# Rural Roads, Climate Change, and the Dynamics of Structural Transformation: Evidence from India

Aparajita Dasgupta* Devvrat Raghav†

Last Updated: September 13, 2024

**Abstract**

Climate change threatens to reverse structural transformation in rural areas by depressing local demand for nontradables, among other channels. In this paper, we ask whether road connectivity can preserve the gains from structural transformation in emerging markets amid a sustained rise in temperature. We do so by exploiting a large-scale rural road construction program in India, which assigned roads to villages whose historical population lay above an arbitrary threshold, in conjunction with a fuzzy difference-in-discontinuity design. We find that road connectivity raises the share of households engaged in cultivation and shifts workers away from farm labour towards the service sector. This is accompanied by an increase in household wealth, which accrues more to cultivators. These effects are moderated by temperatures, and we cannot rule out zero effects on wealth and service sector employment when decadal temperatures increase by 1 percent or more.

*Keywords: Infrastructure, Road access, Climate Shocks, Labour markets, India*

*JEL Classification: O12, O13, O15, O18, Q54, R23*

*Corresponding author; Department of Economics, Ashoka University, India (Email: aparajita.dasgupta@ashoka.edu.in)

†University of Virginia (email: zsa5kb@virginia.edu).

# 1 Introduction

Climate change is a first-order problem for policymakers today. Indeed, global surface temperature has risen faster in 1970-2020 than in any 50-year period in the previous two millenia (Lee et al., 2024). This rise in temperature has received much attention in the economics literature, which has documented their deleterious effects on agricultural yields (Zhao et al., 2017), growth (Dell et al., 2012), labour productivity (Somanathan et al., 2021), and civil conflict (Burke et al., 2015). In recent work, Liu et al. (2023) show using Indian data that rising temperatures can shrink local demand for services and push people back towards agriculture, thereby reversing the process of structural transformation. They also find that these effects are much weaker in areas with historically better access to paved roads and banks.

In this paper, we ask if this impact of paved roads is causal, i.e., whether they can stymie the consequences of rising temperatures on structural transformation. Essentially, we are interested in whether the decline in transport costs and improvement in market-access due to (new) paved roads can weaken the local demand effects of temperature, and if so, to what extent. However, unlike Liu et al. (2023), we limit our attention to rural India, because villages are less likely to be proximate to paved roads and more exposed to the effects of temperature shocks due to the prevalence of agriculture (Colmer, 2021). This choice is appropriate to answer our research question because the local demand effects of temperature manifest only in rural areas (Liu et al., 2023).

We exploit the variation in villages' access to a paved road due to the Pradhan Mantri Gram Sadak Yojana (PMGSY), a large-scale rural road construction program that prioritized the assignment of new roads to villages whose pre-program population lay above an arbitrary threshold, i.e., to villages that were "eligible" for a new road. To estimate the effect of a road, we compare villages with a pre-program population just above the threshold to those that were just below. Because this effect should run counter to that of a temperature increase (see Section 2 for details), we are interested in whether it persists when a village is exposed to a temperature shock, and, if so, to how large a shock.

To measure temperature shocks, we use gridded monthly weather data, spanning 1961-2011 at a spatial resolution of $0.25° \times 0.25°$, from the ERA5 Reanalysis Data. A shock to the decadal mean temperature of a region is defined as the percentage deviation in the previous decade's mean air temperature from the lagged 30-year mean. This is because the latter represents the region's "climate normal", i.e., its expected temperature (Arguez & Vose, 2011). Our measure is defined at the level of a district, which represents a local labour market in India (Allen & Atkin, 2022).

We rely on a fuzzy difference-in-discontinuity design to estimate the effect of road connectivity across the spectrum of decadal temperature shocks. Our strategy combines elements from a fuzzy regression discontinuity (RD) and difference-in-difference (DID) design. As all eligible villages had not received a road by 2012 (Asher & Novosad, 2020), we use a village's *eligibility* for a new road as an instrument for its access to a paved road under PMGSY. We thereby compare outcomes across villages just above the population threshold and those just below. This is the first difference. By interacting the village's access to a paved road with the district-level temperature shock, we compare this difference along the population threshold for groups of villages that suffered the same temperature shock. This is the second difference, which reveals how the effect of road connectivity is moderated by temperature shocks.

We find that the provision of paved roads reduces the employment share of farm labour, while raising the employment shares of both services and agricultural cultivators. Although all three effects are weakened by temperature shocks, the effect on services is especially fleeting: It is attenuated by any temperature shock exceeding one percentage point (the sample median), which equals into a sustained 0.3°C increase for a decade. Thus, the median village experiences an 18 percentage point decline in farm labour's share and a 14 percentage point increase in cultivators' share, but no change in service sector's employment.

Our results on land-use are consistent with the idea that temperature shocks worsen agricultural yields and the option value of agricultural land. In the absence of a temperature shock, we find that road access leads to fewer households selling-off their farmland. However, we observe no change in either mechanization, the adoption of commercial (non-food) crops, or on farm output (as proxied by satellite-based measures of vegetation). Instead, rural roads alter the composition of cultivators by incentivizing smaller landholders to persist in agriculture, whose family members spend additional time on the household farm amid the exodus of farm labour.

Moreover, we find that road connectivity has positive welfare effects, as exemplified by an increase in the household wealth index. However, both effects decline monotonically with temperature and neither persists for villages that experience an above-median temperature shock. This could explain why Asher and Novosad (2020) find a null effect of PMGSY roads on household wealth on the same sample. Our findings also align with predictions from the cannonical two-region model of rural-urban trade, where a fall in trade costs between the rural and urban regions is welfare-improving (Gollin & Rogerson, 2014), but a simultaneous decline in productivity makes the overall effect on welfare ambiguous.

Using the household wealth index in the IHDS data, we also provide suggestive evidence that the increase in cultivators amid road connectivity is not because households remain trapped in agriculture, which is possible if roads increased land prices and prevented cultivators from exiting agriculture by selling their land. We find that cultivator households enjoy a larger increase in wealth, relative to their non-cultivator peers, when their village gains a new road. This wealth premium declines with the intensity of temperature shocks, and disappears when the shock exceeds 0.9 percentage points. Given that we see smaller increases in cultivators' share of employment and more sales of agricultural land amid higher temperature shocks, our results overall suggest that the profitability of agriculture, rather than demand in land markets, is a more likely explanation of the rise in cultivation.

Our paper contributes to three strands of literature. The first strand studies the role of rural transport networks on economic development, such as the recent work on the PMGSY. Evidence suggests that they increase the consumption of non-local goods by rural households Aggarwal (2018); raise middle school enrolment rates (Adukia et al., 2020); improve the uptake of formal healthcare, such as children's vaccinations (Aggarwal, 2021) and reduce fertility (Dasgupta et al., 2024). Asher and Novosad (2020) use a fuzzy RD on the same sample as ours to show that rural roads improve access to transport facilities, like buses, and enable farm labour to exit agriculture. However, they find no effect on household consumption or wealth, unlike prior work that documented a positive link between roads and development outcomes.[1] Garg et al. (2024) show that the farm labour exits lead cultivators to engage more heavily in crop burning as an alternative, which raises PM2.5 pollution and infant mortality in downwind villages. In contrast, Shamdasani (2021) shows that roads reallocate farm labour across nearby markets and reduce search costs for cultivators, which raises investments by cultivators and their use of hired labour.

We make two contributions to the literature on rural roads. First, we show that roads incentivize smaller landholders to remain as cultivators, thereby raising cultivators' share of employment. Consistent with hired labour becoming costlier, we also find that adult household members spend more on the family farm. Second, we show that new roads boost service sector employment and household wealth. However, these effects are strongly attenuated by exposure to high temperature shocks, which could serve to explain some of Asher and Novosad (2020)'s null results, particularly with respect to wealth.

The second strand is concerned with how transport networks foster structural transformation. Theory suggests transport networks can improve market access by lowering mobility costs, both for goods and labour (Michaels, 2008; Gollin & Rogerson, 2014). Empirical evidence supports this assertion, with work on railroads and highways showing that better access to transport can bolster trade, facilitates migration, and boost real incomes (Donaldson, 2018; Morten & Oliveira, 2023).[2] Proximity to transport also raises land prices (Jacoby, 2000; Donaldson & Hornbeck, 2016; Shrestha, 2020) and dampens the volatility in crop

[1]There is both historical and contemporary work that suggests road networks can reduce poverty and bolster economic growth. Examples include contexts such as Bangladesh (Khandker et al., 2009; Khandker & Koolwal, 2011), Ethiopia (Dercon et al., 2009), Tanzania (Castaing Gachassin, 2013), Papua New Guinea (Gibson & Rozelle, 2003), Nepal (Fafchamps & Shilpi, 2013), Indonesia (Gibson & Olivia, 2010), and Vietnam (Mu & Van de Walle, 2011).

[2]The trade-related effects are contingent on factor mobility, with evidence from China showing that railroads had little effect on GDP growth in proximate areas (Banerjee et al., 2020), while highways reduced growth in peripheral regions (Faber, 2014).

prices due to local rainfall shocks (Allen & Atkin, 2022). Nevertheless, because better transportation can improve (non-)agricultural productivity, it can facilitate the transition from farm labour to the non-farm sectors (Gollin & Rogerson, 2014).

However, we show that rural road connectivity has a mixed effect on agricultural participation: It enables exits from farm labour, but also raise the share of households engaged in cultivation, i.e., they increase the number of agricultural "micro-enterprises". This divergence from theory is likely because, in a setting like India, urban labour markets do not have enough vacancies to absorb the rural workforce, even more-so due to non-pecuniary costs of migration (Lagakos et al., 2023).

The final strand of related literature tackles the economic effects of climate change. There is a growing body of evidence on how rising temperatures reduce agricultural yields (Zhao et al., 2017; Vogel et al., 2019), dampen productivity and economic growth (Dell et al., 2012; Bilal & Känzig, 2024), raise the risk of civil conflict (Burke et al., 2015), and threaten to reverse structural transformation through local demand effects over the long term (Liu et al., 2023), even if labour reallocation can mitigate these effects in the short term (Colmer, 2021). We contribute to this literature by showing that last-mile road connectivity can counteract the local demand effects of temperature shocks up to a one percentage point increase beyond the climate normal. Moreover, unlike highways, rural roads do not contribute to deforestation (Asher et al., 2020), which means they do not create the same dichotomy between market access and climate policy (Gollin & Wolfersberger, 2023). Thus, we show that rural roads may represent an indirect policy tool to alleviate some of the burden of rising temperatures. In doing so, we extend the literature to highlight the role of public policies beyond that of direct climate policies, such as pollution legislation (Greenstone & Hanna, 2014), carbon capture and sequestration (Nordhaus, 2019), and carbon pricing (Colmer et al., 2024).

This paper is structured as follows. Section 2 outlines the conceptual framework, Section 3 describes the PMGSY, Section 4 presents our data sources and transformations, Section 5 covers our empirical strategy, Section 6 discusses the results, and Section 7 concludes.

## 2 Conceptual Framework

To fix ideas, consider an economy with two regions – rural and urban.[3] The latter denotes the closest town, which is assumed to be the gateway for the village to the "rest of the world". Economic activity is distributed across three sectors – agriculture, manufacturing, and services. Like Liu et al. (2023), we assume that agricultural and manufacturing output is tradable, but services are not. The sectors are unevenly distributed over space, such that the rural region features only agriculture and services, while the

[3]For examples of the two-region model, see Matsuyama (1992) and Gollin and Rogerson (2014).

urban region contains manufacturing and services.

A standard assumption is that moving goods (and workers) across regions is subject to an iceberg-style transportation cost, i.e., a fraction of the good's value (or worker's wage income) is exhausted during transport (Krugman, 1991; Gollin & Rogerson, 2014). Similarly, the rural region is assumed to be a small open economy that trades with the larger urban region, which means that prices in all rural markets are fixed relative to the urban markets (Matsuyama, 1992). However, costly transportation creates a wedge in prices, thereby preventing equalization over space. Specifically, nominal prices (in equilibrium) are lower in the village.

Climatic conditions often enter the model as shifters of productivity. For one, higher temperatures lead to a decline in agricultural yields (Zhao et al., 2017; Vogel et al., 2019), which is akin to a negative shock to the productivity of agricultural land and physical capital. The heat also lowers labour productivity across sectors (Day et al., 2019; Somanathan et al., 2021; Shayegh et al., 2021; LoPalo, 2023). Therefore, when the village experiences higher-than-usual temperatures, the corresponding drop in agricultural productivity would reduce both aggregate output and agricultural incomes, which would induce a contraction in local demand for services. Because services are non-tradable, the contraction in demand would push labour back towards agriculture (Liu et al., 2023).[4]

One problem with this model of occupational choice is that it does not distinguish between agricultural cultivators and labour. While the former (mostly) own the farms that they work on and earn their profits from the goods market, the latter are hired by cultivators for a fixed wage. This can generate heterogenous responses to shifts in agricultural productivity. If agricultural productivity falls due to high temperature, then workers in the service sector have an incentive to switch to agriculture. In contrast, the effect on cultivators's incentives is ambiguous because a contraction in food supply would raise market prices and the net effect on profits would depend upon the price elasticity in the agricultural goods market.[5]

Improvements in rural road connectivity could attenuate the effect of higher temperatures by reducing transport costs.[6] After all, lower costs effectively expand the output of the agricultural and manufacturing sectors for the same level of inputs. Moreover, lower transport costs could allow rural workers to commute to urban labour markets and earn higher real wages, which would stimulate local demand for services in the village. Thus, any contraction in local demand for services should be relatively muted if the village

[4]If the reduction in food supply makes the food constraint binding, then that would also push workers back towards agriculture (Matsuyama, 1992).

[5]Moreover, amid land market frictions, such as in rural India (Foster & Rosenzweig, 2022), cultivators may be unable to exit agricultural even if they wished to.

[6]For households with vehicles, this reduction would stem from lower travel times on a paved road, relative to a mud (or no) road. For those without vehicles, the availability of an all-weather paved road could enable access to fast(er) and cheap(er) travel by incentivizing both private and public transport providers to expand their routes.

had better road connectivity.

Further, the effect on cultivators is not obvious. On one hand, lower transport costs increase profit margins through the goods market. On the other hand, the increased accessibility of urban labour markets would bid up the wage rate for agricultural labour.[7] If the former is sufficiently stronger than the latter, then cultivation's share of employment in the village in the presence of high temperatures and a new road may even exceed the level that would prevail in an equilibrium with normal temperatures and no road. This could also occur if households reallocate members away from other work towards contributing to the family farm, which would reflect in a relatively higher share of small landholders.[8] Thus, it is possible that road connectivity may not reduce agriculture's share of employment, even amid high temperatures.

Regardless, new paved roads could, theoretically, be a buffer against the deleterious impact of temperature shocks. However, whether road connectivity is sufficient to offset the welfare reductions attributable to a typical temperature shock, and the range of shocks for which this is possible, is an empirical question.

# 3 Program Details

The Pradhan Mantri Gram Sadak Yojana (PMGSY) is a rural road construction program in India, launched in December 2000. Its primary thrust was to provide all-weather paved roads to habitations—hamlets within a village—that did not already have them. Although most roads that were built under the program were replacements for mud roads, i.e., were "new roads", the program did allow for upgradation of existing paved roads. New roads were built in a staggered fashion; the process began in 2001 and the first phase of the program (PMGSY-I) culminated in 2013, having delivered over 483,000 kilometers of paved roads. Figure 3(a) illustrates the timing of new road construction in our analysis sample, wherein we aggregate all PMGSY data to the village-level.

Roads were sanctioned under a two-step procedure. First, block[9] officials identified all habitations whose population, as of the 2001 Population Census, lay above a threshold. This threshold varied over time, amounting to 1000 during 2001-03, 500 during 2004-07, and 250 during 2008-12. If a state had already built roads in habitations above the higher threshold (1000), then it was allowed to immediately use the lower threshold of 500 instead (Asher & Novosad, 2020). Program guidelines also allowed habitations that were

[7] Improved connectivity can also induce workers from nearby villages to commute to the newly-connected village, thereby expanding the supply of labour (Shamdasani, 2021).

[8] The high fixed cost of mechanization prevents small-scale cultivators from using machinery as a substitute for manual labour (Foster & Rosenzweig, 2022)

[9] A block (or a sub-district) is the third level of administration in India, nested within a district (second-level) and a state (first-level). As of the 2011 Population Census, the median block comprised 110 villages and 127,000 people, while the median district contained ten blocks.

within 500 metres of each other to combine their populations into a 'cluster'. Habitations (and clusters) that satisfied the population criterion were placed on the priority list. This list was made available to local elected officials for feedback. It was then passed to district officials, who allocated new roads to habitations on the priority list. Habitations outside the priority list could also gain a new road if they lay along the least-cost path that connected a habitation on the list to its nearby market center or town.

