Road to Productivity: The Effect of Roads on Manufacturing TFP in India

Subham Kailthya* Uma Kambhampati†

Abstract

Higher road density boosts manufacturing total factor productivity (TFP) by lowering transport costs. We exploit exogenous variation in state-centre partisan alignment that asymmetrically stimulates road building in aligned states for identification. Using panel data on Indian manufacturing firms during 1998-2012, we find that, a marginal increase in road density raises TFP by 8%, on average. Its impact, however, exhibits significant heterogeneity: smaller firms, incumbents and urban-based establishments see most pronounced gains. These results are robust to potential violation of the exclusion restriction for a range of plausible priors, a placebo instrument test, and other sensitivity checks.

Keywords: Manufacturing; Total Factor Productivity; Transport Infrastructure; Political Economics; India

JEL codes: D24; L60; L92; R40; D72

*Corresponding author, Department of Economics, University of Warwick, Coventry, UK. Email: subham.kailthya@warwick.ac.uk

†Department of Economics, University of Reading

1 Introduction

Transportation plays a vital role in stimulating economic growth. But, in many emerging economies including India and China, inadequate transport infrastructure hinders manufacturing growth (Jedwab and Moradi 2016, Ghani et al. 2016*a*). The problem of inadequate transportation is pervasive and entrenched. According to the recent World Bank Enterprise Survey, 19% of manufacturing firms globally, 21% of firms in South Asia, and one in ten manufacturing firms in India, report inadequate transport infrastructure as a major obstacle (WB 2014). In response, many emerging countries have embarked on ambitious multi-billion dollar infrastructure projects to jumpstart economic activity.[1] Yet, in spite of such large investments in transportation infrastructure, our understanding of how closing the transport infrastructure gap affects manufacturing growth, especially in emerging countries, is far from complete. In this paper, we provide causal estimates of the effect of roads on manufacturing total factor productivity (TFP) within a developing country context.

To date, two broad approaches to study the economic impact of transportation in developing countries have gained track. The first approach focuses on the economic returns from 'showcase' transport projects such as the Golden Quadrilateral (GQ) project in India or the National Trunk Highway System (NTHS) in China that offer in-depth project-specific insight (see Ghani et al. 2016*a*,*b*, Datta 2012, Asturias et al. 2018, Faber 2014, etc.) or ones that concentrate exclusively on rural roads and its impact on village economies (see Bohlken 2016, Asher and Novosad 2018, among others). Another strand of the literature investigates the relationship between a synthetic index of infrastructure and manufacturing activity (see, for example, Mitra et al. 2002, 2012). While inferences derived from studies that take the first approach are internally valid, they may not be generalizable. On the other hand, evidence based on infrastructure indices, which include not only roads but also railways, electricity, and a whole host of other infrastructure variables offer little policy insight in allocating scarce public resources among competing alternatives.[2]

In this paper, we fill this gap and draw robust generalized insights on the causal effect of roads on economic activity. To do this, we combine establishment-level panel data on Indian manufacturing with data on India's entire road transport network during 1998-2012. An important challenge in estimating the impact of roads on economic activity is that roads are not randomly placed (Banerjee et al. 2012, Shatz et al. 2011, etc.).

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

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

For example, roads might be more prevalent in states that trade more, that is, when roads are built to support trade or, if they trade less, that is, when roads are built to encourage trade (Duranton et al. 2014). Moreover, a state's capacity to invest in roads might be correlated with manufacturing productivity. We address the endogeneity of road placement by exploiting exogenous variation in the partisan alignment of states with the centre that asymmetrically stimulates road building in aligned states.[3] After controlling for endogeneity, we find that, TFP rises by 8%, on average, for every incremental increase in road density – the length of roads, in kilometre, per square kilometre land area of the state. Thus, increasing a state's road density from the 10th percentile to the 90th percentile will raise productivity 1.39 times. The headline estimate, however, exhibits significant heterogeneity: smaller firms, incumbents and urban-based establishments see most pronounced gains in TFP.

These results are robust to several checks: potential violation of the exclusion restriction for a range of plausible priors; a placebo instrument test that mimics state-centre alignment; guarding against outliers; relaxing the model assumptions by adopting a more flexible semi-parametric estimation approach; using alternative methods to estimate TFP; estimating the model on different sub-samples; and, using slight variants of the main instrument as a sensitivity check.

There are two main innovations in this study. First, we use establishment-level panel data to estimate TFP, which is then averaged up to the industry-level by state-year to examine its relationship with road density. Previous studies examining the infrastructure-TFP relationship use cross-section data to estimate TFP, which do not control for unobservable firm-specific heterogeneity (in estimating TFP), and is therefore likely to to be biased (see, for example, Ghani et al. 2016*a*). Secondly, our approach provides a political-economic explanation for differences in manufacturing productivity to emerge due to differential provision of roads across states along the lines of state-centre partisan alignment. Hence, this study sheds light on how electoral incentives in a decentralized democracy might affect the state's capacity to provide critical infrastructure, which in turn influences the scale and growth of manufacturing productivity.

We organize the article as follows: Section 2 provides essential background on the transport infrastructure-productivity relationship and discusses the administration, implementation and financing of road projects in India. Section 3 introduces the dataset. Section 4 specifies a simple theoretical model that outlines the relationship between roads and TFP, discusses how we calculate TFP and presents our identification strategy. Section 5 reports

[3] We focus on roads because it is the predominant mode of transportation accounting for 65% of freight movement and 80% of India's passenger traffic (NHAI cited in Ghani et al. 2016*a*).

regression results on the effect of roads on TFP, while section 6 presents robustness checks. Section 7 concludes.

2 Background

2.1 Transport Infrastructure and Economic Activity

Transport infrastructure affects economic activity in several ways: First, a robust transport network lowers freight rates and improves travel times that help integrate distant markets. The arterial network of roads and railways have played a fundamental role in increasing trade volumes through greater market integration. This relationship has been extensively documented (see, for example, Fernald 1999, Chandra and Thompson 2000, Michaels 2008, Duranton et al. 2014, Holl 2016, Redding and Turner 2015, among others).

Secondly, a reliable transport network affects a firm's logistics throughout the different stages of production. It lowers uncertainty at both the procurement and supply stages thereby reducing the need to hold on to costly inventories (Shirley and Winston 2004, Li and Li 2013, Datta 2012). Low inventory pile-up is particularly relevant for developing countries where inventory levels are typically two to five times higher than in the United States, the developed country benchmark (Guasch and Joseph 2001). Along this line, Gulyani (2001) finds that the 'total logistics cost' of inadequate transport infrastructure in the Indian 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. Thus, lower transport costs lead to larger and spatially diversified markets serviced by manufacturers operating at low levels of inventory. These gains translate into better economic organization and larger allocative efficiency, which enhances welfare (Asturias et al. 2018, Ghani et al. 2016*a*). Rural roads, on the other hand, allow reallocation of workers out of agriculture but has no significant impact on economic opportunities (Asher and Novosad 2018).

The constraints posed by inadequate transportation infrastructure are analysed by Storeygard (2016) who use night-time luminosity as a proxy for the economic size of cities in Africa. Using global oil prices to provide exogenous variation in transport costs, they find that the economic activity of African cities that are further away from roads suffer more than cities that are nearer during high oil price years. In India, Ghani et al. (2016*a*) examine the economic impact of the Golden Quadrilateral (GQ) project and find that industries located nearer to highways produce more output and have higher allocative efficiency. In a similar study, they find that the GQ highway affects rural and urban areas differently. While the entry rates of firms are similar across both rural and urban areas, rural areas experience a larger increase in employment and output (Ghani et al. 2016*b*). Moreover, the gains in allocative efficiency are concentrated in larger states that tend to enjoy a cost advantage (Asturias et al. 2018). This paper contributes to the emerging literature on how transport infrastructure affects economic activity in the context of a rapidly emerging economy.

Next, we provide a brief discussion of the administration, implementation and financing of road infrastructure in India.

2.2 Road Infrastructure in India

India's road network spans more than 5 million kilometres. Of this total, national highways (NHs) constitute about 1.6%, state highways (SHs) another 3.3%, while urban and rural roads constitute about 9.5% and 58.3%, 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 PM's Gram Sadak Yojana (PMGSY).

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 UP, MP, 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. Thus, in 2017-18, proposals for 64,000 kms of road were received from state governments, of which about 10000 km of roads/routes as new NHs (Economic Survey, 2017-18).

Constitutionally, the development of rural roads is the responsibility of the state government in India and thus the central government was not directly involved in the funding of rural road projects. However, from the fifth five-year plan of India, 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 2000, the Government of India initiated the Prime Minister's Gram Sadak Yojana Programme (PMGSY), with the objective of connecting all villages having populations over 500 by the end of 2007.

Thus, the state-centre fiscal dependence in funding road construction presents incentives for the centre to vary the size and timing of federal transfers to favour aligned states with the intention of winning elections. We discuss how we exploit this link to identify the effect of roads in section 4.3.

In the next section, we introduce our data and discuss the construction of key variables.

3 Data

We combine detailed data on organized manufacturing in India with state-level data on transportation infrastructure and socio-demographic characteristics to create a state-industry panel running from 1998 until 2012. We discuss this below.

3.1 Manufacturing Data

We use establishment-level panel data on manufacturing activity during 1998-2012 to estimate TFP. We obtained 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. While ASI covers the entire country, we exclude union territories that do not hold legislative assembly elections from our sample. ASI provides exhaustive data on the book values of a firm'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 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. 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. The sampling design considers the state, sector and 4-digit NIC codes in stratifying the sample. Because of the survey design, we use the sampling multipliers provided by ASI throughout the article and cluster the standard errors of the regression coefficients at the state-industry level (see, Abadie et al. 2017 on clustering as a design problem). Thus, our results are representative of the entire population of manufacturing firms in the organized sector in India.

Then, we construct an unbalanced panel of establishment-level economic activity starting from 1998, when ASI first released firm-level panel data, until 2012. Our dataset includes firm identifiers, which allow us to observe individual firms over time. However, ASI, in compliance with the Collection of Statistics Act, 2008, no longer disseminates district identifiers after 2007 and only the state where a firm is located can be identified. Hence, in this study, we conduct our analysis at the state-industry level. Analysis at this level is appropriate here because important decisions relating to infrastructure provision are taken at this level. Moreover, our covariates relating to transportation infrastructure and socio-demography are available only at the state-year level. Another advantage is that it allows us to consider a longer time-frame (that is, to have larger Ts). This not only increases the precision of our estimates, but also enables us to observe any persistence in impact that changes in road density might exert on TFP. Another practical advantage of this approach is that it is forward compatible as new data is released in future.[4]

[4]While we are aware of the matching algorithm used in Martin et al. (2017) to recover the district identifiers of firms by matching economic variables in the panel data (which do not contain district identifiers) with several rounds of cross-sectional data (which contain district identifiers), it is not suitable for our analysis. This is because: first, our covariates are available only at the state-industry level. It is also at this level that decisions on infrastructure provision are actually made. Secondly, reverse engineering the district identifiers, as in Martin et al. (2017), will constrain data up to the year 2007 (after which ASI discontinued releasing district codes) and result in loss of information (due to the dropping-out of unmatched firms).

