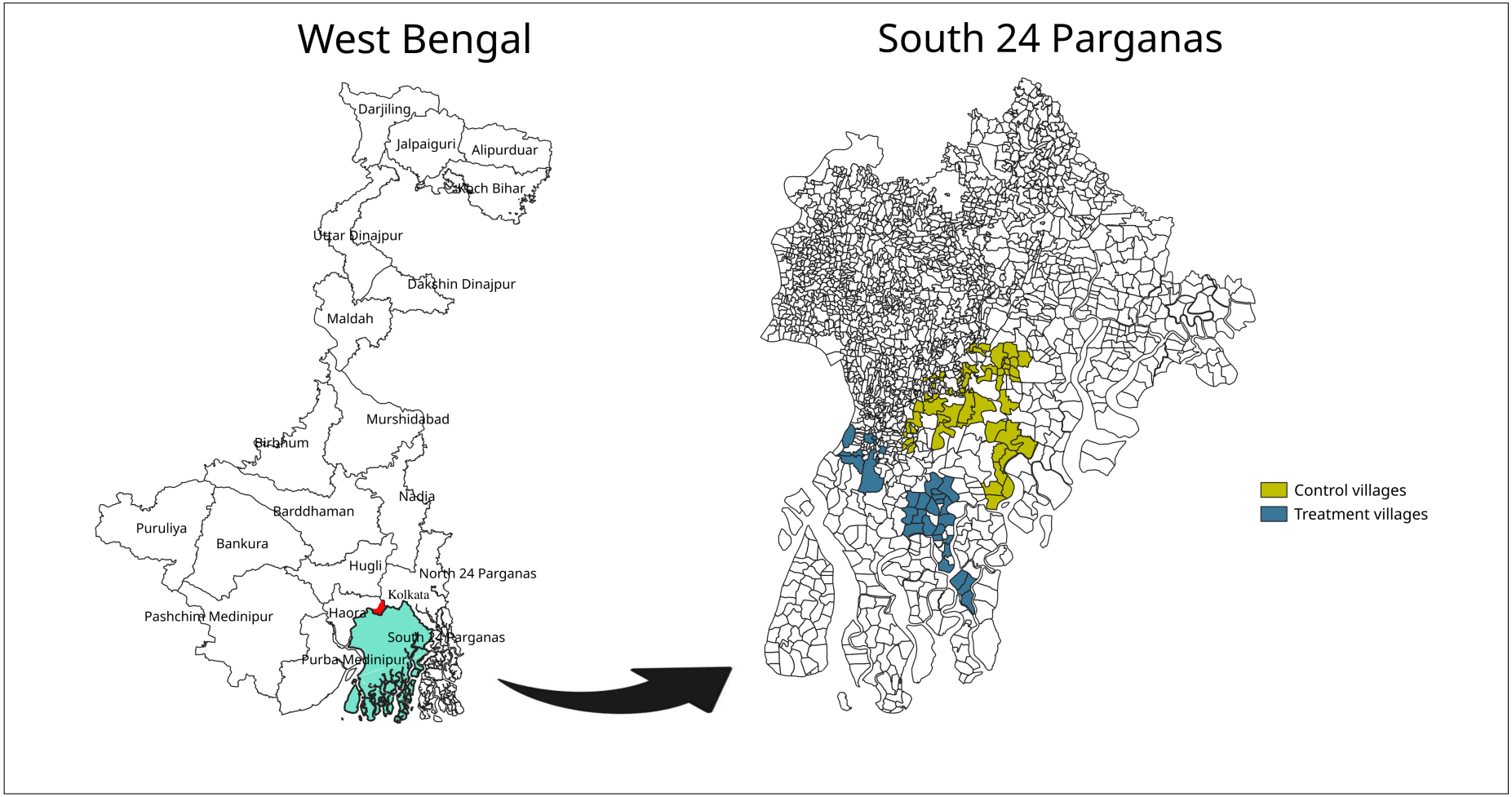
A Background Tables and Figures

Table 6: Cross-country statistics on women justifying domestic violence and experiencing intimate partner violence, 2017

	COUNTRY	Percentage of women who consider that a husband is justified in hitting/beating his wife for at least one specified reason	Percentage of women who ever suffered violence from an intimate partner
(A) LOW-INCOME COUNTRIES	Guinea	92.1	80
	Afghanistan	80.2	60.8
	South Sudan	78.5	NA
	Somalia	75.7	NA
	Congo (Dem Rep)	74.8	50.7
	Chad	73.5	28.6
	Burundi	72.9	46.7
	Mali	72.6	34.6
	Ethiopia	63	28
	Niger	59.6	NA
	Tajikistan	59.6	20.3
	Haiti	58.9	20.8
	Uganda	58.3	49.9
	Tanzania	58	41.7
	Senegal	56.5	78
	Yemen	48.7	67
	Madagascar	45.2	30
	Burkina Faso	43.5	11.5
	Nepal	42.9	25
	Rwanda	41.4	34.4
	Zimbabwe	38.7	35.4
	Benin	36	68.6
	Togo	28.7	22.1
	Mozambique	22.9	21.7
	Malawi	16.3	37.5
	MEDIAN	58.3	35
(B) LOWER MIDDLE-INCOME COUNTRIES	Sri Lanka	53.2	16.6
	Myanmar	51.2	33
	Cambodia	50.4	20.9
	Côte d'Ivoire	47.9	25.9
	Zambia	46.9	42.7
	Pakistan	42.2	85
	Kenya	41.8	39.4
	Uzbekistan	41.5	NA
	Cameroon	36.1	51.1
	Egypt	35.7	NA
	Nigeria	34.7	16.2
	Indonesia	34.5	18.3
	Sudan	34	NA
	Bangladesh	28.3	53.3
	Ghana	28.3	24.4
	Viet Nam	28.2	34.4
	Angola	25.2	34.8
	India	22.1	28.7
	Morocco	22	30
	Tunisia	18.6	20.3
	Bolivia	16.1	64.1
	Philippines	12.9	16.9
	Honduras	12.4	21.6
	Ukraine	2.9	13.2
	MEDIAN	34.25	28.7
(C) UPPER MIDDLE-INCOME COUNTRIES	South Africa	61.2	20.6
	Iraq	54.8	21.2
	Guinea	52.6	56.9
	Algeria	48.2	NA
	Malaysia	41.5	NA
	China	32.7	NA
	Peru	32.2	33.2
	Azerbaijan	28	13.5
	Ecuador	25.2	37.5
	Russia	23.3	19.6
	Iran	21	66
	Jordan	18	23.6
	Kazakhstan	14.2	16.5
	Turkey	13.3	38
	Colombia	11.1	37.4
	Guatemala	11	18
	Thailand	8.6	44.2
	Brazil	8.5	33.5
	Romania	7.5	24
	Mexico	5	14.1
	Belarus	4.1	25
	Cuba	3.9	NA
	Serbia	3.8	23.7
	Dominican Republic	2	20.4
	MEDIAN	16.1	23.85
(D) HIGH-INCOME COUNTRIES	Kuwait	37	NA
	Chinese Taipei	21.6	NA
	Germany	19.6	22
	Korea	18.4	16.5
	Switzerland	15.2	9.8
	Argentina	11.6	NA
	United States	11	35.6
	Chile	10.3	6.7
	United Kingdom	10.2	29
	Sweden	10.2	28
	Spain	9.6	13
	Japan	8.9	15.4
	Hungary	8.7	21
	Poland	7.9	13
	Canada	7.8	1.9
	France	6.6	26
	Netherlands	6.4	25
	Italy	5.3	19
	Australia	3.2	16.9
	Austria	3	13
	Belgium	2	24
	Czech Republic	2	21
	Greece	2	19
	Portugal	2	19
	MEDIAN	8.8	19

Notes. The respondents are nationally representative samples of women 15-49 years of age. The five reasons that are specified for justification of hitting/beating are: the wife burns the food, she argues with her husband, she goes out without telling him, she neglects the children, and she refuses sexual relations with him. Intimate partner violence signifies physical and or sexual violence. Omitted from the table are countries with populations less than 0.1% of the world's total population and countries for which data are not available. *Source.* OECD Gender, Institutions, and Development Database (GID-DB), 2019. https://stats.oecd.org/Index.aspx?DataSetCode=GIDDB2019

Figure 2: **Administrative areas of the survey:** The highlighted area in the map on the left is the district of South 24 Parganas. The highlighted areas in the map on the right are the samples of treatment villages and control villages.



B Robustness Checks

B.1 Various methods and nested covariate sets

Table 7: ATEs (top panel) and ATETs (bottom panel) estimated by various methods using different covariates

	RA		IPW		AIPW		IPWRA		PSMATCH		NNMATCH		LASSO		RFAIPW		RFTMLE	
Number of confounders	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5
Wife's responses																		
Physical abuse	-.10	-.09	-.09	-.09	-.10	-.10	-.10	-.09	-.10	-.10	-.10	-.09	-.11	-.11	-.10	-.10	-.08	-.09
Sexual abuse	-.08	-.08	-.06	-.06	-.07	-.08	-.07	-.07	-.06	-.06	-.05	-.05	-.08	-.08	-.06	-.06	-.06	-.06
Abuse-free relationship	.16	.16	.13	.13	.16	.15	.14	.14	.15	.16	.17	.13	.16	.16	.14	.15	.13	.13
Physical abuse by husband is justified	-.00	-.01	-.02	-.02	-.01	-.01	-.01	-.01	-.01	-.02	-.00	-.01	-.01	-.01	-.02	-.02	-.02	-.02
No knowledge about PDVA	-.13	-.13	-.15	-.15	-.13	-.13	-.14	-.13	-.12	-.12	-.13	-.13	-.13	-.13	-.16	-.16	-.16	-.15
No joint decisions	-.08	-.08	-.08	-.08	-.08	-.08	-.08	-.08	-.06	-.05	-.09	-.08	-.08	-.08	-.08	-.08	-.08	-.08
Report acts of domestic violence	.10	.10	.12	.12	.12	.11	.12	.12	.14	.14	.13	.12	.12	.12	.12	.12	.12	.12
Demolish illegal liquor shops	.17	.18	.19	.19	.18	.17	.18	.18	.18	.19	.20	.20	.18	.18	.19	.19	.19	.19
Husband's responses																		
Physical abuse by husband is justified	-.16	-.16	-.15	-.15	-.15	-.15	-.15	-.15	-.13	-.13	-.13	-.15	-.16	-.16	-.15	-.15	-.15	-.16
No knowledge about PDVA	-.10	-.11	-.14	-.14	-.11	-.11	-.13	-.12	-.14	-.14	-.12	-.13	-.12	-.12	-.12	-.12	-.12	-.12
No joint decisions	-.03	-.03	-.04	-.04	-.03	-.03	-.03	-.03	-.04	-.03	-.04	-.02	-.04	-.04	-.03	-.03	-.03	-.04
Report acts of domestic violence	.05	.05	.05	.05	.05	.05	.05	.05	.05	.04	.07	.05	.05	.05	.05	.05	.04	.05
Demolish illegal liquor shops	.20	.21	.18	.18	.19	.19	.19	.18	.17	.18	.18	.19	.20	.20	.20	.19	.19	.19
Wife's responses																		
Physical abuse	-.13	-.12	-.11	-.11	-.13	-.11	-.11	-.11	-.07	-.08	-.17	-.11	-.12	-.12	-.12	-.12	-.14	-.14
Sexual abuse	-.02	-.03	-.02	-.02	-.02	-.02	-.02	-.01	-.00	-.00	.04	-.01	-.03	-.03	-.04	-.04	-.02	-.02
Abuse-free relationship	.18	.17	.18	.18	.19	.18	.21	.20	.17	.19	.29	.20	.22	.23	.17	.18	.20	.20
Physical abuse by husband is justified	-.03	-.02	-.03	-.03	-.03	-.02	-.03	-.02	-.03	-.04	-.02	-.02	-.03	-.03	-.03	-.03	-.04	-.03
No knowledge about PDVA	-.16	-.16	-.16	-.16	-.16	-.16	-.16	-.15	-.14	-.15	-.13	-.11	-.17	-.17	-.17	-.17	-.20	-.19
No joint decisions	-.05	-.05	-.04	-.04	-.04	-.04	-.03	-.03	-.04	-.01	-.04	-.03	-.05	-.05	-.06	-.06	-.03	-.05
Report acts of domestic violence	.16	.16	.18	.17	.18	.18	.21	.21	.23	.23	.20	.17	.20	.17	.15	.16	.21	.21
Demolish illegal liquor shops	.22	.22	.19	.19	.20	.20	.18	.19	.20	.20	.25	.14	.19	.18	.21	.21	.22	.19
Husband's responses																		
Physical abuse by husband is justified	-.12	-.12	-.07	-.07	-.09	-.08	-.05	-.03	-.03	-.03	.01	-.05	-.11	-.11	-.11	-.12	-.06	-.07
No knowledge about PDVA	-.14	-.15	-.13	-.13	-.13	-.14	-.13	-.12	-.12	-.12	-.13	-.10	-.14	-.15	-.13	-.13	-.13	-.14
No joint decisions	-.03	-.03	-.04	-.04	-.03	-.03	-.03	-.03	-.03	-.03	-.05	-.02	-.04	-.04	-.03	-.04	-.03	-.03
Report acts of domestic violence	.07	.06	.06	.06	.07	.07	.06	.07	.05	.04	.13	.07	.07	.06	.06	.06	.07	.07
Demolish illegal liquor shops	.21	.22	.16	.16	.17	.18	.14	.14	.11	.13	.18	.14	.17	.17	.19	.19	.17	.16

RA: regression adjustment, IPW: inverse probability weighting, IPWRA and AIPW: combination of RA and IPW. PSMATCH and NNMATCH: matching based on, respectively, propensity score and the covariates. LASSO: AIPW where the covariates ultimately adjusted for are selected based on LASSO from up to 3rd order interactions of the covariate sets (W1-W3 and W1-W5 respectively); it uses 10 fold cross-fitting and 15 resamples. STATA commands teffects and telasso were used for these estimators. RFAIPW and RFTML: AIPW and targeted maximum likelihood where the predictions are based on the generalized random forest, and they use the default option in the R package grf (https://grf-labs.github.io/grf/REFERENCE.html).

B.2 Overlap robustness: ATE, ATET or Moving the goalpost

Table 8: **ATE, ATET, effects on trimmed % overlap populations** (R-package: PSweight)

	ATE			ATT			ATO
Estimand	RA	IPW	IPW (10%)	RA	IPW	IPW (10%)	
Wife's responses							
Physical abuse	-.09	-.09	-.08	-.12	-.11	-.10	-.11
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Sexual abuse	-.08	-.06	-.04	-.03	-.02	-.02	-.05
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Abuse-free relationship	.16	.13	.17	.17	.18	.18	.17
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Physical abuse by husband is justified	-.01	-.02	-.02	-.02	-.03	-.03	-.01
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
No knowledge about PDVA	-.13	-.15	-.16	-.16	-.16	-.16	-.15
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
No joint decisions	-.08	-.08	-.06	-.05	-.04	-.04	-.06
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Report acts of domestic violence	.10	.12	.15	.16	.17	.18	.14
	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)	(0.02)
Demolish illegal liquor shops	.18	.19	.20	.22	.19	.18	.18
	(0.05)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)
Husband's responses							
Physical abuse by husband is justified	-.16	-.15	-.11	-.12	-.07	-.05	-.13
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
No knowledge about PDVA	-.11	-.14	-.12	-.15	-.13	-.10	-.12
	(0.03)	(0.02)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
No joint decisions	-.03	-.04	-.03	-.03	-.04	-.03	-.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Report acts of domestic violence	.05	.05	.05	.06	.06	.06	.06
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)
Demolish illegal liquor shops	.21	.18	.18	.22	.16	.16	.19
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)

Standard errors in parentheses. No cell color indicates pvalue $\leq 1\%$, indicates pvalue

IPW (10%) trims propensity score to be $[.1, .9]$ for ATE and $[.1, 1]$ for ATET estimation (Crump et al., 2009). ATO stands for average treatment effect in the overlap population (Li et al., 2018).

- Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. Biometrika, 96: 187–199.
- Li, F., Morgan, K. L., and Zaslavsky, A. M. (2018). Balancing Covariates via Propensity Score Weighting. Journal of the American Statistical Association, 113: 390–400.

B.3 Extended version of the sensitivity analyses to (1)

B.3.1 Conditional c-dependence and breakdown points

Following Masten and Poirier (2018), D_i is conditionally "c-dependent" on $Y_i(j)$ for $j = 0, 1$ given W_i, provided that for all w and y in the support of W_i and $Y_i(j)$, respectively:

$$|P(D_i = 1|Y_i(j) = y, W_i = w) \; - \; P(D_i = 1|W_i = w)| \quad \leq \quad c \; \in [0, 1]. \tag{2}$$

$c = 0$ implies the conditional independence in assumption (1). Larger values of c imply larger violations of (1), with the extreme case of $c \geq \max\{P(D_i = 0|W_i), P(D_i = 1|W_i)\}$ implying no restriction (zero-information) on $P(D_i = 1|Y_i(j) = y, W_i = w)$. Under the assumptions that D_i is conditionally c-dependent on $Y_i(j)$ for $j = 0, 1$ given W_i, Masten and Poirier (2018) provide the upper and lower bounds on ATE and ATET that depend on c, with $c = 0$ signifying point-identification. These bounds are continuous and monotonic in c.

We focus on ATE to fix ideas, although all results are reported for both ATE and ATET. For a given outcome with, e.g., ATE < 0, Masten and Poirier (2020) define the breakdown point c^* as the smallest c-dependence needed to overturn the result that ATE < 0, i.e.,

$$c^* \; = \; \inf_{c \in [0,1]} \{\text{level of } c\text{-dependence such that the upper bound of ATE is} \geq 0\}, \tag{3}$$

with the infimum defined as 1 if the set is empty. To assess whether such a value c^* is reasonable, Masten and Poirier (2020) recommend using the observed covariates to obtain the plausible c-dependence benchmarks as follows:

$$c_k \; = \; \sup_{w_k, w_{-k}} |P(D_i = 1|W_{i,k} = w_k, W_{i,-k} = w_{-k})) \; - \; P(D_i = 1|W_{i,-k} = w_{-k})|$$

where $W_{i,k}$ denotes the k-th covariate and $W_{i,-k}$ denotes the other covariates for different k's. c_k is the additional variation in the conditional probabilities of treatment due to the inclusion of $W_{i,k}$ in the conditioning set. Comparing (3) with (2), we see that in the definition of c_k,

the covariate $W_{i,k}$ takes the role of $Y_i(j)$, and the other covariates $W_{i,-k}$ take the role of W_i. These benchmark c_k's are identifiable from the data. When compared with the breakdown point c^*, these benchmarks help us assess the plausibility of a violation of our selection on observables assumption (1) that is big enough to overturn our causal conclusions.

Table 4 presents the breakdown point c^* for each outcome studied in our paper, using the "tesensitivity" package in STATA.[27] It also reports the elements $W_{i,k}$'s of the observed confounders in W_i, with c_k larger than c^*. Our causal conclusions will not survive if the inclusion of $Y_i(j)$ in the conditioning set already consisting of W_i should have similar impact on the propensity score as the inclusion of $W_{i,k}$ in the conditioning set already consisting of $W_{i,-k}$. For most outcomes, including the key outcome of physical abuse, this requires $Y_i(j)$ to have an impact similar to that of the village's population density i in 1991, or its distance from the nearest town in 1991, which were by *JS*'s account key variables in their selection of the treatment villages. In other words, the unobserved confounder needs to be extremely strong to overturn our causal conclusions.

The breakdown point for the ATE on experiencing no abuse is 0.81. Hence, overturning the result that *JS* increased the proportion of households with no abuse requires $Y_i(j)$ (in 2014) to have, roughly speaking, an impact on the propensity score of the *JS* treatment that is much larger than that of any observed determinants of the treatment that we consider based on *JS*'s account.

There are four outcomes for which our results are much more vulnerable to violations of the selection on observables condition in (1). These are: (i) sexual abuse, (ii) the wife's belief that beating her is justified, (iii) the wife's belief that she does not participate in household decision-making, and (iv) the husband's belief that his wife does not participate in household decision-making. This fragility was expected a priori (even without any formal sensitivity

[27] Masten et al. (2024) propose bootstrap inference for sensitivity analysis with c-dependence. As far as we know (based on our communications on Oct 24, 2025), this is not included in their STATA package "tesensitivity". Since their bootstrap is nonstandard and is computationally intensive (see their Section 6.2), we do not do inference for analysis with c-dependence. We present bootstrap inference for another analysis in the next section.

analysis) because the estimated effects of JS on these outcomes is very small.[28] These are important outcomes by all accounts, and hence the failure of JS to have robust impact on these four outcomes is noteworthy.

ATET is less sensitive than ATE in general, but not for the key outcomes of physical abuse and abuse-free relationship.

The bounds for the ATE's and ATET's as functions of c-dependence along with the benchmarks c_k's and the breakdown point c^* (the intersection with the 0 effect line) are presented in Figures 3-8 using the STATA package "tesensitivity".[29]

B.3.2 Long and short models: Analysis with Riesz representors

Sensitivity analyses proposed by Altonji et al. (2005), Oster (2019), Cinelli and Hazlett (2020), among others, are well-suited for empirical applications that model homogeneous treatment effects, i.e., the so-called partially linear models. Since we allow for heterogeneous effects, i.e., D_i appears by itself and also interacted with W_i in the regression adjustment, we instead follow Chernozhukov et al. (2024)'s approach to what they call the nonparametric model for ATE/ATET to obtain our second set of results on the sensitivity analysis of (1).

Suppose that the selection on observables assumption in (1), i.e., the so-called "short model" where every confounder is observed, is false. Let U_i be the unobserved confounder for village i. We will analyze how badly false the short model has to be for our causal conclusions to be overturned. For concreteness, it is useful to consider Figure 1 and think of U_i as the gender norms in 1991 (Norm1991), the short model as the DAG, i.e., nonparametric structural model, without the red-dashed arrow, and the so-called "long model" as the DAG with the red-dashed arrow.

[28] The baselines for these outcomes were relatively low. Hence there was little room for improvement. All else equal, it is easier to overturn a small effect than a larger effect.

[29] The outcomes take values in $[0, 1]$. With large enough c, the identified sets for ATE(T)s should coincide with the no-information bounds that should necessarily be of length 1 and should contain the 0 effect. However, 0 and/or 1 are often far from the extremes of the observed outcomes (Table 1). We do not use these implausibly conservative extremes, but use the command "tesensitivity" (without change) to bound the ATE(T)s.

