G²LM|LIC
Gender, Growth, and Labour Markets in Low Income Countries Programme

WP

G²LM|LIC Working Paper No. 99 | June 2026

Moving Opportunity Closer:

How Public Transit Transforms Firm Composition and Employment

Akhila Kovvuri (Stanford University)
Karmini Sharma (Imperial College London)

IZA Institute of Labor Economics
Initiated by Deutsche Post Foundation

UKaid
from the British people

G²LM|LIC Working Paper No. 99 | June 2026

Moving Opportunity Closer:

How Public Transit Transforms Firm Composition and Employment

Akhila Kovvuri (Stanford University)
Karmini Sharma (Imperial College London)

G²LM|LIC
c/o IZA – Institute of Labor Economics
Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-510

Email: glm-contact@iza.org

IZA Institute of Labor Economics
Initiated by Deutsche Post Foundation

UKaid from the British people

G²LM|LIC Working Paper No. 99 | June 2026

ABSTRACT

Moving Opportunity Closer:*

How Public Transit Transforms Firm Composition and Employment

Transportation infrastructure can improve workers' access to existing economic opportunities, but it can also reshape economic opportunity itself by influencing where and what kinds of firms locate. This paper studies how public transit infrastructure influences firm location, composition, and employment at the neighborhood level. We construct novel data tracking over one million establishment entries and employ both difference-in-differences and market access specifications, exploiting the phased expansion of the Delhi Metro Rail in India. Transit access increases firm entry near stations, with larger, established retail and service firms locating first and inducing subsequent entry of other firms. These patterns create new economic hubs in peripheral areas, increasing employment per capita, especially for women in a context of low baseline female labor force participation. Counterfactual decompositions using a quantitative spatial model with estimated gender-specific commute elasticities reveal that compositional shifts toward larger establishments and consumer-facing industries that ex-ante employ more women account for the majority of this differential employment effect. Understanding how infrastructure reshapes the demand side of the labor market is thus critical for predicting and enhancing its distributional impacts.

JEL Classification:
O18, R12, J16, J21

Keywords:
transportation infrastructure, firm dynamics, female labor force participation, quantitative spatial economics

Corresponding author:
Akhila Kovvuri
Stanford University
450 Jane Stanford Way
Stanford, CA 94305
Email: akhilajk@stanford.edu

* ∗I am especially grateful to my advisors Melanie Morten, Alessandra Voena, Nick Bloom, and Arun Chandrasekhar for their invaluable support and guidance. I also thank the valuable comments and suggestions of Marcella Alsan, Daliah Al-Shakhshir, Michael Callen, Marcel Fafchamps, Kyra Frye, Jessica Goldberg, Ida Grigoryeva, Florencia Hnilo, Caroline Hoxby, Suhani Jalota, Ryan Louie, Tamri Matiashvili, David McKenzie, Luigi Pistaferri, João Pugliese, Steve Redding, Anirudh Sankar, Isaac Sorkin, Janet Stefanov, Kate Vyborny, Laura Weiwu, and seminar participants at development, labor, spatial/trade, and applied workshops at Stanford, and at the SITE Gender and ACLEC conferences. I am grateful to Resuf Ahmed for support on registration data extraction, to Raj Bhagat Palanichamy and Dan Ma for ward shape files of Delhi, to Cuebiq and the World Bank for mobile phone location data shared through the NetMob 2024 Data Challenge, and to Delhi Metro Rail Corporation officials for explaining and sharing planning documents and data. Laura Lufray provided excellent research assistance. I gratefully acknowledge financial support from the King Center at Stanford University, Stanford's Research, Action, and Impact through Strategic Engagement (RAISE) Doctoral Fellowship, George P. Shultz Dissertation Support Fund at SIEPR, Institute for Humane Studies, G²LM|LIC, and Weiss Fund at the University of Chicago.

1 Introduction

In rapidly urbanizing cities across the world, governments are investing trillions in public transit infrastructure. The conventional rationale focuses on worker access: better transit helps people reach distant jobs. But transit could also fundamentally alter the geography of jobs by influencing where and what kinds of firms locate. It is an open question how such transit-induced changes in firm location affect the spatial distribution of employment. Moreover, workers vary in their propensity to work in different types of firms, raising a second open question: how do transit-induced changes in firm composition translate into heterogeneous employment effects across worker groups?

This paper studies how public transit infrastructure influences firm location, composition, and employment, leveraging novel neighborhood-level data and the phased expansion of the Delhi Metro Rail in India. We construct a dataset of over one million establishment entries to track localized and dynamic firm responses to transit station openings. To address location endogeneity and network effects, we employ difference-in-differences strategies comparing built to planned-but-not-built stations, and market access measures capturing network-wide connectivity changes. Counterfactual decompositions separate employment effects arising from shifts in firm size and industry composition from effects arising within firm types. Using estimated commute elasticities from mobile phone and census data, we further parse residential employment gains into mobility improvements versus local labor demand shifts. To our knowledge, this provides the first comprehensive evidence on how intracity transit reshapes local economic composition and generates heterogeneous employment effects through firm entry dynamics. The findings underscore that firm heterogeneity and boundaries, typically abstracted away in spatial equilibrium models, are critical for understanding distributional impacts.

Our findings reveal that transit access induces firm entry and reshapes the composition of local economic activity. Within months of station openings, new firm registrations increase, with entrants being larger and more organizationally complex. The timing of entry reveals that in areas with limited prior firm entry, larger, established brands enter first, triggering subsequent entry of smaller firms seeking to benefit from spillovers. The key incentive for firm entry appears to be access to consumers: of the 9 additional firms entering per neighborhood in a 6-month period, 7 are in retail or services, consistent with observed increases in foot traffic following station openings measured using mobile phone location data. These changes translate into the creation of new economic hubs around stations, especially in peripheral areas of the city. Employment increases by 177 workers per 100 residents, indicating increasing commercial activity in previously more residential areas.

Notably, we find greater increases in female employment: 28 of these additional workers are women, a large increase given low baseline labor force participation rates in this context. Counterfactual decompositions reveal that shifts toward larger establishments and consumer-facing businesses that ex-ante employ more women account for most of the differential employment effect. We further employ exact hat algebra with the canonical quantitative spatial model (Ahlfeldt, Redding, Sturm, and Wolf, 2015; Heblich, Redding, and Sturm, 2020) to decompose residential employment gains into reduced commuting costs, spatial redistribution of jobs, and shifts in firm composition. This finding has important policy implications: interventions focused solely on reducing women's commuting costs may increase mobility but generate limited employment gains without accompanying increases in jobs that hire women. Understanding how infrastructure reshapes the demand side of the labor market is thus critical for predicting and enhancing its distributional impacts.

To identify intra-city effects of transit infrastructure on firms, we construct the first spatially and temporally granular datasets of establishments for urban India. Through a combination of secondary dataset harmonization, matching algorithms, and scraping, cleaning, and geocoding of administrative sources, we assemble a repeated cross-section of formal and informal establishments from the Economic Census (1990, 2005, and 2013) covering 500 wards,[1] and the flow of new establishment registrations (2011-2024) covering around 1,800 neighborhoods. This database extends beyond the district-level analyses that are typically feasible for firm analyses in Indian cities, increasing the spatial resolution by 200-fold.[2] We complement these establishment data with Population Census data on residential employment and commuting data from census, surveys, and mobile phone location records, enabling estimation of commuting elasticities, counterfactual decompositions of residential employment gains, and descriptive analyses of firm entry mechanisms.

Through these rich spatial data, we identify two key facts that motivate us to study the effect of transit access on the location and composition of firm activity. First, employment rates of residents increase in wards closer to transit stations, especially in the urban periph-

[1] An urban electoral ward or a rural village, spatially similar to an electoral precinct in the United States

[2] Prior studies of firms in India use district-level data from Annual Surveys of Industries or National Sample Surveys (Topalova, 2010; Martin, Nataraj, and Harrison, 2017). Asher, Lunt, Matsuura, and Novosad (2021) enables spatially granular village-level analyses in rural areas (Asher and Novosad, 2020), but intra-city analysis of firms remains data constrained, with Gechter and Tsivanidis (2023) providing a notable exception for Mumbai, India.

ery ($\geq$5 miles from the city center) and particularly for women, aligning with prior work on transit and female employment (Martinez, Mitnik, Salgado, Scholl, and Yañez-Pagans, 2020; Kwon and Lee, 2024; Seki and Yamada, 2025). This is congruent with the theory that improved commuting technology could enable peripheral residents to access employment opportunities in the city center. However, we also observe a second pattern: increased firm entry concentrated around transit stations, rapidly decaying with distance from the stations. This suggests a change in the density of local economic activity in these neighborhoods. In contexts where workers are highly sensitive to commuting distance (Deffebach, Lagakos, Miyauchi, and Yamada, 2025), changes in employment opportunities within one's own neighborhood are likely to influence employment decisions.

To isolate the causal effect of transit access on firm entry and employment, we exploit three features of Delhi Metro's implementation. First, the long planning horizon—initial plans in the 1970s but first operations in 2002—suggests station locations were not based on contemporaneous economic activity. Second, idiosyncrasies in timing meant places planned to receive stations around the same time experienced delays due to funding availability, land acquisition, and engineering constraints, enabling comparisons of treated to not-yet-treated areas. Third, deviations in routes occurred for engineering reasons: for example, Phase 3 lines built along major roadways to minimize costs and disturbance. We use places that were viable according to construction methods and a 2002 master plan but did not receive stations as comparison groups. We implement staggered difference-in-differences designs (Callaway and Sant'Anna, 2021) and long-differences specifications, verifying robustness across alternative comparison groups and distance buffers. Parallel trends hold across all specifications, and the temporal granularity of our firm entry data with exact business commencement dates enables precise identification of responses to station openings.

To capture transit effects beyond localized station areas, we employ two complementary spatial approaches. First, we construct market access measures following Donaldson and Hornbeck (2016) and Tsivanidis (2023) to account for broader connectivity gains across the entire network, given that spatial spillovers are inherent in transit systems. We include falsification tests to verify that changing market access and not endogeneity of placement that is driving the employment effects (Borusyak and Hull, 2023). Second, to quantify the relative importance of firm composition changes versus improved mobility in explaining employment effects, we implement counterfactual decompositions using a gravity-based commuting framework (Ahlfeldt et al., 2015).

We first test whether the spatial correlation between transit stations and new firm registrations reflects a causal effect of transit infrastructure. New firm location decisions reveal the types of establishments drawn to transit access—a first step for understanding how transit reshapes local economic composition, particularly in contexts like India where growth comes primarily from firm entry (Hsieh and Klenow, 2014; Akcigit, Alp, and Peters, 2021; Chatterjee, Krishna, Padmakumar, and Zhao, 2025). We find that transit station openings cause an immediate, significant, and persistent increase in firm entry, with 9 additional firms entering a neighborhood in a 6-month period, compared to the typical entry of 6 firms in comparable areas. The impact decays with distance from stations, with positive spillovers extending 1–2 kilometers. Entrants near transit stations are larger and more specialized, and the increase in firm entry is driven almost entirely by business-to-consumer establishments, with 7 of the 9 new firms being in retail or services serving consumers directly.

Exploiting the temporal granularity of our registration data, we find that larger, established brands enter first near new transit stations, triggering sequential waves of smaller firm entry through customer spillovers. This pattern is especially pronounced in areas with limited baseline firm entry. This is consistent with agglomeration mechanisms in service sectors (Shoag and Veuger, 2018; Benmelech, Bergman, Milanez, and Mukharlyamov, 2019; Leonardi and Moretti, 2023): smaller establishments benefit from customer spillovers generated by larger, branded firms. This dynamic is further supported by increases in foot traffic around new stations measured through mobile phone location data, as well as the lack of entry response from business-to-business firms, whether they hire high- or low-skill labor. This evidence indicates that access to consumers (Miyauchi, Nakajima, and Redding, 2021; Bassi, Kahn, Gracia, Porzio, and Sorin, 2022; Oh and Seo, 2023; Vitali, 2023), rather than to workers, constitutes the primary incentive for firm location near transit stations.

Transit-induced firm entry translates into persistent changes in the stock of establishments and employment density. Using difference-in-differences comparing Phase 2 station areas to not-yet-treated areas, Economic Census data spanning 1990-2013 reveal that the entry of larger firms shifts job composition toward establishments with 47% more employees on average within 1 kilometer of a transit station. On average, there are 177 more employees per 100 residents, indicating that previously residential areas experience increased commercial activity. These employment effects exhibit strong distance gradients, with positive spillovers in neighborhoods within 1-2 kilometers of stations but imprecise and smaller negative effects 2-5 kilometers away, potentially reflecting competition from newly entering firms.

Peripheral areas experience the largest employment gains from transit network expansion. We construct market access measures following (Donaldson and Hornbeck, 2016; Tsivanidis, 2023) to capture how new stations alter connectivity citywide beyond station-adjacent neighborhoods. Areas in the urban periphery experience the largest increases in firms' market access to labor, with greater employment in firms and larger firm size, specifically in retail and services. This pattern aligns with economic decentralization when transit connects previously isolated peripheral areas (Baum-Snow, Brandt, Henderson, Turner, and Zhang, 2017; Baum-Snow, 2020; Nose and Sawada, 2025), and contrasts with transit studies finding employment centralization (Faber, 2014; Heblich et al., 2020).

Transit-induced employment growth occurs in different locations and different types of firms, raising a critical question: which workers benefit? Jobs concentrate near transit stations and shift toward larger, consumer-facing establishments: two changes with potential implications for workforce composition, as workers differ in commute elasticities and firms differ in hiring patterns. We examine whether these spatial and compositional shifts generate differential employment effects by gender, linking back to our motivating fact that labor force participation, especially of women, increases more near transit stations. Of the additional 177 employed per 100 residents, 28 are women. While this represents 17% of total employment gains, this is notable given low baseline female labor force participation in urban India and that other gendered transit interventions in this context, such as free bus travel for women, have not increased female employment (Chen, Coşar, Ghose, Mahendru, and Sekhri, 2024).

What explains women's employment gains near transit stations? While metro access increases entry of female-managed firms (Chiplunkar and Goldberg, 2021; Hunt and Moehling, 2024) and businesses serving female customers (Otterbring, Bhatnagar, Samuelsson, and Borau, 2021; Kelley, Lane, Pecenco, and Rubin, 2023), employment effects are not confined to these categories. Larger firms and consumer-facing industries that enter near transit stations employ more women ex-ante (Miller, Peck, and Seflek, 2022), whether due to reduced discrimination in competitive markets (Black and Brainerd, 2004; Juhn, Ujhelyi, and Villegas-Sanchez, 2014; Araújo, Paz, and West, 2024) or women's preferences for such establishments (Kline, Rose, and Walters, 2022; Schuh, 2024; Larson-Koester, 2020). Critically, female-intensive manufacturing firms (e.g., apparel manufacturing) and business-to-business service firms (e.g., BPO/call-centers) do not respond to metro access, indicating that consumer access, rather than access to female labor, drives firm location decisions near transit stations.

To quantify how much firm compositional shifts versus improved commuting access explain gendered employment gains, we implement two complementary decomposition approaches. First, workplace decompositions reveal that shifts toward larger firms and female-intensive industries account for the majority of increased female employment in establishments near metro stations. Second, we use a quantitative spatial model to decompose residential employment changes into three channels: reduced commuting costs, spatial job redistribution, and firm size and industry compositions shifts. The model incorporates gender-specific commute elasticities estimated from Delhi commuting data, with women 20% more sensitive to commute costs, consistent with estimates from other countries (Le Barbanchon, Rathelot, and Roulet, 2021; Velásquez, 2023). Together, these decompositions demonstrate that the types of firms attracted to transit access, and not just improved worker mobility, determine which workers benefit from infrastructure investment.

This paper contributes to three distinct literatures. First, we contribute to research on firm location decisions and hiring patterns by showing how transportation infrastructure reshapes both where firms locate and which workers they employ. Second, we advance the spatial economics literature on transit infrastructure by documenting, to our knowledge, the first empirical evidence that firm heterogeneity and boundaries matter for distributional employment effects of transit. Third, we extend research on gender and commute by providing direct evidence on the labor demand channel, showing that firm responses to transportation infrastructure create new employment opportunities for women beyond traditional mobility effects. A key methodological contribution is the construction and harmonization of multiple administrative datasets at unprecedented spatial granularity for urban India, enabling analysis of firm and employment responses to transportation infrastructure at the neighborhood level.

Firm Location and Workforce Composition: We build on two strands of literature examining firm location and hiring. Rental costs, wages, and agglomeration economies shape firm sorting across space (Lindenlaub, Oh, and Peters, 2022). Consumer access represents a particularly important location incentive for retail and service firms (Miyauchi et al., 2021; Bassi et al., 2022; Vitali, 2023; Oh and Seo, 2023). A separate literature documents how firm characteristics predict workforce composition across skill, race, and gender dimensions (Holzer and Ihlanfeldt, 1998; Hjort, 2014; Brinatti and Morales, 2021; Kline et al., 2022; Goraya and Ilango, 2024). Gender patterns are particularly pronounced: women concentrate in larger establishments and female-owned firms (Kline et al., 2022; Schuh, 2024; Chiplunkar and Goldberg, 2021; Hunt and Moehling, 2024; Miller et al., 2022). Competitive pressure reduces discriminatory hiring (Black and Brainerd, 2004; Juhn et al., 2014; Araújo et al., 2024), while women may prefer certain establishments due to wages, amenities, or presence of other female workers (Larson-Koester, 2020; Lordan and Pischke, 2022; Schuh, 2024). We link these strands and contribute to this literature by studying how transit infrastructure changes firm location incentives and resulting employment effects. Consumer-facing establishments (retail and services) respond to transit-induced footfall increases, while business-to-business firms and manufacturers do not. Transit also likely affects sorting through rents and competitive pressures, with larger firms entering near stations. In our context, these compositional shifts toward larger, service-oriented firms drive heterogeneous employment effects across worker groups.

Economic Geography of Transit Infrastructure: A substantial literature examines how transit infrastructure affects spatial economic organization (reviewed in Duranton and Puga (2020), Redding and Turner (2015), Redding (2025), Donaldson (2025)). Quantitative spatial equilibrium models have been influential in studying welfare effects of public transit infrastructure, accounting for changes in commute costs and residential sorting (Heblich et al., 2020; Tsivanidis, 2023; Zárate, 2022; Severen, 2023; Balboni, Bryan, Morten, O'Connor, and Siddiqi, 2025). Some of these have emphasized distributional employment effects by worker skill levels (Tsivanidis, 2023) and formality (Zárate, 2022). Recent work has looked at new firm entries and increase in local productivity (Busso and Fentanes, 2024). A central modeling assumption in the canonical spatial model is that firms are boundary-less (with some exceptions: Ahlfeldt, Albers, and Behrens (2022))—employment aggregates through constant elasticity of substitution production functions where 100 firms each employing one worker generates identical distributional effects as one firm employing 100 workers. We contribute by empirically documenting how intracity public transit triggers dynamism in number and types of firms, with the resulting changes in firm size and industry composition influencing the distributional effects on employment.

Commute and Gender: Women face higher commute costs than men, with women willing to trade wages for shorter commutes in developed economies (Nafilyan, 2019; Le Barbanchon et al., 2021; Liu and Su, 2024). Safety concerns, social norms, and household responsibilities amplify these constraints in developing contexts (Kondylis, Legovini, Vyborny, Zwager, and Cardoso De Andrade, 2020; Aguilar, Gutiérrez, and Villagrán, 2021; Borker et al., 2021; Jayachandran, 2021). Recent research examines how transportation interventions affect female mobility and employment. Targeted interventions have addresses this by testing a range of interventions from female-only transportation to bringing jobs to women and have recorded increases women's job search and skill acquisition (Cheema, Khwaja, Naseer, and Shapiro, 2019; Field and Vyborny, 2022; Ho, Jalota, and Karandikar, 2023; Jalota and Ho, 2024; Kapoor and Gade, 2024). Studies of transit infrastructure expansions show mixed employment effects: some find increases in female labor force participation (Martinez et al., 2020; Kwon and Lee, 2024; Seki and Yamada, 2025), while others find limited employment gains from mobility improvements alone (Alam, Cropper, Herrera Dappe, and Suri, 2021; Chen et al., 2024). Kwon and Lee (2024) provides suggestive evidence that inter-city transit increases demand for female workers in peripheral service sectors.

We extend this literature by providing direct evidence on labor demand constraints. While existing work focuses primarily on supply-side frictions—how transportation enables women to reach existing jobs—we show that structural changes in employment opportunities can also drive differential gender effects. Transit stations attract larger establishments and consumer-facing businesses that ex-ante employ more women. Using counterfactual decompositions, we quantify that firm compositional changes account for the majority of observed female employment gains, suggesting supply-side transport interventions alone may have limited impact where labor demand remains concentrated in sectors that under-employ women (Klasen and Pieters, 2015).

2 Context and Institutional Background

2.1 Delhi in the Early 2000s

Delhi experienced rapid demographic expansion from 3 million residents in 1970 to 15 million by 2000, driven by rural-urban migration and natural population increase. This growth occurred alongside limited structural transformation in the city's economic base. Employment patterns circa 2000 reflected a concentration in low-productivity sectors, with the majority of workers employed in retail trade and low-skill manufacturing (Appendix Figure A1a). Self-employment rates remained high at approximately 35%, indicative of a workforce operating predominantly outside formal wage employment. The formal employment sector exhibited a highly skewed firm size distribution: fewer than 2% of establishments employed ten or more workers, yet these larger establishments accounted for one-third of total employment (Appendix Figure A1b). This distribution reflects India's broader pattern of missing medium-sized firms, driven in part by regulatory distortions (Hsieh and Klenow, 2014; Hasan and Jandoc, 2013).

