SOCIAL FRICTIONS TO KNOWLEDGE DIFFUSION:
EVIDENCE FROM AN INFORMATION INTERVENTION

Arthur Alik-Lagrange
Martin Ravallion

ABSTRACT

Does knowledge about antipoverty programs spread quickly within poor communities or are there
significant frictions, such as due to social exclusion? We combine longitudinal and intra-household
observations in estimating the direct knowledge gain from watching an information movie in rural
India, while randomized village assignment identifies knowledge sharing with those in treatment villages
who did not watch the movie. Knowledge is found to be shared within villages, but less so among
illiterate and lower caste individuals, especially when also poor; these groups relied more on actually
seeing the movie. Sizable biases are evident in impact estimators that ignore knowledge spillovers.

Arthur Alik-Lagrange
Toulouse School of Economics
arthur.alik@tse-fr.eu

Martin Ravallion
Center for Economic Research
Georgetown University
ICC 580
Washington, DC 20057
and NBER
mr1185@georgetown.edu

1.   Introduction

The literature on antipoverty programs has often stressed the informational constraints facing policy makers in reliably knowing who is eligible. However, another information problem looms large: that facing potential beneficiaries of such programs. The spread of information about development programs within communities is one factor relevant to the success of those programs. If information disperses quickly then the program will be better able to function, including in recruiting eligible participants. If there are significant frictions to knowledge diffusion—such as due to weak communication links or strategic behavior in whether or how information is conveyed—then the program will work less well, and (in extreme cases) local development efforts may be stalled. There is some supportive evidence from various settings for the view that lack of information is a decisive factor inhibiting successful action by poor people to influence local decision making processes and access the services to which they are entitled.[2] Recognizing this concern, information campaigns are often used by public authorities and non-governmental organizations in conjunction with their main program.[3]

This paper uses an information campaign to identify key aspects of how knowledge is shared within communities, as one element in understanding program efficacy. The paper's premise is that knowledge diffusion about public programs is not an automatic and socially neutral activity but an integral part of the same social and economic processes that perpetuate poverty and inequality. There are foundations for this premise in the literature. It has often been argued that poor minority groups are disadvantaged in the labor market due to their lack of information on job opportunities and weak connections to sources of better information.[4] Economic models have been proposed to explain how this situation can persist, even when efforts are made to equalize opportunities (Loury, 1977). Sociological writings have often emphasized lack of access to knowledge about social programs in creating and maintaining social exclusion (see, for example, Tilly, 2007).

---

[2] "Ignorance" has often been identified as a reason why poor people do not get services intended for all; see, for example, Daponte et al. (1999) with reference to the take up of food stamps in the U.S. Strong "network effects" are often evident in program take-up, and information diffusion is a common explanation. However, also see Aizer and Currie (2004), who find that information sharing played only a limited role in explaining the network effects observed for a U.S. maternal health program.

[3] Useful overviews of the arguments and evidence on factors relevant to the role of information and information campaigns in developing countries can be found in Keefer and Khemani (2005), Khemani (2007) and Mansuri and Rao (2012). The review by La Ferrara (2015) focuses specifically on entertainment media ("edutainment").

[4] See, for example, the discussions in Narayan et al. (2009), World Bank (2011) and Mansuri and Rao (2012).

These issues are relevant to how knowledge about public programs spreads in rural India. The density of a typical village gives ample scope for people to meet and talk about what they have learnt about some new public program. However, social frictions to knowledge diffusion are also likely. Anecdotal observations suggest that knowledge about an initiative coming from outside the village is often controlled by the local elites (who are typically the first contact point in the village) and that they tend to spread knowledge through filters consistent with their own interests. Widespread illiteracy in poor areas supports this selective learning process. In rural India, caste creates special frictions in social interaction and (hence) information diffusion, especially in rural areas. Dalits (also called Scheduled Castes) have faced a long history of discrimination and exclusion; for example, in the majority of Indian villages today, Dalits are not allowed to share food with non-Dalits.[5] It is rare for lower and upper caste families to be close neighbors within the village, and the "Dalit only" area of the village is typically quite well defined and known to all, with strong social pressures (including possible violence) when lower caste people step too far into the domains of other caste groups. The degree of social and political connectivity in this context is greater for more advantaged castes (see, for example, Desai et al., 2010).

In such a setting, knowledge about a new antipoverty program may diffuse rather poorly by word-of-mouth, especially if the information enters via the local elite. Unless a poor, lower-caste, individual comes into direct contact with the source of new information she may come to know little about a program intended to help people like her. Strategic behavior may even act to worsen the information flow. For example, if the program has potentially adverse impacts for the village elite, who are better connected to knowledge sources than the poor, then misinformation may be spread for strategic reasons. Similarly, eligible participants who anticipate rationing of the opportunities announced in an information campaign may rationally choose not to spread the word. Holding a public meeting as part of the campaign might well do a better job of knowledge diffusion than (say) an official letter to the village leader, but it remains unclear just how effective this will be, also noting that socially non-neutral selection processes are likely to be at work in terms of who attends such a meeting and who influences its agenda and discussions; pre-existing inequalities are unlikely to vanish in such a public meeting (Heller and Rao, 2015).

---

[5] A graphic account of the treatment of Dalits in much of rural India can be found in Human Rights Watch (2007).

These observations beg a number of questions: Does information about a public program in rural India spread reliably beyond those who learn about it directly? Does it flow more easily among some groups than others? In particular, are the poor within the village sufficiently well connected socially to tap into the flow of knowledge, or does poverty come with social exclusion, including exclusion from information about programs designed to help poor people?

As this paper shows, an information campaign can throw light on the extent to which information is shared within villages, and so address these questions. The campaign studied here used an entertaining fictional movie to teach people their rights under India's National Rural Employment Guarantee Act (NREGA). NREGA created a justiciable "right-to-work" for all rural households in India. The most direct and obvious way NREGA tries to reduce poverty is by providing extra employment in rural areas on demand. This requires an explicit effort to empower poor people, who are encouraged to take deliberate unilateral actions to demand work on the scheme from local officials. The stipulated wage rate for the scheme is often above the local market wage rate for similar work. Naturally, such a program is seen as a potential threat to the landholding elites in traditional (primarily agricultural) villages; in particular, if the program worked well then it would put upward pressure on local wage rates, reducing the profits of those with larger landholdings.

The setting for this study is rural Bihar—a relatively poor state of about 100 million people in the Northeast of India. In previous research, it was found that most men and three-quarters of women had heard about NREGA, but most were unaware of their precise rights and entitlements under the scheme (Dutta et al., 2014). With the aim of promoting better knowledge about NREGA in this setting, the movie was randomly assigned to sampled villages, with a control group not receiving the movie. Knowledge about NREGA was assessed in both treatment and control villages. Residents were encouraged to watch the movie, but not (of course) compelled to do so. Some watched it and some did not. A previous paper studied the "intent-to-treat" impacts of the movie on knowledge, perceptions and outcomes for program participation (Ravallion et al., 2015). The present paper goes more deeply into the impacts on knowledge, and the channel of that impact—notably whether it was purely through the direct effect of watching the movie or whether it was through knowledge sharing within villages.

The main methodological challenge is how to identify the knowledge gains (if any) for those in the assigned villages who did not actually watch the movie. Such spillover effects are

known to be a source of bias in evaluations of public programs given that one of the key assumptions made in classic evaluation methods is that the comparison group used to infer the counterfactual is not also impacted by the program.[6] As a general proposition, biases can be anticipated for certain impact estimators when this assumption fails to hold. A specific instance is relevant to the evaluation of geographically assigned programs, as studied in this paper. Here the concern is not spillover effects between areas but within them. Many social programs take the form of an individually-assigned intervention (such as a health program, a transfer payment or access to credit). Some people within well-defined geographic areas take up the program and some do not. In evaluating such programs, a seemingly attractive option is to randomly assign access to the program across geographic areas. Take up within areas is endogenous but the randomized assignment across areas can be used as the instrumental variable (IV). Spillover effects are an obvious threat to the validity of this approach and there are various examples in the literature.[7] To the extent that those within an assigned area who do not directly participate in the program are affected (positively or negatively), the exclusion restriction required by the IV estimator will not hold and the estimates will be biased. If the investigator can do a double randomization then the problem can be avoided (Baird et al., 2014).

Methodologically, the paper proposes a solution to the problem of distinguishing the direct effect from the spillover effect through knowledge sharing when double randomization is not a feasible option.[8] We postulate that there is a latent process of knowledge diffusion among households within the village. An individual's knowledge reflects both this process and a latent individual effect representing the individual's "connectedness." The latter is assumed to be time invariant, as it depends on long-standing networks of association between people, reflecting how each individual fits within the village social structure including caste positions and the ability of that individual to process the new information.

The essential idea is to combine a specific double difference method to estimate the direct impact (allowing for endogenous take up), with randomized assignment across villages, which

---

[6] This is sometimes called the stable unit treatment value assumption (SUTVA) following Rubin (1980).
[7] Spillover effects have been identified for: (i) health treatments within schools (Miguel and Kremer, 2004); (ii) schooling in the neighborhoods of transfer recipients (Angelucci and De Giorgi, 2009; Bobonis and Finan, 2009: Bobba and Gignoux, 2014); (iii) local government spending in response to geographically assigned programs (Chen et al., 2009); (iv) crime displacement (Yang, 2008); and (v) a women's empowerment program to encourage child immunization (Janssens, 2011).
[8] For an overview of (experimental and non-experimental) methods for estimating spillover effects see Angelucci and Di Maro (2015). The method used in the present paper is not one of those reviewed by Angelucci and Di Maro.

then identifies the spillover effect. For the purpose of identifying the direct effect, the survey design combined a standard T=2 panel structure with two separate adult interviews to collect data on personal knowledge about NREGA. Having two observations within each household allows us to obtain an estimate that is robust to latent heterogeneity in household factors, similarly to the sibling-difference estimator that has been used in the literature to eliminate a confounding household effect in studying outcomes for children.[9] By exploiting the differences over time, our non-experimental estimator is also robust to time varying individual effects, such as latent social connectivity within the village.

