Propagation and Amplification of Local Productivity Spillovers

Xavier Giroud* Simone Lenzu† Quinn Maingi‡ Holger Mueller§

March 2024

Abstract

The gains from agglomeration economies are believed to be highly localized. Using confidential Census plant-level data, we show that large industrial plant openings not only raise the productivity of local plants but also of distant plants hundreds of miles away, which belong to large multi-plant, multi-region firms that are exposed to the local productivity spillover through one of their plants. This "global" productivity spillover does not decay with distance and is stronger if plants are in industries that share knowledge with each other. To quantify the significance of firms' plant-level networks for the propagation and amplification of local productivity shocks, we estimate a quantitative spatial model in which plants of multi-region firms are linked through shared knowledge. Counterfactual exercises show that while large industrial plant openings have a greater local impact in less developed regions, the aggregate gains are greatest when the plants locate in well-developed regions, which are connected to other regions through firms' plant-level (knowledge-sharing) networks.

Keywords: Productivity spillover; plant-level networks; agglomeration economies.

* Columbia University, NBER, and CEPR. Email: xavier.giroud@gsb.columbia.edu; † New York University. Email: slenzu@stern.nyu.edu; ‡ New York University. Email: rmaingi@stern.nyu.edu; § New York University, NBER, CEPR, and ECGI. Email: hmueller@stern.nyu.edu.
We thank the editor (Dave Donaldson), three anonymous referees, Costas Arkolakis, David Baqaee, Adrien Bilal, Paco Buera, Lorenzo Caliendo, Gabriel Chodorow-Reich, Xavier Gabaix, Cecile Gaubert, Elisa Giannone, Rick Hornbeck, Matthias Kehrig, Nelson Lind, Atif Mian, Joan Monras, Ezra Oberfield, Esteban Rossi-Hansberg, Pete Schott, Amit Seru, Andrei Shleifer, Johannes Stroebel, Amir Sufi, Sharon Traiberman, and seminar participants at Princeton, Harvard, Stanford, Chicago, Columbia, NYU, UCLA, Minnesota, Penn State, WashU, USC, LBS, Emory, U Washington, U Georgia, and many other institutions and conferences for helpful comments. We are grateful to Jun Wong for excellent research assistance. Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this information product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 1908 (CBDRB-FY21-P1908-R9062, CBDRB-FY23-0135).

1 Introduction

At least since Marshall (1890), economists have hypothesized that spatial proximity between firms may generate productivity spillovers. These productive externalities may explain why firms locate near one another, and why state and local governments spend billions of dollars in subsidies to attract firms to their jurisdictions. But how large are these externalities, and how broad is their reach? In this paper, we show that local productivity spillovers can propagate through the entire economy through the plant-level networks of multi-region firms. Building on the empirical framework in Greenstone, Hornbeck, and Moretti (2010, GHM), we show that the opening of large industrial plants ("Million Dollar Plants," "MDPs") not only raises the productivity of local incumbent plants but also of distant plants hundreds of miles away, which belong to large multi-plant, multi-region firms that are exposed to the local spillover through one of their plants. Thus, our results suggest that the (already large) agglomeration spillover effects identified by GHM (i) are undercounted and (ii) accrue to other locations as well. Our results have important policy implications. For one, they imply that local industrial policies aimed at attracting investments, such as MDPs, may have positive effects on productivity in other regions. The traditional view is that such policies, by diverting resources away from other regions, have a zero (or negative) effect on aggregate productivity. Moreover, our results imply that local policymakers are unlikely to fully internalize the productive externalities generated by their investment policies. While these policies may benefit local plants, they also benefit plants outside the policymakers' jurisdictions.

Marshall famously divided agglomeration economies into three broad categories: i) labor market pooling, ii) knowledge spillovers, and iii) input-output linkages. We find no evidence that the productivity gains at either local or distant plants are the result of input, output, or any other trade linkages with the MDP. And while labor market pooling may contribute to the local productivity spillover, it is unlikely that a thicker labor market in the MDP county would raise the productivity of distant plants hundreds of miles away. Knowledge, on the other hand, can be used in local and distant plants alike. Indeed, once it spills over to a firm's local plant, it can be freely shared with other plants of the firm (Markusen, 1984). Consistent with this idea, we find that the productivity gains at distant plants do not decay with distance to the MDP. By contrast, the local productivity spillover decays rapidly with distance: it is strong within a 50-mile radius around the MDP, weaker within 100 miles, and insignificant beyond. Hence, while productivity spillovers between plants of *different* firms are highly localized, they seem to propagate without much friction between different plants of the *same* firm. Finally, and also consistent with knowledge sharing, we find that the "global" productivity spillover is stronger if the MDP and the distant plant are in the same industry or in industries that share knowledge with each other, as measured by mutual R&D flows and patent citations.

To quantify the significance of plant-level networks for the propagation and amplification of productivity spillovers, we develop and estimate a quantitative spatial equilibrium model with goods trade, labor mobility, input-output linkages, plant-level networks, and a rich and realistic geography. While our model builds on the canonical framework developed in Allen and Arkolakis (2014), Ahlfeldt et al. (2015), Redding (2016), and Monte, Redding, and Rossi-Hansberg (2018), we depart from this framework in a number of significant ways.[1] In our model, heterogeneous plants belonging to different sectors produce differentiated goods. Within a region and sector, plants differ in their productivities and parent firms. Plants of the same firm, across regions, are linked through shared knowledge. Specifically, we assume that plant-level productivity depends on local knowledge and knowledge in other regions in which the parent firm operates. This generates productivity linkages across regions and provides a direct mechanism through which productivity shocks in one region propagate to other regions.[2] Despite this complexity, our model admits a simple representation where, within a region and sector, plants' productivities aggregate into a single productivity index. Across regions and sectors, this nests our model into an augmented (economic geography) version of the multi-region, multi-sector model of Caliendo and Parro (2015).

The typical economic geography model focuses on regional outcomes; parameters can thus often be identified using regional aggregates. By contrast, our model features within-region, across-plant heterogeneity. Thus, plant-level micro moments are needed to identify the model's parameters. In our estimation, we target as moments reduced-form difference-in-differences estimates—semi-elasticities of plant-level employment, wages, and productivity with respect to the MDP openings—both for local and distant plants.

[1]Redding and Rossi-Hansberg (2017) provide a taxonomy of the various modeling assumptions and building blocks in quantitative spatial models.

[2]Our model is designed to study productivity spillovers within firms across regions; this sets it apart from models that focus on the evolution of the aggregate productivity distribution (e.g., Lucas and Moll, 2014; Perla and Tonetti, 2014).

To generate model-based estimates that correspond to these reduced-form estimates, we simulate MDP openings in our model economy. This provides us with "pre-" and "post-shock" observations, allowing us to estimate plant-level difference-in-differences regressions that closely mirror those in our reduced-form analysis.

Given our parameter estimates, we undertake counterfactual analyses to quantify the role of within-firm, across location ("global") knowledge sharing for the diffusion and amplification of local productivity shocks. In our first counterfactual, we quantify the aggregate welfare effects of large plant openings, with a focus on the underlying propagation forces. We find that global knowledge sharing and input-output forces have roughly similar amplification effects. Also, they interact in meaningful ways.

Opening a large industrial plant can have a significant impact on a region, especially for less developed regions. However, in the data, almost all of the MDPs open in regions that are already well developed. This raises the question of whether government should intervene to aid less developed regions. In our second counterfactual, we inform this policy debate by randomly assigning large plant openings to more or less developed regions. We find an ambiguous role for regional development. On one hand, the impact on local incumbent plants is greater if the plant opens in a less developed region. In our model, less developed regions have a lower stock of knowledge (recovered from the data), so the relative productivity gains at incumbent plants are higher in those regions. On the other hand, due to global knowledge sharing, the impact on the rest of the economy is greater if the plant opens in a well-developed region, which is connected to other regions through plant-level (knowledge-sharing) networks. We find that the second effect dominates, so the aggregate gains are greatest if the plant opens in a well-developed region, consistent with the observed location choices of the MDPs in the data.

Our paper is related to several strands of literature. First and foremost, our paper contributes to the empirical literature on estimating agglomeration economies. A robust finding in that literature is that these economies are highly localized.[3] We too find that local productivity spillovers, which take place between plants of different firms, decay rapidly with distance. However, we find that global productivity spillovers, which take place between different plants of the same firm, do not decay with distance. Our findings

[3]Numerous empirical studies find that agglomeration economies are highly localized; see the survey articles by Audretsch and Feldman (2004), Duranton and Puga (2004), Rosenthal and Strange (2004, 2020), and Combes and Gobillon (2015). GHM provide well-identified reduced-form estimates of agglomeration spillovers using the MDP openings as natural experiments.

inform the academic debate about the spatial reach of agglomeration economies. In a recent article, Rosenthal and Strange (2020, p. 27) write:

> "Implicit in the idea that spatial concentration increases productivity is another idea: the degree of proximity matters. Agglomeration economies must decay with distance. How close, then, do firms and workers need to be to each other to benefit from agglomeration economies?"

Our paper provides a nuanced answer. On one hand, firms must have a nearby plant to benefit from knowledge spillovers. As we show, the local agglomeration externality is strongly significant only within a 50-mile radius around the MDP. However, not all of a firm's plants need to be located nearby. In fact, it may suffice if only one of the firm's plants is located in close proximity to the MDP. Once the knowledge spills over to that plant, it can be freely shared with other plants of the firm, thus raising the productivity of distant plants hundreds of miles away.

Second, our paper is related to the empirical literature on place-based policies (e.g., Greenstone, Hornbeck, and Moretti, 2010; Busso, Gregory, and Kline, 2013; Kline and Moretti, 2014b; Gaubert, Kline, and Yagan, 2020).[4] A fundamental insight from this literature, well voiced in Glaeser and Gottlieb (2008), is that when agglomeration elasticities are equal across locations, laissez-faire is efficient, and place-based policies are welfare-reducing. In our setting, agglomeration elasticities are *not* equal across locations, because multi-plant, multi-region firms are not evenly spread across space. As a case in point, our counterfactual analysis shows that the (positive) aggregate effects of large industrial plant openings are greatest if the plant opens in a well-developed region, which is connected to many other regions through firms' plant-level networks.

Third, various papers study how shocks propagate across regions or within firms. Caliendo et al. (2018) and Hornbeck and Moretti (2022) show that regional productivity shocks may impact other regions through trade and labor flows. Our paper shows that regional productivity shocks spread throughout the economy through the plant-level networks of multi-region firms. At the firm level, Barrot and Sauvagnat (2016) and Carvalho et al. (2021) show that local shocks caused by natural distasters may spread across regions through firms' supply-chain networks. Our paper studies production networks *within*

[4] Glaeser and Gottlieb (2008), Moretti (2010), Kline and Moretti (2014a), and Neumark and Simpson (2015) discuss the economics of place-based policies. Bartelme et al. (2021) study optimal industrial policies at the sectoral (as opposed to the regional) level.

firms. Also, in our setting, the within-firm diffusion of shocks is not driven by input-output linkages. More closely related to our paper, Ding (2023) studies how shocks spread *within* multi-plant, multi-industry firms. Specifically, he finds that a positive demand shock in one industry of a firm increases its sales in another industry, and the effect is stronger the more the two industries share knowledge inputs. Importantly, Ding provides compelling evidence that knowledge is non-rival within the firm, consistent with the premise of, and results in, our paper. Finally, Giroud and Mueller (2019) study how local demand shocks spread across regions through the internal networks of multi-region, multi-establishment firms. Unlike our paper, which focuses on within-firm knowledge sharing, Giroud and Mueller focus on firm-wide financial constraints as a mechanism that generates linkages between a firm's establishments, and ultimately between regions.

The remainder of this paper is organized as follows. Section 2 presents our main reduced-form results. Section 3 explores potential mechanisms. Section 4 develops a quantitative spatial model in which plants of multi-region firms are linked through shared knowledge. Section 5 contains the structural estimation of the model. Section 6 presents counterfactual analyses. Section 7 concludes.