Transportation infrastructure in the 2000s Delhi relied primarily on walking and bus transit. Census data from 2011 indicate that the median worker commuted less than three miles mile to work, with approximately 45% of workers walking as their primary mode of transport and 30% using buses (Appendix Figure A2a and A2b). Private vehicle ownership remained limited, with motorcycles being the most common private vehicle but mostly only used by men, and cars accessible primarily to higher-income households. This localized commuting pattern suggests that employment opportunities were determined largely by residential location, with workers facing binding constraints in accessing jobs beyond their immediate neighborhoods.

Gender disparities in labor market participation were pronounced. Female labor force participation in Delhi stood at approximately 10% in 2001, compared to 62% for males. Among women who did work, commute distances were systematically shorter than men's, and more reliant on walking (Appendix Figures A2a and A2b). These patterns likely reflect various factors such as household responsibilities, social norms restricting women's mobility, and safety concerns (Jayachandran, 2021; Borker et al., 2021).

2.2 The Delhi Metro Rail System

Against the back drop of increasing congestion, Delhi Metro Rail was initially conceived in a 1970 traffic study, iterated over in the next two decades, with construction finally commencing in 1998 and the first line becoming operational in 2002 (Siemiatycki, 2006; dmrc, 2010; Onishi, 2016). Delhi Metro's construction proceeded continuously over two decades with major operational expansions occurring in distinct phases, evolving from an initial radial network to include circular connectivity that fundamentally altered the city's transit geography. The Appendix Figure 1a shows the evolution of the metro spatially over primarily 3 periods (which we define based on the phase of construction as well as the timeline of the data). The early phase from 2002-2004 connects some of the more central areas of the city, but the majority of the expansion happened in Phase 2 (2005-2011), when the system expanded radially. Phase 3 (2012 onward) added some more radial extensions and circular lines to increase connectivity in peripheries. As shown in Appendix Figure 1b, the network expanded close to about 100 stations within Delhi, with a subsequent acceleration during 2016-2018 bringing the total to 165 stations. We exploit this staggered nature of the expansion to identify the causal effects of transit access.

The nearly three-decade-long planning horizon created temporal disconnects between route decisions and contemporaneous economic conditions. Construction delays arose from extended technical studies, political negotiations, and financing arrangements (Siemiatycki, 2006). Implementation exhibited idiosyncrasies in both station placement and opening timing. Final construction often differed from original plans, with corridors truncated, extended, rerouted, or pushed to subsequent construction phases based on practical constraints. Modifications were implemented to minimize private land acquisition and disturbance to existing properties, as well as to address integration with the existing metro system and economic viability considerations (DMRC, 2011). Timeline variations included both acceleration to meet Commonwealth Games 2011 deadlines (Kayal, 2013) and delays from land acquisition challenges and logistical factors.

These implementation idiosyncrasies provide identifying variation for our analysis. First, many stations planned for simultaneous construction were instead built at different times or never constructed, creating variation in transit access among areas that policymakers initially deemed similar. Second, many expansions, especially in Phase 3, deviated from plans to follow major roadways, prioritizing cost minimization and reduced disturbance. These features enable comparisons between neighborhoods that were deemed suitable for station placement but experienced delays or cancellations due to implementation constraints plausibly orthogonal to firm location decisions.

3 Data

We construct the first spatially and temporally granular dataset of establishment stocks and flows for Delhi, India. This database extends beyond the typical district-level analyses feasible for firm studies in Indian cities, providing both repeated cross-sections of establishment stocks and high-frequency data on firm entry. We assemble two complementary sources of establishment data: the Economic Census (1990, 2005, and 2013), which we expand from 9 urban districts to approximately 450 urban wards or rural villages, and Registrations of Shops and Establishments (2011-2024), a novel dataset covering around 1,800 neighborhoods with date-level temporal resolution. The Economic Census expansion represents a 50-fold increase in spatial units, while the registration data provide an unprecedented level of spatial and temporal granularity for studying firm entry in an Indian city. Through a combination of matching algorithms and web scraping of administrative sources, we create data that enable analysis of both the stock of establishments at a spatially disaggregated level and the flow of new firms at spatially and temporally granular scales rare for the Indian context.

3.1 Economic Census: Establishment Stock (1990, 2005, 2013)

The Economic Census provides repeated cross-sections capturing the universe of non-agricultural establishments in India, including both formal and informal firms. We use data from the years 1990, 2005, and 2013 for our analysis.[3] For each establishment, we observe employment (total and hired workers, disaggregated by gender), gender of the owner, and industry classification codes at the 4-digit level. These data enable us to study firm presence, size, industry composition, and workforce characteristics in response to transit access.

We expand the spatial resolution of the Economic Census in urban Delhi from 9 districts to approximately 450 wards, adapting a matching algorithm developed and used by Gechter and Tsivanidis (2023) in Mumbai. The algorithm leverages a complementary firm-level directory containing exact addresses for establishments with 10 or more employees, which we use to locate the remaining firms at the urban ward level, which is approximately equivalent to an electoral precinct in the United States. Appendix Figure A3 illustrates this expansion: black boundaries show the originally available 9 districts with limited spatial granularity in densely populated city centers, while blue boundaries display Voronoi polygons of the census enumeration units that we aggregate to wards.

3.2 Shops and Establishments Registration: Establishment Flows (2011-2024)

To study how firm location decisions respond to transit infrastructure, we construct a novel dataset of commercial establishment registrations in Delhi. These data come from the Delhi Shops and Establishments Act, which requires most commercial establishments to register with the state's labor department. We create this dataset through comprehensive web scraping of the Delhi government's official registration portal, yielding over one million registration raw records. To our knowledge, this dataset has not been previously utilized in academic research.

The registration data provide several advantages for studying firm entry and composition. First, we observe the exact date of registration and reported date of business commencement, which enables event study analysis around transit station openings. Second, the exact address information allows us to precisely geolocate establishments and measure proximity to transit infrastructure at fine spatial scales. Third, the product or service descriptions in registration records are substantially more granular than standard industry classification

[3] Economic Census was also conducted in 1998 but we have not yet found a way to geolocate the enumeration units in a way that would give us enough spatial granularity in the dense urban parts of the city.

codes. Rather than broad categories such as "retail trade," we observe specific business types like "garment store," "electronic products and components," and so on. Appendix Figure A4 shows the top 20 categories based on number of firm registrations. Such detailed categorization enables us to differentiate between manufacturing and service sectors, between business-to-business and business-to-consumer firms, and between establishments that would typically hire high-skill versus low-skill labor. Fourth, we observe owner and manager names, which enable gender identification. Fifth, we observe the number of male and female employees at the time of registration, which enables us to study gender heterogeneous employment effects.

We aggregate the geocoded establishment registrations to approximately 1,800 neighborhoods covering Delhi using the H3 hexagonal grid system at resolution 8.[4] Each hexagon has an area of approximately 0.075 km^2, corresponding to a radius of roughly 500 meters (a 5 to 10 minute walk from center to boundary). We use hexagons rather than administrative boundaries because hexagons provide uniform spatial units of consistent size and shape across the city, and aggregation to this scale accounts for noise in the geocoding process. For the current analysis, we focus on registrations from 2011 onward for firms that report commencing business operations less than one year prior to registration (Appendix Table A1 shows descriptive statistics for the registration data).

3.3 Additional Data Sources

3.3.1 Population Census

For demographic characteristics and labor force participation at the residence level, we use Population Census data from 1991, 2001, and 2011, analyzed at the harmonized 2011 ward level. Key variables include population, number of workers, literacy rates, and scheduled caste/scheduled tribe population, all disaggregated by gender. These data enable calculation of employment rates by gender and analysis of neighborhood demographic composition. The population census data are sourced from the Census of India and available in tabular format at year-specific administrative units. For rural areas, these units are villages; for urban areas, these are towns or the Delhi Municipal Corporation along with respective electoral wards (or charges in 1991).

[4] https://h3geo.org/

3.3.2 Metro System

We obtain metro station locations, opening dates, and line alignments from official Delhi Metro Rail Corporation announcements and third-party GIS mapping sources. This includes both operational stations and stations that are going to come up in the near future (not-yet-built). For planned-but-never-built stations, we digitize paper maps from a 2002 master plan that outlined the intended metro network through 2021. We obtain road network data from OpenStreetMap, classifying roads as motorways, trunk roads, or primary roads.

Metro exposure, our key treatment variable, is measured at the spatial unit level based on the presence of metro stations within ward boundaries or within specified distances (1km, 2km, etc.) from the centroid of the spatial unit (ward or hexagon). In our base specification for firm entry analysis, we leverage the fact that Phase 3 construction plans were designed to construct lines along major roads to reduce costs and minimize land acquisition and disturbance to residents and businesses. We restrict the sample to neighborhoods within 1 kilometer of major roads and compare neighborhoods within 1km of Phase 3 stations (treated) to those beyond 1km of any station but still near major roads (comparison). This approach enables analysis of how metro accessibility changes over time across different spatial scales while accounting for potential selection in station placement.

3.3.3 Commuting

We use three sources of commuting data. First, the Population Census 2011 provides commute mode and distance to work tabulated at the district-gender level, which we use for descriptive facts about commute in Delhi. Second, the OLA Mobility Survey 2022 is an individual-level dataset with origin and destination neighborhoods for commutes, which we use to calibrate commuting parameters. Third, we obtain mobile phone location data for 2019 from Cuebiq, with density and flows aggregated at the geohash 5 level[5] (4.88km ^{2} grid cells), which we use to measure changes in foot traffic near metro stations.

4 Motivating Facts

4.1 Increase in Labor Force Participation near Metro Stations

We study the effect of metro exposure on the employment rates of residents in a given ward, utilizing Population Census data from 1991, 2001, and 2011 at the 2011 harmonized ward level. We estimate the following specification:

[5] https://www.movable-type.co.uk/scripts/geohash.html

$$Y_{pt} = \beta_1 \text{Treated}_p \times \text{Year2011}_t + \beta_2 \text{Treated}_p \times \text{Year2001}_t + \delta_t + \gamma_p + \epsilon_{pt} \quad (1)$$

where Y_{pt} is the employment outcome (proportion of population employed) in ward p at time t, Treated_p is an indicator equal to 1 if ward p has a metro stop by 2011 and 0 otherwise, Year2011_t and Year2001_t are year indicators, δ_t are year fixed effects, and γ_p are ward fixed effects. We weight the regressions by total population at the ward level, as the data are only available at ward-level aggregates and the outcome variables are proportions. The key coefficient of interest is β_1, which captures the effect of metro access on employment outcomes by 2011, noting that metro roll out started in 2002. Similar to the analysis with the Census of Firms, we restrict our analysis to ward that eventually received metro stations by 2023 and test for parallel trends. We account for spatial autocorrelation using Conley standard errors with a 5km bandwidth, recognizing that employment outcomes in neighboring wards may be correlated due to spillover effects or common unobserved shocks.

Figure 2 presents the estimated treatment effects of metro access on labor force participation at varying distances from transit stations. Panel A shows that overall labor force participation increases in wards closer to metro stations, with the effect declining monotonically with distance. The increase is 5.3 percentage points for wards within 1 kilometer of a station, falling to (an insignificant) 1.3 percentage points for wards 5-10 kilometers away.

This aggregate pattern masks heterogeneity by gender and location. Panel B reveals that the increase in labor force participation is driven primarily by women, with female labor force participation rising by 7.5 percentage points in wards within 1 kilometer of a station. The female employment effect also exhibits a distance gradient, declining at distances beyond 2 kilometers from stations. Panel C demonstrates that these effects are concentrated in peripheral areas: when restricting the sample to wards within 5 kilometers of the central business district, the overall labor force participation effects are larger and more precisely estimated in areas closer to transit stations, consistent with peripheral residents facing higher commute costs that transit access alleviates.

4.2 Increase in Firm Entry near Metro Stations

The increase in labor force participation documented in Section 4.1—particularly pronounced in peripheral areas—could reflect improved commuting access enabling residents to reach employment opportunities in the city center. However, we observe a second spatial pattern that suggests transit infrastructure also reshapes the local economic landscape: firm entry concentrates near metro stations, with this effect decaying rapidly with distance.

Figure 3 presents the total number of firm entries by distance from the nearest transit station. The spatial gradient is pronounced: neighborhoods within 1 kilometer of a station account for over 100,000 firm entries, while those 1-2 kilometers away experience approximately one-third as many entries. This decay continues through intermediate distances, with firm entry declining to minimal levels beyond 5 kilometers from the nearest station. The pattern indicates that transit proximity generates localized effects on firm location decisions, with the influence of station access diminishing sharply beyond a 1-2 kilometer radius.

The concentration of firm entry near transit stations—combined with the labor force participation increases documented in Section 4.1—raises the question of whether transit affects employment primarily through improved worker mobility or through changes in local job opportunities. While employment increases could reflect workers' enhanced ability to commute to distant employment centers, the spatial concentration of firm entry suggests that transit may also generate jobs locally.[6] The clustering of new establishments near stations could arise from endogenous neighborhood characteristics rather than causal effects of transit access, a concern we address through our empirical strategies. Nonetheless, this pattern motivates our analysis of how transit infrastructure affects firm entry, firm and industry composition, and employment.

5 Transit Infrastructure Effects on Firm Entry

5.1 Empirical Strategy

For the firm entry analysis, we leverage the Phase 3 metro expansion (2011-2024) and implement a staggered difference-in-differences design following Callaway and Sant'Anna (2021). Our approach addresses the challenge that metro stations opened at different times across Delhi, requiring methods robust to heterogeneous treatment timing.

Our baseline specification estimates dynamic treatment effects using the following event study framework:

$$Y_{pt} = \alpha + \sum_{k \neq -5} \beta_k \cdot \mathbb{1}[\text{RelTime}_{pt} = k] + \gamma_p + \delta_t + \epsilon_{pt} \tag{2}$$

where Y_{pt} represents firm entry outcomes (number of new firms, employment, or firm char-

[6] We find suggestive evidence of increased correlation between residence employment and workplace employment in peripheral areas.

acteristics) in neighborhood p during 6-month period t. We define neighborhoods using H3 hexagonal spatial indexes at resolution 8, which partition Delhi into approximately 500-meter radius hexagons. This granular geographic unit allows us to precisely measure proximity to metro stations while maintaining computational tractability.

The variable RelTime_{pt} denotes event time relative to metro station opening in neighborhood p, measured in 6-month periods. We bin event time to reduce noise while preserving dynamic patterns. The coefficients $\{\beta_k\}$ trace out the treatment effect trajectory before and after metro opening. We normalize β_{-5} to zero, making period $k = -5$ (2.5 years before opening) the reference period. Treatment begins not at the operational opening date but two years prior, corresponding to the period between the start of testing and when the transit station becomes operational. The end of construction and beginning of testing may generate anticipatory registrations and entries from businesses, as the station's imminent operation becomes apparent. We include neighborhood fixed effects (γ_p) and 6-month period fixed effects (δ_t), and cluster standard errors at the neighborhood level.

In our base specification, we use information from planning documents that detail Phase 3 construction plans and deviations. In many cases, plans were designed to construct lines along major roads to reduce costs and minimize land acquisition and disturbance to residents and businesses. We restrict the sample to neighborhoods within 1 kilometer of major roads (defined as motorways, trunk roads, or primary roads from OpenStreetMap). Within this sample, we compare neighborhoods within 1km of Phase 3 stations (treated) to those beyond 1km of any station but still near major roads (control).

We also use an alternate specification using stations from Delhi's 2002 master plan that were planned but never constructed, as well as Phase 4 stations not yet built by 2024. This addresses concerns that station placement reflects omitted neighborhood characteristics. If our effects persist when comparing to planned-but-not-built locations, it suggests metro impacts operate through actual connectivity rather than unobserved place-based factors.

Our identification strategy relies on two key assumptions:

- Parallel Trends; n the absence of metro station openings, neighborhoods that eventually receive stations would have followed similar trends in firm entry as comparison neighborhoods. We provide evidence for this assumption by: (1) showing insignificant pre-treatment coefficients in our event studies, (2) formally testing joint nullity of pre-treatment effects, and (3) using reasonable comparisons, such as roadways and planned but not built stations.

- No anticipation before testing: We account for potential anticipation once testing begins, but assume that before testing, firms do not enter as construction and delays can hurt business.

5.2 Results

We begin by examining the dynamic response of firm entry to metro station openings. Figure 4a presents event study estimates of treatment effects on the number of new firm registrations in treated neighborhoods relative to comparison neighborhoods in a 6-month period. The average treatment effect on the treated (ATT) is 8.91 additional firms per neighborhood per 6-month period relative to a baseline mean of 6.2 firms in control neighborhoods, representing a 144% increase in the flow of new establishments. We do not observe significant pre-trends. The effect remains elevated and statistically significant at conventional levels throughout the post-treatment period. While the magnitude declines slightly over time, coefficients remain significantly above zero, indicating that metro access generates persistent changes in firm location decisions rather than temporary responses to station openings.

Figure 4b shows that this increase in firm entry is driven almost entirely by establishments serving end consumers (business-to-consumer firms or B2C firms). B2C firms increase by 280.55% (SE: 69.53) relative to a baseline mean of 2.4 firms, while B2B firms show no differential response to transit infrastructure. This pattern is noteworthy given that transit access is typically viewed as a mechanism for firms to access labor pools. Table 1 (Panel A, Col 5) confirms that firms relying on high-skill workers are not more likely to enter near transit stations. These findings link back to the motivating observation that increased customer foot traffic near metro stations provides location advantages for consumer-facing establishments, suggesting that demand-side effects dominate labor access considerations in firm sorting around transit infrastructure.

In addition to entry of more firms, the firms entering near a transit station are 19% larger (more employees) and 94% more likely to be specialized, i.e., the manager being separate from the owner (Table 1, Panel A, Cols 2-3). The results on more firm entries, driven by B2C firms, and more specialized firms also hold with an alternate specification (Table 1, Panel B), using a 2002 map of the planned Delhi metro network (which mapped out the plan for line construction up to 2021).Using planned but not built lines as comparison, there is a 67% increase in the number of new firm registrations within 1km of a transit stations (Table 1, Panel B, Col 1).

The increase in firm entry as well as increased size of firms entering contributes to an increase of 56% in the number employed in new firms, relative to the base in the comparison group (Table 1, Panel A, Col 5). While employment in new firms itself is not directly indicative of increase in employment in general, but the key thing to note here is the composition of industries that employment seems to be absorbed by—larger, more specialized, and likely more productive with higher pay and amenities. In the next section, we look at how these new firm entry results translate to the stock of firms and employment.

6 Transit Infrastructure on Firm Composition and Employment

The sustained entry of new, larger, and more specialized firms near transit stations raises the question of whether these patterns translate into changes in the overall stock of firms and employment. While the firm registration data capture new entrants, they do not reveal net effects on employment once accounting for potential firm exits, relocations, or changes in incumbent firm size. To examine these broader impacts on local economic activity, we turn to establishment-level census data that provide comprehensive snapshots of all firms operating in Delhi at three points in time. We first investigate how metro proximity affects the stock of employment and firms, exploiting the Phase 2 expansion (2005-2013). We then examine whether transit-induced improvements in firm commuter market access—capturing firms' ability to draw workers from across the expanded transit network—generate additional employment effects beyond localized station proximity.

6.1 Metro Proximity and Stock of Employment and Firms

Table 2 presents our main results on how metro station proximity affects employment in firms, average firm size, and firm size distribution. Below, we introduce the main empirical strategy, alternate specifications for robustness, and talk through the key results. Across specifications, we find statistically and economically significant increases in employment and shifts toward larger establishments in wards close to metro stations, with no evidence of differential pre-trends.

6.1.1 Empirical Strategy

Our baseline specification exploits the rollout of Phase 2 metro stations, estimating:

$$Y_{pt} = \beta_1(\text{Phase } 2_p \times \text{Year2013}_t) + \beta_2(\text{Phase } 2_p \times \text{Year2005}_t) + \delta_p + \sum_t \gamma_t' \mathbf{X}_p + \epsilon_{pt} \qquad (3)$$

where Y_{pt} represents employment or firm outcomes in ward p at time $t \in \{1990, 2005, 2013\}$. The variable Phase 2_p is an indicator equal to one if ward p's centroid is within a specified distance of a Phase 2 metro station (opened between 2005-2013) and zero otherwise. We exclude wards within 1 kilometer of Phase 1 stations to focus cleanly on Phase 2 expansions. The coefficient β_1 captures the treatment effect of metro access by 2013, while β_2 tests for differential pre-trends between 1990 and 2005.

We include ward fixed effects (δ_p) to control for time-invariant ward characteristics. The vector $\mathbf{X}_p$ contains flexible year fixed effects: log distance to the city center and log distance to the nearest Phase 1 metro station, each interacted with year indicators ($\sum_t \gamma_t' \mathbf{X}_p$). As the Phase 2 expansion is radial in nature, expanding from the city city center to the peripheries and, by the nature of the metro network, the phase 2 lines expand out of the phase 1 lines, we account for differential time trends in wards by distance to the central business district and already existing metro lines. These controls account for differential trends by distance to the city center and exposure to earlier metro infrastructure. All regressions are weighted by ward area. Standard errors are adjusted for spatial correlation using Conley standard errors with a 5-kilometer cutoff.

