Boston University Global Development Policy Center

HCI WORKING PAPER 010 • 12/2018

Assessing the Spatial Concentration of Indonesia's Manufacturing Sector: Evidence from Three Decades

Samuel Bazzi *is an Assistant Professor of Economics at Boston University and Core Faculty at the Human Capital Initiative at the Global Development Policy Center.*

Alexander D. Rothenberg *is an economist with the RAND Corporation and a professor at the Pardee RAND Graduate School.*

Shanthi Nataraj *is Director of the Labor and Workforce Development Program and a senior economist at the RAND Corperation. She is also a professor at the Pardee RAND Graduate School.*

Amalavoyal Chari *is a Reader in Economics at the University of Sussex.*

ALEXANDER D. ROTHENBERG, SAMUEL BAZZI, SHANTHI NATARAJ, AMALAVOYAL CHARI

Beyond the role of economic forces, many theories of economic geography emphasize the way politics can shape the spacial configuration of economic activity. We investigate the impact of changes in political regimes on industrial concentration using 30 years of data on Indonesian manufacturers. These data span both the reign of Suharto, one of the strongest central governments in Southeast Asia, and its collapse and the subsequent decentralization of power. Using the canonical measure of Ellison and Glaeser, we show that in the mid 1980s, Indonesia's firms exhibited a simlar degree of agglomeration as seen in the United States. Spatial concentration then declined until the 1998 Asian Financial Crisis, and has since begun to rise during the decentralization period. We also measure concentration using the continuous measure developed by Duranton and Overman (2005), and find that the agglomeration exhibited by Indonesian firms is also broadly similar to that documented by Duranton and Overman (2005) for the United Kingdom, although localization drops off more gradually in Indonesia than in the United Kingdom. Using this continuous measure of agglomeration, we identify 32 manufacturing clusters in Indonesia, and investigate the correlates of concentration. We find that the most robust drivers of agglomeration have been natural resources and supply chain linkages, especially with respect to explaining long-term changes in spatial concentration.

1 Introduction

Throughout the world, one of the striking facts about economic geography is that firms and workers tend to cluster, or agglomerate, in certain places. Theorists and empirical researchers have offered many different explanations for spatial concentration, including transport costs (e.g. von Thünen, 1826; Krugman, 1991), natural advantages (e.g. Ellison and Glaeser, 1999), and productivity spillovers. Marshall (1890) emphasizes that these productivity spillovers can take many forms, including (1) direct technology spillovers, (2) labor market pooling, and (3) intermediate input linkages.[1] The idiosyncrasies of history and path dependence probably also play an important role (e.g., Davis and Weinstein, 2002).

Although economic factors are of doubtless importance, politics may also influence the location decisions of firms and workers. Since at least Hoselitz (1955), researchers have emphasized that access to political power, which can encourage favorable policies, provide better information, or enable the extraction of rents, may encourage firms to locate in capital cities, increasing urban primacy. Ades and Glaeser (1995) show empirically that dictatorships have central cities that are, on average, 50 percent larger than their democratic counterparts.

Indonesia represents an interesting test case for examining theories of spatial concentration, particularly as they relate to politics. Because of its unique geography as an archipelago, colonized by the Dutch, spatial inequalities in the concentration of economic activity, employment, and output have been a central feature of Indonesia's economy for centuries. As of 2014, the capital of Jakarta and the surrounding metropolitan region (known as Jabodetabek), constitute the world's second largest urban area, with 30.6 million people, or 12.2 percent of Indonesia's 249.9 million population (DEPKES, 2014). Although Jakarta was the former colonial capital and has been an important city for centuries, its rapid growth after independence may be partly explained by the dictatorship of Suharto, who ruled Indonesia for 31 years with one of the strongest central governments in Southeast Asia. Suharto's New Order regime handed out rents to many politically connected companies (Fisman, 2001), which may have been one force encouraging concentration of economic activities in the capital.

However, over the last two decades, Indonesia has experienced a profound political transformation. The Asian Financial Crisis of 1998, and the subsequent depreciation of the Indonesian Rupiah, brought about the fall of Suharto and resulted in a dramatic transformation of the government. A new period of decentralization has seen district (*kabupaten*) governments becoming more powerful and exerting much greater control over local policies than ever before (Fitrani et al., 2005).

This paper investigates whether these dramatic political changes in Indonesia over the last 30 years led to changes in the spatial distribution of economic activity. In particular, we study the location decisions of large manufacturers in Indonesia.[2] To do so, we use two data sources: (i) a 30 year panel of large manufacturers (*Survei Industri*, or SI), with location information at the district level, and (ii) a 2013 cross section of manufacturers (*Direktori Industri Manufaktur*, or DIM), with address-level location information. The first data source allows us to construct the widely used Ellison and Glaeser (1997) index of industrial concentration and examine how it changes over time and varies by sector. The second

[1] Duranton and Puga (2004) offer an alternative categorization of productivity spillovers, which come from (1) sharing, (2) matching, and (3) learning.

[2] An important limitation of this research is the focus on manufacturing. Although manufacturing is an important source of economic activity and growth, only 20.9 percent of Indonesia's labor force is employed in manufacturing, with 44.8 percent being employed in services, and 34.3 percent employed in agriculture (DEPKES, 2014).

data source enables us to use a different, continuous measure of spatial concentration (Duranton and Overman, 2005), which is seldom used in the literature due to its considerable data requirements.

Using this approach, we find that Indonesia's firms tend to exhibit a similar degree of agglomeration as in the United States and United Kingdom. However, using the Duranton and Overman (2005) measure, we also find that the tendency for agglomeration extends over larger distances than in the United Kingdom. This points to the relatively large clusters of economic activity around a few megacities and especially Jakarta. Interestingly, Indonesian manufacturers exhibited similar levels of spatial concentration as the United States in the mid-1980s as captured by the average Ellison and Glaeser (1997) index across industries. Yet, this degree of concentration in the average industry began to fall in the years leading up to the Asian Financial Crisis. Thereafter, average levels of concentration began to rise again. This pattern goes against the view that democratization and decentralization could be a force for reducing spatial disparities and encouraging economic activity in new areas. However, this could also be explained in part by the persistence of the late 1990s economic shock, leading to greater agglomeration as a means of dealing with risk in both input and output markets.

We also use the comparison across the Duranton and Overman and Ellison and Glaeser (1997) indices in 2012 to show that a well-defined and accurate characterization of industrial concentration requires accounting for not only the distance between spatial units but also the distance between firms within spatial units. This methodological point has useful implications that we draw out by developing a new method for identifying the location of clusters in industries with high degrees of concentration. This exercise delivers useful information for policymakers insomuch as some of these clusters are outside the major centers of production in the largest cities of Java.

Next, we use a host of industry-level covariates to explain patterns of concentration across time and space. We find that the nature of production technologies and output plays a key role in driving agglomeration differences across industries. In particular, we identify higher levels and long-run growth in concentration among industries in which intermediate and natural resource inputs are more important. Moreover, we provide evidence that transport costs as well as technology spillovers are important forces for static albeit not dynamic agglomeration. These results provide an initial window into the drivers of geographic concentration across industries. In other work, we investigate these underlying forces of industrial concentration more formally by means of a new empirical strategy for identifying agglomeration externalities and spillovers (Rothenberg et al., 2016a).

The remainder of the paper is organized as follows. Section 2 provides historical background on the geography of economic activity in Indonesia. Section 3 presents leading indicators of industrial concentration, draws the comparison with developed countries, and identifies the location of different clusters across the archipelago. Section 4 presents descriptive regressions that help explain the patterns of concentration across sectors. Section 5 concludes with a discussion of policy implications and directions for future research.

2 Indonesia's Economic Geography in Historical Context

Spatial disparities in economic activity, employment, and output have been a pervasive feature of Indonesia's economy for centuries. These disparities, interacting together with the significant ethnic and religious diversity across the archipelago, have often threatened to undermine the viability of Indonesia as a nation state. In the colonial period, the Dutch East India Company (VOC) practiced a very uneven development strategy in Indonesia, beginning with their arrival in the early 17th century. The VOC promoted extractive enclaves in the form of plantations and natural resource extraction on the Outer Islands of Sumatra, Sulawesi, and Kalimantan, while at the same time encouraging more balanced, diversified growth in the more densely populated Inner Island of Java.

Before and after World War II, the movement for Indonesia's independence from the Dutch and eventually the Japanese was catalyzed by its focus on "*Bhinneka Tunggal Ikea*", the national motto, which translates to "Unity in Diversity". However, in the 1950s and 1960s, violent separatist movements and conflict between Communist and Islamic political movements threatened to overwhelm Sukarno's early presidency (1945–67) and dissolve the Indonesian nation. Many of these movements were associated with grievances about the continued concentration of political and economic resources on the main islands of Java/Bali at the expense of the rest of the country. In the wake of these recurring upheavals after independence, General Suharto assumed the presidency in 1967, ultimately ruling Indonesia as a dictatorship for over 30 years as one of the strongest and most centralized governments in Southeast Asia. Suppressing separatism in the Outer Islands, often through violent means, was an important feature of Suharto's New Order regime.

However, reducing regional inequality was also a central goal of government policy during the Suharto era. Fueled in part by oil revenue windfalls in the 1970s, the government pushed several flagship development programs, including (i) large-scale population resettlement from rural Java/Bali into new agricultural settlements in the Outer Islands (see Bazzi et al., 2016), (ii) mass primary school construction and water and sanitation infrastructure in lagging regions (Duflo, 2001), block grant transfers to underdeveloped villages (Akita and Szeto, 2000), special economic zones (Rothenberg et al., 2016b), and road construction efforts (Rothenberg, 2013). Although these policies encouraged more even economic development across the archipelago, spatial inequalities remained large and persistent. In the mid-1990s, the per capita regional product of the poorest district was more than 50 times smaller than that of the richest district, and districts at the 90th percentile were around 5 times richer than those at the 10th percentile.

Alongside the New Order development policies, Indonesia experienced rapid economic transformation, growing from one of the poorest countries in Asia to one of the emerging Tigers. Popular press and research articles at the time touted Indonesia as the next emerging giant of Southeast Asia (see, e.g., Hill, 1996). GDP per capita grew by 5.1 percent annually between 1967 and 1997, skyrocketing from USD 550 to 2,433 over that period (as measured in constant 2000 USD). Manufacturing played a central role in this transformation, contributing eight percent of GDP at the beginning of the period and nearly one-quarter at the end (Hill, 2000). Export growth and greater import competition induced by tariff liberalization in the early to mid-1990s played an important role in this process (see Amiti and Konings, 2007). Moreover, new opportunities for agro-industrial development arose in the 1990s as demand grew on world markets for products relying on natural resources such as palm oil, rubber, and lumber that are abundant in Indonesia's Outer Islands. This natural resource boom persisted into the 2000s, leading to greater geographic dispersion in light manufacturing activity across the archipelago as producers tend to locate close to the source of inputs to keep costs down. Nevertheless, much of the non-agro-industrial growth remained concentrated on the island of Java and in particular in the three largest cities of Jakarta, Surabaya and Bandung (see Amiti and Cameron, 2007).

The tremendous economic progress during the Suharto era came to an abrupt and grinding halt with the Asian Financial Crisis in May 1997 and ensuing political turmoil. The Indonesian Rupiah collapsed along with the stock market and the entire financial industry with the government eventually stepping in to liquidate several private banks. Income per capita collapsed by 15 percent, falling to USD 2,083 in 1998 and still further to USD 2,071 in 1999 (in constant 2000 USD). Growth picked up thereafter but took nearly 15 years to get back to trend. The manufacturing sector was hit especially hard as were urban centers more generally.

In the wake of the upheaval in the late 1990s, Indonesia ushered in a new democratic political system with far-reaching institutional reforms. With democracy in place by 1999, the government also embarked on sweeping decentralization reforms beginning in 2001 (see Hill, 2014). The ensuing political transformation devolved considerable new resources and authority that had previously been centralized but was now passed down to the district level. This dramatic wave of reform, often referred to as the "big bang decentralization," was motivated by the aim of more aggressively shifting political resources and economic activity from the core Inner Islands to the peripheral Outer Islands than was achieved during the Suharto era. Policymakers hoped that this process of decentralization combined with democratization would engender new opportunities to stimulate local economic development in areas with underlying potential but a lack of political influence to attract the necessary resources during the authoritarian era.

Overall, Indonesia experienced dramatic changes in its political and economic geography over the last 30 years. In the remainder of the paper, we provide a window into these changes through the lens of the country's firms and workers as they contribute to patterns of spatial concentration.

3 Measuring Spatial Concentration

In this section, we document the spatial concentration of manufacturing activity in Indonesia, focusing both on the current geographic distribution of firms and also on how the concentration of different industries changed over time. To measure spatial concentration, we use two different datasets: (i) the Annual Survey of Manufacturing Establishments (*Survei Tahunan Perusahaan Industri Pengolahan*, or SI) and (ii) Indonesia's 2013 Directory of Manufacturers (*Direktori Industri Manufaktur*, or DIM).

Our first dataset, the SI, aims to be a complete census of manufacturing plants with 20 or more employees. Throughout this paper, we use firms and plants interchangeably as our data cannot distinguish between the two.[3] Conducted annually by Indonesia's Central Statistical Agency (*Badan Pusat Statistik*, or BPS), the survey is very detailed, recording information on plant employment sizes, industry of operation, cost variables, and measures of value added. We also observe the location where each plant operates, but this location information is only available at the district level. We work with over thirty years of annual SI data from 1980–2012.

Our second dataset, the DIM, is also produced by BPS and contains address-level information for the headquarters of nearly 23,000 manufacturing plants in 2013 with more than 20 employees. The data also include the names of firms, current number of employees, the 5-digit code for the industry in which the

[3]Blalock and Gertler (2008) argue that less than 5 percent of firms in the data are organized across multiple plants.

firm operates, and various contact information. These data, which represent a register of the universe of large manufacturing establishments, are publicly available and are produced as part of fielding the SI. However, there are no common firm-level identifiers in the data, meaning that the SI and DIM datasets cannot be reliably linked together.

To understand changes in spatial concentration over time, we require a fixed and unchanging geographic unit of analysis. Many districts in Indonesia (*kabupaten*) were partitioned and divided into new districts as part of the decentralization process noted above (see Bazzi and Gudgeon, 2016). Therefore, we aggregate the spatial units back to the 1990 borders, leaving us with 290 districts, with a median land area of nearly 1,500 square kilometers and a population of roughly 430,000 in 1990. This is a slightly smaller area albeit larger population than the median county in the United States. When we use provinces as the spatial unit of analysis, there are 26 in our sample, with a median land area of nearly 62,000 square kilometers, roughly half the size of the median U.S. state. Although we cannot detect changes in industrial concentration occurring below the district level, the DIM data allow us to compute a highly localized, cross-sectional measure of concentration in 2013.

As with the district codes, we require an industry classification scheme that is consistent over time. We anchor industry codes at the 4-digit level back to the ISIC Rev. 2 system of industrial classifications, which prevailed at the beginning of our study period. Because many indices of spatial concentration are especially noisy for small numbers of firms, we also drop 4-digit industries with fewer than 10 firms. This leaves us with a sample of over 600,000 plant-year observations in the SI panel data, and a sample of over 22,000 firms in the DIM cross-section.

