

Iselin, John; McCulloch, Sean; Ryan, Erica

**Working Paper**

## A Comment on "Climate Change and Labor Reallocation: Evidence from Six Decades of the Indian Census"

I4R Discussion Paper Series, No. 180

**Provided in Cooperation with:**

The Institute for Replication (I4R)

*Suggested Citation:* Iselin, John; McCulloch, Sean; Ryan, Erica (2024) : A Comment on "Climate Change and Labor Reallocation: Evidence from Six Decades of the Indian Census", I4R Discussion Paper Series, No. 180, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/305228>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



No. 180

I4R DISCUSSION PAPER SERIES

# **A Comment on “Climate Change and Labor Reallocation: Evidence from Six Decades of the Indian Census”**

John Iselin

Sean McCulloch

Erica Ryan

October 2024

## I4R DISCUSSION PAPER SERIES

I4R DP No. 180

# **A Comment on “Climate Change and Labor Reallocation: Evidence from Six Decades of the Indian Census**

**John Iselin<sup>1</sup>, Sean McCulloch<sup>2</sup>, Erica Ryan<sup>1</sup>**

*<sup>1</sup>University of Maryland, College Park/USA*

*<sup>2</sup>Brown University, Providence/USA*

OCTOBER 2024

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](#), and [RWI – Leibniz Institute for Economic Research](#), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

### **Editors**

**Abel Brodeur**  
*University of Ottawa*

**Anna Dreber**  
*Stockholm School of Economics*

**Jörg Ankel-Peters**  
*RWI – Leibniz Institute for Economic Research*

# A comment on “Climate Change and Labor Reallocation: Evidence from Six Decades of the Indian Census”\*

John Iselin, Sean McCulloch, Erica Ryan

October 1, 2024

## Abstract

[Liu et al. \(2023\)](#) examines the effect of climate change on labor allocation in India over a long time span. The authors find that rising temperatures are correlated with lower shares of workers in non agricultural sectors. They also identify a likely mechanism: falling agricultural productivity leads to a reduction in demand for non-agricultural goods or services, leading to a reduction in labor demand in non-agricultural sectors. We undertake a reproduction and extension of [Liu et al. \(2023\)](#), and find that we are able to computationally reproduce all the numbers produced by the authors up to marginal differences in the calculation of standard errors. We describe a set of data issues that hindered full reproduction of the original dataset, and, in one case, contradicts a claim of data availability made by the authors. Finally, we test the robustness of the main results to a more consistent use of fixed effects and the use of Poisson regression, following [Chen and Roth \(2024\)](#). The Poisson regression approach does not alter the results, but in several of the new fixed effects specifications the author’s original results are less conclusive and lose statistical significance.

---

\*Authors: Iselin: University of Maryland, College Park. E-mail: [jiselin@umd.edu](mailto:jiselin@umd.edu). McCulloch: Brown University. E-mail: [sean\\_mcculloch@brown.edu](mailto:sean_mcculloch@brown.edu). Ryan: University of Maryland, College Park. E-mail: [emryan@umd.edu](mailto:emryan@umd.edu).

## 1 Introduction

Liu et al. (2023) henceforth LST, investigate the impact of rising temperatures on labor reallocation in India. In the context of developing economies, there are often large gaps in productivity between different parts of the economy. The authors note two particular examples: the gap between agricultural and non-agricultural sectors, and the gap between urban and rural areas. Given these gaps, movement between sectors or areas could lead to substantial productivity gains. The authors ask whether rising temperatures impact this reallocation, with a particular focus on "medium- to long-term climate variations, which capture adaptation or intensification effects more accurately than short-term weather fluctuations." In other words, the authors believe that their approach can more effectively capture effects of climate shifts that look like climate change.

The authors use a district-level panel of weather, Census, and survey data from 1961 through 2011. This includes measures of worker shares across sectors, rates of within-district migration between rural and urban areas, and rates of urbanization, as well as measures of precipitation and temperature that span the 10-years preceding any given census or survey period, designed to capture the long-term effects the authors are interested in. They use two primary methods: a panel fixed-effect model, identified off variation in decade-to-decade temperature and precipitation within a given district (including region-time controls), and a long-difference model, identified off of long-term changes in weather patterns over the entire length of the panel. They find that higher temperatures do appear to reduce the transition from agricultural to non-agricultural sectors, with no matching change in the share of the population living in urban areas. Their results indicate that a 1 degree Celsius increase in temperature leads to a 17 percent increase in the share of the labor force that is in the agricultural sector and an 8.2 percent drop in the share that is in non-agriculture. They also identify a likely mechanism: higher temperatures leads to a reduction in agricultural productivity, reducing demand for non-agricultural goods, thereby reducing *labor demand* in the non-agricultural sector.