Table 2 presents estimates across five specifications that vary the treatment definition and comparison group. In Column 1, we define treatment as proximity within 1 kilometer of a Phase 2 station and compare to all areas beyond 2 kilometers from any Phase 2 station. To account for potential spillover effects that could attenuate our estimates, we exclude wards located 1-2 kilometers from metro stations. This specification provides a cleaner comparison between directly treated areas and those sufficiently distant to be unaffected by proximity benefits.

Column 2 narrows the comparison group to only wards within 1 kilometer of not-yet-built stations (while maintaining the 1km treatment definition and 1-2km spillover exclusion). This approach leverages idiosyncratic variation in construction timing among planned stations while ensuring treated and comparison wards share similar observable characteristics related to metro placement decisions. Column 3 expands the treatment definition to wards within 2 kilometers of a Phase 2 station, with the comparison group being wards within 2 kilometers of not-yet-built stations. Column 4 examines whether areas 2-5 kilometers from stations experience effects, using areas beyond 5 kilometers as the comparison group. Finally, Column 5 uses a continuous specification, replacing the binary treatment indicator with log distance to the nearest Phase 2 station, allowing us to estimate the distance gradient of metro effects. The last three specifications help establish the spatial extent of metro impacts.

Across all three outcome panels—employment per capita (Panel A), log average firm size (Panel B), and share of large firms (Panel C)—we apply these five specifications to assess the robustness of our findings.

6.1.2 Findings

Metro station proximity generates substantial increases in employment density and shifts toward larger establishments. In wards within 1km of a Phase 2 station, employment per capita increases by 0.88 workers per person, which is a ten-fold increase over the mean in the comparison group of 8.3 workers per 100 residents in a ward (Table 2, Panel A, Column 1). Narrowing down to only comparing against not-yet-treated wards, there is a similar but imprecise effect size increase given the smaller sample (Table 2, Panel A, Column 2). When we expand the treatment radius to 2 kilometers (Column 3), the effect size is smaller, the effect size is smaller, with a 3.5 times increase in employment per capita, over the base of 18 workers per 100 residents, suggestive of a decay in the employment effects with distance from the metro station. If positive effects near the metro station are primarily coming from relocation or closing of nearby firms in favor of firms closer to the metro station, we would expect negative employment effects in areas further out from the metro station, but still close enough to have such spillovers. We look at wards within 2-5 kilometers from a metro station to test for this; there is a small negative but statistically insignificant decrease of 0.37 workers per capita in these areas, which could be suggestive of some negative spillovers, but with large standard errors (Table 2, Panel A, Column 4). Column 5 further shows that the employment effects of the metro station decay with distance from the station. The employment effects near stations could point to increase in the number of firms and/or increase in the size of firms, which is what we look at next.

Metro access also induces compositional shifts toward larger establishments. Average firm size increases by 42-47% in wards with metro exposure within 1km, over the base of 2.7-3.8 workers per firm (Table 2, Panel B, Columns 1 and 2). The effects are again smaller accounting for a larger area around the metro station, at 33% increase in firm size within a 2km radius, and still positive but insignificant increase of 13.6% in the 2-5km range (Table 2, Panel B, Columns 3 and 4). In Panel C of Table, we specifically look at the share of firms with 10 or more workers, since this is the employment threshold at which more labor regulations start to apply and thus such firms are more likely to be formal as well as relatively more productive (Hasan and Jandoc, 2013; Amirapu and Gechter, 2020; Rajagopalan and Shah, 2024). Within 1km of metro stations, there is a 5-7 percentage point increase in the share of 10+ worker firms, which is a doubling relative the the baseline share (Table 2, Panel C, Columns 1 and 2). This effect again decays with distance 3.5pp-1.3pp increase on expanding to 2 to 5 kilometers away from the metro station (Table 2, Panel C, Columns 3 and 4).

The decay with distance suggests that metro access affects employment through spatial spillovers extending beyond individual stations. Further, the metro is a transit network, not all areas benefit similarly from changes to the metro system. To account for these, we turn the commuter market access based measure in the next section.

6.2 Commuter Market Access and Firms

6.2.1 Empirical Strategy

To assess how metro expansion shapes employment and firms by changing labor market accessibility, we construct commuter market access (CMA) measures which capture how transit infrastructure affects firms' access to workers and residents' access to jobs through the commuting network (Donaldson and Hornbeck, 2016; Tsivanidis, 2023).

CMA Construction

We calculate Firm Commuter Market Access (FCMA) and Resident Commuter Market Access (RCMA) as sufficient statistics of network connectivity, following the method put forth in Tsivanidis (2023). Appendix Section B goes into more detail on the theoretical framework from which this is derived.

FCMA (Ω_j^F) reflects firms' access to residential labor supply, while RCMA (Ω_i^R) captures residents' access to employment opportunities. These measures satisfy the following system:

$$\Omega_i^R = \sum_{j \neq i} d_{ij}^{-\theta} \frac{L_j^F}{\Omega_j^F}, \quad \Omega_j^F = \sum_{i \neq j} d_{ij}^{-\theta} \frac{L_i^R}{\Omega_i^R} \tag{4}$$

where d_{ij} represents commute costs between locations i and j, L_j^F is employment at location j, L_i^R is residential population at location i, and θ is the commute elasticity. We solve this system iteratively, holding population and employment fixed at baseline (1990) levels to isolate the effect of changing transit infrastructure.

We construct commute costs as $d_{ij} = \exp(\kappa \cdot t_{ij})$, where t_{ij} is travel time in minutes from ward i to ward j using public transport and walking and κ maps commute time to commute cost (we assume $\kappa = 0.01$, based on Ahlfeldt et al. (2015). Travel times are computed using the Fast Marching Method (FMM) algorithm on speed rasters that incorporate metro lines, bus routes, and walking speeds; thus the changing metro network changes travel times between i and j. Appendix Section B explains in detail how the travel speed rasters are constructed and how the FMM algorithm is used to get travel times, all used to get the market access values. We assume the commute elasticity to be $\theta = 3.4$, based on Tsivanidis (2023), but also representative of commute elasticities between 2-4 estimated in an Indian urban city (Mumbai) (Gechter and Tsivanidis, 2023). With the two equations and two unknowns (Ω_i^R and Ω_j^F), we retrieve the measure for Firm's Commuter Market Access which we will use to measure the effects on firms and employment in this section.

Regression Specification

Our baseline specification examines whether changes in ln(FCMA) predict employment and firm composition:

$$Y_{pt} = \beta_1(\Delta \ln(\text{FCMA})_p \times \text{Year2013}_t) + \beta_2(\Delta \ln(\text{FCMA})_{\text{Future},p} \times \text{Year2013}_t) + \delta_p + \gamma_t + \epsilon_{pt} \quad (5)$$

where Y_{pt} represents employment or firm outcomes in ward p at time $t \in \{1990, 2005, 2013\}$. The variable $\Delta \ln(\text{FCMA})_p$ measures the log change in firm commuter market access from 2005 to 2013 due to metro expansion. The coefficient β_1 captures the causal effect of improved access of firms to labor. We include $\Delta \ln(\text{FCMA})_{\text{Future},p}$—the anticipated change in FCMA from 2013-2022 based on planned metro extensions—as a falsification test, with β_2 testing whether future accessibility improvements spuriously predict current outcomes. The regressions include ward (δ_p) and year (γ_t) fixed effects. Standard errors are adjusted for spatial correlation using Conley standard errors with a 5-kilometer cutoff.

Table 3 presents results across five outcomes: total employment in a ward(Column 1), average firm size (Column 2), and the share of firms with 10+ workers in manufacturing (Column 3), high-skill professional services (Column 4), and business-to-consumer industries (Column 5). Panel A estimates the baseline specification. Panel B examines whether employment effects vary in high vs. low rent areas (rents are assessed pre-metro expansion by the city municipal government); for this we interact $\Delta \ln(\text{FCMA})$ with an indicator for ward having above-median rent. Panel C interacts $\Delta \ln(\text{FCMA})$ with an indicator for ward being within 1 kilometer of a metro station, assessing whether proximity to station, beyond the metro network's effect on access, has an additional impact on employment.

6.2.2 Findings

In general, increase in firm commuter market access leads to increase in total employment and a shift toward larger firms, especially those require high-skill labor and/or are consumer-facing. However, there is heterogeneity in the types of firms that respond change in FCMA given underlying land rents and proximity to metro station, with larger business-to-consumer firms more likely to locate in higher rent areas an closer to metro stations.

In Panel A of Table 3, we look at the effects of increase in FCMA and to make the estimates more interpretable, the square brackets report the effect size when FCMA increases from the 25th to 75th percentile (which is a 0.123 log point increase). The total employment in a ward goes up by 16.7% (Panel A, Column 1), with the firm-level data showing that average firm size increases by 6.6% (Panel A, Column 2). Interestingly, manufacturing firms do not respond to increase in FCMA from increasing metro access, even though they labor-intensive (Panel A, Column 3). Instead, it is firms that rely on high-skill workers (such as finance, IT, etc.) and direct consumer interactions (such as retail, personal services) that are more likely to locate in areas with increasing FCMA; a move from 25th to 75th percentile in FCMA is associated with a 0.09 percentage point increase in high-skill firms (over a base of 0.6% of the firms being high skill firms with 10+ workers, this is a 15-fold increase) and with a 0.14 percentage point increase in business-to-consumer (B2C) firms (also a 15-fold increase over the base on 0.9% firms). The falsification test using anticipated future accessibility gains (2013-2022) shows no significant relationship with current outcomes, providing confidence that our estimates capture the causal effect of realized infrastructure improvements rather than confounding trends in areas selected for metro expansion.

In Panel B of Table 3, we look at heterogenous effects of change in FCMA by assessed rents per square foot of wards. The rationale here is that we could explain the increase in firm entry we see along metro stations by the ability for firms to now be able to locate in lower rent areas which experience an increase in market access due to metro connectivity. In this case, we would expect that ward with below-median rent should see greater increases in employment with the same increase in FCMA. However, it is very possible that rents themselves increase due proximity to metro (Suri and Cropper, 2024); this is just a suggestive exercise to explore one potential mechanism through which firms might be more likely to enter in peripheral areas with expansion of metro, to take advantage of places with lower rents and building density. Another potential channel could be that these assessed rents are indicative of the socio-economic status of the residents living in the wards, so that could also influence firm composition.

Higher rent is in general associated with lower employment; an increase in FCMA from 25th to 75th percentile in a below-median ward is associated with a 13.8% increase in employment, thus suggesting that there might be more firms or larger firms taking advantage of (initially) cheaper areas which benefit from metro accessibility (Panel B, Column 1). Change in FCMA in above-median rent areas is not significantly different that that in low-rent areas (if anything, the increase in FCMA just seems to counteract the low employment in high-rent areas). Panel B, Column 2 shows that there are no statistically significant differences in how firm size responds to FCMA changes in high vs. low rent areas; this could be indicative of there being heterogeneity by industry. Manufacturing again does not significantly respond to change in FCMA (Panel B, Column 3). In low-rent areas, the 25th-to-75th percentile FCMA improvement generates a 0.03pp increase in likelihood of a large, high-skilled firm (Panel B, Column 4). Further, the interaction terms reveal that high-rent areas experience additional gains concentrated in skilled service firms: the share of high-skill professional firms increases by an additional 0.23 percentage points (a 58% increase over the baseline of 0.4%), while B2C firms increase by an additional 0.18 percentage points (a 26% increase). This heterogeneity suggests that neighborhoods with higher initial land values (which could be proxies for SES of residents) attract more skilled, service-oriented establishments when labor becomes more accessible, potentially reflecting complementarities between skilled workers, valuable locations, and service sector activity.

Proximity to metro stations amplifies the market access effects for especially consumer-facing firms (Panel C, Column 5). The baseline effect of station proximity itself shows a 0.5 percentage point increase in B2C firms, suggesting stations attract customer-facing establishments through foot traffic or amenity effects beyond pure labor accessibility. A 25th-to-75th percentile FCMA improvement in ward that is further within 1km of a metro stations generates an additional 0.23 percentage point increase in the likelihood of a large, B2C firm, a 25-fold increase over the baseline. Thus, beyond the increased labor accessibility that the metro system provides, B2C firms particularly might benefit more from particularly being closer to a metro station.

These results demonstrate that transit infrastructure affects employment not only through direct station proximity but also through broader improvements in firms' access to labor across the entire commuting network. By the nature of the design of the metro systems, the larger increases in firm market access are in the peripheral areas of the city, rather than the already high-market-access core, thus, the effects point to increasing employment, especially from larger firms in peripheral areas. High-skill industries in particular seem to be very responsive to increased market access to labor. Moreover, there are certain firms that have additional gains from being in the vicinity of a metro station. This could be because metro stations also provide other amenities to surrounding areas such as increase in footfall or customer visibility, which retail firms highly value.

Taken together with the locally concentrated effects of metro stations on increased employment in Section 6.1, the changing firm size and industry composition can have implications for *who* benefits in terms of employment, as firms and industries differ in their workforce composition. This is what we will explore in the next section (Section 7).

7 Transit Infrastructure and Heterogenous Employment Effects

The compositional shifts we document—toward larger establishments and consumer-facing services—raise a critical question: which workers benefit from this transit-induced employment growth? This question is particularly salient in the context of Delhi, where female labor force participation rates remain extremely low at 11% (Section 2). While our motivating fact in Section 4.1 shows that FLFP increases more than male participation near metro stations, it remains unclear whether this reflects improved mobility to existing jobs or changes in job opportunities for women. Understanding this distinction has important policy implications: many interventions focused solely on reducing women's commuting costs—including making bus travel free for women—have not led to increased female employment (Chen et al., 2024). We therefore examine heterogeneous employment effects by gender, with particular attention to the role of firm composition.

Figure 5 provides suggestive evidence that the types of firms locating near metro stations differ systematically in their propensity to hire women. The figure plots business types by their average distance to the nearest metro station and the proportion of female workers they employ. Firms that locate closer to metro stations—including beauty parlors, retail stores, banks, and medical services—tend to employ substantially higher shares of women (above the median of 10.6%), while firms locating farther from stations—such as logistics, manufacturing, and construction—employ predominantly male workforces. This pattern suggests that metro-induced compositional shifts may generate differential employment effects not only through improved commuting access but also by attracting firm types that are ex-ante more likely to hire women.

We test this hypothesis by examining employment effects by gender in newly entering firms near metro stations, followed by counterfactual decompositions to quantify how much of the gender gap reflects compositional shifts versus mobility improvements.

7.1 Female Employment in New Firms

Empirical Approach

Our analysis leverages the Phase 3 metro expansion (2011-2024) using a staggered difference-in-differences design following Callaway and Sant'Anna (2021). We partition Delhi into approximately 500-meter radius neighborhoods using H3 hexagonal spatial indexes and aggregate outcomes to 6-month periods. We compare neighborhoods within 1km of Phase 3 stations (treated) to those beyond 1km of any station but within 1km of major roads (comparison), restricting to road-adjacent neighborhoods since metro construction often followed existing road networks. Treatment begins two years before operational opening to capture anticipatory firm entry during the testing phase. We estimate dynamic treatment effects in event time, normalizing the period 2.5 years before opening to zero, and include neighborhood and time fixed effects with standard errors clustered at the neighborhood level (see Section 5.1 for full specification details).

Findings

Figure 6 presents results on heterogenous employment in new firms due to transit infrastructure access: female employment in newly registered firms increases more than male employment following transit station openings. The event study shows that female employment increases by 98% relative to a baseline mean of 3 workers per neighborhood in a 6-month period, compared to a 45% increase for male employment over a baseline of 18 workers. The gender gap in employment in new firms reduces from 1:6 to 1:4 in favor of women with proximity to transit infrastructure.

These gendered employment gains in new firms mask important heterogeneity in which types of firms are hiring women and whether female employment increases are concentrated in traditionally "female-friendly" establishments or extend more broadly. Building on the patterns shown in Figure 5, we now examine which newly entering firms are contributing to female employment growth. Panel A of Table 4 characterizes the types of firms entering near metro stations, while Panel B compares female versus male employment growth across different firm characteristics.

Metro stations attract industries that historically employ more women. Firms in high metro-propensity industries—those that were closer to Phase 1-2 metro stations during the baseline period and tend to be more consumer-facing and skill-intensive—increase by 48.8% relative to a baseline mean of 10.4 firms per neighborhood (Table 4, Panel A, Column 1). In contrast, firms in low metro-propensity industries show a smaller and statistically insignificant 35.3% increase (Table 4, Panel A, Column 2). These patterns indicate that metro stations disproportionately attract the types of industries that have ex-ante employed more women.

Beyond industry composition, metro access induces marked shifts in firm ownership and management toward greater female representation. Female-managed firms increase by 180% over a baseline mean of only 2.6 such firms per neighborhood (Table 4, Panel A, Column 3). This represents a larger percentage increase than the 46.8% growth in male-managed firms over their baseline of 17.4 firms (Table 4, Panel A, Column 4). We also observe an 86.3% increase in firms with at least one female employee, over a baseline of 11.4 firms (Table 4, Panel A, Column 5), while there is no significant change in firms with no female employees (Table 4, Panel A, Column 6). Female-majority firms—where women comprise more than half the workforce—increase by 184.1%, compared to only a 40.9% increase in male-majority firms (Table 4, Panel A, Columns 7-8). Together, these results indicate that metro stations attract not only consumer-facing industries that value female workers, but also catalyze entry of firms with greater female ownership and representation.

However, female employment gains extend well beyond these "female-friendly" firm types. Panel B of Table 4 compares female versus male employment growth across different firm characteristics, revealing that women's employment increases broadly—including in male-managed, larger, and more productive establishments. Female employment in general industries (not specifically targeting female clientele) increases by 95.8% over a baseline mean of 2.6 female workers per neighborhood (Table 4, Panel B, Column 1). While female employment in female-oriented industries also increases by 93.0%, this effect is imprecise due to the small baseline of 0.2 workers (Table 4, Panel B, Column 2).

Female employment increases even in male-managed firms, where female workers grow by 92.8% over a baseline of 2.1 workers, compared to only a 35.7% (statistically insignificant) increase for male workers in the same firms (Table 4, Panel B, Columns 3-4). This suggests that male-managed firms near metro stations are actively hiring more women, rather than female employment gains being driven solely by female-managed establishments.

Metro access also enables women to access employment in larger, more formal establishments. In firms with 10 or more workers—establishments that are more likely to be formal and productive given India's employment-based regulatory thresholds—female employment increases by 114.6%, translating to 1.6 additional female workers per neighborhood over a baseline mean of 1.4 workers (Table 4, Panel B, Column 5). In contrast, male employment in such firms increases by only 44.4% (statistically insignificant) over a much larger baseline of 7.2 workers (Table 4, Panel B, Column 6). Given that male employment in 10+ worker firms was five times higher than female employment at baseline, these differential growth rates narrow the gender gap to approximately 2-3:1.

Finally, women also gain access to more differentiated and potentially productive firms. In establishments where the owner and manager are different individuals—a marker of organizational complexity and scale—female employment increases by 82.0%, compared to a 54.2% increase for male employment (Table 4, Panel B, Columns 7-8). This difference is notable given that both coefficients are statistically significant, suggesting that metro access enables women to access not just more jobs, but higher-quality jobs in more sophisticated firms.

Taken together, these results demonstrate broad-based female employment gains. Metro stations attract consumer-facing industries and female-managed firms that disproportionately employ women, but female employment increases extend far beyond these firm types. Women gain employment in male-managed firms, large formal establishments, and productive differentiated firms—precisely the types of jobs that offer better pay, stability, and career prospects. The next section explores how this entry of firms translates in the heterogenous employment effects in the stock, as well as how much of the effects on women are driven by the change in firm composition.

7.2 Decomposing Workplace Employment: Within vs. Across Firm Composition

Having established that metro proximity generates increases in female employment alongside compositional shifts toward larger establishments and female-intensive industries, we now investigate the mechanisms underlying these employment gains. Specifically, we decompose the increase in female employment into two channels: changes in firm composition (metro stations attracting firm types that hire more women) versus changes in hiring behavior within existing firm types (firms near metro stations employing more women conditional on firm characteristics).

Table 5 presents the gender-disaggregated employment effects alongside a counterfactual decomposition of female employment changes. Male employment per capita increases by

149 workers per 100 residents in treated wards, representing a ten-fold increase relative to the baseline mean of 14.4 workers, though this effect is imprecise (Column 1). Female employment per capita increases by 29 workers per 100 residents ($p < 0.1$), more than a 20-fold increase over the low baseline mean of just 1.36 female workers per 100 in the population (Column 2)

To understand whether these employment gains reflect changes in which types of firms locate near metro stations versus changes in hiring practices within firm types, we conduct a decomposition. We express female employment in ward p at time t as:

$$\text{FemEmp}_{p,t} = \sum_s \text{Emp}_{s,p,t} \times \text{FemShare}_{s,p,t} \tag{6}$$

where s indexes firm type, defined as the interaction of industry (approximately 80 classifications) and firm size bin. Female employment in a ward equals the sum across all firm types of total employment in each firm type multiplied by the female share of the workforce in that firm type. This decomposition allows us to separate compositional effects (changes in $\text{Emp}_{s,p,t}$ across firm types) from within-type effects (changes in $\text{FemShare}_{s,p,t}$).

We construct two counterfactuals. First, we fix the distribution of employment across firm types at baseline (1990) proportions while allowing total employment and female shares within firm types to change: $\text{Emp}_{s,p,t} = \frac{\text{Emp}_{s,p,1990}}{\text{Emp}_{p,1990}} \times \text{Emp}_{p,t}$. This isolates within-firm-type changes in female employment, such as firms near metro stations hiring more women conditional on industry and size. Second, we fix the female share within each firm type at its baseline level: $\text{FemShare}_{s,p,t} = \text{FemShare}_{s,1990}$. This isolates the contribution of compositional change—if metro access shifts employment toward firm types that historically employed more women, female employment would increase even absent any change in hiring behavior within firm types.

