

Reproducibility Instructions

Seeds of Change: The Effects of the Introduction of *Bt* Cotton in India

November 2025

1. Overview

This document describes the steps and data needed to produce the figures and tables in “Seeds of Change: The Effects of the Introduction of Bt Cotton in India”.

The analysis was performed using Stata 18.0 MP on a Windows PC (64.0 GB RAM). The codes for data cleaning and analysis take less than 10 minutes to run.

In addition, ArcGIS Pro 3.5.3 was used to aggregate districts and to obtain the geospatial characteristics of these aggregated districts.

2. Statement about Rights

I certify that the author of the manuscript has legitimate access to and permission to use the data used in this manuscript.

3. Steps to Produce Figures and Tables

All the required data have been provided in the reproducibility package. Reproducing the figures and tables involves the following steps:

- Modify line 7 in “do/0_main.do” to reflect the working directory.
- Run “do/0_main.do”. This main .do file calls all the individual .do files needed for the analysis.

4. Folder Structure

The structure of the folder is as follows:

- do
- raw
 - Banks
 - DLD
 - Apportioned
 - Unapportioned
 - NSS
 - Others

- dta
 - tmp
- figures
 - agriculture land
 - cotton prices
 - interact_dld
 - interact_nss
 - main_dld
 - main_nss
 - others
 - robust_interact_dld
 - robust_interact_nss
 - robust_main_dld
 - robust_main_nss
- GIS Districts
 - FAO GAEZ
 - geo2_in2009
 - IPUMS
 - output

5. Data Availability and Sources

The following table provides information on the source of each (raw) data set that is used in the analysis.

Name of Dataset	File Name(s)	Link/Access Instructions	Access Date	Data Access
Crosswalk (1966-2017) <i>(International Crops Research Institute for the Semi-Arid Tropics and Tata Cornell Institute)</i>	• raw/DLD/Apportionment.xlsx	This file contains the crosswalk that was digitized from Appendix 3 of the document “District-wise Apportioned data documentation”. A copy of this pdf file has been saved in “raw/DLD”. It can also be downloaded from http://data.icrisat.org/dld/src/support.html .	2024	Open data
District Level Database <i>(International Crops Research Institute for the Semi-Arid Tropics and Tata Cornell Institute)</i>	• raw/DLD/Apportioned/Crops_Area Production Yield.csv	1. Go to http://data.icrisat.org/dld/src/crops.html. 2. Under “Select Categories”, choose “Crops”. 3. Under “Sub Category (Crops)”, select “Area production yield”. 4. Select “Apportioned”. 5. Select all “States”, “Districts”, “Elements”, and “Items”. 6. Click on “Select Years” and select “All Years”. 7. Click on “Export as CSV”.	2024	Open data
	• raw/DLD/Unapportioned/Crops_Area Production Yield.csv	1. Go to http://data.icrisat.org/dld/src/crops.html. 2. Under “Select Categories”, choose “Crops”. 3. Under “Sub Category (Crops)”, select “Area production yield”. 4. Select “Unapportioned”. 5. Select all “States”, “Districts”, “Elements”, and “Items”. 6. Click on “Select Years” and select “All Years”.	2025	Open data

