Economic Geography and Wages*

Mary Amiti
International Monetary Fund, University of Melbourne and CEPR

Lisa Cameron
University of Melbourne

December 10, 2003

Abstract

This paper estimates the agglomeration benefits that arise from vertical linkages between firms. The analysis is based on international trade and economic geography theory developed by Krugman and Venables (1995). We identify the agglomeration benefits off the spatial variation in firm level nominal wages. Unusually detailed intermediate input data allow us to more accurately capture spatial input/output linkages than in previous studies. We take account of the location of input suppliers to estimate cost linkages; and the location of demand from final consumers and other firms to estimate demand linkages. The results show that the externalities that arise from demand and cost linkages are quantitatively important and highly localized. An understanding of the extent and strength of spatial linkages is crucial in shaping policies that seek to influence regional development.

JEL Classifications: F1, L6, R1.

*We would like to thank Bill Griffiths, Gordon Hanson, Keith Head, Russ Hilberry, David Hummels, Wolfgang Keller, Guay Lim, Stephen Redding and Tony Venables for their comments. This paper has been presented at the NBER Summer Institute in Cambridge in 2002, CEPR European Research Workshop in International Trade in Munich 2002, the Empirical Investigations in International Trade workshop in Atlanta 2002, the North-East Universities Development Consortium Conference at Yale University in 2003, New York Federal Reserve, University of Melbourne and the World Bank. We thank seminar paticipants for valuable comments.

1. Introduction

Manufacturing wages vary significantly across regions within countries. For example, in Indonesia's weaving mills industry the average wage paid by a firm at the 90th percentile of the wage distribution in 1996 was more than twice as high as that paid at the tenth percentile (after adjusting for skill differentials). These firms were 518 kilometers apart on the island of Java. Similar patterns are observed for other industries. The existence of such large wage differentials raises the question as to why firms do not relocate to low wage regions and arbitrage these differences away. The reasons we explore in this paper are related to the potential agglomeration benefits they might enjoy from being close to other firms.

Three main sources of externalities arising from geographical agglomerations have been identified by Marshall (1920) - they are (i) input/output linkages[1]; (ii) labor pooling; and (iii) knowledge spillovers. The role of input/output linkages in driving agglomeration of industries and hence wage inequalities has recently been formalized and developed in the international trade and economic geography literature by Krugman and Venables (1995) and Fujita et al (2000). The theory posits that firms benefit from being close to a large supply of intermediate input producers due to savings on transport costs, and from access to a large variety of differentiated inputs, reducing total costs, increasing profits and thus attracting more firms.[2] This gives rise to a cost linkage or supply access effect. Similarly, firms benefit from being close to the markets for their output due to increased demand, giving rise to a demand linkage or market access effect, which also increases profits. We use this theoretic framework to estimate the effect of supply access and market access on firm level wages in Indonesia.[3] Firms in neighboring regions can also benefit from these agglomerations in the form of lower prices for inputs and higher demand for their goods. We estimate how far these pecuniary externalities spread across space.

[1] See Hirschman (1958).

[2] More intense competition in the upstream industry could also lead to lower intermediate input prices and hence more benefits to downstream firms - this would be the case if the upstream industry were oligopolistic instead of monopolistically competitive. (See, for example, Amiti, 2001).

[3] We choose to examine the effects on wages because this variable is likely to be more accurately measured than alternatives such as total factor productivity or profits which rely on a measure of capital stock.

Estimating the benefits of different sources of agglomeration and how far these benefits spread is of particular importance for regional policy development. Governments around the world spend large sums of money in the pursuit of decentralization. This is true in developed countries such as in the European Union, where large amounts of public expenditure are devoted to developing the poorer southern regions. It is also true in developing countries such as Indonesia where decentralization is currently a major political and public policy issue. The concentration of industry on Java has fed into pre-existing sentiments of pro-Java bias, which have fostered movements for greater decentralization. The Indonesian government has been actively pursuing decentralization in an attempt to spread the benefits of industrialization to the other (outer) islands - with limited success. Our study gives an indication of how large the benefits of agglomeration arising from vertical linkages are and is the first to examine how close firms need to be to benefit from these agglomerations. It is the spatial linkages that determine the extent to which the benefits of development spread across space. An understanding of the way in which they operate and how far they spread is crucial when considering policies that seek to influence regional development.

Studying Indonesia gives us access to a rich data set that provides detailed information on firms' vertical linkages with their suppliers and information on firms' geographic locations. Our study is also the first that utilizes detailed firm-level data on connections between potential input suppliers and final producers to examine cost and demand linkages, rather than relying on aggregate input/output tables.[4] We use three waves of Indonesia's manufacturing census, which is a complete enumeration of all firms with 20 or more employees - 1983, 1991 and 1996 and so can examine how geographical links between firms change over a long period of rapid growth.

Indonesia's geography, public policy and political history also make it an interesting laboratory in which to examine the theory. Although its 200 million people are spread over 900 islands and an east-west distance of 5,500 kms, there is large variation in the concentration of workers and manufacturing industry across locations. Manufacturing is

[4]Our dataset includes firm level information on intermediate inputs. We aggregate this up to 5 digit ISIC level giving us input/output relations for over 300 manufacturing industries. The most disaggregated I/O table for Indonesia includes 90 manufacturing sectors.

very heavily concentrated on the island of Java, with about three quarters of non-oil and gas manufacturing located there. Within Java manufacturing is further concentrated in the three main centers of Greater Jakarta, Surabaya and Bandung. See Figure 1. The substantial internal trade costs imposed by the country's geography have played an important role in shaping the country's spatial pattern of industry.

The results show that demand and cost linkages have a significant positive impact on manufacturing wages in Indonesia. An increase in market or supplier access from the 10th to the 90th percentile increases wages by more than 20%. Although firms benefit from vertical linkages, these benefits are highly localized. That is, benefits of agglomeration spread over only a small distance. Only 10% of the benefit of market access spreads beyond 108km and 10% of the benefit of supplier access beyond 262km. These findings highlight why government policies often fail in trying to relocate industry to peripheral areas.

There are a small number of studies that also seek to examine the importance of agglomeration forces arising from vertical linkages. Redding and Venables (2002) use the same theoretical structure as we do but apply it at the country level, and find that supply and market access have a positive effect on cross-country variation in GDP per capita. Their measures of supply and market access are constructed from the coefficients on import and export dummies in a gravity equation. Our disaggregated approach is based on which inputs firms use and hence is likely to more accurately capture vertical linkages between firms. Hanson (1998) estimates a spatial wages equation similar to ours for US manufacturing firms, with the average wage in each county as the dependent variable. He finds evidence supporting the positive benefits of market access but does not attempt to model supply access. His theoretic framework is based on Krugman (1991a) where agglomeration forces arise due to the mobility of workers rather than vertical linkages between firms. Dumais, Ellison and Glaeser (1997) estimate the effect of the three sources of agglomeration on changes in employment using US plant level data at the metropolitan level and state level. They find evidence of all three sources of agglomeration with labor market pooling the strongest. All of their measures only take account of proximity of other firms within the same metropolitan area and ignore distance to neighboring areas. They find that although there is a lot of entry and exit of firms over the period of their study (between 1972 and 1992), aggregate agglomeration patterns are relatively stable.

The existing small body of work on the concentration of industry in Indonesia, although informative, has not specifically examined cost and demand linkages as a source of agglomeration and has largely neglected an examination of the spatial aspects of such linkages. Henderson and Kuncoro (1996) examine firm's location decisions and find that firms strongly prefer locations where there are mature firms in related industries.[5]

Section 2 develops the formal model. Section 3 provides background information on Indonesia and details of the data sources. Section 4 presents the results and section 5 concludes.

2. Theory

We derive our estimating equation from an international trade and economic geography model developed by Krugman and Venables (1995) and extended in Fujita *et al* (1999). It is a model in which vertical linkages between upstream and downstream firms create forces leading to industrial agglomeration. Firms are assumed to compete in a monopolistically competitive environment, where differentiated inputs enter the production function symmetrically and differentiated final goods enter the consumers' utility symmetrically.

