Roads and Capital Misallocation: Evidence from India's Infrastructure Boom*

Anastasia Burya† Martino Pelli‡ Avralt-Od Purevjav§ Jeanne Tschopp¶

February 2026

Abstract

We study the effects of India's road network expansion on capital allocation across firms and aggregate productivity. Focusing on road improvements between 2011 and 2019—largely driven by the National Highways Development Project—we construct a market access measure using historical OpenStreetMap data that captures changes in travel times between postal codes and major cities. Using a staggered difference-in-differences design, we find that improved market access is associated with reductions in capital misallocation and increases in aggregate productivity in treated postal codes, primarily through better capital allocation rather than within-firm efficiency gains. Firms experience a 25% increase in capital, with ex ante high marginal revenue product of capital (MRPK) firms seeing an additional 25% growth and a 45% decline in MRPK, consistent with reduced capital misallocation. We estimate that the Solow residual rises by 2.7-5%, a gain comparable to the lower-bound effects of India's earlier foreign capital liberalization. We show that these allocative gains are driven primarily by reductions in input wedges rather than by changes in markups, indicating that road infrastructure improves efficiency mainly by alleviating physical constraints on input access.

Keywords: Market access, road network, misallocation of capital, aggregate productivity, India.
JEL Codes: R4, O1, L2, D2

---

*We are grateful to FNS [grant 100018_192553/1] for its financial support. We thank, without implicating them, seminar participants at Economics Lunch at University of Basel, and at the 5th Swiss Workshop on Local Public Finance and Regional Economics for helpful comments and suggestions. Sarah Volken and Lea Cristina Schmid provided outstanding research assistance. All remaining errors are ours. The views expressed are those of the authors and do not necessarily reflect the views and policies of the Asian Development Bank (ADB) or the World Bank, their Boards of Governors, or the governments they represent.

†Department of Economics, University of Bern, Schanzeneckstrasse 1, 3001 Bern, Switzerland; anastasia.burya@unibe.ch.

‡Economic Analysis and Operational Support Division (EREA), Economic Research and Development Impact Department (ERDI), Asian Development Bank (ADB); 6 ADB Ave. Mandaluyong City, Metro Manila, 1550 Philippines; mpelli@adb.org.

§The World Bank, Washington DC, United States; apurevjav@worldbank.org.

¶Department of Economics, University of Bern, Schanzeneckstrasse 1, 3001 Bern, Switzerland; jeanne.tschopp@unibe.ch.

1 Introduction

Frictions in input and product markets distort the allocation of productive inputs across firms and reduce aggregate productivity. In the absence of such frictions, marginal revenue products of capital (MRPK) are equalized across firms within an industry. A large literature, however, documents substantial dispersion in MRPK across firms in developing economies, pointing to capital misallocation. In a seminal contribution, Hsieh and Klenow (2009b) show that bringing India's manufacturing sector to the allocative efficiency of the United States would raise total factor productivity (TFP) by 40–60%, highlighting the potentially large aggregate gains from improved allocative efficiency.

Much of the subsequent literature has emphasized financial frictions as a key source of capital misallocation and has studied policies that improve firms' access to capital markets (e.g., Bau and Matray, 2023). However, frictions outside the financial sector may also play an important role.

In this paper we focus on frictions related to transportation infrastructure and study how major highway expansions in India shape the allocation of resources across firms. Poor connectivity raises transport costs, disrupts supply chains, and limits access to input and output markets, increasing investment risk even for firms with access to finance. By improving connectivity, transportation infrastructure can relax effective capital constraints, allowing firms with high returns to capital to expand investment.

Focusing on the capital-deepening margin, we examine whether improved road infrastructure reallocates capital toward ex ante high-MRPK firms and whether these changes are associated with a reduction in capital misallocation and gains in aggregate productivity.

We study this mechanism in the context of India's National Highways Development Project (NHDP), one of the largest road infrastructure programs worldwide. By the early 2000s, poor connectivity was widely recognized as a major constraint on economic growth, even after India's trade liberalization of the 1990s. In response, the Indian government launched the NHDP, leading to a large expansion in highway density between 2011 and 2019. This expansion occurred in two main phases: an initial phase between 2011 and 2013 associated with the completion and opening of the Golden Quadrilateral Highway (GQH), followed by subsequent NHDP phases between 2013 and 2019 that further improved road connectivity across the country.

To quantify changes in connectivity, we construct a time-varying market access index using historical road network data from OpenStreetMap for the period 2011–2019, following Donaldson and Hornbeck (2016) and Baum-Snow et al. (2020). The index measures shortest travel times between postal codes and major Indian cities, accounting for the full road network, including national highways, urban and rural roads, and lower-quality routes. Unlike binary indicators or distance-to-highway measures, this approach provides a continuous, location-specific measure of market access that varies across space and time and captures changes in effective connectivity more precisely.

We estimate the firm-level effects of road network expansion using a rich panel dataset on the financial performance of Indian firms and two complementary difference-in-differences designs. The first exploits changes in market access associated with the completion of the GQH between 2011 and 2013. The second is a staggered design that exploits heterogeneous growth in market access over the full 2011–2019 period, capturing both the initial highway completion and the gradual, uneven improvements of market access under subsequent NHDP investments across space and time.

To address potential identification concerns, we restrict the sample to firms incorporated prior to the announcement of the GQH, mitigating bias from endogenous entry decisions. This restriction focuses the analysis on incumbent firms and likely yields conservative estimates of aggregate gains from improved market access. Firm fixed effects (FE) control for time-invariant unobserved heterogeneity, while industry-year FE absorb common sectoral shocks. We further account for pre-existing regional differences and firm heterogeneity by interacting initial firm characteristics and initial local economic conditions with year FE and by including subdistrict-year FE. We additionally verify balance in pre-treatment firm outcomes across treated and control locations. Finally, a series of diagnostic analyses—including event study plots, a Bacon decomposition, an analysis of the weighting structure, and a stacked regression approach—indicate that biases arising from treatment effect heterogeneity, negative weighting, or differential trends are negligible, supporting the robustness of our empirical strategy.

Our results show substantial capital reallocation following improvements in connectivity. Relative to firms in untreated locations, treated firms increase capital investment, with the largest responses among firms that are ex ante high-MRPK. In the staggered design over the 2011–2019 period, firms located in treated postal codes increase capital by approximately 25%, while ex ante high-MRPK firms experience an additional 25% increase. As capital shifts disproportionately toward these firms, their MRPK declines by roughly 45%, a pattern consistent with a reduction in capital misallocation.

We then quantify the aggregate productivity implications of reduced input misallocation. We measure aggregate productivity using changes in the Solow residual, which reflect both firm-level efficiency (TFPQ) and the allocation of inputs across firms. We find little evidence of improvements in TFPQ, indicating that aggregate gains primarily arise from improved allocation rather than within-firm efficiency. Using the estimated responses of capital and labor, we obtain aggregate productivity gains ranging from 2.7% to 5% among treated firms. These effects are comparable in magnitude to the lower-bound estimates of aggregate productivity gains from foreign capital liberalization in India reported by Bau and Matray (2023), despite important differences in mechanisms, sources of variation—industry-level exposure in their setting versus spatial variation in ours—and timing—the 2001 and 2006 liberalization episodes in their case versus the post-2011 period in ours. This comparison highlights transportation infrastructure as a complementary channel for improving aggregate efficiency by alleviating physical, rather than financial, constraints on production.

Finally, we examine whether the observed allocative gains arise primarily from changes in markups or from changes in pure input wedges, which is important for identifying the channel through which improvements in market access operate. Market access can affect firms either by intensifying competition in final-goods markets, thereby influencing markups, or by improving access to intermediate inputs, lowering effective input costs. We find that the reduction in misallocation is driven mainly by changes in input wedges rather than by changes in markups. In particular, the implied change in markups is modest—about -2.7% on average—whereas pure input wedges adjust substantially: the pure capital wedge falls by about 32%, while the pure labor wedge rises by about 29%, consistent with capital deepening as access to capital and intermediates improves. Taken together, these patterns indicate that improved road connectivity raises allocative efficiency mainly by alleviating physical constraints on input access, rather than through changes in product-market power alone.

A large literature on misallocation has emphasized financial frictions and trade barriers as primary sources of inefficiencies (Banerjee and Moll, 2010; Midrigan and Xu, 2014; Hsieh and Klenow, 2009b; Restuccia and Rogerson, 2008; Epifani and Gancia, 2011; Melitz, 2003; Alcalá and Ciccone, 2004).[1] More recently, Bau and Matray (2023) show that greater access to foreign capital improves capital allocation in Indian firms, suggesting that reducing financial constraints can alleviate misallocation. We contribute to this literature by showing that infrastructure investments can also facilitate capital reallocation, through a different mechanism that operates by relaxing physical constraints on input access.

We also provide new firm-level evidence on the effects of transportation infrastructure. A large literature documents the macroeconomic and spatial effects of transportation infrastructure, emphasizing its role in economic growth, trade, employment, urbanization, and regional development.[2] Firm-level studies show that infrastructure enhances firm productivity (Datta, 2012; Ghani

[1] These studies emphasize that credit constraints significantly distort capital allocation and that firms with better access to financing can accumulate capital more efficiently (Bai et al., 2018). While trade liberalization has been shown to improve resource reallocation and aggregate productivity (Melitz, 2003; Alcalá and Ciccone, 2004), its effects on misallocation are more ambiguous, with some studies finding that market power and markups may counteract efficiency gains (Epifani and Gancia, 2011).

[2] Studies show that improved transportation networks enhance business performance, labor market integration, and trade (Michaels, 2008; Duranton et al., 2014; Datta, 2012). Infrastructure investments shape spatial economic dynamics, affecting urban expansion, decentralization, and land values (Baum-Snow, 2007; Baum-Snow et al., 2016; Atack et al., 2010; Haines and Margo, 2006; Atack and Margo, 2011; Atack et al., 2008). In developing economies, infrastructure fosters regional economic convergence, as seen in the case of railroads reducing inter-regional disparities

et al., 2016; Asturias et al., 2019), facilitates firm entry and competition (Duranton and Turner, 2012; Asturias et al., 2019), and supports firm expansion and industrial growth (Lu, 2018). To our knowledge, however, no study has examined how transportation infrastructure affects the allocation of capital across firms.

We also contribute to the literature on India's road infrastructure by documenting how the effects of different phases of the NHDP unfolded over time. Existing work has focused primarily on the GQH, showing that it improved inventory management and reduced transportation frictions for manufacturing firms (Datta, 2012), increased manufacturing activity and firm productivity (Ghani et al., 2016), and intensified product-market competition (Asturias et al., 2019). In the labor market, road expansions have been shown to contribute to occupational shifts from agriculture to wage employment (Asher and Novosad, 2020).

Finally, we advance empirical measurement by moving beyond the binary indicators (Chandra and Thompson, 2000; Michaels, 2008; Atack et al., 2010; Faber, 2014; Datta, 2012; Ghani et al., 2016) or proximity-based measures (Datta, 2012; Banerjee et al., 2020; Brooks et al., 2021) commonly used in the literature, and instead propose a time-varying, location-specific measure of market access based on travel times.

The remainder of this paper is structured as follows. Section 2 describes the data and key variable measurements. Section 3 relates input price wedges and misallocation. Section 4 presents our empirical approach and Section 5 discusses results. Section 6 quantifies the aggregate productivity effects of reduced misallocation within treated firms and Section 7 assesses the role of shifts in markups and pure input wedges. The last section concludes.

2 Measurements and Data

2.1 Road Network Expansion

By the early 2000s, inadequate road infrastructure was widely recognized as a constraint on India's economic development. National highways represent only a small share of the road network but carry a disproportionate share of freight and passenger traffic; limited lane capacity and poor road quality contributed to congestion, high accident rates, and long travel times. In response, the Gov-

(Donaldson, 2018). However, some studies suggest that infrastructure does not always create new economic activity but may instead reallocate existing activity, particularly when investments favor already developed regions (Chandra and Thompson, 2000; Faber, 2014; Baum-Snow et al., 2020). Highways and intercity networks have been linked to employment growth, labor mobility, and regional expansion, though they may also contribute to economic concentration in developed areas, reinforcing regional inequalities (Duranton and Turner, 2012; Fretz et al., 2022; Faber, 2014). Infrastructure investments further influence labor market transitions, with improved road networks reducing monopsony power and increasing worker mobility (Brooks et al., 2021), while rural road expansions in India have been shown to shift workers from agricultural employment to wage-based jobs (Asher and Novosad, 2020).

ernment of India launched a set of large-scale road programs, most notably the National Highways Development Project (NHDP), alongside complementary initiatives targeting freight corridors and rural connectivity.

The NHDP, launched in 2001, aimed to modernize and expand the national highway system. Its flagship component was the Golden Quadrilateral Highway (GQH), a 5,846-kilometer network connecting Delhi, Kolkata, Mumbai, and Chennai, as well as major secondary hubs such as Ahmedabad, Bengaluru, Jaipur, and Surat. Additional NHDP components included the North-South corridor (Srinagar-Kanyakumari) and the East-West corridor (Silchar-Porbandar). The GQH was implemented under the National Highways Authority of India (NHAI) and represented one of the largest highway projects worldwide.

The timing of highway upgrades under the NHDP is central to our empirical design. National highway expansion accelerated sharply around the completion and opening of the GQH: annual growth in highway density increased from about 0.5% between 2009 and 2011 to about 12% between 2011 and 2013, followed by continued expansion during subsequent NHDP phases. Overall, highway density increased by about 68% between 2011 and 2019.[3]

During the same period, India also increased connectivity through other road programs, including rural roads and economic corridors, contributing to a broad-based expansion of the national road network.[4]

2.2 Market Access Measure

We measure changes in connectivity induced by India's road network expansion using a market access index, following Donaldson and Hornbeck (2016) and Baum-Snow et al. (2020):

$$MA_{pt} = \sum_{c} \frac{N_{c,2011} \times \mathbf{I}(N_{c,2011} \geq 100k)}{\tau_{pct}^{\theta}}, \tag{1}$$

where MA_{pt} denotes market access for postal code p (*pincode* in India) in year t, $N_{c,2011}$ is the population of city c in 2011, and $\mathbf{I}(\cdot)$ is an indicator equal to one for cities with populations above 100,000 (Tier-1 cities under the Reserve Bank of India classification) in 2011. Bilateral trading costs between p and c at time t are denoted by τ_{pct}, and $\theta > 0$ is the distance elasticity governing how rapidly market access declines with trading costs. The index captures access to major economic centers by weighting city populations by bilateral trading costs, so that closer and more

[3]Appendix Figure A.1 shows the evolution of the road network (Panel (a)) and highway density (Panel (b)) over 1951–2021. Panel (a) indicates that between 2001 and 2011 expansions were concentrated in urban and rural roads, whereas the length of national highways increased sharply after 2011.

[4]These include the Pradhan Mantri Gram Sadak Yojana (PMGSY), launched in 2000 to expand rural road connectivity, and the Bharatmala Pariyojana, launched in 2017 to develop economic corridors and feeder routes.

accessible cities contribute more to local market access.

Trading costs are proxied by travel time along the road network and are modeled as:

$$\tau_{pct} = 1 + 0.004\,(time_{pct})^{0.8}, \tag{2}$$

where $time_{pct}$ is the shortest travel time (in seconds) between the centroid of postal code p and city c at time t. Travel times are computed using detailed road network data from OpenStreetMap (OSM) for the period 2011–2019 and reflect all road types, including national highways, urban and rural roads, and lower-quality routes.[5] We use the Open Source Routing Machine (OSRM) to compute shortest-path travel times on a static network, ensuring that changes in $time_{pct}$ reflect infrastructure improvements rather than real-time traffic conditions.[6]

We set the distance elasticity parameter to $\theta = 4$, which lies within the range commonly used in gravity-based market access measures (e.g., Donaldson, 2018; Donaldson and Hornbeck, 2016; Redding and Venables, 2004). Holding θ constant, variation in market access over time is driven entirely by changes in travel times to large cities. Road upgrades that reduce travel times therefore increase MA_{pt}, with larger effects for locations that experience greater improvements in connectivity.

Using this approach, we compute market access for 19,998 postal codes relative to 300 cities over the 2011–2019 period. Compared to the binary indicators or simple distance-to-highway measures commonly used in the infrastructure literature (see, e.g., Chandra and Thompson, 2000; Michaels, 2008; Atack et al., 2010; Faber, 2014; Datta, 2012; Ghani et al., 2016; Banerjee et al., 2020), this time-varying, location-specific measure captures both the spatial heterogeneity and the gradual nature of connectivity improvements generated by India's road investments.