A potential limitation of the SI and DIM data is the omission of manufacturing firms with fewer than 20 employees. This will lead to biased estimates of industrial concentration if smaller firms exhibit different clustering patterns than larger firms. Hsieh and Olken (2014) and Rothenberg et al. (2016c) argue that medium and large sized manufacturing firms represent only a very small portion of all firms, many of which are micro firms with less than 5 employees and are not formally registered. Unfortunately, data limitations make it difficult to study the spatial concentration of micro and small firms (less than 20 employees), but further research is needed.

3.1 Evolution of the Ellison and Glaeser (1997) Index

The literature provides several measures of industrial concentration, but our first set of results focuses on the seminal Ellison and Glaeser (1997) index, defined as follows for a given industry i:

$$\gamma_i = \frac{G - (1 - \sum_\ell x_\ell^2)H}{(1 - \sum_\ell x_\ell^2)(1 - H)} \,, \tag{1}$$

where ℓ indexes locations, $H = \sum_j^J z_j^2$ is a Herfindahl index of the plant j size in terms of employment z for all J plants in the industry, $G = \sum_\ell (x_\ell - s_{\ell j})^2$ measures the sum of square deviations between location ℓ's share of national employment, x_ℓ, and location ℓ's share of employment in industry j, $s_{\ell j}$.[4]

[4] Combes et al. (2008) provides an excellent review of the different measures of spatial concentration. Because the Ellison and Glaeser (1997) index explicitly accounts for industrial concentration (i.e., the extent to which employment is concentrated in a few firms), it is useful for analyzing the SI dataset since many industries are dominated by a small number of large firms, and plant size distributions change significantly over time.

If an industry is perfectly competitive, with a large number of very small plants, H tends to zero, and γ tends to $G/(1-\sum_\ell x_\ell^2)$. In this case, γ purely measures spatial concentration but ignores industrial concentration, a limitation of many existing metrics. More typically, H is positive, and $\gamma = 0$ when firms in the industry concentrate only as much as would be expected under random assignment. Positive values of γ indicate excess spatial concentration, while negative values of γ indicate excess spatial dispersion.

Using the SI data, we compute the spatial concentration of employment, γ, for every 4-digit industry and year. The estimated γ_i can be interpreted as the probability that any pair of plants in industry i choose their locations jointly. Figure 1 depicts how the mean and median of this index evolved across industries over time. In Panel A, we calculate γ using districts as the spatial unit of analysis and find a sharp 40 percent reduction in the index for the average industry, falling from 0.051 in 1984 to 0.032 in 1997. Strikingly, the mean and median γ in 1984 Indonesia are identical to the mean and median γ that Ellison and Glaeser find for the U.S. in 1987.

After the fall of Suharto, the average concentration index begins to rise, and by 2012, the index was back to its levels in the mid-1980s. The median index also shows a decline in concentration in the early 1990s, but it does not change as substantially, suggesting that there is some heterogeneity in the trajectories of different industries over time. In Panel B, using provinces as the spatial unit of analysis, we find a similar pattern of falling concentration in the 1990s and rising concentration after 1998.

That manufacturing industries tended to increase their spatial concentration after the fall of Suharto is striking. Rothenberg (2013) argues that the reduction in spatial concentration that occurred in the 1990s seems to be related to transportation improvements undertaken during the final years of Suharto's New Order regime. Sites which had been previously disconnected experienced greater market access, and this encouraged firms to suburbanize. Yet, the increase in concentration after the fall of Suharto is somewhat puzzling. The political economy literature argues that dictatorships increase spatial concentration, so when they fall, we would expect spatial concentration to decrease. This force for greater dispersion should have been amplified by the process of decentralization. However, another possibility is that the fall of Suharto increased uncertainty and weakened the government, thereby increasing the returns to locating in the capital. There was significant uncertainty about the direction of politics in the years after Suharto, and firms may have located in central cities in order to influence the direction of politics to their advantage.

The patterns of sector-specific concentration measures are largely intuitive. Table 1 depicts the top 20 most (Panel A) and least (Panel B) concentrated industries as captured by γ in 2012.[5] The most concentrated industry, Kapok Manufacturing (ISIC 3216), involves creating woven fiber from the seed pod fluff of a rainforest tree called Ceiba (silk cotton) tree. This tree grows prominently in East Java, and the harvested cotton-like substance can be woven into a fiber that can be used for stuffing and insulation. This example highlights the importance of natural advantages in determining industrial locations, and it also may be useful for explaining the high concentration of Clove Cigarettes (ISIC 3142), Other Tobacco Products (ISIC 3149), and Sugar Products (ISIC 3118). On the other hand, weaving and textile industries (ISIC 3229, ISIC 3212) and handicraft and wood carving (ISIC 3314) are also highly concentrated, which may owe to both natural advantages (access to wood or fibers) but also to labor market pooling. Finally, one of the most concentrated industries is the Radio and TV industry (ISIC 3832), where knowledge

[5] An expanded results table containing all industries can be found in Appendix Table A.1.

spillovers are likely to play a large role.

Looking at changes since the early days of industrialization in 1982, we see sizable reductions in concentration for some industries and increases in others. For example, the dramatic increase in establishments in Motor Vehicle Assembly and Manufacturing (ISIC 3843) is associated with a significant reduction in concentration, pointing to the expansion of this sector into new areas of the country. The same holds for Handicraft and Wood Carving (ISIC 3314), which saw a dramatic drop in γ from 0.49 to 0.06. However, other sectors such as Knitting Mills (ISIC 3213) exhibit a strong increase in concentration alongside a tripling in the number of establishments.

In Panel B, the most dispersed industries include slaughtering and preserving meat (ISIC 3111) and milk products (ISIC 3112), industries that are highly perishable. Glass and Glass Products (ISIC 3621) is another example of a fragile, perishable product that is likely to be damaged when shipped over long, low quality transportation routes. Bricks (ISIC 3641) is the most dispersed industry, likely because of transport costs; the high weight of bricks, relative to their price, would tend to lead them to located in a large number of areas to satisfy demand. A similar story can be told about Cement Products (ISIC 3632) and fixtures made of metal (ISIC 3812).

Together, the varied patterns of change in concentration provide an initial glimpse into the different forces of agglomeration at work across sectors. In Section 4, we explore these forces in a regression framework, but next we develop an alternative measure of concentration.

3.2 The Duranton and Overman (2005) Continuous Approach

While the Ellison and Glaeser (1997) index has several useful properties in analyzing changes in industrial concentration over time, one important limitation is that different choices of the spatial unit of analysis yield different results. For example, in 2012, if we construct γ using districts and compare it to γ using provinces, the spearman rank correlation between these two measures across industries is only 0.62. Another concern is that spatial units are treated symmetrically, and the distances between spatial areas are completely ignored in the calculation. Finally, there is no way to use γ to test for statistical significance.[6]

Duranton and Overman (2005) created an index designed to solve some of these problems with standard indices of spatial concentration. However, to use their approach, the data requirements are substantial. In particular, one needs the latitude and longitude coordinates of every firm in the data. Because such information is not readily available in most firm-level datasets, only a handful of studies have made use of this approach.[7]

We use the 2013 DIM data to construct the Duranton and Overman (2005) index for the first time in Indonesia. The DIM data contain address-level information on the locations of Indonesian manufacturers. To transform this information on addresses into latitude and longitude coordinates, we use the Google Maps API to geolocate each address in the database. Google Maps is quite effective at finding

[6]Another problem with many standard measures of spatial concentration is that spatial aggregation leads to spurious correlations between aggregated variables. This worsens as higher levels of aggregation are considered. This is recognized in the quantitative geography literature as the Modifiable Areal Unit Problem (Yule and Kendall, 1950; Cressie, 1993).

[7]See, for example, Koh and Riedel (2014) on German manufacturing and services, Alfaro and Chen (2014) on multinational firms, Murata et al. (2014) on patent citations and knowledge spillovers, and Kerr and Kominers (2015) on patent data in Silicon Valley.

addresses in Indonesia, especially when postal codes are supplied.[8] Figure 2, Panel A, shows the location of all 22,000+ establishments in DIM. The overlaid district boundaries (in white, based on 2010 district definitions) point to the potential added value of the Duranton and Overman approach (relative to Ellison and Glaeser, for example) in providing a more precise measure of concentration within district and also capturing distances between districts.

Figure 2 also shows the location of establishments in two illustrative industries: wood, bamboo, or rattan furniture production (ISIC 3321) and cooking oil manufacturing (ISIC 3115), which includes the production of a variety of cooking oils, such as palm oil, sesame oil, peanut oil, and castor oil. Each dot on the two maps displays the location of a single firm. Wood furniture production (Panel B) exhibits a significant propensity for clustering. Several centers of production are readily apparent, including those in and around Jakarta, in several cities in Central and East Java, and also in southern Bali. There are also very few wood furniture manufacturers outside of Java and Bali. On the other hand, the locations of cooking oil processors (Panel C), which includes palm oil and coconut oil, are much more diffuse. This is due in large part to the wide geographic scope of the key resource inputs to production, which, in the case of palm oil, must be processed within a day or two of harvest. Moreover, cooking oil production is less capital intensive, and agglomeration economies in cooking oil production are plausibly less important.

The key idea of the Duranton and Overman (2005) index is to measure the distribution of distances between pairs of firms in an industry, and to see how that distribution is different from a counterfactual, or reference, distribution, which would have resulted if firms had randomly chosen locations subject to the geographic constraints of Indonesia. To describe how the index is constructed, first focus on firms in a single industry. Dispensing with earlier notation, suppose that there are N firms in this industry, indexed by $i = 1, 2, ..., N$. For each pair of firms in that industry, we use the latitude and longitude data and Vincenty's (1975) formula to calculate $d_{i,j}$, the distance between that pair of establishments.[9] This generates a total of $N(N-1)/2$ unique bilateral distances.

Next, for each industry's set of pairwise distances, we estimate the kernel density of bilateral distances d:

$$\widehat{K(d)} = \frac{1}{h \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} L_i L_j} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} L_i L_j f\left(\frac{d - d_{i,j}}{h}\right) , \tag{2}$$

where h the bandwidth and $f(\cdot)$ is the kernel density function, and we weight pairwise distance observations by $L_i L_j$, the product of employment totals for firms i and j.[10] By comparing the actual distribution of distances between employees in firms to a *reference* distribution of distances, which is the distance distribution of workers that would have arisen if firms were randomly assigned to locations, we can both detect departures from randomness and also measure the intensity of spatial concentration

[8] As of 2013, Google Maps had made a significant investment in expanding their mapping capabilities in Southeast Asia, and a new version of their mapping software came online. However, depending on how we searched the Google Maps API, such as using addresses with and without postal codes, we obtained slightly different results, so we tried different search options and explored the robustness of our results to different choices. For the most part, the findings are quite similar no matter what the exact search approach. See Appendix A.1 for a more detailed discussion.

[9] Vincenty (1975) proposes iterative techniques to calculate the distance between points on the surface of a spheroid. This assumes that the Earth can be approximated by an oblate spheroid, and is more accurate than other methods, such as great-circle distance, which assumes a spherical Earth.

[10] Because γ is a measure of the spatial concentration of manufacturing employment, we use employee weights to allow for a better comparison between K and γ. Nevertheless, Duranton and Overman (2005) note that the multiplicative weights in (2) assign significant importance to the distances between the largest firms, potentially skewing results.

between firms within an industry.[11] Figure 3 plots $\widehat{K(d)}$ for $d \in [0, 200]$ (in solid blue) for both wood furniture producers (Panel A) and cooking oil manufacturers (Panel B). The density of distances for wood furniture producers contains significant mass at low levels of distance, while the density for palm oil manufacturers is relatively flat, corresponding more closely to a uniform distribution.

These figures also plot 95 percent confidence bands, $\widehat{L(d)}$ and $\widehat{U(d)}$, in grey dotted lines, which represent the bounds of our reference distribution. The reference distribution represents the distribution of distances, weighted by employment, that would have arisen if firms in the industry had chosen locations randomly. To obtain these confidence bands, we first randomly draw locations from the set of all possible sites, sampling from these locations without replacement. After we draw a random set of locations for each firm, we calculate the smoothed density for the resulting distance distribution, weighted by employment, analogous to (2). We repeat the procedure of drawing locations and estimating the counterfactual $\widehat{K(d)}$ 10,000 times. The local confidence bands, $\widehat{U(d)}$ and $\widehat{L(d)}$, which are depicted in the dashed lines of Figure 3, represent the empirical 95 percent confidence intervals at each level of distance. The goal of this procedure is to control for overall industrial agglomeration and to deal with the fact that certain locations may be prohibited for firms, such as residential areas in cities, natural protected areas, or the vast waterways connecting Indonesia's archipelago.

Figure 3 shows that cooking oil production (ISIC 3115) exhibits dispersion until d is approximately 140 km, because $\widehat{K(d)} < \widehat{L(d)}$ for all distances $d \in [0, 140]$. After 140 km, the distribution of distances for cooking oil does not look different from randomness. On the other hand, wood furniture manufacturing shows localization, with $\widehat{K(d)} > \widehat{U(d)}$, at all distances $d \in [0, 40]$. Note that furniture also shows dispersion in a few points around $d = 70$ and $d = 160$, and a second mass of localization after $d = 190$. Because the density of distances must integrate to one, localization at some distances may imply dispersion at other distances, and this is exactly what happens here.

This discussion motivates the following definitions.

Definition 3.1. *If $\widehat{K(d)} > \widehat{U(d)}$ at distance d, the industry exhibits **localization at distance d** (at a 5% confidence level). Intuitively, there is more mass at distance d than we would expect if firms were randomly assigned to locations. However, if $\widehat{K(d)} < \widehat{L(d)}$ for distance d, then the industry exhibits **dispersion at distance d** (at a 5% confidence level). If $\widehat{K(d)} \in \left[\widehat{L(d)}, \widehat{U(d)}\right]$ for distance d, there is no localization or local dispersion at distance d.*

Using these local definitions, we can define global definitions of localization and dispersion:

Definition 3.2. *If $\widehat{K(d)} < \widehat{L(d)}$ for at least some $d \in [0, 200]$, and if $\widehat{K(d)}$ is never greater than $\widehat{U(d)}$, then the industry exhibits **global dispersion** (at a 5% confidence level). If $\widehat{K(d)} > \widehat{U(d)}$ for at least some $d \in [0, 200]$, then the industry exhibits **global localization** (at a 5% confidence level).*

Definition 3.3. ***Index of Localization and Dispersion**: For industry A, our index of localization at distance d, $\Gamma_A(d)$, is given by the following:*

$$\Gamma_A(d) = \max\left(\widehat{K_A(d)} - \widehat{U_A(d)}, 0\right).$$

[11] Note that in estimating these kernel densities, we constrain the density estimates to zero for negative distances by using the approach of "reflecting" the data at the boundaries (Silverman, 1986, p. 30). Ben Jann's `kdens` routine in `STATA` provides the implementation.

