Climate Change and Labor Reallocation: Evidence from Six Decades of the Indian Census

Online Appendices

Maggie Liu
Yogita Shamdasani
Vis Taraz

Appendix A. Theoretical Framework

We develop a simple general equilibrium model of sectoral allocation with three sectors and two regions. We have two goals in developing this model. First, we want to demonstrate that adverse shocks to agricultural productivity can decrease the allocation of labor to the non-agricultural sector via local demand effects. Second, we want to show that high transportation costs can intensify this effect. We develop a model in the spirit of the model developed in Gollin and Rogerson (2014).[1] More broadly, the mechanisms underlying our theoretical model are similar to the mechanisms in many models of structural transformation (Matsuyama, 1992; Kongsamut, Rebelo and Xie, 2001; Gollin, Parente and Rogerson, 2002; Gollin and Rogerson, 2014; Herrendorf, Rogerson and Valentinyi, 2014; Bustos, Caprettini and Ponticelli, 2016).

Our model features three sectors: the agricultural sector (A), the manufacturing sector (M), and the service sector (S). There are two regions: the rural region (R) which produces agriculture and services, and the urban region (U) which produces manufacturing and services. Services are not tradable between the regions, but agricultural and manufacturing goods can be traded. We assume iceberg transportation costs of size q for the agricultural good only. Specifically, we assume that if one unit of the agricultural good is transported to the urban region, there will be fractional losses q so that only $(1-q)$ units will be received.

Individuals in both regions have Stone-Geary preferences over the three goods, which, for tractability, we assume are of the form:

$$u(c_A, c_M, c_S) = ln(c_A - \bar{a}) + ln(c_M - \bar{m}) + ln(c_S)$$

We assume that $\bar{a} > 0$, so that the income elasticity of the agricultural good is less than one. Labor is the only input for all three sectors. Each region has one unit of available labor, and the labor in region J devoted to sector I is denoted by L_I^J. Urban labor is divided between manufacturing and services so that $L_M^U + L_S^U = 1$, and all agricultural goods consumed in the urban region are imported from the rural region. In the rural region, labor is divided between the agricultural and services sectors, so that $L_A^R + L_S^R = 1$ and all manufacturing goods consumed in the rural region are imported from the urban region. We abstract away from the possibility of migration between regions. The production function of the agricultural sector in the rural region is given by:

$$Y_A^R = \theta_A^R (L_A^R)^\beta = \theta_A^R (1 - L_S^R)^\beta$$

where θ_A^R is rural agricultural total factor productivity and β is the elasticity of output with respect to labor. We assume that $\beta < 1$.

[1] Our model differs in three important ways from the model developed in Gollin and Rogerson (2014): we model one rural region instead of two; we shut down migration between the regions; and we model a (non-tradable) service sector in addition to the agricultural and manufacturing sectors.

In the context of our study, shocks to θ_A^R will be driven by higher temperatures. The dependence of agricultural productivity on temperature is well-documented in the empirical literature (Schlenker and Roberts, 2009; Dell, Jones and Olken, 2012; Taraz, 2018). We note that in the Indian context, there is also empirical evidence that higher temperatures reduce non-agricultural productivity (Somanathan et al., 2021), but these reductions are smaller in magnitude relative to impacts on agricultural productivity (Jain, O'Sullivan and Taraz, 2020).[2]

The output of the manufacturing sector in the urban region is given by:

$$Y_M^U = \theta_M^U L_M^U = \theta_M^U (1 - L_S^U)$$

The output of the service sector in each region is given by:

$$Y_S^J = \theta_S^J L_S^J$$

where θ_S^J is the service sector total factor productivity in region J. We further assume that $\theta_S^R = 1$.

To solve for the competitive equilibrium of our model, we note that our model contains no externalities and hence we can apply the First Welfare Theorem. In other words, the competitive equilibrium will be the same as the solution to the Social Planner's problem, and that is what we will solve for, applying equal weights to each region. The maximization problem for the social planner is given by:

$$\max_{c_A^R, c_A^U, c_M^R, c_M^U, c_S^R, c_S^U, L_S^R, L_S^U} ln(c_A^R - \bar{a}) + ln(c_M^R - \bar{m}) + ln(c_S^R) + ln(c_A^U - \bar{a}) + ln(c_M^U - \bar{m}) + ln(c_S^U)$$

where c_I^J be the consumption of good I in region J. This maximization problem is subject to the following four feasibility constraints:

$$\text{(A1)} \qquad \frac{c_A^U}{(1-q)} + c_A^R = \theta_A^R (1 - L_S^R)^{\beta}$$

$$\text{(A2)} \qquad c_M^U + c_M^R = \theta_M^U (1 - L_S^U)$$

$$\text{(A3)} \qquad c_S^R = L_S^R$$

$$\text{(A4)} \qquad c_S^U = \theta_S^U L_S^U$$

Equation (A1) states that the sum of agricultural consumption in the urban region (factoring in transportation costs) and the rural region must equal the output of the agricultural sector in the rural region. Equation (A2) states that the total manufacturing consumption across both regions must be equal to the manufacturing output of the urban region. Equations

[2]This India-specific evidence is consistent with cross-country evidence that the agricultural sector is more sensitive to higher temperatures than the non-agricultural sector in poor countries, but that both respond negatively to higher temperatures (Dell, Jones and Olken, 2012).

(A3) and (A4) state that service consumption in each region must equal service production in that same region, since services are not tradable.

Substituting Equations (A3) and (A4) into the maximization problem, we get a simplified maximization that is subject to Equations (A1) and (A2) only:

$$\max_{c_A^R, c_A^U, c_M^R, c_M^U, L_S^R, L_S^U} ln(c_A^R - \bar{a}) + ln(c_M^R - \bar{m}) + ln(\theta_S^R L_S^R) + ln(c_A^U - \bar{a}) + ln(c_M^U - \bar{m}) + ln(\theta_S^U L_S^U)$$

Manipulating the first order constraints of this maximization problem, we are able to solve for the following expressions for the rural and urban consumption bundles:

$$\begin{aligned}
c_A^R =& \frac{\theta_A^R (1 - L_S^R)^\beta (1 - q) - \bar{a}q}{2(1 - q)} \\
c_A^U =& \frac{\theta_A^R (1 - L_S^R)^\beta (1 - q) + \bar{a}q}{2} \\
c_M^U =& c_M^R = \frac{\theta_M^U (1 - L_S^U)}{2}
\end{aligned}$$

We can then write a further simplified version of the Social Planner's Problem, by substituting in these expressions for c_A^M, c_A^S, c_A^U, and c_S^U:

$$\begin{aligned}
\max_{L_S^R, L_S^U} ln &\left(\frac{\theta_A^R (1 - L_S^R)^\beta (1 - q) - (2 - q)\bar{a}}{2(1 - q)} \right) + ln \left(\frac{\theta_M^U - \theta_M^U L_S^U - 2\bar{m}}{2} \right) + ln \left(L_S^R \right) \\
&+ ln \left(\frac{\theta_A^R (1 - L_S^R)^\beta (1 - q) - (2 - q)\bar{a}}{2} \right) + ln \left(\frac{\theta_M^U - \theta_M^U L_S^U - 2\bar{m}}{2} \right) + ln \left(\theta_S^U L_S^U \right)
\end{aligned}$$

Using the rule for a logarithm of a quotient, we can simplify this maximization problem to:

$$\begin{aligned}
\max_{L_S^R, L_S^U} 2ln &\left(\theta_A^R (1 - L_S^R)^\beta (1 - q) - (2 - q)\bar{a} \right) + 2ln \left(\theta_M^U - \theta_M^U L_S^U - 2\bar{m} \right) + ln \left(L_S^R \right) \\
&+ ln \left(\theta_S^U L_S^U \right) - ln(1 - q)
\end{aligned}$$

Next, we take for the first order conditions for L_S^R and we get:

$$\frac{-2\beta(1 - q)\theta_A^R (1 - L_S^R)^{\beta - 1}}{\theta_A^R (1 - L_S^R)^\beta (1 - q) - (2 - q)\bar{a}} + \frac{1}{L_S^R} = 0$$

Manipulating this expression, we get

$$(1 - L_S^R)^\beta (1 - q) - 2\beta(1 - q)(1 - L_S^R)^{\beta - 1} L_S^R = \frac{\bar{a}(2 - q)}{\theta_A^R} \tag{A5}$$

Now, we can solve for the comparative statics of interest. First, we can solve for the impact of a change in agricultural productivity on rural labor supply to the service sector. Taking the implicit derivative of Equation (A5) with respect to agricultural productivity θ_A^R and rearranging terms, we get:

$$\text{(A6)} \qquad \frac{\delta L_S^R}{\delta \theta_A^R} = \frac{\bar{a}(2-q)}{(\theta_A^R)^2(1-q)} \times \frac{1}{3\beta(1-L_S^R)^{\beta-1} + 2\beta(1-\beta)(1-L_S^R)^{\beta-2}L_S^R}$$

We can now determine the sign of the expression on the right-hand side of Equation (A6). We have assumed that $\bar{a} > 0$ and $q < 1$, so the numerator of the fraction is positive. We have also assumed that $\beta < 1$, and $L_S^R < 1$, so the denominator is also positive. Therefore we have shown that $\frac{\delta L_S^R}{\delta \theta_A^R} > 0$, which means that an increase in agricultural productivity triggers an increase of labor allocated to the service sector (and hence decrease labor supply to the agricultural sector). Conversely, an adverse shock to agricultural productivity will decrease rural service sector employment and increase rural agricultural labor supply.