(Table 5, Column 3) presents results when we hold the firm type distribution constant at 1990 levels, allowing only within-firm-type changes in female shares. The estimated treatment effect is 6.9 workers per 100 residents, but this is statistically insignificant. (Table 5, Column 4) presents results when we hold female shares within firm types constant at 1990 levels, allowing only compositional change. The estimated treatment effect is 10.4 workers per 100 residents ($p < 0.1$), which is closer but yet smaller than the true effect in Column 2.

These results indicate that compositional change is a larger driver of female employment gains near metro stations, though neither channel alone fully accounts for the observed effect, suggesting that the two mechanisms interact to amplify female employment gains.

7.3 Decomposing Residence Employment through Exact Hat Algebra

The decomposition in Section 7.2 establishes that female workplace employment increases primarily reflect compositional shifts across firm types—specifically, toward larger establishments and consumer-facing industries that employ more women ex-ante—rather than changes in hiring behavior within firm types. We now examine how these workplace employment changes translate into residential employment outcomes through the commuting network. Specifically, we ask: to what extent do observed increases in female labor force participation near metro stations reflect (i) reduced commuting costs enabling access to distant jobs, (ii) increased employment levels in accessible locations, or (iii) shifts in the composition of jobs across firm size and industry?

7.3.1 Connecting Workplace to Residence Employment

Workplace employment of gender g in location j can be expressed in terms of firm composition:

$$L_{gj} = \sum_{s} \text{Emp}_{s,j} \times \text{Share}_{g,s,j} \tag{7}$$

where s indexes firm type (industry $\times$ size bin), $\text{Emp}_{s,j}$ denotes total employment in firm type s at location j, and $\text{Share}_{g,s,j}$ is the share of gender g workers in that firm type. This expression makes explicit that workplace employment aggregates across the distribution of firm types.

Labor market clearing requires that workplace employment equals the sum of workers commuting from all residential locations:

$$L_{gj} = \sum_{i} \pi_{ij|i} \cdot R_{gi} \tag{8}$$

where R_{gi} denotes the number of employed residents of gender g living in location i, and $\pi_{ij|i}$ represents the probability that a worker residing in i commutes to workplace j. Following Ahlfeldt et al. (2015), this commuting probability derives from workers' optimal location choices under heterogeneous idiosyncratic preferences distributed Fréchet:

$$\pi_{ij|i} = \frac{\left(\frac{w_j}{d_{ij}}\right)^{\theta_g}}{\sum_{\ell \in \mathbb{N}} \left(\frac{w_\ell}{d_{i\ell}}\right)^{\theta_g}} \tag{9}$$

where w_j denotes the wage at workplace location j, $d_{ij} = \exp(\kappa \cdot t_{ij})$ represents iceberg commuting costs as a function of travel time t_{ij} (in minutes), and θ_g is the commute elasticity governing gender g's sensitivity to commuting costs. The denominator captures residents' total access to employment opportunities across all workplace locations, weighted by wages and commuting costs.

7.3.2 Exact Hat Algebra Framework

To quantify the relative importance of commuting costs, employment levels, and employment composition in generating residential employment changes, we employ an exact hat algebra approach following Heblich et al. (2020). This method enables counterfactual analysis in complex spatial models without requiring full structural estimation of all parameters. Instead, we exploit information about baseline equilibrium outcomes and key elasticities to perform comparative statics between observed states.

We rewrite the labor market clearing condition in terms of relative changes between a baseline year (1990) and counterfactual year (2013):

$$R_{gi,1990} \cdot \hat{R}_{gi} = \sum_j \pi_{ij|i,1990} \cdot L_{gj,1990} \cdot \hat{L}_{gj} \tag{10}$$

where $\hat{x} \equiv x_{2013}/x_{1990}$ denotes the relative change in variable x between baseline and counterfactual years. The baseline commuting probability $\pi_{ij|i,1990}$ is constructed using:

$$\pi_{ij|i,1990} = \frac{\left(\frac{w_{j,1990}}{d_{ij,1990}}\right)^{\theta_g}}{\sum_{\ell \in \mathbb{N}} \left(\frac{w_{\ell,1990}}{d_{i\ell,1990}}\right)^{\theta_g}}$$

The counterfactual commuting probability responds to changes in both wages and commuting costs:

$$\pi_{ij|i}^{\text{CF}} = \frac{\left(\frac{w_{j,1990} \cdot \hat{w}_j^{\text{CF}}}{d_{ij,1990} \cdot \hat{d}_{ij}^{\text{CF}}}\right)^{\theta_g}}{\sum_{\ell \in \mathbb{N}} \left(\frac{w_{\ell,1990} \cdot \hat{w}_\ell^{\text{CF}}}{d_{i\ell,1990} \cdot \hat{d}_{i\ell}^{\text{CF}}}\right)^{\theta_g}}$$

To incorporate agglomeration effects, we allow wages to respond to employment changes according to:

$$\hat{w}_j^{\text{CF}} = \left(\hat{L}_j^{\text{CF}}\right)^{\gamma}$$

where γ represents the agglomeration elasticity. Following estimates from the urban economics literature, we set $\gamma = 0.05$, implying that a 10% increase in employment raises wages by 0.5%. This captures productivity gains from increased employment density through knowledge spillovers, labor market pooling, and input-output linkages, while remaining conservative relative to estimates for developed economies (Heblich et al., 2020).

7.4 Parameter Estimation

Two key parameters require estimation: the time-to-cost parameter κ and the commute elasticity θ_g for each gender $g \in \{\text{male}, \text{female}\}$.

Iceberg Cost Parameter (κ): We set $\kappa = 0.01$ following Ahlfeldt et al. (2015), which maps travel time in minutes to utility costs through the commuting cost function $d_{ij} = \exp(\kappa \cdot t_{ij})$. This value is standard in the urban spatial economics literature (Zárate, 2022; Tsivanidis, 2023).

Commute Elasticity (θ): We estimate gender-specific commute elasticities using a two-step procedure that combines information from mobile phone mobility data and Population Census 2011 commute distance distributions.

First, we estimate the general population commute elasticity using a gravity equation on origin-destination flows from mobile phone location data covering Delhi in 2019. The gravity specification relates bilateral commuting flows to travel times:

$$\Pr(j|i,w) = \exp\left(-\beta \cdot t_{ij} + \gamma_{iw} + \gamma_{jw}\right) \tag{11}$$

where $\Pr(j|i,w)$ is the conditional probability that a worker residing in origin i commutes to destination j in week w, t_{ij} is travel time in minutes, and γ_{iw}, γ_{jw} are origin-week and destination-week fixed effects that absorb all time-varying location-specific characteristics. The coefficient β is the semi-elasticity of commuting flows with respect to travel time, capturing how commuting probability responds to changes in travel costs. Under the Frechet commuting model with iceberg costs $d_{ij} = \exp(\kappa \cdot t_{ij})$, this semi-elasticity comprises two components: $\beta = \kappa \cdot \theta$, where κ translates travel time into utility costs and θ translates utility costs into commuting flows. Given $\kappa = 0.01$ from the literature, we recover $\theta_{\text{general}} = \beta/\kappa$.

Using Poisson pseudo-maximum likelihood estimation (to account for zeros in bilateral trips) on over 4 million origin-destination-week observations across 281 hexagonal neighborhoods (H3-7 resolution, approximately 2 km radius), we estimate $\hat{\beta} = -0.0401$ (SE: 0.0032, significant at the 99% confidence level), yielding $\theta_{\text{general}} = 4.01$.

Second, we estimate the relative commute elasticity between genders using Population

Census 2011 data on commute distance distributions by gender. While these data lack origin-destination flows, the aggregate distance distributions reveal differential distance sensitivity. We estimate gender-specific distance decay rates from the empirical commute distance distributions:

$$\ln(\text{Share}_{d,g}) = \alpha_g + \delta_g \cdot \text{Distance}_d + \epsilon_{d,g} \tag{12}$$

where $\text{Share}_{d,g}$ is the share of gender g workers commuting distance d, and δ_g captures the rate at which commuting probability declines with distance. The ratio of these decay rates provides the relative commute elasticity: $\theta_{\text{female}}/\theta_{\text{male}} = |\delta_{\text{female}}|/|\delta_{\text{male}}| = 1.18$.

Finally, we anchor the relative elasticity to the general population estimate. The mobile phone data represent a population-weighted average of gender-specific elasticities:

$$\theta_{\text{general}} = \alpha_{\text{female}} \cdot \theta_{\text{female}} + (1 - \alpha_{\text{female}}) \cdot \theta_{\text{male}} \tag{13}$$

where $\alpha_{\text{female}} = 0.147$ is the female share of the labor force from Census 2011. Given the relative ratio $\theta_{\text{female}}/\theta_{\text{male}} = 1.18$ and $\theta_{\text{general}} = 4.01$, we solve for levels: $\theta_{\text{male}} = 3.91$ and $\theta_{\text{female}} = 4.61$.

Women are 17.9% more distance-sensitive than men, consistent with gender differences in commuting elasticities estimated in other contexts: Velásquez (2023) finds a 20% difference for single men and women using a spatial equilibrium model in Lima, Peru, while Le Barbanchon et al. (2021) estimates a similar magnitude using a job search model with job application data in France.

Appendix Section C.1 provides complete methodological details on the gravity equation estimation, distance decay calculation, and anchoring procedure.

7.4.1 Counterfactual Scenarios

With estimated parameters (κ, θ_g) and observed baseline conditions $(R_{gi,1990}, L_{gj,1990}, d_{ij,1990})$, we construct four counterfactual scenarios to isolate the contribution of each mechanism:

Counterfactual 1: Commuting Cost Changes Only

We impose the observed reduction in commuting costs from metro expansion ($\hat{d}_{ij}^{CF1} = d_{ij,2013}/d_{ij,1990}$) while holding employment constant at baseline levels ($\hat{L}_{gj}^{CF1} = 1$). This scenario isolates the effect of improved accessibility on residential employment, abstracting from any employment response. Wages remain at baseline ($\hat{w}_j^{CF1} = 1$) since employment does not change.

Counterfactual 2: Commuting Cost and Employment Level Changes

We impose both the observed reduction in commuting costs ($\hat{d}_{ij}^{CF2} = d_{ij,2013}/d_{ij,1990}$) and the observed total employment changes by gender and location ($\hat{L}_{gj}^{CF2} = L_{gj,2013}/L_{gj,1990}$). However, we hold the composition of employment across firm types fixed at baseline proportions:

$$L_{gj,2013}^{CF2} = \sum_s \text{Emp}_{s,j,1990} \times \text{Share}_{g,s,j,2013} \tag{14}$$

This counterfactual captures the combined effect of reduced commuting costs and increased employment levels, but shuts down the compositional shift toward firm types that employ more women. Wages adjust according to $\hat{w}_j^{CF2} = (\hat{L}_j^{CF2})^{0.05}$.

Counterfactual 3: Employment Composition Changes Only

We hold commuting costs constant at baseline ($\hat{d}_{ij}^{CF3} = 1$) but allow employment to change in both level and composition to match observed 2013 values ($\hat{L}_{gj}^{CF3} = L_{gj,2013}/L_{gj,1990}$). This scenario isolates the employment channel—encompassing both increases in total jobs and shifts toward firm types employing more women—while abstracting from improved commuting access. Wages adjust according to $\hat{w}_j^{CF3} = (\hat{L}_j^{CF3})^{0.05}$.

Counterfactual 4: Spatial Distribution of Employment Fixed

We impose the observed reduction in commuting costs ($\hat{d}_{ij}^{CF4} = d_{ij,2013}/d_{ij,1990}$) and allow total city-wide employment to increase to 2013 levels, but hold the spatial distribution of employment constant at baseline proportions:

$$L_{gj}^{CF4} = \frac{L_{g,2013}}{L_{g,1990}} \times \frac{L_{gj,1990}}{L_{g,1990}} \tag{15}$$

where $L_{g,t} = \sum_j L_{gj,t}$ denotes total employment of gender g in year t. This counterfactual captures the effect of reduced commuting costs and aggregate employment growth, but abstracts from the spatial redistribution of jobs documented in Section 6.2. Wages adjust according to $\hat{w}_j^{CF4} = (\hat{L}_j^{CF4})^{0.05}$.

For each counterfactual scenario, we solve for the implied residential employment changes $\hat{R}_{gi}^{CF}$ that satisfy the labor market clearing condition. We then compare these counterfactual predictions to the observed changes $\hat{R}_{gi}^{\text{actual}} = R_{gi,2011}/R_{gi,1991}$. These comparisons reveal which channels account for the largest share of observed residential employment growth, particularly for women in peripheral areas.

7.4.2 Interpreting the Counterfactuals

The counterfactual comparisons provide evidence on three distinct mechanisms through which metro expansion affects residential employment. Counterfactual 1 captures the traditional channel emphasized in the transportation literature: reduced commuting costs enable workers to access distant jobs without requiring changes in employment levels or composition. If this scenario explains a large share of observed residential employment changes, it would suggest that metro's primary effect operates through improved worker mobility.

Counterfactual 2 adds employment level changes while holding firm composition fixed. Comparing CF2 to CF1 reveals the importance of employment growth per se, holding constant the types of firms entering. If CF2 substantially outperforms CF1, it indicates that the creation of new jobs—rather than merely improved access to existing opportunities—drives residential employment gains.

Counterfactual 3 isolates the employment channel (level and composition) while abstracting from commuting improvements. Strong performance of CF3 relative to CF1 would demonstrate that changes in where jobs locate and what types of firms enter matter more than reduced commuting costs. This would be consistent with our findings in Sections 5 and 6 that metro stations attract larger, consumer-facing establishments that employ more women.

Counterfactual 4 examines whether the spatial redistribution of employment documented in Section 6.2 contributes meaningfully beyond aggregate employment growth. If CF4 performs substantially worse than CF3, it would indicate that the creation of new economic hubs in peripheral areas—rather than proportional employment growth everywhere—explains residential employment patterns.

The comparison across these scenarios directly tests whether firm compositional shifts, which we showed account for the majority of female workplace employment gains in Section 7.2, also explain residential employment outcomes through the commuting network. If CF2 performs substantially worse than CF3, it would provide strong evidence that the types of firms entering near metro stations matter as much as the quantity of jobs created.

8 Conclusion

Understanding how firms respond to transportation infrastructure is critical for predicting the full employment effects of transit investments. While conventional analysis focuses on how infrastructure improves workers' access to existing opportunities, our findings demonstrate that transit also fundamentally reshapes the local economic landscape through firm location decisions. Using the phased rollout of Delhi Metro as a natural experiment, we document that public transit operates through two distinct channels: reducing commuting costs for workers and inducing compositional shifts in local firm structure.

Metro access generates 2.5 times more firm entries relative to baseline, with these new establishments being systematically larger and more specialized than incumbents. Employment per capita increases ten-fold in neighborhoods within one kilometer of stations, driven by shifts toward larger establishments and consumer-facing businesses. Average firm size increases 42–47% over baseline, while the share of firms with ten or more workers—the threshold at which formal labor regulations typically apply—doubles from baseline levels. These compositional changes reflect firms' responses to metro-induced increases in foot traffic and consumer access rather than simply relocating existing economic activity.

The firm composition changes generate heterogeneous employment effects across worker types. Employment increases differ by worker characteristics that correlate with propensity to work in different firm types and sensitivity to commuting costs. Our counterfactual decompositions reveal that compositional shifts—the changing mix of firm types locating near metro stations—account for the majority of differential employment effects across worker groups. When we hold the distribution of firm types constant at pre-metro levels, differential employment effects decline by approximately two-thirds, indicating that the types of firms attracted to transit access determine which workers benefit most from infrastructure investment.

These effects concentrate in peripheral areas of the city, where metro access transforms previously disconnected neighborhoods into new economic hubs. Firm entry near metro stations follows a spatial gradient: while transit proximity yields modest advantages in core areas, the differential impact intensifies with distance from the central business district. Beyond ten kilometers from the city center—where neighborhoods without transit access experience minimal firm entry—transit-adjacent neighborhoods maintain entry rates comparable to levels observed in the urban core. This spatial pattern indicates that transportation infrastructure enables economic decentralization by making peripheral locations viable for firm activity that would otherwise concentrate near the city center.

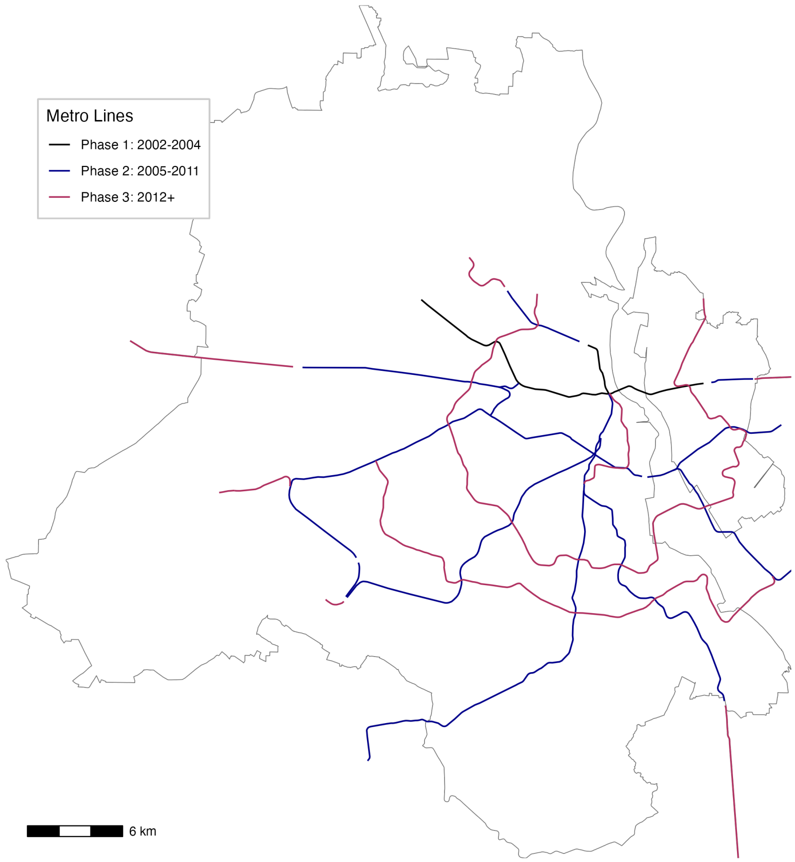
Methodologically, this study demonstrates the value of combining establishment-level data at fine spatial scales with variation in infrastructure timing. Our approach reveals mechanisms that would remain obscured in aggregate analyses: the systematic sorting of firms by type and size around transit stations, the differential employment responses across worker characteristics, and the spatial redistribution of economic activity toward the urban periphery. These patterns highlight that transportation infrastructure effects extend well beyond traditional commuting benefits captured in most policy evaluations.

The dual-channel framework developed here has important implications for transportation and urban development policy. First, our decomposition results indicate that policies focused solely on improving worker mobility—such as subsidized fares or dedicated transport services for specific groups—may generate limited employment gains if they do not also create demand for labor from employers likely to hire particular types of workers. Second, the concentration of firm entry and employment effects near metro stations suggests that complementary land-use policies—such as relaxing zoning restrictions or promoting mixed-use development around transit nodes—could amplify the employment benefits of infrastructure investment by facilitating the entry of larger, more formal establishments. Third, the peripheral concentration of effects demonstrates that infrastructure investments can serve as catalysts for spatial economic transformation, creating employment opportunities in areas distant from traditional economic centers.

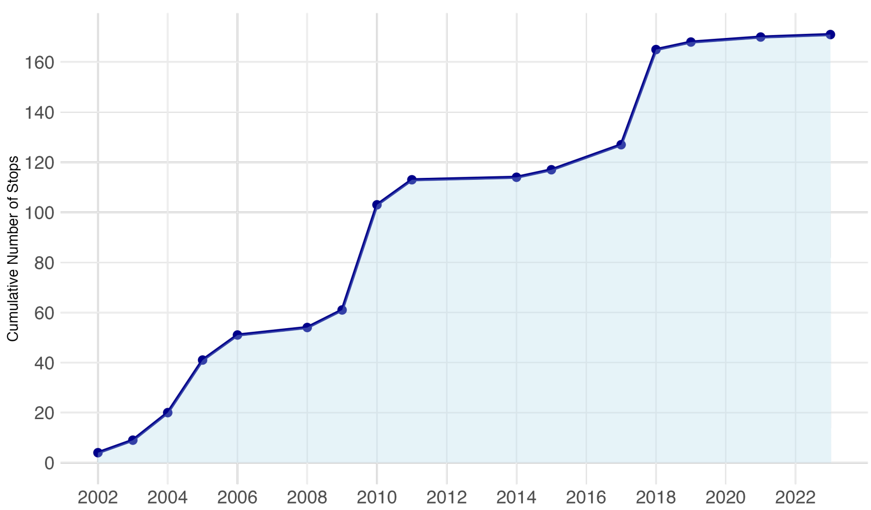
Our findings contribute to understanding how transportation infrastructure shapes local economic development. Rather than simply connecting workers to distant opportunities in city centers, public transit can fundamentally transform neighborhoods by attracting firms that generate new employment locally. This distinction matters for infrastructure planning in rapidly urbanizing developing countries, where investments must both improve connectivity and generate employment growth in peripheral areas experiencing population expansion. Transportation infrastructure, when combined with appropriate complementary policies, can function as an economic development tool that moves opportunity closer to where people live.