2.1. Demand

The utility function U_k, of a representative consumer in district k, is

$$U_k = \prod_i \left(C_k^i\right)^{s^i}, \qquad \sum_i s^i = 1, \tag{2.1}$$

where C_k^i is the sub-utility function i.e. the aggregate consumption of varieties of differentiated industry i goods consumed in district k and s^i is the share of income spent on

[5] Blalock and Gertler (2002) estimate whether supply chains in Indonesia are the conduit for transferring technology from foreign direct investment using aggregate input/output tables. However, they too implicity assume that externalities are bound by district borders (at the provincial level) and hence do not model the spatial dimension.

industry i goods. We denote all location specific variables with subscripts and industry specific variables with superscripts. C_k^i is defined as

$$C_k^i = \left[\sum_{l=1}^{K} \sum_{v=1}^{N_k^i} \left(c_{lk}^{iv} / t_{lk}^i \right)^{\frac{\sigma^i - 1}{\sigma^i}} \right]^{\frac{\sigma^i}{\sigma^i - 1}}, \tag{2.2}$$

where c_{lk}^{iv} is the quantity of a variety v good in industry i produced in district l and consumed in k. N_k^i is the number of varieties of industry i goods produced in district k. The elasticity of substitution between varieties in each industry i is constant, given by $\sigma^i > 1$. The transport cost of shipping a good from district l to k is modelled as Samuelsonian iceberg costs, with $t_{lk}^i \geq 1$.[6] In order to consume one unit of output, consumers must demand t_{lk}^i units because a proportion of imported goods, $1 - \frac{1}{t}$, melts in transit. If $t = 1$ there is free trade and if $t = \infty$ there is no trade. The total transport cost of shipping a good from k to l can be rewritten as a function of distance, d_{kl}, in exponential form as

$$t_{kl}^i = e^{\tau^i d_{kl}}. \tag{2.3}$$

where τ is the elasticity of trade costs with respect to distance. The further a good travels the more of it melts in transit. Dual to the quantity index is the price index, defined as

$$P_k^i = \left[\sum_{l=1}^{K} \sum_{v=1}^{N_k^i} \left(p_l^{iv} t_{lk}^i \right)^{1-\sigma^i} \right]^{\frac{1}{1-\sigma^i}}, \tag{2.4}$$

where p_l^{iv} is the free-on-board producer price.

A consumer's demand functions are derived using two stage budgeting. In stage one, a constant share, s^i, of income, y_l, is allocated to industry i. In stage two, demand functions for a representative consumer located in district k for a variety v in industry i produced in district l is given by maximizing the sub-utility functions, equation 2.2, subject to the budget constraint $s^i y_l$. Aggregating across identical consumers, demand functions are given by

[6] We assume that $t_{kk}^i = 1$.

$$c_{kl}^{iv} = \left(p_k^{iv}\right)^{-\sigma^i} \left(t_{kl}^{iv}\right)^{1-\sigma^i} s^i Y_l \left(P_l^i\right)^{\sigma^i - 1}, \tag{2.5}$$

where Y_l is aggregate income in location l.

2.2. Supply

The production technology in the manufacturing sector consists of a small fixed cost of setting up a plant, F, to produce a variety v. This gives rise to increasing returns to scale technology; and the small size of F ensures that the number of varieties produced is large enough to make oligopolistic interactions negligible. In each industry i, the production function to produce a variety v of a manufactured good is given by

$$\left(L_k^{iv}\right)^{\alpha^i} \left(K_k^{iv}\right)^{\beta^i} \prod_u \left(C_k^u\right)^{\mu^{ui}} = F^i + b^i x_k^{iv}, \qquad \alpha^i + \beta^i + \sum_u \mu^{ui} = 1, \tag{2.6}$$

where L_k^{iv} and K_k^{iv} are the labor and capital[7] amounts required by each firm in industry i to produce output, x_k^{iv}. C_k^u is a quantity index aggregated across varieties of intermediate inputs supplied by industry u, defined analogously to equation 2.2. Hence, industry $u's$ output of intermediate inputs enters the production function of each downstream firm through a CES aggregator as in Ethier (1982). Note that industry i purchases many varieties of inputs from multiple upstream industries.

Profits of a single firm v in district k are given by[8]

$$\pi_k^{iv} = p_k^{iv} x_k^{iv} - \left(w_k^{iv}\right)^{\alpha^i} r_k^{\beta^i} \prod_u \left(P_k^u\right)^{\mu^{ui}} \left[F^i + b^i x_k^{iv}\right]. \tag{2.7}$$

The total cost function comprises a fixed cost, F^i, and a variable cost, b^i. The factor prices are denoted by w_k^{iv}, the wage of an industry i firm in district k, and by r_k, the price of capital in district k (or any other factor of production); and P_k^u is the intermediate input price index of each upstream industry u that supplies inputs to industry i. It is analogous to equation

[7] We allow for more than one primary factor of production in the empirical model as in Amiti (2003).

[8] Each firm produces a distinct variety v. The theory assumes that firms within an industry are symmetric but given that this is not the case in the data we superscript variables by v to allow for variation across firm within an industry.

2.4, with $i = u$. The price index enters the cost function directly. The lower the price of intermediate inputs, the lower the cost of producing industry i goods; and the higher the number of upstream firms, the lower the price index. Being located close to lots of upstream firms also reduces the price index due to savings on transport costs. This has a direct effect on producer prices. The fob producer price is given by profit maximization, which gives the usual marginal revenue equals marginal cost condition, with prices proportional to marginal cost,

$$p_k^{iv} = \left(w_k^{iv}\right)^{\alpha^i} r_k^{\beta^i} \prod_u \left(P_k^u\right)^{\mu^{ui}} b^i \theta^i, \qquad \theta^i = \frac{\sigma^i}{\sigma^i - 1}. \tag{2.8}$$

The mark-up over marginal cost, θ^i, depends on the elasticity of substitution σ^i. Allowing free entry and exit of firms into each industry gives the level of output each firm must produce to just cover fixed costs, and hence make zero profits,

$$x_k^{iv} = \overline{x}^i = \frac{F^i\left(\sigma^i - 1\right)}{b^i}. \tag{2.9}$$

2.3. Aggregate demand

To calculate total demand for industry i goods produced in district k we sum across demand in all districts l,

$$c_k^i = \sum_{l=1}^{K} c_{kl}^i = \left(p_k^i\right)^{-\sigma^i} \sum_{l=1}^{K} \left(t_{kl}^i\right)^{1-\sigma^i} E_l^i \left(P_l^i\right)^{\sigma^i - 1}. \tag{2.10}$$

where $E_l^i = s^i Y_l + \sum_d \mu^{di} N_l^d p_l^d x_l^d$. Demand for industry i goods not only comes from consumers but also from downstream firms, with each firm spending a proportion μ^{di} of its total revenue on intermediate inputs produced by industry i. Demand for intermediate inputs from downstream firms is derived using Shepard's lemma on the price index (as shown in Dixit and Stiglitz, 1977) and they are analogous to equation 2.5 with the term $s^i Y_l$ replaced with $\mu^{di} N_l^d p_l^d x_l^d$. N_l^d is the number of downstream firms in district l.

Substituting prices, expenditure and transport costs (equations 2.8 and 2.3) into the aggregate demand function (equation 2.10) and setting demand equals supply in the product market gives

$$x_k^{iv} = \left(\left(w_k^{iv}\right)^{\alpha^i} r_k^{\beta^i} \prod_u \left(P_k^u\right)^{\mu^{ui}} b^i \theta^i \right)^{-\sigma^i} \tag{2.11}$$
$$\sum_{l=1}^{K} \left\{ e^{-\tau^d \left(\sigma^d - 1\right) d_{kl}} \left(s^i Y_l + \sum_d \mu^{di} N_l^d p_l^d x_l^d \right) \left(P_l^i\right)^{\sigma^i - 1} \right\}.$$

We derive the zero profit wage by imposing the zero profit level of output, substituting for the intermediate input price index, and rearranging. This gives the maximum wage industry i firms can afford to pay,

$$\left(w_k^{iv}\right)^{\alpha^i} = \left(\left(\overline{x}^i\right)^{\frac{1}{\sigma^i}} b^i \theta^i \right)^{-1} r_k^{-\beta^i} \prod_u \left\{ \sum_{l=1}^{K} \sum_{v=1}^{N_k^u} (p_l^{uv})^{1-\sigma^u} e_k^{-\delta_1 d_{kl}} \right\}^{\frac{\mu i}{\sigma - 1}} \tag{2.12}$$
$$\sum_{l=1}^{K} \left\{ e^{-\delta_2 d_{kl}} \left(s^i Y_l + \sum_d \mu^{di} N_l^d p_l^d x_l^d \right) \left(P_l^i\right)^{\sigma^i - 1} \right\}^{\frac{1}{\sigma^i}}.$$

This is the main equation we are interested in. It embodies utility and profit maximization conditions, product market equilibrium and free entry and exit. The first curly brackets represent cost linkages or supplier access (SA), which the theory suggests has a positive effect on wages. Note that we do not have individual prices in our data so the measure we construct (described below) is essentially an inverse proxy of the price index in equation 2.4. It measures the proximity of firms to their potential suppliers - the closer a firm is to its input suppliers the lower its total cost and the higher the zero profit wage. The coefficient on the distance parameter, δ_1, indicates how quickly the externalities arising from proximity to input suppliers diffuse across space. A positive coefficient indicates that firms in close proximity benefit more than those further away. The higher this coefficient the more localized the externalities. The second curly brackets represent demand linkages or market access (MA), which has a positive effect on wages - the closer a firm is to its market, which comprises consumers and other firms that purchase its output, the more profitable it is and hence the higher its zero profit wage. Similarly, the coefficient on distance, δ_2, indicates how far these benefits extend across space.

