Foreign Direct Investment and Local Productivity Spillovers: Evidence from Indonesia*

Taylor Lathrop†
Syracuse University

2025-12-01

Abstract

In this paper, I identify and quantify local spillovers using a rich dataset of Indonesian manufacturing firms with granular location information. Within local clusters of manufacturing employment, I find an elasticity of domestic firm productivity with respect to the average quality of domestic neighbors of 0.025, and an elasticity of domestic firm productivity with respect to the average quality of foreign neighbors of 0.002. For foreign-owned firms, I find little evidence of productivity spillovers from peers of either ownership type. A simple policy simulation indicates that areas of lower manufacturing density gain the most on average from a new high-productivity peer.

JEL Classifications: D24, F21, R11, O14, O33

Keywords: productivity spillovers, foreign direct investment

*I am grateful to Alex Rothenberg, Ryan Monarch, and Shafaat Yar Khan for their guidance. This work greatly benefited from comments from Alfonso Flores-Lagunes, Hugo Jales, Thomas Pearson, Maria Zhu, Kristy Buzard, Abdulaziz Shifa, John Cawley, and participants in the Syracuse University Labor and Trade, Development, and Political Economy groups. All errors remain my own.

†106 Eggers Hall, Syracuse, NY 13244-1020. Email: tlathrop@syr.edu.

1 Introduction

The idea that Foreign Direct Investment (FDI) creates substantial benefits for the host country has long enjoyed widespread popularity among policymakers. In theory, openness to FDI attracts not only direct inflows of foreign capital, but also facilitates technology transfer from foreign to domestic firms via the spillover externalities first proposed by Marshall (1890).

Geographic distance increases transport costs and reduces the likelihood that firms derive benefits from highly-productive firms that are far away from them. On the other hand, firms that are near highly-productive firms may benefit by sharing a labor market with them, supplying their intermediate inputs, or through local knowledge transfers between workers. In pursuit of these spillovers, many developing countries have tried to attract foreign capital by setting up special economic zones, giving away tax concessions, relaxing regulations, and enacting other policies that give preferential treatment to foreign firms.

In spite of this enthusiasm, a large literature on the spillover benefits of FDI has produced mixed results. Studies that have focused on horizontal FDI spillovers and used measures of FDI penetration over broad areas in developing countries have generally found null or negative spillover effects (Aitken and Harrison, 1999; Konings, 2001), though similar studies in developed countries have found positive spillovers (Keller and Yeaple, 2009). Work that has focused on estimating FDI spillovers through vertical linkages (Javorcik (2004) in Lithuania, Blalock and Gertler (2008) in Indonesia), in which foreign firms transfer technology to their upstream or downstream connections, has more consistently found positive FDI spillovers.

Much of the difficulty of accurately estimating spillovers is due to the complex nature of solving the "reflection problem" and credibly disentangling the endogenous, exogenous, and contextual effects described by Manski (1993). Recent work on spillovers between firms has drawn on ideas from the labor economics literature on peer effects to identify endogenous productivity spillovers between firms. The typical method in this literature has been to use a two-step approach: the first step is to estimate total factor productivity via a control function method[1]. The second step uses the productivity estimates recovered from this method as inputs into a process that controls for location fundamentals and other potential confounders. For example, Serpa and Krishnan (2018) use the instrumental variables technique proposed by Bramoullé et al. (2009), using the exogenous characteristics of second-order neighbors to identify endogenous spillovers between U.S. firms. In the Indonesian context, Bazzi et al. (2017) also use the Bramoullé et al. (2009) technique and supply chain networks to identify endogenous spillovers between in-network firms. In a recent working paper, Iyoha (2023) formulates a method to tackle credible estimation of spillovers and network endogeneity simultaneously.

[1] Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg et al. (2015), among others

However, there is still relatively little work that has applied peer effects methods to spillovers at the very local level, particularly in the developing country context. Additionally, because the vast majority of work on productivity spillovers has focused on the direct effect of the productivity of connected firms on own productivity, the effect of nearby firm characteristics or exogenous quality on firm-level productivity is typically controlled for in order to achieve clean identification. However, the spillovers arising from the exogenous characteristics of peers are still an important component of the full spillovers picture within local peer groups. Understanding these exogenous effects would help to contextualize previous work on endogenous effects, and would improve our understanding of the local geography of economic activity.

In this paper, I identify and quantify revenue productivity spillovers at the local level using geography and a high-quality panel dataset of Indonesian manufacturing firms with granular location information. To capture the effect of these relationships at relatively-fine spatial scales, I use data on Indonesian administrative borders and firm-level location information within these borders to assign firms to clusters of manufacturing employment, whose borders are delineated using a density surface defined using a likelihood cross-validated bandwidth and a "watershed segmentation" technique based on Lee and Lee (2024).

Using these local clusters and detailed information on firm-level inputs and financials, I follow the technique used by Baum-Snow et al. (2024) and estimate local revenue spillovers between firms by nonlinear least squares, using the iterative algorithm proposed by Arcidiacono et al. (2012), with a focus on measuring how these spillovers propagate from foreign-owned firms to their local peers. Importantly, this technique focuses only on the recovery of "exogenous" spillovers, in which the exogenous attributes of peers have a causal effect on firm outcomes, by focusing on revenue as an outcome variable and on spillovers that propagate over relatively small spatial scales.

My main identifying assumption is that manufacturers cannot strategically choose the cluster in which they locate to take advantage of cluster-time-specific total factor productivity shocks. At best, they are only able to choose what district (*kabupaten/kota*) they locate in. Additionally, I assume firms lack perfect foresight with respect to the timing of total factor productivity shocks, and would thus be unable to time their entry to positive productivity shocks even in the event that they could choose their peers. In the Indonesian context, the presence of large frictions in real estate markets arising from dueling systems of property rights — the formal system (a holdover from the Dutch colonial era) and the informal *adat* system (Kusno, 2017) — provides support this strategy. Harari and Wong (2025) show that these competing systems make assembling and redeveloping large, contiguous parcels of land difficult. Additionally, the Indonesian government forbids foreigners from owning freehold property outright, a further complication for foreign manufacturers looking to open a branch within the country.

Within local manufacturing clusters, I find evidence of revenue spillovers to domestic firms that differ by peer ownership type. I estimate an average elasticity of domestic firm revenues to average domestic peer quality of 0.025, and a much more modest average elasticity of domestic firm revenues to average foreign peer quality of 0.002. I find that foreign firms make negligible average gains from peers of either ownership type.

In the context of the model, firms naturally have incentives to locate near the highest-quality peers available, but land market frictions limit their ability to precisely choose their location. In spite of this, I find strong evidence of assortative matching between firms on their estimated qualities. High-quality firms tend to locate with other high-quality firms, and low-quality firms with other low-quality firms. However, the assortative matching is not perfect, reinforcing my identification assumptions.

To explore the potential policy implications of my local spillover estimates, I conduct two policy experiments. The first explores how an exogenous productivity increase to any one firm propagates to its peers to determine which areas would obtain the largest average and aggregate productivity gains from such an increase. The second, more directly inspired by the main question of this paper, explores the productivity gains to local firms following the entry of a highly-productive foreign firm, again with the intent of determining which areas would benefit the most. In both cases, there is clear evidence that areas around the urban fringe (where manufacturing density is relatively low) stand to reap the largest benefits.

A large body of literature documents the potential gains to host country firms from FDI spillovers. Historically, a majority of these papers have studied FDI spillovers in the context of developed countries (Caves (1974), Keller and Yeaple (2009), and more recently Amiti et al. (2024), among others). More recently, a literature on FDI spillovers in developing countries has formed. Much of the work in this literature is focused on FDI in China (Lin and Kwan, 2016; Abraham et al., 2010), though some studies have examined other developing countries (Blalock and Gertler, 2008). This paper contributes to this literature by providing new estimates of the effects of FDI in developing countries from the perspective of local spillovers rather than through supply chains, and helps to clarify the extent to which domestic firms in developing countries benefit from co-agglomeration with domestic and foreign peers.

Many recent studies have reinforced labor market pooling and agglomeration as important sources of spillovers between firms, finding that industries that employ similar types of workers tend to benefit from the human capital externalities arising from co-agglomeration (Freedman, 2008; Ellison et al., 2010; Fallick et al., 2006) and emphasizing the importance of interactions between workers in creating those externalities (Ibrahim et al., 2009; Kloosterman, 2008). This paper helps to clarify these dynamics in a setting where foreign and domestic workers and firms share a common area, and adopts a unique strategy for identifying the clusters in which these dynamics operate.

A small but burgeoning literature examines the effects of FDI from a spatial perspective. Tanaka and Hashiguchi (2015) and Lin and Kwan (2016) explore the extent to which FDI presence generates productivity spillovers at the county level in China, and Monastiriotis and Jordaan (2010) document negative FDI spillovers in major urban areas in Greece, with positive spillovers arising in peripheral regions. My focus on Indonesian firms and watershed-based clustering provides a unique setting and a more granular level of agglomeration, and my usage of a structural model of heterogeneous spillovers is unique in this setting, to my knowledge.

The rest of the paper is as follows. Section 2 provides details on the background and institutional setting in Indonesia, describes the datasets used, and describes the method by which I segment groups of Indonesian villages into clusters of manufacturing employment. Section 3 presents some reduced-form evidence to motivate the main investigation of the paper. Section 4 introduces a structural model of FDI spillovers based on Baum-Snow et al. (2024) and Arcidiacono et al. (2012), and section 5 describes an empirical strategy based on the same. Section 6 describes the results. Section 7 explores the extent to which firms are able to match on their quality. Section 8 provides a policy simulation based on the estimated parameters of the model and TFP estimates derived from the model, and section 9 concludes.

2 Background, Setting, and Data

The late 1990s to early 2010s Indonesian economy provides an interesting context for studying FDI and technology transfer, particularly in the manufacturing sector. The world's fourth-largest country by population and fourteenth-largest country by landmass, Indonesia's broad variety of natural resources and ample labor inputs have allowed it to develop a large manufacturing sector spread across many industries. However, the archipelagic nature of its geography and its limited transportation infrastructure create substantial barriers to internal trade across regions, which have in part led to considerable differences in the pattern of manufacturing activity across the country. As a result, by the 1990s, three fourths of all non-petroleum manufacturing activity was located on the island of Java (Amiti and Cameron, 2007). Keenly aware of these regional disparities, the Indonesian government has embarked on several efforts to influence the location decisions of manufacturers and attract foreign capital.[2]

Over the decades leading up to the 1990s, policies towards FDI changed substantially, beginning with the Foreign Investment Law of 1967, which opened many sectors of the economy to limited foreign investment, established a foreign investment advisory board, and provided tax breaks to investors establishing new businesses, with extended tax breaks for firms locating outside of Java. President Suharto's government additionally invested heavily in establishing the city of Batam (located in the Riau Islands and in close proximity to Singapore) as a special economic zone, and

[2]See Rothenberg and Temenggung (2019) for a complete overview.

established the Integrated Economic Development Zone Agency (*Badan Pengembangan Kawasan Pengembangan Ekonomi Terpadu*/BP KAPET) to coordinate the development of selected districts by awarding firms generous tax incentives to locate there, though these efforts were not exclusively focused on attracting foreign investors.

After the 1997 Asian Financial Crisis and the fall of Suharto in 1998, Indonesia embarked on a series of economic reforms aimed at stabilizing its economy, attracting foreign investment, and fostering industrial growth. These reforms were accompanied by broad decentralization reforms that empowered regional policymakers and a renewed focus on developing special economic zones. The success of these policies in attracting firms to districts outside of Java has been mixed (Rothenberg et al., 2025), though the share of manufacturers that are foreign-owned has grown both on and off Java since the 1970s. Given the sizable investments that Indonesia and other developing countries have made towards attracting foreign firms to certain areas, understanding how spillovers propagate between firms within these areas is a topic of considerable importance.

My primary data are drawn from Indonesia's annual survey of manufacturing firms, the *Survei Tahunan Perusahaan Industri Pengolahan* (SI), conducted by Indonesia's Central Bureau of Statistics, *Badan Pusat Statistik* (BPS). The SI is designed to cover all manufacturing firms with twenty or more employees, and it includes detailed information on firm material inputs, electricity consumption, labor count, wage payments, and ownership shares. Firms are assigned a unique, time-invariant identifier that allows for tracking firms over time. BPS takes steps to ensure that the SI is a complete accounting of all currently-operating medium and large manufacturing firms in Indonesia, including sending agents to verify plant closures.

I keep all firms located within Indonesia's urban areas with at least two years of reported data during the sample period, treating the first year that a firm is observed as its entry year, and the last period that it is observed in as its exit year. To determine the boundaries of Indonesia's urban areas, I rely on definitions provided by Civelli et al. (2023). The SI includes ownership shares for domestic private, foreign private, and public ownership. As is standard in the literature, I follow the OECD definition of FDI and treat any firm with more than a 10% foreign ownership share as foreign-owned (OECD, 1996). A number of firms start out as domestically-owned, but experience brownfield investment from foreign entities during the sample period. I treat these firms as domestically-owned until they experience foreign investment, at which point I begin treating them as foreign-owned.

