MIT CEEPR
Center for Energy and Environmental Policy Research

Working Paper Series

Firm Presence, Pollution, and Agglomeration: Evidence from a Randomized Environmental Place-Based Policy

Michael Gechter and Namrata Kala

MAY 2025

CEEPR WP 2025-09

Working Paper Series.

Since 1977, the Center for Energy and Environmental Policy Research (CEEPR) has been a focal point for research on energy and environmental policy at MIT. CEEPR promotes rigorous, objective research for improved decision making in government and the private sector, and secures the relevance of its work through close cooperation with industry partners from around the globe. Drawing on the unparalleled resources available at MIT, affiliated faculty and research staff as well as international research associates contribute to the empirical study of a wide range of policy issues related to energy supply, energy demand, and the environment.

An important dissemination channel for these research efforts is the MIT CEEPR Working Paper series. CEEPR releases Working Papers written by researchers from MIT and other academic institutions in order to enable timely consideration and reaction to energy and environmental policy research, but does not conduct a selection process or peer review prior to posting. CEEPR's posting of a Working Paper, therefore, does not constitute an endorsement of the accuracy or merit of the Working Paper. If you have questions about a particular Working Paper, please contact the authors or their home institutions.

Firm Presence, Pollution, and Agglomeration: Evidence from a Randomized Environmental Place-Based Policy

Michael Gechter* and Namrata Kala†

April 12, 2025

Abstract

Firm location decisions are a key managerial choice, usually optimized over factors like proximity to customers or suppliers. These decisions may also impose externalities on the environment, and on other firms due to competitive or agglomerative forces. The inherent endogeneity of location decisions makes estimating the impact of firm presence difficult. In this paper, we study an environmental place-based policy that randomly moved over 20,000 small firms in New Delhi to industrial areas outside the city over several years. We find that a reduction in firm presence improves air quality, reducing industrial pollution by 8% for the average neighborhood. However, industrial relocation is costly for firms, significantly increasing the probability of firm exit. We combine the exogenous assignment of firms to industrial plots with a model of firms playing a game of incomplete information to estimate the effect of neighborhood composition on firm survival through Marshallian agglomeration forces. We find that proximity to neighboring firms with input-output linkages increases the likelihood of firm survival, and taking these into account while determining firm placement in industrial areas would have halved the costs imposed on firms by the policy. These results provide causal evidence on the trade-offs between firm presence and environmental quality, and show that firm spillovers can be a useful force to minimize the costs on regulated firms.

*MIT Sloan School of Management and Harvard Kennedy School; Y-RISE. Email: mgechter@mit.edu

†MIT Sloan School of Management; BREAD; CEPR; NBER; J-PAL. Email: kala@mit.edu.

We are grateful to the International Growth Center (IGC), MIT Sloan School of Management, the National Science Foundation (NSF), Penn State University, Private Enterprise Development in Low-Income Countries (PEDL) and Structural Transformation and Economic Growth (STEG) for financial support. We would like to thank Jonathan Eaton, Glenn Ellison, Robert Gibbons, Michael Greenstone, Paul Grieco, Dominic Leggett, Melanie Morten, Rohini Pande, Claudia Steinwender, Daniel Xu, and seminar audiences at ASU, Berkeley, ISI Delhi, MIT, NBER, Oxford, Penn State, Stanford, UBC, UChicago, University of Pennsylvania, University of Warwick, and Yale for helpful comments, and Christina Antonakakis, David Cai, Zixu Chen, Caio Figueiredo, Valentina Goetz, Shuhang Lou, Hanna Marsho, Esma Ozer, Garima Singal, and Charlotte Wachter for excellent research assistance. All errors are our own.

1 Introduction

Firm location decisions are one of the most important choices managers make, optimizing factors such as proximity to customers, suppliers, and useful information. At the same time, these decisions may have spillover effects on local neighborhoods, by impacting environmental quality and contributing to local economic activity or agglomeration. For this reason, numerous policies attempt to change location choices of firms. For instance, place-based policies incentivize firms to locate in particular areas and zoning regulations restrict their location choices. Location restrictions that seek to limit pollution exposure in particular also have a long history, starting with the first zoning laws introduced in the early 20th century in New York, in part to improve environmental quality (Wilson et al., 2008). The inherent endogeneity of firm location decisions renders estimating the impact of firm presence on the local economy difficult. Policies that shock firm location decisions can help overcome this but in many cases are bundled with other features. For instance, place-based policies that incentivize firms to locate in certain areas often provide local infrastructure or tax benefits, and may simply displace economic activity rather than creating it.

In this paper we study a policy tool commonly deployed in developing countries to impact firms' location choices: industrial relocation. These policies usually involve the mandated movement of existing firms operating in high-density areas to more remote locations, often, though not always, for environmental reasons. Industrial relocation has been used in many parts of Asia, including China, India, Japan, and South Korea, but little is known about the impacts on relocated firms, and whether such policies achieve their environmental goals. Like many place-based policies, they are often bundled with other policy initiatives, making it difficult to understand their causal impact. Furthermore, they are typically implemented uniformly over an urban area so there is no clear comparison group.

We study a large-scale environmentally-motivated industrial relocation policy in New Delhi, which moved over 20,000 firms to industrial areas outside the city over several years. A unique feature of this policy is that, due to a shortage of industrial plots when relocation began, plot allotment was done via a series of lotteries between 2000 and 2005, with firms actually moving between 2006 and 2010. This generates random variation in firm presence over the time period, between neighborhoods with a greater number of firms receiving a plot earlier in the process, and those with a greater number of firms receiving a plot later in the process (conditional on the total number of firms relocated from a neighborhood). It also generates random variation within the industrial area on who a firm's neighbors are since, conditional on plot size and lottery year, the allocation of firms to plots was also random, something we verify using simulated lotteries. This type of firm, namely, small and medium manufacturing, predominates in developing countries so understanding their emission profile is key to informing policymakers on the potential effects of policies targeting them.

In the first part of the study, we use the random timing of firm removal to estimate the causal impact of firm presence on neighborhood environmental quality, specifically air pollution. We use a relatively fine definition of neighborhood, a 1km by 1km grid-cell, which is the level at which our air pollution data is measured. This allows us to test whether such policies achieve their environmental objectives. A reduction in air pollution should not ex ante be taken for granted because most developing countries have limited regulatory capacity. The relocated firms might move back, be replaced by other polluting firms, or pollution may increase due to, for example, the policy's enabling growing vehicular emissions. We combine several data sources with administrative data on the policy. Controlling for the total number of firms that were relocated in a neighborhood, we compare neighborhoods that on average relocated a higher number of these firms earlier vs. later, with the timing randomly generated by the allotment of plots to firms via lottery.

The results show that the relocation lottery timing has a strong first-stage effect on the time a firm takes possession of their plot in the industrial area. This randomized timing of relocation is what we use to identify the impact of firm presence from sending regions on air pollution. We find that the average neighborhood impacted by the relocation experiences a 1.7 $\mu g/m^3$ drop in particulate matter (PM) levels, i.e. by about 1.6 % of the mean fine PM concentration for the average neighborhood. Since industrial pollution contributes about 20% to Delhi's PM 2.5 (Sharma et al., 2018), relocation reduces industrial pollution in Delhi by 8% for the average neighborhood.

We conduct several robustness checks on this result, some of which we mention here. First, we show that the latitude and longitude of sending regions with early vs. late relocation are similar. That is, neighborhoods' location does not predict early vs. later relocation, as we would expect given that the timing is decided randomly. Second, we construct standard errors using randomization inference, testing the sharp null hypothesis that the actual allocation had no effects. These p-values are similar to their unadjusted counterparts. Third, we conduct a placebo test using firms that had applied for but been deemed ineligible for relocation, and show that their presence in a neighborhood does not predict air pollution changes post the relocation policy. This indicates that our results are not driven by neighborhoods with more early relocation by lotteries also undertaking other regulations to reduce pollution concurrently.

In the second part of the paper, we take advantage of the random assignment of firms to plots (stratified by plot size and year of the lottery) to identify the impact of neighboring firms on firm survival. The largest industrial area, which contained about 15,000 firms had 74 clusters of 150-300 firms, and we use within-cluster variation in the proportion of upstream, downstream and own product firms to study how cluster composition impacts firm survival. This allows us to causally estimate the direction and magnitude of firm spillovers at a fine level of industrial classification, sidestepping the omitted variables that can incentivize firms of certain types to co-locate in observational data. Indeed, the results show that presence of a particular industry can generate in some instances positive and in other instances negative impacts on firms in other industries.[1]

We then model firms' exit decisions as the outcome of the first period of a dynamic game of incomplete information played in each cluster of the largest industrial area, using our causal spillover effect estimates to approximate conditional choice probabilities. We leverage the structure of the model to generate counterfactual exit rates under optimal assignment of firms to plots, using firm spillovers optimally to maximize aggregate firm survival. Compared to the lottery policy's uniform assignment of industries to clusters, the optimal assignment which groups industries with positive spillovers would have increased aggregate firm survival by 12 - 19 percentage points.

Finally, we conduct a back of the envelope cost-benefit exercise for the policy. Relocation reduces air pollution in the sending regions, but is costly for the relocated firms. 74% of firms in the largest industrial area comprising about 15,000 firms had exited by 2018. That is, only 26% had survived. To estimate survival probabilities for relocated firms that were caused by the policy (rather than a "natural" rate of exit), we use the distance firms are relocated. Every kilometer relocated reduces the probability of firm survival by 0.4-1 p.p, indicating the importance of endogenous location choice and the costs of being moved from that location. Using this estimate, we find that relocation caused an increase in relocated firms' exit rate of about 30 percentage points. In a back of the envelope cost-benefit analysis, which converts the reduction in PM levels to the statistical value of lives lost and compares this to costs associated with firm death, we find that the benefits outweigh the costs. Notably, optimal assignment of firms to plots taking into account spillover effects would cut the effect of relocation on firm exit roughly in half.

The paper builds on several literatures, and lies at the intersection of the impact of environmental policies on firms as well as the returns to agglomeration. First, we contribute

[1] This is in contrast to more aggregated agglomeration analyses which are typically unable to differentiate between positive and negative spillovers (e.g. Ahlfeldt et al. (2015)). We show that a firm's position in the input-output matrix relative to its neighbors, along with net competitive effects, explains the large majority of spillover effects it experiences.

to the literature on the impact of environmental regulation targeting firms (Blackman et al., 2010; Chen et al., 2021; De Simone et al., 2024; Do et al., 2018; Fan et al., 2019; Fenske et al., 2023; Foster and Kumar, 2011; Greenstone and Hanna, 2014; Harrison et al., 2015; He et al., 2020; Karplus et al., 2018; Muller and Mendelsohn, 2009; Ryan, 2012; Shapiro and Walker, 2020; Song et al., 2022).[2] Most of these studies focus on natural experiments that evaluate a bundle of pollution reduction measures on environmental quality, and in some instances, firm outcomes. We study a unique experiment allowing us to identify long-term impacts of a widely-used type of environmental regulation, one that forces polluting firms to move out of populated neighborhoods. Such policies are common in low state capacity settings, but relatively little is known about their impacts. Using the randomized timing of removal for identification also allows us to estimate the causal impact of firm presence on environmental quality, as well as test how firms' spillovers on each other can be used to reduce the costs of environmental policies on regulated firms. A related but distinct literature studies how environmental regulation that does not explicitly target firms' location decisions nevertheless impacts entry and exit decisions (Henderson, 1995; Levinson, 1996; List et al., 2003). These include tests of the pollution haven hypothesis, which posits that firms move to areas with lower environmental standards (Brunnermeier and Levinson, 2004; Choi et al., 2023; Jaffe et al., 1995; Millimet and Roy, 2016; Tanaka et al., 2022). In contrast, we study how environmental policies that target firm location impact environmental quality and costs for firms.

Second, we are part of a growing literature investigating the nature of agglomeration economies for firms (the classical foundation of this literature being Rosenstein-Rodan (1943), Murphy et al. (1989), Glaeser et al. (1992), Ciccone and Hall (1996), and Henderson et al. (1995) also see Combes and Gobillon (2015), Rosenthal and Strange (2020), Duranton and Puga (2020), and Bryan et al. (2024) for excellent reviews of empirical applications in this literature). This empirical literature focuses on measuring the magnitude of such effects (Baum-Snow et al., 2023; Combes et al., 2012; Ellison et al., 2010; Gechter and Tsivanidis, 2020; Greenstone et al., 2010; Kline and Moretti, 2014; Leonardi and Moretti, 2023; Tsivanidis, 2023; Vitali, 2022). This literature is largely (though not exclusively) focused on developed countries, though the nature and magnitude of such effects might vary in developing countries due to differences in firm size, market frictions, and other government policies (Gechter and Tsivanidis, 2020; Vitali, 2022). Furthermore, in this literature, experimental variation is (unsurprisingly) almost entirely absent. Nakajima and Teshima (2017) is an exception, using random assignment of firms within a fish market in Tokyo to

[2] Also related is recent experimental work that randomizes particular aspects of environmental regulation such as auditor discretion Duflo et al. (2018) or emissions trading (Greenstone et al., 2022).

investigate the specific channel of shopping externalities. The relocation policy we study affects firms in a variety of manufacturing industries in Delhi, and the random assignment of firms to specific plots in the industrial area allows us to separately identify the effects of these industries on each other's profitability (for instance, upstream vs. downstream relationships), similar to Ellison et al. (2010) but with the benefit of experimental variation. Moreover, we also estimate how firm presence impacts environmental quality.[3]

The rest of the paper is organized as follows. Section 2 describes the context and Delhi's relocation policy. Section 3 details the data used. Section 4 describes the empirical strategy and results for impacts of the relocation policy on the neighborhoods firms left. In Section 5 we discuss our results on the effect of industrial neighborhood composition on firm exit and use these results to allocate firms to neighborhoods. Section 6 describes our back-of-the-envelope cost-benefit analysis. Section 7 concludes.

2 Context and Relocation Policy

2.1 Air Pollution in India: Impacts and Related Policies

Air pollution is a leading risk factor for premature death, accounting for over 8 million deaths in 2021 (HEI, 2021). It is a highly significant public health concern in India, where nearly 100% of the population lives in areas where annual fine PM concentrations exceed WHO recommendations (Greenstone and Fan, 2018). About 1.6 million deaths, 17% of all deaths in India, were attributable to air pollution in 2019 (Pandey et al., 2021), demonstrating the significant health impacts of persistent high pollution exposure.

India has several comprehensive environmental laws, including the Water Act of 1974 and Air Act of 1981. Beginning in the 1980s and 1990s, several pollution reduction policies were initiated, a primary one being the Supreme Court Action Plans (Greenstone and Hanna, 2014) (SCAPs). These plans were comprehensive measures that 17 polluted cities were directed to take by the Supreme Court, and comprised bundles of pollution reduction measures such as fuel switching or firm closures and relocation. Mandates regarding

[3] A second direction in this literature evaluates whether place-based policies such as industrial areas and enterprise zones truly improve firm outcomes, or only displace economic activity from regions not covered by the place-based policy to regions covered by the place-based policy (see Glaeser and Gottlieb (2008) and Neumark and Simpson (2015) for excellent reviews of this literature; other related work studies place-based policies in developed (Busso et al., 2013; Criscuolo et al., 2019; Greenbaum and Engberg, 2004; Hyman et al., 2022; Neumark and Kolko, 2010; O'keefe, 2004) and developing economies (Betcherman et al., 2010; Chaurey et al., 2023; Hasan et al., 2021; Lu et al., 2019)). Our estimates of firm spillovers do not have this concern of displacement causing an upward bias in any firm spillover estimates, since we make use of a large and known set of firms that were moved and co-located.

the location of polluting industries were a primary means to reduce industrial pollution mentioned in these action plans. For instance, 13 of 17 Action Plans mention environmentally driven location mandates like closure or industrial relocation (Harrison et al., 2015). This focus on regulating firms' location choices to reduce industrial pollution continues in the present. The National Clean Air Programme (NCAP) policy is currently the flagship air pollution reduction policy initiative by the central government, and calls for city-specific abatement plans to reduce air pollution. Relocation of polluting industries continues to be a commonly used tool in these plans.

2.2 Relocation Policy In Delhi

In 1999, the Supreme Court mandated the relocation of manufacturing firms in Delhi that were operating in residential areas, with exemptions for certain types of household industries. The Government started developing three industrial areas on the edges of the city to house these firms (a very small fraction of firms, about 2.5%, were given plots in other industrial areas around Delhi). Firms were asked to apply to the program to be allotted a plot in one of these areas. Only firms that had been registered before 1999, were in a residential neighborhood, and were in an industry that was to be relocated were eligible (Singh, 2007).[4] These firms comprised a large range of small manufacturing firms, including automobile parts, food processing, and rubber and plastics producers. Table 4 presents the most frequent goods produced by these firms.

There was some initial confusion about eligibility for relocation, and 50,000 firms applied to the program (because if they were relocation-eligible and did not move, they would have to shut down operations altogether). The government deemed that about 21,000 were actually eligible for relocation and should be assigned plots in the industrial areas (we use firms that applied to the program but were not relocated in a placebo check, which we detail in Section 4). Our primary analysis focuses on randomization within the 21,000 firms, and so is internally valid over the relevant sample of relocation-eligible firms. However, since the number of industrial plots was limited (more were developed over time), plots were allocated via a series of lotteries. The plot assignment lotteries were simple random draws conditional on the year and plot size category from the pool of eligible firms. For instance, the first lottery assigned plots to about 11,000 firms in the first industrial area in the year 2000 in one of four size categories (100 m^2, 150 m^2, 200 m^2, and 250 m^2). The next large lottery, in 2003, assigned plots to 3,687 firms that had not yet

[4]Some neighborhoods were deemed to be exempt from relocation, since they were not adequately residential.

been assigned a plot, and so on. In total, there were 4 large lotteries (with over 2,000 firms each relocated), and many small ones between 2001 and 2015.

Most firms assigned a plot by lottery up to 2004 were assigned to two industrial areas. In 2005, a third industrial area became operational, and the majority of firms assigned a plot in 2005 or later were assigned to this third area. This difference caused a significant delay in a firm moving however- the average year in which a firm relocated early (before 2004) took possession of their plot in the industrial area is in 2005, while the average year in which a firm relocated later (after 2004) received a lease was 2014, a nine-year delay.

Firms that were not allotted an industrial plot in the earlier lotteries could continue operating while they waited for a plot, and once allotted a plot had to move their operations within 3 years. Leases began to be given for the industrial areas in 2006. We will show that firms that "won" the lottery earlier also were able to take possession of their plot earlier (the first step in moving, which was followed by a lease being issued), generating exogenous variation in the timing of their departure from their original location. Given the random timing of the lottery, this generates random variation in firm presence that we use to estimate impacts on pollution (more details are provided in Section 4).

The majority of relocation lotteries were done by 2005, and only about 500-600 firms received an assignment 2010 or later. Each firm was allotted an industrial plot ranging from 28 m^2 (which was a spot in a building housing several small firms) to standalone plots of 250 m^2 - the average plot size was between 100 and 150 m^2. Firms were given concessional loans to allow them to build their factories in the allotted plots, and were given leases for these plots. They were not allowed to sell or rent them, and were technically supposed to continue producing the same products they had done while located within Delhi.[5]

Of the three industrial areas, the largest one (Bawana) housed the majority of relocated firms (over 15,000), and included plots of four size categories (100 m^2, 150 m^2, 200 m^2, and 250 m^2, comprising 50.9%, 26.3%, 5.1% and 17.69% of the data, respectively). We were able to obtain stylized maps of Bawana which we used in combination with Google Earth to get the coordinates of a firm's final allotted plot location. After additionally geocoding their original address, we estimate how far firms were moved. We find that the average firm relocated to Bawana was moved 20.2 km (12.5 miles) from its original location– the 25th percentile is 16.2km (10.1 miles), while the 75th percentile is 27.3km (17 miles).

All the industrial areas were developed by the government body charged with implementing the relocation policy, and within the industrial area the plot allocation was

[5]The 2018 Firm Survival Census in Bawana (described in more detail in the next section) showed that only about 1% of firms changed trade, so this was a relatively rare occurrence.

random as well (conditional on year of lottery won and plot size category. e.g. amongst firms to be assigned a plot of 100 sq m in 2000). Figure A1 shows Landsat images of Bawana. Though somewhat blurry, it shows little to no development. Figure A2 shows Landsat (left panel) and Google Earth (right panel) images. The Google image is clearer, but we include both for completeness. It shows plot boundaries being delineated but no firm movement yet. Finally, Figures A3 and A4 show limited and then significant development of the industrial area in 2005 and 2010, respectively.