Nonparametric point-identification of ATE is ruled out if the short model (1) is false. Under the assumption of conditional independence in the long model, i.e.:

$$Y_i(0) \text{ and } Y_i(1) \text{ are independent of } D_i \text{ conditional on } W_i \text{ and } U_i \text{ for all } i,$$

Chernozhukov et al. (2024) express the partially identified ATE as: $[\tau_L, \tau_U] = \tau \mp \sqrt{B^2}$ where $\tau = E\left[E[Y_i|D_i = 1, W_i] - E[Y_i|D_i = 0, W_i]\right]$ is the ATE under the short model (1),

$$B^2 = \left\{\eta^2_{(Y_i \sim U_i|D_i, W_i)}\right\} \times \left\{\frac{1 - \nu^2_{(D_i \sim U_i|W_i)}}{\nu^2_{(D_i \sim U_i|W_i)}}\right\} \times \left\{E\left[Var(Y_i|D_i, W_i)\right] E\left[\frac{1}{Var(D_i|W_i)}\right]\right\},$$

$$\eta^2_{(Y_i \sim U_i|D_i, W_i)} = 1 - \frac{E\left[Var(Y_i|D_i, W_i, U_i)\right]}{E\left[Var(Y_i|D_i, W_i)\right]}, \quad \text{and} \quad \nu^2_{(D_i \sim U_i|W_i)} = \frac{E\left[1/Var(D_i|W_i)\right]}{E\left[1/Var(D_i|W_i, U_i)\right]}.$$

$\eta^2_{(Y_i \sim U_i|D_i, W_i)}$ and $1 - \nu^2_{(D_i \sim U_i|W_i)}$ are the parameters for the sensitivity analysis of (1) since they are not identifiable from the data because of the presence of the unobserved confounder U_i.[30] $\eta^2_{(Y_i \sim U_i|D_i, W_i)} \in [0, 1]$ measures the additional gain in explanatory power that U_i provides for Y_i, beyond what is already explained by D_i and W_i. $1 - \nu^2_{(D_i \sim U_i|W_i)} \in [0, 1]$ measures the relative gain in the average precision of the treatment D_i model due to U_i. $\eta^2_{(Y_i \sim U_i|D_i, W_i)} = 0$ and $1 - \nu^2_{(D_i \sim U_i|W_i)} = 0$ if "U_i is irrelevant" in the sense that the short model (1) is true. Then ATE is point-identified; otherwise it is partially identified by $[\tau_L, \tau_U]$.[31]

[30] More informatively, $\eta^2_{(Y_i \sim U_i|D_i, W_i)}$ and $1 - \nu^2_{(D_i \sim U_i|W_i)}$ are defined as follows:

$$\eta^2_{(Y_i \sim U_i|D_i, W_i)} = \frac{E\left[\left(E[Y_i|D_i, W_i, U_i] - [E[Y_i|D_i, W_i]\right)^2\right]}{E\left[\left(Y_i - [E[Y_i|D_i, W_i]\right)^2\right]} = 1 - \frac{E\left[Var(Y_i|D_i, W_i, U_i)\right]}{E\left[Var(Y_i|D_i, W_i)\right]},$$

$$\nu^2_{(D_i \sim U_i|W_i)} = \frac{E\left[\left(D_i/P(D_i = 1|W_i) - (1 - D_i)/P(D_i = 0|W_i)\right)^2\right]}{E\left[\left(D_i/P(D_i = 1|W_i, U_i) - (1 - D_i)/P(D_i = 0|W_i, U_i)\right)^2\right]} = \frac{E\left[1/Var(D_i|W_i)\right]}{E\left[1/Var(D_i|W_i, U_i)\right]}.$$

[31] Analogous to the ATE, the partially identified ATET is $\left[\tau_L^{ATET}, \tau_U^{ATET}\right] = \tau^{ATET} \mp \sqrt{B^2_{ATET}}$ where $\tau^{ATET} = E[Y_i|D_i = 1] - E[E[Y_i|D_i = 0, W_i]|D_i = 1]$ is the ATET under the short model (1), and:

$$B^2_{ATET} = \left\{\eta^2_{(Y_i \sim U_i|D_i, W_i)}\right\} \times \left\{\frac{1 - \nu^2_{(D_i \sim U_i|W_i), ATET}}{\nu^2_{(D_i \sim U_i|W_i), ATET}}\right\} \times \left\{E\left[Var(Y_i|D_i, W_i)\right] E\left[\alpha^2(W_i)\right]\right\}$$

When $\widehat{\tau} < 0$ (i.e., our estimated ATE < 0), then the sensitivity analysis studies the plausibility of the values of the sensitivity parameters $\left(\eta^2_{(Y_i \sim U_i | D_i, W_i)}, 1 - \nu^2_{(D_i \sim U_i | W_i)}\right)$ that make $\widehat{\tau}_U \geq 0$. In particular, in Table 5, we report the estimated Robustness Value, which is the value $r_0 \in [0,1]$ such that $\widehat{\tau}_U(r_0, r_0) = 0$ when $\widehat{\tau} < 0$ (alternatively, $\widehat{\tau}_L(r_0, r_0) = 0$ when $\widehat{\tau} > 0$). To account for the estimation error for both τ and the identifiable part of B^2, we do a similar analysis with the upper 95% confidence bound of τ_U when $\widehat{\tau}_U < 0$. The counterpart of the Robustness Value in this context is denoted by Robustness Value (.05) and is defined as $r_0 \in [0,1]$ such that the upper 95% confidence bound is 0 when $\eta^2_{(Y_i \sim U_i | D_i, W_i)} = 1 - \nu^2_{(D_i \sim U_i | W_i)} = r_0$.[32] When $\widehat{\tau}_L > 0$, we look instead at the lower 95% confidence bound of τ_L. These results are reported based on not only the RA estimates as in the rest of the paper, but also the IPW estimates because the Riesz representor, which is fundamental to the construction in Chernozhukov et al. (2024), is directly associated with IPW estimation. The results based on the RA and IPW estimates (and also other estimates) are very similar and convey the same message about the sensitivity of our results to (1).

with $\eta^2_{(Y_i \sim U_i | D_i, W_i)}$ the same as for ATE, but the other terms (due to different Riesz representor) given by:

$$\begin{aligned} \alpha(Z_i) &= \frac{P(D_i = 1|Z_i)}{P(D_i = 1)} \left\{ \frac{D_i}{P(D_i = 1|Z_i)} - \frac{1 - D_i}{P(D_i = 0|Z_i)} \right\} \quad \text{for } Z_i = W_i \text{ or } Z_i = W_i, U_i, \\ \nu^2_{(D_i \sim U_i | W_i), ATET} &= \frac{E\left[\alpha^2(W_i)\right]}{E\left[\alpha^2(W_i, U_i)\right]} = \frac{E\left[P(D_i = 1|W_i)/P(D_i = 0|W_i)\right]}{E\left[P(D_i = 1|W_i, U_i)/P(D_i = 0|W_i, U_i)\right]}. \end{aligned}$$

$\eta^2_{(Y_i \sim U_i | D_i, W_i)}$ and $1 - \nu^2_{(D_i \sim U_i | W_i), ATET}$ are the parameters for the sensitivity analysis of (1) for the estimation of ATET. In the case of ATET, (i) the long and short models do not involve $Y_i(1)$, and (ii) $\eta^2_{(Y_i \sim U_i | D_i, W_i)}$ can be further reduced by taking (i) into account. We abstract from these peripheral matters for simplicity. While, as before, our discussion here focuses on ATE, the results for the sensitivity analysis of ATE and ATET are reported in equal details.

[32] While Chernozhukov et al. (2024) provide standard errors based on the semiparametrically efficient estimators, we use nonparametric percentile bootstrap to obtain the confidence bounds. We justify this as follows. We wish to report this sensitivity analysis with the RA and IPW estimators (result for other estimators are also available from us), since the results while similar are not identical. This means that the standard error based on the efficiency bound formula may be less appropriate with our small sample. On the other hand, formulae for standard errors based on parametric treatment of these estimators are tedious and different for these estimators. We use nonparametric bootstrap to bypass all these.

The Robustness Value and the Robustness Value (.05), based on RA and IPW, for each outcome are reported in Table 5. They suggest that our causal conclusions survive even under large but equal value of the sensitivity parameters.[33]

We report the visual representation of this sensitivity analysis for both ATE and ATET based on, respectively, RA and IPW estimates in Figures 9-20 and 21-32. We follow Imbens (2003), Cinelli and Hazlett (2020) and Chernozhukov et al. (2024) for this visual representation. Taking the sensitivity parameters $\left(\eta^2_{(Y_i \sim U_i | D_i, W_i)}, 1 - \nu^2_{(D_i \sim U_i | W_i)}\right)$ as axes, we plot the contours of: (i) $\widehat{\tau}_U$ and the upper 95% confidence bound of τ_U if $\widehat{\tau} < 0$, and (ii) $\widehat{\tau}_L$ and the lower 95% confidence bound of τ_L if $\widehat{\tau} > 0$.[34] And, similarly for ATET.

How likely is it for the unobserved confounder U_i to have strength on or above the 0 contour line, and thus overturn our results? This is ultimately for each reader to decide. The visual representation in these figures may facilitate that decision.

Benchmarks to further facilitate assessing the plausibility of values of the sensitivity parameters that can overturn our causal conclusions from a non-inferential point of view (in panel (A) of these figures) and from an inferential point of view (in panel (B) of these figures) are obtained as follows. To fix ideas, we present the description of benchmarks for ATE only, while that for ATET follow similarly with its own sensitivity parameters following the discussion in Footnote 31. First, for each observed confounder $W_{i,k}$, and with $W_{i,-k}$ denoting the remaining elements of W_i as before, for $k = 1, \ldots, K$ $(K = 5)$ we estimate:

$$\eta^2_k \;=\; 1 - \frac{E\left[Var(Y_i|D_i, W_i)\right]}{E\left[Var(Y_i|D_i, W_{i,-k})\right]}, \quad \text{and} \quad 1 - \nu^2_k = 1 - \frac{E\left[1/Var(D_i|W_{i,-k})\right]}{E\left[1/Var(D_i|W_i)\right]}$$

resembling the original definitions but with $W_{i,k}$ playing the role of U_i and $W_{i,-k}$ playing

[33] Results under "extreme scenarios" as in Cinelli and Hazlett (2020) (e.g., setting $\eta^2_{(Y_i \sim U_i | D_i, W_i)} = 1$ and studying sensitivity to $1 - \nu^2_{(D_i \sim U_i | W_i)}$), but under Chernozhukov et al. (2024)'s setup, are available from us.

[34] We omit the outcome of response of married women to the question if physical abuse on them is justified. There is no meaningful effect of JS on this outcome (perhaps because the baseline proportion of women who justify such physical abuse is already low) for the sensitivity analysis to make sense.

the role of W_i. Then we take each of the K pairs $(\eta_k^2, 1-\nu_k^2)$ for $k = 1, \ldots, K$ and also a more conservative worst-case measure $(\max_{1\leq k\leq K} \eta_k^2, \max_{1\leq k\leq K}(1-\nu_k^2))$ as benchmarks for our sensitivity parameters. These six $(K+1)$ points are plotted in the contour plots. The regular benchmarks are plotted in black while "the red dot" is the synthetic benchmark, i.e, the worst-case measure. These benchmarks are suggestive reference points for the readers about how much more influential the unobserved confounder U_i needs to be relative to the observed confounders W_i in order to overturn our results on the ATE on each outcome. For completeness of the representation, we also plot, marked in blue, the Robustness Value in the figures in panel (A) and the Robustness Value (.05) in the figures in panel (B).

As far as non-inferential results are concerned, i.e., if we consider $\widehat{\tau}_U$ or $\widehat{\tau}_L$ (panel (A) in Figures 9-20 for RA and 21-32 for IPW), it is evident that the unobserved confounder has to be implausibly influential to overturn the causal conclusion reported in our paper. We should emphasize that while "weak covariates" can artificially make the sensitivity analysis results look "good", our covariates include the population density and remoteness of the villages that were, by JS's own account, the key variables in their selection of the treatment villages. Therefore, based on what we learned from JS about their treatment assignment criteria, there does not seem to be any obvious candidate for unobserved confounder that could be strong enough to overturn the non-inferential results (panel A in the figues) on the RA and IPW estimates of ATEs and ATETs.

The inferential results (panel (B) in Figures 9-32), i.e., the upper 95% confidence bound for τ_U or the lower 95% confidence bound for τ_L, as appropriate, can be overturned by less extreme strength of unobserved confounders. (This is true by definition.) The ATE/ATET on the four outcomes that were fragile in last section continue to be fragile. The analysis here additionally suggests the following about the fragility of the estimates. The RA estimate for ATE and IPW estimate for ATET on the reduction in the proportion of husbands who believe that there is no law against domestic violence can be overturned if the unobserved confounder is as influential as the remoteness of the village, the key factor in JS's selection.

To overturn the effect on the decrease in physical abuse, one of the key outcomes, the unobserved confounder needs to be as influential as the synthetic worst-case benchmark; the IPW estimate of ATET is a little more fragile and can be overturned if the unobserved confounder is as influential as the remoteness of the village. These are unlikely given what we learned from *JS* about their selection criteria. The effect on the other outcomes cannot be overturned by an unobserved confounder that is as strong as the benchmarks, including the worst-case synthetic benchmark.

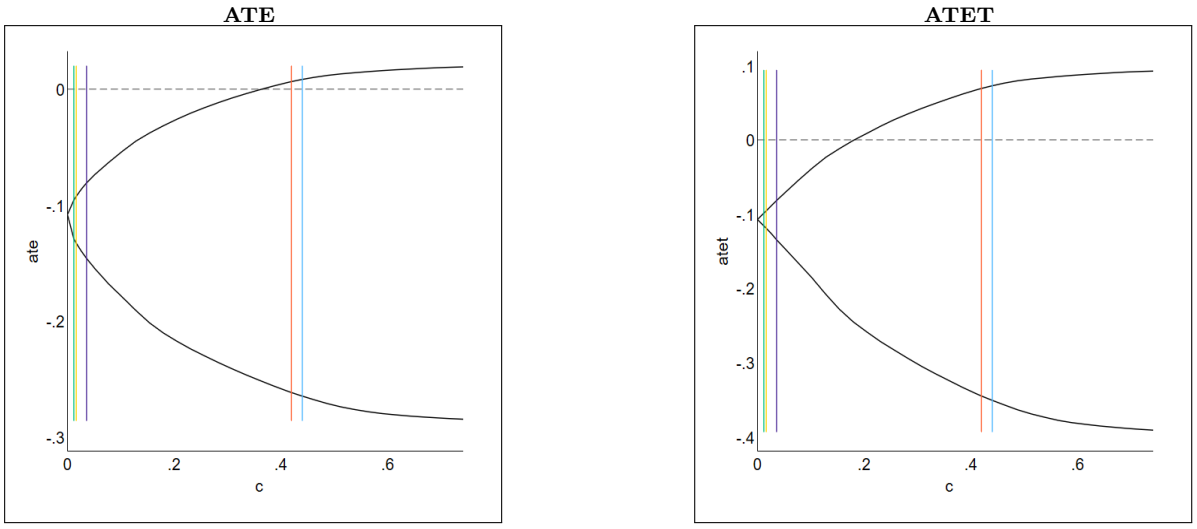
B.3.3 Sensitivity plots

The plots that serve as the extended reference for the discussion of sensitivity analysis in Supplemental Appendices B.3.1 are obtained tesensitivity package in STATA.

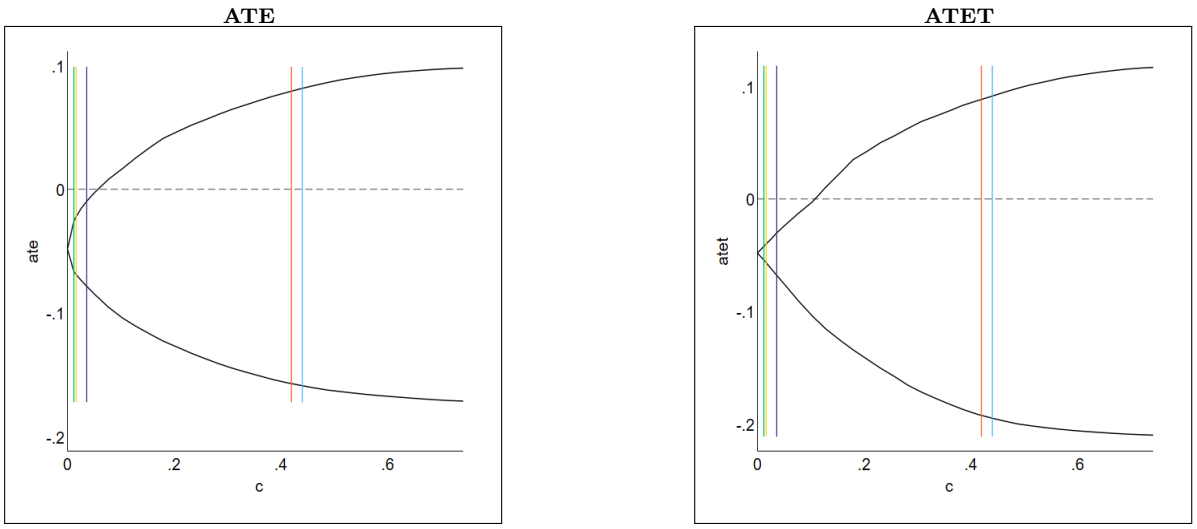
The plots that serve as the extended reference for the discussion of sensitivity analysis in Supplemental Appendices B.3.2 are obtained by our own Matlab code that is available from us. The first set of plots are based on the RA estimates. The second set of plots are based on the IPW estimates. Since we do not assume that the efficiency bound is attained and rather wish to control for the estimation of *parametrically* estimated nuisance parameters, we use non-parametric bootstrap for the inferential plots. RA and IPW represent two extremes about the researchers beliefs about the correctness of the nuisance parameters, respectively, the conditional expectations of potential outcomes and the propensity score. It is reassuring that the sensitivity plots based on both RA and IPW estimates give very similar message about the robustness of our results.