2 Reduced-Form Evidence

2.1 Research Design

We examine how local productivity spillovers propagate across U.S. regions through the plant-level networks of multi-region firms. To identify local productivity spillovers, we build on the natural experiments in GHM, who study the impact of MDP openings on the productivity of incumbent plants. In their setting, plants in (one or more) "runner-up" counties, which narrowly lost the competition, serve as counterfactuals for plants in the "winner" county where the MDP located.[5] We match the MDP openings in the Appendix of Greenstone and Moretti (2003) to plants in the U.S. Census Bureau's Standard Statistical Establishment List (SSEL) based on firm and county name. Exactly as in GHM, we have 11 MDP openings between 1982 and 1985, 18 MDP openings between 1986 and 1989, and 18 MDP openings between 1990 and 1993, adding up to 47 MDP openings.

[5]Winner and runner-up counties are from the reported location rankings of firms in the corporate real estate journal *Site Selection*. The journal includes a regular feature article, *Million Dollar Plants*, that describes where a firm decided to locate a large manufacturing plant. The feature article was last published in 1993.

We use confidential plant-level data from the Census of Manufactures (CMF), the Annual Survey of Manufactures (ASM), and the Longitudinal Business Database (LBD) provided by the U.S. Census Bureau. The CMF and ASM contain information about key plant-level variables, such as shipments, assets, material inputs, employment, payroll, capital expenditures, industry, and location. The LBD contains longitudinal establishment identifiers along with data on employment, payroll, industry, location, and firm affiliation. We first consider the local productivity spillover from the MDP opening on incumbent plants in the winner county. For each MDP opening, we identify all incumbent plants in the winner and corresponding runner-up counties. We use all observations from five years before until five years after the MDP opening, leaving us with 157,000 plant-year observations.[6] We next consider the global productivity spillover on plants outside the winner county that belong to parent firms which have a plant in the winner county before and after the MDP opening.[7] For simplicity, we refer to these plants outside the winner county as "treated plants," though technically they are "indirectly treated," as they belong to firms which are exposed to the MDP opening through one of their plants. To avoid confusion, we refer to incumbent plants in the winner county not as "treated" but simply as "plants in the winner county" "or plants in the MDP county." We use various control groups, leaving us with either 1,407,000, 1,046,000, or 423,000 plant-year observations. We always exclude the MDPs as well as any plants owned by the MDPs' parent firms. The sample period is from 1977 to 1998.

Table 1 provides summary statistics from the year before the MDP opening. Panel A shows plant-level statistics for the local spillover sample consisting of plants in the winner and runner-up counties. Panels B and C show plant- and firm-level statistics for the global spillover sample with 423,000 plant-year observations. This sample consists of (treated) plants outside the winner county that belong to firms with plants in the winner county (treatment group) as well as plants in the same (distant) counties as the treated plants that belong to firms with plants in the runner-up counties (control group).

[6]All sample sizes are rounded to the nearest 1,000 following U.S. Census Bureau disclosure guidelines.

[7]Since we estimate difference-in-differences regressions, we require all plants to have "before" and "after" observations, both in the local and global spillover sample. We do not require plants to be continuously present during the entire 11-year window around the MDP opening. Given that the ASM gets re-sampled every five years, this implies that smaller plants in the ASM, which are sampled probabilistically, will be in our analysis sample for at least five years, e.g., from two years before until two years after the MDP opening. Larger plants are sampled with certainty into the ASM and will be in our analysis sample much longer, often for the entire 11-year window surrounding the MDP opening.

2.2 Local Productivity Spillover

We first consider the local productivity spillover from the MDP opening on incumbent plants in the winner county. We estimate the following specification:

$$y_{ickst} = \xi_c + \xi_k + \xi_{st} + \beta_1 Post_{ct} + \beta_2 (Winner_i \times Post_{ct}) + \varepsilon_{ickst}, \tag{1}$$

where y_{ickst} denotes plant-level productivity (TFP), i denotes counties, c denotes cases, k denotes plants, s denotes industries, t denotes years, $Post_{ct}$ is an indicator for case c that is one from the year of the MDP opening onward, $Winner_i$ is an indicator for the winner county, and ξ_c, ξ_k, and ξ_{st} denote case, plant, and industry × year fixed effects, respectively.[8] A "case" comprises the winner county and associated runner-up counties. The case and plant fixed effects capture time-invariant heterogeneity across cases and plants, respectively. Importantly, the case fixed effects ensure that the impact of the MDP opening on incumbent plants is identified from comparisons within a given winner-loser pair. The industry × year fixed effects capture time-varying shocks at the industry level. Industries are defined at the 3-digit SIC code level. The main coefficient of interest is β_2, which captures the mean change in productivity at plants in the winner county relative to plants in the runner-up counties.

Table 2 presents the results. In this and all other tables, we only report the main coefficient(s) of interest and write "MDP" in lieu of $Winner_i \times Post_{ct}$ for brevity. As column (1) shows, the MDP opening raises the productivity of incumbent plants in the winner county by 4% relative to plants in the runner-up counties.[9] In all our regressions, we weigh observations by plant-level employment from five years before the MDP opening or, if not available, from the earliest available pre-treatment year. As column (2) shows, the result is similar if we do not weigh by employment. Finally, in column (3), we examine if the local productivity spillover decays with distance to the MDP. To this end, we identify all plants within a 250-mile radius around the MDP. We create three dummy variables,

[8] TFP is the estimated residual from a plant-level regression of output on capital, labor, and material inputs (in logs). To allow for different factor intensities across industries and over time, we estimate the regression separately for each 3-digit SIC code industry and year. Accordingly, TFP can be interpreted as the relative productivity of a plant within a given industry and year.

[9] This estimate lies within the range of TFP estimates in GHM (1.46% to 6.13%), albeit it is slightly lower than their baseline estimate (4.77%). While we require plants to be present before and after the MDP opening, GHM require plants to be continuously present in the eight years before the MDP opening. This excludes many smaller plants in the ASM, which are randomly sampled every five years. Table A.1 of Online Appendix A replicates GHM's baseline result using their specification and sample selection procedure.

$(<\ 50\,\text{miles})_k$, $(50\,\text{to}\,100\ \text{miles})_k$, and $(100\,\text{to}\,250\ \text{miles})_k$ indicating whether a plant lies within 50 miles, between 50 and 100 miles, or between 100 and 250 miles of the MDP and interact these dummy variables with both terms in equation (1). As is shown, the local spillover decays rapidly with distance. It is strong within 50 miles of the MDP, much weaker within 100 miles, and insignificant beyond. Hence, productivity spillovers *between* (plants of different) firms are highly localized, consistent with the empirical literature on agglomeration economies.

2.3 Global Productivity Spillover

We next consider the global productivity spillover on (indirectly) treated plants outside the winner county that belong to parent firms with plants in the winner county. We estimate the same difference-in-differences specification as before, except that $Winner_i$ is now an indicator for whether the plant's parent firm has a plant in the winner county before and after the MDP opening, and a "case" includes all treated plants as well as all plants in the corresponding control group.[10]

Table 3 presents the results. In column (1), the choice of control group is motivated by the local spillover analysis. To account for firm-level exposure to unobserved shocks that may affect both winner and runner-up counties, the control group consists of all plants outside the runner-up counties that belong to parent firms with plants in the runner-up counties ("runner-up firms"). As in the local spillover analysis, this specification includes plant, industry × year, and case fixed effects. Thus, we compare treated plants with plants of runner-up firms in the same industry and year, but possibly in different counties.

Borusyak and Hull (2023) point out that omitted variable bias (OVB) may confound network spillover regressions if exposure to the shock is non-random, even if the original shock is exogenous. In our global spillover setting, even if the MDPs are as-good-as randomly assigned, (indirect) exposure to the MDP openings depends on existing firm

[10]For expositional clarity, we estimate the local and global spillover effects in separate regressions; they are based on (almost completely) non-overlapping samples and employ different identification strategies. In principle, they could be jointly estimated in a pooled regression framework, where all variables and fixed effects are interacted with "local" and "global" indicators to preserve the different identification strategies. (Only 0.65% of observations in the pooled sample are in both the local and global spillover sample.) Note that the global spillover regression already fully controls for the local spillover effect by including county × industry × year fixed effects. In the local spillover regression, a small number of observations (0.57%) are also in the global spillover sample, meaning they could be affected by MDP openings in other counties (i.e., from other MDP cases). Removing these overlapping observations has no effect on our estimates.

networks, which are non-random. For example, plants in urban areas tend to be part of larger firm networks, with more connections to other counties, including MDP counties, implying a higher expected treatment exposure. OVB can then arise if urban areas have systematically higher unobserved productivity shocks for reasons unrelated to firm network exposure.[11] In column (2), we control for unobserved factors that may correlate with differences in expected treatment exposure along two dimensions: location and industry. Precisely, the control group consists of all plants of multi-county (MC) firms in the same county as the treated plant, while the inclusion of industry × county × year fixed effects absorbs any unobserved time-varying heterogeneity at the location × industry level that may correlate with expected treatment exposure. Moreover, and importantly, the inclusion of these fixed effects accounts for the possibility of common shocks between the county of the treated plant and the MDP county, thereby addressing concerns that the productivity gains at treated plants may be caused by common regional shocks rather than regional spillovers within firms' plant-level networks.

Finally, in column (3), we further refine the control group by recognizing that the quasi-random assignment of the MDP openings is only with respect to winner and runner-up counties. Precisely, the control group now only includes plants of *runner-up* (MC) firms in the same county as the treated plant. The fixed effects remain the same. Thus, we compare plants in the same county, industry, and year that belong to firms which either have plants in the winner county (treatment group) or in the corresponding runner-up counties (control group). This is our tightest specification and will be our baseline specification in all subsequent global spillover regressions.

As columns (1) to (3) of Table 3 show, the MDP opening raises the productivity of treated plants outside the MDP county by 1.8% to 2% relative to plants in the respective control group. This estimate is stable despite varying control groups and fixed effects.[12] While the gains at treated plants are lower than the corresponding gains at plants in the winner county, Table 1 shows that the typical treated MC firm has about 6.3 plants outside the winner county. Thus, the productivity gains at treated plants accrue to a large number

[11] Borusyak and Hull (2023) give the example of a "deworming RCT" in which individuals' educational outcomes depend on how many of their neighbors are dewormed. Even if the deworming is random, the expected (spillover) treatment depends on how many neighbors an individual has, which is non-random. Individuals in urban areas are likely to have more neighbors, and will thus receive higher spillovers, generating OVB if urban areas exhibit different educational outcomes for reasons unrelated to deworming.

[12] The results are similar if we do not weigh observations by plant-level employment; in the unweighted version of column (3), the coefficient is 0.016 with standard error of 0.005.

of plants in other counties, generating significant aggregate effects (see the quantitative exercise in Section 6.1).

2.4 Treatment Effect Dynamics

Figure 1 shows the dynamics of the treatment effect both for the local and global spillover.[13] There are three takeaways. First, there are no significant differences in pre-trends, providing support for the parallel trends assumption. Second, if the productivity gains propagate through firms' plant-level networks, then the global spillover should set in around the same time as—or at least not before—the local spillover. As is shown, both spillovers become significant one year after the MDP opening. Finally, both spillovers remain large until the end, suggesting the productivity gains are not transitory.

2.5 Employment and Wages

Table 4 studies the impact of the MDP opening on employment and wages. As is shown, employment and wages at plants in the winner county grow by 3.5% and 3.7%, respectively, which is roughly in line with the productivity gains. Likewise, employment at treated plants outside the winner county grows by 1.6%, which is again in line with the productivity gains. By contrast, wages at treated plants only increase by a small amount. This is not surprising. Only a small fraction of the plants in distant counties are treated, putting only mild pressure on local wages. As is shown in Section 5, our model can rationalize this muted wage response along with other central reduced-form moments.

2.6 Extensive Margin

The employment growth in Table 4 is along the intensive margin. Table 5 considers the extensive margin. As columns (1) and (2) show, there is entry of new plants in the winner county, but it is only statistically significant for single-county (SC) plants.[14] The newly

[13] The estimates are obtained using the imputation estimator of Borusyak, Jaravel, and Spiess (2023, BJS). Table A.2 of Online Appendix A shows the BJS estimates side by side with the corresponding OLS estimates.

[14] Columns (1) and (2) consider entry of MC and SC plants separately. If we consider total entry by all (SC and MC) plants in the winner county, we obtain a point estimate of 0.059 (standard error: 0.028), which is lower than the estimate of 0.1255 (standard error: 0.055) in column (1) of Table 9 in GHM. Our sampling and estimation procedures differ from GHM in a number of ways. Notably, GHM use plant-level data from the CMF and exactly two observations per county (one "pre-" and one "post-MDP" CMF; the CMF is conducted every five years in years ending in 2 and 7.) In contrast, we use all available plant-level data (ASM and CMF)

entering SC plants are small: their average size is only 19.9 employees—about half the size of incumbent SC plants—and their total manufacturing output share in the winner county in the five years after the MDP opening is only 0.9%. In columns (3) and (4), we examine if either SC or MC firms with plants in the winner county open or close plants elsewhere. In either case, there is no significant effect on plant openings outside the winner county.