Name of Dataset	File Name(s)	Link/Access Instructions	Access Date	Data Access
		7. Click on "Export as CSV".		
	• raw/DLD/Unapportioned/Census_Operational Holdings.csv	1. Go to http://data.icrisat.org/dld/src/crops.html. 2. Under "Select Categories", choose "Census". 3. Under "Sub Category (Census)", select "Operational Holdings". 4. Select "Unapportioned". 5. Select all "States", "Districts", "Elements", and "Items". 6. Click on "Select Years" and select "All Years". 7. Click on "Export as CSV".	2025	Open data
	• raw/DLD/Unapportioned/Environment_Precipitation.csv • raw/DLD/Unapportioned/Environment_Temperature Maximum.csv • raw/DLD/Unapportioned/Environment_Temperature Minimum.csv • raw/DLD/Unapportioned/Environment_Evapotranspiration.csv	1. Go to http://data.icrisat.org/dld/src/crops.html. 2. Under "Select Categories", choose "Environment". 3. Under "Sub Category (Environment)", select "Precipitation". 4. Select "Unapportioned". 5. Select all "States", "Districts", "Elements", and "Items". 6. Click on "Select Years" and select "All Years". 7. Click on "Export as CSV". 8. Repeat steps 3 to 7 for the following variables: "Temperature Maximum", "Temperature Minimum", and "Evapotranspiration".	2025	Open data
	• raw/DLD/Unapportioned/Infrastructure_Markets Roads.csv	1. Go to http://data.icrisat.org/dld/src/crops.html. 2. Under "Select Categories", choose "Infrastructure". 3. Under "Sub Category (Infrastructure)", select "Market Roads". 4. Select "Unapportioned". 5. Select all "States", "Districts", "Elements", and "Items". 6. Click on "Select Years" and select "All Years".	2025	Open data

Name of Dataset	File Name(s)	Link/Access Instructions	Access Date	Data Access
		7. Click on "Export as CSV".		
	• raw/DLD/Unapportioned/Inputs_Fertilizer Consumption.csv	1. Go to http://data.icrisat.org/dld/src/crops.html. 2. Under "Select Categories", choose "Inputs". 3. Under "Sub Category (Inputs)", select "Fertilizer Consumption". 4. Select "Unapportioned". 5. Select all "States", "Districts", "Elements", and "Items". 6. Click on "Select Years" and select "All Years". 7. Click on "Export as CSV".	2025	Open data
	• raw/DLD/Unapportioned/Irrigation_Sourcewise Irrigated Area.csv	1. Go to http://data.icrisat.org/dld/src/crops.html. 2. Under "Select Categories", choose "Irrigation". 3. Under "Sub Category (Irrigation)", select "Sourcewise Irrigated Area". 4. Select "Unapportioned". 5. Select all "States", "Districts", "Elements", and "Items". 6. Click on "Select Years" and select "All Years". 7. Click on "Export as CSV".	2025	Open data
	• raw/DLD/Unapportioned/Prices_Farm Harvest Prices.csv	1. Go to http://data.icrisat.org/dld/src/crops.html. 2. Under "Select Categories", choose "Prices". 3. Under "Sub Category (Prices)", select "Farm harvest prices". 4. Select "Unapportioned". 5. Select all "States", "Districts", "Elements", and "Items". 6. Click on "Select Years" and select "All Years". 7. Click on "Export as CSV".	2025	Open data
National Sample Survey – Employment and	• raw/NSS/stata12_final_1989-1989(44th)_v2.dta • raw/NSS/stata12_final_2000-1989(55th)_v2.dta • raw/NSS/stata12_final_2004-1989(60th)_v2.dta	These files were downloaded from the following source:	2025	Accessible data

Name of Dataset	File Name(s)	Link/Access Instructions	Access Date	Data Access
Unemployment module (Schedule 10) (India National Sample Survey Office)	• raw/NSS/stata12_final_2007-1989(64th)_v2.dta • raw/NSS/stata12_final_2009-1989(66th)_v2.dta • raw/NSS/stata12_final_2012-1989(68th)_v2.dta • raw/NSS/IND_NSS_68_EMP_2011_2012_clean.dta • raw/NSS/IND_NSS_66_EMP_2009_2010_clean.dta • raw/NSS/IND_NSS_64_EMP_2007_2008_clean.dta • raw/NSS/IND_NSS_62_EMP_2005_2006_clean.dta • raw/NSS/IND_NSS_61_EMP_2004_2005_clean.dta • raw/NSS/IND_NSS_55_EMP_1999_2000_clean.dta • raw/NSS/IND_NSS_43_EMP_1987_1988_clean.dta	Liu, Maggie, Shamdasani, Yogita, and Taraz, Vis. "Data and Code for: Climate Change and Labor Reallocation: Evidence from Six Decades of the Indian Census." Nashville, TN: American Economic Association [publisher], 2023. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2023-04-07. https://doi.org/10.3886/E150441V1 Select "Data/Source/NSS" to access and download the files.		
Global Agro-Ecological Zones (GAEZ) database: Agro-climatic Potential Yield Food and Agriculture Organization (FAO)	• GIS Districts/FAO GAEZ/Irrigated/cott200a_yld_high.tif	File contains the information on the potential yield of cotton under "high" inputs. File can be downloaded from https://s3.eu-west-1.amazonaws.com/data.gaezdev.aws.fao.org/res02/CRUTS32/Hist/6190H/cott200a_yld.tif Accompanying metadata for the file can be found in "GIS Districts/FAO GAEZ/Irrigated/Agro-climatic Potential Yield_High.csv".	2025	Open data
	• GIS Districts/FAO GAEZ/Irrigated/cott200a_yld_low.tif	File contains the information on the potential yield of cotton under "low" inputs. File can be downloaded from https://s3.eu-west-1.amazonaws.com/data.gaezdev.aws.fao.org/res02/CRUTS32/Hist/6190L/cott200a_yld.tif	2025	Open data

