

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Comparative Economics

journal homepage: www.elsevier.com/locate/jce

Road to productivity: Effects of roads on total factor productivity in Indian manufacturing[☆]

Subham Kailthya^{a,*}, Uma Kambhampati^b^a University of Warwick, Coventry, CV4 7AL, UK^b University of Reading, Whiteknights, RG6 6AH, UK

ARTICLE INFO

Keywords:

Market integration
 Total factor productivity
 Transport infrastructure
 Manufacturing industry
 Plant level analysis

ABSTRACT

A denser road network lowers transport costs and stimulates manufacturing total factor productivity (TFP). The placement of roads, however, is likely to be non-random. For identification, we exploit cross-state variation in the strength of centre-state partisan alignment that asymmetrically affects road building in aligned states. Using panel data on manufacturing establishments in India from 1998 to 2012, we find that, a 1% increase in road density raises value-added TFP by about 0.25%, on average. A closer examination reveals that the effect varies by plant characteristics and road type. Younger establishments are more likely to gain from a denser road network with highways playing a prominent role. The results are robust to imperfections in the instrument and to other sensitivity checks.

1. Introduction

Transportation plays a vital role in stimulating economic growth. But, in many emerging economies including India and China, inadequate transportation infrastructure is a common problem that hinders manufacturing growth (Jedwab and Moradi, 2016; Ghani et al., 2016a). A recent World Bank Enterprise Survey (WBES) reports that 19 per cent of manufacturing firms globally, 21 per cent of firms in South Asia, and about 10 per cent of firms in India find inadequate transportation infrastructure as a major obstacle (World Bank, 2014). Not surprisingly, this has spurred governments in many emerging economies to invest in large-scale transportation projects to stimulate economic activity.¹ Yet, our understanding of how closing the transportation infrastructure gap affects manufacturing productivity is far from complete. The knowledge gap is particularly acute for emerging economies where poor connectivity inhibits economic growth. In this paper, we provide causal estimates of the effect of roads on manufacturing total factor productivity (TFP) within a developing country context and show how the effect varies by plant-specific characteristics and road type.

The literature on the economic impact of transportation infrastructure can be organized into three distinct strands. The first focuses on the economic returns from highways connecting nodal cities. Prominent examples of this approach include the study of the Golden Quadrilateral (GQ) highway in India and the National Trunk Highway System (NTHS) in China (Ghani et al., 2016a,b; Datta,

[☆] We thank the editor and three anonymous reviewers for their insightful comments. We benefitted from discussions with Thushyanthan Baskaran and Jayan Jose Thomas on an earlier version of the paper. We thank seminar participants at the Royal Economic Society (2021), Centre for the Study on African Economics (CSAE, 2019), ISI Economic Growth and Development Conference (2019), South Asian Economic Development Conference (SAED, 2019), Cathie Marsh Institute (CMI) at the University of Manchester and brown-bag seminar at the University of Reading for their thoughtful suggestions.

* Corresponding author.

E-mail address: subham.kailthya@warwick.ac.uk (S. Kailthya).

¹ For instance, India's 2017–18 budget announced a record spending of Rs.3.96 trillion (\$59 billion), a 12% annual increase, to build and modernize its transport infrastructure and provide 'renewed impetus' to manufacturing [link].

<https://doi.org/10.1016/j.jce.2021.08.001>

Received 28 February 2019; Received in revised form 20 August 2021; Accepted 24 August 2021

0147-5967/© 2021 Published by Elsevier Inc. on behalf of Association for Comparative Economic Studies.

2012; Asturias et al., 2018; Faber, 2014). The second concentrates on the impact of rural roads on village economies (Aggarwal, 2018; Bohlken, 2019; Asher and Novosad, 2020), while the third creates a synthetic index of infrastructure to examine how the overall infrastructure index affects manufacturing activity (Mitra et al., 2002, 2012). Although these studies provide useful insight, knowledge gaps remain. For instance, while studies that examine the impact of GQ highways are internally valid they may not be generalizable given that GQ highways cover less than 0.1 per cent of India's road network. Moreover, these studies examine the impact of roads on local economic activity at the extensive margin i.e. the additional benefit of new or improved highways, but pays little attention to the impact at the intensive margin i.e. the effect of a denser transportation network which might be illuminating.²

In this paper, we provide causal estimates of the impact of roads on manufacturing TFP. To do this, we combine plant-level panel data on Indian manufacturing with sub-national data on road transportation network, disaggregated by type, from 1998 to 2012. An important challenge in estimating the impact of roads on economic activity is that roads are not randomly placed (Banerjee et al., 2020; Shatz et al., 2011, etc.). For example, roads might be more prevalent in places that trade more i.e. when roads are built to support trade or, if they trade less i.e. when roads are built to encourage trade (Duranton et al., 2014). We address the endogeneity of road placement by exploiting cross-state variation in the strength of centre-state partisan alignment or, *alignment strength* for short, that asymmetrically stimulates road building in aligned states. Here, we consider a state to be aligned with the centre if its ruling party is the same as that governing the centre or if it is a member of the ruling central coalition. The premise is that aligned states face fewer economic and non-economic hurdles in implementing public projects such as roads. Obstacles in the form of delays in land acquisition, obtaining planning permission, vetting of tenders etc. might stall projects entirely. Aligned states, however, face fewer such obstacles and are much more likely to see road projects being implemented in comparison to non-aligned states.

We believe that this is a credible identification strategy for several reasons: First, plant-level productivity is not directly influenced by political-market bargaining that determines alignment strength. Secondly, as far as possible, our electoral outcomes pre-date road building data. In our analysis, the average lag between these variables is two years which attenuates potential concerns around reverse feedback from road building onto electoral outcomes. In addition, elections in India are staggered across groups of states which further reduces the possibility of simultaneity bias. Third, we control for a range of potential confounders in our regression to ensure that omitted variables are not affecting our results. We discuss this further in Section 4.2. However, these approaches cannot rule out endogeneity completely. Hence, we conduct sensitivity checks to ensure that our estimates are robust to potential imperfections in the instrument. To this end, we generate bounds on the treatment effect that relaxes the assumption of strict exogeneity of the instrument with sign restrictions using the approach developed by Nevo and Rosen (2012). Further, we test the sensitivity of our result to plausible violations of the exclusion restriction based on the method by van Kippersluis and Rietveld (2018). We discuss these identification tests in Section 5.2.

To preview our main results, we find that after controlling for endogeneity, a 1% increase in road density raises value-added TFP by about 0.25%, on average. Thus, moving a state from the 10th percentile of the road density distribution to the 50th percentile will raise TFP by about 13%. Furthermore, we show that these gains are due to a reallocation of inputs in favour of higher productivity firms within an industry (see Appendix D). However, a closer examination reveals that the effect of roads on TFP varies by plant-specific characteristics and road type. Younger establishments are more likely to gain from a denser road network with highways playing a prominent role. These results are robust to using alternative measures of TFP, employing the percentage change in road density as the main covariate, estimating the model on alternative sub-samples, guarding against outliers and accounting for spatial spillovers. In addition, to alleviate concerns regarding measurement error arising from using a state-level aggregate measure of road density, we examine the robustness of our results to using district-level road density for a smaller sub-sample for which detailed data was available and find qualitatively similar results.³

Related literature: The existing literature proposes two broad mechanisms by which an improvement in transportation infrastructure stimulates economic activity. First, it reduces trade barriers by lowering transportation cost which leads to greater market integration. The arterial network of roads and railways have played a fundamental role in increasing trade volumes by expanding market access. Donaldson (2015) reviews theoretical and empirical research on the gains from market integration as a result of reduced trade barriers, while Redding and Turner (2015) review the role of transportation costs on the spatial distribution of economic activity. In the context of developing countries, using global oil prices to provide exogenous variation in transport costs, Storeygard (2016) finds that the economic activity of African cities that are further away from roads suffers more than cities that are nearer during high oil price years. In India, Ghani et al. (2016a) utilize the time-variation in GQ highway construction/upgrading and find that industries located nearer to highways produce more output and do so with higher efficiency.

Secondly, a reliable transportation network lowers uncertainty at both the procurement and supply stages thereby reducing the need to hold on to costly inventory (Shirley and Winston, 2004; Li and Li, 2013; Datta, 2012). This is particularly relevant for developing countries where inventory levels are two to five times higher than in the United States (Guasch and Joseph, 2001). In India, Gulyani (2001) finds that the 'total logistics cost' of inadequate transportation infrastructure in the automotive sector is very high, while Datta (2012) estimates that being on the GQ highway reduces inventory holdings by at least six days' worth of production. The combined effect of lower transportation costs and reduced inventory pile-up has a first order impact on

² Hulten et al. (2006) is an exception that examines the impact of national and state highways on Indian manufacturing productivity. However, it does not control for non-random road placement or the effect of spatial spillovers.

³ However, we do not use this as our baseline since it spans only from 2007 to 2010 and covers less than a third of the entire dataset for which information on plant-level TFP is available. We discuss this in more detail in Section 6.

manufacturing productivity. [Asturias et al. \(2018\)](#) document large welfare gains from GQ access due to higher allocative efficiency, labour income and average markups that affect state's terms of trade.

Recent contributions have studied the effect of transportation infrastructure on long-term GDP ([Banerjee et al., 2020](#)), economic activity ([Chandra and Thompson, 2000](#); [Ghani et al., 2016a,a](#)), bilateral trade between cities ([Duranton et al., 2014](#); [Faber, 2014](#)), productivity and allocative efficiency ([Fernald, 1999](#); [Holl, 2016](#); [Lall et al., 2004](#); [Ghani et al., 2016a](#); [Asturias et al., 2018](#)), skill demand in local labour markets ([Michaels, 2008](#)), and urban form ([Baum-Snow et al., 2017](#)), among others. Our paper is closest to [Ghani et al. \(2016a\)](#) and [Asturias et al. \(2018\)](#) in that we study the effect of roads on manufacturing TFP in India. But, while these studies consider exclusively the economic effect of GQ highways that constitute less than 0.1% of India's road network at the extensive margin, we focus on the intensive margin impact of India's entire road transportation network. Furthermore, we document the importance of plant-level micro-structure and road type in affecting the relationship.

The gains from lower transportation costs are not evenly distributed across firms within industries. In fact, the recent literature points out the importance of firm micro-structure in determining how policy distortions affect aggregate productivity ([Restuccia and Rogerson, 2008](#); [Bernard et al., 2012](#); [Melitz, 2003](#)). This means that a fall in transportation costs might result in higher productivity firms to thrive while lower productivity firms shrink or exit altogether. Moreover, when the dispersion of productivity across firms within industries is large, the effect of micro-heterogeneity in affecting aggregate productivity becomes even more salient. [Syverson \(2011\)](#) and [Bartelsman et al. \(2013\)](#), for instance, find enormous and persistent difference in productivity across firms even within narrowly defined industries. Yet, the role of plant-specific heterogeneity in affecting the relationship between falling transportation costs and productivity has received little attention. In contrast to the existing literature, we investigate how plant-specific heterogeneity in terms of size, age and location, affects the relationship between roads and TFP.

When trade barriers fall as a result of improvements in road transportation infrastructure, it reallocates inputs in favour of higher productivity firms. And, this within-industry reallocation affects aggregate productivity. [Hsieh and Klenow \(2009\)](#) notes that eliminating resource misallocation will significantly raise TFP in manufacturing by 86–115 per cent in China, 100–128 per cent in India and 30–43 per cent in the US. To assess the degree of misallocation, [Bartelsman et al. \(2013\)](#) points out that the within-industry covariance between firm size and productivity is a robust measure to examine the impact of misallocation. We therefore construct within-industry covariance between logged TFP and firm size and assess how road transportation infrastructure affects reallocation of resources across plants within the same industry.

Empirical identification of the economic effect of transportation infrastructure is challenging because of endogeneity. In the literature, identification often comes from instrumenting the existing transportation network with historical roads/ railway routes ([Donaldson and Hornbeck, 2016](#); [Donaldson, 2018](#); [Duranton et al., 2014](#); [Agrawal et al., 2017](#)), the distance to straight lines connecting historical cities or terminal nodes ([Banerjee et al., 2020](#); [Ghani et al., 2016a](#); [Asturias et al., 2018](#)), or the construction of least cost path spanning tree network between bilateral pair of nodes ([Faber, 2014](#)). Others have used exogenous variation in oil prices interacted with city-pair distance directly in the regression ([Storeygard, 2016](#)), the duration of exposure of an administrative area to rural roads ([Aggarwal, 2018](#)) or, utilized the population threshold in targeting rural road building for identification ([Asher and Novosad, 2020](#)).⁴ In this paper, our instrumentation strategy relies on the political-economics of road building in a federal democracy. Specifically, we instrument sub-national road density with the strength of centre-state partisan alignment that asymmetrically stimulates road building in aligned states. Moreover, the fact that election cycles in India are staggered across groups of states helps our identification.