Figure 1 maps market access across India in 2011 (Panel (a)) and 2019 (Panel (b)). The color gradient indicates market access levels, MA_{pt}, with lighter colors denoting lower access and darker colors higher access; national highways, including the Golden Quadrilateral, are shown as black lines, and major cities (Tier-1 centers) are marked by red circles proportional to their 2011 population.[7] The figure highlights substantial spatial variation in initial market access and pronounced

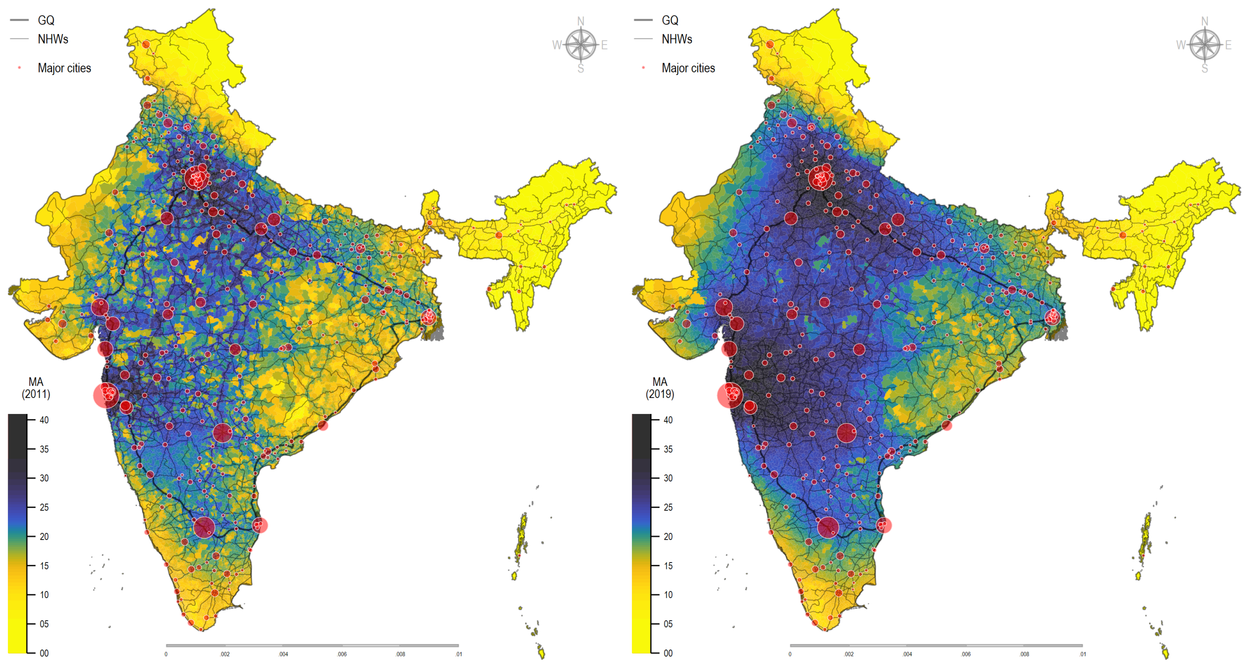
[5]OpenStreetMap (OSM), launched in 2004, relies on crowdsourced data and experienced a gradual expansion in coverage over time. According to Geofabrik, data quality for India prior to 2010 is uneven, particularly for smaller and less frequently used roads, and becomes substantially more reliable thereafter. For this reason, we restrict the construction of our market access measure to the post-2010 period. OSM is widely used in scientific research (see, e.g., Arsanjani et al., 2015).

[6]See Huber and Rust (2016) for details.

[7]Market access is computed using shortest-path travel times on the realized road network rather than geographic distance. Consequently, geographically proximate postal codes may exhibit sharply different market access values if they differ in their connectivity to national highways or feeder roads. Such discontinuities are particularly visible in the pre-expansion period around major metropolitan areas (e.g., south of Delhi and east of Mumbai), where river crossings, topographical constraints, protected land, and regulatory barriers limited road connectivity, forcing indirect travel routes. Because major cities carry substantial economic weight in the market access measure, small differences

Figure 1: Market Access

(a) 2011 (b) 2019



Note: Both panels present visual representations of market access, MA_{pt}, for 2011 (panel a) and 2019 (panel b). The color gradient indicates MA_{pt} levels, with yellow representing lower access and blue representing higher access. National highways, including the GQH, are depicted as black lines, while major cities (Tier-1 centers) are marked by red circles proportional to their population size in 2011.

increases over time, particularly along the GQH and other major highway corridors. These patterns reflect reductions in travel times induced by large-scale road upgrades under the NHDP and form the basis of our empirical analysis.[8]

2.3 Firm-Level Data

Firm- and product-level data come from Prowess dx, a comprehensive panel of Indian firms compiled by the Centre for Monitoring Indian Economy (CMIE). The database contains annual and quarterly financial statements with information on revenues, fixed assets, wage bills, materials,

in travel times translate into large differences in market access.

[8]Appendix Figure A.2 (viewable with Adobe Reader) presents a spatial animation of market access over time. Appendix Figure A.3 shows changes in log market access, $\Delta \ln MA_{pt}$, over two subperiods. The 2011–2013 period, coinciding with the completion of the Golden Quadrilateral, features large increases in market access concentrated along the western GQH corridor connecting major hubs such as Delhi, Mumbai, and Bengaluru. The 2013–2019 period exhibits more geographically diffuse gains, with pronounced increases in central and southwestern regions along corridors linking Mumbai, Bengaluru, and Hyderabad, reflecting subsequent NHDP investments.

product mix, prices, and quantities. Crucially for our analysis, Prowess dx provides detailed firm location data, including postal codes.[9]

Our analysis spans the period 2009–2019. Market access is measured from 2011 onward, reflecting the availability of reliable road network data from OSM, and treatment is defined based on changes in market access between 2011 and 2019, as described in more detail below. We use 2009–2011 as the pre-treatment window. Prior to 2011, national highway expansion was limited—highway density increased by only about 0.5% per year between 2001 and 2011—implying minimal changes in market access during that period. Restricting the analysis to this window also ensures that our identifying variation does not overlap with earlier episodes of financial liberalization in the 1990s and 2000s, which have been shown to affect firm investment and capital allocation (Bau and Matray, 2023).

We focus on a panel of 10,190 firms incorporated before 2000, prior to the announcement of the Golden Quadrilateral project. This restriction mitigates concerns that firm location choices may have been influenced by anticipated road infrastructure investments.[10] Throughout the analysis, nominal variables are converted to real terms using industry-specific deflators from the INDIA KLEMS Database (2020). We further restrict attention to road-dependent industries classified under ISIC Rev. 4 and exclude most service industries.[11]

Marginal Revenue Products As discussed in more detail below, dispersion in marginal revenue products reflects heterogeneity in input price wedges and thus provides information on the extent of resource misallocation. To characterize capital and labor allocation, we compute marginal revenue products following Bau and Matray (2023). We assume a Cobb–Douglas revenue production function,

$$R_{ft} = TFPR_{ft} \prod_{s} y_{fst}^{\alpha_{is}}, \tag{3}$$

where firms f operate in industries indexed by i. R_{ft} denotes firm revenue (value of sales) at time t, y_{fst} is the amount of input $s \in \{k, l, m\}$ used by the firm, where k indexes capital, l labor and m materials. For notational convenience, we write $y_{fkt} = K_{ft}$, $y_{flt} = L_{ft}$, and $y_{fmt} = M_{ft}$. We measure capital K using gross fixed assets, labor L using total employee compensation, and materials M using intermediate inputs. The coefficients α_{is} are industry-specific output elasticities,

[9] Prowess dx covers listed and unlisted public companies as well as private firms of varying sizes and ownership structures, and is widely used in academic research despite not being fully representative of the Indian economy (see, e.g., De Loecker et al., 2016; Goldberg et al., 2010; Bau and Matray, 2023).

[10] As Prowess dx does not permit a reliable analysis of firm exit, we do not study the exit margin. Firms may disappear from the dataset for reasons unrelated to economic shutdown—such as delayed or missing audited financial statements, changes in CMIE's coverage, or temporary breaks in reporting. Firms that exit the dataset are typically small and hold relatively little capital, so their exclusion is unlikely to materially affect our conclusions on capital misallocation.

[11] Appendix Table A.1 lists the industries included in the analysis.

which we estimate in the data and allow to vary across industries, without imposing constant returns to scale. $TFPR_{ft}$ captures firm-specific revenue productivity.

The MRPK is then given by:

$$MRPK_{ft} = \alpha_{ik} \frac{R_{ft}}{K_{ft}}. \tag{4}$$

Because α_{ik} is constant within industry-year cells, it is absorbed by industry-year fixed effects in our empirical specifications. Accordingly, we proxy $MRPK_{ft}$ by the ratio of firm revenue to gross fixed assets,

$$MRPK_{ft} = \frac{R_{ft}}{K_{ft}}. \tag{5}$$

Similarly, the marginal revenue product of labor (MRPL) is computed as the ratio of firm revenue to total labor costs.[12]

Ex Ante Capital and Labor Productivity To study heterogeneity in firms' responses to improved market access, we classify firms based on their ex ante marginal revenue products. Firms are defined as high-MRPK if their average MRPK over the pre-period 2004–2008 exceeds the median within their 4-digit industry; firms below the median are classified as low-MRPK. Averaging over multiple pre-period years reduces measurement error. We apply the same procedure to construct indicators for ex ante high- and low-MRPL firms.

Summary statistics of the main variables, covering the period 2011-2019, are presented in Table A.2 in the appendix.

3 Input Price Wedges and Misallocation

We now clarify how the marginal revenue products measured in Section 2.3 relate to input price wedges and misallocation, and what components of distortions they capture. In line with the existing literature, misallocation arises from firm-level variation in input price wedges, which generates dispersion in marginal revenue products across firms.

The price paid by firm f for input s (capital, labor, or materials) is given by:

$$p_{fs} = (1 + \tilde{\tau}_{fs})p_s, \tag{6}$$

where p_s is the undistorted common input price and $\tilde{\tau}_{fs}$ is a firm-specific wedge, acting as a tax or subsidy and potentially leading to inefficient input allocation in the Arrow–Debreu–McKenzie economy.

[12] Prowess dx does not report employment counts; labor input is therefore measured using total employee compensation.

The first-order conditions (FOCs) of a cost-minimizing firm f imply:

$$(1 + \tilde{\tau}_{fs})p_s = mc_f\, MP_{fs} \quad \forall s, \tag{7}$$

where mc_f is the Lagrange multiplier associated with the cost minimization problem and MP_{fs} denotes the marginal product of input s. By the envelope theorem, mc_f equals the firm's marginal cost. If the firm has pricing power, its markup is given by $\mu_f = p_f/mc_f$, where p_f is the output price. Substituting into (7) yields:

$$p_f MP_{fs} = (1 + \tau_{fs})p_s, \tag{8}$$

where $p_f MP_{fs}$ is the value of the marginal product of input s, $1 + \tau_{fs} = \mu_f(1 + \tilde{\tau}_{fs})$, and τ_{fs} denotes the combined input wedge incorporating both markups and input price distortions (pure wedges).

If both input and output markets were perfectly competitive, markups would be absent and all firms would face the same input prices, implying $(1 + \tau_{fs}) = 1$ and equalization of marginal products across firms. Dispersion in $p_f MP_{fs}$ therefore reflects the presence of wedges and, hence, the extent of misallocation.

Hereafter, we define $MRPS_f \equiv p_f MP_{fs}$ as the marginal revenue product of input s, where $S \in \{K, L, M\}$ denotes capital, labor, or materials. We acknowledge a slight abuse of terminology when firms have pricing power, as changes in input use may also affect output prices.

Building on this discussion, the next section studies how firms' $MRPK$—which reflects the combined effects of markups and input price wedges–evolves following improvements in connectivity, and whether firms with initially high $MRPK$ experience larger capital expansions and sharper subsequent declines in $MRPK$. Such patterns would imply a compression of the $MRPK$ distribution and, hence, a reduction in capital misallocation. We then quantify the resulting aggregate productivity gains and disentangle the relative importance of markup adjustments and changes in pure input wedges.

4 Empirical Approach

To empirically assess these predictions, we estimate empirical specifications in which treatment is defined at the postal-code level based on changes in market access. We exploit two waves of connectivity improvements: the completion of the GQH between 2011 and 2013 and subsequent NHDP phases after 2013. We study the initial wave using a difference-in-differences (DiD) specification based on changes in market access between 2011 and 2013 and the cumulative effects of both waves using a staggered DiD (SDiD) specification over the full 2011-2019 period. Below, we first discuss empirical challenges, then define treatment, and finally present the specifications and results.

4.1 Empirical Challenges

Estimating the effects of improved road infrastructure on firm outcomes raises two main empirical challenges. The first concerns endogenous firm location, as firms—particularly new entrants—may choose to locate in areas with better access to intermediaries and consumers. To mitigate this selection concern, we restrict our analysis to firms incorporated prior to the announcement of the GQH project, i.e., before 2000, and thus operating before the onset of large-scale highway investments.[13] In addition, all specifications include firm FE, which control for time-invariant firm characteristics, including baseline location.

The second empirical challenge is the potential endogeneity of market access. Road investments may be correlated with unobserved, time-varying local economic conditions or policies that also affect firm outcomes. For example, infrastructure improvements may be targeted toward areas experiencing faster growth, leading to biased estimates if such pre-existing trends are not adequately controlled for.

To address this concern, we control for initial local economic conditions using nighttime light intensity at the postal-code level, averaged over the 2004–2008 period and interacted with year FE.[14] We also control for firm age and initial firm size quintile—constructed from average capital over the 2004–2008 period within 1-digit industries—each interacted with year FE, to account for differential adjustment paths across firms.

All specifications further include subdistrict-year FE, which absorb time-varying shocks common to firms located within the same subdistrict, such as local economic conditions and infrastructure investments at the subdistrict or district level. We also include 4-digit industry-year FE to account for industry-specific shocks.

Identification therefore comes from differential changes in market access across firms located in the same subdistrict over time, holding fixed all time-invariant firm characteristics as well as common subdistrict- and industry-level shocks. This variation arises because subdistricts typically

[13]By focusing on firms incorporated before the announcement of the GQH, our analysis abstracts from the effects of road infrastructure on new firm entry and relocation and therefore captures partial-equilibrium responses of pre-existing firms.

[14]Nighttime light data is widely used as a proxy for local economic activity, as regions with higher economic output typically exhibit greater nighttime illumination (Elvidge et al., 1997; Henderson et al., 2012; Chen and Nordhaus, 2011; Pinkovskiy, 2013; Donaldson and Storeygard, 2016; Jean et al., 2016; Tollefsen et al., 2012; Bickenbach et al., 2013; Lall et al., 2017; Tripathy et al., 2016). The data are sourced from Google Earth Engine and originate from the Defense Meteorological Satellite Program's Operational Line-Scan System.

contain multiple postal codes, and changes in market access vary substantially across postal codes within the same subdistrict due to differences in network position and connectivity improvements.

An alternative identification strategy used in the literature exploits planned-route or historical-route instrumental variables.[15] While these IV strategies address concerns related to the non-random placement of infrastructure, they are typically fixed at a single point in time and therefore unsuitable in our setting, which exploits time variation in firms' exposure to road infrastructure. Moreover, many IV approaches focus on aggregate outcomes at the district or county level and are not well suited to capturing the granular firm-level capital adjustments that are central to our analysis.

4.2 Treatment definitions

Treatment in the DiD Our DiD framework defines treatment at the postal-code level based on growth in market access between 2011 and 2013. Postal codes with above-median $\Delta \ln MA_{pt}$ over this period are classified as treated, and firms located in these postal codes are considered exposed to treatment in all years from 2014 onward.

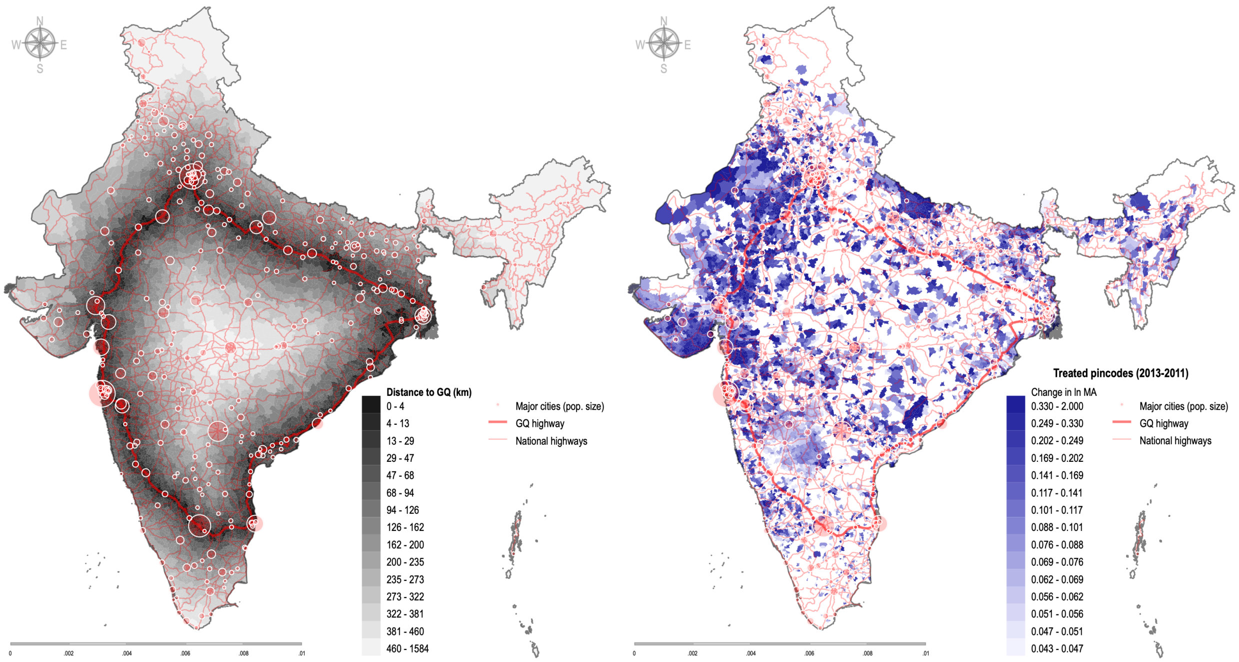
This definition differs from the standard approach in the literature, which typically assigns treatment based on proximity to major highways (Chandra and Thompson, 2000; Michaels, 2008; Atack et al., 2010; Faber, 2014; Datta, 2012; Ghani et al., 2016; Banerjee et al., 2020). As we illustrate below, firms located at similar distances from the GQH can experience markedly different changes in market access, indicating that distance alone is an imperfect proxy for actual exposure to connectivity improvements.

Panel (a) of Figure 2 depicts proximity to the GQH, measured as the shortest distance from each postal code to the highway, with darker shades of gray indicating closer locations. Panel (b) maps changes in market access, $\Delta \ln MA_{pt}$, between 2011 and 2013, displaying only treated postal codes with above-median increases; darker shades of blue correspond to larger gains in market access. In both panels, red lines denote highways, and circles mark major cities, scaled by their 2011 population.

[15] For example, Baum-Snow (2007) uses the 1947 plan of the U.S. interstate highway network to study urban decentralization, while other studies instrument for modern infrastructure using historical transportation networks such as 19th-century railroads. The inconsequential-units approach (Chandra and Thompson, 2000) instead focuses on regions incidentally affected by infrastructure placement.