3.2 Infrastructure Data

Road Density: Road density is the main covariate of interest in this study. We focus on roads since it carries nearly 65% of freight and 80% of all passenger traffic in India (NHAI cited in Ghani et al. 2016*a*). But, using the total length of roads in a state might be problematic because states vary widely in their geographical extent. To overcome this, we focus on road density – the total length of roads, in kilometre, per square kilometre land area of a state – to obtain a normalized metric of access to roads for every state. Focusing on a physical measure rather than the money value to capture market access is advantageous for two main reasons. First, it ensures that our measure is free from distortions due to price volatility. Secondly, a physical measure is less vulnerable to misreporting, which is not an innocuous problem in weak institutional settings that typically lack in transparency and accountability. To compute statewise road density, we obtained data on the total length of roads by state-year from the Ministry of Road Transport and Highways. Summary statistics of our sample indicate that the average road density across states was 4.86 km/sq.km in 2012. Delhi had the highest road density of 21.6 km/sq.km whereas, Jammu and Kashmir had the lowest road density of 0.20 km/sq.km.

Railway Density: 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 operationalise this, we gathered data on the length of railways (in route km) for every state during the study period from the Ministry of Railways and divided it by a state's land area to obtain normalized values of railway density.

Electricity Supply: 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% 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 normalising it by a state's area.

3.3 Socio-Demographic Data

We also include several important socio-demographic variables as controls in our model to reduce omitted variable bias. These include the logarithms of population, literacy rate, the number of main and marginal workers as well as data on total agricultural and industrial workers. Data for these variables are obtained from the 2001 census.

4 Empirical Strategy

4.1 Model Specification

To analyse the effect of roads on manufacturing productivity, we first estimate firm-level productivity as a function of local determinants (see Holl 2016, Martin et al. 2011) using a Cobb-Douglas production function:

$$Y_{it} = A_{it} K_{it}^{\beta_1} L_{it}^{\beta_2} \tag{1}$$

where Y_{it} is the value added output of firm i at time t. L_{it} denotes labour employed in production and, K_{it} is the capital stock. We note that firm i belongs to industry j and is located in state s but we suppress the subscripts for simplicity in the above equation. We then compute average industry-level productivity by state-year as $A_{jst} = 1/n \sum_i A_{ijst}$, which is the average TFP of firms indexed $i = 1 \ldots n$ in industry j within state s at time t. A_{jst} depends on a vector of state-specific characteristics, Z_{st}, and access to roads, $ROAD_{st}$. We adopt this approach because we can only identify the state where a firm is located in ASI panel data, as already mentioned. Since our goal is to examine the effect of roads on industry-level TFP, conducting the analysis at the state-industry level seems to be most appropriate. We model productivity as[5]:

$$A_{jst} = exp[\gamma(ROAD_{st} + Z_{st})] \tag{2}$$

We take log on both sides of eq.(1) to get:

$$y_{it} = \beta_1 k_{it} + \beta_2 l_{it} + \alpha_{jt} \tag{3}$$

[5] Our results hold even if we assume $A_{jst} = \gamma(ROAD_{st})Z_{st}$. This will yield elasticity estimates, which we discuss in section 5.4.

and for industry-level log TFP values from eq.(2) to get,

$$\alpha_{jst} = \gamma ROAD_{st} + \eta_{jst} \tag{4}$$

where lower case letters now denote the log of the respective variables. The term η_{jst} might further include state characteristics, state-industry fixed effects or year effects that affect industry-level productivity along with a random noise term.

We estimate the impact of roads in two steps. First, we estimate firm-level TFP based on eq.(3) and obtain industry-year average TFP by state. In the second step, we use this productivity measure as the dependent variable to estimate the effect of roads in eq.(4).

4.2 Measuring TFP

In this study, we mainly focus on value-added TFP, although we note that our results are almost identical if we use the gross-output definition. 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 TFP as the residual of a Cobb-Douglas production function as:

$$\alpha_{it} = y_{it} - \hat{\beta}_1 k_{it} + \hat{\beta}_2 l_{it} \tag{5}$$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the input coefficients of capital and labour respectively, estimated at the 3-digit NIC industry classification using the Ackerberg et al. (2015) method (henceforth, ACF). We use the ACF method because: first, it accounts for the endogeneity of input choice in estimating productivity. The endogeneity arises because firms 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. Thus, while OP and LP invert investment (OP) and intermediate input (LP) demand functions that are unconditional on labour input, ACF inverts investment or intermediate input that are conditional on labour input to overcome functional dependence and correctly identifies the labour coefficient in the first stage (see Ackerberg et al. 2015 for a detailed discussion on the methodology, and Arnold et al. 2015 for a recent application of this method). We note that our results using the ACF method are consistent with both the LP and the Wooldridge (2009) methods of estimating TFP.

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 firm-level TFP by 3-digit NIC industry classification using 264 thousand observations covering 54 thousand establishments (see Appendix A for the variables used in estimating TFP). Next, we normalize log TFP by dividing it by the employment-weighted average productivity (in logs) for an industry-year.[6] In Figure 1, we plot the log normalized value added TFP (TFP(VA)) for three different years – 1998, 2005 and 2012. We observe that productivity increased consistently between the years with signs of increasing concentration of high-TFP firms in the industry.

[Figure 1 about here.]

We then calculate industry-level averages of value-added TFP. Table 1 presents summary statistics for TFP (see Table 14 in Appendix B for summary statistics relating to the log normalized gross-output TFP values (TFP(GO)). We observe that TFP is about 0.16 at the median for the entire sample. It is higher for small firms (0.20) than large ones (0.13). Younger entrants and firms located in urban areas are also more productive than their respective counterparts. The 90:10 percentile TFP ratio – a measure of dispersion of productivity – is about 3.7, which illustrates significant heterogeneity in TFP. Thus, firms at the 90$^{\text{th}}$ percentile are about four times as productive as those at the 10$^{\text{th}}$ percentile.

[Table 1 about here.]

We now turn to discuss our identification strategy.

4.3 Identification

An important challenge in estimating the impact of roads on economic activity is that roads are non-randomly placed (Banerjee et al. 2012, Shatz et al. 2011, Duranton et al. 2014, etc.). We overcome this inference problem by exploiting exogenous variation in the partisan alignment of states with the centre, which asymmetrically stimulates road building in aligned states. The premise is that, in a federal democracy, transfers favour states that are politically aligned with the central government (see, for example, Sengupta 2011, Arulampalam et al. 2009 etc.). Politically motivated favouritism then

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

disproportionately increases road density in aligned states since 'visible' road projects prove to be electorally rewarding (Mani and Mukand 2007, Wilkinson 2006, Bohlken 2016, etc.). This exogenous variation in a state's road network, after controlling for railways, electricity supply and other state-specific fundamentals, allow us to maintain the exclusion restriction in identifying the causal effect of roads on TFP. Moreover, in section 5.1.4, we test how sensitive the causal effects are to potential violation of the exclusion restriction for a range of plausible priors.

As already discussed in section 2.2, the central government funds the majority of road building projects whereas, the state governments are mainly responsible for its implementation. Hence, by varying the size and timing of federal transfers that favour aligned states, central governments can strategically manipulate the targeting of 'visible' public goods such as roads (in aligned states) with the intention of winning elections, particularly when political competition is high (Arulampalam et al. 2009, Johansson 2003, Baskaran et al. 2015, Bracco et al. 2015).[7] Thus, in decentralized democracies, partisan alignment of lower-level jurisdictions with the centre might lead to 'tactical', rather than 'programmatic' federal transfers that systematically affects the provision of local public goods (Solé-Ollé 2013, Solé-Ollé and Sorribas-Navarro 2008, Sengupta 2011, Khemani 2003).[8]

To operationalize our instrumentation strategy, we define a ruling party as the political party winning the maximum number of seats in a state for each legislative assembly election held between 1998 and 2012. We then construct a binary variable, *Aligned*, that takes a value of one if the state's ruling party is aligned with the centre (or the central coalition) and a value of zero if it is not aligned. To do this, we obtain data on state assembly election results during the period of our study from the Election Commission of India. During this time period, the National Democratic Alliance (NDA) headed by the Bharatiya Janata Party (BJP) was in power at the centre from 1998 until 2003, whereas, the Congress Party-led United Progressive Alliance (UPA) was in power from 2004 to 2014. With coalition governments at the centre, we carefully map the various state parties with the coalition to account for any re-configurations (including parties dropping out

[7] Moreover, in parliamentary systems with single member districts, the chances of winning for the party is closely aligned with the performance of co-partisans which increases public provision that wins votes while reducing the incentive for corruption. As Bohlken (2016) points out, the role of partisan alignment is important because it is able to reconcile 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.

[8] In India, the Finance Commission is responsible for determining the extent of equalization transfers from the centre to the states to neutralize vertical imbalance through a formulaic approach. But as Khemani (2007) notes 'formula-based transfers …have limited success in curbing political influence …the formula itself is determined by a political process or is not binding and leaves room for political discretion'.

during the life of the coalition government[9]). Finally, we drop union territories from our analysis because they do not hold legislative elections.

Another concern in estimating the effect of roads on TFP relates to differences in state-specific characteristics or, differential access to other kinds of infrastructure that might confound the relationship. We address this by first, interacting year with key state characteristics to pick up differential impact along these confounders. Secondly, we control for a state's railway and electricity network. And finally, we soak up unobservable fixed heterogeneity by including state, industry and year fixed effects. Once we include the full set of controls, we are reasonably confident in pinning down the impact of roads on productivity.

In the next section, we discuss the regression results.

5 Results

To conduct our analysis, we carefully construct a weakly balanced panel of industry-level TFP by state-year during 1998-2012. This yields 6,714 observations. In all the regressions, we use the multipliers supplied by ASI and weight the observations by an industry's employment in the initial study period to ensure that our results are representative of the entire population of organized manufacturing firms in India.

5.1 Main Results

Our objective is to estimate the effect of road density on manufacturing TFP. Thus, we estimate an equation of the form:

$$TFP_{jst} = \gamma(ROAD_{st}) + \zeta X_{st} + \theta_j + \eta_s + \phi_t + \mu_{jst} \tag{6}$$

where TFP_{jst} is the productivity of industry j in state s at time t, which corresponds to α_{jst} in eq.(4). $ROAD_{st}$ is the road density for a state in a given year and γ its marginal effect – the main coefficient of interest. X_{st} is a vector of year interacted state-specific characteristics in logs that allow for different time trends according to these characteristics

[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.

along with state-level infrastructure controls. θ_j denotes industry fixed effects; η_s denotes state fixed effects and ϕ_t denotes year effects. μ_{jst} is an error term clustered at the state-industry level to account for possible correlations within a state-industry pair over time that might arise due to state-specific industrial policies and also to account for ASI's sampling design.