There are three main contributions of this study. First, we provide a political-economic basis for understanding the heterogeneity in the sub-national distribution of manufacturing productivity. The mechanism is quite simple: centre-state partisan politics encourages strategic road building in aligned states which, in turn, affects manufacturing productivity. The impact is sequential and we utilize this in our estimation by considering the impact of electoral outcomes that are lagged by about two years, on average, on road building and then estimate the impact of these roads on manufacturing TFP. The observed evidence validates the above mentioned mechanism in the context of India. This is important because the resulting distribution in economic activity is likely to be sub-optimal whenever the observed distribution deviates from an equilibrium based on economic efficiency alone.

Secondly, on the empirical side, we estimate TFP using plant-level panel data that enables us to control for fixed unobservable plant-specific heterogeneity. Previous studies examining the relationship between roads and TFP in India rely on repeated cross-section data that do not control for plant-specific heterogeneity and hence may be biased (see, for example, [Ghani et al., 2016a](#)). In contrast, we control for such heterogeneity and obtain estimates that are closer to the 'true' effect.

Finally, we provide causal estimates of the impact of roads on TFP at the intensive margin. This might be particularly relevant in the case of roads because of its interdependent structure. Furthermore, we highlight the role of plant-level heterogeneity and road type in affecting the relationship.

We organize this paper as follows: Section 2 provides essential background relating to the financing and management of road infrastructure in India. Section 3 introduces the dataset and discusses how we estimate TFP. Section 4 specifies a simple model that outlines the relationship between roads and TFP and presents our identification strategy. Section 5 presents the main regression results and its extensions, while Section 6 reports robustness checks. Section 7 concludes.

⁴ Both [Aggarwal \(2018\)](#) and [Asher and Novosad \(2020\)](#) utilize the programme rollout of the Prime Minister's rural roads (PMGSY) project. While [Aggarwal \(2018\)](#) use the variation in the percentage of population that received a road by district-year to assess the impact of rural roads on the price of consumer goods, consumption variety, education and employment, [Asher and Novosad \(2020\)](#) examines the effect of rural roads on employment opportunities using a fuzzy regression discontinuity design.

2. Road infrastructure in India

In this section, we provide a brief discussion of the administration, implementation and financing of road infrastructure in India which will help in understanding the analysis that follows.

India's road network spans more than 5 million kilometres. Of this total, national highways (NHs) constitute about 1.6 per cent, state highways (SHs) another 3.3 per cent, while urban and rural roads constitute about 9.5 per cent and 58.3 per cent, respectively. Besides these, there are roads that are under the public works department and project roads maintained by different government departments. Each of them, in turn, is developed and managed by a different level of government. The NHs are developed and maintained by the central government through agencies like the National Highways Authority of India (NHAI), while the SHs are the responsibility of the state governments and are maintained by the states' public works departments. Finally, while urban roads are the responsibility of municipalities, rural roads are planned under many national-level rural development and employment schemes including the Minimum Needs Programme, National Rural Employment Programme, Jawahar Rozgar Yojana as well as the Prime Minister's Gram Sadak Yojana (PMGSY) which translates to Prime Minister's rural roads programme.

A significant proportion of government funding for road-building in India comes from the Central Road Fund (CRF) which was created under the Central Road Fund Act, 2000. This allows the Central Government to levy a cess/tax on petrol and high speed diesel. The funds from the CRF are used to develop and maintain national highways, state roads (especially ones that are economically important or that connect states), rural roads (which are developed and maintained by a number of organizations) as well as specific road projects. In addition to the CRF, in 1998, a separate State Road Fund (SRF) was also set up. This was financed from multiple sources: budgetary support from the CRF and state government, direct road user charges from cess on fuel, motor vehicle taxes, fees and tolls, indirect road user charge/tax such as hotel tax and levy on agriculture products, and other resource such as fines, loans etc. These funds are especially significant in the states of Uttar Pradesh, Madhya Pradesh, Kerala, Assam, Karnataka and Rajasthan. In addition, from time to time, the Central Government receives proposals from state governments for certain state roads to be declared as NHs. When these roads are declared as NHs, the state is no longer responsible for financing their development and maintenance. In 2017–18, for instance, proposals for 64,000 kilometres of road were received from state governments, of which about 10,000 kilometres of roads/routes were for new NHs ([Government of India, 2018](#)).

According to the Indian Constitution, the development of rural roads is the responsibility of the state governments and thus the central government was not directly involved in the funding of rural road projects. However, this changed from the fifth five-year plan of India (1974–1978), when the central government started funding rural road projects through various programmes such as the Minimum Needs Programme (MNP), the National Rural Employment Programme (NREP), the Rural Landless Employment Guarantee Programme (RLEGP) and Jawahar Rozgar Yojana (JRY). In December 2000, the central government initiated the Prime Minister's Gram Sadak Yojana Programme (PMGSY), with the objective of connecting all villages having a population over 500 by the end of 2007.

Building roads, however, require coordination between the centre and the states. There are often obstacles relating to land acquisition, obtaining planning permission, vetting of tenders etc. which delay projects and lead to huge cost overruns. The centre can strategically relax such economic and non-economic obstacles in a way that favours aligned states. As a result, aligned states are more likely to see road projects being implemented. This is important for us and we utilize this mechanism in our identification (see Section 4.2). In the next section, we introduce our data and discuss the construction of key variables.

3. Data and measurement

3.1. Data

We combine plant-level data on organized manufacturing in India with data on transportation infrastructure and socio-demographic characteristics from multiple sources. We discuss this below.

Manufacturing Data: To estimate TFP, we obtained plant-level panel data on manufacturing in India from 1998 to 2012. We sourced this data from the Annual Survey of Industries (ASI), a pan-India survey of organized manufacturing establishments administered by the Central Statistical Organization (CSO), Government of India. ASI provides exhaustive data on the book values of a plant's assets and liabilities, employment and labour, receipts and expenses along with several other economic variables for a financial year (e.g. the 1998 survey reports data for 1998–99), but we will only refer to the initial year for simplicity.

ASI covers organized manufacturing establishments or plants registered under the Factories Act, 1948 that employ more than 10 workers if they use electricity in their manufacturing process or 20 workers if they do not. It follows a Circular Systematic sampling design that divides the sampling frame into two sectors: a census sector and a sampling sector.⁵ The sampling design considers the state, sector and 4-digit NIC codes in stratifying the sample. We use the sampling multipliers provided by ASI in all our analysis to ensure that our results are representative of the entire population of organized manufacturing plants in India. We then use this dataset to estimate plant-level TFP from 1998 to 2012 which we describe in Section 3.2.

Road Density Data: Roads are the predominant mode of transportation in India carrying nearly 65% of freight and 80% of all passenger traffic (NHAI cited in [Ghani et al., 2016a](#)). In this study, we consider the log of state-level road density – the ratio of

⁵ The census sector consists of establishments in the states of Manipur, Meghalaya, Nagaland, Tripura and Union Territory of Andaman and Nicobar Islands where all establishments are surveyed. In other states, the census sector includes establishments with more than 100 workers or if they file joint returns i.e. returns for multiple units within a state. The sampling sector, on the other hand, consists of establishments that are not included in the census sector.

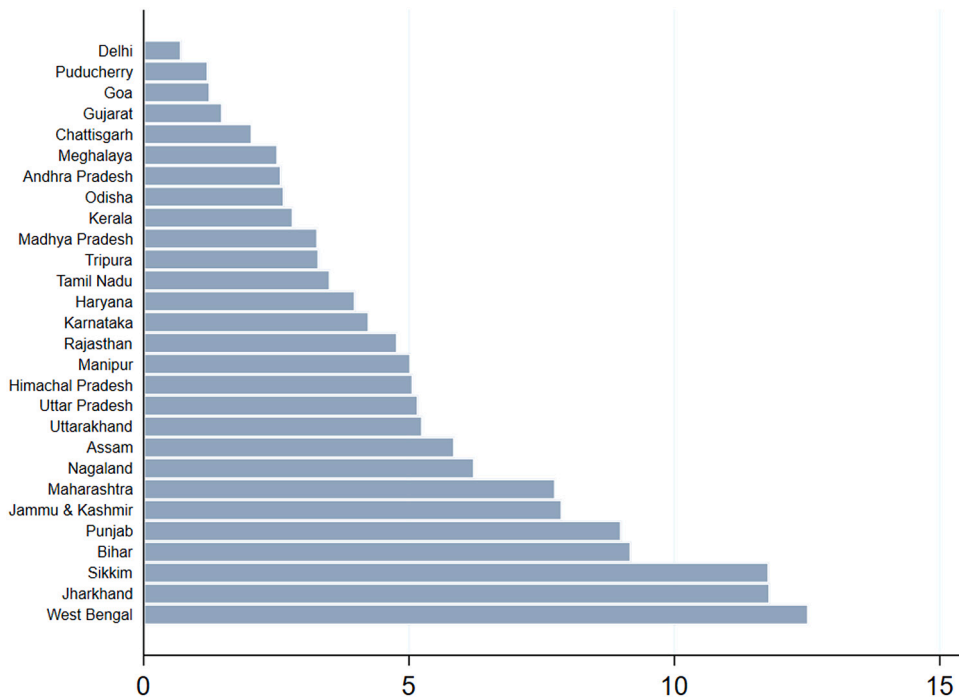


Fig. 1. Statewise percentage change in road density in India, 2002–2012.

the total length of roads (in kilometres) in a state to its land area (in square kilometre) expressed in log scale – as our primary measure of road transportation infrastructure. To this end, we obtained data on road length by state-year from the Ministry of Road Transport and Highways from 1998 to 2012.

The summary statistics indicate that the statewise average road density in our sample was 1.98 km/sq.km with significant variability across states (standard deviation of 3.57 km/sq.km). Delhi was at the top with 19.32 km/sq.km. In contrast, road density was only 0.11 km/sq.km in Jammu Kashmir. This unevenness in road coverage extends onto changes as well. Between 1998–2012, West Bengal recorded the highest increase in road density per sq. km. whereas, it grew the least in Delhi (see Fig. 1). Roads vary in their quality and functions, as already mentioned in Section 2. Hence, it is important to understand how the impact of roads on manufacturing productivity varies by its type. To do this, we gathered data on the length of national highways and state highways and calculated their respective shares by state-year. We also calculated the share of other roads i.e. roads excluding the share of national highways and state highways as a residual.

In addition, to test the sensitivity of our main results to a more granular measure of road density, we construct district-level road density for the year 2007 using GIS data on India's road network obtained from the United Nation's International Steering Committee for Global Mapping (ISCGM).

Railway Density Data: Railways present the main alternative to road transportation in India (see Donaldson, 2018, on the impact of railways on trade in colonial India). Hence, to identify the effect of roads on manufacturing productivity we need to control for a state's railway density in our model. To operationalize this, we gathered data on the length of railways (in route km) for every state-year during the study period from the Ministry of Railways and divided it by a state's land area to construct state-level railway density. Furthermore, in a way analogous to that of roads, we obtained GIS data on railways for the year 2007 from ISCGM and constructed a district-level measure of railway density.

Electricity Supply Data: Another important variable that needs to be controlled for in estimating the relationship between roads and productivity is the availability of uninterrupted electricity. Allcott et al. (2016), for example, find that textile industries in India lose about 5 per cent of output to power blackouts although its effect on productivity is quite small. Moreover, varying the electricity supply might also be an effective political strategy to win elections as shown in Baskaran et al. (2015). We therefore control for electricity supply in our analysis by including the installed power capacity in megawatts by state-year, obtained from the Central Electricity Authority (CEA) after normalizing it by a state's area. For the district-level analysis we obtain information on the percentage of villages that have access to electricity from the 2001 census as a proxy.

Socio-Demographic Data: We also include several important socio-demographic variables interacted with time as controls in our model to reduce omitted variable bias. These include time-interacted log of population, literacy rate, the number of main workers and the number of marginal workers. Data on these socio-demographic variables are obtained from the 2001 census.

Next, we discuss how we measure total factor productivity in manufacturing — the dependent variable in our analysis.

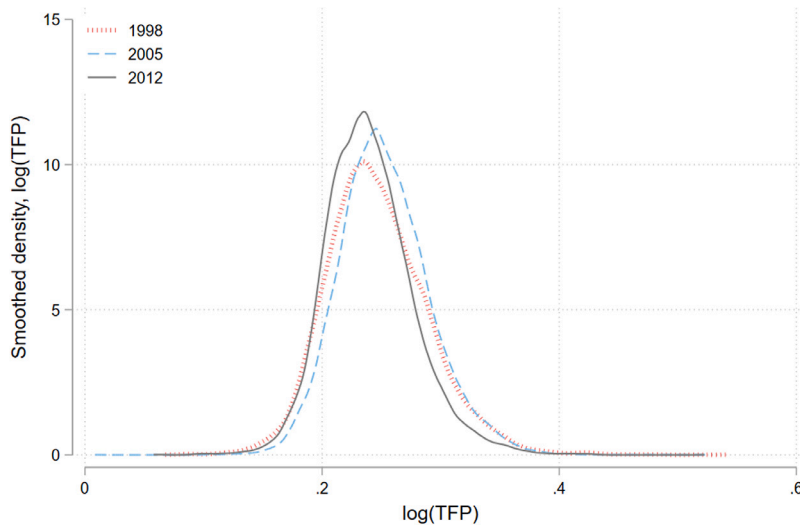


Fig. 2. Empirical density function of TFP in 1998, 2005 and 2012.