Figure 2: Treated Postal Codes based on Distance to the GQH and $\Delta \ln MA_{pt}$ between 2011-2013

(a) Proximity to the GQH (b) Treated postal codes: 2011-2013



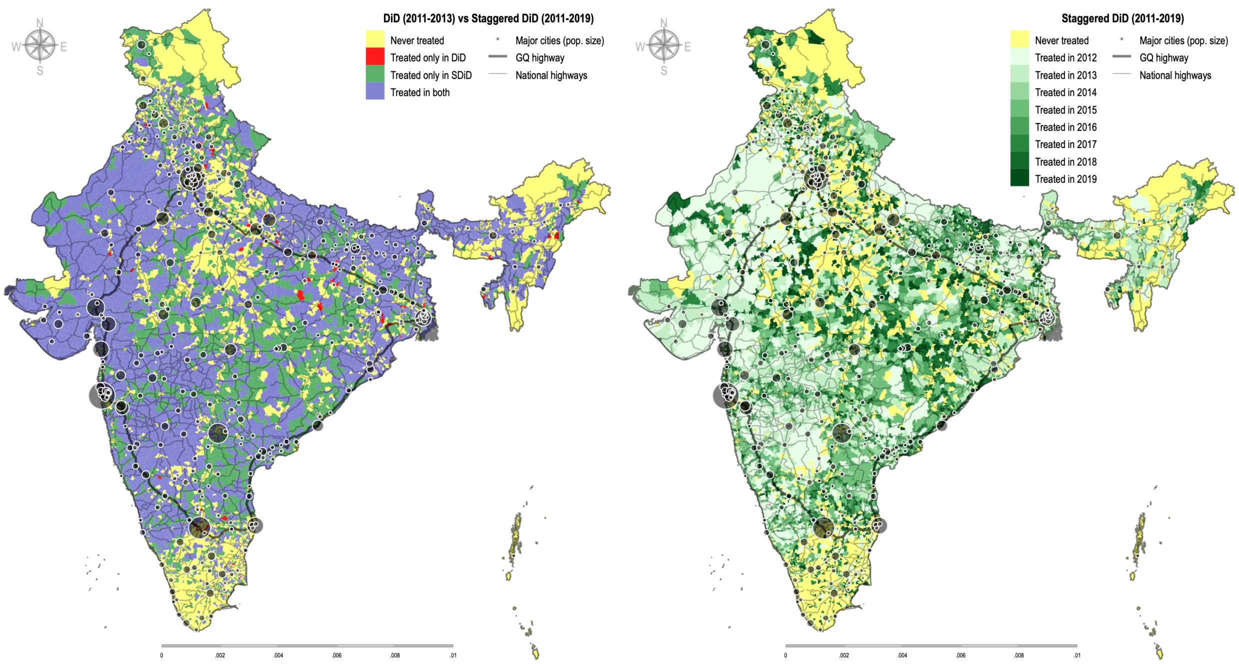
Note: Panel (a) depicts proximity to the GQH, measured as the shortest distance from each postal code to the highway, with darker shades of gray indicating closer locations. Panel (b) maps changes in market access ($\Delta \ln MA_{pt}$) from 2011 to 2013, displaying only treated postal codes with above-median increases, where deeper blue represents larger improvements. The color scale divides the upper half of the distribution of market access changes into 15 equally sized groups (bins), each representing approximately 3.4% of the sample. Lighter shades indicate smaller gains above the median, while darker shades correspond to the largest improvements. Red lines indicate highways, and red circles mark major cities, scaled by population size in 2011.

Comparing panels (a) and (b) highlights a key limitation of distance-based treatment definitions: they rely solely on proximity to the highway, regardless of whether connectivity actually improves. This discrepancy is particularly visible along the northeastern and southeastern segments of the GQH—most notably along the New Delhi–Kolkata and Kolkata–Chennai corridors—where some postal codes located close to the highway exhibit little change in market access. Conversely, several areas in Rajasthan and Punjab, which are relatively distant from the GQH and would therefore be classified as untreated under a distance-based rule, experience substantial increases in market access.

Treatment in the SDiD The DiD specification focuses on changes in market access between 2011 and 2013, corresponding to the completion of the GQH. This window, however, does not capture the substantial connectivity improvements associated with later NHDP phases. We there-

Figure 3: Comparison of Treated Postal Codes in the DiD and SDiD

(a) DiD vs SDiD (b) SDiD by year



Note: Figure 3 compares treatment assignments under the DiD and SDiD by mapping treated postal codes across the two approaches. Red areas are treated only in the DiD, green only in the SDiD, and purple in both. Yellow areas are untreated. Panel (b) shows treatment timing in the SDiD, with yellow for untreated locations and green for treated ones; lighter greens indicate earlier treatments. Black lines represent national highways, and black circles mark major Indian cities, scaled by population size in 2011.

fore complement the DiD analysis with a SDiD approach, which allows postal codes to become treated at different points in time after 2011, capturing the phased and spatially heterogeneous rollout of the NHDP.

Under the SDiD design, a postal code becomes treated in the first year after 2011 in which its cumulative growth in log market access exceeds the median cumulative growth across postal codes over the 2011–2019 period. Treatment is absorbing thereafter.

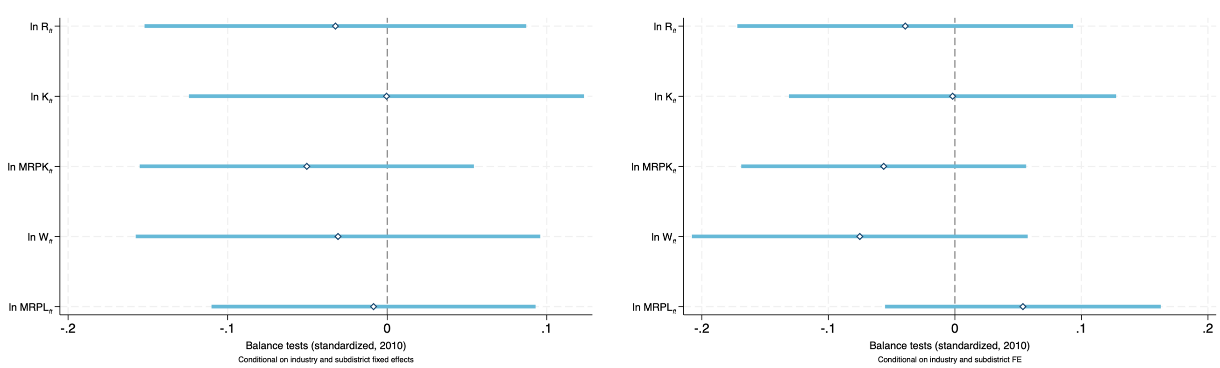
Figure 3 compares treatment assignments under the DiD and SDiD by mapping treated postal codes across the two approaches. Panel (a) shows the geographic distribution of treatment status: red areas are treated only in the DiD, green areas only in the SDiD, purple areas in both, and yellow areas remain untreated. Black lines denote national highways, and black circles mark major cities with populations above 100,000, scaled by their 2011 population. Panel (b) illustrates the timing of treatment in the SDiD, with lighter green shades indicating earlier treatment and darker shades later treatment, highlighting the NHDP's gradual and geographically heterogeneous rollout.

The figure shows that nearly all postal codes treated in the DiD are also treated in the SDiD, especially those affected early by the GQH. Panel (b) further indicates that a substantial share of postal codes classified as controls in the DiD become treated in the SDiD as a result of highway investments in later NHDP phases. Many of these newly treated areas are located farther from the GQH. As a result, differences between the DiD and SDiD estimates primarily reflect the impact of post-2013 infrastructure investments that expand market access beyond the initial GQH corridor.

Balance tests Before turning to the estimation results, we assess whether treated and control firms differ systematically prior to treatment. Balance tests are conducted in 2010, a common pre-treatment year for all units. Conditional on industry and subdistrict FE, we find no systematic differences in pre-treatment outcomes between treated and control firms. Appendix Figure A.4 reports balance tests for additional firm outcomes used in the robustness analysis and similarly shows no systematic pre-treatment differences between treated and control firms.

Figure 4: Balance checks

(a) DiD (b) SDiD



Note: The figure plots the coefficients (dark blue diamonds) and 95% confidence intervals (light blue bars) from regressions of firm-level main outcomes—revenues (R_{ft}), capital (K_{ft}), MRPK ($MRPK_{ft}$), labor input (W_{ft}) and MRPL ($MRPL_{ft}$)—measured in 2010 on a treatment indicator. Regressions are conditional on 4-digit industry FE and subdistrict FE, with standard errors clustered at the postal-code level. All variables are normalized to have mean zero and standard deviation one. Panel (a) corresponds to the DiD design, where treatment is defined based on postal codes that become treated after 2013. Panel (b) corresponds to the SDiD design, where, for balance checks, treatment equals one for ever-treated postal codes and zero for never-treated ones.

4.3 Empirical Strategy

We estimate the effects of improved market access on firm outcomes using the following specification:

$$Y_{ft} = \alpha + \beta_1 T_{pt} + \beta_2 T_{pt} \times MRPK_{f0}^{high} + \boldsymbol{X\gamma} + d_f + d_{it} + d_{ct} + \varepsilon_{ft}, \tag{9}$$

where firm f operates in industry i and is located in postal code p in year $t \in [2009, 2019]$. The dependent variable Y_{ft} is the log of either revenues, capital, or MRPK. The treatment indicator T_{pt} varies across postal codes and over time and is defined according to the DiD or SDiD design described in Section 4.2. In the DiD, T_{pt} equals one for treated postal codes in all years from 2014 onward and zero otherwise. In the SDiD, T_{pt} equals one from the first year a postal code becomes treated and remains one thereafter.

The indicator $MRPK_{f0}^{high}$ equals one for ex ante high-MRPK firms, constructed as described in Section 2.3. The vector $\mathbf{X}$ includes firm age, initial firm size quintile interacted with year FE, and initial nightlights at the postal code level also interacted with year FE. The term d_f denotes a complete set of firm FE, d_{it} industry-year FE (using 4-digit ISIC codes), and d_{ct} subdistrict-year FE. Thus, identification comes from differential changes in market access across postal codes within the same subdistrict over time, after absorbing firm fixed effects and common industry- and subdistrict-level shocks. Standard errors are clustered at the postal code-year level, the unit of measurement for our treatment variable.

Our main coefficients of interest are β_1 and β_2. Without the interaction term, β_1 captures the average effect of improved connectivity on treated firms. Inclusion of the interaction term allows us to examine heterogeneity by firms' initial MRPK levels. In this case, β_1 measures the effect on initially low-MRPK firms, while $\beta_1 + \beta_2$ captures the effect on initially high-MRPK firms.

If improved connectivity relaxes input frictions and induces capital to reallocate toward firms with initially high capital wedges, we expect a stronger increase in capital among these firms, implying $\beta_2 > 0$ when $Y_{ft} = \ln K_{ft}$. Under the same mechanism, and holding revenue responses constant, such reallocation should be associated with a relative decline in MRPK for initially high-MRPK firms, implying $\beta_2 < 0$ when $Y_{ft} = \ln MRPK_{ft}$. Evidence of this pattern would indicate a narrowing dispersion of capital wedges among firms in treated postal codes and suggest a reduction in misallocation.

5 Results

5.1 Capital Adjustments

Table 1 reports the DiD estimates for the completion of the GQH. Panel (a) presents average treatment effects, while Panel (b) allows for heterogeneity by firms' initial MRPK.

On average, the GQH had no statistically significant effect on firm revenues (Panel (a), column 1). However, this average masks substantial heterogeneity. Once we allow for differential responses by initial MRPK, revenues decline significantly for firms with ex ante high MRPK, while remaining unchanged for initially low-MRPK firms (Panel (b), column 4). As discussed in Section Appendix F, this decline reflects adjustments in product-level prices, quantities, and product scope following improved connectivity.

In contrast, capital increases significantly following the GQH, with larger gains among firms with initially high MRPK. Capital rises by 46% for initially low-MRPK firms, with an additional 19% increase for high-MRPK firms (Panel (b), column 5). We also observe a pronounced decline in MRPK following treatment, which is substantially larger among firms with initially high MRPK. In particular, MRPK falls by an additional 42% for initially high-MRPK firms relative to low-MRPK firms (Panel (b), column 6).

Taken together, these patterns indicate that improved market access is associated with stronger capital expansion among firms facing higher initial capital wedges, alongside more muted revenue responses, leading to a compression of MRPK among treated firms. This pattern is consistent with a reduction in capital misallocation following the completion of the GQH.

Table 1: DiD Estimates of the Impact of the GQH

	(1) $\ln R_{ft}$	(2) $\ln K_{ft}$	(3) $\ln MRPK_{ft}$	(4) $\ln R_{ft}$	(5) $\ln K_{ft}$	(6) $\ln MRPK_{ft}$
	(a) Average effects			*(b) Differential effects*		
T_{pt}	-0.173	0.612^{***}	-0.786^{***}	0.007	0.463^{***}	-0.465^{**}
	(0.249)	(0.138)	(0.244)	(0.246)	(0.134)	(0.232)
$\times MRPK_{f0}^{high}$				-0.236^{***}	0.193^{***}	-0.418^{***}
				(0.054)	(0.031)	(0.058)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	45300	45300	45300	45300	45300	45300
R^2	0.94	0.97	0.86	0.94	0.97	0.86

Note: This table presents the DiD estimates of the average effects (Columns 1-3) and the heterogeneous effects (Columns 4-6) of the GQH on log firm-level revenues (R_{ft}), capital (K_{ft}), and MRPK ($MRPK_{ft}$). T_{pt} and $MRPK_{f0}^{high}$ indicate treatment and ex ante high MRPK, respectively. Controls include firm FE, industry-year FE, subdistrict-year FE, firm age, initial firm size interacted with year FE and initial nightlights at the postal code level interacted with year FE. Standard errors clustered at the postal code-year level. Significance: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table 2 reports the SDiD estimates for the NHDP, following the same structure as Table 1. Panel I presents estimates based on two-way fixed effects (TWFE).

Unlike the DiD estimates, the SDiD analysis reveals a positive effect on revenues. Allowing for heterogeneity, revenues increase by approximately 38% for firms with initially low MRPK (p-value = 0.06) and by about 13% for firms with initially high MRPK (column 4). Capital also increases following NHDP investments: firms with ex ante low MRPK increase capital by 25%, with an additional 24% increase for firms with ex ante high MRPK (column 5). MRPK remains unchanged for low-MRPK firms, while declining strongly for firms with high initial MRPK. This pattern indicates that, for low-MRPK firms, capital expansion is largely matched by revenue growth, whereas for high-MRPK firms capital grows more rapidly relative to revenues. Overall, these findings point to a reduction in capital misallocation associated with the NHDP.[16]

In SDiD settings, TWFE estimators may suffer from negative weighting and obscure heterogeneous treatment effects (see e.g., de Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021; Sant'Anna and Zhao, 2020). To assess this concern, we conduct a Bacon decomposition and find that over 90% of the identifying variation arises from comparisons between treated and

[16] Appendix Table A.3 reports results on changes in the variance of MRPK across firms. Consistent with a reduction in capital misallocation, estimated effects are negative in both specifications, although less precisely estimated in the SDiD design, as is common for second-moment outcomes. In economic terms, the DiD estimate corresponds to a decline in the variance of log MRPK of roughly 10% relative to its average pre-treatment level in treated areas, with a comparably sized point estimate in the SDiD specification.

never-treated units, suggesting limited scope for bias from problematic comparisons. We further examine the weighting structure following de Chaisemartin and d'Haultfoeuille (2020) and find that approximately 1.6% of treated group-time cells receive negative weights, accounting for less than 0.2% of the total weight. Finally, event-study plots show no evidence of differential pre-trends (Figure A.5) and, as shown in Section 4.2, treated and control firms are balanced on pre-treatment outcomes conditional on our fixed effects.

To further address concerns related to staggered timing, we also estimate a stacked regression approach (Baker et al., 2022). This approach constructs cohort-specific samples by stacking never-treated observations for each treatment cohort and replaces firm FE with cohort-firm fixed effects while adding cohort-time-to-treatment FE, ensuring that comparisons are made within each cohort while accounting for differential treatment timing. Results are reported in Panel II and are similar, though more precisely estimated, than the TWFE estimates in Panel I. Given these results, we proceed with the TWFE estimator for the remainder of the paper.

Table 2: SDiD Estimates of the Impact of the NHDP

	(1) $\ln R_{ft}$	(2) $\ln K_{ft}$	(3) $\ln MRPK_{ft}$	(4) $\ln R_{ft}$	(5) $\ln K_{ft}$	(6) $\ln MRPK_{ft}$
I.TWFE	*(a) Average effects*			*(b) Differential effects*		
T_{pt}	0.293	0.330**	−0.027	0.378*	0.247**	0.129
	(0.199)	(0.137)	(0.212)	(0.201)	(0.126)	(0.202)
$\times MRPK_{f0}^{high}$				−0.249***	0.242***	−0.457***
				(0.057)	(0.037)	(0.060)
N	44806	44806	44806	44806	44806	44806
R^2	0.94	0.97	0.86	0.94	0.97	0.86
II. Stacked	*(a) Average effects*			*(b) Differential effects*		
T_{pt}	0.391*	0.476***	−0.077	0.478**	0.387***	0.088
	(0.217)	(0.154)	(0.178)	(0.227)	(0.143)	(0.184)
$\times MRPK_{f0}^{high}$				−0.238***	0.246***	−0.452***
				(0.053)	(0.035)	(0.056)
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table presents the SDiD estimates of the average effects (Columns 1-3) and the heterogeneous effects (Columns 4-6) of the NHDP on log firm-level revenues (R_{ft}), capital (K_{ft}), and MRPK ($MRPK_{ft}$). Panel I is estimated using TWFE, while Panel II is estimated using a stacked regression. T_{pt} and $MRPK_{f0}^{high}$ indicate treatment and ex ante high MRPK, respectively. In Panel I, controls include firm FE, industry-year FE, subdistrict-year FE, firm age, initial firm size interacted with year FE and initial nightlights at the postal code level interacted with year FE. In Panel II, we use the same set of controls but replace firm FE with cohort-firm FE and add cohort-time-to-treatment FE. Number of observations is reported for TWFE but omitted for stacked regressions due to cohort stacking. Standard errors clustered at the postal code-year level. Significance: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

5.2 Labor Adjustments

While our primary focus is on the reallocation of capital, the associated increase in capital intensity raises the question of whether firms substitute capital for labor and whether the allocation of labor also improves. To examine this, we study changes in labor inputs—measured by total employee compensation—and in the MRPL. Results are reported in Table 3, with Panel I presenting DiD estimates and Panel II reporting SDiD results.