Our index of dispersion at distance d, $\Psi_A(d)$, is defined as follows:

$$\Psi_A(d) = \begin{cases} \max\left(\widehat{L_A(d)} - \widehat{K_A(d)}, 0\right) & \textit{if } \sum_{d=0}^{200} \Gamma_A(d) = 0 \\ 0 & \textit{otherwise,} \end{cases}.$$

Overall, we find that 48 of the 87 industries (55 percent) deviate from randomness at the 5 percent level of significance. Of these 48 industries, 36 are globally localized at some point over the $d \in [0, 200]$ interval (41 percent of the total), while 12 industries (14 percent of the total) are globally dispersed. These findings represent a somewhat larger tendency for agglomeration and a somewhat weaker tendency for dispersion than in the United Kingdom data analyzed by Duranton and Overman (2005). In that setting, only 52 percent of the industries are localized, while 24 percent are dispersed. We take this comparison a step further in the following subsection.

Before doing so, we compare the continuous measures of concentration with the Ellison and Glaeser γ index in Table 2. Panel A focuses on the top 20 most concentrated indices, ranked by Γ, while Panel B lists the 12 industries that were globally dispersed, ranked by Ψ. A spearman rank correlation of Ellison and Glaeser's (1997) γ and Γ for globally localized industries is only 0.230; a similarly low spearman rank correlation can be found when comparing Ellison and Glaeser's (1997) γ and Ψ for globally dispersed industries. This suggests that departures from randomness are not fully accounted for in the Ellison and Glaeser (1997) index. Some of the most concentrated industries (Kapok Manufacturing, ISIC 3216, Knitting Mills, ISIC 3213, and Clove Cigarettes, ISIC 3142) do appear on both lists, but the most dispersed industry, Manufacture of Cooking Oils (ISIC 3115) is somewhat in the middle of the γ distribution. Note that of the highly concentrated industries, nearly all are significantly localized for all $d \in [0, 100]$. Together, these results suggest that a well-defined and accurate characterization of industrial concentration requires accounting for not only the distance between spatial units but also the distance between firms within spatial units.

3.3 Comparison with the United Kingdom

One useful feature of the Duranton and Overman (2005) index is that it allows us to compare industrial concentration patterns across countries with very different geographies. We offer here a direct comparison of Indonesia and the United Kingdom, two countries at very different stages of development.

Figure 4 plots the number of industries that exhibit global localization (Panel A) or global dispersion (Panel B) at each distance. This figure shows that most of the industries that display localization do so at relatively short distances with $d \in [0, 40]$ (78 out of 84 industries). We also find that the number of industries exhibiting localization begins to taper off after about $d > 50$ or so, a finding that is similar to the United Kingdom results in Duranton and Overman (2005) (reproduced in Panel C). However, the drop off is much more gradual than compared to the United Kingdom, and there is no second peak in the Indonesian data. Another finding, very similar to the United Kingdom data, is that dispersion tends to be uniformly distributed across distance.

In order to measure the extent of localization and dispersion at different distances, we sum the industry-specific indices of localization and dispersion across industries, at each level of distance, to form $\Gamma(d) \equiv \sum_A \Gamma_A(d)$ and $\Psi(d) \equiv \sum_A \Psi_A(d)$. We plot $\Gamma(d)$ and $\Psi(d)$ in Figure 5, Panels A and B, alongside their respective measures from Duranton and Overman (2005). Although these figures are not directly comparable, because Duranton and Overman (2005) have more industries (234) than we have, the shapes are somewhat similar. In particular, we find, as they do, that the intensity of localization is greatest at small distances, and that dispersion does not show much of a relationship with distance. However, as in Figure 4, the decline in localization with respect to distance is much more gradual. This may suggest that spatial concentration in Indonesia, while exhibiting some of the same patterns as an advanced industrialized country like the U.K., displays less intensive spatial clustering at small distances, and greater clustering at low distances.

Figure 6 shows that these average clustering patterns are driven by a small number of industries. We plot the distribution of $\Gamma_A \equiv \sum_{d=0}^{200} \Gamma_A(d)$ and $\Psi_A \equiv \sum_{d=0}^{200} \Psi_A(d)$ for all industries. These highly skewed distributions show that most industries exhibit low levels of localization and dispersion, while only a very small number of industries exhibit high levels of localization or dispersion. Though the magnitudes differ, the general patterns are consistent with Duranton and Overman (2005).

3.4 Identifying the Numbers and Locations of Clusters

Equipped with new estimates of firm-level concentration across Indonesia, we now identify both the number and location of industrial clusters across the archipelago. Recall that in Figure 2, Panel B, we saw that the wood furniture industry (ISIC 3321) appeared to have multiple clusters, in Jakarta, Surabaya, and Central Java. In this subsection, we describe a new, simple approach for how to use estimates of the distance distribution between employees in firms, $\widehat{K(d)}$, to identify the number of clusters and their locations. Again, focusing on the wood furniture industry, Figure 3 shows that $\widehat{K(d)}$ is greater than the upper 95 percent confidence band for distances $d \in [0, 45]$ and after $d = 200$ or so. Figure 7 expands the distance range of Figure 3, Panel A. This figure shows three significant mass point spikes from $d \in [0, 1000]$, one around $d = 0$, the second at $d = 245$, and the third at $d = 450$ or so.

When a distance distribution exhibits spikes like this, they reflect modes of $\widehat{K(d)}$. These modes reflect clustering; we see that lots of pairs of firms have distances of $d = 0$ and $d = 245$. For a globally localized industry, we count the number of clusters in that industry from $d \in [0, 1000]$ as follows:

Definition 3.4. ***Number of Clusters from*** $d \in [0, 1000]$*: If industry A exhibits some degree of global localization and is localized around $d = 0$, the number of peaks of $K(d)$ above $\widehat{U_A(d)}$, defines the number of significant modes of $K(d)$, which in turn represents the number of industrial clusters. A peak above the upper confidence bound at d is defined by $K(d) > K(d - \epsilon) > \widehat{U_A(d)}$ and $K(d) > K(d + \epsilon) > \widehat{U_A(d)}$ for some small $\varepsilon > 0$. We also include a peak if $K(d) > \widehat{U_A(d)}$ around $d = 0$. Let C_A denote the number of such peaks at $d \in [0, 1000]$.*

After estimating the number of clusters, we use k-means clustering to partition firms into their respective clusters, where k is chosen to be equal to C_A.[12] Let the longitude and latitude coordinates of firm i be denoted by (x_i, y_i). If we index the clusters of industry A by $\mathcal{C} = 1, 2, ..., C_A$, we can associate a

[12] K-means clustering is a method of data reduction that is popular in cluster analysis in data mining. The method partitions N observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. We use the random seed method to choose an initial K observations for starting means. See Appendix A.2 for further details.

single, central *cluster location* to each cluster, $(x_{\mathcal{C}}, y_{\mathcal{C}})$, as follows:

$$x_{\mathcal{C}} = \frac{\sum_{i \in \mathcal{C}} L_i x_i}{\sum_{i \in \mathcal{C}} L_i}, \qquad y_{\mathcal{C}} = \frac{\sum_{i \in \mathcal{C}} L_i y_i}{\sum_{i \in \mathcal{C}} L_i}$$

Table 3 reports estimates of the number of clusters for globally localized industries and their locations.[13] Interestingly, most industries have multiple clusters; only 4 of the 32 industries localized around $d = 0$ had 1 cluster. Each of the locations are matched to their nearest large cities.

Most of the industries listed in Table 3 have cluster locations on Java. The capital city, Jakarta, and large nearby cities in the province of West Java are among the most common cluster locations. Yet, most clusters are unique to specific places. These places may be industrially specialized, instead of attracting a diverse set of industries for production. Moreover, there are a number of clusters outside Java, particularly in North Sumatra, another longstanding industrial center in Indonesia. We turn now to an investigation of some of the forces driving these patterns of industrial clustering across time and space.

4 Correlates of Concentration

In the spirit of Rosenthal and Strange (2001), we relate the levels and changes in spatial concentration to industry characteristics in order to better understand the mechanisms leading firms to cluster together. To do so, we estimate sector-level regressions relating the level or change in concentration to industrial characteristics that proxy for different drivers of concentration. We measure these proxies in the years prior to the Asian Financial Crisis. These include:

Natural Advantages. Access to naturally occurring raw materials plays an important role in location choices. For example, several industries in Indonesia, such as Manufacture of Cooking Oils (ISIC 3115), rely heavily on agricultural inputs that are most profitable when sourced from local producers. We proxy for these sector-specific needs with the share of energy and natural resource expenditures in total input costs as well as the total expenditures on water resources.

Transport Costs. It has long been recognized that transport costs play an important role in shaping the location decisions of firms.[14] Unfortunately, there is no simple way to measure the transport costs faced by industries in Indonesia (see Rothenberg, 2013). Ideally, we could use a technological measure of transport costs, which is a function purely of the characteristics of the goods that industries produce, and not the actual transport costs that firms face, which are endogenous to the choices they have already

[13] We only focus on the number of clusters within 1000 km, because in practice this is a sufficiently large band to search for clusters, and peaks typically do not occur after 1000 km. Going from east to west, the longest distance between two points in Indonesia is more than 4,250 km, from East Jambi, Sumatra to Jayapura, Papua. This is over twice as large as the distance between London and Kiev, Ukraine (2,134 km).

[14] However, the exact relationship between spatial concentration and transport costs is unclear. Classic models in urban economics (Alonso, 1964; Mills, 1967; Muth, 1969) and economic geography (Helpman, 1998) predict that lower transport costs induce a dispersion of firms and workers to more peripheral areas, as firms become more able to take advantage of cheaper land and labor markets. On the other hand, lower transport costs make an existing location more profitable, bringing it closer to other markets. Because of this, lower transport costs could intensify the self-reinforcing home market effects that cause agglomerations to form and grow. In the influential core-periphery model of Krugman (1991), reducing trade costs between two regions causes firms to agglomerate, pulling the entire manufacturing sector into one region.

made. As in Rosenthal and Strange (2001), we measure the transport cost sensitivity of industries by using inventories as a proxy for perishability. Our measure is the ratio of the value of inventories to the value of output, averaged across all firms in the SI data from 1992–1996; averaging over multiple years smooths out idiosyncratic variation across industries and firms.

We expect that industries with more highly perishable products (a smaller ratio of inventories to output) will tend to locate closer to their demand sources. With multiple markets, such industries will tend to be dispersed. Industries with less perishable products (a greater ratio of inventories to output) should be able to choose fewer locations for production.

Inter-Industry Inputs. We proxy for the importance of intermediate input linkages using the share of manufactured inputs in total input costs. These supply chain linkages have direct implications for costs and productivity and hence play an important role in models of agglomeration. Indeed, in a companion paper, we investigate the role of intersectoral and spatial linkages in determining the scope of agglomeration spillovers (Rothenberg et al., 2016a).

Labor Market Pooling. As noted by Marshall (1890), an industry benefits from agglomeration because it is more easily able to hire workers with industry-specific skills; these workers tend to be trained in nearby firms that are part of the the same cluster. Unfortunately, it is quite difficult to measure the extent to which industries make use of general or specific skills. However, we expect that industries with a more educated workforce would tend to require more specific skills, whereas industries with a higher share of workers with no schooling would tend to rely on skills that are more easily acquired.

We measure the share of each industry's workers who have no schooling, and the share who have at least a bachelor's degree, using the 1996 SI data (part of the 1996 economic census). We also construct a management ratio, defined as is the fraction of the industry's labor force that is made up of non-production workers. As the share of non-production workers rises, we would expect production to be more advanced, and labor to be more specialized. We expect that the fraction of workers with no schooling should be negatively related to spatial concentration, the fraction of workers with at least a bachelor's degree should be positively related to spatial concentration, and the non-production labor ratio should be positively related to industrial concentration.

Technology Spillovers. One of most highly cited reasons why firms agglomerate is that they benefit from technology spillovers. These spillovers occur because knowledge and ideas can become non-rival, public goods and can increase productivity and reduce costs for all users. We proxy for technology spillovers by measuring the ratio between an industry's spending on research and development and the total value of output. As this ratio increases, we would expect that knowledge spillovers are more important, and spatial concentration should increase.

Politics. We proxy for political involvement in the industry using two measures that capture the importance of state owned enterprises (SOEs) during the heyday of Suharto, 1980–1996. The first measures the average annual share of SOEs in industry value added and the second measures the share of SOEs in the total number of firms in the industry. We expect firms with more intensive state involvement to exhibit excess concentration on account of the rents acquired through relationships with the Suharto government.

Results. In Tables 4 and 5, we find some correlates of industrial concentration consistent with theories of agglomeration. Table 4 relates the above proxies to the Ellison and Glaeser index γ in 2012 at different levels of spatial resolution. Focusing on the more granular district-level results in columns 3 and 4, we find that the nature of production plays a key role in shaping concentration. The importance of inter-industry linkages and natural resource inputs, transport costs proxied by inventory–output ratios, and technology spillovers are consistent with our predictions. However, the correlations with labor market pooling tend to run contrary to our hypotheses: industries with a higher share of workers with no schooling actually exhibit greater concentration, while those with a higher share of workers with Bachelor's degrees exhibit lower concentration. Correlations with political measures are mixed; industries with a higher average share of value added by SOEs exhibit more spatial concentration, but industries with a higher average share of SOEs in the number of firms exhibit lower concentration.

When we turn to a measure of concentration based on the Duranton and Overman (2005) index Γ in 2013, many of these patterns are attenuated (Table 5). Because the Duranton and Overman (2005) index more precisely captures distances between firms than does the coarser Ellison and Glaeser measure, it forces more of the conditional correlations to load on the more localized drivers of concentration such as natural resource and intermediate input intensity as opposed to general drivers associated with, for example, footloose labor and SOEs. While the direction of the effects for natural resources, inter-industry linkages, transport costs and technology spillovers remain the same, the results are no longer statistically significant after controlling for 2-digit ISIC fixed effects. These findings are consistent with those of Rosenthal and Strange (2001), who show that in the United States, higher shares of manufactured inputs and natural resource use are correlated with higher concentration at the state-level, but that the results are not statistically significant at smaller geographic levels. However, unlike Rosenthal and Strange (2001), we do not find a consistent correlation between higher-skilled workers and concentration. This may be due in part to the differing education levels in the populations; Rosenthal and Strange (2001) do not find a consistent relationship between the share of the population with a Bachelor's degree and concentration, but do find that a higher share of the population with a Master's degree is correlated with higher concentration. We similarly find no robust relationship between the share with a Bachelor's degree and concentration when using the Duranton and Overman (2005) index. This finding is also consistent with work by Amiti and Cameron (2007), who show that proximity to input suppliers and final consumers is associated with higher wages among Indonesian firms. While Amiti and Cameron (2007) do find that labor pooling also contributes to higher wages, the magnitude of the effect is much smaller than in the United States, which they argue may be due to the lower levels of skill differentiation in Indonesia.

In Table 6, we measure the correlations between each indicator and the long-difference of γ between 1985 and 2012. Interestingly, we find that inter-industry linkages and natural resource inputs are the only robust predictors of *changes* in industrial concentration over the past 25 years. This suggests that input requirements may have driven the dynamic, long-term incentives for agglomeration, at least over the period spanning Indonesia's period of structural change and transition to democracy. Together, the results in Tables 4–6 provide new insights into the industry-specific drivers of agglomeration across time and space.