The paper finds evidence of spillover effects, which account for about one third of the average impact of the information campaign. While knowledge sharing is evident, poorer people, by various criteria, appear to be less well connected, and so benefit less from the spillover effect—relying more on direct exposure to the intervention. Methodologically, we find that there is a substantial bias in the IV estimator of the average treatment effect on the treated due to a failure of the exclusion restriction, stemming from knowledge spillovers.

The following section describes the setting and data. Section 3 outlines the proposed econometric method. Section 4 presents the results. Section 5 concludes.

2.   The setting and data

Bihar is one of the poorest two or three of India's larger states. Based on official Planning Commission poverty lines for 2009/10, 55% of its rural population of 90 million lived below the poverty line. Although one would hope that NREGA worked well in India's poorest states (where it is presumably needed most), Bihar has one of the lowest participation rates of any state (Dutta et al., 2014). The scheme that implements NREGA in Bihar is the Bihar Rural Employment Guarantee Scheme (BREGS).[10]

The survey data: The data were collected explicitly for the purpose of evaluating the scheme's performance, as documented in Dutta et al. (2014). Two survey rounds were done spanning 150 randomly chosen villages in rural Bihar.[11] The first round (R1) was implemented

---

[9] Early examples of the sibling difference method of addressing household heterogeneity in estimating models for child outcomes include Rosenzweig and Wolpin (1988) and Duncan et al. (1998).

[10] The corresponding national program is called the Mahatma Gandhi National Rural Employment Guarantee Scheme.

[11] The survey instrument is available to readers online at http://explore.georgetown.edu/people/mr1185/. The data are available for replication purposes from the authors.

between May and July of 2009 and the second (R2) during the same months one year later. The timings were chosen for being lean periods for agricultural work, and were thus expected to be peak periods for BREGS. The survey collected information at household and individual levels on a range of characteristics including caste, demographics, asset ownership, consumption, employment and wages, as well as information on BREGS participation, knowledge of NREGA rights and rules, process-related issues and questions related to perceptions about the scheme in the specific village context.

The two-stage sampling design was based on the 2001 Census list of villages. In the first stage, 150 villages were randomly selected from two strata, classified by high and low BREGS coverage based on administrative data for 2008/9. In the second stage, 20 households per village were randomly selected, drawing from three strata based on an initial listing of all village members and a few selected attributes.[12] All summary statistics reported in the study are weighted with appropriate sample weights to be representative at state level and the regressions allow for the survey design.

Our estimation method (discussed in detail in Section 3) requires that we focus on the subsample of panel households in which two members were interviewed in both rounds. This "2x2" subsample has 2,376 individuals. The fact that we exclude households for which only one adult was available for interview suggests the possibility of sample-selection bias. The 2x2 subsample differs from the full sample in some respects. For example, the two-adult panel is more likely to be male headed, less likely to have a widowed respondent, and more likely to be a larger household.[13] We test whether the sample selection affects our results for those evaluation parameters that do not require the balanced panel, which gives a sample of 4,792 individuals.

It is of interest to test for heterogeneity in the direct and spillover effects in the R1 data, though recognizing that stratification comes with a loss of power. In one case we stratify by three groups of households defined by actual or desired BREGS participation in R1, before the information campaign. First, there are the actual participants in BREGS. Second, we identify a group of "excess demanders," defined as those who said they wanted BREGS work but did not get it; past research has indicated substantial un-met demand for work on the scheme (Dutta et

---

[12] With numbers in parentheses, the three strata were those with at least one member who had done public works in the last year (7), those with a member who had engaged in other (non-public works) casual work (7) and all other households (6).

[13] This was tested using a probit for whether the individual was in the balanced panel using covariates from R1. The probit is available from the authors.

al., 2014). The remainder forms the third group, identified as those who were not interested in participating. We expect the information intervention (discussed further below) to have more impact on the second group, since they express a desire to participate—which presumably motivates learning—but are not participating.

We also stratify by four socioeconomic indicators: caste, literacy, landlessness and consumption poverty. Two caste groups are distinguished; the first is the less advantaged group—the Scheduled Castes comprising Dalits and Mahadalits (a local Bihar category, recognized by the state government). The second is mainly Other Backward Castes and the general caste group (not one of the commonly identified poorer caste/tribal groups). "Landlessness" is defined as owning no cultivable land. "Poverty" is identified by a household consumption per person below the R1 median (which closely accords with the official poverty measure for Bihar). The sample is clearly not evenly spread across the combinations of these dimensions, as can be seen from Table 1, which gives the number of sample points in each cell combination. Of those who are both in the lower caste grouping and consumption poor, 86% are landless, and in 57% of cases both persons are also illiterate. Among those who are in our upper caste grouping and not consumption poor, we find that in 41% of cases at least one of the two persons is literate and the household has some land. But the associations are far from perfect; for example, the instances of both adults being illiterate are similar for the two caste groups. Also note that the counts for some cells are quite low, raising the need for caution about inferences for these cells.

Knowledge: In both R1 and R2, the respondents were asked whether they had heard of NREGA and, if so, they were asked 12 questions testing their knowledge of the scheme's functioning and their rights. We call this the NREGA quiz. This was administered separately to one male and one female member of each sampled household when feasible.

As an overall measure of knowledge about the scheme's employment aspects we calculate an "employment knowledge index" as the number of correct answers to the eight employment and wage related questions in the NREGA quiz. The average scores on employment knowledge in R1 is 2.6 for men and 1.5 for women (out of a maximum of 8).

A second measure can be created for knowledge of the facilities and amenities that the scheme mandates must be provided at work sites (daycare, drinking water, shade and first aid kits). Respondents were asked to identify what facilities were supposed to be provided. We call

this the "facilities knowledge index." The mean number of correct answers on the facilities test in R1 was 1.4 and 1.0 for men and women respectively out of a maximum of 4. Clearly knowledge about these aspects of the program is deficient. However, there can be little doubt that knowledge about employment and wages is more important to the lives of poor people in this setting than knowledge about facilities.

It is not easy to do interviews of individuals on their own in this setting. So some cross-effect can be expected. Nonetheless, the scores generally differed between the two respondents, which motivates our identification strategy. In R1, the scores on the employment quiz differed in 68% of cases, though falling to 44% for the facilities knowledge index; Table 2 gives the frequency distribution of the difference in scores. It is clear that the spread is quite large. So there is ample scope for explaining the differences within households in terms of whether the respondent watched the movie, though this is less so for knowledge about facilities than for employment and wages.

Table 3 gives the breakdown of mean scores on our knowledge tests between the three groups. Participants scored better than either excess demanders or others. Participants may well have learnt by participating, so one cannot conclude from Table 3 that knowledge was the cause of participation. We also see from Table 3 that the participants and excess demanders are more likely to be illiterate and to have completed less schooling. There is an association with caste; participants are more likely to be Mahadalits, and the third group (neither participants nor excess demanders) are more likely to be general caste members. There are also signs that BREGS participants tend to have better political connections, as indicated by responses to survey questions on whether the respondent was "close to" each of a series of designated local officials.

The information campaign: The majority of adults are illiterate, so a movie in the local language makes sense as an information tool. Dutta et al. (2014) report pilot tests of alternatives, such as reading a summary of NREGA provisions in gender-specific focal groups, which did not show much promise. But a pilot film (based on TV advertisements for NREGA) did show promise, judged by audience reaction and subsequent discussion in focal groups.[14]

---

[14] Videos have also proven effective in agricultural extension efforts, including in Bihar; see, for example, Kaushal (2015).

So a 20 minute movie was produced to explicitly convey information about rights and entitlements under BREGS.[15] The movie was tailored to Bihar's specific context and program guidelines. Professional actors performed in an entertaining and emotionally engaging story-based plot whose purpose was to provide information on how the scheme works, who can participate and how to go about participating. The movie stresses the fact that all adults are eligible for the scheme and that potential workers need to demand work in order to get it, in addition to providing information on guidelines for time-bound responses from the government on providing work or an unemployment allowance, and for paying wages.

The story line is centered on a temporary migrant worker returning to his village from the city to see his wife and baby daughter. He learns that there is BREGS work available in the village, even though it is the lean season, so he can stay there with his family and friends rather than return to the city to find work. It was intended that the audience would identify strongly with the central characters. While the lead actor is a man, and the main focus is on him throughout, the story line includes a deliberate flow of supporting actors, including women who indicated that the scheme is open to women who are supposed to receive the same wages as me for equal work.

The information campaign was conducted in February-March 2010 in 40 villages randomly selected from the baseline sample of 150 villages. Compliance at the village level was complete. The intervention was done 2-4 months prior to the follow-up survey (R2), so that information had a reasonable time to spread within the treated villages. Given the timing of the seasons and the nature of the program, the expectation was that if the intervention had impact it should be evident within this time period, given that this coincided with the lean season, when demand for BREGS should be high. Around the time that many people would be in need of extra work, the intervention tells them how to go about getting that work. It is also a season in which forgone income from watching the movie is likely to be relatively low.

