# Labor Market Effects of Social Programs: Evidence from India's Employment Guarantee.
# ONLINE APPENDIX

*By* Clément Imbert * and John Papp†

### I. History of Public Works Programs in India

India has a long history of providing public works dating back to British rule. Three large-scale public works programs deserve specific mention. First is the Maharashtra Employment Guarantee Scheme passed in 1976 and still in force today. The NREGA is in part based on the design of the Maharashtra EGS.

Second, the Sampoorn Grameen Rozgar Yojana (SGRY) started in 2001 with the purpose of generating employment across India and was still active until 2008. The total allocation to the SGRY was 35 billion Rupees per year from 2004-2008 (Afridi, 2008).

Finally, the National Food for Work Program was introduced as a pilot for the NREGA in 150 of the phase one districts, with an allocation of 60 billion Rupees in fiscal year 2005-06 (Afridi, 2008). As a comparison, during fiscal years 2006-07 and 2007-08, the allocation for the NREGA was 116 billion Rupees. Confirming existing field observations that the National Food for Work Program was poorly implemented and plagued with massive leakages Dreze (2005), we find little evidence of an increase in public works during this pilot period.

### II. Determinants of Government Employment Provision

The central government funds most of the expenditure for the NREGA (all of labor and 75% of material expenditures). However, the responsibility of implementing the scheme is left to the states and the lower administration levels (districts and village councils). In principle, local officials are meant to respond to worker demand for work, but the process required to provide work requires considerable administrative capacity: selecting public works projects, applying for funds, opening the works, sanctioning expenditures, making payments to workers and suppliers of materials etc. When the scheme started in each district, awareness campaigns also had to be implemented by the administration, sometimes with

* Department of Economics, University of Oxford, Manor Road Building, OX1 3UQ, UK, clement.imbert@economics.ox.ac.uk.
† R.I.C.E, 31 Union Square West Apt 13aa, New York , NY, 10003, johnhpapp@gmail.com

the help of civil society organizations. Depending on the administrative capacity of each state, NREGA implementation was initially more or less successful.

During the period we study, which is immediately after the launch of the scheme, the states of Andhra Pradesh, Chattisgarh, Himachal Pradesh, Madhya Pradesh, Rajasthan, Tamil Nadu and Uttarakhand, which we call "star states" in the analysis provided significantly more employment than other states (see Figure 2). This was partially due to demand for work in these states. However, very poor states such as Bihar, Jharkhand, Orissa, and Uttar Pradesh where demand should be high saw little employment generation. In this second group of states, lack of administrative capacity and rampant corruption hampered public employment delivery, despite large potential demand (Khera, 2011; Dutta et al., 2012). In the 2009-10 NSS employment survey, workers were asked whether they had, and whether they desired NREGA employment. Using answers to these questions, Dutta et al. (2012) confirm that three years after the scheme started, demand for work is still more rationed in the poorest states of India.

In order to investigate the sources of observed disparities in NREGA implementation across states, we use NSS data to regress time spent on public works by rural adults in 2007-08 on the set of district controls presented in Table 1, and plot state-level averages of the residuals in Figure A.1. The seven states we defined as star states all have higher public employment provision than predicted by the model. This finding is consistent with the view that differences in public employment provision across states are due to supply factors (e.g. political will or administrative capacity) rather than demand factors (e.g. poverty or labor market conditions). The state which has lowest public employment provision compared to the predicted value in Figure A.1 is Maharashtra, which had its own employment guarantee since the 1970s and whose government was reluctant to implement NREGA.

## III. Theoretical Appendix

### A. *Utility maximization*

Each household has a utility function $u(c_i, l_i)$ over household consumption $c_i$ and leisure $l_i$. We assume the function is increasing and concave in both arguments. Let $L_i^s$ denote household total labor supply and $D_i$ denote household total labor demand. Household labor supply $L_i^s$ has two components: family labor used for household production $L_i^f$ and wage work supplied by household members to the market $L_i^o$. Household labor demand $D_i$ also has two components: family labor $L_i^f$ and hired by the household $L_i^h$. Households choose $L_i^f$, $L_i^o$ ,$L_i^h$ and $c_i$ to solve the following maximization problem:

$$\max_{c_i, L_i^f, L_i^o, L_i^h} \quad u(c_i, T - L_i^f - L_i^o)$$
$$\text{s. t. } c_i = pWL_i^o + A_iG(L_i^f + L_i^h) - WL_i^h$$

We further impose that the optimal labor quantities $L_i^f$, $L_i^o$ ,$L_i^h$ cannot be negative, and both consumption and leisure must be positive ($c_i > 0$ and $T > L_i^f - L_i^o$). We write the Lagrangian:

$$\mathcal{L} = u(c_i, T - L_i^f - L_i^o) + \lambda(pWL_i^o + A_iG(L_i^f + L_i^h) - WL_i^h - c_i)$$

The Kuhn Tucker conditions write

$$\begin{aligned} u_c' - \lambda \leq 0 \quad & \textit{and} \quad c(u_c' - \lambda) = 0 \\ -u_l' + \lambda pw \leq 0 \quad & \textit{and} \quad L_i^o(u_l' - \lambda pw) = 0 \\ -u_l' + \lambda A_iG' \leq 0 \quad & \textit{and} \quad L_i^f(u_l' - \lambda A_iG') = 0 \\ \lambda(A_iG' - W) \leq 0 \quad & \textit{and} \quad L_i^h(W - A_iG') = 0 \end{aligned}$$

However, we assume that $c_i > 0$ hence the first condition simply yields: $u_c' = \lambda > 0$. We can rewrite the three other conditions using this equality:

$$\begin{aligned} pw \leq \frac{u_l'}{u_c'} \quad & \textit{and} \quad L_i^o(u_l' - \lambda pw) = 0 \\ A_iG' \leq \frac{u_l'}{u_c'} \quad & \textit{and} \quad L_i^f(u_l' - \lambda A_iG') = 0 \\ A_iG' \leq W \quad & \textit{and} \quad L_i^h(W - A_iG') = 0 \end{aligned}$$

There are seven cases to consider depending on whether the optimal $L_i^f$, $L_i^o$ ,$L_i^h$ are null.

**Cases 1** Let us assume that $L_i^o > 0$, $L_i^h > 0$ and $L_i^f > 0$. Then we must have $pw = \frac{u_l'}{u_c'}$ and $W = A_iG'$. However, we also need to have $A_iG' = \frac{u_l'}{u_c'}$. Hence this case is only possible if $p = 1$, i.e. households can be suppliers and buyers of labor at the same time if and only if the labor market is without friction. In the general case with friction, households cannot be on both sides of the market.

**Case 2** we assume that $L_i^o > 0$, $L_i^f = 0$ and $L_i^h = 0$. Then we must have that $pw = \frac{u_l'}{u_c'}$, $A_iG' \leq \frac{u_l'}{u_c'}$ and $A_iG' \leq W$. This case is unlikely. Households cannot not choose to supply labor to the market without producing anything on their farm, because for any $W$ one can find a $L_i^f$ small enough so that the marginal productivity of labor will be higher than $pW$. This is because we assumed that all households are able to produce ($A_i > 0$).

**Case 3** we assume that $L_i^0 = 0$, $L_i^f = 0$ and $L_i^h > 0$. Then we must have that $pw \leq \frac{u_l'}{u_c'}$, $A_i G' \leq \frac{u_l'}{u_c'}$ and $A_i G' = W$. This case is also unlikely. Households will not optimally choose to hire workers without supplying any family labor (i.e. reduce their consumption and devote all their time to leisure), because for any $W$ one could find a $L_i^f$ small enough so that the marginal rate of substitution of consumption to leisure will be higher than $W$.

**Case 4** where $L_i^0 = L_i^f = L_i^h = 0$ is not optimal if $A_i > 0$.

**Case 5** the household is net supplier of labor ($L_i^0 > 0$, $L_i^f > 0$ and $L_i^h = 0$). Then the marginal productivity on the farm is equal to wage labor earnings, which is less than the market wage (i.e. $\frac{u_l'}{u_c'} = A_i G' = pW \leq W$) .

**Case 6** the household is net buyer of labor ($L_i^0 = 0$, $L_i^f > 0$ and $L_i^h > 0$). Then the marginal productivity on the farm is equal to the market wage (i.e. $\frac{u_l'}{u_c'} = A_i G' = W \geq pW$).