Name of Dataset	File Name(s)	Link/Access Instructions	Access Date	Data Access
		Accompanying metadata for the file can be found in "GIS Districts/FAO GAEZ/Irrigated/Agro-climatic Potential Yield_Low.csv".		
India 2009 Second-level Geography (IPUMS International)	• GIS Districts/FAO GAEZ/geo2_in2009.CPG • GIS Districts/FAO GAEZ/geo2_in2009.dbf • GIS Districts/FAO GAEZ/geo2_in2009.prj • GIS Districts/FAO GAEZ/geo2_in2009.sbn • GIS Districts/FAO GAEZ/geo2_in2009.sbx • GIS Districts/FAO GAEZ/geo2_in2009.shp • GIS Districts/FAO GAEZ/geo2_in2009.shp • GIS Districts/FAO GAEZ/geo2_in2009.shx	Files can be downloaded from IPUMS International at https://international.ipums.org/international/gis_yr_specific_2nd.shtml. Scroll to "India" and select "2009" to download the files.	2025	Open data
India 2009-10 Employment Survey (IPUMS International)	• GIS Districts/IPUMS/IPUMS India 2009-10.dta	File can be downloaded from IPUMS International at https://international.ipums.org/international-action/variables/group. Click on "Select Samples" and then "Surveys". Scroll to India and select "2009". Click on "Submit Sample Selections". Select the variables "GEO1_IN2009" and "GEO2_IN2009". These can be found under "Household"-> "Geography: F-N". The metadata for the downloaded data can be found in "GIS Districts/IPUMS/IPUMS India 2009-10.txt".	2025	Accessible data
The Basic Statistical Returns	• raw/Banks/rbi_banks1999.xlsx	The pdf files can be downloaded from the Reserve Bank of India's website. To access the data, go to https://rbi.org.in/Scripts/AnnualPublications.aspx?head=Basic%20Statistical%20Returns.	2025	Open data

Name of Dataset	File Name(s)	Link/Access Instructions	Access Date	Data Access
(Reserve Bank of India)		Select “Archives” on the right-hand side menu. Then select “1999”. Go to “Table No. 2.4 - District and Population Group-Wise Deposits and Credit of Scheduled Commercial Banks”. Download parts 1 to 4. The data in the pdf files were then digitized into the Excel file.		
Adoption of Bt Cotton	• raw/Others/Choudhary and Gaur (2015)	This file contains data that were digitized from some columns in Table 1 of the paper: Choudhary, Bhagirath and Kadambini Gaur, “Biotech cotton in India, 2002 to 2014: Adoption, impact, progress & future,” 2020. SAAA Series of Biotech Crop Profiles.	2025	Open data

6. Description of Codes

The table in this section describes each of the .do files that are used in the analysis.