Figure 1: Delhi Metro Network Expansion

(a) Spatial expansion by phase



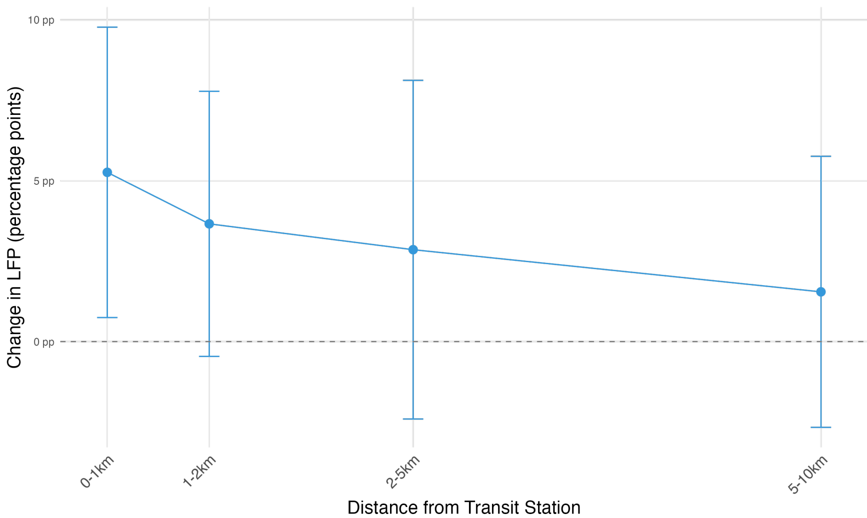
(b) Cumulative transit stations by year



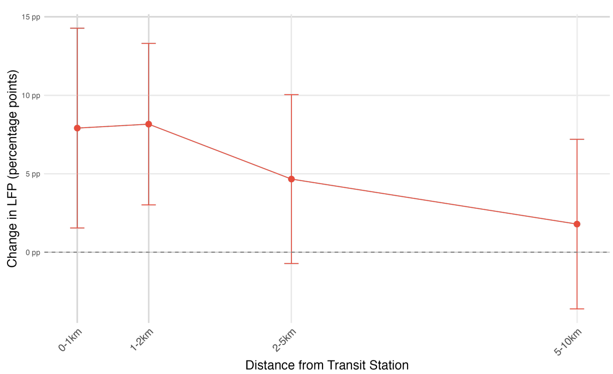
Notes: Transit expansion from Section 2.2. Panel (a) displays the complete Delhi Metro network as of 2024, showing all operational lines and stations. Panel (b) presents the cumulative number of metro stations over time across the three construction phases.

Figure 2: Labor Force Participation by Distance to Transit Stations

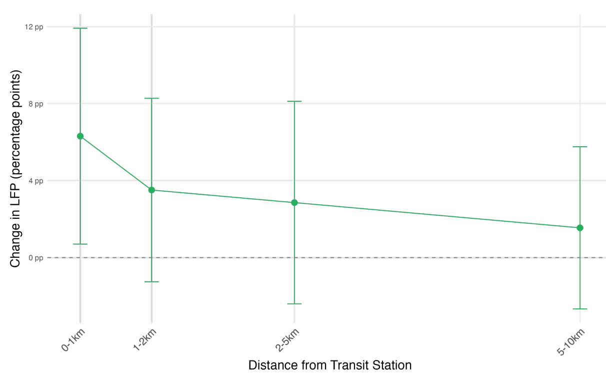
(a) Overall labor force participation



(b) Female labor force participation

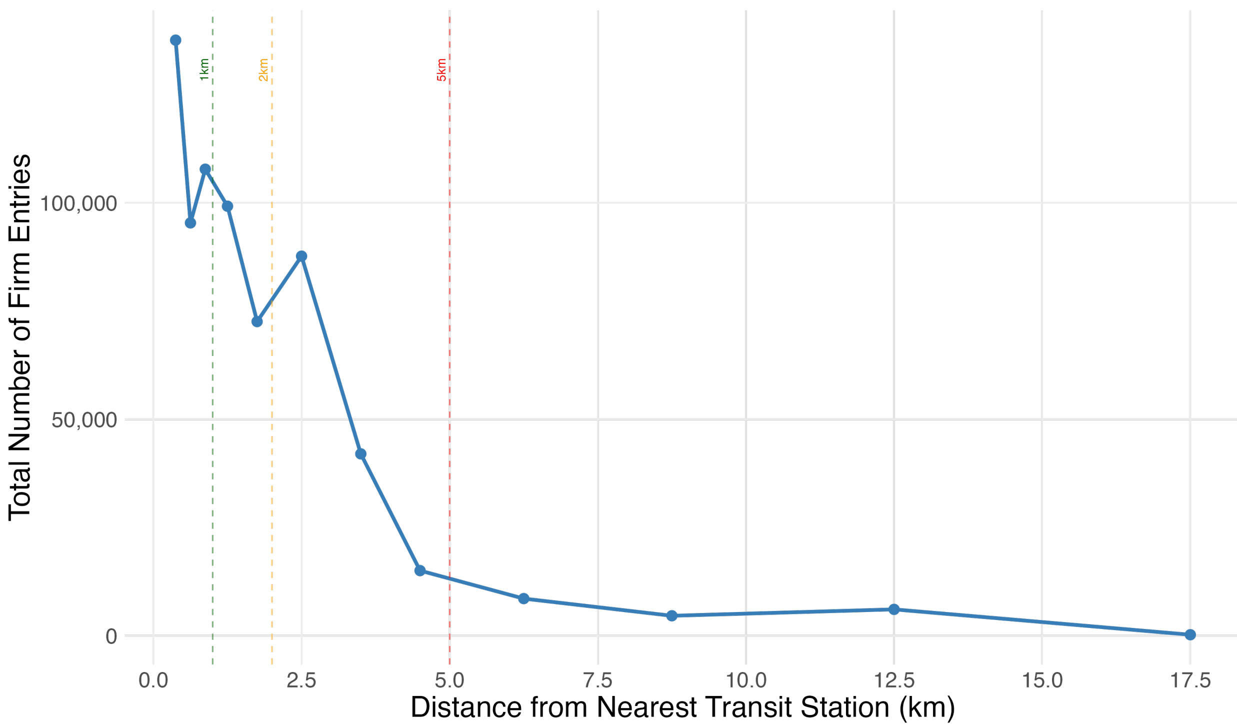


(c) Overall LFP within 5km of CBD



Notes: Estimates of β_1 from equation (1) from Section 4.1, at varying distance intervals from transit stations. Data: Population Census 1991, 2001, and 2011 at the harmonized 2011 ward level. Sample: wards that eventually received metro stations by 2023. The outcome variable is the proportion of the population in the labor force. Panel A presents estimates for the full sample. Panel B restricts the outcome to female labor force participation. Panel C restricts the sample to wards within 5 kilometers of the central business district. Each point represents the estimated treatment effect ($\text{Treated}_p \times \text{Year2011}_t$) for wards within the specified distance from a Phase 1 or Phase 2 metro station (operational by 2011), with comparison wards beyond that distance. Regressions include ward and year fixed effects, and are weighted by 1991 ward population. Vertical bars represent 95% confidence intervals based on Conley standard errors with a 5km bandwidth to account for spatial correlation. The first metro line became operational in 2002.

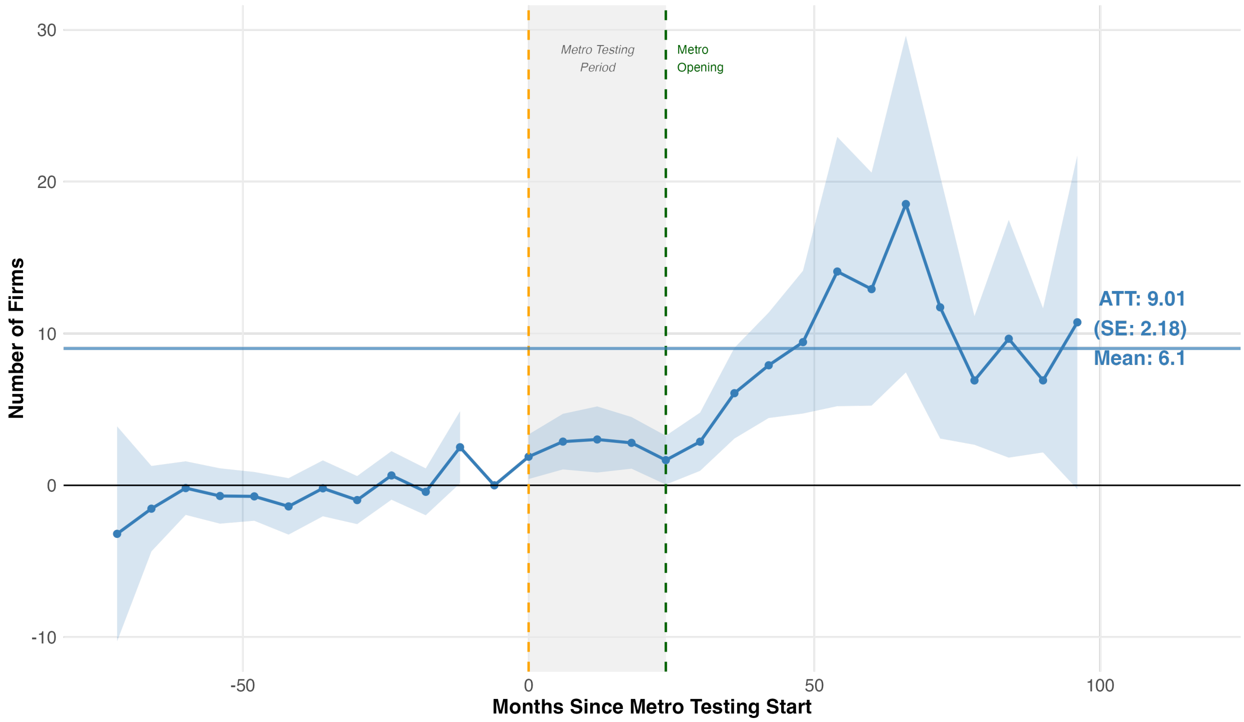
Figure 3: Number of new firm registrations by distance to transit stations



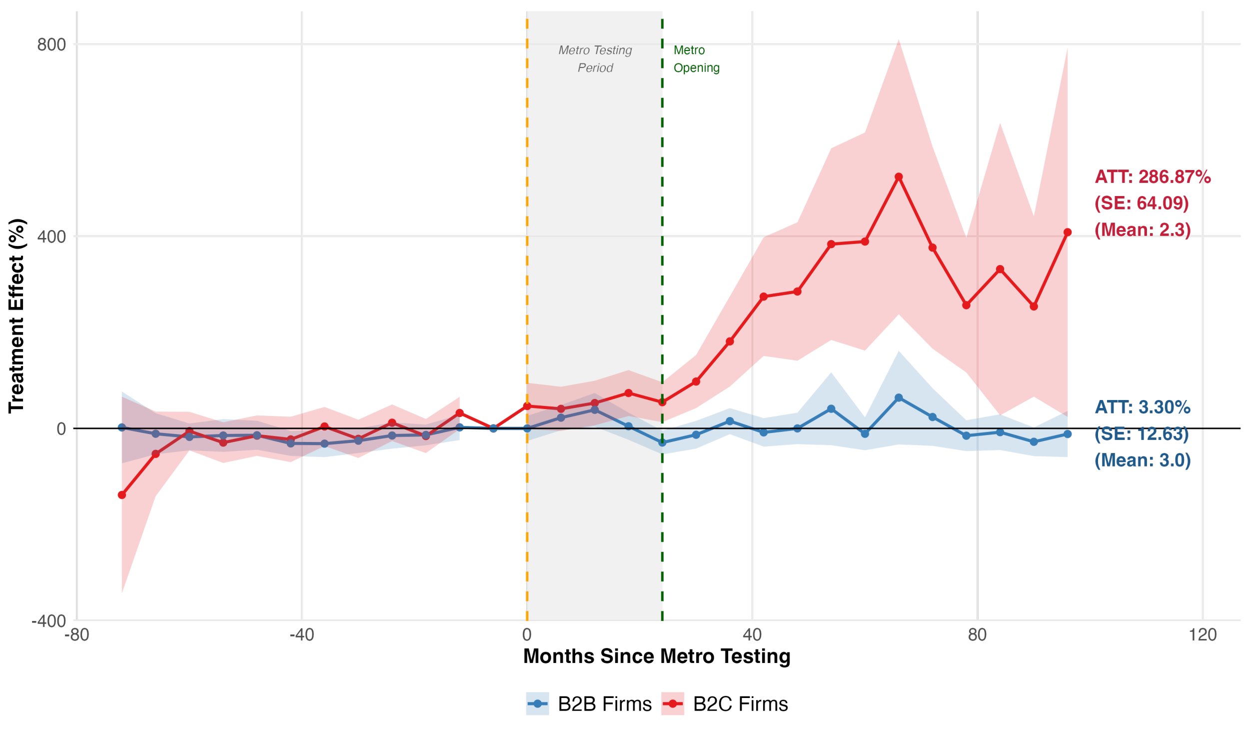
Notes: Descriptive statistics of Registrations of Shops and Establishments Data (2011-2024) from Section 4.2. Sample includes all H3 hexagonal neighborhoods. Vertical dashed lines indicate distance thresholds at 1km, 2km, and 5km. Distance measured from hexagon centroid to nearest operational transit station.

Figure 4: Firm entries near transit stations

(a) Overall increase in firm entries

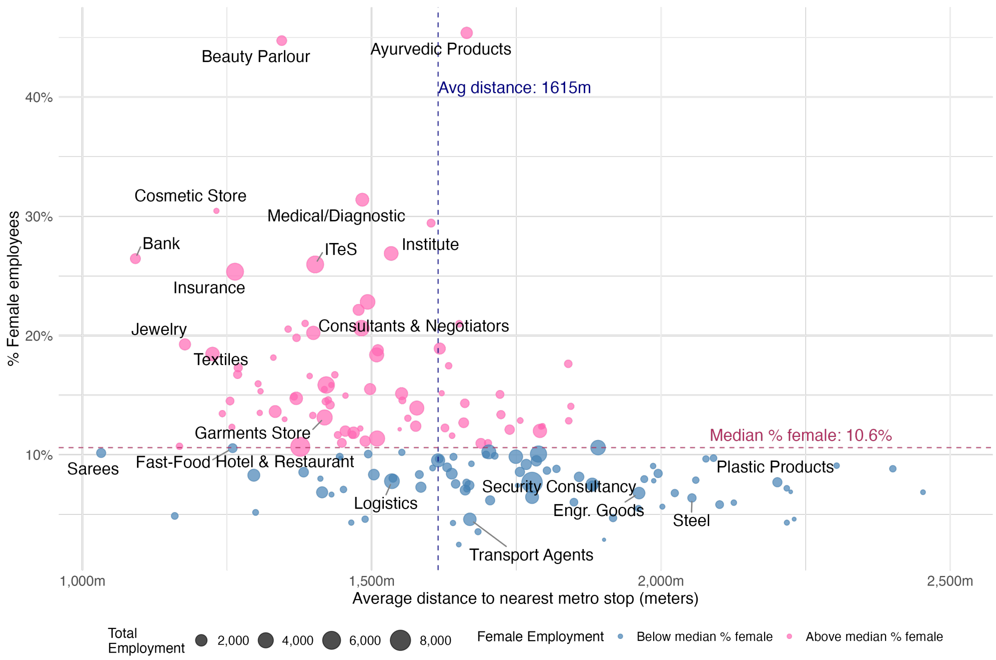


(b) Firm entry by consumer-type



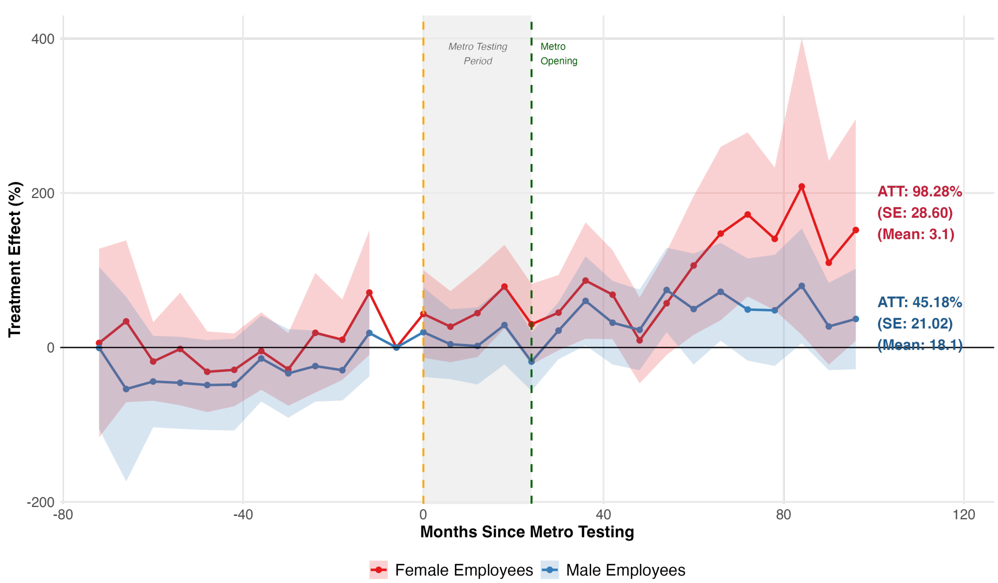
Notes: Staggered DiD estimates from (Callaway and Sant'Anna, 2021) from Section 5.2. Sample: H3 hexagons within 1km of major roads, 6-month periods, 2011-2024. Treatment: transit station within 1km; Comparison: never/not-yet-treated. Shaded: anticipation (4 periods). SE clustered by neighborhood. Panel A: Number of firms. ATT = increase in num. firms; Panel B: Demeaned number of B2C vs. B2B firms (divided by baseline mean of comparison group, 2011-2016). ATT = % change from baseline.

Figure 5: % of industry female, by average distance to transit station



Notes Motivating fact on transit presence and female workforce from Section 7. Data: Shop and establishment registrations 2011-2013. Sample includes industries with ≥50 firms and ≥100 total employees from this time period. Each point represents one industry.

Figure 6: Firm Entry near Transit Stations: Female vs. Male Employment



Notes: Estimates from staggered difference-in-differences (Callaway and Sant'Anna, 2021) from Section 7.1. Data: Shop and establishment registrations 2011-2024; H3 hexagons within 1km of major roads, 6-month periods. Treated: within 1km of transit station; Comparison: never/not-yet-treated neighborhoods. Outcomes normalized by dividing by baseline mean (average across comparison neighborhoods in periods 1-10, 2011-2016); ATT represents % change from baseline. Mean gives is the baseline mean of the original variable. Shaded: anticipation period (4-periods). SE clustered by neighborhood.