Figure 3: Masten-Poirier bounds of ATE and ATET

Physical abuse of wife



Sexual abuse of wife



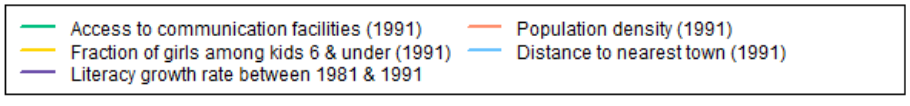


Figure 4: Masten-Poirier bounds of ATE and ATET

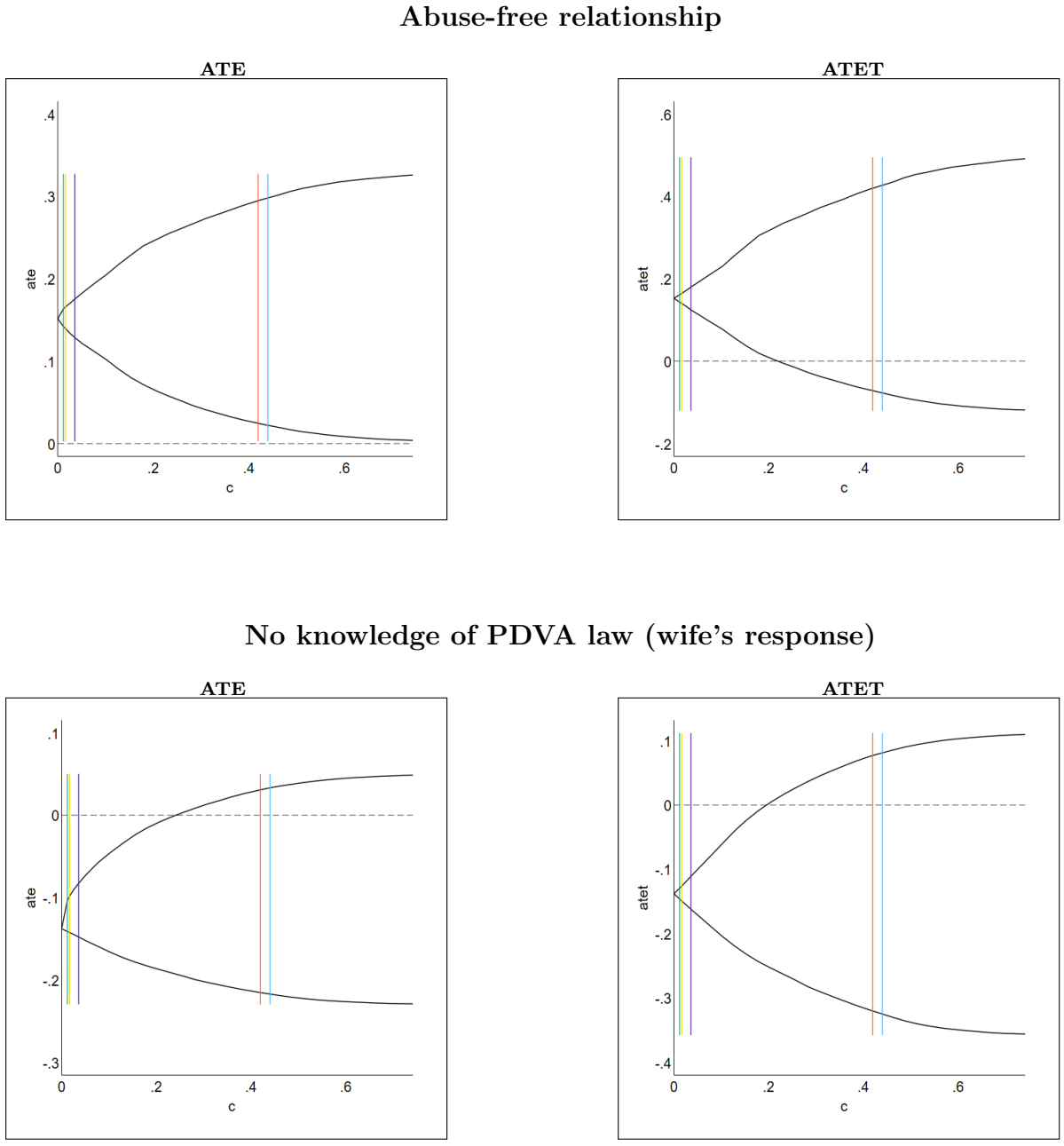
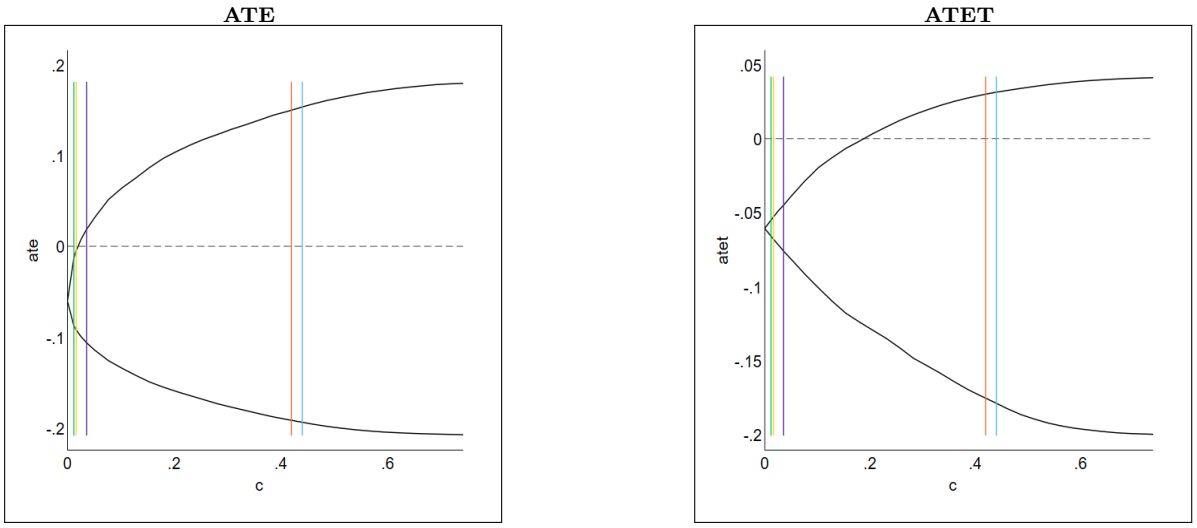
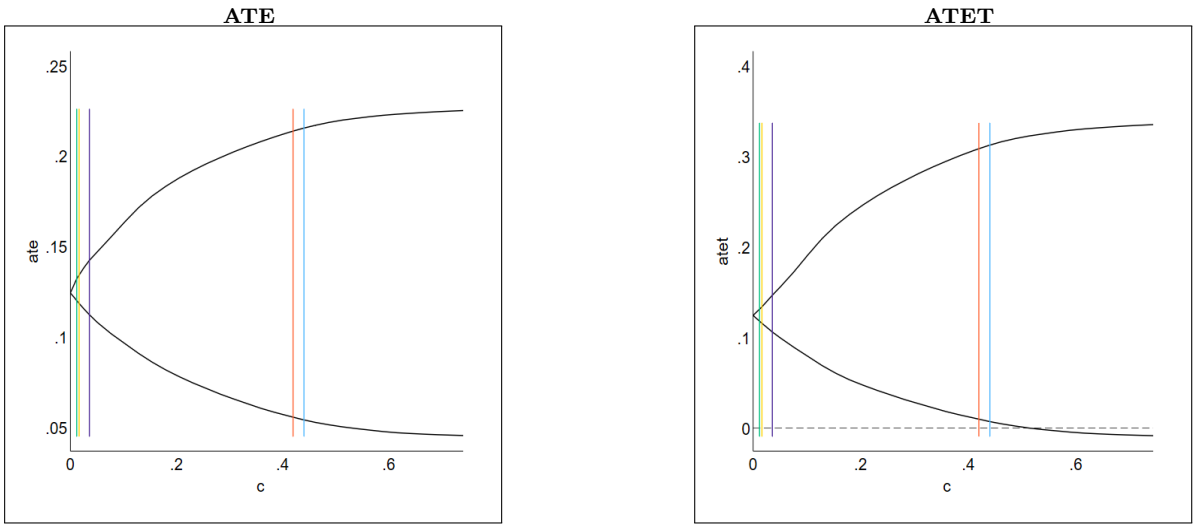


Figure 5: Masten-Poirier bounds of ATE and ATET

Wife does not participate in decision making (wife's response)



Report acts of domestic violence (wife's response)



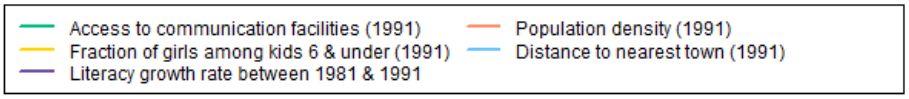
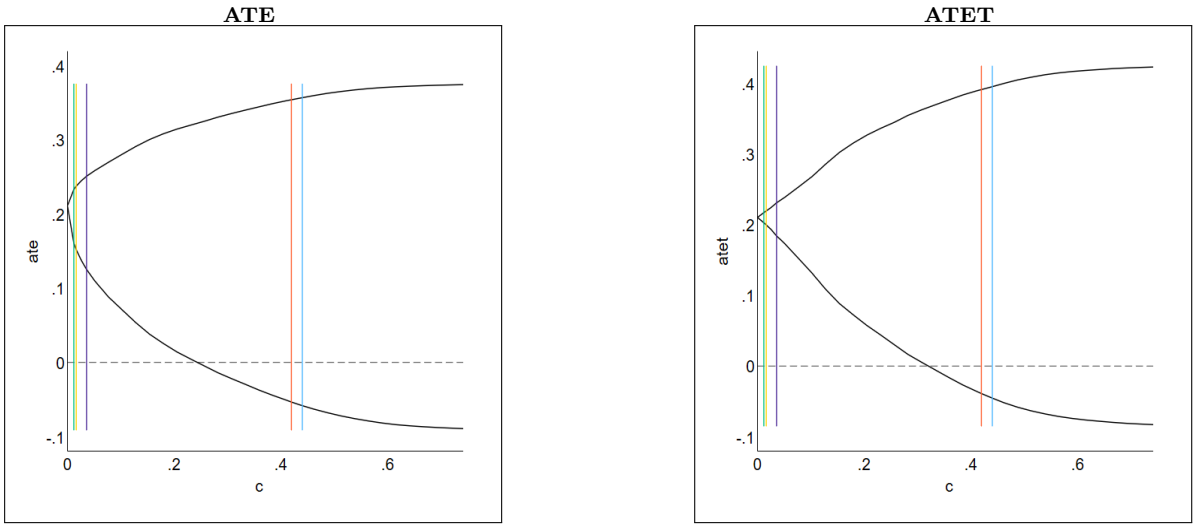
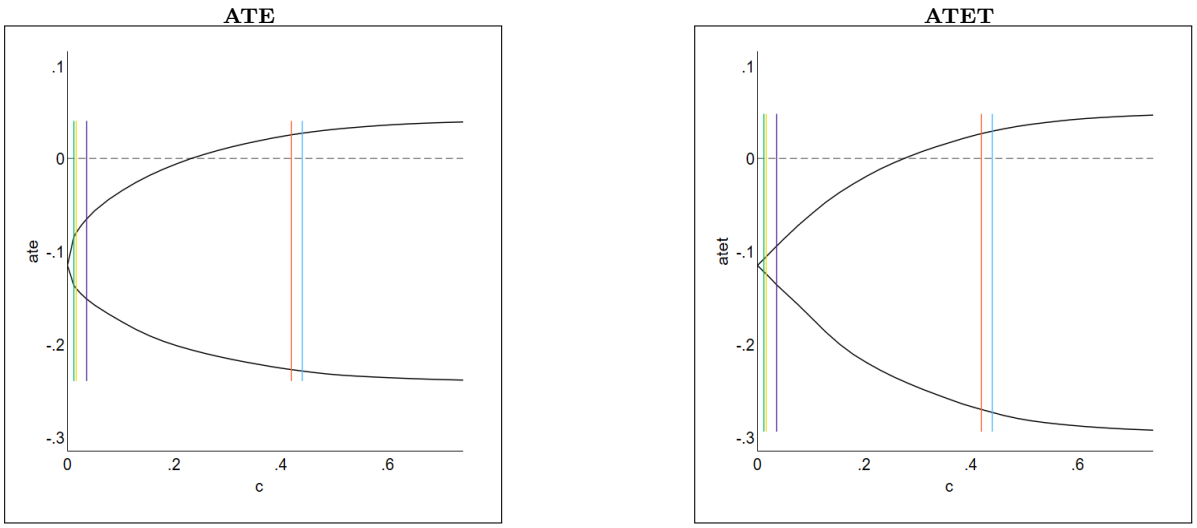


Figure 6: Masten-Poirier bounds of ATE and ATET

Demolish illegal liquor shops (wife's response)



Physical abuse of wife is justified (husband's response)



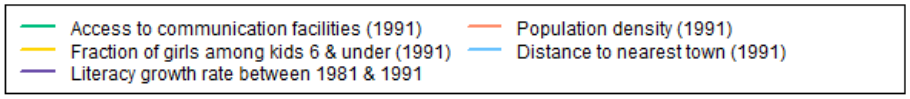
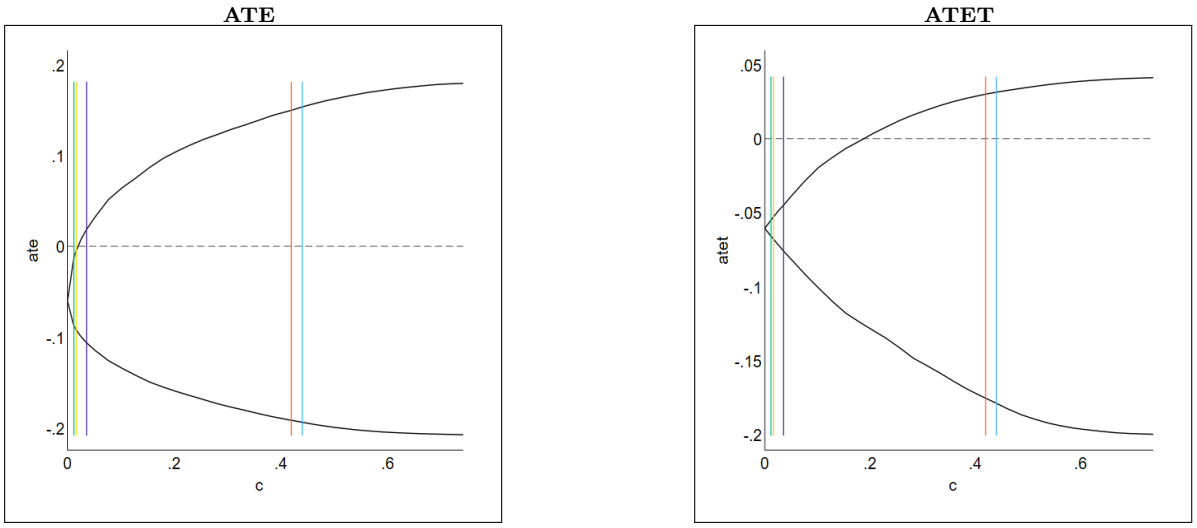
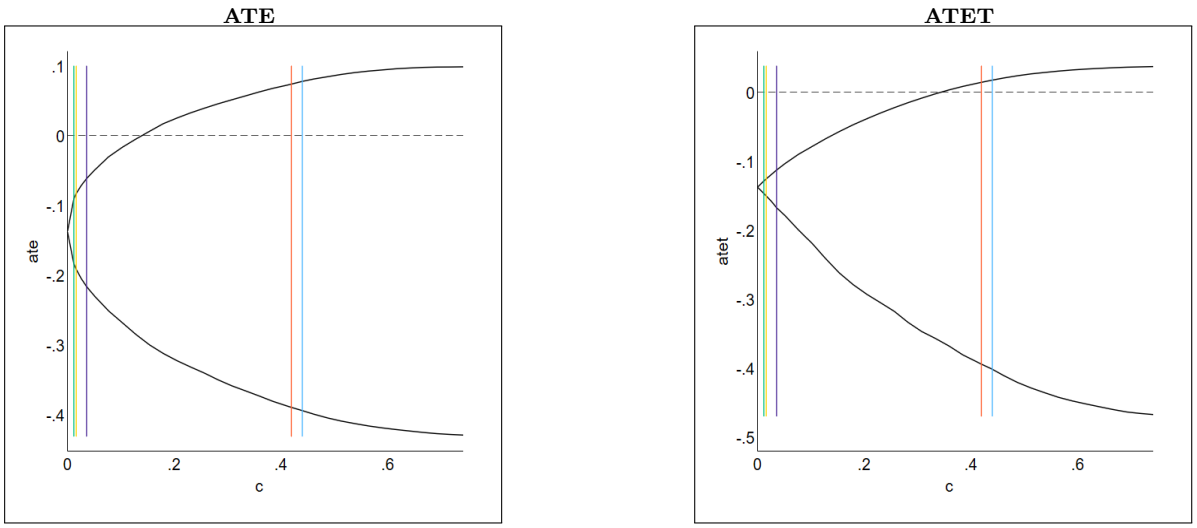


Figure 7: Masten-Poirier bounds of ATE and ATET

Wife does not participate in decision making (husband's response)



No knowledge of PDVA law (husband's response)



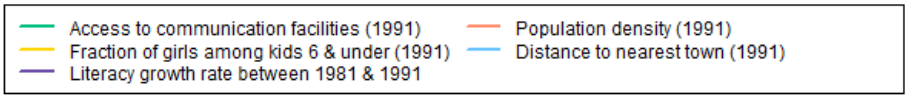
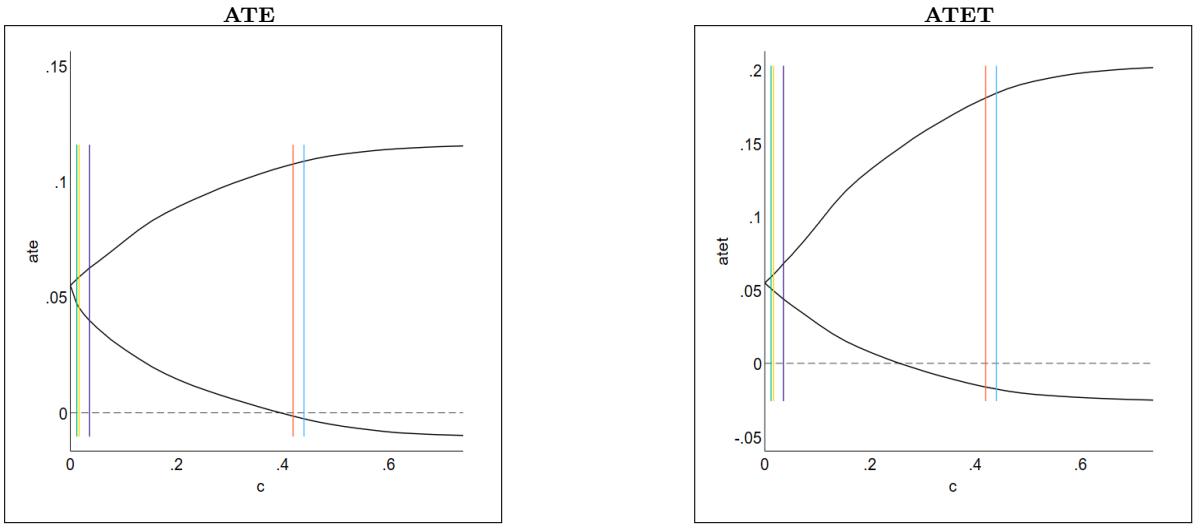
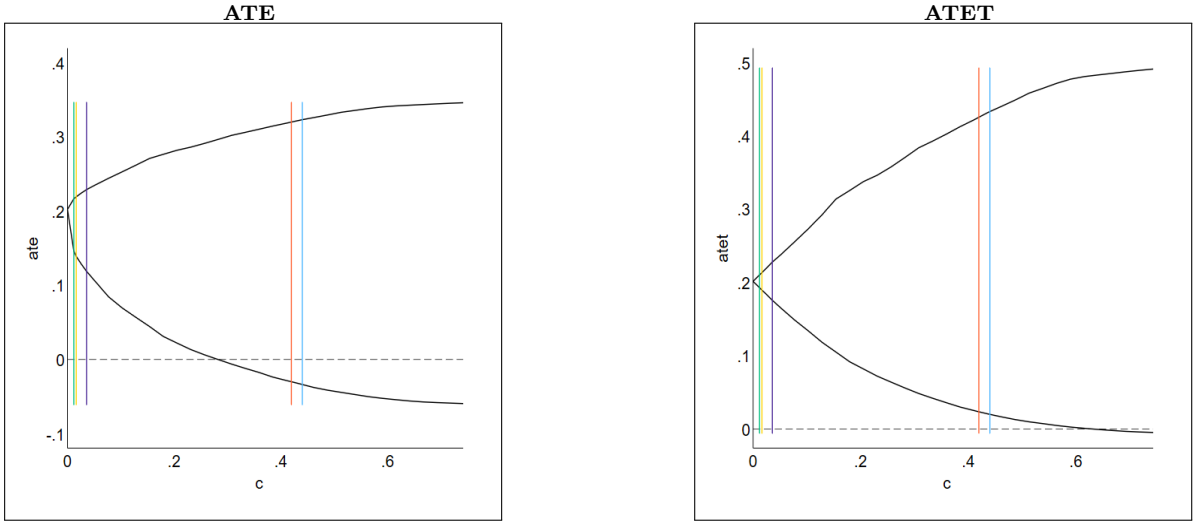


Figure 8: Masten-Poirier bounds of ATE and ATET

Report acts of domestic violence (husband's response)



Demolish illegal liquor shops (husband's response)



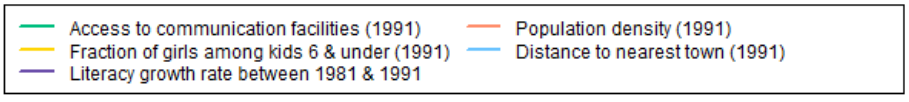
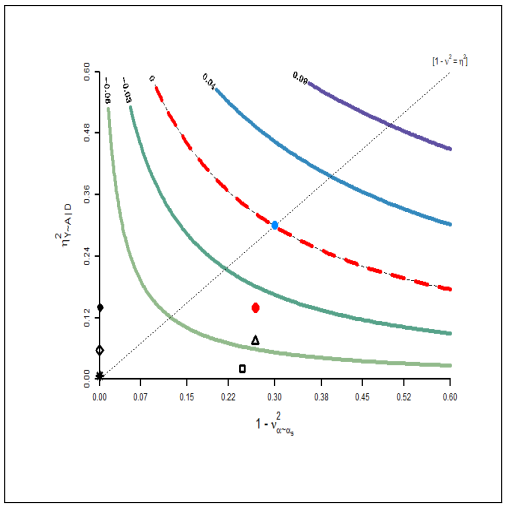


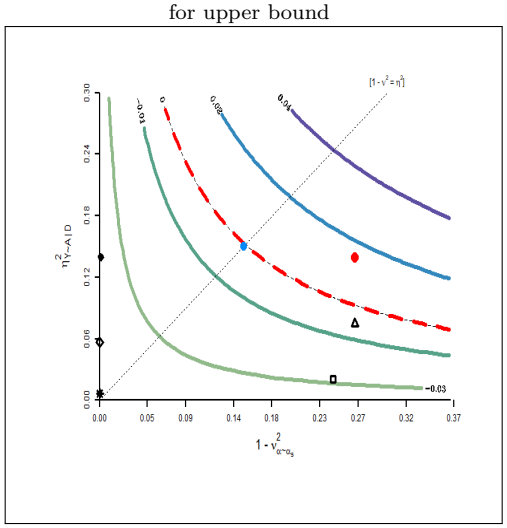
Figure 9: Physical abuse of wife: Contour plots based on Chernozhukov et al. (2024)

ATE: using RA results

(A) Upper bound

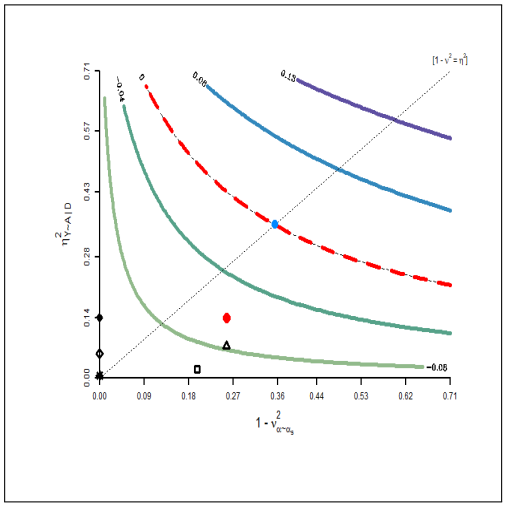


(B) Upper 95 % confidence bound for upper bound

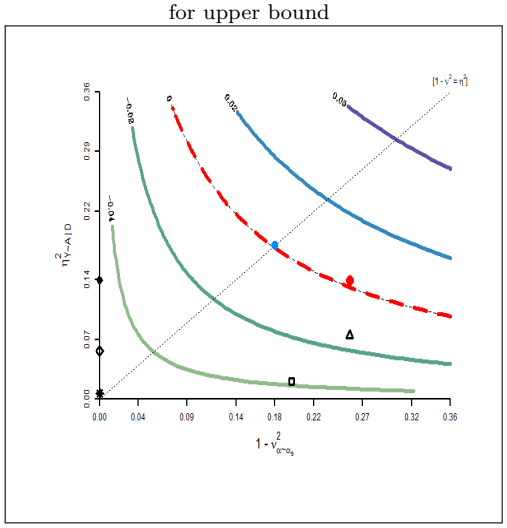


ATET: using RA results

(A) Upper bound



(B) Upper 95 % confidence bound for upper bound



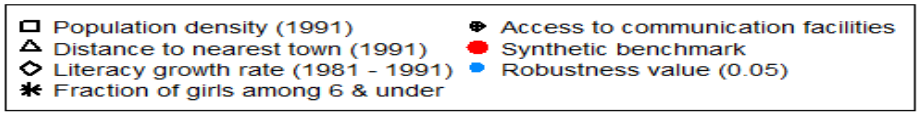
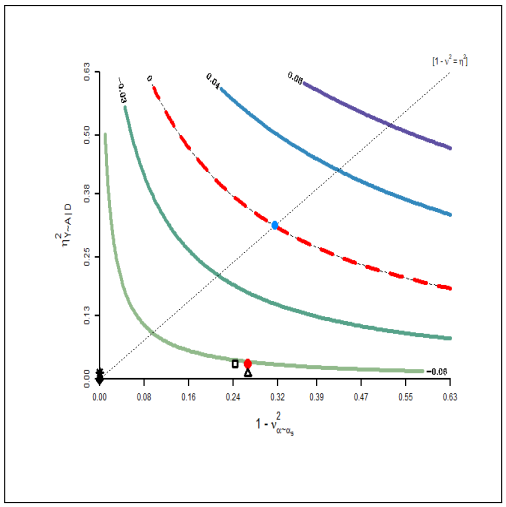