5 Conclusion

In this paper, we calculate the Ellison and Glaeser measure of spatial concentration for Indonesian manufacturing firms, and show that concentration dropped substantially from 1980 until the late 1990s, when Indonesia experienced the Asian Financial Crisis and the fall of the Suharto government. Since that time, spatial concentration has reversed its previous trend and risen steadily. This is contrary to political economy theories that suggest that strong central governments are associated with increased spatial concentration, but it could be consistent with theories that argue that when political regimes grow more uncertain and fragile, as they did during the fall of Suharto, the returns to locating in a central increase.

We also use new geocoded firm location data to construct the continuous measure of spatial concentration developed by Duranton and Overman (2005). In a small extension of this continuous approach, we develop a new technique which allows us to identify the frequency and locations of manufacturing clusters for 32 industries in Indonesia. Comparing our measures of localization and dispersion with those originally reported by Duranton and Overman (2005) for the United Kingdom, we find that the overall patterns are broadly similar, with localization concentrated within relatively short distances, and dispersion uniformly distributed across distance. However, localization drops off more gradually in Indonesia than in the United Kingdom.

Finally, our analysis sheds light on the correlates of concentration in Indonesia during the past 25 years. We find that the most robust drivers of agglomeration have been natural resources and supply chain linkages, especially with respect to explaining long-term changes in spatial concentration. In the companion paper noted above, we investigate the underlying forces of industrial concentration more formally by means of a new identification strategy to estimate agglomeration externalities and productivity spillovers (Rothenberg et al., 2016a).

References

ADES, A. F. AND E. L. GLAESER (1995): "Trade and Circuses: Explaining Urban Giants," *Quarterly Journal of Economics*, 110, 195–227.

AKITA, T. AND J. J. K. SZETO (2000): "Inpres Desa Tertinggal (IDT) Program and Indonesian Regional Inequality," *Asian Economic Journal*, 14, 167–186.

ALFARO, L. AND M. X. CHEN (2014): "The Global Agglomeration of Multinational Firms," *Journal of International Economics*, 94, 263–276.

ALONSO, W. (1964): *Location and Land Use: Toward a General Theory of Land Rent*, Cambridge: Harvard University Press.

AMITI, M. AND L. CAMERON (2007): "Economic Geography and Wages," *Review of Economics and Statistics*, 89, 15–29.

AMITI, M. AND J. KONINGS (2007): "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia," *American Economic Review*, 97, 1611–1638.

BAZZI, S., A. GADUH, A. D. ROTHENBERG, AND M. WONG (2016): "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia," *American Economic Review*, 106, 2658–98.

BAZZI, S. AND M. GUDGEON (2016): "Local Government Proliferation, Diversity, and Conflict," *Working Paper*.

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.

COMBES, P.-P., T. MAYER, AND J.-F. THISSE (2008): *Economic Geography: The Integration of Regions and Nations*, Princeton University Press.

CRESSIE, N. A. C. (1993): *Statistics for Spatial Data*, New York: John Wiley.

DAVIS, D. R. AND D. E. WEINSTEIN (2002): "Bones, Bombs, and Break Points: The Geography of Economic Activity," *American Economic Review*, 92, 1269–1289.

DEPKES (2014): "Estimasi Penduduk Menurut Umur Tanggal Dan Jenis Kelamin Menurut Kabupaten/Kota, Tahun 2014," Technical report, Pusat Data dan Informasi, Kementerian Kesehatan Republik Indonesia.

DUFLO, E. (2001): "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 91, 795–813.

DURANTON, G. AND H. OVERMAN (2005): "Testing for Localisation Using Micro-Geographic Data," *Review of Economic Studies*, 72, 1077–1106.

DURANTON, G. AND D. PUGA (2004): "Micro-foundations of Urban Agglomeration Economies," in *Handbook of Regional and Urban Economics, Vol. 4*, ed. by J. V. Henderson and J. F. Thisse, Elsevier, 2063–2117.

ELLISON, G. AND E. L. GLAESER (1997): "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 105, 889–927.

——— (1999): "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" *American Economic Review*, 89, 311–316.

FISMAN, R. (2001): "Estimating the Value of Political Connections," *American Economic Review*, 91, 1095–1102.

FITRANI, F., B. HOFMAN, AND K. KAISER (2005): "Unity in diversity? The creation of new local governments in a decentralising Indonesia," *Bulletin of Indonesian Economic Studies*, 41, 57–79.

HELPMAN, E. (1998): "The Size of Regions," in *Topics in Public Economics: Theoretical and Applied Analysis*, ed. by D. Pines, E. Sadka, and I. Zilcha, Cambridge University Press: Cambridge, 33–54.

Hill, H. (1996): *The Indonesian economy since 1966: Southeast Asia's emerging giant*, Cambridge University Press: Cambridge.

——— (2000): *The Indonesian Economy*, Cambridge University Press.

——— (2014): *Regional dynamics in a decentralized Indonesia*, Institute of Southeast Asian Studies.

Hoselitz, B. F. (1955): "Generative and Parasitic Cities," *Economic Development and Cultural Change*, 3, 278–294.

Hsieh, C.-T. and B. A. Olken (2014): "The Missing "Missing Middle"," *Journal of Economic Perspectives*, 28, 89–108.

Kerr, W. R. and S. D. Kominers (2015): "Agglomerative Forces and Cluster Shapes," *Review of Economics and Statistics*, 97, 877–899.

Koh, H.-J. and N. Riedel (2014): "Assessing the Localization Pattern of German Manufacturing and Service Industries: A Distance-based Appoach," *Regional Studies*, 48, 823–843.

Krugman, P. (1991): "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483–499.

Marshall, A. (1890): *Principles of Economics*, London: Macmillan and Co.

Mills, E. (1967): "An Aggregative Model of Resource Allocation in a Metropolitan Area," *American Economic Review*, 57, 197–210.

Murata, Y., R. Nakajima, R. Okamoto, and R. Tamura (2014): "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach," *Review of Economics and Statistics*, 96, 967–985.

Muth, R. (1969): *Cities and Housing: the Spatial Pattern of Urban Residential Land Use*, Chicago: University of Chicago Press.

Rosenthal, S. S. and W. C. Strange (2001): "The Determinants of Agglomeration," *Journal of Urban Economics*, 50, 191–229.

Rothenberg, A. D. (2013): "Transport Infrastructure and Firm Location Choice in Equilibrium: Evidence from Indonesia's Highways," Working Paper.

Rothenberg, A. D., S. Bazzi, A. Chari, and S. Nataraj (2016a): "Estimating Productivity Spillovers Using the Structure of Production Networks: Evidence from Indonesia," Working Paper.

——— (2016b): "When Regional Policies Fail: An Evaluation of Indonesia's Integrated Economic Development Zones," Working Paper.

Rothenberg, A. D., A. Gaduh, N. E. Burger, C. Chazali, I. Tjandraningsih, R. Radikun, C. Sutera, and S. Weilant (2016c): "Rethinking Indonesia's Informal Sector," *World Development*, 80, 96–113.

Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis*, New York: Chapman and Hill.

Vincenty, T. (1975): "Direct and Inverse Solutions of Geodesics on the Ellipsoid with Application of Nested Equations," *Survey Review*, 23, 88–93.

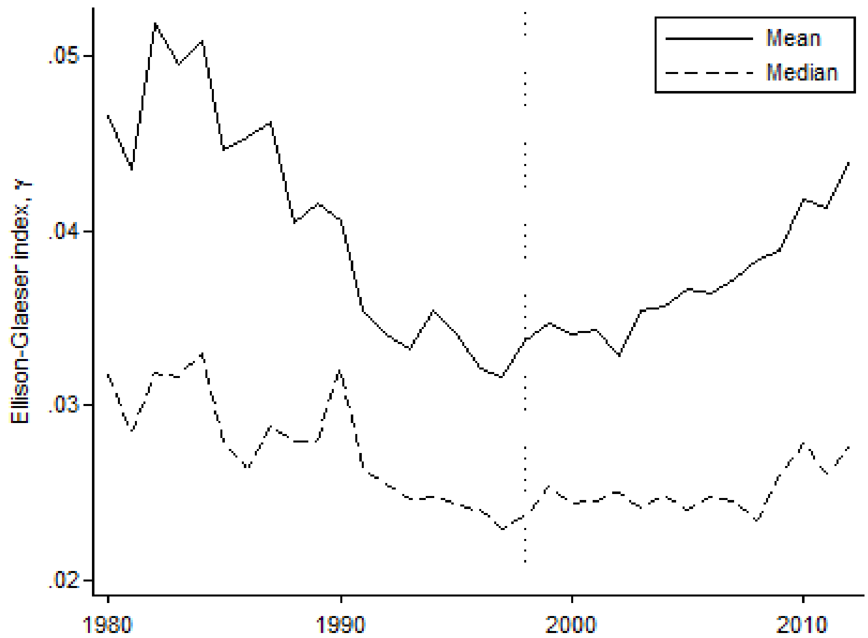
von Thünen, J. H. (1826): *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, Hamburg: Perthes.

Yule, G. U. and M. G. Kendall (1950): *An Introduction to the Theory of Statistics*, London: Griffin.

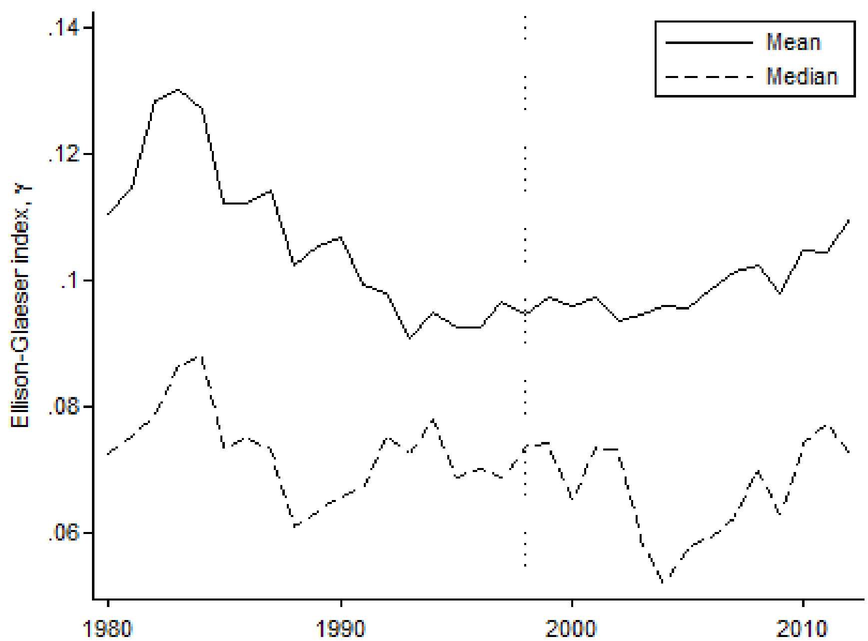
Figures

Figure 1: Trends in Spatial Concentration Across Industries

(A) γ, Using Districts



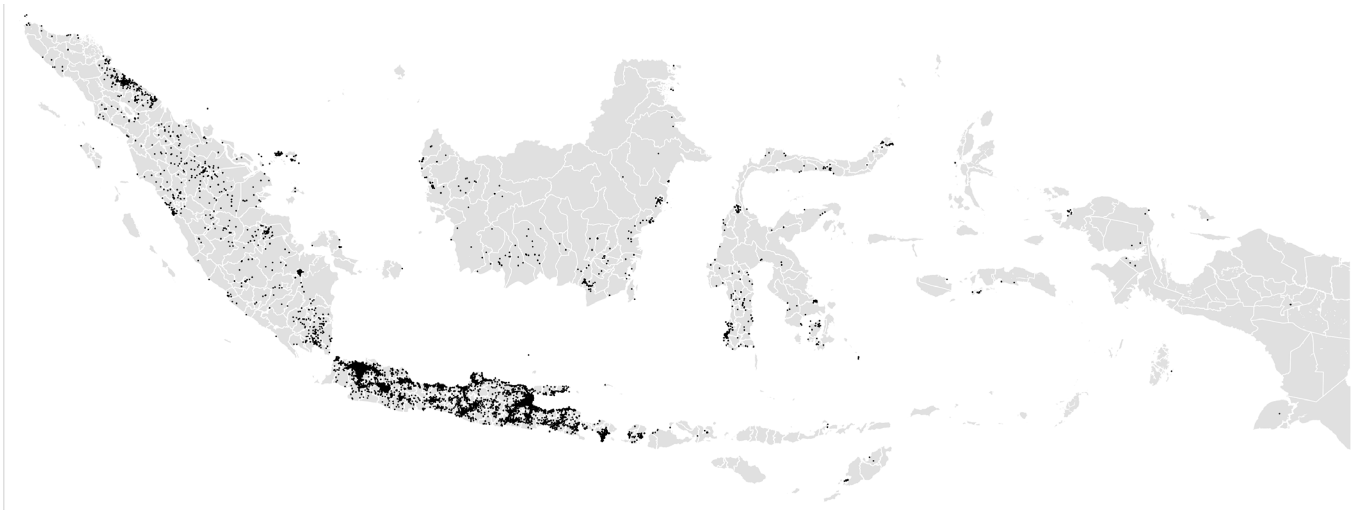
(B) γ, Using Provinces



Source: SI data and authors' calculations. These figures plot the mean (and median) of γ over 4-digit industries with at least 10 firms per industry. Panel A uses districts (*kabupaten*) as the spatial unit of analysis, while Panel B uses provinces as the spatial unit of analysis.