Our basic estimating equation, after taking logs of equation 2.12, becomes

$$\ln w_k^{iv} = \gamma_0 + \gamma_1 * \ln(SA_k^i e^{\delta_1 d_{kl}}) + \gamma_2 * \ln(MA_k^i e^{\delta_{2kl}}) + \gamma_l Z_l + \gamma_i Z_i + \varepsilon_{ik}. \tag{2.13}$$

So the theory posits that wages in location k are a function of supplier access, SA_k^i, and market access, MA_k^i, and the distance parameters, δ_1 and δ_2, as well as industry specific effects Z_i, and location specific effects Z_l. The industry specific effects capture differences in fixed costs, marginal costs and mark-ups. The location specific effects capture differences in prices of immobile factors of production other than labor such as land. We estimate equation 2.13 using non-linear estimation. This enables us to estimate a distance adjusted supplier and market access rather than imposing the distance effect.[9] We will detail how we measure each of these variables below.

Extensions and modifications to the theory Before going to the data with this theory we need to ask how realistic the assumptions of the theory are and whether there are any other important variables omitted that affect wages. First, consider the zero profit assumption. Although firms may not earn zero profits in practice, the relationship in equation 2.13 will still hold provided that wages are an increasing function of profits, which seems likely.

Second, we have allowed wages to vary by industry as well as location whereas the theory does not give any grounds for industry-specific wages. We however cannot ignore that there is significant industry variation in wages within a location. These differences may be explained by standard labor theory variables such as differences in firm size and skill requirements. We add controls of this sort in some of the specifications below. The industry wage differentials may also be driven by differences in the market and supply access of different industries located in the same district. These differences will persist if there are frictions in labor mobility across industries, for example, as a result of industry-specific skill acquisition. The market access and supply access variables vary by 5-digit industry and district.

Third, the theory assumes that labor is completely immobile across locations giving rise to location specific wages. Clearly this is not the case across districts within Indonesia. Provided that there are some frictions in labour mobility between locations then the relationships in 2.13 will hold. This is realistic in the context of Indonesia. Ties to the land are strong and

[9] Other studies usually divide market access proxies, such as GDP, by distance as originally done in Harris (1954). We experimented with modelling transport costs as $t_{kl}^i = (d_{kl})^{-\tau}$ but the exponential functional form we use gives a better fit. The functional form does not affect the other estimated coefficients.

migrating to an industrial center may mean leaving one's own ethnic group and for that reason may be unattractive. Hence, not everyone is willing or able to migrate to the labor markets in industrial centers.[10]

Fourth, other sources of agglomeration such as technological spillovers and labor pooling could give rise to higher wages. We construct variables to capture these effects and include them as additional regressors.

3. Data and Measurement

Our analysis uses firm level data. The geographic unit of analysis is the kabupaten. Indonesia has a five-tiered geographic system – national, provinces, districts (kabupaten), sub-districts (kecamatan) and villages (desa).[11] A map showing the geographic distribution of manufacturing output in 1996 by district is presented in Figure 1. There is little formal sector manufacturing in the eastern islands (Nusa Tenggara Timur, East Timor, Maluku and Irian Jaya) so we drop these regions from our initial sample (and they are not shown on the map). Sulawesi has slightly more in the way of manufacturing and we leave it in because it is a large, important land mass. The figure shows that manufacturing is concentrated largely around Java's urban centers, with some activity in Sumatra, and to a lesser extent Kalimantan. Our sample consists of 210 districts, 88 of which are on the island of Java. These cover an area of 1,375,369 square kilometers, roughly the total land area of Germany, France and Spain together and an east-west distance greater than that from London to Istanbul. As can be seen from Figure 1, there is considerable variability in terms of manufacturing activity within relatively small geographic areas. Much of this variability would be lost if we were to conduct the analysis at a more aggregated level.

[10] Note that economic geography models with labor mobility also lead to agglomeration of industry (see Krugman, 1991b). This is the model used in Hanson (2002), which gives rise to a positive relationship between location-specific nominal wages and market access.

[11] The number of provinces remained constant at 27 over the period of study. A number of kabupaten were split into two or more during the period. We avoid problems associated with changing kabupaten borders by using the kabupaten borders from the earliest year (1983). Urban centers of economic activity are often split off into their own district (called kotamadya) for administrative purposes. We merge all kotamadya that existed in 1983 back into their neighboring kabupaten. Although there is considerable variation in the size of kabupaten across Indonesia, kabupaten size is much more uniform within Java and within the Outer Islands. All but one of our specifications separate out these two regions.

3.1. Sources

Our main data source is the Manufacturing Survey of Large and Medium-sized firms (Survei Industri, SI). This is an annual census of all manufacturing firms in Indonesia with 20 or more employees (N=22,997 in 1996). The SI data capture the formal manufacturing sector - it collects an unusually rich array of firm level data which includes information on firm output, imports, exports, wages, employment by skill level, and foreign ownership.

Most importantly for this study, the SI questionnaire also asks each firm to list all of their individual intermediate inputs and the amount spent on each in rupiah. Although this information is not routinely prepared, it was coded up by the Indonesian Statistical Agency (Badan Pusat Statistic, BPS) and made available to us for the year 1998. For all other years the only available information on inputs is the total expenditure on domestic inputs and imported inputs. We aggregate the 1998 data up within 5 digit industry categories to provide us with a 307 manufacturing input/output table, and assume that the mix of inputs used by industries does not change over our sample period. Combining the input codes with the location codes we are able to link each firm to all potential suppliers in Indonesia and construct the supplier access variable.[12] Similarly in reverse, we can identify the location of firms that are potential purchasers of an industry's output and so construct the market access variables. The 1998 data also lists raw materials used by firms but data at the district level on raw material production is not readily available.[13] The omission of such information would constitute a potentially serious omitted variables problem for industries that are raw materials intensive. For this reason we drop such industries - this includes all food industries (2 digit code=31). Note that data on the "dropped" industries are still used in the construction of the supply and market access variables. For example, the "threads" industry is dropped but these firms supply inputs to the textiles industries and so information on them is used in the calculation of the supply access variable. We also drop the "not elsewhere classified" industries. Our final sample has observations covering

[12] We include inputs of all industries that constitute 10% or more of total intermediate inputs.

[13] It should be possible to construct kabupaten level agricultural output variables from the Agricultural Survey - something we plan to pursue in future work.

172 industries.

In addition to the SI data, we use data on non-oil gross regional domestic product (GRDP) at the district level to construct the regional income data needed for the calculation of the final demand component of the market access variable. These data are also produced by BPS (BPS 1995, 1998, 2000a). The earliest year for which such data are available is 1983. Oil revenues in Indonesia accrue almost entirely to the central government so it is important to net them out when seeking to construct a measure of regional income. Non-oil GRDP figures are published from 1993. For years prior to 1993 we predict district oil revenues from available concurrent provincial figures and subtract this from the GRDP (including oil) data. Final demand shares from Input-Output tables published in BPS (1992, 1997) are applied to the income to construct final consumer demand at the 5 digit industry level.[14]

We construct a measure of skilled labor from the 1995 Intercensal Survey. It is a large household survey (N=216,945) which is conducted at five ten yearly intervals midway between census years. We use information on the educational attainment of the population to control for differences in skill levels across districts.