**Case 7**: the household does not participate to the labor market ($L_i^0 = 0$, $L_i^f > 0$ and $L_i^h = 0$). Then the marginal productivity on the farm is equal to the marginal rate of substitution between consumption and leisure. It is lower than the market wage and higher than labor market earnings (i.e. $\frac{u_l'}{u_c'} = A_i G' \in [pW, W]$).

If $p < 1$ only cases 5, 6 and 7 are possible; households are either labor suppliers, labor buyers or they do not participate to the market. If $p = 1$, cases 1, 5 and 6 are possible and case 7 contracts to a single point: households may be labor sellers, labor buyers, or both.

### *B. Productivity thresholds*

For each value of the wage $W$, let us consider the value of the productivity factor $A_i$ such that labor supply and labor demand from household $i$ are equal:

$$L_i^s(W, A_i G'(D(W, A_i))) \quad = \quad D_i(W, A_i)$$

Let us denote this value $\phi(W)$. Since $L_Y^s \leq 0$ and $D_A(W, A_i) \geq 0$, $\phi(W)$ exists and is unique. Since $L_W^s > 0$ and $D_W(W, A_i) < 0$, the function $\phi(W)$ is strictly increasing in $W$ .

**Proposition 1: A household $i$ is net labor buyer if and only if $A_i > \phi(W)$**

**Proof:** A household with $A_i = \phi(W)$ therefore supplies and demands $D(W, \phi(W))$ labor. Since the marginal cost of hiring labor is $W$ while the marginal value of working in the labor market is $p_i W < W$, the household will always supply labor to its own production function at least up to $D(W, \phi(W))$. Therefore, households with $A_i = \phi(W)$ are neither net labor supplying nor net labor buying households. For $A_i > \phi(W)$, we will have $D(W, A_i) > L^s(W, A_i G'(D(W, A_i)))$, so that the household will be a net labor buyer as long as it can hire labor at $W$ and as long as the marginal value of time is given by $W$ as well. Since net labor buyers supply labor only to their own farm, this will be the case. Net labor buyers will always face an effective marginal wage of $W$. Therefore, if $A_i < \phi(W)$, then $D(W, A_i) < L^s(W, A_i G'(D(W, A_i)))$, so that households will not be net buyers of labor.

**Proposition 2: A household$i$ is net labor supplier if and only if $A_i < \phi(pW) < \phi(W)$**

**Proof:** A household with $A_i = D(pW, \phi(pW))$ will supply and demand $D_w$ units of labor but because $pW < W$ we have $D(pW, \phi(pW)) < D(W, \phi(W))$ and $\phi(pW) < \phi(W)$. For a household with $A_i < \phi(pW)$, we will have $D(pW, A_i) < L^s(pW, A_i G'(D(pW, A_i)))$, so that the household will be a net labor supplier. Net labor suppliers will always face an effective marginal wage of $p_i W$. For a household with $A_i > \phi(pW)$, we will have $D(p_i W, A_i) > L^s(pW, A_i G'(D(p_i W, A_i)))$, so that the household will not be a net labor supplier.

**Proposition 3: For $A_i \in [\phi(pW), \phi(W)]$, household $i$ is neither net supplier or buyer of labor.**

**Proof:** This follows directly from the first two propositions. For $A_i \in [\phi(pW), \phi(W)]$, labor supply and demand $D$ will solve $D = L^s(A_i G'(D), A_i G(D))$. Note that for $A_i \in [\phi(pW), \phi(W)]$, the labor supply and demand will satisfy $A_i G'(D) \in [p_i W, W]$.

Hence the three possible solutions to the utility maximization problem correspond to different values for the productivity factor $A_i$. The most productive households (e.g. those with most land) are net labor buyers and the marginal productivity on their farm is the market wage. The least productive households (e.g. those with little land) are net labor sellers and the marginal productivity on their farm is equal to wage labor earnings $pW$. Households with intermediary levels of productivity will not participate to the market (this last case contracts to a single productivity level if $p = 1$.)

### *C. Compensating Variation Derivation*

In subsection, we prove that the compensating variation which keeps household welfare constant following an increase in public employment provided $dL_g$ is equal to:

$$(1) \qquad -dz_i = \textit{Net Casual Labor Earnings} \times \frac{dW/W}{dL_g} + (W_g - pW) dL_i^g$$

Let us first consider households with low productivity levels $A_i < \phi(pW)$. The equation equating expenditure to income writes

$$e(pW, u_i) = \pi_i(pW) + pWT + (W_g - pW)L_i^g + z_i$$

We derive the change in $z_i$ required to maintain the equality, and therefore maintain the same utility level, following a change in $L_g$. We do this by differentiating Equation III.C with respect to $L_g$:

$$\frac{de(pW, u_i)}{dL^g} = p\pi_i'(pW)\frac{dW}{dL_g} + pT\frac{dW}{dL_g} + (W_g - pW)\frac{dL_i^g}{dL_g} - pL_i^g\frac{dW}{dL_g} + dz_i$$

By the envelope theorem $\frac{de(pW,u_i)}{dW} = p(T - L_i^s)$ and $\pi_i'(pW) = -D_i$. Using these results and re-arranging yields:

$$\begin{aligned} -dz_i &= (L_i^s - L_i^g - D_i)pW\frac{dW/W}{dL_g} + (W_g - pW)dL_i^g \\ &= \textit{Net Casual Labor Earnings} \times \frac{dW/W}{dL_g} + (W_g - pW)dL_i^g \end{aligned}$$

For households with high productivity levels $A_i > \phi(W)$ the equation equating expenditures to income writes:

$$e(W, u_i) = \pi_i(W) + WT + (W_g - W)L_i^g + z_i$$

Using the same demonstration as before, but replacing $p$ with 1, we find that:

$$\begin{aligned} -dz_i &= (L_i^s - L_i^g - D_i)W\frac{dW/W}{dL_g} + (W_g - W)dL_i^g \\ &= \textit{Net Casual Labor Earnings} \times \frac{dW/W}{dL_g} + (W_g - W)dL_i^g \end{aligned}$$

Finally, for households with intermediary productivity levels $A_i \in [\phi(pW), \phi(W)]$, the equation equating expenditures with revenues writes:

$$e(\widetilde{W_i}, u_i) = \pi_i(\widetilde{W_i}) + \widetilde{W_i}T + (W_g - \widetilde{W_i})L_i^g + z_i$$

where $\widetilde{W_i}$ is the shadow wage which does not depend on W. The program only affects households welfare through direct participation, and the compensating variation has the simple form:

$$-dz_i = (W_g - \widetilde{W_i})dL_i^g$$

However, since these households do not buy or sell labor on the market, their net casual labor earnings are zero, and we can also write:

$$-dz_i = \textit{Net Casual Labor Earnings} \times \frac{dW/W}{dL_g} + (W_g - \widetilde{W_i})dL_i^g$$

Which completes our demonstration.

### D. *Impact of Government Hiring on the labor market equilibrium*

The market clearing condition imposes that labor supply of households with low productivity and labor demand of households with high productivity are equal. It writes:

$$p \int_{\underline{A}}^{\phi(pW)} [L_i^s(pW) - D_i(pW) - L_i^g]dA_i = \int_{\phi(W)}^{\overline{A}} [D_i(W) - L_i^s(W) + L_i^g]dA_i \tag{2}$$

To determine the impact on wages of public sector hiring we need to differentiate the market clearing condition with respect to $L^g$. We use Leibnitz integral rule which yields for the left-hand side of equation 2:

$$\begin{aligned}\frac{dp\int_{\underline{A}}^{\phi(pW)}[L_i^s(pW) - D_i(pW) - L_i^g]dA_i}{dL^g} &= [L_i^s(pW,\phi(pW)) - D_i(pW,\phi(pW)) - L_i^g]\phi'\frac{dW}{dL^g} \\ &+p \int_{\underline{A}}^{\phi(pW)} \frac{d[L_i^s(pW) - D_i(pW) - L_i^g]}{dL^g}dA_i\end{aligned}$$

By definition, net labor demand of households with productivity levels $\phi(pW)$ is zero, so that $[L_i^s(pW,\phi(pW)) - D_i(pW,\phi(pW)) - L_i^g] = 0$. Hence the first term is null.

