Can Workfare Programs Moderate Violence? Evidence from India

Thiemo Fetzer *

June 2, 2014

Abstract

Governments in conflict torn states scramble for effective policies to persistently reduce levels of violence. This paper provides evidence that a workfare program that functions as a social insurance, providing employment opportunities in times of need, may be an effective antidote to shut down an important mechanism that drives conflict. By mitigating adverse income shocks, the Indian National Rural Employment Guarantee scheme has been successful in removing the income dependence of insurgency violence and thus, contributes to persistently lower levels of violence.

Keywords: social insurance, conflict, India, insurgency

JEL Codes: D74, Q34, J65

*London School of Economics and STICERD, Houghton Street, WCA2 2AE London. Support from the Kondard-Adenauer Foundation, the Bagri Fellowship of the Asia Research Centre and the LSE-TISS-TATA grant is gratefully acknowledged. I am especially grateful to Gerard Padro-i-Miquel for his continued support. I also would like to thank Tim Besley, Erlend Berg, Jon Colmer, Mirko Draca, Jon de Quidt, Claudia Eger, Christian Fons-Rosen, Maitreesh Ghatak, Benjamin Guin, Sam Marden, Hannes Mueller, Kelly Zhang and seminar participants at LSE, CEP and NEUDC for helpful comments. Special thanks goes to Clement Imbert for kindly sharing some data.

1 Introduction

The World Development Report of 2011 sets out by pointing to the fact 1.5 billion people are living in countries that are severely affected by internal and external conflict. None of these countries has yet achieved even a single Millennium Development Goal. This highlights a well-known fact: conflict is bad. A large literature in economics has tried to assess the true social and economic cost of conflict and the many channels through which it operates, such as by deterring human capital investment (Leon (2009), Akresh and Walque (2008)), affecting time preferences (Voors et al. (2012)), affecting capital investments (Singh (2013)), diverting foreign direct investment (Abadie and Gardeazabal (2008)) or increasing trade costs (Besley et al. (2014)). While understanding the mechanisms through which conflict is costly is important, policy makers are most interested in identifying ways and means through which the dynamics and in particular, the persistence of conflict can be affected. Another long-standing literature in economics and political science has tried to understand whether economic shocks can cause conflict. Many efforts have been made to collect empirical evidence documenting a quite robust relationship between economic shocks and conflict (see e.g. Bazzi and Blattman (2013), Dal Bó and Dal Bó (2011), Dube and Vargas (2013) , Fearon and Laitin (2003) or Miguel et al. (2004)). However, there is surprisingly little work that sheds light on whether public interventions can break the link between economic shocks and conflict.

The key economic mechanism that has been identified is the opportunity cost channel (see Becker (1968), Collier and Hoeffler (1998), and Chassang and Padro-i Miquel (2009) among many). An economic shock puts downward pressure on workers' outside options, which renders joining or supporting insurgency movements incentive compatible.[1] Insurgents draw from this increased support base and are able to affect more violence.

This argument implies that any intervention that smoothes away negative shocks should contribute to weaken the link between economic shocks and conflict, through its stabilising effect on workers outside options. This paper is a contribution to the nascent literature that tries to tackle the question on whether public interventions can achieve this end.

A fundamental challenge is to find a testing ground, since conflict and a functioning state that could provide for such an intervention rarely co-exist. India, however, serves as a unique environment to study this question. Firstly, the coun-

[1] Alternatively, as Vanden Eynde (2011) argues, a low outside option induces insurgents to target civilians as the government may try to hire civilians to become police informers.

try has suffered from many low-intensity intra-state conflicts throughout its history. All these conflicts are endemic, but have all in all a relatively low intensity so that the state still functions on many dimensions. Secondly, India introduced from 2006 onwards a social insurance scheme through the National Rural Employment Guarantee Act (NREGA), that effectively serves as an insurance by providing employment opportunities in times of need. It is the biggest public employment scheme in mankind's history, currently reaching up to 47.9 million rural households annually, generating 210 million person-days of employment.[2] Hence, it is reasonable to assume that the program, due to its scale, may have an impact on the dynamics of conflict.

I make three contributions. This is the first paper to study the relationship between insurgency violence and social insurance in India. In particular, the focus of this paper is not on the levels of violence, but rather on the elasticity of violence with respect to income and how this relationship changes after the introduction of the workfare program.

The second contribution lies in studying the dynamics of rural labour markets in India and how these are affected by social security systems. In particular, I am able to comment on the pass-through of productivity shocks on agricultural wages and how this relationship is cushioned once a stable outside option offers itself to workers. This effect contrasts with a persistent rainfall dependence of agricultural production, highlighting that NREGA may be a substitute to the construction of physical infrastructure. I thus focus on the insurance value of public employment that serves as a income smoothing device and thus, may be a substitute to other forms of insurance.

The third contribution is a methodological one. This is the first paper to use a novel violence dataset that covers the whole of South Asia and has been constructed using scalable Natural Language Processing Tools (presented in Fetzer (2013)). This paper highlights the possibility to use semi-automated machine-learning routines for data cleaning and preparation in a field of economics research, where data availability is always a severe constraint.

The key findings of this paper are as follows. Before the introduction of NREGA, agricultural production, wages and violence in India were strongly rainfall dependent to the present day. This is a surprising finding, since the dependence of Monsoon rainfall should have been weakened through decades worth of investment in physical infrastructure such as damns, irrigation canals, or railroads and roads. Nevertheless, the elasticity between Monsoon rainfall and agricultural

[2]See `http://nrega.nic.in/netnrega/mpr_ht/nregampr.aspx`, accessed on 14.02.2013.

GDP estimated in this paper is actually higher than the one presented in the existing literature derived from historical data. A one percent increase in Monsoon rain, increases agricultural GDP per capita by 0.36%.[3] This relationship between rainfall and agricultural incomes appears to be the driving force behind the strong reduced form relationship between Monsoon rain and conflict in India before the introduction of NREGA.

Following the introduction of NREGA, I highlight in the second step that NREGA appears to have completely removed the relationship between Monsoon rain and conflict. A similar pattern emerges when studying agricultural wages. The introduction of NREGA insulates agricultural wages from shocks, while agricultural output is still very much dependent on Monsoon rainfall. This suggests two things: first, NREGA serves as an effective tool to stabilise agricultural wages and thus incomes; however, it is not able to affect the underlying agricultural production function, at least in the time-period under study.

In the third step, I explore the underlying mechanisms that explain the reduced form findings. I show that NREGA does function as a stabiliser with take-up - both on the extensive, and the intensive margin strongly responding to contemporaneous and lagged rainfall. An 1% lower Monsoon rainfall realisation, increases NREGA participation by 0.2%. These results hold up in an instrumental variables design, suggesting that the elasticity between agricultural GDP and NREGA employment is around -1.6. The reduction in the rainfall dependence of conflict appears to be partly driven by less violence against civilians. This supports recent evidence documented by Vanden Eynde (2011), suggesting that the opportunity cost channel drives violence against civilians who may be tempted to become police informers.

However, my findings do not imply that India has become a more *peaceful* place since the results only suggest that a particular driver of conflict has lost its bite.[4] Nevertheless, despite identification concerns, I provide some tentative evidence that suggests that overall levels of violence, following the introduction of NREGA, have gone down. This again, is driven by lower levels of violence against civilians.

My paper contributes to still nascent but growing literature that evaluates the extent to which public intervention can break the link between economic shocks and violence. Moderating the relationship between conflict and productivity

[3] For comparison, estimates can be found in Jayachandran (2006) or Duflo and Pande (2007).

[4] Khanna and Zimmermann (2013) provide some evidence from a regression discontinuity design that suggests that immediately after the introduction, there has been an increase in the levels of violence. Dasgupta (2014) on the other hand, document lower levels of violence.

shocks requires insulating personal incomes from these shocks. Technologies that can break the link between productivity shocks and incomes can be classified into three categories: (1) physical infrastructure, (2) new production technologies or (3) man made institutions. Most of the empirical literature has focused on evaluating whether these technologies achieve their primary ends: moderating income volatility.[5] Only recently, some studies have emerged that take the results form these papers to study whether they help break the link between productivity and conflict. In the first category falls Sarsons (2011)'s paper, which builds on work by Duflo and Pande (2007) suggesting that the construction of dams moderated wage volatility, but appear not to have moderated Hindu-Muslim riots. Physical infrastructure may prove to be effective only to a limited extent. Hornbeck and Keskin (2011) finds that farmers adjust their production technologies to take advantage of irrigation, which leads to higher production levels but not necessarily lower volatility. In the second category falls the work by Jia (2013), who studies the moderating effect of the drought resistant sweet potato as a new technology on the incidence of riots in historical China. This paper is the first to fall into the third category, evaluating whether a politically created institution such as India's National Rural Employment Guarantee achieves the goal to insulate personal incomes from negative shocks and through that, remove the income dependence of conflict.

My paper also relates to the wider literature on the economics of conflict. Shapiro et al. (2011) study how levels of unemployment affect levels of insurgency violence in Afghanistan, Iraq and the Philippines, finding no support for an opportunity cost channel at work. Iyengar et al. (2011) on the other hand, highlight that increased construction spending seems to cause lower levels of labour intensive violence. Blattman and Annan (2014) present results from a randomised control trial in Liberia, indicating that interventions providing training and capital can greatly increase the opportunity cost of becoming a mercenary and thus, contribute to weaken the relationship between shocks and conflict.[6] A smaller literature studies conflict in India, in particular studying the Maoist movement and the driving forces behind this conflict (see Gomes (2012)). Vanden Eynde (2011)

[5]Duflo and Pande (2007) evaluate the construction of dams and its impact on agricultural production in India. Aggarwal (2014) studies the impact of road construction, while Donaldson (2010) study the impact of railroad construction in colonial India. Burgess and Donaldson (2009) build on that work to study how trade integration may have cushioned the effect of adverse productivity shocks on famine mortality. Another vast literature tries to understand and design effective rainfall or weather insurance schemes (see e.g. Lilleor and Giné (2005) or Cole et al. (2008))

[6]This contrasts with Blattman et al. (2014), who find that a Ugandan employment program, despite large income gains, is correlated with lower levels of aggression or protests.

and Kapur et al. (2012) established that the Naxalite conflict varies systematically with incomes or proxies thereof, suggesting an opportunity cost channel at work. This paper will build on to their work, studying conflict across the whole of India[7] and how the NREGA workfare scheme, by stabilising outside options, removed the opportunity cost channel.

There is also a growing literature that evaluates the NREGA workfare program. Several papers have found that NREGA lead to increases in agricultural wages (Zimmermann (2012), Berg et al. (2012), Imbert and Papp (2012) and Azam (2011)). I will show that the stabilisation in agricultural wages takes place in case of a negative rainfall shock - which corresponds to times when demand for NREGA employment is found to be particularly high. As with any other public works programme, NREGA has been criticised by many stakeholders for its inherent inefficiency and susceptibility to corruption. Indeed, Niehaus and Sukhtankar (2013a) and Niehaus and Sukhtankar (2013b) find evidence of widespread corruption in the system. There are a few papers that have studied NREGA take-up behaviour. Johnson (2009) finds that take-up is highly seasonal and concentrated in the off-season. I confirm his findings, but find evidence that this take-up is driven by rainfall shocks in the preceding growing season, suggesting that agents do use NREGA to smooth consumption when being faced by an adverse shock. This highlights the potential consumption smoothing benefits from public employment (Gruber (1997)). I contribute to the growing literature on NREGA by combining these three observations and linking them to the nature and path of insurgency violence in India.

The paper is organised as follows. The second section provides some background on the context, the workfare program as well, the related literature and a conceptual framework. Section 3 discusses the data used. Section 4 presents the core empirical design and studies the relationship before NREGA was introduced. Section 5 studies NREGA take-up, while section 6 explores the situation after NREGA had been introduced. The last section concludes.

2 Context: Conflict and Insurance in India

Insurgencies in India India serves as a unique testing ground as there are many small-scale insurgencies that have affected India's economic development.[8]

[7]For all the analysis, I exclude Kashmir as this conflict is distinctly more violent and has significant inter state dimensions.

[8]See for example Singh (2013) for estimates of the effect of the Punjab insurgency on local investment. Nilakantan and Singhal (2012) provides some estimates of the economic cost of the Naxalite

The three main conflicts are in the North East of India, mainly comprising the so-called Seven Sister States, the Naxalite Insurgency that stretches through the Red Corridor across the East of India, and thirdly, the conflict in Kashmir. The conflicts can be grouped roughly into movements for political rights (e.g. Assam, Kashmir, Tamil's and Punjab), for social justice (the Naxalite conflict and the conflicts in the North East) and conflict on religious grounds (such in Ladakh [Kashmir] or the various religious conflicts between Muslim, Hindu and Christian groups all over India).

The intensity of these conflicts varies significantly over time. The Kashmir conflict has reduced in intensity significantly, while conflicts in the centre and the North East continue unabatedly. The conflict between the Assamese separatists and the Indian state has been on-going for more than forty years and has lead to a death-toll in excess of 30,000.[9] Concerning Naxalism, there exist no widely acknowledged data on the number of casualties, but the conflict has intensified in recent years with 2010 being considered as one of the bloodiest years ever.[10] It is difficult to study each conflict in isolation, as especially the conflicts in the East and North-East of India are indeed related. The Indian Home Minister e.g. suggests that the northeastern state of Assam has been emerging "as the new theatre of Maoist groups", with collaboration between United Liberation Front of Assam rebel groups and the Naxalites. Hence, studying a conflict in isolation may be insightful, but fails to capture possible broader underlying relationships.

NREGA workfare program The NREG scheme is a country-wide workfare program, which was passed as an act in 2005 and was introduced out from early 2006 onwards.[11] The program was rolled out sequentially in three phases: 200 districts received NREGA from early 2006 onwards, another 130 followed in 2007 and the remaining districts received the scheme in 2008. The exact algorithm used to determine which districts would receive the program first and which ones receive it later is not known. However, it is clear that the roll-out was highly correlated with pre-existing poverty levels and correlates well with an index of backwardness constructed in Planning Commission (2003). This index ranks districts by their backwardness based on the share of scheduled caste / scheduled

conflict in the state of Andhra Pradesh.

[9] See `http://www.globalsecurity.org/military/world/war/assam.htm`, accessed on 14.02.2013.

[10] See `http://www.google.com/hostednews/afp/article/ALeqM5hjfxXhNgGjp9JIyCj2ubaSKI6wIA?docId=CNG.134eae01c393f94f33516bafd808dfc9.371`, accessed on 02.04.2013.

[11] I will use the term NREG Act, NREGA and NREGS interchangeably. Though it is clear that the act is distinct from the scheme introduced under the act. The latter is the subject of this analysis.

tribe population, the levels of agricultural wages and levels of agricultural output per worker. Furthermore, Khanna and Zimmermann (2013) also suggests that the introduction was endogenous to pre-existing levels of insurgency, in particular, Maoist violence. This makes identification of a level effect particularly challenging, though, as will be highlighted in the empirical design, my identification strategy does not rely on the exogeneity of treatment assignment with respect to levels of violence.