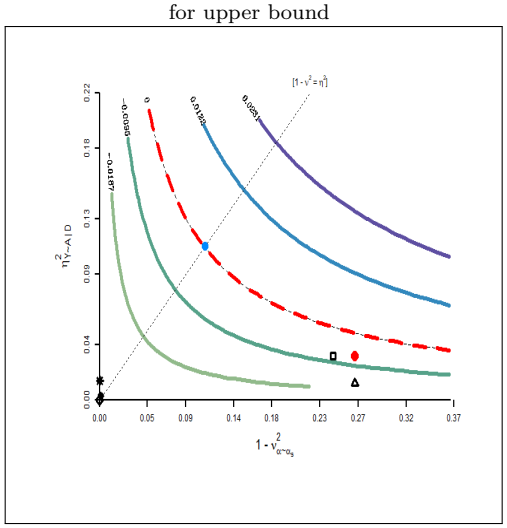
Figure 10: **Sexual abuse of wife: Contour plots based on Chernozhukov et al. (2024)**

ATE: using RA results

(A) Upper bound

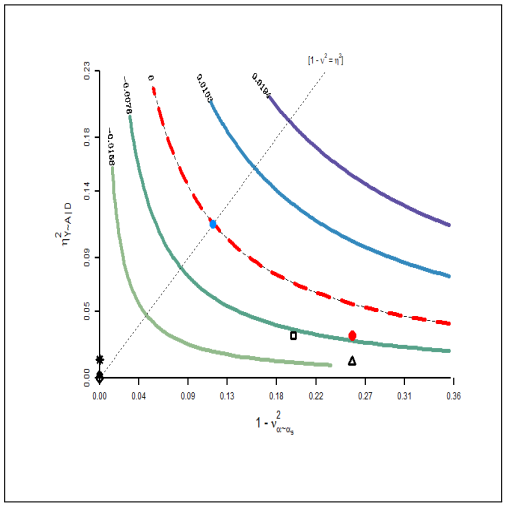


(B) Upper 95 % confidence bound for upper bound

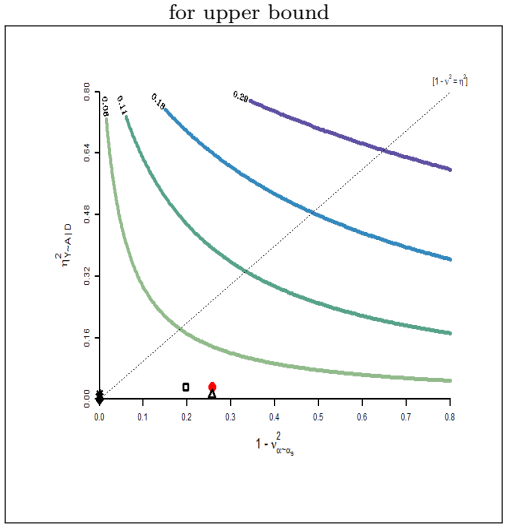


ATET: using RA results

(A) Upper bound



(B) Upper 95 % confidence bound for upper bound



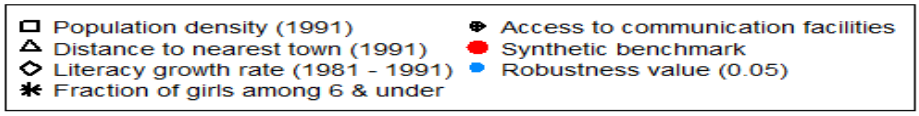
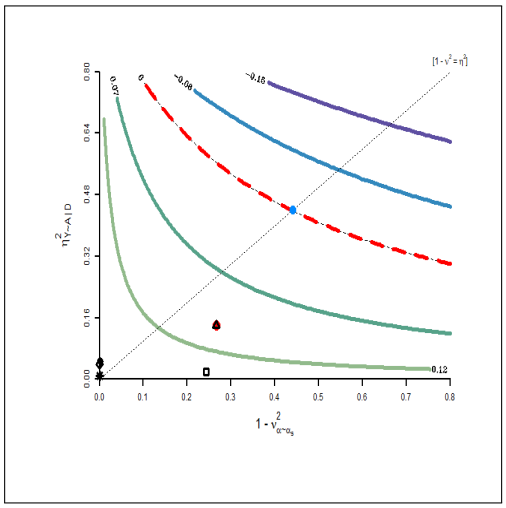


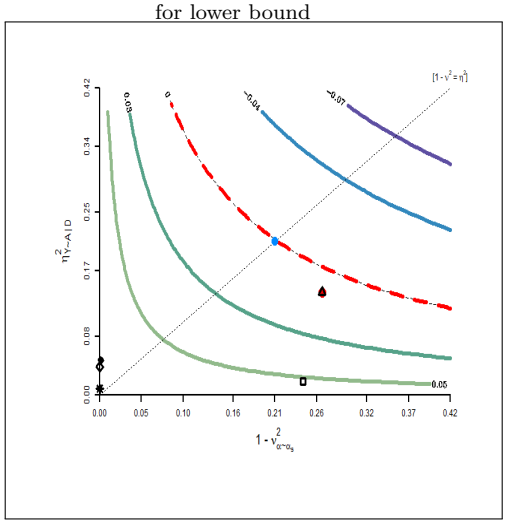
Figure 11: Abuse-free relationship: Contour plots based on Chernozhukov et al. (2024)

ATE: using RA results

(A) Lower bound

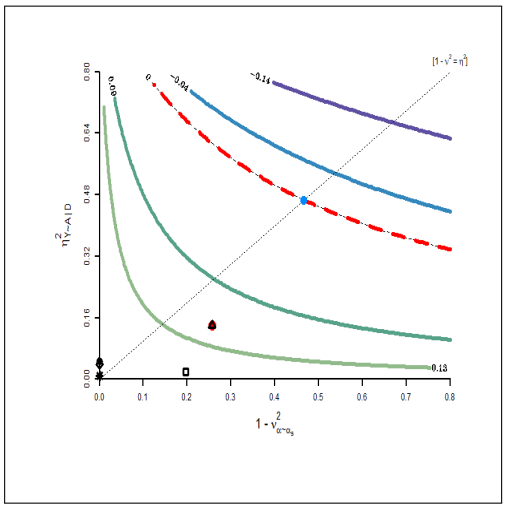


(B) Lower 95 % confidence bound for lower bound

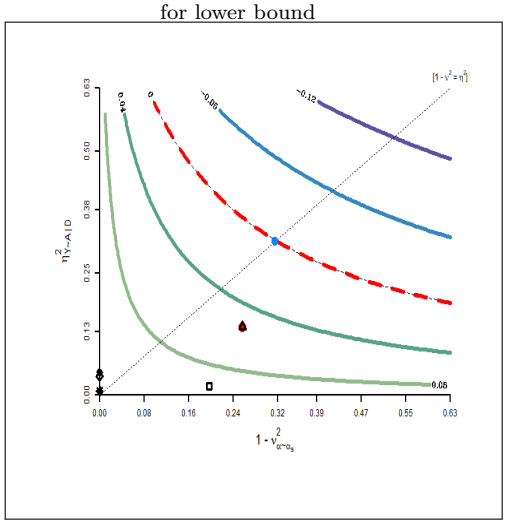


ATET: using RA results

(A) Lower bound



(B) Lower 95 % confidence bound for lower bound



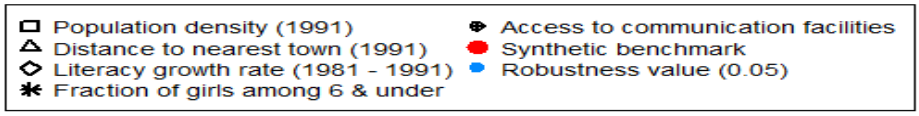
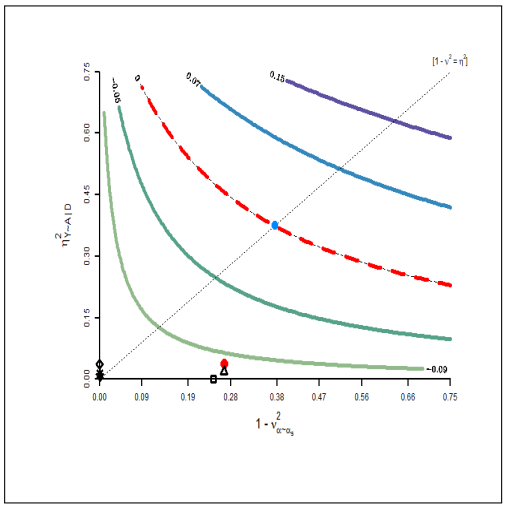


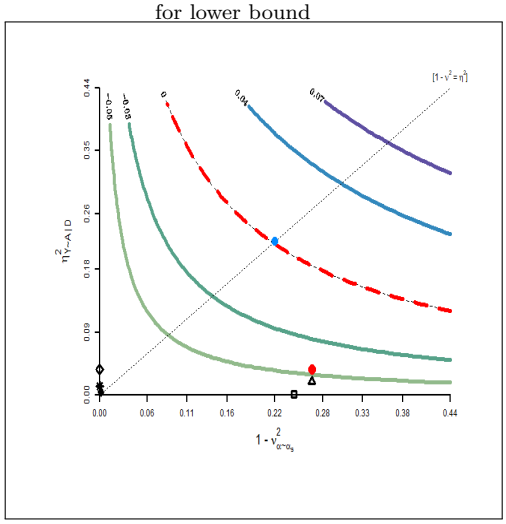
Figure 12: **No knowledge of law against domestic violence (wife's response): Contour plots based on Chernozhukov et al. (2024)**

ATE: using RA results

(A) Lower bound

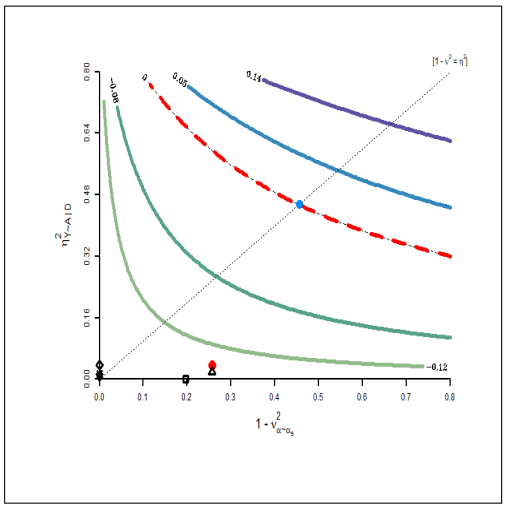


(B) Upper 95 % confidence bound for lower bound

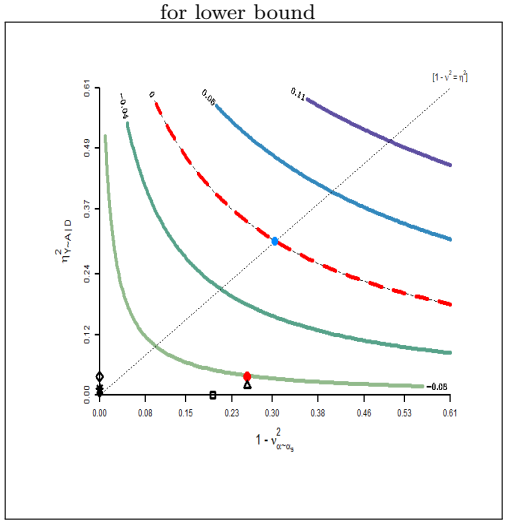


ATET: using RA results

(A) Lower bound



(B) Upper 95 % confidence bound for lower bound



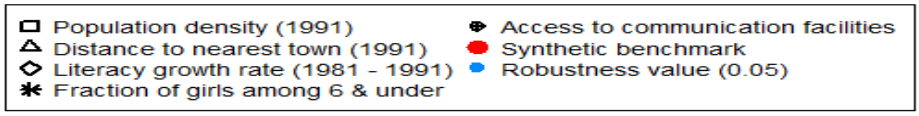
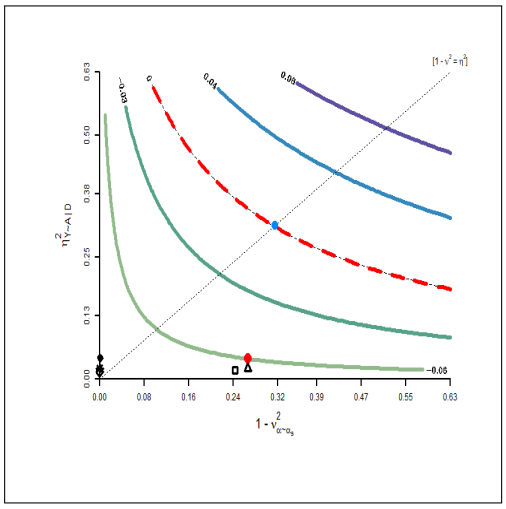


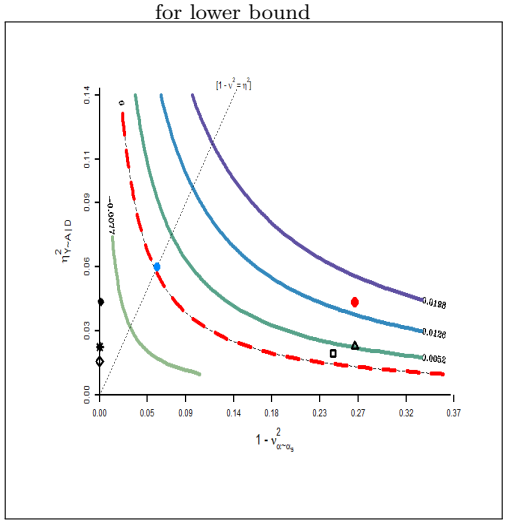
Figure 13: **Wife doesn't participate in decision making (wife's response): Contour plots based on Chernozhukov et al. (2024)**

ATE: using RA results

(A) Lower bound

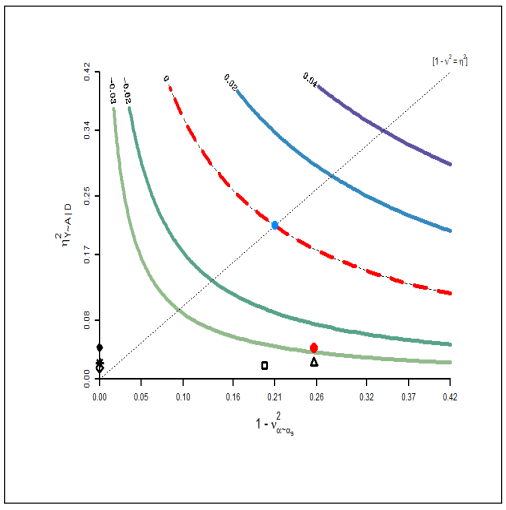


(B) Upper 95 % confidence bound for lower bound

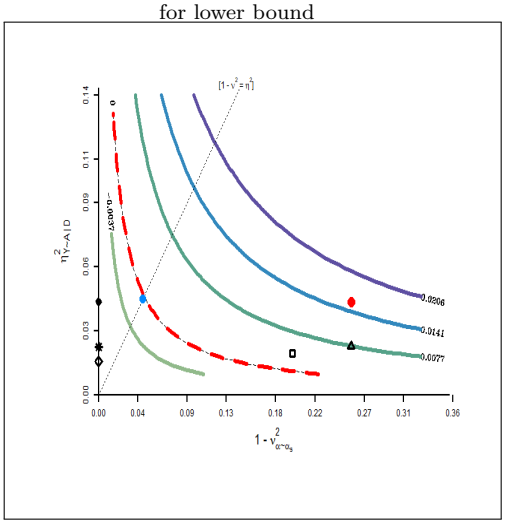


ATET: using RA results

(A) Lower bound



(B) Upper 95 % confidence bound for lower bound

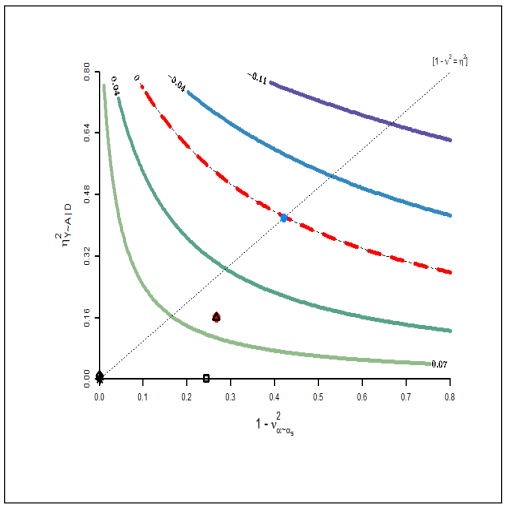


Population density (1991)
Distance to nearest town (1991)
Literacy growth rate (1981 - 1991)
Fraction of girls among 6 & under
Access to communication facilities
Synthetic benchmark
Robustness value (0.05)

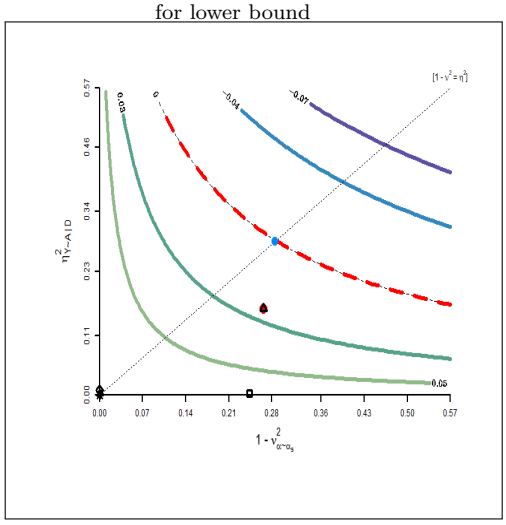
Figure 14: **Report acts of domestic violence (wife's response): Contour plots based on Chernozhukov et al. (2024)**

ATE: using RA results

(A) Lower bound

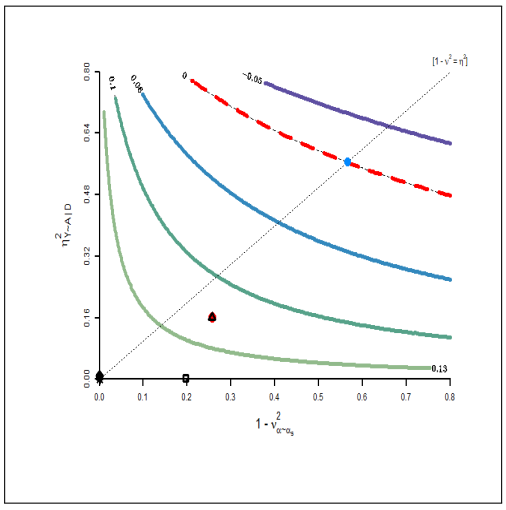


(B) Lower 95 % confidence bound for lower bound

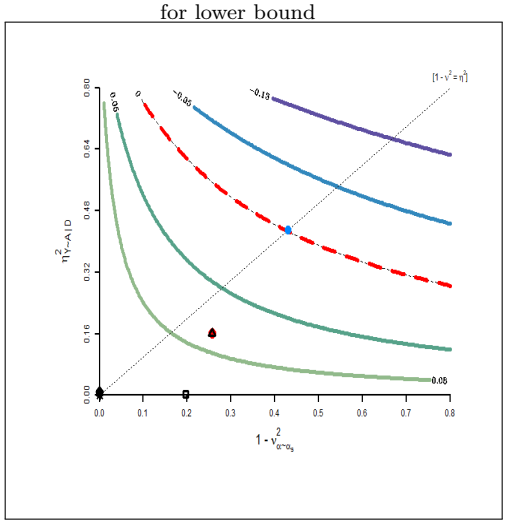


ATET: using RA results

(A) Lower bound



(B) Lower 95 % confidence bound for lower bound

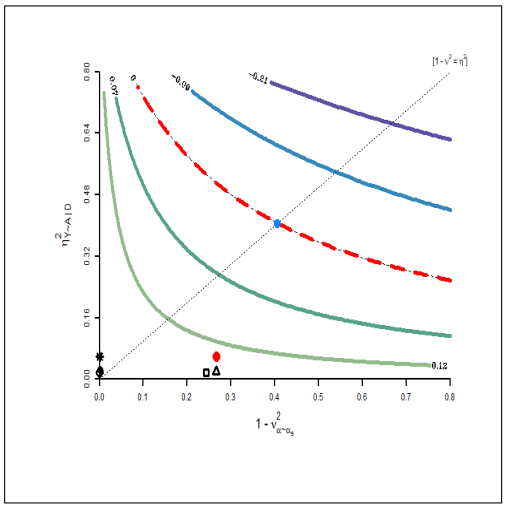


Population density (1991)
Distance to nearest town (1991)
Literacy growth rate (1981 - 1991)
Fraction of girls among 6 & under
Access to communication facilities
Synthetic benchmark
Robustness value (0.05)

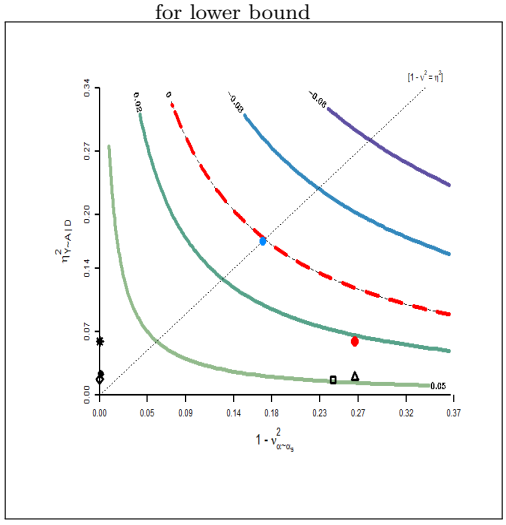
Figure 15: **Demolish illegal liquor shops (wife's response): Contour plots based on Chernozhukov et al. (2024)**