3 Data

3.1 Air Pollution

To estimate impacts on pollution, we need high-resolution data on pollution before the relocation began (in 2000) as well as after. This is challenging, since pollution monitors were relatively sparse for Delhi during this time period, with only one air quality monitor reliably reporting data over the entire this period. We use data on fine particulate matter (PM 2.5, or fine PM) created by van Donkelaar et al. (2016). These data are constructed by combining satellite retrievals of aerosol optical depth, chemical transport modeling, and ground-based measurements (for more details, refer to Van Donkelaar et al. (2021)). They have been extensively in prior work on the impacts of pollution as well as the impact of regulation on pollution, recently by Greenstone and Fan (2018), De Simone et al. (2024), Barrows et al. (2019), and Behrer et al. (2023).

These data have several advantages for our analysis. First, they are available monthly at the 1km by 1km resolution, providing data at a fine spatial resolution and starting before the policy began to be implemented. Second, they cover fine PM, which is an important air pollutant, responsible for large damages to human health globally (McDuffie et al., 2021). We use the 1km by 1km grid cells as our definition of neighborhood, since this is the finest level for which pollution data is available, and create annual measures which is our outcome variable for environmental quality.

3.1.1 Latitude and Longitude of Sending and Receiving Addresses

Geocoding the original addresses using Google's API, we assign a firm to a neighborhood. Combining this information with the timing of when a firm was given a plot, we create a neighborhood-level dataset of the number of firms that were allotted a plot via lottery each year between 2000 and 2005 (95% of firms were assigned a plot via lottery by then), as well as the number of firms whose lease began in each year between 2006 and 2010.

Figure 1 presents the cumulative probability of having won a plot lottery by a given year as well as the probability of having taken possession of the industrial plot by year, showing a positive relationship between the two measures. We present regression results in the next section. We were able to identify the longitude and latitude of plots within Bawana by digitizing maps we acquired from industrial associations, and cross-referencing street names in these stylized maps with Google Earth. Figure A5 shows a photograph of the map for one cluster, and Figure A6 a stylized map of a set of clusters.

3.2 Administrative Data on Relocation Policy

The administrative data on the relocation policy is available from the government body that was responsible for the relocation, the Delhi State Industrial and Infrastructure Development Corporation Ltd. (DSIIDC). The data include firm name, original address from where they were relocated, details such as applicant name, date of the lottery when they were allotted a plot, a free text description of the firm's products entered by the owner, as well as the ID of the plot they were allocated in the industrial area. It also includes details on the timing of the dates when the firm took possession of the industrial area plot and when the firm's lease began.

21,748 firms received a plot assignment in an industrial area and 21,174 have non-missing information on the lottery year of the firm. In Table A1, we show that the availability of lottery year information is not correlated with neighborhood latitude and longitude, i.e. it is uncorrelated with firms' starting geography.[6]

3.2.1 Firm Survival Census in 2018: Bawana Industrial Area

In 2018, DSIIDC conducted a census of all plots in the Bawana industrial area, to determine whether the firm assigned to a plot was still operating there. We classify an assigned firm as being present if it was found in the industrial area, and having exited the market if it was found to be closed in this census, or to have (illegally) rented or sold its plot to another firm (a much smaller fraction, about 1%, changed their trade). This definition gives us the figure of about 26% of firms surviving in the industrial area by 2018.

[6]For another 8.6% of firms, the Google geocoding places them quite outside of Delhi. We show that the propensity for this to happen or for the geocodes to be missing altogether is uncorrelated with the timing of the lottery in Table A6, which we discuss further in Section 3.3.1.

3.2.2 Product Classification

The firm application data include free text information on what firms produce. We determine the 3-digit Annual Survey of Industries Classification Code (ASICC) to assign to each firm a product code through a combination of human judgment and a custom GPT-4 pipeline. We discuss the full details and tailored prompts used in Appendix C. From this exercise, we were able to get narrow (3-digit) assignments for 9,200 of about 15,800 firms in Bawana.

For each firm, we create a measure of what proportion of other firms in their assigned cluster in the industrial area are producing the same product. To understand upstream and downstream linkages, we use the 2010 Annual Survey of Industries (ASI) data. From the 2010 ASI, we first retain only firms producing a single product (to avoid the issue of assigning inputs to multiple outputs, which, like other firm panel datasets, is not available in the ASI), and only keep 3-digit products that have at least five firms producing them. We create product-level expenditure on each input, and generate the proportion of input expenditure for each input.

We call a product x upstream of product y if the input expenditure producers of y spend on x is above the median amount producers of y spend on any input in the 2010 ASI data. For each firm, we create a measure of what proportion of other firms in their assigned cluster in the industrial area are upstream of them. Similarly, we use product-level revenues from other products to generate a measure of downstreamness. We call a product x downstream of product y if producers of y's revenues from selling to producers of x are above the median of producers of y's revenues from selling to any other type of producer (total revenue of each product is taken from the output module, since products could be sold to firms producing other products or directly to customers). For each firm, we create a measure of what proportion of other firms in their assigned cluster in the industrial area are downstream of them. The sum of proportion upstream and proportion downstream give us the proportion of firms in the industrial area that have any input-output linkage with each firm. Our results are robust to alternative definitions of upstream and downstream, including replacing the median with the 75th percentile, which we present in Section 5.2.

3.3 Primary Data Collection

3.3.1 Surveyor Visits to Baseline Addresses of Relocated Firms

To ensure that any geocoding measurement error on assigning firms to neighborhoods is independent of lottery timing, we collected data for about 15,000 firms which were geocoded in-person by surveyors (about 71.4% of the total relocated sample).[7] The probability that geocodes are missing from either of these sources is not correlated with the year of the lottery (see Table A6, and omitting them does not impact the results for pollution.). Moreover, our neighborhood level measure of early relocation, which is the number of firms assigned a plot by lottery by 2004 according to Google maps and the surveyors are highly correlated, with a correlation of 0.97.

4 Impact of Firm Presence on Environmental Quality: Empirical Strategy and Results

4.1 Balance Tests: Firm Original Location

Our use of lottery timing as an exogenous factor in determining the level of firm presence in a neighborhood depends on the fact that lottery winning is truly independent of neighborhood characteristics. To test whether lottery winning time is uncorrelated with firms' starting geography, we run the following pair of regressions at the firm level.

$$\begin{aligned} \text{Initial Longitude}_j &= \alpha + \kappa \text{Year of Lottery Win}_j + \epsilon_j \\ \text{Initial Latitude}_j &= \tilde{\alpha} + \tilde{\kappa} \text{Year of Lottery Win}_j + \tilde{\epsilon}_j \end{aligned} \tag{1}$$

If lottery winning time is unrelated to firm j's original address, κ and $\tilde{\kappa}$ should be estimated as zeros.

The results from estimating Equations (1) are given in Table 1. The coefficient estimates are in degrees, and as expected are very small and statistically insignificant. This indicates that the winners of specific lotteries were not, as expected geographically clustered. We also test for pre-trends in air pollution using event studies, which we discuss in Section 4.3. Third, we simulated each of the yearly lotteries, and compared the actual vs. simulated number of firms relocated early (before 2005) at the neighborhood-level. The actual and simulated number of firms relocated early are highly correlated at the neighborhood-level (correlation of over 0.95).

[7] We sent surveyors to about 15,811 firms but could not find about 1,328 addresses.

4.2 The Impact of Lottery Timing on Movement to the Industrial Area

This section discusses first-stage estimation at the firm-level, to test that the timing of receiving a plot in the industrial area in the lottery process drove the timing of moving to the industrial area. We estimate the following equation for timing outcomes related to movement to the industrial area, analogous to the balance equation (1):

$$Y_j = \alpha + \mu \text{Year of Lottery Win}_j + \epsilon_j \tag{2}$$

where Y_j is the year in which the firm took possession of the industrial plot. This action is the first that firms undertake to begin construction on their assigned plots, and move to the industrial area. The coefficient of interest is μ, which measures the marginal impact of a one-year delay in "winning" the lottery for firm j.

Results from estimating Equation 2 are presented in Table 2. The first two columns show that a year one delay in getting a plot assignment in a lottery causes just over a year's delay in the firm taking possession of the plot. Column 1 presents the results for all firms, and column 2 for firms that had received a plot assignment by 2005 (96% of the data). Columns 3 and 4 show the results using a dummy variable for whether the firm received a plot by 2004 or earlier, with the same samples as in columns 1 and 2, respectively. If a firm did not receive a plot in the earlier lotteries, this delayed it taking possession of the plot by 9 years, indicating a large gap between firm removal for firms that received a plot in an early vs. late (in 2005) lottery (this was because a third industrial area had to be constructed to accommodate these remaining firms). Thus, these results indicate that the lottery timing generated variation in the timing of firm removal from a neighborhood, allowing us to estimate the causal impacts of firm presence on environmental quality.[8] In what follows, we exploit the large delay in movement caused by the relatively small difference in lottery winning time between firms winning by 2004 and those winning after. This creates a large difference in the time firms left their origin neighborhoods, and minimizes the impact of noise in the orthogonal delay process on our estimates.

[8] We also show first stage results at the neighborhood-level i.e. we estimate

$$\begin{aligned}\text{Log(Firms Taken Plot Possession)}_i = &\alpha + \beta \text{Log(Firms Winning a Lottery by 2004)}_i \\ &+ \nu \text{Log(Total Number of Firms Relocated)} + \epsilon_i\end{aligned}$$

where i denotes the neighborhood. We estimate this equation for the year 2006, as well as nine years later when the third industrial area is open, in 2015. Results are presented in Table A5. It shows that a one percent increase in the number of firms winning a lottery early increases the number of firms who took plot possession by nearly one percent in 2006, so an almost one-for-one increase. Nine years later, in 2015, the first stage estimate is about half that, as firms continue to be moved to the industrial areas over time.

4.3 Air Pollution

At the neighborhood level, our primary outcome of interest is air pollution i.e. fine PM concentrations. We estimate a specification that only uses the randomized removal for identification of firm removal on air pollution. We estimate the following event study specification:

$$\begin{aligned}\text{Fine PM}_{it} =& \alpha + \beta_t \sum_{t=1998}^{t=2015, t\neq 2005} \text{Log(Firms Relocated by 2004)}_i \times \mathbb{1}(\text{Year} = t) + \\ & \mu_t \sum_{t=1998}^{t=2015, t\neq 2005} \text{Log(Total Number of Firms Relocated)} \times \mathbb{1}(\text{Year} = t) \\ & + \psi_i + \tau_t + \epsilon_{it},\end{aligned}$$

where i denotes a neighborhood and t the year of measurement. The main coefficients of interest are the β_ts, which measure the marginal impact of a greater number of firms winning a lottery earlier in the process i.e. by 2004, in each year. This timing is exogenous due to the lottery governing the allotment of plots, and the fact that a delay in winning the lottery resulted in a delay in the beginning of a firm's presence in the industrial area on average (as shown in Table 2).

We control flexibly for the total firms relocated by including interactions of this variable with year fixed effects. In addition, we also include neighborhood and year fixed effects (ψ_i and δ_t, respectively). The average year in which firms who had received industrial plots prior to 2005 took possession of their plots is 2005. Therefore, we omit the interaction terms for the year 2005.

We present the event study estimates in Figure 2. Neighborhoods with a greater number of firms winning lotteries by 2004 do not have differential pollution trends before 2005. However, they exhibit lower levels of pollution after, until locations with fewer lottery winners before 2005 catch up by 2012. As a robustness check, results from the same specification, but restricting to neighborhoods where surveyors found at least one origin address of a relocated firm are presented in Table A8, and are consistent with the main sample (Note that while surveyors were only assigned to 70% of the addresses, they were assigned randomly, which implies that these neighborhoods are likely those with more dense economic activity.).[9]

[9]The event study estimates show some evidence of reversion of treatment effects for the last three years, though this is less pronounced when considering neighborhoods with more density of relocated firms (Figure A8). This convergence could also be partially driven by the fact that the first stage is weaker across time,

Aggregating the event study results to a DID, we use the following specification

$$\begin{aligned}\text{Fine PM}_{it} =& \alpha + \beta \text{Log(Firms Relocated by 2004)}_i \times \mathbb{1}(\text{Post 2005}) + \psi_i + \tau_t \\ & + \mu \text{Log(Total Number of Firms Relocated)} \times \mathbb{1}(\text{Post 2005}) + \epsilon_{it}\end{aligned} \tag{3}$$

where the variable definitions are the same as in the previous equation. $\{t > 2005\}$ is a binary variable that takes the value 1 if the year is 2006 or later, and 0 otherwise. We analyzed a fine grid of alternative specifications of Equation 3 using different Box-Cox transformations of Firms Winning a Lottery by 2004 and Total Firms Relocated with λ parameters between 0 (log transformation) and 1 (no transformation). The log transformation consistently maximized goodness-of-fit. As in the previous equation, standard errors are clustered at the neighborhood-level.

Finally, we also run an alternative specification that uses the firms that applied for relocation but were deemed ineligible as a placebo. We estimate the previous equation, and additionally include the interaction of the log of number of ineligible firms and the post-2005 dummy variable. We expect this coefficient to be statistically indistinguishable from zero, since the presence of these firms (which were not removed by the policy) should not systematically change air quality after 2005.

Results are presented in Table 3, with the odd numbered columns showing results from the main specification and even numbered columns additionally including the placebo interactions. The first two columns show results using all neighborhoods, including those with no relocated firms, and the next two columns show results using only neighborhoods with relocated firms. The results are quite consistent across these columns, showing that a one percent increase in firm removal reduces pollution by 0.5 to 0.6 $\mu g/m^3$. Since the mean number of firms relocated is 32.35, this implies a reduction of 0.5*(log(32.35)-log(1)), or 1.7 $\mu g/m^3$ as a result of relocation. This is 1.6 percent of total PM concentrations for the average neighborhood. Industrial pollution contributes about 20% to Delhi's PM 2.5 (Sharma et al., 2018), which implies that relocation reduced industrial pollution in Delhi by 8% in the average affected neighborhood. Columns 5 and 6 report results restricting the sample to neighborhoods with surveyor-identified firms (the analogous DID estimation to the event study presented in Figure A8), with and without the placebo, respectively. Results are consistent with the main specification, although somewhat larger, with a point estimate for β of 1.13 $\mu g/m^3$.

In addition to the clustered OLS results, we also present randomization inference p-values for the main DID coefficient, β, which conducts 1,000 simulations with alternate

as shown by the neighborhood-level first stage estimates in Table A5. The coefficient in 2006 for log(firms relocated by 2004) is nearly one in 2006 (0.997), and nearly half that (0.502) in 2015.

allocations of the number of firms relocated early. The p-values test the sharp null hypothesis of the actual allocation having no treatment effects (so the p-value is the proportion of times the estimated treatment effect under the hypothetical allocation exceeds actual estimated treatment effect) and are also clustered at the neighborhood-level. These are similar to the OLS clustered results.

The placebo results are consistent with the hypothesis that the presence of ineligible firms should not change air quality differentially post-2005. The coefficients on the interaction of the log of the number of ineligible firms and the post-2005 dummy variable in Columns 2, 4, and 6 are, as expected, statistically indistinguishable from zero.

4.4 Additional Specification and Robustness

In this section, we consider an additional specification to estimate the impact of firm presence and present other robustness checks. This additional specification uses the fact that conditional on a baseline number of relocation-eligible firms, neighborhoods with a higher proportion of early plot assignments should see reductions in pollution relative to those with a lower proportion of early plot assignments.

This additional event-study specification is as follows:

$$\begin{aligned}\text{Fine PM}_{it} =& \alpha + \beta_t \sum_{t=1998}^{t=2015, t\neq 2005} \left[\text{Bin}_i \times \text{Prop Firms Lottery by 2004}_i \times \mathbb{1}(\text{Year} = t)\right] \\ &+ \xi_t \sum_{t=1998}^{t=2015, t\neq 2005} \left[\text{Prop Firms Lottery by 2004}_i \times \mathbb{1}(\text{Year} = t)\right] \\ &+ \pi_t \sum_{t=1998}^{t=2015, t\neq 2005} \left[\text{Bin}_i \times \mathbb{1}(\text{Year} = t)\right] + \psi_i + \tau_t + \epsilon_{it},\end{aligned}$$

where the Fine PM variable definition is the same as in the previous specification. Prop Firms Lottery by 2004 is the proportion of firms in neighborhood i that received a plot assignment by 2004. Bin takes on 1 of 4 values. It is 1 if the number of relocation-eligible firms is in the first quartile, 2 if it is in the second, and so on. We also report results from a version of this variable with decile bins. The parameters of interest are the β_ts. The analogous DID specification replaces the year dummies implicit in the definition of the β_t parameters with a post-dummy 2005, as in the main specification.

Event study estimates using 4 bins are presented in Figure A9, and the DID results in Column 1 of Table A2. The event study estimate does not show evidence of a pre-trend before 2005, and shows a negative impact after. The DID estimate shows that conditional on a baseline number of relocation-eligible firms, a marginal increase in the proportion of firms relocated early causes a 1.2 μg/m^3 reduction in fine PM. Normalizing by the standard deviation of proportion relocated early, this would imply that a one standard deviation increase in the proportion of firms relocated early causes a 0.51 μg/m^3 reduction in fine PM. Event study estimates using 10 bins are presented in Figure A10, and the DID results in Column 2 of Table A2. These are consistent with the specification using 4 bins, showing a negative coefficient on the triple interaction that is precisely estimated.

We also show results for the main DID specification without neighborhood and year fixed effects, using the conventional DID specification. These results are presented in Table A4, and (as expected) are nearly identical to the main results. Finally, as an additional robustness check for the specification that only uses the randomized removal for identification, we test for how spatial correlation may impact the analysis, by clustering at a larger geographic level (2km by 2 km) instead of at the neighborhood-level. Results are presented in Table A3, and are consistent with the main results.

In sum, the results show firm removal does causally and significantly reduce air pollution, a primary source of mortality and morbidity. In Section 6, we consider a back of the envelope calculation that considers how the benefits of firm removal compare to the costs, which we focus on next.

5 Agglomeration and Optimal Industrial Area Design

Random assignment of individual plots to firms (within a year and plot size) generates independent variation in the characteristics of a firm's neighbors. In this section, we estimate the effect of the industrial composition of firm i's neighborhood in the industrial area on its long-term survival. We then use these estimates to search for an exit-minimizing alternative to DSIIDC's uniform assignment of firms to plots in the industrial area. We find that roughly half of the effect of relocation on exit could have been avoided with an optimal industrial area design taking into account neighborhood composition effects.

5.1 Balance Tests for Assignment of Plots in the Industrial Area

We first test whether the plot assignments by product are consistent with random allocation. To simulate the original lotteries, for each year of the lottery and plot size category, we pick a random set of firms and then assign them randomly to available plots of that size category. For instance, if x firms were assigned 100 m^2 plots in 2000, we randomly pick x of all the 100 m^2 firms and randomly assign them to the 100 m^2 plots that were assigned in 2000. This generates simulated data on the proportion of firms producing a product in each of the 74 industrial clusters in Bawana. We present the actual and simulated data in Figure 3. The two distributions are extremely similar, and the Kolmogorov–Smirnov test for equality of distribution functions fails to reject that the two are equal (p-value=0.86). As an additional robustness check, we present results using the total number of firms in a cluster producing a product (rather than the proportion). The results are in Figure A7 and also show that the distributions are very similar (The Kolmogorov–Smirnov test for equality of distribution functions has a p-value of 0.92).

5.2 Neighborhood Composition Effects through Net Competition and Input-Output Linkages

We follow Marshall (1920)'s theory of agglomeration in explaining the effect of the composition of the industrial cluster a firm is assigned to on the firm's survival. Marshall's theory is based on the benefits to firms of minimizing transportation costs in accessing customers, inputs, and ideas. Since the relocated firms largely sell to other firms, we measure the differential access to customers generated by the random allocation of a firm's neighbors by the share of downstream neighbors in the firm's cluster. We measure differential access to inputs by the share of upstream neighbors in the cluster. The share of neighbors producing the same good as the firm could have a negative effect on survival through competition, but also a positive agglomeration effect due to transmission of ideas or a thicker market for the kind of labor required to produce the good. The effects we measure are the net of these two forces, so we refer to them as net competitive effects.