Figure 2: Locations of Firms in Indonesia: Selected Industries

(A) 22,000+ ESTABLISHMENTS IN 2013



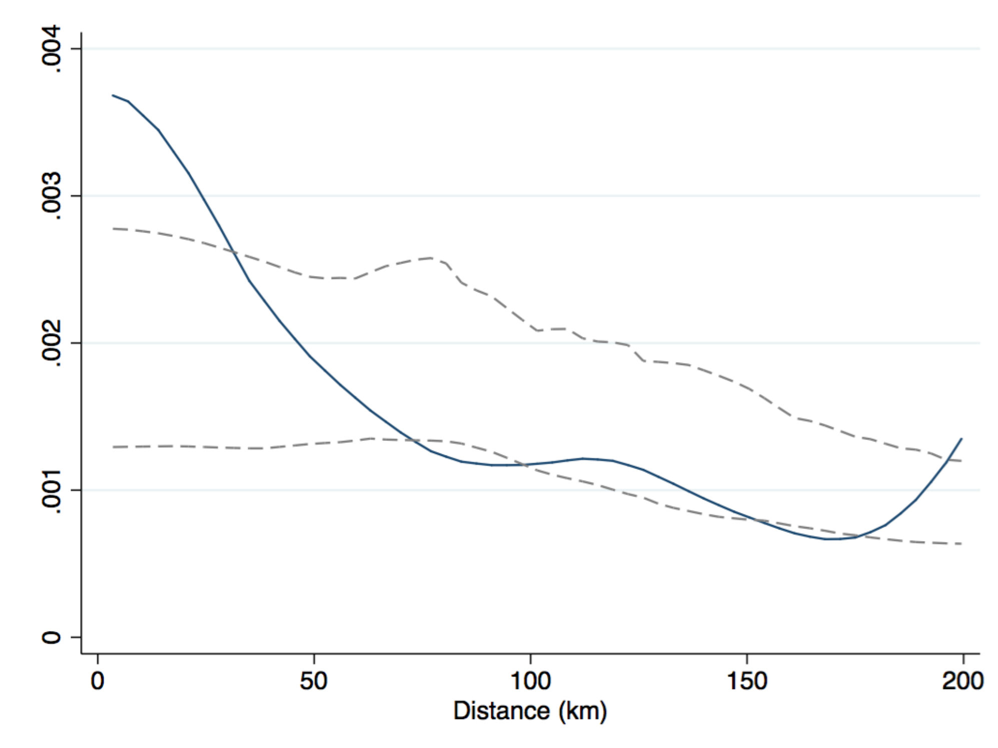
(B) FURNITURE (WOOD, BAMBOO, OR RATTAN) (ISIC 3321)

(C) MANUFACTURE OF COOKING OIL (ISIC 3115)

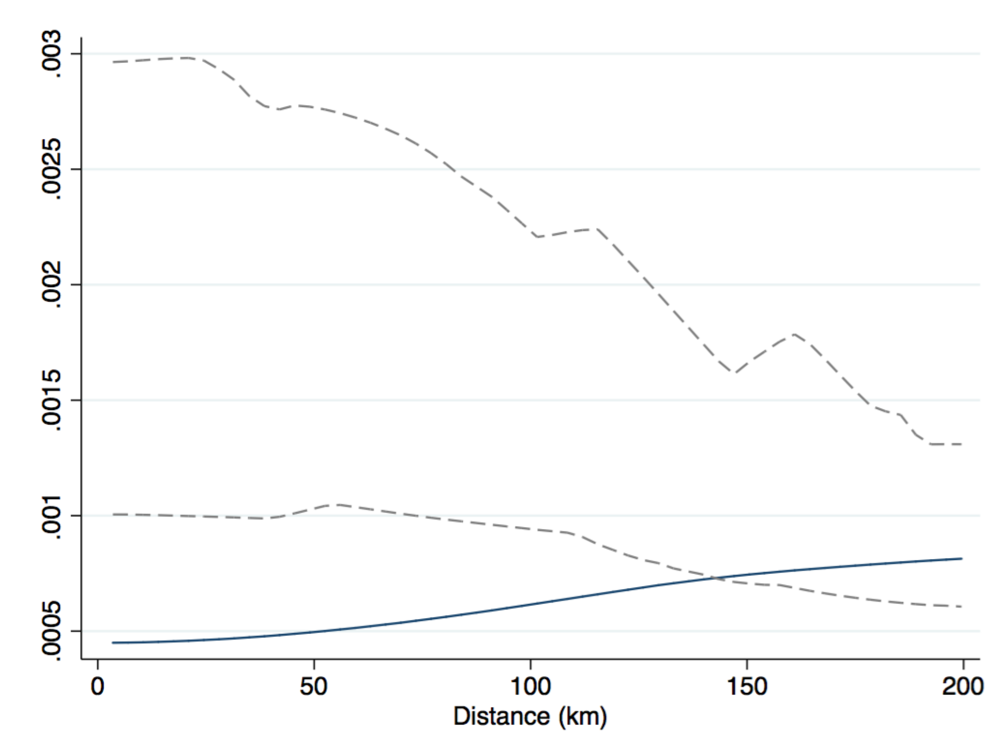
Source: DIBS (2013) and authors' calculations.

Figure 3: K-density and Local Confidence Intervals for Selected 4-Digit Industries

(A) Furniture (Wood, Bamboo, or Rattan) (ISIC 3321)



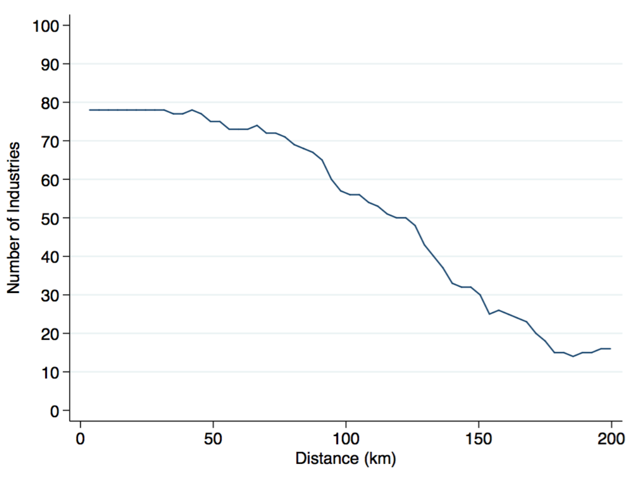
(B) Manufacture of Cooking Oil (ISIC 3115)



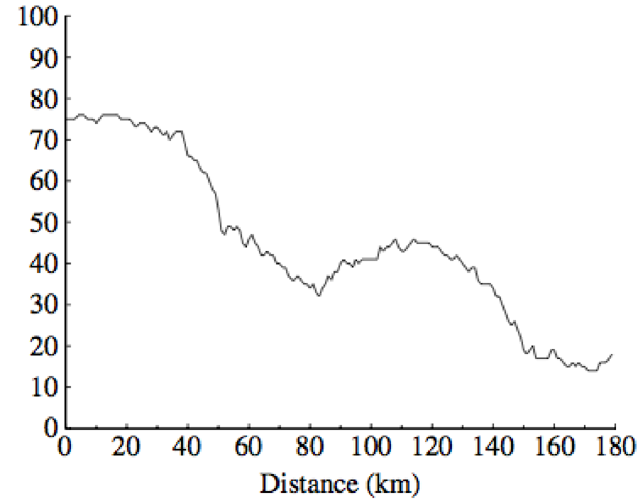
Source: DIBS (2013) and authors' calculations.

Figure 4: Number of Four-Digit Industries with Localization and Dispersion, by km

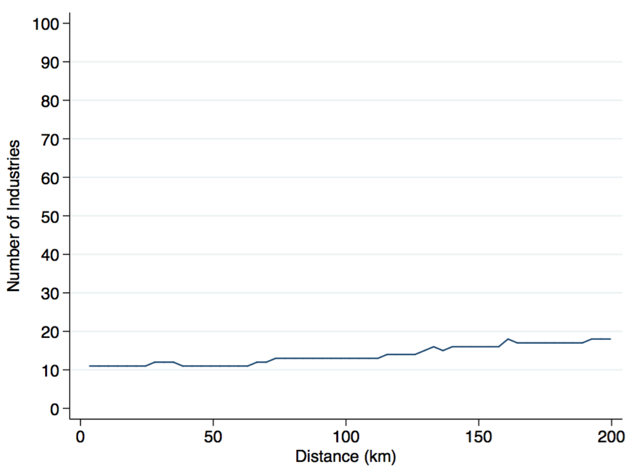
(A) Localization, Indonesia (2013)



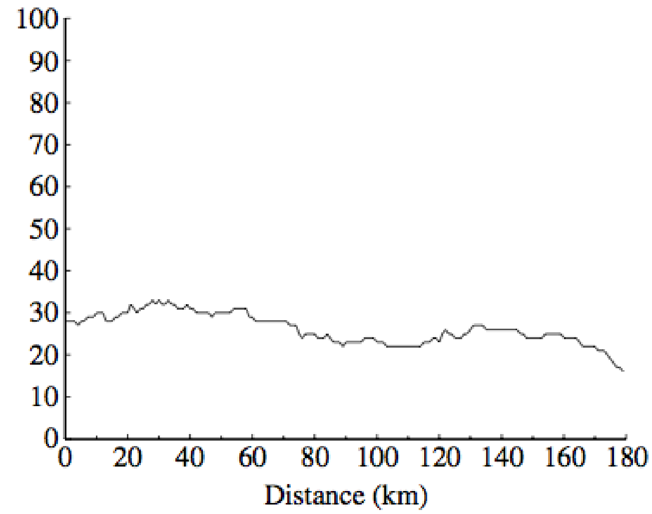
(C) Localization, United Kingdom (1996)



(B) Dispersion, Indonesia (2013)



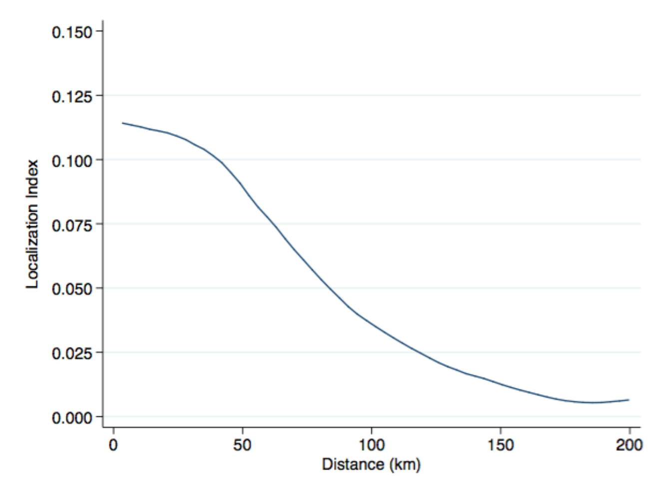
(D) Dispersion, United Kingdom (1996)



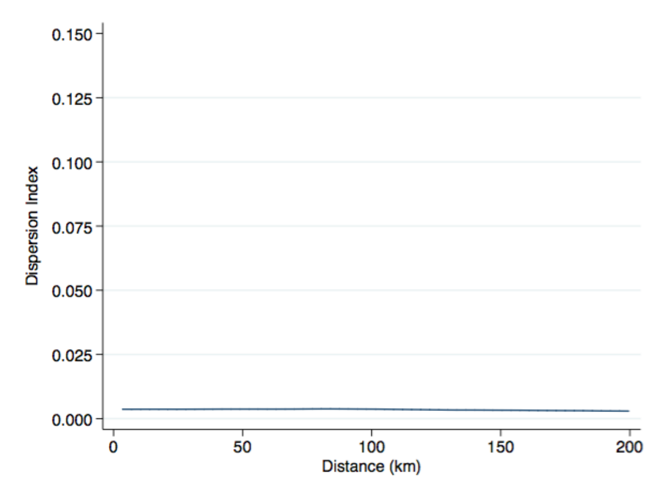
Source: Panel A and B are based on DIBS (2013) data and use authors' calculations. Panel C and D are taken from Duranton and Overman (2005), Figure 3.

Figure 5: Index of Localization and Dispersion by Distance

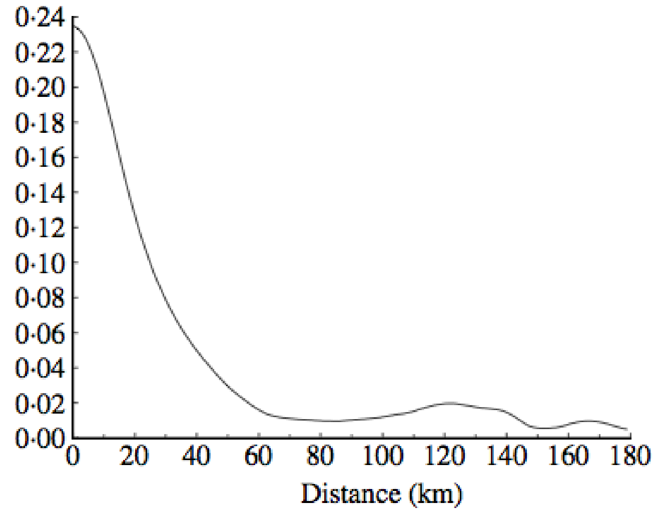
(A) Localization, Indonesia (2013)



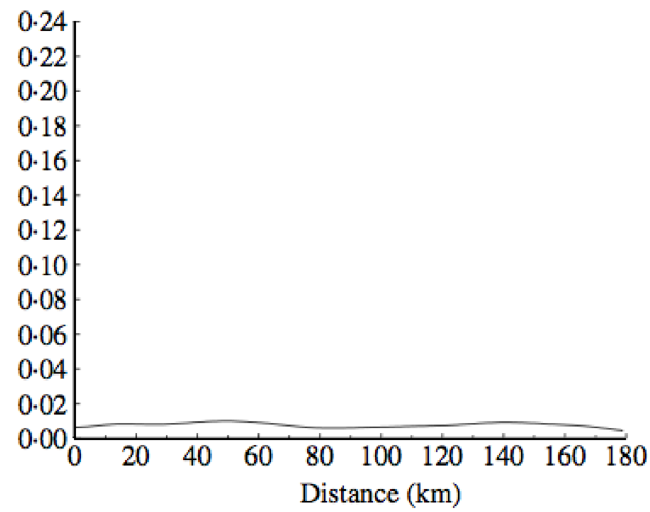
(B) Dispersion, Indonesia (2013)



(C) Localization, United Kingdom (1996)



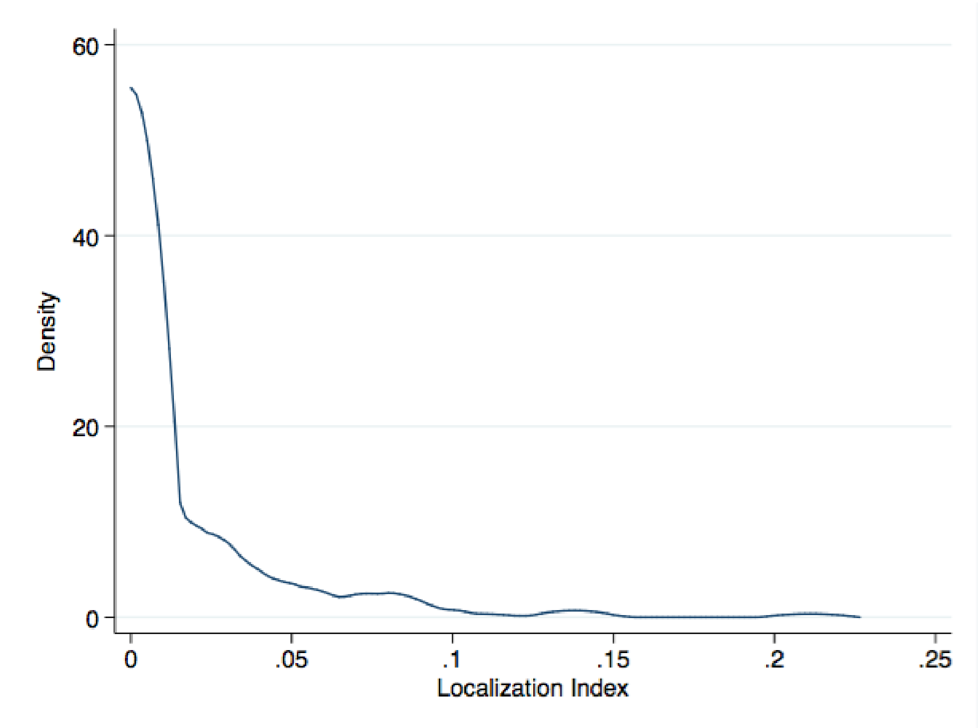
(D) Dispersion, United Kingdom (1996)