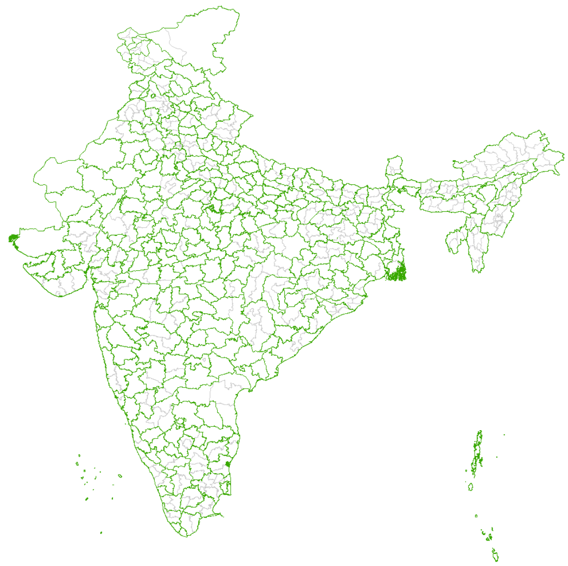
In addition to the direct effect of agricultural productivity shocks on non-agricultural labor supply, we are also interested in the role of transportation costs in modulating this relationship. Looking at Equation (A6), we note that increasing q (in the range $0 < q < 1$) will increase the right-hand side of Equation (A6), and hence $\frac{\delta^2 L_S^R}{\delta \theta_A^R \delta q} > 0$, which means that places with higher transportation costs will face intensified local demand effects.

It is worth noting two important limitations of our model. First, our model does not allow for migration between rural and urban regions. This assumption may be reasonable in the short-term, since the costs of seasonal migration in India is high and workers prefer local public works to migration (Imbert and Papp, 2020). In the medium- to long-term, the level of within-district migration in India is also very low (Kone et al., 2018). Nevertheless, the implications of our model should be caveated when applied to other contexts with higher levels of cross-region migration. On the one hand, local demand effects could be dampened as cross-region migration arbitrages away the difference in agricultural productivity shocks. On the other hand, spatial linkages created by migration could expose regions unaffected by rising temperatures to agricultural risks elsewhere.

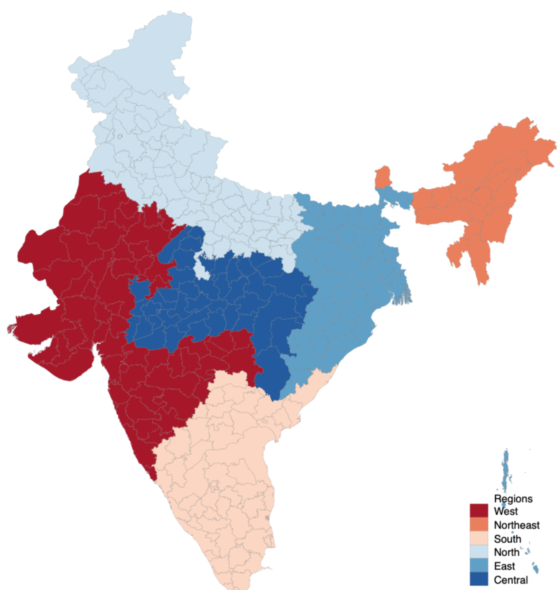
Second, our set-up does not allow for multiple periods, and therefore it cannot provide comparisons of the short- and long-term effects. Without formally modeling multiple periods, the predictions from the above model can be taken as short-term dynamics. To gauge possible intensification or adaptation effects in the longer-term, we posit two possible extensions to the model. The first is to introduce liquidity and mobility costs, which may intensify the effects of rising temperatures over time. If farmers' agricultural incomes are stochastic, it follows that a short duration of high temperatures will reduce farm income, which renders the costs of switching sectors infeasible for a relatively small fraction of farmers. However, a longer period of sustained high temperatures will lead to long-lasting farm income reductions, leaving a much greater fraction of the population not able to afford the liquidity and mobility costs. A second possible extension is to consider costly investment in human capital. Under this set-up, warming affects not only the affordability of switching sectors for *current* workers, but also the human capital investments of *future* workers. A growing literature documents that high temperatures have negative and persistent impacts on human capital.[3] It may be the case that the adverse impacts of warming on structural transformation in the short-term can be compounded by the dampening of human capital accumulation in the longer-term.

[3] For example, Garg, Jagnani and Taraz (2020) find that higher temperatures in India reduce contemporaneous human capital due to an agricultural income channel. Fishman, Carrillo and Russ (2019) find that high temperatures in Ecuador around the time at birth have long-term effects on human capital and earnings productivity that persist into adulthood; Hu and Li (2019) find similar effects looking at China.

Appendix B. Figures and Tables



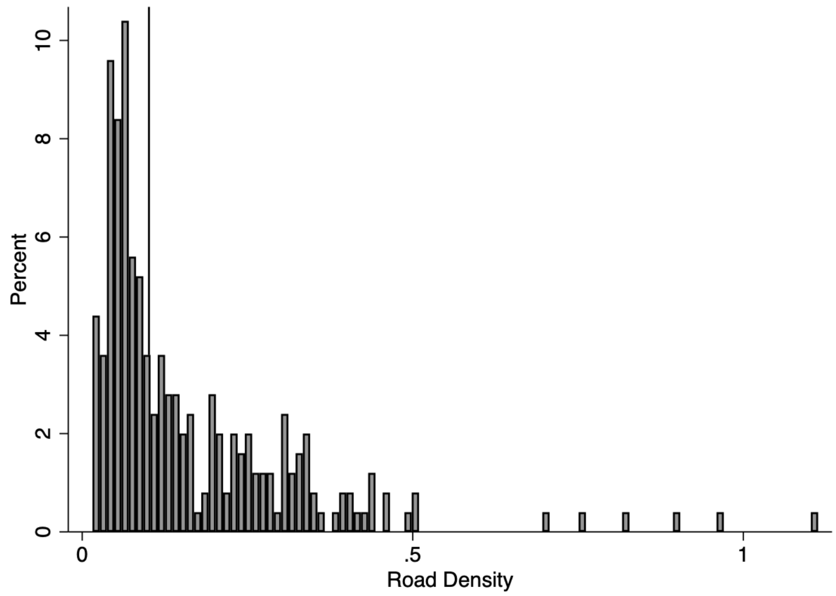
(A) Consistent district boundaries, 1961-2011



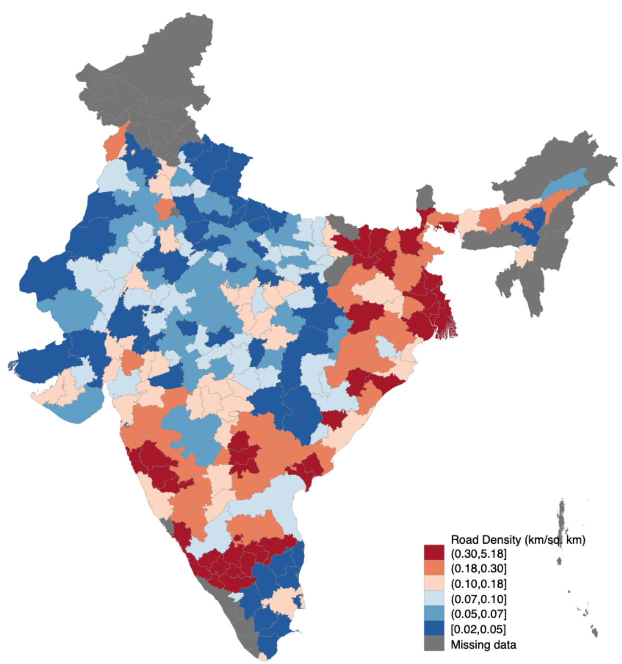
(B) Districts, by region

Figure B1. Analysis Districts

Note: Figure illustrates the 288 consistent district boundaries over 1961-2011 (panel A), and the 287 districts by region used in the analysis (panel B). Lakshadweep is dropped due to lack of weather records. We classify districts into six regions based on the Government of India's administrative regional classification.