The scheme under the NREGA Act is the largest known workfare program, generating 2.76 billion person-days of employment during the financial year 2010/2011. It stipulates that Indians in rural areas are entitled to work 100 days per year on public projects. The program is demand-led, so that the Gram Panchayat has to provide NREGA work if inhabitants require such employment. This implies that - if participating in the program is only attractive, when facing depressed outside options - we would not expect the program to have a sudden impact on violence, but only through the insurance channel. Thus, only when facing a negative income shock, should there be an effect of the program on violence by increasing the opportunity cost of joining or supporting rebel forces.

The implementation of the program is very decentralised. The Gram Panchayat sets up a list of projects. These typically can range anywhere between road construction, well digging, forestry or other forms of micro-irrigation. These projects could thus provide two benefits: first, it could offer the workers a direct benefit through the wage payments and secondly, it could have longer lasting impacts on agricultural productivity.

The NREGA act further requires that 60 percent of the budget for a project be allocated to wages. Also, the use of machines or contractors is prohibited. Workers have to apply for work in NREGA projects by filling out a written application form, after receipt of the application, the Gram Panchayat has to provide work within two weeks after receipt of the application. If the panchayat fails to provide work, a daily unemployment allowance (which is below minimum wage) is to be paid. The projects on which workers are employed have to be in close proximity to the home of the worker (at most 5 km distance) and there is additional renumeration for transportation costs or living expenses, while on the work site.

The wages are to be paid by piece rate (depending on the nature of the project) or through daily wage rate. These have to be above the state-level minimum wage. The NREGA wages must be paid by cheque or by transfer to post-office or bank accounts. In the financial year 2010-2011, expenditure on NREGA reached $ 7.88 billion, thus representing 0.5 per cent of Indian GDP.

3 Conceptual Framework and Hypotheses

It is helpful to develop a very simple conceptual framework to guide the empirical analysis. The opportunity cost argument can be formalised as in Iyengar et al. (2011). Suppose that an individual i maximises a simple utility function $u(C)$ subject to a budget constraint $C = y$. He or she can earn income y from working as agricultural labourer L or by joining the insurgent activity V. The agricultural wage is a function, e.g. of the underlying soil productivity characteristics, but especially the degree of rainfall R, that is $w_L(R)$. This sets up the possibility for there to be an effect of rainfall shocks on incomes, which is a hypothesis to be tested.

Hypothesis 1 *Agricultural wages and output are increasing in the level of Monsoon rainfall.*

The net return of an individual i supporting an insurgency is $w_V - \theta_i$, where the pecuniary income w_V is fixed, but θ_i is a measure of the degree to which an individual supports the objectives of the insurgency group or the degree to which the individual may dislike violence.[12] The θ's are drawn from a distribution with a cumulative distribution function $H(\theta)$. The key decision that a worker takes is whether, or not to participate in an insurgency. The marginal insurgency supporter defines a θ threshold level as $\bar{\theta} = w_V - w_L(R)$ such that all individuals with $\theta < \bar{\theta}$ would support the insurgency. Hence the mass of individuals partaking in the insurgency is given as $H(\bar{\theta})$. Clearly, $\frac{\partial H(\bar{\theta})}{\partial R} = -h'(\bar{\theta})w'(R) < 0$, i.e. a negative rainfall shock will increase the share of the population that participates in the insurgency. Provided the function that generates violence $F(H(\bar{\theta}))$ is increasing in the insurgency movements strength, a low rainfall realisation will induce more violence.

Hypothesis 2 *Insurgency violence is decreasing in the level of Monsoon rainfall.*

In this setup, the insurgency movement is effectively providing insurance as it provides stable wages irrespective of the realisation of rainfall as has been noted before. With the introduction of a workfare program, this role is taken over by the public employment offered by the government and thus, can serve as a means to *stabilise the outside option of workers*. For simplicity, suppose there are two realisations of $R \in \{R_l, R_h\}$ which occur with a probability p and $1 - p$ respectively.

[12] Clearly, w_V may be a function of rainfall as well; however, the assumption here is that insurgencies are able to pool risks somewhat and thus are able to offer relatively more stable wages.

The average $\bar{\theta}$ across districts is given as $E_1(\bar{\theta}) = w_V - E(w_L)$. This measure will change with the introduction of the workfare program. Conceptually, we can think of the workfare program as creating a third sector P for public employment, which pays a fixed minimum wage w_P. Assume that $w_P > w_L(R_l)$ but $w_P <= w_L(R_h)$. There are now two threshold levels for θ, which depend on the state h, l which was drawn.

$$\bar{\theta} = \begin{cases} w_V - w_L(R_h) & \text{if } h \\ w_V - w_P & \text{if } l \end{cases}$$

It is reasonable to assume that for average rainfalls $w_P < w_L(\bar{R})$, but it may well be that for R sufficiently low, the public sector wage is above the wage that would be offered by the agricultural sector. The average θ across districts is given as:

$$E_2(\bar{\theta}) = w_V + pw_L(R_h) + (1-p)w_P$$

clearly, $E_1 > E_2$ provided $w_P > w_L(R_l)$. The latter is true by revealed preference, if take-up on the extensive margin is responsive to rainfall realisations.

Hypothesis 3 *Provided that the wage paid under NREGA is higher than that paid in the agricultural sector due to a bad Monsoon, NREGA participation is negatively correlated with higher levels of Monsoon rainfall.*

From this, it is evident that due to $E_1 > E_2$ after the introduction of the workfare program, there are fewer individuals participating in the insurgency as average incomes are stabilised at a higher level. Since expected incomes are not a function of R_l anymore, this gives rise to the following hypothesis:

Hypothesis 4 *Following the introduction of NREGA, conflict becomes less responsive to Monsoon rainfall.*

In the empirical analysis, I will address each of these hypotheses in turn. With this roadmap in mind, I now turn to discuss the data used in this paper before presenting the empirical specifications and the results.

4 Data

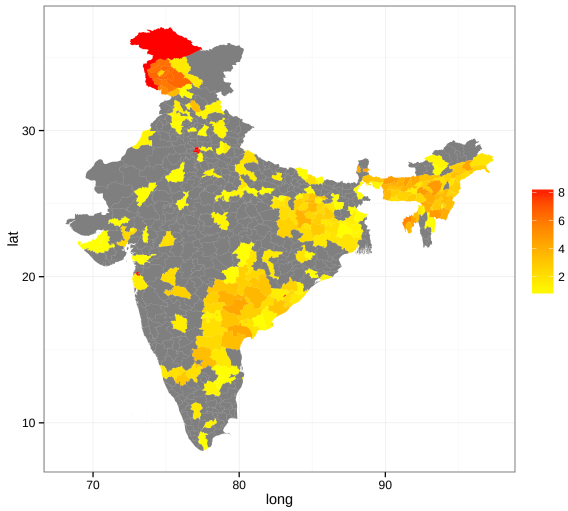
District Level Conflict data The conflict data used in this paper is drawn from the South Asian Terrorism Panel, which has collected newspaper clippings related to conflict across South Asia since the late 1990s. The dataset is extremely rich and complex, covering around 28,000 newspaper clippings for India alone. It is not feasible to hand-code this data without any prior structuring. The lack of structured data has been a key problem plaguing the conflict literature. In Fetzer (2013) I propose a method to use sophisticated natural language processing tools to be applied to the raw newspaper clippings to retrieve core pieces of information that can be transformed into a workable conflict dataset. The idea is simple: the Natural Language Processing routines try to follow the same procedure that humans would use to classify newspaper clippings into incidence counts, by identifying the subject, verb and object which constitute a violent act. Based on a set of verbs that are considered to be indicative of a violent act (this can be a very exhaustive list), machine learning routines then analyse sentences in which such keywords appear, identifying the subject, object, locational and time information in the neighbourhood of that verb. Appendix A.1 provides an example of how the algorithm constructs an incident count based on individual newspaper clipping, while Appendix A.2 compares the dataset to the Global Terrorism Database. The insight is that the semi-automatically retrieved dataset performs extremely well, compared with other violence datasets and even with manually coded data drawn from the same newspaper clippings.

For this study, the main dependent variable is the number of terrorist incidences per district and quarter. This includes all incidences with at least one fatality, but also accounts for general attacks on infrastructure, such as the destruction of telecommunication masts, or attacks without fatalities. These are informative of the insurgents fighting capacity, but are typically not included in other conflict datasets.

The left panel of Figure 1 shows the total number of incidences on a logarithmic scale between 2001 and 2006 across India. It clearly highlights the three major conflict areas: first, the Naxalite conflict, which stretches across India, in the so-called ”red corridor”.[13] The second source of major conflicts occur in the north-east, in the so-called ”7 Sister States”. There, various insurgency outfits seek to obtain independence from the Indian Union. The third major conflict is in the Kashmir region in the north west.

The right panel in Figure 1 plots the intensity of violence after 2006. It becomes clear that violence seems to have become more prevalent across India, in particular in the ”red corridor”. While conflict remained at high levels in the Seven Sister

[13]The states affected include Andhra Pradesh, Maharashtra, Karnataka, Orissa, Bihar, Jharkhand, Chattischargh and West Bengal.



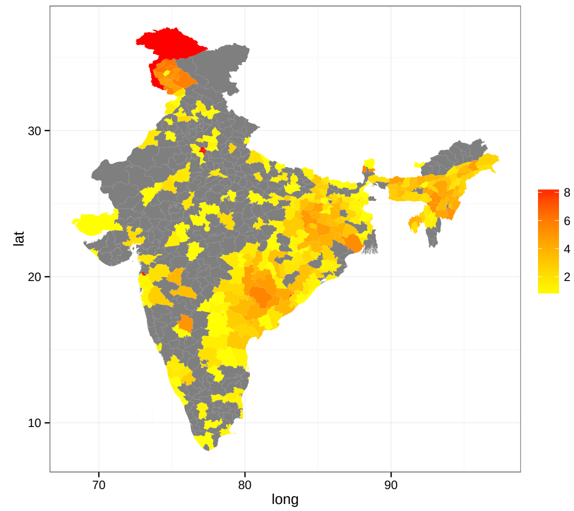


Figure 1: Spatial Dimension of Terrorist Attacks before 2006 (left) and after 2006 (right)

States, there appears to be no geographic between the conflicts there and the red corridor, which coincides well with the anecdotal accounts suggesting that various groups in the northeast work together with the Maoists. The intensification of violence, in particular in the Naxalite conflict and in the North East has also been noted in anecdotal accounts, with 2010 being considered one of the bloodiest years ever.[14] For the main exercises of the paper, I will study India as a whole, but leave out Kashmir, as this conflict has very strong inter-state dimensions (see e.g. Mohan (1992)). The map suggested that violence levels were increasing over time, in particular in the East and North East of India. This is confirmed in Figure 2, which plots the time-series of recorded incidents in the area of study.

Aside from the novel violence dataset, I also invoke a new high resolution observational weather data obtained through remote sensing techniques techniques from a novel precipitation radar, this is described in the next section.

Rainfall data This paper uses data from the Tropical Rainfall Measuring Mission (TRMM) satellite, which is jointly operated by the National Aeronautics and Space Administration (NASA) and the Japan Aerospace and Exploration Agency (JAXA). The satellite carries a set of five instruments to construct gridded rainfall

[14]See `http://www.google.com/hostednews/afp/article/ALeqM5hjfxXhNgGjp9JIyCj2ubaSKI6wIA?docId=CNG.134eae01c393f94f33516bafd808dfc9.371`, accessed on 02.04.2013.

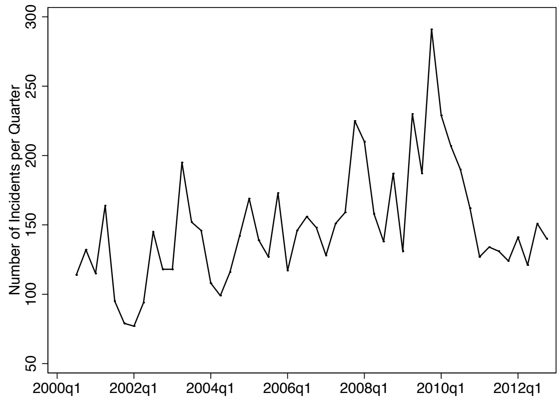


Figure 2: Number of Terrorist Incidents over Time

rates at very high spatial and temporal resolution. Due to the high spatial and temporal resolution it is providing more consistent rainfall estimates than any other available ground based observations and is considered the highest quality rainfall dataset with global coverage that is currently available (Li et al. (2012)). Its adequacy to pick up the spatial heterogeneity in precipitation has been highlighted and verified in the Indian context by Rahman and Sengupta (2007), who have shown that it outperforms e.g. the Global Precipitation Climatology Centre (GPCC) rain gauge analysis data that has been used extensively in economics research.[15] Please consult Appendix A.3 for a more detailed discussion of the data. The daily rainfall from 1998 to 2012 comes at a fine spatial resolution of 0.25 by 0.25 degree grid-cell size, which is converted into overall monthly monthly rainfall in mm.

For the identification, I will focus on the Monsoon season rainfall, which I define in two ways based on the principal crops grown using the state specific Indian crop calendar.[16] I use a narrow- and a broad definition of the Monsoon period. This varies from state to state as the typical onset dates are early May for the north east of India, while the onset may be as late as late June for central India. For most states, the narrow Monsoon-period ranges from June to Septem-

[15]For example by Miguel et al. (2004), Ferrara and Harari (2012) and Kudamatsu et al. (2012). As this data product comes at a coarse spatial resolution, researchers typically apply inverse distance weighting methods to interpolate between grid-points. This typically underestimates the spatial variability or induce variability where there is actually none (see Haberlandt (2007)). This will affect the estimation of extreme values that these papers typically rely on for identification (Skaugen and Andersen (2010)).

[16]In particular the key reference is the crop specific calendar maintained by the Indian Food Security Mission, available via `http://nfsm.gov.in/nfsmmis/RPT/CalenderReport.aspx`, accessed on 12.05.2013.

ber, while the broad ranges from May to November. The narrow Monsoon period accounts for 80% of the annual rainfall, while the broad Monsoon period covers roughly 90% of annual precipitation. For the main analysis I will use the narrower definition that focuses on the key principal crops and use the log of that rainfall. The broader measure and various other transformations are used for robustness checks.

NREGA Participation Data I use the NREGA participation data derived from the so-called Monthly Progress Reports (MPR) from before 2011 and from the Management Information System (MIS) from 2011 onwards. The key variables I study are extensive margin participation as the share of households in a district that participate under NREGA in a given financial year, the days worked per household and the total person days generated. I also obtained data on the number and total cost of ongoing projects, where I classify projects for irrigation purpose specifically.[17]

I study three major margins of NREGA take-up. Firstly, extensive margin participation as the share of households in a district who demand employment. Secondly, intensive margin participation as the log of the number of days worked per household. Last but not least I consider a measure of the number and cost of ongoing NREGA projects.