Note: I used ArcGIS Pro 3.5.3 to manually aggregate the districts that experienced changing boundaries between 1990 to 2011. Data on “potential cotton productivity improvements” and the “total land area” of the aggregated districts were then obtained manually in ArcGIS Pro. These variables have been saved in the reproducibility package. They are found in “GIS Districts/output”. However, please follow these steps if you would like to generate the ArcGIS variables from scratch:

- In the main .do file, “0_main.do”, run the code until line 32. In other words, stop the code after you have finished running the section “***Run do files to create crosswalks across the different datasets”.
- Go to ArcGIS to do the aggregation and generate the relevant variables. Refer to the readme file “GIS Districts/ArcGIS Reproducibility Instructions.pdf” for more information.
- Run the rest of the code from line 49 onward in the “0_main.do” file after you have completed all the steps needed in ArcGIS. In other words, run the code from the section “***Run do files to clean data” onward.

S/N	Name of .do File	Description
1	do/0_main.do	Calls and runs all the .do files needed for the analysis. These are the .do files numbered 2 to 23 in this table.
Self-written programs		
2	do/p1_dld_reg.do	Generates the figures from the regressions involving cotton land share and cotton output (main effects).
3	do/p2_dld_interact.do	Generates the figures from the regressions involving cotton land share and cotton output (heterogeneous effects).
4	do/p3_nss_reg.do	Generates the figures from the regressions involving the share of workers in agriculture (main effects).
5	do/p4_nss_interact_all.do	Generates the figures from the regressions involving employment outcomes (heterogeneous effects).
6	do/p5_nss_interact.do	Generates the figures from the regressions involving the share of workers in agriculture (heterogeneous effects).
7	do/p6_dld_interact_cottonprice.do	Generates the figures from the regressions involving cotton prices (heterogeneous effects).
Crosswalks		
8	do/c1_crosswalk_1966-2017.do	Creates the crosswalk for districts between 1966 and 2017.
9	do/c2_crosswalk_aggregated districts.do	Aggregates the districts to account for changing district boundaries during the period of analysis.
10	do/c3_crosswalk_nss_dld.do	Creates the crosswalk between districts in the NSS data and DLD data.
11	do/c4_crosswalk_banks_dld.do	Creates the crosswalk between districts in the RBI banks data and DLD data. It also cleans and collapses the banks data to aggregated districts.

S/N	Name of .do File	Description
Cleaning		
12	do/c5_clean_faogaez.do	Cleans the geospatial variables (which were obtained manually using ArcGIS Pro 3.5.3).
13	do/c6_clean_dld.do	Cleans the DLD data.
14	do/c7_clean_nss.do	Cleans the NSS data.
15	do/c8_final_dld.do	Prepares the dataset for regressions involving cotton land share and cotton output. It merges all the relevant datasets with the DLD data.
16	do/c9_final_nss.do	Prepares the dataset for regressions involving employment outcomes. It merges all the relevant datasets with the NSS data.
17	do/c10_clean_final_cottonprices.do	Cleans and prepares the dataset for regressions involving cotton prices. It merges all the relevant datasets with the cotton prices data.
Figures		
18	do/s1_bt cotton stats.do	Generates figures on Bt cotton for the “background” section of the paper.
19	do/r1_regressions_dld.do	Generates the raw trends figures for cotton land share and cotton output. It also uses the programs “p1_dld_reg.do” and “p2_dld_interact.do” to generate the figures from the regressions involving these outcomes.
20	do/r2_robustness_dld.do	Uses the programs “p1_dld_reg.do” and “p2_dld_interact.do” to generate the figures from the robustness checks involving cotton land share and cotton output.
21	do/r3_regressions_nss.do	Generates the raw trends figures for the share of workers in agriculture. It also uses the programs “p3_nss_reg.do”, “p4_nss_interact_all.do”, and “p5_nss_interact.do” to generate the figures from the regressions involving employment-related outcomes.
22	do/r4_robustness_nss.do	Uses the programs “p3_nss_reg.do” and “p5_nss_interact.do” to generate the figures from the robustness checks involving the share of workers in agriculture.
23	do/r5_regressions_robustness_cottonprices.do	Generates the raw trends figures for cotton prices. It uses the program “p6_dld_interact_cottonprice.do”. It also generates figures from the regressions and robustness checks involving cotton prices.
ArcGIS analysis		
24	GIS Districts/crosswalk_gis_dld.do	Creates the crosswalk between districts in the IPUMS shapefile and DLD data.