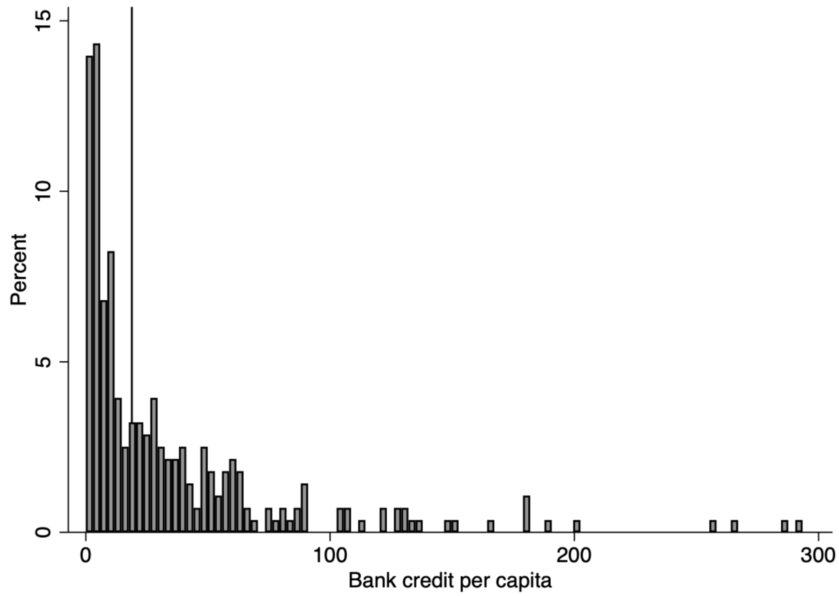
(A) Histogram



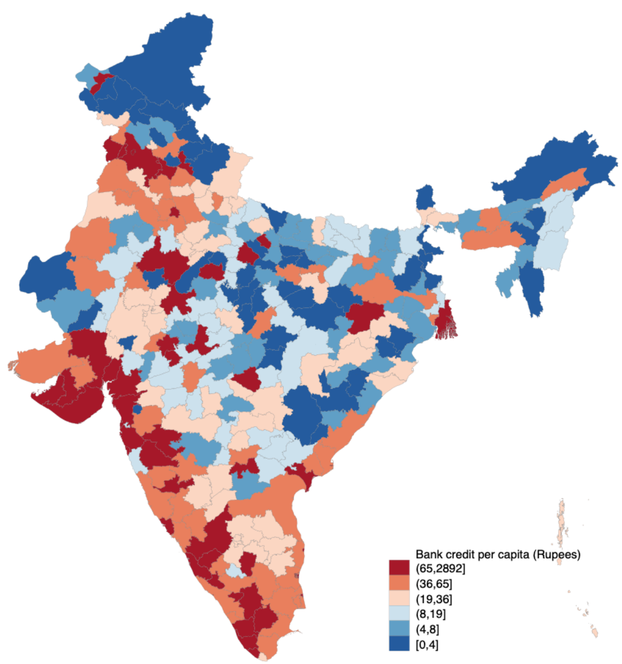
(B) Spatial Distribution

FIGURE B2. Road Density Measure

Note: Figure plots the road density (km/km^2) measure across all districts in panel A, and illustrates the distribution of the same measure across space in panel B. The solid vertical line in panel A denotes the median of the distribution (0.10 km/km^2).



(A) Histogram



(B) Spatial Distribution

FIGURE B3. Bank Credit per Capita Measure

Note: Figure plots the bank credit per capita (Rupees) measure across all districts in panel A, and illustrates the distribution of the same measure across space in panel B. The solid vertical line in panel A denotes the median of the distribution (19 Rupees).

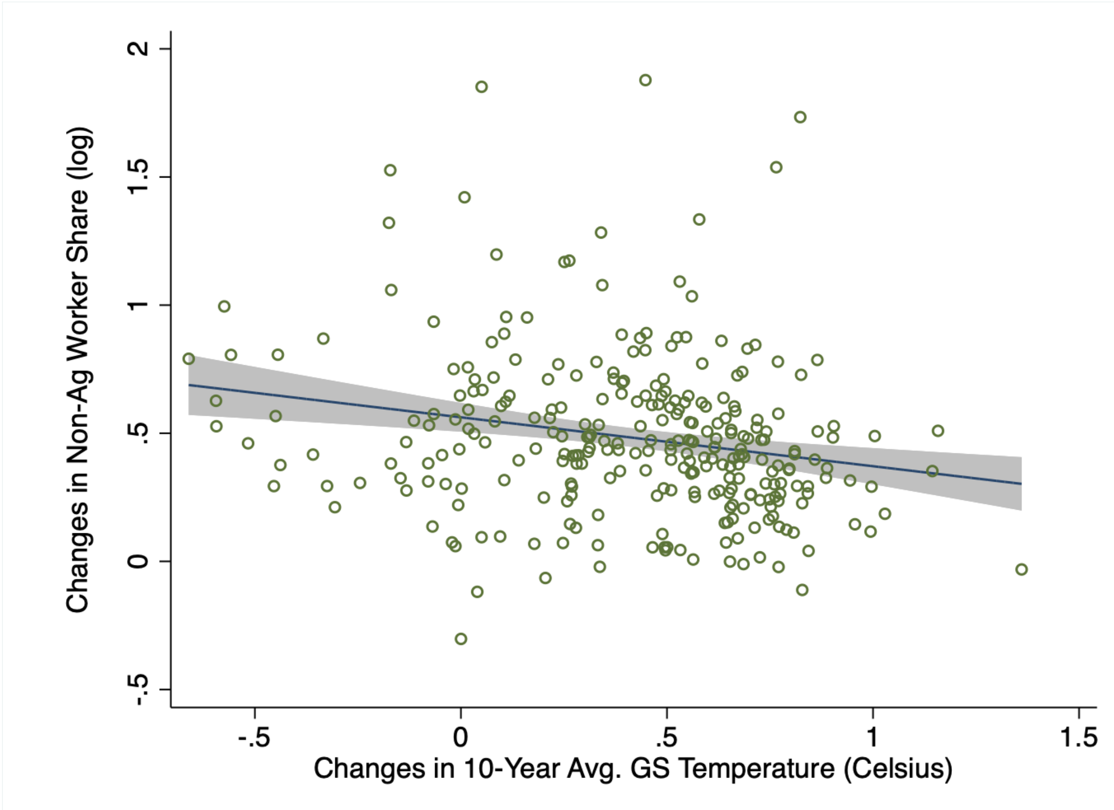


FIGURE B4. Correlation between Long-Run Changes in Temperature and Non-Agricultural Worker Share

Note: Figure plots the relationship between long-run changes in non-agricultural worker share and changes in 10-year average growing season temperature between 1961 and 2011. Data come from district-level panel data constructed from the Indian Census. Each dot represents a district in our sample. For each district, we take the difference in the ten-year average growing season temperature between 1961 and 2011, as well as the difference in the natural log of the non-agricultural worker share between 1961 and 2011, and plot them against one another. A fitted linear regression line and 95% confidence intervals are presented along with the scatter plot.

Table B1. Summary Statistics by Year

Year	1987-88	1993-94	1999-2000	2004-05	2005-06	2007-08	2009-10	2011-12	
NSS Round	43	50	55	61	62	64	66	68	Total
10-Year Avg. GS Temperature (Celsius)	23.95	23.95	23.92	24.04	24.09	24.16	24.17	24.22	24.06
	(3.311)	(3.322)	(3.352)	(3.337)	(3.347)	(3.320)	(3.303)	(3.309)	(3.322)
10-Year Avg. GS Rainfall (100 mm)	1.306	1.167	1.171	1.141	1.114	1.135	1.144	1.170	1.168
	(0.637)	(0.634)	(0.603)	(0.597)	(0.602)	(0.621)	(0.634)	(0.651)	(0.624)
Agricultural Worker Share	0.543	0.650	0.545	0.472	0.382	0.505	0.390	0.363	0.481
	(0.172)	(0.218)	(0.177)	(0.124)	(0.150)	(0.153)	(0.122)	(0.120)	(0.183)
Non-Agricultural Worker Share	0.371	0.311	0.421	0.491	0.572	0.455	0.580	0.606	0.476
	(0.161)	(0.209)	(0.163)	(0.112)	(0.146)	(0.145)	(0.117)	(0.113)	(0.179)
Manufacturing Worker Share	0.0993	0.0818	0.0998	0.119	0.144	0.106	0.114	0.121	0.111
	(0.0644)	(0.0710)	(0.0673)	(0.0634)	(0.0845)	(0.0698)	(0.0674)	(0.0661)	(0.0716)
Services Worker Share	0.220	0.192	0.273	0.307	0.357	0.264	0.347	0.336	0.287
	(0.105)	(0.145)	(0.113)	(0.0808)	(0.0992)	(0.0975)	(0.0865)	(0.0879)	(0.118)
Construction Worker Share	0.0448	0.0294	0.0419	0.0574	0.0611	0.0767	0.110	0.112	0.0667
	(0.0562)	(0.0262)	(0.0280)	(0.0289)	(0.0320)	(0.0458)	(0.0596)	(0.0506)	(0.0516)

Note: Table presents summary statistics for the weather variables and National Sample Survey outcome variables over time for the sample of districts used in regression analysis — this include districts with non-missing observations of non-agricultural shares in all years in the PCA data (N=270). The 50th round in 1993-1994 has incomplete coverage of the urban population — only a quarter of the districts have their urban households represented in the survey. Therefore, the district-level summary statistics in 1993-1994 have relatively higher shares of agricultural workers and lower shares of non-agricultural workers, compared to those in other rounds.

