**Supplementary (For online Publication only)**

Appendix A

Figure A1. Consistent district boundaries, 1961-2021

**

Source: Authors’ construction; Notes: Figure illustrates 280 consistent district boundaries over 1961-2021. In the analysis, only 267 consistent district boundaries have been used as part of the major states of India.

Appendix B

Table B1. Year-wise summary statistics of districts

|  |  |
| --- | --- |
| Variables | Mean (Standard Deviation) |
| 1961 | 1971 | 1981 | 1991 | 2001 | 2011 | 2021 |
| Decadal Population Growth Rate (%) | 20.26 (7.64) | 22.47 (5.18) | 22.36 (5.05) | 21.25 (5.41) | 19.11 (6.58) | 15.62 (6.85) | 14.38 (6.44) |
| Cultivated Land (%) | 53.14 (20.31) | 55.05 (20.33) | 55.4 (20.93) | 55.78 (20.72) | 55.68 (20.63) | 55.3 (20.72) | 55.35 (20.17) |
| Agricultural Population Density | 3.76 (13.31) | 5.48 (29.95) | 8.55 (65.69) | 12.04 (105.06) | 28.26 (350.01) | 55.05 (763.85) | 12.29 (45.49) |
| Log Yield | 6.56 (0.94) | 6.73 (0.81) | 6.92 (0.84) | 7.13 (0.96) | 7.39 (0.91) | 7.62 (0.91) | 7.78 (0.92) |
| Irrigated Area (%) | 10.48 (12.10) | 13.21 (15.06) | 16.61 (17.87) | 20.55 (19.66) | 24.21 (21.94) | 27.71 (22.57) | 30.99 (23.63) |
| Cropping Intensity (%) | 118.58 (16.41) | 122.08 (18.52) | 127.71 (20.39) | 132.46 (24.61) | 137.66 (26.53) | 141.74 (27.23) | 153.19 (31.01) |
| Urbanisation (%) | 15.77 (13.99) | 17.20 (14.27) | 20.09 (14.65) | 22.14 (15.14) | 23.57 (15.67) | 26.43 (17.41) | 28.96 (19.38) |
| Nonfarm Workers (%) | 30.37 (15.17) | 27.87 (14.84) | 31.02 (15.39) | 32.72 (16.03) | 42.56 (17.51) | 45.45 (18.43) | 48.22 (20.12) |

*Source: Authors’ construction*

*Figure B1: Box plot showing the distribution and outliers for variables*

*Source: Authors’ construction*

Notes: PGR = Population Growth Rate; CL = Cultivated Land; Y = Yield of Major Crops; IA = Irrigated Area; CI = Cropping Intensification; U = Urbanisation; NW= Nonfarm Workers

Appendix C

To understand reverse causation, Table C1 places alternatively the population growth rate as a dependent variable of cultivated and agricultural land. After attempting various combinations of independent variables and lags of dependent variables, the following two models passed the specification tests. The models are based on panels up to 2011. In none of the models, the representative land variable was found to be significant.

Table C1. Arellano-Bover/Blundell-Bond linear dynamic model (dependent variable – population growth rate)

|  |  |  |
| --- | --- | --- |
| Variables | Model 1 | Model 2 |
| Population Growth Ratet-1 | 0.429\*\*\* (0.113) | 0.432\*\*\* (0.114) |
| Population Growth Ratet-2 | 0.143\* (0.083) | 0.142\* (0.082) |
| Cultivated Land | -0.042 (0.051) |  |
| Agricultural Land |  | -0.037 (0.051) |
| Agricultural Population Density | -0.002\*\*\* (0.000) | -0.002\*\*\* (0.000) |
| Log Yield | -1.489\*\*\* (0.429) | -1.494\*\*\* (0.430) |
| Urbanisation | -0.024 (0.078) | -0.018 (0.076) |
| Nonfarm Workers | 0.013 (0.049) | 0.017 (0.049) |
| Education | -0.163\*\*\* (0.049) | -0.167\*\*\* (0.049) |
| Constant | 26.104\*\*\* (5.548) | 25.981\*\*\* (5.433) |
| Groups | 265 | 265 |
| Observations | 1056 | 1056 |
| Instruments | 14 | 14 |
| AR (1) | -4.087\*\*\* | -4.106\*\*\* |
| AR (2) | 1.486 | 1.513 |
| Sargan Test | 8.016 | 8.237 |
| Wald Chi2 | 1068.903\*\*\* | 1278.600\*\*\* |