Although the program remains active as of July 2024, we focus only on roads built under PMGSY-I. This is because unlike PMGSY-I, where population was the central criterion, subsequent phases also utilised socioeconomic criteria to determine whether a habitation was to be placed on the priority list.[10] As such, by restricting attention to PMGSY-I roads, we can minimize the role of political economy effects, which are otherwise a feature of infrastructure programs (Blimpo et al., 2013; Burgess et al., 2015). Moreover, not all states complied with the PMGSY-I guidelines. Therefore, we follow Asher and Novosad (2020) and restrict our sample to the six states that did. These complier states, along with their population thresholds (in parentheses), are Chhattisgarh (500, 1000), Gujarat (500), Madhya Pradesh (500, 1000), Maharashtra (500), Odisha (500), and Rajasthan (500).

# 4 Data

Our main dataset is a village-level panel, comprised of 11,188 villages whose baseline population lay within $[c-84, c+84]$, where $c \in \{500, 1000\}$ is the PMGSY population threshold and 84 is the optimal bandwidth (Imbens & Kalyanaraman, 2012) for our difference-in-discontinuity design. For most outcomes, we have data at two points in time — 2001 (the baseline) and 2011 (the endline) — that correspond to Census years. In some cases, we have additional data from the 1991 Population Census. We supplement this dataset with microdata from the India Household and Development Survey (IHDS), which is a nationally-representative panel of 40,000 households who were interviewed in 2004-05 (Round-I) and in 2011-12 (Round-II). The exact data sources, along with their spatio-temporal coverage, are described below.

## 4.1 Data Sources

**Rural roads:** Our data on rural roads originally comes from two sources – the 2001 Population Census and the PMGSY's Online Management, Monitoring, and Accounting System (OMMS). The village directory from the Population Census records whether each village had access to a paved road at the baseline. The OMMS identifies all habitations, and by extension, villages, that received a paved road under the PMGSY, alongside the road's completion date. Although the PMGSY's primary focus was on constructing new

[10] See the official guidelines for further details.

roads, it also upgraded some of the extant stock. However, we limit our attention to new roads. As such, we drop all villages that already had paved roads at the baseline from our sample. A village is considered *treated* if it received a new (all-weather) PMGSY road by 2011.

**Population:** Data on village population is part of the Primary Census Abstract (PCA) of the Population Census for 1991, 2001, and 2011. Moreover, population data from the 2001 Census was used to determine the allotment of new PMGSY roads, with villages receiving priority if their population, as of the 2001 Census, exceeded the PMGSY-specified threshold.

In addition to the total population, the PCA also records the number of residents that are: (i) literate (ii) (fe)male (iii) aged 0-6 years (iv) either Scheduled Caste (SC) or Scheduled Tribe (ST). However, these group-wise population totals are not cross-cutting, i.e., for example, one cannot compute the share of (wo)men that are literate. The age distribution in villages is available only in the Socioeconomic and Caste Census (SECC), which released in 2012. Consequently, we know for each village the number of people in the following age groups: 0-10, 11-20, 21-30, 31-40, 41-50, and 51-60 years.

**Employment:** Our first source of employment data is the PCA, which records the number of people employed in: (i) agricultural cultivation, (ii) agricultural labour (iii) manufacturing, and (iv) other services. An agricultural cultivator (henceforth, just cultivator) is somebody who either owns or rents their farm, and earns profits from the sale of the farm's produce. Agricultural labour, on the other hand, are hired by cultivators for a wage, which can be either cash or in-kind.

The PCA also identifies the number of people that are engaged in agriculture only part-time (between 0-6 months in a year) and dubs them *marginal workers* in agriculture. Others, for whom agriculture is the primary occupation, are known as *main workers.* Therefore, the number of agricultural workers is the sum of main and marginal workers. However, data on marginal workers is available only for 2001 and 2011.

Our second source of data is the Economic Census, which covers all non-agricultural firms in the village. The Economic Census was conducted in 1990, 1998, and 2013, and we match each round to the closest Population Census. Hence, data from 1998 serves as the baseline, while data from 2013 serves as the endline. Notably, the Economic Census records the number of workers in these firms, regardless of their residence. Thus, it complements the data on non-agricultural village employment from the PCA, which covers only residents of the village.

**Agricultural land:** The SECC (2012) contains two measures of land ownership, namely the share of households that own farmland and the share of cultivators that own farmland. The former reflects the overall availability of agricultural land (for rent or otherwise) and the latter represents cultivators' access to such land. We combine these measures into a *farm ownership* using the inverse-covariance-weighting approach of Anderson (2008). To compare villages at the baseline, we use the land ownership rate from the Below Poverty Line (BPL; 2002) Census.

**Agricultural practices:** As before, the SECC (2012) provides our measures of mechanization, namely: (i) the share of households that own irrigation equipment and (ii) the share of households that own mechanical farm equipment, such as tractors and threshers. Using these measures, we construct a *mechanization* index. Moreover, from the 2011 Population Census, we know the three most-commonly-grown crops in most villages. We follow Asher and Novosad (2020) and use this crop schedule to construct three indicators: (i) whether any crop was perishable (ii) whether any crop was non-cereal, and (iii) whether any crop was non-calorie, i.e., was not used for food. We combine these indicators into a *commercial crops* index.

**Vegetation indices:** In the absence of village-level data on agricultural output, we rely on the Enhanced Vegetation Index (EVI) and the Normalized Difference Vegetation Index (NDVI) as proxy measures.[11] We use the mean NDVI and EVI over 2000-2002 for the baseline, and over 2011-2013 for the endline. These are equivalent to the *cumulative* NDVI and EVI used by Asher and Novosad (2020).[12]. Our proxy for farm output is the summary index constructed from the mean NDVI and EVI.

**Household wealth:** The SECC (2012) questionnaire included a module on asset ownership, from which we know the share of households in the village that owned: (i) agricultural land (ii) a motor vehicle (iii) a home (iv) a refrigerator (v) a phone (vi) irrigation equipment (vii) mechanical farm equipment. We combine this module with indicators of whether the household's dwelling has (i) a solid roof and (ii) a solid wall, alongside indicators of whether the household's annual income was between 5000-10000 Indian Rupees (INR) or greater than 10000 INR. All these variables are aggregated using inverse-covariance-weighting to produce an *asset index*.

**Climate:** We obtain data on monthly rainfall and mean surface air temperature for 1961-2011 from the ERA5 Climate Reanalysis data, which comes from the European Centre for Medium-Range Weather Forecasts (ECMWF). Reanalysis datasets, like the ERA5, assimilate sensor data with projections from climate models to reliably produce climate data with fewer spatial or temporal gaps than traditional, sensor-only datasets (Gleixner et al., 2020).[13] By leveraging the ERA5, we continue a trend of recent papers relying on

[11] Asher and Novosad (2020) demonstrate that the NDVI and EVI are both strongly correlated with two other measures of agricultural production, namely: (i) the cereal crop-suitability index from the FAO Global Agro-Ecological Zones (GAEZ) database and (ii) the share of village land that was irrigated, as per the Population Census.

[12] To reduce the influence of other greenery, particularly forest area, on the vegetation indices, we control for the share of village area used for: (i) grazing and (ii) as a forest. Both these shares are computed using the village directory from the Population Census (2001 and 2011)

[13] This is important because traditional datasets that use only sensor outputs, such as the University of Delaware's Gridded Monthly Series (Wilmott & Matsuura, 2018), do not have data on temperature for some of the smaller districts in our sample.

reanalysis data (Foster & Gehrke, 2017; Colmer, 2021).

The ERA5 data is available on a $0.25^{\circ} \times 0.25^{\circ}$ grid. Using the district boundaries from the 2011 Population Census, we aggregate this data for each district by averaging the values of all grid points across all months that lay either on its boundary or within its interior. The result is a district-year panel consisting of annual means of temperature and rainfall.

To measure climate shocks, we first define the expectation with respect to which a shock can be measured. We define the expected temperature (and rainfall) in any district-year as the annual mean for that district over the previous three decades. This choice follows the World Meteorological Organization (WMO)'s advice on using a lagged 30-year period to represent a region's current climate (Arguez & Vose, 2011). As such, the means from this 30-year period represent a district's "climate normal" (Liu et al., 2023).

Our measure of a long-term temperature shock is the percentage deviation of the previous decade's mean temperature from the climate normal (lagged 30-year mean). This measure has a natural centrality – zero represents alignment with expectations, while higher (lower) values represent a hotter-than-expected (cooler-than-expected) decade. Our measure also maintains comparability with previous work, such as Liu et al. (2023), whose use of decadal mean temperature with district fixed effects essentially leverages deviations from the district's long-term (60-year) mean temperature to identify its impact on the labour market, urbanization, and migration.

**IHDS:** We use the India Household and Development Survey (IHDS) to provide supplementary evidence on how road connectivity and temperature could affect (i) land ownership, (ii) household members' time spent on the family farm, and (iii) migration and (iv) household wealth. The IHDS is representative at both the national and state levels, with over 20,000 survey respondents across 500 villages in the six complier states that were surveyed both in 2004-05 and 2011-12. We can identify villages that received a new road during the interim because the questionnaire on village amenities records several dimensions of village infrastructure.

However, the IHDS lacks village-level identifiers, so it cannot be determined with certainty whether a new road originated under the PMGSY. To somewhat assuage this concern, we propose a matching exercise where we would use data from the village questionnaire in 2011-12 to map IHDS villages to the universe of Indian villages in the 2011 Population Census. The matching's success is predicated on amenities that are relatively less prone to measurement error, such as the presence of (i) schools (public, private, primary, secondary) (ii) vocational training institutes (iii) a pucca road (iv) primary health center (v) community center (vi) bus service (private or public) (vii) bank branch.

This process could still leave some IHDS villages with multiple matches. We can solve this by using two variables in the IHDS data that are sourced directly from the 2001 Population Census - the number of households in the village and the village area (in hectares). Specifically, for each IHDS village with multiple matches, we retain only those matches whose number of households is the most similar.[14] If some villages still have several matches, then we can further reduce the set of matches by retaining only those where the village area is the closest. From exploratory work in this regard, we know that this process culminates in a unique match for each IHDS village. Then, we can identify each new road within the IHDS as a PMGSY road if and only if the matched village received a new PMGSY road during 2005-2011.

We discuss the exact matching process in greater depth in the Appendix B, wherein we also present some descriptive statistics that illustrate the quality of this process.

## 4.2 Descriptive Statistics

To anchor our discussion of the results, in Figure 1 we present a series of descriptive statistics that illustrate the secular trends in our sample of villages. Besides confirmation that our sample is comparable to the other villages in the six complier states, there are three highlights:

First, although total employment and agriculture's share of the workforce remained stagnant between 2001-2011, the composition of agricultural employment had changed, driven by a decline in the cultivators' share of employment and a near-equal increase in farm labour's share. This was accompanied by a modest expansion of non-farm village enterprises, amounting to 16 jobs per-1000 residents, with a simultaneous decline in the share of non-hired workers (read: family members) in these enterprises.

Second, rural population grew by an average of 20 percent during both 1991-2001 and 2001-2011, with a similar growth rate in the (mean) number of households. Simultaneously, the average literacy rate rose from one-quarter of the village population to just over half. Diving the untransformed population count by the number of households suggests an average household size of 5-6 people, which matches the average from the household microdata in the IHDS.

Third, consistent with the decline in cultivators' share of employment, the share of households with land ownership declined from by 17 percentage points between 2001 and 2011. However, while there was no change in cultivated area, the share of cultivated area under irrigation rose by 7 percentage points. This suggests that perhaps small(er) landholders, who may have had limited access to irrigation, were the ones driving the exodus from cultivation.

[14] Since the numeric values in the IHDS data may be subject to minor data entry errors, we don't require the match to have exactly the same number of households.

Figure 1: Socioeconomic context over time

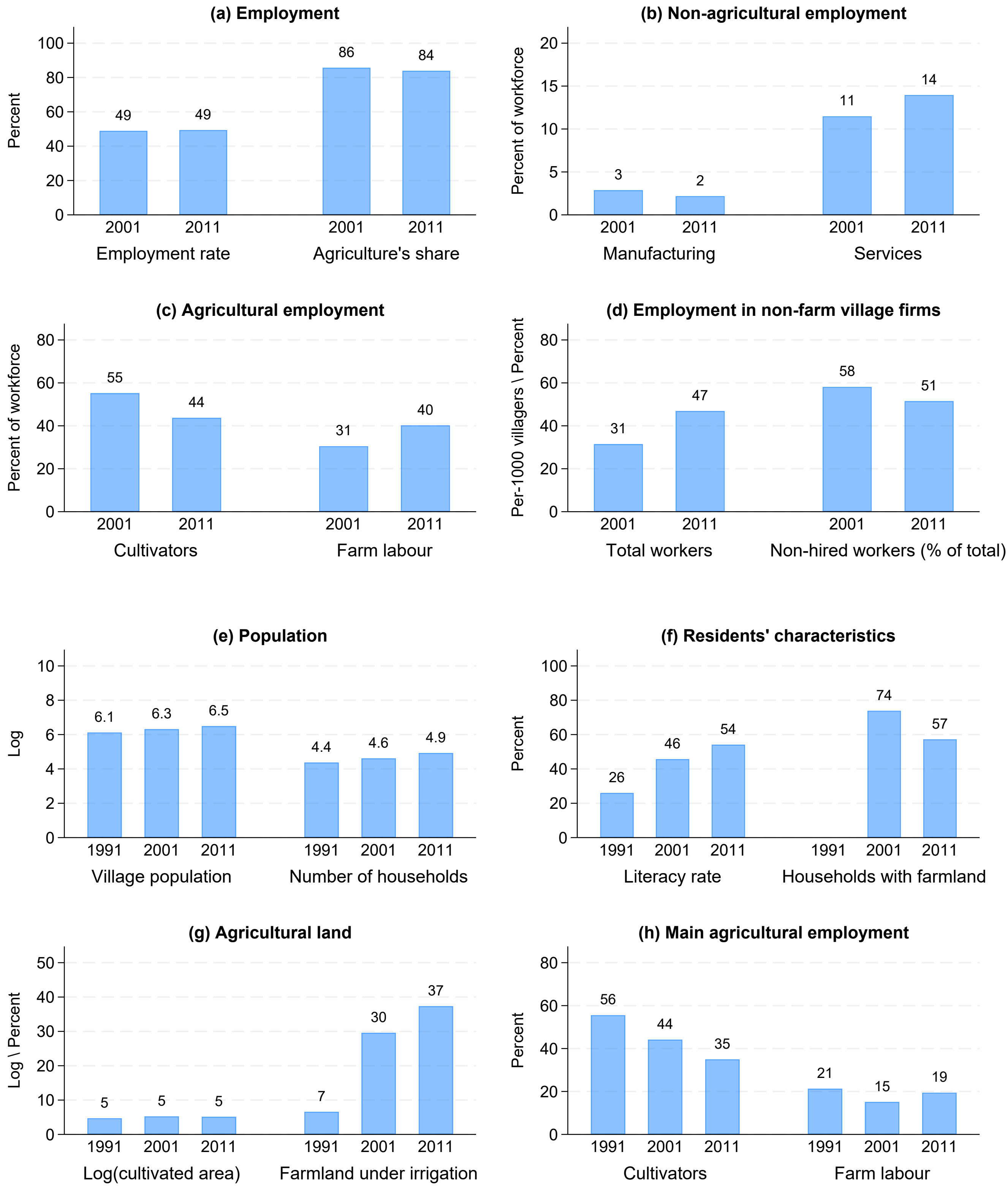
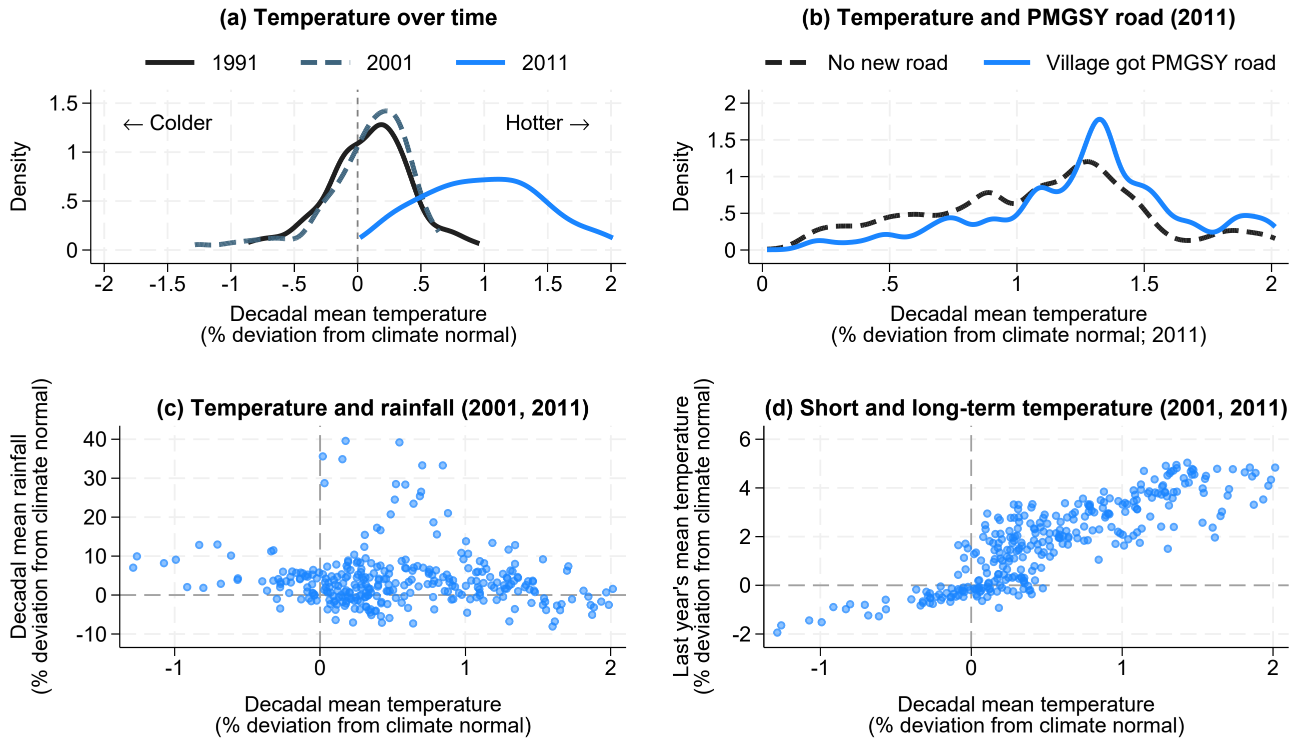