Panel I shows that labor costs decline significantly during the first phase of the NHDP, with larger reductions among firms with low ex ante MRPL. This pattern indicates that labor inputs contract following improved market access, particularly among firms that initially employed relatively unproductive labor. When accounting for later NHDP phases using the SDiD specification (Panel II), estimated effects on labor costs are close to zero and imprecisely estimated. This suggests that adjustments along the labor margin were largely concentrated in the early phase of the NHDP.

Despite these differences in timing, MRPL increases in both phases (Column 4). This rise may reflect changes in product-market markups as well as declines in wage markdowns, possibly due to reduced labor market frictions and strengthen workers' outside options associated with improved connectivity. Additionally, column (4) shows that in both phases, MRPL increases substantially more so for firms with initially low MRPL, leading to a decline the dispersion in MRPL and hence, an improvement in the relative allocation of labor.

Taken together, these findings suggest that labor adjusts primarily in the short run, while capital deepening constitutes the dominant and more persistent response to improved market access.

Table 3: Labor Input Adjustments

	(1) $\ln W_{ft}$	(2) $\ln MRPL_{ft}$	(3) $\ln W_{ft}$	(4) $\ln MRPL_{ft}$
I. DiD: GQH				
	(a) Average effects		*(b) Differential effects*	
T_{pt}	-0.508***	0.318	-0.526***	0.413**
	(0.131)	(0.195)	(0.132)	(0.208)
$\times MRPL_{f0}^{high}$			0.062*	-0.330***
			(0.036)	(0.047)
N	45300	45300	45300	45300
II. SDiD: NHDP				
	(a) Average effects		*(b) Differential effects*	
T_{pt}	-0.041	0.342*	-0.066	0.461**
	(0.109)	(0.174)	(0.111)	(0.192)
$\times MRPL_{f0}^{high}$			0.073*	-0.353***
			(0.039)	(0.043)
N	44806	44806	44806	44806
Controls	Yes	Yes	Yes	Yes

Note: This table presents the average effects (Columns 1-2) and the heterogeneous effects (Columns 3-4) of improved market access on log firm-level labor input (W_{ft}) and MRPL ($MRPL_{ft}$). Panel I shows DiD results and Panel II SDiD results. T_{pt} and $MRPL_{f0}^{high}$ indicate treatment and ex ante high MRPL, respectively. Controls include firm FE, industry-year FE, subdistrict-year FE, firm age, initial firm size interacted with year FE and initial nightlights at the postal code level interacted with year FE. Standard errors clustered at the postal code-year level. Significance: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

5.3 Robustness

Our baseline treatments define exposure using the median change in market access. Appendix Tables A.4 and A.5 explore alternative definitions based on the 40th and 60th percentiles of market access growth. Across these specifications, results for capital and MRPK remain similar to the baseline, indicating that our findings are not sensitive to the choice of cutoff.

Existing work suggests that firms' responses to improvements in market access may be non-linear, with economically meaningful adjustments often emerging only once connectivity gains are sufficiently large (Asturias et al., 2019; Donaldson, 2018; Faber, 2014). Motivated by this evidence, and to mitigate attenuation bias arising from measurement error in continuous market access indices, our baseline analysis relies on a threshold-based treatment that captures sizable changes in connectivity. As a robustness check, Appendix Table A.6 replaces the binary treatment with the continuous log market access measure ($\ln MA_{pt}$). Results remain of the same sign and overall statistical significance as in the baseline DiD and SDiD estimates when market access is measured continuously.[17]

6 Aggregate Effects

This section examines how the firm-level adjustments documented above translate into aggregate productivity gains. We focus on the aggregate effects of the NHDP as a whole, using the SDiD estimates.

To this end, we follow Bau and Matray (2023) and quantify changes in the Solow residual driven by adjustments among treated firms (i.e., firms in treated postal codes). The change in the Solow residual is given by:

$$\Delta \text{Solow}_R = \Delta \text{Net Output}_R - \Delta \text{Net Input}_R, \tag{10}$$

where R denotes the set of treated postal codes. ΔNet Output$_R$ captures the change in output of treated firms, net of the portion reused as inputs within this group. ΔNet Input$_R$ represents the change in net input used by treated firms, excluding the inputs they produce.

In Appendix C, we show that equation (10) can be rewritten, in log-change form, as:

$$\overline{\text{Solow}}_R \approx \underbrace{\sum_i \sum_{f\in(i,R)} \lambda_f \overline{TFPQ}_f}_{\text{I. Changes in production efficiency}} + \underbrace{\sum_i \sum_{f\in(i,R)} \sum_s \lambda_f \alpha_{is} \frac{\tau_{fs}}{1+\tau_{fs}} \overline{y}_{fs},}_{\text{II. Changes in inputs}} \tag{11}$$

where overlined variables $\overline{x}$ represent log-changes, and $s \in \{k, l, m\}$ indexes capital (k), labor (l), and materials (m). λ_f is firm f's sales share in the total net output of treated firms, $TFPQ_f$ denotes firm-level TFP quantity, α_{is} is the output elasticity of industry i with respect to input s, τ_{fs} denotes the firm-level input wedge, and y_{fs} is firm f's usage of input s.

This aggregation formula is similar to those in Bau and Matray (2023), Baqaee and Fahri (2019) and Petrin and Levinsohn (2012). It highlights that aggregate productivity gains can arise both from changes in production efficiency (Component I) and from input reallocation when $\tau_{fs} \neq 0$ (Component II). The reallocation term ($\lambda_f \alpha_{is} \frac{\tau_{fs}}{1+\tau_{fs}} \overline{y}_{fs}$) is positive when inputs expand at firms with positive wedges ($\overline{y}_{fs} > 0$, $\tau_{fs} > 0$) or contract at firms with negative wedges ($\overline{y}_{fs} < 0$, $\tau_{fs} < 0$).

Implementation To quantify changes in aggregate productivity, we assume no misallocation in materials ($\tau_{fm} = 0$), as in Bau and Matray (2023). The remaining components of the Solow

[17] Using the continuous measure reduces the sample by three years and leads to additional losses due to singleton observations. While the implied elasticity of capital with respect to market access is sizable, the average increase in market access between 2011 and 2019 is approximately 5%, implying an average capital increase of about 10%.

residual in equation (11) are either directly observed in the data or estimated using the same SDiD design as in the firm-level analysis, augmented with interaction terms for ex ante high-MRPK and high-MRPL firms. We use this augmented specification throughout this section.

To estimate $\overline{TFPQ}_f$, we run a TFPQ regression with heterogeneous treatment effects by ex ante MRPK and MRPL status; Appendix D describes the measurement of TFPQ, and Appendix Table A.7 reports the results.[18] The estimates are statistically insignificant throughout, indicating that firms adjusted their input mix without experiencing detectable changes in intrinsic production efficiency. Accordingly, any aggregate productivity gains appear to be driven primarily by reduced misallocation rather than improvements in firm-level efficiency. Following Bau and Matray (2023), we therefore abstract from changes in TFPQ in the aggregate accounting, so that the first component of equation (11) is set to zero.

The second component in equation (11) captures the effects of changes in input misallocation for treated firms and requires estimates of α_{is}, $\overline{y}_{fs}$, λ_f and τ_{fs}.

As is standard, industry-level output elasticities, α_{is}, are obtained from production function estimations. Predicted changes in inputs, $\overline{y}_{fs}$, are estimated using the augmented SDiD specification.[19] We then use these estimates to predict changes in inputs for firms that are eventually treated within the treatment window.[20]

The sales share λ_f is computed using data from Prowess dx and the World Input-Output Database (WIOD). We compute λ_f for each year between 2004 and 2008 and use the average over this period.[21] Treated firms' nominal net output is calculated as total nominal output minus the value of output reused as intermediate inputs within the treated group. We proxy the value of output that treated firms reuse as intermediate input as follows:

$$\sum_i \sum_j Output_{iR} \cdot \eta_{ji} \cdot \eta_{jR}, \tag{12}$$

where $Output_{iR}$ denotes the aggregate nominal output of treated firms in industry i, η_{ji} is the share of output from industry i used as an input in industry j (from WIOD), and η_{jR} is the share of treated firms in industry j.[22]

The final inputs required for the second component in equation (11) are the (ex ante) input

[18]Note that, due to a substantial share of missing firm-product prices, the number of observations is lower than in the other firm-level regressions.

[19]Estimates from this specification are reported in columns (2) and (4) of Appendix Table A.8.

[20]Note that this approach abstracts from general equilibrium spillovers from treated to untreated firms.

[21]The WIOD provides World Input-Output Tables and underlying data covering 43 countries, along with a model for the rest of the world, for the period 2000–2014. It includes data for 56 sectors classified according to the International Standard Industrial Classification revision 4 (ISIC Rev. 4) and follows the 2008 System of National Accounts.

[22]Using the share of treated firms' sales in industry j to construct η_{jR} yields nearly identical aggregate productivity estimates.

wedges, τ_{fs}. Following the conventional approach, we estimate these wedges as:

$$\begin{aligned}\tau_{fk} &= \alpha_{ik}\frac{p_i y_f}{rK_f} - 1 \\ \tau_{fl} &= \alpha_{il}\frac{p_i y_f}{wL_f} - 1,\end{aligned} \tag{13}$$

where output elasticities are estimated as described earlier, firm sales ($p_i y_f$), capital and the wage bill (wL_f) are observed in Prowess dx, and the rental rate of capital is set to $r = 0.1$, as in Hsieh and Klenow (2009a).

In the conventional approach, all pre-treatment cross-sectional deviations of expenditure shares from output elasticities are attributed to misallocation. This may overstate aggregate effects in the presence of measurement error in expenditure shares. To address this concern, Bau and Matray (2023) propose a more conservative approach that yields a lower bound on the contribution of each input to aggregate productivity. This approach is valid under the assumption that the policy—here, improvements in MA—(weakly) reduced wedges toward zero, with inputs increasing for firms with ex ante positive wedges and decreasing for firms with ex ante negative wedges. Under this assumption:

$$\sum_i \sum_{f\in(i,R)} \sum_s \lambda_f \alpha_{is} \frac{\tau_{fs}}{1+\tau_{fs}} \overline{y}_{fs} \geq -\sum_i \sum_{f\in(i,R)} \sum_s \lambda_f \alpha_{is} \frac{\Delta\tau_{fs}}{1+\Delta\tau_{fs}} \overline{y}_{fs}, \tag{14}$$

and the right-hand side provides a lower bound on the contribution of input reallocation (Component II).

Our results suggest that this assumption is likely satisfied for capital but unclear for labor in the augmented SDiD specification. We therefore apply the conservative approach only when computing the contribution of capital to aggregate productivity.[23]

In this case, we estimate changes in capital wedges, $\Delta\tau_{fk}$, as:

$$\widehat{\Delta\tau_{fk}} = e^{\log(\widehat{1+\Delta}\tau_{fk})} - 1, \tag{15}$$

where $\log(\widehat{1+\Delta}\tau_{fk}) = \widehat{\beta}_1 T_{pt} + \widehat{\beta}_2 T_{pt} \times MRPK^{high}_{f0} + \widehat{\beta}_3 T_{pt} \times MRPL^{high}_{f0}$, and the $\widehat{\beta}$ coefficients are obtained from the augmented SDiD MRPK regression (Column 3 of Appendix Table A.8).

Results Table 4 reports estimated changes in aggregate productivity, computed using the Solow residual decomposition in equation (11), for treated firms. Columns (1) and (4) report the capital contribution, columns (2) and (5) the labor contribution, and columns (3) and (6) the aggregate effect; columns (4) and (6) use the conservative lower-bound approach.

[23] Applying the conservative approach to labor yields a lower bound for the labor contribution that is close to zero.

We find aggregate productivity gains ranging from 2.67% to 5.08%. Capital contributes between 1.34% and 3.77%, while labor also contributes positively, at around 1.33%.

Interestingly, the aggregate gains under the NHDP are comparable to those estimated by Bau and Matray (2023) following India's foreign capital liberalization in 2001 and 2006 under the conservative approach. While the policy focus (market access vs. deregulation), timing (post-2011 vs. 2001/2006), and source of variation (postal code-year vs. industry-year) differ, the magnitude of the aggregate productivity effects is remarkably similar.

Table 4: Changes in Aggregate Productivity

Conventional (K and W)			Lower-bound (K) & Conventional (W)		
(1) Capital	(2) Labor	(3) Aggregate Effects	(4) Capital	(5) Labor	(6) Aggregate Effects
3.77%	1.33%	5.08%	1.34%	1.33%	2.67%

Note: The table reports estimated aggregate productivity effects of the NHDP for treated firms. Columns (1) and (4) report the capital contribution, columns (2) and (5) the labor contribution, and columns (3) and (6) the aggregate effect; columns (4) and (6) use the conservative lower-bound approach.

7 Mechanisms: Markups versus Pure Input Wedges

Even though combined wedges capture changes in misallocation, they are not directly informative about the mechanism through which these changes operate. By definition, they aggregate the contributions of pure input wedges and markups (overlined variables denote log changes):

$$\underbrace{\overline{(1+\tau_{fs})}}_{\text{Combined Wedge}} = \underbrace{\overline{(1+\tilde{\tau}_{fs})}}_{\text{Pure Wedge}} + \underbrace{\overline{\mu}_f}_{\text{Markup}}$$

Accordingly, the substantial variation in the combined input wedges that we capture may originate from changes in pure input wedges, which reflect improved access to inputs, or from changes in markups, which reflect increased competitive pressures. This section disentangles these two channels and separately quantifies the effects of pure input wedges and markups.

Implementation The changes in combined input wedges, $\overline{(1+\tau_{fs})}$, are estimated from the augmented SDiD specification introduced earlier (columns (3) and (5) of Appendix Table A.8).

To identify markup changes, $\overline{\mu}_f$, we build on the framework of Amiti et al. (2019). We impose two standard assumptions on demand for the final product of the firm f: (1) demand is homogeneous of degree 0 in terms of prices, so that it is not affected if all prices change proportionally; (2) demand is homogeneous of degree 1 in terms of demand shifters, so that if all firms receive the same demand shifter, their demand increases proportionally.

Additionally, we rely on two features of our empirical framework. First, consistent with our earlier findings, we set $\overline{TFPQ}_{ft} = 0$. Second, we assume that treated firms within an industry are exposed to a common demand shock and that the set of treated firms is representative of firms within each industry.

These assumptions and results allow us to obtain the following equation for markup changes (see Appendix E for full derivation details):

$$\overline{\mu}_f = \frac{1-\rho_f}{\rho_f}\left[\underbrace{\left(\frac{1}{\alpha_i}-1\right)\sum_s \alpha_{is}\left(\sum_f \delta_f \overline{y}_{fs} - \overline{y}_{fs}\right)}_{\text{Scale Effect}} + \underbrace{\sum_s \frac{\alpha_{is}}{\alpha_i}\left(\sum_f \delta_f \overline{(1+\tau_{fs})} - \overline{(1+\tau_{fs})}\right)}_{\text{Cost Effect}}\right], \tag{16}$$

where ρ_f denotes the modified pass-through of marginal costs into prices and δ_f is the firm's revenue share within its industry. Modified pass-through is connected to $\tilde{\rho}_f$ pass-through estimated by Amiti et al. (2019) via $\rho_f = \frac{\tilde{\rho}_f(1-\delta_f)}{1-\tilde{\rho}_f\delta_f}$. Lastly, there is a parameter of returns to scale $\alpha_i = \sum_s \alpha_{is}$.

Equation (16) highlights two distinct channels of markup changes. The first term captures adjustments driven by changes in relative production scale within the industry. The second term captures adjustments driven by changes in relative input wedges. Crucially, only idiosyncratic changes affect markups; industry-wide shocks do not.

To identify markups from equation (16), we set the pass-through parameter to $\tilde{\rho}_f = 0.6$, the midpoint of the estimates in Amiti et al. (2019), and calculate revenue shares δ_f directly from the Prowess dx data. Input shares α_{is} are, as before, taken from the production function estimation. Input changes $\overline{y}_{fs}$ and combined wedge changes $\overline{(1+\tau_{fs})}$ are obtained from the estimates in the augmented SDiD specification introduced earlier, allowing for heterogeneous effects by initial MRPK and MRPL (Appendix Table A.8).

Finally, pure wedges $\overline{(1+\tilde{\tau}_{fst})}$ are derived from the identity: $\overline{(1+\tau_{fst})} = \overline{\mu}_{ft} + \overline{(1+\tilde{\tau}_{fst})}$.

Results Table 5 reports the estimated changes in markups, pure input wedges, and combined wedges following improvements in market access. The results are averages for the treated firms and for the groups of treated firms with different initial levels of MRPK and MRPL. Column (1) shows that markups decline modestly, by about 2.7% on average. In contrast, columns (2) and (3) show large changes in pure input wedges: the pure capital wedge declines sharply, while the pure labor wedge increases substantially. Columns (4) and (5) show that combined wedges closely mirror the behavior of pure wedges.

Table 5: Changes in markups, pure and combined input wedges

		Pure wedges		Combined Wedges	
	(1)	(2)	(3)	(4)	(5)
	Markup	Capital	Labor	Capital	Labor
Total	-2.65%	-31.9%	28.7%	-33.5%	26.7%
High MRPK	0.89%	-59.0%	14.5%	-58.2%	15.3%
Low MRPK	-5.77%	-8.0%	41.3%	-13.2%	36.1%
High MRPL	1.52%	-55.5%	8.9%	-53.7%	10.4%
Low MRPL	-7.73%	-3.1%	52.9%	-10.1%	45.5%

Note: The table reports the estimated effects of the NHDP on the markup, pure and combined input wedges. The results are averages for the treated firms and for the groups of treated firms with different initial levels of MRPK and MRPL. Column (1) reports the change in markup, columns (2) and (3) report changes of pure capital and labor wedges respectively, and columns (4) and (5) report changes of combined capital and labor wedges.

These patterns hold across firms with different initial levels of MRPK and MRPL. Even in subsamples where markup adjustments are more pronounced, changes in pure input wedges account for the bulk of the variation in combined wedges. Overall, the reduction in misallocation documented earlier is overwhelmingly driven by improved access to inputs rather than by increased competitive pressure in output markets.