Agricultural Production and Wages In order to test whether NREGA had an impact on the cyclicality of agricultural wages and agricultural production, I construct time-series for the two. To construct agricultural wages, I use Agricultural Wage Data from the Agricultural Wages in India (AWI) series which has been published by the Indian Ministry of Agriculture since 1951. It is unique in offering monthly wage rates by district (sometimes even containing multiple locations per district), and separate wage series for several categories of labour and by gender. The quality of the data is very poor however, with a large number of observations being missing or simply flat wages being reported throughout. In order to increase the signal to noise ratio, I average the data to generate an annual wage series. I detail some of the issues with this dataset in appendix A.6. More reliably measured is agricultural production. I use data on annual district level production collected and published by the Directorate of Economics and Statistics with the Ministry of Agriculture.[18] For every district, I only consider crops

[17]Refer to Appendix A.7 for further discussion of the available NREGA participation data.
[18]This data is available on `http://apy.dacnet.nic.in/cps.aspx`, accessed 14.08.2013.

that have been consistently planted on at least 1000 hectares for the period that the state reports data. I use state-level harvest prices to construct a district level measure of agricultural GDP.

I the next section, I present the empirical strategy before presenting the core results.

5 Empirical Strategy

The aim of the empirical design is to cleanly estimate the changing functional relationships between Monsoon rainfall and agricultural output, wages, the incidence and intensity of violence. To this end, I separately estimate the relationships before the introduction of NREGA and once, for the whole sample after the introduction of NREGA.

The main specification that uses agricultural GDP per capita or wages as left-hand side is:

$$log(y_{dt}) = a_d + b_{pct} + \theta R_{dt} + \mathbf{X'_{pdt}}\text{ß} + \epsilon_{cpdt} \quad (1)$$

where R_{dt} measures contemporaneous Monsoon season rainfall, a_d is a district fixed effect, absorbing any time-invariant district characteristics such as terrain ruggedness or elevation. I construct region and NREGA phase specific time fixed effects in b_{pct}.[19] These demanding time-fixed effects address a key concern as they flexibly control for the fact that the NREGA introduction happened in three distinct phases. Districts in the first phase were poorest and may be subject to distinct shocks or were on distinct (non-linear) trends. These fixed-effects take into account such variation. The matrix X_{pdt} contains a set of district controls that are included in some specifications. These include a set of time-invariant characteristics that have been identified to correlate with the sequence of roll out of NREGA and will be used for robustness interacted with a set of time-fixed effects. These characteristics are identified by exploiting cross-sectional variation across districts estimating:

$$Phase_d = a + \mathbf{H}'_d \beta + u_d \quad (2)$$

where $Phase_d$ is an integer that is either 1, 2 or 3 indicating in which phase a

[19]The geographic regions I consider are the states in the Red Corridor (Andhra Pradesh, Orissa, Bihar, West Bengal, Chhattisgarh, Jharkhand, Karnataka and Maharashtra). The states in the Northeast (Assam, Meghalaya, Sikkim, Tripura, Mizoram, Nagaland and Manipur). The remaining states, mainly in the west of India are contained in its own group.

district received the program and H_d is a matrix for the candidate district characteristics.

For agricultural wages, I include a set of state by NREGA phase specific linear time trends. These become necessary as agricultural wages are increasing dramatically but distinctly for some states in a way that is not captured by the time fixed effects.

The main specification I estimate for conflict is a conditional fixed effect Poisson model as in Santos Silva and Tenreyro (2006). This accounts for the count nature of the conflict data. The specification is:

$$\mathbb{E}(A_{pcdt}) = \delta_d \exp\left(b_{pct} + \eta R_{dpt-1} + \mathbf{X'_{pdt}}\text{ß} + \epsilon_{cpdt}\right) \tag{3}$$

The results are robust to using plain OLS or negative binomial estimators, and I also present results on the incidence of conflict which is simply a linear probability model.[20] Note that rainfall is measured from the preceding calendar year or growing season, which is in line with the existing literature and I confirm that the effect of Monsoon rain on conflict mainly happens with a one year lag.

Following the introduction of NREGA, I essentially estimate the same specifications except that I add an interaction term between the rainfall variable R_{dt} or R_{dt-1} and an NREGA treatment indicator. That is, I construct a dummy variable:

$$T_{dpt} = \begin{cases} 1 & \text{if NREGA available in district } d \text{ at time } t, \\ 0 & \text{else.} \end{cases}$$

Note that by including region by phase- and time fixed effects, the treatment indicator is perfectly collinear with these fixed effect. The variation used to identify the effect comes from within phase-regions over time and thus, I do not live off of variation across districts in different NREGA implementation phases. This is important to bear in mind, as the roll of out NREGA was likely endogenous to pre-existing levels of violence, as has been argued in Zimmermann (2012), which makes it very difficult to exploit variation across NREGA phases.

[20]See table **??** in the appendix for these checks. I use a Pseudo Maximum Likelihood Poisson (PPML) estimator as implemented by Santos Silva and Tenreyro (2006) as it overcomes some of the numerical problems in common implementations in statistical packages such as Stata (see Silva (2011)). The PPML estimator does not require the data to have equi-dispersion. It is consistent, so long as the conditional mean is correctly specified. The estimator is even optimal if the conditional variance is *proportional* to the mean, hence over dispersion is not an issue. Note further that conditional and unconditional likelihood yield identical estimates, but typically the former is chosen as the computation is quicker (Cameron and Trivedi (1999)).

The estimating equation then becomes:

$$log(y_{dt}) = a_d + b_{pct} + \theta R_{dt} + \gamma T_{dt} \times R_{dt} + \mathbf{X'_{pdt}}\text{ß} + \epsilon_{cpdt} \quad (4)$$

while the Conflict regressions are

$$\mathbb{E}(A_{pcdt}) = \delta_d \exp\left(b_{pct} + \eta R_{dpt-1} + \gamma T_{dt} \times R_{dpt-1} + \mathbf{X'_{pdt}}\text{ß} + \epsilon_{cpdt}\right) \quad (5)$$

The identifying assumption for these models is that the timing of the introduction of NREGA in a district was not endogenous to the previously existing relationship between rainfall and conflict. This explicitly allows for the fact that the roll-out likely was endogenous to the levels of violence. In order to control flexibly for the previously existing relationship between Monsoon rain and output, I construct a district specific elasticity θ_d by running

$$log(y_{dt}) = a_d + \theta_d R_{dt} + \nu_{dt} \quad (6)$$

for every district using data from before the introduction of NREGA, where y_{dt} measures agricultural output. I use the estimated elasticities $\hat{\theta}_d$'s as an additional control in $\mathbf{X_{pdt}}$ interacted with a set of time-fixed effects in some specifications.

In order to study the underlying mechanisms, I explore NREGA participation data on the intensive and the extensive margin by estimating:

$$P_{pcdt} = \delta_{d_k} + b_{pct} + \eta R_{dt-1} + \mathbf{X'_{pdt}}\text{ß} + \epsilon_{pcdt} \quad (7)$$

where P_{pcdt} is a measure of intensive- or extensive margin NREGA participation. As the underlying data sources change in a way that systematically varies across districts from 2011 onwards, I include district fixed effects d_k that are different depending on the underlying datasource indexed by k.[21] I also entertain an instrumental variables specification, instrumenting for lagged agricultural output using lagged Monsoon rain.

For Poisson models I present standard errors clustered at the district level. For the linear models, I present standard errors that account for spatial dependence as discussed in Conley (1999).[22] The implicit assumption here is that spatial dependence is linearly decreasing in the distance from district centroids up to a cutoff distance, for which I chose 500 km. Note that some datasets are an unbalanced

[21] Refer to appendix A.7 for more details. The results are robust to using just either part of the data.

[22] I use a routine that iteratively demeans the data before computing the standard errors as in Hsiang (2010). The Stata code for this function is available from my personal website on `goo.gl/ACbuLA`.

panel, in which case the spatial HAC procedure is problematic. For these cases, I present the more conservative standard errors either obtained by clustering at district level or from the Conley routine.[23] I now proceed to present the main results.

6 Results

6.1 Before NREGA: Agriculture, Wages and Violence

This section studies the period before the introduction of NREGA. I restrict the analysis to this period to highlight that the relationship between Monsoon season rainfall, agricultural output, wages and violence had existed well before the introduction of the workfare program. The results from specifications 3 and 4 are presented in Table 1. Columns (1) - (2) study agricultural GDP per capita. Column (2) suggests that a one percent increase in Monsoon season rainfall increases agricultural GDP in that year by 0.36%. The comparison with column (1) which uses the whole annual rainfall highlights that the bulk of the effect of annual rainfall is coming from the Monsoon season.[24] This is a surprising finding, since decades worth of investment in irrigation facilities should have rendered the agricultural output more resilient. In fact, the estimated coefficient here is higher than that found in other previous studies (see for example Jayachandran (2006)), suggesting that this study improves upon the existing work by reducing measurement errors.

Column (3) - (4) performs the same exercise for agricultural wages. The pass through of rainfall variation is statistically significant, but small in size. A 1% increase in rainfall increases agricultural wages by 0.06%. Again the effect is driven almost in its entirety by Monsoon season rainfall.

The last four columns focus on conflict. The estimated coefficients in columns (5)-(6) are elasticities, suggesting that a 1% increase in Monsoon rainfall reduces conflict by 0.87%. The incidence of conflict is also statistically very responsive to Monsoon rainfall variation. Note that the results compare very well with Vanden

[23] All results hold up when clustering at the district level, clustering at the state level is not feasible as there are fewer than 30 clusters in most specifications. An alternative is to cluster at the state by NREGA implementation phase level, most results are robust to clustering at this level. These results are available from the author upon request.

[24] Appendix Table A1 provide some robustness checks adding further temperature controls and other district characteristics and focusing on grain production. Appendix Figure A2 highlights the smooth and monotonous relationship between agricultural GDP and rainfall.

Eynde (2011) who estimates an elasticity between Monsoon rainfall and grain-production of 0.45 and an elasticity of rainfall with respect to civilian casualties of 0.88. In appendix tables A3 and A4 I perform a whole range of robustness checks highlighting that the results are robust to the choice of empirical model, adding a battery of further controls and the choice of rainfall measure to alleviate concerns raised in this literature e.g. by Ciccone (2011). An IV approach using a vegetation index instrumented by rainfall as performed in Kapur et al. (2012) yields very similar results to what they find.

In the next step, I discuss the endogenous nature of the roll-out of NREGA, highlighting however, that it appears not to be endogenous with regard to my identifying assumption.

6.2 NREGA Introduction: Endogeneity of Treatment

As already indicated in section 2, the sequence of the roll out of NREGA is highly endogenous. This is an important caveat to bear in mind when trying to make causal claims exploiting variation stemming from the fact that NREGA was gradually rolled out in different phases. Table 2 confirms that roll-out of NREGA was highly endogenous to a set of district level characteristics, presenting results from specification 2.

It is evident that districts that were violent in 2004 were more likely to receive NREGA in the first rounds. The coefficient is consistently negative, when adding more controls, but remains statistically only marginally significant (which is due to the choice of standard errors). The endogeneity of NREGA roll-out - especially to Naxalite violence - has been highlighted by Zimmermann (2012). Other characteristics that correlate well with the order of roll out are a high population share of scheduled castes or scheduled tribe population. High wages, high agricultural output per capita and a high literacy predict treatment in later rounds. The most important coefficient for my purpose is presented in the second row. Using the constructed district level measure of the elasticity of Monsoon rainfall with respect to agricultural GDP, θ_d as a control. This elasticity measures the local responsiveness of agricultural output to local Monsoon rainfall and is thus, a measure of the extent to which rainfall shocks affect local incomes. In none of the specifications does this measure gain any significance. This gives me confidence that NREGA roll out was not endogenous to the way that rainfall translates into output, while it very well endogenous to production levels and a whole range of other covariates. I will now proceed to present the results indicating how the functional relationship between rainfall and conflict fundamentally changed following the introduction of NREGA.

6.3 After NREGA: Moderation of Violence

Table 3 provides the results indicating how the functional relationship between Monsoon rainfall and agricultural output, wages, violence intensity and incidence changed with the introduction of NREGA. The results are stark. Columns (1) and (2) indicate that the agricultural production function has not fundamentally changed with the introduction of NREGA. The interaction coefficient is positive but insignificant at conventional significance levels, indicating that agricultural output is still highly rainfall dependent. Columns (3) and (4) focus on agricultural wages. The results are stark, indicating that the introduction of NREGA has removed the pass-through of rainfall on agricultural wages, thus insulting the latter from this source of variation. This is not surprising: NREGA is primarily a program to create employment opportunities and thus, may only indirectly affect the underlying agricultural production function, making it more resilient to weather variability due to investment in micro-irrigation facilities.

The last four present the core results. The introduction of NREGA has removed the rainfall dependence of the intensity and incidence of conflict almost throughout (see columns (5)-(8)). Before exploring the underlying mechanisms, I highlight that the results are very robust to alternative ways of looking at the data.

Robustness There are three core robustness checks that I perform for the three main outcome variables. Firstly, I add a set of control variables interacted with a set of time-fixed effects. These control variables include agricultural GDP per capita before 2005, scheduled cast population share, share of literate population, scheduled tribe population share, elevation, household size, the gender gap and most importantly, the estimated elasticity between agricultural output and Monsoon rainfall at district level. As some of these variables varied systematically across the NREGA phases, this allows me to flexibly control for trends that are specific to these variables.

The second set of exercises are placebo checks. First, I study rainfall outside the Monsoon season. This rain only had marginal effect on agricultural output as indicated in section 6.1. Hence, one would not expect that the introduction of NREGA correlates in any significant way with this rainfall variable.

The second placebo moves the NREGA reform three years ahead of time. This is possible as the conflict data begins in mid 2000. This serves as a check to whether the change in the relationship between rainfall, wages and conflict had already happened before NREGA was introduced and thus, serves as a means to check for common trends.

The robustness checks for agricultural wages and output are presented in table 4. For the agricultural wages, I also estimate the interaction effect for wages in harvesting season as opposed to the planting season. This suggests that the moderation effect is coming from the harvesting activity wages, which are relevant after the Monsoon.

The robustness checks for the NREGA effect regressions are presented in table 5. The first column restricts the analysis to the districts that had been violent before NREGA was introduced. The estimated effect is very similar from the main specification. The second and third columns perform the placebo tests as described, while in the fourth column I add the district specific controls interacted with a set of year fixed effects. This is an attempt to control for the set of variables that were driving selection into the different NREGA phases. The estimated coefficients do not change significantly. This is not completely unexpected as the elasticity of income with respect to rainfall, which I argue, is driving the relationship with violence was not a selection criteria. The last column studies contemporaneous Monsoon rain. The coefficients point in similar directions but do not gain significance.