Table 1: Entry of firms

	Num. Firms (1)	Firm Size (2)	Num. Spl (3)	Num. B2C (4)	Num. High-Skill (5)	Total Emp (6)
Panel A: Sample - ≤1km from major roads						
ATT	8.91***	0.79***	2.05***	6.66***	-0.00	12.91**
Effect size (%)	143.5%	19.3%	94.6%	280.6%	-0.4%	55.9%
	(2.09)	(0.27)	(0.51)	(1.53)	(0.15)	(5.02)
Mean	6.21	4.12	2.17	2.37	0.75	23.08
Pre-trend p-val	0.221	0.141	0.184	0.466	0.181	0.263
Treated Nbhds	148	148	148	148	148	148
Total Nbhds	633	633	633	633	633	633
Observations	17091	17091	17091	17091	17091	17091
Panel B: Sample - ≤1km from ever-planned transit station						
ATT	6.61*	0.40	1.66**	5.56**	-0.03	7.79
Effect size (%)	67.2%	10.1%	47.2%	148.8%	-2.7%	21.0%
	(3.54)	(0.33)	(0.75)	(2.26)	(0.19)	(6.68)
Mean	9.83	3.99	3.51	3.74	1.14	37.10
Pre-trend p-val	0.179	0.010	0.659	0.418	0.289	0.731
Treated Nbhds	68	68	68	68	68	68
Total Nbhds	193	193	193	193	193	193
Observations	5211	5211	5211	5211	5211	5211
Panel C: Sample - All (excluding always-treated)						
ATT	13.28***	0.58**	2.40***	9.74***	0.15	16.58***
Effect size (%)	326.1%	14.5%	169.3%	626.8%	30.5%	114.3%
	(2.04)	(0.23)	(0.47)	(1.38)	(0.13)	(4.43)
Mean	4.07	3.97	1.42	1.55	0.49	14.50
Pre-trend p-val	0.000	0.310	0.004	0.019	0.287	0.001
Treated Nbhds	178	178	178	178	178	178
Total Nbhds	1573	1573	1573	1573	1573	1573
Observations	42471	42471	42471	42471	42471	42471

OLS estimates of metro station proximity effects on employment and firms from Section 5.2. Data: Economic Census from years 1990, 2005, 2013. Treatment: proximity of city wards to Phase 2 metro stations (built 2005-2013). Y2013 interactions = treatment effects; Y2005 = pre-trend tests. Column (1): treated = ≤1km from station, comparison = all areas >2km; sample excludes wards 1-2km from station to account for spillovers. Column (2): treated = ≤1km from station, comparison = ≤1km from unbuilt station; sample excludes wards 1-2km from station to account for spillovers. Column (3): treated = ≤2km from station, comparison = ≤2km from unbuilt station. Column (4): treated = 2-5km from Phase 2 station, comparison = >5km. Column (5): log distance to station specification using all areas. All samples exclude wards ≤1km of Phase 1 metro stations. The mean of the dependent variable is the 1990 value of the respective comparison group (for Column (5) - the sample for the mean is above median distance wards); note that mean reported in Panel B is for average employment. The means are taken from respective comparison samples winsorized at the 95th percentile to account for outliers. All regressions include ward, year×log(distance to CBD), and year×log(distance to Phase 1 metro station) fixed effects. Regressions are weighted by ward area. 5km Conley standard errors in (parentheses). $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 2: Effect of Metro Proximity on Stock of Employment and Firms

	(1)	(2)	(3)	(4)	(5)
	Panel A: Outcome = Employment/Population				
Treated=	Metro≤1km	Metro≤1km	Metro≤2km	Metro in 2-5km	Log(Dist)
Comparison=	all	not-yet-built	not-yet-built	>5km	all
Treated×Y2013	0.884*	1.772	0.660*	-0.037	-0.177
	(0.483)	(1.086)	(0.395)	(0.082)	(0.139)
Treated×Y2005	0.252	0.660	0.176	0.054	-0.042
	(0.201)	(0.597)	(0.198)	(0.071)	(0.039)
DepVar Mean (1990)	0.083	0.157	0.184	0.035	0.064
Obs	901	354	675	651	1,139
Wards	349	143	274	248	445
	Panel B: Outcome = Log Average Employment				
Treated=	Metro≤1km	Metro≤1km	Metro≤2km	Metro in 2-5km	Log(Dist)
Comparison=	all	not-yet-built	not-yet-built	>5km	all
Treated×Y2013	0.415***	0.471**	0.332**	0.136	-0.036
	(0.127)	(0.203)	(0.142)	(0.097)	(0.042)
Treated×Y2005	0.059	0.064	0.038	0.295	0.024
	(0.176)	(0.255)	(0.171)	(0.181)	(0.070)
DepVar Mean (1990)	2.740	3.782	3.402	2.346	2.694
Obs	901	354	675	651	1,139
Wards	349	143	274	248	445
	Panel C: Outcome = Share of Large Firms (10+ employees)				
Treated=	Metro≤1km	Metro≤1km	Metro≤2km	Metro in 2-5km	Log(Dist)
Comparison=	all	not-yet-built	not-yet-built	>5km	all
Treated×Y2013	0.052**	0.071**	0.035*	0.013	-0.000
	(0.020)	(0.036)	(0.020)	(0.017)	(0.008)
Treated×Y2005	0.014	0.020	0.015	0.028	0.003
	(0.030)	(0.036)	(0.022)	(0.019)	(0.011)
DepVar Mean (1990)	0.045	0.070	0.055	0.036	0.044
Obs	901	354	675	651	1,139
Wards	349	143	274	248	445

OLS estimates of metro station proximity effects on employment and firms from Section 6.1. Data: Economic Census from years 1990, 2005, 2013. Treatment: proximity of city wards to Phase 2 metro stations (built 2005-2013). Y2013 interactions = treatment effects; Y2005 = pre-trend tests. Column (1): treated = ≤1km from station, comparison = all areas >2km; sample excludes wards 1-2km from station to account for spillovers. Column (2): treated = ≤1km from station, comparison = ≤1km from unbuilt station; sample excludes wards 1-2km from station to account for spillovers. Column (3): treated = ≤2km from station, comparison = ≤2km from unbuilt station. Column (4): treated = 2-5km from Phase 2 station, comparison = >5km. Column (5): log distance to station specification using all areas. All samples exclude wards ≤1km of Phase 1 metro stations. The mean of the dependent variable is the 1990 value of the respective comparison group (for Column (5) - the sample for the mean is above median distance wards); note that mean reported in Panel B is for average employment. The means are taken from respective comparison samples winsorized at the 95th percentile to account for outliers. All regressions include ward, year×log(distance to CBD), and year×log(distance to Phase 1 metro station) fixed effects. Regressions are weighted by ward area. 5km Conley standard errors in (parentheses). $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3: Effect of Changes in Firm Commuter Market Access on Employment and Firms

Outcomes:	(1) Total Emp (Ward)	(2) Avg Firm Size	(3) Manu Firm (10w+)	(4) High-Skill Firm (10w+)	(5) B2C Firm (10w+)
	Panel A: Firm Commuter Market Access (FCMA)				
Δ ln(FCMA)×Y13	1.25**	1.25**	-0.013	0.018***	0.028*
	[16.69%]	[6.60%]	[-0.06pp]	[0.09pp]	[0.14pp]
	(0.562)	(0.593)	(0.041)	(0.006)	(0.014)
Δ ln(FCMA) Future×Y13	0.43	-0.20	-0.014	-0.005	-0.011
	(0.939)	(0.991)	(0.034)	(0.006)	(0.008)
DepVar Mean (1990)	3345.371	3.267	0.030	0.006	0.009
	Panel B: FCMA × Above Median Rent				
HighRent×Y13	-0.31***	-0.03	-0.013	0.002	0.002
	(0.108)	(0.156)	(0.009)	(0.001)	(0.002)
Δ ln(FCMA)×Y13	1.05**	0.97	-0.011	0.005**	0.019
	[13.79%]	[5.09%]	[-0.06pp]	[0.03pp]	[0.10pp]
	(0.433)	(0.868)	(0.056)	(0.003)	(0.013)
Δ ln(FCMA)×HighRent×Y13	0.41	0.70	-0.030	0.046***	0.035***
	[5.20%]	[3.63%]	[-0.16pp]	[0.23pp]	[0.18pp]
	(1.007)	(0.795)	(0.051)	(0.014)	(0.007)
DepVar Mean (1990)	860.298	2.996	0.033	0.004	0.007
	Panel C: FCMA × Has Metro Station in 1km				
Metro×Y13	0.00	0.12	0.003	0.002	0.005***
	(0.226)	(0.146)	(0.008)	(0.002)	(0.002)
Δ ln(FCMA)×Y13	0.50	1.74	0.016	0.013*	0.008
	[6.40%]	[9.31%]	[0.08pp]	[0.06pp]	[0.04pp]
	(0.731)	(1.070)	(0.041)	(0.007)	(0.006)
Δ ln(FCMA)×Metro×Y13	1.45	-0.78	-0.053	0.011	0.046**
	[19.52%]	[-3.89%]	[-0.27pp]	[0.06pp]	[0.23pp]
	(1.381)	(1.154)	(0.078)	(0.012)	(0.019)
DepVar Mean (1990)	2214.234	3.324	0.036	0.006	0.009
Obs	1,037	1,546,559	1,546,559	1,546,559	1,546,559
Wards	375	375	375	375	375

Poisson (columns 1-2) and OLS (columns 3-5) estimates of firm commuter market access (FCMA) changes on employment and firm composition from Section 6.2. Data: Economic Census from years 1990, 2005, 2013. Col (1): total employment at ward level. Col (2): average firm size (employment per firm). Col (3)-(5): binary indicators for firm having 10+ workers in a manufacturing, high-skill professional, or business-to-consumer industry, respectively. Δ ln(FCMA) measures log change in firm commuter market access from 2005 to 2013 due to change in metro network. Δ ln(FCMA) Future measures anticipated future accessibility gains (2013-2022). Panel B interacts Δ ln(FCMA)×Y13 and Δ ln(FCMA) Future×Y13 with indicator for ward having above-median rent per sqft (as assessed by city municipality). Panel C interacts Δ ln(FCMA)×Y13 and Δ ln(FCMA) Future×Y13 with indicator for ward being within 1km of a metro station in 2013. [Brackets] show the predicted effect of moving from the 25th to 75th percentile of FCMA change, which is 0.123 for Col 1 and 0.0512 for Cols 2-5. Effect sizes are in percentage change in the outcome for Poisson models (Cols 1-2), and percentage point change for binary outcomes (Cols 3-5). The dependent variable mean is the 1990 value for the respective control group: Panel A uses wards below median FCMA change; Panel B uses below-median rent wards with below-median FCMA change; Panel C uses wards without metro stations and below-median FCMA change. All regressions include ward and year fixed effects. 5km Conley standard errors in (parentheses). $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 4: Effect of Changes in Firm Commuter Market Access on Employment and Firms

Panel A: Number of New Firms (employment weighted) - contributing to female employment

	Metro-Type		Manager Gender		Firm Workforce		Firm Workforce	
	(1) High	(2) Low	(3) Female Mgr	(4) Male Mgr	(5) $\geq$ 1 women	(6) No women	(7) Fem-Majority	(8) Male-Majority
ATT (% inc.)	48.8*	35.3	180.0***	46.8*	86.3***	26.4	184.1***	40.9*
	(25.3)	(37.4)	(42.5)	(24.5)	(29.1)	(29.9)	(60.3)	(21.8)
Mean	10.43	7.94	2.62	17.41	11.36	11.72	2.58	20.44
Pre-trend p-val	0.105	0.985	0.478	0.236	0.516	0.193	0.311	0.156

Panel B: Number of Employees in New Firms - female vs. male employees

	Female Emp		Male-Managed		10w+ Firm		Specialized Firm	
	(1) Gen Prod	(2) Fem Prod	(3) Fem Emp	(4) Male Emp	(5) Fem Emp	(6) Male Emp	(7) Fem Emp	(8) Male Emp
ATT (% inc.)	95.8***	93.0	92.8**	35.7	114.6***	44.4	82.0***	54.2**
	(28.8)	(59.0)	(37.2)	(27.5)	(36.7)	(43.1)	(30.0)	(22.1)
Mean	2.58	0.21	2.07	13.98	1.43	7.23	1.71	8.72
Pre-trend p-val	0.397	0.527	0.538	0.081	0.744	0.535	0.421	0.525
Neighborhoods	370	370	370	370	370	370	370	370
Observations	9990	9990	9990	9990	9990	9990	9990	9990

From Section 7.1. *Outcomes: Panel A (Number of employment-weighted firms that are...):* (1-2) in high vs. low metro-propensity industries (3-4) Male vs. female-managed; (5-6) With vs. without female employees; (7-8) Female vs. male-majority workforce. *Panel B (Number of employees):* (1-2) Female employees in female-oriented vs. general industries; (3-4) Female vs. male employees in male-managed firms; (5-6) Female vs. male employees in firms with 10+ workers; (7-8) Female vs. male employees in differentiated firms (owner$\neq$manager). ATT from staggered difference-in-differences (Callaway and Sant'Anna, 2021), multiplied by 100 to represent % change relative to baseline. Date: Shops and Establishment Registrations (2011-2024). Sample is at neighborhood-time level, where, with neighborhoods being H3 hexagons within 1km of major roads and time being 6-month periods. Treatment: metro station within 1km; comparison: never-treated and not-yet-treated neighborhoods within 1km of major roads. Baseline mean is the average raw outcome across comparison neighborhoods (never-treated or first-treated after period 2) during 2011-2016 (periods 1-10). Outcome variables are normalized by dividing by this baseline mean, such that ATT represents % change from baseline. Pre-trend p-value from F-test of 8 pre-treatment periods (months -72 to -30). ATT estimated over 4 anticipation periods, event, and 12 post-opening periods. Standard errors clustered by neighborhood. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 5: Decomposition of Gendered Employment Effects

	Employment Effects by Gender and Sector			
	Male Employment	Female Employment	Female Employment Firm Comp Const.	Female Employment Fem Share Const.
	(1)	(2)	(3)	(4)
Treated×Y2013	1.4856	0.2860*	0.0693	0.1040*
	(0.9161)	(0.1708)	(0.0496)	(0.0564)
Treated×Y2005	0.5387	0.1217	0.0256**	0.0249
	(0.5072)	(0.0908)	(0.0120)	(0.0198)
DepVar Mean (1990)	0.1437	0.0136	0.0126	0.0120
Observations	354	354	354	354
Wards	143	143	143	143

OLS estimates of metro station effects on employment from Section 7.2. Treatment: wards ≤1km from Phase 2 stations (built 2005-2013). Comparison: wards ≤1km from unbuilt stations. Y2013 interactions = treatment effects; Y2005 = pre-trend tests. Sample excludes wards 1-2km from built stations and wards ≤1km from Phase 1 stations. Columns show male employment per capita (1), female employment per capita (2), counterfactual female employment holding firm type distribution constant at 1990 levels (3), and counterfactual female employment holding female shares within firm types constant at 1990 levels (4). Column (3) isolates within-firm-type changes in female hiring by fixing the distribution of employment across industry×size bins at baseline proportions. Column (4) isolates compositional effects by fixing the female share within each industry×size bin at baseline levels. Dependent variable means from 1990, winsorized at 95th percentile. Includes ward FE, year FE, year×controls (log distance to CBD, log distance to Phase 1 stations). Weighted by ward area. 5km Conley SEs in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

References

Aguilar, A., E. Gutiérrez, and P. S. Villagrán (2021): "Benefits and unintended consequences of gender segregation in public transportation: Evidence from Mexico City's subway system," *Economic Development and Cultural Change*, 69, 1379–1410.

Ahlfeldt, G. M., T. Albers, and K. Behrens (2022): "A granular spatial model," .

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015): "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 83, 2127–2189.

Akcigit, U., H. Alp, and M. Peters (2021): "Lack of selection and limits to delegation: firm dynamics in developing countries," *American Economic Review*, 111, 231–275.

Alam, M. M., M. L. Cropper, M. Herrera Dappe, and P. Suri (2021): "Closing the Gap: Gender, Transport, and Employment in Mumbai," Tech. rep., The World Bank.

Amirapu, A. and M. Gechter (2020): "Labor Regulations and the Cost of Corruption: Evidence from the Indian Firm Size Distribution," *The Review of Economics and Statistics*, 102, 34–48.

Araújo, B. C., L. S. Paz, and J. E. West (2024): "Cracks in the Glass Ceiling and Gender Equality: Do Exports Shatter the Glass Ceiling?" Tech. rep., National Bureau of Economic Research.

Asher, S., T. Lunt, R. Matsuura, and P. Novosad (2021): "Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in India using the shrug open data platform," *The World Bank Economic Review*, 35, 845–871.

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

Balboni, C., G. Bryan, M. Morten, C. O'Connor, and B. Siddiqi (2025): "Impacts of Infrastructure: People versus Place," .

Bassi, V., M. E. Kahn, N. L. Gracia, T. Porzio, and J. Sorin (2022): "Jobs in the Smog: Firm Location and Workers' Exposure to Pollution in African Cities," Tech. rep., National Bureau of Economic Research.

Baum-Snow, N. (2020): "Urban transport expansions and changes in the spatial structure of us cities: Implications for productivity and welfare," *Review of Economics and Statistics*, 102, 929–945.

Baum-Snow, N., L. Brandt, J. V. Henderson, M. A. Turner, and Q. Zhang (2017): "Roads, railroads, and decentralization of Chinese cities," *Review of Economics and Statistics*, 99, 435–448.

Benmelech, E., N. Bergman, A. Milanez, and V. Mukharlyamov (2019): "The agglomeration of bankruptcy," *The Review of Financial Studies*, 32, 2541–2586.

Black, S. E. and E. Brainerd (2004): "Importing equality? The impact of globalization on gender discrimination," *ILR Review*, 57, 540–559.

Borker, G. et al. (2021): "Safety first: Perceived risk of street harassment and educational choices of women," .

Borusyak, K. and P. Hull (2023): "Nonrandom exposure to exogenous shocks," *Econometrica*, 91, 2155–2185.

Brinatti, A. and N. Morales (2021): "Firm Heterogeneity and the Impact of Immigration: Evidence from German Establishments," Working Paper 21-16, Federal Reserve Bank of Richmond, available at SSRN 4054097. Revise and resubmit at Journal of Political Economy.

Busso, M. and O. Fentanes (2024): "Building up local productivity: Infrastructure and firm dynamics in Mexico," Tech. rep., IDB Working Paper Series.

Callaway, B. and P. H. Sant'Anna (2021): "Difference-in-differences with multiple time periods," *Journal of econometrics*, 225, 200–230.

Chatterjee, S., K. Krishna, K. Padmakumar, and Y. Zhao (2025): "No Country for Dying Firms: Evidence from India," Tech. rep., National Bureau of Economic Research.

Cheema, A., A. I. Khwaja, F. Naseer, and J. N. Shapiro (2019): "Glass walls: Experimental evidence on access constraints faced by women," .

Chen, Y., K. Coşar, D. Ghose, S. Mahendru, and S. Sekhri (2024): "Gender-Specific Transportation Costs and Female Time Use: Evidence from India's Pink Slip Program," Tech. rep., National Bureau of Economic Research.

Chiplunkar, G. and P. K. Goldberg (2021): "Aggregate Implications of Barriers to Female Entrepreneurship," Tech. rep., National Bureau of Economic Research.

Deffebach, P., D. Lagakos, Y. Miyauchi, and E. Yamada (2025): "The spatial distribution of income in cities: New global evidence and theory," *Unpublished manuscript.*

dmrc (2010): "Need for a Metro: History," .

DMRC (2011): "EIA for Phase III Corridors of Delhi Metro," Tech. rep., Delhi Metro Rail Corporation, Metro Bhawan, Fire Brigade Lane, Barakhamba Road, New Delhi-110001.

Donaldson, D. (2025): "Transport Infrastructure and Policy Evaluation," in *Handbook of Regional and Urban Economics*, ed. by D. Donaldson and S. J. Redding, Elsevier, vol. 6, forthcoming. NBER Working Paper No. 34096.

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

Duranton, G. and D. Puga (2020): "The economics of urban density," *Journal of economic perspectives*, 34, 3–26.

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

Field, E. and K. Vyborny (2022): "Women's mobility and labor supply: experimental evidence from Pakistan," *Asian Development Bank Economics Working Paper Series.*

Gechter, M. and N. Tsivanidis (2023): "Spatial spillovers from high-rise developments: Evidence from the Mumbai Mills," .

Goraya, S. S. and A. Ilango (2024): "Identity, Market Access, and Demand-led Diversification," Working Paper.

Hasan, R. and K. Jandoc (2013): "Labor regulations and firm size distribution in Indian manufacturing," *Reforms and economic transformation in India*, 15–48.

Heblich, S., S. J. Redding, and D. M. Sturm (2020): "The making of the modern metropolis: evidence from London," *The Quarterly Journal of Economics*, 135, 2059–2133.

Hjort, J. (2014): "Ethnic Divisions and Production in Firms," *The Quarterly Journal of Economics*, 129, 1899–1946.

Ho, L., S. Jalota, and A. Karandikar (2023): "Bringing Work Home: Flexible Arrangements as Gateway Jobs for Women in West Bengal," Tech. rep., Technical Report, MIT Working Paper.

Holzer, H. J. and K. R. Ihlanfeldt (1998): "Customer Discrimination and Employment Outcomes for Minority Workers," *The Quarterly Journal of Economics*, 113, 835–867.

Hsieh, C.-T. and P. J. Klenow (2014): "The life cycle of plants in India and Mexico," *The Quarterly Journal of Economics*, 129, 1035–1084.

Hunt, J. and C. Moehling (2024): "Do Female–Owned Employment Agencies Mitigate Discrimination and Expand Opportunity for Women?" Tech. rep., National Bureau of Economic Research.

Jalota, S. and L. Ho (2024): "What Works for Her? How Work-from-Home Jobs Affect Female Labor Force Participation in Urban India," *How Work-from-Home Jobs Affect Female Labor Force Participation in Urban India (January 4, 2024)*.

Jayachandran, S. (2021): "Social norms as a barrier to women's employment in developing countries," *IMF Economic Review*, 69, 576–595.

Juhn, C., G. Ujhelyi, and C. Villegas-Sanchez (2014): "Men, women, and machines: How trade impacts gender inequality," *Journal of Development Economics*, 106, 179–193.

Kapoor, R. and S. Gade (2024): "Together in Search: Experimental Evidence from Coordinating Travel Among Women Job-seekers in Urban India," .

Kayal, P. (2013): *Mega-event & Transportation: Analyzing New Delhi & XIX Commonwealth Games, 2010*, Michigan State University. Urban and Regional Planning.

Kelley, E. M., G. Lane, M. Pecenco, and E. Rubin (2023): "Customer discrimination in the workplace: Evidence from online sales," .

Klasen, S. and J. Pieters (2015): "What explains the stagnation of female labor force participation in urban India?" *The World Bank Economic Review*, 29, 449–478.

Kline, P., E. K. Rose, and C. R. Walters (2022): "Systemic discrimination among large US employers," *The Quarterly Journal of Economics*, 137, 1963–2036.

Kondylis, F., A. Legovini, K. Vyborny, A. M. T. Zwager, and L. Cardoso De Andrade (2020): "Demand for safe spaces: Avoiding harassment and stigma," *World Bank Policy Research Working Paper*.

Kwon, E. and N. Lee (2024): "Unintended Pathways: The Impact of High-Speed Rail on Gender Differences in the Local Labor Market-Evidence from South Korea," .

Larson-Koester, M. (2020): "Occupation Gender Segregation: Empirical Evidence from a Matching Model with Transfers," *Job Market Paper*.

Le Barbanchon, T., R. Rathelot, and A. Roulet (2021): "Gender differences in job search: Trading off commute against wage," *The Quarterly Journal of Economics*, 136, 381–426.

Leonardi, M. and E. Moretti (2023): "The agglomeration of urban amenities: Evidence from Milan restaurants," *American Economic Review: Insights*, 5, 141–157.