A similar simplification can be made for $\phi(W)$, while differentiating the right-hand side of equation 2. Hence the derivative of 2 with respect to $L^g$ writes:

$$p \int_{\underline{A}}^{\phi(pW)} [\frac{dL_i^s(pW)}{dL^g} - \frac{dD_i(pW)}{dL^g} - \frac{dL_i^g}{dL^g}]dA_i = \int_{\phi(W)}^{\overline{A}} [\frac{dD_i(W)}{dL^g} - \frac{dL_i^g}{dL^g} - \frac{dL_i^s(W)}{dL^g}]dA_i \tag{3}$$

Let us first consider households which are net labor suppliers ($A_i < \phi(pW)$). Public hiring affects labor supply through its effect on the equilibrium wage and through its effect on non-labor income. We decompose the derivative of $L_i^s$ with respect to $L^g$ in two components:

$$\frac{dL_i^s(pW,y_i)}{dL^g} = \frac{dL_i^s(pW,y_i)}{dW}|_{y_i}\frac{dW}{dL^g} + \frac{dL_i^s(pW,y_i)}{dy_i}\frac{dy_i}{dL^g}$$

where $\frac{dL_i^s}{dW}|_{y_i}$ is the derivative of household $i$'s labor supply with respect to the wage holding non-labor income fixed. The slutsky decomposition yields:

$$\frac{dL_i^s(pW,y_i)}{dW}|_{y_i} \quad = p \quad \frac{dL_i^s}{dW}|_u + \frac{dL_i^s}{dy_i}pL_i^s$$

where $\frac{dL_i^s}{dW}|_u$ is the substitution effect, i.e. the partial derivative of labor supply with respect to the wage holding utility constant. We have that:

$$\begin{aligned}\frac{dy_i^s}{dL_g} &= p\pi_i'(pW)\frac{dW}{dL_g} + (W_g - pW)\frac{dL_i^g}{dL_g} - pL_i^g\frac{dW}{dL_g} \\ &= -pD_i\frac{dW}{dL_g} + (W_g - pW)\frac{dL_i^g}{dL_g} - p\frac{dW}{dL_g}L_i^g\end{aligned}$$

where the second equality follows from the envelope theorem for the profit function $\pi_i'(W) = -D_i$.

Hence, for households with $A_i < \phi(pW)$, we can rewrite the derivative of the labor supply with respect to public hiring as:

$$\frac{dL_i^s(W,y_i)}{dL^g} = p[\frac{dL_i^s}{dW}|_u + \frac{dL_i^s}{dy_i}(L_i^s - D_i - L_i^g)]\frac{dW}{dL_g} + \frac{dL_i^s}{dy_i}(W_g - pW)\frac{dL_i^g}{dL_g}$$

Public hiring affects labor demand only through its effect on the equilibrium wage. Hence the derivative of the labor demand with respect to public hiring writes: $\frac{dD_i(pW)}{dL^g} = pD_i'(pW)\frac{dW}{dL_g}$

Hence, the impact of public sector hiring on the net labor supply of households with $A_i < \phi(pW)$ is given by the following expression:

$$\begin{aligned} p \int_{\underline{A}}^{\phi(pW)} [\frac{dL_i^s(pW)}{dL^g} - \frac{dD_i(pW)}{dL^g} - \frac{dL_i^g}{dL^g}]dA_i &= p^2 \int_{\underline{A}}^{\phi(pW)} [\frac{dL_i^s}{dW}|_u + \frac{dL_i^s}{dy_i}(L_i^s - D_i - L_i^g)]\frac{dW}{dL_g}dA_i \\ &+ p \int_{\underline{A}}^{\phi(pW)} \frac{dL_i^s}{dy_i}(W_g - pW)\frac{dL_i^g}{dL_g}dA_i \\ &- p^2 \int_{\underline{A}}^{\phi(pW)} D_i'(pW)\frac{dW}{dL_g}dA_i - p \int_{\underline{A}}^{\phi(pW)} \frac{dL_i^g}{dL_g}dA_i \end{aligned} \tag{4}$$

Using similar arguments, we can write the impact of public sector hiring on the net labor demand of households with $A_i > \phi(W)$ as:

$$\begin{aligned} \int_{\phi(W)}^{\overline{A}} [\frac{dD_i(W)}{dL^g} + \frac{dL_i^g}{dL^g} - \frac{dL_i^s(W)}{dL^g}]dA_i &= - \int_{\phi(W)}^{\overline{A}} [\frac{dL_i^s}{dW}|_u + \frac{dL_i^s}{dy_i}(L_i^s - D_i - L_i^g)]\frac{dW}{dL_g}dA_i \\ &- \int_{\phi(W)}^{\overline{A}} \frac{dL_i^s}{dy_i}(W_g - W)\frac{dL_i^g}{dL_g}dA_i \\ &+ \int_{\phi(W)}^{\overline{A}} D_i'(W)\frac{dW}{dL_g}dA_i + \int_{\phi(W)}^{\overline{A}} \frac{dL_i^g}{dL_g}dA_i \end{aligned} \tag{5}$$

Plugging equations 4 and 5 into 3 and re-arranging yields:

$$\frac{dW}{dL^g} = \frac{E_1 - E_2}{-E_3 + E_4} \tag{6}$$

Where:

$$E_1 = p \int_{\underline{A}}^{\phi(pW)} \frac{dL_i^g}{dL_g}dA_i + \int_{\phi(W)}^{\overline{A}} \frac{dL_i^g}{dL_g}dA_i$$

is the direct crowding out effect of public employment on wage labor (for the poorest households) and self-employment (for the richest households), $E_1 > 0$

$$E_2 = p \int_{\underline{A}}^{\phi(pW)} [\frac{dL_i^s}{dy_i}(W_g - pW)]\frac{dL_i^g}{dL_g} dA_i + \int_{\phi(W)}^{\overline{A}} [\frac{dL_i^s}{dy_i}(W_g - W)]\frac{dL_i^g}{dL_g} dA_i$$

is the effect on aggregate labor supply through non-labor income $E_2 < 0$. Hence $E_1 - E_2$ is positive as long as the income effect is not positive and large.

$$E_3 = p^2 \int_{\underline{A}}^{\phi(pW)} D'(pW) dA_i + \int_{\phi(W)}^{\overline{A}} D'(W) dA_i$$

is the effect on aggregate labor demand through a change in the wage, $E_3 < 0$.

$$E_4 = p^2 \int_{\underline{A}}^{\phi(pW)} [\frac{dL_i^s}{dW}|_u + \frac{dL_i^s}{dy_i}(L_i^s - D_i - L_i^g)] dA_i + \int_{\phi(W)}^{\overline{A}} [\frac{dL_i^s}{dW}|_u + \frac{dL_i^s}{dy_i}(L_i^s - D_i - L_i^g)] dA_i$$

is the effect on aggregate labor supply through a change in the wage. If leisure is not a luxury good, an increase in the wage should increase labor supply, so that $E_4 > 0$. Hence government hiring increases the equilibrium wage because $E_1 - E_2 > 0$, $-E_3 > 0$ and $E_4 > 0$. The effect is stronger when demand is less elastic (small $-E_3$), when labor supply is less elastic to the wage (small $E_4$).

Assuming that $p = 1$ we obtain the following

$$\frac{dW}{dL^g} = \frac{\int_{\underline{A}}^{\overline{A}} \frac{dL_i^g}{dL_g} dA_i - \int_{\underline{A}}^{\overline{A}} [\frac{dL_i^s}{dy_i}(W_g - W)]\frac{dL_i^g}{dL_g} dA_i}{-\int_{\underline{A}}^{\overline{A}} D'(W) dA_i + \int_{\underline{A}}^{\overline{A}} [\frac{dL_i^s}{dW}|_u + \frac{dL_i^s}{dy_i}(L_i^s - D_i - L_i^g)] dA_i}$$

From this equation, we see that an increase in government hiring will raise wages as long as the income effect is not positive and larger than one ($\int_{\underline{A}}^{\overline{A}} L_{y_i}^s (W_g - W) dA_i < 1$). The increase will be larger if demand is less elastic (small $-D'(W)$) or if labor supply is less elastic (small $\int_{\underline{A}}^{\overline{A}} \big(\frac{dL_i^s}{dW}|_u + \frac{dL_i^s}{dy_i}(L_i^s - L_i^g - D_i)\big) dA_i$).

In the special case where $p = 1$, the model indicates how empirical estimates of the impact of government hiring on employment and wages can be used to compute the labor demand elasticity. In a frictionless labor market, the change in aggregate private sector employment can be written as: $\frac{dD}{dL^g} = D'(W)\frac{dW}{dL^g}$, where $D'(W) = \int_{\underline{A}}^{\overline{A}} D_i'(W) dA_i$. Hence, in this framework, we can compute the elasticity of labor demand as the ratio of the percentage change in the wage divided by the percentage change in employment.