NREGA Effect over Time All in all, these results suggest that the relationship between rainfall and violence changes after the introduction of NREGA. This suggests that there is some effect of the NREGA on the dynamics of violence. A key concern with the above specification however is, that the relationship between rainfall and violence may have been changing over time, independently from the introduction of NREGA - i.e. there could be time-specific changes to the way that rainfall translates into violence, that are independent of NREGA, but may be picked up by the interaction term. In order to address this concern, I estimate a very flexible specification, where I allow the effect of rainfall on violence to be a different for each quarter of each year.[25]

The specification I estimate is

[25] That is to say since the sample period is 2000q2 to 2012q4, I estimate 50 individual rainfall effect coefficients.

$$
\begin{aligned}
\mathbb{E}(A_{pcdt}) &= \delta_d \exp(b_{pct} + \alpha T_{dpt} + \sum_t \eta_t R_{dt-1} \\
&+ \sum_{p=1}^{3} \eta_p R_{dt-1} + \sum_{p=1}^{3} \gamma_p T_{dpt} P_p R_{dt-1} + \mathbf{X'_{dt}}\text{ß} + \epsilon_{cpdt})
\end{aligned}
$$

This specification still allows for the estimation of a phase-specific rainfall-effect η_p and also a phase-specific NREGA effect γ_p, as the way that rainfall translates into violence may be different across phases, which is not picked up by the simple time specific effects η_t which are homogeneous across the three phases. The results from this specification are best presented graphically. Figure 3 plots the overall effect by NREGA phase, which is simply the linear constraint: $\hat{\eta}_t + \hat{\eta}_p + \hat{\gamma}_p T_{pt}$.

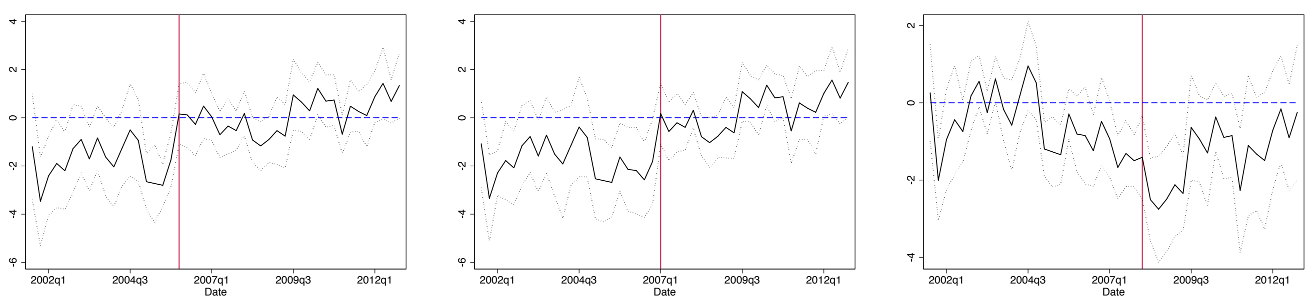


Figure 3: The Effect of Monsoon rain on Violence over Time for Phase 1, Phase 2 and Phase 3 districts from left to right.

It becomes evident that this more demanding specification confirms the previous findings that suggest that the relationship between rainfall and violence has changed after the introduction of NREGA. The graphs for districts in the first and second phase look very similar with negative overall effects before NREGA and insignificant effects afterwards, suggesting that the overall effect is driven by districts in the first two phases, which were - poorer on average - and thus, are the districts where one would expect NREGA to have the largest effect.

I now explore whether the reduced form findings can be reconciled with evidence on NREGA take-up, the targeting of violent acts and present some estimates of an overall level effect.

7 Mechanisms

7.1 NREGA Take-up Behavior

I proceed by studying whether NREGA take-up behaviour follows a similar pattern suggested by an opportunity cost argument. If NREGA provides a safe outside option in dire times, this should be reflected in increased take-up following an adverse shock. The results are presented in table 6. The fist column measures the log of total person-days in employment created in a financial year. The elasticity between rainfall and participation is strongly negative. This overall take-up effect is decomposed into extensive- and intensive margin participation in columns (2) and (3). The extensive margin measures the share of households who participate. Since the program is provided on a per-household level, this is the correct way to measure extensive margin participation. Column (2) suggests that a one percent increase in rainfall reduces the share of households who participate by 0.05% and intensive margin participation by 0.118%. The instrumental variables result suggest a unit elasticity between agricultural GDP per capita and intensive margin NREGA participation. This high elasticity could be driven through a general equilibrium effect, as low production drives up prices and may actually depress real incomes, leading to additional demand for NREGA employment through that general equilibrium effect. Columns (4) and (5) look at how the costs of active NREGA projects in a district at the end of a financial year respond to passed Monsoon realisation. The point estimates for both, costs on irrigation projects and all projects are very similar to the costs of overall participation. Since at least 60% of the costs must be budgeted to cover labour expenses, the similarity of the coefficients with the coefficient in column (1) is very plausible. The similarity of the coefficients in column (4) and (5) suggests furthermore, that adverse rainfall realisation do not predict NREGA project activity for irrigation purposes differentially.

Robustness In table 7 I present some robustness checks of the relationship between rainfall and NREGA take-up. In particular, in the first two columns I constrain the analysis to the years where the data-source is common pre 2011. The coefficients are slightly higher but very similar to the previous findings. Columns (3) and (4) include some further controls. In column (3) its most notable that rainfall outside the Monsoon has only a very weak or insignificant effect; this is as expected as rainfall outside the Monsoon season did not seem to predict wages or production strongly. Contemporaneous Monsoon, however, has a strong effect on

NREGA take-up as well as lagged Monsoon. This is due to the way that NREGA data is reported on a financial-year calendar with goes from April to March and so, contemporaneous Monsoon rain may affect take-up up to March of the subsequent year. Column (4) includes the set of district specific controls identified as important selection criteria for the roll-out interacted with a set of time-fixed effects. The coefficient drops quite a lot, but still remains significant at the 5% level. In column (5) and (6) I focus on take-up by scheduled cast/ scheduled tribe populations. Since the Naxalites recruit some of their supporters from among these populations, its important to see whether the take-up by these subpopulations follows a similar pattern. The results confirm that this is indeed the case.

In order to corrobate these findings on take-up, I estimate the above specification using the constructed monthly participation data from the reported monthly cumulative figures. This data is quite noisy due to reporting lags, especially in the earlier years. Months with missing data are dropped.

I estimate the following specification:

$$P_{pcdt} = \delta_{d_k} + b_{pct} + \sum_{i=1}^{12} \eta_i R_{dt} + \epsilon_{pcdt} \tag{8}$$

This provides a set of coefficients η_i that can be plotted to trace out the impact of Monsoon season rainfall on NEGA participation across different months from the beginning of calendar year up until its end. The results are depicted in figure 4.

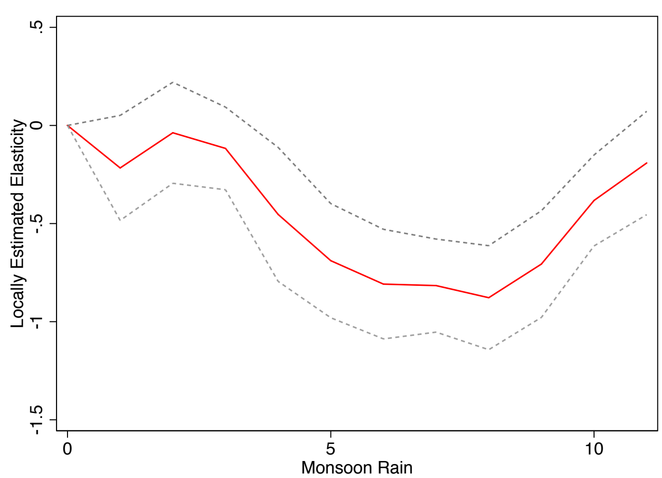


Figure 4: NREGA Month-on-Month Take-up and Monsoon Rainfall

The graph suggests that Monsoon rainfall begins to reduce NREGA participation from May onwards. As Monsoon onset in some parts of India is as early as May, this makes sense. The effect of Monsoon rainfall on participation is strongest from August onwards, as the Monsoon begins to withdraw. The results are very suggestive: in the high agricultural season, fed by good rainfall, there are ample employment opportunities available, which reduces NREGA participation. This effect persists even into the Rabi season in November and December.

In the next section, I explore what types of violence appear to become less rainfall dependent with the introduction of NREGA.

7.2 Targets of Violence

A question that has started to arise in the economic analysis of the drivers of conflict lies behind who is the actual target of violence. Vanden Eynde (2011) argues that civilians, facing an income shock, find themselves torn between becoming police informers, which offers some economic benefits. This comes however, at a cost, as insurgents react with more violence against civilians. Hence, a natural question that arises is what type of violence is particularly income-dependent and how does this change following the introduction of the employment guarantee. The conflict data allows a rough classification of the subject of violent activities into groups: civilians, security forces and terrorists. In Fetzer (2013) I highlight how this is done with the aid of humans to classify ambiguous cases.

Columns (1) to (3) of table 8 performs the analysis of the NREGA effect, breaking up the violence data into the different classes. The pattern that emerges is very suggestive. While all types of violence is responsive to lagged Monsoon, the moderating effect of NREGA is most strongly seen for violence targeted against civilians in column (1). In column (2), the NREGA effect is visible as well; however, the sum of the two coefficients actually is positive, which could suggest that violence against security forces is becoming pro-cyclical as opposed to counter-cyclical. The sum of the two coefficients is insignificant however. The third column looks at incidences where the subject of the incidence were terrorists. There appears to be only a weak moderating effect of NREGA.

Column (4) presents results on the share of incidences with subjects classified as civilians. The coefficients confirm what columns (2) and (3) indicate: NREGA could help bring civilians out of *the line of fire*. As a consequence, this could free up resources for the insurgents that were previously used to extort civilians and allow increased targeted violence against the state and its institutions.

While this is an open question that remains to be addressed by future research, I attempt to combine the preceding findings to provide a rough estimate of the overall level effect of NREGA.

7.3 Separating Level and Dynamic Effects of NREGA

The preceding results suggested that NREGA does have a moderating effect on the cyclical nature of violence, in particular, the violence targeted against civilians. However, the existing literature evaluating the economic impacts of NREGA also indicate strong increases in wage levels.[26] In the context of the conceptual framework, such an increase in wage levels can be seen as an increase in the returns to labour in both, good- and bad states of the world. This - of course - does have an independent level effect on violence, as it shifts the overall participation constraint. However, the underlying opportunity cost mechanism is still intact, suggesting that there is an independent effect stemming from the stabilisation of agricultural incomes in the bad state It is challenging to identify a level effect, due to the endogeneity of the roll-out. More importantly, any study focusing on the level effect by interpreting the NREGA treatment indicator as such, actually finds a mixture between the level effect due to higher wage levels irrespective of the state of the world, and the effect stemming from a reduced income elasticity of conflict.

The following exercises aim to highlight the importance of taking into account the insurance mechanism that is the focus of this paper. I estimate the following specifications while imposing various constraints:

$$\mathbb{E}(A_{cdt}) = \delta_d \exp\left(b_{ct} + \alpha T_{dt} + \eta R_{dpt-1} + \gamma T_{dt} \times R_{dt-1} + \mathbf{X'_{dt}}\text{ß} + \epsilon_{cdt}\right) \quad (9)$$

where b_{ct} are now region by time fixed effects, rather than region by phase and time fixed effects. This set of fixed effects allows the estimation of the parameter α, which can be interpreted as the level effect of NREGA if we are willing to assume that the roll-out of NREGA was exogenuous.

I estimate a constrained version of specification 9, requiring that $\eta = \gamma$. In this case, I force the effect of rainfall to be the same before and after the introduction of NREGA. I also estimate a specification with the constraint $\eta = \gamma = 0$, which effectively means not controlling for rainfall. The key question is how this will affect the estimated coefficient $\hat{\alpha}$. In both cases, the coefficient $\hat{\alpha}$ should overstate the effect of NREGA in absolute value.

[26]See Zimmermann (2012), Berg et al. (2012), Imbert and Papp (2012) and Azam (2011).

The results are presented in table 9. The first column presents the constrained regression where I do not control for rainfall. The level effect coefficient is negative and statistically significant. In the second column, I control for rainfall, which renders the coefficient slightly larger in absolute value. The third column is the unconstrained coefficient, allowing the functional relationship between rainfall and conflict to change with the introduction of NREGA. The interesting observation is that the coefficient on the level effect goes down and is estimated relatively imprecisely, moving from a p-value close to 0.001 to p-value of 0.45. This suggests that the dynamic effect of NREGA, operating by mitigating income shocks, is being partially captured in estimates of $\hat{\alpha}$, when one does not explicitly control for this important economic channel through which NREGA operates.

Despite this, any estimate of $\hat{\alpha}$ is plagued by the fact that the NREGA introduction was endogenous to violence levels and many other observable and unobservable covariates. Thus, any estimate of a effect should be taken with a grain of salt. Nevertheless, in table 10 I present results of the level effect, controlling explicitly for rainfall and its interaction with NREGA. The results I find are broadly consistent with Dasgupta (2014), who estimate effects of NREGA on levels of Maoist violence.

The first column presents the basic level effect estimate of contemporaneous treatment. The second column adds lagged effects of the NREGA treatment indicator, suggesting that the first lag is highly significant. The point estimate suggest that the introduction of NREGA reduced levels of violence by between 30% to 50%. The third column adds the district characteristics interacted with time-fixed effects. The estimated effect increases in absolute value. Columns (4)-(9) explore the heterogeneity of the estimated effect by interacting the treatment indicator with a set of district-characteristics that have been identified to matter for the sequence of roll out. Important covariates are the age only statistically significant heterogeneity is for the scheduled tribe population share, suggesting that the level effect is weaker for districts with a high scheduled caste population share. Furthermore, indicative is the coefficient on average household size. This suggests that the level effect is significantly weaker for districts with a larger average household size. This is not too unsurprising, since the NREGA program provides an allowance for 100 days of work *per household*. Hence, larger households are disadvantaged in that respect. Column (8) interacts the treatment indicator with the log of the mean level of agricultural GDP per capita before 2005. The coefficient is insignificant. While the results on the dynamics of conflict do not square with Khanna and Zimmermann (2013), the estimated level effects do stand at odds with the ones estimated in their paper. Clearly, they focus on the short-run effects of the scheme's introduction which may have lead to more police presence and thus, more violence targeted against police. Table 11 presents results when estimating the level effect of NREGA, splitting up the attacks into ones with civilian, security force or insurgent subjects. As suggested by the results on the dynamics of conflict, the level effect appears to come in its entirety from less violence targeted against civilians.

8 Conclusion

This paper has set out to investigate the impact of the NREGA Workfare Program on the dynamics of violence in Indian intra-state conflicts. I find that the income dependency of violence has decreased significantly following the introduction of the public employment scheme, suggesting that one of the key drivers of insurgency violence can be moderated through the effective introduction and provision of social insurance. This indicates that a possible tool to affect the dynamics of violence in conflict torn areas is the introduction of a social insurance system that provides stable outside options in times of need. The key design feature that enables NREGA to function as such is, that it is entirely demand driven. The then shows that the observed NREGA effects are plausible when studying NREGA participation data; furthermore, there appears to be a general equilibrium effect of NREGA on agricultural wages as well - stabilising wages when these otherwise would be depressed due to adverse weather conditions.