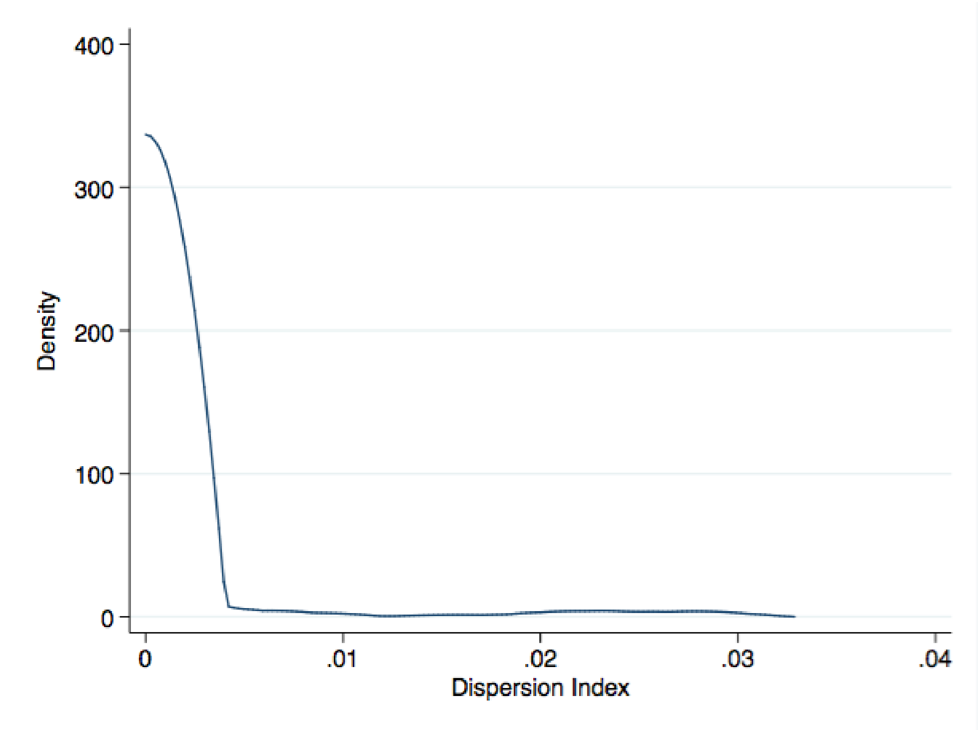
Source: Panel A and B are based on DIBS (2013) data and use authors' calculations. Panel C and D are taken from Duranton and Overman (2005), Figure 3.

Figure 6: Distribution of Localization and Dispersion Across Industries

(A) Localization

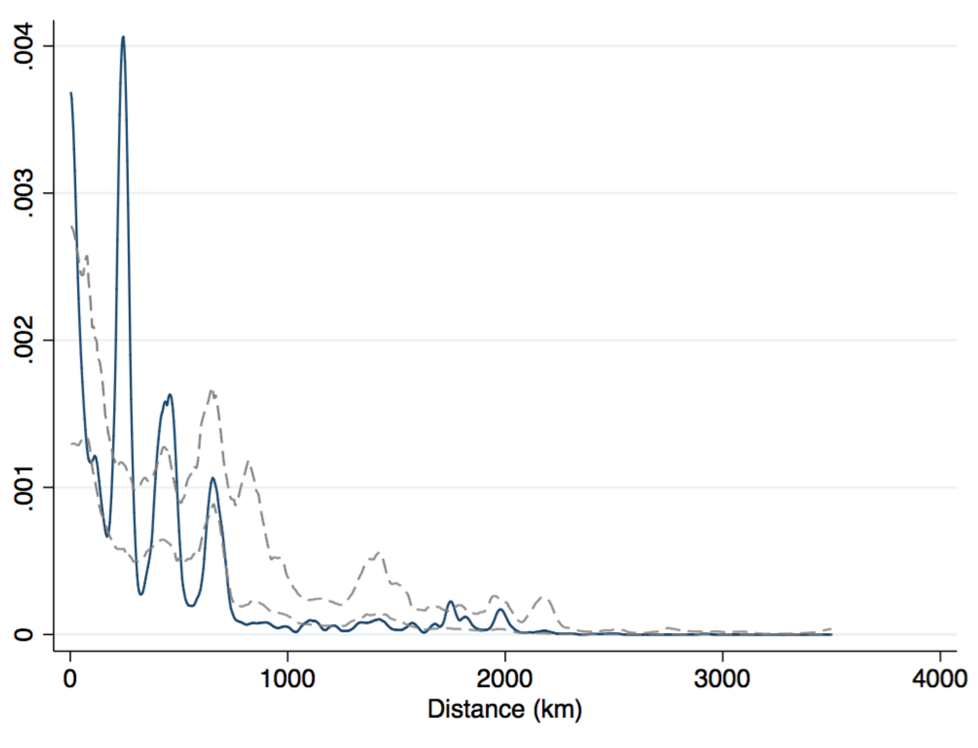


(B) Dispersion



Source: DIBS (2013) and authors' calculations.

Figure 7: K-density and Local Confidence Intervals for Furniture (Wood, Bamboo, or tan) (ISIC 3321)



Source: DIBS (2013) and authors' calculations.

Tables

Table 1: Most and Least Concentrated Industries, Using the Ellison and Glaeser (1997) Index

Panel A: Most Concentrated		1982		1992		2002		2012	
ISIC	Description	γ	N	γ	N	γ	N	γ	N
3216	Kapok Manufacturing	0.217	39	0.275	80	0.321	67	0.654	126
3906	Manufacture of Stationairies			0.025	68	0.056	124	0.240	184
3121	Tapioca Flour, Sago, Cassava Flour, and Other Flour	0.105	144	0.079	210	0.104	195	0.195	189
3841	Ship Building and Repairing	0.119	54	0.046	164	0.039	147	0.174	132
3642	Roofing Tiles	0.143	151	0.173	621	0.217	718	0.163	768
3832	Radio, Tv, Cassettes, Etc.	0.039	35	0.042	114	0.147	218	0.154	213
3633	Lime	0.085	69	0.035	94	0.031	99	0.123	98
3213	Knitting Mills	0.051	120	0.026	247	0.120	312	0.113	333
3312	Wooden Boxes and Containers	0.130	12	0.030	34	0.061	48	0.098	39
3142	Clove Cigarettes	0.041	175	0.059	154	0.055	207	0.070	349
3611	Ceramics and Porcelain	-0.002	25	0.036	82	0.069	84	0.068	58
3843	Motor Vehicle Assembly and Manufacturing	0.145	55	0.058	206	0.076	275	0.068	311
3229	Wearing Apparels Not Elsewhere Classified	0.129	16	0.012	60	0.006	74	0.067	80
3149	Manufacture of Other Tobacco Products	0.024	20	0.050	27	0.044	25	0.065	21
3122	Tea Processing	0.151	56	0.141	201	0.029	205	0.062	209
3314	Handicraft and Wood Carving	0.486	34	0.257	71	0.032	91	0.062	77
3523	Household Cleaning and Cosmetics	0.069	99	0.042	140	0.055	125	0.061	133
3116	Grain and Bean Products (Rice, Coffee)	0.010	227	0.083	520	0.092	512	0.061	654
3845	Bicycle, Becak Assembly/Manufacturing	0.061	33	0.031	68	0.007	68	0.060	42
3212	Made Up Textile Goods Except Wearing Apparel	0.014	130	0.029	347	0.049	367	0.057	425
Panel B: Least Concentrated		**1982**		**1992**		**2002**		**2012**	
ISIC	**Description**	γ	N	γ	N	γ	N	γ	N
3641	Bricks	0.097	77	-0.001	16	-0.064	38	-0.084	34
3902	Musical Instruments					-0.050	20	-0.041	34
3214	Carpets and Rugs			-0.068	10	-0.037	16	-0.038	18
3233	Products of Leather and Substitutes			0.006	98	0.030	121	-0.021	151
3112	Milk Products	0.036	27	0.018	32	0.025	41	-0.008	47
3909	Other Manufacturing Industries	0.009	42	0.035	72	0.021	109	-0.007	82
3131	Alcoholic Liquors							-0.005	12
3119	Chocolate Powder, Chocolate, and Sugar Products	-0.006	60	-0.004	98	-0.003	102	-0.001	87
3411	Paper and Paper Products	0.015	55	0.010	122	-0.007	132	-0.000	149
3529	Chemical Products Not Elsewhere Classified	-0.028	53	0.016	95	-0.003	141	0.000	140
3126	Coffee Powder and Fried (?)	-0.013	16	-0.002	27	0.001	40	0.001	60
3541	Products of Petroleum Refineries					-0.008	21	0.001	34
3811	Agricultural, Carpentry, and Metal Cutting Tools	0.040	110	0.026	154	0.018	221	0.002	184
3111	Slaughtering and Preserving Meat	0.001	14	0.002	35	-0.004	35	0.004	64
3632	Goods Made From Cement	0.001	286	0.002	385	0.002	304	0.004	305
3812	Furniture and Fixtures Primarily of Metal Products	0.018	48	0.009	75	0.006	128	0.005	149
3117	Products from Flour	0.006	379	0.004	665	0.001	814	0.006	956
3215	Cordage and Twine	-0.009	18	-0.017	36	-0.002	35	0.006	60
3560	Plastic Wares	0.033	282	0.013	802	0.007	1090	0.011	1252
3113	Canning, Preserving, Processing of Fruits / Vegetables			0.012	43	-0.008	62	0.011	85

Authors' calculations.

Table 2: Comparing Duranton and Overman and Ellison and Glaeser Measures of Industrial Concentration

Panel A: Most Concentrated		**2012**		**Localized at $d \in$ ...**							
ISIC	**Description**	γ	N	Γ	≤ 5	$(5, 10]$	$(10, 25]$	$(25, 50]$	$(50, 75]$	$(75, 100]$	N
3216	Kapok Manufacturing	0.654	126	0.156	1	1	1	1	0	0	122
3844	Motor Cycles and Three-Wheeled Motor Vehicles	0.029	146	0.116	1	1	1	1	1	1	158
3214	Carpets and Rugs	-0.038	18	0.107	1	1	1	1	1	1	21
3142	Clove Cigarettes	0.070	349	0.092	1	1	1	1	1	1	348
3843	Motor Vehicle Assembly and Manufacturing	0.068	311	0.090	1	1	1	1	1	1	285
3213	Knitting Mills	0.113	333	0.082	1	1	1	1	1	1	352
3513	Resin, Plastic Material, and Synthetic Fibre	0.039	64	0.075	1	1	1	1	1	1	86
3240	Footwear	0.045	433	0.071	1	1	1	1	1	1	444
3833	Manufacture of Electrical Apparatus and Supplies	0.046	304	0.064	1	1	1	1	1	1	273
3832	Radio, Tv, Cassettes, Etc.	0.154	213	0.062	1	1	1	1	1	1	158
3642	Roofing Tiles	0.163	768	0.058	1	1	1	1	1	1	733
3212	Made Up Textile Goods Except Wearing Apparel	0.057	425	0.056	1	1	1	1	1	1	540
3904	Toys Manufacturing	0.032	119	0.053	1	1	1	1	1	1	118
3141	Drying and Processing Tobacco	0.039	527	0.050	0	0	0	0	0	0	485
3221	Wearing Apparels	0.026	1951	0.047	1	1	1	1	1	1	1754
3523	Household Cleaning and Cosmetics	0.061	133	0.037	1	1	1	1	1	1	126
3820	Repair Including Machineries and Sewing Repair	0.048	357	0.034	1	1	1	1	1	1	133
3521	Paint, Varnisher, Laquers	0.015	122	0.027	1	1	1	1	1	1	117
3811	Agricultural, Carpentry, and Metal Cutting Tools	0.002	184	0.024	1	1	1	1	1	1	249
3611	Ceramics and Porcelain	0.068	58	0.023	1	1	1	1	1	1	47
Panel B: Most Dispersed		**2012**		**Dispersed at $d \in$ ...**							
ISIC	**Description**	γ	N	Ψ	≤ 5	$(5, 10]$	$(10, 25]$	$(25, 50]$	$(50, 75]$	$(75, 100]$	N
3115	Manufacture of Cooking Oils	0.031	623	0.016	1	1	1	1	1	1	516
3117	Products from Flour	0.006	956	0.010	0	0	0	1	1	1	917
3114	Canning, Preserving, Processing of Seafood	0.024	888	0.010	1	1	1	1	1	1	647
3311	Saw Mills and Wood Processing	0.027	733	0.007	1	1	1	1	1	1	647
3621	Glass and Glass Products	0.013	80	0.005	0	0	0	0	0	0	145
3841	Ship Building and Repairing	0.174	132	0.005	1	1	1	1	1	1	125
3631	Cement	0.015	23	0.004	1	1	1	1	1	1	17
3113	Canning, Preserving, Processing of Fruits / Vegetables	0.011	85	0.002	1	1	1	1	1	1	97
3552	Industrial Rubber	0.022	277	0.001	1	1	1	1	1	1	247
3710	Basic Iron and Steel	0.016	245	0.001	0	0	0	0	0	0	218
3313	Bambo, Rattan, and Willow Pleats	0.019	148	0.001	1	1	1	1	0	0	153
3411	Paper and Paper Products	-0.000	149	0.000	0	0	0	0	0	0	145

Authors' calculations.