In the present report prepared for the Institute for Replication ([Brodeur et al. 2024](#)), we investigate whether their analytical results are computationally reproducible, describe select issues with data availability, and test the robustness of their

results to: (1) using a more consistent set of fixed effects and (2) using a Poisson regression approach to assess the effect of climate on migration. We were able to successfully reproduce LST’s results in tables 1 through 6, with one exception. The Conley standard errors in our reproduction, produced using the *reg2hdfespatial* Stata command, are close but not identical to those reported in the original paper, as shown in [Table 2](#). This issue emerges - we suspect, from using the updated version of the ado file available online, rather than the older version used by the authors (which is included in the replication package but is no longer available at the provided link). We are able to perfectly replicate the Conley standard errors using the version of the ado file provided in the replication package, as shown in [Table 3](#).

We make two modifications to the paper as presented by LST. First, Tables 2, 4, 5 and 6 (their district-year panel specifications) do not utilize a consistent set of fixed effects. While in all regression results they include district and year FEs, in Tables 2 and 4 they include specifications with region-year linear trends and region-year fixed effects, while in Tables 5 and 6 they include specifications with region-year linear trends and state-year linear trends. This change corresponds with the shift from Census data to Survey data, but to our knowledge the authors do not justify this change in the article. We report results for Table 2 with all three specifications included (see [Table 3](#)). We find that the inclusion of state-year linear trends reduces the magnitude and statistical significance of the effect of temperature on the agricultural and non-agricultural labor shares: by over half for the former. The effect of temperature on urbanization almost quadruples, and becomes statistically significant, but the effects on migration do not change. These results imply that the variation the authors are relying on for their estimates persists when controlling for region-specific trends, but not for state-specific trends.

Second, the authors specify migration as the log of the migration share  $\ln Y_{jsrt} = \ln(0.01 + Y_{jsrt})$ , to account for 0s. Following [Chen and Roth \(2024\)](#), we reproduce the migration results in Table 2 using Poisson regression, as shown in [Table 4](#). We find that the sign of the coefficients on temperature change, switching from negative to positive (of a similar magnitude) but the results remain statistically indistinguishable from 0.

## 2 Computational Reproducibility

To computationally reproduce the analysis in [Liu et al. \(2023\)](#), we used the replication package provided by the authors.<sup>1</sup> The package includes most, but not all, of the required data - the Census files for 2001 and 2011 are held on a separate Dropbox, located [here](#). The package includes all of the code required to produce both the analysis dataset and the results in the published paper. It also includes two ado files that cannot be downloaded directly via *ssc install* via Stata. We were able to reproduce the analysis in the paper from the raw data, with some small exceptions, which will be discussed below. See Table 1 for details.

### 2.1 Data Issues

We describe in this section a set of minor errors in either coding or data cleaning that we uncovered while reproducing the study. These were, by and large, minor and rarely resulted in a deviation from the results presented in the paper. Rather, they either required minor changes to make the code run (ex. variables stored as strings when the code refers to them as integers) or reflect some difficulty in our ability to reconstruct the raw data as provided.

First, after downloading the 2001 and 2011 Census files stored on the separate Dropbox, we noticed that the variables in the dataset *udir.dta* are all stored as strings. This results in the do-file *lst-dm-01c-make\_pca\_2001\_2011.do* breaking when the resulting state identifiers are merged onto the 2001 and 2011 Census files. This was easily fixed by modifying the code to update the variables to numeric.

Second, in the final analysis file *lst\_analysis\_PCA288.dta*, there is one district-year observation (Andaman and Nicobar Islands in 1961) with a negative agricultural labor share. This is because the value for total population is stored as 0, resulting in the denominator of the share variable being a negative number (see line 124 in the do-file *lst-dm-05a-create-lst\_analysis\_PCA288.do*). However, this does not impact the main analysis, as the impacted district is dropped in the balanced panel.