Lindenlaub, I., R. Oh, and M. Peters (2022): "Firm sorting and spatial inequality," Tech. rep., National Bureau of Economic Research.

Liu, S. and Y. Su (2024): "The geography of jobs and the gender wage gap," *Review of Economics and Statistics*, 106, 872–881.

Lordan, G. and J.-S. Pischke (2022): "Does Rosie like riveting? Male and female occupational choices," *Economica*, 89, 110–130.

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

Martinez, D. F., O. A. Mitnik, E. Salgado, L. Scholl, and P. Yañez-Pagans (2020): "Connecting to economic opportunity: The role of public transport in promoting women's employment in Lima," *Journal of Economics, Race, and Policy*, 3, 1–23.

Miller, C., J. Peck, and M. Seflek (2022): "Missing women, integration costs, and big push policies in the Saudi labor market," *American Economic Journal: Applied Economics*, 14, 51–77.

Miyauchi, Y., K. Nakajima, and S. J. Redding (2021): "The economics of spatial mobility: Theory and evidence using smartphone data," Tech. rep., National Bureau of Economic Research.

Nafilyan, V. (2019): "Gender differences in commute time and pay," *Office for National Statistics, London, United Kingdom, available at https://www. ons. gov. uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/articles/genderdifferencesincommutetimeandpay/2019-09-04# toc (9 September 2020).*

Nose, M. and Y. Sawada (2025): "From Battlefield to Marketplace: Industrialization via Interregional Highway Investments in the Greater Mekong Sub-Region," Tech. rep., Institute for Economics Studies, Keio University.

Oh, R. and J. Seo (2023): "What causes agglomeration of services? theory and evidence from seoul," *Yale University, mimeograph.*

Onishi, Y. (2016): "Breaking Ground: A narrative on the making of Delhi Metro," *Japan International Cooperation Agency.*

Otterbring, T., R. Bhatnagar, P. Samuelsson, and S. Borau (2021): "Positive gender congruency effects on shopper responses: field evidence from a gender egalitarian culture," *Journal of Retailing and Consumer Services*, 63, 102738.

Rajagopalan, S. and K. Shah (2024): "Why Indian Firms Don't Scale: Labor Edition," *India Political Economy Program: Research Paper.*

Redding, S. J. (2025): "Evaluating transport improvements in spatial equilibrium," Tech. rep., National Bureau of Economic Research.

Redding, S. J. and M. A. Turner (2015): "Transportation costs and the spatial organization of economic activity," *Handbook of regional and urban economics*, 5, 1339–1398.

Schuh, R. (2024): "Miss-allocation: The value of workplace gender composition and occupational segregation," *FRB of New York Staff Report.*

Seki, M. and E. Yamada (2025): "Gendered Heterogeneous Effects of Urban Public Transportation on Employment: Evidence from the Delhi Metro," .

Severen, C. (2023): "Commuting, labor, and housing market effects of mass transportation: Welfare and identification," *Review of Economics and Statistics*, 105, 1073–1091.

Shoag, D. and S. Veuger (2018): "Shops and the city: Evidence on local externalities and local government policy from big-box bankruptcies," *Review of Economics and Statistics*, 100, 440–453.

Siemiatycki, M. (2006): "Message in a metro: building urban rail infrastructure and image in Delhi, India," *International journal of urban and regional research*, 30, 277–292.

Suri, P. and M. Cropper (2024): "What are the Benefits of Metro Rail in Mumbai, India?" *Available at SSRN 4910918*.

Topalova, P. (2010): "Factor immobility and regional impacts of trade liberalization: Evidence on poverty from India," *American Economic Journal: Applied Economics*, 2, 1–41.

Tsivanidis, N. (2023): "Evaluating the impact of urban transit infrastructure: Evidence from Bogota's Transmilenio," *University of California, Berkeley.*

Velásquez, D. (2023): "Transit Infrastructure, Couples' Commuting Choices, and Gender Earnings Inequality," *Unpublished manuscript.*

Vitali, A. (2023): "Consumer search and firm location: Theory and evidence from the garment sector in Uganda," IGC.

Zárate, R. D. (2022): "Spatial misallocation, informality, and transit improvements: Evidence from mexico city," .

Appendix

A Appendix Tables and Figures

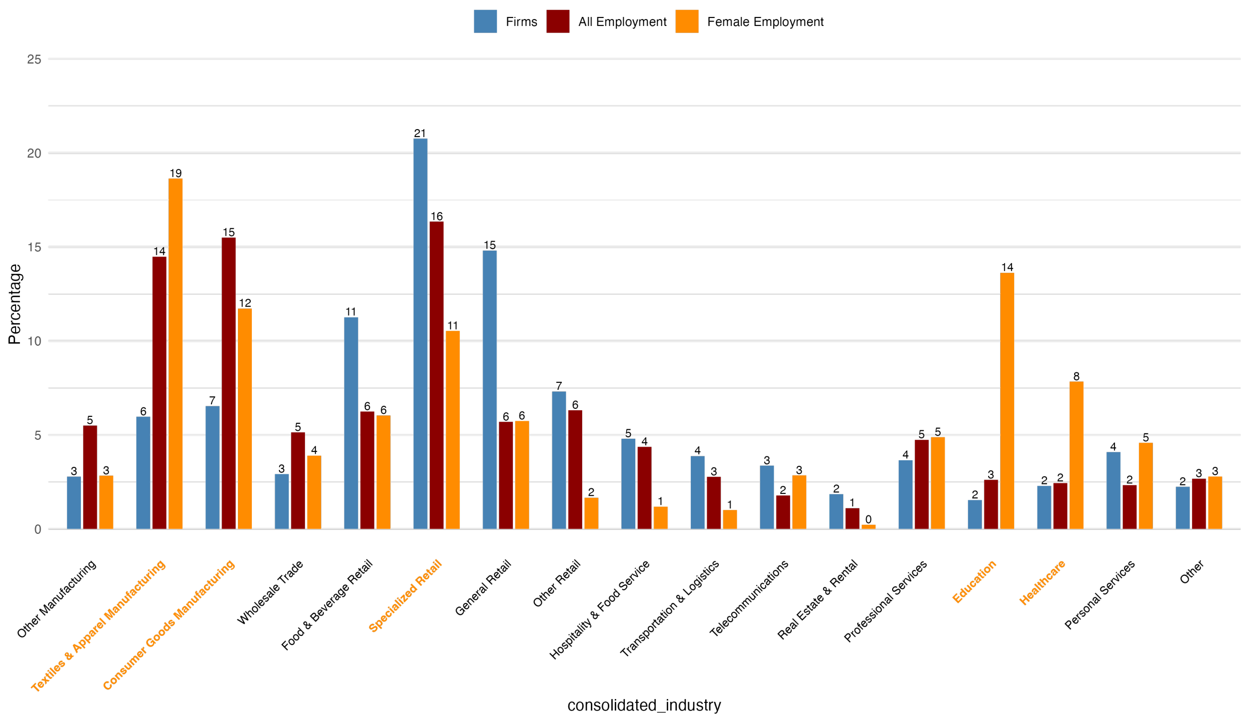
Table A1: Establishment Registrations: Descriptive Statistics

Variable	Mean	Median	SD
Panel A: Firm Characteristics			
Total Employees	2.77	1.00	8.20
Proportion Female Employees	0.11	0.00	0.22
B2C Business (=1)	0.65	1.00	0.48
High-Skill Business (=1)	0.09	0.00	0.28
Female Owner (=1)	0.14	0.00	0.34
Female Manager (=1)	0.14	0.00	0.34
Different Owner & Manager (=1)	0.28	0.00	0.45
Observations: 683,938 firms			
Panel B: Neighborhood Characteristics			
Population in 2011	7974.86	4063.98	9070.08
Distance to Nearest Major Road (m)	1654.38	1186.57	1511.17
Distance to Nearest Metro Stop in 2011 (m)	5092.82	3655.12	4370.92
Has Metro Stop within 1km in 2011	0.17	0.00	0.38
Observations: 1,895 hexagons			
Panel C: Neighborhood-6month Characteristics			
Number of Firms	13.23	1.00	35.74
Number of Firms (10+ employees)	0.65	0.00	1.62
Total Employment	36.55	2.00	84.23
Female Employment	5.63	0.00	14.99
Observations: 51,165 hexagon-periods			

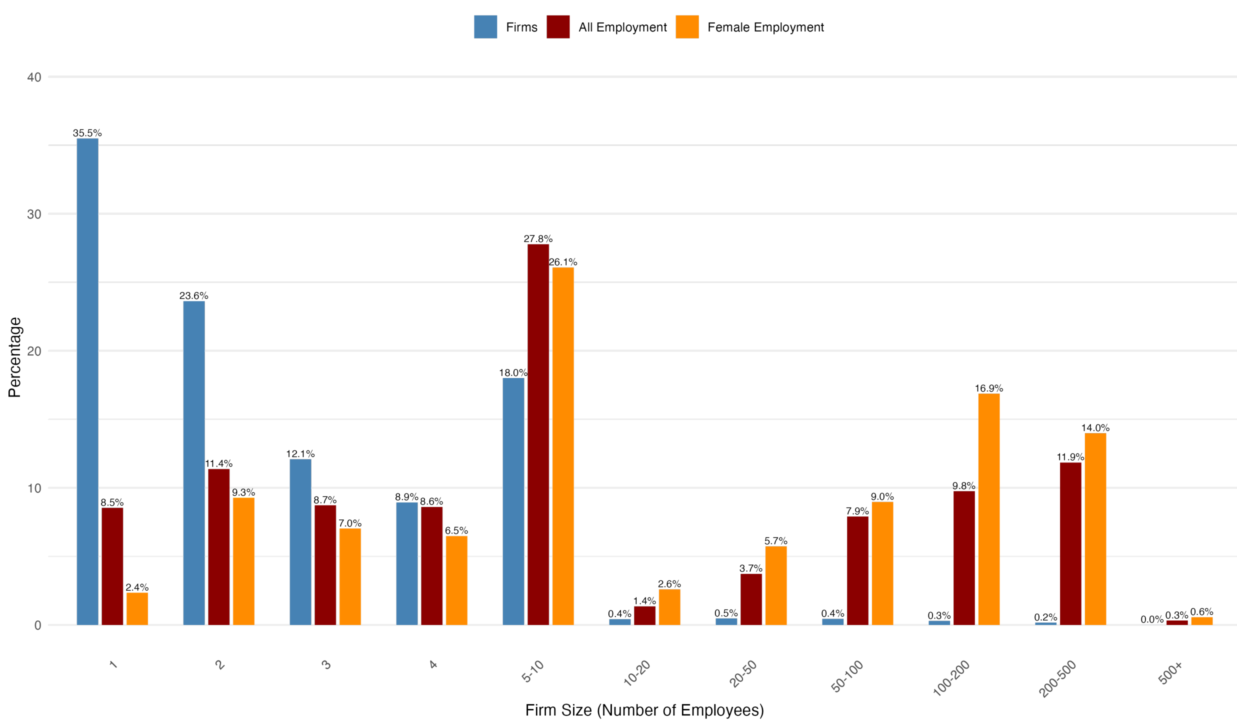
Data: Registrations of Shops and Establishments (2011-2024). Panel A: firm-level statistics for firms established within one year prior to registration. B2C Business = firms selling goods or providing services directly to end consumers. High-Skill Business = firms requiring specialized skills or professional qualifications. Female Owner/Manager based on name classification. Different Owner & Manager indicates specialized management structure. Panel B: neighborhood-level statistics. Neighborhoods defined as H3 hexagons (resolution 8, approximately 500-meter radius). Distance to Nearest Major Road calculated from hexagon centroid to motorways, trunk roads, or primary roads (OpenStreetMap). Panel C: neighborhood-6month period level statistics. Data aggregated to 6-month periods (2011-2024) across 1,895 hexagons.

Figure A1: Employment distribution by industry and firm size in 2005

(a) By industry category



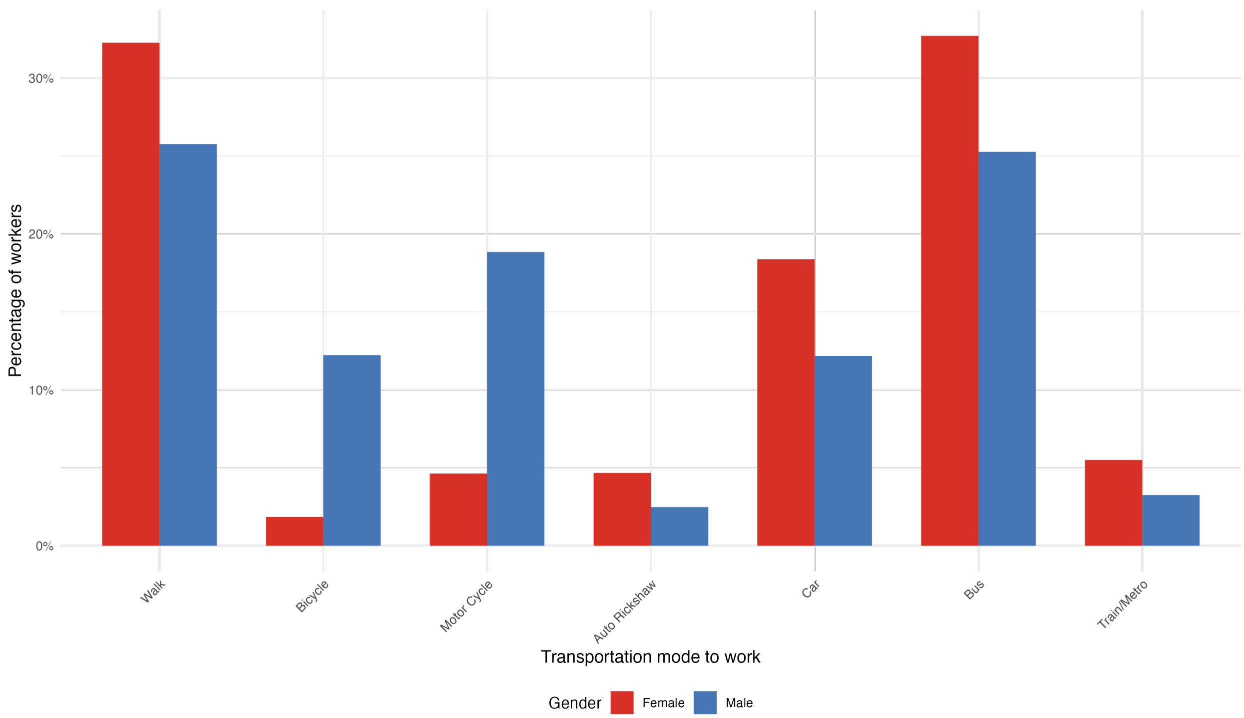
(b) By firm size



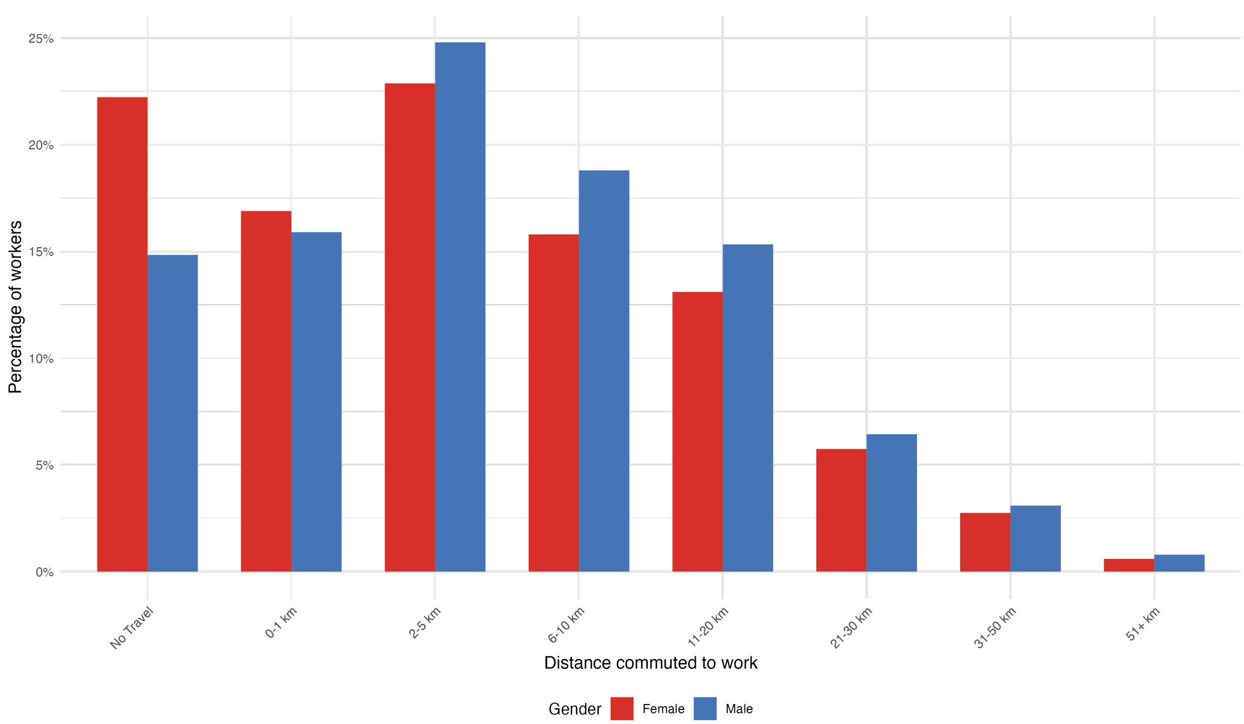
Notes: Data from Economic Census 2005. Panel (a) presents the distribution of firms, total employment, and female employment across industry categories. Panel (b) presents the distribution across firm size bins measured by number of employees. Bars represent the percentage of all firms (blue), total employment (red), or female employment (orange) in each category.

Figure A2: Employment distribution by industry and firm size in 2005

(a) Commute mode, by gender

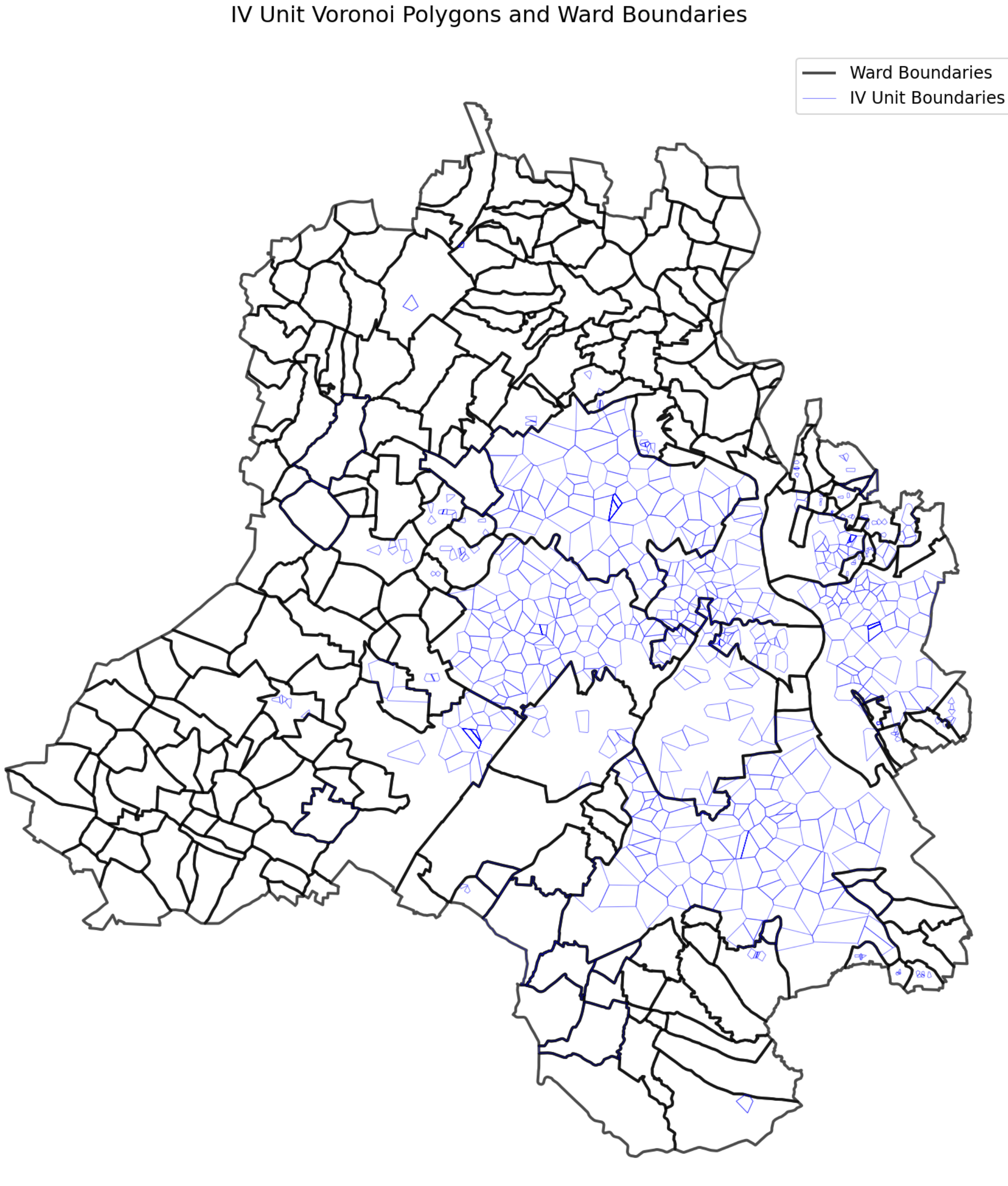


(b) Commute distance, by gender



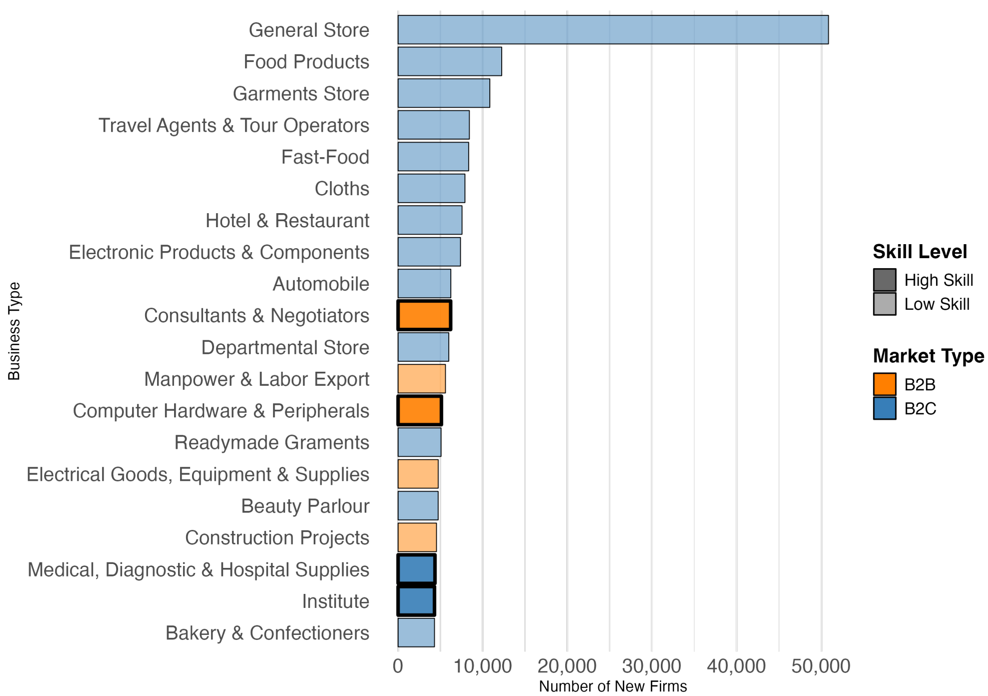
Notes: Data from Population Census 2011. Panel (a) presents the distribution of firms, total employment, and female employment across industry categories. Panel (b) presents the distribution across firm size bins measured by number of employees. Bars represent the percentage of all firms (blue), total employment (red), or female employment (orange) in each category.