Figure plots the empirical density function of log value-added TFP in 1998, 2005 and 2012. TFP is normalized by dividing it by the employment-weighted average productivity for an industry-year.

3.2. Measuring TFP

In this paper, we mainly focus our discussion on value-added TFP. However, as shown below, our results are similar if we instead consider the gross-output definition of TFP. We focus on multifactor productivity because differences in TFP reflect shifts in the production isoquants with higher-TFP producers on a higher isoquant than lower-TFP producers (Syverson, 2011). We estimate plant-level TFP as the residual from a Cobb–Douglas production function as in Eq. (4) below, at the 3-digit NIC industry classification using the Akerberg et al. (2015) method (henceforth, ACF).⁶

To estimate TFP, we first deflate sales, labour, raw material, energy, and capital to arrive at their respective real values. Since the book value of capital is measured at historic costs, we use the perpetual inventory method (PIM) that accounts for differences in the vintage of capital stock. We then estimate plant-level TFP by 3-digit NIC industry classification using 264 thousand observations covering 54 thousand plants (see Appendix A for the variables used in estimating TFP). Next, we normalize logged TFP by dividing it by the employment-weighted average productivity (in logs) for an industry-year.⁷

Fig. 2 plots the density distribution of log normalized value-added TFP (TFPVA) across all manufacturing plants during 1998, 2005 and 2012, while Fig. 3 plots the corresponding empirical cumulative distribution functions (CDFs). Table 12 in Appendix E, shows results from pairwise comparisons of the equality of the CDFs using the Kolmogorov–Smirnov (KS) test. The result from the KS test and the distribution plots show that while manufacturing TFP declined between 1998 and 2005, it bounced back and increased between 2005 and 2012.

Table 1 refers to summary measures of TFP for all the plants in our sample during 1998–2012. It shows that the average difference in logged total factor productivity (TFP) between a plant at the 90th percentile of an industry’s TFP distribution and a plant at the 10th percentile is 0.383. This corresponds to a TFP ratio of 1.47 ($=e^{0.383}$). Thus, plants at the 90th percentile produce, on average, almost fifty per cent more output with the same set of observable inputs than plants at the 10th percentile. In addition, the range’s standard deviation across industries is 0.12 implying a high dispersion in TFP across plants.

Further, disaggregating TFP by a plant’s age, size and location is revealing. Incumbents – plants that are more than 5 years old – are slightly more productive than younger plants particularly beyond the 75th percentile. Classifying plants based on their asset-size shows that smaller plants – those with fixed assets less than the industry-year median value – are more productive than larger plants that fall above this threshold. In terms of location, we observe that urban-based plants are more productive than rural-based plants.

⁶ There are two main reasons for choosing the ACF method in estimating TFP: First, the ACF method accounts for the endogeneity of input choice in estimating productivity. The endogeneity arises because firms/ plants can observe their productivity before choosing the level of inputs, which leads to correlations between inputs and productivity. Secondly, the ACF method overcomes the functional dependence problem that affects identification of the labour coefficient in Olley and Pakes (1996) (OP) and Levinsohn and Petrin (2003) (LP) methods. The OP and LP invert investment (OP) and intermediate input (LP) demand functions that are unconditional on labour input. The ACF method, however, inverts the investment or intermediate input that are conditional on labour input to overcome functional dependence and hence correctly identifies the labour coefficient in the first stage (see Akerberg et al., 2015 for a detailed discussion on the methodology, and Arnold et al., 2015 for a recent application of this method).

⁷ This approach is followed in Ghani et al. (2016a). We adopt this approach for comparability and note that our results hold even without normalizing the TFP values.

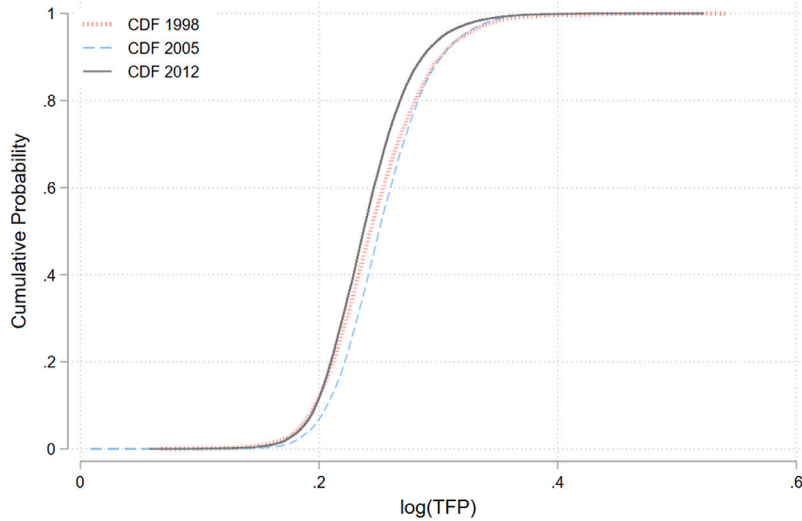


Fig. 3. Empirical CDF of TFP in 1998, 2005 and 2012.

Figure plots the empirical cumulative distribution function (CDF) of log value-added TFP in 1998, 2005 and 2012. TFP is normalized by dividing it by the employment-weighted average productivity for an industry-year.

Table 1
Productivity in Indian Manufacturing, 1998–2012.

| | Obs. | Mean | p10 | p25 | p50 | p75 | p90 |
|---------------------|--------|-------|-------|-------|-------|-------|-------|
| All | 258918 | 0.201 | 0.059 | 0.070 | 0.088 | 0.317 | 0.442 |
| <i>By age:</i> | | | | | | | |
| Young | 43194 | 0.196 | 0.060 | 0.071 | 0.087 | 0.301 | 0.436 |
| Incumbent | 215724 | 0.202 | 0.058 | 0.070 | 0.088 | 0.320 | 0.443 |
| <i>By size:</i> | | | | | | | |
| Small | 128366 | 0.218 | 0.059 | 0.071 | 0.103 | 0.339 | 0.451 |
| Large | 130552 | 0.184 | 0.058 | 0.070 | 0.084 | 0.256 | 0.433 |
| <i>By location:</i> | | | | | | | |
| Rural | 101993 | 0.181 | 0.056 | 0.068 | 0.082 | 0.274 | 0.422 |
| Urban | 156925 | 0.215 | 0.061 | 0.073 | 0.095 | 0.336 | 0.455 |

Notes: Table shows the dispersion of log value-added Total Factor Productivity (TFP) in Indian manufacturing during 1998–2012 estimated using the Akerberg, Caves and Frazer (2015) method. Plant-level log(TFP) values are normalized by dividing it by employment-weighted average productivity for an industry-year. Row ‘all’ in the table corresponds to estimates for the full sample. ‘Young’ relates to plants aged less than 5 years while ‘incumbents’ correspond to plants aged 5 years and above. ‘Small’ corresponds to plants with fixed-assets less than the industry-year median value whereas, ‘large’ refers to plants with fixed-assets more than the industry-year median value. ‘Rural’ denotes plants located in rural areas (as defined by ASI) whereas, ‘urban’ corresponds to plants in urban areas.

4. Empirical strategy

4.1. Model specification

As mentioned earlier, in this paper, we analyse the effect of roads on manufacturing productivity. To do this, we first estimate productivity as a function of local determinants using a Cobb–Douglas production function (see [Holl, 2016](#); [Martin et al., 2011](#)):

$$Y_{it} = A_{it} K_{it}^{\beta_1} L_{it}^{\beta_2} \quad (1)$$

where Y_{it} is the value-added output of the i th plant at time t . L_{it} denotes labour employed in production and K_{it} is the stock of capital. We note that plant i belongs to industry j and is located in district d within state s but we suppress the subscripts for simplicity. Productivity, A_{it} , depends on access to roads, R_{st} (or R_{dt} when we observe roads at the district level), and a vector of state-specific characteristics, X_{st} (or X_{dt} at the district level). We model plant-level productivity as:

$$A_{it} = R_{st}^{\gamma} X_{st}^{\delta} \quad (2)$$

Then, we log-linearize Eq. (1) and retrieve plant-level logged TFP, α_{it} as:

$$y_{it} = \beta_1 k_{it} + \beta_2 l_{it} + \alpha_{it} \quad (3)$$

$$\alpha_{it} = y_{it} - \beta_1 k_{it} - \beta_2 l_{it} \quad (4)$$

and express logged TFP as a linear function of road density and other determinants:

$$\alpha_{it} = \gamma r_{st} + \psi_{it} \quad (5)$$

where lower case letters denote the log of the respective variables. The term ψ_{it} is a composite term that includes random noise along with plant and time fixed effects, state characteristics and other factors that might affect productivity.

Thus, we estimate the impact of roads in two steps. First, we estimate productivity from Eq. (4). In the second step, we use the estimated plant-level logged TFP as the dependent variable in a regression to estimate the effect of roads, the parameter γ , in Eq. (5).

We now turn to discuss our identification strategy.

4.2. Identification

An important challenge in estimating the impact of roads on economic activity is that road placement is likely to be non-random (Banerjee et al., 2020; Shatz et al., 2011; Duranton et al., 2014, etc.). That is, roads may be more prevalent in areas which have higher economic activity or, if economic activity is low and roads are built to encourage economic activity. We overcome this inference problem by exploiting cross-state variation in the strength of centre-state partisan alignment over multiple election cycles that asymmetrically stimulates road building in aligned states. The rest of this section lays out our identification strategy.

Our identification strategy utilizes the political-economics of road building in India. As already mentioned in Section 2, the central government funds the majority of road building projects whereas, the state governments are responsible for its implementation. Because of this vertical dependence, the centre can strategically relax the economic and non-economic obstacles associated with implementing public projects that favour road building in aligned states, for example.⁸ It is well documented that in decentralized democracies, partisan alignment of lower-level jurisdictions with the centre often leads to ‘tactical’, rather than ‘programmatic’ federal transfers that systematically affect the provision of local public goods (Solé-Ollé, 2013; Solé-Ollé and Sorribas-Navarro, 2008; Sengupta, 2011; Khemani, 2003).

To operationalize our instrumentation strategy, we construct a variable *alignment strength* by multiplying the seat share of a state’s ruling party i.e. the party winning the maximum number of seats in a state legislative assembly election, by plus one if the state’s ruling party is aligned with the centre and by minus one if it is not aligned with the centre. To this end, we obtain data on all state assembly election results from the Election Commission of India spanning the period of our study. The political climate in India has witnessed significant changes during this period. The National Democratic Alliance (NDA), a coalition headed by the nationalist Bharatiya Janata Party (BJP) was at the centre from 1998 until 2003. From 2004 until 2014, it was the Congress Party-led coalition – the United Progressive Alliance (UPA) – that held the centre. We carefully map the various state parties with the coalition at the centre to account for any re-configurations over time as well as parties dropping out during the life of the coalition government.⁹ Finally, we drop from our analysis union territories that do not hold legislative elections and state-year pairs that were under the direct rule of the President for at least 100 days in a given year.

In all our regressions using state-level road density, we control for several observable state-specific characteristics to maintain the conditional exclusion restriction assumption. We interact key state-specific characteristics with a time trend to pick up differential impact along these confounders and include it in our empirical model. In addition, we control for the time-varying density of a state’s railway and electricity network. Finally, we soak up unobservable time-invariant plant-level factors by including plant and time fixed effects. Once we include the full set of controls, we are reasonably confident in pinning down the impact of roads on productivity.

However, we face two main threats to our identification strategy. First, electoral outcomes might be affected by road building due to reverse feedback. To alleviate this concern, we use alignment strength that is lagged by two years, on average. In addition, elections in India are staggered across groups of states which further reduces the possibility of simultaneity bias. And, as mentioned above, we control for a range of potential confounders in our regression to ensure that omitted variables are not affecting our results. However, we cannot completely rule out the potential endogeneity of the instrument. To this end, we use the method developed by Nevo and Rosen (2012) to generate bounds on the impact of roads even if the instrument is itself potentially endogenous. We discuss this in Section 5.2.1 where we find that our main result is robust to relaxing the assumption of strict instrument exogeneity.

Secondly, the exclusion restriction might not be met. This might happen if, for instance, the instrumental variable – alignment strength – were to affect productivity through some other channel besides road building even after including the full set of controls. We consider two plausible violations that might arise: alignment strength might affect productivity through a clientelistic relationship between the centre and the states or, due to industry-specific lobbying. In Section 5.2.2, we show using the method developed by van Kippersluis and Rietveld (2018), that our main result is robust to such plausible violations in the exclusion restriction.

In the next section, we discuss the regression results.

⁸ As Wilkinson (2006) documents, when Indian states hold competitive elections politicians announce several infrastructural projects with the aim of building political support and to raise money for campaign finance. Arulampalam et al. (2009), Johansson (2003), Baskaran et al. (2015) and Bracco et al. (2015) etc. show the electoral incentives of targeting public goods while Bohlken (2019) highlights the importance of partisan alignment in reconciling the incentives for private rents (that ministers keep to themselves, as argued in Lehne et al., 2018) and providing public goods and controlling wider corruption in infrastructure projects.