### E. Impact on Household Consumption

In this section, we derive the impact of a workfare program on household consumption. The impact on consumption is different from the impact on welfare because it also includes labor supply effects. Household consumption is given by:

$$c_i = \pi_i(\widetilde{W}_i) + \widetilde{W}_i L_i^s(\widetilde{W}_i, y_i) + (W_g - \widetilde{W}_i)L_i^g \tag{7}$$

Assuming a small change in $L^g$ ($\{L_i^g\}$), we totally differentiate 7 to obtain:

$$\begin{aligned} \frac{dc_i}{dL^g} &= (W_g - \widetilde{W}_i)\frac{dL_i^g}{dL^g} \\ &+ \widetilde{W}_i L_{yi}^s (W_g - \widetilde{W}_i)\frac{dL_i^g}{dL^g} \\ &+ (L_i^s - D_i - L_i^g)\frac{dW}{dL^g} \\ &+ \widetilde{W}_i \Big[\frac{dL_i^s}{dW}|_u + L_{y_i}^s(L_i^s + T - L_i^g - D_i)\Big]\frac{dW}{dL^g} \end{aligned}$$

The first term is the income gain due to participation in public works. The impact of this increase in income on labor supply is captured by the second term. It is negative if leisure is a normal good. Together, these first two terms yield the increase in consumption that would be observed by matching participants and non-participants in program areas.

The two last terms express the "indirect benefit", i.e. income gains accruing to households through equilibrium effects. The third term is the change in income due to the equilibrium change in the wage (holding labor supply constant). The last term captures the labor supply response due to the change in income from the equilibrium change in the wage. It is composed of a positive substitution effect and an income effect, which could be negative for households that are net buyers of labor.

### F. Imperfect Competition

We assume that the marginal productivity of labor is equal to the wage rate. Some have noted the presence of market power on the part of employers Binswanger and Rosenzweig (1984). If employers have market power then government hiring may actually increase private sector wages *and* employment. We refer the interested reader to Basu, Chau and Kanbur (2009), who provide a full analysis. Here, we sketch the main intuition and discuss the implications for the interpretation of the empirical results. A monopsonistic employer with production function $F(L)$ facing an inverse labor supply curve $W(L)$ sets the wage and employment such that:

$$F'(L^*) \quad = \quad W(L^*) + W'(L^*)L^* \tag{8}$$

This is the well-known result that the marginal productivity of labor will be above the wage rate if employers exercise their market power. The extent of the distortion depends on the slope of the labor supply curve ($W'(L)$). If the selection rule used by the government to hire workers under the workfare program shifts $W'(\cdot)$ down (makes labor supply more elastic), then all things equal, $L^*$ must increase to maintain the equality in Equation 8. Since the workfare program also reduces the available workforce, the net effect on private sector work is ambiguous.

For the present analysis, the important issue is whether, given the rise in wages due to the program, Equation 1 still captures the welfare impact of the program under imperfect competition. For labor suppliers, the welfare impact is the same. For labor buyers, however, Equation 1 no longer correctly captures the welfare impact of the program since the welfare impact now depends on how the inverse labor supply function changes, which in turn will be a function of the particular rationing rule used by the government.

## IV. Data Appendix

### A. *National Sample Survey Organisation: Employment Surveys*

**Sample:** The main data source used in this paper is the National Sample Survey rounds 55, 61, 64 and 66. These surveys are conducted on an irregular basis roughly every two years. They are "thick" rounds, with a sample size of roughly 70 thousand rural households.[1] The surveys are stratified by urban and rural areas of each district. The survey is conducted from July to June, and in each district, surveying is divided into four sub-rounds each lasting three months. Although the sample is not technically stratified by sub-round, the NSSO states that it attempts to distribute the number of households surveyed evenly within each district sub-round.

Table A.1 presents evidence on how the sample is distributed throughout the years in practice. For employment outcomes, a district is missing in a given quarter if no household was interviewed. From Table A.1 we see that in 1999-00, between 16 and 20 districts were missing per quarter, which is due to two separate issues. First, five districts were not at all surveyed, second, despite NSSO effort to distribute interviews in a given district during the whole year a few districts have been surveyed in some quarters only. This methodological issue was fixed in the later rounds, as can be seen in Table A.1. In 2004-05 and 2007-08, which are the years we use in our main specification, we have observations for almost all 497

[1] Two additional Rounds 60 and 62, have been conducted which we do not use in the analysis, because they are "thin" rounds, with roughly 35 thousand rural households.

district-quarters but one or two. In 2009-10, four districts are missing because they could not be matched unambiguously with 2007-08.

For casual wages, a district is missing in a given quarter if no household was surveyed or if no prime-age adult reported doing casual work in the past week. As a result the proportion of missing observations is larger for wages than for the employment variables. The fraction of missing observations is as high as seven percent for the first quarters of the survey year 1999-00, but not more than four percent for the years 2004-05 and 2007-08. One might worry that by reducing private employment the program may increase the probability that a district is missing in a given quarter. However, this does not seem to be a major concern given that the fraction of early districts among non-missing observations is constant across quarters.

**Outcomes:** Our main outcomes are individual measures of employment and wages, which are constructed as follows. The NSS Employment Survey includes detailed questions about the daily activities for all persons over the age of four in surveyed households for the most recent seven days. We compute for each person the fraction of days in the past seven days spent in each of four mutually exclusive activities: non-government work, public works, not in the labor force, and unemployed. Individuals who worked in casual labor over the past seven days are asked their total earnings from casual labor. For each individual we compute average earnings per day worked in casual labor. We perform a similar computation using days spent doing salaried labor to construct our measure of daily salaried earnings.

**Individual controls** For the purpose of our analysis of the impact of NREGA on casual labor earnings, we include workers characteristics as controls in our main specification. Individual controls include dummy variables for age groups (31-40, 41-50, 51-60), education levels (below primary, primary, middle, secondary or higher), caste (ST, SC, OBC), religion (Muslim, Other), gender (Female) and marital status (Married). The omitted category is single, illiterate, hindu males, aged 18-30 and belonging to general caste.

### *B. District Controls*

Table 1 provides a list of district controls and their sources. Here, we describe how the district controls are constructed.

**Census** A number of the district controls are computed from the primary census abstract of 2001. In all cases, we use information for rural areas only, which we then aggregate to the district level. We compute "fraction of scheduled tribes" and "fraction of scheduled castes" by dividing by total population. "Population density" is obtained by dividing total population by total area. "Literacy rate," "male labor force participation ratio" and "female labor force participation ratio" are respectively computed by dividing the number of literate persons, of male workers and of female workers respectively by total population aged six and over. “Fraction of labor force in agriculture” is obtained by dividing the number of rural individuals who report working as cultivators or agricultural laborers as their main or secondary occupation by the total number of workers. Finally, we use information from the census village directory to compute “irrigated cultivable land per capita” and “non irrigated cultivable land per capita.”

**Agricultural Productivity:** We compute agricultural productivity per worker for each agricultural year in each district using two sources of data. First, the Ministry of Agriculture publishes yearly data on output and harvest prices of 36 grain and cash crops in every district [2]. This allows us to compute the value of agricultural production for every district-year. Second, we use National Sample Survey data to estimate the number of (self employed and wage) workers active in agriculture for every district-year. NSS survey years match exactly the Ministry of Agriculture definition of agricultural years (July-June). Hence, dividing output value by the number of agricultural works yields agricultural productivity per worker for each NSS survey year.

**Rainfall** To control for monthly rainfall at the district level over the period 1999-2010, we use data from the Tropical Rainfall Measuring Mission (TRMM), which is a joint mission between NASA and the Japan Aerospace Exploration Agency (JAXA). The TRMM Muli-Satellite Precipitation Analysis provides rainfall data for every three hours at a resolution of 0.25 by 0.25 degree grid-cell size. Rainfall measurement are made by satellite and calibrated using monthly rain gauge analysis data from the Global Precipitation Climatology Project (GPCP).[3] The data is then scaled up to obtain mean monthly rainfall for every cell (see (Fetzer, 2013) for more details). There are on average 6 grid-cells per district. We compute cumulative rainfall in each district-month as the sum of rainfall since July 1st, and express it as percentage deviation from the 1998-2011 mean for this district-month.