Figure 2 summarizes our measure of temperature shocks. Panel (a) indicates that temperatures have been rising over time, relative to the climate normal. The shock is (approximately) normally-distributed around

Figure 2: Distribution of temperature shocks




zero in 1991, suggesting that the median district had experienced a decade where the mean temperature was in-line-with the (historical) expectation. In contrast to the modest rightward shift in this distribution in 2001, the sharp change in 2011 suggests that the 2002-2011 decade was non-trivially hotter than the previous two. Indeed, the median district experienced a decadal mean temperature that was 1% above the expectation, which translates into an average increase of 0.26°C throughout the decade. Moreover, the distribution's support in 2011 does not include zero, i.e., every single district in our sample experienced a hotter-than-expected decade in 2002-2011, albeit to different degrees.

Panel (b) shows that both villages that received a PMGSY road by 2011 and those that did not are similarly well-represented throughout the distribution of district-level temperature shocks. As such, our discussion of the buffer effects of road connectivity is not limited only to contexts where temperature shocks are relatively less (or more) pronounced.

Panel (c) reveals that though the decadal temperature shocks are uncorrelated with decadal rainfall shocks, reflecting the fact that the sustained increase in temperatures is not a byproduct of a decline in rainfall. Contrariwise, the decadal temperature shock is positively correlated with a similar measure of short-term temperature – the percentage deviation of the previous year's temperature from the climate normal. This suggests that it may be difficult to separate the effect of long-term temperature from that of the short-term.

However, in figure A1 (see Appendix A), we plot the spatial distribution of short-term and long-term (decadal) temperature shocks at the endline and make two observations. First, districts that suffer stronger long-term shocks are not necessarily the ones that faced an unusually-hot year in 2010. Second, the decadal temperature shock is not systematically correlated with the intensity of road connectivity—the share of villages that received a PMGSY road by the endline—in the district.

# 5 Empirical Strategy

In the six states that complied with the guidelines, the PMGSY prioritized the assignment of new paved roads to villages that were eligible, i.e., whose baseline—as of the 2001 Population Census—population was above certain time-varying thresholds. Because very few villages with population close to 250 had received a PMGSY road by 2011, we focus on villages close to either the 1000 or 500-people threshold. We follow Asher and Novosad (2020) and match each village to the closest threshold. As such, villages with population between 200 and 750 are matched to the 500 threshold, while those with population between 750 and 1300 are matched to the 1000-people threshold.[15]

Although the PMGSY used sharp population thresholds, not all eligible villages had received a road by the time our outcome data was collected (2011-13, depending upon the census). Thus, eligibility for a road under the PMGSY is an imperfect predictor of having gotten a new road. This makes a fuzzy regression discontinuity design well-suited to estimating the local average treatment effect (LATE) of rural roads, which is identified by villages at (or near) the population threshold that were both eligible to receive a new road and did, i.e., are 'compliers' in the instrument variables (IV) parlance.

However, we are interested in whether the effect of rural roads, which runs contrary to that of (positive) temperature shocks, can counterveil the latter. This is akin to asking if the LATE of rural roads is moderated by villages' exposure to temperature shocks, and if so, then does the LATE persist even if temperature shocks are large. To answer this question, we turn to a fuzzy difference-in-discontinuity (RD-DID) design, which blends elements from regression discontinuity (RD) and difference-in-difference (DID) (Galindo-Silva et al., 2018; Takahashi, 2024). This boils down to interacting the village-level PMGSY road indicator with our district-level measure of temperature shocks, and adding the interaction as another to-be-instrumented variable in the fuzzy RD. As one would expect, the instrument for this interaction is an interaction of the road-eligibility indicator with the measure of temperature shocks. Thus, the second "difference" in our RD-DID is the difference between groups of villages that suffer different levels of

[15]This matching determines the value of the running variable for each village, which is defined as the nearest threshold minus the baseline village population.

temperature shocks.

## 5.1 Estimating Equations

Although most of our outcomes are observed over two rounds, some, most notably household wealth, are observed only post-treatment. As such, we use two variations of the fuzzy RD-DID. The first variation is used for outcomes where pre-treatment data is available, which we use to estimate a fuzzy RD-DID in *changes*, i.e., we first-difference both the outcome and the temperature shock measure. This variation corresponds to the following system of equations:

$$\begin{aligned}\Delta Road_{vd} = \alpha_0 &+ \alpha_1 Above_{vd,2001} + \alpha_2 \Delta Temp_d + \alpha_3 \Delta Temp_d \times Above_{vd,2001} \\ &+ \alpha_4 Left_{vd,2001} + \alpha_5 Right_{vd,2001} + \alpha_6 Left_{vd,2001} \times \Delta Temp_d \\ &+ \alpha_7 Right_{vd,2001} \times \Delta Temp_d + \mathbf{X}_{v,2001}\gamma + \zeta_{vd} \qquad (1)\\ \Delta Y_{vd} = \beta_0 &+ \beta_1 \Delta Road_{vd} + \beta_2 \Delta Temp_d + \beta_3 \Delta Road_{vd} \times \Delta Temp_d \\ &+ \alpha_4 Left_{vd,2001} + \alpha_5 Right_{vd,2001} + \alpha_6 Left_{vd,2001} \times \Delta Temp_d \\ &+ \alpha_7 Right_{vd,2001} \times \Delta Temp_d + \mathbf{X}_{v,2001}\psi + \epsilon_{vd} \qquad (2)\end{aligned}$$

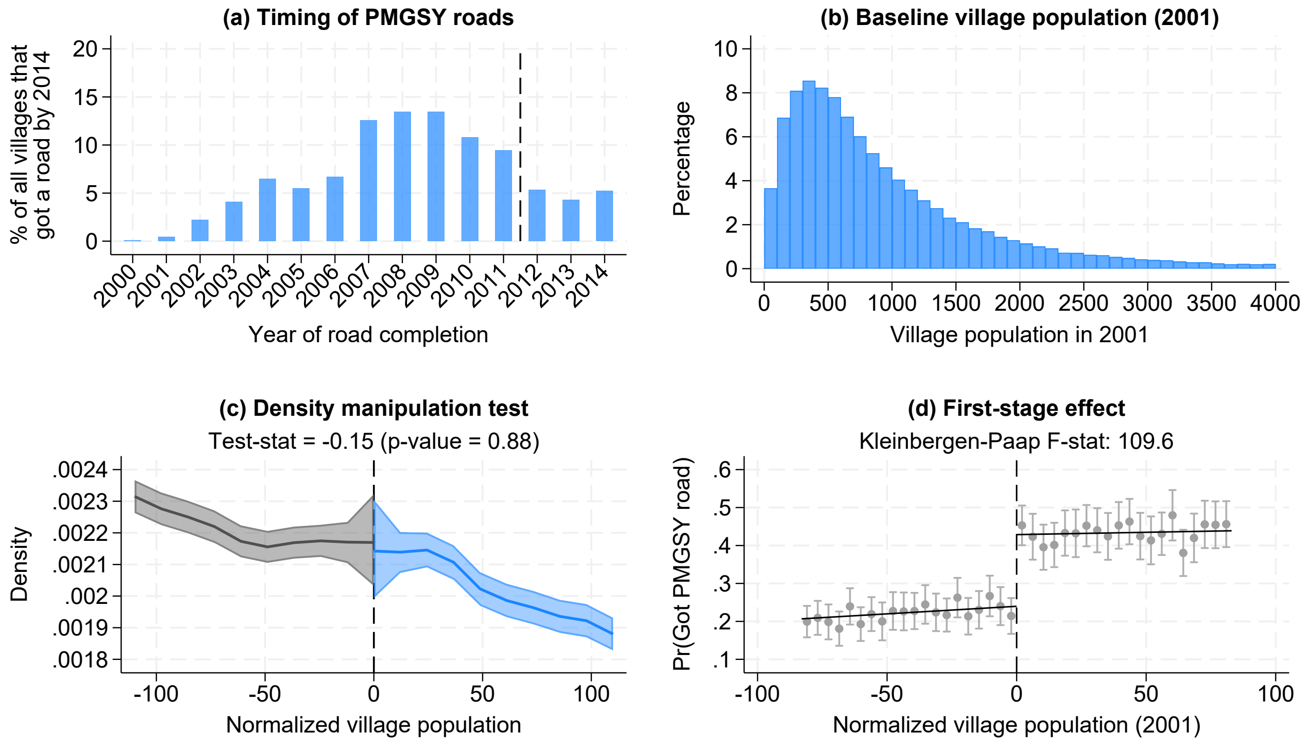
where equation (1) represents one of the two first-stage equations[16] and equation (2) is the second-stage. $\Delta Y_{vd}$ is the first-difference of the outcome for village $v$ in district $d$, i.e., the change from 2001 to 2011. $\Delta Road_{vd}$ is an indicator that takes value 1 if village $v$ in district $d$ received a new paved road between 2001 and 2011 (inclusive), and 0 otherwise. $\Delta Temp_d$ is the change in the decadal temperature shock in district $d$ between 2001 and 2011. $\beta_1$ and $\beta_3$ are the coefficients of interest. The former represents the marginal effect of roads for villages in a district with no temperature shocks, while the latter describes the change in this effect from a 1 percentage point increase in the mean decadal temperature.

$Above_{vd,2001}$ is an indicator that equals 1 when the village's baseline population ($Pop_{vd,2001}$) lay above the nearest PMGSY threshold ($c$), and 0 otherwise. $Left_{vd,2001} = (Pop_{vd,2001} - c)$ is the linear polynomial to the left of (below) the population threshold. Likewise, $Right_{vd,2001} = Above_{vd,2001} \times (Pop_{vd,2001} - c)$ represents the linear polynomial to the right of (above) the population threshold. The interaction terms $(Left \times \Delta Temp)_{vd}$ and $(Right \times \Delta Temp)_{vd}$ allow the effect of temperature to vary smoothly with population on either side of the threshold. The inclusion of these four terms is what allows $\beta_1$ and $\beta_5$ to identify the level effects on the outcome trend for villages at the population threshold.

$\mathbf{X}_{vd,2001}$ is a vector of village-level covariates, measured at the baseline. The covariates include indicators

[16]The other first-stage equation has $\Delta Road_{vd} \times \Delta Temp_d$ as the outcome variable and the same regressors as equation (1).

Figure 3: Validating the identification strategy




for the presence of a medical center, primary school, electrification, distance from the nearest town (in kilometers), the literacy rate, the share of households that belong to scheduled castes, and the share of households that earned over 250 INR per-month. This vector also includes the first-difference of the rainfall shock. Observe that the first-differencing means that the covariate vector is implicitly interacted with an endline (2011) indicator.

The second variation of the fuzzy RD-DID differs in two respects. First, both the outcome and temperature shock are expressed in *levels*, not changes. Compared to equations (1) and (2), we replace $\Delta Road_{vd}$ with $Road_{vd,2011}$ and $\Delta Temp_d$ with $Temp_{d,2011}$. Second, we include district-level fixed effects to partial-out the influence of time-invariant geographic and institutional confounders; this is already accomplished by the first-differencing in fuzzy RD-DID in changes.

Like Asher and Novosad (2020), we use a triangular kernel with a data-derived optimal bandwidth of 84, obtained from the procedure of Imbens and Kalyanaraman (2012). Standard errors are clustered at the village level to allow for serial correlation. Our results are robust to the use of standard errors that account for spatio-temporal dependence upto 200 KM (Conley, 1999). They are also robust to a change in the bandwidth (from 60 through 110) and the use of a rectangular kernel.

Table 1: Sample balance at baseline (2001)

| | Sample Mean (S.D.) | | | | RDD | |
|---|---|---|---|---|---|---|
| | No road | | Got road | | Estimate (S.E.) | |
| Primary school available | 0.95 | (0.21) | 0.97 | (0.18) | -0.0046 | (0.0078) |
| Medical center available | 0.15 | (0.35) | 0.18 | (0.39) | -0.016 | (0.015) |
| Electricity available | 0.43 | (0.50) | 0.43 | (0.50) | 0.024 | (0.020) |
| Distance from nearest town (in km) | 25.9 | (21.8) | 29.1 | (24.4) | -1.17 | (0.88) |
| Land irrigated (share of village area) | 0.28 | (0.29) | 0.28 | (0.29) | 0.0053 | (0.012) |
| Log(village land area) | 5.08 | (0.77) | 5.35 | (0.82) | 0.0039 | (0.032) |
| Literate households (share of total) | 0.46 | (0.16) | 0.45 | (0.15) | 0.0052 | (0.0063) |
| Scheduled Caste households (share of total) | 0.15 | (0.18) | 0.13 | (0.16) | 0.0033 | (0.0071) |
| Own farmland (share of households) | 0.72 | (0.24) | 0.79 | (0.21) | -0.0059 | (0.0096) |
| Land irrigated (share of village area) | 0.28 | (0.29) | 0.28 | (0.29) | 0.0053 | (0.012) |
| Subsistence farmers (share of households) | 0.41 | (0.26) | 0.50 | (0.27) | -0.0024 | (0.011) |
| Earn $\geq$ USD 4 per month (share of households) | 0.76 | (0.28) | 0.79 | (0.26) | -0.0071 | (0.011) |
| Per-capita workers in non-farm village firms (1998) | 0.034 | (0.11) | 0.025 | (0.052) | 0.00095 | (0.0032) |
| Main cultivators' share of employment | 0.42 | (0.23) | 0.48 | (0.24) | -0.010 | (0.0087) |
| Main services' share of employment | 0.092 | (0.12) | 0.088 | (0.11) | 0.0011 | (0.0047) |
| Main farm labour's share of employment | 0.16 | (0.18) | 0.12 | (0.16) | 0.010 | (0.0069) |
| Log(cultivated area) | 5.16 | (0.75) | 5.38 | (0.80) | -0.016 | (0.033) |
| Vegetation index (mean; 2000-01) | 0.11 | (0.93) | -0.28 | (1.05) | -0.048 | (0.040) |
| Villages | 7582 | | 3606 | | 11188 | |

**Notes**: This table reports baseline sample means and results of a regression discontinuity design (RDD) estimate of differences in these baseline means across villages that received a new PMGSY road (treated) and those without a new road (control). In the first two columns, the standard deviations are reported in parentheses. In the third column the standard errors (clustered at the village level) are reported in parentheses. All data comes from 2001, unless otherwise indicated.

## 5.2 Identifying Assumptions

Before proceeding to the results, it is worth establishing whether our setting satisfies the identifying assumptions of a fuzzy RD-DID. Given the absence of a confounding policy that takes effect at the same population threshold, there are only three assumptions that must be satisfied. The first assumption is monotonicity, i.e., the probability of receiving a road should increase discontinuously at the PMGSY's population threshold. Figure 3(d) shows that this is indeed true; the probability of receiving a new paved road rises by 21 percentage points at the threshold.

The second assumption is imperfect control, i.e., villages should not be able to control their baseline population once the PMGSY is announced. Validating this assumption boils down to testing for manipulation, i.e., testing if the density of the running variable—the difference between a village's baseline population and the nearest population threshold—is continuous at the population thresholds.