⁹ For instance, the All India Anna Dravida Munnetra Kazhagam (AIADMK), a leading regional party in Tamil Nadu was part of the NDA alliance in parliamentary elections in 1998 but withdrew its support a year later leading to the BJP government’s collapse and an early re-election in 1999 where AIADMK realigned with the Congress Party. We update the ruling coalition to account for any changes at the centre.

5. Results

To conduct our analysis, we construct a weakly balanced panel of over 205,000 establishments during 1998–2012. In all regressions, we use the multiplier supplied by ASI to ensure that our results are representative of the entire population of organized manufacturing plants in India.

5.1. Main results

Our objective is to estimate the effect of road density on manufacturing productivity. Thus, we estimate the following regression:

$$\log(TFP_{it}) = \gamma \log(R_{st}) + \delta X_{st} + \theta_i + \phi_t + \mu_{it} \quad (6)$$

where $\log(TFP_{it})$ is the logged productivity of plant i at time t , which corresponds to α_{it} in Eq. (5). $\log(R_{st})$ is the log of road density with γ the main coefficient of interest. X_{st} is a vector of state-specific characteristics (in logs) from census 2001 interacted with time to allow for differential time-trends according to these characteristics. Besides, we include time-varying infrastructure controls to purge the effect of other key infrastructure – railways and electricity access. θ_i denotes plant fixed effects while ϕ_t denotes time fixed effects. μ_{it} is an error term clustered at the state-industry-year level to account for possible correlations across plants within a state-industry-year block.

However, as noted earlier, roads are likely to be non-randomly placed. Roads might be placed to support economic activity which would positively bias γ , the ols estimate of the coefficient on road density in Eq. (6). The bias would however be negative if roads were instead placed in remote areas with underdeveloped manufacturing to promote economic activity. In the Indian context, the recent impetus on rural roads construction under the Prime Minister's flagship rural roads scheme (PMGSY), which coincide with the period of our study, provides a strong case for our prior that there is negative selection in road placement (see, for example, Aggarwal (2018) and Asher and Novosad (2020), among others).¹⁰ We test this formally using the method proposed by Oster (2019) which confirms our prior intuition that negative selection affects road placement in our data.¹¹ We therefore use an IV estimator that accounts for endogeneity in road placement.

5.1.1. First stage estimates

For *alignment strength* to be a valid instrument, it must be: (a) strongly associated with the endogenous treatment variable – logged road density; (b) be unrelated with other confounders; and (c) satisfy the exclusion restriction i.e. it should affect the outcome variable, logged TFP, only through the road density channel.

Our first stage regression is of the form:

$$\log(R_{st}) = \delta AS_{st} + \zeta X_{st} + \theta_i + \phi_t + v_{it} \quad (7)$$

where $\log(R_{st})$ denotes logged road density in state s at time t . AS_{st} indicates the strength of centre-state partisan alignment. v_{it} is the error term. θ_i and ϕ_t are the plant and year fixed effects, respectively.

Table 2 shows results from the first-stage regression. Column 1 regresses the endogenous variable, logged road density, on the instrument, *alignment strength*, after conditioning on plant and year fixed effects. Column 2 additionally includes year interacted state controls and infrastructure controls to the model estimated in column 1. The results from both columns 1 and 2 show that a marginal increase in *alignment strength* is associated with a 0.03% increase in log road density and this effect is statistically significant at the 1% level. We conduct the cluster-robust Kleibergen–Paap F-test to ensure that the instrument is not weak. The results in Table 3 confirms that instrument validity is not a concern in our analysis.

To get a clearer picture, Fig. 4 plots the relationship between the seat share of a state's ruling party on the x -axis and the residualized logged road density on the y -axis for aligned and non-aligned states, respectively. The residualized logged road density is the residual obtained from regressing logged road density on year fixed effects. The left panel plots the relationship between seat share and residualized log road density for aligned states whereas, the panel on the right shows the relationship for non-aligned states. It is clear from the left panel that the centre rewards states with a large majority for their support. Moreover, the steeper gradient for aligned states means that the centre actively rewards its core supporters.

5.1.2. Second stage estimates (IV2SLS)

In the second stage, we estimate the following equation by IV2SLS:

$$\log(TFP_{it}) = \gamma \widehat{\log(R_{st})} + \delta X_{st} + \theta_i + \phi_t + \mu_{it} \quad (8)$$

which plugs-in the estimated $\widehat{\log(R_{st})}$ from Eq. (7) to yield consistent estimates of γ , the elasticity of TFP with respect to road density.

¹⁰ PMGSY, announced in December 2000 aimed at bringing all villages with a population of at least 500 within reach of the nearest market by all-weather roads by 2007.

¹¹ The method proposed by Oster (2019) helps to understand the degree of selection due to omitted variables. The test is based on the assumption that the relationship between the treatment and the observables can inform the relationship between the treatment and the unobservables. The intuition behind this method is that omitted variable bias is proportional to coefficient movements once controls are included, but only when such coefficient movements are scaled by the movement in the model R-squares. To be precise, the test shows that the unobservables would need to be -0.11 as important as the observables to produce a treatment effect of zero.

Table 2

First stage: Effect of IV on road density.

| Dependent variable: log(Road density) | (1) | (2) |
|---------------------------------------|----------------------|----------------------|
| Alignment strength (AS) | 0.0003*** (0.000) | 0.0003*** (0.000) |
| Year FE | Yes | Yes |
| Plant FE | Yes | Yes |
| Year \times State controls | No | Yes |
| Infrastructure controls | No | Yes |
| RMSE | 0.176 | 0.149 |
| Observations | 206425 | 205351 |

Notes: Table shows first stage estimates from regressing log(Road density) on Alignment strength (AS) conditional on the full set of controls. AS is the seat share of a state's ruling party multiplied by 1 if it is aligned with the centre and it is multiplied by -1 if it is not aligned with the centre. See Table 3 for the full list of controls. Standard errors clustered at the state-industry-year level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

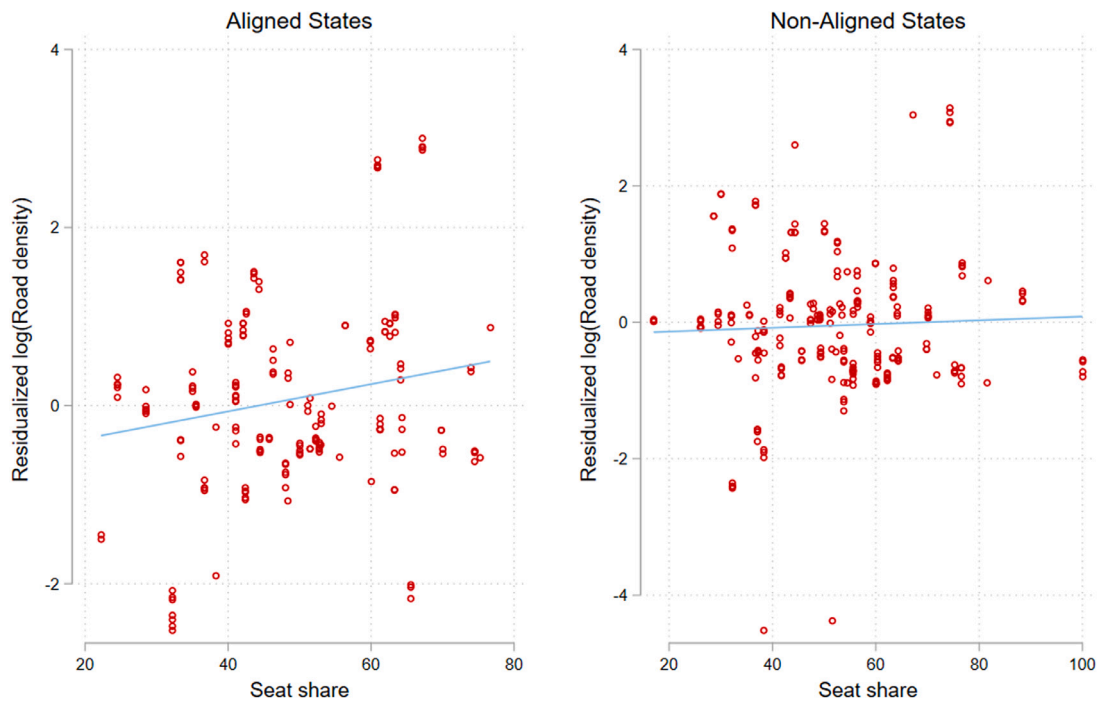


Fig. 4. Relationship between the seat share of a state's ruling party and residualized log(Road density). The latter is the residual obtained from regressing log(Road density) on year fixed effects. The panel on the left plots the relationship for aligned states whereas, the panel on the right relates to non-aligned states.

Table 3 presents the estimated effect of road density on TFP, the coefficient γ . Column 1 in Table 3 regresses logged road density on logged value-added TFP (TFPVA) using the fixed-effect least-squares method. Column 4 does the same except that it uses logged gross-output TFP (TFPGO) as the dependent variable. The coefficients on logged road density in columns 1 and 4 are both positive and statistically significant at the 1% level. However, least squares estimates are biased when road placement is endogenous.

Columns 2 and 3 in Table 3 present results from estimating Eq. (8) by the instrumental variable two-stage least squares (IV2SLS) method. Column 2 instruments logged road density with alignment strength and includes a set of plant and year fixed effects. Column 3 additionally includes year interacted state-specific characteristics and controls for a state's railway density and power infrastructure. Columns 5 and 6, present IV estimates with TFPGO as the dependent variable and are otherwise similar to columns 2 and 3. The results show that a 1% increase in road density raises valued-added manufacturing TFP by about 0.25% (column 3) and a slightly larger effect of 0.26% when we consider gross-output TFP (column 6). Thus, moving a state from the 10th percentile of the road density distribution to the 50th percentile will result in raising TFP by about 13%. In all regressions we cluster the standard errors at the state-industry-year level. Moreover, as we show next, these results are robust to imperfections in the instrument and to a battery of other sensitivity checks.

Table 3
Effect of road density on TFP: Plant-level analysis.

| Dependent variable: log(TFP) | TFP (Valued added) | | | TFP (Gross output) | | |
|------------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| log(Road density) | 0.014*** (0.004) | 0.239*** (0.061) | 0.252*** (0.067) | 0.014*** (0.004) | 0.250*** (0.063) | 0.262*** (0.069) |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Plant FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year \times State controls | Yes | No | Yes | Yes | No | Yes |
| Infrastructure controls | Yes | No | Yes | Yes | No | Yes |
| RMSE | 0.139 | 0.143 | 0.141 | 0.138 | 0.142 | 0.141 |
| Method | OLS | IV | IV | OLS | IV | IV |
| KP F-test | | 32.39 | 31.09 | | 32.39 | 31.09 |
| Clusters | 6390 | 6477 | 6349 | 6390 | 6477 | 6349 |
| Observations | 207826 | 206425 | 205351 | 207826 | 206425 | 205351 |

Notes: In cols.(1)–(3) the dependent variable is log value-added TFP whereas, in cols.(4)–(6) it is log gross-output TFP both estimated using the ACF method. All regressions include a constant term, plant fixed effects and year fixed effects. Cols.(1) and (4) show results from an ols panel fixed effects model. Cols.(2) and (5) show results from a panel IV2SLS model where log(Road density) is instrumented by *Alignment strength* (AS). Cols.(3) and (6) show results from a panel IV2SLS model that additionally controls for time-interacted state characteristics (in logs): population, literacy, total main and marginal workers, and state-level time-varying infrastructure controls: railway density and access to electricity. RMSE = Root Mean Squared Error; KP F-test is the Kleibergen–Paap weak identification test of instrument validity. Standard errors clustered at the state-industry-year level. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table 4
Effect of road density on TFP: Imperfect IV.

| Variable | Lower bound(CI) | LB(Estimator) |
|-------------------|-----------------|---------------|
| log(Road density) | [0.092 | (0.278 |

Notes: One sided bound computed using the Imperfect IV method by [Nevo and Rosen \(2012\)](#). The direction of correlation between the endogenous variable and the unobservable is assumed to be negative.

5.2. Identification tests

As mentioned earlier, there remains some possibility that our instrument is not entirely exogenous. To address this we have taken two steps already: our electoral outcomes are lagged by two years on average to reduce potential concerns regarding reverse feedback. Secondly, we included a list of controls to reduce bias due to omitted variables. In addition to these two steps, we undertake two empirical tests to ensure that our results are robust to imperfections in the instrument: First, we implement the Imperfect IV test to check the sensitivity of the IV estimate to the instrument itself being potentially endogenous (Section 5.2.1). And secondly, we test for plausible violations in the exclusion restriction (Section 5.2.2). Taken together, we present compelling evidence that our results are reliable.

5.2.1. Imperfect instrumental variable (IIV)

One concern with the IV2LS estimate is that the instrument might not be strictly exogenous which violates the assumption of strict instrument exogeneity. In our case, this might arise if the instrument, AS , were to be correlated with the unobserved error, μ i.e. if $Cov(AS, \mu) \neq 0$ in Eq. (8).