However, as observed before, roads are non-randomly placed. They might be strategically placed to support trade, or to promote trade. In the former case, eq(6) will overestimate γ, the impact of road density on TFP whereas, it will be biased downward in the latter case. Hence, to yield unbiased and consistent estimates, we instrument road density with state-centre partisan alignment.

5.1.1 First Stage Estimates

For alignment to be a valid instrument, it must be: (a) strongly associated with the endogenous treatment variable, road density, that is, have a non-zero first stage; (b) be unrelated with other confounders; and (c) satisfy the exclusion restriction, that is, it should affect the outcome variable, TFP, only through the road density channel.

How does partisan alignment with the centre affect a state's road density? We first explore this graphically and then provide formal regression results. Figure 2 plots the average annual change in road density between aligned and non-aligned states. It shows that, on average, aligned states witnessed larger changes in road density than non-aligned states.

[Figure 2 about here.]

Our first stage regression is:

$$ROAD_{st} = \delta A_{st} + \kappa X_{st} + \rho_j + \psi_s + \lambda_t + \nu_{jst} \tag{7}$$

where $ROAD_{st}$ denotes road density of state s at time time t. A_{st} indicates whether the state is aligned with the centre. ν_{jst} is the error term while ρ_j, ψ_s and λ_t are the industry, state and year fixed effects, respectively.

Table 2 presents first stage regression results that formally test the relationship between partisan alignment and road density. Column 1 regresses road density on $Aligned_{st}$, after conditioning on state, industry and year fixed effects whereas, column 2 additionally includes year interacted state controls and infrastructure controls. It shows that road density is 0.10-0.17 points higher when a state's ruling party is aligned with the centre. It is statistically significant at the 1% level of significance, which demonstrates that the instrument is highly relevant. However, if the instrument is weakly associated with the endogenous treatment variable, it will bias IV estimates in the direction of OLS (Bound et al. 1995). To rule this out, we conduct the cluster-robust Kleibergen-Paap weak identification test. Table 3 presents results from weak identification tests, which show that our instrument is strongly associated with the endogenous treatment variable. We test the sensitivity of our results to potential violations of the exclusion restriction for a range of plausible priors in section 5.1.4.

[Table 2 about here.]

5.1.2 Second Stage Estimates (IV2SLS)

Our second stage regression is:

$$TFP_{jst} = \gamma(\widehat{ROAD}_{st}) + \zeta X_{st} + \theta_j + \eta_s + \phi_t + \mu_{jst} \tag{8}$$

which plugs-in the estimated $\widehat{ROAD}_{st}$ from eq.(7) to yield consistent estimates of γ, the marginal effect of road density on TFP.

Table 3 presents results from estimating the marginal effect of road density, γ. Columns 1 and 4 in Table 3 present fixed effect least squares regression estimates. While the coefficient of road density is positive, it is not statistically significant. But, as already discussed, fixed effect estimates are likely to be biased due to the non-randomness of road placement. In fact, the OLS estimates are likely to be downward biased since the major thrust of the government has been to extend connectivity to remote habitations where manufacturing TFP is typically low (see Table 1 for differences in rural-urban manufacturing TFP). Moreover, as already mentioned, rural roads constitute about 58% of India's road network which is likely to downward bias OLS estimates.

Columns 2 and 3 in Table 3 present results from an instrumental variable two-stage least squares (IV2SLS) regression model, where we regress value-added TFP on road density, and where the latter is instrumented by the political alignment of a state's ruling party with the centre. While column 1 includes industry, state and year fixed effects, column 2 contains the full set of controls, which additionally includes year interacted state-specific characteristics and infrastructure controls. Columns 5 and 6, present IV estimates relating to gross-output TFP and are similar in every other way to columns 2 and 3. These results indicate that an additional km of road per sq.km area leads to about an 8% point increase in TFP. Moreover, increasing a state's road density from the 10$^{\text{th}}$ percentile to the 90$^{\text{th}}$ percentile will raise productivity 1.39 times ($\approx \exp(0.364)/\exp(0.032)$).

The standard errors are clustered at the state-industry level and the use of sampling weights ensures that these results are nationally representative.

[Table 3 about here.]

5.1.3 Placebo Instrument Test

One concern is whether political alignment indeed affects road density or, if it is merely an artefact of the data. To dispel this concern we conduct a placebo test motivated by Chetty et al. (2009) and Martin et al. (2017), where we randomly assign alignment status to a state during a year and use the constructed variable as our placebo instrument for road density. *Ex-ante*, we would expect placebos to yield insignificant estimates. We conduct 100 placebo runs in total where each draw is from a uniform distribution that preserves the prior that, at any given time, a randomly chosen state is aligned with the centre. Table 4 presents results from one such draw where we find that the placebo instrument is not informative, that is, it has no significant effect on road density and its effect on TFP is negative and not statistically significant, even after conditioning on the full set of controls.

[Table 4 about here.]

Table 5 summarises the results from running the placebo test 100 times for TFP(VA) and TFP(GO), respectively. We get an insignificant result 92 times out of 100 for TFP(VA). For TFP(GO), it is 91 times out of 100. Thus, we can safely reject the effect of placebo instruments – the randomly assigned alignment status – in favour of true alignment status. Only 4 times out of 100 runs do we get a significant positive coefficient for the effect of road density on TFP(VA) at the 5% level of significance (5 times for TFP(GO)) whereas, the coefficient has incorrect sign and is negative in 4 cases, each for TFP(VA) and TFP(GO).

[Table 5 about here.]

[Figure 3 about here.]

Figure 3 plots the empirical cumulative distribution function (CDF) for the 100 placebo tests for TFP(VA). The true coefficients, as shown by the vertical line, is towards the right of the CDFs implying that the placebo test was successful.

5.1.4 Potential Violation of Exclusion Restriction

The exclusion restriction implies that the direct effect of the instrument on the outcome variable, that is, the parameter λ in eq.(9), is strictly zero.

$$TFP_{jst} = \gamma(ROAD_{st}) + \lambda A_{st} + \epsilon_{jst} \tag{9}$$

In the above equation, A_{st} denotes whether state s is politically aligned with the centre at time t. TFP_{jst} is the productivity in the $\mathrm{j^{th}}$ industry and $ROAD_{st}$ denotes road density. ϵ_{jst} is a composite term which includes the full set of fixed effects, control variables and random error.

The exclusion restriction, however, cannot be formally tested as the errors are unobservable. One way to check for potential violation of the exclusion restriction is to run auxiliary regressions as an informal test (see Bound and Jaeger 2000, Altonji et al. 2005, Angrist et al. 2010). The idea is that in a sub-sample for which the first stage (the effect of the IV on the endogenous treatment variable) is zero – the zero-first-stage sub-sample – the reduced form (the effect of the IV on the outcome variable) should also be zero, to satisfy the exclusion restriction. The 'plausibly exogenous' method proposed by Conley et al. (2012) allows one to test the sensitivity of the IV estimator to a range of plausible priors regarding the violation of the exclusion restriction. But, there is little guidance on how to choose the priors. van Kippersluis and Rietveld (2018) extend this approach and show that using the priors from the zero-first-stage sub-sample is an informed way to deal with potential violations of the exclusion restriction. Here, we apply the method suggested in van Kippersluis and Rietveld (2018) to check the robustness of our IV estimates. We identify the sub-sample of manufacturing firms falling under the 3-digit NIC classification 'Manufacturing not elsewhere classified (n.e.c)' as our zero-first-stage sub-sample.

Panels B and C in Table 6 present results from reduced form regressions (that is, the effect of alignment on TFP) and first-stage regressions (that is, the effect of the instrument, aligned, on road density), respectively. We observe that the coefficient on alignment at the first stage (Panel C) and the reduced form (Panel B) for the sub-sample of manufacturing firms classified as 'manufacturing n.e.c' is negligibly small and is not statistically significant.

We then proceed to estimate 'plausibly exogenous' regressions under two separate approaches. The first approach plugs-in the estimator, $\hat{\lambda}$, from the zero-first-stage sub-sample – manufacturing n.e.c – to examine how the causal effect of interest, γ, changes under a plausible violation of the exclusion restriction. The second approach incorporates uncertainty around $\hat{\lambda}$ by specifying non-zero elements in its variance-covariance matrix.

Following the rule of thumb in Imbens and Rubin (2015), who suggest 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.

Panel A in Table 6 presents results from 'plausibly exogenous' regressions. It shows that the effect of road density on TFP is statistically significant and slightly larger in magnitude than the IV2SLS estimates in Table 3. This suggests that our IV estimates are robust to potential violation of the exclusion restriction and any bias that might arise from plausible priors about λ, will put an upward pressure on γ.

[Table 6 about here.]

In the next section, we discuss the extent to which the effect of roads on TFP differ by age, size and location of firms.

5.2 Impact Heterogeneity

Our results so far clearly indicate that higher road density translates into positive gains in manufacturing TFP. What is not clear, however, is whether the impact differ by an firm's size, age or location, a question that might be of particular relevance to policymakers. In this section, we present results that reveal the heterogeneity in the impact of roads along these axes.

5.2.1 Young vs Incumbent Firms

We begin by examining whether roads differentially affect young firms relative to incumbents. We classify young firms as those established in the last four years (relative to the survey-year) whereas, firms above this threshold are the incumbents. We then aggregate TFP(VA) by a firm's age at the state-industry level for every year in the data following this classification. The number of observations is smaller than the full sample as some industries do not have firms in a given year that belongs to the young category or the incumbent category, as the case maybe.

[Table 7 about here.]

Table 7 presents results for young and incumbent firms in the industry in Panels A and B respectively. We focus on columns 3 and 6 that correspond to the full-specification model for TFP(VA) and TFP(GO), respectively. The results indicate that a marginal increase in road density has a significant positive impact on TFP for both young and incumbent firms. However, the magnitude of impact is larger for incumbents. A marginal increase in road density results in a 8.2% point increase in TFP(VA) for incumbents (Column 3, Panel B in Table 7) compared to an 8.1% point increase for younger entrants (Column 3, Panel A in Table 7). This is hardly surprising because younger entrants have already factored in location decisions at the entry-stage. Thus, additional productivity gains from higher road density for younger firms, once they have established, are relatively small. In comparison, incumbents can now restructure their operation in response to higher road density, which gets reflected in higher productivity gains.