To show that there is no evidence of manipulation, we proceed in two steps. First, in Figure 3(b) we plot the density of baseline village population for all villages in the six states. Then, we use the (non-parametric) local polynomial density estimator of Cattaneo et al. (2020) with valid confidence bands coming from Cattaneo et al. (2021). Figure 3(c) plots the estimated density, alongside the confidence bands. Visually, there is no evidence of a discontinuity. To formally test the null hypothesis of a continuous density, Cattaneo et al. (2020) provide a test statistic that follows the Binomial distribution. In our setting, the test statistic equals -0.15, with a "$p$-value" of 0.88, i.e., we cannot reject the null.

Third, potential outcomes should be continuous at the population threshold.[17] Following Asher and Novosad (2020) and Garg et al. (2024), we test the plausibility of this assumption by checking if either the outcome variable(s) or the relevant covariates, measured at the baseline, vary discontinuously at the population thresholds. Table 1 provides the sample means at the baseline for both covariates and outcomes, and confirms that neither exhibit a discontinuity around the thresholds.

# 6 Results

## 6.1 Employment

Table 2: Impact of roads and temperature on employment

| | Employment | | Share in workforce | | |
|---|---|---|---|---|---|
| | Village | Village firms | Cultivation | Services | Farm labour |
| Decadal Temperature | -0.0051 | -0.044** | 0.015 | -0.011** | 0.031*** |
| | (0.0068) | (0.020) | (0.011) | (0.0049) | (0.0078) |
| | [0.0073] | [0.022] | [0.013] | [0.0049] | [0.010] |
| Decadal Rainfall | -0.0015*** | -0.0017** | 0.0011*** | -0.0011*** | 0.0018*** |
| | (0.00030) | (0.00083) | (0.00043) | (0.00024) | (0.00035) |
| | [0.00035] | [0.00091] | [0.00047] | [0.00030] | [0.00052] |
| Districts | 177 | 176 | 177 | 177 | 177 |
| Sample Mean (2011) | 0.49 | 0.066 | 0.34 | 0.12 | 0.22 |
| First-Differenced | Yes | Yes | Yes | Yes | Yes |
| Sample Period | 2001-2011 | 2001-2011 | 2001-2011 | 2001-2011 | 2001-2011 |

**Notes**: *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. | The standard errors in parentheses are clustered at the district level. The standard errors in square brackets are adjusted for spatio-temporal autocorrelation up to 200 KM (Conley, 1999; Collela et al., 2020) | Sample Period denote the Census years, such that 2001-2011 means that the outcome is observed in the 2001 and 2011 Censuses.

Table 2 presents the effect of long-term temperature shocks on both total employment and the sectoral distribution of workers. We construct the estimation sample by aggregating village-level employment data to the district, using the village population as a weight, and then first-differencing it. The result is a district-level cross-section, where each outcome corresponds to the growth rate between 2001 and 2011.

[17]For outcomes where we have multiple rounds of data, this assumption could be relaxed to require continuity in the first-difference (trends) of potential outcomes.

We additionally control for the share of villages in the district that had paved roads at the baseline, and the decadal temperature shock in 1991. Standard errors are clustered at the district level. For sector-wise employment we consider only main workers; this helps verify whether the findings of Liu et al. (2023) carry-over to our sample.

We find that temperature shocks affect neither overall employment nor main cultivators' share of total employment, with both (non-)effects being precisely estimated. The latter could reflect their occupational immobility due to frictions in land markets. However, a one percentage point decadal temperature shock induces a contraction in non-farm village firms that amounts to 11 jobs in a village with 1000 people. This shock also reduces the service sector's share of employment by 1.4 percentage points and raises farm labour's share of employment by 2.3 percentage points. Our results are consistent with Liu et al. (2023)'s assertion that by reducing agricultural incomes, temperature shocks drive-down local demand and, in the absence of alternatives, force people back into agricultural labour. In line with previous work, we also find that unlike temperature, rainfall shocks have much weaker effects on employment and the sectoral distribution of workers in India (Burgess et al., 2017; Colmer, 2021; Liu et al., 2023).

Table 3: Impact on employment; heterogeneity by temperature

| | Cultivation | | Services | | Farm labour | |
|---|---|---|---|---|---|---|
| | Main | Marginal | Main | Marginal | Main | Marginal |
| New road | 0.35** | 0.19* | 0.20** | -0.097 | -0.14 | -0.48** |
| | (0.18) | (0.11) | (0.098) | (0.077) | (0.17) | (0.21) |
| | [0.17] | [0.11] | [0.11] | [0.078] | [0.17] | [0.21] |
| New road x Temperature | -0.17* | -0.11** | -0.13** | 0.045 | 0.11 | 0.24** |
| | (0.096) | (0.057) | (0.055) | (0.042) | (0.093) | (0.11) |
| | [0.095] | [0.057] | [0.060] | [0.043] | [0.092] | [0.11] |
| Linear combination | 0.14** | 0.046 | 0.040 | -0.039 | 0.0066 | -0.18** |
| | (0.069) | (0.046) | (0.033) | (0.028) | (0.060) | (0.073) |
| | [0.066] | [0.043] | [0.035] | [0.027] | [0.060] | [0.074] |
| Villages | 11188 | 11188 | 11188 | 11188 | 11188 | 11188 |
| Control Mean (2011) | 0.34 | 0.076 | 0.097 | 0.044 | 0.20 | 0.22 |
| Control Trend (2001-11) | -0.087 | -0.021 | 0.0051 | 0.019 | 0.040 | 0.052 |
| Temperature Mean (2001-11) | 1.26 | 1.26 | 1.26 | 1.26 | 1.26 | 1.26 |
| First-Differenced | Yes | Yes | Yes | Yes | Yes | Yes |
| Sample Period | 2001-11 | 2001-11 | 2001-11 | 2001-11 | 2001-11 | 2001-11 |

**Notes**: *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. | The standard errors in parentheses are clustered at the village level. The standard errors in square brackets are adjusted for spatio-temporal autocorrelation up to 50 KM (Conley, 1999; Collela et al., 2020). | Sample Period denote the Census years, such that 2001-11 means that the outcome is observed in the 2001 and 2011 Censuses. *Temperature Mean (2001-11)* denotes mean of the first-differenced district-level temperature shock (i.e., the average change in the shock from 2001 to 2011) among villages that received a road. The *linear combination* is defined as the coefficient on *New road* plus the product of the coefficient on the interaction term and the mean temperature shock.

Next, we turn to the question of whether road connectivity can buffer the effects of temperature. We focus particularly on agriculture and services' share of employment, but, unlike in Table 2, we now distinguish the effect on main and marginal employment, which allows us to provide a richer understanding of the impact of roads. Table 3 presents the temperature-moderated effect of rural roads on sectoral employment shares for all sectors except manufacturing, which together represent 98% of all employment.

The first column suggests in the absence of a temperature shock, main cultivators' share of employment would be 35 percentage points higher. A likely explanation is that new roads reduced transport costs to nearby towns, thereby allowing cultivators to earn higher prices in those non-hyperlocal markets. This is plausible because rural roads improved the availability of public transport (Asher & Novosad, 2020) and over 85% of villages in the sample were not proximate to an agricultural market at the baseline.

However, a one percentage point temperature shock halves the effect of a new road. This is in line with previous findings that high temperature shocks depress agricultural yields and incomes (Dell et al., 2012), which creates an incentive for cultivators to sell their land and switch occupations. Moreover, the significant interaction term suggests that improved connectivity may have partially alleviated land market frictions, perhaps by increasing the non-farm-related demand for land or increasing the option value of agricultural land for non-cultivation uses. We discuss this in detail in the next section.

For concreteness, we also compute the marginal effect of roads for the 'average' treated village. First, we calculate the mean of the temperature shock in 2011 for villages that got a road. Then, we evaluate the linear combination of the two coefficients by multiplying the interaction term with the shock mean (1.26 percent). For this average village, the estimated effect on main cultivators' employment share was 14 percentage points relative to the control group. This amounts to a net increase of 4.3 percentage points, once we subtract the control group's outcome trend (an 8.7 percentage point reduction). The positive net effect indicates that access to non-hyperlocal agricultural markets can counteract the income effect of a temperature increase to some degree.

The effect on marginal cultivators' share is smaller and indistinguishable from a null effect once the temperature shock exceeds one percentage point. This disparity suggests that it is the primary earners of the household, who would almost always be classified as main workers, that respond more to changes in local connectivity. In a typical case of a nuclear household engaged in cultivation, one might expect the men be main workers while women and children to be marginal workers.[18] If so, then we should expect to a see a rise in main cultivators' share of employment for men, but not for women. In column (3) of Table A2, we show that this is indeed the case.

If road connectivity increases agricultural incomes, then it can stimulate local demand for non-tradables,

[18] The Population Census records them as workers even if they spend most of their time on other non-employment activities.

which can explain why main service workers' share of employment grows when the new roads are not accompanied by a temperature shock. However, this impact is fleeting: In any district that experienced greater than a one percentage point temperature shock, the effect of roads is statistically insignificant. When disaggregating the effect by gender in column (1) of Table A3, we find that employment in services rises only for women, but does so even when the temperature shock exceeds one percentage point.[19]

But that is not to say that only women exit farm labour. The sixth column of Table 3 demonstrates that the overall employment share of marginal farm labour declines by 18 percentage points even when the village is faced with the mean level of the temperature shock. This implies a net decrease of 12 percentage points after adding the control group's trend, which is a non-trivial effect. Table A3 complements these findings and shows that they apply to both men and women. Overall, the results indicate that road connectivity provides a sufficient buffer against the backsliding into farm labour that arises due to rising temperatures.

## 6.2 Agriculture

Table 4: Impact on other farm outcomes; heterogeneity by temperature

| | Land ownership | Mechanization | Commercial crops | Output |
|---|---|---|---|---|
| New road | 2.32** | -0.038 | 1.26 | -0.21 |
| | (1.07) | (0.65) | (1.14) | (0.35) |
| | [1.14] | [0.66] | [1.08] | [0.40] |
| New road x Temperature | -1.41** | 0.018 | -0.74 | 0.087 |
| | (0.68) | (0.42) | (0.69) | (0.19) |
| | [0.73] | [0.42] | [0.66] | [0.23] |
| Linear combination | 0.54** | -0.016 | 0.33 | -0.10 |
| | (0.26) | (0.16) | (0.28) | (0.13) |
| | [0.27] | [0.17] | [0.27] | [0.14] |
| Villages | 11184 | 11184 | 8080 | 11089 |
| Control Mean (2011) | -0.040 | 0.027 | 0.083 | 0.091 |
| Control Trend (2001-11) | | | | -0.015 |
| Temperature Mean (2011) | 1.26 | 1.26 | 1.26 | 1.26 |
| District FE | Yes | Yes | Yes | No |
| First-Differenced | No | No | No | Yes |
| Sample Period | 2011 | 2011 | 2011 | 2001-11 |

**Notes**: *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. | The standard errors in parentheses are clustered at the village level. The standard errors in square brackets are adjusted for spatio-temporal autocorrelation up to 50 KM (Conley, 1999; Collela et al., 2020). | Sample Period denote the Census years, such that 2001-11 means that the outcome is observed in the 2001 and 2011 Censuses. *Temperature Mean (2001-11)* denotes mean of the first-differenced district-level temperature shock (i.e., the average change in the shock from 2001 to 2011) among villages that received a road. The *linear combination* is defined as the coefficient on *New road* plus the product of the coefficient on the interaction term and the mean temperature shock.

[19]Given that women represent a minority of the rural workforce in our sample, any employment effects that are driven mostly by women are more likely to be masked in the aggregated results.

It is worth examining why road connectivity increased cultivators' share of employment, because the underlying mechanism can be informative about the associated welfare effects. An optimistic interpretation of this increase is that households that would otherwise have exited cultivation due to low profits instead choose to remain because road connectivity increases the market price of their produce, and thus, their welfare. A pessimistic interpretation is that because rural roads can raise land prices (Jacoby, 2000; Shrestha, 2020), they could exacerbate existing land market frictions (Foster & Rosenzweig, 2022) and prevent cultivators from selling their land and exiting agriculture. If so, households would be trapped in agriculture, which lowers their welfare relative to an equilibrium where they could switch occupation.

Absent granular data on land prices, we turn to land ownership in order to identify the more plausible interpretation. The first column of Table 4 shows that the land ownership index[20] rises in response to road connectivity, but less-so amid temperature shocks. This effect heterogeneity implies that households are less likely to retain land when agriculture becomes less profitable. Further, because temperature shocks affect agricultural productivity more than the non-farm sectors (Dell et al., 2012), they should reduce the land's option value when used for agriculture vis-a-vis other purposes. Thus, cultivators' incentive to sell their land becomes stronger relative to non-cultivators' incentive to purchase land. Because households respond to this by reducing farmland ownership, it is plausible that their decision to do otherwise in the absence of a temperature shock is due to agriculture's increased profitability, rather than prohibitive market frictions.

In contrast, we find that road connectivity had no effect on the ownership of mechanical farm equipment among households, i.e., there was no change in mechanization. Similarly, there is no visible effect on the adoption of commercial (non-food) crops, even though prior work suggests that improvements in market access tend to alter the crop mix (Allen & Atkin, 2022). One possibility is that road connectivity changes the composition of cultivators, and that different subgroups of cultivators make different (possibly opposite) choices in response to connectivity. This is plausible, especially because we observe no effect of roads on agricultural production—as proxied by the vegetation index—despite the presence of more cultivators.

To address the possibility of heterogenous effects, we rely on the household-level IHDS microdata to see how farm inputs respond to new roads. In lieu of our preferred difference-in-discontinuity design, which is infeasible on the IHDS due to the small number of villages near the population threshold, we use a difference-in-difference with a household-year as the unit of analysis. Given the threats to identification, we include both household and state-year fixed effects to partial-out time-invariant differences across households and the impact of concurrent state-specific policies, respectively.

[20]As described in the Section 2, this index comprises (i) the share of households that own farmland and (ii) the share of village area used for cultivation.

Table A1 presents the results from the IHDS data. As a sanity check, first we verify if the effect on farm ownership is qualitatively similar; this is indeed the case. We find that road connectivity induces members of land-owning households to spend more time working on the family farm, and that this effect is dampened by temperature shocks. This indicates that cultivators substitute hired labour with family members' time, rather than physical capital, when the supply of farm labour becomes limited due to road connectivity. We also find that cultivators' expenditure on farm inputs (seeds, fertilizer, equipment rents etc.) declines amid road connectivity, but less-so when temperature shocks occur. In conjunction with the observed changes in farm ownership rates, this suggests that rural roads slow down the rate of agglomeration in agriculture by incentivizing small-scale landholders to continue with cultivation, thereby changing the composition of cultivators.

### 6.3 Welfare

Much of our prior discussion hinges on the assumption that better roads raise rural incomes. This is concerning because Asher and Novosad (2020) found, using the same sample as ours, that new roads did not raise either consumption or asset ownership overall. But it is possible that the welfare improvements of road connectivity were counterbalanced by temperature shocks. Before we can test this claim, we must first rule out differential outmigration from villages that experienced temperature shocks. After all, if road connectivity led the worst-affected households to migrate away, and if this did not happen in unconnected villages, then households in the former group would be positively-selected.

In Table 5 we first determine what roads and temperature do to the composition of village population. The first column shows that population growth was not dissimilar between treated and control villages, even when temperature shocks are accounted for. In Table A4, we show that there was no change in the age distribution, either.[21] There was also no tangible differences in the share of kids aged 0-6 years, which is indicative of whether young families moved out. Similarly, there is no evidence of systematic migration by gender or caste groups. We can now discuss the effects on household welfare. In the absence of a temperature shock, new roads could have increased the wealth (asset ownership) index by nearly 1.5 standard deviations. This effect remains large (0.6 standard deviations) and statistically significant ($p = 0.078$) for a temperature shock as large as 0.9 percentage points, i.e., until the 25th percentile of the shock distribution. Conversely, this implies that 75% of villages with newfound road connectivity lay in districts that experienced temperature shocks that were large enough to render the welfare effects indistinguishable from zero. As such, it is unsurprising that the treatment effect for the average village

[21] But this is not because temperature shocks, by themselves, have no effect on migration. In Table A5, we find that districts that experienced higher temperature shocks had their age distribution skewed leftwards, i.e., adults, particularly those older than 30 years, made up a smaller share of the population.