Figure A3: Result of matching algorithm to increase spatial granularity of the Economic Census



Notes: Data from Economic Census 2005. Black boundaries represent the originally available spatial units in the raw data. Blue boundaries represent Voronoi polygons of census enumeration units, which are aggregated to approximately 450 urban wards for analysis.

Figure A4: Distribution of Firm Registrations by Business Type



Notes: Data from Delhi Shops and Establishments Registrations, 2011-2024. Figure displays the 20 most common business types by number of new firm registrations. Bars are color-coded by market type (B2C in blue, B2B in orange) and shading indicates typical skill level of workforce (light for low-skill, dark for high-skill).

B Construction of Commuter Market Access Measures

B.1 Canonical Quantitative Spatial Model

To get the sufficient statistics for market access, we follow Tsivanidis (2023) in utilizing the canonical quantitative spatial model from Ahlfeldt et al. (2015). The city consists of discrete locations indexed by i for residence and j for employment. A worker o chooses a location i in which to live and a location j in which to work.

Preferences: Worker o derives utility from consumption of a freely traded numeraire good (c_{ijo}) and residential floorspace (ℓ_{ijo}), along with location-specific residential amenities B_i:

$$U_{ijo} = B_i z_{ijo} \left(\frac{c_{ijo}}{\beta}\right)^{\beta} \left(\frac{\ell_{ijo}}{1-\beta}\right)^{1-\beta} \frac{1}{d_{ij}}$$

where $\beta \in (0,1)$ governs the expenditure share on the numeraire good; z_{ijo} represents an idiosyncratic component of utility specific to worker o choosing residence i and workplace j, reflecting individual-level factors that make certain residence-workplace combinations more attractive; and $d_{ij} \geq 1$ captures iceberg commuting costs that rise with travel time τ_{ij} between locations i and j, parameterized as $d_{ij} = e^{\kappa \tau_{ij}}$.

The idiosyncratic utility component z_{ijo} is drawn from an independent Fréchet distribution:

$$F(z_{ijo}) = e^{-T_i E_j z_{ijo}^{-\theta}}, \quad T_i, E_j > 0, \quad \theta > 1$$

The scale parameter $T_i > 0$ shifts the mean utility from residing in location i; the scale parameter $E_j > 0$ shifts the mean utility from employment in location j; and the shape parameter $\theta > 1$ governs the heterogeneity in idiosyncratic utility draws across residence-workplace pairs, with higher values implying less dispersion.

Budget constraint and optimization: Workers inelastically supply one unit of labor. Setting the final good as numeraire ($p = 1$) and applying first-order conditions for utility maximization subject to the budget constraint, optimal demands for worker o residing in i and employed in j are:

$$c_{ijo} = \beta w_j$$

$$\ell_{ijo} = (1-\beta)\frac{w_j}{Q_i}$$

Here w_j denotes the wage at employment location j, while commuting costs enter through reduced utility rather than reduced income. The term Q_i is the rental price of residential floorspace in location i. Land rents are assumed to accrue to absentee landlords outside the city.

Indirect utility: Substituting these optimal choices back into the utility function yields indirect utility:

$$U_{ijo} = \frac{B_i z_{ijo} w_j Q_i^{\beta-1}}{d_{ij}}$$

This expression reveals that commuting costs effectively reduce the utility value of wage income, which can equivalently be interpreted as a reduction in effective labor units supplied.

Given the indirect utility expression, workers select residence-workplace pairs to maximize U_{ijo}. With Fréchet-distributed idiosyncratic shocks, standard discrete choice results yield the probability that a worker chooses to live in location i and commute to work in location j:

$$\pi_{ij} = \frac{T_i E_j \left(d_{ij} Q_i^{1-\beta}\right)^{-\theta} (B_i w_j)^{\theta}}{\sum_{r=1}^{S} \sum_{s=1}^{S} T_r E_s \left(d_{rs} Q_r^{1-\beta}\right)^{-\theta} (B_r w_s)^{\theta}}$$

Since we cannot isolate T_i from B_i and E_j from w_j, we let the residential amenities and workplace wages absorb the respective parameters. The conditional probability that a worker residing in i chooses to work in j simplifies to:

$$\pi_{ij|i} = \frac{(w_j/d_{ij})^{\theta}}{\sum_s (w_s/d_{is})^{\theta}}$$

Workers are attracted to employment locations offering high wages net of commuting costs. Define the denominator as **Residential Commuter Market Access** (RCMA):

$$\Omega_i^R \equiv \sum_s (w_s/d_{is})^{\theta}$$

which summarizes residents' access to well-paying jobs from location i. Higher RCMA indicates better connectivity to high-wage employment opportunities.

Aggregating labor supply: Total labor supply to employment location j equals the sum of workers commuting from all residential locations:

$$L_j^F = \sum_i \pi_{ij|i} L_i^R = \sum_i \frac{(w_j/d_{ij})^{\theta}}{\Omega_i^R} L_i^R = w_j^{\theta} \sum_i d_{ij}^{-\theta} \frac{L_i^R}{\Omega_i^R}$$

Define **Firm Commuter Market Access** (FCMA) as:

$$\Omega_j^F \equiv \sum_i d_{ij}^{-\theta} \frac{L_i^R}{\Omega_i^R}$$

so that $L_j^F = w_j^\theta \Omega_j^F$. This expression reveals that labor supply to location j depends on both the wage offered and the firm's access to residential labor pools through the commuting network. Locations with high FCMA attract more workers conditional on wages because they are well-connected to populous residential areas where workers have limited alternative employment options (low Ω_i^R).

For empirical implementation, we hold population (L_i^R) and employment (L_j^F) fixed at baseline 1990 levels, abstracting from equilibrium responses in wages and population. This isolates the pure effect of changing commute costs from infrastructure investment. Rearranging the labor supply equation and substituting back into the RCMA definition yields the system:

$$\Omega_i^R = \sum_{j \neq i} d_{ij}^{-\theta} \frac{L_j^F}{\Omega_j^F}, \quad \Omega_j^F = \sum_{i \neq j} d_{ij}^{-\theta} \frac{L_i^R}{\Omega_i^R}$$

where we relabel the dispersion parameter as θ (the commute elasticity) for consistency with the empirical literature.

The two equations jointly determine the two unknowns (Ω_i^R and Ω_j^F) for each location. Intuitively, residential locations with better access to employment (high RCMA) provide workers more outside options, reducing the effective labor supply to any single workplace. Conversely, employment locations with better access to residential areas (high FCMA) draw from larger labor pools. The system captures these interdependencies: each location's market access depends on the market access of all other locations through the commuting network.

We parameterize commute costs as $d_{ij} = \exp(\kappa \cdot t_{ij})$, where t_{ij} is travel time in minutes and $\kappa = 0.01$ (Ahlfeldt et al., 2015). We set $\theta = 3.4$ following Tsivanidis (2023), consistent with commute elasticity estimates from Indian urban contexts (Gechter and Tsivanidis, 2023). We exclude own-location from summation ($i \neq j$) since workers do not commute within their residence ward.

Critically, time variation in CMA arises solely through changes in $d_{ij}(t)$: as the metro network expands, travel times t_{ij} fall, commute costs decline, and market access rises. Since log changes $\Delta \ln(\text{FCMA})$ are invariant to the normalization, this is the estimable object we bring to the data.

B.2 Data Construction and Implementation

We implement a four-step computational pipeline to construct commuter market access measures. The geographic unit is Population Census 2011 wards (approximately 450 urban wards in Delhi). We calculate measures for four time periods: 1990, 2005, 2011, and 2022, capturing different phases of metro expansion.

Step 1: Constructing Bus Network Lines

We rely on travel times by public transit means, primarily the metro network (once functional), the bus network, and walking. We already construct the metro network in the method described in the data appendix. Here, we describe how we construct the bus network. The key assumption is that the bus network remains relatively stable across the study period; metro expansion is the primary source of variation.

- Data source: GTFS (General Transit Feed Specification) bus transit data for Delhi.
- Methodology: We extract bus routes, stops, and stop sequences from GTFS files and build route geometries by connecting consecutive stops in each trip. We clean the network by identifying and removing outlier segments based on: (1) invalid travel times ($\leq$0 or >30 minutes between stops), (2) implausible distances (>5km straight-line or $\leq$50m), and (3) unrealistic speeds (<3 km/h or >80 km/h). We split trips at outlier segments to create continuous valid route segments.
- Output: Cleaned bus line network shapefile with segments, travel times, and average speeds.

Step 2: Creating Travel Speed Rasters

Using the metro and bus routes, we create a 50-meter resolution raster grid covering Delhi plus a 1km buffer. We assign speeds to each grid cell based on infrastructure availability:

Infrastructure Type	Speed (km/h)	Speed (m/s)
Metro lines	45	12.5
Bus routes	15	4.17
Walking (baseline)	5	1.39
Outside Delhi	0.001	0.0003

For grid cells with multiple infrastructure types, we assign the maximum speed available. Year-specific rasters incorporate the metro network as it expands: 1990 has no metro (bus and walking only); 2005 includes Phase 1 metro lines (2002-2005 openings); 2011 includes Phase 1 and Phase 2 metro lines; and 2022 includes the full planned network through Phase 3.

The output is in the format of year-specific speed rasters (GeoTIFF format) representing maximum achievable speed at each location.

Step 3: Computing Ward-to-Ward Travel Times Using Fast Marching Method Fast Marching Method (FMM) algorithm is an efficient numerical method for solving the Eikonal equation: $|\nabla T| = 1/F$, where T is the travel time field and F is the speed field (from Step 2 rasters). The algorithm computes minimum travel time from a source to all grid cells. We implement FMM using the scikit-fmm Python package. The procedure:

1. Load year-specific speed raster
2. Convert ward polygons to centroid points
3. For each origin ward centroid:
    - Run FMM to compute travel time surface to all grid cells
    - Extract travel time to all destination ward centroids
4. Create pairwise origin-destination travel time matrix
5. Handle edge cases: infinite travel times set to maximum observed finite time

We parallelize the process across origin wards when possible. A typical computation involves approximately 450 origins × 450 destinations = 200,000 ward-pairs per year. The output is in the format of year-specific ward-to-ward travel time matrices

Step 4: Calculating Commuter Market Access Measures We use as inputs the ward-level employment from Economic Census (1990, 2005, 2013); ward-level residential population from Population Census (1991, 2001, 2011); and year-specific travel time matrices from Step 3. Then we use the following method to solve the system of equations involving the firm and resident commuter market access:

1. Initialize: $\Omega_i^R(0) = 1$ for all i; $\Omega_j^F(0) = 1$ for all j
2. For iteration $t = 1$ to max_iterations:
    (a) Update RCMA: $\Omega_i^R(t) = \sum_{j \neq i} d_{ij}^{-\theta} \times L_j^F / \Omega_j^F(t-1)$
    (b) Update FCMA: $\Omega_j^F(t) = \sum_{i \neq j} d_{ij}^{-\theta} \times L_i^R / \Omega_i^R(t)$
    (c) Check convergence: $||\Omega(t) - \Omega(t-1)|| <$ tolerance
    (d) If converged, stop; else continue

Implementation details: Commute costs are $d_{ij} = \exp(0.01 \times t_{ij})$. Parameters are $\theta = 3.4$, tolerance $= 10^{-6}$, and max iterations $= 1000$. Population (L_i^R) and employment (L_j^F) are held fixed at baseline 1990 levels for all years. Only commute costs d_{ij} vary across years based on infrastructure changes. We exclude own-ward from summation ($i \neq j$, $j \neq i$).

The constructed FCMA measures enter the main regression specification (Section 6.2) as:

$$Y_{pt} = \beta_1(\Delta \ln(\text{FCMA})_p \times \text{Year2013}_t) + \beta_2(\Delta \ln(\text{FCMA})_{\text{Future},p} \times \text{Year2013}_t) + \delta_p + \gamma_t + \epsilon_{pt}$$

where $\Delta \ln(\text{FCMA})_p$ captures the log change in firm commuter market access from 2005 to 2013 due to metro expansion. The coefficient β_1 identifies the causal effect of improved labor market accessibility on employment and firm outcomes. The inclusion of $\Delta \ln(\text{FCMA})_{\text{Future},p}$ as a falsification test—measuring anticipated changes from 2013-2022 based on planned extensions—ensures that results reflect actual infrastructure impacts rather than endogenous station placement.

C Construction of Counterfactuals

C.1 Estimation of Gender-Specific Commute Elasticities

C.1.1 Mobile Phone Data: Structure and Censoring

Data Source and Spatial Coverage: We use mobile phone location data from Cuebiq covering Delhi in 2019, aggregated to H3 hexagonal grid cells at resolution 7 (approximately 5.2 km^2 area, 2 km radius). The dataset contains 281 hexagons covering Delhi and a 20 km buffer. For each origin-destination-week triplet (i, j, w) where $i, j \in \{1, \dots, 281\}$ and $w \in \{1, \dots, 52\}$, we observe the number of individuals whose primary nighttime location is hexagon i and primary daytime location is hexagon j during week w.

Data Censoring: To protect privacy, Cuebiq censors the data: origin-destination-week cells with fewer than 10 trips are suppressed and reported as missing. This creates a selective missingness problem where low-flow pairs (typically long-distance commutes or peripheral locations) are systematically censored. Of the $281 \times 281 \times 52 = 4,105,092$ possible origin-destination-week combinations, approximately 1.9 million observations (46%) are censored.

Imputation Strategy: We impute censored flows using exponential decay based on travel time. The imputation proceeds in three steps:

Step 1: Estimate decay parameter from uncensored data. Using observed (non-censored) flows, we estimate:

$$\ln(\text{TripCount}_{ijw}) = \lambda_0 - \lambda \cdot \text{TravelTime}_{ij}^{\text{norm}} + \gamma_i + \gamma_j + \epsilon_{ijw}$$

where $\text{TravelTime}_{ij}^{\text{norm}}$ is travel time normalized to [0,1] range and γ_i, γ_j are origin and destination fixed effects. This yields $\hat{\lambda} = 2.18$.

Step 2: Predict censored flows. For censored observations, we impute:

$$\text{TripCount}_{ijw}^{\text{imputed}} = 9 \times \exp(-\hat{\lambda} \cdot \text{TravelTime}_{ij}^{\text{norm}})$$

where 9 is the upper bound of the censoring range [0,9]. This approach assumes censored flows follow the same distance-decay pattern as observed flows but fall below the 10-trip threshold.

C.1.2 Construction of Conditional Probabilities

For each origin-week (i, w), we calculate total outflow:

$$\text{TotalOutflow}_{iw} = \sum_j \text{TripCount}_{ijw}^{\text{imputed}}$$

The conditional probability of commuting from i to j in week w is:

$$\Pr(j|i, w) = \frac{\text{TripCount}_{ijw}^{\text{imputed}}}{\text{TotalOutflow}_{iw}}$$

By construction, $\sum_j \Pr(j|i, w) = 1$ for each origin-week pair. This satisfies the normalization constraint required for PPML estimation with origin-week fixed effects.

C.1.3 Travel Time Calculation

Travel times t_{ij} are computed using the Fast Marching Method (FMM) on 2019 speed rasters that incorporate metro lines (45 km/h), bus routes (15 km/h), and walking (5 km/h). For each hexagon pair, we calculate the minimum travel time from origin centroid to destination centroid. The travel time matrix is symmetric and time-invariant (held fixed at 2019 network configuration).

C.1.4 Spatial Correlation

Since hexagons are spatially clustered, commuting flows from nearby origins to the same destination are likely correlated. We account for this using Conley (1999) standard errors with a 5 km cutoff. For each origin hexagon, we identify all other origins within 5 km (using centroid-to-centroid Euclidean distance) and allow arbitrary correlation in residuals among these spatial neighbors. The Conley procedure requires origin coordinates (latitude, longitude), which we extract from hexagon centroids.

C.1.5 Census Distance Data: Structure and Coding

Data Source: Population Census 2011 Transportation Tables (DDW-0700B-28) provide, for each district-sector, the number of workers by gender commuting within seven distance bins: no travel, 0-1 km, 2-5 km, 6-10 km, 11-20 km, 21-30 km, 31-50 km, and 51+ km. We aggregate across Delhi's 11 districts to obtain city-level distributions by gender.

Distance Assignment: We assign each bin a representative distance in kilometers:

Bin	Midpoint (km)	Rationale
No travel	—	Excluded from regression
0-1 km	0.5	Midpoint of range
2-5 km	3.5	Midpoint of range
6-10 km	8.0	Midpoint of range
11-20 km	15.5	Midpoint of range
21-30 km	25.5	Midpoint of range
31-50 km	40.5	Midpoint of range
51+ km	60.0	Approximate (51+75)/2

The "no travel" category is excluded from decay estimation since these workers do not face commuting costs. The 51+ km bin is assigned 60 km based on the assumption that few workers commute beyond 75 km in this context.

Share Calculation: For each gender g and distance bin d, we calculate:

$$\text{Share}_{dg} = \frac{\text{Workers}_{dg}}{\sum_{d'} \text{Workers}_{d'g}}$$

where the sum in the denominator excludes "no travel." This yields the share of commuting workers (conditional on commuting) in each distance bin.

C.1.6 Distance Decay Estimation

We estimate gender-specific decay rates via weighted OLS:

$$\ln(\text{Share}_{dg}) = \alpha_g + \delta_g \cdot \ln(\text{Distance}_d) + \epsilon_{dg}$$

with weights equal to Workers_{dg}, the number of workers in each bin. Weighting ensures that distance bins with more workers (shorter distances) receive greater influence in the regression, reflecting higher precision in those cells. The log-log specification implies constant elasticity of commuting share with respect to distance.

Seven observations per gender (7 distance bins) yield 14 total observations. Despite the small sample, the R-squared exceeds 0.97 for both genders, indicating strong log-linear fit. Standard errors are heteroskedasticity-robust.

C.1.7 Anchoring Procedure

The census provides relative elasticity $\theta_f/\theta_m = |\delta_f|/|\delta_m|$ but not absolute levels. To recover θ_f and θ_m, we impose the constraint that the population-weighted average equals the mobile phone estimate:

$$\theta_{\text{general}} = \alpha_f \cdot \theta_f + (1 - \alpha_f) \cdot \theta_m$$

where α_f is the female labor force share. We calculate:

$$\alpha_f = \frac{\sum_{\text{all districts}} \text{Workers}_f}{\sum_{\text{all districts}} \text{Workers}_{\text{total}}} = 0.147$$

Given $\theta_f = r \cdot \theta_m$ where $r = 1.178$, we substitute into the weighted average:

$$\theta_{\text{general}} = \alpha_f \cdot (r \cdot \theta_m) + (1 - \alpha_f) \cdot \theta_m \tag{16}$$

$$= \theta_m \cdot [\alpha_f \cdot r + (1 - \alpha_f)] \tag{17}$$

Solving: $\theta_m = \theta_{\text{general}}/[\alpha_f \cdot r + (1 - \alpha_f)]$ and $\theta_f = r \cdot \theta_m$.