7. Lists of Exhibits

The tables below list the figures that appear in the paper and the corresponding .do files that produce them.

Main Paper

S/N	Exhibit Name	Output Filename(s)	Code
1	Figure 1	figures/others/raw_btarea.png figures/others/raw_btfarmers.png	s1_bt cotton stats.do
2	Figure 2	Not applicable as this figure is not a data output.	
3	Figure 3	figures/main_dld/raw_shareland_cotton.png figures/main_dld/main_shareland_cotton.png figures/main_dld/raw_prod_cotton.png figures/main_dld/main_prod_cotton.png figures/main_dld/raw_lprod_cotton.png figures/main_dld/main_lprod_cotton.png	r1_regressions_dld.do
4	Figure 4	figures/main_nss/raw_sec_prin_ag.png figures/main_nss/main_ag_employment.png	r3_regressions_nss.do
5	Figure 5	figures/interact_nss/interact_migland_ag_employment.png figures/interact_nss/interact_migland_notag_employment.png figures/interact_nss/migland_subsmpl_Gujarat_ag_employment.png figures/interact_nss/interact_migland_notemployed.png	r3_regressions_nss.do
6	Figure 6	figures/interact_dld/interact_migland_shareland_cotton.png figures/interact_dld/interact_migland_prod_cotton.png figures/interact_dld/interact_migland_lprod_cotton.png	r1_regressions_dld.do
7	Figure 7	figures/interact_dld/raw_interact_labland_shareland_cotton.png figures/interact_dld/interact_labland_shareland_cotton.png figures/interact_dld/raw_interact_labland_prod_cotton.png figures/interact_dld/interact_labland_prod_cotton.png figures/interact_dld/raw_interact_labland_lprod_cotton.png figures/interact_dld/interact_labland_lprod_cotton.png	r1_regressions_dld.do
8	Figure 8	figures/interact_nss/raw_interact_labland_sec_prin_ag.png figures/interact_nss/interact_labland_ag_employment.png	r3_regressions_nss.do

S/N	Exhibit Name	Output Filename(s)	Code
9	Figure 9	figures/cotton prices/raw_cottonprice.png figures/cotton prices/cottonprice.png figures/cotton prices/raw_logcottonprice.png figures/cotton prices/logcottonprice.png	r5_regressions_robustness_cottonprices.do
10	Figure 10	figures/cotton prices/raw_interact_labland_cottonprice.png figures/cotton prices/interact_labland_cottonprice.png figures/cotton prices/raw_interact_labland_logcottonprice.png figures/cotton prices/interact_labland_logcottonprice.png	r5_regressions_robustness_cottonprices.do
11	Table 1	Not applicable as this table is not a data output. It reports numbers that appear in Blaise et al. (2014).	
12	Table 2	Not applicable as this table is not a data output. It reports numbers that appear in Blaise et al. (2014).	

Online Appendix

S/N	Exhibit Name	Output Filename(s)	Code
1	Figure B1	figures/robust_main_dld/controls_climate_shareland_cotton.png figures/robust_main_dld/subsmpl_exNorth_shareland_cotton.png figures/robust_main_dld/controls_nonparm_shareland_cotton.png figures/robust_main_dld/subsmpl_nochange_shareland_cotton.png figures/robust_main_dld/subsmpl_Gujarat_shareland_cotton.png figures/robust_main_dld/subsmpl_outliers_shareland_cotton.png	r2_robustness_dld.do
2	Figure B2	figures/robust_main_dld/controls_climate_prod_cotton.png figures/robust_main_dld/subsmpl_exNorth_prod_cotton.png figures/robust_main_dld/controls_nonparm_prod_cotton.png figures/robust_main_dld/subsmpl_nochange_prod_cotton.png figures/robust_main_dld/subsmpl_Gujarat_prod_cotton.png figures/robust_main_dld/subsmpl_outliers_prod_cotton.png	r2_robustness_dld.do
3	Figure B3	figures/robust_main_dld/controls_climate_lprod_cotton.png figures/robust_main_dld/subsmpl_exNorth_lprod_cotton.png figures/robust_main_dld/controls_nonparm_lprod_cotton.png figures/robust_main_dld/subsmpl_nochange_lprod_cotton.png figures/robust_main_dld/subsmpl_Gujarat_lprod_cotton.png figures/robust_main_dld/subsmpl_outliers_lprod_cotton.png	r2_robustness_dld.do