Double randomization was not feasible. In some settings one could imagine randomly assigning tickets for seeing the movie. However, that was not considered a reasonable approach in this context. There would be (justifiable) concerns within a treatment village about why some people got a ticket and some did not. There was no plausible reason for rationing, such as due to

---

[15] The movie was commissioned by Dutta et al. and was produced by the local NGO, Praxis—the Institute for Participatory Practices. The movie can be seen here.

budget or capacity constraints. And most villages lack a venue that allows exclusion based on not having a ticket.

The treatment and control samples are well balanced. The differences in sample means of the village variables used in the analysis (including village means of household and individual variables) are only statistically significant at the 5% level for three variables out of the 70 tested; Ravallion et al. (2015) provides details.[16]

There was negligible spillover effect on the control villages; only 12 households (0.4%) in the R2 sample from control villages reported that they were aware that a film on BREGS had been shown elsewhere. However, spillover effects within the treatment villages are likely.

The movie was shown in two separate locations in each treatment village over one or two days. Typically, it was projected in common areas, such as an open ground, school building, or community hall. The screenings were in an open space about half the time; school buildings were the venue for about half the remainder. At each location, the film was screened twice, followed by a question and answer session and distribution of one-page flyers that pictorially illustrated the main entitlements and processes under NREGA. On arriving in each village, efforts were made to announce and advertise the upcoming screenings in advance. Local officials such as the Mukhiya and Sarpanch, opposition leaders and local BREGS officials were invited to attend.

The movie was a big event in the treatment villages. On average, about 365 people attended either screening, roughly evenly split between the two screenings. About two-thirds of those attending were men. Only in 11% of the showings did people say that the information provided was not new. The average discussion time after the movie was 38 minutes and the movie was deemed by the facilitators to have generated a "lot of discussion" in 29% of the showings. Based on our survey, 86% of men and 77% of women in the treatment villages were aware that the movie had been shown. 55% of men in the sample had actually seen the movie, as compared to 43% of women. 27% of men and 33% of women had not seen the movie but reported that they had discussed it with others in the village.

Table 4 provides descriptive statistics for the balanced sample of individuals in treatment villages on knowledge of the movie and the incidence of watching the movie broken down by BREGS participation status, caste, literacy, landholding and poverty, all measured in R1. 85%

---

[16] Of course, some significant differences using conventional "one-at-a-time" tests are to be expected by chance even when fully randomized.

knew about the movie, and this varies little across the groups. 42% decided to watch the movie, and this varies with individual, household and village characteristics. BREGS participants in R1 were more likely to do so; illiterate individuals were less likely to do so.

3.   The model

Knowledge about NREGA for person i in household j in village v at time t, is denoted k_{ijvt}. The baseline survey is at $t=1$, including the NREGA quiz. The movie is then shown and the re-survey, with the quiz, is done at $t=2$. The movie is randomly assigned to villages, with the assignment denoted M_{vt}, with $M_{v2}=1$ if village v got the movie and $M_{v2}=0$ otherwise; by definition, $M_{v1}=0$ for all v. The incidence of watching the movie is denoted m_{ijvt} with $m_{ijv2}=1$ if person i in household j and village v saw the movie and zero otherwise; obviously $m_{ijv1}=0$.

The task of the empirical analysis is to identify how this information intervention affected knowledge about NREGA, and how the movie's impact varied by socioeconomic group. The latter is one element of the time-invariant individual characteristic, denoted δ_i. The individual effect δ_i can be taken to reflect how much each individual is able to tap into the general spread of knowledge given his or her social position. One can think of this as the person's "social connectivity" within the village. Some relevant characteristics are observed (such as the person's caste or education) and some are not (such as the person's cognitive ability). We also postulate the existence of a latent household and village effect, η_{jvt}, that influences knowledge. Similarly to the individual effects, some of these variables are observed and some are not. Household wealth and education variables fall under this heading. (These can vary over time, such as due to household income shocks.)

Combining these observations we postulate a knowledge production function of the form:

$$k_{ijvt}=k(m_{ijvt},M_{vt},\delta_i,\eta_{jvt}) \tag{1}$$

Notice that M_{vt} appears here to reflect knowledge spillovers, i.e., knowledge gains for someone who lives in a village where the movie was shown, but did not necessarily watch the movie. Our task is to estimate a linearized version of equation (1). We treat δ_i and η_{jvt} as unobserved, although we will use the stratifications described in the last section when studying the

heterogeneity in the impact of the movie on knowledge; specifically we will study whether the knowledge effects of m_{ijvt} and M_{vt} vary with certain observable dimensions of δ_i and η_{jvt}.

Consider any δ_i that increases knowledge. How might it alter the impact of the movie on knowledge? The direction of such "cross-effects" is theoretically ambiguous. We might reasonably expect education (say) to enhance knowledge generally and to enhance the impact of a new source of information on knowledge.[17] By contrast, as argued in the introduction, greater social inclusion—more connections within village networks—is likely to enhance knowledge but presumably diminish the knowledge gain from watching the movie; it will be the socially excluded for whom watching the movie will bring the greater gains in knowledge.

Watching the movie is, of course, a matter of personal choice so it is potentially endogenous to knowledge about NREGA. There can be a direct utility gain from watching the movie—its entertainment value—in addition to the benefit from the knowledge gained. (A person may watch the movie solely for its entertainment value.) In deciding whether or not to watch the movie a person weighs the expected benefit against the cost. The expected benefit depends on a set of fixed personal characteristics not all of which are observed. So we can write the utility function as:

$$u_{ijvt} = u(m_{ijvt}, k_{ijvt}, \delta_i, \eta_{jvt}) \tag{2}$$

When δ_i is low the person will have a higher expected benefit from watching the movie, but a lower level of knowledge about programs such as NREGA. The benefits also depend on the time-varying household and village characteristics. The cost of watching the movie is mainly the opportunity cost of time, which will tend to be lower for the poor, and especially the underemployed poor, who are likely to be interested in NREGA. There can also be idiosyncratic factors that influence the costs of watching the movie (such as the need to attend to a sick family member). For the present purpose, we can think of the cost of watching the movie as including an additive random variable with distribution function F_{ijvt}. The probability of watching the movie is then a function of the utility gain, which depends on M_{vt}, δ_i and η_{jvt}, namely:

$$\Pr(m_{ijvt} = 1) = F_{ijvt}[u(1, k(1,1,\delta_i,\eta_{jvt}),\delta_i,\eta_{jvt}) - u(0, k(0,1,\delta_i,\eta_{jvt}),\delta_i,\eta_{jvt})] \text{ for } M_{vt} = 1 \tag{3.1}$$

$$= 0 \text{ for } M_{vt} = 0 \tag{3.2}$$

---

[17] For example, it was found that those amongst poor workers randomly assigned access to a training program who had more education tended to gain more from the training (Galasso et al., 2004).

Later we will use a probit motivated by this model in one of the comparison impact estimators.

Our main task is to estimate the direct and indirect effect of the movie on knowledge, as represented in equation (1). On linearizing the knowledge production function $k(.)$ we have a switching regression for $t=2$:

$$
\begin{aligned}
k_{ijv2} &= (\alpha_1 + \beta m_{ijv2} + \delta_i + \eta_{jv2} + \varepsilon_{ijv2})M_{vt} \\
&+ (\alpha_0 + \delta_i + \eta_{jv2} + \varepsilon_{ijv2})(1 - M_{vt}) \\
&= \alpha_0 + (\alpha_1 - \alpha_0)M_{v2} + \beta m_{ijv2} + \delta_i + \eta_{jv2} + \varepsilon_{ijv2}
\end{aligned}
\tag{4}
$$

For $t=1$, we have:

$$
k_{ijv1} = \alpha_0 + \eta_{jv1} + \delta_i + \varepsilon_{ijv1}
\tag{5}
$$

Given that the movie screenings are randomly assigned across villages it can be assumed that $Cov(M_{v2}, \eta_{jv2}) = Cov(M_{v2}, \delta_i) = Cov(M_{v2}, \varepsilon_{ijvt}) = 0$ (for $t=1, 2$) A key identifying assumption is that the endogeneity of watching the movie stems solely from the two effects, δ_i, and η_{jvt}. Specifically, we assume that $Cov(m_{ijv2}, \varepsilon_{ijvt}) = 0$ but that $Cov(m_{ijv2}, \eta_{jv2}) \neq 0$ and $Cov(m_{ijv2}, \delta_i) \neq 0$ (given equation (3)).