3 Analysis of Mechanisms

According to Marshall (1890), agglomeration economies can be divided into three broad categories: labor market pooling, knowledge spillovers, and input-output linkages. Based on measures of economic distance between the MDP and the incumbent plants, GHM conclude that the local productivity spillover is consistent with either labor market pooling or knowledge externalities, but not with input-output linkages.[15]

3.1 Knowledge Sharing

While labor market pooling and knowledge externalities may both contribute to the local productivity spillover, it is unlikely that a larger labor market in the winner county would affect the productivity of distant plants hundreds of miles away. Knowledge, on the other hand, can be used in local and distant plants alike. Indeed, once the knowledge spills over to a firm's local plant, it can be freely shared with other plants of the firm (Markusen, 1984). To explore this idea, we examine if the productivity gains at treated plants become weaker as the distance to the MDP increases. In Table 6, we exclude all plants within 100 miles, 250 miles, or 500 miles, or within the same state or Census division as the MDP. As is shown, the estimates are stable and practically identical to the original estimate in Table 3. Hence, unlike the local productivity spillover, which takes place between (plants of) different firms, the global productivity spillover, which takes place between different plants of the same firm, does not decay with geographical distance.

---

from five years before until five years after the MDP opening, resulting in a county-year sample that is five times larger than GHM's county-year sample. When replicating GHM's extensive-margin result using their sampling and estimation procedures, we obtain a point estimate of 0.121 (standard error: 0.056). Using a two-sample Z-test, we find that this "replication coefficient" is not statistically different from our coefficient of 0.059 at conventional levels (Z-score: 0.99; p-value: 0.322).

[15] "Thus, the data fail to support the types of stories in which an auto manufacturer encourages (or even forces) its suppliers to adopt more efficient production techniques" (GHM, p. 576). Table A.3 of Online Appendix A confirms that input-output linkages play no significant role for the local productivity spillover.

Table 7 provides additional evidence consistent with knowledge sharing as a mechanism for the global productivity spillover.[16] In column (1), we interact both terms in equation (1) with an indicator for whether the treated plant is in the same 4-digit SIC code industry as the MDP. Plants in the same 4-digit SIC code industry produce similar goods and use similar production processes and therefore likely draw on similar knowledge.[17] As is shown, the global productivity spillover is stronger if the treated plant and the MDP are in the same industry. In columns (2) and (3), we interact both terms in equation (1) with measures of mutual knowledge flows at the industry-pair level from Ellison, Glaeser, and Kerr (2010). "Mutual R&D flows" captures how R&D in one industry flows out to benefit another industry; "mutual patent citations" captures the extent to which technologies associated with one industry cite technologies associated with another industry. As is shown, the global productivity spillover is stronger if the treated plant and the MDP are in industries that share knowledge with each other.

3.2 Trade with the MDP

If the MDP buys inputs from, or sells goods to, a local plant in the winner county, it may also find it easier to buy from, or sell to, other plants of the same firm in distant counties.[18] Exploring trade linkages in the local spillover setting is challenging, as the MDP and the incumbent plants are in the same county. By contrast, in the global spillover setting the MDP and the treated plants are in different counties—the average distance is 612.5 miles—providing us with informative tests to explore the trade channel.

Table 8 shows the results. In columns (1) and (2), we interact both terms in equation (1) with measures of input or output linkages between the industry of the treated plant

[16]In Table A.3 of Online Appendix A we apply the empirical tests in Table 7 to the local productivity spillover, with analogous results.

[17]The 4-digit SIC code classification is extremely fine; it comprises 459 manufacturing industries in the CMF/ASM. For example, "nitrogenous fertilizers" (SIC 2873), "phosphatic fertilizers" (SIC 2874), and "fertilizers, mixing only" (SIC 2875) all have *different* 4-digit SIC codes.

[18]Trade could result in higher measured TFP due to mismeasurement. For instance, if a plant has increasing returns to scale but is wrongly assumed to have constant returns to scale, an increase in observed TFP might be (incorrectly) attributed to an increase in fundamental productivity. GHM perform a battery of robustness tests to assuage concerns regarding TFP mismeasurement. Section VII.F of their paper specifically addresses the role of output prices and demand effects. In the global spillover setting, the main concern is that the TFP gains at treated plants might be driven by an increase in demand *from the MDP*. A general increase in demand from the winner county (e.g., due to the MDP opening) for goods produced in the treated plant's county cannot explain the global spillover result, as the control group consists of plants in the same distant county, industry, and year as the treated plant.

and the industry of the MDP. In column (3), we interact both terms in equation (1) with a measure of how tradable the treated plant's industry is. The idea is that, if the channel is trade, the global spillover should be stronger if the treated plant's industry is more tradable. To measure an industry's tradability, we use its geographical Herfindahl index based on the argument that less tradable industries (e.g., cement) are more geographically dispersed (Mian and Sufi, 2014). In column (4), we interact both terms in equation (1) with a measure of exports from the treated plant's county to the winner county (from the Commodity Flow Survey). The idea is that, if the channel is trade, the global spillover should be stronger if the treated plant is in a county that exports more to the winner county. Finally, in column (5), we interact both terms in equation (1) with the geographical distance between the treated plant and the MDP. According to the gravity equation, trade flows should be declining in distance. Column (6) is similar to column (5), except that we use shipments in lieu of TFP as the dependent variable. As is shown, all interaction terms are statistically insignificant, suggesting that our results are not driven by trade with (or demand from) the MDP.

3.3 Investment in Productivity

Competition with the MDP in the labor or product market may induce incumbent firms to invest in productivity. Other plants of these firms, in distant counties, may also benefit. Under this alternative channel, entry by the MDP would still have a causal effect on plant-level productivity. However, the effect would not be "external" to the plants (as in the case of knowledge spillovers) but rather "internal," in the sense that it would be the outcome of within-firm decisions, e.g., to invest in R&D.

Table 9 considers the issue of internal versus external effects. Our first set of tests is based on the premise that, if the effect was internal, investment responses should be heterogeneous: SC plants, smaller plants (or plants of smaller firms), and plants of financially constrained firms should invest less and experience smaller productivity gains as a result. In columns (1) and (2), we interact both terms in equation (1) with an MC dummy and plant size (number of employees), respectively. (Using firm size yields similar results.) In columns (3) and (4), we merge the local spillover sample with Compustat and interact both terms in equation (1) with measures of firms' financial constraints: the KZ-index of Kaplan and Zingales (1997) and the SA-index of Hadlock and Pierce (2010). All interaction terms are statistically insignificant, which is consistent with the effect being external.[19] Our second set of tests focuses on R&D (column (5)) and innovation (column (6)) at the firm level. An increase in either of those margins would be consistent with the effect being internal. In column (5), we merge the local spillover sample with Compustat; in column (6), we merge it with both Compustat and the USPTO patent database. As is shown, there is no statistically significant effect of the MDP opening on firm-level R&D or patenting activity, which is again consistent with the effect being external.[20]

4 Theoretical Framework

We develop a quantitative spatial model to quantify the impact of knowledge sharing through plant-level networks on sub-regional, regional, and aggregate outcomes. In our model, heterogeneous plants belonging to different sectors produce differentiated goods. Within a region and sector, plants differ in their productivities and parent firms. Plants can be either stand-alone ("single-county plants" or "SC plants") or belong to parent firms with plants in other locations ("multi-county plants" or "MC plants"). Plants' productivities depend on local knowledge. MC plants' productivities additionally depend on knowledge in the other regions in which the parent firm has plants; this generates direct productivity linkages across regions. Despite this complexity, our model admits a simple representation whereby, within a region and sector, plants' productivities aggregate into a single productivity index. Across regions and sectors, this nests our model into an augmented version of the multi-region, multi-sector model of Caliendo and Parro (2015).

4.1 Primitives

Our model economy consists of N heterogeneous regions ("locations" or "counties") which interact through trade in goods markets and labor mobility. Locations, denoted by $i, n \in \mathcal{N}$, exogenously differ from one another with regard to land supply, amenities, and the spatial

[19]Consistent with the interaction term on plant size being statistically insignificant, the weighted and unweighted coefficient estimates in Table 2 are similar and not statistically different from each other.

[20]Merging the local spillover sample with Compustat—or Compustat and the USPTO patent database—reduces the number of sample firms. One concern is that the local spillover effect is insignificant in these smaller samples due to lack of statistical power. This is not the case. When we estimate the local spillover regression on the plant-year samples that correspond to the firm-year samples in columns (5) and (6), we obtain coefficients of 0.040 (standard error: 0.019) and 0.039 (standard error: 0.017), respectively.

allocation of intermediate goods producers ("plants"). Plants are organized into J networks, denoted by j, k, which we call firms. J^{SC} firms consist of a single plant ("single-county firms" or "SC firms"); J^{MC} firms have plants in multiple counties ("multi-county firms" or "MC firms"). Each county has at least one plant, and each MC firm has at most one plant per county. Each plant belongs to one of S sectors, denoted by $s, t \in \mathcal{S}$. Denote the set of locations with one or more plants in sector s as $\mathcal{N}_s$, the set of sectors with one or more plants in location n by $\mathcal{S}_n$, the set of locations where firm j has a plant by $\mathcal{E}_j$, the set of plants in location i by $sj \in \mathscr{E}_i$, and the set of plants in location i and sector s by $j \in \mathscr{E}_{is}$. We refer to individual plants by their location-sector-firm tuple.[21]

4.2 Consumer Preferences

Workers are mobile and endowed with one unit of labor each that is inelastically supplied. Worker v working for plant $\{n, s, j\}$ earns wage w_{nsj} and derives utility from goods consumption (C_v), land use (h_v), and plant-level idiosyncratic amenities (b_{nsjv}):

$$u_{nsjv} = b_{nsjv} C_v^{\alpha} h_v^{1-\alpha}, \tag{2}$$

where $\alpha \in (0, 1)$ and b_{nsjv} is drawn from a multivariate Fréchet distribution given by:

$$\mathbb{P}\left(\bigcap_{n \in \mathcal{N}} \bigcap_{sj \in \mathscr{E}_n} \left\{b_{nsj} \leq t_{nsj,}\right\}\right) = exp\left\{-\sum_{n \in \mathcal{N}} \left(\sum_{sj \in \mathscr{E}_n} (B_n B_s)^{\frac{1}{1-\rho}} t_{nsj}^{-\frac{\epsilon}{1-\rho}}\right)^{1-\rho}\right\}, \tag{3}$$

for all $\left\{t_{nsj}\right\}_{n \in \mathcal{N}: sj \in \mathscr{E}_n} \in [0, \infty)^{\sum_{n \in \mathcal{N}} |\mathscr{E}_n|}$. The amenity scale parameters B_n and B_s index the average draws of idiosyncratic utility for plants in location n and sector s, respectively. The amenity shape parameter $\epsilon > 1$ controls the dispersion in idiosyncratic draws across locations. The amenity correlation parameter $\rho \in [0, 1)$ controls the strength of the correlation of within-location, across-plant idiosyncratic utility draws.

Consumers have Cobb-Douglas preferences over final goods from each sector:

[21] We take the structure of MC firms' plant-level networks as given, consistent with the evidence in Section 2.6 showing that these networks do not adjust in the response to the MDP openings. Modeling the choice of number, size, and location of a firm's plants in spatial equilibrium requires solving a complex combinatorial choice problem. For recent advances, see, e.g., Arkolakis, Eckert, and Shi (2021), who solve the plant location choice problem when there are positive or negative complementarities between plants, and Oberfield et al. (2023), who solve the limit problem in which the firm chooses a density rather than a discrete set of plants.

$$C_v = \Pi_{s \in \mathcal{S}} c_{vs}^{\kappa_s}, \tag{4}$$

where c_{vs} is the amount of sector s's final good consumed by consumer v.