Third, there were several broken links in the ReadMe file of the replication package, which are enumerated below:

---

<sup>1</sup>See <https://www.openicpsr.org/openicpsr/project/150441/version/V1/view>.

- On page 1, the link to the Conley spatial HAC standard errors ADO file.
- On pages 2 and 7, the link to the authors' website.
- On pages 2, 7, and 8 the link to the Census of India's Website.
- On page 7 the link to the SAS commands used to create the data dictionary for Vanneman's data.
- On page 9 the link to the VDSA data.
- On page 9 the links to Willmott and Matsuura (2014).

Fourth, the authors claim that data from the 1971 Census on the share of male population who are rural-to-urban intra-district migrants are unavailable. This is consistent with the raw data provided by Vanneman and Barnes (2000), which is the source file used by [Liu et al. \(2023\)](#). However, we were able to locate a set of PDFs with the missing data, accessible via the [Census of India Catalog](#). For the 1971 Census, the required data is available at the district level, but is stored in state-level PDFs. Due to time constraints, we have not put together the data, but the data are available.<sup>2</sup>

We also experienced the following additional issues when trying to download the raw source data from the original sources:

- **PCA: 2001 - 2011:** `pca_census_01_india.dta`, `pca_census11_india.dta`, and `udir.dta` are only available through their dropbox, so we were not able to replicate these data files from the relevant raw source data. For the other files sourced through the census we were able to download all files and confirm they match the supplied raw data.
- **PCA: Crosswalk and Map:** We were not able to find the Administrative Atlas of India on the Indian Census website.
- **NSS - CES:** The files downloaded directly from the website are not stored in the same file format as the supplied raw data. As a result, we were not able to confirm that the source data matches the raw data provided.

---

<sup>2</sup>To find, use the document search option, search for Migration tables, Part II-D, and limit the search to the 1971 Census. The series number should vary across states. The following [link](#) is to the example file for the Punjab.

- **NSS - EMP:** Similar to the CES data, the source data from the website are not stored in the same file format as the supplied raw data. As a result, we were not able to confirm that the source data matches the raw data provided.
- **VDSA:** The data available on the website are at subject/district level, whereas the raw data available in the replication package is aggregated into a single data file. As a result, we were not able to confirm the validity of this data set.
- **EOPP:** This data was easy to navigate, but the ReadMe file was not clear on which file to download. We were able to verify which file to download after checking the files available in the replication package.
- **RBI:** The link provided takes you to a 200+ page pdf file. The provided data in the replication package is an excel file, with no instructions on how they coded the data from the pdf. As a result, we were unable to verify the raw source data.

## 2.2 Pre-analysis Plan

The authors did not register a pre-analysis plan.

## 3 Robustness Reproduction

We chose to test the robustness of the results in two ways. First, we reproduce Tables 2 and 5 using the full set of region or state fixed effects and linear trends. Second, we replicate the results in Table 2 for the migration outcome variable using a Poisson regression, to avoid the issues raised by [Chen and Roth \(2024\)](#). The decision to test the use of Poisson regression as a robustness check was made after reading the paper but prior to observing the code and data, while the choice to test using region or state fixed effects and linear trends was made during the review process.

### 3.1 Region and State Fixed Effects and Linear Trends

In tables 2 and 4 of LST, the authors present alternative estimates including region-year trends versus region-year fixed effects. In tables 5 and 6, the authors present

alternative estimates including region-year trends versus state-year trends. It is unclear why the authors did not include a consistent set of alternative specifications throughout all the tables, e.g. either region-year trends and region-year FE throughout or region-year trends and state-year trends throughout.

We reproduce Table 2 in LST with additional specifications that include state-year trends so results under all three sets of controls are presented. The results of this extension are presented in [Table 3](#). We use the exact final analysis data file that was used to create Table 2 of [Liu et al. \(2023\)](#). Columns 1, 3, 4, 6, 7, 9, 10, and 12 perfectly replicate the results of Table 2. The remaining columns (2, 5, 8, and 11) add to the results the specification with state-specific linear trends. For these specifications, the estimates have similar patterns qualitatively for agricultural and nonagricultural labor share but are significantly smaller in magnitude compared to the estimates reported in the paper (e.g., 0.18 versus 0.06 for agricultural labor share). These estimates are no longer statistically significant at conventional levels. On the other hand, there is a larger, negative coefficient for the urbanization rate when controlling for state-year trends which is statistically significant at the 5% level.