Clearly the η_{jvt} effect cannot be eliminated using individual panel data alone. Our identification strategy exploits the fact that we have two observations of individual knowledge within the sampled households, as well as panel data. This feature of the study design was introduced to obtain a consistent estimate of the effect of an individual watching the movie. We can then net this effect out of the data on knowledge and exploit the randomized assignment of M_{vt} to identify the external effect of the movie on knowledge amongst those who need not have seen the movie. Taking the difference over time (Δ) eliminates the individual knowledge effect, but still leaves the time-varying household-village knowledge effect:

$$
\Delta k_{ijv2} = \eta_{jv2} - \eta_{jv1} + (\alpha_1 - \alpha_0)M_{v2} + \beta m_{ijv2} + \Delta \varepsilon_{ijv2}
\tag{6}
$$

(Noting that $M_{v1} = m_{ijv1} = 0$.) Taking the difference between the two adults interviewed in each household:

$$\nabla\Delta k_{ijv2} = \beta\nabla m_{ijv2} + \nabla\Delta\varepsilon_{ijv2} \tag{7}$$

Here ∇ denotes the difference between two adults in the same household. Recalling that it is assumed that $Cov(m_{ijv2}, \varepsilon_{ijvt}) = 0$ (and hence $Cov(\nabla m_{ijv2}, \Delta\nabla\varepsilon_{ijv2}) = 0$), OLS applied to (7) gives a consistent estimate of β, denoted $\hat{\beta}$. On replacing β by $\hat{\beta}$, equation (7) becomes:

$$k_{ijv2} - \hat{\beta}m_{ijv2} = \alpha_0 + (\alpha_1 - \alpha_0)M_{v2} + \eta_{jv2} + \delta_i + \varepsilon_{ijv2} \tag{8}$$

Given that the movie is randomly assigned, we can then estimate (8) consistently by OLS to obtain α_0 and α_1, and (hence) identify the knowledge spillover effect.

Notice that the fact that our balanced panel comprises one male and one female in almost all households means that we cannot meaningfully identify the gender difference in the direct impact of watching the movie, though we can test for a gender difference in the spillover effect.

We define the following evaluation parameters based on the above model:

$$ATET = E\big[k_{ijv2}\big|M_{v2} = 1, m_{ijv2} = 1\big] - E\big[k_{ijv2}\big|M_{v2} = 0, m_{ijv2} = 0\big]$$

$$= \alpha_1 - \alpha_0 + \beta \tag{9.1}$$

$$SE = E\big[k_{ijv2}\big|M_{v2} = 1, m_{ijv2} = 0\big] - E\big[k_{ijv2}\big|M_{v2} = 0, m_{ijv2} = 0\big]$$

$$= \alpha_1 - \alpha_0 \tag{9.2}$$

$$ITT = E\big[k_{ijv2}\big|M_{v2} = 1\big] - E\big[k_{ijv2}\big|M_{v2} = 0\big]$$

$$= \alpha_1 - \alpha_0 + \beta E\big[m_{ijv2}\big|M_{v2} = 1\big] \tag{9.3}$$

These are standard parameters adapted to the existence of spillover effects. *ATET* is the impact on knowledge of watching the movie relative to the control villages where the movie was not shown. This combines the effect of seeing the movie (β) and living in a village where the movie was shown ($\alpha_1 - \alpha_0$). *SE* is the indirect effect alone—the knowledge spillover. *ITT* is the usual intent-to-treat parameter.

We will compare our estimates with a number of other approaches found in the literature. One of these is the OLS regression model motivated by the switching regression in (4). We will

do this with and without controls. Another is propensity score matching (PSM), which makes fewer assumptions than OLS—in particular it does not assume a linear parametric model for knowledge. PSM has been used often in estimating *ATET* and it is proposed by Angelucci and De Giorgi (2009) as a candidate for estimating spillover effects. We can write the PSM estimators for *SE* and *ATET* as:

$$SE = E\big[k_{ijv2}\big|M_{v2} = 1, m_{ijv2} = 0, p(X_{ijv2})\big] - E\big[k_{ijv2}\big|M_{v2} = 0, m_{ijv2} = 0, p(X_{ijv2})\big] \quad (10.1)$$

$$ATET = E\big[k_{ijv2}\big|M_{v2} = 1, m_{ijv2} = 1, p(X_{ijv2})\big] - E\big[k_{ijv2}\big|M_{v2} = 0, m_{ijv2} = 0, p(X_{ijv2})\big] \quad (10.2)$$

where $p(X_{ijv2})$ is the conditional probability for an individual in movie villages to watch the movie as a function of a vector of round 1 individual, household and village characteristics, X_{ijv2}. (This can be motivated by (3.1) and (3.2) but replacing δ_i, η_{jvt} by the vector of observables X_{ijvt}.) This estimator differs from the naïve OLS with controls but relies on the same assumption that the endogeneity issue can be fully addressed in controlling for observed characteristics. Assuming that part of the decision to watch the movie is actually due to unobserved characteristics a bias will remain in the matching estimate.

Another estimator of interest is:

$$LATE = ITT/E\big[m_{ijv2}\big|M_{v2} = 1\big] \quad (11)$$

This is the usual local average treatment effect; in this case, it is equivalent to estimating the average treatment effect of watching the movie using the randomized assignment at village level as the IV. This might help address the endogeneity issue, but it is expected to deliver biased estimates in the presence of knowledge spillovers in movie-villages.

Before implementing our preferred method, we ran a Monte-Carlo simulation to test the identification strategy and the estimation procedure on artificial data for which we know the true values of the parameters of interest. Full details are given in Appendix 1. The results give one confidence that the proposed method is to be preferred when there are both an endogeneity issue and spillovers. Next we apply it to our data.

16

4. Results

Column (1) Table 5 gives our preferred estimates for the employment knowledge scores, with $\hat{\beta}$ from estimating equation (7) and $\hat{\alpha}_1 - \hat{\alpha}_0$ from (8). We find significant effects of the movie on knowledge and these effects are both direct ($\beta > 0$) and indirect ($\alpha_1 - \alpha_0 > 0$). The total effect (*ATET*) is slightly more than one correct answer on the employment questions ($\hat{\alpha}_1 - \hat{\alpha}_0 + \hat{\beta}$=1.14), which represents about a one third increase in the number of correct answers in the control group ($\hat{\alpha}_0 = 3.46$).[18] About two thirds is the direct effect of watching the movie and one third is indirect, via the intra-village knowledge spillovers.

For comparison, Table 5 gives results for various alternative estimators. Column (2) gives the naïve estimator, using OLS on the switching regression (equation 1). This method incorrectly attributes the impact fully to the direct effect of watching the movie, and substantially overestimates that effect ($\hat{\beta} = 1.55$). Column (3) gives the naïve estimate with controls for individual and household characteristics; this is similar to the results for Column (2).

Column (4) gives the PSM estimates. We use a probit to predict the probability of watching the movie, $\hat{p}(X)$, for all individuals in movie and non-movie villages. In Appendix 3, we report the coefficients of the probit; the estimates reveal some reasonably intuitive selection mechanisms. In Figure 1 we provide the densities of the predicted propensity score in movie and non-movie villages which have a large region of common support. We insure common support by dropping individuals j in non-movie villages such that $\hat{p}(X_j) < \min(\hat{p}(X_i))$ for i in movie villages, and individuals i in movie villages such that $\hat{p}(X_i) > \max(\hat{p}(X_j))$ for j in non-movie villages. We then match nearest neighbors in term of $\hat{p}(X_i)$. We do this by matching one-by-one each individual watching the movie with one counterpart in non-movie villages, and then matching each individual not watching the movie in a movie village with one counterpart in the non-movie villages. We then compute $\widehat{SE}$ and $\widehat{ATET}$ reported in Tables 5 and 8. The PSM estimates are quite similar to OLS. Neither method accords well with our preferred method.

---

[18] All estimates of ATET are based on the sum $\hat{\alpha}_1 - \hat{\alpha}_0 + \hat{\beta}$. In the case of our preferred estimate we report bootstrapped standard errors for 500 replications given that the coefficients come from different regressions on different samples. For OLS estimates we report the s.e. and p values for a t test of the coefficients linear combination.

Column (5) gives the IV estimate of the internal effect ($\hat{\beta}$) assuming no spillovers, and using the randomized assignment at village level as the IV. This is similarly biased to the OLS estimate. Column (6) gives the ITT estimate.

Recall that we are constrained to using the balanced panel of two individuals per household. To test for sample selection bias, we re-estimated ITT, the naïve OLS, the naïve OLS with controls, the PSM and the IV estimator on the full sample. The detailed results are reported in Appendix 2. The estimates are very similar to those we obtained for the selected subsample of two panel individuals.

Now we turn to the stratified results. In Table 6 we stratify the results for $\hat{\beta}$, $\hat{\alpha}_1 - \hat{\alpha}_0$ and ATET ($\hat{\alpha}_1 - \hat{\alpha}_0 + \hat{\beta}$). First, we give a split by gender. As noted, we cannot differentiate the direct effect by gender given that we rely on having two individuals interviewed in each household, typically one male and one female. We can, however, differentiate the spillover effect. We find that this is almost identical between men and women.

Next, Table 6 gives the results by R1 participation status—namely BREGS participants in R1, excess demanders in R1 (who wanted work on BREGS but did not get it) and the rest. We find that the significant direct effect of watching the movie is confined to the excess demanders, and for them it entails an extra two correct answers to the questions on employment provisions of the scheme. The spillover effect is not, however, found in this group but in the other two. In further calculations (not reported but available on request) we also found that the naïve estimator performs quite well for the excess demanders, capturing 90% of the direct effect of watching the movie on employment knowledge. The IV estimator does less well. However, we found that the naïve estimator performs poorly for the other two groups, substantially overestimating the direct effect, and underestimating the spillover effect.

Table 6 gives stratifications by the various poverty indicators. Considering caste first, we find that the direct effect is among the "lower caste" (Mushar/Mahadalit/SC/ST) group, not among the other castes, while the spillover effect is found for the latter. The weaker spillover effect for the lower caste group is consistent with the hypothesis that they are less well connected to the information flows within the village.

Similarly, illiterate, landless and consumption-poor individuals had strong direct effects of watching the movie, but saw far weaker spillover effects. Interestingly ATETs are not very

different between the two caste groups. But this hides a marked difference in how much the impact is direct versus indirect.