4.3 Production Technology

Plants produce intermediate goods with technology $q_{isj} = z_{isj} {l_{isj}}^{\gamma_s} {m_{isj}}^{1-\gamma_s}$, where z_{isj}, l_{isj}, and m_{isj} denote plant $\{i, s, j\}$'s productivity, labor, and materials, respectively. Materials are a Cobb-Douglas aggregator of sectors' final goods, $m_{isj} = \Pi_{t \in \mathcal{S}} q_{isjt}^{\delta_{st}}$, where q_{isjt} is the total amount of sector t's final good used in production by plant $\{i, s, j\}$ and $\{\delta_{st}\}_{\{s,t\} \in \mathcal{S}^2}$ are input-output weights with $\sum_{t \in \mathcal{S}} \delta_{st} = 1 \; \forall s \in \mathcal{S}$. We assume that plants are monopolistically competitive in goods markets—that is, they do not internalize their impact on other plants' prices nor their own impact on local price indices—and take all factor prices as given, which implies that they set a constant net markup of $\mu_{isj} = \frac{1}{\omega - 1}$ over marginal cost.[22,23]

Intermediate goods are tradable. Goods trade is subject to bilateral "iceberg" trade costs such that $\tau_{ni} \geq 1$ units must be shipped from location i in order for one unit to arrive in location n. In our structural estimation, we parameterize trade costs as a constant elasticity function of distance: $\tau_{ni} = \tau_{in} = dist_{ni}^{\psi}$.

Final goods are nontradable. Each location has a representative final goods producer for each sector. The final goods producer in sector s uses intermediate goods from all plants in sector s as inputs into a nested CES production technology:

$$q_{nis} = \left(\sum_{j \in \mathscr{E}_{is}} q_{nisj}^{\frac{\omega-1}{\omega}} \right)^{\frac{\omega}{\omega-1}} \tag{5}$$

[22]Under our modeling assumptions, plants to do not internalize their impact on either goods prices or wages. In principle, we could allow for a limited degree of internalization. For example, we could allow plants to internalize their impact on local price indices. While this adds complexity to the model, it would not materially affect the estimation or counterfactuals. Likewise, we could allow for a limited degree of labor market power under assumptions that maintain the tractability of Cobb-Douglas production by generating constant wage markdowns below the marginal product of labor. While introducing such constant markdowns complicates all expressions involving Cobb-Douglas production, they ultimately drop out when γ_s is calibrated to match plants' expenditure shares.

[23]We assume that plants rebate profits to consumers through a proportional dividend d such that a consumer working at plant $\{i, s, j\}$ has total income $w_{isj}(1 + d)$ and $d = \frac{\sum_{n \in \mathcal{N}} \sum_{sj \in \mathscr{E}_n} \mu_{nsj} \left(\frac{w_{nsj} l_{nsj}}{\gamma_s} \right)}{\sum_{n \in \mathcal{N}} \sum_{sj \in \mathscr{E}_n} w_{nsj} l_{nsj}}$. While the dividend appears in some formulas, such as aggregate welfare, this formulation ensures that it does not affect labor supply.

and

$$q_{ns} = \left(\sum_{i \in \mathcal{N}_s} q_{nis}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}, \tag{6}$$

where q_{nisj} is the amount of plant $\{i, s, j\}$'s good used by the final goods producer in sector s and location n, and q_{ns} is the total output of sector s's final good in location n.

4.4 Knowledge and Productivity

We assume plants' productivities depend on knowledge and endogenous agglomeration economies that depend on local population size L_i:

$$z_{isj} = \tilde{z}_{isj} k_{isj} L_i^{\beta}, \tag{7}$$

where k_{isj} is plant-specific knowledge, $\tilde{z}_{isj}$ is plant-specific fundamental productivity, and L_i^{β} represents classical, local agglomeration economies.[24] We assume plants learn from their neighbors, through local knowledge K_i, as well as from other plants in their firm networks. Specifically, and building on Markusen (1984), we assume plants draw on firm-wide shared knowledge (or "knowledge capital") K_j. Plants' knowledge is given by:

$$k_{isj} = K_i^{1-\theta} K_j^{\theta}, \tag{8}$$

where local knowledge K_i is an aggregator of plants' knowledge in location i, $K_i = \left(\sum_{sj \in \mathscr{E}_i} k_{isj}\right)^{\zeta}$, and firm-wide knowledge K_j depends on local knowledge in all of the firm's locations, $K_j = \Pi_{i \in \mathcal{E}_j} K_i$.[25] Note that inserting $K_j = \Pi_{i \in \mathcal{E}_j} K_i$ in equation (8) yields $k_{isj} = K_i K_{j,-i}^{\theta}$, where $K_{j,-i}^{\theta} \equiv \Pi_{n \in \mathcal{E}_j \setminus \{i\}} K_n$ is an aggregator of knowledge in the firm's other locations. For example, if firm j has plants in locations 1 and 2, then plant $\{1, s, j\}$'s knowledge is given by $k_{1sj} = K_1 K_2^{\theta}$. Crucially, this implies that setting θ equal to zero in our counterfactuals does not affect the weight on *local* knowledge in the plant's knowledge technology. Also, our knowledge technology implies that the impact of a local knowledge shock on firm-wide knowledge is invariant to how many plants the firm has; it is motivated

[24] Plants' output q_{isj} is thus a function of labor and knowledge, as in Caliendo and Rossi-Hansberg (2012) and Caliendo, Monte, and Rossi-Hansberg (2015).

[25] Table A.4 of Online Appendix A shows that, in our reduced-form setting, both the local and global productivity spillover increase with MDP size, consistent with our modeling assumption of knowledge values being additive in the local knowledge index, $\sum_{sj \in \mathscr{E}_i} k_{isj}$.

by the fact that in our reduced-form global spillover analysis the interaction term between the MDP dummy and the number of the firm's plants is insignificant.[26] Finally, for parsimony, our knowledge technology abstracts from certain aspects of heterogeneity documented in Section 3. Specifically, Table 7 shows that the global spillover varies at the industry-pair level based on whether plants are in the same industry or in industries linked through mutual R&D flows or mutual patent citations. Table A.3 of Online Appendix A shows similar results for the local spillover. While such heterogeneity across industry pairs may be important for certain types of counterfactuals, it is unlikely to have a material impact on the counterfactuals undertaken in Section 6.

For SC plants, equation (7) reduces to $z_{isj} = \tilde{z}_{isj} K_i L_i^{\beta}$. In counties with only SC plants, our productivity process is therefore similar to that in standard models with local agglomeration economies. By contrast, counties with MC plants are connected to other counties through a knowledge-sharing network allowing for direct productivity spillovers across locations. The key parameter which controls the strength of within-firm, across-location spillovers is the global knowledge-sharing parameter θ. Ceteris paribus, a higher value of θ puts more "weight" on knowledge in the firm's other locations and thus generates larger global spillovers in response to local knowledge shocks.[27] The parameter ζ, on the other hand, controls the strength of local knowledge spillovers *across* firms; it is critical in pinning down local plants' productivity responses to entry by the MDP.

4.5 Solution to Consumer Problem

4.5.1 Goods and Housing

Consider consumer v in location n. Standard Cobb-Douglas demand results imply that $c_{vs} p_{ns} = \kappa_s x_v$, where c_{vs} is the amount of sector s's final good consumed by consumer v, x_v is the amount consumer v spends on goods, and p_{ns} is the price of sector s's final good in location n. It follows that the local consumption price index is given by $P_n = \Pi_{s \in \mathcal{S}} p_{ns}^{\kappa_s}$.

We assume that land is inelastically supplied and owned by immobile landlords who receive rents R_n from workers as income and consume their local consumption bundle (see

[26] In our global spillover regression (column (3) of Table 3), interacting the MDP dummy with the log number of the treated firm's plants yields a coefficient of 0.020 (0.008) for the MDP dummy and −0.002 (0.004) for the interaction term (standard errors in parentheses).

[27] Our knowledge technology admits higher-order spillovers: local knowledge shocks spread to distant plants through their firms' networks (through K_j); from there, the shocks continue to spread to the (distant) plants' local neighbors (through K_i), and so on.

Monte, Redding, and Rossi-Hansberg, 2018). Equation (2) implies that consumers spend a fraction $(1-\alpha)$ of their income on land.[28] Thus, total expenditure on land is the product of average wages W_n, local population L_n, and the gross dividend $(1+d)$. Land market clearing implies that equilibrium land rents are given by:

$$R_n = (1-\alpha)\frac{W_n L_n (1+d)}{H_n}. \tag{9}$$

4.5.2 Location Choice and Welfare

Each worker chooses a plant that maximizes her utility. Worker v's indirect utility from working for plant $\{n, s, j\}$ is given by $b_{nsjv}\frac{w_{nsj}(1+d)}{P_n^{\alpha} R_n^{1-\alpha}}$. In the spirit of McFadden (1978), worker v faces a nested choice; we can decompose this choice problem into a choice of location-sector pair ns and, within a given location-sector pair, a choice of plant $j \in \mathscr{E}_{ns}$. Applying results from Lind and Ramondo (2023), we show in Online Appendix B.1 that the labor share of location-sector pair ns is given by:

$$\frac{L_{ns}}{\bar{L}} = \frac{B_n B_s^{\frac{1}{1-\rho}} \left(\frac{\left(W_{ns}^b\right)^{\frac{1}{1-\rho}} \left(W_n^b\right)^{-\frac{\rho}{1-\rho}}}{P_n^{\alpha} R_n^{1-\alpha}} \right)^{\epsilon}}{\sum_{i\in\mathcal{N}} B_i \left(\frac{W_i^b}{P_i^{\alpha} R_i^{1-\alpha}} \right)^{\epsilon}}, \tag{10}$$

where the "amenity wages" W_n^b and W_{ns}^b are aggregators of plant-level wages:

$$W_n^b := \left(\sum_{s\in\mathcal{S}_n} B_s^{\frac{1}{1-\rho}} \left(W_{ns}^b\right)^{\frac{\epsilon}{1-\rho}} \right)^{\frac{1-\rho}{\epsilon}} \tag{11}$$

and

$$W_{ns}^b := \left(\sum_{j\in\mathscr{E}_{ns}} w_{nsj}^{\frac{\epsilon}{1-\rho}} \right)^{\frac{1-\rho}{\epsilon}}. \tag{12}$$

We show in Online Appendix B.1 that within-location-sector (supply-side) labor shares are given by:

[28] See Davis and Ortalo-Magné (2011) for evidence in support of the constant housing expenditure share implied by the Cobb-Douglas representation in equation (2).

$$\frac{l^S_{nsj}}{L_{ns}} = \frac{w_{nsj}^{\frac{\epsilon}{1-\rho}}}{\sum_{k\in\mathscr{E}_{ns}} w_{nsk}^{\frac{\epsilon}{1-\rho}}}, \tag{13}$$

where l^S_{nsj} represents labor supplied to plant $\{n, s, j\}$ given plant-level wages $\{w_{nsk}\}_{k\in\mathscr{E}_{ns}}$. Plants face an upward-sloping labor supply curve. To attract additional workers with lower idiosyncratic preference draws, real wages $w_{nsj}/(P_n^\alpha R_n^{1-\alpha})$ must increase. When plant $\{n, s, j\}$'s real wage increases, if $\rho > 0$, it attracts workers from both within location n and other locations. As $\rho \to 1$, within-location preferences become perfectly correlated.

Finally, we show in Online Appendix B.1 that average realized utility (or welfare) is given by:

$$\bar{U} = (1+d)\,\Gamma\!\left(\frac{\epsilon-1}{\epsilon}\right)\left[\sum_{n\in\mathcal{N}} B_n \left(\frac{W_n^b}{P_n^\alpha R_n^{1-\alpha}}\right)^{\epsilon}\right]^{\frac{1}{\epsilon}}, \tag{14}$$

where $\Gamma(\cdot)$ denotes the gamma function.