Table B2. Effect of Rising Temperatures on Primary and Secondary Occupations

	Agricultural Labor Share					
	Total		Main		Marginal	
Panel A: Main and Marginal Employment in Census	(1)	(2)	(3)	(4)	(5)	(6)
T	0.198	0.176	0.155	0.143	0.518	0.473
	(0.069)***	(0.073)**	(0.069)**	(0.072)**	(0.128)***	(0.136)***
	[0.088]**	[0.089]**	[0.097]	[0.092]	[0.219]**	[0.214]**
P	-0.086	-0.051	0.011	0.060	-0.088	-0.016
	(0.070)	(0.072)	(0.067)	(0.065)	(0.125)	(0.124)
	[0.103]	[0.106]	[0.101]	[0.097]	[0.243]	[0.231]
Region-year trends	Y	N	Y	N	Y	N
Region-year FE	N	Y	N	Y	N	Y
Observations	1,290	1,290	1,290	1,290	1,290	1,290

	Ag Worker Share				Non-Ag Worker Share			
	Primary Occupation		Secondary Occupation		Primary Occupation		Secondary Occupation	
Panel B: Primary and Secondary Employment in NSS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T	0.191	0.213	0.211	0.132	-0.076	-0.202	-0.335	-0.223
	(0.061)***	(0.054)***	(0.079)***	(0.080)*	(0.067)	(0.062)***	(0.166)**	(0.181)
	[0.074]***	[0.070]***	[0.082]**	[0.082]	[0.085]	[0.071]***	[0.177]*	[0.187]
P	-0.086	-0.110	0.079	0.042	-0.058	0.073	-0.341	-0.184
	(0.119)	(0.096)	(0.086)	(0.101)	(0.093)	(0.098)	(0.210)	(0.230)
	[0.137]	[0.126]	[0.095]	[0.099]	[0.125]	[0.140]	[0.221]	[0.238]
Region-year trends	Y	N	Y	N	Y	N	Y	N
State-year trends	N	Y	N	Y	N	Y	N	Y
Observations	2,120	2,120	2,072	2,072	2,120	2,120	2,072	2,072

Note: The dependent variables in Panel A are the natural logarithm of the shares of total agricultural laborers (columns 1-2), of main agricultural laborers (columns 3-4), and of marginal agricultural laborers (columns 5-6). Temperature and precipitation are decadal averages of the past ten growing seasons. The sample in Panel A comes from district-level data constructed from the Indian Census. The sample is restricted to districts for which the dependent variable is non-missing in all years, and excludes data from 1991 as counts of main and marginal agricultural workers are not available in that year. The dependent variables in Panel B are the natural logarithm of the share of workers engaged in agriculture as a primary occupation (columns 1-2), engaged in agriculture as a secondary occupation (columns 3-4), engaged in non-agriculture as a primary occupation (columns 5-6), and engaged in non-agriculture as a secondary occupation (columns 7-8). The sample in Panel B comes from district-level data aggregated from the National Sample Survey. The sample is restricted to districts for which i) the dependent variables are non-missing in all years in the NSS data, and ii) the non-agricultural share variable is non-missing across all years in the PCA data. All columns in both panels include district and year fixed effects. We present standard errors clustered by district in parentheses, and Conley standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B3. Heterogeneous Effect of Rising Temperatures by Long-Run Temperature

	Ag Labor Share		Non-Ag Worker Share		Urbanization		Migrant Share	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T x Less Hot District	0.156	0.170	-0.049	-0.071	0.036	0.050	-0.078	-0.103
	(0.088)*	(0.090)*	(0.047)	(0.044)	(0.058)	(0.060)	(0.070)	(0.074)
	[0.114]	[0.110]	[0.050]	[0.048]	[0.062]	[0.058]	[0.066]	[0.066]
T x Hot District	0.239	0.203	-0.097	-0.079	-0.047	-0.043	0.082	0.103
	(0.087)***	(0.089)**	(0.044)**	(0.042)*	(0.064)	(0.069)	(0.101)	(0.103)
	[0.088]***	[0.087]**	[0.048]**	[0.041]*	[0.060]	[0.063]	[0.102]	[0.104]
Region-year trends	Y	N	Y	N	Y	N	Y	N
Region-year FE	N	Y	N	Y	N	Y	N	Y
Observations	1,548	1,548	1,620	1,620	1,596	1,596	1,350	1,350

Note: The dependent variable is the natural logarithm of the share of agricultural laborers in Columns (1) and (2), of the share of non-agricultural workers in Columns (3) and (4), of urbanization rates in Columns (5) and (6), and of the share of intra-district migrants in Columns (7) and (8). Temperature is the decadal average of the past ten growing seasons. "Hot District" is a binary variable that takes the value 1 if a district's average growing season temperature for the period 1901-2014 is above the median for that period; "Less Hot District" takes the value 1 if a district's average temperature for 1901-2014 is below the median. Data are district-level panel data constructed from the Indian Census. We restrict our sample to districts for which the dependent variable is non-missing in all years. All columns include district fixed effects and year fixed effects and control for decadal precipitation interacted with the hot and less hot district dummies. We present standard errors clustered by district in parentheses, and Conley standard errors that allow for spatial correlation up to 500km and arbitrary serial correlation in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B4. Effect of Rising Temperatures using National Sample Survey

	Ag Worker Share		Non-Ag Worker Share	
Panel A: Panel specification	(1)	(2)	(3)	(4)
T (Decadal Average)	0.283	0.281	-0.122	-0.248
	(0.063)***	(0.059)***	(0.065)*	(0.061)***
	[0.083]***	[0.077]***	[0.087]	[0.074]***
P (Decadal Average)	-0.014	-0.023	-0.048	0.103
	(0.133)	(0.146)	(0.094)	(0.096)
	[0.154]	[0.161]	[0.122]	[0.135]
Region-year trends	Y	N	Y	N
State-year trends	N	Y	N	Y
Observations	2,120	2,120	2,120	2,120
Panel B: Short and long-term effects	(1)	(2)	(3)	(4)
Current Year T	-0.086	-0.061	0.122	0.099
	(0.025)***	(0.030)**	(0.023)***	(0.024)***
	[0.032]***	[0.037]*	[0.033]***	[0.032]***
Decadal Average T	0.285	0.270	-0.123	-0.239
	(0.063)***	(0.061)***	(0.063)*	(0.061)***
	[0.080]***	[0.075]***	[0.078]	[0.069]***
Region-year trends	Y	N	Y	N
State-year trends	N	Y	N	Y
Observations	2,120	2,120	2,120	2,120

Note: The dependent variable is the natural logarithm of the share of individuals engaged in agriculture in Columns (1) and (2), and of the share of individuals engaged in non-agricultural sectors in Columns (3) and (4). Data are district-level panel data aggregated from the National Sample Survey. In Panel A, temperature and precipitation measures are decadal averages of the past ten growing seasons, and in Panel B, they are current growing season monthly averages and decadal averages of the past ten growing seasons. All columns include district and year fixed effects. We restrict our sample to districts for which the dependent variable is non-missing in all years. In addition, we restrict our sample to districts with non-missing observations of non-agricultural shares across all years in the PCA data. We present standard errors clustered by district in parentheses, and Conley standard errors that allow for spatial correlation up to 500km and arbitrary serial correlation in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B5. Effect of Rising Temperatures on Agricultural Yields

	Agricultural Yields	
Panel A: Main Effects	(1)	(2)
T	-0.0840	-0.0606
	(0.0173)***	(0.0115)***
	[0.0205]***	[0.0203]***
P	0.0906	0.1506
	(0.0115)***	(0.0133)***
	[0.0145]***	[0.0142]***
Region-time trends	Y	N
Region-decade FE	N	Y
Observations	11,860	11,860
Panel B: By Road Density	(1)	(2)
T	-0.094	-0.088
	(0.022)***	(0.022)***
	[0.025]***	[0.024]***
T x High Road Density	0.010	0.006
	(0.025)	(0.025)
	[0.028]	[0.027]
Region-year trends	Y	N
Region-decade FE	N	Y
P-val of sum, cluster	0.0001	0.0001
P-val of sum, Conley	0.0009	0.0012
Observations	11,860	11,860
Panel C: By Bank Credit per Capita	(1)	(2)
T	-0.101	-0.096
	(0.017)***	(0.017)***
	[0.021]***	[0.021]***
T x High Bank Credit	0.032	0.031
	(0.025)	(0.024)
	[0.025]	[0.024]
Region-year trends	Y	N
Region-decade FE	N	Y
P-val of sum, cluster	0.0047	0.0084
P-val of sum, Conley	0.0084	0.0129
Observations	11,860	11,860

Note: The dependent variable is the natural logarithm of aggregate yields. Temperature and precipitation are annual averages over the growing season months. We use annual data from VDSA spanning 1966 to 2010. All columns include district and year fixed effects. Panels B and C control for precipitation and precipitation interacted with the heterogeneity measure. We present standard errors clustered by district in parentheses, and Conley standard errors that allow for spatial correlation up to 500km and arbitrary serial correlation in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B6. Effect of Rising Temperatures on Rural & Urban Non-Agricultural Worker Share

	Non-Agricultural Worker Share			
	Rural		Urban	
	(1)	(2)	(3)	(4)
T	-0.084	-0.106	-0.007	-0.016
	(0.040)**	(0.038)***	(0.011)	(0.012)
	[0.060]	[0.055]*	[0.017]	[0.016]
P	-0.015	0.014	-0.031	-0.009
	(0.042)	(0.043)	(0.014)**	(0.014)
	[0.061]	[0.055]	[0.014]**	[0.014]
Region-year trends	Y	N	Y	N
Region-year FE	N	Y	N	Y
Observations	1,608	1,608	1,596	1,596