5.2.2 Small vs Large Firms

Here, we consider if the effect of higher road density differs by firm-size. To do this, we classify firms as small if the real value of their fixed assets (land, buildings, machinery etc.) is lower than the industry's median value and we consider them as large if their fixed assets exceed the median value. We then aggregate the TFP(VA) by firm size at the state-industry level for every year in the data resulting in about 6,300 observations that relate to small firms and 6,100 observations for large firms.

[Table 8 about here.]

Table 8 presents coefficient estimates of road density on TFP for small and large firms in Panels A and B, respectively. We find that, at the margin, higher road density benefits small firms more than they do large ones. For small firms, increasing road density raises TFP(VA) by 9.5% percentage points (Column 3, Panel A in Table 8), whereas, for large firms, TFP rises by 6.5% points (Column 3, Panel B in Table 8). That productivity gains are higher for small firms might be due to the organizational flexibility of small-sized firms. In contrast, large firms face rigidities in restructuring production, which might constrain the scope for productivity gains that higher road density has to offer.

5.2.3 Rural vs Urban Sub-Samples

In this section, we examine if a denser road network affects firms in rural states differently than firms in urban states. For the purpose of this study, we define a state as rural if its urban population is less than the median value of urban population for all states in the sample.

[Table 9 about here.]

In Table 9, we examine the differential impact of roads on value-added TFP between firms located in rural states (Panel A) and those in urban states (Panel B). Further, we disaggregate the impact of firm's location by its size and age (see Table 15 in Appendix B for the corresponding impact on TFP(GO)). In each column, we condition the estimates on the full set of controls. We find that, higher road density increases TFP in rural states by 8.5% points (Panel A, Column 1 in Table 9) whereas, the increase is about 21% points for urban states (Panel B, Column 1 in Table 9). The larger effects of roads on firms located in urban states might be explained by the presence of complementarities in production that urbanisation offers, which allow firms in predominantly urban states to amplify the benefits arising from higher road density.

Considering the impact of roads on TFP by a combination of size, age and location characteristics of firms, we find that the largest impact of roads is on younger entrants in urban states, where value-added TFP rises by 41.6% points (Panel B, column 4) followed by small-sized firms in urban states where TFP rises by 29.8% points (Panel B, column 2). Thus, the effect of roads on manufacturing TFP is considerably heterogeneous and differs by age, size and location of firms.

In the next section, we turn to results from long-differenced regressions.

5.3 Long-Differenced Results

Thus far, we consider the impact of roads on manufacturing productivity using state-industry panel data during 1998-2012. However, building roads takes time and road infrastructure might need to reach a tipping point before it begins to affect productivity. On the other hand, it is possible that the impact of roads is only transitory, spurring a short burst in productivity as roads are built/ upgraded, but decays over time as firms readjust their operations and internalise the benefits. To investigate this, we consider the long horizon impact of road density on industry-level TFP by long-differencing our dependent variable, TFP(VA) and TFP(GO), as well as our main covariate, road density, over a span of 14 years. We do this by simply subtracting the values of these variables in 1998 from their corresponding values in 2012. Differencing in this way takes out time invariant factors at the state-industry level. Because our dependent variable is in logs, we can interpret the long-differenced dependent variable as the growth rate in productivity. We then regress the long-differenced variables on the change in road density between 1998 and 2012, which we can write as:

$$\Delta TFP_{js} = \beta(\Delta ROAD_s) + \delta X_s + \theta_j + u_{js} \quad (10)$$

where the prefix Δ denotes that the variable is long-differenced and X_s denotes a vector of state-specific variables. In eq.(10), we examine the impact of a marginal increase in the change in road density between 1998 and 2012, $\Delta ROAD_s$, on the growth rate of productivity, ΔTFP_{js} . We instrument $\Delta ROAD_s$ by the maximum of the consecutive number of years for which a state was aligned with the centre between 1998 and 2012.

[Table 10 about here.]

Table 10 shows results from these estimations. In columns 1 and 2, the dependent variables are respectively $\Delta TFP(VA)$ and $\Delta TFP(GO)$, calculated using the ACF method. After controlling for endogeneity, we find that a marginal increase in the change in road density leads to about an 10% point increase in the TFP growth rate. Thus, increasing road density not only affects the scale of productivity (see Table 3), but also its growth rate (see Table 10).[10]

5.4 Elasticity Estimates

The above results show how road density affects manufacturing productivity either on an annual basis using panel data or, in the longer term, by using long-differenced regressions. Another useful and perhaps more intuitive way to understand the results is to look at elasticity – the percentage change in TFP due to a one percent increase in road density. To obtain elasticity estimates we regress value-added TFP (in logs) on log (Road density) where we instrument the latter with *Aligned* and include the full list of controls. Table 11 presents elasticity estimates. We find that overall, a one percent increase in road density leads to a 11% point increase in TFP. Mirroring earlier results from our semi-log model, the elasticity estimates reveal that the TFP of small firms, incumbents and rural-based establishments are more elastic to changes in road density than their respective counterparts. While the productivity of large-sized firms are the least elastic to road density (8.8%), urban-based firms are the most elastic (32.2%).

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

[Table 11 about here.]

[10]Columns 3 and 4 in Table 10 show that the long-differenced effects of roads on TFP estimated using the LP method are similar to that obtained by employing the ACF method in columns 1 and 2.

6 Robustness

Alternative measure of road density: Here, we treat road density as a a binary variable, *high road density*, that takes a value of 1 if road density exceeds two-thirds of the road density distribution during that year, and is zero otherwise. Relaxing the functional form specification in this way allows us to capture discontinuities in the effect. Table 12, columns 3 and 6, regress high road density dummy on TFP(VA) and TFP(GO) respectively, after including the full set of controls and instrumenting the endogenous treatment variable by *Aligned*. The first stage results show that our instrument is appropriate (see Table 16 in the Appendix B). We estimate the effect of high road density to be about 25%, which is high in comparison to the benchmark estimates in Table 3.

[Table 12 about here.]

Alternative Estimates of TFP: Throughout this article, we mainly focus on TFP estimated by the ACF method since it avoids the functional dependence problem that affects labour coefficients in Olley and Pakes (1996) type methods, as already mentioned. Column 1 in Table 13 presents results from a regression where the dependent variable – TFP (value added) – is estimated using the Levinsohn and Petrin (2003) method (see Table 17 in Appendix B for first stage results). We find that the effect of roads on TFP is robust to the choice of method in estimating productivity. Importantly, the results are quite similar to those obtained by the ACF method.[11]

Treating Outliers: One important issue in working with establishment-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 firm-level log TFP before obtaining industry averages, and then re-run our estimates. Column 2 in Table 13 presents results after 1%/99% winsorization of the dependent variable. We observe that, even after treating outliers in this way, our results remain almost identical and hence our results are robust to to outliers.

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 establishments. One example of this is Uttarakhand, which implemented tax incentive schemes to attract industries that affected

[11]We note that our findings are almost identical even if we use the GMM estimation approach suggested by Wooldridge (2009), but do not show the results to conserve space.

productivity (see Chaurey 2016). To ensure that our results are robust to these changes, we exclude these six states and re-estimate our regressions. Column 3 in Table 13 shows that excluding these six states only marginally lowers the effect of roads on productivity.

Focusing on 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 testing how the causal effect of roads changes when we concentrate on major Indian states.[12] Column 4 in Table 13 shows that the effect size is almost 2.5 times as large as the benchmark estimates in Table 3 for this sub-sample of major states. One plausible reason for this might be the availability of complementary factor inputs in major states, which magnify the effect of roads on TFP.

Alternative Instruments: Thus far, we instrument road density with *Aligned*, a binary indicator for whether a state's ruling party is aligned with the centre during a year. But, does the share of seats of the aligned ruling party in a state affect the provision of roads? To examine this, we interact *Aligned* with the seat share of the ruling party and use this as our instrument (see column 4 in Table 13). This approach effectively weights aligned parties by the proportion of seats they won in the respective legislative assemblies in a state-year. Besides this, it is also important to examine the effect of alignment during an election year. To do this, we interact *Aligned* with a dummy for an election year (see column 5 in Table 13) and use it as an instrument for road density.

We observe that: first, both the alternative instruments are informative (see Table 17 in Appendix B); and, second, the results in columns 5 and 6 in Table 13 are qualitatively similar to the main results in column 3 in Table 3. In fact, as expected, the coefficient on road density with 'Aligned x Election year dummy' as instrument (IV-A) is marginally higher than the benchmark. In contrast, accounting for the seat share of aligned ruling-parties across states (IV-A) slightly lowers the estimate of roads.

[Table 13 about here.]

7 Conclusion

In this paper, we examine the causal effect of roads on manufacturing productivity using panel data from India during 1998-2012. Higher road density reduces transport costs,

[12]We define major states to 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.

which improves firm-logistics, reduces inventory pile-up and extends market size. These gains get reflected into higher manufacturing TFP.

An important challenge in estimating the effect of roads on TFP is that road placement is non-random. For example, roads might be more prevalent in states that trade more, that is, when roads are built to support trade, or if they trade less, that is, when roads are built to encourage trade. Moreover, a state's capacity to invest in roads might be correlated with its manufacturing output. We overcome this inference problem by exploiting exogenous variation in the timing and duration of state-centre partisan alignment across multiple election cycles that are staggered across states. The premise is that in a federal democracy, vertical transfers are likely to favour states that are aligned with the centre, which asymmetrically stimulates road building, a 'visible' public good, in aligned states, after controlling for other infrastructure variables such as railways and electricity supply. We also include state, industry and year fixed effects along with key socio-demographic characteristics and infrastructure variables to reduce omitted variables bias.

After controlling for endogeneity and including the full set of controls, we find that a marginal increase in road density raises TFP by 8%, on average. In elasticity terms, this translates to a 11% point increase in TFP for every 1% point increase in road density. Thus, increasing a state's road density from the 10^{th} percentile to the 90^{th} percentile will raise productivity 1.39 times. These results are robust to potential violations of the exclusion restriction for a range of plausible priors, a placebo instrument test that randomly aligns states with the centre, using a semi-parametric approach to quantify impact, to the choice of methods in estimating TFP, guarding against outliers by winsorizing the dependent variables, running regressions on different sub-samples as well as using slight variants of the main instrument. The effect of roads on TFP, although persistent, is significantly heterogeneous: smaller firms, incumbents and urban-based establishments see most pronounced gains.

We present two main innovations in this study. First, we estimate TFP using establishment-level panel data, after controlling for fixed firm-specific unobservable heterogeneities that previous studies do not accommodate. Secondly, we provide a political-economic explanation for differences in manufacturing productivity to emerge due to differential provision of roads across states along the lines of state-centre partisan alignment. This study therefore sheds light on the way electoral incentives in a decentralized democracy might affect the provision of critical infrastructure, with important implications on manufacturing productivity.

Thus, to conclude, we find that higher road density increases manufacturing productivity. We provide evidence that the relationship is causal. However, the impact of roads on manufacturing TFP differs by age, size and location of firms. Moreover, higher road density affects not only the scale, but also the growth rate of productivity. These findings together suggest that addressing the shortfall in transport infrastructure can unlock potentially large economic dividends for rapidly emerging economies.