Measuring where firms are located and who their neighbors are is a key part of this analysis. For each firm, the SI provides a code that identifies its location at all administrative levels, down to the village (*desa* or *kelurahan*) level.[3] Additionally, BPS provides GIS shapefiles that identify the

[3] To the extent that multiplant firms exist in the dataset, they are not identifiable. Poczter et al. (2014) suggest that multiplant firms constitute less than 5% of the sample, however.

borders of each administrative unit. Indonesia's village-level administrative divisions have changed considerably over the years - the same *desa* code may identify different areas in different years. To account for these changes, I use a crosswalk provided by BPS to map firms to geographic locations that are consistent over the studied time period. These village-level crosswalks are only available from 1996 onward. I further restrict the sample to begin in 1999 to avoid the macroeconomic effects of the Asian Financial Crisis of 1997 and the subsequent administrative changes that resulted from the fall of the Suharto government in 1998. The sample period ends in 2010, the last year of available data.[4]

2.1 Cluster Segmentation

Obtaining a unique solution to my empirical model requires grouping firms into mutally-exclusive, non-overlapping geographic peer groups. Delineating spatial clusters of economic activity presents a unique challenge, as economic activity may not be neatly constrained to government-defined administrative boundaries. Thus, it may be desirable to define spatial clusters based on data. In the broad spatial literature, this is typically done using a density-based approach — that is, selecting a density threshold and defining blocks by areas that fall above this threshold.

However, identification of clusters depends crucially on the selected density threshold, and there may not be a "one size fits all" solution; a single density threshold that identifies separate clusters in high-density areas may miss clusters in low-density areas, whereas a threshold that identifies low-density area clusters may oversmooth high-density areas into single clusters. Lee and Lee (2024) point out this problem and propose a core-based approach that identifies the centroids of clusters and then proceeds to identify their boundaries. I adapt their approach to create clusters of manufacturing employment in Indonesian cities. Figure 1 provides a visual example of manufacturing employment watershed segmentation using the Jakarta metropolitan area — Figure 1a shows Jakarta's village-level administrative boundaries, with each village shaded by its level of manufacturing employment.

Using a shapefile provided by BPS of Indonesia's village boundaries in 2000, I use GIS software to create a raster image of 1×1 kilometer cells as shown in Figure 1b. The value within each raster cell is an average of the manufacturing employment population of each village overlapping the cell, weighted by their areas within the cell. I then employ a spatial quartic kernel (Silverman, 1986) to smooth this raster surface into a density surface, as shown in Figure 1c.[5] Using the quartic kernel, the density of employment at each (x, y) location is calculated as:

[4] More recent versions of SI exist, but they do not contain *desa*-level location indentifiers.

[5] This is done by converting each raster cell to a point with a value of the cell, which differs slightly from the approach used by Lee and Lee (2024), who run the raster image through a kernel-based smoothing filter to create a density surface.

$$d_{x,y} = \frac{1}{h^2} \sum_{i:dist_i(x,y)\leq h} \left[\frac{3}{\pi} \times pop_i \left(1 - \left(\frac{dist_i}{h} \right)^2 \right)^2 \right]$$

where i is a raster cell, $dist_i$ is the distance of i from point (x, y), pop_i is the weighted average manufacturing population in raster cell i, and h is the bandwidth. Only cells within a distance of h from point (x, y) are included. The selection of the bandwidth h is crucial; the size of the bandwidth will directly impact the size and number of clusters, and care is needed in selecting it to avoid under- or oversmoothing the density surface. In my primary specification, I use likelihood crossed-validation (Baddeley et al., 2015) to select a bandwidth h of roughly 2.6 kilometers. In alternative specifications, I also manually select several alternative bandwidths to explore how different cluster sizes affect the magnitude and sign of spillovers.

After creating the density surface, I invert the surface so that the local areas of highest manufacturing employment density become local minima. This also creates "ridges" at the areas of lowest density, i.e. the boundaries between clusters. Using this inverted density surface as a digital elevation model (DEM), I calculate a flow direction raster using the D8 algorithm proposed by O'Callaghan and Mark (1984). This algorithm takes each point of the DEM and determines which of its eight neighboring points are downhill from it. Cells with no downhill neighbors correspond to local minima. I apply a basin delineation tool to the flow direction raster, which automatically selects "pour points" (areas of lowest elevation/highest flow accumulation) and uses the ridges surrounding these pour points to create watersheds. The resulting watersheds are shown in Figure 1d. I assign firms to watersheds based on the location of the centroid of their corresponding villages, dropping any firms that do not share a cluster with any others.[6]

3 Motivating Evidence

In this section, I motivate the main exercise of this paper by exploring the differences between foreign and domestic firms and providing some reduced form evidence of the effects of foreign entry on domestic firms in Indonesia. Foreign-owned firms differ from domestically-owned firms in a number of important ways. Table 1 presents the results of several simple fixed effects least squares regressions of firm characteristics on foreign status, controlling for firm size. Foreign-owned status is associated with significantly higher wage rates, outputs, revenues, and total factor productivity.[7]

To begin exploring the impact on domestic manufacturers when these foreign firms set up near

[6] This requires transforming the watershed raster image back into a polygon-based shapefile. This process leaves behind some pixel-sized artifacts on the edges of some watersheds, which I eliminate by merging them into neighboring polygons that have the longest shared border with them.

[7] The TFP that I use here is backed out of the model - calculation occurs in Section 8.1

them, I estimate the following multiple-period differences-in-differences model:

$$y_{ict} = \varphi_{ct} + \beta D_{ict} + u_{ict}$$

where y_{ict} is some outcome variable of interest for domestic firm i in cluster c in year t, φ_{ct} is a vector of cluster and time fixed effects, D_{ict} is a binary treatment indicator, and u_{ict} is an error term. Treatment for domestic firms occurs when they go from having no foreign peers within their clusters to having at least one foreign peer. Because foreign firms enter at different points in the sample, treatment for different units occurs at different times. I follow the Callaway and Sant'Anna (2021) method of estimating differences-in-differences with multiple periods via a double-robust inverse probability weighted estimator. The control group is all domestic firms that have yet to be treated in each period. I exclude from the sample any domestic firms who are treated before the beginning of my sample period and any domestic firms who experience a reversal in treatment (i.e. the departure of all foreign neighbors).

Table 2 presents results of this estimation for the full sample, with outcome variables log revenue, log output, log labor, and log TFP.[8] Though I find little effect of foreign entry on revenues, output, or labor with this specification, I do find that foreign entry is correlated with a significant boost in residual total factor productivity of 0.127 for domestic incumbents. To investigate these effects further, I estimate the same model for each two-digit ISIC manufacturing industry. Table 3 presents these results. Though foreign entry produces unclear effects at best in some industries, it appears to be more impactful in food processing, metalworking, and to some extent chemical manufacturing.

Finally, it may be the case that foreign entry causes domestic firms to enter as well, perhaps in the hopes of entering supply chains or sharing labor pools with foreign firms. It may also be the case that foreign entry crowds out domestic firms, causing them to exit. To investigate whether foreign entry is associated with arrivals or departures of domestic firms within the cluster in which they locate, I estimate several cluster-level fixed effects least squares specifications, regressing total number of domestic firms, percent change in domestic firms, log number of domestic entrants, and log number of domestic exits on a binary indicator for new foreign presence where there previously was none. As above, I exclude clusters that start the sample period with foreign presence and clusters that lose all foreign presence during the sample period. The results of this specification are presented in Table 4. Clusters with foreign presence tend to have around 1.4 times as many domestic firms, and also tend to experience more churn post-foreign entry, with 39% more domestic entrants and 37% more domestic exits than clusters that do not experience foreign entry.

Summarizing, the reduced-form evidence of the effects of foreign entry on domestic firms is somewhat muddy. The entry of a foreign peer is associated with productivity gains for domestic

[8] Figure A1 presents event study plots of the same multi-period differences-in-differences regressions.

firms on average, but foreign firms also tend to enter into more populous clusters with higher churn. The mechanics of why foreign entry is associated with productivity gains are unclear. To provide clarity on any potential channel through which these productivity increases occur, some additional structure is needed.

4 Structural Model of TFP Spillovers

The model closely follows the setup of Baum-Snow et al. (2024), with the distinguishing feature that it allows for heterogeneity in spillovers by firm type. The key feature of the model is that a firm's productivity depends on its own fixed effect, location fundamentals, and weighted averages of the fixed effects of its domestic- and foreign-owned peers. A firm's peers are determined by its location. Urban Indonesian villages are grouped into local clusters c of manufacturing employment, and firms are assigned to these clusters based on the geographical location of the centroid of the village they are located in. Firms do not have full control over their location - though they may choose which district C they locate in, they cannot choose a particular cluster within that district due to commercial real estate frictions that commonly occur within cities.

In each year t, each firm i in industry k within cluster c chooses variable inputs to maximize their short-run profit function:

$$\pi_{i,c,k,t} = p_{C,k,t} A_{i,c,k,t} L_{i,c,k,t}^{\theta_k} - w_{C,k,t} L_{i,c,k,t} - F_{i,c,k,t} \tag{1}$$

Firms choose their variable inputs $L_{i,c,k,t}$ conditional on their location - input prices $w_{C,k,t}$ are determined yearly at the district-industry level. I also invoke the assumption of perfect competition, so output prices $p_{C,k,t}$ are also determined at the district-industry level and may be controlled for with district-industry-year fixed effects. θ_k represents the variable input share, which varies by industry. $F_{i,c,k,t}$ are capital and land inputs, which are fixed in the short run. Following from equation (1), an expression for log revenue is:

$$\ln R_{i,c,k,t} = \ln p_{C,k,t} + \ln A_{i,c,k,t} + \theta_k L^{\star}_{i,c,k,t} \tag{2}$$

where $L^{\star}_{i,c,k,t}$ is firm i's optimal demand for variable inputs. Substituting in this equation yields an alternative expression for log revenue that will be useful later:

$$\ln R_{i,c,k,t} = \frac{\theta_k}{1-\theta_k} \ln \theta_k + \frac{1}{1-\theta_k} \ln p_{C,k,t} + \frac{1}{1-\theta_k} \ln A_{i,c,k,t} - \frac{\theta_k}{1-\theta_k} \ln w_{C,k,t} \tag{3}$$

I assume that the main object of interest, domestic firm i's total factor productivity $A_{i,c,k,t}$, follows the following data-generating process:

$$\ln A_{i,c,k,t} = \alpha_i^A + \phi_{C,k,t}^A + \frac{\beta_{\text{Dom}}^A}{M_{c,t,g(i)} - 1} \left(\sum_{j \in \mathbb{N}_{c,t,g(i) \sim i}} \alpha_j^A \right) + \frac{\beta_{\text{For}}^A}{M_{c,t,g'}} \left(\sum_{j \in \mathbb{N}_{c,t,g'}} \alpha_j^A \right) + \varepsilon_{i,c,k,t}^A \tag{4}$$

In words, domestic firm i's productivity depends on its own fixed effect α_i, which is a time-invariant measure of firm quality, district-industry-year fixed effects $\phi_{C,k,t}$, and linear-in-means aggregations of the fixed effects of its peers by ownership type. $M_{c,t,g(i)}$ is the number of firms in the same cluster as i and of the same ownership type (domestic) as firm i, denoted $g(i)$. $\mathbb{N}_{c,t,g(i)\sim i}$ is the set of such firms within the cluster excluding i. β_{Dom} therefore measures the elasticity of firm i's productivity to the average quality of its domestically-owned peer firms. Similarly, $M_{c,t,g'}$ is the number of firms in the same cluster as i whose ownership type (foreign) differs from i, $\mathbb{N}_{c,t,g'}$ is the set of all such firms within the cluster, and β_{For} measures the elasticity of firm i's productivity to the average quality of its peer firms of different (foreign) ownership type.

There is a separate, equivalent data-generating process for foreign firm total factor productivity:

$$\ln A_{i,c,k,t} = \alpha_i^A + \phi_{C,k,t}^A + \frac{\beta_{\text{For}}^A}{M_{c,t,g(i)} - 1} \left(\sum_{j \in \mathbb{N}_{c,t,g(i) \sim i}} \alpha_j^A \right) + \frac{\beta_{\text{Dom}}^A}{M_{c,t,g'}} \left(\sum_{j \in \mathbb{N}_{c,t,g'}} \alpha_j^A \right) + \varepsilon_{i,c,k,t}^A \tag{5}$$

For the remainder of the paper, I will focus on the equations for domestic firms, but all calculations and equations apply for foreign firms as well.

As in Baum-Snow et al. (2024), my reduced-form estimating equation for spillovers is based on revenue spillovers that can be related back to TFP spillovers rather than on TFP spillovers directly. There are several advantages to choosing revenue as the dependent variable in the estimating equation. Control function approaches to TFP estimation use lagged measures of labor, capital, and materials as instruments, but these objects (particularly capital) are reported less consistently in SI than revenues, which are consistently reported for the vast majority of firms. Additionally, these methods are unable to estimate TFP for firms in their first year of operation. Since identifying variation in the empirical exercise comes from changes in peer group composition resulting from entry and exit of firms, a control function approach is not well-suited to this exercise.