### 3.2 Poisson regression

The results of this extension are presented in [Table 4](#). We use the exact final analysis data file that was used to create Table 2 of [Liu et al. \(2023\)](#). Columns 1 and 2 replicate perfectly columns 7 and 8 of that table, with the exceptions of the Conley standard errors. Columns 3 and 4 report the results of the exact same specifications, but run using a Poisson regression. The observation count is different due to the *ppmlhdfc* command in Stata dropping five observations as either singletons or separated by a fixed effect. In the Poisson regression, the sign on the temperature results switches, but is statistically indistinguishable from 0. The results imply that a 10 percent increase in Temperature increases the rate of rural to urban migration rate by approximately 1.6 percent, though these new results do not rule out a null effect.

## 4 Conclusion

We replicated the "Climate Change and Labor Reallocation: Evidence from Six Decades of the Indian Census." We found up to very minor coding errors and differences in the calculation of standard errors that all results could be replicated from the provided data.

Reproducing the raw data was challenging, largely due to outdated links and changing file structure. However, what raw data we were able to reproduce matched the provided raw data exactly or the authors' description of any data changes. It is worth noting that future work could extend this analysis using data on rural-to-urban migration from the 1970s Census not utilized by the authors.

We pursued two robustness checks. First, we tested the log outcome variables using a Poisson regression. We found that results using these specifications were similar, measuring a null effect. Second, we estimated the coefficients in table 2 of LST under additional specifications that control for state-by-year linear trends, which were included as controls in other tables presented in the paper, but omitted from the main results. We found results were sensitive to the sets of region- or state-level fixed effects or linear trends included. In particular, the coefficient estimates on labor sector allocation decreased significantly in magnitude while remaining similar qualitatively when using state-level linear trends, as opposed to the regional-level controls included by the authors. While not compiled in this report, future replication efforts could focus on reproducing the other tables in LST with all three sets of controls.

## References

- Brodeur *et al.*: 2024, Mass Reproducibility and Replicability: A New Hope, *I4R Discussion Paper Series* .
- Chen, J. and Roth, J.: 2024, Logs with zeros? some problems and solutions, *Quarterly Journal of Economics* **139**(2), 891–936.
- Liu, M., Shamdasani, Y. and Taraz, V.: 2023, Climate change and labor reallocation: Evidence from six decades of the indian census, *American Economic Journal: Economic Policy* **15**(2), 395–423.

## 5 Tables

Table 1: Replication Package Contents and Reproducibility

Replication Package Item	Fully	Partial	No
Raw data provided		✓	
Analysis data provided	✓		
Cleaning code provided	✓		
Analysis code provided	✓		
Reproducible from raw data	✓*		
Reproducible from analysis data	✓		

*Notes:* This table summarizes the replication package contents contained in [Liu et al. \(2023\)](#). The asterisk on the reproducible from Raw data denotes the minor required changes to the raw files provided.

Table 2: Replicating Table 2, standard errors with updated reg2hdfespatial

	Ag. labor share		Nonag. worker share		Urbanization		Migrant share	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Temperature	0.181 (0.059)*** (p: 0.002) [0.079]** (p: 0.022)	0.157 (0.062)** (p: 0.011) [0.081]* (p: 0.052)	-0.077 (0.033)** (p: 0.020) [0.037]** (p: 0.039)	-0.086 (0.031)*** (p: 0.006) [0.038]** (p: 0.022)	-0.021 (0.042) (p: 0.613) [0.045] (p: 0.636)	0.001 (0.045) (p: 0.986) [0.044] (p: 0.985)	-0.013 (0.059) (p: 0.830) [0.063] (p: 0.843)	-0.018 (0.064) (p: 0.783) [0.073] (p: 0.810)
Perspiration	-0.153 (0.059)** (p: 0.010) [0.117] (p: 0.192)	-0.081 (0.060) (p: 0.177) [0.110] (p: 0.461)	-0.026 (0.030) (p: 0.381) [0.040] (p: 0.519)	-0.001 (0.030) (p: 0.962) [0.035] (p: 0.967)	0.025 (0.042) (p: 0.543) [0.047] (p: 0.586)	0.000 (0.044) (p: 0.999) [0.048] (p: 0.999)	-0.016 (0.050) (p: 0.750) [0.060] (p: 0.790)	-0.000 (0.058) (p: 0.995) [0.070] (p: 0.996)
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

Notes: This table replicates the results in [Liu et al. \(2023\)](#) Table 2 with updated versions of the Stata packages handling computation of Conley standard errors. Standard errors clustered at the district level in parentheses, and Conley standard errors in brackets, with the respective p-values below each.