The paper contributes to the growing literature that evaluates how infrastructure, technology or other types of institutions can moderate the links between income and criminal activity in general. While a vast literature has emerged that tries to evaluate the insulating effects of physical infrastructure on incomes, the literature that evaluates its implications for conflict is still at an early stage.

There are some important open questions however. If NREGA drove up the opportunity cost of conflict and thus, the implicit wages for insurgents, does this induce insurgents to shift away from labour intensive means to inflict violence towards more capital intensive ones? This has been studied by Iyengar et al. (2011) in the context of a labour market intervention in Iraq. Since NREGA has been identified to drive up wage levels, this is an important question to be explored further. Similarly, as it appeared that the dynamic as well as the level effect is mainly driven by less violence against civilians, what are the effects on violence against the Indian state or its security forces? Not having to inflict violence against civilians in order to prevent them turning into police informants could free up significant resources for the insurgents, enabling to direct more violence against the state. In fact, there is some anecdotal evidence suggesting that Naxalites are increasingly targeting urban populations.[27]

Last but not least, the NREGA program may have implications for the insurgents extortion base as well. This has not been explored in this paper, but the evidence collected by Vanden Eynde (2011) suggests that violence in places with a stable tax base has distinct patterns for the types of violence inflicted. Stabilised rural incomes could indirectly, by stabilising the extortion base, strengthen the insurgents fighting capacity.

[27] See for example `http://articles.economictimes.indiatimes.com/2013-08-13/news/41375368_1_urban-areas-cpi-organisations`, accessed 12.12.2013 or Magioncalda (2010).

References

Abadie, A. and J. Gardeazabal (2008, January). Terrorism and the world economy. *European Economic Review 52*(1), 1–27.

Aggarwal, S. (2014). Do Rural Roads Create Pathways Out of Poverty? Evidence from India. (April).

Akresh, R. and D. D. Walque (2008). Armed conflict and schooling: Evidence from the 1994 Rwandan genocide. *mimeo* (April).

Azam, M. (2011). The Impact of Indian Job Guarantee Scheme on Labor Market Outcomes: Evidence from a Natural Experiment. *SSRN Electronic Journal*.

Bazzi, S. and C. Blattman (2013). Economic Shocks and Conflict: Evidence from Commodity Prices. *mimeo*.

Becker, G. (1968). Crime and punishment: An economic approach. *Journal of Political Economy 76*(2), 169–217.

Berg, E., S. Bhattacharyya, R. Durgam, and M. Ramachandra (2012). Can Rural Public Works Affect Agricultural Wages? Evidence from India. *CSAE Working Paper*.

Besley, T., T. Fetzer, and H. Mueller (2012). The Welfare Cost of Lawlessness: Evidence from Somali Piracy. *mimeo*.

Besley, T., H. Mueller, and T. Fetzer (2014). The Welfare Cost of Lawlessness: Evidence from Somali Piracy. *mimeo*.

Blattman, C. and J. Annan (2014). Can employment reduce lawlessness and rebellion? Experimental evidence from an agricultural Intervention in a fragile state. (212).

Blattman, C., N. Fiala, and S. Martinez (2014, December). Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda. *The Quarterly Journal of Economics 129*(2), 697–752.

Blattman, C. and E. Miguel (2009). Civil War. *NBER Working Paper 14801*.

Burgess, R. and D. Donaldson (2009). Can openness mitigate the effects of weather shocks? evidence from indias famine era. (1938).

Burnicki, A. C., D. G. Brown, and P. Goovaerts (2007, May). Simulating error propagation in land-cover change analysis: The implications of temporal dependence. *Computers, Environment and Urban Systems 31*(3), 282–302.

Cameron, A. C. and P. K. Trivedi (1999). Essentials of Count Data Regression. *mimeo*, 1–17.

Chassang, S. and G. Padro-i Miquel (2009). Economic shocks and civil war. *Quarterly Journal of Political Science*, 1–20.

Ciccone, A. (2011). Economic shocks and civil conflict: A comment. *American Economic Journal: Applied Economics* (February).

Cole, S., J. Tobacman, and P. Topalova (2008). Weather insurance: Managing risk through an innovative retail derivative. *mimeo*.

Collier, P. and A. Hoeffler (1998). On economic causes of civil war. *Oxford economic papers 50*, 563–573.

Conley, T. (1999). GMM estimation with cross sectional dependence. *Journal of econometrics 92*, 1–45.

Dal Bó, E. and P. Dal Bó (2011, August). Workers, Warriors, and Criminals: Social Conflict in General Equilibrium. *Journal of the European Economic Association 9*(4), 646–677.

Dasgupta, A. (2014). Can Anti-poverty Programs Reduce Conflict ? India s Rural Employment Guarantee and Maoist Insurgency.

Dee D. P., Uppala S. M., Simmons A. J., Berrisford P., Poli P., Kobayashi S., Andrae U., Balmaseda M. A., Balsamo G., Bauer P., Bechtold P., Beljaars A. C., L. M. van de Berg, Bidlot J., Bormann N., Delsol C., Dragani R., Fuentes M., Geer A. J., Haimberger L., Healy S. B., Hersbach H., Hólm E. V., Isaksen L., Kållberg P., Köhler M., Matricardi M., McNally A. P., Monge-Sanz B.M., Morcrette J.J., Park B.K., Peubey C., de Rosnay P., Tavolato C., J. N. Thépaut, and F. Vitart (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society 137*, 553–597.

Donaldson, D. (2010). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *NBER Working Paper 16487*.

Dube, O. and J. Vargas (2013). Commodity price shocks and civil conflict: Evidence from Colombia. *Review of Economic Studies*.

Duflo, E. and R. Pande (2007, May). Dams. *The Quarterly Journal of Economics 122*(2), 601–646.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association 87*(420), 998–1004.

Fearon, J. D. and D. D. Laitin (2003, March). Ethnicity, Insurgency, and Civil War. *American Political Science Review 97*(01), 75.

Ferrara, E. L. and M. Harari (2012). Conflict , Climate and Cells: A disaggregated analysis. *mimeo*.

Fetzer, T. R. (2013). Developing a Micro-Level Conflict Data Set for South Asia. (2010), 1–17.

Gomes, J. (2012). The political economy of the Maoist conflict in India: an empirical analysis. (March), 1–54.

Gruber, J. (1997). The Consumption Smoothing Benefits of Unemployment Insurance. *American Economic Review 87*(1), 192–205.

Haberlandt, U. (2007, January). Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *Journal of Hydrology 332*(1-2), 144–157.

Hornbeck, R. and P. Keskin (2011). The evolving impact of the Ogallala Aquifer: Agricultural adaptation to groundwater and climate. *NBER Working Paper*.

Hsiang, S. M. (2010, August). Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proceedings of the National Academy of Sciences of the United States of America 107*(35), 15367–72.

Imbert, C. and J. Papp (2012). Equilibrium Distributional Impacts of Government Employment Programs: Evidence from India s Employment Guarantee. *mimeo*.

Iyengar, R., J. Monten, and M. Hanson (2011). Building Peace: The Impact of Aid on the Labor Market for Insurgents. *NBER Working Paper 17297*.

Jayachandran, S. (2006). Selling labor low: Wage responses to productivity shocks in developing countries. *Journal of Political Economy* (January), 1–60.

Jia, R. (2013). Weather Shocks , Sweet Potatoes and Peasant Revolts in Historical China. *Economic Journal*.

Johnson, D. (2009). Can Workfare Serve as a Substitute for Weather Insurance? The Case of NREGA in Andhra Pradesh. *Centre for Micro Finance Working Paper* (32).

Kapur, D., K. Gawande, and S. Satyanath (2012). Renewable Resource Shocks and Conflict in India's Maoist Belt. *mimeo*.

Khanna, G. and L. Zimmermann (2013). Guns and Butter ? Fighting Violence with the Promise of Development. (January).

Kudamatsu, M., T. Persson, and D. Strömberg (2012). Weather and Infant Mortality in Africa. *mimeo*.

Leon, G. (2009). Civil conflict and human capital accumulation: The long term effects of political violence in Peru. *mimeo*.

Leung, Y., Y. Ge, J. Ma, and J. Wang (2005). Measurement Errors and their Propagation in the Registration of Remote Sensing Images. In *Developments in Spatial Data Handling*, pp. 285–297. Springer Berlin Heidelberg.

Li, X.-H., Q. Zhang, and C.-Y. Xu (2012, March). Suitability of the TRMM satellite rainfalls in driving a distributed hydrological model for water balance computations in Xinjiang catchment, Poyang lake basin. *Journal of Hydrology 426-427*, 28–38.

Lilleor, H. and X. Giné (2005). Weather insurance in semi-arid India. *mimeo*, 1–24.

Magioncalda, W. (2010). A Modern Insurgency: Indias Evolving Naxalite Problem. Technical Report 140, Center for Strategic & International Studies, Washington, DC.

Miguel, E., S. Satyanath, and E. Sergenti (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy 112*(4).

Ministry of Rural Development (2009). Joint Convergence Guideline. Technical report, Ministry of Rural Development.

Mohan, A. (1992). The historical roots of the Kashmir conflict. *Studies in Conflict & Terrorism 15*(4), 283–308.

Niehaus, P. and S. Sukhtankar (2013a, November). Corruption Dynamics: The Golden Goose Effect. *American Economic Journal: Economic Policy 5*(4), 230–269.

Niehaus, P. and S. Sukhtankar (2013b, August). The marginal rate of corruption in public programs: Evidence from India. *Journal of Public Economics 104*, 52–64.

Nilakantan, R. and S. Singhal (2012). Naxalite Insurgency and the Economic Benefits of a Unique Robust Security Response .

Planning Commission (2003). Identification of Districts for Wage and Self employment programmes. Technical report, Planning Commission, New Delhi.

Rahman, H. and D. Sengupta (2007). Preliminary Comparison of Daily Rainfall from Satellites and Indian Gauge Data. *mimeo*.

Santos Silva, J. and S. Tenreyro (2006). The Log of Gravity. *The Review of Economics and Statistics 88*(4), 641–658.

Sarsons, H. (2011). Rainfall and conflict. *mimeo* (October).

Shapiro, J. N., M. Callen, E. Berman, and J. Felter (2011). Do Working Men Rebel? Insurgency and Unemployment in Afghanistan, Iraq, and the Philippines. *Journal of Conflict Resolution 55*(4), 496–528.

Silva, J. M. C. S. (2011). poisson : Some convergence issues. *mimeo*, 1–8.

Singh, P. (2013, January). Impact of Terrorism on Investment Decisions of Farmers: Evidence from the Punjab Insurgency. *Journal of Conflict Resolution*.

Skaugen, T. and J. Andersen (2010, July). Simulated precipitation fields with variance-consistent interpolation. *Hydrological Sciences Journal 55*(5), 676–686.

Smith, S. R., M. A. Bourassa, and M. Long (2011, July). Pirate attacks affect Indian Ocean climate research. *Eos, Transactions American Geophysical Union 92*(27), 225.

Vanden Eynde, O. (2011). Targets of Violence: Evidence from India's Naxalite Conflict. *mimeo* (November), 1–50.

Voors, M. J., E. E. M. Nillesen, P. Verwimp, E. H. Bulte, R. Lensink, and D. P. V. Soest (2012, April). Violent Conflict and Behavior: A Field Experiment in Burundi. *American Economic Review 102*(2), 941–964.

Zimmermann, L. (2012). Labor market impacts of a large-scale public works program: evidence from the Indian Employment Guarantee Scheme. *mimeo*.

Tables for the Main Text

Table 1: Before the Introduction of NREGA: Reduced Form Relationship between Rainfall, Agricultural Production, Wages and Violence

	Agricultural GDP		Agricultural Wages		Violence Intensity		Violence Incidence	
	(1) Annual Rain	(2) Monsoon	(3) Annual Rain	(4) Monsoon	(5) Annual Rain	(6) Monsoon	(7) Annual Rain	(8) Monsoon
Rainfall	0.537***	0.362***	0.060**	0.058***	-0.989**	-0.866***	-0.020**	-0.022***
	(0.114)	(0.086)	(0.025)	(0.018)	(0.421)	(0.270)	(0.009)	(0.007)
Observations	3239	3239	1419	1419	2841	2841	12657	12657
Number of Districts	471	471	314	314	148	148	543	543
Estimation	OLS	OLS	OLS	OLS	Poisson	Poisson	OLS	OLS

Notes: All regressions include region-phase-time fixed effects and district fixed effects. Agricultural GDP is in logs and measured in per capita terms using the 2001 Census population data. Agricultural wages is the average annual field worker wages in logs. Columns (5) - (7) use the one year lagged values of rainfall. For the linear models, standard errors are adjusted to reflect spatial dependence as modelled in Conley (1999). Spatial autocorrelation is assumed to linearly decrease in distance up to a cutoff of 500 km. District distances are computed from district centroids. Poisson regressions present standard errors clustered at the district level, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2: NREGA Introduction: Endogeneity of NREGA Rollout

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Any Violence	-0.474	-0.394	-0.370	-0.425	-0.237	-0.287	-0.166	-0.247
	(0.353)	(0.278)	(0.284)	(0.304)	(0.234)	(0.219)	(0.165)	(0.219)
Elasticity θ		-0.014	-0.016	0.001	-0.003	-0.006	-0.015	0.010
		(0.024)	(0.024)	(0.020)	(0.019)	(0.019)	(0.018)	(0.016)
Agricultural GDP per Capita			0.072	0.097**	0.112*	0.125*	0.143**	0.150***
			(0.095)	(0.048)	(0.065)	(0.068)	(0.057)	(0.043)
Mean Agricultural Wage				0.962***				0.388***
				(0.240)				(0.124)
Share Literate					0.030***	0.029***	0.028***	0.028***
					(0.001)	(0.001)	(0.003)	(0.010)
Scheduled Cast Share						-0.008	-0.030***	-0.019***
						(0.008)	(0.007)	(0.006)
Scheduled Tribe Share							-0.014***	-0.017***
							(0.004)	(0.005)
Number of Districts	544	470	470	213	470	470	470	213

Notes: Cross sectional regressions of district level controls on treatment-phase indicator. All time varying measures such as violence and agricultural GDP were measured in 2004, when the Planning Commission presented its latest report on backwardness. Standard errors are adjusted to reflect spatial dependence as modelled in Conley (1999). Spatial autocorrelation is assumed to linearly decrease in distance up to a cutoff of 500 km. District distances are computed from district centroids, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: After the Introduction of NREGA: Reduced Form Relationship between Rainfall, Agricultural Production, Wages and Violence