We consider the following partially-linear model of how these cluster composition measures affect the survival of firm i in cluster k, with the parameters of interest being $\alpha_{upstream}$, $\alpha_{downstream}$, and α_{own}:

$$\begin{aligned} Active_{ik} =& \alpha_{upstream} ShareUpstream_{ik} + \alpha_{downstream} ShareDownstream_{ik} + \alpha_{own} ShareOwn_{ik} \\ &+ g\left(\{1\{product_i = m\}\}_{m=1}^{M}, \{1\{Lottery_{ik} = l\}\}_{l=1}^{L}, ShareMissing_k\right) + \epsilon_{ik}. \end{aligned} \tag{4}$$

$Active_{ik}$ is an indicator for firm ik's being found active in the 2018 census of the Bawana industrial area. $ShareUpstream_{ik}$ is the share of cluster k's assigned firms who report producing a good upstream of i's product. $ShareDownstream_{ik}$ is the share of cluster k's assigned firms who report producing downstream of i's product. $ShareOwn_{ik}$ is the share of block k's assigned firms producing the same product as i. $\{1\{product_i = m\}\}_{m=1}^{M}$ is a vector of product indicators, with M being the total number of products. $\{1\{Lottery_i = l\}\}_{l=1}^{L}$ is a set of dummy variables denoting the specific lottery firm i won to secure a plot in the industrial area, i.e. a year-plot size dummy.[10] $ShareMissing_k$ is the share of firms assigned to cluster k for which a product code could not be assigned.[11]

Given that we have 180 distinct 3-digit ASICC codes being produced in Bawana and 35 lotteries, we take steps to reduce dimension to proceed. We do so using Belloni et al. (2013)'s post-double-LASSO method. We approximate the function $g(\cdot,\cdot,\cdot)$ of product and lottery indicators and $ShareMissing_k$ in Equation (4) with fixed effects for each product and lottery, and a linear function of $ShareMissing_k$. We select which fixed effects to include, and whether $ShareMissing_k$ should be included, by running LASSO regressions of our approximation to Equation (4) and the treatment variables $ShareUpstream_{ik}$, $ShareDownstream_{ik}$, $ShareOwn_{ik}$ as specified below.

$$\begin{aligned}Active_{ik} =&\alpha_{upstream} ShareUpstream_{ik} + \alpha_{downstream} ShareDownstream_{ik} + \alpha_{own} ShareOwn_{ik}\\&+\sum_{m=1}^{M}\gamma_{active,m}1\{Product_{ik}=m\}+\sum_{l=1}^{L-1}\lambda_{active,l}1\{Lottery_{ik}=l\}+\psi_{active}ShareMissing_k\\&+r_{active,i}+\epsilon_{ik}\end{aligned}\tag{5}$$

$$\begin{aligned}ShareUpstream_{mk} =&\gamma_{upstream,m}+\sum_{l=1}^{L-1}\lambda_{upstream,l}1\{Lottery_{ik}=l\}+\psi_{upstream}ShareMissing_k\\&+r_{upstream,mk}+\nu_{upstream,mk}\end{aligned}\tag{6}$$

$$\begin{aligned}ShareDownstream_{mk} =&\gamma_{downstream,m}+\sum_{l=1}^{L-1}\lambda_{downstream,l}1\{Lottery_{ik}=l\}+\psi_{downstream}ShareMissing_k\\&+r_{downstream,mk}+\nu_{downstream,mk}\end{aligned}\tag{7}$$

$$ShareOwn_{mk} =\gamma_{own,m}+\sum_{l=1}^{L-1}\lambda_{own,l}1\{Lottery_{ik}=l\}+\psi_{own}ShareMissing_k+r_{own,mk}+\nu_{own,mk}.\tag{8}$$

$(r_{active,i}, r_{upstream,mk}, r_{downstream,mk}, r_{own,mk})$ represent errors due to the linear approximations. Following Belloni et al. (2013) we assume sparsity: the number of nonzero terms in

[10]It is possible that the size of the industrial cluster matters in forming relationships between firms. In specifications interacting the share variables with the log of number of firms in a cluster we do not find statistically significant heterogeneous effects, hence this choice of specification. These results are omitted for brevity but available upon request.

[11]The correlations between the proportion of firms with missing products and proportion of upstream, downstream, or own products are all small, less than 0.1.

Equations (5), (6), (7), and (8) needed to make the approximation errors small is also small relative to the sample size. The product fixed effects $\gamma_{upstream,m}$, $\gamma_{downstream,m}$, and $\gamma_{own,m}$ in the LASSO regressions for $ShareUpstream_{mk}$, $ShareDownstream_{mk}$, and $ShareOwn_{mk}$ control for the overall share of the product, as well as the share of firms upstream and downstream of the product among all relocated firms. The other effects are included for completeness, almost never selected by LASSO, and have no substantial impact in the handful of cases in which they are.

Results are presented in Table 5, where we have divided the composition variables by their standard deviation. Column 1 shows that a one standard deviation increase in the proportion of firms with any input-output linkage in a cluster increases the probability of firm survival by 4 p.p., about 15% relative to the mean survival rate. In contrast, a one standard deviation increase in the proportion of firms producing the same product has a small, negative, and imprecisely estimated impact on survival. The second column shows the impact of input linkages and output linkages separately. Both upstream and downstream linkages matter, but the point estimate on upstream linkages is bigger, about 5 p.p, or 20% relative to mean survival rate (however, we cannot reject that these two coefficients are equal). We are able to reject that upstream linkages have the same effect as own-product firms, and the coefficient on own product firms is, as in Column 1, small and negatively impacts firm survival. These results show the significant impact that input-output linkages can have in improving firm survival. Columns 3 and 4 present results using the 75th percentile upstream and downstream threshold rather than the median. Results are similar Columns 1 and 2, respectively. Thus, input-output linkages have significant spillovers on firms' long-term survival probability. As described in the next subsection these estimates allow us to generate counterfactual estimates of how much the costs to firms can be used by leveraging these spillovers.

The specification above assumes that relevant firm spillovers occur through Marshallian forces: input-output linkages and net competitive effects. Is this consistent with the data? In Appendix B, we estimate a much more general model allowing for arbitrary and asymmetric spillovers from the share of producers of each product n on the exit decision of a producer of product m. In this model, each of these spillovers can be determined by the input-output relationship between m and n, by net competition, or by features idiosyncratic to the pair. To address the high dimension of the model, we take a Bayesian approach to inference and find a posterior mean magnitude of the idiosyncratic effects that makes up only 4% of the magnitude of all forces driving spillover effects, with a 90% credible interval of [0.004, 0.138]. This shows that Marshallian forces, as we have modeled them in this section, are indeed driving the large majority of spillover effects across firms.

5.3 Optimal Assignment

5.3.1 Model of Firm Exit Decisions

We can use the spillover form estimates from Section 5.2, which estimate the impacts of proximity of firms producing the same product as well as firms with input-output linkages, to evaluate the aggregate effect of alternatives to the uniform assignment of firms to plots in the largest industrial area. To do so, we model firms' decision to remain active in each cluster of the Bawana industrial area as the first period of a dynamic game of incomplete information. Uniquely for the dynamic games literature, we observe the initial condition of each game as the random assignment of firms to the cluster (see e.g. Berry and Compiani (2021) for issues arising from the unobservability of the initial state in dynamic games).

Firms are defined by the same characteristics as in Equation (4): the ASICC code of the product they produce and the year-size-allotted combination of the lottery they won. Size allotted proxies for firm size. In addition to these observed characteristics, following Benkard et al. (2020) we assume each firm experiences a profit shock ν_{ik} with support $\mathcal{V}$, which is unobserved to us as researchers and to the other firms. We denote the initially-assigned vector of producer shares for cluster k by $s_k \in \Delta^{M-1}$, where Δ^{M-1} is the $(M-1)$-dimensional simplex. This vector of shares constitutes the state variable in the model summarizing conditions following entry of the assigned firms in cluster k. Given s_k, $Product_{ik}$, $Lottery_{ik}$, and ν_{ik}, firm ik chooses whether to remain in operation or exit.

The strategy for firm i is the function $\sigma_{ik} : \Delta^{M-1} \times M \times L \times \mathcal{V} \to \{0, 1\}$, where 0 represents the choice to exit and 1 the choice to stay active. Conditional on the set of firms assigned to each cluster, we assume that a pure strategy Bayes Nash Equilibrium (BNE) exists and that the same BNE is played across all clusters in the data. Under this setup the conditional probability of choosing to stay active is given by the following expression,

$$
\begin{aligned}
&P(Active_{ik} = 1 | s_k, Product_{ik} = m, Lottery_{ik} = l) \\
&= \int \{\nu : \sigma_{ik}(s_k, m, l, \nu) = 1\} dG(\nu | s_k, Product_{ik} = m, Lottery_{ik} = l),
\end{aligned}
$$

the measure of the set of ν values such that firm i chooses to stay active. Uniform random assignment of firms to plots also means regions of the support of the conditioning set of the choice probabilities will be well-represented, which can be an issue in applications of Benkard et al. (2020)'s methodology. More importantly, random assignment of firms to clusters make the conditional choice probabilities causal in the sense that they reflect the expectation of *potential* $Active_{ik}$ with s_k set to s, $Product_{ik}$ set to m, and $Lottery_{ik} = l$. Differences in conditional choice probabilities thus represent treatment effects, not selection.

We can therefore arrive at counterfactual exit rates under the same equilibrium played in the data by varying the values of the conditioning variables in estimates of $P(Active_{ik} = 1|s_k, Product_{ik} = m, Lottery_{ik} = l)$. Our estimates from Table 5 yield linear approximations to these conditional probabilities which we will now use to arrive at a set of assignments of firms to clusters which maximize firm survival.

5.3.2 Optimal Counterfactual Assignment

To solve for the optimal survival-maximizing assignment of firms to clusters, we formulate the following optimization program:

$$\max_{s\in[0,1]^{(M-1)\times K}} s' \left(\begin{bmatrix} \frac{1}{p_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{p_2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{p_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{p_K} \end{bmatrix} \otimes \delta \right) s$$

$$\text{subject to} \quad \sum_{m=1}^{M-1} s_{mk} \leq p_k \; \forall k \qquad \sum_{k} s_{mk} = S_m \; \forall m. \tag{9}$$

p_k is the fraction of plots in cluster k, S_m the total share of firms of industry m, and s_{mk} the share of product m in cluster k times p_k. δ is the matrix of effects of assigned ASICC producer shares on the probability of each type of producer remaining active, with rows identifying the product the affected firm manufactures and columns the share of each product. We arrive at each product-pairwise effect δ_{mn} by taking the Marshallian effects from Table 5 and applying them to each product pair according to their input-output relationship. For aggregate exit, only the average pairwise spillover effect matters so that we can set $\delta_{nm} = \delta_{mn} = 1/2(\delta_{mn} + \delta_{nm})$ in the δ matrix we use in optimization.

We work with all 180 3-digit product codes with 28 or more firms assigned, which together make up 90% of all firms assigned a 3-digit code, and an "other" category. Figure 4 plots the matrix of net spillover effects across these product codes, and Figure 5 the matrix of spillover effects after symmetrizing as in the previous paragraph. Problem (9) is not convex because δ is unrestricted, and in fact the matrix in Figure 5 is indefinite. To solve the problem, we perform a log barrier reformulation where for each value of the barrier parameter we solve a Karush-Kuhn-Tucker system using Newton's method. We implement this in KNITRO, using its algorithm for selecting initial values and stopping when the probability of finding an unobserved solution reached 1%.

Figure 6 illustrates our optimal assignment. The shares assigned to each cluster by product are indicated by the color of the four squares surrounding each intersection of gray lines, with the horizontal lines representing products and the vertical lines representing industry clusters. The numbers represented by each row add to 1. Comparing with Figure 5, we see that products with input-output linkages implying a net positive cross-industry effect on the probability of remaining active through 2018 tend to be grouped together in clusters. For instance, product codes 561 and 571 ("printed books, newspaper, periodicals, note books, register etc and other printed matters" and "packing materials made of paper", respectively), whose net spillover implies a strong positive effect on the probability of remaining active, are often in the same cluster.

Relative to uniform assignment, the optimal assignment illustrated in Figure 6 decreases aggregate exit by 19.2 percentage points. This is number is robust to alternative definitions of upstream and downstream. If we call a product x downstream of product y if y's revenues from selling to x are above the 75th percentile of y's revenues from selling to any other type of producer, instead of the median, and do the same for the definition of upstream, our assignment reduces aggregate exit by 12.3 percentage points.

We also check whether the performance of optimal assignment is reliant on extrapolation beyond the range of producer shares of clusters we observe in the data. If we further constrain the optimal assignment problem so that at most 9% of the producers of a product are assigned to any one cluster (the maximum we observe in the data), aggregate exit falls by 15 percentage points relative to uniform assignment. Therefore, leveraging firm spillovers emerges as a robust solution to mitigate a substantial amount of the destructive effects of industrial relocation.

6 Back of the Envelope Cost-Benefit Analysis

6.1 Benefits

In this section, we construct a back-of-the-envelope (BOTE) estimate of the impacts of firm relocation (we note here that these are indicative estimates, and rely on prior work to estimate how pollution reductions impact mortality). To estimate the benefits of the program, we consider benefits from avoided mortality as a result of reduced pollution exposure, which is a first-order benefit of pollution reduction. These are a function of three factors. The first is the impact of firm removal on air pollution, which we have already estimated. We use the most conservative coefficient for log(firms receiving a plot assignment by 2004) of 0.5 μg/m^3 from Table 3, and assign each 1 km2 neighborhood with a relocated firm a pollution reduction of log(total firms removed) times 0.5 (since it is a pollution change of (log(total firms removed)-log(1))*0.5 μg/m^3). This has the benefit that neighborhoods that removed more firms experience a greater increase in the health benefits. The second factor is the population exposed. We assign each neighborhood the 2011 Population Census population density of the Census ward it lies in- the average density for neighborhoods with any relocated firms is 19,677.16 people per sq km, which reflects the large numbers of people exposed to pollution in Delhi. The third factor is the mapping between the change in exposure to mortality risk.

To translate the change in exposure to reduction in mortality risk, we use estimates from prior work. There is a range of these estimates, and we consider two possibilities. The first estimate relies on Pope III et al. (2020) which conduct a meta-analysis. The studies they include from Asia show that a 10 $\mu g/m^3$ reduction in exposure to fine PM reduces all-cause mortality risk by 5% (1.05 hazard ratio). Scaling that linearly gives a mortality reduction estimate of 0.5% per μg/m^3. Aggregating up the reduced mortality risk in each grid square by multiplying these three factors yields an estimate of 82,823.47 lives saved annually. Using a $700,000 Value of Statistical Life (VSL) (Majumder et al., 2018), the mortality benefits of the policy are about $57.98 billion. The second possibility is that the exposure-to-mortality relationship is concave, especially at high levels of pollution (Cropper and Park, 2022), and so a small reduction in pollution does not have any discernible health impacts, in which case the program did not have any benefits. We discuss each case in the benefit cost comparison below.

How do the estimates from the first case change once we account for the increase in pollution in the Bawana area? If we assign all the firms to Bawana, using the estimated reduction in pollution from firm removal would imply that air quality worsened by 0.5*(log(21,000)-log(1)), or 4.98 $\mu g/m^3$. The population of Bawana in 2011 was about 73,000, and applying the mortality reduction estimate from the first possibility above gives an impact of 1,754.89 lives lost. This reduces the benefits of the policy to fall to 81,068.58 lives ($56.75 billion). Thus, while the benefits of the policy using the (Pope III et al., 2020) estimates are still substantial, the distributional effects are also significant due to the increased pollution burden in Bawana.

6.2 Costs: Impact on Relocated Firms

In this section, we consider how the relocation policy affected firms that were relocated outside the city. First, for this back-of-the-envelope calculation, we use a counterfactual survival probability using distance that a firm was relocated. The probability firms survived in the industrial area in the long run is strongly decreasing in the distance they were moved, highlighting that location choice is a consequential endogenously chosen parameter by firms.

Only about 26% of firms are still present and functioning in the industrial area by 2018. To estimate a BOTE counterfactual death rate for these firms, we take two approaches - the first is to estimate how the survival rate varies as a function of the distance which the firm was moved. In Figure 7, we present evidence that the further a firm was moved, the less likely it is to survive, and Table 6 presents regression results for this relationship. Column 2 of Table 6 includes neighborhood fixed effects, comparing firms from nearby baseline locations. Each kilometer relocated lowers the probability of firm survival by between 0.4 and 1.4 percentage points. The latter estimate includes neighborhood fixed effects, while the former does not. This implies that for the average firm, which was relocated 20.37 kilometers, the reduction in the probability of firm survival is 28.42 percentage points higher than a firm that was not relocated at all (using the second estimate). This is a substantial reduction in firm survival, and similar in magnitude to the 26% survival rate. [12]

Next, we use the larger firm exit estimate of attributing 28.42 percentage points lower survival probability to the policy to estimate the costs on firms (to get a conservative benefit-cost ratio). This greater exit risk is over 12 years (the average firm takes possession of the plot in 2006, and the exit data was collected in 2018), yielding an annual probability of 2.4% lower risk of survival due to the policy. We use panel data from Center of Monitoring the Indian Economy (CMIE) (Center for Monitoring the Indian Economy, 2021) to impute firm sales and salaries (this is likely an overestimate, since CMIE is more likely to include larger firms). Restricting the sample to the state of Delhi and to the year 2006 when the average firm took possession of their industrial plot, the median firm has \$2.9 million in sales, and pays \$87,700 in wages annually. This implies that the policy caused \$1.46 billion in lost sales and \$44.2 million in lost wages, for a total damage estimate of \$1.51 billion in lost sales and wages per year (for the 21,000 relocated firms).

[12] Alternatively, we could use survival rates estimated for Indian firms more broadly - for instance, Sengupta and Singh (2019) find that the probability of firm survival in India over 20 years for registered firms is about 50%, about a reduction of 2.5% each year for survival. Therefore, seventeen years after the policy, about 42.5% of these firms should have survived. As we show below, our cost-benefit estimates do not change significantly if we use this higher exit rate.

6.3 Comparison of Benefits and Costs

Compared to the benefits of reduced mortality, using the first elasticity estimate, the benefits are 36.15 times higher than the costs to firms. Thus, while the relocation policy caused firms to exit at a faster rate and consequently led to lost sales and wages, the benefits of improved air quality were substantially higher.

How do these cost-benefit estimates compare with other pollution reduction possibilities in India? First, let us consider mitigating damages from coal-fired power plants, which claim 84,650 lives in India each year. Installing scrubbers avoids 72% of deaths in the first year (Cropper et al., 2019), saving 60,948 lives. India has 204 GW of coal capacity (Ministry of Power, 2022); therefore, on average, scrubbers in a 500 MW plant would save 150 lives per year, at a cost of $1.2 billion over 20 years, or 60 million per year. These estimates imply that the cost per life saved is around $400,000. Another pollution reduction possibility is crop residue burning in India, which causes 86,000 premature deaths in India annually (Lan et al., 2022). Jack et al. (2022) find that payment for ecosystem services to incentivize farmers to not burn residue is effective in reducing burning, and applying those treatment effects leads to a cost of life saved of $2,930. Industrial relocation is in the middle of these policies in terms of cost-effectiveness, saving 81,068.58 lives per year at a cost of 1.51 billion, which yields a cost of life saved estimate of $18,626.21. Taking firm spillovers into account with our exit-reducing assignment cuts the effect of firm relocation on exit roughly in half, bringing the cost per life saved down to $9683.16. Thus, leveraging firms' economic spillovers caused by input-output linkages can significantly reduce the costs of such policies on firms.[13]

7 Conclusion

Firm location decisions have important spillovers to the neighborhoods they locate in. These spillovers can be positive, generating employment and knowledge flows, or negative, such as increasing pollution exposure. We find that the presence of the polluting firms studied in this paper negatively impacts neighborhood-level ambient environmental quality in New Delhi. The removal, however, also impacts the relocated firms, substantially decreasing their survival probabilities. These survival probabilities could have been significantly increased by relocating firms taking into account firm spillovers in the

[13] If on the other hand, pollution reductions did not impact health meaningfully, while the program's benefit-cost estimate would be negative, the point about using spillovers to reduce costs on relocated firms still stands.

relocated industrial areas, indicating that such spillovers can be a powerful force to reduce costs on relocated firms.

Did this firm removal persist in the longer-term? In 2021-2022, we sent surveyors to the original addresses of relocated firms (details of this data collection are in Section 3.3.1), where they identified what was present at the original address. Results are summarized in Table 7, and show that *any* firm was present in less than 10% of the locations. We also asked the surveyors to collect information on whether the firm could possibly be the same as the relocated firm, using firm name and other characteristics to identify this. This happened for less than a third of all firms, or about 3% of all observations. The largest category of land use was residential buildings (in 45% of locations), followed by mixed use residential with retail shops and retail shops alone (21% of locations). The rest were split between construction or empty lots (2.7%), warehouses (4%), or were locked during multiple visits (7.7%).[14] Overall, firm removal seems permanent, leading the way for largely residential and retail uses to replace these firms.

Removal of firms may have important equity implications, by increasing commuting costs or moving costs for workers, as well as impacting the affordability of a neighborhood. Furthermore, other policy choices may affect the impacts on firms and environment-for instance, which types of industries should be relocated to provide the maximum environmental benefits and minimize costs to firms and workers, and whether lump sum transfers instead of allocating them land in a fixed place is better for firms' survival. These, and related questions, remain interesting questions for future work.

[14] Surveyors were unable to locate about 10% of locations.

References

Ahlfeldt, G. M., Sturm, D. M., Redding, S. J., and Wolf, N. (2015). The Economics of Density: Evidence from the Berlin Wall. *Econometrica*, 83(6):2127–2189.