Table 3: Number of Clusters and Cluster Locations

ISIC	Description	C_A	Γ	Cluster Locations
3216	Kapok Manufacturing	3	0.156	C. Java (Jepara), E. Java (Banyuwangi, Pasuruan)
3844	Motor Cycles and Three-Wheeled Motor Vehicles	2	0.116	DKI Jakarta, E. Java (Mojokerto)
3214	Carpets and Rugs	4	0.107	W. Java (Bandung, Depok, Karawang), C. Java (Blora)
3142	Clove Cigarettes	4	0.092	C. Java (Kudus), E. Java (Pamekasan, Jombang), N. Sumatra (Pematangsiantar)
3843	Motor Vehicle Assembly and Manufacturing	2	0.09	DKI Jakarta, E. Java (Nganjuk)
3213	Knitting Mills	3	0.082	W. Java (Cimahi), C. Java (Sragen), Banten (Tangerang)
3513	Resin, Plastic Material, and Synthetic Fibre	3	0.075	W. Java (Karawang), E. Java (Mojokerto), Banten (Tangerang)
3240	Footwear	2	0.071	E. Java (Sidoarjo), Banten (Tangerang)
3833	Manufacture of Electrical Apparatus and Supplies	2	0.064	W. Java (Karawang), Riau (Karimun)
3832	Radio, Tv, Cassettes, Etc.	2	0.062	W. Java (Subang), Aceh (Simeulue)
3642	Roofing Tiles	4	0.058	W. Java (Bogor, Majalengka), E. Java (Sampang), N. Sumatra (Deli Serdang)
3212	Made Up Textile Goods Except Wearing Apparel	3	0.056	W. Java (Bandung), E. Java (Ngawi), Riau (Bengkalis)
3904	Toys Manufacturing	1	0.053	W. Java (Subang)
3221	Wearing Apparels	4	0.047	DKI Jakarta, C. Java (Pemalang, Semarang), E. Java (Probolinggo)
3523	Household Cleaning and Cosmetics	2	0.037	W. Java (Indramayu), N. Sumatra (Labuhanbatu)
3820	Repair Including Machineries and Sewing Repair	1	0.034	W. Java (Karawang)
3521	Paint, Varnisher, Laquers	1	0.027	W. Java (Indramayu)
3811	Agricultural, Carpentry, and Metal Cutting Tools	2	0.024	DKI Jakarta, E. Java (Surabaya)
3611	Ceramics and Porcelain	2	0.023	DKI Jakarta, E. Java (Sidoarjo)
3233	Products of Leather and Substitutes	2	0.016	E. Java (Mojokerto), Banten (Tangerang)
3813	Structural Metal Products	2	0.016	DKI Jakarta, E. Java (Surabaya)
3211	Spinning, Weaving, and Finished Textiles	4	0.014	W. Java (Purwakarta), C. Java (Temanggung), E. Java (Sampang), N. Sumatra (Padang Lawas)
3560	Plastic Wares	3	0.012	DKI Jakarta, E. Java (Bojonegoro), N. Sumatra (Labuhanbatu)
3412	Containers and Boxes of Paper and Paper Board	2	0.01	E. Java (Mojokerto), Lampung (Lampung Tengah)
3901	Jewelry	3	0.01	C. Java (Brebes), E. Java (Banyuwangi), N. Sumatra (Simalungun)
3420	Printing, Publishing and Allied Industries	2	0.008	DKI Jakarta, E. Java (Gresik)
3125	Krupuk, Emping, Karak and Other Chips	3	0.006	W. Java (Kuningan), E. Java (Pasuruan), N. Sumatra (Karo)
3321	Furniture Made of Wood, Bamboo, Or Rattan	4	0.005	W. Java (Karawang), C. Java (Demak), E. Java (Sampang), N. Sumatra (Labuhanbatu Utara)
3229	Wearing Apparels Not Elsewhere Classified	3	0.003	W. Java (Bandung Barat), C. Java (Klaten), E. Java (Pasuruan)
3850	Measuring, Optical and Photographic Equip.	1	0.003	W. Java (Karawang)
3149	Manufacture of Other Tobacco Products	2	0.002	C. Java (Banjarnegara), E. Java (Batu)
3633	Lime	2	0.000	W. Java (Bogor), E. Java (Lamongan)

Authors' calculations.

Table 4: Explaining the Level of Spatial Concentration in 2012 Based on the Ellison and Glaeser (1997) γ Index

	Spatial Unit of Analysis: Province		District	
	(1)	(2)	(3)	(4)
Net Productivity Measure (Avg., 1990-1996)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)
Management Ratio (Avg., 1980-1996)	3.186 (1.529)**	2.701 (1.418)*	0.164 (0.182)	0.404 (0.234)*
Share of workers with no schooling (1996)	0.392 (0.414)	0.332 (0.589)	0.418 (0.216)*	0.696 (0.287)**
Share of workers w/ at least bachelors degree (1996)	-17.530 (8.379)**	-17.530 (8.228)**	-2.140 (0.789)***	-2.595 (0.994)**
Total water expenses (1996)	0.000 (0.000)**	0.000 (0.000)**	0.000 (0.000)	0.000 (0.000)
Avg Share of Value Added, SOEs (1980-1996)	1.615 (0.612)**	1.718 (0.682)**	0.483 (0.279)*	0.486 (0.211)**
Avg Share of Number of Firms, SOEs (1980-1996)	-5.235 (2.298)**	-5.195 (2.397)**	-1.015 (0.455)**	-1.154 (0.441)**
Inventory / Output (Avg., 1992-1996)	1.045 (0.575)*	1.092 (0.616)*	1.007 (0.321)***	0.893 (0.223)***
RND Expenses / Output (Avg., 1992-1996)	25.363 (53.832)	25.739 (55.520)	33.477 (13.856)**	41.483 (16.315)**
Natural Resources, % of Raw Material Value (2000)	1.107 (0.709)	0.915 (0.774)	1.189 (0.424)***	0.770 (0.268)***
Energy, % of Raw Material Value (2000)	0.941 (0.689)	0.591 (0.812)	1.106 (0.396)***	0.610 (0.290)**
Manufactured Inputs, % of Raw Material Value (2000)	1.203 (0.683)*	0.998 (0.770)	1.165 (0.401)***	0.678 (0.274)**
N	93	93	93	93
Adjusted R^2	0.405	0.376	0.221	0.359
F-Stat	2.214	.	5.765	.
2-Digit ISIC FE	No	Yes	No	Yes

Authors' calculations. Robust standard errors in parentheses. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 5: Explaining the Level of Spatial Concentration in 2013 Based on the Duranton and Overman (2005) Γ Index

	(1)	(2)
Net Productivity Measure (Avg., 1990-1996)	-0.000 (0.000)	0.000 (0.000)
Management Ratio (Avg., 1980-1996)	-0.064 (0.087)	-0.092 (0.106)
Share of workers with no schooling (1996)	0.065 (0.076)	0.153 (0.098)
Share of workers w/ at least bachelors degree (1996)	-0.035 (0.544)	0.201 (0.533)
Total water expenses (1996)	0.000 (0.000)	0.000 (0.000)
Avg Share of Value Added, SOEs (1980-1996)	0.013 (0.116)	0.005 (0.107)
Avg Share of Number of Firms, SOEs (1980-1996)	-0.145 (0.185)	-0.129 (0.194)
Inventory / Output (Avg., 1992-1996)	0.117 (0.098)	0.099 (0.089)
RND Expenses / Output (Avg., 1992-1996)	4.968 (10.074)	8.411 (9.969)
Natural Resources, % of Raw Material Value (2000)	0.236 (0.119)*	0.123 (0.103)
Energy, % of Raw Material Value (2000)	0.208 (0.111)*	0.073 (0.113)
Manufactured Inputs, % of Raw Material Value (2000)	0.256 (0.111)**	0.121 (0.108)
N	87	87
Adjusted R^2	0.005	0.126
F-Stat	2.465	.
2-Digit ISIC FE	No	Yes

Authors' calculations. Robust standard errors in parentheses. */**/*** denotes significant at the 10% / 5% / 1% levels.

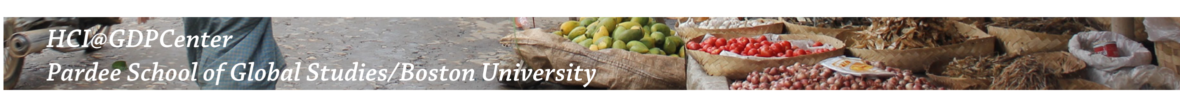


Table 6: Explaining Changes in Spatial Concentration 1985–2012 based on the Ellison and Glaeser (1997) γ Index

Dependent Variable: $\Delta\ \gamma$	Spatial Unit of Analysis			
	Province		District	
	(1)	(2)	(3)	(4)
Net Productivity Measure (Avg., 1990-1996)	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Management Ratio (Avg., 1980-1996)	1.642	1.661	-0.100	0.182
	(1.101)	(1.146)	(0.184)	(0.273)
Share of workers with no schooling (1996)	0.506	0.597	0.096	0.329
	(0.416)	(0.459)	(0.280)	(0.309)
Share of workers w/ at least bachelors degree (1996)	-6.707	-6.384	0.929	0.530
	(5.653)	(6.176)	(0.971)	(1.129)
Total water expenses (1996)	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Avg Share of Value Added, SOEs (1980-1996)	0.483	0.618	0.351	0.303
	(0.696)	(0.891)	(0.310)	(0.272)
Avg Share of Number of Firms, SOEs (1980-1996)	-2.475	-2.579	-0.354	-0.478
	(1.616)	(1.855)	(0.479)	(0.498)
Inventory / Output (Avg., 1992-1996)	0.110	-0.007	0.029	-0.039
	(1.028)	(1.189)	(0.406)	(0.410)
RND Expenses / Output (Avg., 1992-1996)	-12.272	-19.390	-4.038	8.856
	(61.369)	(70.441)	(21.864)	(21.979)
Natural Resources, % of Raw Material Value (2000)	5.192	4.952	2.757	2.553
	(1.318)***	(1.777)***	(0.502)***	(0.508)***
Energy, % of Raw Material Value (2000)	5.479	5.085	2.849	2.749
	(1.530)***	(1.989)**	(0.466)***	(0.561)***
Manufactured Inputs, % of Raw Material Value (2000)	5.194	4.950	2.714	2.461
	(1.303)***	(1.794)***	(0.485)***	(0.504)***
N	88	88	88	88
Adjusted R^2	0.246	0.171	0.290	0.378
F-Stat	69.675	.	80.789	.
2-Digit ISIC FE	No	Yes	No	Yes

Authors' calculations. Robust standard errors in parentheses. */**/*** denotes significant at the 10% / 5% / 1% levels.

A Appendix

A.1 Geocoding the 2013 Directory of Manufacturers

To measure the extent of spatial concentration of manufacturing in Indonesia, we use detailed location information on large manufacturers from Indonesia's 2013 Directory of Manufacturers (*Direktori Industri Manufaktur*, abbreviated as DIM). The DIM dataset, produced by BPS and made publicly available, contains address information for the headquarters of 23,122 manufacturing firms with more than 20 employees.

To use the address information to obtain latitude and longitude coordinates for each of these firms, we used a STATA routine that sends address strings to the Google's Geocoding API Version 3 and returns coordinate information. The STATA routine is `geocode3`.

We found that our results depended partly on the way address-level information was passed to the API. In particular, we tried different combinations of passing the names of company, address, and zip together, or just the address and zip alone. We are exploring the robustness of our findings to different mapping APIs (MapQuest and Nokia Here).

In the end, 22,071 coordinates were generated successfully, accounting for over 96% of the total. The remaining firms for which we could not find addresses were due to the invalidity of the address or limitations of the Google Map. Four coordinates seem to be outliers and we will further examine them later.

The only thing need to notice is that the Google Geocoding API has a daily query limit of 2500 per IP-address. When the limit is reached, the STATA will return "`Over_Query_Limit`" and stop the program. Because of this, it took us about 9 days to retrieve the coordinates of every firm.

A.2 K-Means Clustering Algorithm

To describe the K-means clustering algorithm, let $(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)$ denote the set of locations of firms in industry A, and let C_A denote the number of clusters, identified by the process described in Section 3.4. The clustering algorithm proceeds as follows:

1. First, we choose C_A initial locations for the clusters by randomly picking C_A firms from the set of total firms.
2. Each other firm is assigned to the initial cluster point that it is closest to, using Euclidean distances to determine which cluster is closest.
3. The centroid of each of the resulting C_A clusters becomes the new mean.
4. Repeat steps 2 and 3 until convergence.

A.3 Appendix Tables and Figures

Table A.1: Most and Least Concentrated Industries, Using the Ellison and Glaeser (1997) Index

		1982		1992		2002		2012	
ISIC	Description	γ	N	γ	N	γ	N	γ	N
3111	Slaughtering and Preserving Meat	0.001	14	0.002	35	-0.004	35	0.004	64
3112	Milk Products	0.036	27	0.018	32	0.025	41	-0.008	47
3113	Canning, Preserving, Processing of Fruits / Vegetables			0.012	43	-0.008	62	0.011	85
3114	Canning, Preserving, Processing of Seafood	0.105	55	0.022	320	0.017	542	0.024	888
3115	Manufacture of Cooking Oils	0.021	125	0.063	255	0.057	324	0.031	623
3116	Grain and Bean Products (Rice, Coffee)	0.010	227	0.083	520	0.092	512	0.061	654
3117	Products from Flour	0.006	379	0.004	665	0.001	814	0.006	956
3118	Sugar Products	0.006	58	0.012	106	0.012	104	0.030	115
3119	Chocolate Powder, Chocolate, and Sugar Products	-0.006	60	-0.004	98	-0.003	102	-0.001	87
3121	Tapioca Flour, Sago, Cassava Flour, and Other Flour	0.105	144	0.079	210	0.104	195	0.195	189
3122	Tea Processing	0.151	56	0.141	201	0.029	205	0.062	209
3123	Manufacture of Ice	0.004	133	0.014	182	0.016	174		
3124	Soybean Products	0.003	91	0.020	237	0.077	256	0.033	327
3125	Krupuk, Emping, Karak and Other Chips	0.105	87	0.050	399	0.047	554	0.042	924
3126	Coffee Powder and Fried (?)	-0.013	16	-0.002	27	0.001	40	0.001	60
3127	Seasoning	0.052	22	0.039	172	0.020	249	0.017	328
3128	Cattle Food	0.028	37	0.013	100	0.017	94	0.039	103
3129	Other (Food) Products	0.037	66	0.024	24				
3131	Alcoholic Liquors							-0.005	12
3132	Wine	0.048	11	-0.042	13	0.053	14		
3134	Soft Drinks and Carbonated Waters	0.012	73	0.012	153	0.031	233	0.019	305
3141	Drying and Processing Tobacco	0.129	351	0.077	694	0.041	563	0.039	527
3142	Clove Cigarettes	0.041	175	0.059	154	0.055	207	0.070	349
3143	Cigarettes Manufacturing	0.027	29	0.040	31	0.016	28	0.043	38
3149	Manufacture of Other Tobacco Products	0.024	20	0.050	27	0.044	25	0.065	21
3211	Spinning, Weaving, and Finished Textiles	0.030	1422	0.042	1287	0.048	1098	0.053	1289
3212	Made Up Textile Goods Except Wearing Apparel	0.014	130	0.029	347	0.049	367	0.057	425
3213	Knitting Mills	0.051	120	0.026	247	0.120	312	0.113	333
3214	Carpets and Rugs			-0.068	10	-0.037	16	-0.038	18
3215	Cordage and Twine	-0.009	18	-0.017	36	-0.002	35	0.006	60
3216	Kapok Manufacturing	0.217	39	0.275	80	0.321	67	0.654	126
3219	Textile Not Elsewhere Classified			0.005	33	0.009	105	0.035	209
3221	Wearing Apparels	0.067	228	0.035	1651	0.038	1832	0.026	1951
3229	Wearing Apparels Not Elsewhere Classified	0.129	16	0.012	60	0.006	74	0.067	80
3231	Leather Tanneries and Finishing	-0.006	38	0.013	68	0.005	62	0.024	69
3233	Products of Leather and Substitutes			0.006	98	0.030	121	-0.021	151
3240	Footwear	-0.013	59	0.060	320	0.046	346	0.045	433
3311	Saw Mills and Wood Processing	0.037	473	0.034	1106	0.033	1218	0.027	733
3312	Wooden Boxes and Containers	0.130	12	0.030	34	0.061	48	0.098	39
3313	Bambo, Rattan, and Willow Pleats	0.103	16	0.086	59	0.024	99	0.019	148
3314	Handicraft and Wood Carving	0.486	34	0.257	71	0.032	91	0.062	77
3319	Wood Products Not Elsewhere Classified	-0.009	39	0.019	168	0.016	192	0.019	127
3321	Furniture Made of Wood, Bamboo, Or Rattan	0.010	130	0.017	705	0.027	1295	0.040	1223
3323	All Kinds of Matras			-0.024	12	-0.012	23	0.032	78
3411	Paper and Paper Products	0.015	55	0.010	122	-0.007	132	-0.000	149
3412	Containers and Boxes of Paper and Paper Board	0.040	23	0.031	98	0.024	163	0.017	233
3419	Pulp and Paper Products Not Elsewhere Classified	-0.018	10	0.012	30	0.030	42	0.034	51
3420	Printing, Publishing and Allied Industries	0.034	293	0.028	514	0.025	580	0.016	540

Authors' calculations.