ATE: using RA results

(A) Lower bound

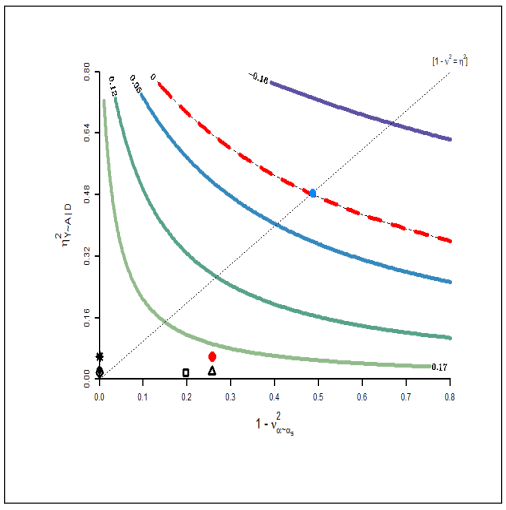


(B) Lower 95 % confidence bound for lower bound

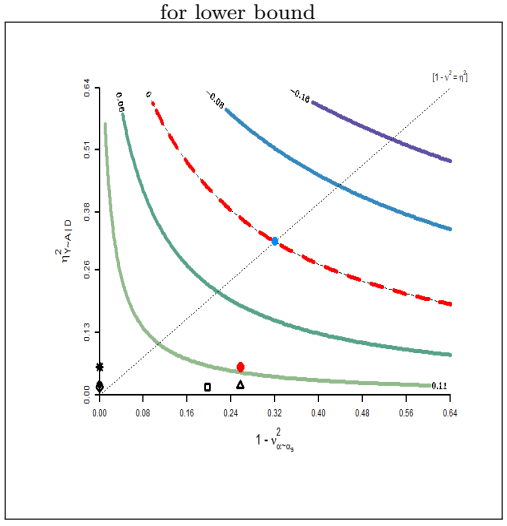


ATET: using RA results

(A) Lower bound



(B) Lower 95 % confidence bound for lower bound



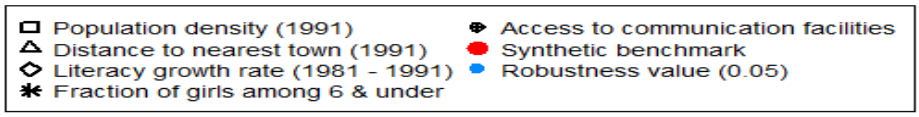
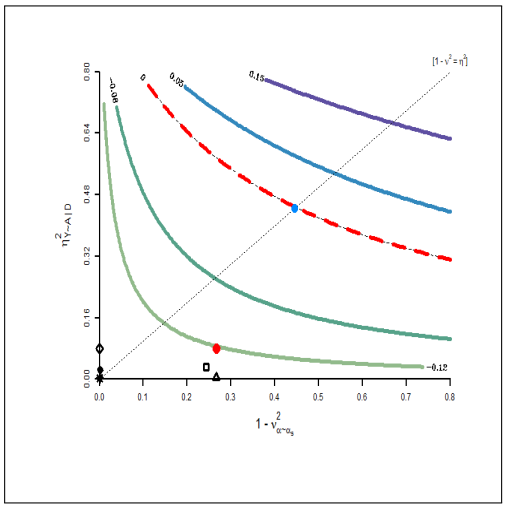


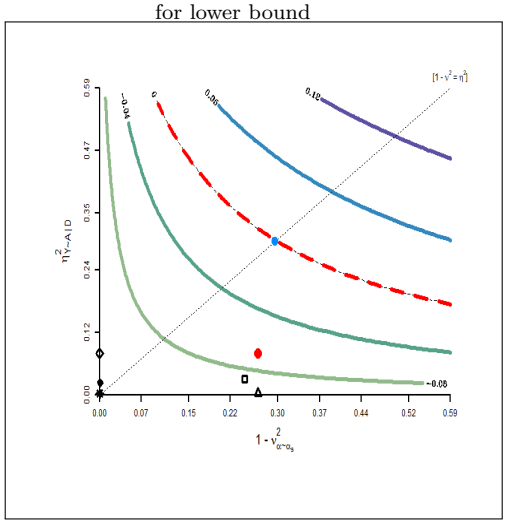
Figure 16: **Wife beating is justified (husband's response): Contour plots based on Chernozhukov et al. (2024)**

ATE: using RA results

(A) Lower bound

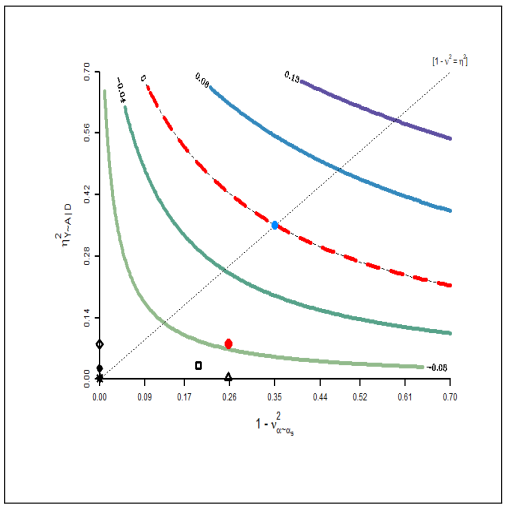


(B) Upper 95 % confidence bound for lower bound

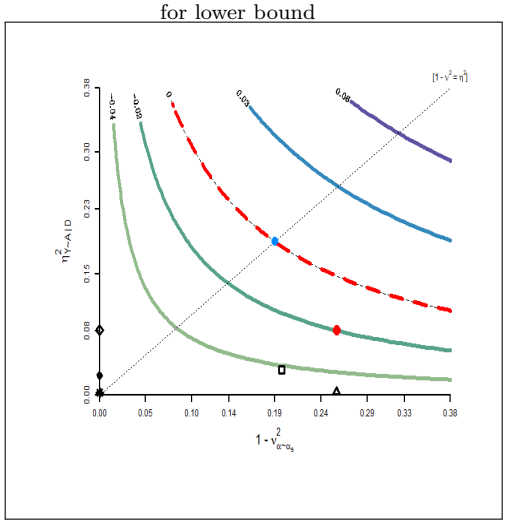


ATET: using RA results

(A) Lower bound



(B) Upper 95 % confidence bound for lower bound



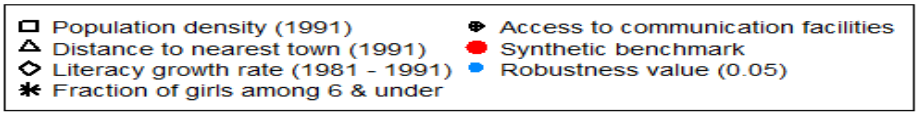
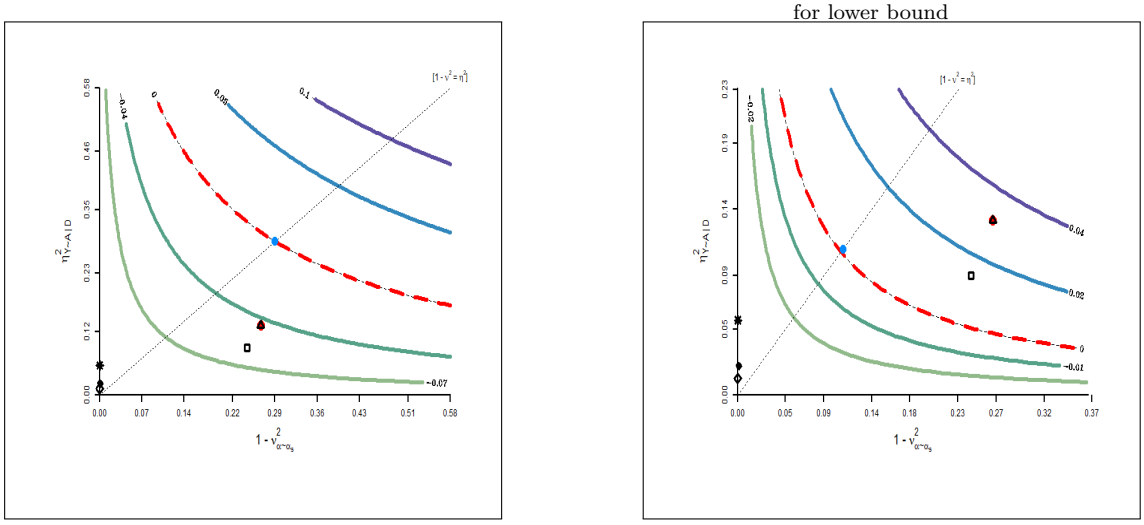


Figure 17: No knowledge of law against domestic violence (husband's response): Contour plots based on Chernozhukov et al. (2024)

ATE: using RA results

(A) Lower bound

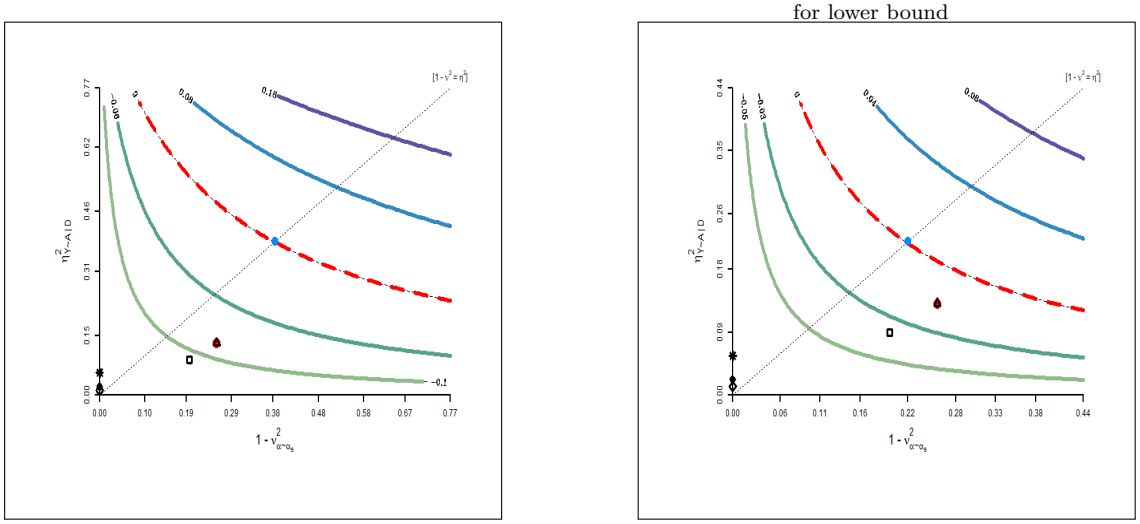
(B) Upper 95 % confidence bound for lower bound



ATET: using RA results

(A) Lower bound

(B) Upper 95 % confidence bound for lower bound



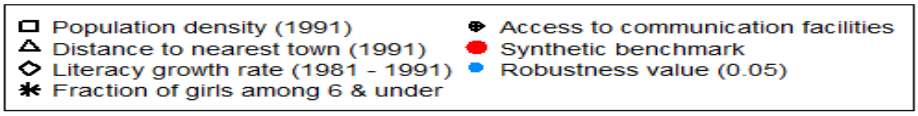
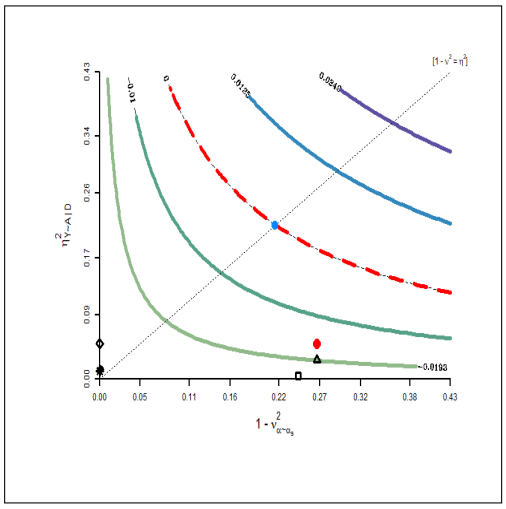


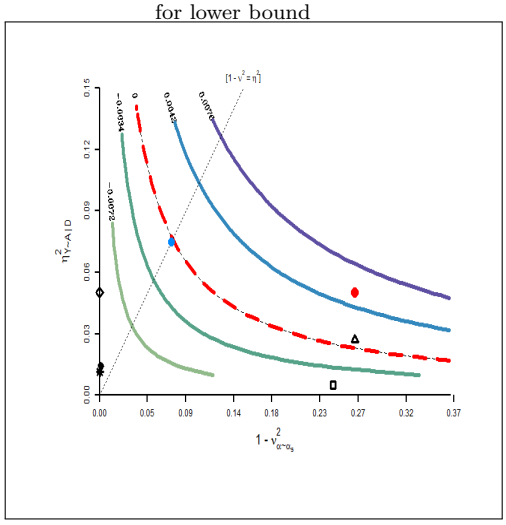
Figure 18: **Wife doesn't participate in decision making (husband's response): Contour plots based on Chernozhukov et al. (2024)**

ATE: using RA results

(A) Lower bound

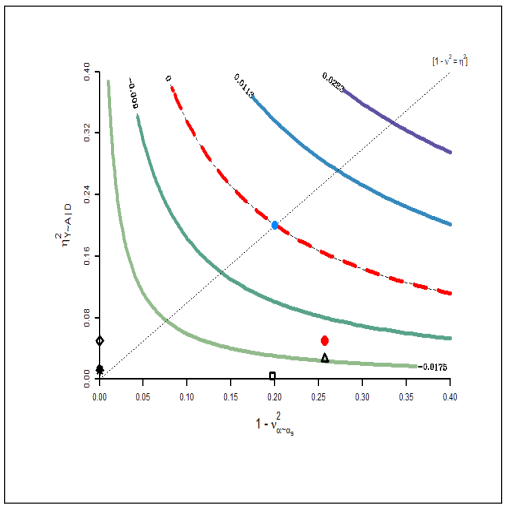


(B) Upper 95 % confidence bound for lower bound

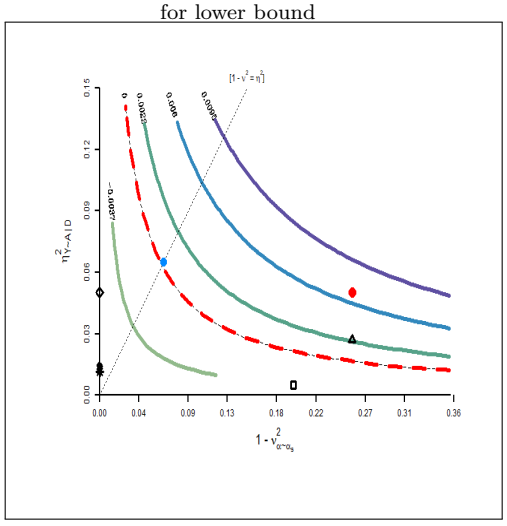


ATET: using RA results

(A) Lower bound



(B) Upper 95 % confidence bound for lower bound



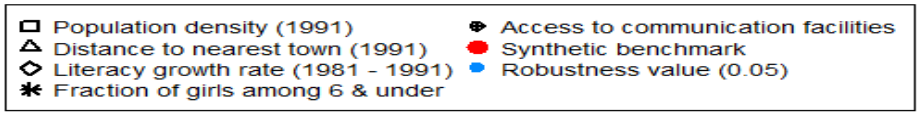
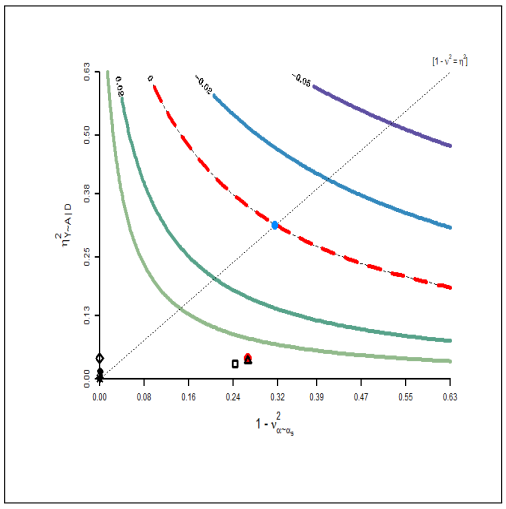


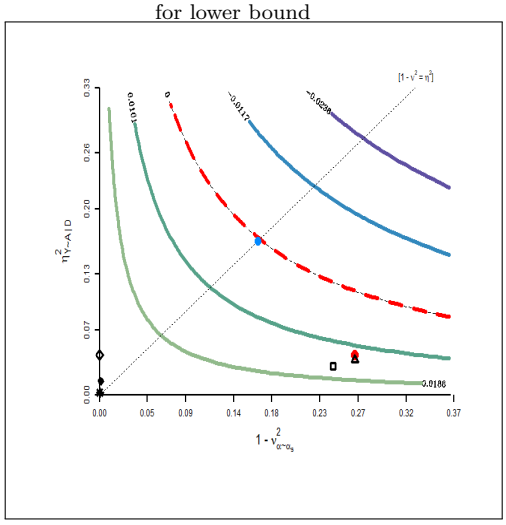
Figure 19: Report acts of domestic violence (husband's response): Contour plots based on Chernozhukov et al. (2024)

ATE: using RA results

(A) Lower bound

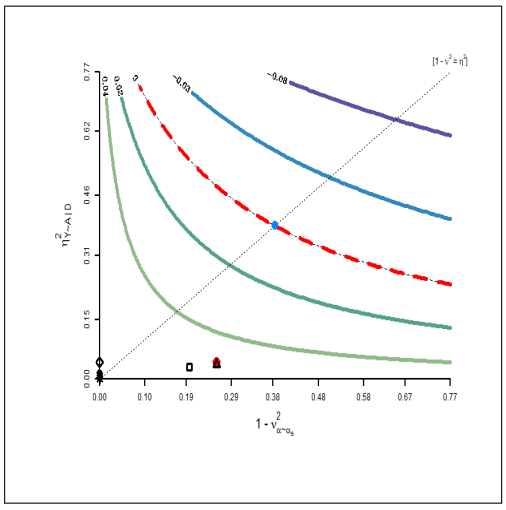


(B) Lower 95 % confidence bound for lower bound

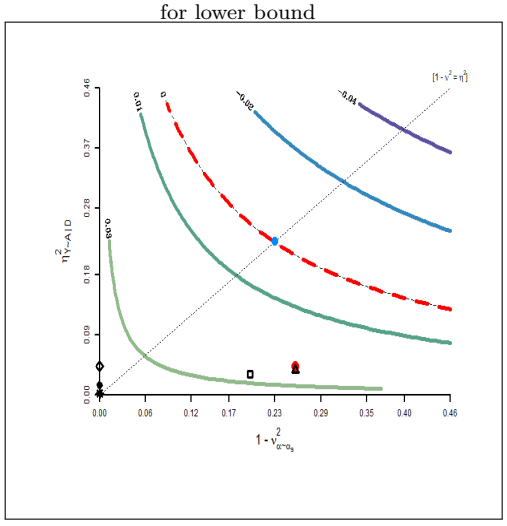


ATET: using RA results

(A) Lower bound



(B) Lower 95 % confidence bound for lower bound



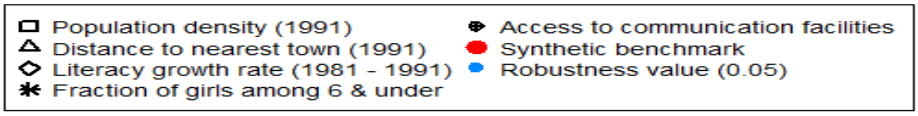
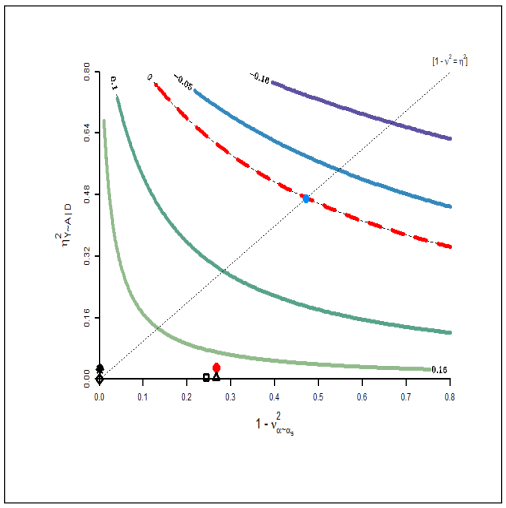


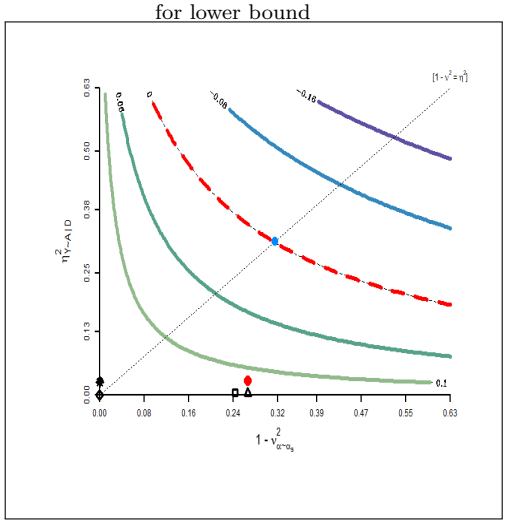
Figure 20: **Demolish illegal liquor shops (husband's response): Contour plots based on Chernozhukov et al. (2024)**