4.6 Solution to Producer Problem

4.6.1 Plant Production

We can decompose plant-level production decisions into within- and across-location-sector production. We first show in Online Appendix B.1 that within-location labor demand in sector s satisfies:

$$\frac{l^D_{isj}}{L_{is}} = \frac{z_{isj}^{\omega-1} w_{isj}^{\gamma_s(1-\omega)-1}}{\sum_{k\in\mathscr{E}_{is}} z_{isk}^{\omega-1} w_{isk}^{\gamma_s(1-\omega)-1}}, \tag{15}$$

where L_{is} is total employment in location i and sector s and l^D_{isj} is total labor demand by plant $\{i, s, j\}$.[29]

We next characterize across-location-sector production. Let $\tilde{p}_{is}$ be the Free on Board price of one unit of the cost-minimizing bundle of intermediate goods from plants in

[29]Equation (15) implies that labor demand is strictly greater than zero, which ensures that all plants produce in equilibrium. This differs from models of granular firms in international trade (e.g., Eaton, Kortum, and Sotelo, 2012; Gaubert and Itskhoki, 2021), where only a subset of firms produces in equilibrium. In our (domestic) setting, it is crucial that all plants produce in equilibrium in order to mimick the economic footprints of plant-level networks in the Census data.

location i and sector s. Moreover, define the location-sector "productivity index":[30]

$$\mathcal{MC}_{is} := \frac{\omega}{\omega - 1}\left(\sum_{j\in\mathscr{E}_{is}}\left(\frac{w_{isj}^{\gamma_s}}{z_{isj}W_{is}^{\gamma_s}}\right)^{1-\omega}\right)^{\frac{1}{1-\omega}}. \tag{16}$$

We show in Online Appendix B.1 that, in equilibrium, the Free on Board Price satisfies:

$$\tilde{p}_{is} = \mathcal{MC}_{is}W_{is}^{\gamma_s}\left(p_{is}^{m}\right)^{1-\gamma_s}, \tag{17}$$

where $p_{is}^{m} = \Pi_{t\in S}p_{it}^{\delta_{st}}$ is the materials price index.

4.6.2 Final Goods Production

Local aggregate demand for final goods consists of demand from local consumers, local landlords, and local plants. Cobb-Douglas production and consumption imply that local aggregate demand for final goods satisfies:

$$q_{ns}p_{ns} = \sum_{jt\in\mathscr{E}_n} w_{ntj}l_{ntj}\left(\kappa_s\left(1+d\right) + \delta_{ts}\frac{1-\gamma_t}{\gamma_t}\right). \tag{18}$$

Also, CES demand results imply that the final goods price index is given by:

$$p_{ns} = \left(\sum_{i\in\mathcal{N}_s}\left(\tau_{ni}\tilde{p}_{is}\right)^{1-\eta}\right)^{\frac{1}{1-\eta}}. \tag{19}$$

Moreover, they imply that final goods producers in location n have expenditure shares:

$$\frac{X_{nis}}{X_{ns}} = \left(\frac{\tilde{p}_{is}\tau_{ni}}{p_{ns}}\right)^{1-\eta}, \tag{20}$$

where X_{ns} is total expenditure and X_{nis} is expenditure on intermediate goods from location i. Combining equations (18) and (20) with Cobb-Douglas production, imposing that plant-level revenue equals plant-level income from sales, substituting in markups, and summing over all plants in sector s and region n gives:

$$\frac{W_{ns}l_{ns}}{\gamma_s}\frac{\omega}{\omega-1} = \sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{S}_i} W_{it}L_{it}\left(\frac{\tilde{p}_{is}\tau_{ni}}{p_{ns}}\right)^{1-\eta}\left(\kappa_s\left(1+d\right) + \delta_{ts}\frac{1-\gamma_t}{\gamma_t}\right). \tag{21}$$

[30] While the formula for $\mathcal{MC}_{is}$ includes relatives wages, those are, in equilibrium, solely a function of productivity; hence the name "productivity index."

4.7 General Equilibrium

Define the following endogenous objects: plant-level knowledge $\boldsymbol{k} := \left\{k_{nsj}\right\}_{n\in\mathcal{N}:js\in\mathscr{E}_n}$, within-location-sector labor shares $\boldsymbol{l} := \left\{\frac{l_{nsj}}{L_{ns}}\right\}_{n\in\mathcal{N}:s\in\mathcal{S}_n:j\in\mathscr{E}_{ns}}$, within-location-sector relative wages $\boldsymbol{w} := \left\{\frac{w_{nsj}}{W_{ns}}\right\}_{n\in\mathcal{N}:s\in\mathcal{S}_n:j\in\mathscr{E}_{ns}}$, county-sector level labor $\boldsymbol{L} := \{L_{ns}\}_{n\in\mathcal{N}:s\in\mathcal{S}_n}$, county-sector level average wages $\boldsymbol{W} := \{W_{ns}\}_{n\in\mathcal{N}:s\in\mathcal{S}_n}$, and county-level average wages $\boldsymbol{W}_n := \{W_n\}_{n\in\mathcal{N}}$. Moreover, define the following exogenous fundamentals: land endowments $\boldsymbol{H} := \{H_n\}_{n\in\mathcal{N}}$, amenity scale parameters $\boldsymbol{B} := \left\{\{B_n\}_{n\in\mathcal{N}}, \{B_s\}_{s\in\mathcal{S}}\right\}$, plant-level fundamental productivity $\boldsymbol{z} := \left\{\tilde{z}_{isj}\right\}_{i\in\mathcal{N}:js\in\mathscr{E}_i}$, plant-level networks $\boldsymbol{\mathscr{E}} := \{\mathscr{E}_i\}_{i\in\mathcal{N}}$, and bilateral trade costs $\boldsymbol{\tau} := \{\tau_{ni}\}_{\{n,i\}\in\mathcal{N}^2}$. Finally, define the set of "sectoral parameters" $\Omega := \left\{\{\kappa_s, \gamma_s\}_{s\in\mathcal{S}}, \{\delta_{st}\}_{\{s,t\}\in\mathcal{S}^2}\right\}$.

We are ready to characterize equilibria of our model. Equilibrium knowledge, labor allocations, and wages, $\{\boldsymbol{k}, \boldsymbol{l}, \boldsymbol{w}, \boldsymbol{L}, \boldsymbol{W}\}$, are pinned down by equation (8) (knowledge equilibrium condition), (13) and (15) (within-location-sector equilibrium conditions), and (10) and (21) (across-location-sector equilibrium conditions).[31] Further, if $\{\boldsymbol{k}, \boldsymbol{l}, \boldsymbol{w}, \boldsymbol{L}, \boldsymbol{W}\}$ solves all five equations, there exists a unique equilibrium with those knowledge values, labor allocations, and wages and, given fundamentals $\{\boldsymbol{H}, \boldsymbol{B}, \boldsymbol{z}, \boldsymbol{\mathscr{E}}, \boldsymbol{\tau}\}$, we can recover the equilibrium values of all endogenous objects in closed form. We thus refer to equilibria by their corresponding knowledge values, labor allocations, and wages, $\{\boldsymbol{k}, \boldsymbol{l}, \boldsymbol{w}, \boldsymbol{L}, \boldsymbol{W}\}$.

Proposition 1. *Given parameter values $\{\alpha, \beta, \epsilon, \zeta, \eta, \theta, \rho, \omega, \Omega\}$ and fundamentals $\{\boldsymbol{H}, \boldsymbol{B}, \boldsymbol{z}, \boldsymbol{\mathscr{E}}, \boldsymbol{\tau}\}$, and given parametric restictions, there exists a unique vector $\{\boldsymbol{k}, \boldsymbol{l}, \boldsymbol{w}\}$ which is consistent with an equilibrium of the model.*

Proof. See Online Appendix B.2.[32] □

Using Proposition (1), we show in Online Appendix B.2 that plant-level knowledge and fundamental productivity, $\{\boldsymbol{k}, \boldsymbol{z}\}$, uniquely aggregate into a location-sector productivity index that enters into the across-location-sector equilibrium conditions and thereby pins down $\{\boldsymbol{L}, \boldsymbol{W}\}$. Furthermore, we show in Online Appendix B.2 that the

[31]The knowledge equilibrium condition (8) is pinned down by firms' plant-level networks, rather than by firms' production choices. That said, for expositional simplicity, we include the knowledge equilibrium condition in the full set of model equilibrium conditions rather than defining a separate condition for a knowledge fixed point.

[32]We use results from Allen, Arkolakis, and Li (2022) to derive the parametric conditions in Proposition 1. These conditions hold in the model economy in Section 5 under the estimated parameter values.

across-location-sector equilibrium conditions are isomorphic to corresponding conditions in a version of Caliendo and Parro (2015) augmented with mobile labor across regions, idiosyncratic preferences over locations and sectors, a land market, and classical, local agglomeration economies.

4.8 Model Inversion

The following proposition shows that the model can be inverted to recover unique (up to a normalization) values of $\{\boldsymbol{B}, \boldsymbol{z}\}$ and $\{\boldsymbol{k}, \boldsymbol{w}, \boldsymbol{W}\}$ that are consistent with the observed distribution of economic activity.

Proposition 2. *Given parameter values* $\{\alpha, \beta, \epsilon, \zeta, \eta, \theta, \rho, \omega, \Omega\}$, *fundamentals* $\{\boldsymbol{H}, \boldsymbol{\mathscr{E}}, \boldsymbol{\tau}\}$, *and observed data* $\{\boldsymbol{l}, \boldsymbol{L}, \boldsymbol{W}_n\}$, *and given parametric restrictions, there exist unique (up to a normalization) unobserved fundamentals* $\{\boldsymbol{B}, \boldsymbol{z}\}$ *and plant-level knowledge and wages* $\{\boldsymbol{k}, \boldsymbol{w}, \boldsymbol{W}\}$ *that rationalize the data as an equilibrium of the model.*

Proof. See Online Appendix B.2.[33] □

5 Structural Estimation

5.1 Model Economy

We simulate an economy with a large number of locations, plants, and firms that mirrors the geography of production networks in U.S. Census data. Each location corresponds to a specific county. We assign each county its actual geographical coordinates, land area, manufacturing employment and wages, and number of plants in each sector using 1982 Census data.[34] Sectors are based on 2-digit SIC code manufacturing industries. We assign plants to firms based on the joint distribution of firm size and spatial dispersion in firms' plant-level networks. We provide more details in Online Appendix C.1.

5.2 Parameters and Identification

We divide the parameters into two sets. The first set consists of parameters which we calibrate using additional data and values from the literature. We set the expenditure

[33] The parametric restrictions in Proposition 2 are the same as in Proposition 1.

[34] We use data from 1982 to ensure that the MDPs are not included in the baseline economy.

share on housing, $1 - \alpha$, equal to 0.3 to match the BLS Consumer Expenditure Survey, the elasticity of trade costs to distance, ψ, equal to $1.29/(\eta - 1)$ to match the elasticity of trade flows to distance in Monte, Redding, and Rossi-Hansberg (2018), and the elasticity of local firm productivity to agglomeration, β, equal to 0.023 (Gaubert, 2018).[35] Finally, we set the sectoral parameters Ω equal to consumption and input-output weights in the BEA Input-Output Accounts data and value-added shares from the NBER CES Manufacturing Industry database.

The second set of parameters, $\{\zeta, \theta, \eta, \epsilon, \omega, \rho\}$, consists of parameters which we estimate by targeting all six central moments from our reduced-form analysis: semi-elasticities of plant-level employment, wages, and productivity with respect to the MDP openings, both for local and distant plants. To obtain model-based estimates that correspond to these reduced-form estimates, we simulate the 47 MDP openings in our model economy and compute the new equilibrium. This procedure provides us with a plant-level data set consisting of "pre-" and "post-MDP opening" observations, allowing us to estimate plant-level difference-in-differences regressions that mirror those in our reduced-form analysis. We provide further details in Online Appendix C.2.

There is a tight link between the structural parameters and the economic forces that govern the productivity, employment, and wage responses to the MDP openings in our model. First, the local knowledge sharing parameter ζ governs the magnitude of the local productivity spillover; given the local spillover, the global knowledge sharing parameter θ pins down the value of the global productivity spillover. Second, the parameters η and ϵ control the across-region-sector labor demand and supply elasticities, respectively. Given the magnitudes of the productivity spillovers, η and ϵ thus primarily influence the local employment and wage responses. Finally, the parameters ω and ρ pin down the within-region-sector labor demand and supply elasticities, respectively; in our model, this implies ω and ρ jointly control the global employment and wage responses.

5.3 Estimation

Our estimation targets all six central moments from our reduced-form analysis: the local productivity response (Table 2, column (1)), the local employment and wage responses

[35] Given this value of beta, the inclusion of classical, local agglomeration economies in our model is largely immaterial. Without those, our model can still match all six reduced-form moments exactly with only minimal changes to the parameter values.

(Table 4, columns (1) and (2)), the global productivity response (Table 3, column (3)), and the global employment and wage responses (Table 4, columns (3) and (4)). As our model is just identified, the difference-in-differences regression coefficients from our estimated model exactly match the corresponding reduced-form coefficients.