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 3: Replicating Table 2, adding in state-year linear trends

	Ag. labor share			Nonag. worker share			Urbanization			Migrant share		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Temperature	0.181 (0.059)*** (p: 0.002) [0.077]** (p: 0.018)	0.063 (0.056) (p: 0.259) [0.072] (p: 0.382)	0.157 (0.062)** (p: 0.011) [0.076]** (p: 0.038)	-0.077 (0.033)** (p: 0.020) [0.038]** (p: 0.045)	-0.064 (0.038)* (p: 0.091) [0.040] (p: 0.105)	-0.086 (0.031)*** (p: 0.006) [0.035]** (p: 0.015)	-0.021 (0.042) (p: 0.613) [0.046] (p: 0.645)	-0.081 (0.041)** (p: 0.047) [0.041]** (p: 0.049)	0.001 (0.045) (p: 0.986) [0.046] (p: 0.986)	-0.013 (0.059) (p: 0.830) [0.062] (p: 0.840)	0.036 (0.056) (p: 0.525) [0.062] (p: 0.563)	-0.018 (0.064) (p: 0.783) [0.068] (p: 0.797)
Precipitation	-0.153 (0.059)** (p: 0.010) [0.098] (p: 0.120)	-0.093 (0.057) (p: 0.102) [0.092] (p: 0.309)	-0.081 (0.060) (p: 0.177) [0.093] (p: 0.386)	-0.026 (0.030) (p: 0.381) [0.036] (p: 0.472)	-0.033 (0.029) (p: 0.259) [0.035] (p: 0.351)	-0.001 (0.030) (p: 0.962) [0.032] (p: 0.965)	0.025 (0.042) (p: 0.543) [0.044] (p: 0.564)	0.011 (0.039) (p: 0.767) [0.040] (p: 0.774)	0.000 (0.044) (p: 0.999) [0.045] (p: 0.999)	-0.016 (0.050) (p: 0.750) [0.058] (p: 0.783)	-0.042 (0.049) (p: 0.397) [0.054] (p: 0.441)	-0.000 (0.058) (p: 0.995) [0.064] (p: 0.996)
Region-year trends	Y	N	N	Y	N	N	Y	N	N	Y	N	N
State-year trends	N	Y	N	N	Y	N	N	Y	N	N	Y	N
Region-year FE	N	N	Y	N	N	Y	N	N	Y	N	N	Y
Observations	1,548	1,548	1,548	1,620	1,620	1,620	1,596	1,596	1,596	1,350	1,350	1,350

Notes: Columns (1), (3), (4), (6), (7), (9), (10), and (12) replicate the results in [Liu et al. \(2023\)](#) Table 2. The remaining columns run models that include state-specific linear trends, rather than region-specific trends or region-year fixed effects. Each model includes fixed effects for year and district. Standard errors clustered at the district level in parentheses, and Conley standard errors in brackets, and below each are the corresponding p-values

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 4: Migration Results with OLS and Poisson Regression

	OLS		Poisson	
	(1)	(2)	(3)	(4)
Temperature	-0.0126 (0.0588) (p: 0.830)	-0.0176 (0.0638) (p: 0.783)	0.1587 (0.1563) (p: 0.310)	0.1636 (0.1636) (p: 0.317)
Precipitation	-0.0160 (0.0504) (p: 0.750)	-0.0004 (0.0582) (p: 0.995)	-0.0486 (0.1198) (p: 0.685)	0.0218 (0.1283) (p: 0.865)
Region-year trends	Yes	No	Yes	No
Region-year fixed effects	No	Yes	No	Yes
Observations	1350	1350	1345	1345

*Notes:* Columns (1) and (2) replicate Columns (7) and (8) of [Liu et al. \(2023\)](#) Table 2. Columns (3) and (4) use the *ppmlhdfc* Stata command to run a Poisson pseudo-likelihood regression. Each model includes fixed effects for year and district. Standard errors clustered at the district level in parentheses, with p-values below each.

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$