**Other district controls** ”Pre-election year” is a dummy for whether state assembly or Panchayati Raj (local) elections are to be held in the following year. To construct this control, we used online reports from the Electoral Commission of India[4] and from the State Election Commissions of each states. “PMGSY Road Construction” is an estimate of the number of km of road built under the national rural roads construction program Pradhan Mantri Gram Sadak Yozna. We use online reports on each road built under the scheme to compute for each district quarter the average number of km completed per quarter over the last

[2] Data is available at `http://eands.dacnet.nic.in/`.
[3] Data is available at `http://trmm.gsfc.nasa.gov/`
[4] `http://www.eci.nic.in/ecimain1/index.aspx`

five quarters.[5]

### C. ARIS-REDS Household Hired Labor

For our calibration exercise in Section VI, we require estimates of labor hired by households, information which is not available in the NSS Employment Surveys. For this reason, we use the ARIS-REDS survey data, collected by the National Council of Applied Economic Research (Delhi) in 1999-00.[6] The ARIS-REDS survey covers a nationally representative rural sample of Indian households, with detailed information on household expenditures, on household members' employment income and on operating costs of households' farm and non-farm businesses.

For each household, we sum all income earned by prime-age household members from casual labor and total labor costs for farm and non-farm businesses. For each consumption quintile, we then compute the total casual payments as a fraction of total casual earnings for all households across all quintiles. Let $e_t^q$ and $p_t^q$ denote casual earnings and casual payments, respectively, for households in consumption quintile $q$ at date $t$. We compute for each quintile $f_{2000}^q = \frac{p_{2000}^q}{\sum_q e_{2000}^q}$. The resulting fractions are reported in the sixth row of Table 6. As expected the fraction of total casual earnings paid by households in the lower quintiles is much lower than the fraction paid by households in the upper quintiles. These fractions sum to less than one across consumption quintiles because some casual labor earnings come from urban employers.

In order to estimate casual labor payments by households of each consumption quintile in 2004-05, we make the assumption that casual labor payments made by each consumption quintile as a fraction of total earnings is constant over time, i.e. $f_{2005}^q = f_{2000}^q$. We then multiply total casual labor earnings from the NSS Employment Survey by the fractions in row six for each consumption quintile to obtain our estimate of casual labor payments by quintile: $\widehat{p_{2005}^q} = f_{2005}^q * \sum_q e_{2005}^q$. Our estimates are shown in row seven of Table 6

### D. Weighting

The NSSO provides sample weights which ensure that the weighted mean of each outcome is an unbiased estimate of the average of the outcome for the population National Sample Survey Office (2010). For the purpose of our analysis, we re-weight observations so that the sum of all weights within each district is constant over time and proportional to the rural population of the district as estimated from the NSS Employment Surveys. When we use NSSO survey weights without reweighting, the results are almost identical to our main results (results not shown). As compared to using ordinary least squares without any weighting,

[5] http://pmgsy.nic.in/
[6] http://adfdell.pstc.brown.edu/arisredsdata/readme.txt

our approach allows us to make sure that our results are not driven by smaller districts with few observations for casual wages. More concretely, let $w_i$ be the weight for person $i$, and let $\Omega_{dt}$ be the set of all persons surveyed in district $d$ at time $t$. Then the new weight for person $i$ is $w_i \times \frac{\omega_d}{\sum_{i \in \Omega_{dt}} w_i}$ where $\omega_d$ is the population weight for district $d$.

We also present estimates of our main specification without using any sample weight (see A.3). The estimated wage effects increase, which suggests that smaller districts experience larger changes in wages. Perhaps surprisingly, standard errors decrease slightly as compared to the estimation with sample weights. Whether the use of weights enhances precision or not depends on the variance structure of the error term (Solon, Haider and Wooldridge, 2013). On the one hand, smaller districts have fewer observations per quarter, hence taking into account differences in sample size across districts may increase precision. On the other hand, since labor market outcomes are highly correlated within districts, the district error may represent a large share of the variance of the error term, and the use of weights may harm precision. Following Solon, Haider and Wooldridge (2013) suggestion, we implemented a Breush Pagan test by regressing the squared error term on the inverse of the district sample size. The test confirms both the presence of heteroskedasticity, and the importance of the district error in the variance of the error term, with the latter effect dominating the former.

### *E. Construction of District Panel*

During the period covered by the analysis, some districts split while other districts merged together. Constructing the district panel requires matching districts both over time as well as across data sets. Fortunately, the NSS district definitions for surveying stayed constant from 2004 to 2008, despite splits and merges. We therefore use the NSS district definitions from this period and match other data sets to these. We first match the NSS 1999-2000 to 2004-05 and 2007-08 data. All districts could be matched between the two surveys but for five districts missing in 1999-00. However about fifty of them had split between 1999-00 and 2005-05. We adopt the following procedure If a given district has split in x districts (x is most of the time equal to two, sometimes three), we duplicate observations from that district x times so that one set of observation can be matched with one of the newly created district. In order to keep the total weight of that district constant, we divide each weight in the 1999-00 data-set by x. We next match the NSS 2009-10 data: all districts but four could be matched unambiguously with districts in NSS 2004-05 and 2007-08 data. In two occurrences, two districts were split to create a third one, making it impossible to match observations from the new districts to a specific district. We remove these districts from 2009-10 data. We further match NSS data with Census 2001 survey, NREGA phases 2005, ARIS-REDS 1999-00 survey, PMGSY road construction data from 2001 to 2010

## V. Alternative specifications

Berg et al. (2013) and Zimmermann (2013) estimate the labor market impact of the program using empirical strategies different from ours. In this section, we describe how we apply their strategy to our data and compare the resulting estimates with their findings.

### A. Berg et al. (2013)

Berg et al. (2013), use monthly wage time series over the whole 2000-2010 period to estimate the effect of the program using two alternative specifications. The first is a difference-in-differences strategy similar to ours. The second is a trend break model. In order to compare our results with Berg et al. (2013), we use all four survey years (from 1999-00 to 2009-10) from the NSS data and estimate the effect of the program using two different specifications.

First, We estimate our main specification without controls and using the four rounds:

$$Y_{idt} = \beta T_{dt} + \eta_t + \mu_d + \varepsilon_{idt}$$

This specification estimates the program impact using two difference-in-differences. The impact of the program is identified based on the difference between changes in outcomes in early districts and in late districts between 2004-05 and 2007-08 and the difference between changes in outcomes in late districts and in early districts between 2007-08 and 2009-10. The results are presented in the first column of table A.8. We find a significant increase in time spent on public works, a significant decrease in private sector work, and a significant 3.4% increase in casual wages. Findings from this specification are consistent with our main results and our estimates for the 2007-08 to 2009-10 period presented in Table A.5.

Second, we redefine the treatment variable $T_{dt}$ as the number of months since the program was launched in district $d$. We also include a district specific time trend $\delta_d$ and estimate the following equation:

$$Y_{idt} = \beta T_{dt} + \delta_d t + \eta_t + \mu_d + \varepsilon_{idt}$$

This specification identifies the program effect as a break in trends when the program was launched. As Table A.8 shows, the estimates provide strong evidence that the program had a positive effect on time spent in public works, a negative effect on time spent in private sector employment, and a positive effect on casual wages, with an estimated effect of 0.27% per month. Adding district-specific trends changes the magnitude of the coefficients but not their sign or their statistical significance. These results are close to Berg et al. (2013).

### B. Zimmermann (2013)

Zimmermann (2013) uses a regression discontinuity design to identify the effect of the program on employment and wages. The selection of early districts was based on a backwardness ranking made by the Planning Commission for an earlier program (Planning Commission, 2003). Hence within each state, and taking the number of early districts as given, one can use each district's backwardness rank to predict its assignment to early or late phases. One can then estimate the effect of the program by comparing 2007-08 outcomes between early and late phase districts close to the cut-off, controlling for the backwardness rank.

The identifying assumption of this regression discontinuity framework is that absent the program, districts to the left and the right of the cut-off would have had the same labor market outcomes. An important threat to this strategy is manipulation of the assignment of districts to early and late phases. In this context, it seems unlikely that the backwardness index itself was manipulated, since it was defined years before NREGA was invented. However, the number of early districts in each state (and hence the state level cut-off) was the result of an intense political bargain, and is unlikely to be exogenous (Gupta, 2006). We hence have some concerns regarding the validity of the regression discontinuity approach.