To see how revenue spillovers may proxy for TFP spillovers, first consider the following reduced form equation for estimating revenue spillovers, which mirrors the data generating process for TFP:

$$\ln R_{i,c,k,t} = \alpha_i^R + \phi_{C,k,t}^R + \frac{\beta_{\text{Dom}}^R}{M_{c,t,g(i)} - 1} \left(\sum_{j \in \mathbb{N}_{c,t,g \sim i}} \alpha_j^R \right) + \frac{\beta_{\text{For}}^R}{M_{c,t,g'}} \left(\sum_{j \in \mathbb{N}_{c,t,g'}} \alpha_j^R \right) + \varepsilon_{i,c,k,t}^R \tag{6}$$

Using equations (3) and (6), the relationship between the reduced-form expression for revenue and the data generating process for domestic firm i's TFP is

$$\begin{aligned} &\alpha_i^R + \frac{\beta_{\text{Dom}}^R}{M_{c,t,g(i)} - 1} \left(\sum_{j \in \mathbb{N}_{c,t,g(i) \sim i}} \alpha_j^R \right) + \frac{\beta_{\text{For}}^R}{M_{c,t,g'}} \left(\sum_{j \in \mathbb{N}_{c,t,g'}} \alpha_j^R \right) + \varepsilon_{i,c,k,t}^R \\ &\quad = \frac{1}{1-\theta_k} \left[\alpha_i^A + \frac{\beta_{\text{Dom}}^A}{M_{c,t,g(i)} - 1} \left(\sum_{j \in \mathbb{N}_{c,t,g(i) \sim i}} \alpha_j^A \right) + \frac{\beta_{\text{For}}^A}{M_{c,t,g'}} \left(\sum_{j \in \mathbb{N}_{c,t,g'}} \alpha_j^A \right) + \varepsilon_{i,c,k,t}^A \right] \end{aligned} \tag{7}$$

Assuming normal distribution of the error terms and setting $\alpha_i^R = \frac{1}{1-\theta_k}\alpha_i^A$, this expression shows that the revenue spillovers β_{Dom}^R and β_{For}^R directly measure productivity spillovers β_{Dom}^A and β_{For}^A if all firms have the same variable input share.

For firms in the manufacturing sector, this is likely too strong an assumption. Instead, in the main specification, I allow the variable input share θ_k to vary by industry and use $(1-\theta_k)\ln R_{i,c,k,t}$ as the dependent variable. To calculate θ_k, I sum up the payments to labor (wages) for each industry-year to obtain an industry-year level variable input share $\theta_{k,t}$. I then average these industry-year level shares over all years to obtain an estimate of θ_k.

Thus far, I have assumed a perfectly competitive framework. To test the strength of this assumption, I also allow for monopolistic competition. Baum-Snow et al. (2024) showed that, by following the setup in De Loecker (2011) and allowing all firms in industry k to have the same markups and the same output demand elasticities η_k, the structural equation for firm log revenues equation (3) becomes:

$$\begin{aligned} \ln R_{i,c,k,t} = &\frac{1+\eta_k}{\eta_k(1-\theta_k) - \theta_k} \ln A_{i,c,k,t} - \frac{\theta_k(1+\eta_k)}{\eta_k(1-\theta_k) - \theta_k} \ln w_{C,k,t} \\ &- \frac{\theta_k(1+\eta_k)}{\theta_k - \eta_k(1-\theta_k)} \ln \left(\frac{1+\eta_k}{\eta_k} \theta_k \right) + \frac{\ln X_{k,t} + \zeta_{i,c,k,t}}{\theta_k - \eta_k(1-\theta_k)} \end{aligned} \tag{8}$$

where $X_{k,t}$ is an industry-time effect and $\zeta_{i,c,k,t}$ is a demand shock.

By the same process as above, the relationship between the reduced-form estimating equation for revenues and the structural data generating process for total factor productivity under monopolistic competition is:

$$\alpha_i^R + \frac{\beta_{\text{Dom}}^R}{M_{c,t,g(i)}-1}\left(\sum_{j\in\mathbb{N}_{c,t,g(i)\sim i}} \alpha_j^R\right) + \frac{\beta_{\text{For}}^R}{M_{c,t,g'}}\left(\sum_{j\in\mathbb{N}_{c,t,g'}} \alpha_j^R\right) + \varepsilon_{i,c,k,t}^R$$
$$= -\frac{1+\eta_k}{\theta_k-\eta_k(1-\theta_k)}\alpha_i^A + \frac{1+\eta_k}{\theta_k-\eta_k(1-\theta_k)}\left[\frac{\beta_{\text{Dom}}^A}{M_{c,t,g(i)}-1}\left(\sum_{j\in\mathbb{N}_{c,t,g(i)\sim i}} \alpha_j^A \frac{\theta_k-\eta_k(1-\theta_k)}{1+\eta_k}\right)\right.$$
$$\left. + \frac{\beta_{\text{For}}^A}{M_{c,t,g'}}\left(\sum_{j\in\mathbb{N}_{c,t,g'}} \alpha_j^A \frac{\theta_k-\eta_k(1-\theta_k)}{1+\eta_k}\right) + \frac{1+\eta_k}{\theta_k-\eta_k(1-\theta_k)}\varepsilon_{i,c,k,t}^A + \frac{\zeta_{i,c,k,t}}{\theta_k-\eta_k(1-\theta_k)}\right]$$

Though this is a considerably more complicated version of equation (7) above, the interpretation is similar: if all firms have the same variable input share θ_k and face the same demand elasticity η_k, or if all firms are in the same industry k, then revenue spillovers measure total factor productivity spillovers without any adjustment necessary. If not, then using $\frac{\eta_k(1-\theta_k)-\theta_k}{1+\eta_k}\ln R_{i,c,k,t}$ as the outcome variable will allow for measuring total factor productivity spillovers instead.

To calculate η_k, I start by calculating each firm's markup over marginal costs as the variable input share θ_k times the sum of firm-level revenues in industry k in year t divided by the sum of payments to labor and materials in industry k in year t. I then average this object over all years to obtain η_k.

5 Empirical Strategy

5.1 Threats to identification

Manski (1993) showed that simultaneity in the actions of individuals in peer groups causes the expected mean outcome for a group to be perfectly collinear with its mean characteristics, making it difficult to disentangle endogenous and exogenous spillovers. Much of the peer effects literature is dedicated to solving this "reflection problem" in order to credibly estimate endogenous spillovers, in which an individual's own outcome is directly influenced by the outcomes of its peers. This paper instead attempts to isolate the "exogenous" effects of the characteristics of peer firms on a firm's outcome - specifically, the effect of the average quality of peer firms, as measured by their time-invariant fixed effects. Focusing on revenue as the outcome rather than TFP during estimation also avoids the potential endogeneity of firm's input choices that is typically problematic in productivity estimation, though backing TFP out of the model from revenue spillovers does require some assumptions about demand.

Given the importance of the composition of firms within peer groups in identifying spillovers, it is important to ensure that the results are not simply driven by "correlated effects," in which peers experience similar outcomes due to prior similarities in their environments. Controlling for firm-level fixed effects controls for this type of sorting. Additionally, the inclusion of district-year fixed effects controls for district-level trends that may affect firm outcomes.

My key identifying assumption is that these trends are unrelated to the changes in peer group composition within clusters used for identification. This assumption would be violated if firms could forsee which clusters would experience productivity shocks or other changes in location fundamentals correlated with productivity increases, and could move directly there. However, this is not so easily done, particularly for manufacturing firms that have significant spatial requirements, and particularly in the Indonesian context. Land markets complicated by dueling systems of property rights make assembling large, contiguous parcels of land a challenge for land developers in Indonesian cities (Harari and Wong, 2025), and such frictions tend to reduce the magnitude of agglomeration effects (Brooks and Lutz, 2016). Additionally, restrictions on foreign land ownership make entry even more complicated for foreign firms.

Beyond these factors, the ability of firms to time their entry is an important consideration. If my assumption that firms are unable to choose their own cluster was violated, then potential entrants would still need to time their entry well if they wanted to take advantage of local productivity shocks. Because of the large fixed costs necessary to open a new factory and the difficulty of assembling parcels of land for new construction in dense urban areas, it is unlikely that potential entrants would be able to coordinate their entry with productivity shocks even if they were able to observe them.

Since firms obtain larger spillovers when their peers are of higher average quality, there is an incentive for firms to try to locate in areas with high-quality peers. Ideally, the same frictions that prevent firms from choosing locations to obtain positive productivity shocks would also prevent firms from choosing their peers, so that we would mostly observe random sorting of firms on quality. However, I will show below that there is evidence for assortative matching between firms on quality, both at the high and low ends of the quality spectrum, indicating that firms are at least somewhat able to sort into clusters with other similar-quality firms. Even so, the inclusion of the firm fixed effects in the regression should account for this sorting, just as the inclusion of the area-year fixed effects accounts for changes in location fundamentals.

The boundaries of the clusters are another important component. By assumption, firms only experience local spillovers from their peer firms within the same cluster, and not from any firms outside of the cluster. This assumption would likely be violated if firms located near cluster boundaries were closer to one another than to the other firms within their respective clusters, which could easily occur if clusters were delineated by area or administrative boundaries. By using density-based clustering with a watershed segmentation algorithm, the areas near cluster boundaries are also the areas of lowest manufacturing density, and so the likelihood of firms being located near boundaries and spilling over onto one another is greatly reduced.

5.2 Estimating equation

My estimation strategy follows Arcidiacono et al. (2012), who develop an algorithmic approach that approximates nonlinear least squares to estimate spillovers from panel data. The estimating equation for domestic firm i is:

$$\ln R_{i,c,k,t} = a_i + \phi_{C,k,t} + \frac{\beta_{\text{Dom}}}{M_{c,t,g(i)} - 1} \left(\sum_{j \in \mathbb{N}_{c,t,g(i) \sim i}} a_j \right) + \frac{\beta_{\text{For}}}{M_{c,t,g'}} \left(\sum_{j \in \mathbb{N}_{c,t,g'}} a_j \right) + \varepsilon_{i,c,k,t} \tag{9}$$

Here, $\phi_{C,k,t}$ is some combination of local area, year, and industry fixed effects; my primary specification includes district-year and industry-year fixed effects. In robustness checks, I also estimate a specification with district-industry-year fixed effects and a specification that uses subdistrict(*kecamatan*) fixed effects rather than district ones, though the resulting standard errors are considerably higher.

a_i is the firm-level fixed effect, which may be decomposed into common and idiosyncratic components $a_i = \bar{\alpha} + \alpha_i$. $\bar{\alpha}$ is a baseline level of quality common to all firms. Because $\frac{1}{M_{c,t,g(i)}-1} \sum_{j \in \mathbb{N}_{c,t,g(i) \sim i}}$ and $\frac{1}{M_{c,t,g'}} \sum_{j \in \mathbb{N}_{c,t,g'}}$ always sum to 1, this component is not directly identifiable, but may be factored out as $\bar{\alpha}(1 + \beta_{\text{Dom}} + \beta_{\text{For}})$. In the estimation procedure, I add this term to the location fundamentals term $\phi_{C,k,t}$. Thus, the estimating equation becomes

$$\ln R_{i,c,k,t} = \alpha_i + \tilde{\phi}_{C,k,t} + \frac{\beta_{\text{Dom}}}{M_{c,t,g(i)} - 1} \left(\sum_{j \in \mathbb{N}_{c,t,g(i) \sim i}} \alpha_j \right) + \frac{\beta_{\text{For}}}{M_{c,t,g'}} \left(\sum_{j \in \mathbb{N}_{c,t,g'}} \alpha_j \right) + \varepsilon_{i,c,k,t} \tag{10}$$

where $\tilde{\phi}_{C,k,t} = \phi_{C,k,t} + \bar{\alpha}(1 + \beta_{\text{Dom}} + \beta_{\text{For}})$.

α_i is the critical component of this analysis - it is a time-invariant, firm-specific measure of firm quality, and the key predictor variable in the model is a linear-in-means aggregate of this object among domestic and foreign peer firms. Identifying variation comes from changes in the within-cluster composition of domestic and foreign firms over time, rather than from changes in the quality or the decisions of peer firms (beyond their decisions to enter or exit). This is also where the use of panel data is key - α_i, which is fixed over time, could not be identified in cross-sectional data.

5.3 Estimation algorithm

Arcidiacono et al. (2012) proved that under linear-in-means weighting schemes, β_{Same} and β_{Diff} are identifiable via non-linear least squares as long as at least one cluster experiences a change in composition over the sample period. Further, if each cluster has one firm with at least two non-missing outcomes over the sample period, all fixed effects may be identified jointly with the betas.

The estimation algorithm has two repeated steps. The first step is to take firm fixed effects α as given, and obtain estimates of β_{Dom}, β_{For}, and $\tilde{\phi}_{C),k,t}$. The second step takes these parameters as given and obtains new estimates of the firm fixed effects using updating equations derived from the nonlinear least squares estimator, which solves:

$$\min_{\alpha_i,\tilde{\phi}_{C,k,t},\beta} \sum_t \sum_i \left[\ln R_{i,c,k,t} - \alpha_i - \tilde{\phi}_{C,k,t} - \frac{\beta_{\text{Dom}}}{M_{c,t,g(i)} - 1} \sum_{j \in \mathbb{N}_{c,t,g\sim i}} \alpha_j - \frac{\beta_{\text{For}}}{M_{c,t,g'}} \sum_{j \in \mathbb{N}_{c,t,g'}} \alpha_j \right]$$

Differentiation with respect to α_i gives the updating equations for α_i. I continue to iterate over these steps until the process converges to the values that minimize the sum of squared errors above - I set this point as the point at which the maximum percentage change in the peer effect elasticities from the previous iteration is 0.0001 percent. To calculate standard errors, I perform a wild bootstrap procedure (Davidson and MacKinnon, 2006) by transforming the residuals from estimation of equation (10) by a random variable with mean 0 and variance 1. For the main specification, I use the Rademacher distribution, which simply takes the values 1 and -1 with equal probability and ensures symmetric distribution of the residuals.[9] After obtaining the transformed residuals, I use them to generate new values of the dependent variable and re-estimate the model. I calculate the standard errors of β_{Dom} and β_{For} as the standard deviation of all converged β bootstrap estimates.

6 Results

My main results are presented in Table 5. Each column corresponds to log revenue under different assumptions. The first column, my preferred specification, presents results when the outcome is log revenue adjusted by $(1 - \theta_k)$. For domestic firms, I find a significant effect of 0.025 for β_{Dom}, and a much smaller and effect of 0.002 for β_{For}. For foreign firms, the coefficient of β_{Dom} is 0.007, and the coefficient on β_{For} is 0.000, but neither result is signficant.