Barrows, G., Garg, T., and Jha, A. (2019). The health costs of coal-fired power plants in india.

Baum-Snow, N., Pavan, R., and Gendron-Carrier, N. (2023). Local Productivity Spillovers.

Behrer, A. P., Choudhary, R., and Sharma, D. (2023). Air pollution reduces economic activity.

Belloni, A., Chernozhukov, V., and Hansen, C. (2013). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2):608–650.

Benkard, C., Bodoh-Creed, A., and Lazarev, J. (2020). Simulating the dynamic effects of horizontal mergers: Us airlines. *Working Paper*.

Berry, S. T. and Compiani, G. (2021). Empirical Models of Industry Dynamics with Endogenous Market Structure. *Annual Review of Economics*, 13:309–334.

Betcherman, G., Daysal, N. M., and Pagés, C. (2010). Do employment subsidies work? evidence from regionally targeted subsidies in turkey. *Labour Economics*, 17(4):710–722.

Blackman, A., Lahiri, B., Pizer, W., Planter, M. R., and Piña, C. M. (2010). Voluntary environmental regulation in developing countries: Mexico's clean industry program. *Journal of Environmental Economics and Management*, 60(3):182–192.

Brunnermeier, S. B. and Levinson, A. (2004). Examining the evidence on environmental regulations and industry location. *The Journal of Environment & Development*, 13(1):6–41.

Bryan, G., Frye, K., and Morten, M. (2024). Spatial economics for low-and middle-income countries.

Busso, M., Gregory, J., and Kline, P. (2013). Assessing the incidence and efficiency of a prominent place based policy. *American Economic Review*, 103(2):897–947.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1).

Center for Monitoring the Indian Economy (2021). Prowessdx database.

Chaurey, R., Abeberese, A. B., and Menon, R. (2023). Place-based policies and migration: Evidence from india. *Available at SSRN 4697765*.

Chen, Q., Chen, Z., Liu, Z., Serrato, J. C. S., and Xu, D. Y. (2021). Regulating conglomerates: Evidence from an energy conservation program in china.

Choi, J., Hyun, J., Kim, G., and Park, Z. (2023). Trade policy uncertainty, offshoring, and the environment: Evidence from us manufacturing establishments. *Offshoring, and the Environment: Evidence from US Manufacturing Establishments (February 1, 2023)*.

Ciccone, A. and Hall, R. E. (1996). Productivity and the Density of Economic Activity. *American Economic Review*, 86(1):54–70.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, 80(6):2543–2594.

Combes, P.-P. and Gobillon, L. (2015). The Empirics of Agglomeration Economies. In *Handbook of Regional and Urban Economics*, chapter 5, pages 247–348. Elsevier.

Criscuolo, C., Martin, R., Overman, H. G., and Van Reenen, J. (2019). Some causal effects of an industrial policy. *American Economic Review*, 109(1):48–85.

Cropper, M. and Park, Y. (2022). The health benefits of air pollution control in india. *Indian Economic Review*, 57(2):245–263.

Cropper, M. L., Guttikunda, S., Jawahar, P., Lazri, Z., Malik, K., Song, X.-P., and Yao, X. (2019). Applying benefit-cost analysis to air pollution control in the indian power sector. *Journal of Benefit-Cost Analysis*, 10(S1):185–205.

De Simone, R., Naaraayanan, S. L., and Sachdeva, K. (2024). Opening the brown box: Production responses to environmental regulation. *Available at SSRN 4707581*.

Do, Q.-T., Joshi, S., and Stolper, S. (2018). Can environmental policy reduce infant mortality? evidence from the ganga pollution cases. *Journal of Development Economics*, 133:306–325.

Duflo, E., Greenstone, M., Pande, R., and Ryan, N. (2018). The value of regulatory discretion: Estimates from environmental inspections in india. *Econometrica*, 86(6):2123–2160.

Duranton, G. and Puga, D. (2020). The Economics of Urban Density. *Journal of Economic Perspectives*, 34(3):3–26.

Ellison, G., Glaeser, E. L., and Kerr, W. R. (2010). What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review*, 100(June):1195–1213.

Fan, H., Zivin, J. S. G., Kou, Z., Liu, X., and Wang, H. (2019). Going green in china: Firms' responses to stricter environmental regulations. Technical report, National Bureau of Economic Research.

Fenske, J., Haseeb, M., and Kala, N. (2023). How rules and compliance impact organizational outcomes: Evidence from delegation in environmental regulation. Technical report, National Bureau of Economic Research.

Foster, A. and Kumar, N. (2011). Health effects of air quality regulations in Delhi, India. *Atmospheric Environment*, 45(9):1675–1683.

Gechter, M. and Tsivanidis, N. (2020). Spatial spillovers from urban renewal: Evidence from the mumbai mills redevelopment. Technical report, Working Paper.

Glaeser, E., Kallal, H., Scheinkman, J., and Shleifer, A. (1992). Growth in Cities. *Journal of Political Economy*, pages 1126–1152.

Glaeser, E. L. and Gottlieb, J. D. (2008). The economics of place-making policies. Technical report, National Bureau of Economic Research.

Greenbaum, R. T. and Engberg, J. B. (2004). The impact of state enterprise zones on urban manufacturing establishments. *Journal of Policy Analysis and Management*, 23(2):315–339.

Greenstone, M. and Fan, C. Q. (2018). Introducing the air quality life index. *AQLI Annual Report*.

Greenstone, M. and Hanna, R. (2014). Environmental regulations, air and water pollution, and infant mortality in india. *American Economic Review*, 104(10):3038–72.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, 118(3):536–598.

Greenstone, M., Pande, R., Sudarshan, A., and Ryan, N. (2022). The benefits and costs of emissions trading: Experimental evidence from a new market for industrial particulate emissions. *Working paper*.

Harrison, A., Hyman, B., Martin, L., and Nataraj, S. (2015). When do firms go green? comparing price incentives with command and control regulations in india.

Hasan, R., Jiang, Y., and Rafols, R. M. (2021). Place-based preferential tax policy and industrial development: Evidence from india's program on industrially backward districts. *Journal of Development Economics*, 150:102621.

He, G., Wang, S., and Zhang, B. (2020). Watering down environmental regulation in china. *The Quarterly Journal of Economics*, 135(4):2135–2185.

HEI (2021). State of global air 2024.

Henderson, J. V. (1995). Effects of air quality regulation.

Henderson, V., Kuncoro, A., and Turner, M. (1995). Industrial development in cities. *Journal of Political Economy*, 103(5):1067–1090.

Hyman, B. G., Freedman, M., Khanna, S., and Neumark, D. (2022). Firm responses to state hiring subsidies: Regression discontinuity evidence from a tax credit formula. Technical report, National Bureau of Economic Research.

Jack, B. K., Jayachandran, S., Kala, N., and Pande, R. (2022). Money (not) to burn: payments for ecosystem services to reduce crop residue burning. Technical report, National Bureau of Economic Research.

Jaffe, A. B., Peterson, S. R., Portney, P. R., and Stavins, R. N. (1995). Environmental regulation and the competitiveness of us manufacturing: what does the evidence tell us? *Journal of Economic literature*, 33(1):132–163.

Karplus, V. J., Zhang, S., and Almond, D. (2018). Quantifying coal power plant responses to tighter so2 emissions standards in china. *Proceedings of the National Academy of Sciences*, 115(27):7004–7009.

Kline, P. and Moretti, E. (2014). Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority. *Quarterly Journal of Economics*, 129(1):275–331.

Lan, R., Eastham, S. D., Liu, T., Norford, L. K., and Barrett, S. R. (2022). Air quality impacts of crop residue burning in india and mitigation alternatives. *Nature communications*, 13(1):6537.

Leonardi, M. and Moretti, E. (2023). The Agglomeration of Urban Amenities: Evidence from Milan Restaurants. *American Economic Review: Insights*, 5(2):141–157.

Levinson, A. (1996). Environmental regulations and manufacturers' location choices: Evidence from the census of manufactures. *Journal of public Economics*, 62(1-2):5–29.

List, J. A., Millimet, D. L., Fredriksson, P. G., and McHone, W. W. (2003). Effects of environmental regulations on manufacturing plant births: evidence from a propensity score matching estimator. *Review of Economics and Statistics*, 85(4):944–952.

Lu, Y., Wang, J., and Zhu, L. (2019). Place-based policies, creation, and agglomeration economies: Evidence from china's economic zone program. *American Economic Journal: Economic Policy*, 11(3):325–360.

Majumder, A., Madheswaran, S., et al. (2018). *Value of statistical life in India: A hedonic wage approach*. Institute for Social and Economic Change.

Marshall, A. (1920). *Principles of Economics: An Introductory Volume*. Macmillan and Co., London.

McDuffie, E., Martin, R., Yin, H., and Brauer, M. (2021). Global burden of disease from major air pollution sources (gbd maps): a global approach. *Research Reports: Health Effects Institute*, 2021.

Millimet, D. L. and Roy, J. (2016). Empirical tests of the pollution haven hypothesis when environmental regulation is endogenous. *Journal of Applied Econometrics*, 31(4):652–677.

Muller, N. Z. and Mendelsohn, R. (2009). Efficient pollution regulation: getting the prices right. *American Economic Review*, 99(5):1714–1739.

Murphy, K. M., Shleifer, A., and Vishny, R. W. (1989). Industrialization and the big push. *Journal of Political Economy*, 97(5):1003–26.

Nakajima, K. and Teshima, K. (2017). Identifying Neighborhood Effects among Firms: Evidence from Location Lotteries of the Tokyo Tsukiji Fish Market. Technical report, Hitotsubashi University Working Paper.

Neumark, D. and Kolko, J. (2010). Do enterprise zones create jobs? evidence from california's enterprise zone program. *Journal of Urban Economics*, 68(1):1–19.

Neumark, D. and Simpson, H. (2015). Place-based policies. In *Handbook of regional and urban economics*, volume 5, pages 1197–1287. Elsevier.

O'keefe, S. (2004). Job creation in california's enterprise zones: a comparison using a propensity score matching model. *Journal of Urban Economics*, 55(1):131–150.

Pandey, A., Brauer, M., Cropper, M. L., Balakrishnan, K., Mathur, P., Dey, S., Turkgulu, B., Kumar, G. A., Khare, M., Beig, G., et al. (2021). Health and economic impact of air pollution in the states of india: the global burden of disease study 2019. *The Lancet Planetary Health*, 5(1):e25–e38.

Pope III, C. A., Coleman, N., Pond, Z. A., and Burnett, R. T. (2020). Fine particulate air pollution and human mortality: 25+ years of cohort studies. *Environmental research*, 183:108924.

Rosenstein-Rodan, P. N. (1943). Problems of industrialisation of eastern and south-eastern europe. *The economic journal*, 53(210-211):202–211.

Rosenthal, S. S. and Strange, W. (2020). How Close Is Close? The Spatial Reach of Agglomeration Economies. *Journal of Economic Perspectives*, 34(3):27–49.

Ryan, S. P. (2012). The costs of environmental regulation in a concentrated industry. *Econometrica*, 80(3):1019–1061.

Sengupta, R. and Singh, M. (2019). Firm survival in India: Status of firms formed over the last 30 years. *Ideas for India*.

Shapiro, J. S. and Walker, R. (2020). Is air pollution regulation too lenient? evidence from us offset markets. Technical report, National Bureau of Economic Research.

Sharma, S., Bawase, M., Ghosh, P., Saraf, M., Goel, A., Suresh, R., Datta, A., Jhajhjra, A., Kundu, S., Sharma, V., et al. (2018). Source apportionment of pm2. 5 and pm10 of delhi ncr for identification of major sources. *The Energy Resources Institute, Delhi and Automative Research Association of India*.

Singh, V. (2007). *Delhi Master Plan 2021*. Rupa Publications.

Song, R., Xia, F., and Zhang, B. (2022). Pollution taxes as a second-best: Accounting for multidimensional firm heterogeneity in environmental regulations.

Tanaka, S., Teshima, K., and Verhoogen, E. (2022). North-south displacement effects of environmental regulation: The case of battery recycling. *American Economic Review: Insights*, 4(3):271–288.

Tsivanidis, N. (2023). Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio. *Working Paper*.

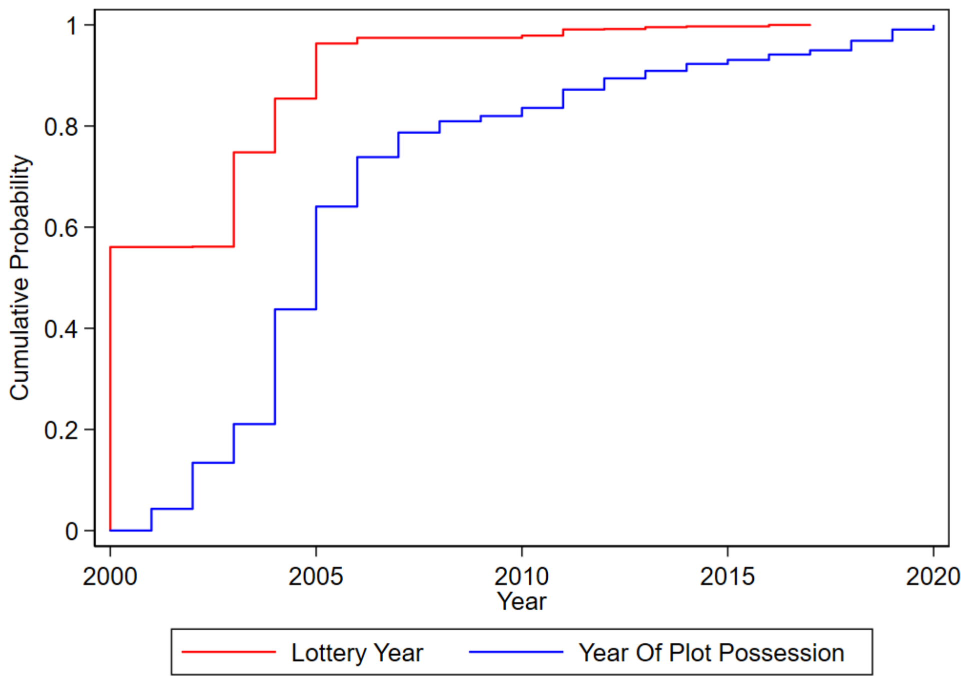
Van Donkelaar, A., Hammer, M. S., Bindle, L., Brauer, M., Brook, J. R., Garay, M. J., Hsu, N. C., Kalashnikova, O. V., Kahn, R. A., Lee, C., et al. (2021). Monthly global estimates of fine particulate matter and their uncertainty. *Environmental Science & Technology*, 55(22):15287–15300.

van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Levy, R. C., Lyapustin, A., Sayer, A. M., and Winker, D. M. (2016). Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors. *Environmental Science and Technology*, 50:3762–3772.

Vitali, A. (2022). Consumer search and firm location: Theory and evidence from the garment sector in uganda. *Work. Pap*.

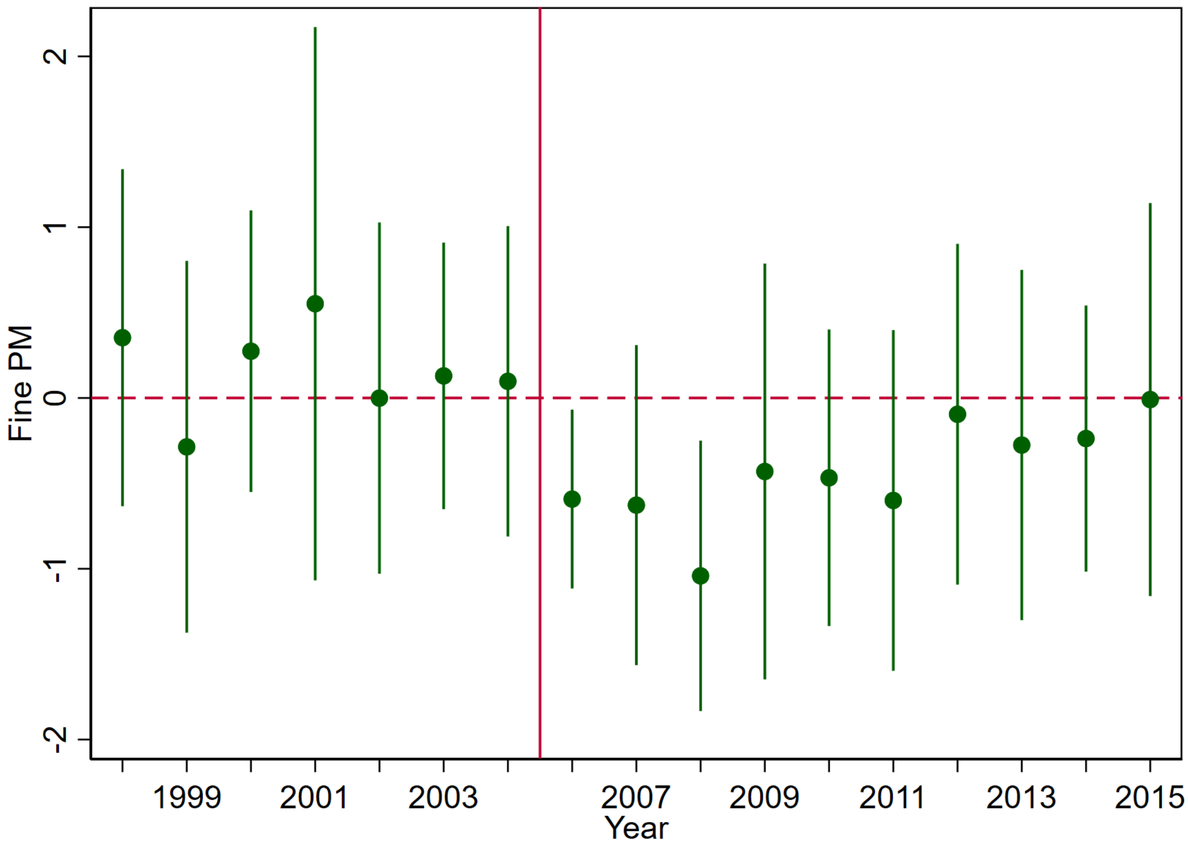
Wilson, S., Hutson, M., and Mujahid, M. (2008). How planning and zoning contribute to inequitable development, neighborhood health, and environmental injustice. *Environmental Justice*, 1(4):211–216.

Figure 1: Timing of Plot Lottery and Plot Possession in the Industrial Area



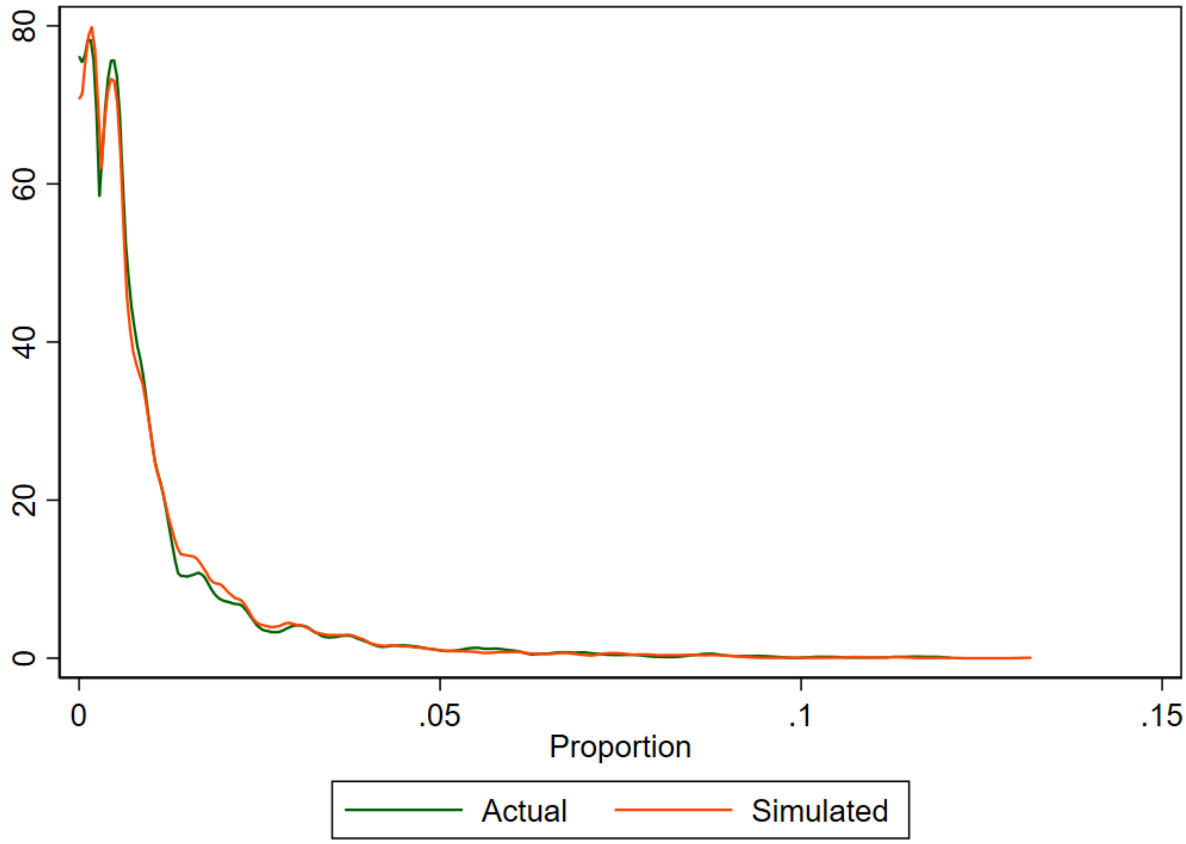
Notes: The blue line depicts the cumulative distribution function of the year in which the firm took possession of the plot in the industrial area. The red line depicts the cumulative distribution function of the year in which the firm was assigned a plot in the industrial area by lottery.