Given that these poverty dimensions are correlated it is of interest to also see the parameter estimates for the various combinations of characteristics. As we saw in Table 1, some cell combinations are sparse. Four cells have less than 33 observations, although all the rest have about 100 or more. We dropped the four cells with few observations (one quarter of the combinations)—all of which were lower caste households with land, which is unusual (Table 1).

Table 7 gives the key parameter estimates for each combination of the four poverty dimensions (excluding the four cells with too few observations). The direct effect of watching the movie is especially strong for the consumption poor when they are also of lower caste (by our grouping). The positive spillover effect is especially strong for non-poor upper caste individuals who are not landless; knowledge appears to be shared well within this more advantaged group. A significant negative spillover effect emerges for many of those who are both illiterate and landless. This is suggestive of the presence of strategically placed misinformation for these groups.

Table 8 gives the results for the facilities knowledge score (corresponding to Table 5 for employment). Our preferred estimates in Column (1) do not indicate any significant direct or indirect effect of the movie on knowledge about facilities. By contrast the OLS and PSM estimators suggest a significant positive effect of watching the movie, although this vanishes once one addresses the endogeneity using the randomized village assignment as the IV.

When we stratify by the three aforementioned participation groups we again find a significant direct effect for the excess demanders but not others (Table 9). We find no difference according to caste. Given these results, we do not provide the analogous table to Table 6 for the facilities index. The only notable point was that we found a significant spillover effect on facilities knowledge for poor, lower-caste, landless households when at least one of the two individuals interviewed was literate.

5. Conclusions

The extent to which information about public programs is shared within villages can matter greatly to access to a program and hence its impact. The paper has proposed a method of identifying such knowledge spillover effects for an individualized treatment when double

randomization is not feasible. Differences in learning between individuals within the same household are used to identify the direct impact of the campaign while randomized assignment at village level is used to identify the spillover effect.

We have applied this method to data on the use of an entertaining "information movie" that aimed to teach poor people in rural Bihar their rights under India's National Rural Employment Guarantee Act. We find that about one third of the average impact on knowledge about the key employment and wage aspects of NREGA was due to the knowledge spillover— the knowledge gains to people who did not actually watch the movie but lived in a village where it was shown. We find substantial biases in impact estimation methods that ignore spillover effects.

A key finding is that the knowledge diffusion process is far weaker for disadvantaged groups, defined in terms of caste, landholding, literacy or consumption poverty. For poor people, it appears that the direct effect of watching the movie is all that really matters to learning about NREGA. There is also some indication of negative spillover effects for illiterate and landless households, suggesting the spread of misinformation. These results are consistent with the view that poverty persists in part at least through weak social connectivity, leading to limited gains to poor people from knowledge diffusion—including knowledge about public programs intended to help make them less poor. Public information campaigns in this setting need to be targeted to poor groups, rather than relying on existing knowledge diffusion processes within villages.

References

Aizer, Anna and Janet Currie, 2004, "Networks or Neighborhoods? Correlations in the Use of Publicly Funded Maternity Care In California," *Journal of Public Economics* 88(12), 2573–2585.

Angelucci, Manuela, and Giacomo De Giorgi, 2009, "Indirect Effects of an Aid Program: How do Cash Injections Affect Ineligibles' Consumption?" *American Economic Review* 99(1): 486-508.

Angelucci, Manuela, and Vincenzo Di Maro, 2015, "Program Evaluation and Spillover Effects," Policy Research Working Paper 7243, World Bank (*Journal of Development Effectiveness*, forthcoming).

Baird, Sarah, Aislinn Bohren, Craig McIntosh and Berk Özler, 2014, "Designing Experiments to Measure Spillover Effects," Policy Research Working Paper 6824, World Bank.

Bobba, Matteo, and Jeremie Gignoux, 2014, "Policy Evaluation in the Presence of Spatial Externalities: Reassessing the Progresa Program." PSE Working Papers n2011-37.

Bobonis, Gustavo and Frederico Finan, 2009, "Neighborhood Peer Effects in Secondary School Enrollment Decisions," *Review of Economics and Statistics* 91: 695–716.

Chen, Shaohua, Ren Mu and Martin Ravallion, 2009, "Are There Lasting Impacts of Aid to Poor Areas? Evidence from Rural China," Journal of Public Economics 93: 512-528.

Daponte, Beth, Seth Sanders, Lowell Taylor, 1999, "Why do Low-Income Households Not Use Food Stamps? Evidence from an Experiment," *Journal of Human Resources* 34(3): 612–628.

Desai, Sonalde, Amaresh Dubey, Brijlal Joshi, Mitali Sen, Abusaleh Shariff, and Reeve Vanneman, 2010, *Human Development in India: Challenges for a Society in Transition*. New Delhi: Oxford University Press.

Duncan, G. J., W.J. Yeung, J. Brooks-Gunn and J.R. Smith, 1998, "How Much does Childhood Poverty affect the Life Chances of Children?" *American Sociological Review* 63:406-423.

Dutta, Puja, Rinku Murgai, Martin Ravallion and Dominique van de Walle, 2014, *Right-to-Work? Assessing India's Employment Guarantee Scheme in Bihar.* Equity and Development Series, World Bank.

Galasso, Emanuela, Martin Ravallion and Agustin Salvia, 2004, "Assisting the Transition from Workfare to Work: Argentina's *Proempleo* Experiment," *Industrial and Labor Relations Review* 57(5): 128-142.

Heller, Patrick and Vijayendra Rao, 2015, "Deliberation and Development," In *Deliberation and Development: Rethinking the Role of Voice and Collective Action in Unequal Societies*, edited by Patrick Heller and Vijayendra Rao, Washington DC, World Bank.

Human Rights Watch, 2007, *Hidden Apartheid: Caste Discrimination Against India's "Untouchables,"* Human Rights Watch and Center for Human Rights and Global Justice, New York University School of Law, New York.

Janssens, Wendy, 2011, "Externalities in Program Evaluation: The Impact of a Women's Empowerment Program on Immunization," *Journal of the European Economic Association* 9(6): 1082–1113.

Kaushal Kundan Kumar, 2015, "A Good Farmer and Neighbor," Digital Green Blog, Patna, Bihar, India.

Keefer, Philip and Stuti Khemani, 2005, "Democracy, Public Expenditures, and the Poor: Understanding Political Incentives for Providing Public Services," *World Bank Research Observer* 20 (1): 1–28.

Khemani, Stuti, 2007, "Can Information Campaigns Overcome Political Obstacles to Serving the Poor?" In *The Politics of Service Delivery in Democracies: Better Access for the Poor*, ed. S. Devarajan and I. Widlung, 56–69. Stockholm: Ministry of Foreign Affairs, Government of Sweden.

La Ferrara, Eliana, 2015, "Mass Media and Social Change: Can We Use Television to Fight Poverty?" Working Paper 564, IGIER, Università Bocconi.

Loury, Glenn, 1977. "A Dynamic Theory of Racial and Income Differences." In P. Wallace and A. La Mond (eds) *Women, Minorities, and Employment Discrimination*, Lexington, MA: Lexington Books.

Mansuri, Ghazala and Vijayendra Rao, 2013, *Localizing Development: Does Participation Work?* Washington DC: World Bank.

Miguel, Edward and Michael Kremer, 2004, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica* 72(1): 159-217.

Narayan, Deepa, Lant Pritchett and Soumya Kapoor, 2009, *Moving Out of Poverty. Success from the Bottom Up.* Washington DC: World Bank.

Ravallion, Martin, Dominique van de Walle, Rinku Murgai and Puja Dutta, 2015, "Empowering Poor People through Public Information? Lessons from a Movie in Rural India," *Journal of Public Economics* 132: 13-22.

Rosenzweig, Mark, and Kenneth Wolpin, 1988, "Heterogeneity, Intrafamily Distribution, and Child Health." *Journal of Human Resources* 23(4):437-61.

Rubin, Donald, 1980, "Discussion of the Paper by D. Basu," *Journal of the American Statistical Association* 75: 591-593.

Tilly, Charles, 2007, "Poverty and the Politics of Exclusion," Background paper for World Bank Study: *Moving Out of Poverty*, Columbia University.

World Bank, 2011, *Poverty and Social Exclusion in India*, Washington DC: World Bank.

Yang, Denis, 2008, "Can Enforcement Backfire? Crime Displacement in the Context of Customs Reform in the Philippines," *Review of Economics and Statistics* 90(1): 1-14.

Appendix 1: Simulations to test the estimation method

We simulate $\widetilde{k_{ijv2}}$ using the following data generating process:

$$\widetilde{k_{ijv2}} = \alpha_0 + (\alpha_1 - \alpha_0)M_{v2} + \beta m_{ijv2} + \eta.\,gender_i + \varepsilon_{ijv2}$$

$$k_{ijv1} = \alpha_0 + \eta.\,gender_i + \varepsilon_{ijv2}$$

Where $\varepsilon_{ijv2} \sim N(0, \sigma^2)$ is a random draw in each replication, $gender_i$ is 1 for male and -1 for female, such that $E[\eta.\,gender_i + \varepsilon_{ijv2}] = 0$, and (for the purpose of the simulation) it is assumed that neither the gender nor ε_{ijv2} is observed by the econometrician. The values of $\alpha_0 = E[k_{ijv2}|M_{v2} = 0]$, $\sigma^2 = Var[k_{ijv2}|M_{v2} = 0]$ and $\beta = 1$ are fixed over the following simulations:

Simulation 1: no spillover and no endogeneity: $\alpha_1 = \alpha_0, \eta = 0$

Simulation 2: no spillover and endogeneity: $\alpha_1 = \alpha_0, \eta = 2$

Simulation 3: spillover and no endogeneity: $\alpha_1 = \alpha_0 + 2, \eta = 0$

Simulation 4: spillover and endogeneity: $\alpha_1 = \alpha_0 + 2, \eta = 2$

For each of these simulations we did an OLS estimate of the intention to treat effect (*ITT*), our preferred estimator (*pref*), the naïve OLS estimate (*naïve*), the omniscient OLS estimate where one is able to control for the unobserved source of endogeneity (*omni*), and the *LATE*, for a 2SLS IV where the dummy "watched the movie" is instrumented by a randomly assigned dummy variable, "movie shown in the village". These were compared to the true values of β, SE and $ATET$. For each simulation we ran 500 replications.