	Agricultural GDP		Agricultural Wages		Violence Intensity		Violence Incidence	
	(1) Annual Rain	(2) Monsoon	(3) Annual Rain	(4) Monsoon	(5) Annual Rain	(6) Monsoon	(7) Annual Rain	(8) Monsoon
Monsoon Rain	0.504***	0.374***	0.064***	0.062***	-1.195***	-1.330***	-0.031**	-0.035***
	(0.099)	(0.078)	(0.021)	(0.019)	(0.445)	(0.306)	(0.015)	(0.013)
NREGA x Monsoon Rain	-0.076	-0.132	-0.089***	-0.086***	1.100**	1.098***	0.032***	0.032***
	(0.091)	(0.083)	(0.017)	(0.016)	(0.509)	(0.388)	(0.012)	(0.010)
F-Test	3.68***	2.52***	-1.2	-1.3	-.41	-1.11	.06	-.21
Observations	4480	4480	2455	2455	8868	8868	25521	25521
Number of Districts	471	471	336	336	217	217	543	543
Estimation	OLS	OLS	OLS	OLS	Poisson	Poisson	OLS	OLS

Notes: All regressions include region-phase-time fixed effects and district fixed effects. Columns (5) - (7) use the one year lagged values of rainfall. For the linear models, standard errors are adjusted to reflect spatial dependence as modelled in Conley (1999). Spatial autocorrelation is assumed to linearly decrease in distance up to a cutoff of 500 km. District distances are computed from district centroids. Poisson regressions present standard errors clustered at the district level. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: After the Introduction of NREGA: Robustness of the Moderating Effect of NREGA on Agricultural Production and Wages

	Agricultural GDP			Agricultural Wages				
	(1) Placebo Rain	(2) Placebo Reform	(3) Controls	(4) Placebo Rain	(5) Placebo Reform	(6) Controls	(7) Harvesting	(8) Planting
Monsoon	0.073	0.364***	0.349***	-0.015	0.049	0.058***	0.051**	0.030
	(0.049)	(0.079)	(0.078)	(0.012)	(0.032)	(0.019)	(0.022)	(0.026)
NREGA x Monsoon	0.037	-0.020	-0.092	0.044***	-0.018	-0.098***	-0.069***	-0.040
	(0.028)	(0.062)	(0.067)	(0.012)	(0.026)	(0.017)	(0.027)	(0.031)
Observations	4480	4480	4480	2455	2455	2428	1987	1987
Number of Districts	471	471	471	336	336	331	280	280
Estimation	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS

Notes: All regressions include region-phase-time fixed effects and district fixed effects. Columns (1) and (4) use the rainfall outside the Monsoon season for a placebo test. Columns (2) and (5) present the results when shifting the reform 3 years ahead of time. Column (3) and (6) include a set of district characteristics interacted with a full set of time fixed effects as well as the NREGA treatment indicator. The district characteristics include the log of agricultural GDP per capita for the years prior to 2005, scheduled cast population share, share of literate population, scheduled tribe population share, household size and the gender gap. Standard errors are adjusted to reflect spatial dependence as modelled in Conley (1999). Spatial autocorrelation is assumed to linearly decrease in distance up to a cutoff of 500 km. District distances are computed from district centroids, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: After the Introduction of NREGA: Robustness of the Moderating Effect of NREGA on Violence

	(1) Previously Violent	(2) Placebo Rain	(3) Placebo Reform	(4) Controls	(5) Contemporaneous
Monsoon Rain	-1.289***	-0.067	-0.918**	-1.508***	-0.798***
	(0.311)	(0.258)	(0.406)	(0.257)	(0.309)
NREGA x Monsoon	1.044***	0.104	0.334	1.770***	0.515
	(0.402)	(0.310)	(0.400)	(0.332)	(0.363)
District Controls	No	No	No	Yes	No
Observations	6194	8868	8868	8733	9716
Number of Districts	151	217	217	214	222

Notes: All regressions are estimated using a pseudo-maximum likelihood estimator, whose moment conditions coincide with a Poisson model. The dependent variable throughout is the number of terrorist incidences. All regressions include region-phase-time fixed effects and district fixed effects. Columns (1) restricts the analysis to all districts that were violent before the introduction of NREGA. Column (2) uses the rain outside the Monsoon season as a placebo, while column (3) estimates the effect of a false NREGA reform three years before the actual one. Column (4) includes a set of district characteristics interacted with a full set of time fixed effects as well as the NREGA treatment indicator. The district characteristics include the log of agricultural GDP per capita for the years prior to 2005, scheduled cast population share, share of literate population, scheduled tribe population share, household size and the gender gap. Column (5) uses contemporaneous Monsoon rainfall. Robust standard errors clustered at the district level are given in the parentheses with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: NREGA Introduction: Relationship between Rainfall and NREGA Takeup

	NREGA Takeup			Costs of Active Projects	
	(1) Overall	(2) Extensive Margin	(3) Intensive Margin	(4) All Projects	(5) Irrigation Projects
Reduced Form:					
Monsoon rain	-0.216***	-0.055***	-0.118**	-0.231**	-0.237**
	(0.076)	(0.014)	(0.056)	(0.102)	(0.104)
Observations	3066	3060	3066	2825	2741
Number of Districts	538	537	538	501	500
Instrumental Variables:					
Agricultural GDP per	-1.647***	-0.312***	-1.063***	-1.484**	-1.891**
Capita	(0.468)	(0.104)	(0.372)	(0.616)	(0.794)
First Stage	22.7	22.6	22.7	18.2	16.8
Mean of Dependent Variable	14.7	.351	3.66	7.19	5.82
Observations	1664	1662	1664	1477	1414
Number of Districts	455	455	455	408	397

Notes: All regressions include region-phase-time fixed effects and district fixed effects. Column (1) uses the log of the cumulative person days until the end of a financial year as dependent variable, while column (2) uses the share of households that demanded employment as a dependent variable. The dependent variable in column (3) is the log of the number of days per household. Columns (4) and (5) use the log of the total costs of ongoing projects. All rainfall variables are lagged. The instrumental variables result use the lagged Monsoon rainfall as instrument for lagged agricultural output. "First Stage" is an F statistic for weak identification, reporting the minimum of either the Cragg-Donald or Kleibergen-Paap test statistic. Standard errors are adjusted to reflect spatial dependence as modelled in Conley (1999). Spatial autocorrelation is assumed to linearly decrease in distance up to a cutoff of 500 km. District distances are computed from district centroids, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: NREGA Introduction: Relationship between Rainfall and NREGA Takeup

	Just MPR Data		Controls		Scheduled Caste / Tribe Takeup	
	(1) Overall	(2) Projects	(3) Weather	(4) Controls	(5) Overall	(6) Weather
Lagged:						
Monsoon rain	-0.316***	-0.276**	-0.318***	-0.325***	-0.171**	-0.274***
	(0.092)	(0.136)	(0.084)	(0.087)	(0.078)	(0.091)
Outside Monsoon rain			-0.080*	-0.078**		-0.077*
			(0.042)	(0.036)		(0.039)
Hot days			-0.710	-0.837		-0.644
			(0.724)	(0.528)		(0.684)
Contemporaneous:						
Monsoon rain			-0.428***	-0.471***		-0.412***
			(0.109)	(0.098)		(0.112)
Hot days			-0.442	-0.628		0.031
			(0.640)	(0.498)		(0.619)
Outside Monsoon			-0.023	-0.045		-0.044
			(0.045)	(0.046)		(0.041)
District Controls	No	No	No	Yes	No	Yes
Observations	2079	1845	3066	3066	3061	3061
Number of Districts	535	479	538	538	538	538

Notes: All regressions include region-phase-time fixed effects and district fixed effects. Column (1) and (2) restrict the sample to cover only the period before the MIS data collection system was implemented. Column (3) adds contemporaneous weather controls in addition to the lagged ones. Column (4) and (6) include a set of district characteristics interacted with a full set of time fixed effects as well as the NREGA treatment indicator. The district characteristics include the log of agricultural GDP per capita for the years prior to 2005, scheduled cast population share, share of literate population, scheduled tribe population share, household size and the gender gap. Column (5) and (6) restrict the analysis to estimate the responsiveness of overall participation by scheduled caste or scheduled tribes. Standard errors are adjusted to reflect spatial dependence as modelled in Conley (1999). Spatial autocorrelation is assumed to linearly decrease in distance up to a cutoff of 500 km. District distances are computed from district centroids, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8: NREGA and the Targets of Violence

	(1) Civilians	(2) Security Forces	(3) Terrorists	(4) Share of Civilian
Monsoon	-1.517***	-1.076***	-1.033***	-0.134**
	(0.335)	(0.396)	(0.362)	(0.053)
NREGA x Monsoon	1.445***	1.419***	0.596	0.094**
	(0.407)	(0.505)	(0.399)	(0.041)
NREGA				-0.686**
				(0.287)
Observations	7894	5111	6220	2521
Number of Districts	197	136	150	217

Notes: All regressions are estimated using a pseduo-maximum likelihood estimator, whose moment conditions coincide with a Poisson model. The dependent variable in columns (1)-(3) is the number of incidents where the subject of the incident has been coded to be either a civilian, security force or terrorist. Column (4) uses the share of incidents with civilian targets. Regressions (1) - (3) include region by phase-time fixed effects and district fixed effects, while column (4) uses time- and district fixed effects. Robust standard errors clustered at the district level are given in the parentheses with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 9: Dynamic Versus Direct Level Effect of NREGA

	(1) $\eta = \gamma = 0$	(2) $\eta = \gamma$	(3) Unconstrained
NREGA	-0.427***	-0.481***	-0.334**
	(0.164)	(0.174)	(0.167)
Monsoon Rain		-0.766***	-1.609***
		(0.233)	(0.343)
NREGA x Monsoon			1.305***
			(0.351)
Observations	9597	9597	9597
Number of Districts	217	217	217

Notes: All regressions are estimated using a pseduo-maximum likelihood estimator, whose moment conditions coincide with a Poisson model. The dependent variable is the number of violent incidences per quarter. All regressions include time fixed effects, district fixed effects. The first column does not control for rainfall, while the second column constraints the rainfall coefficient to be the same before, and after the introduction of NREGA. Robust standard errors clustered at the district level are given in the parentheses with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 10: Level Effect of NREGA

	Level Effect Estimates			Heterogeneity of Level Effect					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
NREGA	-0.334**	-0.263*	-0.492**	-0.949***	-0.756***	-0.480**	-0.479***	-0.463**	-0.462*
	(0.167)	(0.148)	(0.197)	(0.231)	(0.266)	(0.204)	(0.173)	(0.197)	(0.263)
$\text{NREGA}_{\text{t-1}}$		-0.468**							
		(0.194)							
Heterogeneity: NREGA $\times$									
Scheduled Tribe				0.246***					0.300***
				(0.063)					(0.100)
Scheduled Caste					-0.434				0.867*
					(0.323)				(0.483)
Literacy						-0.541			-0.238
						(0.850)			(0.784)
Agricultural GDP Before 2005								1.063	2.498*
								(1.519)	(1.397)
Householdsize							3.164***		3.067***
							(1.163)		(1.051)
NREGA Dynamic Effect									
Monsoon Rain	-1.609***	-1.572***	-1.858***	-1.846***	-1.885***	-1.856***	-2.002***	-1.840***	-1.875***
	(0.343)	(0.337)	(0.422)	(0.412)	(0.420)	(0.419)	(0.425)	(0.421)	(0.414)
NREGA x Monsoon	1.305***	1.337***	1.584***	1.571***	1.660***	1.568***	1.852***	1.550***	1.585***
	(0.351)	(0.356)	(0.485)	(0.461)	(0.475)	(0.479)	(0.487)	(0.486)	(0.467)
District Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9597	9597	8248	8248	8248	8248	8248	8248	8248
Number of Districts	217	217	188	188	188	188	188	188	188

Notes: All regressions are estimated using a pseudo-maximum likelihood estimator, whose moment conditions coincide with a Poisson model. The dependent variable is the number of violent incidences per quarter. All regressions include region by time fixed effect and district fixed effects. "District Controls" includes a full set of cross sectional district characteristics interacted with time fixed effects. The district characteristics include the log of agricultural GDP per capita for the years prior to 2005, scheduled cast population share, share of literate population, scheduled tribe population share, household size and the gender gap. Robust standard errors clustered at the district level are given in the parentheses with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 11: Level Effect of NREGA and Targets of Violence

	(1) Overall	(2) Civilians	(3) Security Forces	(4) Terrorists
NREGA	-0.573***	-0.657***	-0.194	-0.436
	(0.200)	(0.208)	(0.263)	(0.266)
District Controls	Yes	Yes	Yes	Yes
Observations	8836	7896	5170	5922
Number of Districts	188	168	110	126

Notes: All regressions are estimated using a pseduo-maximum likelihood estimator, whose moment conditions coincide with a Poisson model. The dependent variable in columns (1)-(3) is the number of incidents where the subject of the incident has been coded to be either a civilian, security force or terrorist. All regressions include time- and district fixed effects. “District Controls” includes a full set of cross sectional district characteristics interacted with time fixed effects. The district characteristics include the log of agricultural GDP per capita for the years prior to 2005, scheduled cast population share, share of literate population, scheduled tribe population share, household size and the gender gap. Robust standard errors clustered at the district level are given in the parentheses with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix Tables

Table A1: Before the Introduction of NREGA: Robustness of Relationship between Weather Variables and Agricultural Output

	Agricultural GDP			Grain Production		
	(1) Outside Monsoon	(2) Temperature	(3) Controls	(4) Outside Monsoon	(5) Temperature	(6) Controls
Monsoon	0.364***	0.357***	0.293***	0.369***	0.367***	0.234***
	(0.086)	(0.086)	(0.075)	(0.076)	(0.076)	(0.057)
Outside Monsoon	0.122**			0.114***		
	(0.051)			(0.043)		
Hotdays		-1.047			-0.290	
		(0.653)			(0.509)	
Vegetation Index			6.621***			7.394***
			(1.072)			(1.000)
Nightlights			-0.061			-0.061
			(0.207)			(0.133)
District Controls	No	No	Yes	No	No	Yes
Observations	3239	3239	3239	3196	3196	3196
Number of Districts	471	471	471	464	464	464

Notes: All regressions include region-phase-time fixed effects and district fixed effects. Hotdays measures the number of days per year with average temperatures above 30°. Vegetation Index is the annual MODIS Satellite Normalised Vegetation Index that may serve as a proxy for forest cover or agricultural productivity. Nightlights measures the share of the district that emits stable night lights in a given year. "District Controls" includes a full set of cross sectional district characteristics interacted with time fixed effects. The district characteristics include the log of agricultural GDP per capita for the years prior to 2005, scheduled cast population share, share of literate population, scheduled tribe population share, household size and the gender gap. Standard errors are adjusted to reflect spatial dependence as modelled in Conley (1999). Spatial autocorrelation is assumed to linearly decrease in distance up to a cutoff of 500 km. District distances are computed from district centroids, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2: Before the Introduction of NREGA: Robustness of Relationship between Weather Variables and Agricultural Wages