Note: The dependent variable is the natural logarithm of the share of rural non-agricultural workers in Columns (1) and (2), and the natural logarithm of the share of urban non-agricultural workers in Columns (3) and (4). Temperature and precipitation are decadal averages of the past ten growing seasons. Data are district-level panel data constructed from the Indian Census. We restrict our sample to districts for which the dependent variable is non-missing in all years. All columns include district and year fixed effects. We present standard errors clustered by district in parentheses, and Conley standard errors that allow for spatial correlation up to 500km and arbitrary serial correlation in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B7. Effect of Rising Temperatures on Cultivator & Agricultural Worker Share

Panel A: Cultivator Share	(1)	(2)	(3)	(4)	(5)	(6)
T	0.090	0.116	0.036	0.075	0.077	0.105
	(0.071)	(0.074)	(0.055)	(0.057)	(0.039)**	(0.039)***
	[0.060]	[0.059]**	[0.053]	[0.051]	[0.043]*	[0.037]***
T x High Road Density			0.006	-0.013		
			(0.074)	(0.074)		
			[0.066]	[0.063]		
T x High Bank Credit					-0.008	-0.018
					(0.113)	(0.114)
					[0.099]	[0.100]
P-val of sum, cluster			0.445	0.264	0.546	0.444
P-val of sum, Conley			0.438	0.226	0.451	0.343
Observations	1,620	1,620	1,620	1,620	1,620	1,620
Panel B: Ag Labor + Cultivator Share						
T	0.072	0.093	0.083	0.121	0.115	0.141
	(0.036)**	(0.038)**	(0.045)*	(0.048)**	(0.039)***	(0.039)***
	[0.043]*	[0.045]**	[0.038]**	[0.038]***	[0.049]**	[0.049]***
T x High Road Density			-0.037	-0.066		
			(0.061)	(0.062)		
			[0.048]	[0.048]		
T x High Bank Credit					-0.094	-0.103
					(0.062)	(0.060)*
					[0.062]	[0.058]*
Region-year trends	Y	N	Y	N	Y	N
Region-year FE	N	Y	N	Y	N	Y
P-val of sum, cluster			0.303	0.231	0.690	0.481
P-val of sum, Conley			0.271	0.170	0.703	0.498
Observations	1,548	1,548	1,548	1,548	1,548	1,548

Note: The dependent variable is the natural logarithm of the share of cultivators in Panel A, and of the share of agricultural workers (agricultural laborers and cultivators) in Panel B. Temperature is the decadal average of the past ten growing seasons. All columns include district and year fixed effects. We restrict our sample to districts for which the dependent variable is non-missing in all years. *High Road Density* is a binary variable that takes the value 1 if the district has above median road density at baseline. *High Bank Credit* is a binary variable that takes the value 1 if the district has above median bank credit per capita at baseline. Data are district-level panel data constructed from Indian Census. Columns (1) and (2) control for decadal precipitation. Columns (3) and (4) control for decadal precipitation interacted with the high road density dummy and include high road density-by-year fixed effects. Columns (5) and (6) control for decadal precipitation interacted with the high bank credit dummy and include high bank credit-by-year fixed effects. We present standard errors clustered by district in parentheses, and Conley standard errors that allow for spatial correlation up to 500km and arbitrary serial correlation in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B8. Effect of Rising Temperatures on Crop Area Shares

	Dry Season Crop Area		Labor-Intensive Crop Area	
	(1)	(2)	(3)	(4)
T	-0.0038	-0.0040	-0.0094	-0.0079
	(0.0027)	(0.0017)**	(0.0030)***	(0.0019)***
	[0.0035]	[0.0035]	[0.0041]**	[0.0041]***
P	0.0078	0.0074	0.0104	0.0061
	(0.0021)***	(0.0019)***	(0.0027)***	(0.0020)***
	[0.0028]***	[0.0028]**	[0.0041]**	[0.0042]**
Region-year trends	Y	N	Y	N
Region-decade FE	N	Y	N	Y
Observations	11,705	11,705	11,705	11,705

Note: The dependent variable is the share of crop area planted with dry season crops in Columns (1) and (2), and the share of crop area planted with labor-intensive crops in Columns (3) and (4). Temperature and precipitation are decadal averages of the past ten growing seasons. We use annual data from VDSA spanning 1966 to 2010. The dry season (*rabi*) crops are wheat, pearl millet, barley, chickpea, pigeon pea, rapeseed and mustard seed, linseed, and sunflower. The labor-intensive crops are defined to be those that require 700 or more average person-hours per hectare, which are rice, groundnut, cotton, and sugarcane (FICCI, 2015). All columns include district and year fixed effects. We present standard errors clustered by district in parentheses, and Conley standard errors that allow for spatial correlation up to 500km and arbitrary serial correlation in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B9. Robustness to Alternate Samples and Variable Definitions

	Ag Labor Share			Non-Ag Worker Share			Urbanization			Migrant Share		
Panel A: Unbalanced Panel	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
T	0.206**	0.337***	0.160*	-0.069*	-0.140**	-0.135**	0.017	-0.185**	-0.006	0.006	-0.152**	0.012
	(0.094)	(0.097)	(0.095)	(0.037)	(0.056)	(0.057)	(0.045)	(0.082)	(0.068)	(0.061)	(0.077)	(0.066)
T x High Road Density		-0.375***			0.125*			0.229**			0.223**	
		(0.114)			(0.066)			(0.095)			(0.107)	
T x High Bank Credit			0.046			0.095			-0.006			-0.034
			(0.164)			(0.062)			(0.082)			(0.095)
P-val: T + T x High Road D.		0.601			0.706			0.458			0.380	
P-val: T + T x High Bank C.			0.146			0.306			0.820			0.790
Observations	1691	1498	1691	1705	1498	1705	1689	1494	1689	1411	1242	1411
Panel B: Grid Point Average T & P												
T	0.138**	0.310***	0.271***	-0.0723**	-0.118**	-0.151***	0.0229	-0.185**	-0.00590	0.00941	-0.133*	0.0286
	(0.0601)	(0.0903)	(0.0671)	(0.0312)	(0.0572)	(0.0488)	(0.0447)	(0.0831)	(0.0687)	(0.0640)	(0.0743)	(0.0672)
T x High Road Density		-0.314***			0.119*			0.248**			0.238**	
		(0.108)			(0.0682)			(0.0992)			(0.108)	
T x High Bank Credit			-0.258**			0.123**			0.00858			-0.0505
			(0.106)			(0.0578)			(0.0863)			(0.101)
P-val: T + T x High Road D.		0.954			0.975			0.319			0.225	
P-val: T + T x High Bank C.			0.882			0.446			0.962			0.810
Observations	1524	1434	1524	1578	1434	1578	1560	1428	1560	1325	1195	1325
Panel C: Log T & P												
ln T	2.535***	3.320***	2.587***	-1.803***	-1.902***	-2.285***	-0.732	-2.187***	-1.122*	-0.915	-1.517***	-0.615
	(0.642)	(1.174)	(0.707)	(0.338)	(0.358)	(0.470)	(0.755)	(0.499)	(0.590)	(0.572)	(0.443)	(0.444)
ln T x High Road Density		-4.252**			1.731*			3.428**			3.450*	
		(2.008)			(1.035)			(1.542)			(1.942)	
ln T x High Bank Credit			-1.709			1.398			0.899			-0.948
			(2.249)			(0.997)			(1.436)			(2.062)
P-val: ln T + ln T x High Road D.		0.581			0.859			0.403			0.313	
P-val: ln T + ln T x High Bank C.			0.686			0.325			0.865			0.453
Observations	1542	1458	1542	1614	1458	1614	1596	1452	1596	1345	1210	1345

Note: The dependent variable is the natural logarithm of the share of agricultural laborers in Columns (1) and (2), of the share of non-agricultural workers in Columns (3) and (4), of urbanization rates in Columns (5) and (6), and of the share of intra-district migrants in Columns (7) and (8). Temperature is the decadal average of the past ten growing seasons. *High Road Density* is a binary variable that takes the value 1 if the district has above median road density at baseline. Data are district-level panel data constructed from the Indian Census. The samples in Panels B and C are restricted to districts for which the dependent variable is non-missing in all years. All columns control for decadal precipitation and include district and region-by-year fixed effects. Columns (2), (5), (8), and (11) also control for decadal precipitation interacted with the high road density dummy and include high road density-by-year fixed effects. *High Bank Credit* is a binary variable that takes the value 1 if the district has above median bank credit per capita at baseline. Columns (3), (6), (9), and (12) control for decadal precipitation interacted with the high bank credit dummy and include high bank credit-by-year fixed effects. We present standard errors clustered by district in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B10. Effect of Rising Temperatures: Distributed Lagged Averages