The following table gives the results. Simulation 1 indicates that, in the absence of spillover or endogeneity, all strategies give good estimates. Our preferred estimate had a larger standard error, which is to be expected since we are estimating it on a sample of size N/2 (noting that the first stage is a household level regression). From Simulation 2 we found that, in the absence of a spillover effect but with endogeneity, the naïve estimate is biased, while our preferred estimator, the IV and omniscient estimates are all unbiased. With a spillover effect but

no endogeneity, the IV estimate is upward biased, while our preferred estimate, the naïve estimate and the omniscient are all unbiased. Finally, Simulation 4 indicates that with a spillover and endogeneity, the IV and naïve estimates are biased, while both our estimate and the omniscient estimate are unbiased.

		Simulation 1: $\alpha_1 = \alpha_0, \eta = 0$			
true values	β	1			
	SE	0			
	$ATET$	1			
estimates		Mean	Std. Err.	[95% Conf. Interval]	
	ITT	0.41	0.00	0.40	0.41
	β^{pref}	1.01	0.02	0.98	1.04
	SE^{pref}	-0.01	0.01	-0.02	0.01
	$ATET^{pref}$	1.01	0.01	0.99	1.03
	β^{naive}	1.00	0.01	0.99	1.02
	SE^{naive}	0.00	0.00	-0.01	0.01
	$ATET^{naive}$	1.00	0.01	0.99	1.01
	β^{omni}	1.00	0.01	0.99	1.02
	SE^{omni}	0.00	0.00	-0.01	0.01
	$ATET^{omni}$	1.00	0.01	0.99	1.01
	$LATE$	1.00	0.01	0.98	1.01
		Simulation 2: $\alpha_1 = \alpha_0, \eta = 2$			
true values	β	1			
	SE	0			
	$ATET$	1			
estimates		Mean	Std. Err.	[95% Conf. Interval]	
	ITT	0.41	0.00	0.40	0.41
	β^{pref}	1.01	0.02	0.98	1.04
	SE^{pref}	-0.01	0.01	-0.02	0.01
	$ATET^{pref}$	1.01	0.01	0.99	1.03
	β^{naive}	1.56	0.01	1.55	1.58
	SE^{naive}	-0.23	0.00	-0.24	-0.22
	$ATET^{naive}$	1.33	0.01	1.32	1.34
	β^{omni}	1.00	0.01	0.99	1.02
	SE^{omni}	0.00	0.00	-0.01	0.01
	$ATET^{omni}$	1.00	0.01	0.99	1.01
	$LATE$	1.00	0.01	0.98	1.01

Simulation 3: $\alpha_1 = \alpha_0 + 2, \eta = 0$					
true values	β	1			
	SE	2			
	$ATET$	3			
estimates		Mean	Std. Err.	[95% Conf. Interval]	
	ITT	2.41	0.00	2.40	2.41
	β^{pref}	1.01	0.02	0.98	1.04
	SE^{pref}	1.99	0.01	1.98	2.01
	$ATET^{pref}$	3.01	0.01	2.99	3.03
	β^{naive}	1.00	0.01	0.99	1.02
	SE^{naive}	2.00	0.00	1.99	2.01
	$ATET^{naive}$	3.00	0.01	2.99	3.01
	β^{omni}	1.00	0.01	0.99	1.02
	SE^{omni}	2.00	0.00	1.99	2.01
	$ATET^{omni}$	3.00	0.01	2.99	3.01
	$LATE$	5.91	0.01	5.89	5.93
Simulation 4: $\alpha_1 = \alpha_0 + 2, \eta = 2$					
true values	β	1			
	SE	2			
	$ATET$	3			
estimates		Mean	Std. Err.	[95% Conf. Interval]	
	ITT	2.41	0.00	2.40	2.41
	β^{pref}	1.01	0.02	0.98	1.04
	SE^{pref}	1.99	0.01	1.98	2.01
	$ATET^{pref}$	3.01	0.01	2.99	3.03
	β^{naive}	1.56	0.01	1.55	1.58
	SE^{naive}	1.77	0.00	1.76	1.78
	$ATET^{naive}$	3.33	0.01	3.32	3.34
	β^{omni}	1.00	0.01	0.99	1.02
	SE^{omni}	2.00	0.00	1.99	2.01
	$ATET^{omni}$	3.00	0.01	2.99	3.01
	$LATE$	5.91	0.01	5.89	5.93

Appendix 2: Robustness to sample selection.

The following tables give the coefficients for the full sample.

	Employment knowledge index				
	Naïve	Naïve + controls	PSM	IV	ITT
$(\alpha_1 - \alpha_0)$	0.16	0.30***	0.14		0.55***
	(0.12)	(0.11)	(0.11)		(0.11)
β	1.43***	1.21***			
	(0.17)	(0.15)			
α_0	3.26***	3.26***	3.12***	3.26***	3.26***
	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)
ATET/LATE	1.59***	1.51***	1.37***	1.83***	
	(0.15)	(0.14)	(0.16)	(0.25)	
N	3681	3425	2878/3073	3681	3749

	Facilities knowledge index				
	Naïve	Naïve + controls	PSM	IV	ITT
$(\alpha_1 - \alpha_0)$	-0.17**	-0.03	-0.02		0.02
	(0.09)	(0.09)	(0.09)		(0.08)
β	0.70***	0.54***			
	(0.13)	(0.12)			
α_0	2.24***	2.24***	2.06***	2.24***	2.24***
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
ATET/LATE	0.53***	0.51***	0.45***	0.27	
	(0.11)	(0.17)	(0.11)	(0.19)	
N	3681	3425	2878/3073	3681	3749

Note: Robust standard errors in parentheses for OLS and ITT; for PSM Abadie-Imbens robust standard errors; for preferred estimate se(ATET) is bootstrapped standard error (500 replications). * p<0.10, ** p<0.05, *** p<0.01, controls include individual, household and village characteristics observed in round 1. The estimate of α_0 for OLS with controls is evaluated at the mean points of the controls.

Appendix 3: Probit for the probability of watching the movie

Controls for R1		
Age	0.10*	(0.05)
Square of age (/100)	-0.08	(0.06)
Education		
Literate (less than class 5)	0.27	(0.21)
Class 5 pass (Primary)	0.38	(0.31)
Class 8 pass (middle)	0.67*	(0.34)
Class 10th pass (Secondary)	0.77**	(0.38)
Class 12 pass (Higher scondary)	0.54	(0.45)
More than higher secondary	1.55**	(0.68)
Relation to head		
Household head	0.10	(0.47)
Spouse of household head	-0.13	(0.41)
Marital status		
married	0.14	(0.80)
widow or divorced	-0.88	(0.89)
Gender	-0.25	(0.55)
Pearlin index interacted with (male)	0.04	(0.07)
Pearlin index interacted with (female)	0.01	(0.06)
Household characteristics		
ln(hhsize)	-0.30	(0.76)
ln(hhsize) sq	0.18	(0.24)
Share of male adults	-0.43	(0.53)
Share of female adults	-0.42	(0.67)
Share of elderly	-1.22	(0.86)
Share of children younger than age	-0.19	(0.49)
Male household head	-0.31	(0.41)
Age of household head	-0.03	(0.06)
Age of household head squared	0.00	(0.00)
Max education in household		
Literate (less than class 5)	-0.19	(0.22)
Class 5 pass (Primary)	-0.48*	(0.27)
Class 8 pass (middle)	-0.35	(0.27)
Class 10th pass (Secondary)	-0.26	(0.33)
Class 12 pass (Higher scondary)	-0.20	(0.35)
More than higher secondary	-0.91*	(0.49)
Social group		
SC	0.46	(0.31)
Mahadalit	0.62	(0.39)
OBC	-0.06	(0.24)