S/N	Exhibit Name	Output Filename(s)	Code
4	Figure B4	figures/robust_main_nss/controls_climate_ag_employment.png figures/robust_main_nss/subsmpl_exNorth_ag_employment.png figures/robust_main_nss/controls_nonparm_ag_employment.png figures/robust_main_nss/subsmpl_nochange_ag_employment.png figures/robust_main_nss/subsmpl_Gujarat_ag_employment.png figures/robust_main_nss/subsmpl_outliers_ag_employment.png	r4_robustness_nss.do
5	Figure B5	figures/agriculture land/aglandarea_controls_nonparm.png figures/agriculture land/logaglandarea_controls_nonparm.png figures/agriculture land/aglandarea_share_controls_nonparm.png	r2_robustness_dld.do
6	Figure B6	figures/robust_interact_dld/labland_controls_climate_shareland_cotton.png figures/robust_interact_dld/labland_subsmpl_exNorth_shareland_cotton.png figures/robust_interact_dld/labland_controls_nonparm_shareland_cotton.png figures/robust_interact_dld/labland_subsmpl_nochange_shareland_cotton.png figures/robust_interact_dld/labland_subsmpl_Gujarat_shareland_cotton.png figures/robust_interact_dld/labland_subsmpl_outliers_shareland_cotton.png	r4_robustness_nss.do
7	Figure B7	figures/robust_interact_dld/labland_controls_climate_prod_cotton.png figures/robust_interact_dld/labland_subsmpl_exNorth_prod_cotton.png figures/robust_interact_dld/labland_controls_nonparm_prod_cotton.png figures/robust_interact_dld/labland_subsmpl_nochange_prod_cotton.png figures/robust_interact_dld/labland_subsmpl_Gujarat_prod_cotton.png figures/robust_interact_dld/labland_subsmpl_outliers_prod_cotton.png	r4_robustness_nss.do
8	Figure B8	figures/robust_interact_dld/labland_controls_climate_lprod_cotton.png figures/robust_interact_dld/labland_subsmpl_exNorth_lprod_cotton.png figures/robust_interact_dld/labland_controls_nonparm_lprod_cotton.png figures/robust_interact_dld/labland_subsmpl_nochange_lprod_cotton.png figures/robust_interact_dld/labland_subsmpl_Gujarat_lprod_cotton.png figures/robust_interact_dld/labland_subsmpl_outliers_lprod_cotton.png	r4_robustness_nss.do
9	Figure B9	figures/robust_interact_nss/labland_controls_climate_ag_employment.png figures/robust_interact_nss/labland_subsmpl_exNorth_ag_employment.png figures/robust_interact_nss/labland_controls_nonparm_ag_employment.png figures/robust_interact_nss/labland_subsmpl_nochange_ag_employment.png figures/robust_interact_nss/labland_subsmpl_Gujarat_ag_employment.png figures/robust_interact_nss/labland_subsmpl_outliers_ag_employment.png	r4_robustness_nss.do

S/N	Exhibit Name	Output Filename(s)	Code
10	Figure B10	figures/cotton prices/cottonprice_nochange.png figures/cotton prices/labland_subsmpl_nochange_cottonprice.png figures/cotton prices/logcottonprice_nochange.png figures/cotton prices/labland_subsmpl_nochange_logcottonprice.png	r5_regressions_robustness_cottonprices.do