ATE: using RA results

(A) Lower bound

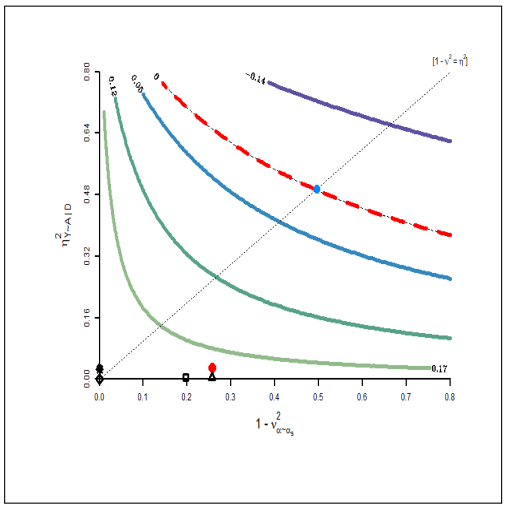


(B) Lower 95 % confidence bound for lower bound

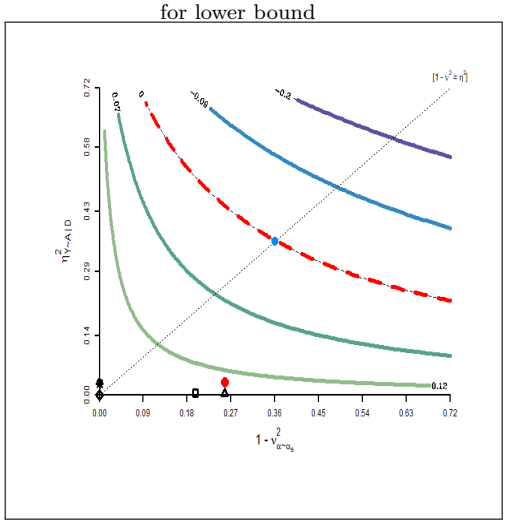


ATET: using RA results

(A) Lower bound



(B) Lower 95 % confidence bound for lower bound



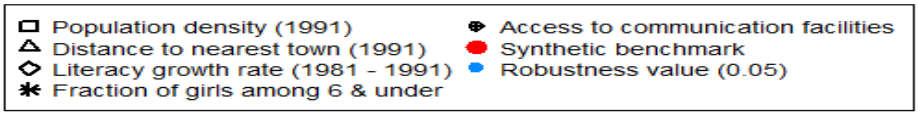
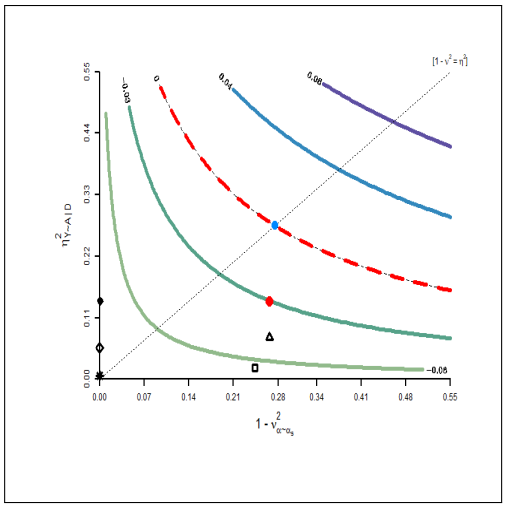


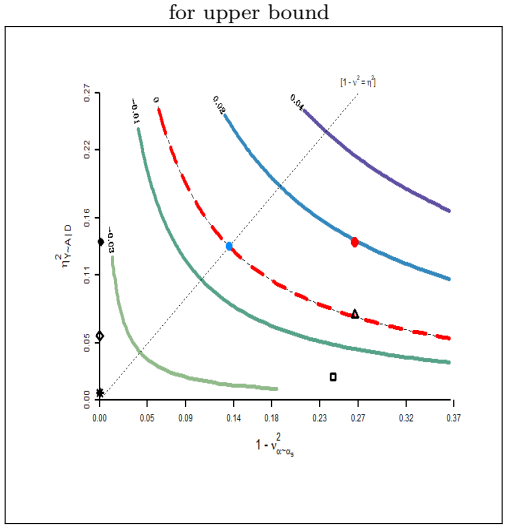
Figure 21: **Physical abuse of wife: Contour plots based on Chernozhukov et al. (2024)**

ATE: using IPW results

(A) Upper bound

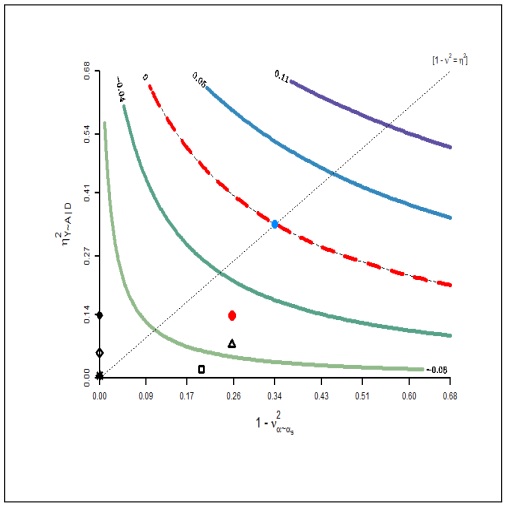


(B) Upper 95 % confidence bound for upper bound

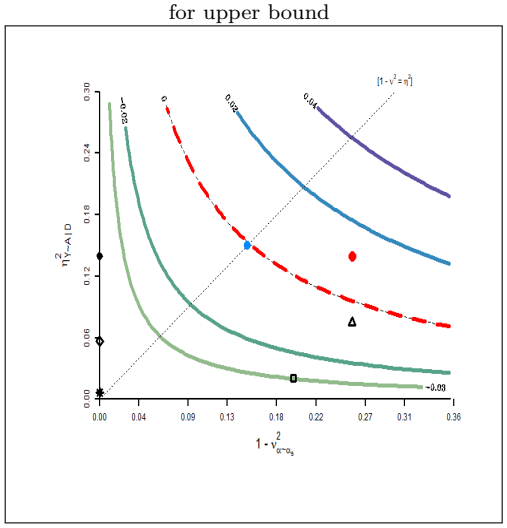


ATET: using IPW results

(A) Upper bound



(B) Upper 95 % confidence bound for upper bound



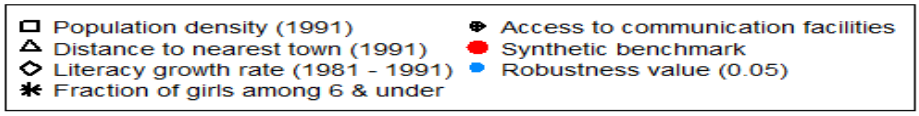
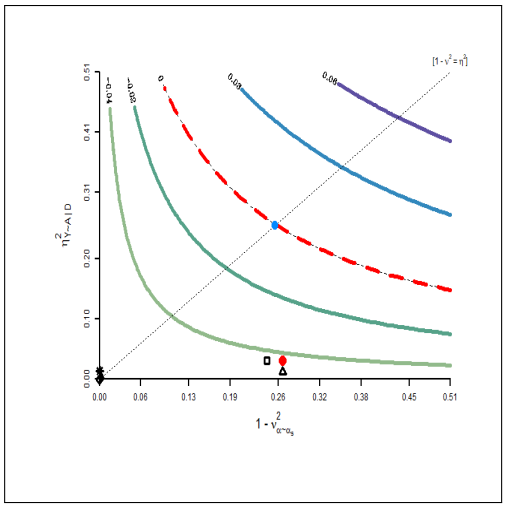


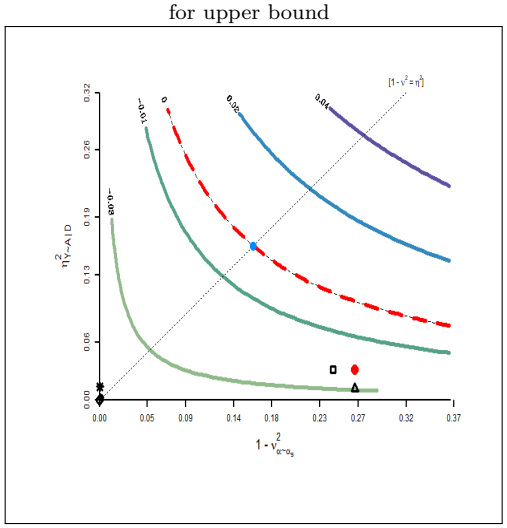
Figure 22: Sexual abuse of wife: Contour plots based on Chernozhukov et al. (2024)

ATE: using IPW results

(A) Upper bound

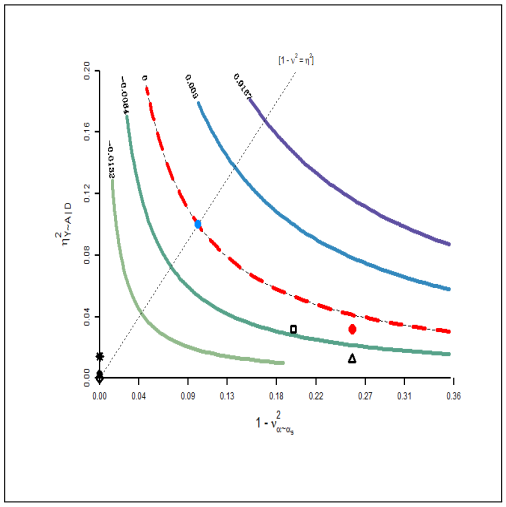


(B) Upper 95 % confidence bound for upper bound

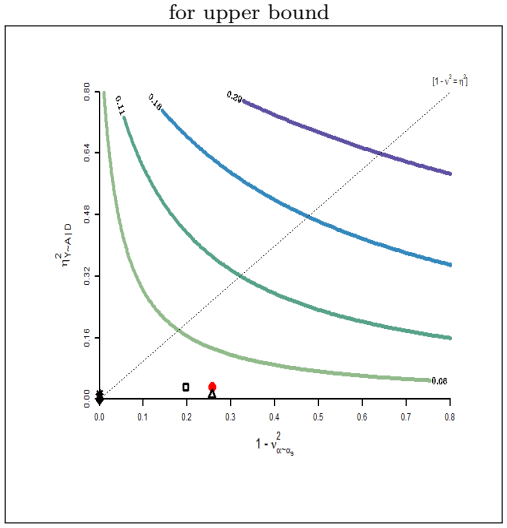


ATET: using IPW results

(A) Upper bound



(B) Upper 95 % confidence bound for upper bound



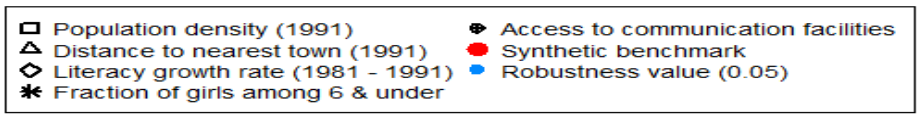
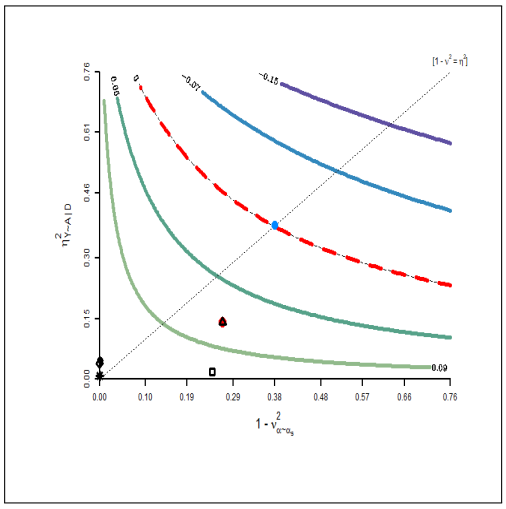


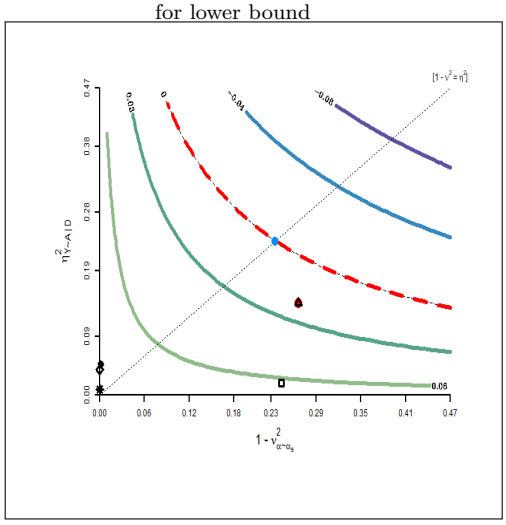
Figure 23: **Abuse-free relationship: Contour plots based on Chernozhukov et al. (2024)**

ATE: using IPW results

(A) Lower bound

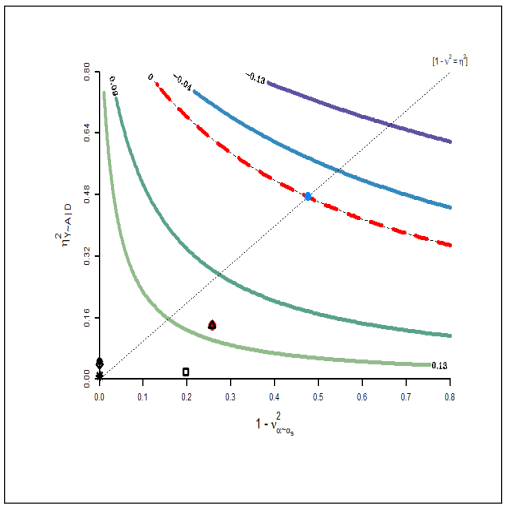


(B) Lower 95 % confidence bound for lower bound

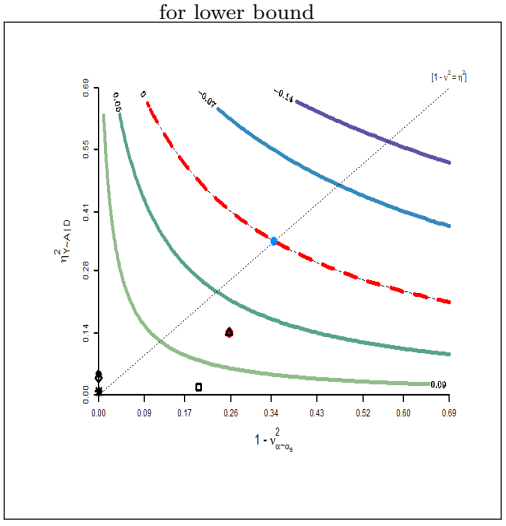


ATET: using IPW results

(A) Lower bound



(B) Lower 95 % confidence bound for lower bound

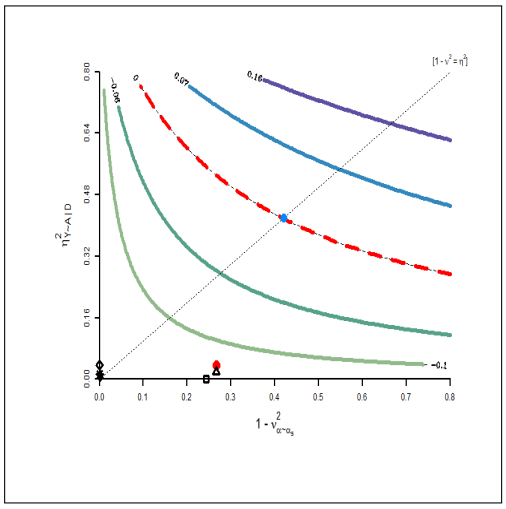


Population density (1991)
Distance to nearest town (1991)
Literacy growth rate (1981 - 1991)
Fraction of girls among 6 & under
Access to communication facilities
Synthetic benchmark
Robustness value (0.05)

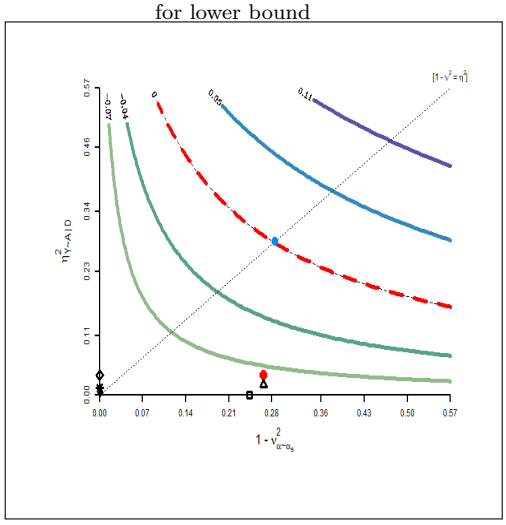
Figure 24: **No knowledge of law against domestic violence (wife's response): Contour plots based on Chernozhukov et al. (2024)**

ATE: using IPW results

(A) Lower bound

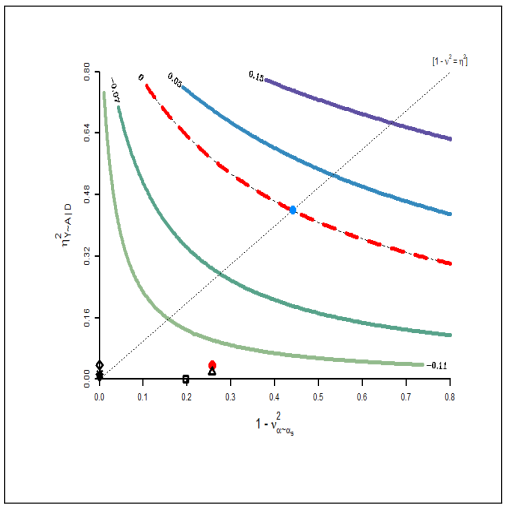


(B) Upper 95 % confidence bound for lower bound

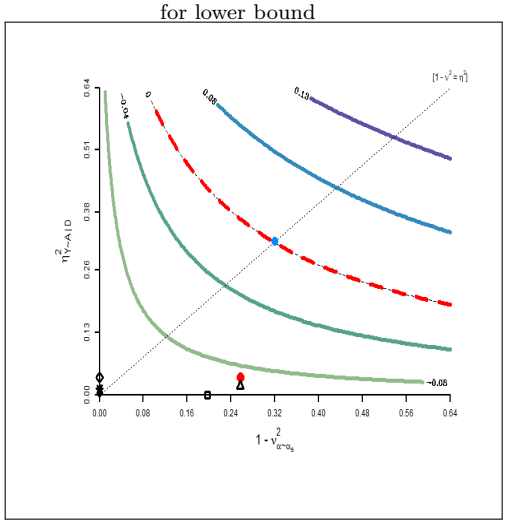


ATET: using IPW results

(A) Lower bound



(B) Upper 95 % confidence bound for lower bound



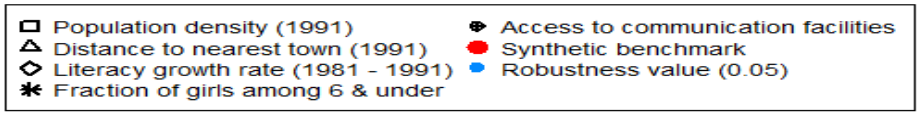
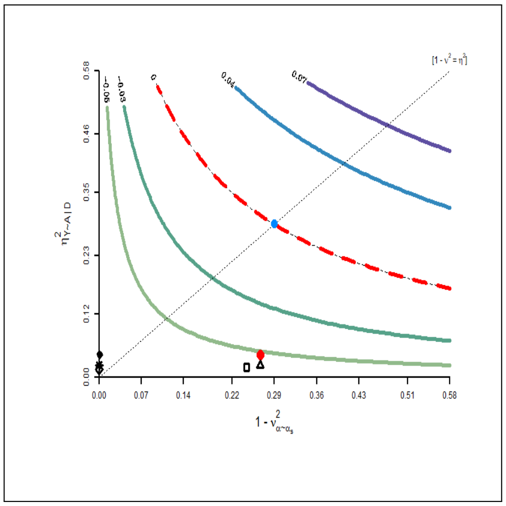


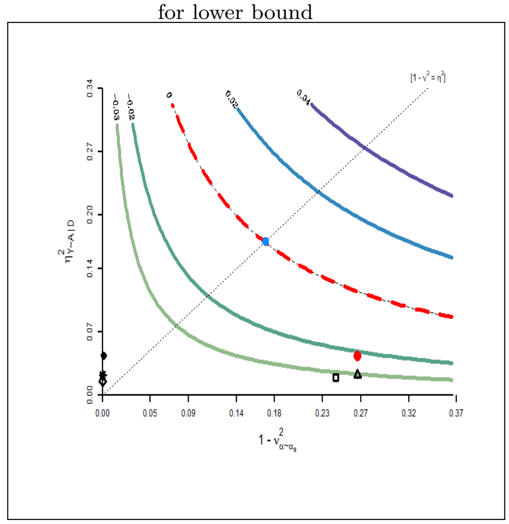
Figure 25: Wife doesn't participate in decision making (wife's response): Contour plots based on Chernozhukov et al. (2024)

ATE: using IPW results

(A) Lower bound

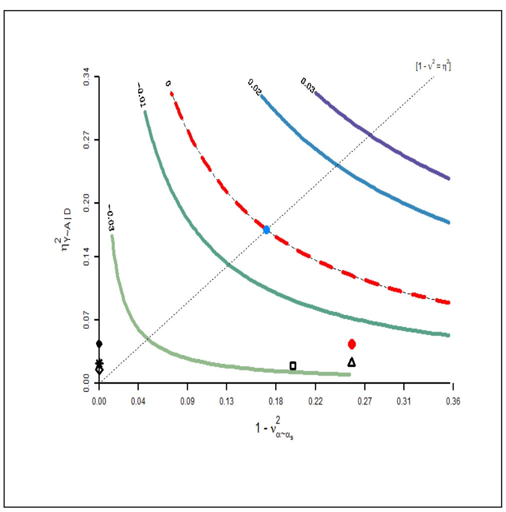


(B) Upper 95 % confidence bound for lower bound

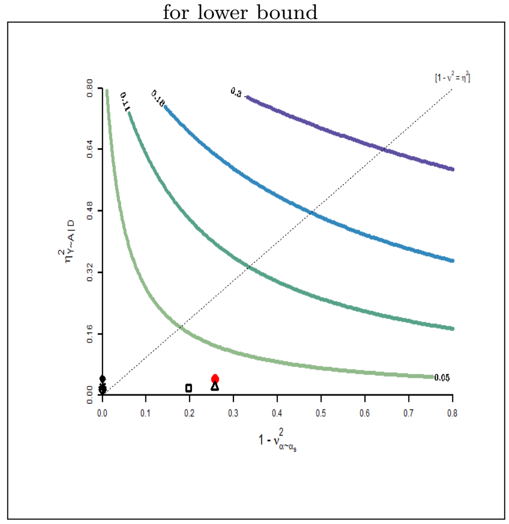


ATET: using IPW results

(A) Lower bound



(B) Upper 95 % confidence bound for lower bound



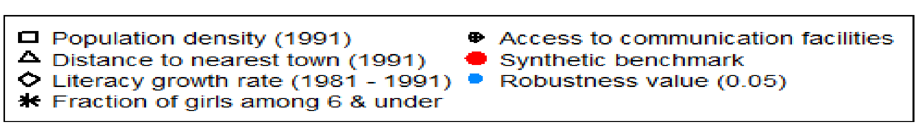
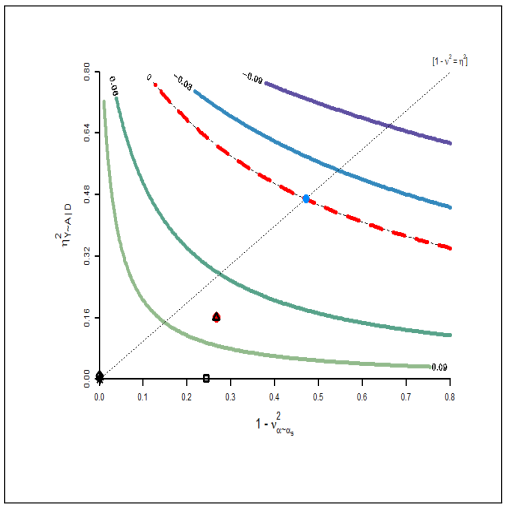