ST	1.12**	(0.53)
Hindu	-0.11	(0.31)
Asset-house index	-0.05	(0.05)
Asset-house index squared	-0.01	(0.02)
Asset-house index cubed (/100)	0.10	(0.17)
Land owned	0.04	(0.14)
Land owned squared	0.01	(0.01)
BPL ration card	-0.07	(0.15)
Related\close:		
Mukhiya or sarpanch of panchayat	0.01	(0.19)
Ward member or panch of panchayat	0.11	(0.17)
PO/BDO of block	-0.16	(0.76)
any political worker	-0.30	(0.41)
Household voted in panchayat election	0.09	(0.29)
Household suffered a shock in past year:		
Accident	0.22	(0.15)
Illness	0.88***	(0.32)
Job loss	-0.29	(0.23)
Natural disaster	-0.17	(0.24)
Other	0.96***	(0.28)
Men in household do casual work	0.10	(0.18)
Women in household do casual work	0.17	(0.17)
No male migration	-0.04	(0.14)
Regular-salaried worker in household	-0.69*	(0.35)
Village and GP characteristics		
GP has a panchayat bhawan	0.23	(0.28)
Mukhiya's age	0.00	(0.02)
Mukhiya lives in village	0.28	(0.35)
Mukhiya is male	-0.02	(0.22)
Mukhiya completed class 5	-0.16	(0.51)
Mukhiya is a farmer	0.27	(0.38)
Mukhiya held a GP post in the past	-0.32	(0.28)
Mukhiya's family held a GP post in the past	-0.64*	(0.38)
Mukhiya is a contractor	-0.03	(0.49)
Listing: share of SC households	0.31	(0.74)
Listing: share of OBC households	0.92	(1.00)
Village is predominantly Hindu	-0.29	(0.25)
Village is electrified	-0.07	(0.30)
Village has a pucca road	-0.18	(0.38)
Within 5 km of bus stop	-0.03	(0.43)
Within 5 km of GP	0.35	(0.57)
Within 5 km of town	0.15	(0.51)

Within 5 km of block headquarters	0.34	(0.42)
Post office in village	-0.16	(0.27)
Nonagricultural enterprises within 5 km of village	0.12	(0.48)
ln(mean asset index in village)	-0.51	(1.16)
Asset inequality in village	1.90	(2.65)
Listing: share households with kutcha house	-1.00	(0.92)
Listing: share BPL households	1.44	(1.33)
Flood in village in past year	-0.33	(0.50)
Drought in village in past year	-0.19	(0.29)
Any shock in village in past year	-0.66	(0.45)
Good relations among village social groups	-0.27	(0.35)
Self-help group in village	-0.01	(0.24)
Civil society organization in village	0.23	(0.33)
BREGS Vigilance Committee in village	0.09	(0.51)
Other associations in village	0.27	(0.34)
Constant	-1.99	(2.49)
N	877	
Pseudo R²	20.74%	

Note: The Pearlin index is a slightly simplified individual-specific version of the Pearlin Mastery scale, which is a measure of the extent to which individuals perceive themselves to be in control of factors that affect their lives.

Table 1: Distribution of sampled individuals across characteristics

		Literacy and land				Total
		Both illiterate + landless	Both illiterate + some land	1 or 2 literate + landless	1 or 2 literate + some land	
	Lower caste + poor	212	22	110	30	374
	Lower caste + nonpoor	172	30	98	32	332
Caste and poverty	Upper caste + poor	190	136	156	236	718
	Upper caste + nonpoor	204	168	186	394	952
Total		778	356	550	692	2376

Table 2: Difference in scores on the NREGA quiz

Absolute difference in scores between two adults in the same household	Employment/work knowledge (frequency, %)		Facilities knowledge (frequency, %)	
	R1	R2	R1	R2
0	31.67	23.22	55.65	49.93
1	25.93	31.85	18.20	25.58
2	16.20	21.41	12.65	16.22
3	10.65	12.34	7.69	5.84
4	9.01	6.49	5.82	2.44
5	4.57	3.12	n.a.	n.a.
6	1.74	1.04	n.a.	n.a.
7	0.21	0.52	n.a.	n.a.
8	0.02	n.a.	n.a.	n.a.

Table 3: Descriptive statistics by round 1 BREGS participation status

	BREGS participant			Excess demand			Rest			Total		
	N	mean	se(mean)	N	mean	se(mean)	N	mean	se(mean)	N	mean	se(mean)
Employment knowledge index	728	3.426	0.065	1112	2.630	0.054	616	2.394	0.071	2456	2.779	0.037
Facilities knowledge index	728	2.512	0.051	1112	2.023	0.041	616	1.73	0.048	2456	2.073	0.027
Gender	728	0.500	0.019	1112	0.500	0.015	616	0.500	0.02	2456	0.500	0.01
Education												
Illiterate	721	0.792	0.015	1109	0.717	0.014	615	0.48	0.02	2445	0.67	0.01
Literate (less than class 5)	721	0.116	0.012	1109	0.109	0.009	615	0.126	0.013	2445	0.116	0.006
Class 5 pass (Primary)	721	0.039	0.007	1109	0.061	0.007	615	0.093	0.012	2445	0.064	0.005
Class 8 pass (middle)	721	0.024	0.006	1109	0.049	0.006	615	0.115	0.013	2445	0.061	0.005
Class 10th pass (Secondary)	721	0.018	0.005	1109	0.047	0.006	615	0.103	0.012	2445	0.055	0.005
Class 12 pass (Higher scondary)	721	0.010	0.004	1109	0.009	0.003	615	0.046	0.008	2445	0.020	0.003
More than higher secondary	721	0.002	0.002	1109	0.008	0.003	615	0.036	0.008	2445	0.014	0.002
Caste/tribe group												
Mahadalit	728	0.128	0.012	1112	0.037	0.006	616	0.016	0.005	2456	0.055	0.005
Scheduled Tribe (ST)	728	0.033	0.007	1112	0.031	0.005	616	0.001	0.001	2456	0.023	0.003
Scheduled Caste (SC) (other)	728	0.326	0.017	1112	0.261	0.013	616	0.064	0.01	2456	0.223	0.008
Other Backward Castes (OBC)	728	0.471	0.019	1112	0.559	0.015	616	0.657	0.019	2456	0.563	0.01
General caste	728	0.043	0.008	1112	0.112	0.009	616	0.262	0.018	2456	0.136	0.007
Religion												
Hindu	726	0.962	0.007	1112	0.874	0.010	616	0.847	0.015	2454	0.890	0.006
Muslim	726	0.038	0.007	1112	0.126	0.010	616	0.153	0.015	2454	0.110	0.006
Close to...												
Ward member or panch of panchayat	728	0.621	0.018	1112	0.434	0.015	616	0.400	0.020	2456	0.475	0.010
Mukhiya or sarpanch of panchayat	728	0.479	0.019	1112	0.281	0.013	616	0.293	0.018	2456	0.338	0.010
Other elected member	728	0.154	0.013	1112	0.066	0.007	616	0.066	0.010	2456	0.09	0.006
BREGS worksite mate	728	0.258	0.016	1104	0.031	0.005	612	0.021	0.006	2444	0.09	0.006
Monitoring-vigilance committee member of BREGS	726	0.020	0.005	1104	0.006	0.002	610	0.005	0.003	2440	0.01	0.002

Gram Rozgar Sewak of panchayat	726	0.150	0.013	1102	0.042	0.006	612	0.067	0.010	2440	0.079	0.005
Gram Panchat Sewak/Panchayat Sachiv	726	0.075	0.010	1104	0.041	0.006	612	0.058	0.009	2442	0.055	0.005
Revenue official of Panchayat	726	0.025	0.006	1102	0.041	0.006	612	0.013	0.005	2440	0.029	0.003
PO/Block Dev. Officer of block	728	0.017	0.005	1112	0.003	0.002	616	0.019	0.006	2456	0.011	0.002
Any political worker	728	0.096	0.011	1112	0.013	0.003	616	0.006	0.003	2456	0.033	0.004
House asset index	724	-0.830	0.059	1112	-0.417	0.067	614	1.538	0.103	2450	0.021	0.048

Note: sample of households with two panel individuals.

Table 4: Descriptive statistics on compliance and knowledge about the movie, split by round 1 characteristics

Knows about the movie?

BREGS status	N	mean	se(mean)
Participant	204	0.89	0.02
Excess demand	255	0.82	0.02
Rest	168	0.83	0.03
Caste			
Mahadalit/SC/ST	160	0.93	0.02
OBC or General	430	0.82	0.02
Literacy			
One or both literate	314	0.84	0.02
Both illiterate	276	0.85	0.02
Landholding			
Has some land	258	0.86	0.02
Landless	332	0.84	0.02
Poverty			
Poor	314	0.86	0.02
Non-Poor	276	0.83	0.02
Total	590	0.85	0.01

Watched the movie?

BREGS status	N	mean	se(mean)
Participant	204	0.55	0.03
Excess demand	255	0.42	0.03
Rest	168	0.33	0.04
Caste			
Mushar/Mahadalit/SC/ST	160	0.54	0.04
OBC or General	430	0.38	0.02
Literacy			
One or both literate	314	0.45	0.03
Both illiterate	276	0.39	0.03
Landholding			
Has some land	258	0.41	0.03
Landless	332	0.44	0.03
Poverty			
Non-Poor	314	0.42	0.03
Poor	276	0.43	0.03
Total	590	0.42	0.02

Note: sample of households with two panel individuals in villages where the movie was shown. Poor individuals are in households characterized by a monthly per capita expenditure below the whole population median in Round 1 (585 Rupees).

Table 5: Coefficients for employment knowledge index

	(1)	(2)	(3)	(4)	(5)	(6)
	Preferred	Naïve OLS	OLS with controls	PSM	IV	ITT
$(\alpha_1 - \alpha_0)$	0.36***	0.03	0.24	0.22		0.69***
	(0.13)	(0.15)	(0.15)	(0.14)		(0.13)
β	0.77**	1.57***	1.32***			
	(0.36)	(0.20)	(0.19)			
α_0	3.46***	3.46***	3.46***	3.21***	3.46***	3.46***
	(0.08)	(0.08)	(0.08)	(.09)	(0.08)	(0.08)
ATET/LATE	1.14***	1.59***	1.56***	1.35***	1.63***	
	(0.24)	(.17)	(.17)	(0.18)	(0.29)	
N	1188/2376	2376	2219	1894/2015	2376	2376

Note: Robust standard errors in parentheses for OLS and ITT; for PSM Abadie-Imbens robust standard errors; for preferred estimate se(ATET) is bootstrapped standard error (500 replications). * p<0.10, ** p<0.05, *** p<0.01, controls include individual, household and village characteristics observed in round 1. The estimate of α_0 for OLS with controls is evaluated at the mean points of the controls.