This conclusion is consistent with our product-level evidence. As shown in Appendix F, for the SDiD, we find zero or mildly positive effects for revenues and quantities, and noisy negative effects on prices. Both these findings are consistent with the effects of declining production costs, where we would expect to observe an increase in quantities and decline in prices. Product-level regressions additionally show strong effects on product scope and turnover. Together, these results suggest that improvements in market access primarily operate by lowering effective input costs and relaxing firm-level constraints, rather than by intensifying price competition in final goods markets.

8 Conclusion

This paper shows that improvements in market access driven by highway expansions can substantially reduce input misallocation and raise aggregate productivity. Focusing on connectivity gains induced by India's NHDP, we document large reallocations of inputs across firms as transport frictions decline.

At the firm level, these adjustments are concentrated in capital. Improvements in market access lead to significant increases in capital use, with markedly stronger responses among firms that were ex ante high-MRPK. This pattern is consistent with a reduction in capital misallocation, as previously constrained firms expand toward more efficient input combinations.

At the aggregate level, we find that the productivity gains associated with the NHDP are sizable and comparable in magnitude to the lower-bound effects of foreign capital liberalization documented in the literature, despite operating through distinct mechanisms. While financial liberalization relaxes borrowing constraints directly, improvements in connectivity reduce physical barriers to production and facilitate the reallocation of inputs across firms. This highlights the complementary role of infrastructure investment and financial access in improving resource allocation and boosting aggregate productivity.

Finally, we investigate the mechanisms underlying the reduction in misallocation by decomposing changes in wedges into product-market markups and pure input wedges. We find that the estimated decline in misallocation is driven primarily by changes in pure wedges, highlighting input reallocation—rather than increased product-market competition—as the main channel through which improved connectivity under the NHDP affects productivity

More broadly, these findings highlight reductions in spatial frictions as an important and complementary margin—alongside financial access—through which misallocation can be reduced and aggregate productivity increased.

References

Ackerberg, D.A., K. Caves, and G. Frazer, "Identification Properties of Recent Production Function Estimators," *Econometrica*, November 2015, *83*, 2411–2451.

Alcalá, Francisco and Antonio Ciccone, "Trade and productivity," *The Quarterly journal of economics*, 2004, *119* (2), 613–646.

Amiti, Mary, Oleg Itskhoki, and Jozef Konings, "International shocks, variable markups, and domestic prices.," *The Review of Economic Studies*, 2019, *86* (6), 2356–2402.

Arsanjani, Jamal Jokar, Alexander Zipf, Peter Mooney, and Marco Helbich, "OpenStreetMap in GIScience," *Lecture notes in geoinformation and cartography*, 2015, p. 324.

Asher, Sam and Paul Novosad, "Rural roads and Local Economic Development," *American Economic Review*, 2020, *110* (3), 797–823.

Asturias, Jose, Manuel García-Santana, and Roberto Ramos, "Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*, 2019, *17* (6), 1881–1940.

Atack, Jeremy and Robert A Margo, "The impact of access to rail transportation on agricultural improvement: The American Midwest as a test case, 1850–1860," *Journal of Transport and Land Use*, 2011, *4* (2), 5–18.

_ , Fred Bateman, Michael Haines, and Robert A. Margo, "Did Railroads Induce or Follow Economic Growth?: Urbanization and Population Growth in the American Midwest, 1850–1860," *Social Science History*, 2010, *34* (2), 171–197.

_ , Michael R Haines, and Robert A Margo, "Railroads and the Rise of the Factory: Evidence for the United States, 1850-70," Technical Report, National Bureau of Economic Research 2008.

Bai, John, Daniel Carvalho, and Gordon M Phillips, "The impact of bank credit on labor reallocation and aggregate industry productivity," *The Journal of Finance*, 2018, *73* (6), 2787–2836.

Baker, Andrew C., David F. Larcker, and Charles C.Y. Wang, "How Much Should We Trust Staggered Difference-In-Differences Estimates?," *Journal of Financial Economics*, 2022, *144* (2), 370–395.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "On the road: Access to transportation infrastructure and economic growth in China," *Journal of Development Economics*, 2020, *145*, 102442.

Banerjee, Abhijit V and Benjamin Moll, "Why does misallocation persist?," *American Economic Journal: Macroeconomics*, 2010, *2* (1), 189–206.

Baqaee, D. and E. Fahri, "A Short Note on Aggregating Productivity," Technical Report, National

Bureau of Economic Research 2019.

Bau, Natalie and Adrien Matray, "Misallocation and capital market integration: Evidence from India," *Econometrica*, 2023, *91* (1), 67–106.

Baum-Snow, Nathaniel, "Did Highways Cause Suburbanization?*," *The Quarterly Journal of Economics*, 05 2007, *122* (2), 775–805.

_ , J Vernon Henderson, Matthew A Turner, Qinghua Zhang, and Loren Brandt, *Highways, market access and urban growth in China*, SERC, Spatial Economics Research Centre, 2016.

_ , _ , _ , _ , and _ , "Does investment in national highways help or hurt hinterland city growth?," *Journal of Urban Economics*, 2020, *115*, 103124.

Bickenbach, F., E. Bode, M. Lange, and P. Nunnenkamp, "Night lights and regional GDP," *Weltwirtschaftliches Archiv*, 2013, *149* (4), 715–729.

Brooks, Wyatt J, Joseph P Kaboski, Illenin O Kondo, Yao Amber Li, and Wei Qian, "Infrastructure investment and labor monopsony power," *IMF Economic Review*, 2021, *69*, 470–504.

Chandra, Amitabh and Eric Thompson, "Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system," *Regional science and urban economics*, 2000, *30* (4), 457–490.

Chen, X. and W. D. Nordhaus, "Using luminosity data as a proxy for economic statistics," *Proceedings of the National Academy of Sciences*, 2011, *108* (21), 8589–8594.

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

de Chaisemartin, C. and X. d'Haultfoeuille, "Two-way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

Donaldson, D. and A. Storeygard, "The view from above: Applications of satellite data in economics," *Journal of Economic Perspectives*, 2016, *30* (4), 171–198.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, April 2018, *108* (4-5), 899–934.

_ and Richard Hornbeck, "Railroads and American economic growth: A "market access" approach," *The Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Duranton, Gilles and Matthew A. Turner, "Urban Growth and Transportation," *The Review of Economic Studies*, 03 2012, *79* (4), 1407–1440.

_ , Peter M. Morrow, and Matthew A. Turner, "Roads and Trade: Evidence from the US," *The Review of Economic Studies*, 04 2014, *81* (2), 681–724.

Elvidge, C. D., K. E. Baugh, E. A. Kihn, H. W. Kroehl, and E. R. Davis, "Mapping city lights with nighttime data from the DMSP operational linescan system," *Photogrammetric Engineering & Remote Sensing*, 1997, *63* (6), 727–734.

Epifani, Paolo and Gino Gancia, "Trade, markup heterogeneity and misallocations," *Journal of International Economics*, 2011, *83* (1), 1–13.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *The Review of Economic Studies*, 03 2014, *81* (3), 1046–1070.

Fretz, Stephan, Raphaël Parchet, and Frédéric Robert-Nicoud, "Highways, market access and spatial sorting," *The Economic Journal*, 2022, *132* (643), 1011–1036.

Ghani, Ejaz, Arti Grover Goswami, and William R Kerr, "Highway to success: The impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing," *The Economic Journal*, 2016, *126* (591), 317–357.

Goldberg, P. K., A. K. Khandelwal, N. Pavcnik, and P. Topalova, "Multiproduct Firms and Product Turnover in the Developing World: Evidence from India," *Review of Economics and Statistics*, 2010, *92(4)*, 1042–1049.

Goodman-Bacon, A., "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Haines, Michael R and Robert A Margo, "Railroads and local economic development: The United States in the 1850s," 2006.

Henderson, V., A. Storeygard, and D. N. Weil, "Measuring economic growth from outer space," *American Economic Review*, 2012, *102* (2), 994–1028.

Hsieh, C.-T. and P.J. Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 2009, *4* (124), 1403–1448.

Hsieh, Chang-Tai and Peter J Klenow, "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 2009, *124* (4), 1403–1448.

Huber, Stephan and Christoph Rust, "Calculate travel time and distance with OpenStreetMap data using the Open Source Routing Machine (OSRM)," *The Stata Journal*, 2016, *16* (2), 416–423.

Jean, N., M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, 2016, *353* (6301), 790–794.

Lall, S. V., J. V. Henderson, and A. J. Venables, *Africa's cities: Opening doors to the world*, World Bank Publications, 2017.

Levinsohn, J. and A. Petrin, "Estimating Production Functions Using Inputs to Control for Unobservables," *The Review of Economic Studies*, 2003, *70* (2), 317–341.

Loecker, J. De, P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik, "Prices, Markups, and Trade Reform," *Econometrica*, 2016, *84* (2), 445–510.

Lu, Will Jianyu, "Transport, Infrastructure and Growth: Evidence from Chinese Firms," *Available at SSRN 3289842*, 2018.

Melitz, Marc J, "The impact of trade on intra-industry reallocations and aggregate industry productivity," *econometrica*, 2003, *71* (6), 1695–1725.

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

Midrigan, Virgiliu and Daniel Yi Xu, "Finance and misallocation: Evidence from plant-level data," *American economic review*, 2014, *104* (2), 422–58.

Olley, G.S. and A. Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, November 1996, *64* (6), 1263–1297.

Petrin, A and J. Levinsohn, "Measuring Aggregate Productivity Growth using Plant-level Data," *The RAND Journal of Economics*, 2012, *4* (43), 705–725.

Pinkovskiy, M. L., "Economic discontinuities at borders: Evidence from satellite data on lights at night," Technical Report, Massachusetts Institute of Technology 2013.

Redding, Stephen and Anthony J Venables, "Economic geography and international inequality," *Journal of international Economics*, 2004, *62* (1), 53–82.

Restuccia, Diego and Richard Rogerson, "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic dynamics*, 2008, *11* (4), 707–720.

Sant'Anna, P. H. C. and J. Zhao, "Doubly Robust Difference-in-Differences Estimators," *Journal of Econometrics*, 2020, *219* (1), 101–122.

Tollefsen, D. F., H. Strand, and H. Buhaug, "PRIO-GRID: A unified spatial data structure," *Journal of Peace Research*, 2012, *49* (2), 363–374.

Topalova, P. and A. Khandelwal, "Trade Liberalization and Firm Productivity: The Case of India," *The Review of Economics and Statistics*, August 2011, *93* (3), 995–1009.

Tripathy, R. R., H. Sajjad, and C. D. Elvidge, "Nighttime light imagery as proxy measure of economic conditions," *Remote Sensing*, 2016, *8* (5), 334.

Online Appendix for:

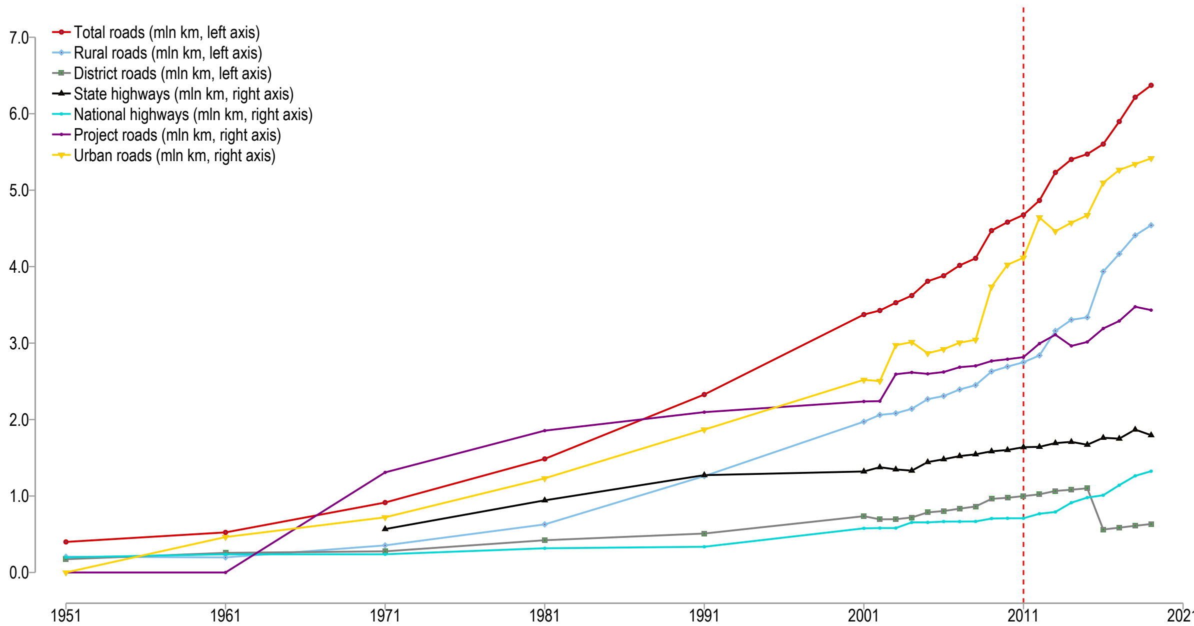
Roads and Capital Misallocation: Evidence from India's Infrastructure Boom

Anastasia Burya (University of Bern)
Martino Pelli (Asian Development Bank)
Avralt-Od Purevjav (The World Bank)
Jeanne Tschopp (University of Bern)

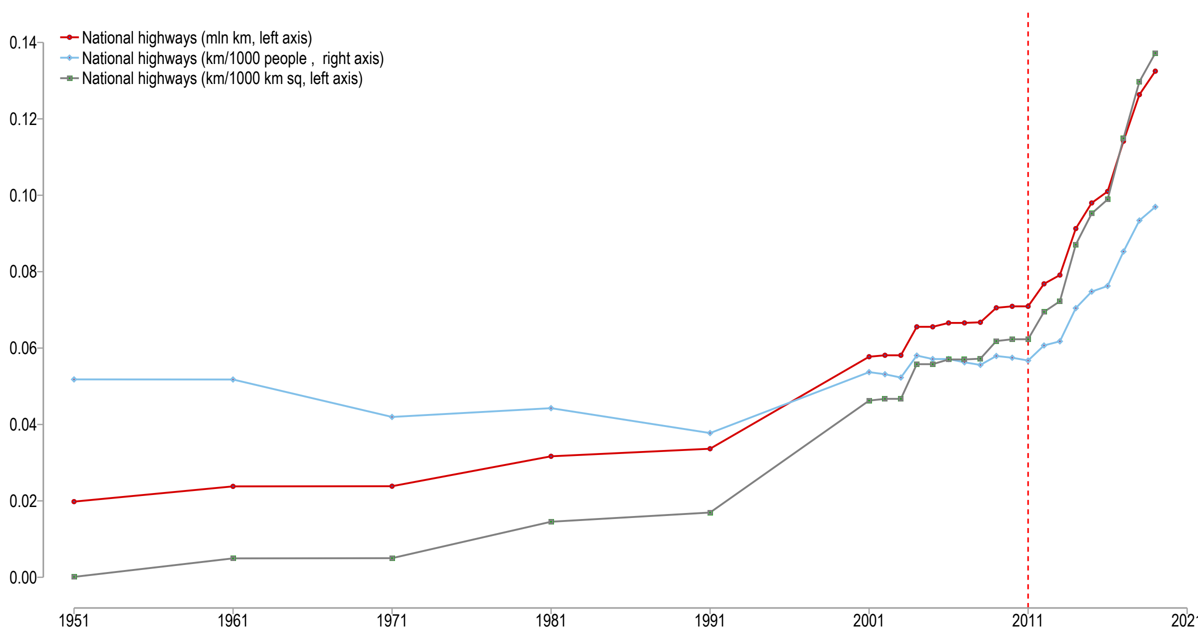
Appendix A Figures

Figure A.1: Road Network and Highway Density

(a) Road Network



(b) Highway Density



Source: Ministry of Road Transport and Highways (MoRTH), Government of India, 2021.

Figure A.2: Spatial Animation of Market Access, 2011-2019.