	Annual Wages			Seasonal Wages	
	(1) Outside Monsoon	(2) Temperature	(3) Controls	(4) Harvesting	(5) Planting
Monsoon	0.058***	0.060***	0.053***	0.046*	0.010
	(0.019)	(0.018)	(0.019)	(0.028)	(0.017)
Outside Monsoon	0.006				
	(0.011)				
Hotdays		0.265			
		(0.162)			
Vegetation Index			-0.170		
			(0.256)		
Nightlights			-0.119*		
			(0.065)		
District Controls	No	Yes	Yes	No	No
Observations	1419	1419	1419	1387	1195
Number of Districts	314	314	314	318	260

Notes: All regressions include region-phase-time fixed effects and district fixed effects. Hotdays measures the number of days per year with average temperatures above 30°. Vegetation Index is the annual MODIS Satellite Normalised Vegetation Index that may serve as a proxy for forest cover. Nightlights measures the share of the district that emits stable night lights in a given year. "District Controls" includes a full set of cross sectional district characteristics interacted with time fixed effects. The district characteristics include the log of agricultural GDP per capita for the years prior to 2005, scheduled cast population share, share of literate population, scheduled tribe population share, household size and the gender gap. Standard errors are adjusted to reflect spatial dependence as modelled in Conley (1999). Spatial autocorrelation is assumed to linearly decrease in distance up to a cutoff of 500 km. District distances are computed from district centroids, stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3: Before the Introduction of NREGA: Robustness of Relationship between Weather Variables and Conflict

	Robustness to Choice of Empirical Model				Robustness to Controls		
	(1) Poisson-IV	(2) Poisson	(3) Negative Binomial	(4) OLS	(5) Non-Monsoon Rain	(6) Further Weather	(7) District Controls
Lagged:							
Agricultural GDP	-2.427**						
	(0.989)						
Monsoon Rain		-0.855***	-0.796***	-0.091***		-0.962***	-0.656**
		(0.272)	(0.205)	(0.031)		(0.296)	(0.289)
Outside Monsoon Rain					-0.205	-0.273	-0.174
					(0.211)	(0.212)	(0.215)
Hotdays						2.212	1.646
						(1.853)	(2.128)
Contemporaneuous:							
Monsoon Rain						0.303	0.384
						(0.276)	(0.319)
Hotdays						2.778	1.489
						(2.083)	(1.905)
District Trends	No	No	No	No	No	No	Yes
Observations	2213	2841	3312	12657	2841	2841	2630
Number of Districts	120	148	148	543	148	148	143

Notes: All regressions include region-phase-time fixed effects and district fixed effects. Column (1) presents the results of an IV regression, instrumenting lagged agricultural GDP per capita with lagged Monsoon rainfall. Column (3) presents a negative-binomial model with bootstrapped standard errors. Hotdays measures the number of days per year with average temperatures above 30°. "District Controls" includes a full set of cross sectional district characteristics interacted with time fixed effects. The district characteristics include the log of agricultural GDP per capita for the years prior to 2005, scheduled cast population share, share of literate population, scheduled tribe population share, household size and the gender gap. Robust standard errors clustered at the district level are given in the parentheses with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4: Before the Introduction of NREGA: Robustness of Relationship between Weather Variables and Conflict

	Rainfall Measures		Temperature		Vegetation Index	
	(1) Deficiency	(2) Normalised Deficiency	(3) Days above 30○	(4) Temperature	(5) OLS	(6) IV
Monsoon rain	1.033***	-0.192***				
	(0.281)	(0.046)				
Days above 30°			2.740			
			(1.770)			
Average Monsoon Temperature				0.514**		
				(0.261)		
Vegetation					-13.534**	-83.822***
					(5.914)	(26.670)
Observations	2841	2841	2841	2841	2841	2841
Number of Districts	148	148	148	148	148	148
Estimation	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson

Notes: All regressions include region-phase-time fixed effects and district fixed effects. Hotdays measures the number of days per year with average temperatures above 30°. Column (6) presents an IV model, instrumenting the lagged vegetation index with lagged monsoon rainfall as suggested by Kapur et al. (2012). Robust standard errors clustered at the district level are given in the parentheses with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A5: NREGA Introduction: Relationship between Rainfall and NREGA Takeup

	MPR Data Ongoing Project Costs		Temperature, Trends and Controls			Other Project Measures	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All Projects	Irrigiation Projects	Trends	Nightlights	Weather Controls	Active Costs	Active Count
Monsoon rain	-0.244**	-0.270**	-0.153*	-0.211***	-0.292***	-0.134**	-0.071
	(0.111)	(0.132)	(0.079)	(0.082)	(0.091)	(0.057)	(0.056)
Days above 30°					-2.174***		
					(0.758)		
Night Lights				-0.058			
				(0.249)			
Monsoon rain (contemporaneuous)					-0.228**		
					(0.101)		
Days above 30° (contemporaneuous)					-1.263**		
					(0.600)		
District Time Trend	No	No	Yes	No	No	No	No
Observations	1849	1775	2897	2897	2897	2897	2893
Number of Districts	474	461	529	529	529	529	529

Notes: All regressions include region-phase-time fixed effects and district fixed effects. “First Stage” is an F statistic for weak identification, reporting the minimum of either the Cragg-Donald or Kleibergen-Paap test statistic. Columns (5) - (7) use the one year lagged values of rainfall. Robust standard errors clustered at the district level are given in the parentheses with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A Online Appendix

A.1 Conflict Data

Empirical research on the economics of conflict almost always suffer from severe data limitations. This lies in the nature of the subject of study, that typically places that exhibit conflict are only weakly institutionalised with little official report of violence and little press and media coverage. Blattman and Miguel (2009)'s review cites that the correlation across different civil war datasets ranges from 0.42 to 0.96, which may be the reason why empirical results are often not reproducible using similar identification strategies, but different datasets or variable definitions (e.g. Ciccone (2011)).

There exists no broad conflict dataset that covers India or South East Asia as a whole. This gap was filled through the violence dataset introduced in Fetzer (2013). This paper documents the process through which in the Indian context 28,638 newspaper reports were transformed into a workable conflict dataset using both machine-learning, semi-automated coding techniques and scalable manual hand-coding methods.[28] This section sketches the semi-automated process through which the daily newspaper clippings are transformed (more details are provided in Fetzer (2013)). A typical sample may look as follows:

> Two unidentified terrorists massacred six members of a family and left a seventh injured at Mangnar Top, Poonch district, on December 31, 2001. Local residents refused to cremate the bodies of the slain victims, insisting that a Union Minister should visit the area and take notice of the increasing terrorist violence there.

The semi-automated routine defines a terrorist-incident as an Event-tuple, $E = \{L, T, V, S, O\}$ defined by a location L, a date or time of the event T, a verb V that indicates the type of violent act, and the verb's associated subject S, the perpetrator of the act and the object O that was subjected to the act V. The semi-automated routine tries to fill all these elements of the tuple for each sentence using common machine-learning algorithms implemented in natural language processing packages.

In the above text-snippet, only one sentence satisfies the requirement of all

[28] The raw material was a set of 28,638 newspaper clippings collected by the Institute for Conflict Management in New Delhi through the South Asian Panel on Terrorism (SATP) since 2001, see `http://www.satp.org`, accessed in October 2012.

elements being present, yielding:

$$\underbrace{\text{Two unidentified terrorists}}_{\text{subject}}\ \overbrace{\text{massacred}}^{\text{verb}}\ \overbrace{\text{six members of a family}}^{\text{object}}$$

$$\text{and}\ \overbrace{\text{left}}^{\text{verb}}\ \overbrace{\text{a seventh}}^{\text{object}}\ \overbrace{\text{injured}}^{\text{verb}}$$

$$\overbrace{\text{at Mangnar Top, Poonch district, on December 31, 2001.}}^{\text{time \& location}}$$

which is transformed into:

$$\begin{aligned} E_1 &= \{'\text{Mangar Top Poonch}','\text{December 31 2001}',\\ &\quad '\text{massacre}','\text{two unidentified terrorists}',\\ &\quad '\text{six members of a family at Mangnar Top, Poonch district}'\} \end{aligned}$$

An incident is counted as long as all pieces of information can be deduced from the underlying sentence. This is essentially mimicking the process through which humans would code this data manually. An exhaustive list of verbs is used to spot events and a sentence is normalised to contain at most one event. The individual elements of the tuple E are then transformed by assigning labels to the snippets indicating whether the actor was a terrorist, security force or a civilian and similarly for who subjected to the act V.[29]

The data has been evaluated in Fetzer (2013) and correlates very well with hand-coded data. The correlation between this automatically retrieved data and the hand-coded data for the Naxalite conflict used by Vanden Eynde (2011) is at least 93%.

A.2 Comparison of Results with Global Terrorism Database

This section highlights that the results obtained in my paper can not be replicated when studying the conflict for India contained in the Global Terrorism Database (GTD) collected by National Consortium for the Study of Terrorism and Responses to Terrorism at the University of Maryland. This database has been used in more than 30 journal publications and thus, serves as an interesting testing ground.

[29] Note that in the sentence there exists a further event $E_2 = \{'\text{Mangar Top Poonch}','\text{December 31 2001}','\text{left}','\text{two unidentified terrorists}','\text{a seventh injured at Mangnar Top, Poonch district}'\}$. As described in Fetzer (2013), a sentence will be counted as containing information of at most one incident.

To begin with, I estimate the main specifications using the number of terrorist incidences in the global terrorism database as a left-hand side.

Table A6: NREGA Effect in the GTD and Fetzer (2013) dataset

	Fetzer (2013) Dataset			Global Terrorism Database		
	(1) Pre NREGA	(2) Dynamic	(3) Level	(4) Pre NREGA	(5) Dynamic	(6) Levels
Monsoon	-0.866***	-1.330***	-0.680***	-0.985	-1.338*	-1.062**
	(0.270)	(0.306)	(0.261)	(0.684)	(0.764)	(0.462)
NREGA x Monsoon		1.098***			0.359	
		(0.388)			(0.676)	
NREGA			-0.540***			-1.098
			(0.166)			(1.264)
Observations	2841	8868	10199	851	5268	5268
Number of Districts	148	217	217	57	186	186

Notes: All regressions are estimated using a pseduo-maximum likelihood estimator, whose moment conditions coincide with a Poisson model. Regressions in columns (1)-(2) and (4)-(5) include region-phase-time fixed effects as well as district fixed effects, while results for columns (3) and (6) come from a regression with time- and district fixed effects. The dependent variable is the number of incidences per district and quarter. Robust standard errors clustered at the district level are given in the parentheses with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Columns (1)-(3) study the dataset used in this paper, while columns (4)-(6) use the GTD database. In column (4) it becomes obvious that in the GTD data, there appears to be no statistically significant correlation between rainfall and conflict, while there is a strong documented in the Fetzer (2013) data in column (1). The geographic coverage of the GTD dataset is also a lot more limited before the introduction of NREGA, with only 57 districts reported as having violent incidences before NREGA was introduced while there are almost three times as many districts reported in the other datasets. The moderating effect of NREGA is seen only in column (2), but not in column (5), albeit the coefficient is positive.

This is a source of concern unless the data coverage in the two datasets varies systematically in a way that is correlated with rainfall variation. A simple way to answer this question is to evaluate the two datasets by regressing one on the other and seeing what are the chances of an incident reported in one dataset to be represented in the other dataset and how this relationship varied over time, with rainfall variation and with the interaction of NREGA.

First, I explore the simple relationship over time by estimating:

$$GTD_{dt} = \delta_d + b_{rt} + \sum_{t=2000}^{2010} \gamma_t A_{dt} + \epsilon_{dt}$$

I plot the coefficients γ_t in Figure A1.

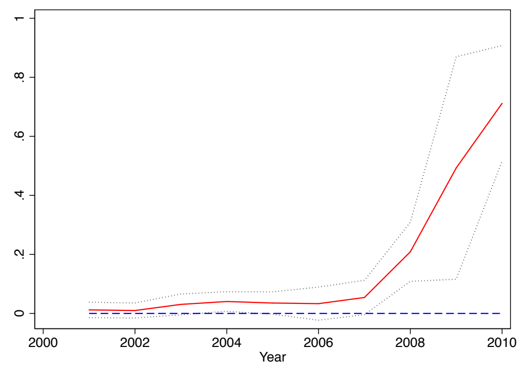


Figure A1: Relationship between Fetzer (2013) and GTD Data over Time

The specification, by using district- and region x time fixed effects takes out any fixed- conflict region and time varying reporting differences, while the district fixed effects remove any time-invariant district specific reporting biases. The coefficients paint a very stark picture: the datasets do not compare well at all before 2007. The good news is that the coefficients are consistently positive, suggesting that the overall correlation is positive. However, the point estimates are very small and only sometimes statistically significantly different from zero. This suggests that int he earlier years it is extremely unlikely for an incident captured in one dataset to appear in the other. In more recent years, the data become increasingly similar.

Why have the two datasets converged? It appears that the underlying data source in the GTD database has evolved significantly over time. Since 2008, the SATP reports feed into the GTD database, while before that the GTD database was mainly fed by newswire services. By 2010, more than 53% of the incidences in the GTD database were directly referenced with a report from the SATP newspaper clippings dataset. This is clearly, a lower bound since for many reports in the GTD dataset one can manually find references in the SATP dataset, but not necessarily vice versa.

While the level of violence reported in the GTD database seems to be significantly lower for early years, it is important for the identification whether this mismatch in reporting is correlated with rainfall realisations.

In order to explore this, I measure the differences and the absolute value of the differences between the two datasets and run the three specifications from above again.

The results are presented in table A7. The coefficients suggest that a positive rainfall realisation in the preceding month is significantly correlated with a lower reporting difference, i.e. implying that the mismatch between the Fetzer (2013) dataset and the GTD dataset is smaller. This highlights that reporting is likely to be endogenous to past weather and thus, past income realisations. While this is something that can fundamentally, not be checked, I believe that this is more likely to be a problem for the GTD database, where reporting has been found to correlate with Foreign Direct Investment in Fetzer (2013). The introduction of NREGA appears to have further reduced the mismatch between the two datasets.

Table A7: Evolution of Reporting Differences between GTD and Fetzer (2013) datasets

	Reporting Difference			Absolute Value of Reporting Difference		
	(1) Pre NREGA	(2) Dynamic	(3) Level	(4) Pre NREGA	(5) Dynamic	(6) Levels
Monsoon	-0.078**	-0.090**		-0.107***	-0.136***	
	(0.032)	(0.036)		(0.030)	(0.034)	
NREGA x Monsoon		0.051			0.060	
		(0.042)			(0.043)	
NREGA		-0.398	-0.048		-0.503*	-0.094*
		(0.269)	(0.055)		(0.278)	(0.050)
Observations	12657	25521	27693	12657	25521	27693
Number of Districts	543	543	543	543	543	543

Notes: All regressions are simple linear regressions with time- and district fixed effects. Robust standard errors clustered at the district level are given in the parentheses with stars indicating *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

If we take this and the previous results together, this suggests that there is some systematic differences to the GTD dataset which correlates with rainfall in a systematic way and the introduction of NREGA may have lead to a moderation of this reporting difference. Since the two datasets appear to be converging over time and the coverage of the GTD dataset actually expanding, it seems reasonable to conclude that the SATP data source on which the Fetzer (2013) dataset is a more consistent way to measure conflict.