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 and value-added output were converted into real terms by deflating the 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 (see also Topalova and Khandelwal (2011) and Kathuria and Sen (2014) for application of this method). We apply the perpetual inventory method (PIM) and adjust the book value of capital to reflect replacement cost instead of historic cost in which they are measured. 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.

B Tables

[Table 14 about here.]

[Table 15 about here.]

[Table 16 about here.]

[Table 17 about here.]

References

Abadie, A., Athey, S., Imbens, G. W. and Wooldridge, J. (2017), When should you adjust standard errors for clustering?, Technical report, National Bureau of Economic Research.

Ackerberg, D. A., Caves, K. and Frazer, G. (2015), 'Identification properties of recent production function estimators', *Econometrica* **83**(6), 2411–2451.

Allcott, H., Collard-Wexler, A. and O'Connell, S. D. (2016), 'How do electricity shortages affect industry? evidence from india', *American Economic Review* **106**(3), 587–624.

Altonji, J. G., Elder, T. E. and Taber, C. R. (2005), 'An evaluation of instrumental variable strategies for estimating the effects of catholic schooling', *Journal of Human resources* **40**(4), 791–821.

Angrist, J., Lavy, V. and Schlosser, A. (2010), 'Multiple experiments for the causal link between the quantity and quality of children', *Journal of Labor Economics* **28**(4), 773–824.

Arnold, J. M., Javorcik, B., Lipscomb, M. and Mattoo, A. (2015), 'Services reform and manufacturing performance: Evidence from india', *The Economic Journal* **126**(590), 1–39.

Arulampalam, W., Dasgupta, S., Dhillon, A. and Dutta, B. (2009), 'Electoral goals and center-state transfers: A theoretical model and empirical evidence from India', *Journal of Development Economics* **88**(1), 103–119.

Asher, S. and Novosad, P. (2018), *Rural roads and local economic development*, The World Bank.

Asturias, J., García-Santana, M. and Ramos, R. (2018), 'Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India', *Journal of the European Economic Association* .

Balakrishnan, P., Pushpangadan, K. and Babu, M. S. (2000), 'Trade liberalisation and productivity growth in manufacturing: Evidence from firm-level panel data', *Economic and Political weekly* pp. 3679–3682.

Banerjee, A., Duflo, E. and Qian, N. (2012), On the road: Access to transportation infrastructure and economic growth in China, Technical report, National Bureau of Economic Research.

Baskaran, T., Min, B. and Uppal, Y. (2015), 'Election cycles and electricity provision: Evidence from a quasi-experiment with Indian special elections', *Journal of Public Economics* **126**, 64–73.

Bloomberg (2017), 'Modi to spend $59 billion to upgrade india's infrastructure'.
URL: *https://www.bloomberg.com/news/articles/2017-02-01/modi-plans-59-billion-rail-road-push-as-bombardier-ge-invest*

Bohlken, A. T. (2016), Development or rent-seeking: How political influence shapes infrastructure provision in India, Technical report, Mimeo.

Bound, J. and Jaeger, D. A. (2000), Do compulsory school attendance laws alone explain the association between quarter of birth and earnings?, *in* 'Research in labor economics', Emerald Group Publishing Limited, pp. 83–108.

Bound, J., Jaeger, D. A. and Baker, R. M. (1995), 'Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak', *Journal of the American statistical association* **90**(430), 443–450.

Bracco, E., Lockwood, B., Porcelli, F. and Redoano, M. (2015), 'Intergovernmental grants as signals and the alignment effect: Theory and evidence', *Journal of Public Economics* **123**, 78–91.

Chandra, A. and Thompson, E. (2000), 'Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system', *Regional Science and Urban Economics* **30**(4), 457–490.

Chaurey, R. (2016), 'Location-based tax incentives: Evidence from India', *Journal of Public Economics* .

Chetty, R., Looney, A. and Kroft, K. (2009), 'Salience and taxation: Theory and evidence', *American economic review* **99**(4), 1145–77.

Conley, T. G., Hansen, C. B. and Rossi, P. E. (2012), 'Plausibly exogenous', *Review of Economics and Statistics* **94**(1), 260–272.

Datta, S. (2012), 'The impact of improved highways on Indian firms', *Journal of Development Economics* **99**(1), 46–57.

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

Duranton, G., Morrow, P. M. and Turner, M. A. (2014), 'Roads and trade: Evidence from the US', *Review of Economic Studies* **81**(2), 681–724.

Faber, B. (2014), 'Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System', *Review of Economic Studies* **81**(3), 1046–1070.

Fernald, J. G. (1999), 'Roads to prosperity? assessing the link between public capital and productivity', *American economic review* **89**(3), 619–638.

Ghani, E., Goswami, A. G. and Kerr, W. R. (2016*a*), 'Highway to success: The impact of the Golden Quadrilateral project for the location and performance of indian manufacturing', *The Economic Journal* **126**(591), 317–357.

Ghani, E., Goswami, A. G. and Kerr, W. R. (2016*b*), 'Highways and spatial location within cities: Evidence from India', *The World Bank Economic Review* **30**(Supplement_1), S97–S108.

Guasch, L. J. and Joseph, K. (2001), 'Inventories in developing countries', *World Bank Policy Research Woeking Paper 2552* .

Gulyani, S. (2001), 'Effects of poor transportation on lean production and industrial clustering: Evidence from the Indian auto industry', *World development* **29**(7), 1157–1177.

Holl, A. (2016), 'Highways and productivity in manufacturing firms', *Journal of Urban Economics* **93**, 131–151.

Hulten, C. R., Bennathan, E. and Srinivasan, S. (2006), 'Infrastructure, externalities, and economic development: a study of the Indian manufacturing industry', *The World Bank Economic Review* **20**(2), 291–308.

Imbens, G. W. and Rubin, D. B. (2015), *Causal inference in statistics, social, and biomedical sciences*, Cambridge University Press.

Jedwab, R. and Moradi, A. (2016), 'The permanent effects of transportation revolutions in poor countries: Evidence from Africa', *Review of economics and statistics* **98**(2), 268–284.

Johansson, E. (2003), 'Intergovernmental grants as a tactical instrument: empirical evidence from Swedish municipalities', *Journal of Public Economics* **87**(5-6), 883–915.

Kathuria, V. and 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.

Khemani, S. (2007), The political economy of equalization transfers, *in* ‘Fiscal Equalization’, Springer, pp. 463–484.

Lehne, J., Shapiro, J. N. and Eynde, O. V. (2018), ‘Building connections: Political corruption and road construction in India’, *Journal of Development Economics* **131**, 62–78.

Levinsohn, J. and Petrin, A. (2003), ‘Estimating production functions using inputs to control for unobservables’, *The Review of Economic Studies* **70**(2), 317–341.

Li, H. and Li, Z. (2013), ‘Road investments and inventory reduction: Firm level evidence from China’, *Journal of Urban Economics* **76**, 43–52.

Mani, A. and Mukand, S. (2007), ‘Democracy, visibility and public good provision’, *Journal of Development Economics* **83**(2), 506–529.

Martin, L. A., Nataraj, S. and Harrison, A. E. (2017), ‘In with the big, out with the small: Removing small-scale reservations in India’, *American Economic Review* **107**(2), 354–86.

Martin, P., Mayer, T. and Mayneris, F. (2011), ‘Spatial concentration and plant-level productivity in France’, *Journal of Urban Economics* **69**(2), 182–195.

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

Mitra, A., Sharma, C. and Véganzonès-Varoudakis, M.-A. (2012), ‘Estimating impact of infrastructure on productivity and efficiency of Indian manufacturing’, *Applied Economics Letters* **19**(8), 779–783.

Mitra, A., Varoudakis, A. and Veganzones-Varoudakis, M.-A. (2002), ‘Productivity and technical efficiency in Indian states’ manufacturing: the role of infrastructure’, *Economic development and cultural change* **50**(2), 395–426.

Olley, G. S. and Pakes, A. (1996), ‘The dynamics of productivity in the telecommunications equipment industry’, *Econometrica* **64**(6), 1263–1297.

Redding, S. J. and Turner, M. A. (2015), Transportation costs and the spatial organization of economic activity, *in* ‘Handbook of regional and urban economics’, Vol. 5, Elsevier, pp. 1339–1398.

Sengupta, B. (2011), ‘Provision of public goods in a federal economy: The role of party politics’, *European Journal of Political Economy* **27**(1), 104–119.

Shatz, H. J., Kitchens, K. E. and Rosenbloom, S. (2011), *Highway infrastructure and the economy: implications for federal policy*, Rand Corporation.

Shirley, C. and Winston, C. (2004), 'Firm inventory behavior and the returns from highway infrastructure investments', *Journal of Urban Economics* **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. and Sorribas-Navarro, P. (2008), 'The effects of partisan alignment on the allocation of intergovernmental transfers. differences-in-differences estimates for Spain', *Journal of Public Economics* **92**(12), 2302–2319.

Storeygard, A. (2016), 'Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa', *The Review of Economic Studies* **83**(3), 1263–1295.

Syverson, C. (2011), 'What determines productivity?', *Journal of Economic literature* **49**(2), 326–65.

Topalova, P. and Khandelwal, A. (2011), 'Trade liberalization and firm productivity: The case of India', *Review of economics and statistics* **93**(3), 995–1009.

van Kippersluis, H. and Rietveld, C. A. (2018), 'Beyond plausibly exogenous', *The Econometrics Journal* **21**(3), 316–331.

WB (2014), World bank enterprise survey, Technical report, World Bank.

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', *Economics Letters* **104**(3), 112–114.