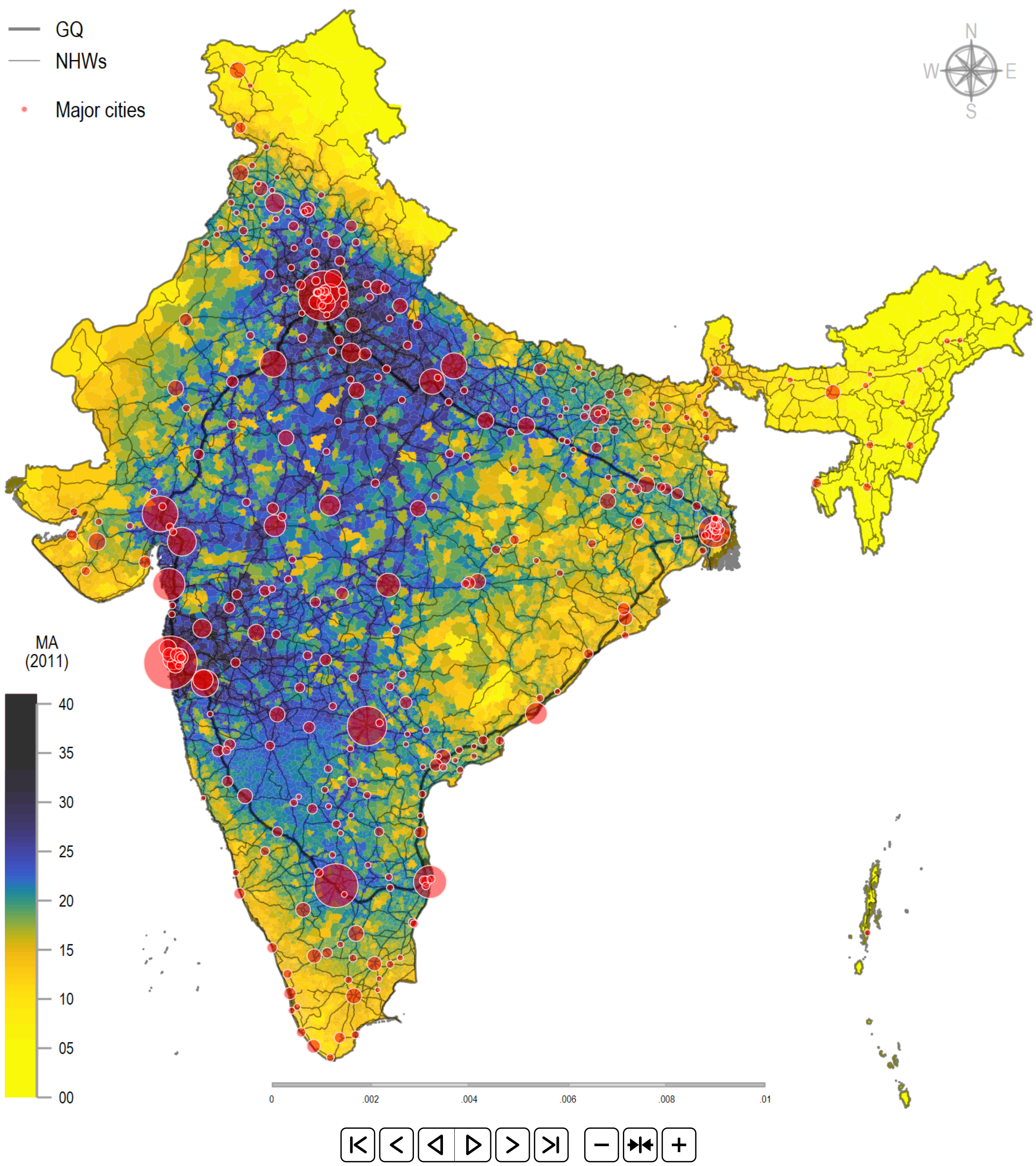
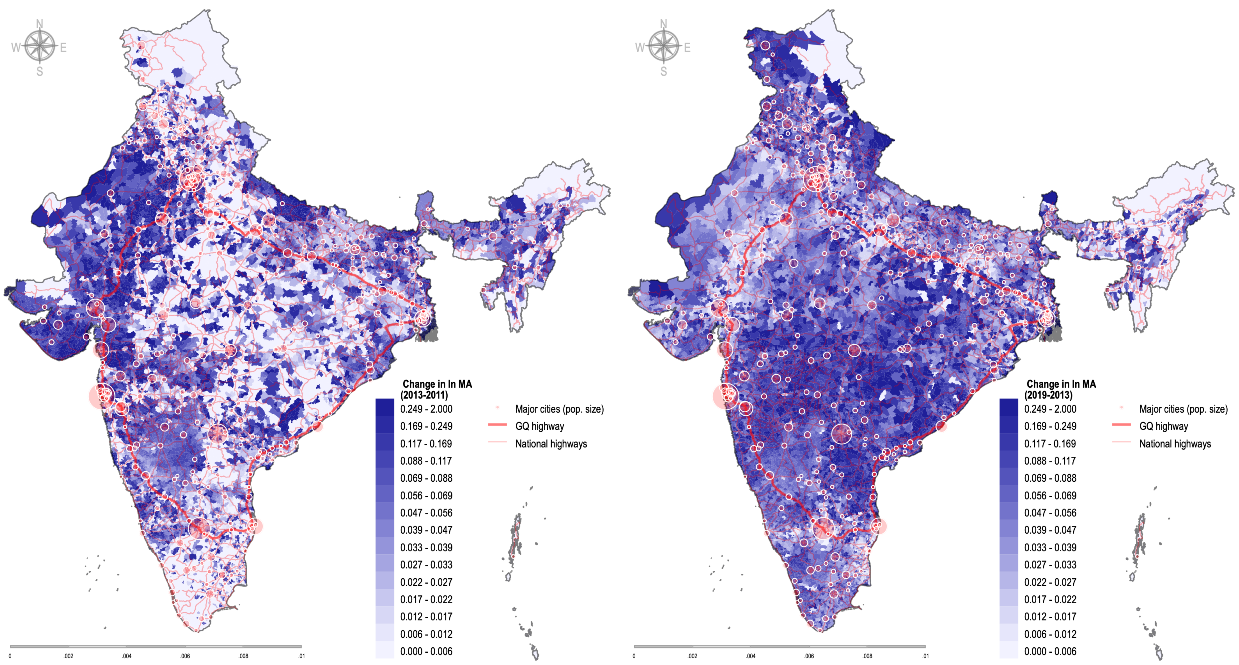


Figure A.3: Changes in (log) Market Access

(a) GQH phase: 2011-2013

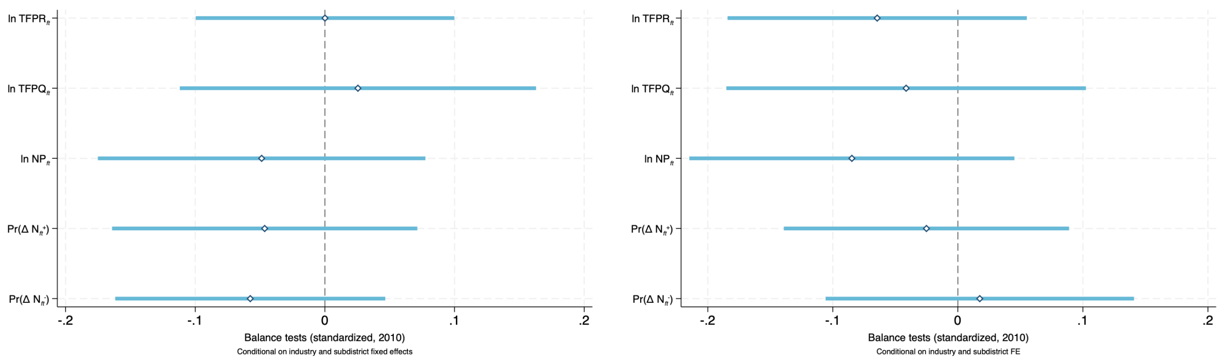
(b) Later phases of the NHDP: 2013-2019



Note: Figure A.3 illustrates changes in $\ln MA_{pt}$ for the periods 2011–2013 (Panel (a)) and 2013–2019 (Panel (b)). The blue gradient indicates changes in market access. The color scale is held constant across both panels, based on the full distribution of market access changes from the 2011–2013 period. It is divided into 15 equally sized groups (bins), each representing approximately 6.7% of the sample over that period. Lighter shades indicate smaller changes, while darker shades correspond to larger gains. The number of negative log changes in Panel (a) represents less than 0.01% of the observations and is not represented. There are no negative log changes in Panel (b). Red lines represent highways. Major cities (Tier-1 centers) are shown as red circles, proportional to their 2011 population size.

Figure A.4: Balance checks

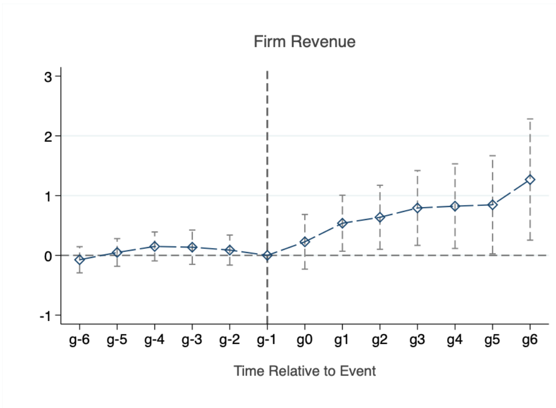
(a) DiD (b) SDiD



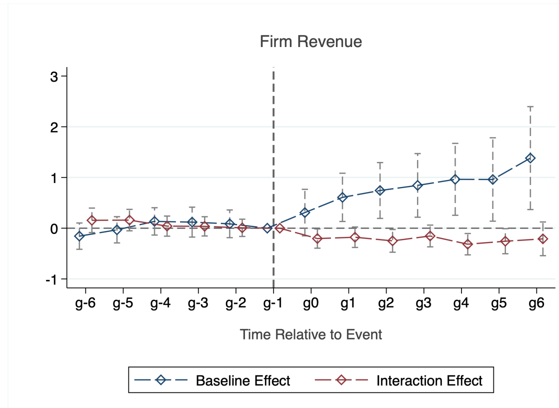
Note: The figure plots the coefficients (dark blue diamonds) and 95% confidence intervals (light blue bars) from regressions of firm-level outcomes measured in 2010 on a treatment indicator. $\ln TFPR_{ft}$ denotes log TFP revenue, $\ln TFPQ_{ft}$ denotes log TFP quantity, $\ln NP_{ft}$ is the log number of products, ΔN^{+}_{ft} captures the probability of product addition, and ΔN^{-}_{ft} captures the probability of product deletion. Regressions are conditional on 4-digit industry FE and subdistrict FE, with standard errors clustered at the postal-code level. All variables are normalized to have mean zero and standard deviation one. Panel (a) corresponds to the DiD design, where treatment is defined based on postal codes that become treated after 2013. Panel (b) corresponds to the SDiD design, where treatment equals one for ever-treated postal codes and zero for never-treated ones.

Figure A.5: Event Study Plots

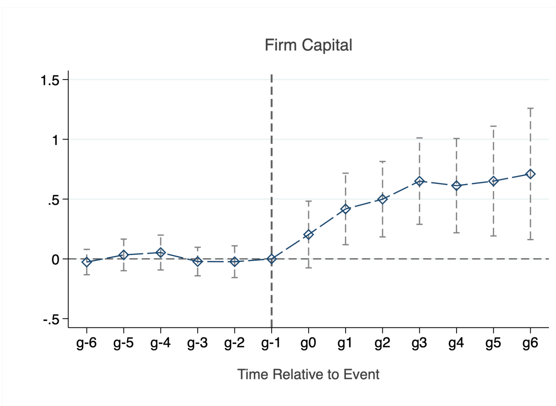
(a) No Interaction



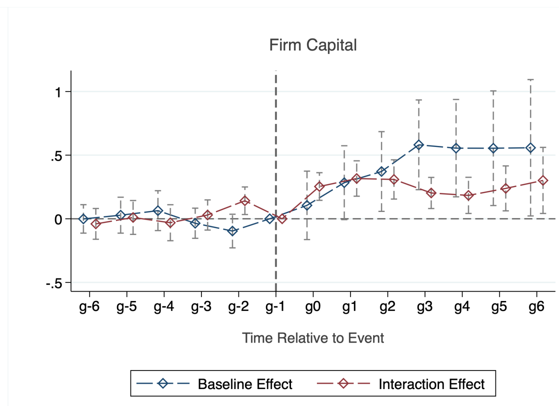
(b) With Interaction



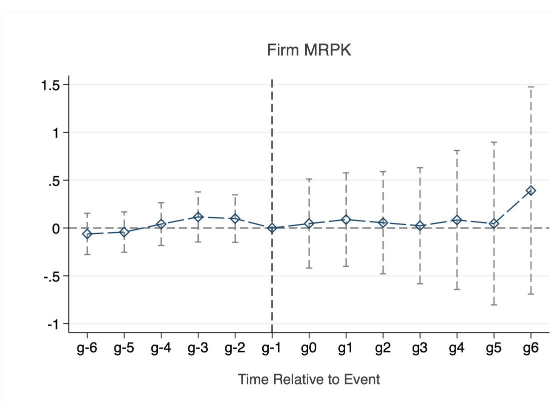
(c) No Interaction



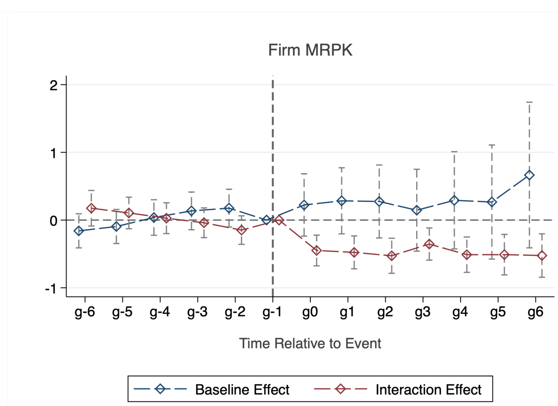
(d) With Interaction



(e) No Interaction



(f) With Interaction



Note: Each panel reports event study estimates from TWFE regressions similar to those in Table 2, examining the dynamic effects of market access on log firm revenues (Panels (a) and (b)), log firm capital (Panels (c) and (d)), and log firm MRPK (Panels (e) and (f)). Panels (a), (c), and (e) present estimates without the interaction term, showing the baseline treatment effect (blue line), while Panels (b), (d), and (f) include an interaction between treatment and high-MRPK firms (red line). The horizontal axis shows time relative to treatment, with $g0$ indicating the year of treatment. The vertical axis displays the estimated treatment effects by event time. Standard errors are clustered at the postal code-year level. 95% confidence intervals.

Appendix B Tables

Table A.1: ISIC Rev.4 1-digit Industries

Included Industries	
A	Agriculture, forestry, and fishing
B	Mining and quarrying
C	Manufacturing
D	Electricity, gas, steam, air conditioning supply
E	Water supply; sewerage and waste management
F	Construction
G	Wholesale and retail trade; repair of motor vehicles and motorcycles
Excluded Industries	
I	Accommodation and food service activities
J	Information and communication
K	Financial and insurance activities
L	Real estate activities
M	Professional, scientific, and technical activities
N	Administrative and support service activities
O	Public administration and defence; compulsory social security
P	Education
Q	Human health and social work activities
R	Arts, entertainment and recreation
S	Other service activities
T	Activities of households as employers; undifferentiated goods- and services producing activities of households for own use
U	Activities of extraterritorial organizations and bodies diversified

Note: This table lists the industries included in our analysis, primarily road-dependent sectors.

Table A.2: Summary Statistics

Variable	N	Mean	SD	Min	Median	Max
$\ln R_{ft}$	53983	2.11	2.35	-6.10	2.34	7.25
$\ln K_{ft}$	53983	1.27	2.15	-8.77	1.37	5.37
$\ln W_{ft}$	53983	-0.65	2.20	-6.59	-0.55	4.84
$\ln MRPK_{ft}$	53983	0.85	1.59	-5.40	0.85	5.95
$\ln MRPL_{ft}$	53983	2.76	1.31	-2.58	2.69	6.65
$\ln MA_{pt}$	41419	3.33	0.29	1.72	3.37	3.68

Note: The table shows summary statistics of the main variables, covering the period 2011-2019, as described in the main text.

Table A.3: Variance of Log Firm-level MRPK

	(1) DiD: GQH	(2) SDiD: NHDP
T_{pt}	-0.295**	-0.237
	(0.130)	(0.224)
Controls	Yes	Yes
N	2792	4113
R^2	0.87	0.74

Note: This table reports DiD (Column 1) and SDiD (Column 2) estimates for the variance of log MRPK across firms. In column (1), the variance is computed at the 4-digit industry $\times$ year $\times$ treatment-group level. The cell-level variance is regressed on the treatment indicator (T_{pt}), controlling for industry-year and industry-treatment-group FE. In column (2), the variance is computed at the industry-cohort-year level and regressed on the treatment indicator (T_{pt}), controlling for industry-year and industry-cohort FE. In both specifications, the variance is adjusted for small-sample bias; observations are weighted by the number of firms used to compute each cell-level variance, and standard errors are clustered at the industry level. Significance: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.4: Alternative Threshold: p40

	(1)	(2)	(3)	(4)	(5)	(6)
	$\ln R_{ft}$	$\ln K_{ft}$	$\ln MRPK_{ft}$	$\ln R_{ft}$	$\ln K_{ft}$	$\ln MRPK_{ft}$
I. DiD: GQH	*(a) Average effects*			*(b) Differential effects*		
T_{pt}	-0.172	0.614^{***}	-0.787^{***}	-0.041	0.460^{***}	-0.503^{**}
	(0.249)	(0.138)	(0.244)	(0.244)	(0.135)	(0.229)
$\times MRPK_{f0}^{high}$				-0.178^{***}	0.209^{***}	-0.384^{***}
				(0.036)	(0.025)	(0.041)
N	45310	45310	45310	45310	45310	45310
R^2	0.94	0.97	0.86	0.94	0.97	0.86
	(1)	(2)	(3)	(4)	(5)	(6)
II. SDiD: NHDP	*(a) Average effects*			*(b) Differential effects*		
T_{pt}	0.086	0.043	0.043	0.169	-0.071	0.235
	(0.180)	(0.148)	(0.260)	(0.174)	(0.149)	(0.248)
$\times MRPK_{f0}^{high}$				-0.202^{***}	0.275^{***}	-0.467^{***}
				(0.052)	(0.033)	(0.055)
N	43777	43777	43777	43777	43777	43777
R^2	0.94	0.97	0.86	0.94	0.97	0.87
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table presents the average effects (Columns 1-3) and the heterogeneous effects (Columns 4-6) of improved market access on log firm-level revenues (R_{ft}), capital (K_{ft}), and MRPK ($MRPK_{ft}$). Panel I shows DiD results and Panel II SDiD results. T_{pt} and $MRPK_{f0}^{high}$ indicate treatment and ex ante high MRPK, respectively. Relative to the baseline definition, treatment is defined using the 40th percentile of market access growth. Controls include firm FE, industry-year FE, subdistrict-year FE, firm age, initial firm size interacted with year FE and initial nightlights at the postal code level interacted with year FE. Standard errors clustered at the postal code-year level. Significance: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.5: Alternative Threshold: p60

	(1) $\ln R_{ft}$	(2) $\ln K_{ft}$	(3) $\ln MRPK_{ft}$	(4) $\ln R_{ft}$	(5) $\ln K_{ft}$	(6) $\ln MRPK_{ft}$
I. DiD: GQH	*(a) Average effects*			*(b) Differential effects*		
T_{pt}	-0.729***	-0.053	-0.675***	-0.611***	-0.164	-0.454**
	(0.225)	(0.139)	(0.235)	(0.220)	(0.139)	(0.226)
$\times MRPK^{high}_{f0}$				-0.238***	0.226***	-0.447***
				(0.058)	(0.034)	(0.064)
N	45310	45310	45310	45310	45310	45310
R^2	0.94	0.97	0.86	0.94	0.97	0.86
II. SDiD: NHDP	*(a) Average effects*			*(b) Differential effects*		
T_{pt}	-0.034	0.274**	-0.303	0.027	0.204	-0.180
	(0.180)	(0.134)	(0.193)	(0.182)	(0.133)	(0.192)
$\times MRPK^{high}_{f0}$				-0.130*	0.150**	-0.262***
				(0.078)	(0.059)	(0.086)
N	44643	44643	44643	44643	44643	44643
R^2	0.94	0.97	0.86	0.94	0.97	0.86
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table presents the average effects (Columns 1-3) and the heterogeneous effects (Columns 4-6) of improved market access on log firm-level revenues (R_{ft}), capital (K_{ft}), and MRPK ($MRPK_{ft}$). Panel I shows DiD results and Panel II SDiD results. T_{pt} and $MRPK^{high}_{f0}$ indicate treatment and ex ante high MRPK, respectively. Relative to the baseline definition, treatment is defined using the 60th percentile of market access growth. Controls include firm FE, industry-year FE, subdistrict-year FE, firm age, initial firm size interacted with year FE and initial nightlights at the postal code level interacted with year FE. Standard errors clustered at the postal code-year level. Significance: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.6: Misallocation of Capital: Continuous Treatment (2011-2019)