The linear-in-means coefficients can be interpreted as meaning that a doubling of average domestic peer quality within Indonesian urban manufacturing employment clusters leads to a 2.5 percent increase in firm revenue for domestic firms. Similarly, a doubling of average foreign peer quality leads to a 0.2 percent increase in revenue for domestic firms. Neither effect will significantly

[9] Davidson and Flachaire (2008) show that symmetrically-distributed residuals typically perform better in wild bootstrap estimation than asymmetrically-distributed ones.

impact the revenues of foreign firms, at least through the lens of local, exogenous spillovers.

My estimate for domestic-to-domestic spillovers among Indonesian manufacturers is considerably larger than the estimates between Canadian high-skill services firms in Baum-Snow et al. (2024), though my foreign-to-domestic spillovers is closer in magnitude. Compared to the estimates of endogenous spillovers between Indonesian firms in Blalock and Gertler (2008) and Bazzi et al. (2017), my exogenous spillovers estimates are smaller. These results should be understood in the context of the estimation strategy, which focuses only on the recovery of exogenous spillovers - my results cannot rule out the presence of endogenous spillovers, either at very local scales or very broad ones, and therefore does not contradict prior findings that Indonesian manufacturers are able to make large productivity gains by entering supply chain networks with high productivity firms.

That the foreign-to-domestic spillover effect is significantly smaller than the domestic-to-domestic spillover effect is an interesting result. Though foreign firms in the sample tend to be more productive and generate more revenue, these benefits only seem to spill over to their domestic neighbors very slightly. However, being located near a high productivity domestic firm generates sizable benefits for other domestic firms.

My estimates in the second column for domestic-to-domestic spillovers using unadjusted log revenue, 0.027, differs only slightly from my estimate using adjusted revenue. There is also little difference in the foreign-to-domestic estimate. Also, the domestic-to-foreign and foreign-to-foreign estimates remain insignificant. The same is true for my estimates using adjusted log revenue under the assumption of monopolistic competition, which are not statistically different from my adjusted log revenue results. This pattern remains largely true for other dependent variables that are highly correlated with revenue or that have the same structural relationship with total factor productivity through the model.

Given these estimated effects and the local nature of the clusters, I interpret these results as reflecting both knowledge transfers between workers and labor market sharing. Domestic Indonesian firms are likely predominantly staffed by Indonesian workers. Though Indonesia is home to a broad array of ethnicities and languages, Indonesians share a common workplace language in Bahasa Indonesia. Linguistic and cultural differences between foreign and domestic workers may help explain why domestic firms that are able to expropriate gains from their domestic neighbors are not able to do so as easily with their foreign neighbors. Additionally, foreign firms in the sample pay considerably higher wages and employ a larger amount of workers. It may be the case that foreign firms attract high-productivity workers from domestic firms with these higher wages, thereby counteracting whatever positive knowledge spillovers may result from their presence.

7 Assortativity and Matching in the Network of Firms

A key identifying assumption in the empirical work is that firms are unable to enter in ways that allow them to take advantage of changes in location fundamentals correlated with productivity. Ideally, this assumption would also account for firm sorting on quality. If firm sorting into locations were purely random, I would expect to find little relationship between the quality of any given firm and the quality of its peers. However, if firms are able to engage in assortative matching, then I should frequently observe high-quality firms locating in clusters with other high-quality firms. In this section, I will present evidence on the extent of such assortative matching on quality between firms.

Figure 2 presents a binned scatterplot of firm-level quality (the firm fixed effect, α) and the highest level of quality in the firm's cluster. Figure 3 presents the same by ownership status. In each case, I exclude the firms with the highest quality in each cluster. In each case, there appears to be a clear positive correlation between firm quality and quality of the the firm with the highest quality in each cluster. Some degree of positive correlation is to be expected here, since no included firm will have a maximum peer quality lower than its own quality, but the degree to which high-quality firms are able to match to even higher-quality firms appears to be substantial, both overall and between firms of different ownership types. At the lower tail, it appears that lower-quality domestic firms are generally more able to match with medium-quality peers than lower-quality foreign firms.

To explore further, I turn to measures of homophily from graph theory. I treat the universe of Indonesian manufacturing firms as a single network, and clusters as groups in which all of the firms are connected via "edges" to other firms in the same cluster. I then calculate the Pearson correlation coefficient for the alphas across the edges, also known as the numeric assortativity coefficient (Newman, 2003; Hagberg et al., 2008). The numeric assortativity coefficient takes values between -1 and 1, where -1 corresponds to a maximally dissasortive network in which firms only connect to peers of different quality, and 1 corresponds to a maximally assortative network in which firms only connect to peers of the same quality. I find a numeric assortativity coefficient of 0.5011, indicating fairly strong assortativity on quality across Indonesian manufacturing firms.

The assortativity coefficient might be driven by high-quality firms matching with other high-quality firms, low-quality firms matching with other low-quality firms, or assortative matching across the firm quality spectrum. It would be nice to know something about assortativity at the local or cluster level. Unfortunately, assortativity coefficients are not well-behaved for complete networks in which every node is connected to every other node (Mironov and Prokhorenkova, 2024), as is the case within my clusters of manufacturers. Instead, I partition the sample into high- and low-quality firms, where a firm is assigned to the high-quality group if it has above-median quality. I then construct a measure of how dissimilar the distribution of high- and low-quality firms within a cluster is to the population distribution (which is naturally 50-50). In particular, I calculate

$$r_c = \frac{N_{h,c}}{N_c} - \frac{N_{l,c}}{N_c} \tag{11}$$

where $N_{h,c}$ and $N_{l,c}$ are the number of high-quality and low-quality firms in cluster c, respectively, and N_c is the total number of firms in c. r_c, then, takes values from -1 to 1, where a cluster with an r_c of 0 would have an even split of high- and low-quality firms, as in the sample at large, and a cluster with an r_c of 1 would have only high-quality firms.

Figure 4 presents a binned scatter plot of r_c as a percentage against the maximum firm quality in each cluster. An estimated local polynomial regression of degree 3 is overlaid. This figure sheds some light on the forces driving the strong assortativity coefficient I obtained for Indonesia at large. At the upper tail of r_c, where the proportion of high-quality firms approaches 100%, the corresponding maximum cluster-level firm qualities are generally high, and at the lower tail of r_c, the corresponding maximum cluster-level firm qualities are generally low.

The reason for co-location among high-quality firms is fairly straightforward to understand. In the model, firms have more to gain from higher-quality peers, so there are natural incentives for high-quality firms to seek one another out. More surprising is that I also observe evidence of assortative matching between low-quality firms, who do not have incentives through the model to cluster together. This may be due to co-location among high-quality firms leading to higher land rents that in turn drive lower-quality firms (which tend to also be lower-revenue firms) to locate in less expensive locations.

8 Policy Simulations

In this section, I conceptualize two potential policy interventions and their effects on firms at the local level. To begin, I calculate firm-level total factor productivity under the assumption of industry-specific variable input shares. I then use these productivity measures to explore a policy intervention that directly increases the productivity of a single firm, and another that explores the local spillovers effects of the entry of a highly-productive foreign firm. For each exercise, I restrict my attention to 2009, the year with the most firms in the sample, though my productivity calculations are done for the full sample.

8.1 Calculation of TFP

I begin by constructing a measure of residual TFP using the model. As in the main specification, I relax the assumption that all firms in the sample share the same variable input share and assume that firms within the same 2-digit ISIC industry face the same output elasticities and variable input shares. Then, for each industry k, input factor $f \in (L, K, B, M)$ (labor, capital, buildings, and materials), and year t, I calculate the output elasticity with respect to factor f in year t, $\theta^f_{k,t}$ as an aggregate of payments to f within industry k divided by the aggregate of total costs within industry k. I average the industry-level output elasticities over the years to obtain θ^f_k. I then calculate residual log TFP for each firm as:

$$\ln A_{i,c,k,t} = \ln R_{i,c,k,t} - \theta^L_k \ln L_{i,c,k,t} - \theta^K_k \ln K_{i,c,k,t} - \theta^B_k \ln B_{i,c,k,t} - \theta^M_k \ln M_{i,c,k,t}$$

In robustness checks, I also estimate yearly firm-level total factor productivity using control function methods (Wooldridge, 2009; Ackerberg et al., 2015) for use in these exercises. The results I obtain using control-function productivity estimates are quite similar to those I obtain using residual total factor productivity as calculated above.

8.2 Exogenous Productivity Increase of One Firm

In the first policy simulation exercise, I construct a scenario in which the Indonesian government is able to exogenously increase the productivity of a single firm in the sample by 10 percent (perhaps by hiring a management consultancy firm), with the goal of creating the largest possible gains to aggregate productivity. Which firms or areas would make the best targets? For each firm, I simulate a 10% increase in that firm's productivity and calculate the changes in aggregate and average productivity within that firm's peer group, using the spillover parameter estimates from my primary results.

Figure 5 and Figure 6 display two heat maps resulting from this exercise, aggregated from the firm level to subdistrict-level averages. For ease of viewing, I restrict the sample to Java - the trends throughout Indonesia are similar.[10] Figure 5 displays the subdistrict-level average increase to peer firm productivity from a 10% increase in firm-level productivity, while Figure 6 displays the subdistrict-level average of the aggregate increase in peer firm productivity within a cluster.

Taken together, the two maps reveal an interesting trend - Central urban areas unsurprisingly accrue the highest aggregate gains (as more firms naturally locate in more populous areas), with the typical cluster in these areas experiencing aggregate log productivity gains of 0.01 to 0.015. However, the areas with the highest average gains are typically city outskirts, where firms make average log productivity gains of up to 0.005 (one-third to one-half of the aggregate gains made in city center clusters) all on their own. Figure 7 plots the relationship between cluster-level average productivity spillovers and the number of firms in the village, and Figure 8 does the same using the number of workers in the cluster. Each plot includes estimated local polynomial regression lines, with 95% confidence intervals in grey bands around each line. Though the relationship in

[10]See Appendix A for Indonesia-wide images.

Figure 8 is noisier, in each case there is a clear negative trend, reinforcing the takeaway that firms in relatively less dense areas are able to make larger gains on average when the average productivity of their neighbors increases.

8.3 Entry of a Highly-Productive Foreign Firm

In the second policy simulation exercise, inspired by developing country efforts to attract foreign direct investment in particular, I examine the potential effects of a highly-productive foreign firm on local domestic firm productivity. For each cluster, I simulate the entry of a foreign firm with a total factor productivity of one standard deviation greater than the mean productivity for foreign firms. Then, I calculate the change in average spillovers from foreign to domestic firms in each cluster by multiplying the average productivity of foreign firms by my estimated elasticity of domestic productivity to average foreign peer productivity, 0.002.

The results of this exercised are summarized for the island of Java by Figure 9. As with the previous exercise, it seems that the areas furthest from the urban core have the most to gain from the entry of a new, highly-productive foreign peer, while areas nearer to the urban core make little gains. Indeed, in some cases where the average productivity of foreign firms is already very high, the spillover effect of a new foreign entry is actually negative, since it reduces the average foreign productivity.

For Indonesia at large, the areas that gain the most from foreign entry are (perhaps unsurprisingly) the areas that have no foreign presence to start with - these areas are mostly on the outer islands like Kalimantan and Sulawesi. Outside of these areas, areas with few foreign firms generally have the greatest opportunities to gain from foreign entry - the low concentration of foreign firms means that a new foreign entry potentially boosts the local average foreign productivity considerably. In dense urban areas with many manufacturing firms, the potential gains are minimal. These denser areas also tend to hold many high-productivity foreign firms, so the effect of new highly-productive foreign entry is minimized both by the high number of foreign firms and the high average foreign productivity that these areas begin with.

9 Conclusion

This paper provides new estimates of heterogeneous productivity spillovers at local scales. When firms are grouped by ownership status and spillovers are allowed to differ by the ownership status of the target and peer firms, spillovers are mostly found to occur between domestic firms. Domestic firms do benefit from the presence of higher productivity peers of either type, though they benefit much more from highly-productive domestic peers than highly-productive foreign ones. Foreign firms do not expropriate local productivity spillovers from peer firms of either type on average.

These results are remarkably robust to alternative to alternative specifications, with highly similar results occurring when variable input shares are not allowed to vary, or when using a monopolistic competition framework.

There is strong evidence for assortative matching between firms on firm-level quality. High-quality firms tend to sort into areas with other high-quality firms - these areas also tend to be the most dense areas in terms of number of manufacturing workers and number of firms. Low-quality firms also tend to match with one another, often in less dense areas. As a result, firms in these less dense areas have the most to gain from an increase in the productivity of their peers, whether that boost comes from an exogenous productivity increase to one peer or from the entry of a new, highly-productive foreign firm. On the other hand, firms in high-productivity, dense areas have little to gain from a boost in peer productivity.

Understanding the way that spillovers function between firms is an important step in understanding the mechanisms of growth in developing countries. Location matters for integration into supply chains, market access, and labor pooling. This paper sheds some light on the role that foreign investment and industrial agglomeration play in development at the local level, and provides evidence of the importance of considering heterogeneity in ownership type when thinking about local spillovers. Though foreign firms tend to employ more workers, pay higher wages, and have higher levels of productivity, their presence is less impactful to their domestic peers than the presence of highly-productive domestic firms.

References

Abraham, F., J. Konings, and V. Slootmaekers (2010): "FDI spillovers in the Chinese manufacturing sector: Evidence of firm heterogeneity 1," *Economics of Transition*, 18, 143–182.

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

Aitken, B. J. and A. E. Harrison (1999): "Do domestic firms benefit from direct foreign investment? Evidence from Venezuela," *American economic review*, 89, 605–618.