Our estimated parameter values are informative about key economic forces in our model. As for the local and global knowledge sharing parameters, we obtain estimates of $\hat{\zeta} = 0.046$ (0.012) and $\hat{\theta} = 0.63$ (0.20), respectively (GMM standard errors in parentheses). Thus, we can comfortably reject the null of no inter- or intra-firm knowledge sharing. As for the parameters controlling across-region-sector labor demand and supply, our estimates are $\hat{\eta} = 3.3$ (0.76) and $\hat{\epsilon} = 3.4$ (1.4), respectively. Hence, we cannot reject the null that $\eta = 4$ (e.g., Bernard et al., 2003; Broda and Weinstein, 2006; Redding, 2016) or $\epsilon = 3$ (e.g., Redding, 2016; Bryan and Morten, 2019). Finally, as for the parameters controlling within-region-sector labor demand and supply, we obtain estimates of $\hat{\omega} = 2.1$ (0.40), indicating relatively inelastic labor demand, and $\hat{\rho} = 0.57$ (0.35), indicating highly, albeit not perfectly, elastic labor supply.

6 Counterfactuals

6.1 Propagation Forces

In our first counterfactual, we quantify the aggregate effects of the 47 MDP openings under various assumptions about the underlying propagation forces. There are two main propagation forces in our model: input-output linkages and within-firm, across-location ("global") knowledge sharing.[36] In our model, we can turn input-output forces off by setting the parameter vector γ_s to one and turn global knowledge sharing off by setting the parameter θ to zero. In this first counterfactual, we either turn both forces off, turn only input-output forces on, turn only global knowledge sharing on, or turn both forces on. We hold all other parameters at their estimated or calibrated values and follow the steps in our estimation procedure described in Section 5.

While MDP openings constitute large regional shocks, they constitute relatively small shocks at the aggregate level. When both propagation forces are turned off,

[36]While input-output linkages are not the main source of the local productivity spillover (see Section 3), they can amplify the local productivity spillover and thereby have significant aggregate effects.

aggregate welfare increases by 0.0038%. When only global knowledge sharing or only input-output forces are turned on, the welfare gains increase by a factor of 2.35 and 2.96, respectively. Hence, global knowledge sharing and input-output linkages have roughly similar amplification effects. However, the two forces also interact in meaningful ways. When both forces are turned on, the welfare gains increase by a factor of 6.94, which is almost double the two marginal effects combined (594% vs. 135% + 196%). Intuitively, input-output forces amplify the welfare gains from increases in productivity, including productivity increases due to global spillovers.

6.2 Plant Openings and Regional Development

Opening a large industrial plant can have a significant impact on a region, especially for smaller and less developed regions. It can boost employment, raise productivity, and spur industrial activity. In extreme cases, it can help a lagging economy escape a "poverty trap" equilibrium (Kline, 2010). However, in the data, winner counties are seldom small or underdeveloped: compared to the rest of the economy (but not compared to the runner-up counties), they have higher incomes and income growth, higher population and population growth, and a higher share of labor in manufacturing (GHM, 2010). Indeed, 40 of the 47 MDP counties in our data are either in the highest or second highest population quintile.

If industrial plants tend to locate in regions that are already well developed, the question is whether the government should intervene to aid less developed regions.[37] To inform this policy debate, we randomly assign MDP openings to more or less developed regions and study their local impact as well as their impact on the rest of the economy. We always use the same MDP; it is a representative MDP based on plant size and industry from the set of 47 MDP openings. We proxy for regional development using population size. Specifically, we sort counties into population quintiles and assign the MDP to a 10% random sample of counties from each quintile. Sorting by population divides counties into rural vs. urban areas. Also, population is highly correlated with income, manufacturing employment, and other measures of regional development. To ensure that our results are not driven by outlier counties with excessively high MDP employment shares, we require that the MDP's county-industry employment share lies within the 95th percentile of its empirical distribution based on the 47 MDP counties. This eliminates extremely rural

[37] One example of a policy proposal towards this goal consists of national government interventions in local governments' subsidies to firms; see Slattery and Zidar (2020) for a review of local subsidy policies.

counties with no, or hardly any, pre-existing employment in the MDP's industry. As in our first counterfactual, we turn global knowledge sharing on and off; otherwise, we follow the estimation procedure described in Section 5.

Figure 2 shows the impact of an MDP opening on real value added (VA), defined as the (inflation-adjusted) value of output minus the cost of materials.[38] In Panel A, the red bars show the local impact on plants in the MDP county (excluding the MDP itself); the blue bars show the impact on plants in the rest of the economy.[39] A light (dark) color indicates that global knowledge sharing is turned off (on). As the MDP opening has a stronger relative impact on the MDP county than on the rest of the economy, we use separate Y-axes for the MDP county (left) and the rest of the economy (right). To study the role of regional development, we show results separately for each population quintile. In the lowest quintile, the MDP opens in a less developed region; in the highest quintile, it opens in a well-developed region.

As is shown in Panel A, the local impact of the MDP opening is declining in the level of regional development of the MDP county. In our model, less developed regions have a lower stock of knowledge (recovered from the data), so the relative gain in productivity, and ultimately in real VA, at local incumbent plants is higher in these regions. This is true regardless of whether global knowledge sharing is turned on. In fact, it makes little difference if global knowledge sharing is turned on: the general equilibrium (feedback) effect of global knowledge sharing on the MDP county is small, which is why the light and dark red bars are practically identical.

In stark contrast, turning on global knowledge sharing makes a big difference for the impact of the MDP opening on the rest of the economy. When global knowledge sharing is turned off, the MDP opening causes a decline in real VA in the rest of the economy; this is the familiar result that local investment policies can have negative effects on other regions. By contrast, when global knowledge sharing is turned on, real VA in the rest of the economy *increases*: the MDP opening now (also) raises the productivity of plants in distant

[38]As is common in spatial models with labor mobility and Fréchet preferences over locations, in our model, ex-ante welfare—that is, worker-level expected utility prior to the realization of idiosyncratic Fréchet draws—is equalized across regions in equilibrium. For this reason, we focus on real VA as a meaningful and policy-relevant measure of the heterogeneous regional effects of large plant openings.

[39]The local impact captures spillovers from the MDP opening on incumbent plants as well as any general equilibrium effects on those plants. We exclude the MDP itself from the local impact because it scales with the size of the local economy in a way that is mechanical: adding an MDP of a given size to a "small" county (in terms of real VA) has a larger relative effect than does adding the same MDP to a "large" county.

regions, which are connected to the MDP county through plant-level (knowledge-sharing) networks. Moreover, the gains in the rest of the economy are increasing in the level of regional development of the MDP county: more developed regions have more MC plants and thus more plant-level network connections with other regions.[40]

Panel A of Figure 2 illustrates the ambiguous role of regional development. On one hand, the impact of the MDP opening on the local economy is stronger when the MDP county is *less* developed. On the other hand, with global knowledge sharing turned on, the positive effect on the rest of the economy is stronger when the MDP county is *well* developed. Panel B shows the aggregate effect on the MDP county (excluding the MDP itself) and the rest of the economy combined when global knowledge sharing is turned on. As can be seen, the pattern is similar to that in Panel A for the rest of the economy. Intuitively, the MDP county is small relative to the rest of the economy; the impact on the latter therefore dominates. Precisely, if the MDP opens in a less developed region (lowest quintile), aggregate real VA increases by 0.012%. In contrast, if the MDP opens in a well-developed region (highest quintile), aggregate real VA increases by 0.039%. Thus, the aggregate gains are greatest if the MDP opens in a well-developed region—which is connected to other regions through plant-level (knowledge-sharing) networks—consistent with the observed location choices of the MDPs in the data.

7 Conclusion

The gains from agglomeration economies are thought to be highly localized. In this paper, we show that local productivity spillovers can propagate through the entire economy through the plant-level networks of multi-region firms. Specifically, building on the empirical framework in GHM, we show that large industrial plant openings not only raise the productivity of local incumbent plants but also of distant plants hundreds of miles away, which belong to large multi-plant, multi-region firms that are exposed to the local productivity spillover through one of their plants. Consistent with a knowledge-sharing channel, this "global" productivity spillover does not decay with distance and is stronger if

[40]Regional development may also correlate with other forces in the model, such as input-output linkages. However, when we turn global knowledge sharing off, all those other forces remain the same, yet the positive effect of regional development disappears (light blue bars). Hence, we conclude that the positive effect of regional development is *because* more developed regions have more MC plant-level network connections, and not because of input-output linkages or other forces in the model.

plants are in industries that share knowledge with each other.

To quantify the significance of firms' plant-level networks for the propagation and amplification of local productivity shocks, we develop and estimate a quantitative spatial model in which plants of multi-region firms are linked through shared knowledge. In our estimated model, input-output linkages and within-firm, across-region ("global") knowledge sharing have quantitatively similar effects. We finally use our estimated model to study the implications of regional development for the impacts of large industrial plant openings. While large industrial plant openings have a greater local impact in less developed regions, the aggregate gains are greatest when the plants locate in well-developed regions, which are connected to other regions through firms' plant-level (knowledge-sharing) networks.

References

Ahlfeldt, Gabriel, Stephen Redding, Daniel Sturm, and Nikolaus Wolf, 2015, The Economics of Density: Evidence from the Berlin Wall, Econometrica 83, 2127–2189.

Allen, Treb, and Costas Arkolakis, 2014, Trade and the Topography of the Spatial Economy, Quarterly Journal of Economics 129, 1085–1140.

Allen, Treb, Costas Arkolakis, and Xiangliang Li, 2022, On the Equilibrium Properties of Network Models with Heterogeneous Agents, mimeo, Dartmouth College.

Arkolakis, Costas, Eckert, Fabian Eckert, and Rowan Shi, 20221, Combinatorial Discrete Choice, mimeo, Yale University.

Audretsch, David, and Maryann Feldman, 2004, Knowledge Spillovers and the Geography of Innovation, in: Handbook of Regional and Urban Economics, Volume 4, Vernon Henderson and Jacques-François Thisse (eds.). Amsterdam: North-Holland.

Barrot, Jean-Noël, and Julien Sauvagnat, 2016, Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks, Quarterly Journal of Economics 131, 1543–1592.

Bartelme, Dominick, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare, 2021, The Textbook Case for Industrial Policy: Theory Meets Data, mimeo, University of Michigan.

Bernard, Andrew, Jonathan Eaton, Bradford Jensen, and Samuel Kortum, 2003, Plants and Productivity in International Trade, American Economic Review 93, 1268–1290.

Borusyak, Kirill, and Peter Hull, 2023, Non-Random Exposure to Exogenous Shocks, Econometrica 91, 2155–2185.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, 2023, Revisiting Event Study Designs: Robust and Efficient Estimation, forthcoming, Review of Economic Studies.

Broda, Christian, and David Weinstein, 2006, Globalization and the Gains from Variety, Quarterly Journal of Economics 121, 541–585.

Bryan, Gharad, and Melanie Morten, 2019, The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia, Journal of Political Economy 127, 2229–2268.

Busso, Matias, Jesse Gregory, and Patrick Kline, 2013, Assessing the Incidence and Efficiency of a Prominent Place Based Policy, American Economic Review 103, 897–947.

Caliendo, Lorenzo, and Esteban Rossi-Hansberg, 2012, The Impact of Trade on Organization and Productivity, Quarterly Journal of Economics 127, 1393–1467.

Caliendo, Lorenzo, and Fernando Parro, 2015, Estimates of the Trade and Welfare Effects of NAFTA, Review of Economic Studies 82, 1–44.

Caliendo, Lorenzo, Ferdinando Monte, and Esteban Rossi-Hansberg, 2015, The Anatomy of French Production Hierarchies, Journal of Political Economy 123, 809–852.

Caliendo, Lorenzo, Fernando Parro, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte, 2018, The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy, Review of Economic Studies 85, 2042–2096.

Carvalho, Vasco, Makoto Nirei, Yukiko Saito, and Alireza Tahbaz-Salehi, 2021, Supply Chain Disruptions: Evidence from the Great East Japan Earthquake, Quarterly Journal of Economics 136, 1255–1321.

Combes, Pierre-Philippe, and Laurent Gobillon, 2015, The Empirics of Agglomeration Economies, in: Handbook of Regional and Urban Economics, Volume 5A, Gilles Duranton, Vernon Henderson, and William Strange (eds.). Amsterdam: North-Holland.