	Ag Labor Share		Non-Ag Worker Share		Urbanization		Migrant Share	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average (T_0 to T_{-2})	0.023	0.068	-0.048	-0.067	0.064	0.061	0.029	-0.036
	(0.043)	(0.044)	(0.024)**	(0.026)***	(0.028)**	(0.032)*	(0.054)	(0.063)
	[0.076]	[0.074]	[0.034]	[0.033]**	[0.033]**	[0.036]*	[0.058]	[0.059]
Average (T_{-3} to T_{-6})	0.088	0.094	0.009	-0.024	-0.058	-0.055	-0.035	0.034
	(0.031)***	(0.037)**	(0.015)	(0.018)	(0.022)***	(0.029)*	(0.042)	(0.052)
	[0.053]*	[0.056]*	[0.022]	[0.022]	[0.028]**	[0.033]*	[0.052]	[0.057]
Average (T_{-7} to T_{-9})	0.076	-0.020	-0.054	0.006	-0.019	0.005	0.000	-0.022
	(0.035)**	(0.044)	(0.023)**	(0.025)	(0.023)	(0.028)	(0.039)	(0.047)
	[0.060]	[0.067]	[0.028]*	[0.030]	[0.027]	[0.031]	[0.051]	[0.057]
Region-year trends	Y	N	Y	N	Y	N	Y	N
Region-year FE	N	Y	N	Y	N	Y	N	Y
Observations	1,548	1,548	1,620	1,620	1,596	1,596	1,350	1,350

Note: The dependent variable is the natural logarithm of the share of agricultural laborers in Columns (1) and (2), of the share of non-agricultural workers in Columns (3) and (4), of urbanization rates in Columns (5) and (6), and of the share of intra-district migrants in Columns (7) and (8). Data are district-level panel data constructed from the Indian Census. We restrict our sample to districts for which the dependent variable is non-missing in all years. All columns include district and year fixed effects. All regressions control for the corresponding distributed lagged averages for precipitation. We present standard errors clustered by district in parentheses, and Conley standard errors that allow for spatial correlation up to 500km and arbitrary serial correlation in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B11. Robustness to Controlling for Time-Varying Covariates

	Ag Labor Share		Non-Ag Worker Share		Urbanization		Migrant Share	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T	0.159	0.129	-0.054	-0.068	-0.007	0.013	0.014	0.015
	(0.056)***	(0.058)**	(0.032)*	(0.031)**	(0.044)	(0.047)	(0.056)	(0.062)
	[0.077]**	[0.074]*	[0.040]	[0.034]**	[0.049]	[0.049]	[0.061]	[0.068]
P	-0.125	-0.053	-0.026	-0.000	0.044	0.020	-0.038	-0.004
	(0.054)**	(0.053)	(0.030)	(0.031)	(0.042)	(0.042)	(0.045)	(0.050)
	[0.096]	[0.087]	[0.038]	[0.033]	[0.048]	[0.046]	[0.061]	[0.064]
High-yielding-variety area	0.155	0.147	0.087	0.106	0.074	0.071	-0.214	-0.211
	(0.066)**	(0.067)**	(0.048)*	(0.049)**	(0.059)	(0.062)	(0.058)***	(0.062)***
	[0.077]**	[0.077]*	[0.056]	[0.055]*	[0.054]	[0.055]	[0.061]***	[0.064]***
Labor regulation strictness index	-0.077	-0.071	0.002	0.008	0.003	-0.001	0.044	0.045
	(0.021)***	(0.023)***	(0.012)	(0.013)	(0.013)	(0.015)	(0.016)***	(0.018)**
	[0.030]**	[0.033]**	[0.015]	[0.014]	[0.015]	[0.016]	[0.021]**	[0.022]**
Road density	-0.126	-0.143	0.022	0.021	0.030	0.044	0.027	-0.005
	(0.042)***	(0.031)***	(0.012)*	(0.014)	(0.042)	(0.046)	(0.109)	(0.122)
	[0.053]**	[0.050]***	[0.018]	[0.018]	[0.039]	[0.041]	[0.091]	[0.099]
Number of markets	0.003	0.002	-0.002	-0.002	0.000	0.000	-0.003	-0.002
	(0.006)	(0.005)	(0.001)*	(0.001)	(0.004)	(0.003)	(0.008)	(0.008)
	[0.004]	[0.004]	[0.001]*	[0.001]	[0.003]	[0.002]	[0.006]	[0.006]
Number of banks	-0.049	-0.053	-0.001	-0.002	-0.008	-0.005	-0.007	-0.006
	(0.012)***	(0.013)***	(0.006)	(0.006)	(0.008)	(0.008)	(0.013)	(0.013)
	[0.011]***	[0.011]***	[0.005]	[0.005]	[0.008]	[0.008]	[0.012]	[0.012]
Region-year trends	Y	N	Y	N	Y	N	Y	N
Region-year FE	N	Y	N	Y	N	Y	N	Y
Observations	1,546	1,546	1,546	1,546	1,529	1,529	1,283	1,283

Note: The dependent variable is the natural logarithm of the share of agricultural laborers in Columns (1) and (2), of the share of non-agricultural workers in Columns (3) and (4), of urbanization rates in Columns (5) and (6), and of the share of intra-district migrants in Columns (7) and (8). Temperature and precipitation are decadal averages of the past ten growing seasons. Data are district-level panel data constructed from the Indian Census. We restrict our sample to districts for which the dependent variable is non-missing in all years. All columns include district and year fixed effects. All time-varying covariates are decadal averages. District-level high-yielding variety area, road density, and number of markets are from the VDSA data set. The state-level labor market strictness index, from Besley and Burgess (2004), ranges from 3 to -3; positive values denote states that are more rigid (pro-worker); negative values denote states that are more flexible (pro-employer). The number of banks per district is measured in 100's and is from Fulford (2013). See Appendix C for more details. We present standard errors clustered by district in parentheses, and Conley standard errors that allow for spatial correlation up to 500km and arbitrary serial correlation in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE B12. Effect of Rising Temperatures using Long-Difference Specification with Alternative End Points

	Ag Labor Share	Non-Ag Worker Share	Urbanization	Migrant Share
	(1)	(2)	(3)	(4)
T	0.2768	-0.1376	0.0177	-0.3828
	(0.1011)***	(0.0551)**	(0.0876)	(0.1593)**
	[0.2243]	[0.0587]**	[0.1298]	[0.1878]**
P	-0.9920	0.1561	-0.3520	0.6279
	(0.3688)***	(0.1987)	(0.2972)	(0.5536)
	[0.4429]**	[0.1828]	[0.2864]	[0.8081]
Region FE	Y	Y	Y	Y
Observations	258	270	266	264

Note: The dependent variable is the share of agricultural laborers in Column (1), the share of non-agricultural workers in Column (2), urbanization rates in Column (3), and the share of intra-district migrants in Column (4). The dependent variable in each column is the difference (in natural logarithm) of an outcome between two 20-year periods, 1961-1971 and 2001-2011. The outcome in 1961-1971 are calculated as the average of 1961 and 1971 decadal observations, and that in 2001-2011 are calculated as the average of 2001 and 2011. The independent variables are differences in average growing-season temperature and precipitation over the same periods. Data are district-level data constructed from the Indian Census. We present standard errors in parentheses, and Conley standard errors that allow for spatial correlation up to 500km and arbitrary serial correlation in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B13. Heterogeneous Effect of Rising Temperatures with Alternate Thresholds

	Ag Labor Share		Non-Ag Worker Share		Urbanization		Migrant Share	
Panel A: Road Network Density	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T	0.329***	0.291***	-0.154**	-0.156***	-0.210**	-0.159**	-0.233***	-0.063
	(0.112)	(0.099)	(0.067)	(0.052)	(0.092)	(0.080)	(0.087)	(0.095)
T x (Road Density > 40th pct)	-0.301**		0.113		0.229**		0.305***	
	(0.121)		(0.075)		(0.105)		(0.107)	
T x (Road Density > 60th pct)		-0.290**		0.155**		0.191*		0.112
		(0.118)		(0.066)		(0.097)		(0.119)
P-val: T + T x High Road D.	0.677	0.996	0.266	0.994	0.739	0.583	0.324	0.458
Observations	1458	1458	1458	1458	1452	1452	1210	1210
Panel B: Bank Credit Per Capita	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T	0.298***	0.212***	-0.181***	-0.156***	-0.037	-0.079	-0.045	-0.008
	(0.077)	(0.064)	(0.051)	(0.041)	(0.077)	(0.062)	(0.074)	(0.072)
T x (Bank Credit > 40th pct)	-0.224**		0.143**		0.036		0.026	
	(0.106)		(0.059)		(0.091)		(0.098)	
T x (Bank Credit > 60th pct)		-0.149		0.126**		0.118		-0.026
		(0.119)		(0.056)		(0.084)		(0.112)
P-val: T + T x High Bank C.	0.373	0.552	0.308	0.484	0.988	0.533	0.812	0.728
Observations	1548	1548	1620	1620	1596	1596	1350	1350