The Imperfect Instrumental Variable (IIV) approach by [Nevo and Rosen \(2012\)](#) allows us to estimate bounds on the effect of logged road density (the treatment effect) by replacing the strict instrumental exogeneity assumption with an assumption about the sign of the correlation between AS and μ . Specifically, we assume that: (a) the correlation between $\log(R_{st})$ and μ is negative. To fix our ideas, we can think of μ as a latent variable that captures the level of local development or bureaucratic efficiency. Given the recent impetus on connecting remote habitations where manufacturing is typically underdeveloped such as under the Prime Minister's rural-roads project (PMGSY), it might be a reasonable assumption that $Cov(\log(R_{st}), \mu)$ is negative (i.e. roads are provided in places with lower levels of development or weaker bureaucratic efficiency) and that AS is less correlated with unobserved μ than $\log(R_{st})$.

Table 4 shows that the estimated lower bound on logged road density is 0.28% (with a confidence interval of 0.092) which is within some margin of the IV estimates in either column 3 or column 6 in Table 3. Thus, it shows that our main IV estimate is robust to the instrument itself being potentially endogenous.

5.2.2. Plausibly exogenous regression

The second test is an indirect one to check the sensitivity of the IV estimate to potential violations of the exclusion restriction. The exclusion restriction implies that the direct effect of the instrument on the outcome variable i.e. the parameter λ in Eq. (9)

Table 5
Results from Plausibly exogenous regression.

| Dependent variable: | log(TFP) |
|---|--|
| | Panel A: Effect of log(Road density) on log(TFP) |
| Plausibly exogenous | 0.656*** (0.000) |
| Plausibly exogenous (with uncertainty) | 0.656*** (0.000) |
| | Panel B: Effect of Aligned × Seat Share on log(TFP) |
| Reduced form (full sample) | 0.0001*** (0.0000) |
| Direct effect (zero first-stage group) | -0.0001 (0.0001) |
| | N = 1,642 |
| Direct effect (remaining sample) | 0.0001*** (0.0000) |
| | N = 114,705 |
| | Panel C: Effect of Aligned × Seat share on log(Road density) |
| First stage (full sample) | 0.0003*** (0.000) |
| First stage (zero first-stage group) | 0.0002 (0.0002) |
| First stage (remaining sample) | 0.0001*** (0.0000) |

Notes: Standard errors in parenthesis clustered at the state-industry-year level. The row plausibly exogenous is estimated by plugging-in the mean, μ_λ from the zero-first-stage sub-group and assumes the standard deviation $\Omega_\lambda = 0$. The row ‘with uncertainty’ follows Imben’s rule and uses $\Omega_\lambda = (0.125 \sqrt{S_0^2 + S_{-0}^2})^2$ where S_0 and S_{-0} are the sample standard deviations of the zero-first-stage sub-sample and its complement set, respectively. AS = Aligned × Seat Share (see Table 3). $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

below, is strictly zero.

$$\log(TFP_{it}) = \gamma \log(R_{st}) + \lambda AS_{st} + \varepsilon_{it} \quad (9)$$

In the above equation, AS_{st} is the instrument, alignment strength. $\log(TFP_{it})$ is the logged productivity of the i^{th} plant and $\log(R_{st})$ denotes logged road density. ε_{it} is a composite term which includes the full set of fixed effects, control variables and random error.

For any IV, one cannot formally test if the exclusion restriction is satisfied as the errors are unobserved. However, one indirect way to check for this is an informal test (see Bound and Jaeger, 2000; Altonji et al., 2005; Angrist et al., 2010).¹² The idea is as follows: In a sub-sample for which the effect of alignment strength (the IV) on logged road density (the endogenous treatment variable) is zero (i.e. the first stage is zero), the effect of the IV on the logged TFP (i.e. the reduced form) should also be zero if the exclusion restriction is met. As van Kippersluis and Rietveld (2018) note, even though this informal test cannot verify if the exclusion restriction is met, it ‘builds confidence’ that the IV satisfies the exclusion restriction and is a ‘convincing piece of evidence’.¹³

Here, we consider two plausible threats to the exclusion restriction. First, it is possible that alignment strength affects productivity due to a clientelistic relationship between the centre and the states. Secondly, the instrument might affect productivity through industry-specific lobbying. To check how sensitive our IV estimates are to plausible violations of the exclusion restriction, we select a sub-sample for which λ ought to be zero in the first-stage equation. An appropriate sub-sample where this should hold in principle is one where: (a) elections were closely contested i.e. where the margin of victory of the ruling party in terms of vote share was less than 5 per cent which rules out clientelism as a motive; and, (b) manufacturing firms that are a heterogeneous group and hence lacks the cohesiveness to lobby effectively. To satisfy the latter restriction, we select a sub-sample of manufacturing plants classified as ‘Manufacturing not elsewhere classified (n.e.c)’ according to its 3-digit NIC code. The sub-sample that meets both these restrictions forms our zero-first-stage sub-sample.

Panel B in Table 5 presents results from reduced form regressions whereas, Panel C relates to first-stage regression results. We observe that the coefficient on alignment strength at the first stage (Panel C) and the reduced form (Panel B) for the zero-first-stage sub-sample is negligibly small and not statistically different from zero. The row plausibly exogenous in Panel A is estimated by

¹² Informal tests that run auxiliary regressions have been used to check the validity of the instrument by Altonji et al. (2005) who investigate whether religious affiliation is a valid instrument for attending Catholic schools. To do this, they look at a sub-sample of public eighth graders among which almost nobody subsequently attends a Catholic school. They find direct effect of the IV on the outcome variable for the sub-sample which leads them to conclude that the IV is inappropriate. On the other hand, Angrist et al. (2010) use twin births and same-sex siblings as IVs for the number of children in assessing the quantity/quality trade-off and find that certain ethnicities and mothers who bore their first child at a young age are less affected by the IVs and hence satisfies the zero first stage condition.

¹³ Importantly, van Kippersluis and Rietveld (2018) show that the estimated direct effect of the IV on the outcome variable from the zero-first-stage sub-sample, $\hat{\lambda}$, can act as a prior on the distribution of λ assuming that it follows a normal distribution with mean $\mu_\lambda = \hat{\lambda}$ and variance Ω_λ . Thus, van Kippersluis and Rietveld (2018) extend the ‘plausibly exogenous’ method developed by Conley et al. (2012) in providing guidance on which prior to use in sensitivity checks.

Table 6
Effect of road density on TFP by plant type.

| Dependent variable: log(TFPVA) | Incumbent (1) | Size (2) | Rural (3) |
|--------------------------------|---------------------|------------------|---------------------|
| log(Road density) | 0.294*** (0.086) | 0.138 (0.128) | 0.218*** (0.060) |
| log(Road density) × Incumbent | -0.060* (0.035) | | |
| log(Road density) × Size | | 0.006 (0.007) | |
| log(Road density) × Rural | | | 0.086 (0.071) |
| Year FE | Yes | Yes | Yes |
| Plant FE | Yes | Yes | Yes |
| Year × State controls | Yes | Yes | Yes |
| Infrastructure controls | Yes | Yes | Yes |
| RMSE | 0.141 | 0.134 | 0.142 |
| KP F-test | 13.72 | 13.12 | 5.91 |
| Clusters | 6349 | 6341 | 6349 |
| Observations | 205351 | 203242 | 205351 |

Notes: The dependent variable is value-added log(TFP) in all the three columns of the table. Col.(1) interacts log(Road density) with an *incumbent* dummy, where an incumbent is an establishment aged five years or more. Col.(2) interacts log(Road density) with plant *size* measured as the value of a plant's logged fixed-assets in real terms. Col.(3) interacts log(Road density) with an indicator for *rural* as classified by ASI. In all regressions, log(Road density) is instrumented by 'Aligned × Seat Share' (AS) and the interaction terms in each column are instrumented by AS times the respective covariate. Cols.(1)–(3) respectively control for incumbent status, size and rural location. All regressions include a constant term, plant fixed effects and year fixed effects and control for time-interacted state characteristics (in logs): population, literacy, total main and marginal workers, and state-level time-varying infrastructure: railway density and access to electricity. RMSE = Root Mean Squared Error; KP F-test is the Kleibergen–Paap weak identification test of instrument validity. Standard errors clustered at the state-industry-year level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

plugging-in the mean, μ_λ from the zero-first-stage sub-group and assumes the standard deviation $\Omega_\lambda = 0$. It shows that the effect of logged road density on logged TFP is statistically significant and much larger than our IV2SLS estimates in Table 3. This suggests that our IV estimates are robust to plausible violation of the exclusion restriction and that such violations will inflate our IV estimates in Table 3.¹⁴

5.3. Extensions

In this section, we consider the heterogeneity in the impact of roads on manufacturing TFP by plant-specific characteristics (Section 5.3.1), and by road type (Section 5.3.2).

5.3.1. Impact heterogeneity by plant characteristics

Plants differ widely in their productivity even within narrowly defined industries (Syverson, 2011; Bartelsman et al., 2013). This is true in our data where a plant at the 90th percentile of the productivity distribution produces, on average, almost fifty per cent more output with the same set of inputs than plants at the 10th percentile. Thus, the effect of policy distortions is likely to vary across heterogeneous producers (Restuccia and Rogerson, 2008). This means that the gains from a fall in transportation costs will be different across producers which vary in age, size and location. In this section, we examine the role of plant-specific heterogeneity in affecting the relationship between roads and TFP.

Young versus incumbent plants: We begin by examining whether roads affect younger plants, i.e. establishments under the age of 5, differently than incumbents that fall above this age threshold. We hypothesize that the incremental TFP gains from a denser road network are greater for young plants than it is for incumbents. To examine this hypothesis we create a variable, *incumbent*, that takes a value of one if the establishment had been in existence for five years or more. The incumbent dummy variable takes a value of zero for young plants formed less than five years ago.

Column 1 in Table 6 interacts logged road density with the *incumbent* dummy. Because road density is endogenous, so is the interaction term. Hence, we interact the incumbent dummy with the IV and use this as an instrument for the interaction term.

¹⁴ In addition, the row 'with uncertainty' in Panel A of Table 5 specifies non-zero elements in its variance-covariance matrix. Following the rule of thumb in Imbens and Rubin (2015), which suggests that the normalized difference between the treatment and control group in a regression setting should not exceed one-quarter, we specify $\Omega_\lambda = (0.125 \sqrt{S_0^2 + S_{-0}^2})^2$, where S_0 and S_{-0} are the sample standard deviations of the zero-first-stage sub-sample and its complement set, respectively. The effect continue to be statistically significant at the 1 per cent level of significance reinforcing our estimates in the first row of the table.

Table 7
Effect of road density on TFP by road type.

| Dependent variable: log(TFPVA) | National highway (1) | State highway (2) | Other (3) |
|--------------------------------|-------------------------|----------------------|----------------------|
| log(Road density) | 2.334* (1.388) | 0.167* (0.088) | 1.134*** (0.291) |
| log(Road density) x NH | 0.460 (0.285) | | |
| log(Road density) x SH | | 0.091*** (0.033) | |
| log(Road density) x Other | | | -0.495*** (0.145) |
| Year FE | Yes | Yes | Yes |
| Plant FE | Yes | Yes | Yes |
| Year × State controls | Yes | Yes | Yes |
| Infrastructure controls | Yes | Yes | Yes |
| RMSE | 0.162 | 0.138 | 0.135 |
| KP F-test | 1.62 | 2.04 | 0.88 |
| Clusters | 6349 | 4542 | 4542 |
| Observations | 205351 | 155601 | 155601 |

Notes: The dependent variable is value-added log(TFP) in all the columns of the table. Col.(1) interacts log(Road density) with the logged share of *national highways* (NH) within a state's road network; col.(2) with the logged share of *state highways* (SH); and, col.(3) with the logged share of roads excluding state and national highways from the total (Other). In all regressions, log(Road density) is instrumented by 'Aligned × Seat Share' (AS) and the interaction term in each column is instrumented by AS times the respective covariate. All regressions include a constant term, plant fixed effects and year fixed effects and control for time-interacted state characteristics (in logs): population, literacy, total main and marginal workers, and state-level time-varying infrastructure: railway density and access to electricity. RMSE = Root Mean Squared Error; KP F-test is the Kleibergen–Paap weak identification test of instrument validity. Standard errors clustered at the state-industry-year level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

After controlling for endogeneity, the coefficient on the interaction term shows that the marginal effect of road density on TFP for an incumbent is about 0.06% lower than it is for young plants. Thus, young plants are likely to benefit more from a denser road network than incumbent plants due to lower frictions. Incumbents, over a number of years will have already set up structures and processes, invested in supply chains and private transport systems, built relational networks, and established their markets which would cumulatively work to bypass the need for good transportation infrastructure. Thus, while improved infrastructure will reduce the cost of such adjustments, it will have a smaller impact on incumbent productivity relative to younger establishments.

Small versus large plants: Here, we consider the effect of plant-size in affecting the relationship between roads and TFP. We define size by the value of a plant's fixed assets. Larger plants enjoy higher economies of scale than smaller plants. This implies that, at the margin, the TFP gains from a denser road network is likely to be greater for larger plants in comparison to their smaller counterpart. To test this, we interact logged road density with the logged real value of a plant's fixed assets.