Figure 2: Effect of Firm Removal on Air Pollution (Fine PM, $\mu g/m^3$)



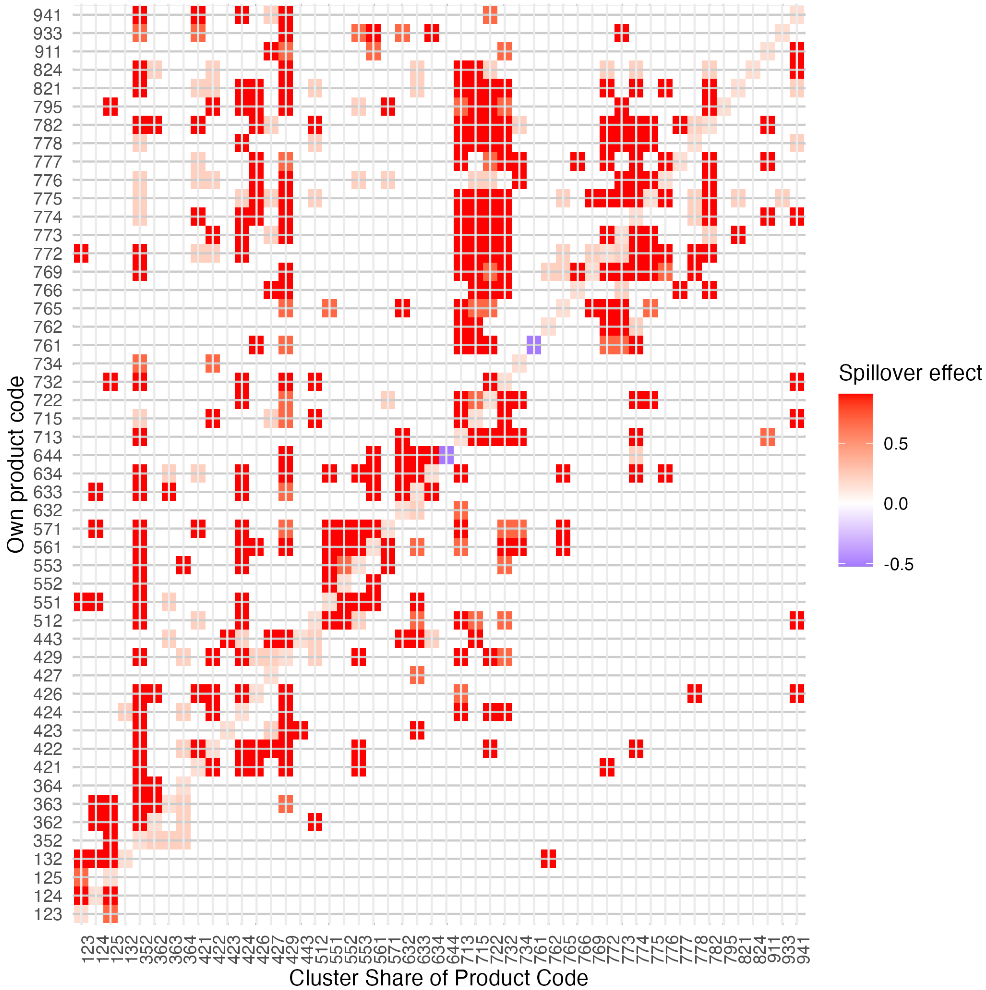
Notes: Interactions of year fixed effects with Log(firms relocated by 2004) shown in the figure. 2005 is Omitted. Grid ID and year fixed effects, as well as interactions of year fixed effects with the log of total firms relocated included. Bars represent 90% confidence intervals.

Figure 3: Actual vs. Simulated Plot Assignments in the Industrial Area



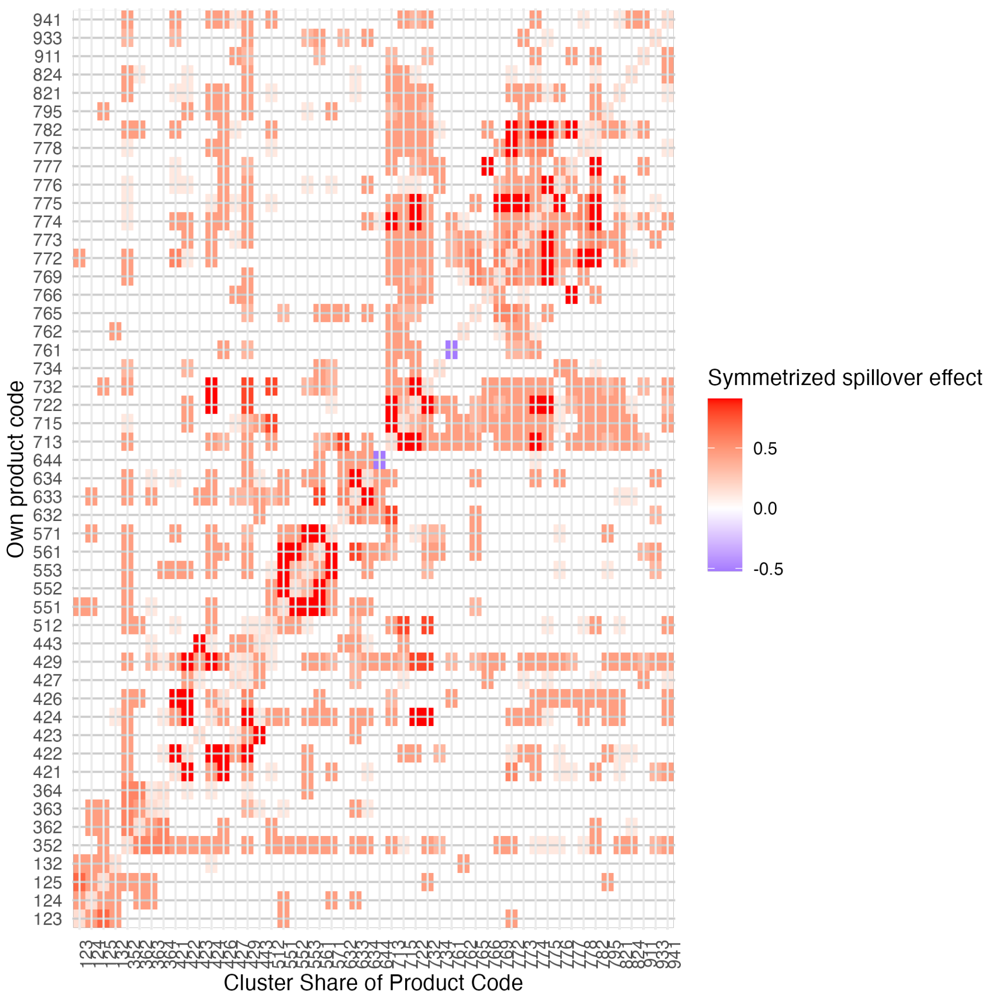
Notes: The orange line depicts the kernel density of the block-product distribution (proportion of each product in each block) generated by the simulated lotteries, while the green line depicts this distribution for actual plot assignments. To simulate the original lotteries, for each year of the lottery and plot size category, we randomly pick firms (each lottery randomly picks the number of firms by size category) and randomly assign them to available plots of that size category (for instance, if x firms were assigned 100 m^2 plots in 2000, we randomly pick x of all the 100 m^2 firms and randomly assign them to the 100 m^2 plots that were assigned in 2000).

Figure 4: Effect of the Industrial Cluster Share of Producers of Each Good on Firm Survival, by Own Good



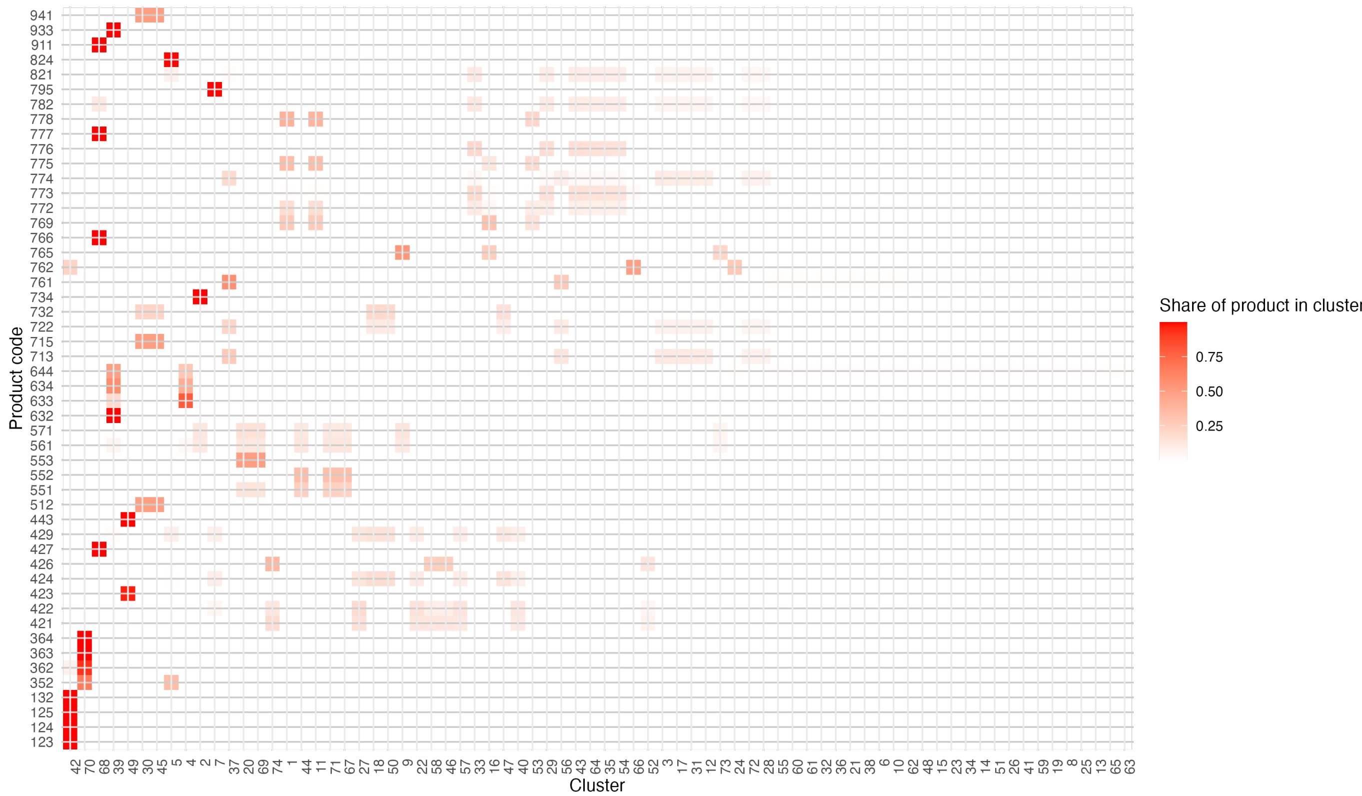
Notes: the color of the four squares around each intersection of gray lines represents the effect of the share in an industrial cluster of the product code corresponding to the vertical gray line on the survival of a firm producing the good corresponding to the horizontal line. Survival is a dummy variable that takes the value 1 if the firm was operating in the largest industrial area, Bawana, in 2018, and 0 otherwise. Effects combine estimates from Table 5 with input-output relationship between individual product codes described in the paper.

Figure 5: Effect of the Industrial Cluster Share of Producers of Each Good on Firm Survival, Symmetric



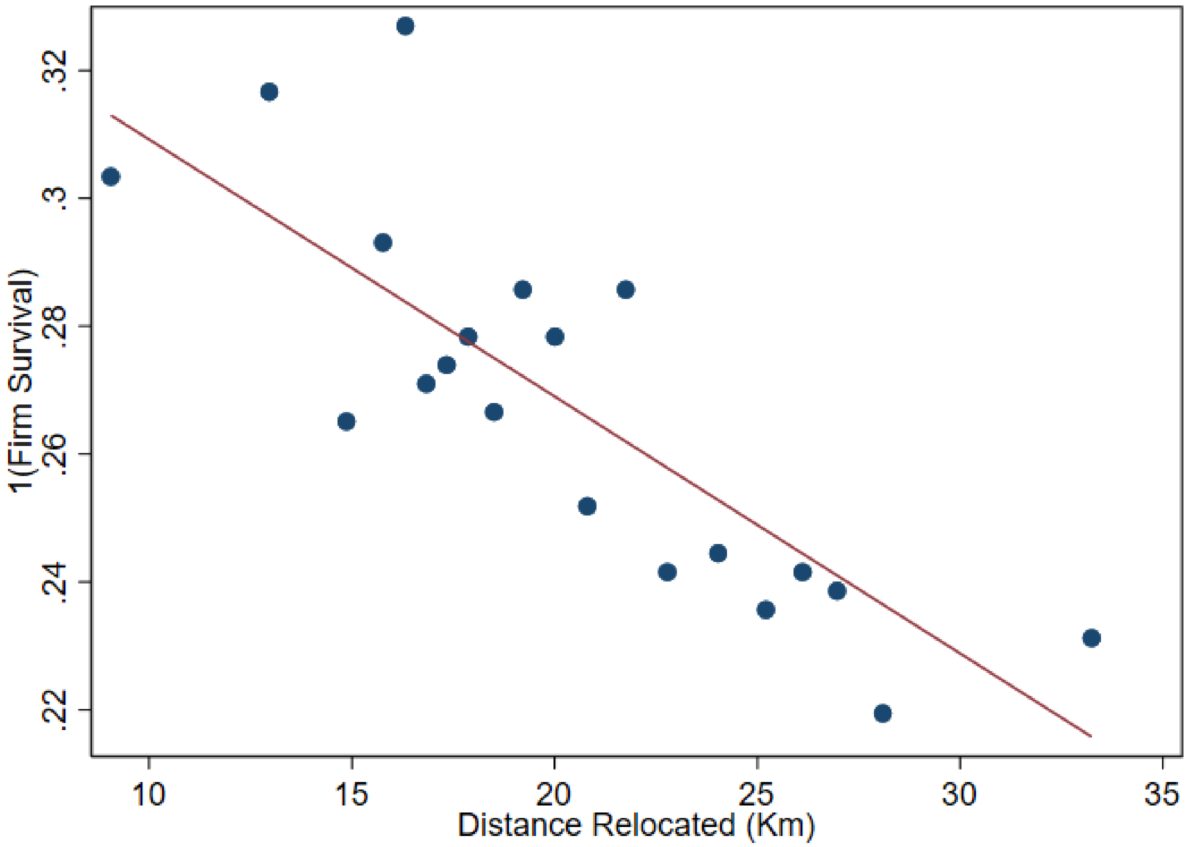
Notes: the color of the four squares around each intersection of gray lines represents the effect of the share in an industrial cluster of the product code corresponding to the vertical gray line on the survival of a firm producing the good corresponding to the horizontal line. Effects δ_{mn}, where m denotes the own product represented by the horizontal gray line and n the cluster share of the product represented by the vertical gray line from Figure 4 after making $\delta_{nm} = \delta_{mn} = 1/2(\delta_{mn} + \delta_{nm})$ so that the matrix is symmetric. Spillover effects are on survival, which is a dummy variable that takes the value 1 if the firm was operating in the largest industrial area, Bawana, in 2018, and 0 otherwise.

Figure 6: Optimal Product Composition of Clusters in the Industrial Area



Notes: the color of the four squares around each intersection of gray lines represent the share of the product represented by the horizontal gray line which is optimally assigned to the industrial cluster represented by the vertical gray line. The shares add up to 1 across rows. There are 74 industrial clusters in the largest industrial area.

Figure 7: Binscatter Plot: Distance Relocated and Firm Exit



Notes: Distance Relocated is the distance between the firm's original address geocoded using Google Maps and the assigned plot in the industrial area. 1(Firm Survival) is a dummy variable that takes the value 1 if the firm was operating in the industrial area in 2018, and 0 otherwise.

Table 1: Balance by Year of Lottery: Longitude and Latitude

	(1) Latitude	(2) Longitude
Lottery Year	0.0000932	-0.000906
	(0.000326)	(0.00148)
Constant	28.48****	79.00****
	(0.652)	(2.960)
Mean of Dependent Variable	28.66	77.19
R-Squared	0.0000	0.0000
N	19,508	19,508

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$. Latitude and longitude are the latitude and longitude of the firm's origin address as provided by Google Maps. Lottery Year is the year in which a firm received their plot assignment.

Table 2: First Stage of Lottery Timing Impacting Firm Movement to Industrial Area

	Year of Plot Posession	
	(1)	(2)
Lottery Year	1.150****	
	(0.0110)	
Lottery Year $\geq$ 2005		9.942****
		(0.0693)
Constant	-295.2****	2004.7****
	(22.02)	(0.0186)
Lottery Year	All	
Mean of Dependent Variable	2006.16	2006.16
R-Squared	0.4577	0.6443
N	19568	19568

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$. Year of Plot Possession is the year in which the firm took possession of the plot in the industrial area. Lottery Year is the year in which a firm received their plot assignment.

Table 3: Impact of Random Firm Removal on Air Pollution (Fine PM, μg/m^3)

	Fine PM (μg/m^3)					
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Firms Relocated by 2004) $\times$ $\mathbb{1}$(Year>2005)	-0.662**	-0.589*	-0.577*	-0.563*	-1.132***	-1.137***
	(0.306)	(0.314)	(0.321)	(0.325)	(0.381)	(0.408)
Log(Ineligible Firms) $\times$ $\mathbb{1}$(Year>2005)		0.157		0.0585		-0.00986
		(0.0999)		(0.115)		(0.164)
Constant	105.8****	105.7****	106.9****	106.9****	106.7****	106.7****
	(0.0152)	(0.0173)	(0.0494)	(0.0548)	(0.0692)	(0.0782)
Neighborhoods	All		With Relocated Firms		With Surveyor Identified Relocated Firms	
Randomization Inference P-Value	0.06	0.11	0.05	0.00	0.00	
Mean of Dependent Variable	106	106	107	107	107	107
R-Squared	0.9695	0.9695	0.9767	0.9767	0.9792	0.9792
N	24930	24930	10692	10692	7326	7326

Notes: Standard errors clustered at the neighborhood (1km by 1km grid cell) level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$. Fine PM is the annual concentration of fine particular matter in μg/m^3 at the neighborhood-year level. Firms Relocated by 2004 is the total number of firms in a neighborhood who have been assigned an industrial plot in a lottery by 2004. Ineligible Firms is the total number of firms in a neighborhood who applied for relocation, but were deemed to be ineligible for relocation by the government. Randomization Inference P-Value is the randomization inference p-value from randomly assigning Log (Firms Relocated by 2004) in 1,000 simulations.

Table 4: Most Common Products in the Largest Industrial Area

Product Name	Number of Firms
Motor vehicles (Passengers/goods transportation & special purpose vehicles)	1239
Printed books, newspaper, periodicals, notebooks, register etc. & other printed matters	770
Packing materials made of paper	614
Lamp, filament, electrodes, anodes/connectors, fittings & parts	487
Bags/boxes/panels/containers of plastic/pvc	401
tubes/pipes/basin & sanitary fittings of plastic/pvc	321
Film (non-sensitive/photographic)/foil/rolls/tape/rope of plastic/pvc & related materials	314
Articles, parts of plastic/pvc n.e.c	294
Copper and copper alloy, worked	266
Finished products of iron/steel	257
Wooden (incl plywood) furniture, boxes (incl packing box) and other wooden articles	248
Electrical motors, generators, transformer, power pack (this incl pump set fitted with electric motor)	226
Footwear plastic/pvc	210
Audio/video/sound apparatus & parts	182
Aluminium and aluminium alloys, worked	180

Notes: The product names are three digit ASICC codes (from 2010 product codes for the ASI). The second column shows the number of firms that produce products with the closest match with that ASICC code according to ChatGPT (please see Appendix C for details on the matching procedure).