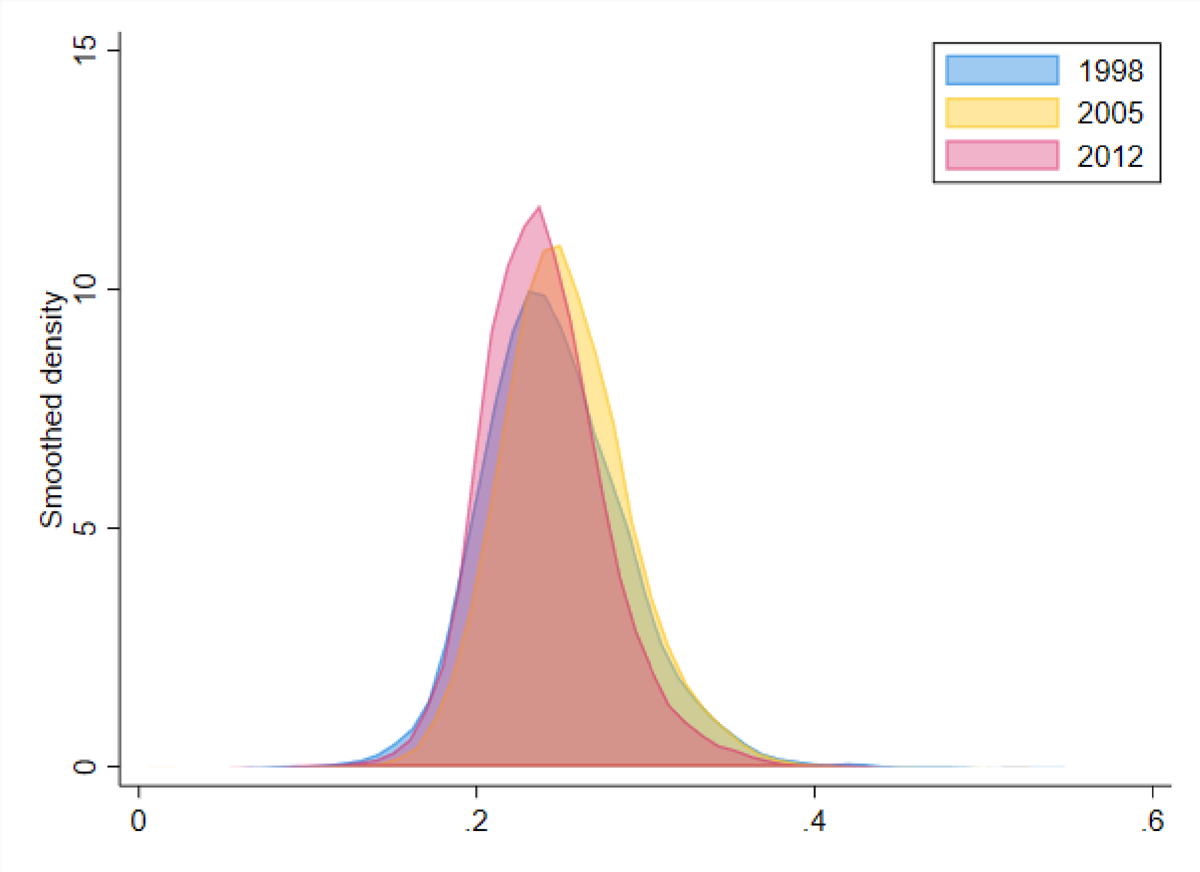

Figure 1: TFP in 1998, 2005 and 2012.
Figure plots the density of log value-added TFP in 1998, 2005 and 2012. TFP is normalized by dividing it by employment-weighted average productivity for an industry-year.

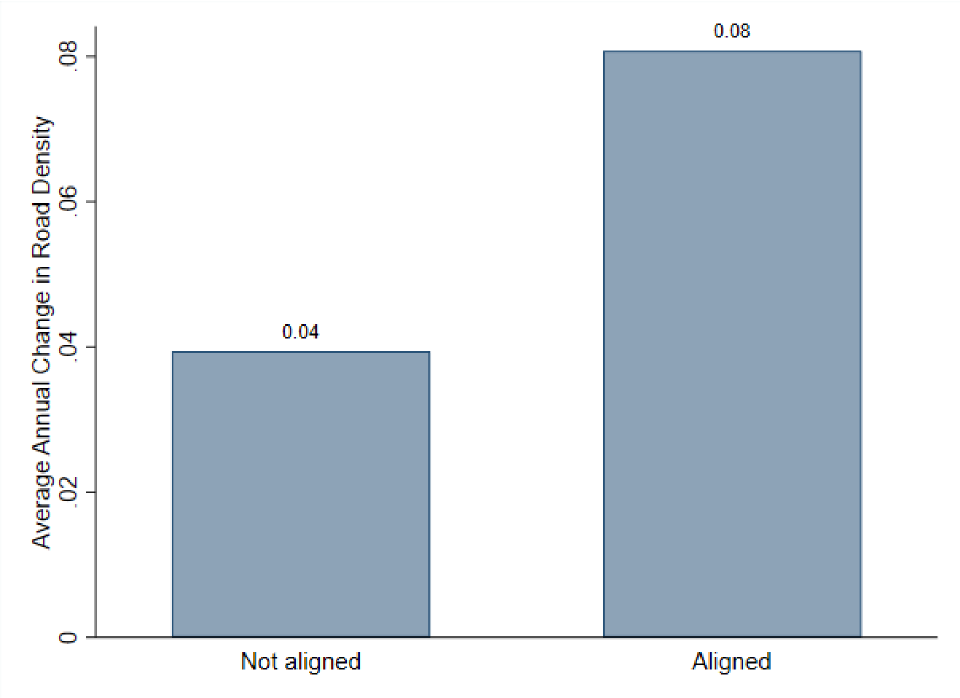

Figure 2: Political Alignment and Change in Road Density.
Figure plots the average year-on-year change in road density across Indian states by alignment status during 1998-2012.

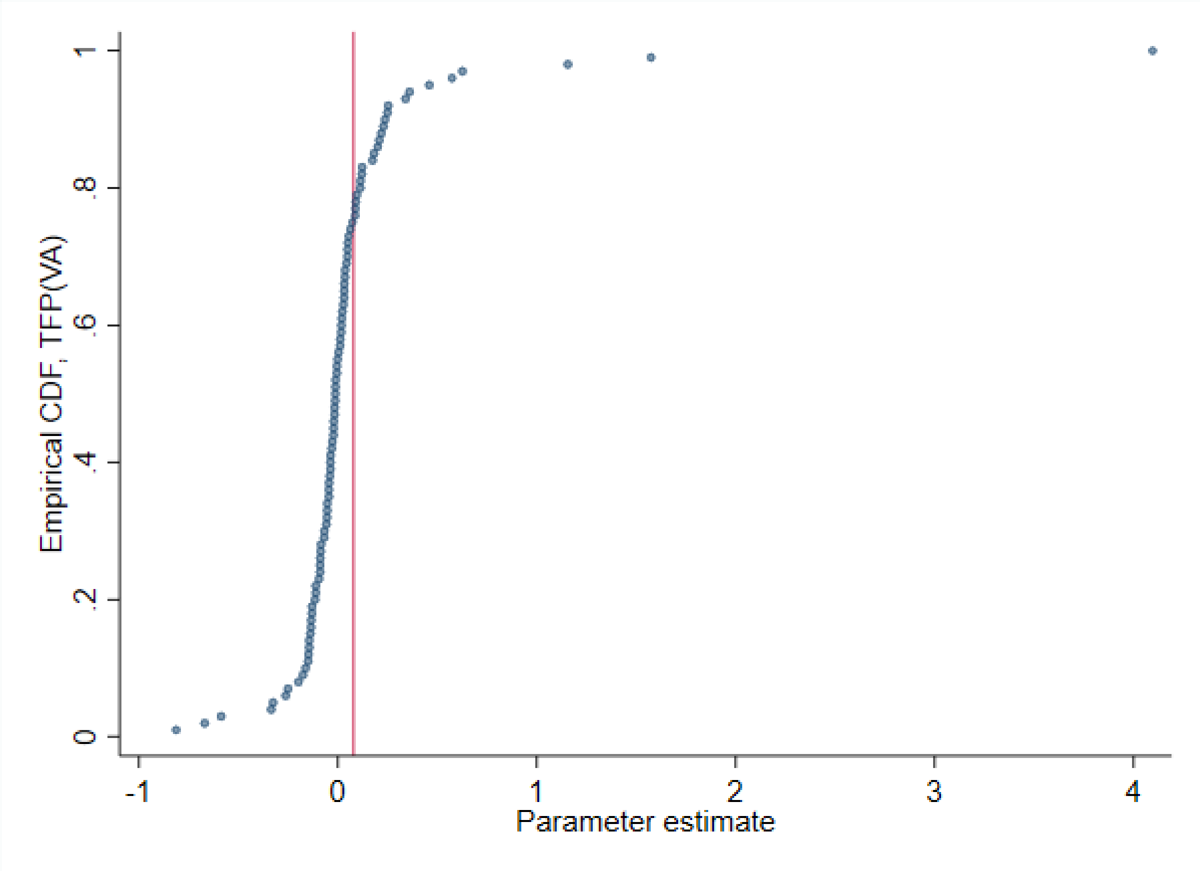

Figure 3: Figure shows empirical CDF distribution of 100 placebo runs and their true coefficients for value-added TFP (in logs). The vertical lines show the coefficients on true alignment status.

Table 1: TFP in Indian Manufacturing (Value Added), 1998-2012

Groups	Obs.	Mean	p10	p25	p50	p75	p90
All	6714	0.162	0.070	0.097	0.153	0.212	0.262
By Age:							
Young	5071	0.169	0.068	0.087	0.147	0.220	0.299
Incumbent	6595	0.163	0.069	0.095	0.155	0.216	0.267
By Size:							
Small	6333	0.202	0.069	0.101	0.192	0.277	0.345
Large	6173	0.128	0.069	0.085	0.116	0.155	0.201
By Location:							
Rural	3203	0.138	0.065	0.080	0.117	0.177	0.239
Urban	3511	0.186	0.086	0.129	0.187	0.230	0.274
TFP_{2012}/ TFP_{1998}	298	1.334	0.759	0.953	1.204	1.589	2.057

Notes: Table shows the dispersion of log value-added Total Factor Productivity (TFP) in Indian manufacturing during 1998-2012 estimated using the ACF method (Ackerberg et al. 2015). The firm-level log TFP values are normalized by dividing it by employment-weighted average productivity (in logs) for an industry-year. Row 'All' corresponds to all observations in our sample. 'Young' includes industry-level estimates of only firms aged < 5 years whereas, 'Incumbents' include those that are aged ≥ 5 years. 'Small' includes firms in the industry with fixed-assets lower than the median fixed assets in the industry, while 'Large' includes those with fixed assets greater than the industry median value. 'Rural' shows TFP for industries located in rural states, where a rural state is one with urbanisation rate lower than the median value across all states in our sample. 'Urban' shows TFP for industries located in states where the urbanisation rate exceeds the median value. TFP_{2012}/ TFP_{1998} is the ratio of log TFP values in 2012 and 1998.