Table 5: Impact on population and welfare; heterogeneity by temperature

| | Total | Population share | | | Asset Index |
|---|---|---|---|---|---|
| | Log(Population) | Kids (0-6) years | Male | SC or ST | |
| New road | -0.060 | 0.012 | 0.023 | 0.034 | 1.48** |
| | (0.11) | (0.023) | (0.014) | (0.060) | (0.68) |
| | [0.11] | [0.023] | [0.014] | [0.063] | [0.67] |
| New road x Temperature | 0.011 | -0.015 | -0.015* | -0.023 | -1.09** |
| | (0.059) | (0.012) | (0.0076) | (0.034) | (0.44) |
| | [0.061] | [0.013] | [0.0076] | [0.037] | [0.43] |
| Linear combination | -0.047 | -0.0075 | 0.0044 | 0.0054 | 0.32 |
| | (0.040) | (0.0089) | (0.0052) | (0.020) | (0.24) |
| | [0.039] | [0.0092] | [0.0051] | [0.019] | [0.24] |
| Villages | 11188 | 11188 | 11188 | 11188 | 11184 |
| Control Mean (2011) | 6.44 | 0.15 | 0.51 | 0.46 | -0.024 |
| Temperature Mean (2001-11) | 1.26 | 1.26 | 1.26 | 1.26 | 1.07 |
| District FE | No | No | No | No | Yes |
| First-Differenced | Yes | Yes | Yes | Yes | No |
| Sample Period | 2001-11 | 2001-11 | 2001-11 | 2001-11 | 2011 |

**Notes**: *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. | The standard errors in parentheses are clustered at the village level. The standard errors in square brackets are adjusted for spatio-temporal autocorrelation up to 50 KM (Conley, 1999; Collela et al., 2020). | Sample Period denote the Census years, such that 2001-11 means that the outcome is observed in the 2001 and 2011 Censuses. *Temperature Mean (2001-11)* denotes mean of the first-differenced district-level temperature shock (i.e., the average change in the shock from 2001 to 2011) among villages that received a road. The *linear combination* is defined as the coefficient on *New road* plus the product of the coefficient on the interaction term and the mean temperature shock.

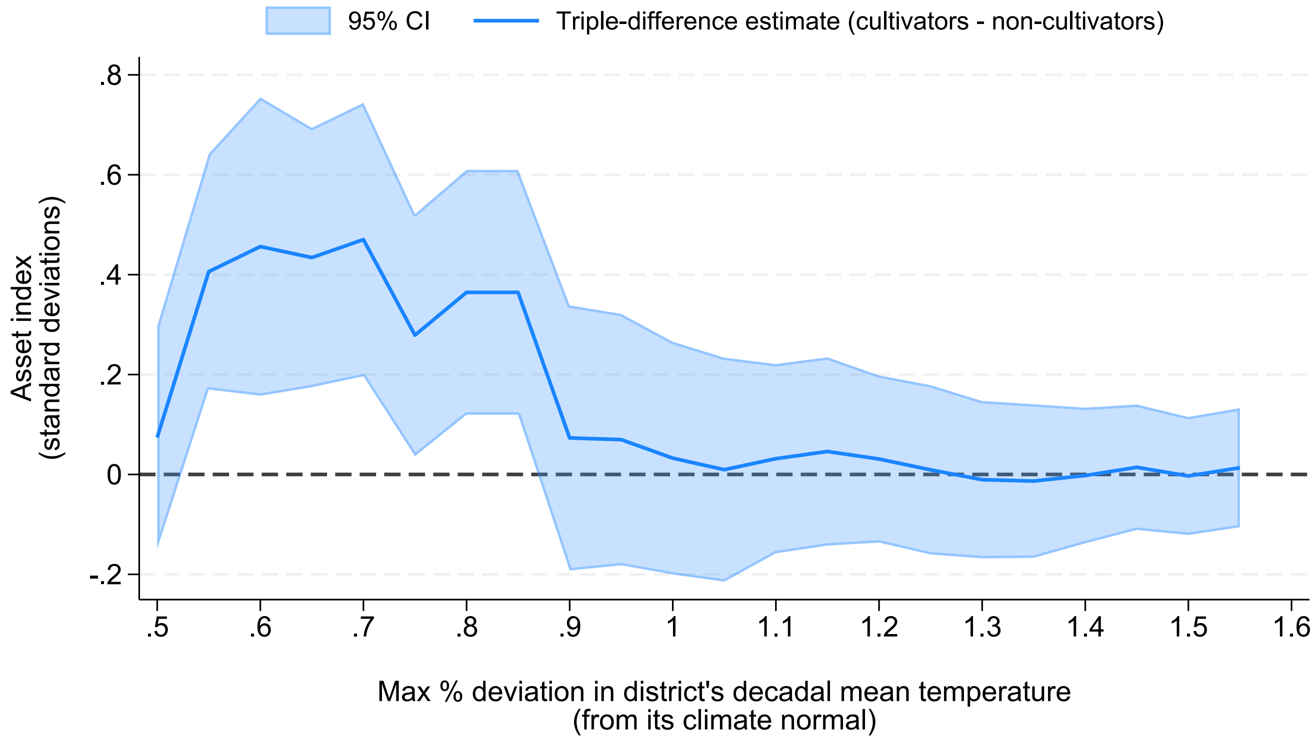
is small and statistically insignificant, as is evident both from the linear combination and the results of Asher and Novosad (2020).

The household wealth index is also well-suited to shed light on whether our optimistic or pessimistic interpretation of the effect of roads is more likely. Suppose that the optimistic interpretation is true, and cultivators remain in agriculture because it is more profitable. If so, then road connectivity should have at least as large an effect on income and wealth for cultivators as non-cultivator households. Therefore, if we compare the difference in trends between cultivators and non-cultivators across treated and control villages, then the gap between cultivators and non-cultivators should either grow or remain unchanged in the treated villages, relative to the control.[22] If it did shrink, then it is more likely that cultivators did not exit agriculture because their lack of occupational mobility was exacerbated by roads. Formally, we would like to test:

$$\Delta Y_{hvd} = \phi_0 Cultivator_h + \phi_1 \Delta Road_{vd} + \phi_2(\Delta Road_{vd} \times Cultivator_h) + \mathbf{X}_{vd}\nu + \gamma_s + \epsilon_{hvd} \quad (3)$$

[22] This hypothesis assumes that cultivators, who are predominantly landowners, are *ex ante* wealthier on average, which we find to be the case in the IHDS.

Figure 4: Differential effect of road connectivity on cultivators' relative welfare




where $h$ indexes household, $v$ indexes the village, and $d$ indexes the district. As before, $\Delta Y_{hvd}$ denotes the first-difference of the outcome. $\Delta Road_{vd}$ denotes the change in treatment status, and $Cultivator_h$ is an indicator of whether the household is engaged primarily in cultivation. $\gamma_s$ is a state-year fixed effect, accounting for state-level policy changes that could target specific groups. $\mathbf{X}_{hvd}$ is a vector of baseline household controls and time-varying village controls. Since outcomes are likely correlated within a village, the standard errors would be clustered at the village level.

However, because the effect of roads on income ($\phi_1$ and $\phi_2$) are strongly-moderated by temperature, we would expect any growth in the gap ($\phi_2$) to be limited to regions that experienced small temperature shocks. This motivates a quadruple-difference-in-difference, which is perhaps not an elegant solution. Instead, we test the triple-difference specification hinted above, but on a sample limited to districts that experienced temperature shocks below a certain limit. Then, we can incrementally expand the sample by raising the limit on the temperature shock and re-estimate the equation. If the gap ($\phi_2$) is affected by temperature, then it should be significant only in the early subsamples that exclude districts with large temperature shocks.

In Figure 4, we plot the results of this exercise. We find that cultivators benefit relatively more from roads when temperature shocks are under 0.8 percentage points, but this "cultivator-premium" disappears amid exposure to higher temperature shocks. However, on average, cultivators are never left worse-off than non-cultivators, which suggests that they remain in agriculture not because of immobility, but incentives.

### 6.4 Robustness

To gauge the robustness of our results, we begin by varying the bandwidth parameter and using a rectangular kernel. Akin to Asher and Novosad (2020), we choose three evenly-spaced bandwidths near the optimal (84), namely 70, 90, and 110. In Figure A3, we show that our results are qualitatively similar for all three bandwidths when using a triangular kernel, although the effect on main cultivators' share is only significant at the 10% level for the smallest bandwidth. All results are stronger when a rectangular kernel is used, regardless of the choice of bandwidth.

Another concern is that the outcome density in the neighbourhood of the population threshold may be nonlinear, which could be mischaracterized by a linear polynomial as a discontinuity. To address this, it is informative to visualise the outcome density along the running variable. But recall that equation (3) involves an interaction between the road dummy the district-level temperature shock, which means the effect of a new road road on the outcome varies over temperature. To visualise this effect with a quadratic local polynomial, we would require a three-dimensional plot, where the running variable and temperature would represent separate axes.[23] As one might suspect, this is not quite feasible.

In lieu of visualising both the level effect of roads and the interaction term, we visualise the level effect from a model that omits all interaction terms, i.e., one that estimates the LATE. Because the main concern is that there are false positives, we limit attention to all outcomes where the LATE is either statistically significant for the full sample or for a subset of the sample that did not experience very large temperature shocks. The latter group includes main services' share of employment and the asset ownership index, and any effect of a road on them is estimated using only districts where the temperature shock ($\Delta Temp_d$) was under 1 percentage point (the median). In Figures A4 and A5, we show that the degree of the local polynomial does not change our results.[24]

An additional concern with the use of interaction terms in an RD design is the imposition of a parametric form for heterogenous effects, which means the RD-DID design does not offer the same robustness (to misspecification) guarantees of the usual non-parametric RD. Hsu and Shen (2019) show that if the policy's (rural roads) true effect is zero and the model is misspecified, then the interaction term has a tendency to over-reject the null hypothesis of homogenous effects. In our setting, this concern extends to key

[23] In that case, the temperature-moderated effect of a new road could be expressed as a curve in the outcome-temperature space. However, it is unclear how one might illustrate the three-dimensional equivalent of a confidence interval.

[24] All level effects that were statistically insignificant with a linear polynomial also remain as such when a quadratic polynomial is used.

outcomes where the LATE is insignificant for the full sample, but the coefficients suggest the presence of heterogenous treatment effects. As such, in Figure A5, we use the test of Hsu and Shen (2019) to determine whether there truly are heterogenous effects. We find strong evidence against the null hypothesis of homogenous effects. Thus, our results do not conflate a zero effect with heterogenous treatment effects.[25]

Next, we adopt the placebo exercise of Asher and Novosad (2020) to test the hypothesis that our results may be incorrectly attributing the underlying differences in outcomes near the population threshold to road connectivity. Specifically, we test the reduced-form effect of being above the PMGSY-specified population threshold on the probability of receiving a new road and on outcomes where the LATE is non-zero, just as above (with the same restrictions on the temperature shock). We compare the reduced-form effect for our main sample and a "placebo sample" where program guidelines would be irrelevant, i.e., where the first-stage effect would be zero. Our placebo sample consists of villages in two groups: (i) those with nearly 1000 people in states that only used lower thresholds, i.e., Gujarat, Maharashtra, Odisha, and Rajasthan (ii) those with with nearly 1000 people in seven non-complier states, namely Andhra Pradesh, Assam, Bihar, Jharkhand, Karnataka, Uttar Pradesh, and Uttarakhand. Table A3 presents the results of this exercise. Reassuringly, we find no first-stage effect or statistically significant reduced-form effects on the placebo sample, unlike the main sample.

# 7 Conclusion

Broadly, climate change remains a first-order problem for policymakers today, as exemplified by the fact that air surface temperatures have risen faster since 1970 than in any 50-year period in the previous two millenia (Lee et al., 2024). Much of the related discussion is rightfully centred-around mitigation strategies to limit the adverse effects on productivity (Taraz, 2018; Somanathan et al., 2021; Liu et al., 2023), welfare (Dell et al., 2012), and civil conflict (Burke et al., 2015), among others. In this project, we ask whether rural road connectivity can preserve the gains from structural transformation in emerging markets in the face of long-term temperature shocks.

We find that road connectivity can successfully counteract the effects of long-term temperature shocks that amount to one percentage point or under. For instance, villages with a new road, even in districts with a one percentage point temperature shock, exhibited a 14 percentage point rise in cultivators' share and an 18 percentage point decline in farm labour's share, with no adverse effects on the service sector's share. Moreover, the increase in cultivation does not appear to be driven by occupational immobility and land

[25] The only exception is marginal cultivators' employment share, where we cannot reject the null hypothesis of a homogenous zero effect.

market frictions. Rather, the evidence suggests that cultivators remain in farming because it becomes more profitable. We also demonstrate that in the absence of temperature shocks, new roads lead to greater asset ownership, despite no evidence of increased agricultural investments. Though these effects are attenuated by temperature shocks, they are sufficiently large to counteract the deleterious impacts of temperature on the average treated village.

Our results demonstrate the potential of last-mile road connectivity as a tool for policymakers in the emerging markets. Place-based policies, such as infrastructure, can be effective due to the spatially-uneven effects of rising temperatures, which place a disproportionate burden on rural regions that are mostly reliant on agriculture. Amid these sustained pressures, reduced transport frictions and increased market access may be a potential pathway to alleviating some of the burden from those that are most-vulnerable in the context of climate change.

## References

Adukia, A., Asher, S., & Novosad, P. (2020). Educational investment responses to economic opportunity: evidence from indian road construction. *American Economic Journal: Applied Economics*, *12*(1), 348–76.

Aggarwal, S. (2018). Do rural roads create pathways out of poverty? evidence from india. *Journal of Development Economics*, *133*, 375-395. doi: https://doi.org/10.1016/j.jdeveco.2018.01.004

Aggarwal, S. (2021). The long road to health: Healthcare utilization impacts of a road pavement policy in rural india. *Journal of Development Economics*, *151*, 102667.

Allen, T., & Atkin, D. (2022). Volatility and the gains from trade. *Econometrica*, *90*(5), 2053–2092.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, *103*(484), 1481–1495.

Arguez, A., & Vose, R. S. (2011). The definition of the standard wmo climate normal: The key to deriving alternative climate normals. *Bulletin of the American Meteorological Society*, *92*(6), 699–704.

Asher, S., Garg, T., & Novosad, P. (2020). The ecological impact of transportation infrastructure. *The Economic Journal*, *130*(629), 1173–1199.

Asher, S., & Novosad, P. (2020). Rural roads and local economic development. *American economic review*, *110*(3), 797–823.

Banerjee, A., Duflo, E., & Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in china. *Journal of Development Economics*, *145*, 102442.

Bilal, A., & Känzig, D. R. (2024). The macroeconomic impact of climate change: Global vs. local temperature. *Unpublished*.

Blimpo, M. P., Harding, R., & Wantchekon, L. (2013). Public investment in rural infrastructure: Some political economy considerations. *Journal of African Economies*, *22*(suppl_2), ii57–ii83.

Burgess, R., Deschenes, O., Donaldson, D., & Greenstone, M. (2017). Weather, climate change and death in india. *Unpublished*, 577–617.

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., & Padró i Miquel, G. (2015). The value of democracy: evidence from road building in kenya. *American Economic Review*, *105*(6), 1817–1851.

Burke, M., Hsiang, S. M., & Miguel, E. (2015). Climate and conflict. *Annual Review of Economics*, *7*(1), 577–617.

Castaing Gachassin, M. (2013). Should i stay or should i go? the role of roads in migration decisions. *Journal of African Economies*, *22*(5), 796–826.

Cattaneo, M. D., Jansson, M., & Ma, X. (2020). Simple local polynomial density estimators. *Journal of the*

*American Statistical Association*, *115*(531), 1449–1455.

Cattaneo, M. D., Jansson, M., & Ma, X. (2021). Local regression distribution estimators. *Journal of econometrics*, 105074.

Colmer, J. (2021). Temperature, labor reallocation, and industrial production: Evidence from india. *American Economic Journal: Applied Economics*, *13*(4), 101–124.

Colmer, J., Martin, R., Muûls, M., & Wagner, U. J. (2024). Does pricing carbon mitigate climate change? firm-level evidence from the european union emissions trading system. *Review of Economic Studies*, rdae055.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of econometrics*, *92*(1), 1–45.

Dasgupta, A., Karandikar, A., & Raghav, D. (2024). Road access, fertility, and child health in rural india. *Population and Development Review*.

Day, E., Fankhauser, S., Kingsmill, N., Costa, H., & Mavrogianni, A. (2019). Upholding labour productivity under climate change: an assessment of adaptation options. *Climate policy*, *19*(3), 367–385.

Dell, M., Jones, B. F., & Olken, B. A. (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics*, *4*(3), 66–95.

Dercon, S., Gilligan, D. O., Hoddinott, J., & Woldehanna, T. (2009). The impact of agricultural extension and roads on poverty and consumption growth in fifteen ethiopian villages. *American Journal of Agricultural Economics*, *91*(4), 1007–1021.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, *108*(4-5), 899–934.

Donaldson, D., & Hornbeck, R. (2016). Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, *131*(2), 799–858.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from china's national trunk highway system. *Review of Economic Studies*, *81*(3), 1046–1070.

Fafchamps, M., & Shilpi, F. (2013). Determinants of the choice of migration destination. *Oxford Bulletin of Economics and Statistics*, *75*(3), 388–409.