Amiti, M. and L. Cameron (2007): "Economic geography and wages," *The Review of Economics and Statistics*, 89, 15–29.

Amiti, M., C. Duprez, J. Konings, and J. Van Reenen (2024): "FDI and superstar spillovers: Evidence from firm-to-firm transactions," *Journal of International Economics*, 152, 103972.

Arcidiacono, P., G. Foster, N. Goodpaster, and J. Kinsler (2012): "Estimating spillovers using panel data, with an application to the classroom," *Quantitative Economics*, 3, 421–470.

Baddeley, A., E. Rubak, and R. Turner (2015): *Spatial Point Patterns: Methodology and Applications with R*, Chapman and Hall/CRC Press.

Baum-Snow, N., N. Gendron-Carrier, and R. Pavan (2024): "Local productivity spillovers," *American Economic Review*, 114, 1030–1069.

Bazzi, S., A. V. Chari, S. Nataraj, and A. D. Rothenberg (2017): "Identifying productivity spillovers using the structure of production networks," .

Blalock, G. and P. J. Gertler (2008): "Welfare gains from foreign direct investment through technology transfer to local suppliers," *Journal of international Economics*, 74, 402–421.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of econometrics*, 150, 41–55.

Brooks, L. and B. Lutz (2016): "From today's city to tomorrow's city: An empirical investigation of urban land assembly," *American Economic Journal: Economic Policy*, 8, 69–105.

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

Caves, R. E. (1974): "Multinational firms, competition, and productivity in host-country markets," *Economica*, 41, 176–193.

Civelli, A., A. Gaduh, A. D. Rothenberg, and Y. Wang (2023): "Urban sprawl and social capital: Evidence from Indonesian cities," *The Economic Journal*, 133, 2110–2146.

Davidson, R. and E. Flachaire (2008): "The wild bootstrap, tamed at last," *Journal of Econometrics*, 146, 162–169.

Davidson, R. and J. G. MacKinnon (2006): "Bootstrap methods in econometrics," .

De Loecker, J. (2011): "Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity," *Econometrica*, 79, 1407–1451.

Ellison, G., E. L. Glaeser, and W. R. Kerr (2010): "What causes industry agglomeration? Evidence from coagglomeration patterns," *American Economic Review*, 100, 1195–1213.

Fallick, B., C. A. Fleischman, and J. B. Rebitzer (2006): "Job-hopping in Silicon Valley: Some evidence concerning the microfoundations of a high-technology cluster," *The review of economics and statistics*, 88, 472–481.

Freedman, M. L. (2008): "Job hopping, earnings dynamics, and industrial agglomeration in the software publishing industry," *Journal of Urban Economics*, 64, 590–600.

Hagberg, A., P. J. Swart, and D. A. Schult (2008): "Exploring network structure, dynamics, and function using NetworkX," Tech. rep., Los Alamos National Laboratory (LANL), Los Alamos, NM (United States).

Harari, M. and M. Wong (2025): "Colonial legacy and land market formality," *Journal of Urban Economics*, 149, 103789.

Ibrahim, S. E., M. H. Fallah, and R. R. Reilly (2009): "Localized sources of knowledge and the effect of knowledge spillovers: an empirical study of inventors in the telecommunications industry," *Journal of Economic Geography*, 9, 405–431.

Iyoha, E. (2023): "Estimating Productivity in the Presence of Spillovers: Firm-level Evidence from the US Production Network," Working paper.

Javorcik, B. S. (2004): "Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages," *American economic review*, 94, 605–627.

Keller, W. and S. R. Yeaple (2009): "Multinational enterprises, international trade, and productivity growth: firm-level evidence from the United States," *The review of economics and statistics*, 91, 821–831.

Kloosterman, R. C. (2008): "Walls and bridges: knowledge spillover between 'superdutch'architectural firms," *Journal of Economic Geography*, 8, 545–563.

Konings, J. (2001): "The effects of foreign direct investment on domestic firms: Evidence from firm-level panel data in emerging economies," *Economics of transition*, 9, 619–633.

Kusno, A. (2017): "Power and time turning: The capital, the state, and the kampung in Jakarta," in *Cities and Power*, Routledge, 63–73.

Lee, S. and S. H. Lee (2024): "Cities as Inverted Watersheds," Working paper.

Levinsohn, J. and A. Petrin (2003): "Estimating production functions using inputs to control for unobservables," *The review of economic studies*, 70, 317–341.

Lin, M. and Y. K. Kwan (2016): "FDI technology spillovers, geography, and spatial diffusion," *International Review of Economics & Finance*, 43, 257–274.

Manski, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, 60, 531–542.

Marshall, A. (1890): *The Principles of Economics*, McMaster University Archive for the History of Economic Thought.

Mironov, M. and L. Prokhorenkova (2024): "Revisiting graph homophily measures," *arXiv preprint arXiv:2412.09663*.

Monastiriotis, V. and J. A. Jordaan (2010): "Does FDI promote regional development? Evidence from local and regional productivity spillovers in Greece," *Eastern Journal of European Studies*, 1, 139.

Newman, M. E. (2003): "Mixing patterns in networks," *Physical review E*, 67, 026126.

O'Callaghan, J. F. and D. M. Mark (1984): "The extraction of drainage networks from digital elevation data," *Computer vision, graphics, and image processing*, 28, 323–344.

OECD (1996): "OECD Benchmark Definition of Foreign Direct Investment," .

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

Poczter, S., P. Gertler, and A. D. Rothenberg (2014): "Financial crisis and productivity evolution: Evidence from Indonesia," *The World Economy*, 37, 705–731.

Rothenberg, A. D. and D. Temenggung (2019): *Place-based policies in Indonesia: A critical review*, World Bank.

Rothenberg, A. D., Y. Wang, and A. V. Chari (2025): "When Regional Policies Fail: An Evaluation of Indonesia's Integrated Economic Development Zones," *Journal of Development Economics*, 176.

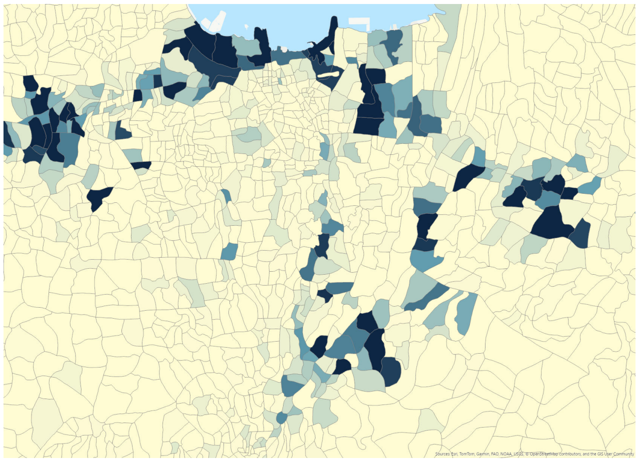
Serpa, J. C. and H. Krishnan (2018): "The impact of supply chains on firm-level productivity," *Management Science*, 64, 511–532.

Silverman, B. (1986): "Density estimation for statistics and data analysis," *Monographs on Statistics and Applied Probability.*

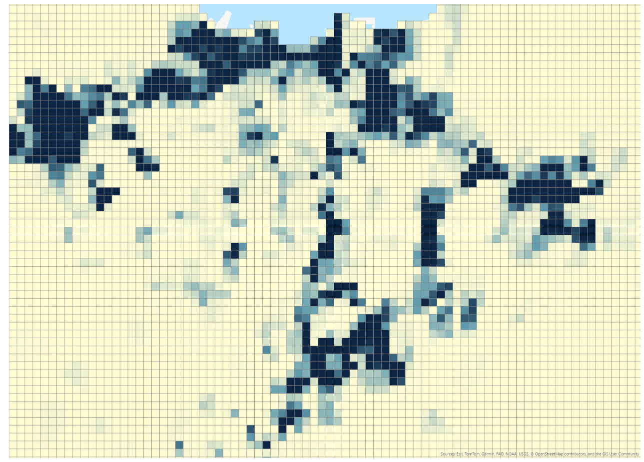
Tanaka, K. and Y. Hashiguchi (2015): "Spatial spillovers from foreign direct investment: Evidence from the Yangtze River Delta in China," *China & World Economy*, 23, 40–60.

Wooldridge, J. M. (2009): "On estimating firm-level production functions using proxy variables to control for unobservables," *Economics letters*, 104, 112–114.

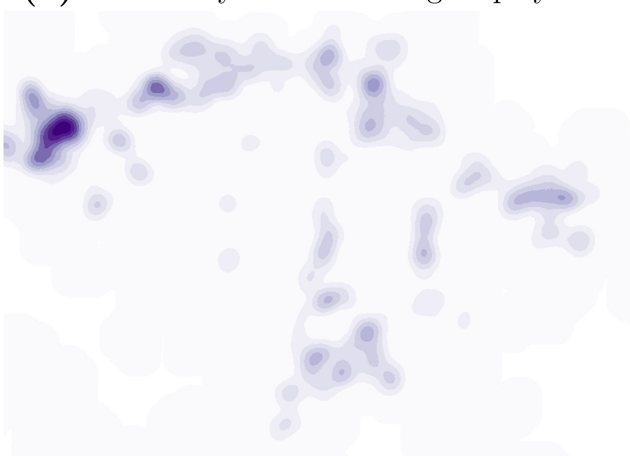
Figure 1: Construction of manufacturing employment watersheds, Jakarta



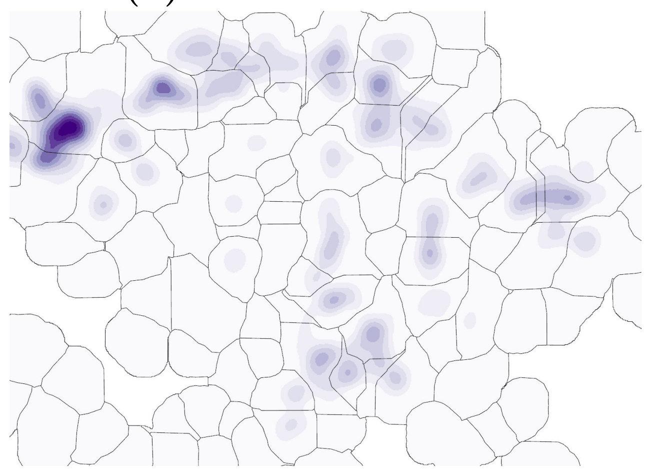
(a) Jakarta by manufacturing employment



(b) 1km × 1km raster cells



(c) Kernel smoothing (Gaussian kernel)



(d) Calculated employment watersheds

Notes: This figure displays the process for delineating watersheds of manufacturing employment in Indonesia, using the Jakarta metropolitan area as an example.

Table 1: Fixed-effects Least Squares: Exploring Foreign Status

Dependent variable:	Log Wage Rate (1)	Log Output (2)	Log Revenue (3)	TFP (4)
Foreign-owned (0 1)	0.663	1.946	1.968	0.575
	(0.010)	(0.012)	(0.012)	(0.020)
Year FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
N firm-year obs	186,228	281,097	280,120	280,507

Notes: This table presents estimates of a fixed-effect least squares regression of selected firm-level characteristics on foreign ownership status. Standard errors in parentheses. City and year fixed effects are included.

Table 2: Callaway-Sant'Anna method; ATT on characteristics of domestic firms, full sample

Industry	Log Revenue (1)	Log Output (2)	Log Labor (3)	Log TFP (4)
ATT	-0.007	-0.008	0.011	0.127
	(0.041)	(0.039)	(0.017)	(0.086)
Firm FE	Yes	Yes	Yes	Yes
Cluster FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes

Notes: This table presents estimates of the average treatment effect on the treated from a staggered differences-in-differences specification, where the dependent variables are characteristics of Indonesian domestically-owned firms. Treatment occurs when a domestic firm goes from having no foreign peers to having at least one foreign peer. Firms that are treated pre-panel are excluded, as are firms that that go from treated to untreated over the sample period. Standard errors in parentheses.

Table 3: Callaway-Sant'Anna method; ATT on characteristics of domestic firms, by industry

Industry	Log Revenue (1)	Log Output (2)	Log Labor (3)	Log TFP (4)
Food processing	0.141	0.115	0.051	0.075
	(0.081)	(0.074)	(0.034)	(0.119)
Textiles	-0.064	-0.057	-0.055	0.212
	(0.097)	(0.098)	(0.051)	(0.278)
Furniture	-0.089	-0.066	0.024	0.323
	(0.079)	(0.079)	(0.038)	(0.124)
Paper	0.102	0.059	0.140	-0.233
	(0.166)	(0.162)	(0.073)	(0.471)
Chemicals	-0.131	-0.133	0.026	-0.078
	(0.104)	(0.098)	(0.044)	(0.121)
Ceramics	0.082	0.079	0.038	-0.461
	(0.136)	(0.135)	(0.063)	(0.548)
Metalworking	-1.281	0.103	0.207	0.458
	(0.140)	(0.031)	(0.002)	(0.182)
Electronics	-0.092	-0.077	-0.039	-0.048
	(0.163)	(0.158)	(0.059)	(0.361)
Firm FE	Yes	Yes	Yes	Yes
Cluster FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes

Notes: This table presents estimates of the average treatment effect on the treated from a staggered differences-in-differences specification, where the dependent variables are characteristics of Indonesian domestically-owned firms. Results are presented for each 2-digit ISIC manufacturing industry. Treatment occurs when a domestic firm goes from having no foreign peers to having at least one foreign peer. Firms that are treated pre-panel are excluded, as are firms that that go from treated to untreated over the sample period. Standard errors in parentheses.