Davis, Morris, and François Ortalo-Magné, 2011, Household Expenditures, Wages, Rents, Review of Economic Dynamics 14, 248–261.

Ding, Xiang, 2023, Industry Linkages from Joint Production, mimeo, Georgetown University.

Duranton, Gilles, and Diego Puga, 2004, Micro-Foundations of Urban Agglomeration Economies, in: Handbook of Regional and Urban Economics, Volume 4, Vernon Henderson and Jacques-François Thisse (eds.). Amsterdam: North-Holland.

Eaton, Jonathan, Samuel Kortum, and Sebastian Sotelo, 2012, International Trade: Linking Micro and Macro, NBER Working Paper 17864.

Ellison, Glenn, Edward Glaeser, and William Kerr, 2010, What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns, American Economic Review 100, 1195–1213.

Gaubert, Cecile, 2018, Firm Sorting and Agglomeration, American Economic Review 108, 3117–3153.

Gaubert, Cecile, and Oleg Itskhoki, 2021, Granular Competitive Advantage, Journal of Political Economy 129, 871–939.

Gaubert, Cecile, Patrick Kline, and Danny Yagan, 2020, Place-Based Redistribution, mimeo, UC Berkeley.

Giroud, Xavier, and Holger Mueller, 2019, Firms' Internal Networks and Local Economic Shocks, American Economic Review 109, 3617–3649.

Glaeser, Edward, and Joshua Gottlieb, 2008, The Economics of Place-Making Policies, Brookings Papers on Economic Activity, Spring, 155–239.

Greenstone, Michael, and Enrico Moretti, 2003, Bidding for Industrial Plants: Does Winning a 'Million Dollar Plant' Increase Welfare? NBER Working Paper 9844.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti, 2010, Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings, Journal of Political Economy 118, 536–598.

Hadlock, Charles, and Joshua Pierce, 2010, New Evidence on Measuring Financial Constraints: Moving Beyond the KZ Index, Review of Financial Studies 23, 1909–1940.

Hornbeck, Richard, and Enrico Moretti, 2022, Estimating Who Benefits from Productivity Growth: Local and Distant Effects of City TFP Shocks on Wages, Rents, and Inequality, forthcoming, Review of Economics and Statistics.

Kaplan, Steven, and Luigi Zingales, 1997, Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints? Quarterly Journal of Economics 112, 169–215.

Kline, Patrick, 2010, Place Based Policies, Heterogeneity, and Agglomeration, American Economic Review: Papers and Proceedings 100, 383–387.

Kline, Patrick, and Enrico Moretti, 2014a, People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs, Annual Review of Economics 6, 629–662.

Kline, Patrick, and Enrico Moretti, 2014b, Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority, Quarterly Journal of Economics 129, 275–331.

Lind, Nelson, and Natalia Ramondo, 2023, Trade with Correlation, American Economic Review 103, 317–353.

Lucas, Robert, and Benjamin Moll, 2014, Knowledge Growth and the Allocation of Time, Journal of Political Economy 122, 1–51.

Markusen, James, 1984, Multinationals, Multi-Plant Economies, and the Gains from Trade, Journal of International Economics 16, 205–226.

Marshall, Alfred, 1890. Principles of Economics. London: Macmillan.

McFadden, Daniel, 1978, Modelling the Choice of Residential Location, in: Spatial Interaction Theory and Planning Models, Anders Karlqvist, Lars Lundqvist, Folke Snickars, and Jörgen Weibull (eds.). Amsterdam: North-Holland.

Mian, Atif, and Amir Sufi, 2014, What Explains the 2007–2009 Drop in Employment? Econometrica 82, 2197–2223.

Monte, Ferdinando, Stephen Redding, and Esteban Rossi-Hansberg, 2018, Commuting, Migration, and Local Employment Elasticities, American Economic Review 108, 3855–3890.

Moretti, Enrico, 2010, Local Labor Markets, in: Handbook of Labor Economics, Volume 4B, Orley Ashenfelter and David Card (eds.). Amsterdam: North-Holland.

Neumark, David, and Helen Simpson, 2015, Place-Based Policies, in: Handbook of Regional and Urban Economics, Volume 5B, Gilles Duranton, Vernon Henderson, and William Strange (eds.). Amsterdam: North-Holland.

Oberfield, Ezra, Esteban Rossi-Hansberg, Pierre-Daniel Sarte, and Nicholas Trachter, 2023, Plants in Space, forthcoming, Journal of Political Economy.

Perla, Jesse, and Christopher Tonetti, 2014, Equilibrium Imitation and Growth, Journal of Political Economy 122, 52–76.

Redding, Stephen, 2016, Goods Trade, Factor Mobility and Welfare, Journal of International Economics 101, 148–167.

Redding, Stephen, and Esteban Rossi-Hansberg, 2017, Quantitative Spatial Economics, Annual Review of Economics 9, 21–58.

Rosenthal, Stuart, and William Strange, 2004, Evidence on the Nature and Sources of Agglomeration Economies, in: Handbook of Regional and Urban Economics, Volume 4, Vernon Henderson and Jacques-François Thisse (eds.). Amsterdam: North-Holland.

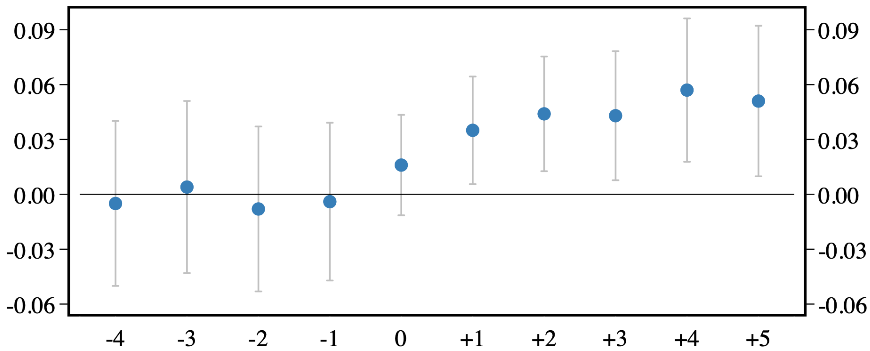
Rosenthal, Stuart, and William Strange, 2020, How Close Is Close? The Spatial Reach of Agglomeration Economies, Journal of Economic Perspectives 34, 27–49.

Slattery, Cailin, and Owen Zidar, 2022, Evaluating State and Local Business Incentives, Journal of Economic Perspectives 34, 90–118.

Figure 1: Treatment Effect Dynamics

This figure shows the coefficient estimates from columns (2) and (4) in Table A.2 of Online Appendix A along with 95% confidence intervals. The estimates are obtained using the imputation estimator of Borusyak, Jaravel, and Spiess (2023). The base year is τ = -5.

Panel A: *Local Productivity Spillover*



Panel B: *Global Productivity Spillover*

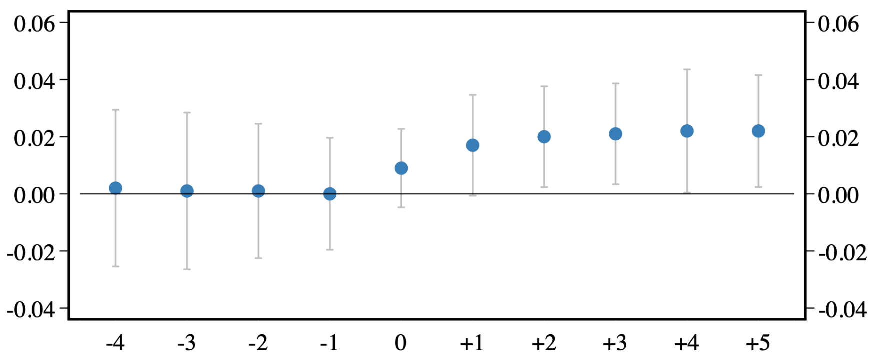
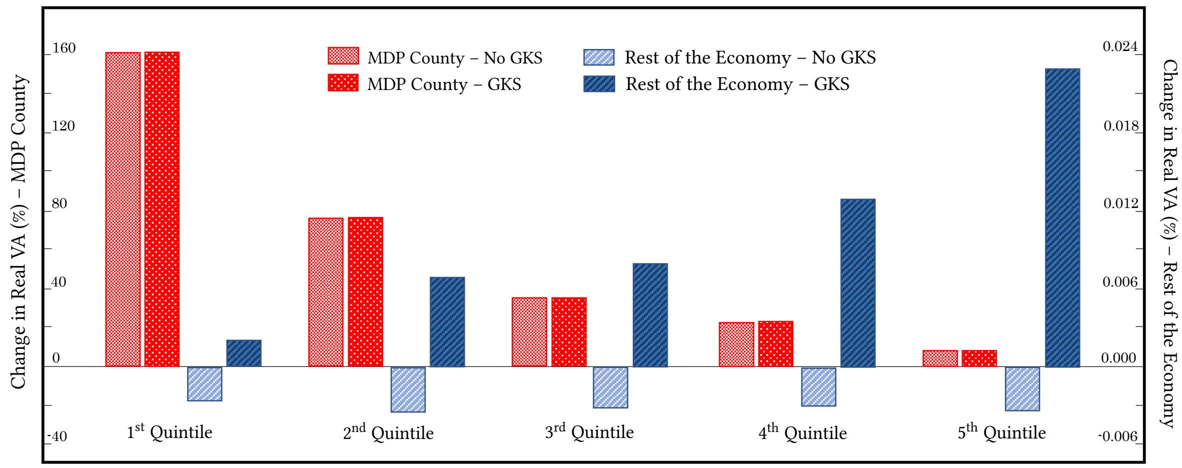


Figure 2: Plant Openings and Regional Development

This figure shows the percent change in real value added (VA) from MDP openings in more or less developed regions. MDPs are randomly assigned to counties sorted into population quintiles. In Panel A, the red bars show the local impact on the MDP county; the blue bars show the impact on the rest of the economy. A light (dark) color indicates that global knowledge sharing (GKS) is turned off (on). The left Y-axis pertains to the MDP county; the right Y-axis pertains to the rest of the economy. In Panel B, the (blue) bars show the aggregate impact on the entire economy. The MDP itself is excluded from both the local and aggregate impact.

Panel A: Effect on MDP County and Rest of the Economy



Panel B: Aggregate Effect

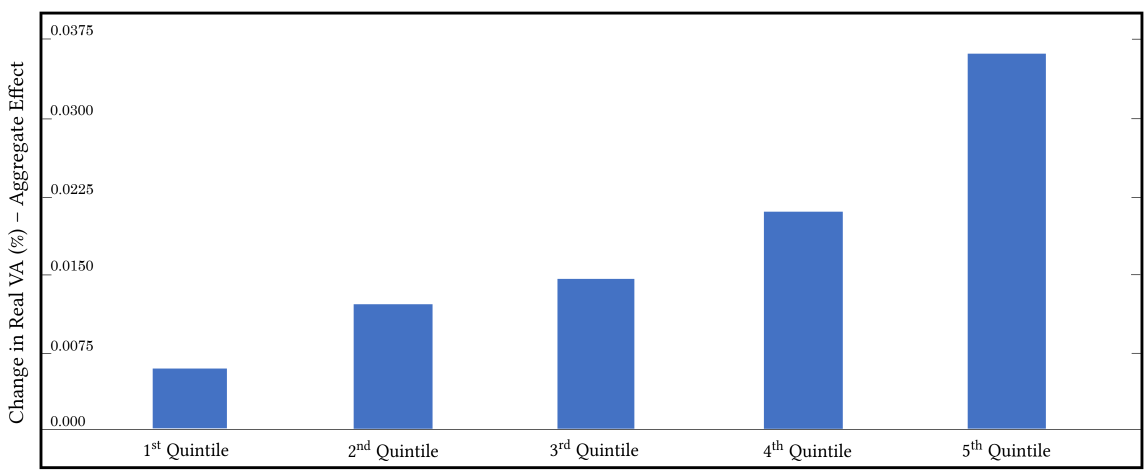


Table 1: Summary Statistics

Panel A provides plant-level statistics for the local spillover sample. Panel B provides plant-level statistics for the global spillover sample consisting of 423,000 plant-year observations. Panel C provides firm-level statistics for the parent firms associated with the plants in Panel B. Column (4) reports p-values of the difference between columns (2) and (3). Wages are in \$1,000. All statistics are from the year prior to the MDP opening. Standard deviations are in parentheses. The sample period is from 1977 to 1998.