Table 6: Coefficients for employment knowledge index, split by round 1 characteristics

	β	$(\alpha_1 - \alpha_0)$	ATET
Gender			
Male	0.77**	0.36**	1.13***
	(0.36)	(0.17)	(0.27)
Female	0.77**	0.36**	1.14***
	(0.36)	(0.18)	(0.26)
NREGS status			
participant	-0.20	0.55**	0.35
	(0.63)	(0.22)	(0.40)
excess demander	1.98***	-0.26	1.72***
	(0.44)	(0.19)	(0.32)
rest	-0.40	1.20***	0.80**
	(0.45)	(0.25)	(0.39)
Caste			
Mahadalit/SC/ST	1.47***	-0.10	1.37***
	(0.44)	(0.22)	(0.32)
OBC or General	0.45	0.55***	1.00***
	(0.48)	(0.15)	(0.34)
Literacy			
Both illiterate	1.47**	-0.03	1.43***
	(0.69)	(0.18)	(0.48)
One or both literate	0.31	0.67***	0.98***
	(0.33)	(0.18)	(0.27)
Landholding			
Landless	1.06**	0.00	1.07***
	(0.51)	(0.17)	(0.35)
Has some land	0.33	0.82***	1.15***
	(0.39)	(0.19)	(0.32)
Poverty			
Poor	0.83**	0.29*	1.12***
	(0.41)	(0.17)	(0.30)
Non-poor	0.73	0.40**	1.13***
	(0.55)	(0.18)	(0.37)

Note: Standard errors in parentheses, se (β) and se $(\alpha_1 - \alpha_0)$ are robust standard errors, se(ATET) are bootstrapped standard errors (500 replications), * p<0.10, ** p<0.05, *** p<0.01. Poor households are characterized by a monthly per capita expenditure below the whole population median in Round 1 (585 Rupees).

Table 7: Parameter estimates for combinations of round 1 poverty dimensions

			Literacy and land			
			Both illiterate + landless	Both illiterate + some land	1 or 2 literate + landless	1 or 2 literate + some land
	Lower caste + poor	β	2.36*** (0.42)		1.66*** (0.42)	
		$\alpha_1 - \alpha_0$	-0.64** (0.31)	n.a.	0.88* (0.53)	n.a.
	Lower caste + nonpoor	β	-0.12 (0.58)		1.56 (0.93)	
		$\alpha_1 - \alpha_0$	0.16 (0.42)	n.a.	0.58 (0.47)	n.a.
Caste and Poverty	Upper caste + poor	β	0.58 (1.55)	2.55** (0.75)	-0.66 (0.42)	-0.49 (0.71)
		$\alpha_1 - \alpha_0$	0.55 (0.45)	0.26 (0.44)	1.06** (0.42)	0.44 (0.36)
	Upper caste + nonpoor	β	1.84 (2.07)	-0.81 (0.71)	0.85 (0.74)	0.05 (0.46)
		$\alpha_1 - \alpha_0$	-0.91** (0.44)	1.06*** (0.37)	-0.22 (0.44)	1.17*** (0.30)

Note: Estimates are only provided when the cell size was deemed adequate; the four missing cells had sample sizes of 32 or less. Standard errors in parentheses, se (β) and se $(\alpha_1 - \alpha_0)$ are robust standard errors; * p<0.10, ** p<0.05, *** p<0.01

Table 8: Coefficients for facilities knowledge index

	(1)	(2)	(3)	(4)	(5)	(6)
	Preferred	Naïve	Naïve With controls	PSM	IV	ITT
$(\alpha_1 - \alpha_0)$	-0.02	-0.22*	-0.08	0.03		0.09
	(0.10)	(0.12)	(0.12)	(0.10)		(0.10)
β	0.26	0.73***	0.54***			
	(0.31)	(0.17)	(0.15)			
α_0	2.32***	2.32***	2.32***	2.20***	2.32***	2.32***
	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)
ATET/LATE	0.24	0.51 **	0.46***	0.54***	0.21	
	(0.22)	(.14)	(.14)	(0.13)	(0.24)	
N	1188/2376	2376	2219	1894/2015	2376	2376

Note: Robust standard errors in parentheses for OLS and ITT; for PSM Abadie-Imbens robust standard errors; for preferred estimate se(ATET) is bootstrapped standard error (500 replications). * p<0.10, ** p<0.05, *** p<0.01, controls include individual, household and village characteristics observed in round 1.

The estimate of α_0 for OLS with controls is evaluated at the mean points of the controls.

Table 9: Coefficients for facilities knowledge index, split by round 1 characteristics

	β	$(\alpha_1 - \alpha_0)$	ATET
Gender			
Male	0.26	0.02	0.28
	(0.31)	(0.15)	(0.25)
Female	0.26	-0.07	0.19
	(0.31)	(0.14)	(0.23)
BREGS status			
participant	-0.99	0.08	-0.92**
	(0.66)	(0.18)	(0.39)
excess demander	1.16***	-0.18	0.98***
	(0.38)	(0.15)	(0.29)
rest	-0.07	0.31	0.24
	(0.20)	(0.19)	(0.26)
Caste			
Mahadalit/SC/ST	0.32	0.11	0.43
	(0.38)	(0.17)	(0.27)
OBC or General	0.24	-0.07	0.16
	(0.42)	(0.13)	(0.32)
Literacy			
Both illiterate	0.10	-0.13	-0.03
	(0.65)	(0.14)	(0.44)
One or both literate	0.37	0.06	0.43
	(0.28)	(0.14)	(0.24)
Landholding			
Has some land	0.24	-0.06	0.18
	(0.48)	(0.13)	(0.34)
Landless	0.30	0.02	0.32
	(0.25)	(0.16)	(0.25)
Poverty			
Poor	-0.10	0.14	0.04
	(0.32)	(0.13)	(0.22)
Non-poor	0.54	-0.16	0.38
	(0.47)	(0.15)	(0.35)

Note: Standard errors in parentheses, se (β) and se $(\alpha_1 - \alpha_0)$ are robust standard errors, se(ATET) are bootstrapped standard errors (500 replications), * p<0.1, ** p<0.05, *** p<0.01. Poor households are characterized by a monthly per capita expenditure below the whole population median in Round 1 (585 Rupees).

Figure 1: Density of predicted propensity scores in movie and non-movie villages
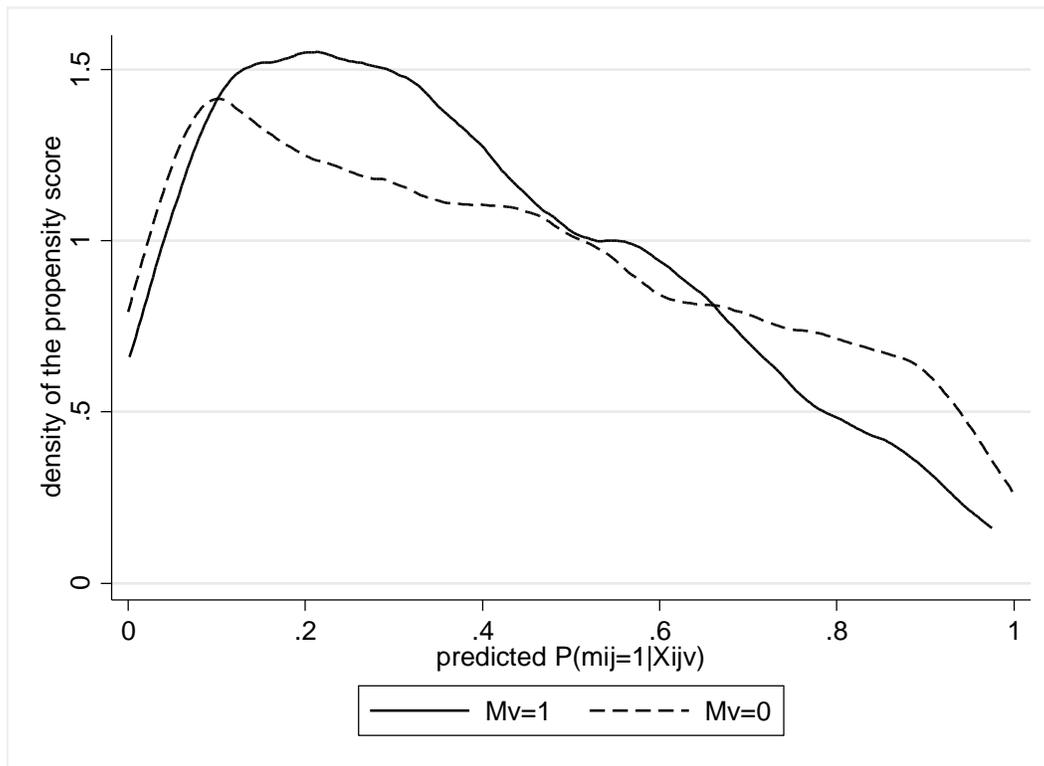
Note: Univariate kernel density estimation.