BPS(2000b) provides information on land utilization in Indonesia. From this we construct a variable for the percentage of the district's potentially arable land that is not covered with housing and another for the percentage of district land area that is swamp. We use these to proxy for the cost of immobile factors of production and location amenity.

Finally, distances between districts were calculated using ArcView's GIS technology with a district level map of Indonesia. We construct pairwise measures of the greater circle distance between the geographic center of each location. We thus end up with 210 distance variables (in kilometers). The distances range from a minimum of 6.2 km between North Jakarta and Central Jakarta to a maximum of 3,304 km from Aceh Besar in the north-western tip of Sumatra to Sangihe Talaud in the far north-east of North Sulawesi.

[14] The input-output tables have a total of 90 sectors in 1995 and 87 manufacturing sectors in 1990. These are more aggregated than the 5-digit ISIC industry categories. We apportion the final demand shares between 5-digit industries on the basis of the value of national output (net of exports) of each 5 digit industry.

3.2. Measurement

The dependent variable - the average firm wage - is constructed by dividing each firm's annual wage bill (in Rupiah) by the average number of workers employed over that 12 month period. We then convert this to a daily wage assuming a six day working week. These data produce a wage distribution similar to that for formal sector workers in the most commonly used source of Indonesian wage data, the Labor Force Survey (*Sakernas*).[15] The supplier access variable is calculated from firms' self-reported value of output in rupiah; and the market access variable is calculated from firm's self-reported total expenditure on intermediate inputs.

Supplier Access The supplier access effect comes through the price indices of intermediate inputs, P_k^u, defined as in equation 2.4 with $i = u$. Since individual input price data are unavailable we approximate the cost linkages as follows:

$$SA_k^i = \sum_l^K \left[\left(\sum_u^U a^{ui} \phi_l^u \right) e^{-\delta_1 . d_{kl}} \right], \text{ where } \phi_l^u = \frac{X_l^u}{X^u} = \frac{1}{X^u} \sum_{v=1}^{N_l} x_l^{uv} p_l^{uv}. \tag{3.1}$$

The term ϕ_l^u is the total value of intermediate inputs produced by industry u in district l, X_l^u, divided by the total produced in Indonesia, X^u. We know where in Indonesia these inputs are produced, however we do not know exactly from which location these inputs are purchased so our measure represents potential suppliers rather than actual suppliers. Although we do not have individual prices, the cost linkages are still well-represented in equation 3.1 since this 'price index' is lower the higher the share of intermediate inputs that are produced in close proximity. The share of intermediate inputs are weighted by the share of industry u in the total cost of industry i inputs, a^{ui}.

Market Access The market access variable is given by:

$$MA_k^i = \sum_{l=1}^K \left[\left(\frac{s^i Y_l + \sum_d^D a^{di} I_l^d}{TD^i} \right) e^{-\delta_2 . d_{kl}} \right]. \tag{3.2}$$

[15] Alatas and Cameron (2003) compare kernel density estimates of the wage distribution from both sources for the Jakarta area and find them to be similar.

The inner bracketed term sums demand across all downstream firms and consumers in location l that demand industry i goods. Total demand from downstream firms is defined as the total expenditure of downstream firms in district l on intermediate inputs, I_l^d, times the share of downstream firms' intermediate input expenditure that is spent on industry i goods, a^{di} (which equals $\mu^{di} N_l^d p_l^d x_l^d$ in equation 2.12). This, scaled by total demand in Indonesia by firms and consumers, TD^i, is distance adjusted (in the same way as the supply access variable) so that demand from consumers within the same district receives a higher weighting than demand from locations further away. The size of the distance adjustment is empirically determined.

International trade Treating international demand and supply in the same way as their domestic counterparts would require detailed production data and demand patterns for all countries that trade with Indonesia. These data are unavailable at a sufficiently disaggregated level so we begin by simply adding controls to the wage equation for the share of the firm's output that is exported and the share of the firm's inputs that are imported. We then try an alternative specification that is more closely aligned with the theory. In this specification we model the rest of the world (ROW) as being in one geographic location and then distance to the ROW varies across Indonesia only via a 'distance to port' component which we define as being distance to the closest port, d_p. That is, the market access term becomes

$$MA_k^i = \sum_{l=1}^{K} \left[\left(\frac{s^i Y_l + \sum_d^D a^{di} I_l^d}{TD^i} \right) e^{-\delta_2 . d_{kl}} + \gamma_x . exshare . e^{-\delta_x d_{kp}} \right]. \tag{3.3}$$

where $exshare$ is the percentage of the firm's output that is exported. We allow exports to have a different effect on wages than domestic demand via γ_x and we estimate the parameter on distance to the nearest port (δ_x).[16] For the supply access variables we treat imported inputs as a separate industry - on the basis of quality differences between imported and domestic inputs. This requires a separate term for all imported inputs, thus adding the

[16] In this specification the domestic demand term is deflated by $(1 - exshare)$ so it represents the share of total (international and domestic) demand that comes from each kabupaten.

share of imported inputs, exponentially weighted by the distance to the closest port as an explanatory variable. We find that the coefficients on domestic supplier and market access are not affected by this alternative treatment of trade so we then proceed with the simpler specification.

Labor Pooling To examine the effects of labor pooling we follow Dumais, Ellison and Glaeser (1997) and construct an index that captures the similarity of firm f in district k's labor requirements to the requirements of other firms in the same district. The index is calculated as

$$LP_k^f = -\sum_s (L^{fs} - \sum_{j \neq i} \frac{E_k^j}{E_k - E_k^f} L^{js})^2, \tag{3.4}$$

where $L^{fs}$is the fraction of firm f's labor force that has education level s, E_k^f is the number of workers in firm f, and E_k is the total number of workers in district k. The index thus compares the educational composition of firm f's workforce with the education composition of other firms in the same district. The education categories are no education, primary education, lower secondary high school, upper secondary high school and tertiary educated. The index is a sum of squared deviations measure. The higher the value of the index, the better the match between the firm's education composition and that of surrounding firms. The maximum value of zero indicates a perfect match.[17] A pooled market for specialized worker skills benefits workers and firms. Krugman (1991b) shows that it is more profitable for firms to locate where there is a pooled market for skills despite competition from other firms for workers because the benefits of a more efficient labor force outweigh the competition effects. Hence, we hypothesize that the index will have a positive effect on wages.

Technological and Knowledge Spillovers We measure the effect of technology spillovers by proximity to other firms within the same 5 digit category - ie the number of firms in the same industry in every district, distance adjusted in the same way as the linkage variables. The more firms in close proximity with related technology the more likely

[17]We calculated this measure at the provincial and kabupaten level. The provincial level variable gave a better fit.

there "will be ideas in the air" that a firm can learn from. However, in addition to capturing spillovers (which would allow firms to pay higher wages), this variable may pick up the "competition effect" i.e. it could be seen as an inverse proxy of the price index, P_l^i, of substitute goods in equation 2.12 hence putting downward pressure on firms' profits and their ability to pay high wages. Thus, a priori the direction of this variable's impact is ambiguous.

Ideally, one would have access to a technology flow matrix or to research and development stock measures in order to properly capture the effects of technological spillovers. Dumais, Ellison and Glaeser (2002) rely on a technology flow matrix published in 1974. We do not follow their approach because the matrix is too aggregated for our purposes with categories not easily matched to ours and we expect that technology flows would have changed considerably since 1974. Keller (2002) uses R&D expenditure to estimate technological spillovers on productivity levels in nine OECD countries. In Indonesia, it is more likely that new knowledge from R&D is imported rather than coming from domestic R&D - given that less than 10% of the firms in our sample invested in any form of R&D in 1996; and of those that do, the median expenditure is less than US$3000 per annum. [18]

We also construct a measure of market share to capture the competition effect more directly. It is defined as the ratio of a firm's output to the 5-digit industry total. We hypothesize that this variable should be positive because an increase in competition (lower market share) reduces profits and hence wages.[19]

4. Results

4.1. Preliminary Examination of the Data

Our initial sample covers 13,472 firms from 172 industries located across 177 different districts.[20] Of these firms, 11,361 are on the island of Java and 2,111 in the Outer Islands. We

[18] Note that the highest R&D industries in Indonesia are also not those identified by Keller (2002) as high R&D. Even if expenditures were more substantial we would be unable to construct an R&D stock variable as in that study because R&D data is only available since 1995. Estimating benefits of knowledge spillovers via imports and foreign direct investment is beyond the scope of this paper.