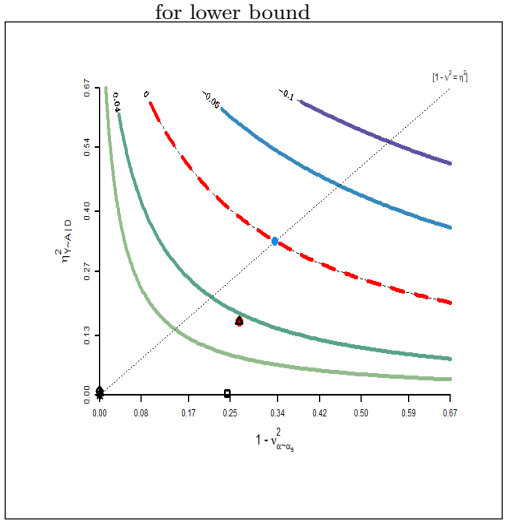
Figure 26: **Report acts of domestic violence (wife's response): Contour plots based on Chernozhukov et al. (2024)**

ATE: using IPW results

(A) Lower bound

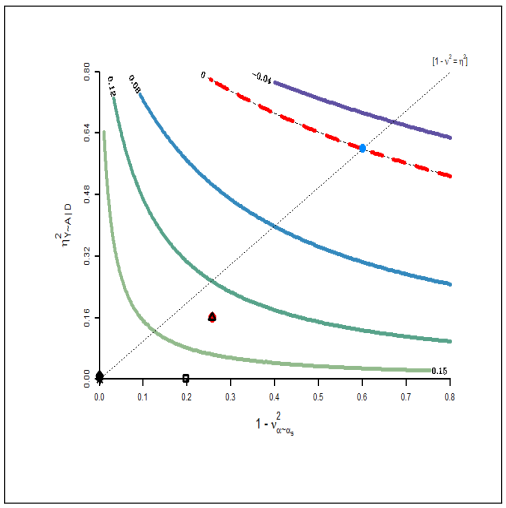


(B) Lower 95 % confidence bound for lower bound

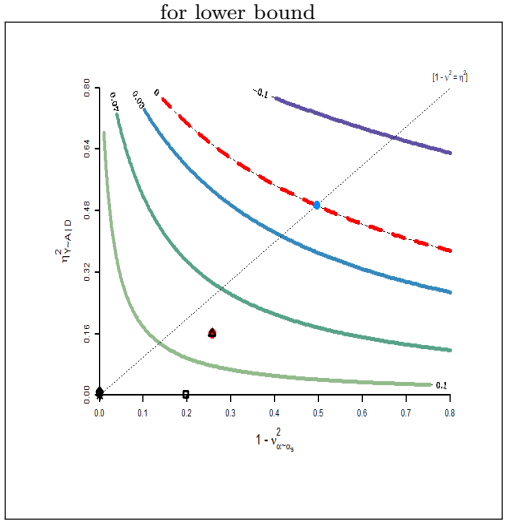


ATET: using IPW results

(A) Lower bound



(B) Lower 95 % confidence bound for lower bound

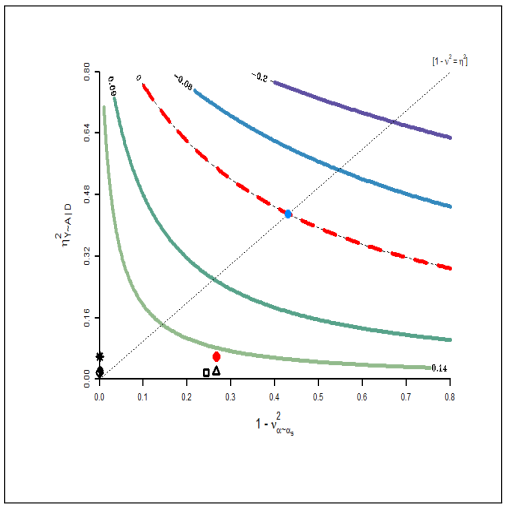


Population density (1991)
Distance to nearest town (1991)
Literacy growth rate (1981 - 1991)
Fraction of girls among 6 & under
Access to communication facilities
Synthetic benchmark
Robustness value (0.05)

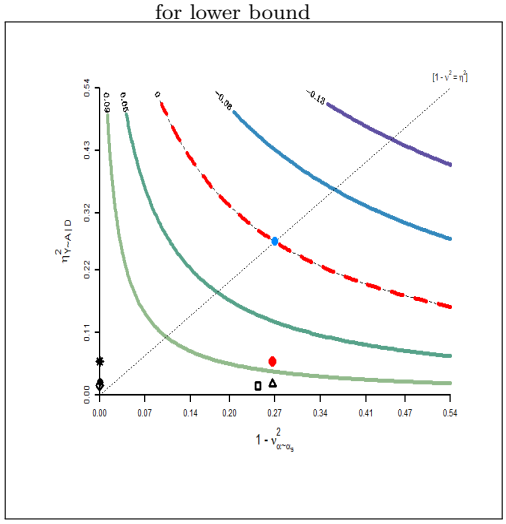
Figure 27: Demolish illegal liquor shops (wife's response): Contour plots based on Chernozhukov et al. (2024)

ATE: using IPW results

(A) Lower bound

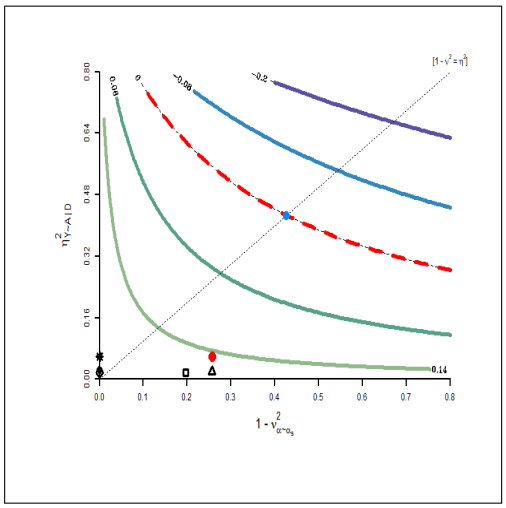


(B) Lower 95 % confidence bound for lower bound

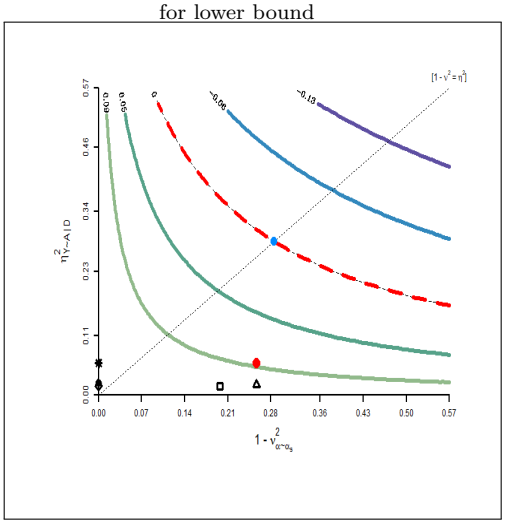


ATET: using IPW results

(A) Lower bound



(B) Lower 95 % confidence bound for lower bound



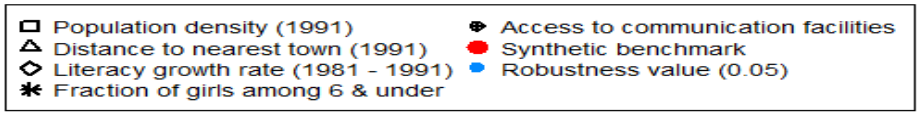
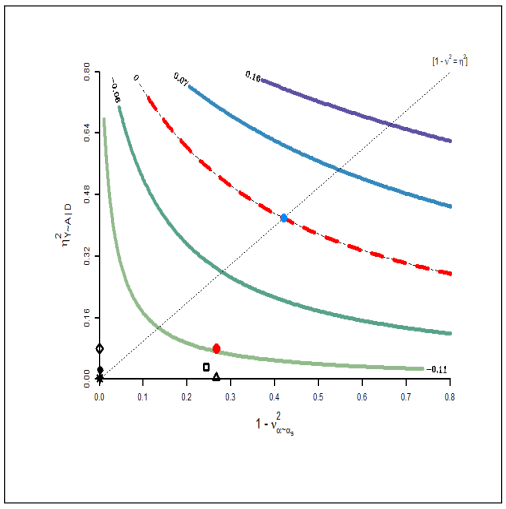


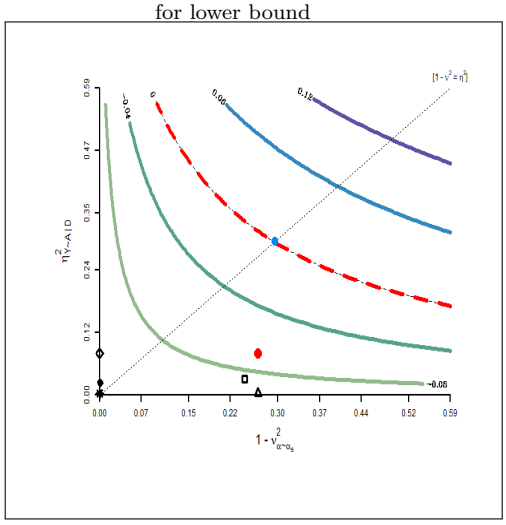
Figure 28: **Wife beating is justified (husband's response): Contour plots based on Chernozhukov et al. (2024)**

ATE: using IPW results

(A) Lower bound

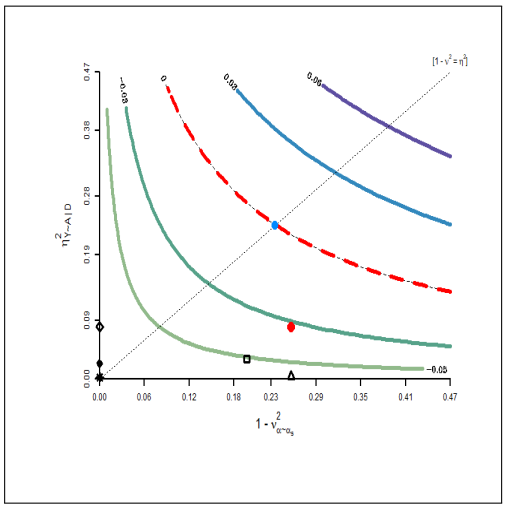


(B) Upper 95 % confidence bound for lower bound

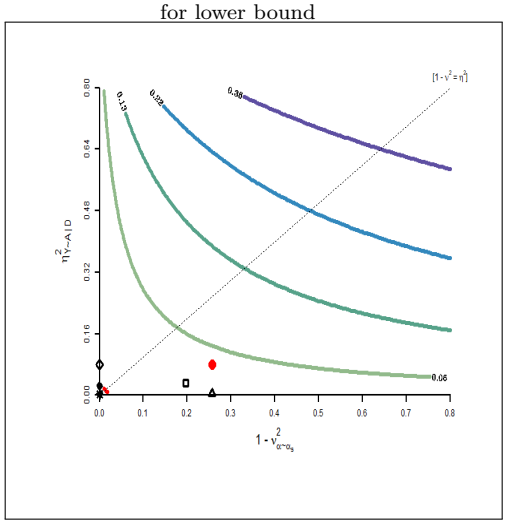


ATET: using IPW results

(A) Lower bound



(B) Upper 95 % confidence bound for lower bound

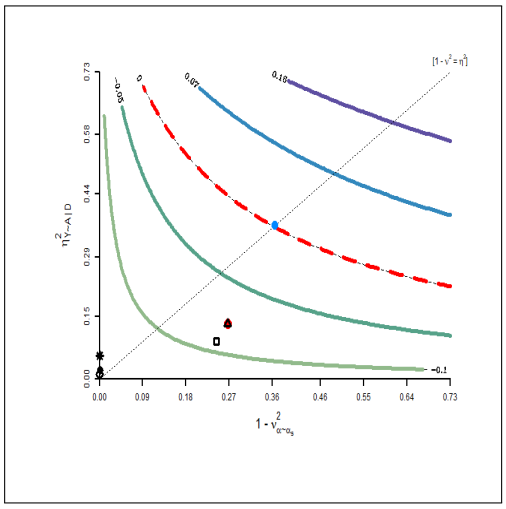


Population density (1991)
Distance to nearest town (1991)
Literacy growth rate (1981 - 1991)
Fraction of girls among 6 & under
Access to communication facilities
Synthetic benchmark
Robustness value (0.05)

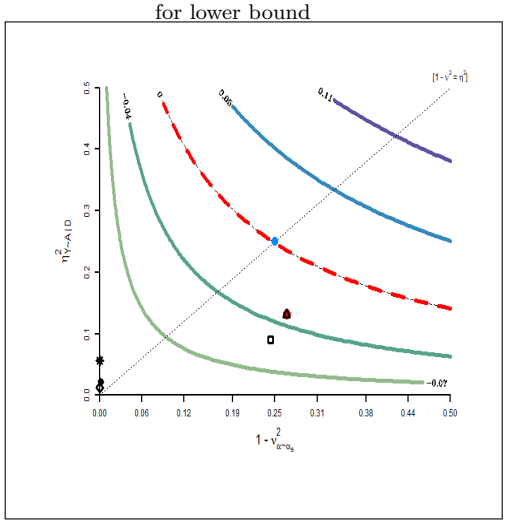
Figure 29: No knowledge of law against domestic violence (husband's response): Contour plots based on Chernozhukov et al. (2024)

ATE: using IPW results

(A) Lower bound

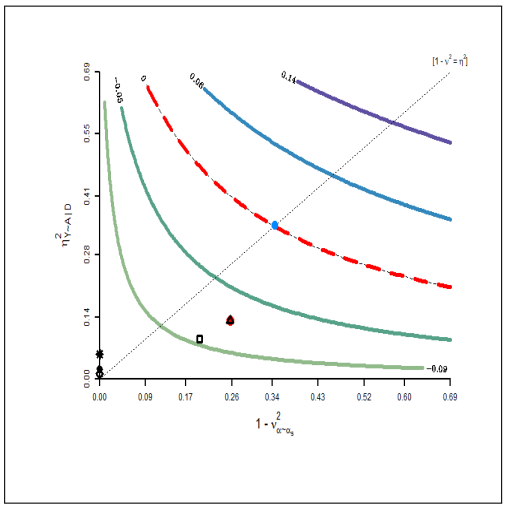


(B) Upper 95 % confidence bound for lower bound

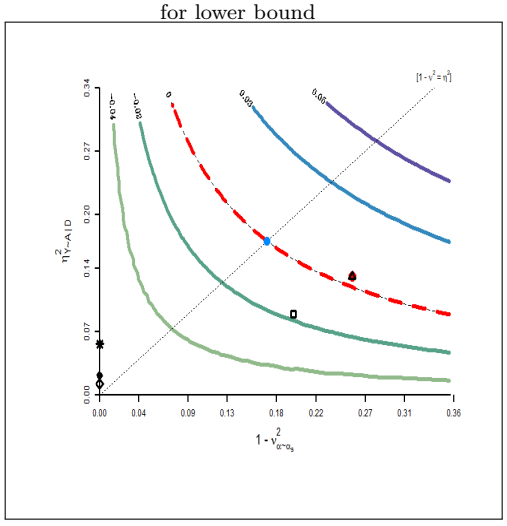


ATET: using IPW results

(A) Lower bound



(B) Upper 95 % confidence bound for lower bound



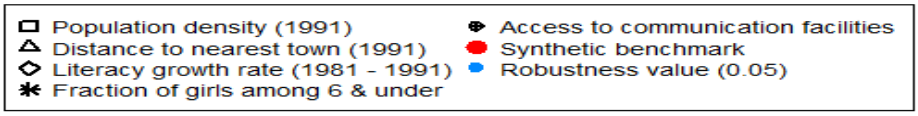
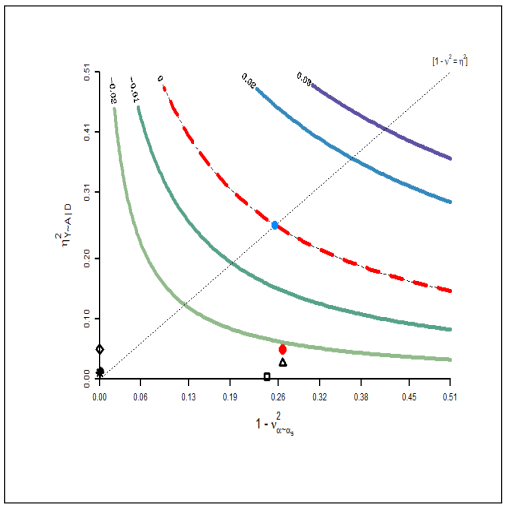


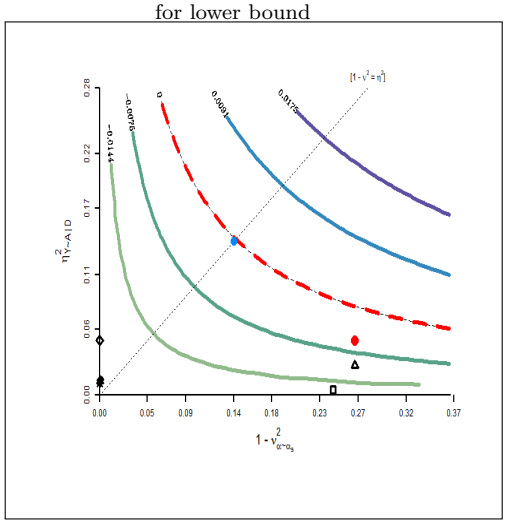
Figure 30: **Wife doesn't participate in decision making (husband's response): Contour plots based on Chernozhukov et al. (2024)**

ATE: using IPW results

(A) Lower bound

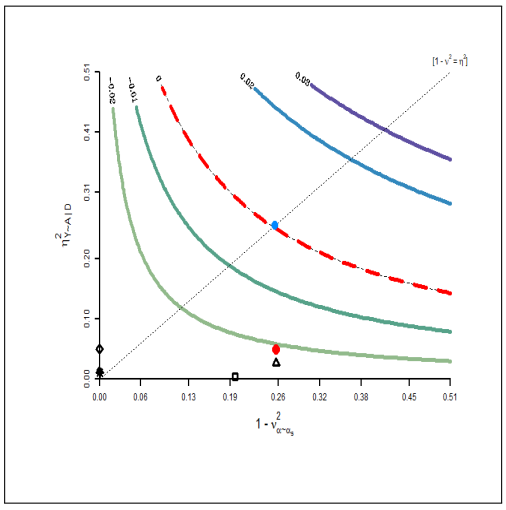


(B) Upper 95 % confidence bound for lower bound

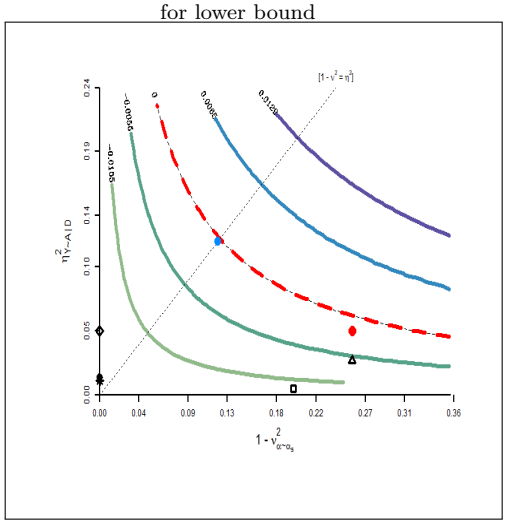


ATET: using IPW results

(A) Lower bound



(B) Upper 95 % confidence bound for lower bound



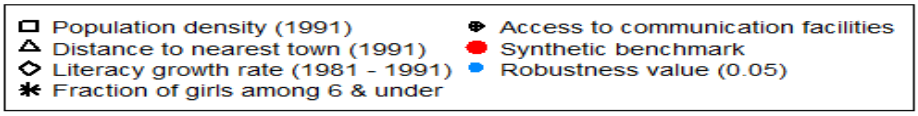
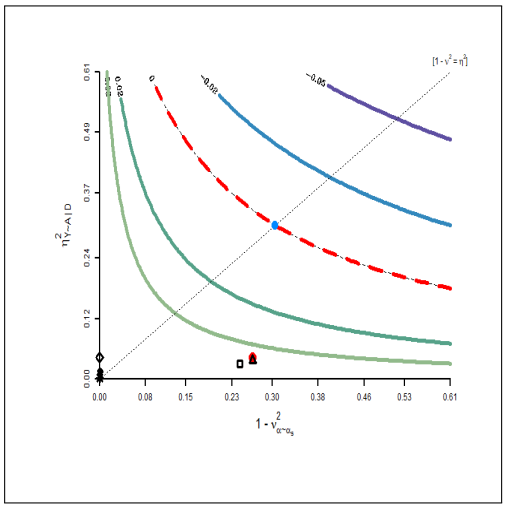


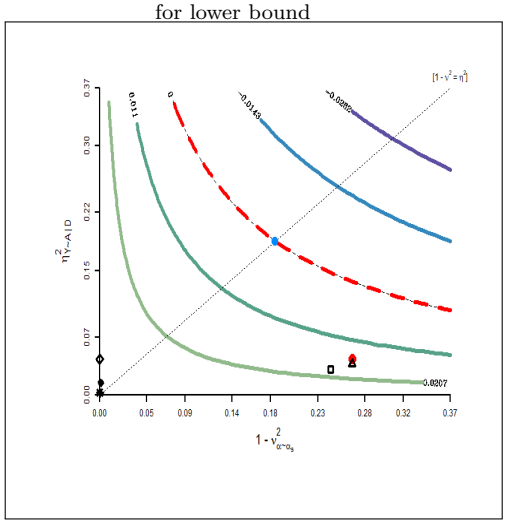
Figure 31: Report acts of domestic violence (husband's response): Contour plots based on Chernozhukov et al. (2024)