Column 2 in Table 6 shows that, after controlling for endogeneity by instrumenting the interaction term with the product of plant size and the IV, larger plants benefit an additional 0.006% at the margin although this does not turn out to be statistically significant.

Rural versus urban based plants : The effect of roads on manufacturing TFP might also vary by a plant's location. Specifically, we examine if a denser road network affects plants located in rural areas differently than urban-based plants. As the summary statistics in Table 1 shows, TFP is, on average, lower for plants located in rural areas than those in urban areas. Hence, relaxing the transportation constraint might boost the productivity of rural-based plants because of their low initial levels of TFP.

We follow ASI's classification of establishments into rural and urban in our analysis. We define a *rural* dummy that takes a value of one if the plant is in a rural location while it takes a value of zero if the plant is in an urban location. Like before, we interact rural with logged road density and instrument it with the product of the IV and the rural dummy and focus on its coefficient. Because we include plant fixed effects in our model, our variation comes from either plants changing their location from rural to urban or from a re-classification of rural locations into urban or vice-versa during the period of our study. Column 3 in Table 6 presents IV estimates. It shows that the effect of the interaction term is positive but not statistically significant.¹⁵

Thus, we find that the productivity gains due to a denser road network are higher for younger plants than it is for incumbents. Larger plants and plants located in rural areas see increased productivity gains although neither of them turn out to be statistically significant.

Next, we focus on the heterogeneous effect of road type on manufacturing TFP.

¹⁵ As there may not be much variation in the rural dummy within plants over time, we ran a separate regression with industry fixed effects instead of plant fixed effects. Although the coefficient on the rural dummy reverses in sign, it remains statistically insignificant at the 10 per cent level of significance.

Table 8
Alternative measures of TFP and road density.

| | Levinsohn & Petrin method (1) | Woolridge method (2) | District-level log(Road density) (3) | Change in log(Road density) (4) |
|------------------------------|-------------------------------|----------------------|--------------------------------------|---------------------------------|
| log(Road density) | 0.258*** (0.068) | 0.258*** (0.068) | 0.115*** (0.033) | |
| Δ log(Road density) | | | | 0.775* (0.410) |
| Year FE | Yes | Yes | Yes | Yes |
| Plant FE | Yes | Yes | No | Yes |
| Industry FE | No | No | Yes | No |
| Region FE | No | No | Yes | No |
| Year \times State controls | Yes | Yes | No | Yes |
| Infrastructure controls | Yes | Yes | Yes | Yes |
| RMSE | 0.141 | 0.141 | 0.214 | 0.163 |
| KP F-test | 18.00 | 18.00 | 111.60 | 3.77 |
| Clusters | 6349 | 6349 | 10665 | 6349 |
| Observations | 205351 | 205351 | 66296 | 205351 |

Notes: The dependent variables are value-added log(TFP) estimated by the Levinsohn & Petrin method in col.(1), the Woolridge method in col.(2), and ACF method in cols.(3) and (4). Col.(3) presents IV2SLS estimates using district-level road density while col.(4) regresses value-added log(TFP) on Δ log(Road density) and includes the full set of controls. Δ is the first-difference operator. In all regressions, the endogenous covariate is instrumented by 'Aligned \times Seat Share'. All regressions include a constant term. All regressions, except col. (3), include plant fixed effects and year fixed effects and control for time-interacted state characteristics (in logs): population, literacy, total main and marginal workers, and state-level time-varying infrastructure: railway density and access to electricity. The regression in col.(3) includes region, industry and year fixed effects and controls for a plant's age, rural/urban location, plus key district-level variables: logged railway density, area, logged population and the share of villages in a district that have access to electricity. RMSE = Root Mean Squared Error; KP F-test is the Kleibergen–Paap weak identification test of instrument validity. Standard errors clustered at the state-industry-year level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.3.2. Impact heterogeneity by road type

As discussed in Section 2, India's road network is made up of different types of roads — national highways (NH), state highways (SH), and other roads that include rural, urban and project roads. National highways span across state borders and connect major state capitals, ports, industrial centres etc. On the other hand, state highways form the arterial roads within a state that connect national highways, important cities and district headquarters. We group the remaining as other roads which provide last-mile connectivity linking villages and markets. Thus, different types of roads differ in the services they provide. In this section, we focus on the relationship between the type of road and its effect on manufacturing TFP.

Columns 1 in Table 7 interacts logged road density with logged share of national highways within a state's road network while Columns 2 and 3 relate to state highways and other roads respectively. All regressions control for logged road density instrumented by alignment strength and the interaction term in each column is instrumented by the product of the respective covariate and the IV.

The interaction with NH share is large but statistically insignificant (column 1), We find that, after controlling for endogeneity, a marginal increase in NH share raises TFP by 0.46%, although it is not statistically significant (column 1). By contrast, a marginal increase in the SH share raises TFP by 0.09% and this is statistically significant at the 1% level. The share of other roads i.e. the share of roads excluding NHs and SHs is associated with a decline in manufacturing TFP. One main reason for this might be that ASI data includes only manufacturing establishments in the organized sector which are more likely to be within reach of state or national highways. Another point is that 'other' roads are likely to be of poorer quality (e.g. unsurfaced roads) and states with a higher share of poor quality roads are likely to be associated with lower manufacturing productivity. Moreover, other roads consist of a mix of different kinds of roads which vary in their functions and are maintained by disparate organizations making its interpretation less than straightforward (see Section 2).

In the next section, we present results from different robustness checks.

6. Robustness

In this section, we check the sensitivity of the main result to alternative measures of TFP, to different measures of road transportation, estimating the model on different sub-samples, guarding against outliers, and accounting for spatial spillovers.

Alternative measures of TFP: Throughout this paper, we mainly focus on TFP estimated by the Akerberg et al. (2015) method since it avoids the functional dependence problem that affects labour coefficients in Olley and Pakes (1996) type methods, as already mentioned in Section 3.2. Column 1 in Table 8 presents results from regressing logged road density on logged value-added TFP estimated using the Levinsohn and Petrin (2003) method, while in column 2 the dependent variable is logged value-added TFP

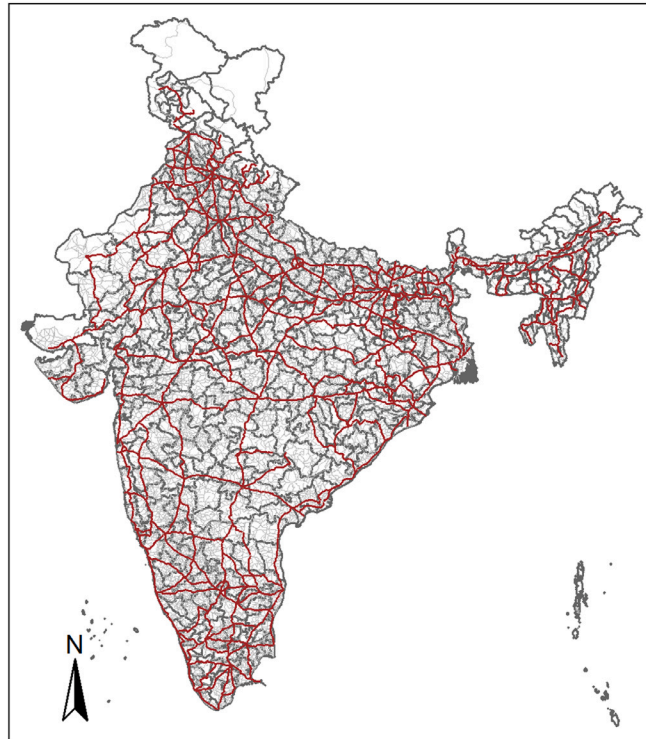


Fig. 5. India's Road Network, 2007

The brown lines indicate primary roads while the grey lines indicate secondary and other roads. The black lines indicate district boundaries. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

estimated by the one-step [Wooldridge \(2009\)](#) method. The effect size from columns 1 and 2 in [Table 8](#) are identical. They are also qualitatively similar to the IV estimates in columns 3 and 6 in [Table 3](#). Thus, our results are robust to the choice of method in estimating TFP.

Alternative measures of road density: Here we examine the robustness of our results to two alternative measures of road density. First we consider a more granular district-level measure of road density. We then consider the effect of using the annual percentage change in road density.

Within-state variability in the distribution of roads might lead to measurement error when using an aggregate state-level measure of road density. Thus, we consider a more granular district-level measure of road density for a smaller sub-sample for which data was available. Our analysis using detailed data on road transportation infrastructure is however restricted to the four years from 2007 to 2010 because district-level road density data were available only for the year 2007 while district identifiers in the matched ASI plant-level data extended only up to the year 2010. We discuss details relating to data construction in [Appendix B](#). The map in [Fig. 5](#) overlays India's road network onto district boundaries that correspond to the 2001 census. [Fig. 6](#), on the other hand, maps the constructed districtwise total road density for the year 2007.

While we trade-off increased granularity against a smaller sample size, these estimates allow us to consider the robustness of our results to a more disaggregated specification. After combining plant-level panel data with district-level roads we regress logged TFP on district-level log road density from 2007 to 2010 and include region,¹⁶ industry and year fixed effects along with controls for a plant's age, its rural/urban location and key district-level variables such as logged railway density, area, logged population and the share of villages in a district that have access to electricity.

Column 3 in [Table 8](#) presents IV estimates from using district-level road density from 2007 to 2010. It shows that a 1% increase in road density raises TFP by 0.12% and this effect is statistically significant at the 1% level. The result using district-level road density is qualitatively similar but much smaller than the estimates in [Table 3](#) using state-level road density. This suggests that our results using state-level road density can be thought of as an upper bound but remain firmly robust to concerns regarding measurement error.

¹⁶ We constructed four regional dummies: North, South, East and West. We have included regional dummies in the regression because district-level roads do not vary during 2007–2010 and our instrument, alignment strength, is measured at the state-level.

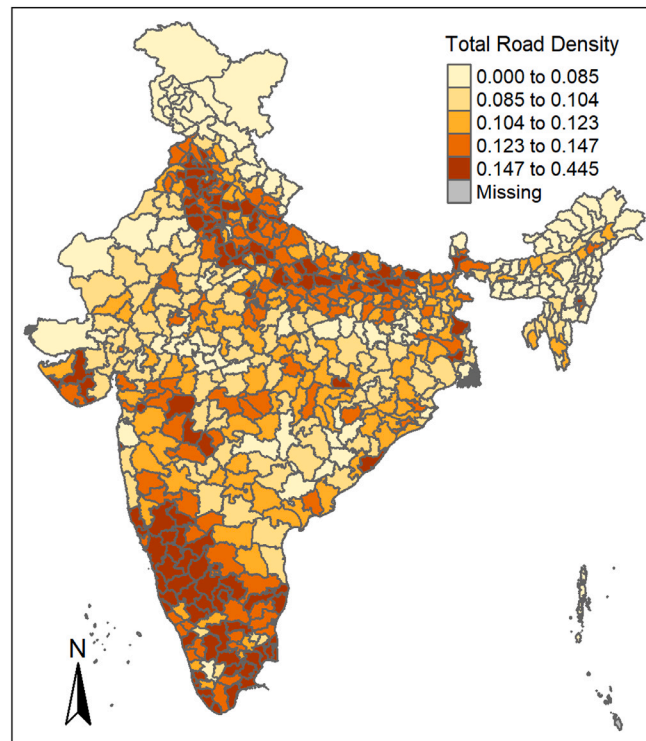


Fig. 6. Districtwise Total Road Density in India, 2007.

Additionally, in column 4, [Table 8](#), we regress logged value-added TFP on the first-difference of logged state-level road density to observe the impact of a change in log road density (i.e. the approximate growth rate of road density) on manufacturing TFP. The result shows that, after controlling for endogeneity, a change in log road density positively affects plant-level TFP.

Excluding Delhi: Delhi has a road density of about 19.3 km/sq.km, the highest for a state in our sample.¹⁷ In comparison, the national average stood at less than 2 km/sq.km. Hence, it is important that we investigate whether the inclusion of Delhi unduly influences our result. Column 1, [Table 9](#) shows that dropping Delhi from the analysis yields an effect size of nearly 0.30%, which is slightly larger but qualitatively similar to the main result in column 3, [Table 3](#). This might perhaps be due to Delhi suffering from the negative effects of congestion. However, the estimates show that the main result is robust to excluding Delhi from the analysis.

Excluding split states: In 2000, three new states – Jharkhand, Chhattisgarh and Uttarakhand – were carved out of Bihar, Madhya Pradesh and Uttar Pradesh, respectively. It is possible that changes in state boundaries rework the institutional mechanisms governing states, which influence manufacturing TFP. For instance, the affected states might adopt industrial policies to attract new firms. One example of this is Uttarakhand, which implemented tax incentive schemes to attract industries that affected productivity (see [Chaurey, 2017](#)). To ensure that our results are robust to these changes, we exclude these six states and re-estimate our regressions. Column 2, [Table 9](#) shows that excluding these six states does not affect our main result.