Foster, A. D., & Gehrke, E. (2017). Start what you finish! ex ante risk and schooling investments in the presence of dynamic complementarities. *Unpublished*.

Foster, A. D., & Rosenzweig, M. R. (2022). Are there too many farms in the world? labor market transaction costs, machine capacities, and optimal farm size. *Journal of Political Economy*, *130*(3), 636–680.

Galindo-Silva, H., Some, N. H., & Tchuente, G. (2018). Fuzzy difference-in-discontinuities: Identification theory and application to the affordable care act. *arXiv preprint arXiv:1812.06537*.

Garg, T., Jagnani, M., & Pullabhotla, H. K. (2024). Rural roads, farm labor exits, and crop fires. *American Economic Journal: Economic Policy*, *16*(3), 420–450.

Gibson, J., & Olivia, S. (2010). The effect of infrastructure access and quality on non-farm enterprises in rural indonesia. *World Development*, *38*(5), 717–726.

Gibson, J., & Rozelle, S. (2003). Poverty and access to roads in papua new guinea. *Economic development and cultural change*, *52*(1), 159–185.

Gleixner, S., Demissie, T., & Diro, G. T. (2020). Did era5 improve temperature and precipitation reanalysis over east africa? *Atmosphere*, *11*(9), 996.

Gollin, D., & Rogerson, R. (2014). Productivity, transport costs and subsistence agriculture. *Journal of Development Economics*, *107*, 38–48.

Gollin, D., & Wolfersberger, J. (2023). Agricultural trade and deforestation: the role of new roads. *Unpublished*.

Greenstone, M., & Hanna, R. (2014). Environmental regulations, air and water pollution, and infant mortality in india. *American Economic Review*, *104*(10), 3038–3072.

Hsu, Y.-C., & Shen, S. (2019). Testing treatment effect heterogeneity in regression discontinuity designs. *Journal of Econometrics*, *208*(2), 468–486.

Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of economic studies*, *79*(3), 933–959.

Jacoby, H. G. (2000). Access to markets and the benefits of rural roads. *The economic journal*, *110*(465), 713–737.

Khandker, S. R., Bakht, Z., & Koolwal., G. B. (2009). The poverty impact of rural roads: Evidence from bangladesh. *Economic Development and Cultural Change*, *57*(4), 685-722.

Khandker, S. R., & Koolwal, G. B. (2011). Estimating the long-term impacts of rural roads: a dynamic panel approach. *World Bank Policy Research Working Paper*(5867).

Krugman, P. (1991). Increasing returns and economic geography. *Journal of political economy*, *99*(3), 483–499.

Lagakos, D., Mobarak, A. M., & Waugh, M. E. (2023). The welfare effects of encouraging rural–urban migration. *Econometrica*, *91*(3), 803–837.

Lee, H., Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P., … Ruane, A. C. (2024). Climate change 2023 synthesis report summary for policymakers. *CLIMATE CHANGE 2023 Synthesis Report: Summary for Policymakers.*

Liu, M., Shamdasani, Y., & 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.

LoPalo, M. (2023). Temperature, worker productivity, and adaptation: evidence from survey data production. *American Economic Journal: Applied Economics*, *15*(1), 192–229.

Matsuyama, K. (1992). Agricultural productivity, comparative advantage, and economic growth. *Journal of economic theory*, *58*(2), 317–334.

Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, *90*(4), 683–701.

Morten, M., & Oliveira, J. (2023). The effects of roads on trade and migration: Evidence from a planned capital city. *American Economic Journal: Applied Economics*.

Mu, R., & Van de Walle, D. (2011). Rural roads and local market development in vietnam. *The Journal of Development Studies*, *47*(5), 709–734.

Nordhaus, W. (2019). Climate change: The ultimate challenge for economics. *American Economic Review*, *109*(6), 1991–2014.

Raghav, D. (2024). *Boulevard of dreams: Rural roads and temporary migration.* (Unpublished)

Shamdasani, Y. (2021). Rural road infrastructure & agricultural production: Evidence from india. *Journal of Development Economics*, *152*, 102686.

Shayegh, S., Manoussi, V., & Dasgupta, S. (2021). Climate change and development in south africa: the impact of rising temperatures on economic productivity and labour availability. *Climate and Development*, *13*(8), 725–735.

Shrestha, S. A. (2020). Roads, participation in markets, and benefits to agricultural households: Evidence from the topography-based highway network in nepal. *Economic Development and Cultural Change*, *68*(3), 839–864.

Somanathan, E., Somanathan, R., Sudarshan, A., & Tewari, M. (2021). The impact of temperature on productivity and labor supply: Evidence from indian manufacturing. *Journal of Political Economy*, *129*(6), 1797–1827.

Takahashi, M. (2024). Difference-in-discontinuities design. *Unpublished*.

Taraz, V. (2018). Can farmers adapt to higher temperatures? evidence from india. *World Development*, *112*, 205–219.

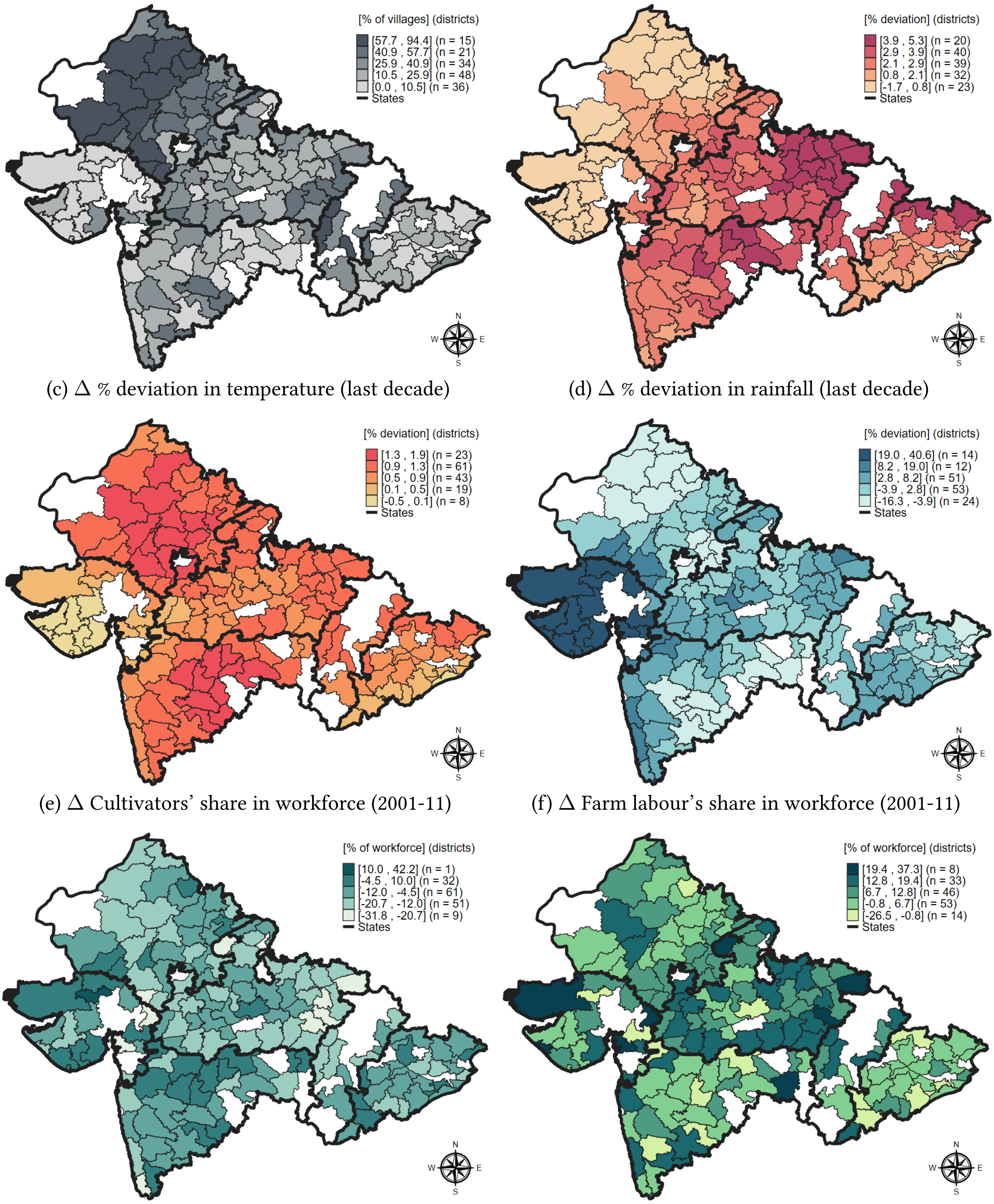
Vogel, E., Donat, M. G., Alexander, L. V., Meinshausen, M., Ray, D. K., Karoly, D., … Frieler, K. (2019). The effects of climate extremes on global agricultural yields. *Environmental Research Letters*, *14*(5), 054010.

Wilmott, C. J., & Matsuura, K. (2018). *Terrestrial air temperature and precipitation: Monthly time series (1900 - 2017).* Retrieved from http://climate.geog.udel.edu/html_pages/README.ghcn_ts2.html.

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., … others (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of sciences*, *114*(35), 9326–9331.

# Appendix A

Figure A1: Roads, weather, and agricultural work across space




(a) Villages (in sample) with PMGSY roads

(b) Δ % deviation in temperature (last year)

(c) Δ % deviation in temperature (last decade)

(d) Δ % deviation in rainfall (last decade)

(e) Δ Cultivators' share in workforce (2001-11)

(f) Δ Farm labour's share in workforce (2001-11)

Figure A2: Cultivators' share in the SECC versus Population Census (binned scatterplot)

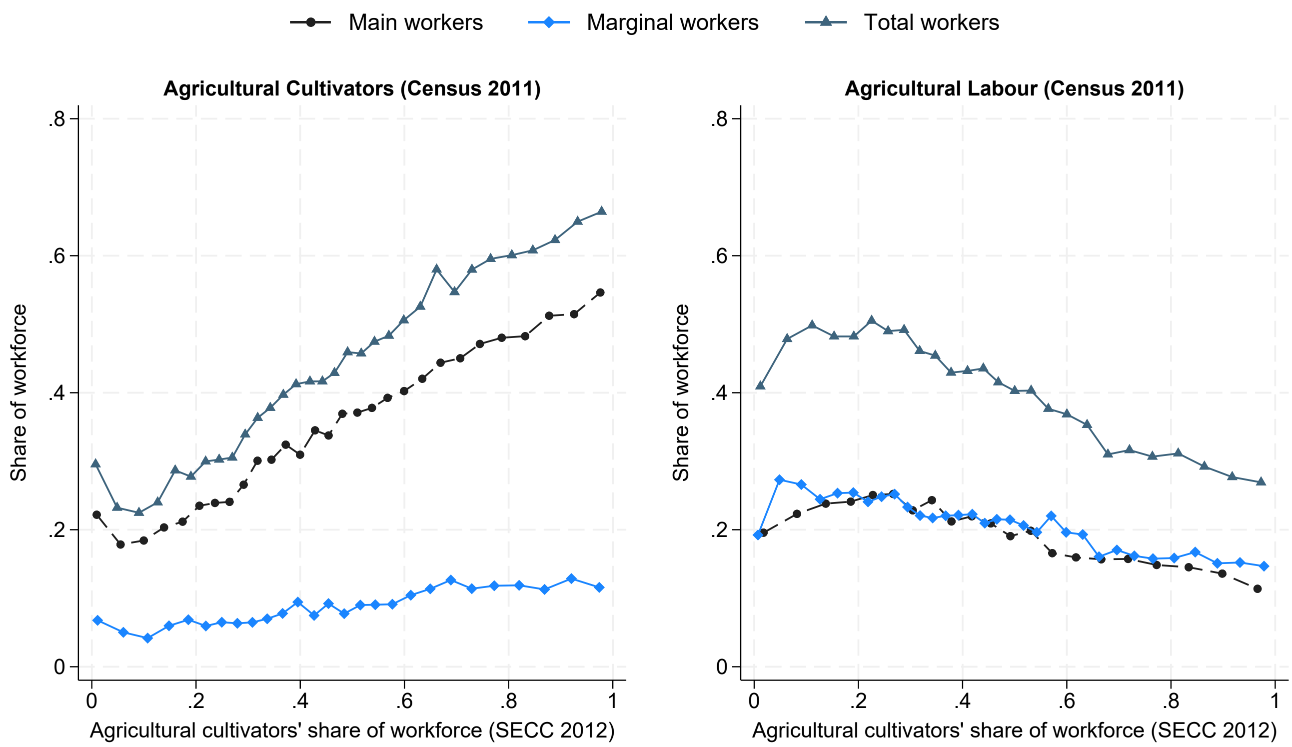


Figure A3: Robustness to alternate bandwidths and kernel

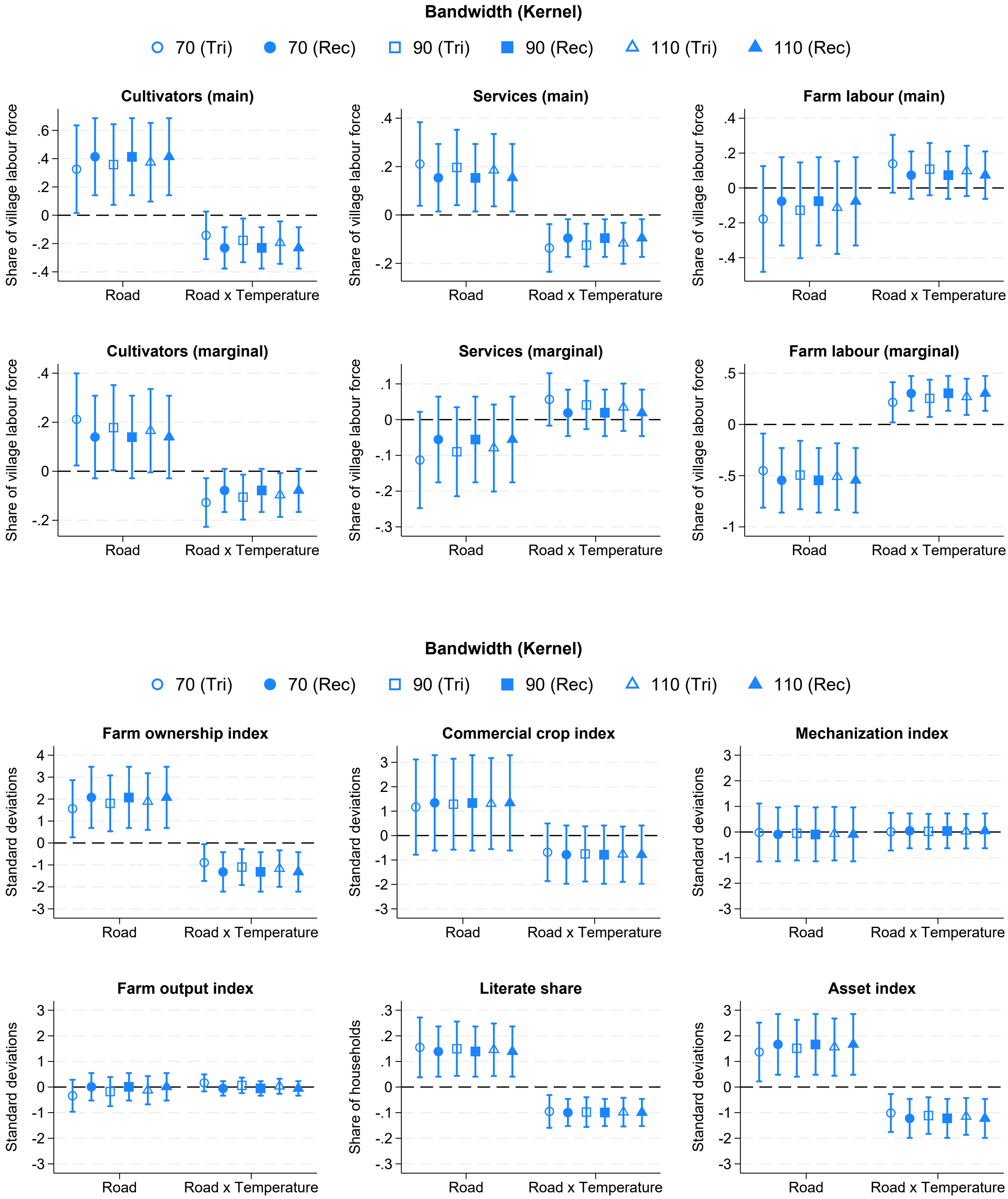


Figure A4: Robustness to order of local polynomial (linear)