[19] However, it should be noted that Nickell (1996) shows that increased competition leads to increased productivity in the UK, which would then likely lead to an increase in wages.

[20] 38 of the 210 kabupaten do not have firms in the industries included in our sample.

examine linkages between these firms and firms in the full range of 210 districts and 307 industries. Table 1 presents summary statistics of the data. Manufacturing industry is very agglomerated in Indonesia, obviously in Java and also within Java. In 1996, 82% of formal sector manufacturing output was produced in Java, 40.2% within Greater Jakarta, and 46.8% in the three main manufacturing centers of Greater Jakarta, Bandung and Surabaya. The share of output being produced in Java has not changed dramatically over time. It was 80.5% in 1983 but within Java it has become more concentrated - only 38.7% was produced in the major centers in 1983. Similar patterns are seen for individual industries. The garment industry is the largest industry in our sample (in terms of the number of firms). It is highly concentrated in Java (96.3%), with 69.9% of total production occurring in the Jakarta region (up from 63.8% in 1983). Hence it appears that even as travel and communication across space become more efficient, industry has continued to become more localized. The means of the market and supply access variables are lower in the Outer Islands owing to its lesser industrialization and also its lower population density. Java constitutes only 6.6% of the Indonesian land mass but 60% of its population. There are 900 people per square kilometer versus 44.2 in the Outer Islands. In 1996 64% of Indonesian non-oil GDP was produced in Java.

Average wages do not differ markedly between Java and the Outer Islands. Wages are generally higher in the areas where industry is clustered but there are exceptions. For example, wages are relatively high in parts of Kalimantan and Sulawesi where there is not much manufacturing. The raw within-district correlation between wages and the linkage variables shows a positive relationship as hypothesized, with a correlation of 0.053 and 0.198 for market access and supply access respectively. And the correlation between the own district supplier and market access variables is only 0.23. This low correlation enables us to overcome a concern that has arisen in previous studies where supplier and market access variables have been highly correlated. As a result of being able to accurately pinpoint the location of suppliers and also to identify suppliers at the 5-digit level, we are able to separately and precisely estimate the two different - and sometimes competing- vertical linkages.

4.2. Formal Results

Equation 2.13 is estimated using non-linear least squares. All standard errors have been corrected for clustering by 5-digit industry using a generalization of the White method.[21] We include locational dummies for the islands of Sumatra, Kalimantan and Sulawesi in all estimations and also a dummy for Jakarta to take account of the benefits of being located close to the central government. Our industry controls are at the two digit level and are relative to the textiles, clothing, footwear and leather industry. We include more disaggregated industry controls in further specifications below. Table 2 presents the results for the whole of Indonesia, and Java and the Outer Islands separately. The results for Indonesia as a whole (column 1) show that demand and cost linkages have a positive and strongly significant effect, as predicted by the theory. Both the coefficients on distance (δ) and the coefficients on the distance-adjusted supply and market access variables (γ) are significant. These variables explain 29% of the variation in log wages. Column 2 presents the results for Java. The coefficients here are also positive and significant, and the γ's are larger, suggesting that the agglomeration externalities are quantitatively more important in Java than in Indonesia as a whole. The results show that a distance-adjusted increase of 10% in supplier access increases wages by 1.03% and a 10% increase in market access allows firms to increase wages by 2.2%.

The parameters on distance, δ, indicate how quickly the market and supply access spillovers decay with distance. If $\delta = 0$, then an increase in the externality in one district has the same effect on wages in all districts in Indonesia, regardless of how far they are from the source. If $\delta = \infty$ then an increase in the externality in location l will have no effect on wages in district k ($k \neq l$) – all effects are completely localized which means that firms only benefit from demand and supply within their own district. To examine how far the benefits of market access and supply access spread we use Keller's (2002) approach and calculate at what distance from the district are 90% of the effects of the district's externality dissipated. This involves finding the D^* that satisfies $0.1 = e^{-\delta D^*}$. The results from column

[21] See Rogers(1993).

(2) indicate that both effects are highly localized with only 10% of the market access benefit spreading beyond 85 kms; and the supplier access benefit spreading a little further with 10% of the benefit going beyond 231 kilometers.

Column 3 presents the results for the Outer Islands. In sharp contrast to Java, all of the market access and supply access parameters are statistically insignificant for the Outer Islands. The Outer Islands are much more sparsely populated and much less industrialized than Java. In 1996 there were only 4,339 formal sector firms in the outer regions (or 0.003 firms per square kilometer) compared with 18,506 (0.145 per square kilometer) in Java and many of these were involved in the processing of natural products like wood and rubber.

The linkage terms in the first three columns include links to firms on all islands. In Column (4), Table 2 we re-estimate the equation for Java but now exclude links to the Outer Islands. The results show that linkages to the Outer Islands do not generate agglomeration externalities for firms on Java - the coefficients in columns 2 and 4 are almost identical. These results underpin the difficulty the Indonesian government has experienced in trying to move industry to the outer regions. Not only is the very small number of firms in these regions a problem, the Outer Islands are so far from Java so as to not benefit from the existence of the Javanese markets and suppliers.[22]

The coefficients on the percentage of output exported and the percentage of inputs imported are positively signed and significant in all of the specifications, confirming that the more internationally focused firms pay higher wages. To check that these results are not sensitive to the way trade is included, we re-estimate column (4) with the alternative treatment of international trade (described above) and report the results in the final column. Prior to 1985 Indonesian government regulation forced all international shipping through one of four ports - Tanjung Priok (Jakarta) and Surabaya in Java, Belawan in North Sumatra and Ujungpandang in Sulawesi. Since 1985 investment in port infrastructure has remained centered on these four ports and they remain the most important gateways for international freight. We include imports as a separate term, adjusted by distance to the nearest of these port; and we include exports inside the market access term, also adjusted by distance to the

[22] The insignificance of the linkage variables for outer islands persists with the inclusion of further controls.

nearest port. Both the exports and imports terms remain significant. It is difficult to interpret the coefficient on distance as a spread of externalities given that the distances are only to the port and not to the trading partner but the statistically significant estimate of δ_X as 0.55 shows that exporting firms benefit from being close to a port. The distance coefficient on imports, δ_m, is 0.44 but insignificant, suggesting that access to imports is unaffected by a firm's location within Java. Note that these firms do not necessarily import the goods themselves, they may buy imported inputs from an importing agent and hence being close to a port may be less vital.

The estimates of the domestic supply and market access parameters are almost completely unchanged by the new treatment of trade - the coefficient on supply access is slightly higher and the one on market access slightly lower but both fall well-within the 95% confidence interval of the column (4) estimates. Both the import and export terms remain significant. Given that this more complicated alternative specification does not affect the market and supply access parameters, subsequent specifications will use the simpler specification.

Note that the Jakarta dummy is insignificant in Columns (2) and (4). Thus, having controlled for the market and supplier access that Jakarta provides, there are no additional benefits from being in the nation's capital. Below we restrict our attention to more closely characterizing the linkages on Java (excluding linkages to the Outer Islands).

4.3. Additional controls

Table 3 examines whether the results for Java are robust to the addition of further controls.

Other sources of agglomeration In Column 2 of Table 3 we add variables that attempt to capture the other forces of agglomeration - labor pooling and technological spillovers. The labor pooling index is strongly significant and positive, suggesting that firms benefit from the presence of other firms that use a similar mix of skills and as a result will be more productive and pay higher wages. To capture technological spillovers we include the number of firms in the firm's own 5-digit industry. This is calculated for each district and then distance-weighted in the same way as the market and supply access variables. It is negatively signed and significant indicating that proximity to other firms in the same industry reduces the zero profit wage. It may be that the benefits of spillovers are offset by competition effects (even though we have controlled for competition by also including the market share variable - the firm's share of Java-wide same 5-digit industry output - which has a positive and significant coefficient as hypothesized), or that spillovers arise through other channels not captured by this variable. The coefficient on distance, δ_3, is insignificant indicating that competition comes from firms with equal force from any location within Java.

Industry and firm specific variables The industry dummies are intended to capture differences between fixed and variable costs and industry mark-ups. The results so far include controls at the 2 digit level, however these industry differences may persist within the 2 digit categories and so Column 3 of Table 3 presents the results with 3-digit industry dummies. The coefficients on the linkage terms only change very slightly.[23] The spillover variable is now insignificant so we drop this variable from subsequent specifications.