Considering a sub-set of major states: It is possible that relatively smaller states are driving our results. Even though we control for the endogeneity of road placement, it might be worth exploring the sensitivity of the main result to dropping the smaller states from our sample. Column 3, [Table 9](#) shows that the effect size for a sub-sample of 18 major states¹⁸ is slightly larger than the IV estimates from the full sample in column 3, [Table 3](#). Hence, we can rule out that the fringe states are affecting our results.

Treating TFP outliers: One important issue in working with plant-level data is that outliers might unduly influence our estimates. To avoid this, we winsorize the top 1% and the bottom 1% of the distribution of plant-level logged TFP and then re-run our regressions. Column 4, [Table 9](#) presents results after 1%/99% winsorization of the dependent variable. It shows that a 1% increase in road density raises TFP by about 0.22%, which is slightly smaller than the main IV estimates in column 3, [Table 3](#). This shows that our results are robust to outliers.

Spatial spillover: The relationship between road density and TFP might be affected by spatial spillovers. If spatial spillovers are negative it will upward bias the estimates whereas, if spatial spillovers are positive it will downward bias the estimates ([Donaldson, 2015](#)). We discuss how spatial spillovers might affect estimates in greater detail in [Appendix C](#).

¹⁷ In terms of productivity, however, Delhi ranks fifth among the twenty-eight states that we analyse during the period 1998–2012.

¹⁸ Major states include Punjab, Haryana, Delhi, Rajasthan, Uttar Pradesh, Bihar, West Bengal, Jharkhand, Odisha, Chhattisgarh, Madhya Pradesh, Gujarat, Maharashtra, Andhra Pradesh, Karnataka, Goa, Kerala and Tamil Nadu.

Table 9
Different sub-samples and treating outliers.

| Dependent variable: log(TFPVA) | Excludes Delhi | Excludes split states | Major states | Winsorized |
|--------------------------------|---------------------|-----------------------|---------------------|---------------------|
| | (1) | (2) | (3) | (4) |
| log(Road density) | 0.298*** (0.089) | 0.254*** (0.067) | 0.301*** (0.097) | 0.223*** (0.061) |
| Year FE | Yes | Yes | Yes | Yes |
| Plant FE | Yes | Yes | Yes | Yes |
| Year × State controls | Yes | Yes | Yes | Yes |
| Infrastructure controls | Yes | Yes | Yes | Yes |
| RMSE | 0.143 | 0.141 | 0.147 | 0.127 |
| KP F-test | 13.19 | 18.26 | 11.15 | 17.99 |
| Clusters | 6071 | 6298 | 4826 | 6353 |
| Observations | 199404 | 204455 | 187325 | 205582 |

Notes: The dependent variable is value-added log(TFP) for all the columns in the table. In col.(1), log(TFP) is regressed on log(Road density) for a sub-sample that excludes Delhi; col.(2) excludes the six states affected by state re-organization in November 2000; col.(3) considers 18 major states; whereas, col.(4) considers a 1%/99% winsorized log(TFP) as the dependent variable. In all regressions, log(Road density) is instrumented by 'Aligned × Seat Share'. All regressions include a constant term, plant fixed effects and year fixed effects and control for time-interacted state characteristics (in logs): population, literacy, total main and marginal workers, and state-level time-varying infrastructure: railway density and access to electricity. RMSE = Root Mean Squared Error; KP F-test is the Kleibergen–Paap weak identification test of instrument validity. Standard errors clustered at the state-industry-year level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 10
Including spatially lagged road density.

| Dependent variable: | log(TFPVA) |
|------------------------------------|---------------------|
| log(Road density) | 0.082* (0.048) |
| Spatially lagged log(Road density) | 0.205*** (0.049) |
| Year FE | Yes |
| Plant FE | Yes |
| Year × State ch. | Yes |
| Infrastructure controls | Yes |
| RMSE | 0.128 |
| KP F-test | 5.53 |
| Clusters | 4511 |
| Observations | 170202 |

Notes: The dependent variable is value-added log(TFP). Table shows plant-level analysis for 18 major states after controlling for the spatial lag of log(Road density). log(Road density) is instrumented by 'Aligned × Seat Share' and the spatial lag of log(Road density) is instrumented by the spatial lag of 'Aligned × Seat Share'. The regression includes a constant term, plant and year fixed effects and controls for time-interacted state characteristics (in logs): population, literacy, total main and marginal workers, and state-level infrastructure controls: railway density and access to electricity. Spatial lag of log(Road density) is a row-normalized rook contiguity measure that denotes spatial interaction amongst states. RMSE = Root Mean Squared Error; KP F-test is the Kleibergen–Paap weak identification test of instrument validity. Standard errors clustered at the state-industry-year level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 10 shows that the elasticity of TFP with respect to road density is about 0.082 after accounting for spatial spillover. This is smaller than the IV estimate in column 3, Table 3. The coefficient on the spatial lag term is about 0.21%. Thus, although we find evidence of positive spillovers from a denser road network, the direct effect is robust to accounting for spatial spillovers.

7. Conclusion

In this paper, we examine the causal effect of roads on manufacturing productivity in India using plant-level panel data from 1998 to 2012. A denser road network reduces transport costs and stimulates economic activity which leads to higher manufacturing TFP. However, the placement of roads is likely to be non-random. We overcome the inference problem by using an instrumental variable strategy where we instrument logged road density with cross-state variation in the strength of centre-state partisan alignment over multiple election cycles that are staggered across states. Furthermore, we include plant and year fixed effects and control for socio-demographic characteristics and infrastructure variables to reduce bias from omitted variables.

After controlling for endogeneity, we find that a 1% increase in road density raises manufacturing value-added TFP by about 0.25%. On closer examination, we find that the gains from a denser road network vary by plant-specific attributes and road type. Younger plants are more likely to benefit from a denser road network with highways playing a prominent role. We test the sensitivity of our IV estimates to potential imperfections in the instrument, to using alternative measures of TFP, employing the percentage change in logged road density, estimating the model on different sub-samples, guarding against outliers and accounting for spatial spillovers. To rule out measurement error in using an aggregate state-level measure of road density, we conduct additional analysis using district-level road density for a smaller sub-sample for which data was available and find the results to be qualitatively similar.

Overall, we find that higher road density improves manufacturing total factor productivity. We provide evidence that this relationship is causal. The impact of roads on manufacturing TFP, however, differs by plant-specifics and road type. These findings together suggest that addressing the shortfall in transportation infrastructure can unlock potentially large economic gains.

Appendix A. Variables used to estimate TFP

In estimating TFP, we use the following variables: gross value of output or value added output, total man-days worked, raw materials, power and fuel. The values of gross output and value-added output were converted into real terms by deflating the respective nominal values by industry-specific wholesale price indices (WPI) whereas, expenses on raw materials and power and fuels were deflated by overall WPI. The price indices were obtained from the Office of the Economic Adviser, GoI (see <http://eaindustry.nic.in/home.asp>).

We follow the methodology in Balakrishnan et al. (2000) to measure capital stock which is the method adopted in Topalova and Khandelwal (2011) and Kathuria and Sen (2014). We apply the perpetual inventory method (PIM) and adjust the book value of capital to reflect its replacement cost instead of its historic cost in which they are reported.

To arrive at a measure of capital stock at replacement costs for a base year, we first assume that our base year is 2006. This choice is driven by the fact that we have maximum observations for that particular year. We then compute a revaluation factor assuming that the life of a machine is twenty years, and both the price of capital and the growth of investment changes at a constant rate throughout the assumed twenty years lifetime of capital stock. We use the revaluation factor to convert base year capital to capital at replacement cost in current prices. We then deflate the current value by a deflator based on Gross Fixed Capital Formation (GFCF) series obtained from the Ministry of Statistics and Programme Implementation (MOSPI), GoI. Finally, we obtain the capital stock for every period by summing over investments in subsequent years.

Appendix B. Constructing data for district-level analysis

To analyse the effect of district-level roads on manufacturing TFP, we first obtained GIS data on India's road network for the year 2007 developed by the United Nation's International Steering Committee for Global Mapping (ISCGM).¹⁹ We then overlaid the GIS roads data onto district boundaries and calculated the sum of all road lengths falling within each district boundary. To obtain district-level road density, we divided the total roads length for each district by its respective area.

Next, we constructed a panel data on plant-level economic activity containing district identifiers. However, ASI panel data does not provide information on district identifiers. To get information on the district-level location of plants we separately obtained annual cross-section ASI data which contain district identifiers and matched them with ASI panel data.²⁰ However, some observations were lost in the matching process and information on district identifiers extended only up to the year 2010. Thus, we ended up with a panel data containing both plant and district identifiers from 2007 to 2010.

Appendix C. Spatial spillover

The displacement effect, whereby economic activity relocates from one location to another is one way by which improved transportation infrastructure generates spatial spillovers. As Agrawal et al. (2017) notes, the benefits from improvements in transportation could, in principle, be zero-sum, such that the gains from one location are offset by losses in another location due to a reallocation of economic activity.

Recent research that explicitly accounts for spatial spillovers has examined the effect of spillovers from interstate highways on regional productivity (Holtz-Eakin and Schwartz, 1995), on innovation in the US (Agrawal et al., 2017), and on inventory levels in China (Li and Li, 2013). To isolate the direct effect of roads from spatial spillovers these studies include a spatial lag of highways – the highways in neighbouring counties or provinces – in the empirical model. Here, we adopt a similar approach and consider the robustness of the estimated effect of roads to including a spatially lagged road density term.

To check for the presence of spatial spillovers, we first construct a spatial weight matrix, \mathbf{W} , that accounts for the pattern of spatial interactions among neighbours (the states in this case). We consider two states as neighbours if they share the same border and then row-normalize the spatial weight matrix, \mathbf{W} . We then construct a spatially lagged road density term (in logs), $\mathbf{W}\log(R_{st})$, by pre-multiplying statewise logged road density by the spatial weight matrix and include this term in a regression of the form:

$$\log(TFP_{it}) = \gamma \log(R_{st}) + \rho \mathbf{W}\log(R_{st}) + \delta X_{st} + \theta_i + \phi_i + \mu_{it} \quad (10)$$

¹⁹ This was the only open source GIS data on roads that covered the period of our study and contained detailed information on roads disaggregated by type. The site also contains GIS data for 2016 which is beyond the period of our study.

²⁰ We are only aware of Martin et al. (2017) who use a similar strategy to study the effect of de-reservation policy in Indian manufacturing.

Table 11
Effect of road density on allocative efficiency.

| Dependent variable: | Long-differenced OP covariance |
|------------------------------------|-----------------------------------|
| $\Delta \log(\text{Road density})$ | 0.0089 (0.014) |
| Industry FE | Yes |
| Region FE | Yes |
| RMSE | 0.416 |
| KP F-test | |
| Clusters | 25 |
| Observations | 350 |

Notes: The dependent variable is the long-differenced Olley–Pakes covariance term. The dependent variable, Δ OP covariance (3-year moving average), is regressed on $\Delta \log(\text{Road density})$ and includes industry and region fixed effects. Δ is the long-difference operator which subtracts the value of the respective variable in 2000 from its corresponding value in 2012. Standard errors are clustered at the state level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

where $W\log(R_{st})$ is the spatial lag of logged road density and ρ is the spatial correlation coefficient. We then estimate this model on a sample of 18 major Indian states that account for about 85 per cent of the land area covered by the 28 states in this study. We focus on major states to ensure that geographically smaller states do not unduly influence our results. Moreover, as we noted earlier, dropping smaller states do not alter the result in any meaningful way (see column 2, Table 9).

Appendix D. Allocative efficiency

Here, we present results relating to the effect of roads on the within-industry reallocation of resources across producers. This is important because misallocation of resources across producers affects aggregate productivity. Its effects are enormous and persistent (Hsieh and Klenow, 2009, 2014; Syverson, 2011). Thus, policies that improve allocative efficiency by reallocating resources away from low productivity firms and towards high productivity firms will increase welfare. In this section, we analyse if higher road density affects the allocative efficiency of industries at the 3-digit NIC level.

To do this, we obtain a measure of allocative efficiency by computing the Olley and Pakes (1996) (OP) covariance term for each state-industry pair by year during 1998–2012. First, we define industry j 's logged labour productivity $\omega_j = \sum_{i=1}^{K_j} \theta_{ij} \omega_{ij}$, where K_j denotes the number of plants in the industry, ω_{ij} is the logged value-added per worker of plant i in industry j and θ_{ij} is plant i 's share in the industry as measured by its employment share in the industry. Following Olley and Pakes (1996), we then decompose an industry's labour productivity as:

$$\omega_j = \bar{\omega}_j + \sum_{i=1}^{K_j} (\theta_{ij} - \bar{\theta}_j)(\omega_{ij} - \bar{\omega}_j) \quad (11)$$

where $\bar{\omega}_j$ is the unweighted average of plant-level logged valued-added per worker and $\bar{\theta}_j$ is the unweighted average of plant's industry share. The second term is the OP covariance term which measures the sample covariance between productivity and firm-size and captures allocative efficiency.