ATE: using IPW results

(A) Lower bound

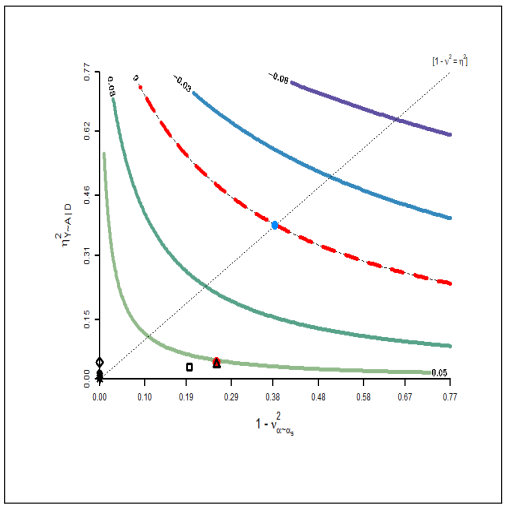


(B) Lower 95 % confidence bound for lower bound

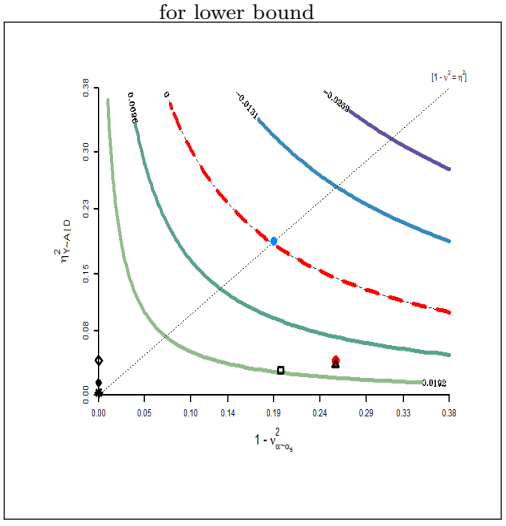


ATET: using IPW results

(A) Lower bound



(B) Lower 95 % confidence bound for lower bound



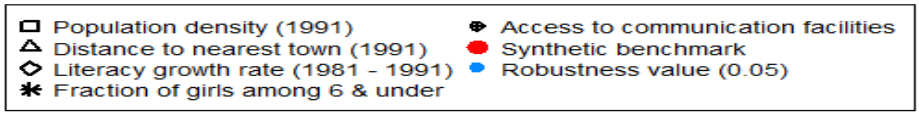
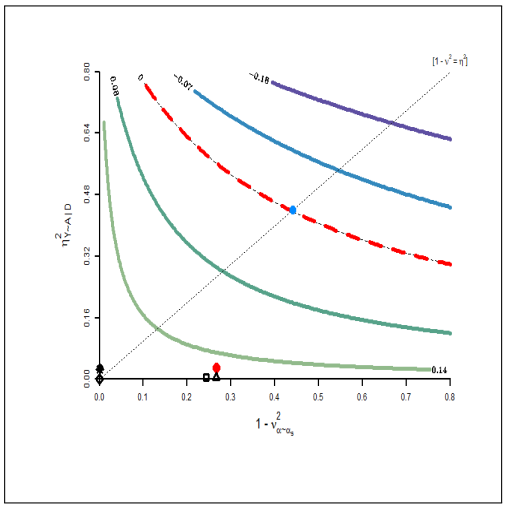


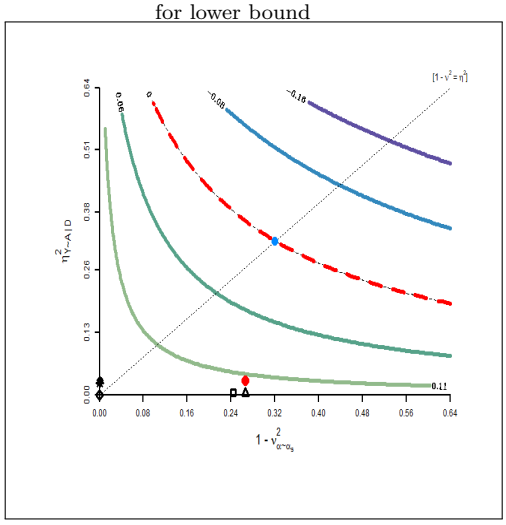
Figure 32: Demolish illegal liquor shops (husband's response): Contour plots based on Chernozhukov et al. (2024)

ATE: using IPW results

(A) Lower bound

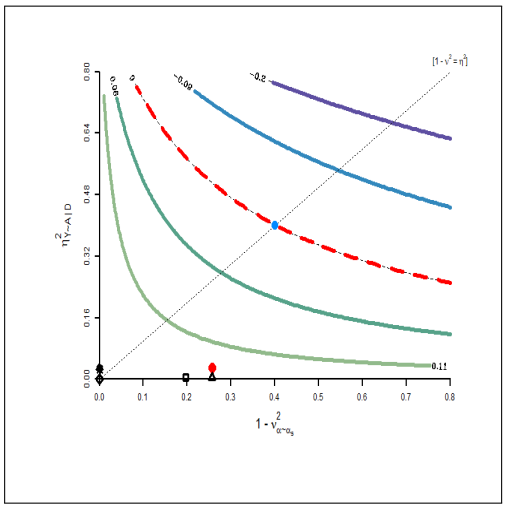


(B) Lower 95 % confidence bound for lower bound

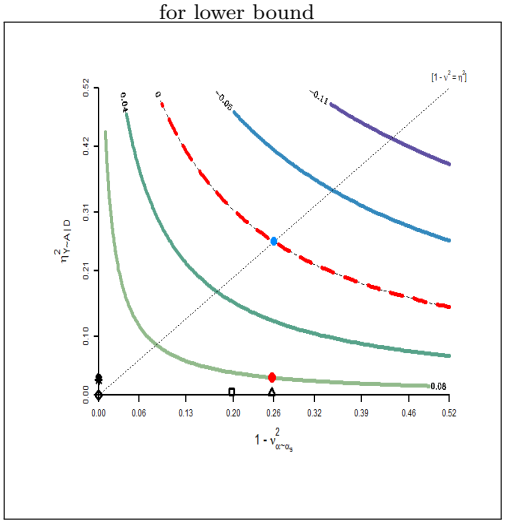


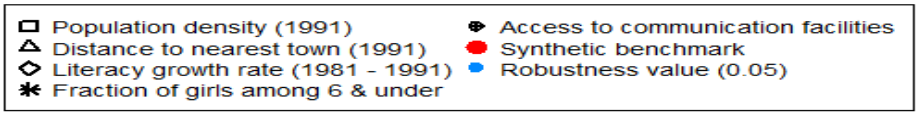
ATET: using IPW results

(A) Lower bound



(B) Lower 95 % confidence bound for lower bound





C Synopses of Four Jana Sanskriti plays

Shonar Meye (Golden Girl)

Shonar Meye dramatizes the plight of many young women in India. It portrays the life of a girl from childhood to adulthood. The writing of *Shonar Meye* was the result of 5 years of intensive work in remote areas of the Ganges Delta in West Bengal. At this time, *JS* conducted theater workshops with many groups of villagers. The play is based on the discussions at these workshops in which individuals shared their experiences of oppression and their daily challenges. Though written almost 15 years ago, it remains an extremely relevant play at the time of this writing.

In the play, Ram babu is a middle-class villager who lives with his wife, son, and daughter. Ram babu favors his son over his daughter because he believes that his son will look after him and his wife in their old age but his daughter will get married and leave for her in-laws' home. His daughter wants to study, but her family will not support her education. The family makes her spend most of her time on daily household chores. Before marriage, prospective in-laws inspect the girl to check whether she is physically suitable to marry their son. The girl passes the inspection, and Ram babu arranges for his daughter to marry the handsome son of a well-to-do family. The groom's family demand a dowry of 10,000 rupees and 110 grams of gold. They ask Ram babu to arrange the dowry by the time of the marriage. Ram babu decides to sell his land and take a bank loan for the marriage, but fails to pay the dowry by the time of the marriage. The groom's father threatens Ram babu and his family that they will not be able to see their daughter again until he satisfies all the dowry demands. The daughter faces the wrath of her husband's parents because of her father's inability to meet the dowry demands. She has to work very hard. If she makes even a small mistake, she is beaten. The play ends when the daughter confronts her oppressors.

Ekti Meyer Kahini (Story of a Girl)

This is another play that depicts the different stages in the lives of women: the period before marriage, the arrangement of the marriage, and life after marriage. The first part of the play highlights gender inequality. The protagonist, Sankari, is a 13-14 years old daughter of a poor agricultural worker. She has an elder brother who is married. Sankari wants to study, but because of her family's poverty she is unable to do so. Her brother and sister-in-law want her to get married and leave for her in-laws' house as soon as possible. The second part of the play showcases dowry-related problems and the lack of choice that Sankari has about when and whom to marry. Sankari's father pays a hefty dowry at the time of her marriage. The last part of the play focuses on the ill effects of early marriage on young girls (13-14 years old) and how the central characters of the family into which a young girl marries (the mother-in-law and husband) can become tools of oppression. As punishment for mistakes she made doing household chores, Sankari is sent back to her natal family to bring money to meet the medical expenses that her in-laws paid for when she fell sick. At her father's house, the situation is no better for Sankari: her brother and sister-in-law harass her and put pressure on her and her father for her to go back to her in-laws' house. Sankari knows that if she returns to her husband's home without the money her in-laws demand, she will be killed. The play ends as she sees her dilemma – whether to return to her husband's home or stay at her father's home and try and earn a living for herself.

Hay Re Mod (The Curse of Alcohol)

In early 2005, *JS* organized a sit-in protest against the illegal production and sale of liquor in the region in West Bengal where the troupes perform. Villagers blocked the main highway that connects Kolkata to the Ganges Delta. This was the start of an anti-liquor campaign. There is a strong nexus between politicians, illegal liquor shop owners, local government officials, and the police. Prasad Sarkar, one of the protesters, explained that the cause of the protest police corruption: "You are spineless policemen. You find

our work illegal [i.e., blocking the highway], and you don't notice (chokhe pore na) the illegal production of liquor because it is in your self-interest" (Da Costa, 2010). Women and adolescent children bear the brunt of the consequences of alcohol abuse in increased domestic violence and children forced to drop out of school because their families are short of funds. The play *Hay Re Mod* (The Curse of Alcohol) presents these problems through the narrative of a woman named Naina. She has two school-going sons and two married brothers-in-law. Naina's husband is a drunkard. He spends all his earnings on alcohol and contributes nothing to run the household or to buy books and school supplies for their sons. From time to time, she has to borrow from her neighbors and do odd jobs outside her home to support the household. When Naina asks him for money, he becomes violent and beats her. Naina complains to the head of the village government (the Panchayat). She even goes to the police station to register a complaint. Naina laments that the police "take bribes behind the scenes (*pechon theke ghoosh*), that is why they cannot find a solution to our problems" (Da Costa 2010). The husband's brothers, too, plead with the husband to stop drinking, but their efforts fail. The play ends with the brothers-in-law moving out of the house and Naina taking her husband to a barely functional rural hospital to get him treated for nausea, abdominal pain, and dizziness – all effects of consuming illicit liquor.

Reference

Da Costa, Dia. 2010. "Have they Disabled Us? Liquor Production and Grammars of Material Distress" in *Development Dramas. Reimagining Rural Political Action in Eastern India.* Routledge: Delhi.

Int Bhata (The Brick Factory)

In *Int Bhata*, the brick-factory owner promises the workers overtime pay to persuade them to work late to complete a large sales order for bricks. When they have completed the order, he refuses to pay over-time wages. As he points out, he doesn't need to: "There are

so many workers like you roaming around for work." Later in the evening, the owner comes to the home of one of the workers, Phulmoni, to demand sex. When she refuses to continue the sexual relationship that she has had with him in exchange for loans to her husband, the factory owner threatens to have her husband jailed if he does not repay the loans that very evening. She gives in to the factory owner's demand for sex. Her husband comes home and discovers her in the arms of the factory owner. In the next scene, the villagers find her guilty of dishonoring it by adultery and cane her as punishment.

At the end of the play, two actors speak to the audience:

First Person: "Hunger caused Phulmoni to go to the city for work. Taking advantage of her poverty, the owner forced himself on her. Phulmoni was judged guilty."

Second Person: "But the owner is the guilty one. Who will punish him?"

In this drama, the belief that the adulterous wife violated her marital obligations is misguided. The drama is designed to evoke collective representations of the role of a wife and to convince the viewers that there is something wrong with the representations in particular, they lead to oppression.

D Sampling Procedure

At the time of our survey in 2014, *JS* used to regularly perform in 125 villages in three administrative blocks in the South 24 Parganas district of the state of West Bengal in India. These villages are the treated villages. To avoid contaminating our estimates by spillovers, we decided to sample the control villages not from these same three blocks, but from three neighboring administrative blocks. Our sampling strategy was as follows.

- From each of the 3 control blocks and 3 treatment blocks, we randomly sampled between 1 and 7 Gram Panchayats (GPs).[35]

[35] A GP is the lowest unit of rural government. Each GP normally includes several villages (called "census villages" in the Census of India).

- From each control GP, we sampled census villages with probabilities proportional to the 1991 population.

- From each treated GP, we randomly sampled census villages from a list provided by JS of the villages where it had regularly performed since 2004.

- From each census village, we randomly sampled either one or two polling booths from the 2014 electoral list.

- From the electoral list for each selected polling booth, we randomly sampled 15-35 households in each control village, and 35 households in each treatment village. We used the electoral list because a voting card is a proof of identity held by most residents of at least 18 years of age (the minimum voting age in India).

- The field workers were charged with the task of interviewing one married couple in each household. Each investigator was given details (name, gender, age, and the household head's or husband's name) of the members of households to be interviewed and a list of possible replacement households if no eligible married woman was present and willing to participate in the survey. With the help of a family member, on arrival at the household, investigators determined whether an eligible married woman was present. If more than one such woman was present, the investigator-team randomly chose one. In only rare cases did an eligible household member refuse to cooperate.

- The team sought to interview the husband of the selected female respondent wherever he may have been at the time of the wife's interview. The team achieved this in 99% of the cases. For the remaining 1% of the married women, another married man of the same or neighboring household was interviewed. The tables in this paper report data only for married couples.

E Survey Questions and Instructions to the Field Investigators

Although the questions on the incidence of abuse, i.e., the ones immediately below, were only asked toward the end of the interview (and in the absence of the husband) to help the female respondent feel comfortable responding to them, we present here the questions on the various outcomes in the same order that they appear in our description in Section 4.

Questions on abuse, asked to the female respondents only

I am going to ask you some questions about your relationship with your husband. Does your husband:		
	Yes	No
1. *Emotional abuse*:		
a. Say or do something to humiliate you in front of others?	1	2
b. Threaten to hurt or harm you or someone close to you?	1	2
c. Insult you or make you feel bad about yourself?	1	2
2. Physical abuse:		
a.Throw something at you? Twist your arm or pull your hair? Slap you?	1	2
b. Punch you with his fist or something that could hurt you? Try to choke you or burn you? Threaten or attack you with a knife or a gun or any other weapon?	1	2
3. Sexual abuse:		
a. Physically force you to have sexual intercourse with him even when you did not want to?	1	2

Instructions to field investigators on how to ask these questions:

1. a. Suppose your husband does not like the food that you have cooked or he is unhappy about something that you have done. Say, when you go outside the house to fetch water, someone from your parents" house calls you. You receive the call and talk to the person for a few minutes. Observing this, does your husband get annoyed? Does your husband scold you publicly in the presence of your friends and neighbors?

   b. Does your husband threaten you or somebody close to you? Suppose you have burnt the food because you were attending to your child who was crying. Does your

husband threaten you by saying that if you repeat this (burn the food), I will accost your brother when I meet him in the market and physically harm him?

c. Does your husband tell you that you are useless, illiterate, and ignorant? Does he compare you with your neighbor's wife, saying that not only is she beautiful, but she also runs her household efficiently, helps her children with their homework, and earns money by making hand-made dolls during her spare time?

2. a. & b. Does your husband physically abuse you? Suppose your husband returns home after working in the fields, takes a shower, and sits down for his afternoon meal. The food is not yet ready. He throws whatever he finds within his reach at you. If the meal is too hot, he pulls your hair and slaps you. This is one instance when the husband physically abuses his wife. The husband could also be frustrated that his wife did not bring adequate dowry during their marriage. This frustration is reignited when his friends discuss the huge amount of dowry that another friend of theirs got. Does your husband take out his frustration by physically abusing you? Does the physical abuse take extremely violent forms like your husband strangling you or burning you by pushing you towards the open flame?

3. (The field investigators are trained to modulate their voice to a very low pitch and whisper) Sister, I am going to ask you a personal question concerning your sexual relationship with your husband. Please do not take it otherwise. We are only trying to assess whether you face such problems or have faced them in the past. Suppose one day, you are feeling feverish or exhausted because your child has been crying the whole day and you had a few relatives of your husband over for a meal. At night, your husband may want to indulge in sexual activities with you, but you are not enthusiastic about it. You express your feelings to your husband. Is it the case that in spite of your unwillingness to indulge in sexual activities, you husband tries to force you?

Questions on attitudes towards wife beating

In your opinion, is a husband justified in hitting or beating his wife in the following situations?			
	Yes	No	DK
a. If she goes out without telling him	1	2	9
b. If she neglects the house or the children	1	2	9
c. If she argues with him	1	2	9
d. If she refuses to have sex with him	1	2	9
e. If she shows disrespect to her in-laws	1	2	9
f. If he suspects her of being unfaithful	1	2	9
g. If she is does not have a male child	1	2	9

Instructions given to the field investigators on how to ask these questions:

1. Suppose while her husband is at work, a woman finds that she is missing an essential ingredient for cooking that is available in the neighborhood store. Or suppose that she gets news that her mother, who lives in a neighboring village, has fallen down and sprained her ankle. She goes to the store / visits her parental home. When her husband returns home, she informs him that she has gone to the store/visited her natal home. Suppose that, on hearing this, the husband angrily questions her: Why did she go out of the house without asking him? Why hadn't she arranged in advance to have all essential ingredients at home? Since her father and brother are living in her natal home, why did she have to rush to attend to her mother? If the annoyance expressed by the husband leads to him to assault his wife, is he justified?

2. A couple have 3 (replace with the correct number) children. There is a lot of housework that the wife needs to complete before her husband returns from the fields at lunchtime. The wife has bathed her children and dressed them in fresh clothes. After some time, she finds them playing in the dirt. While the wife is finishing her cooking, she finds that her infant child has started to cry; or that her children are quarreling and crying. Her husband comes home and blames her. He accuses her of neglecting the household and the children. If the annoyance expressed by the husband leads him to assault his wife, is he justified?

3. Suppose a wife fails to serve a meal on time because she had to prepare her children for school, clean the house, and complete the cooking. Her husband returns home from work (in the fields) and finds that his wife has not finished cooking. The wife tries to explain to her husband why she had been delayed in cooking the meal on that day. But he does not listen and accuses her of neglecting her household chores. He blames her parents for not teaching her how to run a household efficiently etc. This leads to an argument between the two and finally the husband assaults his wife. Is this action by the husband justified? (voice ofthe investigator should be very soft and sympathetic while asking this question)

4. Suppose one day, a wife is feeling feverish or exhausted because her child has been crying the whole day and she also had to serve a lavish meal to a few relatives of her husband. At night, her husband may want to indulge in sexual activities with his wife, but she denies him. She explains her reasons to her husband. If the husband gets angry with his wife for denying him his rights and assaults her, is he justified?

5. Suppose a woman's child is crying and her mother-in-law accuses her that she does not take care of her child properly, does not feed the child on time, etc. Due to incessant crying of the child, the wife's frustration builds up and she talks back to her mother-in-law. The mother-in-law takes this as an insult and complains to her son when he returns home during lunchtime. He gets angry with his wife for insulting his mother and physically assaults her. Is he justified in his actions?

6. Suppose a wife meets a male relative/acquaintance from her natal village when going to the market (or suppose she discusses with her child's male teacher her child inattentiveness towards studies). Whenever her husband sees his wife talking with another man, he grows suspicious that she is cheating on him. Because of his suspicions, he scolds her, verbally abuses her, and finally resorts to physical violence. Do you think the husband is justified in his actions?

7. Suppose a couple already has 3 children (replace with the correct number of children). The wife does not want any more children because she realizes that their current household financial situation will not allow them to provide the additional child with adequate nutrition and education. But her husband wants another child because they only have daughters. He wants a son to carry his lineage. The wife tries to reason with her husband. This leads to a verbal argument between the two. Finally, the husband assaults his wife in anger. Do you think the husband is justified in his actions?

Questions on decision making in the household

(No special instructions to the field investigators except for reading out the options)

I am going to ask you about who makes the decisions in the following situations					
Who decides about:	Respondent	Husband (wife)	Both	Family	Other
a. Children's education	1	2	3	4	6
b. The family's health care needs	1	2	3	4	6
c. Purchasing major household items	1	2	3	4	6
d. Purchasing minor household items	1	2	3	4	6
e. Visits to the wife's family or relatives	1	2	3	4	6
f. Marriage of your children	1	2	3	4	6
g. Number of children to have	1	2	3	4	6
h. Use of contraceptives	1	2	3	4	6

Questions on awareness about laws against domestic violence

No special instructions were given to the field investigators when asking these questions.

Please tell me if you know about the following			
	Yes	No	DK
a. Do you know that a woman can file a complaint in the police station against any male perpetrator or any female or male relatives of the husband who has perpetrated domestic violence against her	1	2	9
b. Do you know that a woman who is a victim of domestic violence can register a case if she is restrained from entering any portion of the shared household in which she resides	1	2	9
c. Do you know that a woman who is a victim of domestic violence can claim monetary compensation for medical treatment, loss of earnings, and maintenance for herself and for her children from her husband?	1	2	9

Questions on community actions

No special instructions given to the field investigators when asking these questions.

Please tell me if you would willingly participate in the actions described below			
	Yes	No	DK
a. Participate in demolishing an illegal liquor shop in your neighborhood/village?	1	2	9
b. Report to anyone an act of domestic violence that you have witnessed?	1	2	9