	(1)	(2)	(3)	(4)
Panel A:	All	Winner	Loser	p-value (2) − (3)
Employees	141.7	146.3	139.8	0.377
	(571.4)	(589.3)	(562.8)	
Wages	39.5	41.5	38.7	0.454
	(852.7)	(877.2)	(763.9)	
TFP	0.002	0.003	0.001	0.672
	(0.586)	(0.610)	(0.551)	
Panel B:	All	Treated	Control	p-value (2) − (3)
Employees	268.2	272.6	266.3	0.482
	(846.7)	(903.4)	(821.8)	
Wages	35.9	34.3	36.5	0.535
	(202.2)	(311.5)	(162.9)	
TFP	0.016	0.017	0.016	0.903
	(0.640)	(0.653)	(0.637)	
Panel C:	All	Treated	Control	p-value (2) − (3)
Employees	1,988	1,968	1,997	0.834
	(6,702)	(6,862)	(6,548)	
Plants	7.4	7.3	7.5	0.661
	(10.9)	(10.6)	(11.0)	
Counties	5.4	5.3	5.5	0.532
	(7.7)	(7.2)	(7.8)	
States	2.7	2.6	2.8	0.448
	(2.8)	(2.6)	(2.9)	

Table 2: Local Productivity Spillover

The dependent variable is TFP at the plant level. MDP is an indicator for the winner county that is one from the year of the MDP opening onward. In column (3), (< 50 miles), (50 to 100 miles), and (100 to 250 miles) are indicators for whether a plant lies within 50 miles, between 50 and 100 miles, and between 100 and 250 miles, respectively, of the MDP. Only the main coefficients of interest are shown. Except for column (2), observations are weighted by plant-level employment. Standard errors are double clustered at the county and year level. The sample period is from 1977 to 1998.

	TFP		
		Unweighted	Distance
	(1)	(2)	(3)
MDP	0.040	0.038	
	(0.016)	(0.014)	
MDP × (< 50 miles)			0.043
			(0.015)
MDP × (50 to 100 miles)			0.027
			(0.014)
MDP × (100 to 250 miles)			0.011
			(0.010)
Plant FE	Yes	Yes	Yes
Industry × year FE	Yes	Yes	Yes
Case FE	Yes	Yes	Yes
R-squared	0.88	0.82	0.86
Observations	157,000	157,000	2,209,000

Table 3: Global Productivity Spillover

The dependent variable is TFP at the plant level. MDP is an indicator for whether the plant's parent firm has a plant in the winner county before and after the MDP opening. The indicator is one from the year of the MDP opening onward. In column (1), the control group consists of all plants of runner-up firms. In column (2), the control group consists of all plants of MC firms in the same county as the treated plant. In column (3), the control group consists of all plants of runner-up firms in the same county as the treated plant. In all three columns, the sample is restricted to MC plants outside the winner and runner-up counties. Only the main coefficients of interest are shown. Observations are weighted by plant-level employment. Standard errors are double clustered at the county and year level. The sample period is from 1977 to 1998.

	TFP		
	(1)	(2)	(3)
MDP	0.018	0.020	0.018
	(0.007)	(0.008)	(0.008)
Plant FE	Yes	Yes	Yes
Industry × year FE	Yes	-	-
Industry × county × year FE	-	Yes	Yes
Case FE	Yes	-	Yes
Control group	Plants of runner-up firms	Plants of MC firms in same county	Plants of runner-up firms in same county
R-squared	0.87	0.86	0.88
Observations	1,407,000	1,046,000	423,000

Table 4: Employment and Wages

This table presents variants of the regressions in column (1) of Table 2 (columns (1) and (2)) and column (3) of Table 3 (columns (3) and (4)) in which the dependent variable is either employment (columns (1) and (3)) or wages (columns (2) and (4)) at the plant level. Only the main coefficients of interest are shown. Observations are weighted by plant-level employment. Standard errors are double clustered at the county and year level. The sample period is from 1977 to 1998.

	Local Spillover		Global Spillover	
	Employment	Wages	Employment	Wages
	(1)	(2)	(3)	(4)
MDP	0.035	0.037	0.016	0.002
	(0.013)	(0.016)	(0.007)	(0.004)
Plant FE	Yes	Yes	Yes	Yes
Industry × year FE	Yes	Yes	-	-
Industry × county × year FE	-	-	Yes	Yes
Case FE	Yes	Yes	Yes	Yes
R-squared	0.97	0.80	0.98	0.58
Observations	157,000	157,000	423,000	423,000

Table 5: Extensive Margin

This table presents variants of the regression in column (1) of Table 2. In columns (1) and (2), the dependent variable is the logarithm of the number of MC and SC plants, respectively, at the county level. In columns (3) and (4), the dependent variable is the logarithm of one plus the number of plants in other counties of MC and SC firms, respectively, which have plants in the winner or runner-up counties prior to the MDP opening. Only the main coefficients of interest are shown. Observations are weighted by number of plants per county (columns (1) and (2)) or firm-level employment (columns (3) and (4)). Standard errors are double clustered at the county and year level (columns (1) and (2)) or firm and year level (columns (3) and (4)). The sample period is from 1977 to 1998.

	Number of plants in winner vs. loser counties		Number of plants in other counties	
	MC plants	SC plants	MC firms	SC firms
	(1)	(2)	(3)	(4)
MDP	0.036	0.066	0.005	0.002
	(0.026)	(0.030)	(0.014)	(0.014)
County FE	Yes	Yes	-	-
Firm FE	-	-	Yes	Yes
Year FE	Yes	Yes	-	-
Industry × year FE	-	-	Yes	Yes
Case FE	Yes	Yes	Yes	Yes
R-squared	0.99	0.99	0.24	0.63
Observations	1,000	1,000	76,000	81,000

Table 6: Distance to the MDP

This table presents variants of the regression in column (3) of Table 3 in which treated plants in close proximity to the MDP are excluded from the sample. Only the main coefficients of interest are shown. Observations are weighted by plant-level employment. Standard errors are double clustered at the county and year level. The sample period is from 1977 to 1998.

	TFP				
	Excluding plants within 100 miles of MDP	Excluding plants within 250 miles of MDP	Excluding plants within 500 miles of MDP	Excluding plants in MDP state	Excluding plants in MDP Census division
	(1)	(2)	(3)	(4)	(5)
MDP	0.018	0.017	0.018	0.018	0.018
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
Plant FE	Yes	Yes	Yes	Yes	Yes
Industry × county × year FE	Yes	Yes	Yes	Yes	Yes
Case FE	Yes	Yes	Yes	Yes	Yes
R-squared	0.88	0.88	0.89	0.88	0.88
Observations	402,000	365,000	286,000	395,000	345,000

Table 7: Knowledge Flows

This table presents variants of the regression in column (3) of Table 3. In column (1), both terms in equation (1) are interacted with a dummy variable indicating whether the treated plant is in the same 4-digit SIC code industry as the MDP. In columns (2) and (3), both terms in equation (1) are interacted with measures of mutual R&D flows and patent citations, respectively, between the industry of the treated plant and the industry of the MDP. The measures are the unidirectional measures $\text{Tech}_{ij} \equiv \max\{\text{TechIn}_{i\leftarrow j}, \text{TechOut}_{i\rightarrow j}\}$ and $\text{Patent}_{ij} \equiv \max\{\text{PatentIn}_{i\leftarrow j}, \text{PatentOut}_{i\rightarrow j}\}$ from Ellison, Glaeser, and Kerr (2010). Only the main coefficients of interest are shown. Observations are weighted by plant-level employment. Standard errors are double clustered at the county and year level. The sample period is from 1977 to 1998.

	TFP		
	Same industry	Mutual R&D flows	Mutual patent citations
	(1)	(2)	(3)
MDP	0.017	0.015	0.013
	(0.008)	(0.008)	(0.007)
MDP × knowledge flows	0.012	0.533	0.356
	(0.005)	(0.263)	(0.175)
Plant FE	Yes	Yes	Yes
Industry × county × year FE	Yes	Yes	Yes
Case FE	Yes	Yes	Yes
R-squared	0.88	0.88	0.88
Observations	423,000	423,000	423,000

Table 8: Trade with the MDP

This table presents variants of the regression in column (3) of Table 3. In columns (1) and (2), both terms in equation (1) are interacted with measures of input and output flows, respectively, between the industry of the treated plant and the industry of the MDP. The measures are the unidirectional measures $\text{Input}_{ij} \equiv \max\{\text{Input}_{i\leftarrow j}, \text{Input}_{i\rightarrow j}\}$ and $\text{Output}_{ij} \equiv \max\{\text{Output}_{i\leftarrow j}, \text{Output}_{i\rightarrow j}\}$ from Ellison, Glaeser, and Kerr (2010), where $\text{Input}_{i\leftarrow j}$ ($\text{Output}_{i\rightarrow j}$) denotes industry i's inputs (outputs) that come from (are sold to) industry j, normalized by industry i's revenues. In column (3), both terms in equation (1) are interacted with a measure of tradability of the treated plant's industry. The measure is the industry's geographical Herfindahl index from Appendix Table I of Mian and Sufi (2014). In column (4), both terms in equation (1) are interacted with a measure of exports from the treated plant's county to the winner county. The measure is the value of shipments from the treated plant's county to the winner county, normalized by the value of shipments to the winner county, from the Commodity Flow Survey. In columns (5) and (6), both terms in equation (1) are interacted with the geographical distance between the treated plant and the MDP. In column (6), the dependent variable is shipments at the plant level. Only the main coefficients of interest are shown. Observations are weighted by plant-level employment. Standard errors are double clustered at the county and year level. The sample period is from 1977 to 1998.

	TFP					Shipments
	Input flows	Output flows	Tradability	Exports	Distance	Distance
	(1)	(2)	(3)	(4)	(5)	(6)
MDP	0.017	0.017	0.017	0.018	0.019	0.029
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.013)
MDP × input/output flows	0.263	0.138				
	(0.432)	(0.250)				
MDP × tradability			0.019			
			(0.034)			
MDP × exports				0.031		
				(0.092)		
MDP × distance					-0.001	-0.000
					(0.004)	(0.010)
Plant FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry × county × year FE	Yes	Yes	Yes	Yes	Yes	Yes
Case FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.88	0.88	0.88	0.88	0.88	0.96
Observations	423,000	423,000	423,000	423,000	423,000	423,000

Table 9: Investment in Productivity

This table presents variants of the regression in column (1) of Table 2. In columns (1) and (2), both terms in equation (1) are interacted with an MC dummy and plant size, respectively. Plant size is the number of employees of the plant in the year before the MDP opening. In columns (3) and (4), both terms in equation (1) are interacted with firm-level measures of financial constraints (FC). The sample is restricted to firms in Compustat. In column (3), FC is the KZ-index of Kaplan and Zingales (1997). In column (4), FC is the SA-index of Hadlock and Pierce (2010). In column (5), the dependent variable is R&D scaled by assets at the firm level. The sample is restricted to firms in Compustat with non-missing R&D values. In column (6), the dependent variable is the logarithm of one plus the number of patents at the firm level. The sample is restricted to firms in the merged Compustat-USPTO patent database. Only the main coefficients of interest are shown. Observations are weighted by plant-level employment (columns (1) to (4)) or firm-level employment (columns (5) and (6)). Standard errors are double clustered at the county and year level (columns (1) to (4)) or firm and year level (columns (5) and (6)). The sample period is from 1977 to 1998.

	TFP				R&D	Innovation
	MC dummy	Plant size	KZ-index	SA-index		
	(1)	(2)	(3)	(4)	(5)	(6)
MDP	0.047	0.043	0.041	0.040	0.001	0.011
	(0.018)	(0.020)	(0.018)	(0.017)	(0.001)	(0.009)
MDP × MC	-0.008					
	(0.017)					
MDP × plant size		-0.001				
		(0.006)				
MDP × FC			-0.001	-0.001		
			(0.002)	(0.004)		
Plant FE	Yes	Yes	Yes	Yes	-	-
Firm FE	-	-	-	-	Yes	Yes
Industry × year FE	Yes	Yes	Yes	Yes	Yes	Yes
Case FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.88	0.88	0.89	0.89	0.70	0.91
Observations	157,000	157,000	42,000	42,000	15,000	40,000