Table 4: Cluster panel analysis of foreign presence - fixed effects least squares

Dependent variable	(1)
Log N domestic firms	1.433
	(0.022)
Pct. Δ domestic firms	0.021
	(0.001)
Log N new domestic entrants	0.391
	(0.032)
Log N domestic exits	0.376
	(0.028)
Year FE	Yes
City FE	Yes

Notes: This table presents results from the estimation of several local-area level fixed-effects least squares regressions, controlling for year and city-level fixed effects. Standard errors in parentheses.

Table 5: Main Results: Spillovers by Firm Ownership Type

Domestic firm results	Log Adj. Revenue (Perf. Comp.) (1)	Log Revenue (2)	Log Adj. Revenue (Monop. Comp.) (3)
Average domestic peer FE	0.025	0.027	0.028
	(0.010)	(0.013)	(0.007)
Average foreign peer FE	0.002	0.005	0.003
	(0.002)	(0.003)	(0.002)
Foreign firm results			
Average domestic peer FE	0.007	0.007	0.008
	(0.015)	(0.017)	(0.009)
Average foreign peer FE	0.000	0.005	0.001
	(0.003)	(0.008)	(0.004)
Firm FE	Yes	Yes	Yes
Industry-year FE	Yes	Yes	Yes
Area-year FE	Yes	Yes	Yes
N firm-year obs	219,184	219,283	200,118

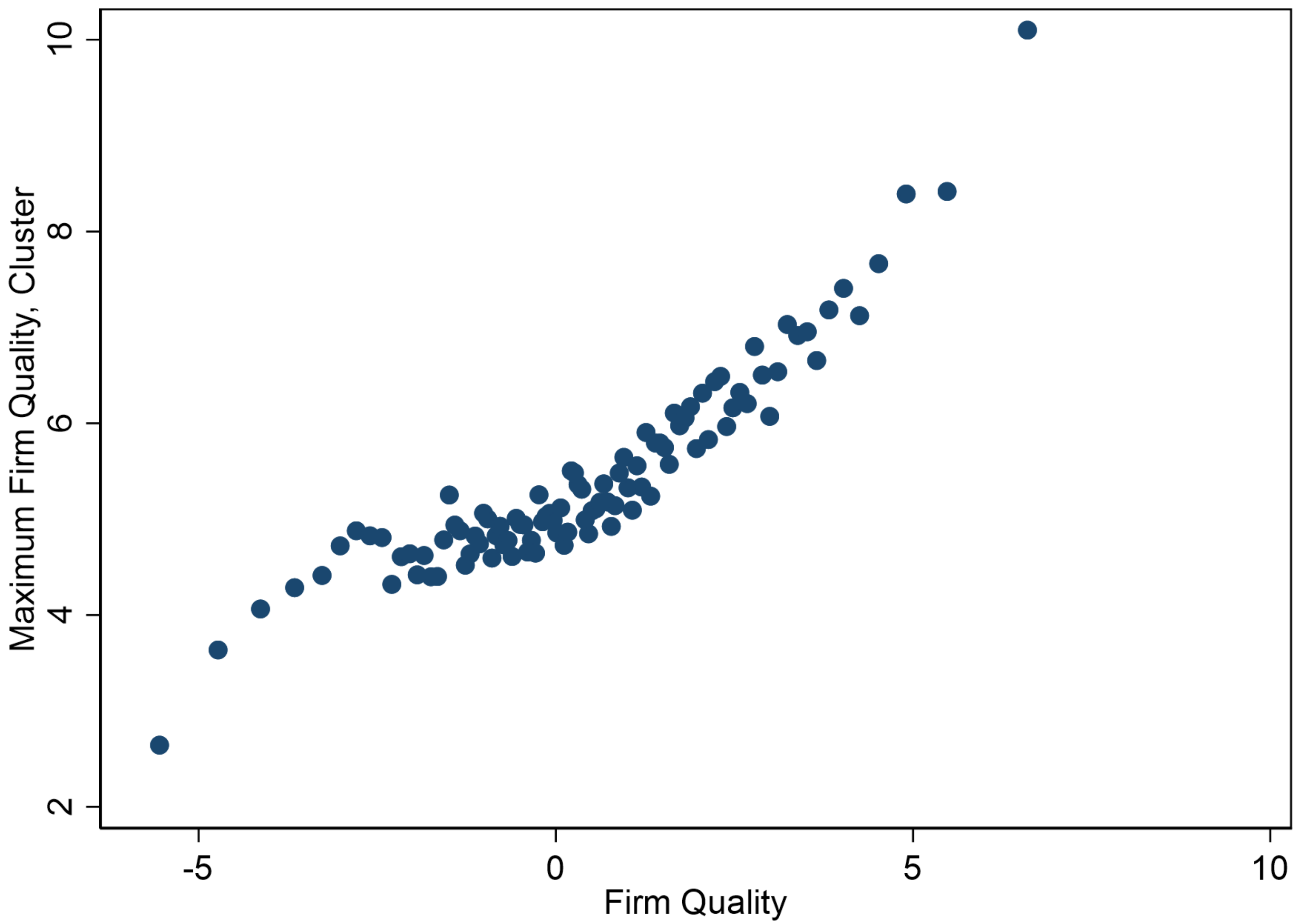
Notes: This table presents results from estimation of the primary estimating equation (10). Bootstrap standard errors are in parentheses. To reduce calculation time, 100 bootstrap replications are performed for the primary specification only - all other standard errors are calculated with 10 bootstrap replications.

Table 6: Additional Results: Spillovers by Firm Ownership Type

Log Output **Domestic firm results**	Log Output (1)	Log Value Added (2)	Log Output/Worker (3)
Average domestic peer FE	0.028	0.026	0.032
	(0.007)	(0.006)	(0.010)
Average foreign peer FE	0.004	0.002	0.007
	(0.003)	(0.002)	(0.003)
Foreign firm results			
Average domestic peer FE	0.008	0.011	0.003
	(0.011)	(0.012)	(0.004)
Average foreign peer FE	0.003	0.004	0.007
	(0.007)	(0.007)	(0.010)
Firm FE	Yes	Yes	Yes
Industry-year FE	Yes	Yes	Yes
Area-year FE	Yes	Yes	Yes
N firm-year obs	219,688	219,658	219,267

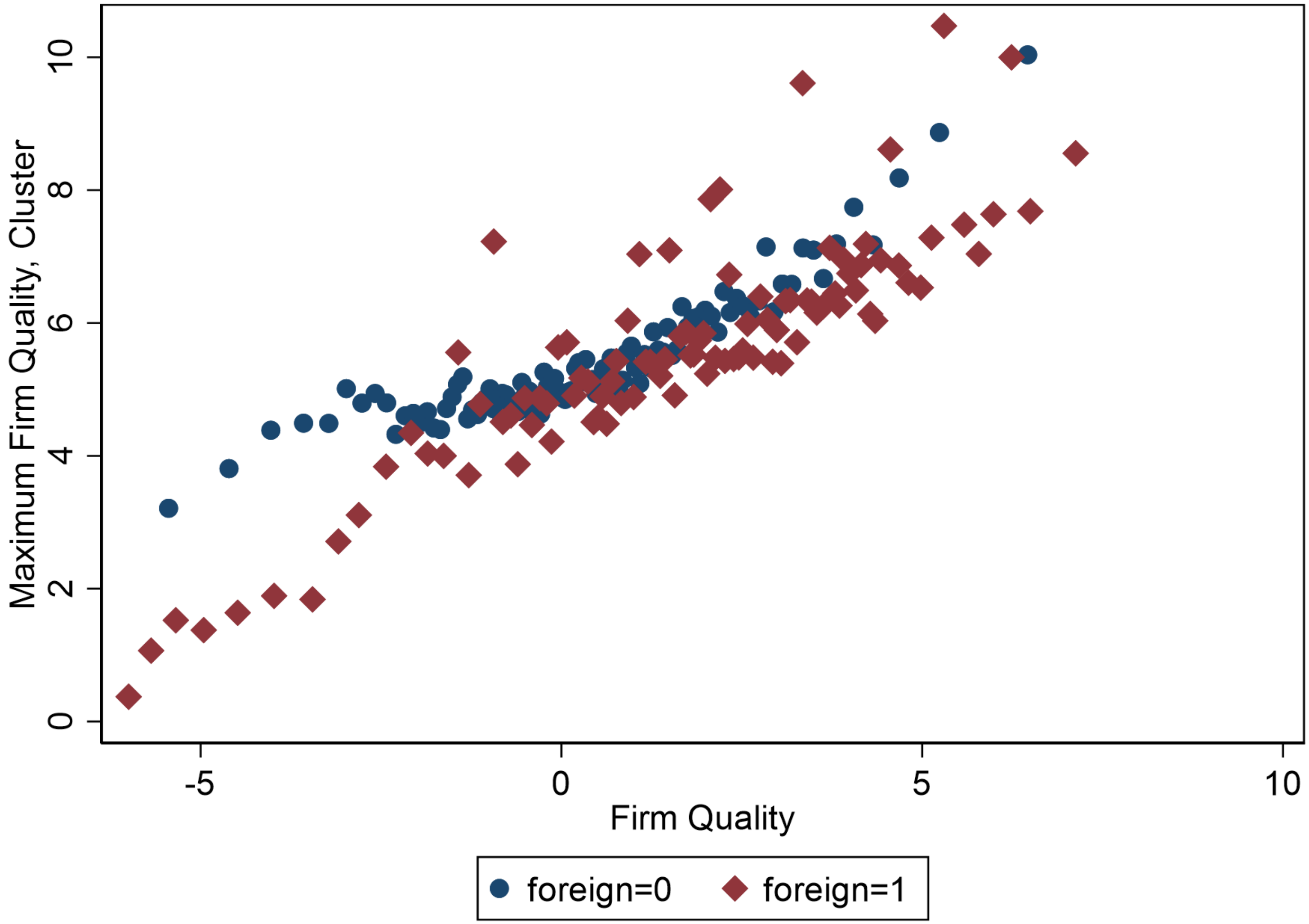
Notes: This table presents results from estimation of the primary estimating equation (10) on additional dependent variables. Bootstrap standard errors are in parentheses. All other standard errors are calculated with 10 bootstrap replications.

Figure 2: Firm-level quality (alpha) versus maximum firm quality, cluster



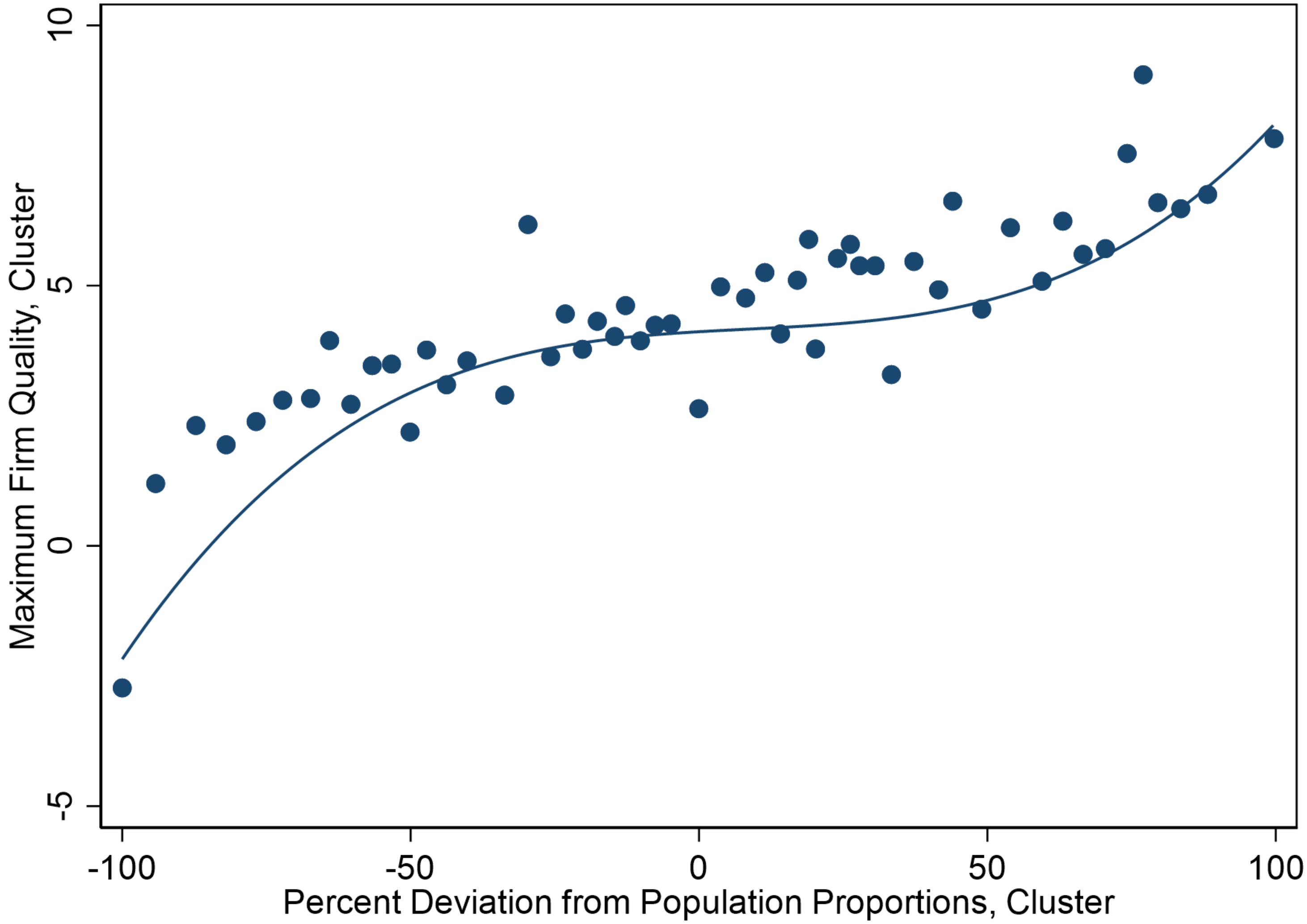
Notes: This figure displays a binned scatterplot of the firm-level quality measure α from the model against the maximum level of α within each firm's cluster. For each cluster, the firm with the maximum level of quality is excluded.