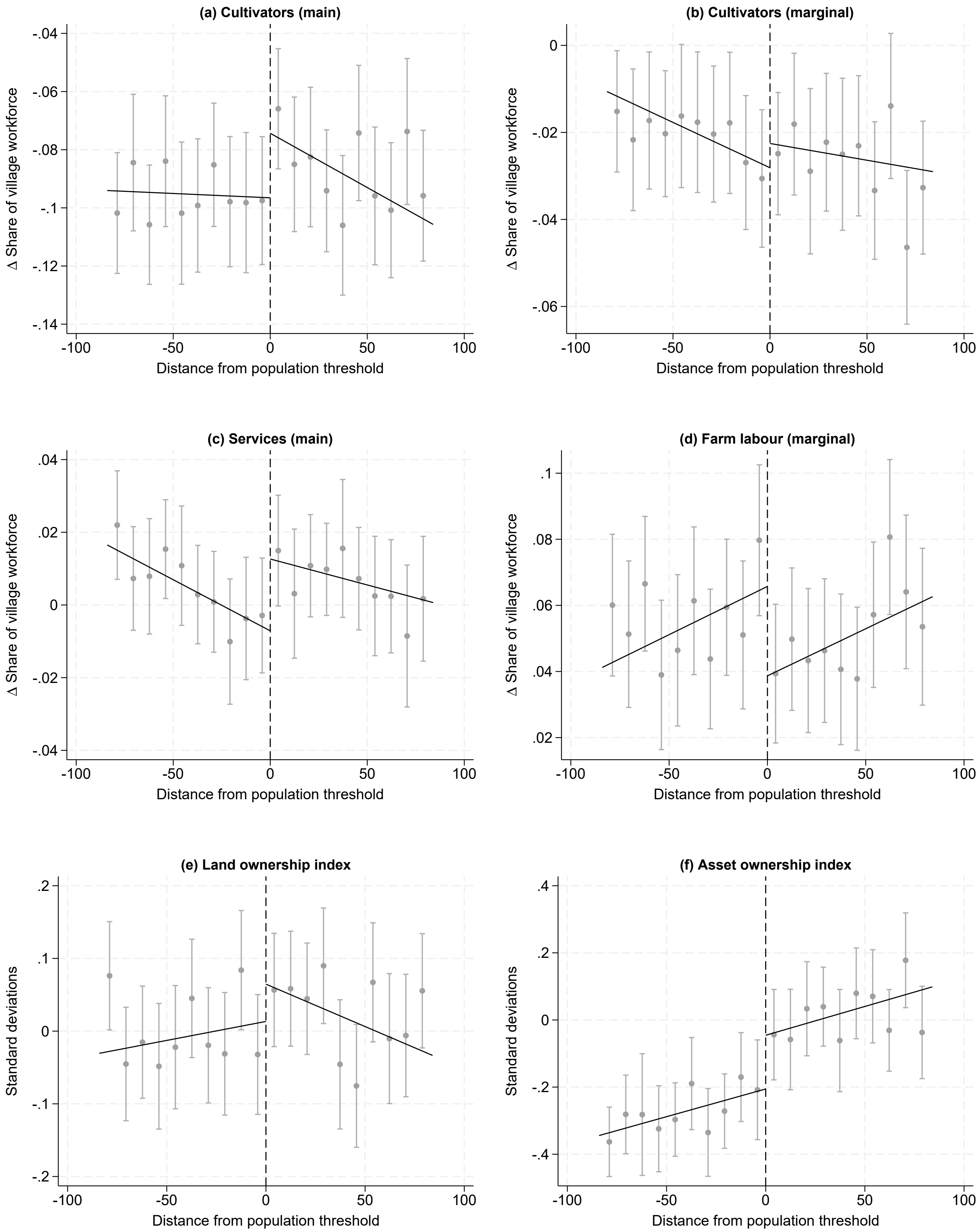


Figure A5: Robustness to order of local polynomial (quadratic)

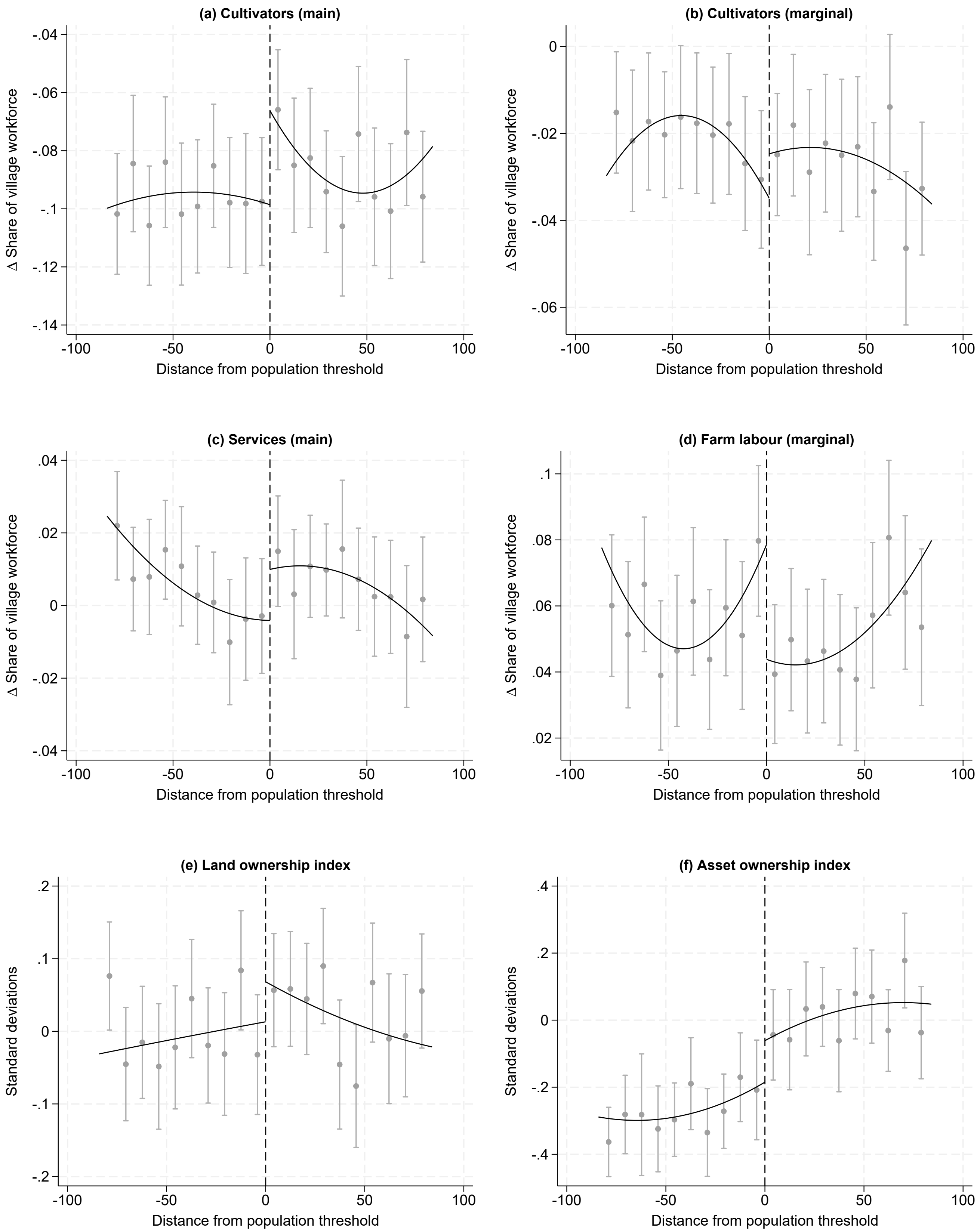


Figure A6: Testing for heterogenous treatment effects of rural roads (Hsu & Shen, 2019)

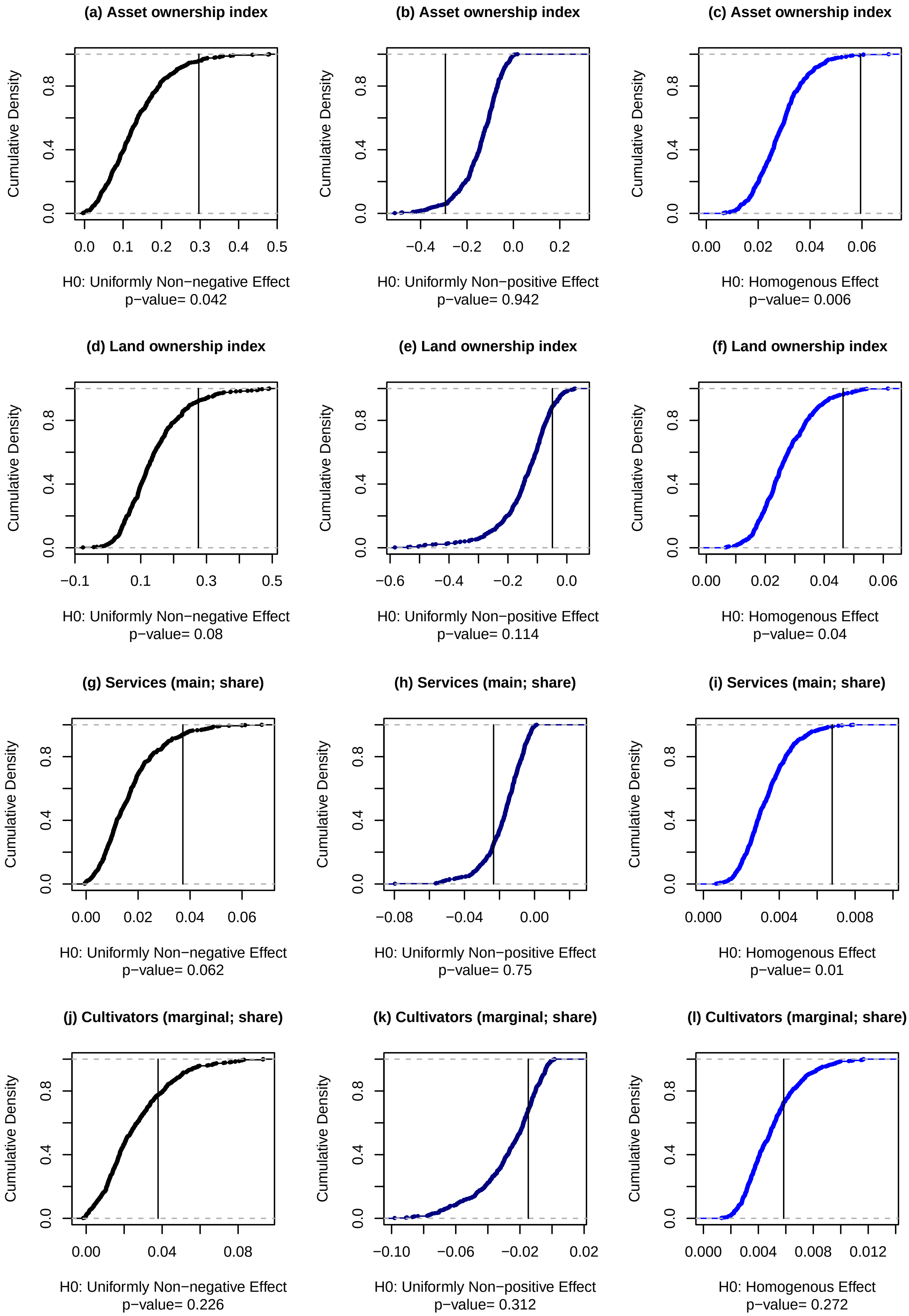


Table A1: Impact on household occupation and welfare; heterogeneity by temperature

| | Household farm | | | Migration | Welfare |
|---|---|---|---|---|---|
| | Have any? | Work on it? | Input Spending | Non-Resident | Assets |
| New Road | 0.19** | 0.27** | -1.05** | 0.10 | 0.42** |
| | (0.091) | (0.12) | (0.41) | (0.11) | (0.21) |
| New Road × Temperature | -0.19** | -0.25** | 0.73** | -0.096 | -0.37** |
| | (0.082) | (0.11) | (0.34) | (0.10) | (0.18) |
| Linear Combination | -0.021 | 0.0089 | -0.27** | -0.0012 | 0.019 |
| | (0.032) | (0.036) | (0.11) | (0.034) | (0.052) |
| Observations | 6664 | 16371 | 4344 | 6664 | 6662 |
| Control Mean | 0.80 | 0.66 | -0.043 | 0.23 | -0.60 |
| Temperature Mean (2001-11) | 1.07 | 1.05 | 1.07 | 1.07 | 1.07 |
| Unit of Analysis | Household | Person | Household | Household | Household |
| Village FE | Yes | Yes | Yes | Yes | Yes |
| State-Year FE | 2005-12 | 2005-12 | 2005-12 | 2005-12 | 2005-12 |

**Notes**: *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. | Standard errors are clustered at the district level. | The sample is a balanced panel of adults aged 18-60 in person-level regressions and households in household-level regressions.

Table A2: Farm employment; heterogeneity by temperature and gender

| **Panel (a): Men** | Services | Cultivation | | Farm labour | |
|---|---|---|---|---|---|
| | Main | Main | Marginal | Main | Marginal |
| New road | 0.13 | 0.40** | -0.015 | -0.025 | -0.39** |
| | (0.097) | (0.18) | (0.081) | (0.16) | (0.17) |
| New road x Temperature | -0.072 | -0.21** | 0.0089 | 0.048 | 0.19** |
| | (0.051) | (0.094) | (0.040) | (0.085) | (0.086) |
| Linear combination | 0.039 | 0.13* | -0.0037 | 0.035 | -0.15** |
| | (0.037) | (0.071) | (0.034) | (0.061) | (0.062) |
| Villages | 11188 | 11188 | 11188 | 11188 | 11188 |
| Control Mean (2011) | 0.12 | 0.42 | 0.050 | 0.21 | 0.15 |
| Temperature Mean (2001-11) | 1.26 | 1.26 | 1.26 | 1.26 | 1.26 |
| First-Differenced | Yes | Yes | Yes | Yes | Yes |
| Sample Period | 2001-11 | 2001-11 | 2001-11 | 2001-11 | 2001-11 |
| **Panel (b): Women** | Services | Cultivation | | Farm labour | |
| | Main | Main | Marginal | Main | Marginal |
| New road | 0.31** | 0.29 | 0.16 | -0.064 | -0.53* |
| | (0.13) | (0.20) | (0.18) | (0.19) | (0.27) |
| New road x Temperature | -0.17** | -0.15 | -0.095 | 0.061 | 0.27* |
| | (0.071) | (0.10) | (0.090) | (0.100) | (0.14) |
| Linear combination | 0.089* | 0.11 | 0.042 | 0.013 | -0.20* |
| | (0.047) | (0.085) | (0.076) | (0.073) | (0.10) |
| Villages | 11150 | 11150 | 11150 | 11150 | 11150 |
| Control Mean (2011) | 0.088 | 0.20 | 0.11 | 0.19 | 0.32 |
| Temperature Mean (2001-11) | 1.26 | 1.26 | 1.26 | 1.26 | 1.26 |
| First-Differenced | Yes | Yes | Yes | Yes | Yes |
| Sample Period | 2001-11 | 2001-11 | 2001-11 | 2001-11 | 2001-11 |

**Notes**: *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. | Standard errors (listed in parentheses) are clustered at the village level. | Sample Period denote the Census years, such that 2001-11 means that the outcome is observed in the 2001 and 2011 Censuses. *Temperature Mean (2001-11)* denotes mean of the first-differenced district-level temperature shock (i.e., the average change in the shock from 2001 to 2011) among villages that received a road. The *linear combination* is defined as the coefficient on *New road* plus the product of the coefficient on the interaction term and the mean temperature shock.