Table 5: Reduced Form Impacts of Firm Spillovers

	Firm Survival			
	(1)	(2)	(3)	(4)
Proportion Upstream or Downstream (Std)	0.0444*			
	(0.0237)			
Proportion Own Type (Std)	-0.0229	-0.0219		
	(0.0140)	(0.0140)		
Proportion Upstream (Std)		0.0542**		
		(0.0212)		
Proportion Downstream (Std)		0.0252		
		(0.0257)		
Proportion Upstream or Downstream (Std)			0.0548**	
			(0.0222)	
Proportion Own Type (Std)			-0.0262	-0.0258
			(0.0168)	(0.0162)
Proportion Upstream (Std)				0.0455**
				(0.0184)
Proportion Downstream (Std)				0.00908
				(0.0203)
Threshold	Median		75th Percentile	
P-Val: Any Linkage=Own	0.052*		0.029**	
P-Val: Upstream=Own		0.014**		0.022**
P-Val: Downstream=Own		0.150		0.249
P-Val: Upstream=Downstream		0.431		0.226
Mean of Dependent Variable	.26	.26	.26	.26
N	9036	9011	9036	9011

Notes: Robust standard errors clustered at the block-level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$. Firm Survival is a dummy variable that takes the value 1 if the firm was operating in the industrial area in 2018, and 0 otherwise. All Proportion variables are standardized by dividing by their respective standard deviation. Post double LASSO (Belloni et al., 2013) estimates, selecting product fixed effects, a dummy variable for each lottery (lottery-year and plot-size combination), as well as proportion of plots with missing plot assignments in LASSO regressions with the dependent and Proportion variables as outcomes.

Table 6: Impacts of Distance Relocated on Firm Survival

	(1)	(2)
	$\mathbb{1}$(Firm Survival)	
Distance Relocated	-0.00402***	-0.0141***
	(0.000643)	(0.00479)
Constant	0.349***	0.554***
	(0.0137)	(0.0974)
Mean of the Dependent Variable	0.267	0.268
N	13580	13431
R2	0.00287	0.0365

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$. Firm survival is a dummy variable that takes the value 1 if the firm is operating in the industrial area (Bawana) in 2018, and 0 otherwise. Distance relocated is the distance in kilometers between the firm's original address and the plot it is assigned in the industrial area.

Table 7: Descriptive Evidence From Surveyor Visits in 2021/2022

What is Located at Original Address	Proportion
Residential	0.452
Commercial and Residential/Market	0.21
Firm	0.099
Locked building	0.077
Warehouse	0.04
Empty Plot/Under Construction	0.027

Notes: 15,756 addresses were visited in total by surveyors. Surveyors were asked to reach the origin address, and record what they found there e.g. a firm, residence etc. If there was a firm, they administered a short survey to elicit the firm's name, size, and products sold. 10% of addresses were not found by surveyors. See Section 3.3.1 for details of the data collection.

Appendices

A Appendix Figures and Tables

Figure A1: Bawana Industrial Area-2000

Figure A2: Bawana Industrial Area-2001

(a) LANDSAT

(b) Google

Figure A3: Bawana Industrial Area-2005

Figure A4: Bawana Industrial Area-2010

Figure A5: Map of an Industrial Area Cluster (Pocket)

Figure A6: Map of the Largest Industrial Area: Bawana

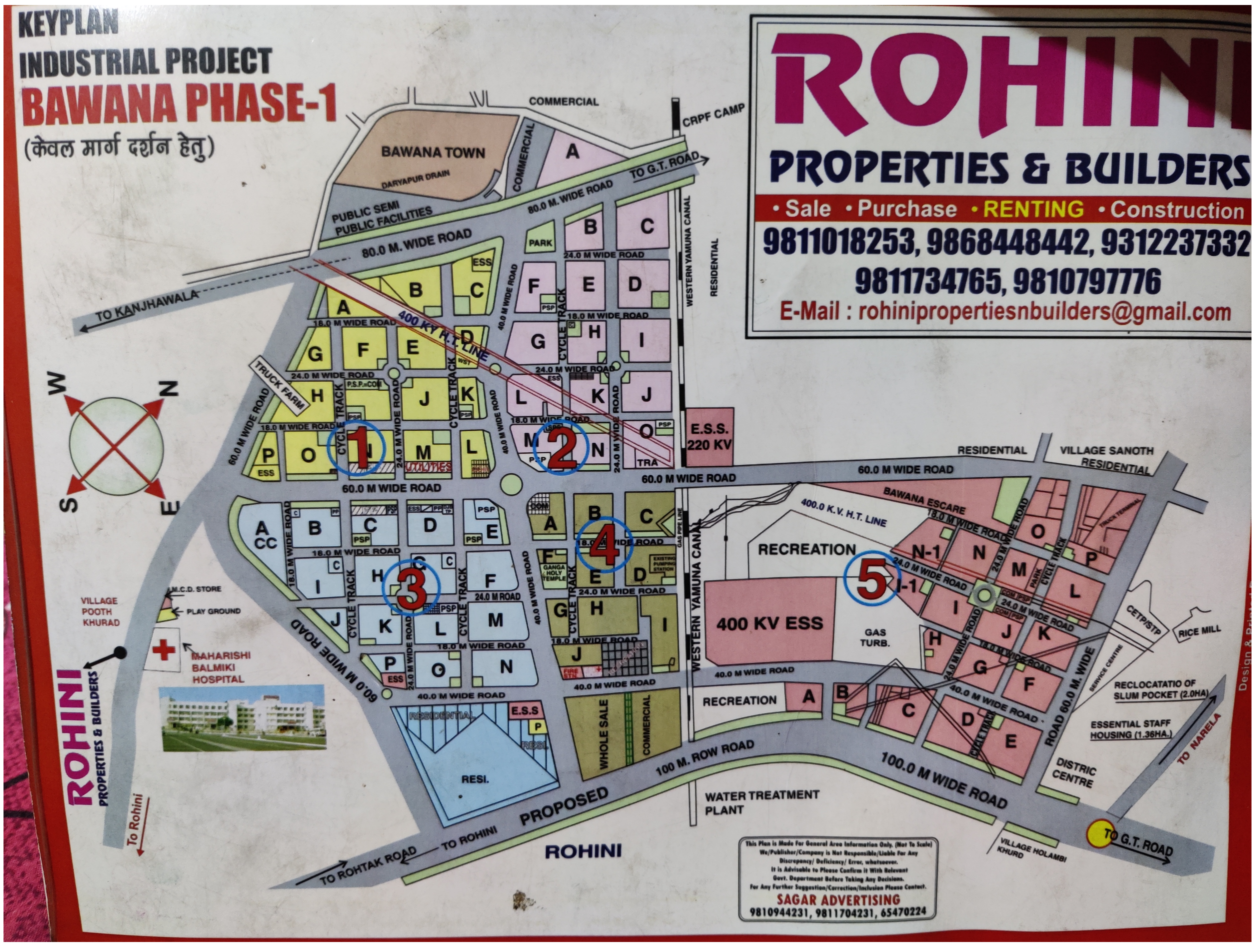
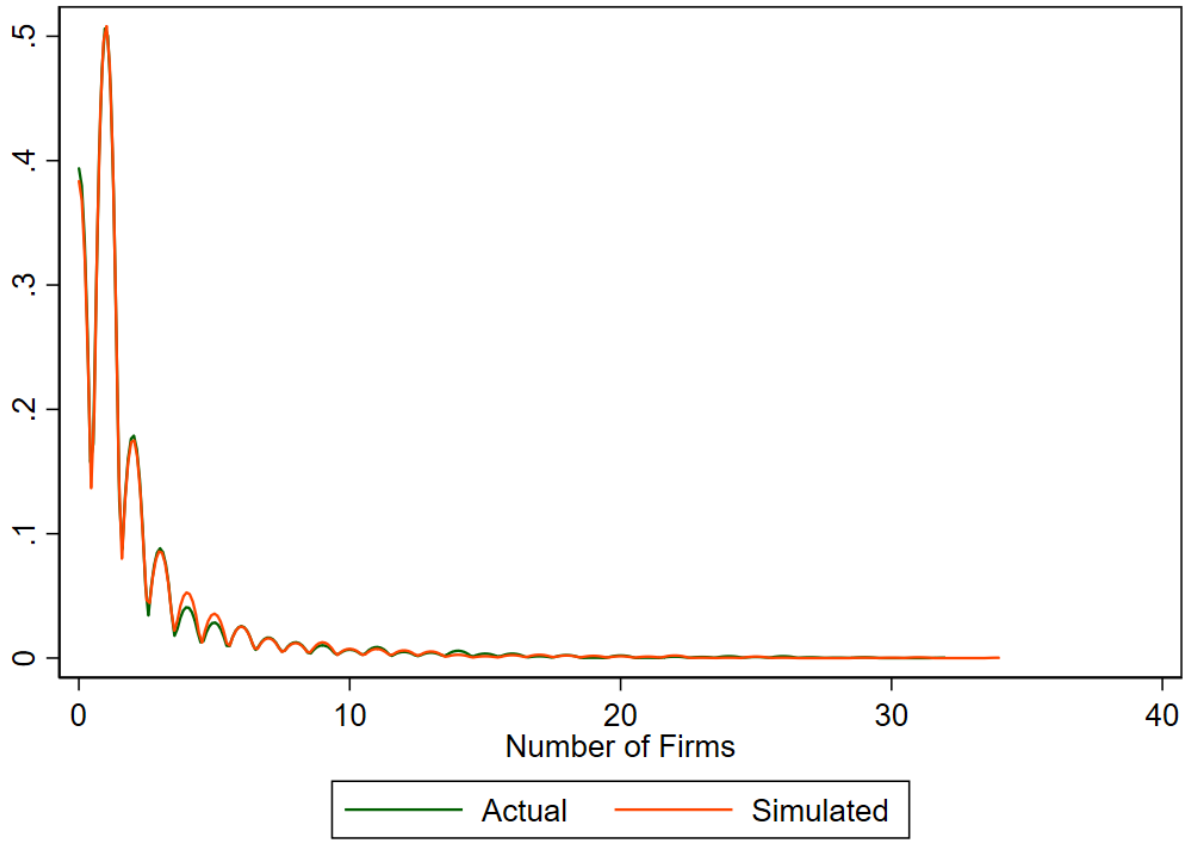
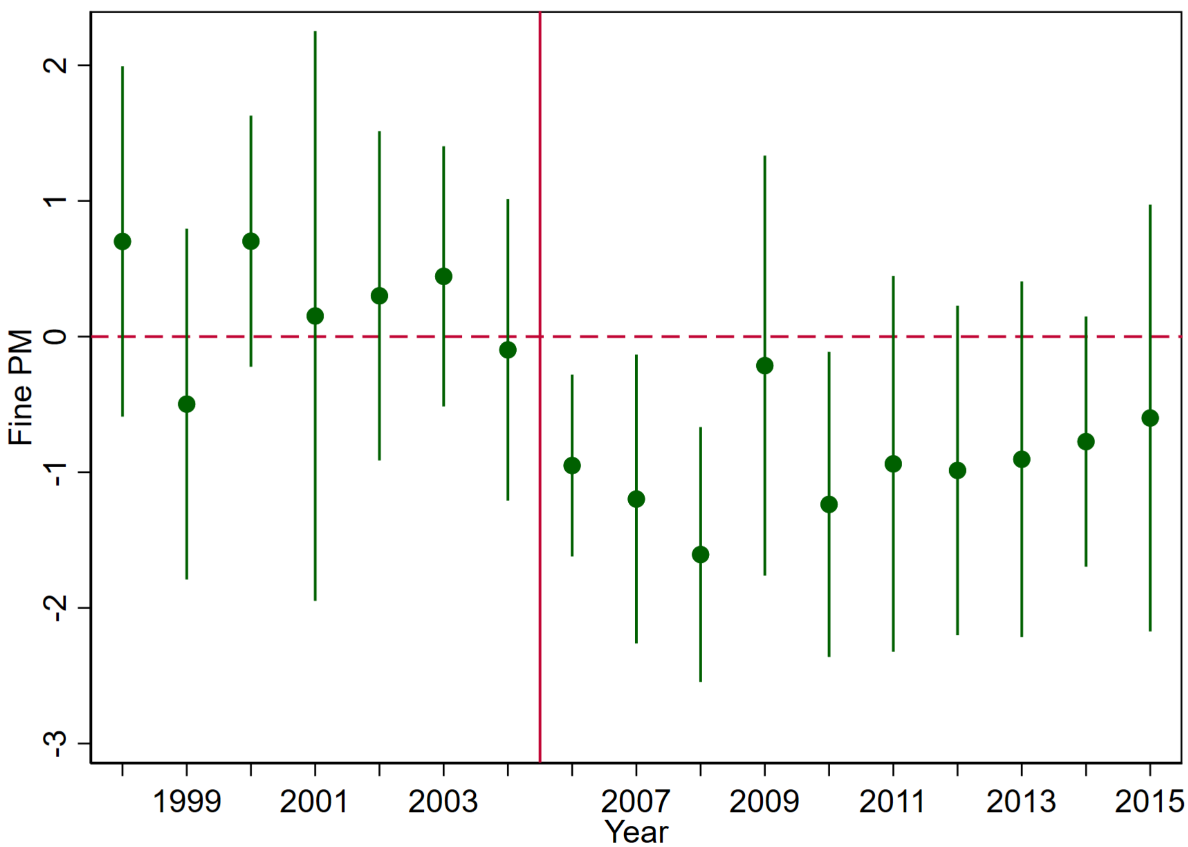


Figure A7: Actual vs. Simulated Plot Assignments in the Industrial Area: Number of Firms Producing Each Product in Each Block



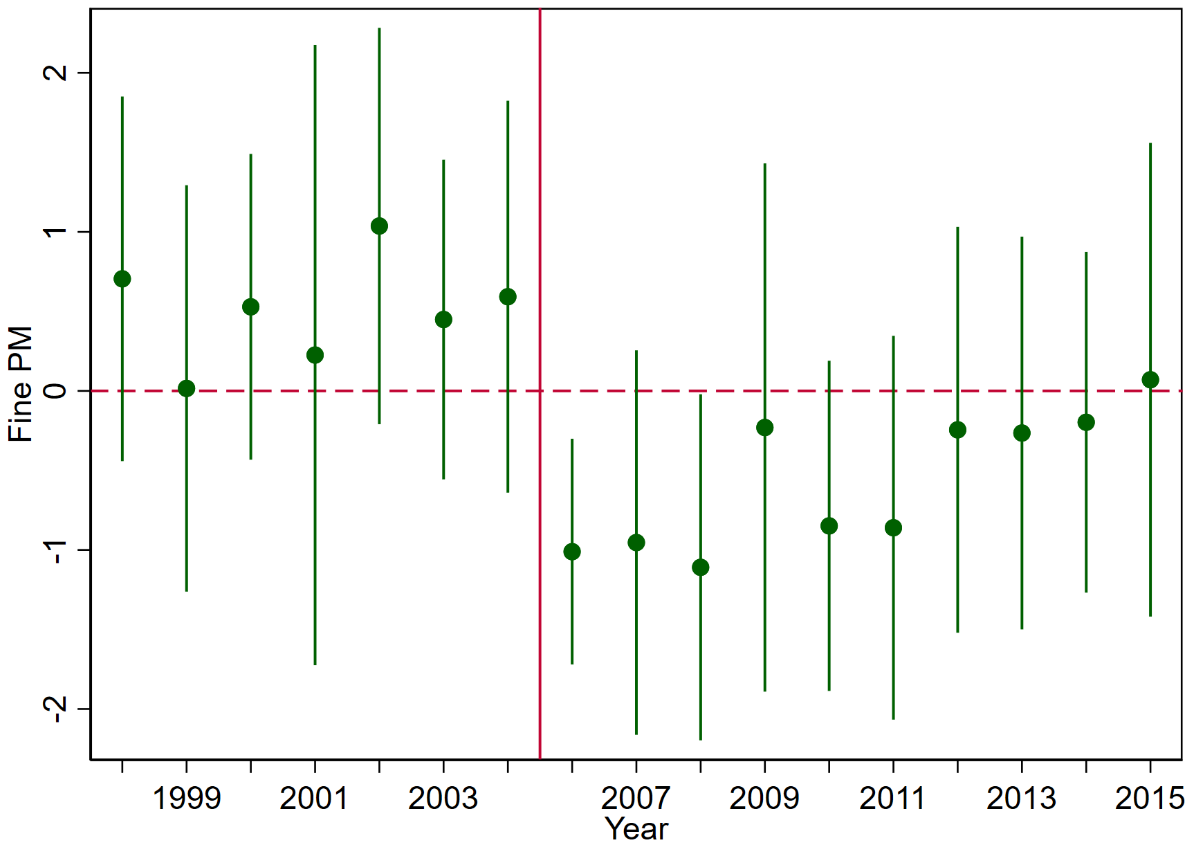
Notes: The orange line depicts the kernel density of the block-product distribution (number of each product in each block) generated by the simulated lotteries, while the green line depicts this distribution for actual plot assignments. To simulate the original lotteries, for each year of the lottery and plot size category, we randomly pick firms (each lottery randomly picks the number of firms by size category) and randomly assign them to available plots of that size category (for instance, if x firms were assigned 100 m^2 plots in 2000, we randomly pick x of all the 100 m^2 firms and randomly assign them to the 100 m^2 plots that were assigned in 2000).

Figure A8: Effect of Firms Leaving on Air Pollution (Fine PM, μg/m^3): Neighborhoods with Surveyor-Indentified Firms



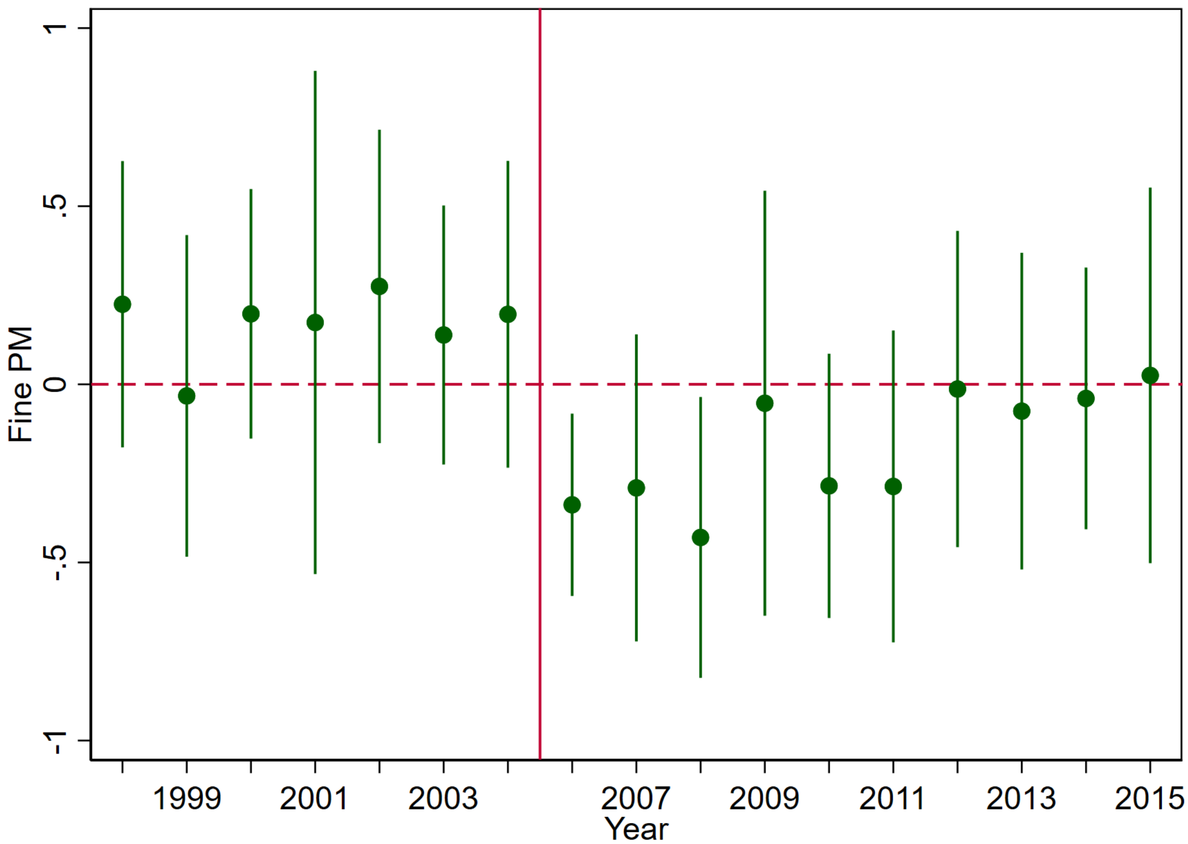
Notes: Interactions of year fixed effects with Log(firms relocated by 2004) shown in the figure. 2005 is Omitted. Grid ID and year fixed effects, as well as interactions of year fixed effects with the log of total firms relocated included. Bars represent 90% confidence intervals.