Note: The dependent variable is the natural logarithm of the share of agricultural laborers in Columns (1) and (2), of the share of non-agricultural workers in Columns (3) and (4), of urbanization rates in Columns (5) and (6), and of the share of intra-district migrants in Columns (7) and (8). Temperature is the decadal average of the past ten growing seasons. In Panel A, we use alternate heterogeneity dummies which take the value of 1 if the district's baseline road density is above the 40th or 60th percentile, depending on the column. Data are district-level panel data constructed from the Indian Census. We restrict our sample to districts for which road density/bank credit data is non-missing and the dependent variable is non-missing in all years. All columns include district fixed effects, region-by-year fixed effects, and high road density-by-year fixed effects. We control for decadal precipitation and decadal precipitation interacted with the road density dummy. In Panel B, we use alternate heterogeneity dummies which take the value of 1 if the district's baseline bank credit per capita is above the 40th or 60th percentile, depending on the column. All columns include district fixed effects, region-by-year fixed effects, and high bank credit-by-year fixed effects. We control for decadal precipitation and decadal precipitation interacted with the high bank credit dummy. We present standard errors clustered by district in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B14. Heterogeneity by Road Density: Interacting Other Baseline Characteristics with Temperature

Panel A: Ag Labor Share	(1)	(2)	(3)	(4)	(5)
T	0.355	0.557	0.420	0.389	0.352
	(0.099)***	(0.103)***	(0.116)***	(0.134)***	(0.153)**
	[0.118]***	[0.133]***	[0.141]***	[0.166]**	[0.179]**
T x High Road Density	-0.362	-0.387	-0.345	-0.353	-0.391
	(0.114)***	(0.116)***	(0.115)***	(0.113)***	(0.123)***
	[0.128]***	[0.128]***	[0.127]***	[0.124]***	[0.137]***
Controls	None	I	II	III	IV
Observations	1,458	1,458	1,428	1,428	1,344
Panel B: Non-Ag Worker Share					
T	-0.137	-0.115	-0.069	-0.007	-0.048
	(0.058)**	(0.063)*	(0.060)	(0.062)	(0.070)
	[0.052]***	[0.063]*	[0.063]	[0.066]	[0.066]
T x High Road Density	0.119	0.117	0.083	0.100	0.127
	(0.068)*	(0.067)*	(0.062)	(0.061)	(0.064)**
	[0.062]*	[0.060]*	[0.059]	[0.057]*	[0.058]**
Controls	None	I	II	III	IV
Observations	1,458	1,458	1,428	1,428	1,344
Panel C: Urbanization					
T	-0.190	-0.121	-0.066	-0.037	-0.225
	(0.083)**	(0.090)	(0.097)	(0.114)	(0.125)*
	[0.068]***	[0.075]	[0.084]	[0.100]	[0.107]**
T x High Road Density	0.231	0.224	0.185	0.191	0.208
	(0.098)**	(0.098)**	(0.097)*	(0.095)**	(0.089)**
	[0.084]***	[0.084]***	[0.086]**	[0.085]**	[0.074]***
Controls	None	I	II	III	IV
Observations	1,452	1,452	1,422	1,422	1,344
Panel D: Migrant Share					
T	-0.149	-0.135	-0.195	-0.243	-0.348
	(0.078)*	(0.086)	(0.110)*	(0.127)*	(0.161)**
	[0.075]**	[0.097]	[0.122]	[0.130]*	[0.159]**
T x High Road Density	0.227	0.222	0.221	0.208	0.168
	(0.108)**	(0.110)**	(0.116)*	(0.114)*	(0.120)
	[0.100]**	[0.102]**	[0.109]**	[0.109]*	[0.111]
Controls	None	I	II	III	IV
Observations	1,210	1,210	1,190	1,190	1,120

Note: The dependent variable is the natural logarithm of the share of agricultural laborers in Panel A, of the share of non-agricultural workers in Panel B, of urbanization rates in Panel C, and of the share of intra-district migrants in Panel D. Temperature is the decadal average of the past ten growing seasons. *High Road Density* is a binary variable that takes the value 1 if the district has above median road density at baseline. All columns include district, region-by-year and high road density-by-year fixed effects. We control for decadal precipitation and decadal precipitation interacted with the road density dummy. Data are district-level panel data constructed from the Indian Census. We restrict our sample to districts for which road density data is non-missing and the dependent variable is non-missing in all years. In Columns (2) through (5) we cumulatively add other dummy controls interacted with decadal temperature and with decadal precipitation, to test the stability of our road density heterogeneity coefficient. These controls are I: above median bank credit per capita at baseline; II: I and above median male agricultural wages at baseline; III: II and above median long-run temperature; IV: III and above median proportion irrigated land at baseline. We present standard errors clustered by district in parentheses, and Conley standard errors that allow for spatial correlation up to 500km and arbitrary serial correlation in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B15. Heterogeneity by Bank Credit: Interacting Other Baseline Characteristics with Temperature

Panel A: Ag Labor Share	(1)	(2)	(3)	(4)	(5)
T	0.271	0.562	0.402	0.354	0.285
	(0.070)***	(0.097)***	(0.111)***	(0.134)***	(0.155)*
	[0.082]***	[0.118]***	[0.125]***	[0.149]**	[0.169]*
T x High Bank Credit	-0.217	-0.219	-0.206	-0.208	-0.243
	(0.110)**	(0.106)**	(0.098)**	(0.097)**	(0.114)**
	[0.111]*	[0.112]*	[0.109]*	[0.109]*	[0.123]**
Controls	None	I	II	III	IV
Observations	1,548	1,458	1,428	1,428	1,344
Panel B: Non-Ag Worker Share					
T	-0.157	-0.194	-0.144	-0.079	-0.146
	(0.048)***	(0.066)***	(0.064)**	(0.068)	(0.078)*
	[0.062]**	[0.070]***	[0.061]**	[0.063]	[0.070]**
T x High Bank Credit	0.106	0.104	0.058	0.061	0.084
	(0.059)*	(0.066)	(0.064)	(0.062)	(0.067)
	[0.066]	[0.075]	[0.068]	[0.067]	[0.074]
Controls	None	I	II	III	IV
Observations	1,620	1,458	1,428	1,428	1,344
Panel C: Urbanization					
T	-0.042	-0.190	-0.146	-0.127	-0.356
	(0.069)	(0.080)**	(0.096)	(0.115)	(0.127)***
	[0.068]	[0.076]**	[0.083]*	[0.098]	[0.108]***
T x High Bank Credit	0.030	0.099	0.070	0.079	0.092
	(0.087)	(0.090)	(0.090)	(0.091)	(0.079)
	[0.078]	[0.095]	[0.088]	[0.088]	[0.079]
Controls	None	I	II	III	IV
Observations	1,596	1,452	1,422	1,422	1,344
Panel D: Migrant Share					
T	-0.013	-0.035	-0.105	-0.160	-0.250
	(0.069)	(0.095)	(0.120)	(0.143)	(0.176)
	[0.084]	[0.117]	[0.134]	[0.145]	[0.166]
T x High Bank Credit	-0.029	-0.094	-0.134	-0.134	-0.192
	(0.099)	(0.109)	(0.105)	(0.105)	(0.100)*
	[0.114]	[0.124]	[0.124]	[0.124]	[0.120]
Controls	None	I	II	III	IV
Observations	1,350	1,210	1,190	1,190	1,120

Note: The dependent variable is the natural logarithm of the share of agricultural laborers in Panel A, of the share of non-agricultural workers in Panel B, of urbanization rates in Panel C, and of the share of intra-district migrants in Panel D. Temperature is the decadal average of the past ten growing seasons. *High Bank Credit* is a binary variable that takes the value 1 if the district has above median road density at baseline. All columns include district, region-by-year and high bank credit-by-year fixed effects. We control for decadal precipitation and decadal precipitation interacted with the bank credit dummy. Data are district-level panel data constructed from the Indian Census. We restrict our sample to districts for which bank credit data is non-missing and the dependent variable is non-missing in all years. In Columns (2) through (5) we cumulatively add other dummy controls interacted with decadal temperature and with decadal precipitation, to test the stability of our bank credit heterogeneity coefficient. These controls are I: above median road density at baseline; II: I and above median male agricultural wages at baseline; III: II and above median long-run temperature; IV: III and above median proportion irrigated land at baseline. We present standard errors clustered by district in parentheses, and Conley standard errors that allow for spatial correlation up to 500km and arbitrary serial correlation in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix C. Data

Section II summarizes the data used in the analysis. In this data appendix, we provide additional information on the various data sources as well as the construction of key variables.

C.1. **Census Data.** The Primary Census Abstract (PCA) and the Migration (D-series) data tables in the Indian Population Census are the sources of decadal district-level data on demographic and economic indicators (population, worker counts by categories, migrant counts) from 1961 to 2011. For the years 1961-1991, we use data from Vanneman and Barnes (2000), and for the years 2001 and 2011, we use data from the Census website.[4]

The Census classifies workers into four categories: cultivators, agricultural laborers, workers in household industry and other workers. A cultivator is defined as a worker who is "engaged in cultivation of land owned or held from Government or held from private persons or institutions for payments in money, kind or share." An agricultural laborer is defined as a worker who "works on another person's land for wages in money or kind or share. She or he has no risk in the cultivation, but merely works in another person's land for wages." Household industry is defined as "an industry conducted by one or more members of the household at home or within the village in rural areas and only within the precincts of the house where the household lives in urban areas." Other worker is defined as a "worker other than cultivator, agricultural laborer or worker in household industry." Examples of other workers include work in the public sector, manufacturing, construction, trade, business etc.[5]

We construct agricultural labor share as the total count of agricultural laborers divided by total workers. We construct non-agricultural worker share as the sum of workers across two categories – household industry workers and other workers – divided by total workers. Note that the agricultural labor share is not a perfect complement to the non-agricultural worker share as we do not include cultivators when constructing the agricultural labor share. We construct cultivator share as the total count of main cultivators divided by total workers. Note that our measure of cultivator share excludes marginal cultivators (those who "worked for less than six months in the reference period") because we do not have this data for 1961-1991.