Figure 3: Firm-level quality (alpha) versus maximum firm quality, cluster, by foreign status



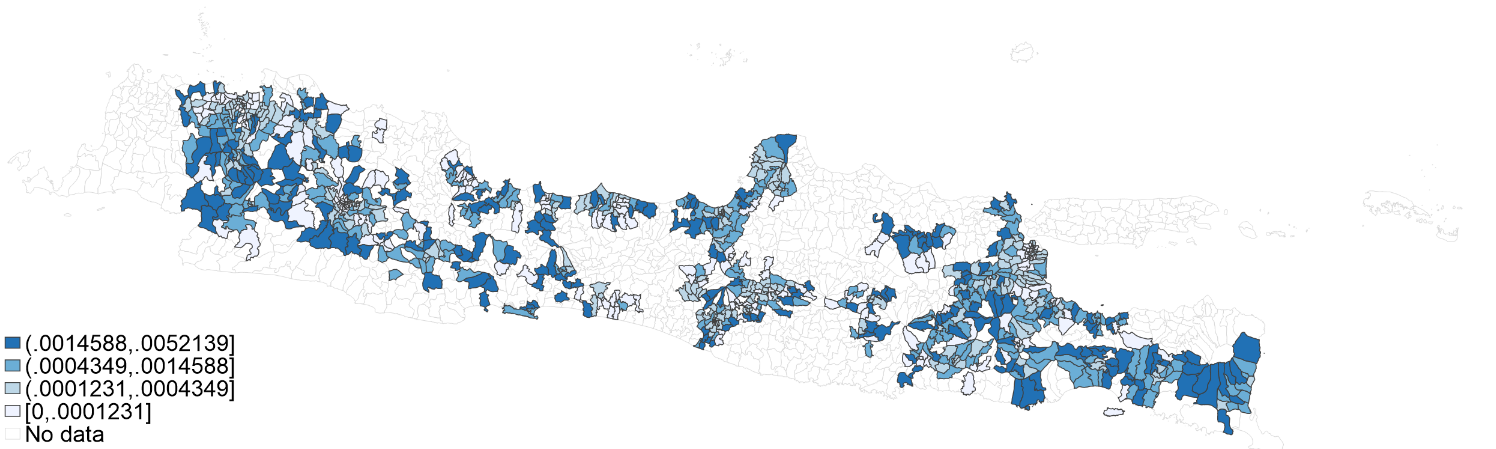
Notes: This figure displays a binned scatterplot of the firm-level quality measure α from the model against the maximum level of α within each firm's cluster, by foreign ownership status. Domestically-owned firms are blue circles, and foreign-owned firms are red diamonds. For each cluster, the firm with the maximum level of quality is excluded.

Figure 4: Percent Deviation from Population Proportion versus maximum firm quality, cluster



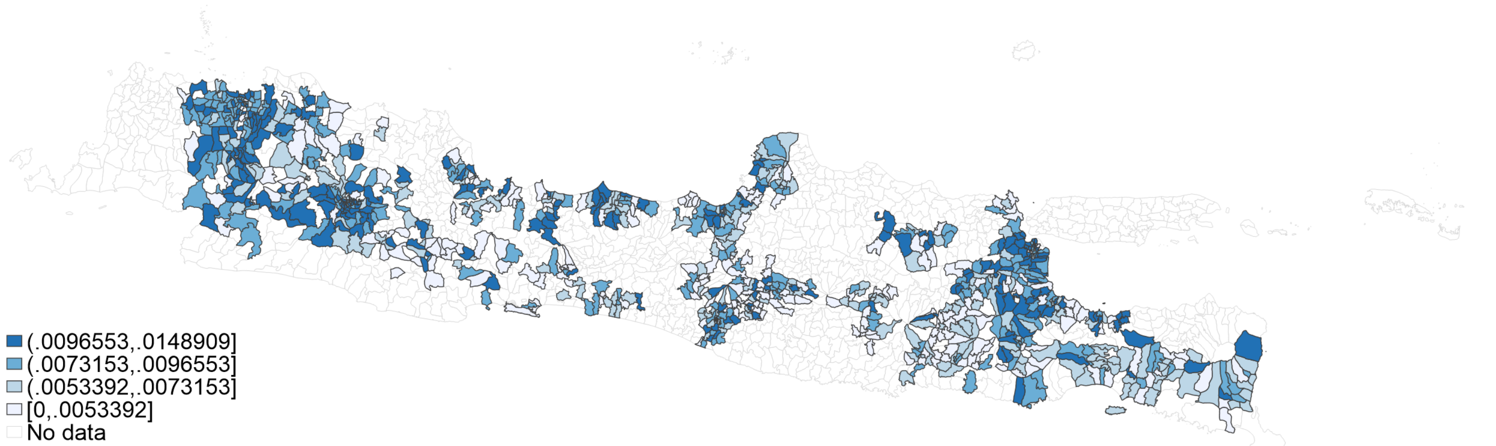
Notes: This figure displays a binned scatterplot of the cluster-level dissimilarity measure r_c against the maximum level of α within cluster. A local polynomial line of degree three is overlaid.

Figure 5: Subdistrict-level averaged average gains to peer productivity, Java



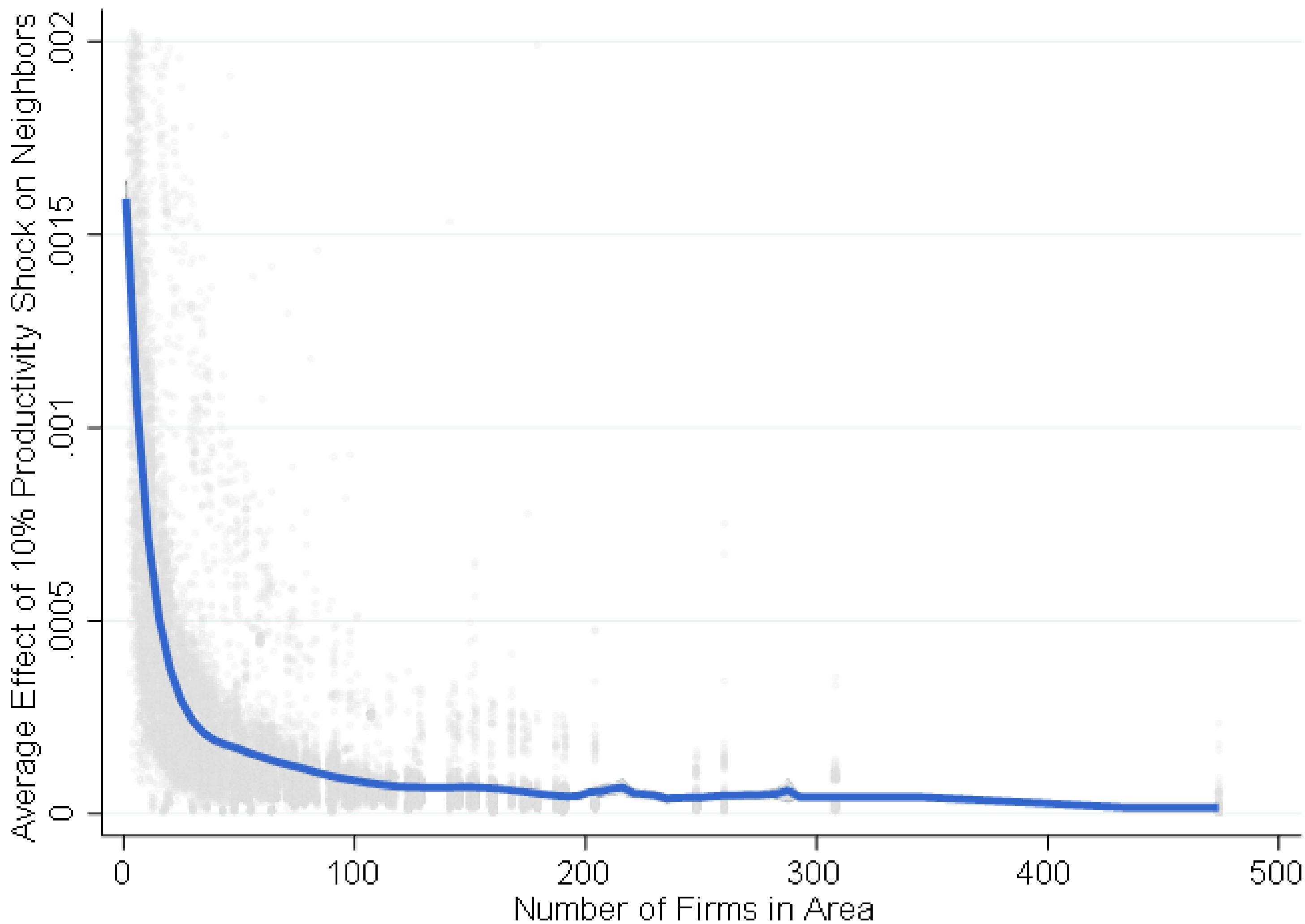
Notes: This figure displays Java's urban subdistricts (*kecamatan*) shaded by the subdistrict-level average of the cluster-level average gains to peer productivity from the first policy experiment.

Figure 6: Subdistrict-level averaged aggregate gains to peer productivity, Java



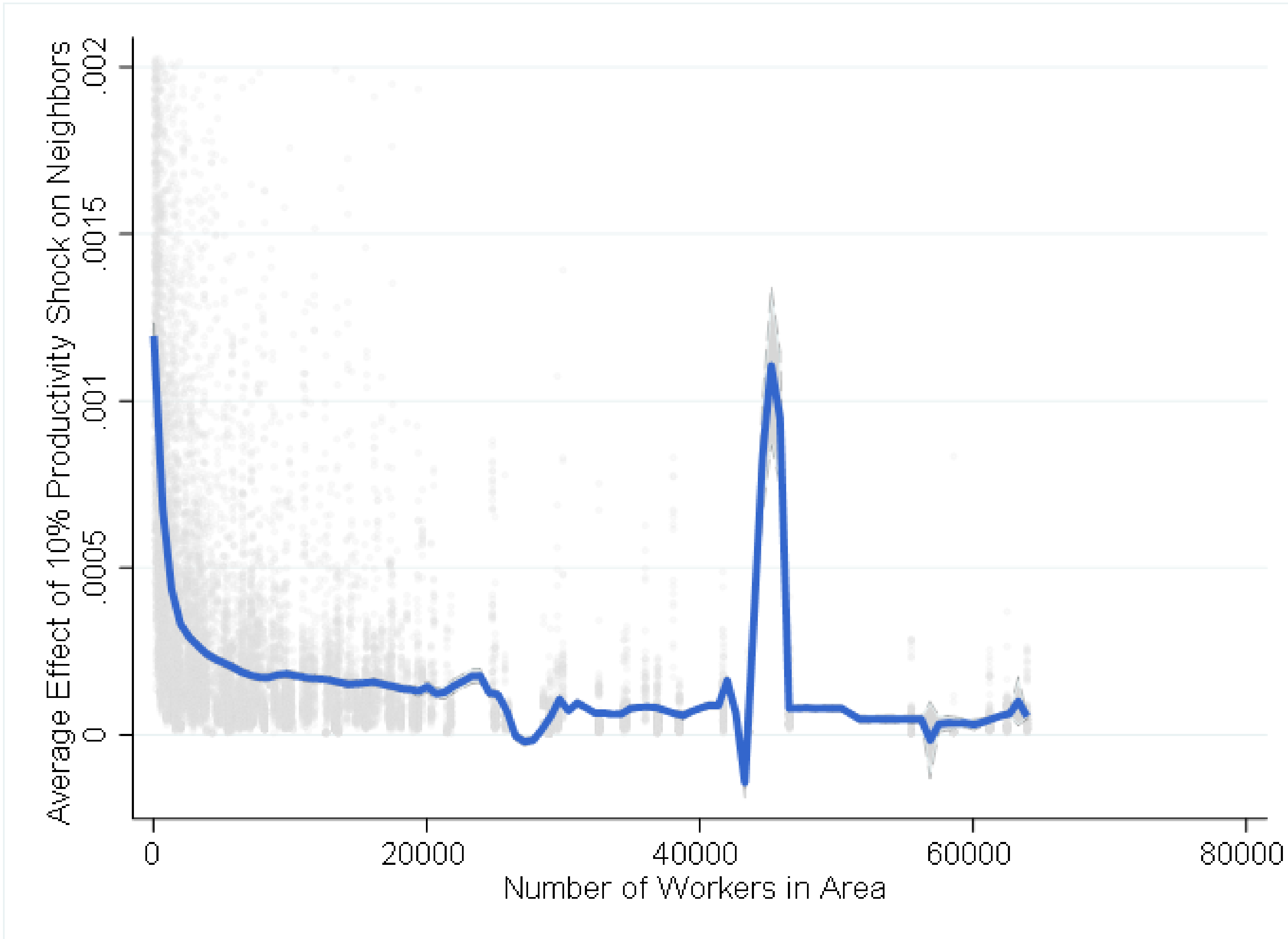
Notes: This figure displays Java's urban subdistricts (*kecamatan*) shaded by the subdistrict-level average of the cluster-level aggregate gains to peer productivity from the first policy experiment.