Industry wage differentials are known to exist for a number of reasons that are not in the theoretical model and that have not so far been controlled for - such as differences in human capital requirements and differing firm characteristics. Column 4 adds these additional controls. Specifically, the percentage of workers that are tertiary educated, high school educated and female, firm size (number of workers), the percentage of government ownership and the percentage of foreign ownership in the firm. In addition we control for the education attainment of the population within each district. The variable *skill* is calculated from the 1995 Intercensal Survey and is the percentage of a district's population that has at least a high school education. Adding these controls increases the adjusted R^2 from 0.37 (with the 3 digit dummies) to 0.47. All of the additional controls are strongly statistically significant and are signed as expected. For example, a one percentage point increase in the percentage of workers who are female decreases average firm wages by 0.32%. The coefficients on the

[23] We also estimated the equations with 4-digit dummies (not reported here). The coefficients on the linkage terms, and the estimates of the δs were the same as with the 3-digit industry dummies.

market and supply access variables remain statistically significant and are slightly smaller in magnitude.

Location specific effects A potential concern with our estimates is that we may be picking up a relationship that is being driven by a third omitted variable that is correlated with both wages and the linkage variables. For example, it may be that firms are attracted to districts which have good existing infrastructure such as roads, telecommunications and a skilled workforce or that are attractive to live in and that wages are bid up in these areas. We have already controlled for the skill level of the population, now we add controls for exogenous amenity. Previous studies have used variables reflecting the weather of locations - following Roback (1982) average temperature, humidity and wind speed are typically used. These variables do not adequately capture differences in exogenous amenity in Java which are almost invariably hot and humid.[24] Instead, to capture exogenous amenity we have included a dummy variable for whether the district is on the coast, the distance to the closest major port and another measuring the percentage of the district's area that is swamp land. We also include a measure of the percentage of potentially arable land that is not housing as an inverse proxy of the price of immobile factors and hence expect this variable to have a positive effect on wages. All these variables are at the district level.

Column 5 controls for exogenous amenity in one further way. We include the total number of formal sector manufacturing firms in each district as an explanatory variable. This variable reflects the attractiveness of a district to firms (including pre-existing infrastructure) so we would expect it to be positively signed. To reduce the possibility of this variable being correlated with the error term we lag it 10 years. This takes us back to the early stages of Java's rapid industrialization. The number of formal sector firms almost doubled in Java between 1986 and 1996 (10,159 compared to 18,506).

All of the additional variables are signed as expected but only being on the coast and the number of firms in 1986 are statistically significant. The ten-year lagged number of firms

[24] Bandung is an exception to this. Its maximum temperatures hover around the mid 20's (celsius), compared to the low 30's for most other locations. In the sensitivity analysis we experiment with dropping Bandung and the results are not sensitive to its exclusion.

is an important determinant of wages, but the extent of a district's industrialization is not driving our supply and market access results. The coefficients on the linkage terms remain significant and the point estimates remain similar in magnitude.

Interpretation of Magnitudes Column (6) presents our preferred specification. It drops the insignificant location-specific variables. Market and supply access have a significant positive effect on wages of similar magnitude: an increase in supply access of 10% increases wages by 0.9% and an increase in market access of 10% increases wages by 1.5%. Most of this benefit dissipates over a short distance: only 10% of the benefit of market access spreads farther than 108 km and only 10% of the benefit of supply access spreads beyond 262 km. Another way of examining the magnitude of the effects is to analyze the effect of reducing distance between all districts. For example, suppose all districts were 20% closer to each other than they are now. Our results indicate that the resulting improved supplier access would lead to an average increase in wages of 1.7% and a maximum of 7.2%; and the improved market access would lead to an average increase of 2.9%, with a maximum of 13.1%.

To examine the relative magnitude of the different sources of agglomeration we consider how an increase in each variable from the 10th to the 90th percentile affects wages. We find that market access has the largest average effect on wages of 26.6%; then supplier access with an average of 21.8%; and labor pooling the smallest effect of 11.9%. Similarly, increasing each variable by an average of 10 percentiles, 20 percentiles or 25 percentiles shows the linkage variables to have the largest effect. For example, the results from increasing variables by an average of 25 percentiles are as follows: market access increases wages by 9.6%; supplier access by 8.4%; and labor pooling by 3.1%.[25] This contrasts with Dumais, Ellison and Glaeser (2002) that finds labor pooling to have the largest effect in the US. Labor pooling may be less important in a developing country because skills are not as differentiated as in a developed country.

[25] This is calculated by averaging the effect of an increase from the 25th to the 50th percentile and from the 50th to the 75th percentile. This is consistent with the elasticities. A 10% increase in labor pooling results in a 0.09% increase in wages which is significantly smaller than the market and supply access effects.

4.4. Sensitivity Tests

Table 4 presents the results of a number of sensitivity tests to explore the possibility of endogeneity arising from reverse causality. That is, we are concerned that the location of firms, and hence the patterns of supply and market access may be determined by wages, rather than the reverse. First, following the approach of Hanson (2000) and Keller (2002) we re-estimate the equation with the full set of controls but dropping districts that individually constitute more than 2% of Indonesia's GDP. This drops the main industrial centers of Jakarta, Surabaya and Bandung. Wages in these large centers of economic activity are the most likely to determine location patterns both within these centers and in neighboring districts. Hence the sensitivity of the results to dropping these observations gives us an indication of the extent of any endogeneity bias in our results. Dropping these cities also reduces the possibility of simultaneity bias arising from natural geographic features in these locations that may explain agglomeration - for example, Jakarta and Surabaya's natural harbours and Bandung's elevated position.

Second, in a similar vein, we drop the own district component of the market and supplier access variables. If the linkage terms were a function of wages then this is more likely to be the case for own district effects.

Third, we lag both the linkage variables five years. This reduces the possible correlation between the error term and these variables. However, to the extent that these variables are correlated over time any endogeneity that exists will persist.

Finally, we drop observations on industries for whom more than 20% of inputs come from within their own 5 digit industry. This reduces the scope for reverse causality coming through the supply access variable and also ensures that the variable is indeed picking up vertical linkages rather than horizontal spillovers.

The estimates of all four market access and supply access parameters $(\gamma_{1,}\gamma_{2,}\delta_{1,}\delta_2)$ are robust to all of these sensitivity tests. The coefficients remain significant. The point estimates in many cases are almost exactly the same and where they differ they lie well-within the 95 percentile confidence interval of the original estimates. Having established that the results are robust we now examine changes over time.

4.5. Changes Over Time

We compare results for 1983 and 1991 with 1996 in Table 5. Summary statistics are presented in Table A1. Some of the control variables are not available in the earlier years so we also present results for 1996 with a smaller, comparable set of regressors. The supply access estimates are significant in all years and very stable across time. The market access parameters are stable from 1991 to 1996. However, the coefficient γ_2 is significantly smaller in 1983 (0.09 compared with 0.19 in 1996 and 1991). This suggests that market access has become more important over time. The point estimate is much higher in 1983 than in the later years (decreasing from 7.2 in 1983 to 3.4 in 1991 and further to 2.6 in 1996). This suggests that the market access externality may have become less localized over time. However the estimate for 1983 is estimated so imprecisely that its 95% confidence interval encompasses the estimates from the other years.

As transport infrastructure and telecommunications improvements take place one might expect that externalities arising from agglomeration benefits would spread over longer distances. However, as technologies become more advanced and products become more sophisticated the need for face to face communication becomes more important making externalities even more localized. These two offsetting effects may explain why the spread of the supply access externality has remained constant over time while the market access effect may have become more diffused. Given that a large part of the market access component comprises final demand from consumers, where face to face contact between producers and consumers is not so important, the fall in transport costs may dominate the effect.[26]

The stability of the results across time is significant in two senses. First in terms of the robustness of our results - the variables for 1991and 1983 were constructed from a completely separate set of data and produce very similar estimates. Second, in a substantive sense -

[26] These findings are consistent with the international trade and distance literature. For example, Berthelon and Freund (2003) find that the effect of distance on international trade has not changed for 75% of industries but has become more important for 25% of industries, suggesting that these industries trade less with more distant countries than they did 20 years ago.

even though Indonesia experienced dramatic change between 1983 and 1996 in terms of improvements in infrastructure, the effects of supplier access remained largely unchanged, with some increase in the market access effect. This is consistent with findings of studies such as Dumais, Ellison and Glaeser (2002) which show that although there is a large amount of individual entry and exit of firms over time, the overall patterns of agglomeration are persistent.