Table 11 presents results from a regression where the long-differenced 3-year moving average of OP covariance between 2000 and 2012 is regressed on the long-differenced logged road density during the same period after including industry and region fixed effects. The coefficient on the first-difference of logged density is 0.009 but it is not statistically significant. Our result is somewhat comparable to Asturias et al. (2018), who find that GQ highways improve the OP covariance term by 0.037 during 2001–2006 but are not statistically significant. Thus, the result indicates that higher road density is associated with a positive reallocation of resources from low productivity plants to high productivity plants within industries during the period under study.

Appendix E. K-S test

Here, we present results from the Kolmogorov–Smirnov test which examines the pairwise equality of the empirical CDFs of TFPVA for the years 1998, 2005 and 2012.

Table 12
Kolmogorov–Smirnov test: CDF of log(TFP).

| F(X) vs G(X) | max{F(X) - G(X)} (1) | min{F(X) - G(X)} (2) | Combined (3) |
|---------------------------------|-------------------------|-------------------------|-----------------|
| $F(X_{1998})$ vs. $G(X_{2005})$ | 0.0897 [0.000] | -0.0050 [0.718] | 0.0897 [0.000] |
| $F(X_{1998})$ vs. $G(X_{2012})$ | 0.0102 [0.241] | -0.0846 [0.000] | 0.0846 [0.000] |
| $F(X_{2005})$ vs. $G(X_{2012})$ | 0.0004 [0.995] | -0.1503 [0.000] | 0.1503 [0.000] |

Notes: Table shows results from Kolmogorov–Smirnov test for equality of cumulative distribution functions F(X) and G(X), where X denotes log(TFP). The first row in the table compares the CDF of log(TFP) in 1998 with the CDF of log(TFP) in 2005. Col.(1) tests the hypothesis that $F(X) \leq G(X)$, col.(2) tests $G(x) \leq F(x)$, while col.(3) tests $F(X) \neq G(X)$. P-values in square brackets.

References

- Akerberg, D.A., Caves, K., Frazer, G., 2015. Identification properties of recent production function estimators. *Econometrica* 83 (6), 2411–2451.
- Aggarwal, S., 2018. Do rural roads create pathways out of poverty? Evidence from India. *J. Dev. Econ.* 133, 375–395.
- Agrawal, A., Galasso, A., Oettl, A., 2017. Roads and innovation. *Rev. Econ. Stat.* 99 (3), 417–434.
- Allcott, H., Collard-Wexler, A., O’Connell, S.D., 2016. How do electricity shortages affect industry? Evidence from India. *Amer. Econ. Rev.* 106 (3), 587–624.
- Altonji, J.G., Elder, T.E., Taber, C.R., 2005. An evaluation of instrumental variable strategies for estimating the effects of Catholic schooling. *J. Hum. Resour.* 40 (4), 791–821.
- Angrist, J., Lavy, V., Schlosser, A., 2010. Multiple experiments for the causal link between the quantity and quality of children. *J. Labor Econ.* 28 (4), 773–824.
- Arnold, J.M., Javorcik, B., Lipscomb, M., Mattoo, A., 2015. Services reform and manufacturing performance: Evidence from India. *Econ. J.* 126 (590), 1–39.
- Arulampalam, W., Dasgupta, S., Dhillon, A., Dutta, B., 2009. Electoral goals and center-state transfers: A theoretical model and empirical evidence from India. *J. Dev. Econ.* 88 (1), 103–119.
- Asher, S., Novosad, P., 2020. Rural roads and local economic development. *Amer. Econ. Rev.* 110 (3), 797–823.
- Asturias, J., García-Santana, M., Ramos, R., 2018. Competition and the welfare gains from transportation infrastructure: Evidence from the golden quadrilateral of India. *J. Eur. Econom. Assoc.*
- Balakrishnan, P., Pushpangadan, K., Babu, M.S., 2000. Trade liberalisation and productivity growth in manufacturing: Evidence from firm-level panel data. *Econ. Political Weekly* 3679–3682.
- Banerjee, A., Duflo, E., Qian, N., 2020. On the road: Access to transportation infrastructure and economic growth in China. *J. Dev. Econ.* 102442.
- Bartelsman, E., Haltiwanger, J., Scarpetta, S., 2013. Cross-country differences in productivity: The role of allocation and selection. *Amer. Econ. Rev.* 103 (1), 305–334.
- Baskaran, T., Min, B., Uppal, Y., 2015. Election cycles and electricity provision: Evidence from a quasi-experiment with Indian special elections. *J. Public Econ.* 126, 64–73.
- Baum-Snow, N., Brandt, L., Henderson, J.V., Turner, M.A., Zhang, Q., 2017. Roads, railroads, and decentralization of Chinese cities. *Rev. Econ. Stat.* 99 (3), 435–448.
- Bernard, A.B., Jensen, J.B., Redding, S.J., Schott, P.K., 2012. The empirics of firm heterogeneity and international trade. *Annu. Rev. Econ.* 4 (1), 283–313.
- Bohlken, A.T., 2019. Development or rent seeking? How political influence shapes public works provision in India. *British J. Political Sci.* 1–22.
- Bound, J., Jaeger, D.A., 2000. Do compulsory school attendance laws alone explain the association between quarter of birth and earnings? *Res. Labor Econ.* 83–108.
- Bracco, E., Lockwood, B., Porcelli, F., Redoano, M., 2015. Intergovernmental grants as signals and the alignment effect: Theory and evidence. *J. Public Econ.* 123, 78–91.
- Chandra, A., Thompson, E., 2000. Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Reg. Sci. Urban Econ.* 30 (4), 457–490.
- Chaurey, R., 2017. Location-based tax incentives: Evidence from India. *J. Public Econ.* 156, 101–120.
- Conley, T.G., Hansen, C.B., Rossi, P.E., 2012. Plausibly exogenous. *Rev. Econ. Stat.* 94 (1), 260–272.
- Datta, S., 2012. The impact of improved highways on Indian firms. *J. Dev. Econ.* 99 (1), 46–57.
- Donaldson, D., 2015. The gains from market integration. *Annu. Rev. Econ.* 7 (1), 619–647.
- Donaldson, D., 2018. Railroads of the Raj: Estimating the impact of transportation infrastructure. *Amer. Econ. Rev.* 108 (4–5), 899–934.
- Donaldson, D., Hornbeck, R., 2016. Railroads and American economic growth: A “market access” approach. *Q. J. Econ.* 131 (2), 799–858.
- Duranton, G., Morrow, P.M., Turner, M.A., 2014. Roads and trade: Evidence from the US. *Rev. Econom. Stud.* 81 (2), 681–724.
- Faber, B., 2014. Trade integration, market size, and industrialization: evidence from China’s National Trunk Highway System. *Rev. Econom. Stud.* 81 (3), 1046–1070.
- Fernald, J.G., 1999. Roads to prosperity? Assessing the link between public capital and productivity. *Amer. Econ. Rev.* 89 (3), 619–638.
- Ghani, E., Goswami, A.G., Kerr, W.R., 2016a. Highway to success: The impact of the golden quadrilateral project for the location and performance of Indian manufacturing. *Econ. J.* 126 (591), 317–357.
- Ghani, E., Goswami, A.G., Kerr, W.R., 2016b. Highways and spatial location within cities: Evidence from India. *World Bank Econ. Rev.* 30 (Supplement_1), S97–S108.
- Guasch, L.J., Joseph, K., 2001. Inventories in Developing Countries. World Bank Policy Research Working Paper 2552.
- Gulyani, S., 2001. Effects of poor transportation on lean production and industrial clustering: Evidence from the Indian auto industry. *World Dev.* 29 (7), 1157–1177.
- Holl, A., 2016. Highways and productivity in manufacturing firms. *J. Urban Econ.* 93, 131–151.
- Holtz-Eakin, D., Schwartz, A.E., 1995. Spatial productivity spillovers from public infrastructure: Evidence from state highways. *Int. Tax Public Finance* 2 (3), 459–468.
- Hsieh, C.-T., Klenow, P.J., 2009. Misallocation and manufacturing TFP in China and India. *Q. J. Econ.* 124 (4), 1403–1448.
- Hsieh, C.-T., Klenow, P.J., 2014. The life cycle of plants in India and Mexico. *Q. J. Econ.* 129 (3), 1035–1084.
- Hulten, C.R., Bennathan, E., Srinivasan, S., 2006. Infrastructure, externalities, and economic development: a study of the Indian manufacturing industry. *World Bank Econ. Rev.* 20 (2), 291–308.
- Imbens, G.W., Rubin, D.B., 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

- Government of India, 2018. Economic Survey 2017-18.
- Jedwab, R., Moradi, A., 2016. The permanent effects of transportation revolutions in poor countries: Evidence from Africa. *Rev. Econ. Stat.* 98 (2), 268–284.
- Johansson, E., 2003. Intergovernmental grants as a tactical instrument: Empirical evidence from Swedish municipalities. *J. Public Econ.* 87 (5–6), 883–915.
- Kathuria, V., Sen, K., 2014. *Productivity in Indian Manufacturing: Measurement, Method and Analysis*. Routledge.
- Khemani, S., 2003. *Partisan Politics and Intergovernmental Transfers in India*. The World Bank.
- van Kippersluis, H., Rietveld, C.A., 2018. Beyond plausibly exogenous. *Econom. J.* 21 (3), 316–331.
- Lall, S.V., Shalizi, Z., Deichmann, U., 2004. Agglomeration economies and productivity in Indian industry. *J. Dev. Econ.* 73 (2), 643–673.
- Lehne, J., Shapiro, J.N., Eynde, O.V., 2018. Building connections: Political corruption and road construction in India. *J. Dev. Econ.* 131, 62–78.
- Levinsohn, J., Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. *Rev. Econom. Stud.* 70 (2), 317–341.
- Li, H., Li, Z., 2013. Road investments and inventory reduction: Firm level evidence from China. *J. Urban Econ.* 76, 43–52.
- Martin, P., Mayer, T., Mayneris, F., 2011. Spatial concentration and plant-level productivity in France. *J. Urban Econ.* 69 (2), 182–195.
- Martin, L.A., Nataraj, S., Harrison, A.E., 2017. In with the big, out with the small: Removing small-scale reservations in India. *Amer. Econ. Rev.* 107 (2), 354–386.
- Melitz, M.J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71 (6), 1695–1725.
- Michaels, G., 2008. The effect of trade on the demand for skill: Evidence from the interstate highway system. *Rev. Econ. Stat.* 90 (4), 683–701.
- Mitra, A., Sharma, C., Véganzonés-Varoudakis, M.-A., 2012. Estimating impact of infrastructure on productivity and efficiency of Indian manufacturing. *Appl. Econ. Lett.* 19 (8), 779–783.
- Mitra, A., Varoudakis, A., Véganzonés-Varoudakis, M.-A., 2002. Productivity and technical efficiency in Indian states' manufacturing: the role of infrastructure. *Econom. Dev. Cult. Chang.* 50 (2), 395–426.
- Nevo, A., Rosen, A.M., 2012. Identification with imperfect instruments. *Rev. Econ. Stat.* 94 (3), 659–671.
- Olley, G.S., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. *Econometrica* (ISSN: 00129682, 14680262) 64 (6), 1263–1297.
- Oster, E., 2019. Unobservable selection and coefficient stability: Theory and evidence. *J. Bus. Econom. Statist.* 37 (2), 187–204.
- Redding, S.J., Turner, M.A., 2015. Transportation costs and the spatial organization of economic activity. In: *Handbook of Regional and Urban Econ.* vol. 5, Elsevier, pp. 1339–1398.
- Restuccia, D., Rogerson, R., 2008. Policy distortions and aggregate productivity with heterogeneous establishments. *Rev. Econ. Dyn.* 11 (4), 707–720.
- Sengupta, B., 2011. Provision of public goods in a federal economy: The role of party politics. *Eur. J. Political Econ.* 27 (1), 104–119.
- Shatz, H.J., Kitchens, K.E., Rosenbloom, S., 2011. *Highway Infrastructure and the Economy: Implications for Federal Policy*. Rand Corporation.
- Shirley, C., Winston, C., 2004. Firm inventory behavior and the returns from highway infrastructure investments. *J. Urban Econ.* 55 (2), 398–415.
- Solé-Ollé, A., 2013. Inter-regional redistribution through infrastructure investment: Tactical or programmatic? *Public Choice* 156 (1–2), 229–252.
- Solé-Ollé, A., Sorribas-Navarro, P., 2008. The effects of partisan alignment on the allocation of intergovernmental transfers. Differences-in-differences estimates for Spain. *J. Public Econ.* 92 (12), 2302–2319.
- Storeygard, A., 2016. Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa. *Rev. Econom. Stud.* 83 (3), 1263–1295.
- Syverson, C., 2011. What determines productivity? *J. Econ. Lit.* 49 (2), 326–365.
- Topalova, P., Khandelwal, A., 2011. Trade liberalization and firm productivity: The case of India. *Rev. Econ. Stat.* 93 (3), 995–1009.
- Wilkinson, S.I., 2006. *The politics of infrastructural spending in India*. Department of Political Science, University of Chicago, Mimeo 31.
- Wooldridge, J.M., 2009. On estimating firm-level production functions using proxy variables to control for unobservables. *Econom. Lett.* 104 (3), 112–114.
- World Bank, 2014. *World Bank Enterprise Survey*.