Table A.1: Most and Least Concentrated Industries, Using the Ellison and Glaeser (1997) Index (continued)

ISIC	Description	1982 γ	1982 N	1992 γ	1992 N	2002 γ	2002 N	2012 γ	2012 N
3511	Basic Chemicals Except Fertilizer	0.006	65	0.014	201	0.041	268	0.030	291
3512	Fertilizer			0.011	24	0.032	58	0.015	90
3513	Resin, Plastic Material, and Synthetic Fibre	-0.048	16	0.082	44	0.046	64	0.039	64
3514	Mosquito Pesticide, Inscence Coil and Spray	-0.037	29	0.025	40	-0.022	37	0.017	40
3521	Paint, Varnisher, Laquers	0.122	34	0.025	79	0.018	115	0.015	122
3522	Pharmaceuticals and Herbal Drugs	0.046	141	0.055	225	0.046	229	0.040	219
3523	Household Cleaning and Cosmetics	0.069	99	0.042	140	0.055	125	0.061	133
3529	Chemical Products Not Elsewhere Classified	-0.028	53	0.016	95	-0.003	141	0.000	140
3541	Products of Petroleum Refineries					-0.008	21	0.001	34
3542	Lubricating Oil							0.017	12
3544	Products of Coal								
3551	Tyres and Tubes	-0.008	38	0.049	53	0.012	64	0.037	56
3552	Industrial Rubber	0.059	99	0.032	288	0.030	257	0.022	277
3559	Rubber Products Not Elsewhere Classified	0.018	63	0.005	140	0.016	147	0.039	145
3560	Plastic Wares	0.033	282	0.013	802	0.007	1090	0.011	1252
3611	Ceramics and Porcelain	-0.002	25	0.036	82	0.069	84	0.068	58
3621	Glass and Glass Products	0.049	43	0.030	50	0.003	72	0.013	80
3631	Cement			0.004	11	0.007	17	0.015	23
3632	Goods Made From Cement	0.001	286	0.002	385	0.002	304	0.004	305
3633	Lime	0.085	69	0.035	94	0.031	99	0.123	98
3641	Bricks	0.097	77	-0.001	16	-0.064	38	-0.084	34
3642	Roofing Tiles	0.143	151	0.173	621	0.217	718	0.163	768
3690	Other Non Metallic Mineral Products	0.019	45	0.033	202	0.021	275	0.017	279
3710	Basic Iron and Steel	0.018	37	0.033	116	0.018	186	0.016	245
3811	Agricultural, Carpentry, and Metal Cutting Tools	0.040	110	0.026	154	0.018	221	0.002	184
3812	Furniture and Fixtures Primarily of Metal Products	0.018	48	0.009	75	0.006	128	0.005	149
3813	Structural Metal Products	0.031	65	0.032	145	0.028	207	0.020	231
3814	Steel Containers (All Kinds)	0.071	41	0.047	15				
3819	Metal Products Not Elsewhere Classified	0.018	90	0.021	285	0.015	450	0.021	460
3820	Repair Including Machineries and Sewing Repair	0.047	123	0.014	248	0.030	327	0.048	357
3831	Batteries	-0.015	15	0.004	15				
3832	Radio, Tv, Cassettes, Etc.	0.039	35	0.042	114	0.147	218	0.154	213
3833	Manufacture of Electrical Apparatus and Supplies	0.008	64	0.024	200	0.041	309	0.046	304
3841	Ship Building and Repairing	0.119	54	0.046	164	0.039	147	0.174	132
3843	Motor Vehicle Assembly and Manufacturing	0.145	55	0.058	206	0.076	275	0.068	311
3844	Motor Cycles and Three-Wheeled Motor Vehicles	0.169	13	0.063	38	0.079	99	0.029	146
3845	Bicycle, Becak Assembly/Manufacturing	0.061	33	0.031	68	0.007	68	0.060	42
3846	Manufacture of Motor Vehicles Body and Equip.	0.004	47	0.027	18				
3850	Measuring, Optical and Photographic Equip.	0.032	27	0.026	66	0.058	65	0.025	68
3901	Jewelry	0.161	12	0.025	73	0.020	100	0.038	144
3902	Musical Instruments					-0.050	20	-0.041	34
3903	Sporting and Athletic Goods	0.264	14	-0.011	37	0.027	39	0.025	72
3904	Toys Manufacturing			0.038	77	0.005	88	0.032	119
3906	Manufacture of Stationairies			0.025	68	0.056	124	0.240	184
3909	Other Manufacturing Industries	0.009	42	0.035	72	0.021	109	-0.007	82

Authors' calculations.

Table A.2: Comparing Duranton and Overman and Ellison and Glaeser Measures of Industrial Concentration

		2012		2013							
ISIC	Description	γ	N	Γ	≤ 5	$(5,10]$	$(10,25]$	$(25,50]$	$(50,75]$	$(75,100]$	N
3111	Slaughtering and Preserving Meat	0.004	64	0.000	0	0	0	0	0	0	56
3112	Milk Products	-0.008	47	0.000	0	0	0	0	0	0	44
3113	Canning, Preserving, Processing of Fruits / Vegetables	0.011	85	0.000	0	0	0	0	0	0	97
3114	Canning, Preserving, Processing of Seafood	0.024	888	0.000	0	0	0	0	0	0	647
3115	Manufacture of Cooking Oils	0.031	623	0.000	0	0	0	0	0	0	516
3116	Grain and Bean Products (Rice, Coffee)	0.061	654	0.000	0	0	0	0	0	0	575
3117	Products from Flour	0.006	956	0.000	0	0	0	0	0	0	917
3118	Sugar Products	0.030	115	0.000	0	0	0	0	0	0	114
3119	Chocolate Powder, Chocolate, and Sugar Products	-0.001	87	0.000	0	0	0	0	0	0	89
3121	Tapioca Flour, Sago, Cassava Flour, and Other Flour	0.195	189	0.000	0	0	0	0	0	0	188
3122	Tea Processing	0.062	209	0.000	0	0	0	0	0	0	207
3124	Soybean Products	0.033	327	0.000	0	0	0	0	0	0	311
3125	Krupuk, Emping, Karak and Other Chips	0.042	924	0.006	1	1	1	1	0	0	920
3126	Coffee Powder and Fried (?)	0.001	60	0.000	0	0	0	0	0	0	51
3127	Seasoning	0.017	328	0.000	0	0	0	0	0	0	338
3128	Cattle Food	0.039	103	0.000	0	0	0	0	0	0	99
3131	Alcoholic Liquors	-0.005	12	0.000	0	0	0	0	0	0	11
3134	Soft Drinks and Carbonated Waters	0.019	305	0.000	0	0	0	0	0	0	261
3141	Drying and Processing Tobacco	0.039	527	0.050	0	0	0	0	0	0	485
3142	Clove Cigarettes	0.070	349	0.092	1	1	1	1	1	1	348
3143	Cigarettes Manufacturing	0.043	38	0.000	0	0	0	0	0	0	28
3149	Manufacture of Other Tobacco Products	0.065	21	0.002	1	1	1	1	1	1	25
3211	Spinning, Weaving, and Finished Textiles	0.053	1289	0.014	1	1	1	1	0	1	1290
3212	Made Up Textile Goods Except Wearing Apparel	0.057	425	0.056	1	1	1	1	1	1	540
3213	Knitting Mills	0.113	333	0.082	1	1	1	1	1	1	352
3214	Carpets and Rugs	-0.038	18	0.107	1	1	1	1	1	1	21
3215	Cordage and Twine	0.006	60	0.000	0	0	0	0	0	0	80
3216	Kapok Manufacturing	0.654	126	0.156	1	1	1	1	0	0	122
3219	Textile Not Elsewhere Classified	0.035	209	0.000	0	0	0	0	0	0	254
3221	Wearing Apparels	0.026	1951	0.047	1	1	1	1	1	1	1754
3229	Wearing Apparels Not Elsewhere Classified	0.067	80	0.003	1	1	1	1	0	0	68
3231	Leather Tanneries and Finishing	0.024	69	0.000	0	0	0	0	0	0	69
3233	Products of Leather and Substitutes	-0.021	151	0.016	1	1	1	1	1	1	126
3240	Footwear	0.045	433	0.071	1	1	1	1	1	1	444
3311	Saw Mills and Wood Processing	0.027	733	0.000	0	0	0	0	0	0	647
3312	Wooden Boxes and Containers	0.098	39	0.000	0	0	0	0	0	0	32
3313	Bambo, Rattan, and Willow Pleats	0.019	148	0.000	0	0	0	0	0	0	153
3314	Handicraft and Wood Carving	0.062	77	0.000	0	0	0	0	0	0	95
3319	Wood Products Not Elsewhere Classified	0.019	127	0.000	0	0	0	0	0	0	140
3321	Furniture Made of Wood, Bamboo, Or Rattan	0.040	1223	0.005	1	1	1	1	0	0	1180
3323	All Kinds of Matras	0.032	78	0.000	0	0	0	0	0	0	68

Authors' calculations.

Table A.2: Comparing Duranton and Overman and Ellison and Glaeser Measures of Industrial Concentration (continued)

		2012		2013							
ISIC	Description	γ	N	Γ	≤ 5	$(5, 10]$	$(10, 25]$	$(25, 50]$	$(50, 75]$	$(75, 100]$	N
3411	Paper and Paper Products	-0.000	149	0.000	0	0	0	0	0	0	145
3412	Containers and Boxes of Paper and Paper Board	0.017	233	0.010	1	1	1	1	1	0	230
3419	Pulp and Paper Products Not Elsewhere Classified	0.034	51	0.000	0	0	0	0	0	0	63
3420	Printing, Publishing and Allied Industries	0.016	540	0.008	1	1	1	1	1	0	508
3511	Basic Chemicals Except Fertilizer	0.030	291	0.000	0	0	0	0	0	0	247
3512	Fertilizer	0.015	90	0.000	0	0	0	0	0	0	86
3513	Resin, Plastic Material, and Synthetic Fibre	0.039	64	0.075	1	1	1	1	1	1	86
3514	Mosquito Pesticide, Inscence Coil and Spray	0.017	40	0.000	0	0	0	0	0	0	36
3521	Paint, Varnisher, Laquers	0.015	122	0.027	1	1	1	1	1	1	117
3522	Pharmaceuticals and Herbal Drugs	0.040	219	0.000	0	0	0	0	0	0	88
3523	Household Cleaning and Cosmetics	0.061	133	0.037	1	1	1	1	1	1	126
3529	Chemical Products Not Elsewhere Classified	0.000	140	0.000	0	0	0	0	0	0	135
3541	Products of Petroleum Refineries	0.001	34	0.000	0	0	0	0	0	0	51
3542	Lubricating Oil	0.017	12	0.000	0	0	0	0	0	0	13
3544	Products of Coal			0.000	0	0	0	0	0	0	10
3551	Tyres and Tubes	0.037	56	0.003	0	0	0	0	0	0	52
3552	Industrial Rubber	0.022	277	0.000	0	0	0	0	0	0	247
3559	Rubber Products Not Elsewhere Classified	0.039	145	0.000	0	0	0	0	0	0	127
3560	Plastic Wares	0.011	1252	0.012	1	1	1	1	1	0	1139
3611	Ceramics and Porcelain	0.068	58	0.023	1	1	1	1	1	1	47
3621	Glass and Glass Products	0.013	80	0.000	0	0	0	0	0	0	145
3631	Cement	0.015	23	0.000	0	0	0	0	0	0	17
3632	Goods Made From Cement	0.004	305	0.000	0	0	0	0	0	0	271
3633	Lime	0.123	98	0.000	1	1	1	1	0	0	97
3641	Bricks	-0.084	34	0.000	0	0	0	0	0	0	40
3642	Roofing Tiles	0.163	768	0.058	1	1	1	1	1	1	733
3690	Other Non Metallic Mineral Products	0.017	279	0.000	0	0	0	0	0	0	263
3710	Basic Iron and Steel	0.016	245	0.000	0	0	0	0	0	0	218
3811	Agricultural, Carpentry, and Metal Cutting Tools	0.002	184	0.024	1	1	1	1	1	1	249
3812	Furniture and Fixtures Primarily of Metal Products	0.005	149	0.000	0	0	0	0	0	0	156
3813	Structural Metal Products	0.020	231	0.016	1	1	1	1	1	0	293
3819	Metal Products Not Elsewhere Classified	0.021	460	0.000	0	0	0	0	0	0	380
3820	Repair Including Machineries and Sewing Repair	0.048	357	0.034	1	1	1	1	1	1	133
3832	Radio, Tv, Cassettes, Etc.	0.154	213	0.062	1	1	1	1	1	1	158
3833	Manufacture of Electrical Apparatus and Supplies	0.046	304	0.064	1	1	1	1	1	1	273
3841	Ship Building and Repairing	0.174	132	0.000	0	0	0	0	0	0	125
3843	Motor Vehicle Assembly and Manufacturing	0.068	311	0.090	1	1	1	1	1	1	285
3844	Motor Cycles and Three-Wheeled Motor Vehicles	0.029	146	0.116	1	1	1	1	1	1	158
3845	Bicycle, Becak Assembly/Manufacturing	0.060	42	0.000	0	0	0	0	0	0	38
3850	Measuring, Optical and Photographic Equip.	0.025	68	0.003	1	1	1	1	1	1	77
3901	Jewelry	0.038	144	0.010	1	1	1	1	1	0	233
3902	Musical Instruments	-0.041	34	0.000	0	0	0	0	0	0	31
3903	Sporting and Athletic Goods	0.025	72	0.006	0	0	0	0	0	0	66
3904	Toys Manufacturing	0.032	119	0.053	1	1	1	1	1	1	118
3906	Manufacture of Stationairies	0.240	184	0.000	0	0	0	0	0	0	150
3909	Other Manufacturing Industries	-0.007	82	0.000	0	0	0	0	0	0	15

Authors' calculations.

BOSTON UNIVERSITY
Global Development Policy Center

Boston University
53 Bay State Road
Boston, MA 02215

gdp@bu.edu
www.twitter.com/gdpc_bu
www.bu.edu/gdp

The Human Capital Initiative (HCI) is a research inititiative at Boston University's Global Development Policy Center. The GDP Center is a University wide center in partnership with the Frederick S. Pardee School for Global Studies. The Center's mission is to advance policy-oriented research for financial stability, human wellbeing, and environmental sustainability.

www.bu.edu/gdp

ACKNOWLEDGEMENT

We are grateful for financial support from Private Enterprise Development for Low-Income Countries (PEDL), a joint research initiative of the Centre for Economic Policy Research (CEPR) and the Department For International Development (DFID). Kun Gu provided excellent research assistance. All errors remain our own.