5. Conclusions

This paper examines the benefits of agglomeration arising from vertical linkages between firms. Using firm level data for Indonesia from 1996, 1991 and 1983, we show that firms benefit greatly from proximity to a large supply of inputs and good market access. Firms with the best supply or market access can afford to pay more than 20% higher wages than those with the poorest access. Labor pooling is less quantitatively important and we were unable to identify any positive effects from technology spillovers. These results are robust to controlling for more standard explanations of wage variation such as skill levels and firm size; and infrastructure variables. The results are also robust to a set of sensitivity tests designed to test the extent of endogeneity of the market access and supply access variables.

Further, we found that the benefits of vertical linkages are highly localized. Firms do benefit from vertical linkages but not if they are located in the periphery. Firms located 108 kms (262 kms) from the market (supply) access agglomeration receive only 10% of the benefit of those located at the source. We show that firms located in Indonesia's Outer Islands are too far away to benefit from the agglomeration of industries on the main island of Java.

The large agglomeration benefits arising from vertical linkages combined with the high localization of the benefits can explain why firms are reluctant to relocate to low wage areas. These results also underscore the difficulty governments around the world have in generating economic growth in far flung regions - where the citizens are often the poorest and benefit the least from economic growth. Although our results are based on Indonesian data, they clearly have more general implications. Large regional inequalities are a world-wide phenomenon and governments continue to spend large sums of money to try to attract firms to poorer regions. Given the size of the estimated agglomeration externalities, our results suggest that overcoming the attraction of existing agglomerations is likely to continue to be a difficult task.

References

[1] Alatas, Vivi and Lisa Cameron (2003). The Impact of Minimum wages on Employment in a Low Income Country: An Evaluation Using the Difference-in-Difference Approach. *World Bank Policy Research Working Paper* 2985, 1-31.

[2] Amiti, Mary (2003). Location of Vertically Linked Industries: Agglomeration versus Comparative Advantage.*European Economic Review*, forthcoming.

[3] Amiti, Mary (2001). Regional Specialisation and Technological Leapfrogging. *Journal of Regional Science*, 41(1), 149-172.

[4] Amiti, Mary and Lisa Cameron (2002), Demand Linkages and the Spatial Distribution of Employment and Wages in a Developing Country. University of Melbourne, mimeo.

[5] Blalock, Garrick and Paul Gertler (2003). Technology from Foreign Direct Investment and Welfare Gains through the Supply Chain. Cornell University, mimeo.

[6] Badan Pusat Statistik (2000a). Produk Domestik Regional Bruto Kabupaten/Kotamadya di Indonesia 1995-1998, Jakarta.

[7] Badan Pusat Statistik (2000b). Luas Lahan Menurut Penggunaannya di Indonesia, 1999, Jakarta.

[8] Badan Pusat Statistik (1997). Tabel Input-Output Indonesia, 1995. Jilid: III, Jakarta.

[9] Biro Pusat Statistik (1998a). Produk Domestik Regional Bruto Kabupaten/Kotamadya di Indonesia 1993-1996, Jakarta.

[10] Biro Pusat Statistik (1995). Produk Domestik Regional Bruto Kabupaten/Kotamadya di Indonesia 1983-1993, Jakarta.

[11] Biro Pusat Statistik (1992). Tabel Input-Output Indonesia, 1990. Jilid: III, Jakarta.

[12] Berthelon, Matias and Caroline Freund (2003). On the Conservation of Distance, World Bank mimeo.

[13] Dixit, Avinash K. and Joseph Stiglitz (1977). Monopolistic Competition and Optimum Product Diversity. *American Economic Review* 67(3), 297-308.

[14] Dumais, Guy, Glenn Ellison and Edward L. Glaeser (2002). Geographic Concentration as a Dynamic Process. *Review of Economics and Statistics* LXXXIV(2), 193-204.

[15] Ethier, W. (1982). National and International Returns to Scale in the Modern Theory of of International Trade. *American Economic Review* 72, 389-405.

[16] Fujita, Masahisa , Paul Krugman and Anthony J. Venables (1999). *The Spatial Economy: Cities, Regions and International Trade*. Cambridge: MIT Press.

[17] Hanson, Gordon (2000). Market Potential, Increasing Returns, and Geographic Concentration", revised version of NBER Working Paper No. 6429 (1998).

[18] Harris, C., (1954). The Market as a Factor in the Localization of Industry in the United States. *Annals of the Association of American Geographers* 64, 315-348.

[19] Henderson, J. Vernon and Ari Kuncoro(1996), Industrial Centralization in Indonesia, *World Bank Economic Review* 10(3), 513-40.

[20] Hirschman, A. (1958). *The Strategy of Economic Development.* New Haven, GT: Yale University Press.

[21] Keller, Wolfgang (2002). Geographic Location of International Diffusion. *American Economic Review*, 92(1), 120-42.

[22] Krugman, Paul (1991a). Increasing Returns to Scale and Economic Geography. *Journal of Political Economy* 99(31), 483-499.

[23] Krugman, Paul (1991b). *Geography and Trade.* The MIT Press.

[24] Krugman, Paul. and Anthony J. Venables (1995). Globalization and the Inequality of Nations. *Quarterly Journal of Economics* 110, 857-880.

[25] Marshall, A. (1920). *Principles of Economics.* London: Macmillan.

[26] Nickell, Stephen J. (1996). Competition and Corporate Performance. *Journal of Political Economy* 104: 724-746.

[27] Redding, Steve and Anthony J. Venables. (2000) Economic Geography and International Inequality. CEPR Discussion Paper 2568.

[28] Rogers, W.H. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin* 13: 19-23. Reprinted in Stata Technical Bulletin Reprints, vol. 3, 88-94.