We first assess whether the algorithm accurately predicts whether a district is in early or late phases. Since the ranking is only available for 17 states, Himachal Pradesh and Uttarkhand are excluded from the sample. The prediction is accurate for 95% phase 1 districts, 81% of phase 2 districts and 84% of phase 3 districts. This suggests that the Planning Commission ranking was not perfectly followed for the assignment of districts into implementation phases. Political considerations likely explain why there was imperfect compliance, and why the regression discontinuity design is "fuzzy" (Gupta, 2006).

We follow Zimmermann (2013) and control for the outcome level at baseline (in 2004-05) $Y_{ds}^{05}$ and state fixed effects $\mu_s$ in the specification. We also restrict the sample to phase two and three districts (phase 1 districts are far from the cut-off) and estimate different polynomials of the district rank $R_{ds}$ to the left and to the right of the state specific cutoff $\kappa_s$. If $Y_{ids}^{08}$ denotes the outcome for individual $i$ in district $d$ and state $s$ in year 2007-08, the estimating equation is:

$$\begin{aligned}
Y_{ids}^{08} &= \beta T_{ds} + \delta_0 Y_{ds}^{05} + \delta_1 R_{ds} * (R_{ds} > \kappa_s) + \delta_2 R_{ds} * (R_{ds} < \kappa_s) \\
&+ \delta_3 (R_{ds})^2 * (R_{ds} > \kappa_s) + \delta_4 (R_{ds})^2 * (R_{ds} < \kappa_s) + \mu_s + \varepsilon_{id}
\end{aligned}$$

Table *A.9* presents the estimated program impact using this approach. We focus here on the flexible specification, which allows for a different slope to the right and to the left of the cutoff, which is Zimmermann (2013)'s preferred specification. We find a positive but insignificant effect of the program on time spent on public works (0.51 and 0.35 percentage points for the linear and quadratic specification respectively), a negative but insignificant effect on time spent doing private sector work (-0.8 and -1.5 percentage points), and positive effects on private sector wages (6 and 11%). These estimates are reasonably close to those of our preferred specification, and never significantly different from them. The estimation is however very noisy, and none of these estimates is significant.

## REFERENCES

**Afridi, Farzana.** 2008. "Can Community Monitoring Improve the Accountability of Public Officials." *Economic and Political Weekly*, 35–40.

**Basu, Arnab K., Nancy H. Chau, and Ravi Kanbur.** 2009. "A Theory of Employment Guarantees: Contestability Credibility and Distributional Concerns." *Journal of Public economics*, 93(3-4): 482–497.

**Berg, Erlend, Sambit Bhattacharyya, Rajasekhar Durgam, and Manjula Ramachandra.** 2013. "Can Rural Public Employment Schemes Increase Equilibrium Wages? Evidence from a Natural Experiment inIndia."

**Binswanger, H P, and M R Rosenzweig,** ed. 1984. *Contractural Arrangements, Employment, and Wages in Rural Labor Markets in Asia.* Yale University Press.

**Dreze, Jean.** 2005. "Loot For Work Programme." *The Times of India.*

**Dutta, Puja, Rinku Murgai, Martin Ravallion, and Dominique van de Walle.** 2012. "Does India's employment guarantee scheme guarantee employment ?" The World Bank Policy Research Discussion Paper 6003.

**Fetzer, Thiemo.** 2013. "Can Workfare Programs Moderate Violence? Evidence from India." Manuscript.

**Gupta, Santanu.** 2006. "Were District Choices for NFFWP Appropriate?" *Journal of Indian School of Political Economy*, 18(4): 641–648.

**Khera, Reetika,** ed. 2011. *The Battle for Employment Guarantee.* Oxford University Press.

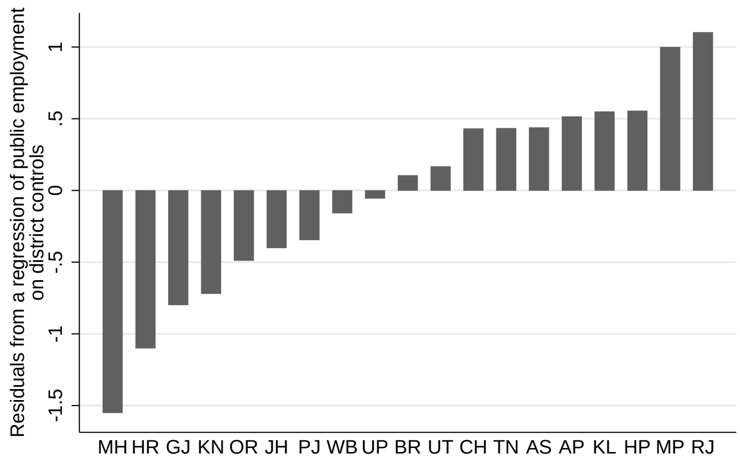
**National Sample Survey Office.** 2010. "Employment and Unemployment Sitation in India 2007-08." NSS 64th Round, July 2007 to June 2008, Appendix B.

**Planning Commission.** 2003. "Identification of Districts for Wage and Self Employment Programmes." Report of the Task Force.

**Solon, Gary, Steven J. Haider, and Jeffrey Wooldridge.** 2013. "What Are We Weighting For?" National Bureau of Economic Research, Inc NBER Working Papers 18859.

**Zimmermann, Laura.** 2013. "Why Guarantee Employment? Evidence from a Large Indian Public-Works Program."

Figure A.1. Unexplained heterogeneity in employment provision across States




Source: National Sample Survey (NSS) 2007-2008.
The sample is composed of all rural adults in early phase districts.
We compute state-level averages of residuals from a regression of time spent on public works on district controls presented in Table 1. All estimates are computed using sample weights.

Table A.1—Balance of District Panel

| | Q3 Jul-Sep (1) | Q4 Oct-Dec (2) | Q1 Jan-Mar (3) | Q2 Apr-Jun (4) |
|---|---|---|---|---|
| *Employment Variables* | | | | |
| 1999-00 | 478 | 478 | 483 | 482 |
| 2004-05 | 497 | 496 | 494 | 495 |
| 2007-08 | 496 | 497 | 495 | 497 |
| 2009-10 | 495 | 495 | 494 | 495 |
| *Casual Wages* | | | | |
| 1999-00 | 462 | 465 | 474 | 473 |
| 2004-05 | 478 | 480 | 479 | 481 |
| 2007-08 | 480 | 483 | 487 | 483 |
| 2009-10 | 475 | 475 | 476 | 479 |

Each cell shows the number of districts with non-missing observations per district-quarter. There are 497 districts in the panel. The NSS attempts to survey an equal number of villages in each districts during each quarter. During thick rounds (1999-2000, 2004-05, 2007-08, 2009-10), this is generally possible. Casual wages are only available for district-quarters during which at least one respondent reports working in casual labor. Five districts were not surveyed in 1999-2000.

Table A.2—Main Specification estimated without controls

| | Public | Private | Unemployed | Not in Labor Force | Log Deflated Daily Casual Earnings | Log Deflated Daily Casual Earnings |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Program X Dry | 0.964*** | -1.947*** | 1.623*** | -0.640* | 0.0353* | 0.0460** |
| | (0.246) | (0.640) | (0.452) | (0.369) | (0.0197) | (0.0232) |
| Program X Rainy | 0.206*** | -0.00801 | 0.653 | -0.851** | 0.00496 | 0.0215 |
| | (0.0768) | (0.586) | (0.455) | (0.391) | (0.0198) | (0.0260) |
| Observations | 356,636 | 356,636 | 356,636 | 356,636 | 64,167 | 64,167 |
| District Controls | No | No | No | No | No | Yes |
| Worker Controls | No | No | No | No | No | No |

Each column presents the results of a separate regression. All regressions include district and year-quarter fixed effects. In columns 1 through 5, the sample is composed of all adults aged 18 to 60 interviewed from July 2004 to June 2005 and from July 2007 to June 2008. The unit of observation is a person. The outcomes are defined as in Table 3. Program is a dummy variable equal to one for early districts during July 2007 to June 2008. Dry is a dummy variable equal to one for the first two quarters of the year. Rainy is a dummy variable equal to one for the second two quarters of the year. District controls are listed in Table 1. District controls that do not vary over time are interacted with a dummy for 2007-08 (post-program). Worker controls are listed in the notes of Table 3. All estimates are computed using sampling weights. Standard errors in parentheses are adjusted for correlation of the errors at the district level. ***, **, and * indicate significance at the 1, 5, and 10% levels.