*Source: Authors’ construction*

*Notes: dependent variable: Population Growth Rate (%); \*\*\* p<.01, \*\* p<.05, \* p<.1; Standard Errors are in parenthesis*

Appendix D

For the robustness of the results in Table 3 of the main paper, we have run the same set of models using the full sample, which consists of 280 districts over the panel. The results show, in all the models, that the population growth rate significantly affects cultivated land, which re-affirms our earlier models with major states of India.

Table D1. Results of Arellano-Bover/Blundell-Bond linear dynamic panel regression model showing the effect of population growth rate on cultivated land – Full Sample

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variables | Model 1a | Model 1b | Model 2a | Model 2b | Model 3a | Model 3b | Model 4a | Model 4b | Model 5a | Model 5b |
| Cultivated Landt-1 | 0.551\*\*\* (0.089) | 0.529\*\*\* (0.079) | 0.501\*\*\* (0.101) | 0.487\*\*\* (0.087) | 0.552\*\*\* (0.11) | 0.52\*\*\* (0.098) | 0.511\*\*\* (0.101) | 0.504\*\*\* (0.084) | 0.499\*\*\* (0.096) | 0.494\*\*\* (0.096) |
| Population Growth Rate | 0.105\*\*\* (0.027) | 0.100\*\*\* (0.021) | 0.093\*\*\* (0.026) | 0.096\*\*\* (0.021) | 0.097\*\*\* (0.025) | 0.110\*\*\* (0.031) | 0.073\*\* (0.031) | 0.073\*\* (0.034) | 0.079\*\*\* (0.025) | 0.103\*\*\* (0.031) |
| Agricultural Population Density |  |  | 0.000\*\* (0.000) | 0.000\*\*\* (0.000) | -0.012\*\*\* (0.002) | -0.012\*\*\* (0.002) | 0.000\* (0.000) | 0.000\*\* (0.000) | -0.012\*\*\* (0.002) | -0.012\*\*\* (0.001) |
| Log Yield |  |  |  |  | 0.697\*\* (0.352) | 0.882\*\* (0.365) |  |  | 0.668\*\* (0.335) | 0.947\*\*\* (0.34) |
| Irrigated Area |  |  |  |  | -0.037 (0.031) | -0.031 (0.032) |  |  | 0.010 (0.033) | -0.012 (0.034) |
| Cropping Intensity |  |  |  |  | -0.005 (0.013) | -0.006 (0.01) |  |  | -0.007 (0.013) | -0.007 (0.01) |
| Urbanization |  |  |  |  |  |  | -0.019 (0.047) | 0.020 (0.023) | -0.027 (0.04) | 0.018 (0.023) |
| Nonfarm Workers |  |  |  |  |  |  | -0.015 (0.022) | -0.028 (0.018) | -0.033 (0.022) | -0.031\* (0.018) |
| Constant | 22.682\*\*\* (5.145) | 23.967\*\*\* (4.388) | 25.834\*\*\* (5.806) | 26.571\*\*\* (4.802) | 19.571\*\*\* (6.797) | 19.703\*\*\* (6.219) | 26.777\*\*\* (5.769) | 26.713\*\*\* (4.784) | 24.163\*\*\* (5.995) | 21.418\*\*\* (6.226) |
| Groups | 280 | 280 | 278 | 279 | 278 | 276 | 278 | 279 | 278 | 279 |
| Observations | 1397 | 1677 | 1385 | 1663 | 1381 | 1658 | 1385 | 1663 | 1381 | 1658 |
| Instruments | 10 | 12 | 11 | 13 | 14 | 16 | 13 | 15 | 16 | 18 |
| AR (1) | -2.430\*\* | -2.129\*\* | -2.314\*\* | -2.079\*\* | -2.495\*\*\* | -2.203\*\* | -2.353\*\* | -2.120\*\* | -2.368\*\* | -2.162\*\* |
| AR (2) | -0.979 | -0.052 | -0.236 | 0.191 | -0.105 | 0.285 | -0.266 | 0.209 | -0.080 | 0.265 |
| Sargan Test | 5.186 | 5.752 | 4.062 | 4.378 | 4.363 | 6.636 | 3.918 | 3.784 | 2.997 | 4.801 |
| Wald Chi2 | 41.155\*\*\* | 67.074\*\*\* | 58.396\*\*\* | 111.718\*\*\* | 274.127\*\*\* | 357.701\*\*\* | 50.045\*\*\* | 106.570\*\*\* | 292.832\*\*\* | 604.793\*\*\* |