The Census further splits agricultural laborers into two categories: main and marginal agricultural laborers. A main agricultural laborer is defined as a worker who "worked for more than six months in the reference period." A marginal agricultural laborer is defined as a worker who "worked for less than six months in the reference period." We construct main

[4]`https://censusindia.gov.in/pca/`

[5]Details on these categories can be found in the codebook of Vanneman and Barnes (2000) available at `http://vanneman.umd.edu/districts/codebook/laborforce.html`, as well as in the 2011 Census meta data available at `https://www.censusindia.gov.in/2011census/HLO/Metadata\_Census\_2011.pdf`.

(marginal) agricultural labor share as the main (marginal) count of agricultural laborers divided by total workers.

Turning to urbanization and migration, we construct urbanization share as the total urban population divided by total population. The Census definition of urban areas has stayed largely consistent since the 1961 census. Urban areas constitutes (a) Statutory Towns: all places with a municipality, corporation, cantonment board or notified town area committee, etc., (b) Census Towns: all places which satisfied the following criteria: i) A minimum population of 5,000; ii) At least 75 per cent of the male main working population engaged in non-agricultural pursuits; and iii) A density of population of at least 400 persons per sq. km., and (c) Adjoining Outgrowths: a viable unit such as a village or a hamlet (part of a village) contiguous to a town and posses urban features in terms of infrastructure and amenities.

For migration, the Census classifies an individual as an intra-district migrant "if the place in which he is enumerated during the census is other than his place of immediate last residence," and if the last residence is within the same district of his/her current residence. We construct migrant share as the total count of intra-district rural-to-urban male migrants divided by total male population. We consider male migration only as a majority of female migration in India is for marriage, which is outside the scope of our study. We do not have this measure for 1971 due to missing data on migrant counts in Vanneman and Barnes (2000).

C.2. **National Sample Survey Data.** The Consumer Expenditure (Schedule 1) and Employment and Unemployment (Schedule 10) modules of the National Sample Survey, collected by the National Sample Survey Office (NSSO), are the sources of nationally representative data on demographic and social characteristics, consumption patterns as well as labor market behavior at the individual- and household-level in India.

We use eight rounds of the Employment and Unemployment schedule, spanning the years 1987 to 2012.[6] The time period covered in each round corresponds to the agricultural year from July to the following June. More specifically, the data covers the following time periods (with round number reported in parentheses): July 1987- June 1988 (43rd), July 1993 - June 1994 (50th), July 1999 - June 2000 (55th), July 2004 - June 2005 (61st), July 2005 - June 2006 (62nd), July 2007 - June 2008 (64th), July 2009 - June 2010 (66th), July 2011 - June 2012 (68th).

We restrict the sample to include individuals aged 14 – 65 who participate in the labor force. We use a series of questions regarding individual-level employment activities — this includes employment status and industry of the main activity — during a seven-day reference period. We use industry information from the reference week to classify individuals

[6]They can be downloaded from http://microdata.gov.in/nada43/index.php/catalog/EUE.

as working in agricultural and non-agricultural sectors.[7] We aggregate individual-level data to construct the following employment shares at the district level: the share of the labor force who are engaged in agriculture, the share of the labor force who are engaged in non-agriculture, and the share of the labor force who are engaged in manufacturing, services, and construction. We also use information on each individual's principal and subsidiary employment during the year to construct the following employment shares at the district level: shares of the labor force engaged in agriculture as a primary and secondary occupation, and shares of the labor force engaged in non-agriculture as a primary and secondary occupation.

There is one caveat with the NSS sample described above. The 50th round in 1993-1994 has incomplete coverage of the urban population — most of the districts have their rural households represented, but only a quarter of the districts have their urban households represented in this survey round. We present results using the complete set of NSS rounds in the paper; these results are robust to the exclusion of the 50th round.[8]

In addition, we use six rounds of the Consumption schedule, spanning the years 1993 to 2012.[9] More specifically, the data covers the following time periods (with round number reported in parentheses): July 1993 - June 1994 (50th), July 1999 - June 2000 (55th), July 2004 - June 2005 (61st), July 2007 - June 2008 (64th), July 2009 - June 2010 (66th), July 2011 - June 2012 (68th). We aggregate household-level data to construct the following annual consumption per capita measures at the district level: total consumption, food consumption, and non-food consumption. The consumption variables are adjusted to 2005 base prices using separate purchasing-power-parity conversion rates for urban and rural areas at the state level — this teases out temporal and spatial price level differences such that our consumption variables are comparable across all states, across urban and rural areas, and over all time periods.

C.3. **Weather Data.** The Terrestrial Precipitation: Monthly Time Series (1900–2014), version 4.01, and the companion Terrestrial Air Temperature data set (Matsuura and Willmott, 2015a,b) is the source of gridded monthly-level data on temperature and precipitation from 1951-2011.

We first construct district-level weather data by taking the weighted average of all grid points within 100 kilometers of each district's centroid, using weights that are the inverse of the squared distance between the grid point and the district centroid. This inverse distance weighting method is also used in Burgess et al. (2017) and Taraz (2018). We calculate average temperature and precipitation during the main agricultural growing season (June

[7]The agricultural sector includes sub-sectors such as crop and animal production, hunting and related service activities, forestry and logging, and fishing and aquaculture. The non-agricultural sectors include mining and quarrying, manufacturing, construction, and services.
[8]These results are available upon request.
[9]They can be downloaded from http://microdata.gov.in/nada43/index.php/catalog/CEXP.

through February) as these have the greatest impacts on agriculture. Next, we aggregate the growing season weather variables to ten-year averages.

C.4. **Infrastructure and Yields Data.** The Village Dynamics in South Asia (VDSA) Meso dataset, compiled by researchers at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT, 2015), consists of a large set of demographic, socioeconomic and agro-ecological variables at the district level. It covers 19 major agricultural states in India at an annual frequency from 1966 to 2010.[10]

The VDSA Meso dataset is the source of district-level data on total length of roads in kilometers. The underlying sources of the VDSA roads data are the annual State Statistical Abstracts. We construct a baseline road infrastructure density measure as the total length of roads in kilometers in each district in 1970 – the earliest year for which this data is available – divided by the total surface area, computed in ArcGIS using the consistent district boundaries illustrated in Appendix Figure B1A. Data on length of roads is missing for 15 districts in 1970. Furthermore, we are unable to construct a road density measure for 22 additional districts since coverage in the VDSA Meso dataset is limited to nineteen states in India.

The VDSA Meso dataset is also the source of annual district-level data on crop yields. The underlying sources of the VDSA data on yields are state-level agricultural agencies such as the Directorate of Agriculture and the Directorate of Agriculture and Food Production. We construct an annual yield measure that aggregates yields across all the crops in VDSA that have non-missing price data, using 1966-1970 crop prices as weights. The crops included are rice, wheat, sugarcane, cotton, groundnut, sorghum, maize, pearl millet, finger millet, barley, chickpeas, pigeon pea, sesame, rapeseed and mustard, castor, and linseed.

In one of our robustness tests, we use data on high yielding variety areas from VDSA. Specifically, we control for the fraction of area that is grown with high yield varieties, as a fraction of the total cultivated area in that district in that year.

C.5. **Bank Credit Data.** The Basic Statistical Returns (BSR) reports, collected by the Reserve Bank of India, is the source of district-level bank credit.[11] The Basic Statistical Returns System was launched in 1971 with the goal of creating a database of scheduled commercial banks. We use the 1972 BSR reports (the earliest available), and digitize Table 2.2, which contains district-wise statistics on the number of functioning offices of scheduled commercial banks, aggregate deposits, and total credit (advances) from all offices as of the last Friday in December 1972. The coverage of data in the 1972 BSR report is 98.7% of

[10]The states covered in the data base are Andhra Pradesh, Assam, Bihar, Chhattisgarh, Gujarat, Haryana, Himachal Pradesh, Jharkhand, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh, Uttarakhand, and West Bengal.

[11]These reports are available at `https://dbie.rbi.org.in/DBIE/dbie.rbi?site=publications`.

aggregate deposits and 98.6% of total credit. We construct a baseline bank credit per capita measure as the total bank credit in a district divided by its total population in 1971.

C.6. **Other Data.** We draw on two other data sources in our robustness tests. First, we use state-level data on labor regulation strictness from Besley and Burgess (2004). The index ranges from 3 to -3; positive values denote states that are more rigid (pro-worker); negative values denote states that are more flexible (pro-employer). It is based on a tabulation of state-level amendments to the Industrial Disputes Act of 1947, which regulates trade unions, arbitration, and procedures to be followed in the case of an industrial disputes. Because different states passed different amendments to this Act at different points in time, the index from Besley and Burgess (2004) provides a measure of labor regulation with both spatial and temporal variation.

Second, we use data on the number of banks in each district in each year, as a proxy for financial development of the district. The data on the number of banks is based on bank opening data from the Reserve Bank of India, as compiled by Fulford (2013).