Table A.3—Main Specification estimated without Sample Weights

| | Public | Private | Unemployed | Not in Labor Force | Log Deflated Daily Casual Earnings |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Program X Dry | 1.110*** | -1.402** | 0.449 | -0.157 | 0.0679*** |
| | (0.254) | (0.639) | (0.470) | (0.376) | (0.0207) |
| Program X Rainy | 0.466*** | 0.265 | -0.206 | -0.526 | 0.0563** |
| | (0.159) | (0.624) | (0.496) | (0.371) | (0.0219) |
| Observations | 356,636 | 356,636 | 356,636 | 356,636 | 64,167 |
| District Controls | Yes | Yes | Yes | Yes | Yes |
| Worker Controls | No | No | No | No | Yes |

Each column presents the results of a separate regression. All regressions include district and year-quarter fixed effects. In columns 1 through 5, the sample is composed of all adults aged 18 to 60 interviewed from July 2004 to June 2005 and from July 2007 to June 2008. The unit of observation is a person. The outcomes are defined as in Table 3. Program is a dummy variable equal to one for early districts during July 2007 to June 2008. Dry is a dummy variable equal to one for the first two quarters of the year. Rainy is a dummy variable equal to one for the second two quarters of the year. District controls are listed in Table 1. District controls that do not vary over time are interacted with a dummy for 2007-08 (post-program). Worker controls are listed in the notes of Table 3. All estimates are computed without sampling weights. Standard errors in parentheses are adjusted for correlation of the errors at the district level. ***, **, and * indicate significance at the 1, 5, and 10% levels.

Table A.4—Main Specification controlling for changes in outcomes between 1999-00 and 2004-05

| | Public | Private | Unemployed | Not in Labor Force | Log Deflated Daily Casual Earnings |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Program X Dry | 1.163*** | -1.666** | 0.329 | 0.235 | 0.0543** |
| | (0.297) | (0.731) | (0.505) | (0.443) | (0.0210) |
| Program X Rainy | 0.457*** | 0.274 | -0.630 | -0.0305 | 0.0386 |
| | (0.174) | (0.740) | (0.548) | (0.536) | (0.0235) |
| Observations | 356,636 | 356,636 | 356,636 | 356,636 | 64,167 |
| District Controls | Yes | Yes | Yes | Yes | Yes |
| Worker Controls | No | No | No | No | Yes |

Each column presents the results of a separate regression. All regressions include district and year-quarter fixed effects. Each regression also includes the change in average outcome in the district between 1999-00 and 2004-05. The sample is composed of all adults aged 18 to 60 interviewed from July 2004 to June 2005 and from July 2007 to June 2008. The outcomes are defined as in Table 3. Program is a dummy variable equal to one for early districts during July 2007 to June 2008. Dry is a dummy variable equal to one for the first two quarters of the year. Rainy is a dummy variable equal to one for the second two quarters of the year. District controls are listed in Table 1. District controls that do not vary over time are interacted with a dummy for 2007-08 (post-program). Worker controls are listed in the notes of Table 3. All estimates are computed using sampling weights. Standard errors in parentheses are adjusted for correlation of the errors at the district level. ***, **, and * indicate significance at the 1, 5, and 10% levels.

Table A.5—Changes in outcomes in late relative to early phase districts between 2007-08 and 2009-10, when the program is extended to late phase districts

| | Public | Private | Unemployed | Not in Labor Force | Log Deflated Daily Casual Earnings |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Program | 0.00569*** | -0.0106** | 0.00838** | -0.00297 | 0.0340** |
| | (0.00139) | (0.00488) | (0.00362) | (0.00299) | (0.0171) |
| Program X Dry | 0.00948*** | -0.0203*** | 0.0132*** | -0.00191 | 0.0489** |
| | (0.00247) | (0.00612) | (0.00412) | (0.00344) | (0.0191) |
| Program X Rainy | 0.00190** | -0.000908 | 0.00353 | -0.00403 | 0.0191 |
| | (0.000741) | (0.00547) | (0.00409) | (0.00396) | (0.0190) |
| Observations | 356,636 | 356,636 | 356,636 | 356,636 | 64,167 |
| District Controls | No | No | No | No | No |
| Worker Controls | No | No | No | No | No |
| | Public | Private | Unemployed | Not in Labor Force | Log Deflated Daily Casual Earnings |
| | (1) | (2) | (3) | (4) | (5) |
| Program | 0.00804*** | -0.00710 | -0.00101 | 0.000695 | 0.0496** |
| | (0.00215) | (0.00662) | (0.00482) | (0.00435) | (0.0202) |
| Program X Dry | 0.0114*** | -0.0163** | 0.00349 | 0.00201 | 0.0568*** |
| | (0.00291) | (0.00734) | (0.00504) | (0.00448) | (0.0207) |
| Program X Rainy | 0.00425** | 0.00336 | -0.00612 | -0.000810 | 0.0412* |
| | (0.00172) | (0.00747) | (0.00551) | (0.00541) | (0.0232) |
| Observations | 356,636 | 356,636 | 356,636 | 356,636 | 64,167 |
| District Controls | Yes | Yes | Yes | Yes | Yes |
| Worker Controls | No | No | No | No | Yes |

Each column presents the results of a separate regression. All regressions include district and year-quarter fixed effects. Each regression includes 1999-00 to 2004-05 changes in outcome in each district. The sample is composed of all adults aged 18 to 60 interviewed from July 2004 to June 2005 and from July 2007 to June 2008. The outcomes are defined as in Table 4 and 5. Program is a dummy variable equal to one for early districts during July 2007 to June 2008. Dry is a dummy variable equal to one for the first two quarters of the year. Rainy is a dummy variable equal to one for the second two quarters of the year. District controls are listed in Table 1. District controls that do not vary over time are interacted with a dummy for 2007-08 (post-program). Worker controls are listed in the notes of Table 5. Star states is a dummy variable equal to one for districts within star states. Other states is a dummy variable equal to one for districts that are not in star states. See Table 2 for a description of star states. All estimates are computed without sampling weights. Standard errors in parentheses are adjusted for correlation of the errors at the district level. ***, **, and * indicate significance at the 1, 5, and 10% levels.

Table A.6—Main Specification controlling for state specific time effects

| | Public | Private | Unemployed | Not in Labor Force | Log Deflated Daily Casual Earnings |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Program X Dry | 1.083*** | -1.864** | 0.833 | -0.0520 | 0.0424** |
| | (0.285) | (0.743) | (0.544) | (0.465) | (0.0198) |
| Program X Rainy | 0.333 | 0.170 | -0.127 | -0.376 | 0.0101 |
| | (0.208) | (0.785) | (0.581) | (0.531) | (0.0224) |
| Observations | 356,636 | 356,636 | 356,636 | 356,636 | 64,167 |
| District Controls | Yes | Yes | Yes | Yes | Yes |
| Worker Controls | No | No | No | No | Yes |

Each column presents the results of a separate regression. All regressions include district and year-quarter fixed effects, as well as a dummy for each state interacted with a dummy for 2007-08.B41 The sample is composed of all adults aged 18 to 60 interviewed from July 2004 to June 2005 and from July 2007 to June 2008. The outcomes are defined as in Table 3. Program is a dummy variable equal to one for early districts during July 2007 to June 2008. Dry is a dummy variable equal to one for the first two quarters of the year. Rainy is a dummy variable equal to one for the second two quarters of the year. District controls are listed in Table 1. District controls that do not vary over time are interacted with a dummy for 2007-08 (post-program). Worker controls are listed in the notes of Table 3. All estimates are computed using sampling weights. Standard errors in parentheses are adjusted for correlation of the errors at the district level. ***, **, and * indicate significance at the 1, 5, and 10% levels.