Figure A9: Impact of Random Firm Removal on Air Pollution (Fine PM, $\mu g/m^3$): Triple Interaction Between Baseline Number of Firms and Proportion Lotteried Early, 4 Bins



Notes: Triple interactions presented of three variables. The first is a variable that takes the value 1 if the baseline total number of firms relocated is in the first quartile, 2 if it is the second, and so on. The second is the proportion of firms in the neighborhood that had received an industrial plot by 2004. The third is a dummy variable that takes the value 1 if the year is 2006 or later. Grid ID and year fixed effects, as well as all double interactions included. Bars represent 90% confidence intervals.

Figure A10: Impact of Random Firm Removal on Air Pollution (Fine PM, $\mu g/m^3$): Triple Interaction Between Baseline Number of Firms and Proportion Lotteried Early, 10 Bins



Notes: Triple interactions presented of three variables. The first is a variable that takes the value 1 if the baseline total number of firms relocated is in the first decile, 2 if it is the second, and so on. The second is the proportion of firms in the neighborhood that had received an industrial plot by 2004. The third is a dummy variable that takes the value 1 if the year is 2006 or later. Grid ID and year fixed effects, as well as all double interactions included. Bars represent 90% confidence intervals.

Table A1: Balance by Whether Lottery Year is Missing: Longitude and Latitude

	(1) Latitude	(2) Longitude
1(Lottery Year Missing)	-0.000228	-0.00567
	(0.00546)	(0.0150)
Constant	28.66****	77.19****
	(0.000926)	(0.00405)
Mean of Dependent Variable	28.66	77.18
R-Squared	0.0000	0.0000
N	24373	24373

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$. Latitude and longitude are the latitude and longitude of the firm's origin address as provided by Google Maps. 1(Lottery Year Missing) is a binary variable that takes the value 1 if the year of the firm's assignment by lottery is missing, and 0 otherwise.

Table A2: Impact of Random Firm Removal on Air Pollution (Fine PM, μg/m^3): Triple Interaction Between Baseline Number of Firms and Proportion Lotteried Early

	Fine PM (μg/m^3)	
	(1)	(2)
Triple Interaction(Bins=4)	-1.009**	
	(0.396)	
4 Bins X Post	0.793**	
	(0.317)	
Proportion 2004 $\times$ Post	1.245*	0.588
	(0.640)	(0.453)
Triple Interaction(Bins=10)		-0.325**
		(0.143)
10 Bins X Post		0.255**
		(0.114)
Constant	106.4****	106.6****
	(0.289)	(0.206)
Mean of Dependent Variable	106.89	106.89
R-Squared	0.9767	0.9767
N	10,692	10,692

Notes: Standard errors clustered at the neighborhood (1km by 1km) grid cell level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$. Fine PM is the annual concentration of fine particular matter in μg/m^3 at the neighborhood-year level. Triple Interaction (Bins=4) is the triple interaction between three variables. The first is a variable (labeled "4 Bins" in the table) that takes the value 1 if the baseline total number of firms relocated is in the first quartile, 2 if it is the second, and so on. The second is the proportion of firms in the neighborhood that had received an industrial plot by 2004 (labeled "Proportion 2004" in the table). The third is a dummy variable that takes the value 1 if the year is 2006 or later (labeled "Post" in the table). Triple Interaction (Bins=10) is the triple interaction between three variables. The first is a variable (labeled "10 Bins" in the table) that takes the value 1 if the baseline total number of firms relocated is in the first decile, 2 if it is the second, and so on. The second is the proportion of firms in the neighborhood that had received an industrial plot by 2004 (labeled "Proportion 2004" in the table). The third is a dummy variable that takes the value 1 if the year is 2006 or later (labeled "Post" in the table).

Table A3: Impact of Random Firm Removal on Air Pollution (Fine PM, μg/m^3): Cluster at 2km X 2km Level

	Fine PM (μg/m^3)					
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Firms Relocated Per by 2004) $\times$ $\mathbb{1}$(Year>2005)	-0.465	-0.447	-0.577*	-0.563*	-1.145***	-1.148***
	(0.330)	(0.331)	(0.338)	(0.340)	(0.388)	(0.404)
Log(Ineligible Firms) $\times$ $\mathbb{1}$(Year>2005)		0.0817		0.0585		-0.00576
		(0.112)		(0.118)		(0.163)
Constant	106.8****	106.8****	106.9****	106.9****	106.8****	106.8****
	(0.0439)	(0.0545)	(0.0615)	(0.0708)	(0.0877)	(0.102)
Neighborhoods	All		With Relocated Firms		With Surveyor Identified Relocated Firms	
Mean of Dependent Variable	107	107	107	107	107	107
R-Squared	0.9755	0.9755	0.9767	0.9767	0.9794	0.9794
N	13032	13032	10692	10692	7146	7146

Notes: Standard errors clustered at the 2km by 2km grid cell level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$. Fine PM is the annual concentration of fine particular matter in μg/m^3 at the neighborhood -year level. Firms Relocated by 2004 is the total number of firms in a neighborhood who have been assigned an industrial plot in a lottery by 2004. Ineligible Firms is the total number of firms in a neighborhood who applied for relocation, but were deemed to be ineligible for relocation by the government. Randomization Inference P-Value is the randomization inference p-value from randomly assigning Log (Firms Relocated by 2004) in 1,000 simulations.

Table A4: Impact of Random Firm Removal on Air Pollution (Fine PM, μg/m^3): Without Neighborhood and Year Fixed Effects

	Fine PM (μg/m^3)					
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Firms Relocated Per by 2004) $\times$ $\mathbb{1}$(Year>2005)	-0.662**	-0.589*	-0.577*	-0.563*	-1.132***	-1.137***
	(0.306)	(0.314)	(0.321)	(0.325)	(0.381)	(0.407)
Log(Ineligible Firms) $\times$ $\mathbb{1}$(Year>2005)		0.157		0.0585		-0.00986
		(0.0999)		(0.115)		(0.164)
Constant	95.44****	95.34****	96.79****	96.68****	95.96****	95.68****
	(0.151)	(0.154)	(0.330)	(0.350)	(0.439)	(0.445)
Neighborhoods	All		With Relocated Firms		With Surveyor Identified Relocated Firms	
Mean of Dependent Variable	107	107	107	107	107	107
R-Squared	0.3924	0.3933	0.3860	0.3862	0.3884	0.3892
N	24930	24930	10692	10692	7326	7326

Notes: Standard errors clustered at the neighborhood (1km by 1km grid cell) level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$. Fine PM is the annual concentration of fine particular matter in μg/m^3 at the neighborhood -year level. Firms Relocated by 2004 is the total number of firms in a neighborhood who have been assigned an industrial plot in a lottery by 2004. Ineligible Firms is the total number of firms in a neighborhood who applied for relocation, but were deemed to be ineligible for relocation by the government.

Table A5: First Stage of Lottery Timing Impacting Firm Movement to Industrial Area: Neighborhood-Level

	(1) Log(Firms Given Plot Possession By 2006)	(2) Log(Firms Given Plot Possession) By 2015)
Log(Firms Assigned Plot in Lottery by 2004)	0.997****	0.502****
	(0.0437)	(0.0467)
Constant	-0.112****	-0.00445
	(0.0169)	(0.00409)
Year	2006	2015
Mean of Dependent Variable	23.88	28.93
R-Squared	0.9818	0.9955
N	555	579

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$. Firms Given Plot Possession by 2006 is the total number of firms in a neighborhood who have taken possession of their industrial plot by 2006. Firms Given Plot Possession by 2015 is the total number of firms in a neighborhood who have taken possession of their industrial plot by 2015. Firms Relocated by 2004 is the total number of firms in a neighborhood who have been assigned an industrial plot in a lottery by 2004. All regressions include controls for the log of total number of firms relocated in a neighborhood.

Table A6: Correlation Between Missing Geocoding with Timing of Firm Lottery

	(1) Surveyor Geocode Missing	(2) Google Geocode Missing
Lottery Year	-0.00125	-0.000170
	(0.00115)	(0.000930)
Constant	2.601	0.421
	(2.303)	(1.862)
Mean of the Dependent Variable	0.0980	0.0800
N	15010	20186
R2	0.0000749	0.00000167

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$. Restricted to lottery years 2005 and earlier, since over 90% of firms received a plot by 2005. Surveyor Geocode Missing is a dummy variable that takes the value 1 if the surveyor was unable to find the address, and 0 otherwise. Google Geocode missing is a dummy variable that takes the value 1 if Google Maps was unable to find a geocode for the firm within the greater Delhi area, and 0 otherwise.

B Fully Heterogeneous Spillover Specification

In this section, we develop an asymmetric, fully heterogeneous counterpart of Equation (4). Our aim here is to investigate the extent to which spillovers as whole can be explained by the Marshallian heterogeneity we allow for in Equation (4). We find that the large majority of heterogeneity in spillovers is indeed driven by Marshallian relationships between the goods firms produce.

Our fully-heterogeneous model estimates the following equation.

$$Active_{ik} = 1\left\{\sum_{m=1}^{M}\kappa_m 1\{product_i = m\} + \sum_{m,n\neq(M,M)}\delta_{mn}1\{product_i = m\}\overline{1\{product_j = n\}}_k + \varepsilon_{ik} \geq 0\right\} \tag{10}$$

where ε_{ik} follows a logistic distribution. The spillovers here are asymmetric because the effect of the share of n-producers in industrial cluster k depends on what i produces.

As in Section 5.3, product indices $\{m : m < M\}$ correspond to 3-digit product codes with 28 or more firms assigned, and M to all other products. We could estimate Equation (10), and then run a meta-regression of the estimated $\hat{\delta}_{mn}$ parameters on indicators for input-output relationships between each mn pair. However, since the dimension of the δ vector is the square of the number of products we consider (180 distinct products) minus one, the individual parameters would be very imprecisely-estimated.

We instead take a Bayesian approach to combine the two steps and regularize our estimates. We specify

$$\delta_{mn} = \alpha^{mn}_{upstream}1\{n \to m\} + \alpha^{mn}_{downstream}1\{n \leftarrow m\} + \alpha^{m}_{own}1\{n = m\} + \xi_{nm}$$

where $1\{n \to m\}$ indicates that n is upstream of m and $1\{n \leftarrow m\}$ that n is downstream of m. We use the definitions of upstream and downstream described in Section 3.2.2. We assume κ_m, $\alpha^{mn}_{upstream}$, $\alpha^{mn}_{downstream}$, α^{m}_{own}, ξ_{nm} are normally distributed with means μ_κ, $\mu_{upstream}$, $\mu_{downstream}$, μ_{own},and standard deviations σ_κ, $\sigma_{upstream}$, $\sigma_{downstream}$, σ_{own}. ξ_{nm} is normally distributed with mean 0 and standard deviation σ_ξ. We give the mean parameters independent Normal(0, 10) priors and the standard deviation parameters independent LogNormal(0, 0.75) priors.

We estimate the model using Hamiltonian Monte Carlo in Stan (Carpenter et al., 2017) and calculate the posterior share of the variance in heterogeneous spillover effects δ_{nm} attributable to features other than input-output linkages and net competitive effects between n and m. Table A7 shows the results, comparing the posterior distribution of the

Table A7: Share of Variation in Firm Spillovers Due to Non-Input-Output Effects

	Mean	Quantiles 5%	25%	75%	95%
Posterior	0.125	0.020	0.057	0.170	0.302
Prior	0.583	0.073	0.328	0.856	0.972

Notes: Quantiles of the share of variation in the effect of the share of producers of product n in a product-m-producer's block in the industrial area which is unexplained by the input-output relationship between n and m. 3000 burn-in Hamiltonian Monte Carlo draws and 2000 samples from the posterior per chain. 4 chains. R-hat criterion indicating good chain mixing lies below 1.005 for all parameters.

share to its prior. We find that Marshallian forces explain the large majority of variation in the δ_{mn} parameters. The posterior mean for the share of the variance in δ_{mn} explained by factors other than input-output relationships is 0.125. The 90% credible interval for the share of variation in spillovers due to non-Marshallian factors ranges only from 0.025 to 0.302. This is as compared with the 0.583 mean share due to non-Marshallian features over 10,000 draws from the prior , with a 5% to 95% interquantile range of 0.073 to 0.972. Therefore, a large fraction of spillovers are indeed mediated through Marshallian forces, which motivates our main spillover specifications.

The results in Table A7 depend on how frequently we classify product pairs as upstream and downstream from one another. An alternative way of evaluating the importance of non-input-output-based spillovers is to compute the posterior distribution of $\frac{\sigma_\xi}{\sigma_{upstream}+\sigma_{downstream}+\sigma_{own}+\sigma_\xi}$. We do this in Table A8. The 90% credible interval for this quantity ranges only from 0.003 to 0.129. The posterior mean of the share is 0.04. This is as compared to the prior mean of 0.25, with the 5% to 95% interquantile range of the prior distribution ranging from 0.009 to 0.777.

Table A8: Relative Magnitude of Variation in Firm Spillovers Due to Non-Input-Output Effects

	Mean	Quantiles 5%	25%	75%	95%
Posterior	0.043	0.004	0.013	0.053	0.138
Prior	0.251	0.009	0.053	0.389	0.780

Notes: Quantiles of the share of variation in the effect of the share of producers of product n in a product-m-producer's block in the industrial area which is unexplained by the input-output relationship between n and m. 3000 burn-in Hamiltonian Monte Carlo draws and 2000 samples from the posterior per chain. 4 chains. R-hat criterion indicating good chain mixing lies below 1.005 for all parameters.

C Matching Information on Firm Products to Annual Survey of Industries Commodity Classification (ASICC) Codes

C.1 Procedure

Before any analysis we created a custom pipeline using the GPT-4-turbo model to assign Annual Survey of Industries Commodity Classification (ASSIC) codes to firms based on the goods they report producing. We performed prompt engineering to guide the model using a set of detailed instructions that included accounting for potential typos and nuanced elements like distinguishing between manufacturing a particular product vs. manufacturing the machinery that produces that product. The output consists of a single product code assignment or a "missing" label for cases where GPT could not find a good match. The details of the prompts used are included in the next section. These requests are submitted to the GPT-4-turbo model using OpenAI's API, and the model's responses were parsed and integrated back into the output data. GPT models are inherently non-deterministic. To minimize randomness and enhance reproducibility, we set the temperature parameter, which measures response entropy, to 0.

C.2 Human Evaluation

To evaluate the results, we randomly selected samples of unique product descriptions and compared them to human assignments to ASICC codes done by research assistants. We compared the human assignments with the results of different versions of the GPT model concluding that GPT model assignment produced a reasonable and efficient fit for nearly all cases and was more consistent than a team of research assistants. We also used the human evaluation to extensively test different input prompts in search of best matching. This resulted in an extensive initial prompt that covers the edges cases in which GPT-4 performed poorly.

C.3 Output Processing

We process the assignment requests using the Batch method of OpenAI's API. We download the resulting output as JSONL files and extract the assignment answer for each firm. GPT provides up to 5 product codes in order from most to least preferred. We pick the most preferred code.

C.4 Scripts for GPT-4 API prompting

The following is the full initial prompt provided to GPT-4:

```
You are a helpful assistant who helps match product descriptions to ASICC product codes.
We wish to match each product description to one or multiple ASICC product codes I will provide,
and you should select the best match if there is one.
Please note that we distinguish products and machines to produce the products.
We consider that there might be typos in the product descriptions,
and we do the matching based on the corrected product descriptions.
Please pay attention to the noun part of the ASICC product code description,
and ensure the selected ASICC product codes match the product description well.
Please pay attention to the material of the product, when it is available.
If the product description contain multiple products,
you can choose a best one that includes all of them if possible,
or you can choose different codes for them separately.
If you are NOT SURE, you should PRINT None.

Here are some examples for guidance:

### EXAMPLES ###

If the item is general, and different versions of it will lead to codes with different first
    digits, then we should NOT assign a match.
E.g. "Steel Furniture" should be assigned to "715 manufacture of iron/alloy steel",
"Plastic Furniture" should be assigned to "421 bags/boxes/panels/containers of plastic/pvc",
while "FURNITURE" or "Furniture Chairs" SHOULD NOT HAVE AN ASSIGNMENT as the material is not
    clear.

If the item is general, but we can find a best higher level code, we should assign it to the
    highest level code.
E.g. "Garments" or "Readymade garments" should be assigned to "6 textile and textile articles",
"Hardware Goods" should be matched to "75 non-electrical machine tools & general purpose
    machineries and components and parts thereof [no sub-group formation]",
"Steel Fabrication" should be matched to "71 iron & steel (incl stainless steel) & articles
    thereof",
"Paper products" should be matched to "55 paper and paper board".
```

GENERAL TERMS SUCH AS "WIRE and CABLES" or "Wire Drawing" or "MACHINERY" SHOULD BE MATCHED to "7 base metals, products thereof & machinery equipment and parts thereof, excluding transport equipment" as their material is not specified,
"Steel Wires and Cables" should be matched to "71 iron & steel (incl stainless steel) & articles thereof",
"Copper Wire" should be matched to "72 copper, nickel & zinc articles thereof",
"Electrical Wires" should be matched to "774 lamp, filament, electrodes/anodes/connectors, fittings & parts",
"PVC Wires" should be matched to "424 film (non-sensitive/photographic)/foil/rolls/tape/rope of plastic/pvc & related materials".

Auto is mostly like to be related to motor vehicles, so "Auto Parts" should be matched to "821 motor vehicles - passengers/goods transportation and special purpose vehicles",
"Motor Parts" should be matched to "821 motor vehicles - passengers/goods transportation and special purpose vehicles",
"Rear Window Mirror" should be matched to "941 glass and glassware";

A vague action such as "Stamping" should not have a match because it is not clear about the products,
an action that is more specifc about a product can be matched based on the product it produced, for example:
"Electroplating" or "Electroplating job to order" should be matched to "71 iron & steel (incl stainless steel) & articles thereof",
"PRINTING" SHOULD BE MATCHED to "561 printed books, newspaper, periodicals, note books, register etc and other printed matters",
"Dyeing of Cotton Cloth" should be matched to "63 cotton, cotton yarn and fabrics" because the product is cotton cloth, not dyeing material.

We also make the distinction to the following to general groups:
"Electrical Goods" or "Electric Components" etc. should be matched to "77 electrical & electronic machinery & equipment incl parts (excl medical & non-conventional energy equipment)",
while "Electronic Goods" should be matched to "78 electronics equipment & parts excl bio-medical equipment".
"Kitchen Wares" should be matched to "74 misc. manufacture of base metals, n.e.c [no sub-group formation]".
"Corrugated Boxes" or "Card Board Boxes" should be matched to "571 packing materials made of paper [paper used for packing is shown under respective heading under division 55]",
while "Corrugated Paper Rolls" should be matched to "552 paper (uncoated) used for newsprint and for other special purpose excl writing/printing".
"pvc Compound" should be matched to "42 plastic, pvc articles incl packaging products and footwear plastic or pvc [note: toys will go sub-division 933]".
"DIE", as a machine tool to cut, SHOULD be matched to "74 misc. manufacture of base metals, n.e.c [no sub-group formation]".
"Engineering Goods" or "Engineering Work" should be matched to "75 non-electrical machine tools & general purpose machineries and components and parts thereof [no sub-group formation]".