Figure 7: Cluster-level average gains to peer productivity versus number of firms in cluster



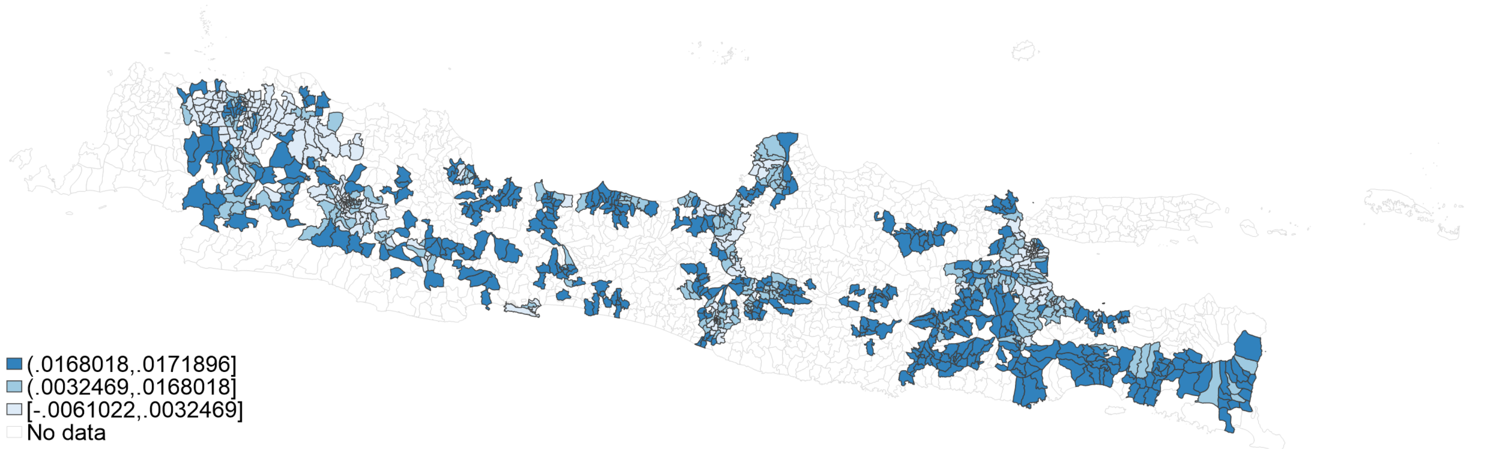
Notes: This figure displays a scatterplot of the cluster-level average gains to peer productivity from a 10% shock to a the productivity of a single firm versus the number of firms in the cluster. A local polynomial line of degree three is overlaid.

Figure 8: Cluster-level average gains to peer productivity versus number of workers in cluster



Notes: This figure displays a scatterplot of the cluster-level average gains to peer productivity from a 10% shock to a the productivity of a single firm versus the number of workers in the cluster. A local polynomial line of degree three is overlaid.

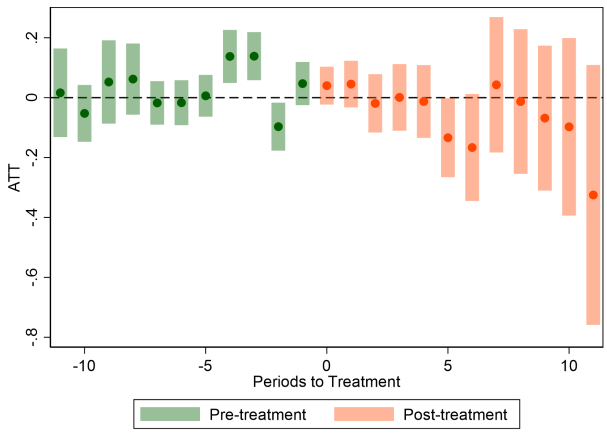
Figure 9: Subdistrict-level averaged gains to peer productivity from additional foreign firm, Java



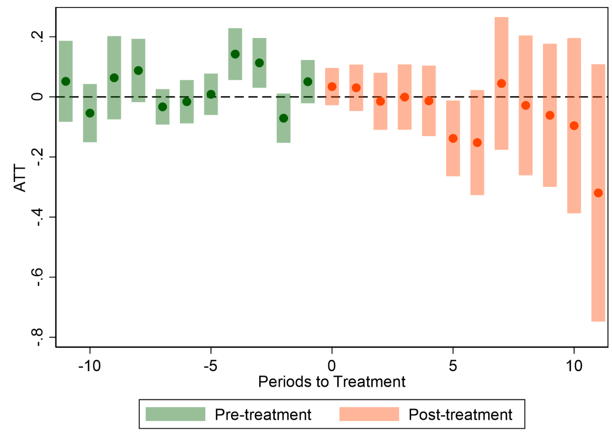
Notes: This figure displays Java's urban subdistricts (*kecamatan*) shaded by the subdistrict-level average of the cluster-level average gains to peer productivity from the second policy experiment.

A Additional tables and figures

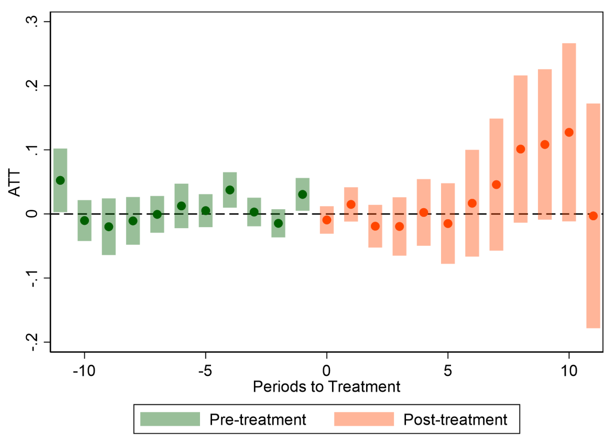
Figure A1: Event Study Plots, Callaway-Sant'Anna Method



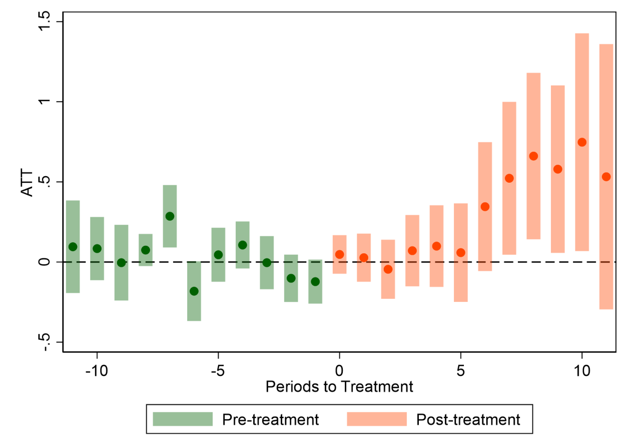
(a) Log Revenue



(b) Log Output



(c) Log Labor



(d) TFP

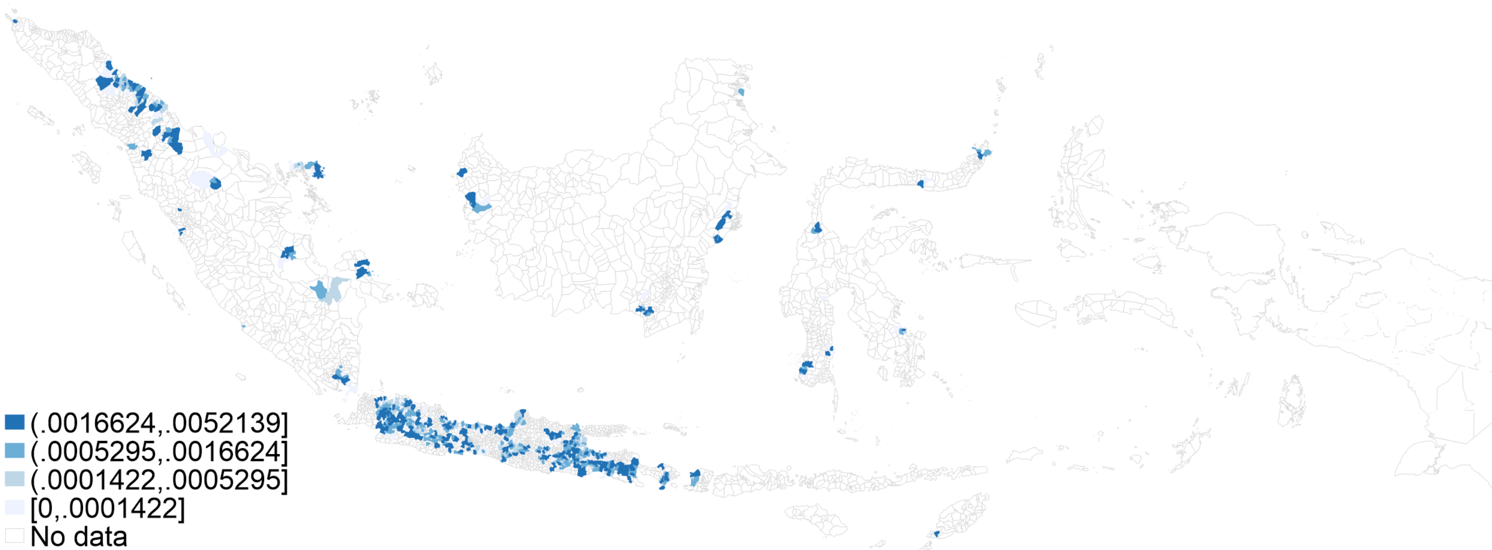
Notes: This figure several event study plots using the Callaway-Sant'Anna multiperiod differences-in-differences method. Point estimates are in circles, and standard standard errors are shaded bars. Green bars and circles are pre-treatment, and orange bars and circles are post-treatment. The dependent variables are listed under each sub-figure, and are measured at the firm level. Treatment is the entry of a foreign firm in a cluster where there previously were none. The dependent variables are only measured for domestic firms.

Table A.1: Robustness: Spillovers by ownership type, *kecamatan*-level fixed effects

Outcome	**Log Revenue**
Domestic Firms	(1)
Avg. domestic peer FE	0.021
	(0.022)
Avg. foreign peer FE	0.003
	(0.003)
Foreign Firms	
Avg. domestic peer FE	0.005
	(0.039)
Avg. foreign peer FE	0.011
	(0.007)
Firm FE	Yes
Industry-year FE	Yes
Area-year FE	Yes
N firms	27,457
N firm-year obs	197,253

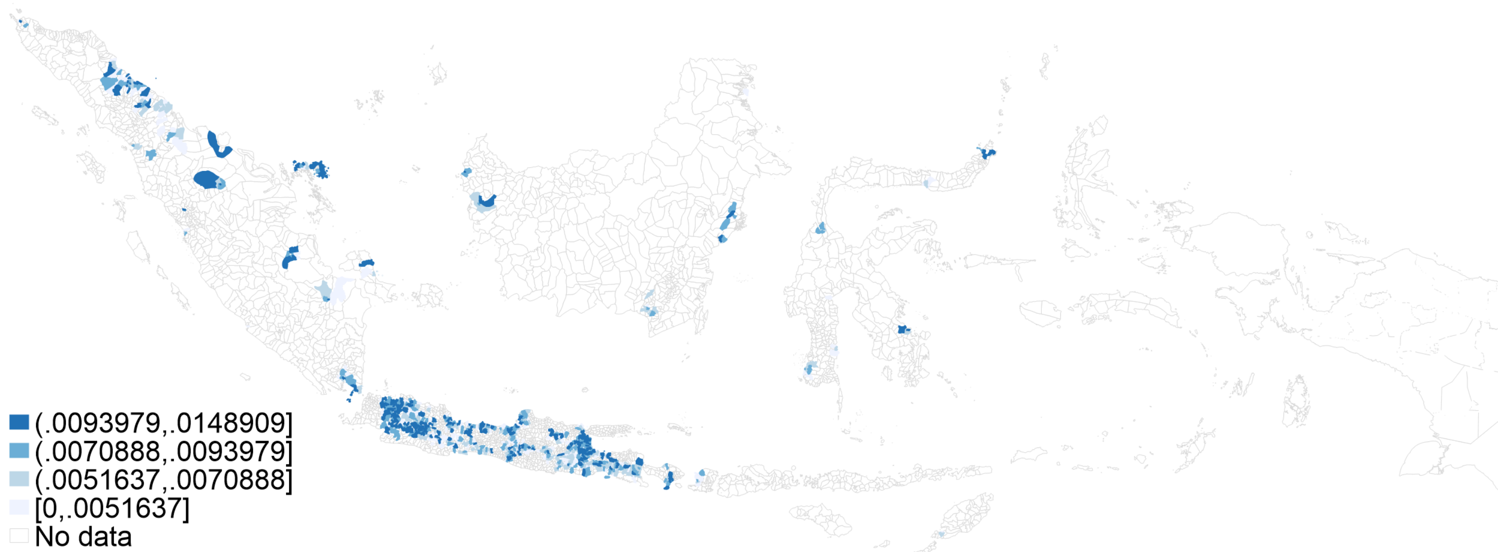
Notes: This table presents results from estimation of the primary estimating equation (10), using alternative subdistrict-level fixed effects. Bootstrap standard errors are in parentheses.

Figure A2: Subdistrict-level averaged average gains to peer productivity, Indonesia-wide



Notes: This figure displays Indonesia's urban subdistricts (*kecamatan*) shaded by the subdistrict-level average of the cluster-level average gains to peer productivity from the first policy experiment.

Figure A3: Subdistrict-level averaged aggregate gains to peer productivity, Indonesia-wide



Notes: This figure displays Java's urban subdistricts (*kecamatan*) shaded by the subdistrict-level average of the cluster-level aggregate gains to peer productivity from the first policy experiment.