Table A.7—Program effect on wages for men and women

| | Log Deflated Casual Wages | | |
|---|---|---|---|
| | **Whole sample** | **Male Workers** | **Female Workers** |
| | (1) | (2) | (3) |
| Program | 0.0202 | 0.0177 | 0.0481*** |
| | (0.0127) | (0.0184) | (0.0152) |
| Program X Dry | 0.0353** | 0.0321** | 0.0630*** |
| | (0.0151) | (0.0151) | (0.0235) |
| Program X Rainy | 0.00496 | 0.00285 | 0.0347 |
| | (0.0152) | (0.0153) | (0.0225) |
| Observations | 44,278 | 19,889 | 64,167 |
| District Controls | No | No | No |
| Workers Controls | No | No | No |
| | **Whole sample** | **Male Workers** | **Female Workers** |
| | (4) | (5) | (6) |
| Program | 0.0403*** | 0.0421** | 0.0292 |
| | (0.0166) | (0.0221) | (0.0201) |
| Program X Dry | 0.0488*** | 0.0516*** | 0.0381 |
| | (0.0162) | (0.0177) | (0.0250) |
| Program X Rainy | 0.0304 | 0.0313 | 0.0186 |
| | (0.0187) | (0.0199) | (0.0277) |
| Observations | 44,278 | 19,889 | 19,889 |
| District Controls | Yes | Yes | Yes |
| Workers Controls | Yes | Yes | Yes |

Each column presents results from a separate regression. All regressions include district and year-quarter fixed effects. The sample is composed of all adults aged 18 to 60 interviewed from July 2004 to June 2005 and from July 2007 to June 2008. Program is a dummy variable equal to one for early districts during July 2007 to June 2008. Dry is a dummy variable equal to one for the first two quarters of the year. Rainy is a dummy variable equal to one for the second two quarters of the year. District controls are listed in Table 1. District controls that do not vary over time are interacted with a dummy for 2007-08 (post-program). All estimates are computed using sampling weights. Standard errors in parentheses are adjusted for correlation of the errors at the district level. ***, **, and * indicate significance at the 1, 5, and 10% levels.

Table A.8—Trend break specification

| | PUBLIC WORKS | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Program | 0.00425*** | | |
| | (0.00133) | | |
| Months in the Program | | 0.000126*** | 6.75e-05*** |
| | | (2.39e-05) | (2.21e-05) |
| Observations | 692,651 | 692,651 | 692,651 |
| District Trends | No | No | Yes |
| | **Private Sector Work** | | |
| | (1) | (2) | (3) |
| Program | -0.00973** | | |
| | (0.00421) | | |
| Months in the Program | | -0.000168*** | -0.000148*** |
| | | (4.81e-05) | (4.34e-05) |
| Observations | 692,651 | 692,651 | 692,651 |
| District Trends | No | No | Yes |
| | **Log Deflated Casual Wages** | | |
| | (1) | (2) | (3) |
| Program | 0.0363** | | |
| | (0.0141) | | |
| Months in the Program | | 0.00352*** | 0.00268*** |
| | | -0.000136 | -0.000115 |
| Observations | 125,339 | 125,339 | 125,339 |
| District Trends | No | No | Yes |

Each column presents results from a separate regression. All regressions include district and year-quarter fixed effects. The sample is composed of all adults aged 18 to 60 interviewed from July 1999 to June 2000, from July 2004 to June 2005, from July 2007 to June 2008 and from July 2009 to June 2010. Program is a dummy variable equal to one for early districts during July 2007 to June 2010 and for late districts during July 2009 to June 2010. "Months in the Program" is equal to the number of months since NREGA was launched, i.e. February 2006, April 2007 and April 2008 for districts in first, second and third phase respectively. The specification is described in Section A.5 in Appendix. No control is included. All estimates are computed using weights proportional to district population. Standard errors in parentheses are adjusted for correlation of the errors at the district level. ***, **, and *

Table A.9—Regression discontinuity

| **Panel A** | Dependent Variable: Public Sector Work | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Predicted NREGA | 0.00367 | 0.00309 | 0.00507 | 0.00238 | 0.00346 |
| | (0.00250) | (0.00315) | (0.00357) | (0.00334) | (0.00524) |
| Observations | 1,063 | 1,063 | 1,063 | 1,063 | 1,063 |
| **Panel B** | Dependent Variable: Private Sector Work | | | | |
| | (1) | (2) | (3) | (4) | (5) |
| Predicted NREGA | -0.00611 | -0.00416 | -0.00849 | -0.00483 | -0.0151 |
| | (0.00642) | (0.00881) | (0.0112) | (0.00939) | (0.0183) |
| Observations | 1,063 | 1,063 | 1,063 | 1,063 | 1,063 |
| **Panel C** | Dependent Variable: Log Deflated Daily Casual Earnings | | | | |
| | (1) | (2) | (3) | (4) | (5) |
| Predicted NREGA | -0.126*** | -0.00295 | 0.0630 | -0.00517 | 0.113 |
| | (0.0261) | (0.0366) | (0.0474) | (0.0410) | (0.0885) |
| Observations | 872 | 872 | 872 | 872 | 872 |
| Baseline Control | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes |
| Linear Slope | No | Yes | Yes | Yes | Yes |
| Quadratic Slope | No | No | No | Yes | Yes |
| Flextible Specification | No | No | Yes | No | Yes |

Each column presents results from a separate regression. A unit of observation is a district-quarter. The sample is composed of all adults aged 18 to 60 interviewed from July 2007 to June 2008 living in second and third NREGA phase districts. Predicted NREGA is a dummy variable equal to one if the district rank according to the Planning Commission Backwardness Index is lower than the state specific cut-off for early phases. The specification is described in Section A.5 in Appendix. Flexible Specification allows for different slopes to the right and to the left of the cutoff. All estimates are computed using weights proportional to district population. Standard errors in parentheses are adjusted for correlation of the errors at the district level. ***, **, and * indicate significance at the 1, 5, and 10% levels.

Table A.10—Outcomes by Consumption Quintile

| | Public | Private | Unemployed | Not in Labor Force | Log Deflated Daily Earnings |
|---|---|---|---|---|---|
| | (1) | (2) | (5) | (6) | (7) |
| Program X Dry X Quintile 1 | 1.928*** | -2.768* | 1.823* | -0.984 | 0.0633* |
| | (0.700) | (1.457) | (1.075) | (0.802) | (0.0353) |
| Program X Dry X Quintile 2 | 0.956** | -0.887 | 0.490 | -0.560 | 0.0589* |
| | (0.443) | (1.100) | (0.775) | (0.680) | (0.0318) |
| Program X Dry X Quintile 3 | 1.271*** | -1.119 | -1.059 | 0.907 | -0.0125 |
| | (0.321) | (1.127) | (0.807) | (0.692) | (0.0322) |
| Program X Dry X Quintile 4 | 0.831*** | -1.072 | -0.0343 | 0.275 | 0.0546 |
| | (0.223) | (0.975) | (0.686) | (0.702) | (0.0410) |
| Program X Dry X Quintile 5 | 0.616*** | -0.325 | -0.574 | 0.283 | -0.0538 |
| | (0.218) | (1.085) | (0.636) | (0.939) | (0.0582) |
| Program X Rainy X Quintile 1 | 0.607*** | -0.371 | -0.0489 | -0.187 | 0.0342 |
| | (0.231) | (1.379) | (1.195) | (0.780) | (0.0370) |
| Program X Rainy X Quintile 2 | 0.634*** | -0.583 | -0.417 | 0.367 | 0.0359 |
| | (0.199) | (1.014) | (0.829) | (0.634) | (0.0338) |
| Program X Rainy X Quintile 3 | 0.691*** | 1.222 | -1.411** | -0.502 | -0.0329 |
| | (0.198) | (0.968) | (0.717) | (0.666) | (0.0367) |
| Program X Rainy X Quintile 4 | 0.421** | 0.788 | -0.517 | -0.692 | 0.0219 |
| | (0.209) | (0.965) | (0.734) | (0.671) | (0.0394) |
| Program X Rainy X Quintile 5 | 0.368** | 1.867 | -1.786*** | -0.449 | -0.0304 |
| | (0.168) | (1.253) | (0.638) | (1.112) | (0.0651) |
| Observations | 356,636 | 356,636 | 356,636 | 356,636 | 64,167 |
| District x Quintile FE | Yes | Yes | Yes | Yes | Yes |
| Quarter x Year FE | Yes | Yes | Yes | Yes | Yes |
| District Controls | Yes | Yes | Yes | Yes | Yes |
| Worker Controls | No | No | No | No | Yes |

The unit of observation is a person. The sample uses all persons 18 to 60. The outcomes are defined as in Table 3. Program is a dummy variable equal to one for early districts during July 2007 to June 2008. Dry is a dummy variable equal to one for the first two quarters of the year. Rainy is a dummy variable equal to one for the second two quarters of the year. Quintile 1 to 5 are dummy variables equal to one if the individual is in a household with expenditure in that quintile. Quintile 1 is the poorest quintile. District controls are listed in Table 2. District controls that do not vary over time are interacted with a dummy for 2007-08 (post-program). Worker controls are listed in the notes of Table 3. The sample includes all observations from July 2004 to June 2005 and from July 2007 to June 2008. All estimates are computed using sampling weights. ***, **, and * indicate significance at the 1, 5, and 10% levels.