Table 2: First Stage: Effect of Alignment on Road Density

Dependent variable: Road Density	IV-I (1)	IV-II (2)
Aligned	0.170***	0.103***
	(0.03)	(0.02)
Industry FE	Yes	Yes
State FE	Yes	Yes
Year FE	Yes	Yes
Year x State controls	No	Yes
Infrastructure controls	No	Yes
Observations	6667	6535
RMSE	0.346	0.314

Notes: Table shows first stage estimates from regressing road density on *Aligned* conditional on the full set of controls. See Table 3 for the full list of controls. Standard errors clustered at state-industry level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Effect of Road Density on TFP: Main Results

	Valued added			Gross output		
	FE	IV-I	IV-II	FE	IV-I	IV-II
Dependent variable: Log TFP	(1)	(2)	(3)	(4)	(5)	(6)
Road density	0.003	0.071***	0.081***	0.003	0.072***	0.082***
	(0.00)	(0.02)	(0.02)	(0.00)	(0.02)	(0.02)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Year x State controls	No	No	Yes	No	No	Yes
Infrastructure controls	No	No	Yes	No	No	Yes
Observations	6714	6667	6535	6714	6667	6535
RMSE	0.058	0.063	0.061	0.056	0.061	0.059
KP-F (Instrument validity)		40.69	30.84		40.69	30.84

Notes: In cols.(1)-(3) the dependent variable is log value-added TFP, while in cols.(4)-(6) it is log gross-output TFP, both estimated using ACF method. FE shows results from a panel fixed effects model; IV-I from a panel IV2SLS model, where road density is instrumented by *Aligned*, and controls for state, industry and year fixed effects. IV-II shows results from a panel IV2SLS model that additionally controls for time-interacted state characteristics (in logs) – population, literacy, total main and marginal workers, total main workers in agriculture and industry – and infrastructure controls that include railway and electricity density. Road density is the total state-wide road length (in km) divided by its land area (in sq. km). RMSE=Root Mean Squared Error; KP-F stats is the Kleibergen-Paap weak identification test of instrument validity. Standard errors clustered at the state-industry level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Placebo Instrument Test

Dependent variable: Log TFP	TFP(VA) (1)	TFP(GO) (2)	First Stage (3)
Road density	-0.152	-0.162	
	(0.28)	(0.29)	
Aligned (Placebo)			-0.009
			(0.01)
Industry FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Year x State controls	Yes	Yes	Yes
Infrastructure controls	Yes	Yes	Yes
Observations	6582	6582	6582
RMSE	0.074	0.075	0.315
KP-F (Instrument validity)	0.433	0.433	

Notes: Results from a single placebo test where we instrument a state's road density with a randomly assigned alignment status drawn from a uniform distribution that preserves the prior that, at any given time, a state is aligned with the centre. Placebos were assigned 100 times. This table shows one example of a placebo run. We include the full set of controls. See Table 3 for the full list of conditioning variables. RMSE=Root Mean Squared Error; KP-F stats is the Kleibergen-Paap weak identification test of instrument validity. Standard errors clustered at the state-industry level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Summary Statistics of Placebo Tests

Variable	Above 0 (1)	Below 0 (2)	Insignificant (3)
TFP(VA)	4	4	92
TFP(GO)	5	4	91

Notes: In cols.(1) and (2), we show the number of runs when each outcome of interest was above or below 0 and significant at the 5% level, respectively, whereas in col.(3), we show the number of runs when the result did not turn out to be significant at the 5% level. Placebos were assigned 100 times.

Table 6: Plausibly exogenous tests

Dependent variable:	TFP(VA)	TFP(GO)
	Panel A: Effect of Road Density on TFP	
Plausibly exogenous	0.113***	0.128***
	(0.023)	(0.023)
Plausibly exogenous	0.113***	0.128***
(with uncertainty)	(0.027)	(0.026)
	Panel B: Effect of Alignment on TFP	
Reduced form	0.008***	0.008***
(full sample)	(0.002)	(0.002)
Direct Effect	-0.002	-0.003
(zero first-stage group)	(0.010)	(0.009)
	N = 300	
Direct Effect	0.008***	0.009***
(remaining sample)	(0.008)	(0.002)
	N = 6233	
	Panel C: Effect of Alignment on Road Density	
First stage	0.103***	0.103***
(full sample)	(0.019)	(0.019)
First stage	0.052	0.052
(zero first-stage group)	(0.049)	(0.049)
First stage	0.103***	0.103***
(remaining sample)	(0.019)	(0.019)

Notes: Standard errors in parenthesis clustered at the state-industry level. The row plausibly exogenous is estimated by plugging-in the mean, μ_λ from the zero-first-stage sub-group and assumes $\Omega_\lambda = 0$. The row 'with uncertainty' 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Effect of Roads on TFP: Young vs. Incumbent

	Valued added			Gross output		
	FE	IV-I	IV-II	FE	IV-I	IV-II
Dependent variable: Log TFP	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Young Firms Aged < 5 years					
Road density	0.005	0.050*	0.081*	0.006	0.050*	0.080*
	(0.01)	(0.03)	(0.04)	(0.01)	(0.03)	(0.04)
Observations	5071	5046	4982	5071	5046	4982
RMSE	0.088	0.089	0.089	0.087	0.088	0.088
KP-F (Instrument validity)		33.66	25.90		33.66	25.90
	Panel B: Incumbent Firms Aged ≥ 5 years					
Road density	0.003	0.070***	0.082***	0.003	0.071***	0.083***
	(0.00)	(0.02)	(0.03)	(0.00)	(0.02)	(0.03)
Observations	6595	6548	6417	6595	6548	6417
RMSE	0.061	0.065	0.063	0.059	0.063	0.061
KP-F (Instrument validity)		38.99	28.75		38.99	28.75
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Year x State controls	No	No	Yes	No	No	Yes
Infrastructure controls	No	No	Yes	No	No	Yes

Notes: In cols.(1)-(3) the dependent variable is log value-added TFP while in cols.(4)-(6) it is log gross-output TFP both estimated using ACF method. FE shows results from a panel fixed effects model; IV-I from a panel IV2SLS model, where road density is instrumented by alignment, and controls for state, industry and year fixed effects. IV-II shows results from a panel IV2SLS model that additionally controls for time-interacted state characteristics (in logs) – population, literacy, total main and marginal workers, total main workers in agriculture and industry – and infrastructure controls that include railway and electricity density. Panel A shows results that correspond to young firms aged < 5 years in the industry whereas, Panel B shows results for incumbent firms in the industry aged ≥ 5 years. Road density is the total state-wide road length (in km) divided by its land area (in sq. km). RMSE=Root Mean Squared Error; KP-F stats is the Kleibergen-Paap weak identification test of instrument validity. Standard errors clustered at the state-industry level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Effect of Roads on TFP: Small vs. Large

	Valued added			Gross output		
	FE	IV-I	IV-II	FE	IV-I	IV-II
Dependent variable: Log TFP	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Small Firms with Fixed Assets $<$ Industry Median					
Road density	0.006	0.082***	0.095***	0.006*	0.084***	0.096***
	(0.00)	(0.02)	(0.03)	(0.00)	(0.02)	(0.03)
Observations	6333	6288	6176	6333	6288	6176
RMSE	0.080	0.085	0.083	0.078	0.083	0.080
KP-F (Instrument validity)		40.86	30.29		40.86	30.29
	Panel B: Large Firms with Fixed Assets $\geq$ Industry Median					
Road density	0.000	0.051***	0.065***	0.000	0.050***	0.068***
	(0.00)	(0.02)	(0.02)	(0.00)	(0.02)	(0.03)
Observations	6173	6129	6020	6173	6129	6020
RMSE	0.050	0.053	0.053	0.049	0.052	0.052
KP-F (Instrument validity)		36.68	27.32		36.68	27.32
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Year x State controls	No	No	Yes	No	No	Yes
Infrastructure controls	No	No	Yes	No	No	Yes

Notes: In cols.(1)-(3) the dependent variable is log value-added TFP while in cols.(4)-(6) it is log gross-output TFP both estimated using ACF method. FE shows results from a panel fixed effects model; IIV-I from a panel IV2SLS model, where road density is instrumented by alignment, and controls for state, industry and year fixed effects. IV-II shows results from a panel IV2SLS model that additionally controls for time-interacted state characteristics (in logs) – population, literacy, total main and marginal workers, total main workers in agriculture and industry – and infrastructure controls that include railway and electricity density. Panel A shows results that correspond to small firms in the industry with fixed assets lower than the industry-year median value of fixed assets whereas, Panel B shows results for large firms in the industry with fixed assets that exceed the industry-year median value of fixed assets. Road density is the total state-wide road length (in km) divided by its land area (in sq. km). RMSE=Root Mean Squared Error; KP-F stats is the Kleibergen-Paap weak identification test of instrument validity. Standard errors clustered at the state-industry level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Effect of Roads on TFP (Value added): Rural vs. Urban

Dependent variable: Log TFP	All (1)	Small (2)	Large (3)	Young (4)	Incumbent (5)
	Panel A: Rural sample				
Road density	0.085**	0.088**	0.081**	0.046	0.090**
	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)
Observations	3024	2790	2928	2639	2931
RMSE	0.056	0.074	0.049	0.080	0.060
KP-F (Instrument validity)	17.45	17.16	14.29	17.60	15.47
	Panel B: Urban sample				
Road density	0.206***	0.298***	0.119	0.416**	0.180***
	(0.08)	(0.11)	(0.07)	(0.20)	(0.07)
Observations	3511	3386	3092	2343	3486
RMSE	0.079	0.114	0.060	0.146	0.075
KP-F (Instrument validity)	15.91	16.04	13.07	7.91	15.56
Industry FE	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Year x State controls	Yes	Yes	Yes	Yes	Yes
Infrastructure controls	Yes	Yes	Yes	Yes	Yes

Notes: Columns indicate log value-added TFP for: All firms in the industry (col.(1)), small firms (col.(2)), large firms (col.(3)), young firms (col.(4)) and incumbent firms (col.(5)). Panel A shows results for industries located in rural states, where a rural state is one where the urban population is less than the median urban population of all states in our sample. Panel B shows results for urban-based firms, that is, firms in states exceeding the median urban population. Road density is the total state-wide road length (in km) divided by its land area (in sq. km). The coefficients are from separate regressions of value-added TFP on road density for the respective sub-samples as in column headings. All regressions instrument road density with alignment and control for year-interacted state characteristics (in logs) – population, literacy, total main and marginal workers, total main workers in agriculture and industry – a state's railway and electricity density along with state, industry and year fixed effects. RMSE=Root Mean Squared Error; KP-F stats is the Kleibergen-Paap weak identification test of instrument validity. Standard errors clustered at the state-industry level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Effect of Roads on TFP: Long-Differenced Results

	ACF Method		LP Method		
Dependent variable:	ΔTFP (VA)	ΔTFP (GO)	ΔTFP (VA)	ΔTFP (GO)	First Stage
Long-differenced Log TFP	(1)	(2)	(3)	(4)	(5)
Δ Road density	0.100**	0.099**	0.101**	0.098**	
	(0.05)	(0.04)	(0.05)	(0.04)	
$CAY_{98\text{-}12}$					0.087***
					(0.01)
Industry FE	Yes	Yes	Yes	Yes	Yes
State controls	Yes	Yes	Yes	Yes	Yes
Infrastructure controls	Yes	Yes	Yes	Yes	Yes
Observations	298	298	298	298	298
RMSE	0.091	0.090	0.091	0.090	0.470
KP-F (Instrument validity)		10.03			