Table A3: First-stage and reduced-form effect on the main and placebo samples

| **Panel (a): Main sample** | New road (Received) | Cultivation (Main) | Services (Main) | Farm labour (Marginal) | Land (Index) | Asset (Index) |
|---|---|---|---|---|---|---|
| Above threshold | 0.19*** | 0.022** | 0.020** | -0.027** | 0.069* | 0.13** |
| | (0.018) | (0.011) | (0.0079) | (0.011) | (0.036) | (0.056) |
| Villages | 11188 | 11188 | 4728 | 11188 | 11184 | 3828 |
| Control Mean (2011) | 0 | 0.34 | 0.11 | 0.22 | -0.040 | -0.21 |
| District FE | No | No | No | No | Yes | Yes |
| First-Differenced | Yes | Yes | Yes | Yes | No | No |
| Sample Period | 2001-11 | 2001-11 | 2001-11 | 2001-11 | 2011 | 2011 |
| Restricted Sample | No | No | Yes | No | No | Yes |
| Max. Temperature Shock | | | 1.0 | | | 1.0 |
| **Panel (b): Placebo** | New road (Received) | Cultivation (Main) | Services (Main) | Farm labour (Marginal) | Land (Index) | Asset (Index) |
| Above threshold | -0.00041 | 0.012 | -0.0029 | 0.0022 | -0.0027 | -0.011 |
| | (0.013) | (0.010) | (0.0060) | (0.0097) | (0.037) | (0.023) |
| Villages | 13955 | 13955 | 11254 | 13955 | 13501 | 9237 |
| Control Mean (2011) | 0 | 0.32 | 0.14 | 0.17 | 0.21 | -0.31 |
| District FE | No | No | No | No | Yes | Yes |
| First-Differenced | Yes | Yes | Yes | Yes | No | No |
| Sample Period | 2001-11 | 2001-11 | 2001-11 | 2001-11 | 2011 | 2011 |
| Restricted Sample | No | No | Yes | No | No | Yes |
| Max. Temperature Shock | | | 1.0 | | | 1.0 |

**Notes**: *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. | The standard errors in parentheses are clustered at the village level. Panel (a) presents the regression discontinuity estimate of being eligible for a PMGSY road on employment shares (same as Table 2) and asset ownership (same as Table 5), i.e., the reduced-form effect of a road. For outcomes where a null effect was observed at the average level of the temperature shock (services' employment share and asset ownership), the sample is limited to districts that did not suffer a temperature shock of greater than 0.8 percentage points (the bottom quartile). All columns include the usual controls (see Section 5.2) and a triangular kernel with the optimal bandwidth (84) and a linear polynomial. Panel (b) uses a placebo sample to estimate the same specification. The placebo sample contains: (1) villages near the thresholds in states that did not comply with program rules and (2) villages close to the 1000-people threshold in complier states that used only the 500-people threshold.

Table A4: Impact on age distribution within villages

| | Age group (share of population) | | | | | |
|---|---|---|---|---|---|---|
| | 1-10 | 11-20 | 21-30 | 31-40 | 41-50 | 51-60 |
| New road | 0.00044 | -0.0044 | 0.0045 | -0.011 | -0.0017 | 0.012 |
| | (0.027) | (0.023) | (0.019) | (0.018) | (0.016) | (0.014) |
| New road x Temperature | 0.0014 | 0.00026 | -0.0041 | 0.0087 | 0.00019 | -0.0064 |
| | (0.015) | (0.013) | (0.011) | (0.010) | (0.0091) | (0.0079) |
| Linear combination | 0.0022 | -0.0041 | -0.00066 | 0.000045 | -0.0014 | 0.0040 |
| | (0.0085) | (0.0072) | (0.0061) | (0.0057) | (0.0050) | (0.0042) |
| Villages | 11184 | 11184 | 11184 | 11184 | 11184 | 11184 |
| Control Mean (2011) | 0.23 | 0.24 | 0.19 | 0.15 | 0.12 | 0.073 |
| Temperature Mean (2001-11) | 1.26 | 1.26 | 1.26 | 1.26 | 1.26 | 1.26 |
| District FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Sample Period | 2011 | 2011 | 2011 | 2011 | 2011 | 2011 |

**Notes**: *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. | Standard errors (listed in parentheses) are clustered at the village level. | Data on the population's age distribution is from the Socioeconomic and Caste Census (SECC), which was released in 2012. *Temperature Mean (2001-11)* denotes mean of the first-differenced district-level temperature shock (i.e., the average change in the shock from 2001 to 2011) among villages that received a road. The *linear combination* is defined as the coefficient on *New road* plus the product of the coefficient on the interaction term and the mean temperature shock.

Table A5: Impact of temperature on age distribution within villages

| | Age group (share of population) | | | | | |
|---|---|---|---|---|---|---|
| | 1-10 | 11-20 | 21-30 | 31-40 | 41-50 | 51-60 |
| Temperature | 0.030*** | 0.011*** | -0.013*** | -0.011*** | -0.0098** | -0.0077*** |
| | (0.0090) | (0.0036) | (0.0025) | (0.0031) | (0.0039) | (0.0029) |
| Rainfall | 0.0012 | -0.000067 | -0.00017 | -0.00082* | -0.00045 | 0.00035 |
| | (0.0014) | (0.00054) | (0.00033) | (0.00048) | (0.00072) | (0.00045) |
| Villages | 11188 | 11188 | 11188 | 11188 | 11188 | 11188 |
| Control Mean (2011) | 0.23 | 0.24 | 0.19 | 0.15 | 0.12 | 0.073 |
| Temperature Mean (2011) | 1.13 | 1.13 | 1.13 | 1.13 | 1.13 | 1.13 |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Sample Period | 2011 | 2011 | 2011 | 2011 | 2011 | 2011 |

**Notes**: *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. | Standard errors (listed in parentheses) are clustered at the district level. | Data on the population's age distribution is from the Socioeconomic and Caste Census (SECC), which was released in 2012. *Temperature Mean (2001-11)* denotes mean of the first-differenced district-level temperature shock (i.e., the average change in the shock from 2001 to 2011). The *linear combination* is defined as the coefficient on *New road* plus the product of the coefficient on the interaction term and the mean temperature shock.

# Appendix B - IHDS Matching

The matching follows the procedure of Raghav (2024) and proceeds in four stages:

**Stage 0:** Each IHDS village $v$ is initially matched with all villages in its district that had the same broad population category. The population category is a ternary variable that equals 1 if $v$ had less than 1000 people as of 2001, equals 2 if $v$ had between 1001 and 5000 people, and equals 3 otherwise. All population figures are sourced directly from the 2001 Population Census, but only the censored version is included in the IHDS.[26] As shown in the top-left subplot of Figure B1(a), the number of matches for each $v$ is very large at this stage, with a median of 500.

**Stage 1:** We use a subset of the village amenities, whose availability was recorded in both the 2011 Population Census and IHDS-II (2011-12), to reduce the number of matches. For each IHDS village $v$ and amenity $\alpha$, if $\alpha$ was (was not) available in $v$, then we drop all matches of $v$ that did not (did) have $\alpha$ according to the Population Census.

For this strict matching-on-amenities, we chose amenities whose availability was relatively difficult to mismeasure and unlikely to significantly change between 2011 and 2012.[27] These amenities are: (i) public primary school (ii) private primary school (iii) public middle school (iv) private middle school (v) public secondary school (vi) community centre (vii) bus service (viii) mobile phone coverage, and (ix) paved road.

There exist amenities that could be informative in this regard, but not for a strict matching. For instance, the IHDS records a primary health centre or a bank branch as being available in a village even if it may be located quite far away, such as 10 kilometers or more. However, if the IHDS reports the absence of these amenities, then it quite plausible that the "true" SHRUG match should also record these amenities as unavailable in the Population Census, especially if the latter does not consider far away amenities to be available within the village. In a similar vein, the presence of a community health workers (ASHAs) in villages was increasing during the 2010s because of India's National Rural Health Mission (NRHM); some villages could conceivably have an ASHA in the IHDS (in 2012) but not in the Population Census (in 2011). Thus, for these three examples, we employ a weak matching-on-amenities: if the IHDS village $v$ did not have it, then any SHRUG matches must not, either.

Some amenities that would fit the above criteria are deliberately excluded; they are instead used to evaluate the quality of this matching at the end. Still, this matching-on-amenities greatly reduces the amount of matches for each village $v$: the median number of stage-1 matches is 26, although the top-right subplot of Figure B1(a) reveals that a non-trivial share of IHDS villages still have over 100 matches.

[26]This is to preserve the respondents' anonymity by making it very difficult to identify the villages.

[27]This is important because over 90% of villages in the IHDS were surveyed in 2012.

**Stage 2:** The IHDS additionally sources two more village-level variables directly from the 2001 Population Census – the number of households and the total village area (in square hectares). The latter is reported at a precision of two decimal places. For each IHDS village $v$, we check if any of its stage-1 matches has exactly the same number of households and village area. we find that 20 percent of IHDS villages have a unique stage-1 match, the "perfect match", that satifies this criterion. In such cases, we drop any stage-1 matches that aren't the perfect match, i.e., $v$ is mapped only to its perfect match. As indicated by the bottom-left subplot of Figure B1(a), the median number of such stage-2 matches is 18.

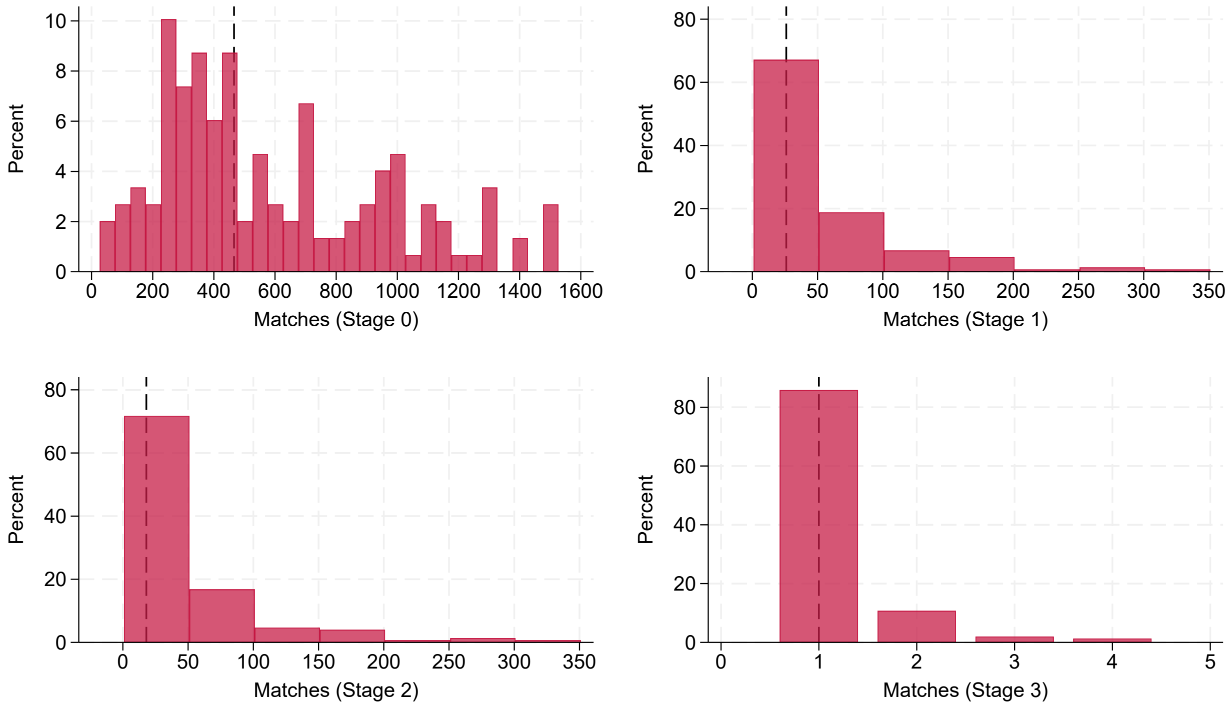
**Stage 3:** Given that 80% of IHDS villages do not have a unique stage-2 match, it is possible that either the number of households or village area are recorded with some, possibly minor, errors in the IHDS. we assume that the latter is more likely to be the source of such errors. Consequently, for each village without a perfect match, we keep only the stage-2 match that had the closest number of households. For nearly half of these IHDS villages, the remaining stage-2 match had exactly the same number of households. Overall, as illustrated in the bottom-right subplot of Figure B1(a), about 85% of all IHDS villages have a unique match at the end of stage-3.

**Stage 4:** For the remaining 15% of IHDS villages, we retain the stage-3 match whose total area is closest to that of the IHDS village. This produces a unique match for these remaining villages, i.e., all IHDS villages have a unique stage-4 match in the SHRUG data.

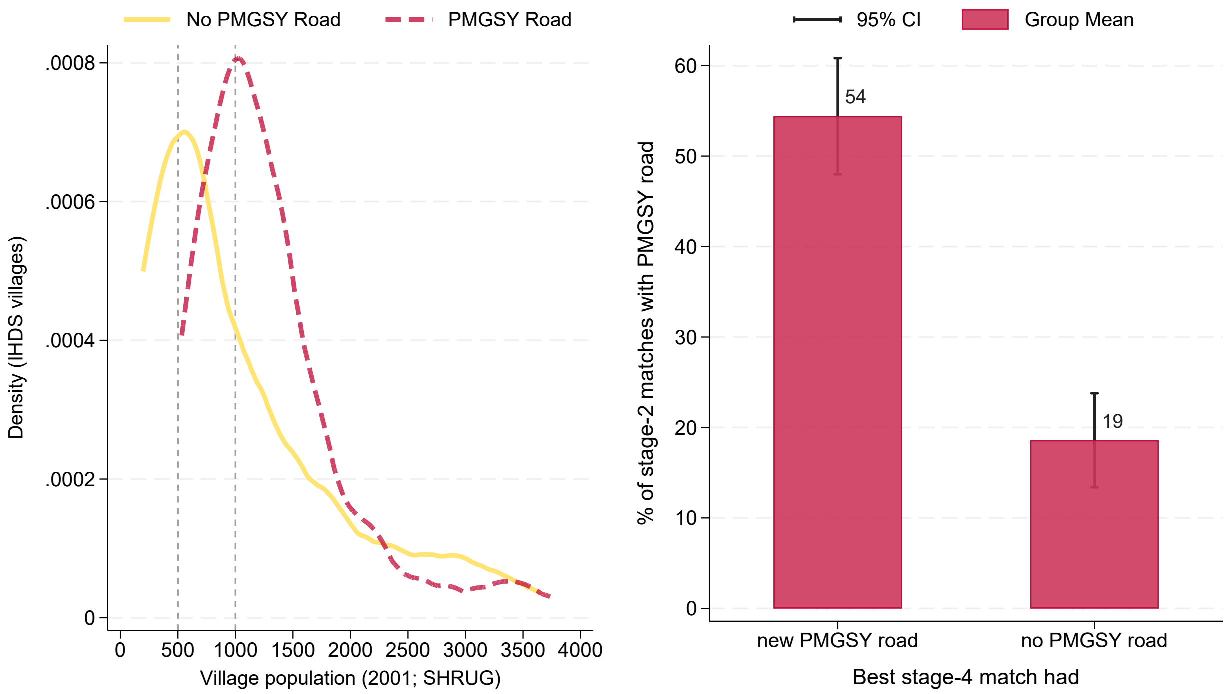
Table B1 compares the availability of village amenities in the IHDS villages and their best SHRUG matches. The congruence in amenities—the data on these amenities comes from IHDS-II (2011-12) and the 2011 Population Census—suggests that the matching can indicate which new road in the IHDS villages came from the PMGSY.

Figure B1: Summarization of matching quality (IHDS and SHRUG)

(a) Number of SHRUG matches throughout matching process




(b) Quality of best stage-4 SHRUG match




**Notes:** In panel (a), the dashed line represents the median number of SHRUG matches for any given IHDS village. In panel (b), the line with short dashes denotes the PMGSY population thresholds as of 2003 (1000) and 2007 (500). In panel (c), the lines denote the 95% confidence interval for the group means.

Table B1: Balance table - IHDS villages and best stage-4 SHRUG matches

| **Amenity** | Available in village | | Data consistency between sources |
|---|---|---|---|
| | IHDS | Best SHRUG match | |
| | $100\,\Pr(Y=1)$ | $100\,\Pr(X=1)$ | $100\,\Pr(Y=X)$ |
| Primary School (Govt.) | 99.41 | 99.41 | 100 |
| Primary School (Pvt.) | 20.12 | 20.12 | 100 |
| Middle School (Govt.) | 66.27 | 66.27 | 100 |
| Middle School (Pvt.) | 13.02 | 13.02 | 100 |
| Secondary School (Govt.) | 20.12 | 20.12 | 100 |
| Secondary School (Pvt.)$^{\dagger}$ | 3.55 | 3.55 | 94.08 |
| Primary Health Centre$^{\dagger\dagger}$ | 2.96 | 2.96 | 100 |
| Health Dispensary$^{\dagger}$ | 8.28 | 2.96 | 91.12 |
| Community Health Worker (ASHA)$^{\dagger\dagger}$ | 94.67 | 98.82 | 93.49 |
| Anganwadi Centre$^{\dagger}$ | 98.22 | 98.82 | 97.04 |
| Community Centre | 0 | 0 | 100 |
| Bus Service | 45.56 | 45.56 | 100 |
| Bank Branch ($\leq$ 1 KM away)$^{\dagger\dagger}$ | 7.69 | 9.47 | 95.86 |
| Electricity Available$^{\dagger}$ | 98.2 | 97.35 | 95.3 |
| Mobile Phone Coverage | 100 | 100 | 100 |
| Paved Road (2011) | 80.47 | 80.47 | 100 |

**Notes**: This table is constructed using the set of villages that did not already have roads at the baseline. $^{\dagger}$ denotes an amenity that was not used to match IHDS and SHRUG villages. $^{\dagger\dagger}$ denotes an amenity that was used for a partial match, i.e., all potential SHRUG matches were dropped if they had the amenity but the IHDS village did not.