	(1) $\ln R_{ft}$	(2) $\ln K_{ft}$	(3) $\ln MRPK_{ft}$	(4) $\ln R_{ft}$	(5) $\ln K_{ft}$	(6) $\ln MRPK_{ft}$
	(a) Average effects			*(b) Differential effects*		
$\ln MA_{pt}$	-0.975	2.723^{***}	-3.475^{***}	-0.756	2.113^{***}	-2.655^{**}
	(0.911)	(0.813)	(1.137)	(0.949)	(0.814)	(1.184)
$\times MRPK_{f0}^{high}$				-0.358	0.997^{***}	-1.340^{**}
				(0.451)	(0.238)	(0.522)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	34300	34300	34300	34300	34300	34300
R^2	0.95	0.98	0.88	0.95	0.98	0.88

Note: This table presents the average effects (Columns 1-3) and the heterogeneous effects (Columns 4-6) of improved market access on log firm-level revenues (R_{ft}), capital (K_{ft}), and MRPK ($MRPK_{ft}$). Relative to the baseline definition, treatment is continuous and given by the log market access ($\ln MA_{pt}$). $MRPK_{f0}^{high}$ indicate ex ante high MRPK. Controls include firm FE, industry-year FE, subdistrict-year FE, firm age, initial firm size interacted with year FE and initial nightlights at the postal code level interacted with year FE. Standard errors clustered at the postal code-year level. Significance: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.7: Log Firm-Level Total Factor Productivity Quantity (TFPQ)

	(1)	(2)	(3)	(4)
	DiD: GQH		SDiD: NHDP	
T_{pt}	-1.028	-1.005	-1.208	-1.319
	(1.382)	(1.411)	(1.238)	(1.266)
$\times MRPK_{f0}^{high}$		-0.022		0.241
		(0.313)		(0.275)
$\times MRPL_{f0}^{high}$		-0.004		-0.011
		(0.311)		(0.295)
Controls	Yes	Yes	Yes	Yes
N	20606	20606	20426	20426
R^2	0.81	0.81	0.81	0.81

Note: This table reports the average and heterogeneous effects of improved market access on log firm-level TFPQ. Columns 1-2 show DiD results and Columns 3-4 SDiD results. T_{pt}, $MRPK_{f0}^{high}$ and $MRPL_{f0}^{high}$ indicate treatment, ex ante high MRPK and ex ante high MPRL, respectively. Controls include firm FE, industry-year FE, subdistrict-year FE, firm age, initial firm size interacted with year FE and initial nightlights at the postal code level interacted with year FE. Standard errors clustered at the postal code-year level. Significance: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.8: Heterogeneity by Both Firm Ex ante MRPK and Ex ante MRPL

	(1) $\ln R_{ft}$	(2) $\ln K_{ft}$	(3) $\ln MRPK_{ft}$	(4) $\ln W_{ft}$	(5) $\ln MRPL_{ft}$	(6) $\ln EI_{ft}$	(7) $\ln MI_{ft}$
I. DiD: GHQ	*Differential effects*						
$T_p \times P_\tau$	0.008	$0.463{*}{*}{*}$	$-0.464{*}$	$-0.530{*}{*}{*}$	$0.514{*}{*}$	$-0.375{*}$	0.141
	(0.253)	(0.132)	(0.242)	(0.134)	(0.209)	(0.196)	(0.341)
$\times MRPK_{f0}^{high}$	$-0.156{*}{*}{*}$	$0.145{*}{*}{*}$	$-0.296{*}{*}{*}$	0.007	$-0.154{*}{*}{*}$	$0.105{*}$	$-0.109{*}$
	(0.054)	(0.031)	(0.056)	(0.034)	(0.044)	(0.054)	(0.058)
$\times MRPL_{f0}^{high}$	$-0.214{*}{*}{*}$	$0.131{*}{*}{*}$	$-0.330{*}{*}{*}$	0.060	$-0.272{*}{*}{*}$	-0.072	-0.049
	(0.061)	(0.032)	(0.058)	(0.039)	(0.045)	(0.056)	(0.054)
N	45300	45300	45300	45300	45300	40530	33342
R^2	0.94	0.97	0.86	0.97	0.87	0.96	0.94
II. SDiD: NHDP	*Differential effects*						
T_p	$0.434{*}{*}$	$0.220{*}$	0.205	-0.058	$0.487{*}{*}$	-0.063	-0.212
	(0.211)	(0.124)	(0.212)	(0.109)	(0.194)	(0.224)	(0.256)
$\times MRPK_{f0}^{high}$	$-0.170{*}{*}{*}$	$0.202{*}{*}{*}$	$-0.349{*}{*}{*}$	-0.035	$-0.112{*}{*}{*}$	0.070	$-0.188{*}{*}$
	(0.057)	(0.035)	(0.059)	(0.038)	(0.041)	(0.051)	(0.057)
$\times MRPL_{f0}^{high}$	$-0.242{*}{*}{*}$	$0.121{*}{*}{*}$	$-0.331{*}{*}{*}$	$0.084{*}{*}$	$-0.316{*}{*}{*}$	-0.071	0.028
	(0.062)	(0.035)	(0.058)	(0.041)	(0.043)	(0.056)	(0.055)
N	44806	44806	44806	44806	44806	40115	32976
R^2	0.94	0.97	0.86	0.97	0.87	0.96	0.94
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table presents the heterogeneous effects of improved market access on log firm-level revenues (R_{ft}), capital (K_{ft}), MRPK ($MRPK_{ft}$), labor input (W_{ft}), MRPL ($MRPL_{ft}$), energy inputs (EI_{ft}) and material inputs (MI_{ft}). Panel I show DiD results and Panel II SDiD results. T_{pt}, $MRPK_{f0}^{high}$ and $MRPL_{f0}^{high}$ indicate treatment, ex ante high MRPK and ex ante high MPRL, respectively. Controls include firm FE, industry-year FE, subdistrict-year FE, firm age, initial firm size interacted with year FE and initial nightlights at the postal code level interacted with year FE. Standard errors clustered at the postal code-year level. Significance: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table A.9: Product Mix & Product-Level Analysis

	Firm				Product		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$\ln R_{ft}$	$\ln NP_{ft}$	$\Pr(\Delta N^{+}_{ft})$	$\Pr(\Delta N^{-}_{ft})$	$\ln R_{\iota ft}$	$\ln P_{\iota ft}$	$\ln Q_{\iota ft}$
I. DiD: GQH							
	(a) Average effects						
T_{pt}	-0.173	-0.023	-0.743	-0.268	0.356	0.275	0.080
	(0.249)	(0.185)	(0.557)	(0.378)	(0.380)	(0.493)	(0.619)
	(b) Differential effects by ex ante MRPK						
T_{pt}	0.007	-0.020	-0.888	-0.375	0.455	0.236	0.219
	(0.246)	(0.186)	(0.573)	(0.384)	(0.404)	(0.491)	(0.635)
$\times MRPK^{high}_{f0}$	-0.236***	-0.003	0.188	0.140	-0.338*	0.133	-0.470*
	(0.054)	(0.030)	(0.133)	(0.093)	(0.195)	(0.215)	(0.245)
N	45300	44115	44115	44115	44920	44920	44920
II. SDiD: NHDP							
	(a) Average effects						
T_{pt}	0.293	-0.012	0.838**	-0.797**	0.073	-0.450	0.523
	(0.199)	(0.210)	(0.427)	(0.325)	(0.310)	(0.634)	(0.554)
	(b) Differential effects by ex ante MRPK						
T_{pt}	0.378*	-0.028	0.738*	-0.879***	0.182	-0.564	0.746
	(0.201)	(0.209)	(0.437)	(0.322)	(0.324)	(0.639)	(0.572)
$\times MRPK^{high}_{f0}$	-0.249***	0.045	0.292**	0.241***	-0.338*	0.355*	-0.692***
	(0.057)	(0.029)	(0.147)	(0.090)	(0.192)	(0.186)	(0.237)
N	44806	43641	43641	43641	44265	44265	44265
Product × Firm FEs	No	No	No	No	Yes	Yes	Yes
Other FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the average effects (Panel (a)) and heterogeneous effects (Panel (b)) of improved market access on log firm-level revenues (R_{ft}), on the number of products (NP_{ft}), and the probability of product addition (ΔN^{+}_{ft}) and deletion (ΔN^{-}_{ft}). The table also reports effects on log firm-product-level revenues ($R_{f\iota t}$), prices ($P_{f\iota t}$), and quantities ($Q_{f\iota t}$). Panel I shows DiD results and Panel II SDiD results. T_{pt} and $MRPK^{high}_{f0}$ indicate treatment and ex ante high MRPK, respectively. Controls include industry-year FE, subdistrict-year FE, firm age, initial firm size interacted with year FE and initial nightlights at the postal code level interacted with year FE. Additionally, firm-level regressions include firm FE, while product-level regressions add firm-product FE. Standard errors clustered at the postal code-year level. Significance: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Appendix C Derivation of the Aggregation Formula

Below, we derive the aggregation formula used to quantify how reducing misallocation affects aggregate productivity among treated firms. We measure this effect through changes in the Solow residual, ΔSolow_R, driven by variations in treated firms (i.e., firms in treated postal codes). Our derivation follows Bau and Matray (2023), except that we examine misallocation reduction at the postal code level rather than by industry, as can be seen from Table A.10.

Table A.10: Treated and Untreated Regions, Industries, and Firms

	Untreated Region $\backslash R$ **(all untreated postal codes)**	**Treated Region** R **(all treated postal codes)**
Industry 1	$\dots$, firm $f \in (1, \backslash R), \dots f' \in (1, \backslash R), \dots$	firm $f \in (1, R), \dots, f' \in (1, R), \dots$
$\vdots$	$\vdots$	$\vdots$
Industry i	$\dots$, firm $f \in (i, \backslash R) \longleftarrow f' \in (i, \backslash R)$	firm $f \in (i, R), \dots, f' \in (i, R), \dots$
$\vdots$	$\vdots$	$\vdots$
Industry I	$\dots$, firm $f \in (I, \backslash R), \dots f' \in (I, \backslash R), \dots$	firm $f \in (i = I, R), \dots, f' \in (I, R)$

Note: The table illustrates treated and untreated regions, industries, and firms. A firm can sell (buy) its output (inputs) within the same industry in the untreated region ($\backslash R$), across industries in the untreated region ($\backslash R$), or to (from) firms in any industry in the treated region (R).

Specifically, the change in the Solow residual is given by:

$$\Delta\text{Solow}_R = \Delta\text{Net Output}_R - \Delta\text{Net Input}_R, \tag{A.1}$$

where R denotes a treated region, encompassing all treated postal codes. $\Delta\text{Net Output}_R$ captures the net output change of treated firms, net of the portion reused as inputs within these firms. $\Delta\text{Net Input}_R$ represents the change in net input used by treated firms, excluding the inputs they produce.

For clarity, but without loss of generality, we derive the Solow residual considering only output wedges. In particular, we follow Bau and Matray (2023) which treat the input of each firm as being the output of a fictitious middleman intermediary who buys the input and sells it to the firm with some output wedge. As discussed by the authors, this can be done because input wedges are a special case of output wedges; hence, this approach is isomorphic to having an input wedge on the original firm.

Change in treated firms' net output (ΔNet Output$_R$ in equation (A.1)) The change in log total nominal net output in the treated region, valued at fixed prices is:

$$
\begin{aligned}
\overline{C}_R &= \frac{\Delta PC_R}{PC_R} \\
&= \frac{1}{PC_R}\sum_i \sum_{f\in(i,R)} p_f \Delta c_f \\
&= \frac{1}{PC_R}\sum_i \sum_{f\in(i,R)} p_f c_f \overline{c}_f \\
&= \sum_i \sum_{f\in(i,R)} \frac{p_f c_f}{PC_R}\overline{c}_f,
\end{aligned}
\tag{A.2}
$$

where overlined variables $\overline{x}$ denote log-changes, PC_R represents total nominal net output in the treated region. For firms in these treated regions, net output c_f is given by:

$$
c_f = y_f - \sum_j \sum_{\tilde{f}\in(j,R)} y_{\tilde{f},f}, \tag{A.3}
$$

where y_f denotes firm f's output, and $y_{\tilde{f},f}$ represents the inputs purchased by firm $\tilde{f}$ from firm f (i.e., the portion of y_f used by firm $\tilde{f}$).[24]

Using equation (A.3), we can derive an expression for $\overline{c}_f$ (again, for firms in treated regions):

$$
\begin{aligned}
\overline{c}_f &= \frac{\Delta c_f}{c_f} \\
&= \frac{1}{c_f}\left(\Delta y_f - \sum_j \sum_{\tilde{f}\in(j,R)} \Delta y_{\tilde{f},f}\right) \\
&= \frac{y_f}{c_f}\overline{y}_f - \sum_j \sum_{\tilde{f}\in(j,R)} \frac{y_{\tilde{f},f}}{c_f}\overline{y}_{\tilde{f},f}
\end{aligned}
\tag{A.4}
$$

Finally, substituting equation (A.4) into (A.2), we obtain the following expression:

$$
\overline{C}_R = \sum_i \sum_{f\in(i,R)} \lambda_f \overline{y}_f - \sum_i \sum_{f\in(i,R)} \sum_j \sum_{\tilde{f}\in(j,R)} \frac{p_f y_{\tilde{f},f}}{PC_R}\overline{y}_{\tilde{f},f}, \tag{A.5}
$$

[24] Using equation (A.3), the total nominal net output in the treated region is:

$$
PC_R = \sum_i \sum_{f\in(i,R)} p_f c_f = \sum_i \sum_{f\in(i,R)} p_f \left(y_f - \sum_j \sum_{\tilde{f}\in(j,R)} y_{\tilde{f},f}\right).
$$

where $\lambda_f \equiv \frac{p_f y_f}{PC_R}$ represents firm f's output share in treated firms nominal net output.

Change in treated firms' net input (ΔNet Input$_R$ in (A.1)) Meanwhile, the change in the log treated firms' net input, valued at fixed prices, is given by:

$$
\begin{aligned}
\sum_i \sum_{f\in(i,R)} \sum_j \sum_{\tilde{f}\in(j,\setminus R)} \Delta\left(p_{\tilde{f}} y_{f,\tilde{f}}\right)\frac{1}{PC_R} &= \sum_i \sum_{f\in(i,R)} \sum_j \sum_{\tilde{f}\in(j,\setminus R)} \frac{p_{\tilde{f}}}{p_f y_f}\frac{p_f y_f}{PC_R} y_{f,\tilde{f}} \overline{y}_{f,\tilde{f}} \\
&= \sum_i \sum_{f\in(i,R)} \lambda_f \sum_j \sum_{\tilde{f}\in(j,\setminus R)} \frac{p_{\tilde{f}} y_{f,\tilde{f}}}{p_f y_f}\overline{y}_{f,\tilde{f}} \qquad \text{(A.6)}
\end{aligned}
$$

In what follows, we rewrite equation (A.6) to explicitly depend on firms' markups, μ_f, and productivity, $TFPQ_f$. To do so, assume the production function of firm f takes the following form:

$$
y_f = TFPQ_f g_f\left(\{y_{f,\tilde{f}}\}_{\tilde{f}}\right),
$$

such that:

$$
\overline{y}_f = \overline{TFPQ}_f + \sum_j \sum_{\tilde{f}} \frac{\partial \ln g_f}{\partial \ln y_{f,\tilde{f}}}\overline{y}_{f,\tilde{f}} \qquad \text{(A.7)}
$$

Morevoer, the FOCs of a cost-minimizing firm are:

$$
p_{\tilde{f}} = mc_f TFPQ_f \frac{\partial g_f}{\partial y_{f,\tilde{f}}}, \qquad \text{(A.8)}
$$

where mc_f is the Lagrange multiplier (the marginal cost) and, again, $p_{\tilde{f}}$ the price of the input bought from firm $\tilde{f}$. If the firm has pricing power, the markup is $\mu_f = \frac{p_f}{mc_f}$ and equation (A.8) can be rearranged as follows:

$$
\begin{aligned}
\frac{p_{\tilde{f}} y_{f,\tilde{f}}}{p_f y_f} &= \frac{1}{\mu_f}\varepsilon_{g_f, y_{f,\tilde{f}}} \\
&= \frac{1}{\mu_f}\frac{\partial \ln g_f}{\partial \ln y_{f,\tilde{f}}} \qquad \text{(A.9)}
\end{aligned}
$$

Substituting equation (A.9) into (A.7), we obtain:

$$
\overline{y}_f = \overline{TFPQ}_f + \mu_f \sum_j \sum_{\tilde{f}} \frac{p_{\tilde{f}} y_{f,\tilde{f}}}{p_f y_f}\overline{y}_{f,\tilde{f}} \qquad \text{(A.10)}
$$

Note that the second term of the equation can be decomposed into a summation over treated firms and another over untreated firms. Rearranging terms, we can rewrite equation (A.10) as follows:

$$\sum_{j}\sum_{\tilde{f}\in(j,\backslash R)}\frac{p_{\tilde{f}}y_{f,\tilde{f}}}{p_f y_f}\overline{y}_{f,\tilde{f}}=\frac{1}{\mu_f}\left(\overline{y}_f-\overline{TFPQ}_f\right)-\sum_{j}\sum_{\tilde{f}\in(j,R)}\frac{p_{\tilde{f}}y_{f,\tilde{f}}}{p_f y_f}\overline{y}_{f,\tilde{f}} \tag{A.11}$$

Finally, substituting equation (A.11) into (A.6), we obtain an expression for the change in the log treated firms' net input, valued at fixed prices:

$$\begin{aligned}\sum_{i}\sum_{f\in(i,R)}\sum_{j}\sum_{\tilde{f}\in(j,\backslash R)}\Delta\left(p_{\tilde{f}}y_{f,\tilde{f}}\right)\frac{1}{PC_R} &= \sum_{i}\sum_{f\in(i,R)}\lambda_f\left[\frac{1}{\mu_f}\left(\overline{y}_f-\overline{TFPQ}_f\right)-\sum_{j}\sum_{\tilde{f}\in(j,R)}\frac{p_{\tilde{f}}y_{f,\tilde{f}}}{p_f y_f}\overline{y}_{f,\tilde{f}}\right]\\ &= \sum_{i}\sum_{f\in(i,R)}\frac{\lambda_f}{\mu_f}\left(\overline{y}_f-\overline{TFPQ}_f\right)-\sum_{i}\sum_{f\in(i,R)}\sum_{j}\sum_{\tilde{f}\in(j,R)}\frac{p_{\tilde{f}}y_{f,\tilde{f}}}{PC_R}\overline{y}_{f,\tilde{f}}\end{aligned} \tag{A.12}$$

Solow residual Combining equations (A.5) and (A.12), we obtain the following expression for the Solow residual:

$$\begin{aligned}\overline{\text{Solow}}_R &= \sum_{i}\sum_{f\in(i,R)}\lambda_f\overline{y}_f-\sum_{i}\sum_{f\in(i,R)}\frac{\lambda_f}{\mu_f}\left(\overline{y}_f-\overline{TFPQ}_f\right)\\ &= \sum_{i}\sum_{f\in(i,R)}\lambda_f\overline{TFPQ}_f+\sum_{i}\sum_{f\in(i,R)}\lambda_f\left(1-\frac{1}{\mu_f}\right)\left(\overline{y}_f-\overline{TFPQ}_f\right)\end{aligned} \tag{A.13}$$

where the second terms in equations (A.5) and (A.12) cancel out.