END OF THE EXAMPLES

Here is the list of codes to select from:

###BEGINNING of industry codes, separated by new lines###

One digit codes

1 Animal, Vegetable, Horticulture, Forestry Products, Beverages, Tobacco And Pan Masala And Non-Edible Water/Spirit & Alcohol Chiefly Used In Industry
2 Ores, Minerals, Mineral Fuels, Lubricants, Gas & Electricity
3 Chemical And Allied Products
4 Rubber, Plastic, Leather & Products Thereof
5 Wood, Cork, Thermocol & Paper And Articles Thereof
6 Textile And Textile Articles
7 Base Metals, Products Thereof & Machinery Equipment And Parts Thereof, Excluding Transport Equipment
8 Railways, Airways, Ships & Road Surface Transport And Related Equipment & Parts
9 Other Manufactured Articles And Services N.e.c
End of one digit codes

Two digits codes
11 Animal & Food Products Of Animal Origin
12 Fruits, Vegetables, Cereals & Pulses And Other Vegetable Produces Like Lac, Gum Etc. And Preparation Thereof
13 Other Food And Edible Products & Preparations
14 Products Of Horticulture Incl Tissue Culture; Hydroponics And Forestry
15 Beverages, Tobacco And Pan Masala
16 Miscellaneous Non-Edible Substances For Industrial Use
21 Salts, Sulpher, Plastering Materials, Lime & Cement
22 Ores (Base Metal), Slag, Ash
23 Mineral Fuels: Oils, Products & By-Products
24 Gas (Fuel) Natural And Manufactured
25 Electrical Energy
29 Non-Metallic Mineral, Mineral Products Refractories
31 Inorganic Chemical, Compound Of Precious Materials Etc
32 Organic Chemicals
33 Pharmaceutical And Medical Products Excl Dental Materials
34 Fertilizer/Pesticides/Plant Protection Materials
35 Dyeing, Tanning, Colouring, Ink & Paint Etc
36 Essential Oil, Cosmetics & Perfumes, Dental Materials, Wax Polishing/Cleaning Materials
37 Explosives, Pyrotechnic Products, Matches And Certain Combustible Products, Pyrophoric Alloys
38 Photographic, Cinematographic Goods & Other Photo-Sensitive Materials [This Excl Apparatus Equipment And Parts Thereof]
39 Misc Chemical Goods Incl Albuminoidal Substances Modified Starches (Excl Sub-Division 124) And Glues, Enzymes
41 Rubber And Manufacture Of Rubber [This Incl Synthetic Rubber] [No Sub-Division Formation]
42 Plastic, Pvc Articles Incl Packaging Products And Footwear Plastic Or Pvc [Note: Toys Will Go Sub-Division 933]
43 Hides, Skins And Leather
44 Article Of Leather, Sadlary And Harness, Travel Goods, Hand Bags And Similar Products Of Animal Gut
45 Fur Skins And Artificial Fur & Articles Thereof
49 Misc Products/Articles Of Rubber/Pvc/Leather Etc [This Will Incl Mixed Products Of Rubber/Leather Etc.]
51 Wood And Pulp Of Wood And Products Thereof Excl Forestry
52 Cork, Thermocols And Articles Thereof
53 Manufacture Of Straw, Basketware And Wickerware
54 Pulp Of Wood & Other Fibrous Materials
55 Paper And Paper Board
56 Printed Books, Newspaper, Periodicals, Registers Etc
57 Packing Materials And Containers Of Paper
61 Silk And Silk Products [This Includes Art Silk]

62 Wool/Animal Hair, Yarn & Fabrics
63 Cotton, Cotton Yarn And Fabrics
64 Synthetic (Man-Made) And Mixed Textiles
65 Jute, Coir & Other Natural Fibre Yarn & Textile
66 Special Woven Fabric And Articles Thereof [No Sub-Group Formation]
67 Floor Coverings Etc [No Sub-Group Formation]
68 Impregnated, Coated, Covered, Laminated Textile Fabrics [No Sub-Group Formation]
69 Misc Textile Yarn, Fabrics, Articles & Waste N.e.c [No Sub-Group Formation]
71 Iron & Steel (Incl Stainless Steel) & Articles Thereof
72 Copper, Nickel & Zinc Articles Thereof
73 Aluminium, Tin, Lead & Other Base Metals & Articles Thereof
74 Misc. Manufacture Of Base Metals, N.e.c [No Sub-Group Formation]
75 Non-Electrical Machine Tools & General Purpose Machineries And Components And Parts Thereof [No Sub-Group Formation]
76 Non-Electrical Industry Specific Equipment/Machineries Incl Parts Thereof
77 Electrical & Electronic Machinery & Equipment Incl Parts (Excl Medical & Non-Conventional Energy Equipment)
78 Electronics Equipment & Parts Excl Bio-Medical Equipment
79 Special Purpose Machineries And Equipment And Parts [This Encl Medical/Bio-Medical/Surgical Laboratory/Body Fitness Machinery & Parts Thereof]
81 Railway/Metro-Railway/Trams & Rolling Stock
82 Road Surface Vehicle Excl Railways & Parts N.e.c
84 Human Being/Animal Driven Cart Etc
85 Aircraft And Spacecraft, Ship & Associated Transports
91 Optical, Photographic, Watch And Other Precision Equipment, Musical Instruments And Parts Thereof
92 Precious & Semi-Precious Stone/Jewellary; Works Of Art And Other Decorative Items N.e.c
93 Sports, Athletic Goods, Toys & Amusements Equipment And Human Safety Articles
94 Glass And Fibre Glass, Ceramic, Porcelain, Asbestos And Cement And Articles Thereof Incl Articles Of Stone
95 Miscellaneous Classified Manufacturing Articles And Parts Thereof
96 Misc Non Classified Manufactured Articles & Parts
97 Service Related Items
99 Miscellaneous Items Of Asi N.e.c
End of two digits codes

Three digits codes
111 Live Animals Chiefly For Food
112 Meat & Meat Products Edible
113 Fish & Fish Products Incl. Crustaceans & Other Aquatic Invertebrates Chiefly Used For Human Consumption. [This Sub-Division Does Not Incl. Marine Mammals, Caviar Or Caviar Substitutes Prepared From Fish Eggs]
114 Dairy Products, Poultry, Birds, Egg, Honey, Other Edible Products Of Animal Origin
115 Animal Fats & Oils And Cleavages Products Incl. Edible Fats, Animal Waxes [This Chapter Does Not Incl. Cocoa, Butter Fat, Fatty Acid]
116 Preparation Used For Animal Feeding
119 Other Produce Of Animal Origin
121 Edible Fruits & Nuts; Edible Vegetables And Certain Roots And Tubes And Their Products. [This Sub-Division Incl Both Fresh & Preserved Fruits & Nuts & Vegetables Incl. Cashewnut/Coconut Kernel]
122 Tea, Mate, Coffee And Spices
123 Cereals (Including Rice) And Pulsed Unmilled
124 Products And Milling Industries; Malt & Malted Milk, Starches, Inulin, Wheat Gluten Etc.; Flour & Flakes Of Potato And Dried Leguminous Vegetables. [This Sub-Division Excludes Rice]
125 Vegetable Oils & Fats & Their Cleavage Products And Vegetable Waxes

126 Oil Seed, Oilgenous Fruits, Misc Grains, Seeds And Misc Fruits; Edible Flour And Meals Of Oil Seed. [This Heading Incl Palm Nuts And Kernals Mango Kernels But Not Cashewnut Or Cocnut Kernels Which Will Fall Under 121]
127 Lac Gum, Resins And Other Vegetable Saps & Extracts
128 Residues & Waste Prepared Animal Fodder Excl Oilseed
129 Vegetable Products N.e.c Incl Waste, Fodder, Vegetable Plaiting Like Baboos, Bamboo Canes Split Used For Plaiting
131 Sugar, Molasses, Khandsari, Gur [This Sub-Division Does Not Incl Chemically Pure Sugars But Will Include Sacrose, Lactose, Maltose & Glucose Excl Medicaments Under Division 33]
132 Chocolate, Cocoa & Cocoa Preparations & Sugar Confectionery
133 Macaroni, Noodles, Couscous & Similar Products
134 Bakery Products
135 Fruit Juices & Vegetable Juices & Syrup, Pickles Chutneys Fruit Pulp Etc
139 Miscellaneous Edible Preparations And Food Products N.e.c [This Incl Food Colouring & Flavouring Material, Food Preservatives But Excl Yeasts Used For Laboratory Use]
141 Live Trees, Plants, Bulbs, Roots, Cut Flower & Ornamental Foliage
142 Produce Of Hydroponics
143 Produce Of Tissue Culture
144 Forestry & Logging Products
145 Produce Of Seri Culture
151 Beverages, Spirits And Vinegar
152 Soft Drinks, Mineral Water And Other Edible Preparations Of Water
153 Tobacco And Manufactured Tobacco Substitutes
154 Ice Including Edible Ice Products
156 Pan And Pan Masala And Related Ingredients
161 Natural Water & Ice For Industrial Use [This Excl Water For Clinical Use]
163 Misc Non-Edible Substances Of Animal/Vegetable Origin N.e.c
211 Salts, Sulpher, Lime, Stone, Granites & Marble
213 Gypsum, Plaster
214 Clay, Kaolin, Earth, Graphite, Sand, Quartz
215 Abrasive (Excluding Quartz) [Quartz Classified Under Sub-Division 214]
219 Crude Minerals
221 Iron Ores & Concentrates
222 Copper Ores & Concentrate
223 Aluminium Ores & Concentrates
224 Nickel Ores & Concentrates
225 Ash And All Residues Of Base Metals
226 Slags - All Types
227 Ores Of Precious And Semi-Precious Metals
229 Misc Ores & Concentrates N.e.c
231 Coal, Lignite: Products & By-Products
232 Petroleum Products & By-Products
233 Diesel Products & By-Products
234 Nuclear Fuel Incl Radio-Actives & Isotopes Compound; Alloys; Dispersions And Mixtures
239 Mineral Fuels, Solvent, Lubricants & Oils N.e.c
241 Gas (Fuel) Natural And Manufactured
251 Electrical Energy
291 Refractory Bricks, Blocks, Tiles & Similar Refractory Products [Ceramic Products Classified Under 943]
311 Inorganic Elements, Excluding Base Metals, Rare Gas,Halides And Sulphur Compunds
312 Inorganic Acids, Oxygen Compounds, Carbonates & Carbide
313 Sodium And Potassium Compounds
314 Inorganic Alkali And Compounds Thereof
315 Inorganic Halogen And Sulphur Compounds Of Metals/Non-Metals
316 Inorganic Nitrogen And Phosphorus Compounds

317 Inorganic Gases Including Rare Gases In Different States
319 Misc. Inorganic Compounds N.e.c
331 Vitamins
332 Hormones, Glands And Products Thereof
333 Antisera, Blood Fractions, Immulogical Products And Vaccines
334 Anti-Biotics & Preparations Thereof
335 Alkaloids & Preparations Thereof
336 Anti-Psychotic/Sedative/Hypnotic Medicines
337 Pharmaceutical Products Incl Family Planning Goods
338 Homeopath, Ayurvedic & Unani Medicines
339 Other Medicaments And Preparations Thereof
341 Organic Manures/Fertilizer
342 Inorganic Fertilizer [Crude Minerals Used As Fertilizer Is Classified Under Sub-Division 219]
343 Misc Fertilizer/Manures
344 Insecticides/Pesticides/Weedicites
349 Misc Chemical And Related Materials Used In Agriculture
351 Dyeing, Tanning Materials And Their Derivatives
352 Colouring Materials, Ink, Artist Ink, Paint Etc [This Excl Food Colours Vide Sub-Division 139]
353 Polishing (Non-Wax), Emulsifying, Distemper And Related Materials
361 Essential Oil And Essence
362 Cosmetics, Perfumes
363 Soap, Detergent, Whitening & Cleaning Materials [This Excl Medicated Products]
364 Wax & Polishing Materials
365 Dental & Dentistry Materials Excl Wax
371 Propellant/Explosive Powder
372 Detonator/Percussions Fuse Etc
373 Fire Works, Signaling Flare & Other Polytechnic Articles
374 Ferro-Cerium & Related Products
381 Photographic Film, Paper And Plates
382 Cinematography/Video Film, Cassette
383 X-Ray Film & Plates N.e.c
384 Misc Photosensitive Non-Photographic Film, Plates, Cloth, Drums Etc
389 Misc Photography, Cinematography Supplies
421 Bags/Boxes/Panels/Containers Of Plastic/Pvc
422 Tubes/Pipes/Basin & Sanitary Fittings Of Plastic/Pvc
423 Footwear Plastic/Pvc
424 Film (Non-Sensitive/Photographic)/Foil/Rolls/Tape/Rope Of Plastic/Pvc & Related Materials
425 Sheet/Linear/Cloth/Laminated Sheet Of Plastic/Pvc
426 Misc Articles Of Plastic/Pvc
427 Bakellite, Acrylic & Similar Items & Articles Thereof
429 Articles, Parts Of Plastic/Pvc N.e.c
431 Hides
432 Skins
433 Leather
441 Leather Bags, Cases, Purse & Other Novelty Items
442 Leather Apparels
443 Leather Footwear & Parts Thereof
449 Misc Leather Manufactured Items
451 Fur Skins And Articles Thereof
452 Artificial Fur And Articles Thereof
511 Wood & Wood By-Products [This Excl Forestry & Logging Products Classified Under Sub-Division 144]
512 Wooden (Incl Plywood) Furniture, Boxes (Incl Packing Box) And Other Wooden Articles
513 Charcoal

519 Miscellaneous Wooden Products
521 Cork And Articles Thereof
522 Thermocol And Products Thereof
531 Manufacture Of Straw, Basketware, Wickerware, Caneware
541 Pulp Of Wood & Other Fibrous Materials
551 Paper Used For Writing/Printing/Graphic Design/Computer Stationary
552 Paper (Uncoated) Used For Newsprint And For Other Special Purpose Excl Writing/Printing
553 Paper Coated, Cellulose Wading, Impregnated
554 Craft Paper And Paper For Special Use
555 Boards, Paper Boards All Kind
561 Printed Books, Newspaper, Periodicals, Note Books, Register Etc And Other Printed Matters
571 Packing Materials Made Of Paper [Paper Used For Packing Are Shown Under Respective Heads Under Division 55]
573 Misc Articles Of Paper Used As Container Etc
611 Silk Worms Incl Cocoons, Raw Silk, Raw Silk Waste
612 Silk & Art Silk Yarn Incl Silk Thread & Waste Thereof
613 Silk Fabrics Incl Silk Waste Fabrics
614 Silk Made Up Articles Incl Apparel
619 Other Silk Textile Goods
621 Wool Incl Fine, Raw, Coarse; Waste Of Animal Hair Incl Fur [Fur & Articles Thereof Classified Under Division 45]
622 Yarn Of Wool/Animal Hair Excl Fur [Fur & Articles Thereof Classified Under Division 45]
623 Fabrics Made Of Wool/Animal Hair & Waste Thereof [This Sub-Division Excludes Fur And Articles Thereof]
624 Made Up Articles Of Wool, Animal Hair Incl Apparel [This Sub-Division Excludes Fur And Articles Thereof]
629 Other Textile Products, Made Of Wool & Animal Hair.
631 Ginned Cotton, Cotton, Cotton Waste (Raw)
632 Cotton Yarn, Fibre Incl Cotton Thread
633 Cotton Fabrics Incl Cotton Hosiery Fabrics
634 Made Up Articles Of Cotton Incl Apparel
639 Other Cotton Textile Goods N.e.c
641 Man-Made Fibre Raw, Tow Of Viscose, Polyester, Filaments, Chips, Powder And Waste
642 Man-Made Yarn/Fibre And Waste Thereof
643 Fabrics Of Man-Made Fibre
644 Made-Up Articles Of Man-Made Fabrics
649 Man-Made Fibre Articles N.e.c
651 Raw Fibre Of Jute, Coir, Sisal, Hemp, Mista & Other Natural Fibre And Waste Thereof
652 Processed Fibre/Yarn Of Jute, Coir, Sisal, Hemp, Mista And Other Natural Fibre
653 Fabrics, Cloth Of Jute, Coir, Sisal, Hemp, Mista Etc
654 Man-Made Articles Natural Fibre
659 Other Jute And Natural Fibre Goods N.e.c
711 Pig Iron/Ferro Alloy Etc In Primary Form
712 Semi-Finished Products Of Iron/Steel
713 Finished Products Of Iron/Steel
714 Stainless Steel In Primary & Finished Forms
715 Manufacture Of Iron/Alloy Steel
721 Copper & Copper Alloy, Refined Or Not, Unwrought
722 Copper And Copper Alloy, Worked
723 Nickel And Nickel Alloys, Refined Or Not, Unwrought
724 Nickel And Nickel Alloys, Worked
725 Zinc And Zinc Alloys, Refined Or Not, Unwrought
726 Zinc And Zinc Alloys, Worked
731 Aluminium And Aluminium Alloys, Unwrought
732 Aluminium And Aluminium Alloys, Worked

733 Tin And Tin Alloys Unwrought
734 Tin And Tin Alloys Worked & Tin Plates Iron/Steel Products
735 Lead And Lead Alloys, Unwrought
736 Lead And Lead Alloys, Worked
737 Other Base Metals & Alloys Thereof, Unwrought
738 Other Base Metals & Alloys Thereof, Worked
761 Agricultural & Forestry Machineries/Parts Thereof
762 Food, Beverages & Tobacco Processing Machineries & Parts
763 Minings, Quarrying & Metallurgical Machineries/Parts
764 Construction/Cement Machineries & Parts
765 Textile, Leather & Rubber Processing, Paper Printing Machineries & Parts Thereof
766 Non-Electrical Domestic/Office Appliances & Parts
767 Chemical/Plastic/Glass/Weapon/Ammunition Machineries And Parts Thereof
768 Lift And Lifting Equipment, Fixed Or Mobile & Parts Thereof
769 Misc Non-Electrical Machineries And Parts Thereof, N.e.c
771 Electrical Machinery/Equipment
772 Electrical Motors, Generators, Transformer, Power Pack [This Incl Pump Set Fitted With Electric Motor]
773 Switch, Switch-Gear, Control Panel, Circuit Breakers Etc And Parts Thereof
774 Lamp, Filament, Electrodes/Anodes/Connectors, Fittings & Parts
775 Measuring/Controlling/Regulating Instruments
776 Battery, Accumulators, Cells And Parts Thereof
777 Domestic And Office Electrical Equipment
778 Electro Magnet, Fans, Armature, Coils & Electro-Magnetic Equipment
779 Electrical Equipment, Parts And Accessories, N.e.c
781 Telephone/Telecommunication/Transmission Equipment
782 Audio/Video/Sound Apparatus & Parts
783 Computer & Computing Equipment & Peripherals & Parts
784 Electronic Valves/Tubes & Components
785 Electronic Cards & Its Components
789 Other Electronic Components & Parts
791 Non-Conventional Energy Generation Equipment
792 Machineries For Re-Cycling Materials
793 Special Application Equipment, Nuclear Reactor Etc & Parts
794 Fabricated & Structural Construction & Machinery
795 Medical/Bio-Medical, Surgical, Laboratory And Health Fitness Equipment & Parts
811 Railways, Metro-Railways, Trams & Rolling Stock
812 Railways Rolling Stock N.e.c
813 Metro Railways And Tramways & Rolling Stock
821 Motor Vehicles - Passengers/Goods Transportation And Special Purpose Vehicles
822 Chassis/Body - Motor Vehicles, Parts N.e.c
823 Motor Cycle, Scooter, Moped & Parts N.e.c
824 Non-Motorized Cycles/Wheel Chairs & Parts N.e.c
845 Cart, Animal Driven/Hand-Driven
851 Aircraft/Helicopter & Other Flying Machines
852 Ships, Boats And Other Waterways Transports
911 Optical & Photographic Equipment & Parts
912 Precision Equipment For Measuring Solar Other Non-Conventional Energy
913 Survey, Design & Laboratory Equipment & Parts
914 Clock Scientific/Quartz And Parts Thereof
915 Musical/Fine Arts Instruments & Parts Thereof
921 Precious & Semi-Precious Stone/Jewellary Etc
922 Works Of Art Including Imitation, Fancy And Decorative Items
923 Hair, Fine Fibre, Natural/Artificial & Articles Thereof
931 Sports And Athletic Accessories

```
932 Human Safety Articles & Parts Thereof
933 Toys And Amusement Articles
934 Arms & Ammunitions & Parts Thereof
935 Domestic/Industrial Safety Aricles & Parts Thereof
941 Glass And Glassware
942 Fibre Glass & Articles Thereof
943 Ceramic, Porcelain And Articles Thereof
944 Asbestos, Cement And Articles Thereof
951 Currency Coins
952 Common Purpose Stationary Items
954 Advertising Materials/Boards (Completed) N.e.c
961 Manufactured Items & Parts N.e.c
971 Sales, Maintenance, Repair & Trade Services
972 Hotel And Restaurant Services
973 Communication, Transport & Storage Services
974 Real Estate, Leasing & Rental Services, Financial Intermediation Services
975 Legal, Accounting, Auditing & Book Keeping Services
976 Film/Video/Advertising/Decorative Industry Services
977 Architectural, Engineering & Other Technical/Consultancy Services
978 Services Provided By Extra Territorial Organisation
979 Community, Social, Personal Services & Misc
992 Unspecified Inputs/Outputs
999 Total Consumption/Control Totals
## End of Three digits codes ##

### END of codes to select from###

"The codes provided are hierarchical in terms of digits. Codes with 3 digits
are more specific than codes with 2 digits. You should prefer to choose the
finest code possible.
```

We follow this with a message asking for the assignment given the particular firm description.

```
The company we are looking at produces: FIRM DESCRIPTION.
Which are the best matches from the codes PROVIDED earlier?
Please provide up to five (5) matches, separated by ';' and list them in order from best to worse
    .
You do not need to provide a code if the product is too vague.
If not possible to select any option, you should PRINT 'None'.
```

Contact.

MIT CEEPR Working Paper Series
is published by the MIT Center for Energy and Environmental Policy Research from submissions by affiliated researchers.
For inquiries and/or for permission to reproduce material in this working paper, please contact:

General inquiries: ceepr@mit.edu
Media inquiries: ceepr-media@mit.edu

MIT CEEPR
Center for Energy and
Environmental Policy Research

MIT Center for Energy and
Environmental Policy Research
Massachusetts Institute of Technology
77 Massachusetts Avenue, E19-411
Cambridge, MA 02139-4307
USA

ceepr.mit.edu

MASSACHUSETTS INSTITUTE OF TECHNOLOGY