A.3 TRMM Rainfall Data

This paper is the first one in economics to use data from the Tropical Rainfall Measuring Mission (TRMM) satellite, which is jointly operated by the National Aeronautics and Space Administration (NASA) and the Japan Aerospace and Exploration Agency (JAXA). The satellite carries a set of five instruments to construct gridded rainfall rates at very high spatial and temporal resolution.

Remotely sensed weather data is an important source of data, in particular, for less developed countries, where observational data is scarce. This is particularly relevant in the case of India, where observational weather may vary dramatically in systematic ways. There are three main drawbacks. First, most observations come from rain gauges, where measurements are taken once a day. Climatologist are concerned about rain gauges in particular in tropical- or subtropical areas, since most rainfall is convective. Such convective rainfalls are highly local, generating intermittent and scattered rainfall, which may not be picked up using rain gauges, if the network is not spatially fine enough. The TRMM satellite orbits the earth every 90 minutes, thus providing multiple observations each day. An alternative is to consider data from weather radars. Rainfall radar may provide estimates for rainfall in a radius of 200 km around the station, however it is unreliable for distances in excess of 200 km. In the Indian case, rainfall radar data is not made available and would be problematic, since most reporting radar stations are clustered along the coast. The third general concern regarding observational weather data is the fact that reporting may be endogenous e.g. to violence or other variables that are correlated with the dynamics of violence. This has been found to be the case in hilly regions. This has been highlighted recently by Smith et al. (2011), who show that Somalian piracy has generated a "black hole" in the Indian ocean, where observational weather data from merchant vessels is not available anymore, as vessels take routes avoiding piracy infested areas.[30]

The TRMM Multi-Satellite Precipitation Analysis provides daily rainfall from 1998 to 2012 at a fine spatial resolution of 0.25 by 0.25 degree grid-cell size. The data from the various instruments aboard the satellite are cleaned and calibrated using additional data from the accumulated Climate Assessment and Monitoring System (CAMS) and the Global Precipitation Climatology Centre (GPCC) rain gauge analysis data. The output of the algorithm are 3-hourly rainfall rates for that time-period. This is then scaled up to obtain monthly mean precipitation rates, which in turn are transformed into overall monthly rainfall.

Any data product derived from remote sensing suffer from a problem known as "error propagation" (see Leung et al. (2005)). As the raw data is transformed in the analytical process, simple small measurement errors may be propagated due

[30]See Besley et al. (2012) for a discussion. Another example is the case of Vanden Eynde (2011), who had to merge several districts together in order to obtain consistent rainfall estimates, since many stations simply fail to report rainfall estimates. Most of these stations are located in places with conflict or in newly created districts or states.

to the mathematical and numerical transformations. As remote sensing products are using some machine-learning algorithms to classify observations, the error propagation is specific to the algorithm used for classification (see Burnicki et al. (2007)). In order to take this problem into account, I obtain the PERSIANN precipitation dataset that is based on the TRMM Multi-Satellite Data as well. This dataset is used to confirm the main results of this paper and can address the potential issues regarding error-propagation since the type of error propagation is specific to which algorithm is used.[31]

A.4 Temperature Reanalysis Data

As a solution to the problem of limited data availability for ground measurements, I construct temperature readings from a gridded daily reanalysis dataset that uses remote sensing data and sophisticated climate models to construct daily temperature on a 0.75° (latitude) x 0.75° (longitude) grid (equivalent to 83km x 83km at the equator).[32] The ERA-Interim reanalysis is provided by the European Centre for Medium-Term Weather Forecasting (ECMWF).[33] As the grid is significantly coarser than the rainfall data, I construct inverse distance weighted daily mean temperatures for all grid points within 100 km of the geographic centre of each district. The weighting used is the inverse of the distance squared from the district centroid.

A.5 Agricultural Production and State Level Harvest Prices

For every district, I only consider crops that have been consistently planted on at least 1000 acres for the period that the state reports agricultural production to the data dissemination service of the Directorate of Economics and Statistics with the Ministry of Agriculture.[34] This leaves the following crops: bajra, barley, castorseed, chilly, cotton, gram, groundnut, jowar, jute, linseed, maize, mesta, potato, ragi, rapeseed, rice, sesamum, sugarcane, tobacco, tumeric, tur-arhar and wheat.

For each of these crops, I obtained state-level farm harvest prices to compute a district level measure of the agricultural GDP. Unfortunately, district level harvest prices were not available throughout or only for a limited number of crops that

[31] The TRMM data is from the algorithm 3B-43 (version 7), while the latest PERSIANN precipitation data was obtained from `http://chrs.web.uci.edu/persiann/` on 22.04.2013.

[32] To convert degrees to km, multiply 83 by the cosine of the latitude, e.g at 40 degrees latitude 0.75 x 0.75 cells are 83 x cos(40) = 63.5 km x 63.5 km.

[33] See Dee et al. 2011 for a detailed discussion of the ERA-Interim data.

[34] This data is available on `http://apy.dacnet.nic.in/cps.aspx`, accessed 14.12.2013.

did not match well with the actual planted crops. For that reason, I stuck with the state-level prices. The resulting dataset is an unbalanced panel, since not all states consistently report data to the Ministry of Agriculture information systems.

Linear Relationship between Agricultural Output and Monsoon Rain Many papers on the relationship between agricultural incomes and violence use different transformations on the rainfall variable. A common form that these take is to considering only rainfall below or above a certain threshold as constituting a negative productivity shock. In the case of India, the relationship between rainfall and output is however fairly monotone. To highlight this, I estimated the production function using local-linear regression method developed by Fan (1992). I first demean the data by the region-time fixed effects as well as the district-fixed effects and then, estimate the following local linear model:

$$\min_{\eta} \sum_{i=1}^{n} (\tilde{y}_i - (R - \tilde{R}_i)'\theta)^2 K(\frac{\tilde{R} - \tilde{R}_i}{h_n})$$

where $\tilde{y}_i$ and $\tilde{R}_i$ are the residuals after removing district- and time fixed effects. The rainfall variable $\tilde{R}$ is evaluated at 50 grid points around which a linear regression is estimated. This provides a sequence of estimates of θ that can be plotted. The results are depicted in Figure A2.

It becomes clear that the gradient suggests a monotonically increasing relationship between agricultural income and rainfall. The standard errors are estimated using a cluster-bootstrap procedure and the results indicate that one can not rule out that abundant rainfall is correlated with lower agricultural production as well. However, it is unlikely that the existing rainfall data is able to pick up local flash floods sufficiently well, as the spatial resolution is simply too coarse.

A.6 Agricultural Wages in India

This appendix describes the process of how the agricultural wage data was cleaned and put in shape for the analysis in the paper. The data is of variable and sometimes questionable quality.

The raw data gives monthly wages for male, female and children, broken into skilled- and unskilled agricultural labour and different types of labour. The types of skilled labour are blacksmith, carpenter and cobbler, while unskilled labour combines ploughman, reaper/harvester, sower, weeder, other agricultural labour.

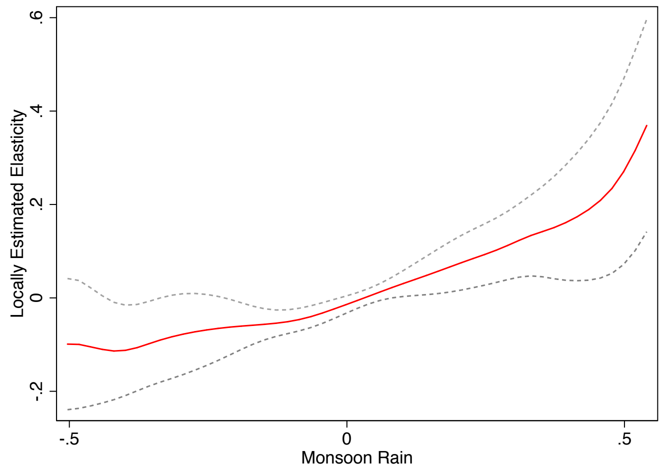


Figure A2: Local Linear Regression Relationship between Monsoon Rainfall and Agricultural GDP

In some states, these separate unskilled labour categories are not reported[35], but rather, a category "Field Labour Wages" is reported. This is conceived to be an average of the different categories.

In some districts these wages are reported throughout the year, while in others the wages are reported only in the parts of the year, when particular activities are actually carried out (i.e. sowing wages in the early Kharif season of May, June and July), while harvesting wages are reported in the fall of a given year.

After digitising and entering the raw data, I proceed to construct a quarterly level agricultural field-labour wage as my main dependent variable.

For each district, there may be multiple wage-observations in case there are multiple reporting centres. I generate a balanced panel requiring each quarter of the year to have at least one non-missing observation of agricultural wages belonging to the particular category of unskilled labour. I then construct the simple average across these wage-observations.

There are advantages and disadvantages to this approach. In particular, by construction, this implies that within a year, some field labour wage observations are noisier then others. This can be taken into account by adequately weighting the observations.

As an alternative, I can impose the requirement that there be at least one observation for each different unskilled labour category within a quarter. This condition is very stringent, as it fails to recognise the types of agricultural activi-

[35]The states for which this is the case are Andhra Pradesh, Karnataka and Maharashtra.

ties that are pursued during a year. This approach reduces the number of districts significantly, but the results remain the same.

The southwest monsoon typically enters the mainland over Kerala in the first week of June. It moves northward to cover the whole of India by mid-July. It starts withdrawing from mid-September. The southwest monsoon is critical to the development of Indian agricultural production. The southwest monsoon provides 80 percent of India's total precipitation and is critical to the development of its major food and commercial crops such as rice, coarse grains, pulses, peanuts, soybeans and cotton. Planting of the largely rainfed Kharif (monsoon season) crops, which include rice, sorghum, corn, millet, peanut, soybean and cotton will begin after the monsoon firmly establishes itself over the major producing states and planting will continue through July and early August. Farmers in the northern rice surplus states of Punjab and Haryana, where irrigation is available, often complete rice transplanting prior to the monsoon arrival. This season's pre-monsoon, or early season rains in central, south and east India should provide a favorable early season planting conditions for rice, soybeans, sorghum and corn. The country's economy is to a large extent dependent on monsoon rains.

A.7 NREGA Data Sources and Roll Out

The data for the roll-out of NREGA come from the Ministry of Rural Development, which is responsible for administering the scheme. The sequence of roll-out was highly endogenous to a set of district level backwardness characteristics, such as the share of scheduled caste, scheduled tribe population, baseline agricultural productivity, literacy and existing levels of conflict. This becomes obvious when considering Figure A3. This picture highlights that a lot of districts in the east of India received NREGA in the first round. A lot of these districts did suffer from Naxalite violence. As discussed in the main body, I do not require exogeneity of treatment to levels of violence for my empirical design.

There are two main sources for data on NREGA take-up. These are the district-level monthly-progress reports (MPR) and data coming from the Management Information System (MIS). The latter is a completely non-paper based system that has only become mandatory to use in the financial year but was still not fully operational until 2010-2011.

There are a lot of issues regarding the reliability of either datasets, as there is quite some mismatch between the two datasets, especially in the earlier years

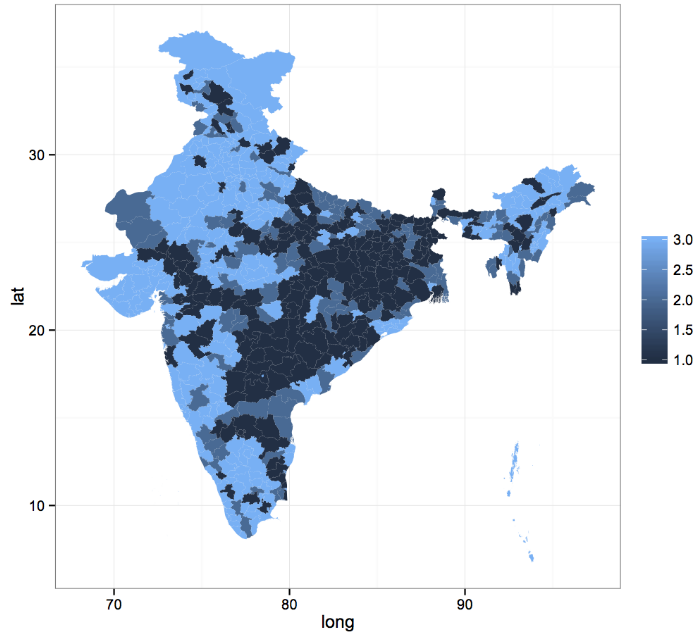


Figure A3: Phases of the NREGA Roll-out across India

when the MIS was introduced.[36] This may be due to partial compliance in the MIS after it had been introduced, but could be also because the MPR system is more subject to manipulation. It is difficult to asses the underlying divergence in the two databases.

The MPR data is available continually from 2006 to the financial year 2010-2011, from which point onwards I rely on data from the MIS.[37] The format of the reports has changed considerably, with the major break occurring in 2011. This is partly due to the evolving nature of NREGA. Ministry of Rural Development (2009) details that several programs by the Ministry of Water Resources are to be joined with the NREGA by 2011. An important part of this program are rural sanitation projects that are funded by the Ministry of Water Resources for a set of targeted districts. This implies that there are district-specific breaks in the NREGA data. In the empirical specifications which combine data from before and after 2011, I flexibly control for these breaks by introducing a district specific fixed effect.

[36]See for example mismatch between MIS data and National Sample Survey returns data highlighted by `http://www.indiatogether.org/2013/jun/gov-nregs.htm`,accessed on 12.06.2013.

[37]Thanks to Clement Imbert for sharing NREGA MPR data for the earliest years.

I focus on five key variables: for the take-up I study cumulative person days provided, cumulative number of (distinct) households provided employment as well as the number of days per household at the district level. I also look at the number of person days for scheduled caste and scheduled tribe populations, as well as the share of person days that accrue to females.

For the NREGA project measures, I study the total cost or number of ongoing projects at the end of each financial year. I consider all projects together or specifically, only projects that are catered towards irrigation.[38]

Despite having access to NREGA for many months in a financial year, I only study the reported metrics at the end of each financial year (that is March of each calendar year). This becomes necessary as there are significant reporting delays which induce large jumps in the cumulative month on month measures which are less likely driven by participation, but more likely due to reporting issues.

I construct the NREGA take-up, participation and project data to match the Monsoon calendar as in the main exercises. Since the financial year commences each April, the contemporaneous Monsoon variable is more likely to be significant, as it may drive take-up during from September onwards.

[38] The categories in the data that are consistently reported are: "Micro Irrigation Works","Drought Proofing","Water Conservation and Water Harvesting","Provision of Irrigation facility to Land Owned by Scheduled Caste/ Scheduled Tribe".