*Source: Authors’ construction*

*Notes: dependent variable: Cultivated Land (%); \*\*\* p<.01, \*\* p<.05, \* p<.1; Standard Errors are in parenthesis*

Appendix E

In the theoretical framework, we argue that the percentage use of cultivated land will decline after the population growth rate starts declining and the growth rate in agricultural intensification exceeds the population growth rate. Figure E1 shows the yearly mean values of three variables i.e., population growth rate (%), percentage cultivated land, and growth in yield of major crops (%), overall districts from 1961-71 to 2011-21. From 1971 to 81, the population growth rate started declining, and in that period, growth in the yield of major crops exceeded the population growth rate. This fulfills the two conditions we have postulated. Following this, the percentage of cultivated land started declining from 1981-91.

Figure E1: Trend in Population Growth Rate, Growth in Agricultural Yield and Cultivated Land Use

Source: Authors’ construction

Notes: Values are the year-wise mean of district-level measures of the three variables.

Appendix F

We have implemented a machine learning approach to compare in-sample and out-of-sample errors for different predictors. Since we are interested in changes in land use and population growth over time, we have used the first difference variables rather than the original variables. We randomly split the samples into two sets, with 80% in the training set and 20% in the test set. The panel used is up to 2011. We have used five predictors: ordinary least squares (OLS), principal component analysis (PCA), neural networks, random forest, and boosting. All five models were run for the training set, and then the mean square error (MSE) for the training and test samples was calculated.

Table F1 shows the MSE of the training and test samples for the five predictors. The results indicate a very small difference in MSEs, between the training and test samples. The training sample MSE is lowest for the random forest model, while the test sample MSE is lowest for the OLS model. Finally, we have shown the OLS model with the full sample in Table F2, and the results demonstrate that the population growth rate significantly affects cultivated land use.

Table F1: Mean Square Error of Training and Test Samples

|  |  |  |
| --- | --- | --- |
| Predictor | Train Sample MSE | Test Sample MSE |
| Ordinary Least Square | 25.139 | 24.388 |
| Principal Component Analysis | 25.286 | 24.974 |
| Neural Network | 25.608 | 24.587 |
| Random Forest | 13.118 | 24.707 |
| Boosting | 24.688 | 25.509 |

*Source: Authors’ construction;*

Table F2: Results of ordinary least square regression showing the effect of population growth on cultivated land with the first difference

|  |  |
| --- | --- |
| Variables | Model 1 |
| Population Growth Rate | 0.043\* (0.023) |
| Agricultural Population Density | -0.018\*\*\* (0.006) |
| Log Yield | 0.637\*\* (0.294) |
| Irrigated Area | 0.061\*\*\* (0.022) |
| Cropping Intensity | -0.017 (0.012) |
| Urbanization | -0.039 (0.043) |
| Nonfarm Workers | -0.113\*\*\* (0.024) |
| Constant | 0.675\*\*\* (0.202) |
| No. of Observation | 1318 |
| Wald Chi2 | 8.031\*\*\* |

*Source: Authors’ construction*

*Notes: dependent variable: Cultivated Land (%); \*\*\* p<.01, \*\* p<.05, \* p<.1; Standard Errors are in parenthesis*