Notes: Cols.(1) and (2) relate to long-differenced log value-added and gross-output TFP between 1998 and 2012 respectively, estimated using Ackerberg, Caves and Frazer (2015) method. Cols.(3) and (4) relate to long-differenced log value-added and gross-output TFP respectively, estimated using the Levinsohn and Petrin (2003) method for the same time period. Col.(5) shows first stage results. ΔRoad density is the change in road density between 1998 and 2012 where road density is defined as total state-wide road length (in km) divided by its land area (in sq. km). We instrument road density with the maximum of the consecutive number of aligned year between 1998-2018 ($CAY_{98\text{-}12}$). All regressions control for industry fixed effects and state characteristics – population, literacy, total main and marginal workers and total main workers in agriculture and industry – and difference in rail and electricity density between 1998 and 2012. RMSE=Root Mean Squared Error; KP-F stats is the Kleibergen-Paap weak identification test of instrument validity. Standard errors clustered at the state-industry level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Elasticity Estimates: Road Density and TFP (Value added)

Dependent variable: Log TFP	All (1)	Rural (2)	Urban (3)	Young (4)	Incumbent (5)	Small (6)	Large (7)
log(Road density)	0.108***	0.116**	0.322***	0.105*	0.107***	0.124***	0.088***
	(0.03)	(0.05)	(0.12)	(0.06)	(0.03)	(0.04)	(0.03)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year x State controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Infrastructure controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6535	3024	3511	4982	6417	6176	6020
RMSE	0.058	0.055	0.071	0.088	0.060	0.080	0.050
KP-F (Instrument validity)	66.46	36.79	23.69	51.99	63.54	65.21	58.69

Notes: Dependent variable is log TFP(VA) for the sub-sample indicated in columns (1-7). The results are from a panel IV2SLS model that regresses TFP(VA) on log(Road density), where the latter is instrumented by *Aligned*. All regressions control for time-interacted state characteristics (in logs) – population, literacy, total main and marginal workers, total main workers in agriculture and industry – a state's railway and electricity density and include state, year and industry fixed effects. RMSE=Root Mean Squared Error; KP-F stats is the Kleibergen-Paap weak identification test of instrument validity. Standard errors clustered at the state-industry level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Effect of High Road Density on TFP (Value added and Gross output)

	Valued added			Gross output		
	FE	IV-I	IV-II	FE	IV-I	IV-II
Dependent variable: Log TFP	(1)	(2)	(3)	(4)	(5)	(6)
High Road density	0.002	0.481*	0.250**	0.002	0.490*	0.255**
	(0.00)	(0.26)	(0.11)	(0.00)	(0.26)	(0.11)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Year x State controls	No	No	Yes	No	No	Yes
Infrastructure controls	No	No	Yes	No	No	Yes
Observations	6714	6667	6535	6714	6667	6535
RMSE	0.058	0.132	0.083	0.056	0.133	0.082
KP-F (Instrument validity)		5.40	9.13		5.40	9.13

Notes: In cols.(1)-(3) the dependent variable is log value-added TFP, while in cols.(4)-(6) it is log gross-output TFP both estimated using ACF method. FE shows results from a panel fixed effects model; IV-I from a panel IV2SLS model, where road density is instrumented by alignment, and controls for state-industry and year fixed effects. IV-II shows results from a panel IV2SLS model that additionally controls for time-interacted state characteristics (in logs) – population, literacy, total main and marginal workers, total main workers in agriculture and industry – and infrastructure controls that include railway and electricity density. *High Road density* is a dummy variable that takes a value of 1 if a state's road density exceeds two-third of the road density distribution in a state-year across all states in the sample, and is zero otherwise. KP-F stats is the Kleibergen-Paap weak identification test of instrument validity. Standard errors clustered at the state-industry level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 13: Robustness: Effect of Roads on TFP

Dependent variable: Log TFP	LP(VA) (1)	Winsorized sample (2)	No split states (3)	Major states (4)	IV-A (5)	IV-B (6)
	Panel A: TFP (Value added)					
Road density	0.079***	0.080***	0.077***	0.193**	0.055***	0.097***
	(0.02)	(0.02)	(0.02)	(0.09)	(0.01)	(0.03)
RMSE	0.059	0.059	0.060	0.084	0.058	0.063
	Panel B: TFP (Gross output)					
Road density	0.083***	0.079***	0.079***	0.200**	0.056***	0.100***
	(0.02)	(0.02)	(0.02)	(0.09)	(0.01)	(0.03)
RMSE	0.059	0.058	0.059	0.084	0.056	0.062
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Year x State controls	Yes	Yes	Yes	Yes	Yes	Yes
Infrastructure controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6535	6538	6428	4965	6535	6535
KP-F (Instrument validity)	30.84	30.84	31.17	9.86	63.94	38.68

Notes: The dependent variable in col.(1) is log value-added TFP estimated using the LP method (Levinsohn and Petrin 2003). In col.(2), we winsorize the top 1% and bottom 1% of log value-added TFP estimated using the ACF method. In col.(3), we exclude the states of Bihar, Jharkhand, Madhya Pradesh, Chhattisgarh, Uttar Pradesh and Uttarakhand that were re-organized in 2000 from our analysis; col.(4) concentrates on major Indian states, while cols.(5) and (6), present results from regressions where we instrument road density with 'Aligned x Seat share' and 'Aligned x Election year dummy', respectively. Panel A relates to value-added TFP whereas, panel B relates to gross-output TFP. Road density is the total state-wide road length (in km) divided by its land area (in sq. km). All regressions include the full set of controls – time-interacted state-specific variables, a state's railway and electricity density along with state, year and industry fixed effects. KP-F stats is the Kleibergen-Paap weak identification test of instrument validity. Standard errors clustered at the state-industry level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 14: TFP in Indian Manufacturing (Gross Output), 1998-2012

Groups	Obs.	Mean	p10	p25	p50	p75	p90
All	6714	0.161	0.070	0.096	0.153	0.211	0.260
By Age:							
Young	5071	0.168	0.069	0.087	0.147	0.219	0.298
Incumbent	6595	0.163	0.069	0.096	0.156	0.214	0.266
By Size:							
Small	6333	0.198	0.067	0.100	0.190	0.273	0.337
Large	6173	0.130	0.071	0.087	0.120	0.158	0.203
By Location:							
Rural	3203	0.137	0.065	0.080	0.118	0.176	0.239
Urban	3511	0.185	0.086	0.132	0.187	0.228	0.274
TFP_{2012}/ TFP_{1998}	298	1.336	0.776	0.966	1.233	1.602	1.977

Notes: This table shows the dispersion of log gross-output Total Factor Productivity (TFP) in Indian manufacturing during 1998-2012 estimated using the Ackerberg, Caves and Frazer (2015) method. The firm-level log TFP values are normalized by dividing it by employment-weighted average productivity (in logs) for an industry-year. Row 'All' corresponds to all observations in our sample. 'Young' includes industry-level estimates of only firms aged < 5 years, whereas 'Incumbents' include those that are aged ≥ 5 years. 'Small' includes firms in the industry with fixed-assets lower than the median fixed assets in the industry, while 'Large' includes those with fixed assets greater than the industry median value. 'Rural' shows TFP for industries located in rural states where a rural state is defined to have urbanisation rate lower than the median value across all states in our sample. 'Urban' shows TFP for industries located in states where the urbanisation rate exceeds the median value. TFP_{2012}/ TFP_{1998} is the ratio of log TFP values in 2012 and 1998.

Table 15: Effect of Roads on TFP (Gross output): Rural vs. Urban

Dependent variable: Log TFP	All (1)	Small (2)	Large (3)	Young (4)	Incumbent (5)
			Panel A: Rural sample		
Road density	0.081**	0.079**	0.082**	0.042	0.086**
	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)
Observations	3024	2790	2928	2639	2931
RMSE	0.053	0.071	0.049	0.076	0.057
KP-F (Instrument validity)	17.45	17.16	14.29	17.60	15.47
			Panel B: Urban sample		
Road density	0.212***	0.300***	0.129*	0.421**	0.186***
	(0.08)	(0.10)	(0.08)	(0.20)	(0.07)
Observations	3511	3386	3092	2343	3486
RMSE	0.080	0.113	0.062	0.148	0.076
KP-F (Instrument validity)	15.91	16.04	13.07	7.91	15.56
Industry FE	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Year x State controls	Yes	Yes	Yes	Yes	Yes
Infrastructure controls	Yes	Yes	Yes	Yes	Yes

Notes: Columns indicate log gross-output TFP for: All firms in the industry (col.(1)), small firms (col.(2)), large firms (col.(3)), young firms (col.(4)) and incumbent firms (col.(5)). Panel A shows results for industries located in rural states, where a rural state is one where the urban population is less than the median urban population of all states in our sample. Panel B shows results for urban-based firms, that is, firms in states exceeding the median urban population. Road density is the total state-wide road length (in km) divided by its land area (in sq. km). The coefficients are from separate regressions of gross-output TFP on road density for the respective sub-samples as in column headings. All regressions instrument road density with alignment and control for year-interacted state characteristics (in logs) – population, literacy, total main and marginal workers, total main workers in agriculture and industry – a state's railway and electricity density along with state, industry and year fixed effects. RMSE=Root Mean Squared Error; KP-F stats is the Kleibergen-Paap weak identification test of instrument validity. Standard errors clustered at the state-industry level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 16: First Stage: Effect of Alignment on High Road Density Dummy

Dependent variable: High Road Density Dummy	IV-I (1)	IV-II (2)
Aligned	0.025**	0.033***
	(0.01)	(0.01)
Industry FE	Yes	Yes
State FE	Yes	Yes
Year FE	Yes	Yes
Year x State controls	No	Yes
Infrastructure controls	No	Yes
Observations	6667	6535
RMSE	0.246	0.248

Notes: First stage estimates from regressing High Road density on *Aligned* conditional on the full set of controls. High Road density is a dummy variable that takes a value of 1 if road density in a state exceeds the 60^{th} percentile of the overall road density, and zero otherwise. Standard errors clustered at the state-industry level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 17: First Stage (Robustness): Effect of Alignment on Road Density

Dependent variable: Road Density	LP(VA) (1)	Winsorized sample (2)	No split states (3)	Major states (4)	IV-A (5)	IV-B (6)
Aligned	0.094***	0.094***	0.094***	0.041***		
	(0.02)	(0.02)	(0.02)	(0.01)		
Aligned x Seat Share					0.003***	
					(0.00)	
Aligned x Election Year						0.102***
						(0.02)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Year x State controls	Yes	Yes	Yes	Yes	Yes	Yes
Infrastructure controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6020	6020	5919	4706	6020	6020
RMSE	0.306	0.306	0.307	0.295	0.302	0.307

Notes: Results from first stage estimates that regress road density on alignment. Standard errors clustered at the state-industry level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.