Finally, substituting (A.9) and (A.10) into the second component of equation (A.13), and defining $\alpha_{f,\tilde{f}}$ as the output elasticity with respect to the input purchased from $\tilde{f}$, equation (A.13) becomes:

$$\overline{\text{Solow}}_R = \sum_{i}\sum_{f\in(i,R)}\lambda_f\overline{TFPQ}_f+\sum_{i}\sum_{f\in(i,R)}\sum_{j}\sum_{\tilde{f}}\lambda_f\alpha_{f,\tilde{f}}\left(1-\frac{1}{\mu_f}\right)\overline{y}_{f,\tilde{f}} \tag{A.14}$$

Rewriting the Solow residual as a function of the combined input wedges As in Bau and Matray (2023), we rewrite equation (A.14) using input wedges instead of output wedges. In the presence of input wedges, the FOCs for the cost-minimizing firm become:

$$
\begin{aligned}
p_{\tilde{f}} &= \frac{p_f}{\mu_f(1+\tilde{\tau}_{f,\tilde{f}})} TFPQ_f \frac{\partial g_f}{\partial y_{f,\tilde{f}}} \\
&= \frac{p_f}{(1+\tau_{f,\tilde{f}})} TFPQ_f \frac{\partial g_f}{\partial y_{f,\tilde{f}}},
\end{aligned} \tag{A.15}
$$

where $(1+\tilde{\tau}_{f,\tilde{f}})$ represents the (pure) input wedge that firm f faces for the good purchased from $\tilde{f}$, and $(1+\tau_{f,\tilde{f}}) = \mu_f(1+\tilde{\tau}_{f,\tilde{f}})$ is the combined input wedge. Thus, incorporating input wedges implies dividing the marginal product by the combined wedge instead of the output wedge (the markup).

Consequently, accounting for input wedges, the Solow residual becomes:

$$
\begin{aligned}
\overline{\text{Solow}}_R &= \sum_i \sum_{f\in(i,R)} \lambda_f \overline{TFPQ}_f + \sum_i \sum_{f\in(i,R)} \sum_j \sum_{\tilde{f}} \lambda_f \alpha_{f,\tilde{f}} \left(1 - \frac{1}{1+\tau_{f,\tilde{f}}}\right) \overline{y}_{f,\tilde{f}}, \\
&= \sum_i \sum_{f\in(i,R)} \lambda_f \overline{TFPQ}_f + \sum_i \sum_{f\in(i,R)} \sum_j \sum_{\tilde{f}} \lambda_f \alpha_{f,\tilde{f}} \left(\frac{\tau_{f,\tilde{f}}}{1+\tau_{f,\tilde{f}}}\right) \overline{y}_{f,\tilde{f}},
\end{aligned} \tag{A.16}
$$

where μ_f in equation (A.14) has been replaced by the combined wedge.

Solow residual as a function of capital, labor and materials Having rewritten the Solow residual in terms of input wedges, we now express equation (A.16) as a function of firm-level capital (k), labor (l), and materials (m) wedges, treating each input-firm combination as a producer. This yields:

$$
\overline{\text{Solow}}_R \approx \sum_i \sum_{f\in(i,R)} \lambda_f \overline{TFPQ}_f + \sum_i \sum_{f\in(i,R)} \sum_{s\in\{K,L,M\}} \lambda_f \alpha_{is} \left(\frac{\tau_{fs}}{1+\tau_{fs}}\right) \overline{y}_{fs}, \tag{A.17}
$$

where we have assumed output elasticity to be industry-specific, and where $y_{f,s}$ represents firm f's input s. This aggregation formula aligns with those in Bau and Matray (2023), Baqaee and Fahri (2019), and Petrin and Levinsohn (2012).

Appendix D TFPQ Estimation

To measure firm productivity—a key input for estimating changes in the Solow residual—we use quantity-based total factor productivity (TFPQ). While revenue-based total factor productivity (TFPR) is commonly used in productivity studies, TFPQ is preferable for assessing production efficiency because it is unaffected by firms' pricing strategies, thereby isolating productivity from market power effects. Hence, a higher TFPQ indicates a firm's ability to produce more output with the same level of inputs. In our context, where market access is likely to affect firm prices, TFPQ provides a more reliable measure of productivity effects.

We compute log firm-level TFPQ as follows:

$$\ln TFPQ_{ft} = \ln TFPR_{ft} - \ln \bar{\bar{p}}_{ft}, \tag{A.18}$$

where $\bar{\bar{p}}_{ft}$ denotes the sales-share-weighted average price of a firm's products.[25]

TFPR estimates are obtained from the residuals of a production function regression in which firm-level output is regressed on inputs such as labor, capital, and intermediate inputs. A key challenge in this setting is the presence of unobserved productivity shocks that may be correlated with input choices. To address this concern, the literature adopts semi-parametric control function approaches that use input demand functions as proxies for unobserved productivity (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015).

We estimate $TFPR_{ft}$ following Levinsohn and Petrin (2003), a method widely used in related studies (e.g., Goldberg et al., 2010; Topalova and Khandelwal, 2011; Bau and Matray, 2023):

$$R_{ft} = \gamma_0 + \gamma_1 W_{ft} + \gamma_2 K_{ft} + \mathbf{IN}\boldsymbol{\gamma} + \omega_{ft} + \epsilon_{ft}, \tag{A.19}$$

where W_{ft} denotes labor costs, R_{ft} revenues, and K_{ft} gross fixed assets. $\mathbf{IN}$ is a vector of intermediary inputs, including expenditures on power, fuel, and raw materials. All variables, including the unobserved productivity term ω_{ft}, are in logs and deflated using the INDIA KLEMS two-digit industry-level intermediate input price index (ISIC Rev. 4).

TFPR estimates are then derived by subtracting the firm's predicted output from its actual output at time t. For estimation, we rely on the elements of $\mathbf{IN}$ as proxies for ω_{ft}, excluding industries with fewer than 30 firms. Productivity is estimated by either pooling all data or industry-by-industry, based on one-digit ISIC rev.4 classifications.

[25] In Prowess dx, firm-product prices are missing for a substantial number of observations. To address this, when computing $\bar{\bar{p}}_{ft}$, we impute missing prices using the median subdistrict-product-unit-year price. If this median price is unavailable, we progressively aggregate to higher geographical levels in the following order: district-product-unit-year, state-product-unit-year, and national-product-unit-year. If prices remain missing, we restart the procedure at the subdistrict-product-year level (omitting the unit dimension) and follow the same sequence of geographical aggregation.

Appendix E Derivation of Markup Changes

In order to identify markup changes, we start by considering demand for final goods of the firm f: $Y_f = D(\{p_f\}; \{\varepsilon_f\})$, with $\{p_f\}$ is a vector of prices of all the firms in the market, and $\{\varepsilon_f\}$ is a vector of demand shifters. We impose two assumptions: (1) demand D is homogeneous of degree 0 in terms of prices, so that it is not affected if all prices change proportionally; (2) demand D is homogeneous of degree 1 in terms of demand shifters, so that if all firms receive the same demand shifter, their demand increases proportionally. In this case, we can note, by Euler homogeneous function theorem, that $\sum_k D''_{p_f p_k} p_k = 0$ and $\sum_k D''_{p_f \varepsilon_k} \varepsilon_k = D'_{p_f}$. Let prices be set by the Lerner formula: $p_f = \frac{\xi_f}{\xi_f - 1} mc_f$, with demand elasticity $\xi_f = -\frac{D'_{p_f} p_f}{D}$. Denote additionally elasticity of demand elasticity of the firm f to the price of the firm k: $\eta_{fk} = (\xi_f)'_{p_k} \frac{p_k}{\xi_f}$; and to the demand shifter of the firm k: $\psi_{fk} = (\xi_f)'_{\varepsilon_k} \frac{\varepsilon_k}{\xi_f}$. Using this definition and properties of the demand function defined above, note: $\sum_k \eta_{fk} = 0$ and $\sum_k \psi_{fk} = 0$. We consider the log-differences of the prices (henceforth overlined variables $\overline{x}$ represent log-changes) and obtain a formula similar to Amiti et al. (2019):

$$\overline{p}_f = \frac{\xi_f - 1}{\xi_f - 1 + \eta_{ff}} \overline{mc}_f + \sum_{k \neq f} \frac{-\eta_{fk}}{\xi_f - 1 + \eta_{ff}} \overline{p}_k + \sum_k \frac{\psi_{fk}}{\xi_f - 1 + \eta_{ff}} \overline{\varepsilon}_k$$

Keep following Amiti et al. (2019) and impose an additional assumption that only the following aggregate of the competitor prices matters for the price changes: $\overline{p}_{-f} = \frac{\sum_{k \neq f} \delta_k \overline{p}_k}{1 - \delta_f}$, with δ_f revenue market share. Denote pass-through of marginal costs into prices as $\tilde{\rho}_f = \frac{\xi_f - 1}{\xi_f - 1 + \eta_{ff}}$, and effects of the demand shocks as $\tilde{\gamma}_{fk} = \frac{\psi_{fk}}{\xi_f - 1 + \eta_{ff}}$. In this case, the log-changes of prices can be simplified to:

$$\overline{p}_f = \tilde{\rho}_f \, \overline{mc}_f + (1 - \tilde{\rho}_f) \overline{p}_{-f} + \sum_k \tilde{\gamma}_{fk} \overline{\varepsilon}_k$$

For convenience, define the aggregate industry price as $\overline{P}_i = \sum_f \delta_f \overline{p}_f$, with i coding the industry, where the firm f operates. For this case, the log-change in price can be rewritten to become:

$$\overline{p}_f = \rho_f \, \overline{mc}_f + (1 - \rho_f) \overline{P}_i + \sum_k \gamma_{fk} \overline{\varepsilon}_k$$

$$\text{where } \overline{P}_i = \sum_f \delta_f \overline{p_f}, \quad \rho_f = \frac{\tilde{\rho}_f (1 - \delta_f)}{1 - \tilde{\rho}_f \delta_f}, \quad \text{and } \gamma_{fk} = \frac{\tilde{\gamma}_{fk} (1 - \delta_f)}{1 - \tilde{\rho}_f \delta_f}$$

Note that coefficients in front of the marginal costs and competitor price aggregate still sum to one, and $\sum_k \gamma_{fk} = 0$. This means that a cost-push shock common across firms would increase all the prices proportionally; and common across firms demand shifter would have no direct effect on prices.

Using the definition of the markup: $\overline{p}_f = \overline{\mu}_f + \overline{mc}_f$ and the equation above, we can derive:

$$\overline{\mu}_f = (1-\rho_f)\left(\frac{\sum_f \delta_f \rho_f \, \overline{mc}_f}{\sum_f \delta_f \rho_f} - \overline{mc}_f\right) + (1-\rho_f)\frac{\sum_f \delta_f \sum_k \gamma_{fk}\overline{\varepsilon}_k}{\sum_f \delta_f \rho_f} + \sum_k \gamma_{fk}\overline{\varepsilon}_k$$

Note that markup for the individual firm changes due to deviation of the marginal costs from the weighted average, so that cost shocks that affect all the firms in the industry similarly will have no effect on it. For instance, in our setting, prices of inputs are common within industries, so their changes will not affect markups, while changes in input wedges will be consequential, as they are firm-specific.

The next step is to derive the changes in marginal costs implied by the change in input wedges and demand. Denote each of the inputs used for production by the firm f as y_{fs}, where s indexes capital, labor and materials. From the production function defined in equation (6) and standard cost minimization problem obtain marginal costs:

$$mc_f = Y_f^{\frac{1}{\alpha_i}-1} TFPQ_f^{\frac{-1}{\alpha_i}} \prod_s [(1+\tilde{\tau}^s_{ft})p_{is}]^{\frac{\alpha_{is}}{\alpha_i}} \alpha_{is}^{\frac{-\alpha_{is}}{\alpha_i}}$$

$$\text{where } Y_f = TFPQ_f \prod_s y_{fs}^{\alpha_{is}} \text{ and } \alpha_i = \sum_s \alpha_{is}$$

Using the equation above, and taking into account that common industry-wide shocks do not affect markups, and using additionally, the empirical result that there was no $TFPQ$ changes due to policy, so that $\overline{TFPQ}_{ft} = 0$, we can arrive at the equation connecting wedges and markups:

$$\begin{aligned}
\overline{\mu}_f =&(1-\rho_f)\left(\frac{1}{\alpha_i}-1\right)\left(\frac{\sum_f \delta_f \rho_f \overline{y}_f}{\sum_f \lambda_f \rho_f} - \overline{y}_f\right) + \\
&+(1-\rho_f)\sum_s \left(\frac{\alpha_{is}}{\alpha_i}\right)\left(\frac{\sum_f \delta_f \rho_f \overline{(1+\tilde{\tau}_{fs})}}{\sum_f \delta_f \rho_f} - \overline{(1+\tilde{\tau}_{fs})}\right) + \\
&+(1-\rho_f)\frac{\sum_f \delta_f \sum_k \gamma_{fk}\overline{\varepsilon}_k}{\sum_f \delta_f \rho_f} + \sum_k \gamma_{fk}\overline{\varepsilon}_k
\end{aligned} \tag{A.20}$$

Lastly, we adjust the equation above to implement it in the data. We note that we only observe combined wedges and substitute in the definition $(1+\tau_{fst}) = \mu_{ft}(1+\tilde{\tau}_{fst})$.

The log-changes of the variables are going to come from the SDiD regressions, and, as such, they represent the changes between treatment and control group. We assume that firms in the treatment group receive the common demand shock ε^R, and that treatment group is representative within the industry. Under these assumptions, we require $\sum_{k\in(i,R)} \gamma_{fk} = \sum_{k\in(i,\setminus R)} \gamma_{fk}$, and recall the property $\sum_k \gamma_{fk} = 0$ to obtain: $\sum_{k\in(i,R)} \gamma_{fk}\varepsilon^R = 0$. Therefore, in our setting, direct demand shifter effect is zero.

For the estimation we use the pass-through parameter $\tilde{\rho} = 0.6$ as a mid-point of the estimates from Amiti et al. (2019). Finally, we implement the following equation in the data:

$$\overline{\mu}_f = \frac{1-\rho_f}{\rho_f}\left[\left(\frac{1}{\alpha_i}-1\right)\sum_s \alpha_{is}\left(\sum_f \delta_f \overline{y}_{fs} - \overline{y}_{fs}\right) + \sum_s \left(\frac{\alpha_{is}}{\alpha_i}\right)\left(\sum_f \delta_f \overline{(1+\tau_{fs})} - \overline{(1+\tau_{fs})}\right)\right].$$

Appendix F Product-Market Changes

This section examines how improved market access affects firms' product mix, focusing on revenues, prices, and quantities at the product level.[26] Using product-level data from Prowess dx, we estimate regressions analogous to the firm-level specifications, replacing firm FE with firm–product FE. This identification strategy exploits within firm–product variation, particularly in prices and quantities, while accounting for product entry and exit.[27]

Results are reported in Table A.9. Column (1) reproduces the firm-level baseline revenue estimates for comparison. Columns (2)–(4) report firm-level regressions for the log number of products, and the probabilities of product addition and deletion, estimated using a linear probability model. Columns (5)–(7) present firm–product regressions with log revenues, prices, and quantities as outcomes.

Panel I reports DiD estimates. Panel I(a) shows no statistically significant average effects across outcomes. Panel I(b), however, reveals heterogeneity by initial MRPK. While product scope remains largely unchanged for ex ante high-MRPK firms, these firms experience a decline in per-product quantities, leading to lower product-level and firm-level revenues. This suggests that the revenue decline observed in column (1) is driven by contractions at the intensive margin rather than by changes in product scope.

Panel II presents SDiD estimates and points to more pronounced adjustments along the product dimension. As shown in Panel II(a), columns (3) and (4), firms are 84% more likely to introduce new products and 80% less likely to discontinue existing ones following NHDP-induced improvements in MA. These patterns are mirrored in Panel II(b) and are consistent with the higher firm-level revenues observed in column (1). Ex ante high-MRPK firms exhibit a stronger tendency to add products and a smaller reduction in product deletions relative to other firms. In addition, these firms experience a modest decline in product prices—statistically significant at the 10% level—suggesting lower markups, alongside a mild increase in per-product quantities.

[26]Unit prices are computed as total sales divided by total quantities sold. A limitation of the product-level data is inconsistency in measurement units across firms and, at times, within firms over time. We address this by standardizing units across and within firms and excluding firm–product pairs that report changes in unit types.

[27]Product-level data contain many missing values, resulting in fewer observations for firm–product revenues, prices, and quantities than for firm-year outcomes.