Figure 1: Geographic Distribution of Formal Sector Manufacturing Output, 1996

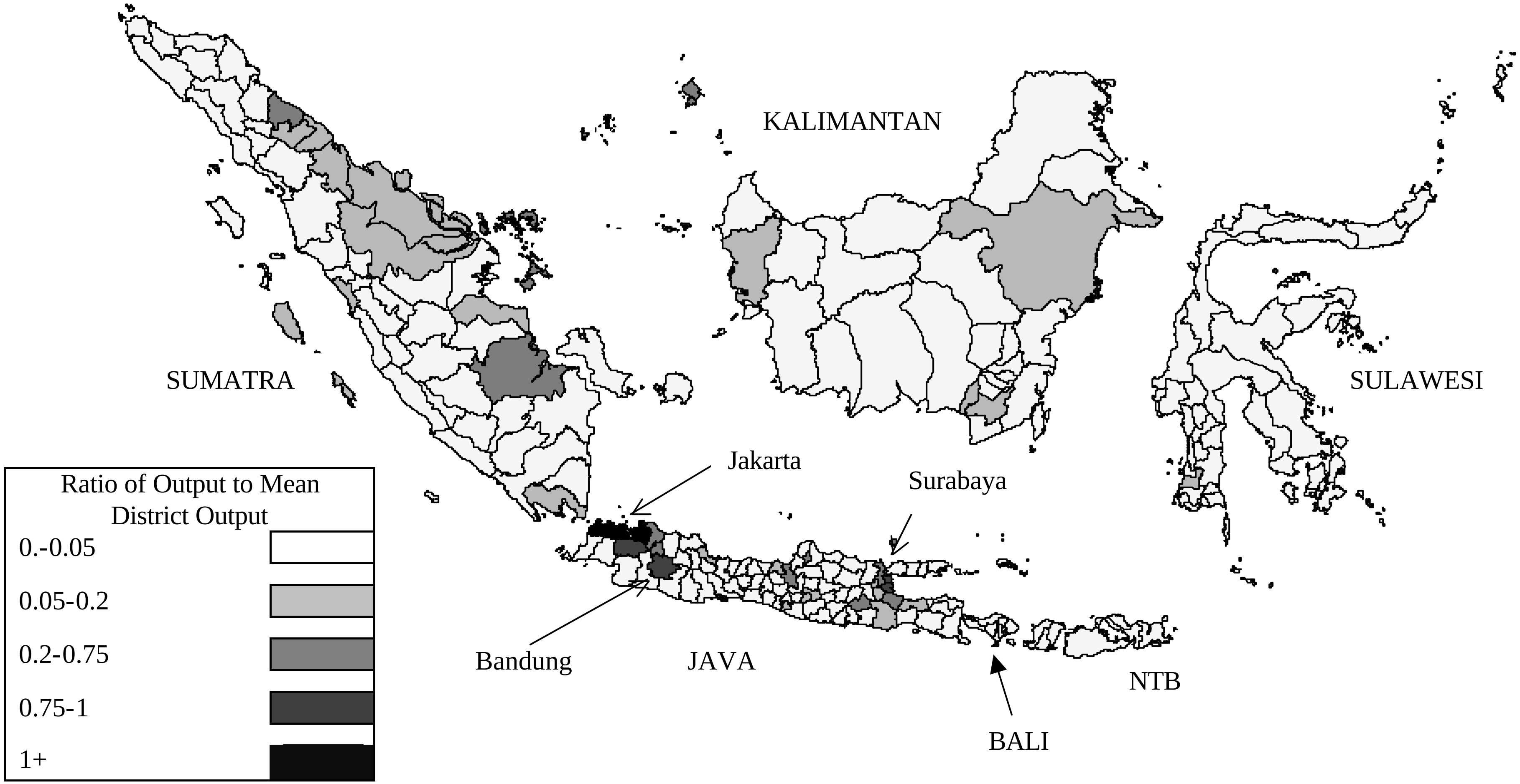

Table 1: Summary Statistics

	Indonesia				Java				Outer Islands			
	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
wage	7905.53	6226.97	920.85	51877.92	7893.89	6245.25	928.81	51877.92	7968.18	6128.72	920.85	50399.16
supplier access	0.05	0.08	0	1	0.05	0.08	0	1	0.02	0.04	0	0
market access	0.02	0.04	0	1	0.02	0.04	0	1	0.01	0.02	0	0
imports	0.10	0.26	0	1	0.11	0.25	0	1	0.09	0.26	0	1
exports	0.17	0.34	0	1	0.14	0.32	0	1	0.30	0.42	0	1
size	206.21	594.75	12	23516	205.28	613.45	12.00	23516	211.22	481.90	14	5184
foreign ownership	0.05	0.19	0	1	0.05	0.19	0	1	0.06	0	0	1
govt ownership	0.01	0.12	0	1	0.01	0.11	0	1	0.03	0.15	0	1
female participation	0.37	0.30	0	1	0.37	0.30	0	1	0.40	0.31	0	1
high school education	0.31	0.27	0	1	0.29	0.26	0	1	0.39	0.28	0	1
tertiary education	0.03	0.06	0	0.93	0.03	0.06	0	0.93	0.03	0.06	0	0.81
population skill level	0.36	0.13	0.09	0.62	0.36	0.14	0.09	0.62	0.35	0.12	0.12	0.56
labour pooling	-0.03	0.04	-0.39	0.00	-0.03	0.04	-0.36	0.00	-0.03	0.04	-0.39	0
spillovers	50.66	95.24	1	393	56.79	101.68	1	393	17.67	31	1	128
competition	0.01	0.07	0	1	0.0145	0.0662	0	1				
firms86	338.05	299.88	0	1143	374.24	305.30	2	1143	143.29	165	0	450
coast	0.67	0.47	0	1	0.62	0.48	0	1	0.91	0	0	1
swamp	0.03	0.04	0	0.60	0.02	0.04	0	0.14	0.05	0.07	0	0.60
land	0.59	0.20	0	0.96	0.56	0.20	0.06	0.96	0.73	0.12	0	0.96
skill	0.36	0.13	0	0.62	0.36	0.14	0.09	0.62	0.35	0.12	0	0.56
port	132.20	158.57	0	944.18	97.27	97.86	0	350.90	320.23	258.88	0	944.18
# industries	172				170				128			
# kabupatens	177				87				90			
N	13472				11361				2111			

TABLE 2: BASIC SPECIFICATION

	INDONESIA	JAVA	OUTER ISLANDS	JAVA	JAVA ALTERNATIVE TRADE
	(1)	(2)	(3)	(4)	(5)
Supply Access - γ_1	0.0556	0.1031	0.0159	0.0994	0.1201
	(0.0191)	(0.0239)	(0.0098)	(0.0233)	(0.0198)
Distance, km/100 - δ_1	1.7899	0.9962	3.0841	1.0654	0.9061
	(0.7069)	(0.2329)	(2.4139)	(0.2636)	(0.1877)
Market Access - γ_2:	0.1071	0.2224	0.0022	0.2215	0.2022
	(0.028)	(0.0395)	(0.0194)	(0.0389)	(0.0342)
Distance, km/100 - δ_2	2.8104	2.7127	5.4849	2.6943	2.782
	(1.2288)	(0.4206)	(49.3683)	(0.4108)	(0.4863)
Exports	0.3348	0.2561	0.4611	0.2559	0.3805
	(0.0587)	(0.0417)	(0.0649)	(0.0417)	(0.0848)
Distance to port/100, km - δ_X					0.5581
					(0.1141)
Imports	0.4059	0.38	0.3151	0.3806	0.5265
	(0.0869)	(0.0942)	(0.0723)	(0.0942)	(0.1015)
Distance to port, km/100 - δ_M					0.4478
					(0.5909)
Region Dummies:					
Sumatra	0.3414		0.0801		
	(0.0679)		(0.0688)		
Kalimantan	0.5191		0.2356		
	(0.0955)		(0.0939)		
Sulawesi	0.2682		-0.2134		
	(0.0966)		(0.0838)		
Jakarta	0.1124	-0.0337		-0.0316	-0.0322
	(0.0316)	(0.0309)		(0.0312)	(0.0312)
Industry Dummies:					
Wood/Furniture	0.1207	0.2297	0.1918	0.2282	0.2104
	(0.039)	(0.0327)	(0.098)	(0.0329)	(0.0365)
Paper/Printing	0.3681	0.3643	0.5494	0.3636	0.3634
	(0.03)	(0.0248)	(0.0933)	(0.025)	(0.0231)
Chemicals/Plastics	0.3052	0.3273	0.3942	0.327	0.3200
	(0.0721)	(0.0711)	(0.106)	(0.0712)	(0.0684)
Non-metallic Minerals	0.1874	0.2266	0.3351	0.2258	0.2267
	(0.0367)	(0.0312)	(0.0824)	(0.0313)	(0.0277)
Metals	0.5573	0.5047	0.5397	0.5044	0.5173
	(0.1126)	(0.1114)	(0.1419)	(0.1114)	(0.1086)
Machinery and Components	0.3847	0.3563	0.6174	0.3557	0.3471
	(0.0487)	(0.0398)	(0.1094)	(0.0401)	(0.0332)
Other	0.0437	0.0447	0.2501	0.0444	0.0405
	(0.0509)	(0.0503)	(0.0827)	(0.0505)	(0.0468)
Constant	8.9272	9.3125	8.3897	9.3082	9.2800
	(0.0648)	(0.0627)	(0.1307)	(0.0625)	(0.0602)
Linkage Variables Coverage:	Indonesia	Indonesia	Indonesia	Java	Java
RSS	3736.3	2926.7	571.9	2927.8	2912.9
R-squared	0.29	0.33	0.35	0.33	0.33
N	13472	11361	2111	11361	11361

* Standard errors in parentheses.

Table 3: Estimates for Java

	Basic	+ spillovers + competiton	+3 digit	+ firm characteristric	+exog. amenity +initial firms	preferred specification
	(1)	(2)	(3)	(4)	(5)	(6)
Supply Access - γ_1	0.0994	0.1232	0.1338	0.1029	0.0876	0.093
	(0.0233)	(0.0235)	(0.0223)	(0.0172)	(0.0189)	(0.0193)
Distance, km/100 - δ_1	1.0654	0.936	0.8709	0.8993	0.9177	0.8771
	(0.2636)	(0.1925)	(0.1706)	(0.1602)	(0.3665)	(0.1703)
Market Access - γ_2:	0.2215	0.1903	0.1874	0.1399	0.1371	0.1450
	(0.0389)	(0.0327)	(0.034)	(0.0327)	(0.0289)	(0.0329)
Distance, km/100 - δ_2	2.6943	3.3643	3.5493	2.4598	2.2128	2.1368
	(0.4108)	(0.4812)	(0.4972)	(0.6391)	(0.5924)	(0.5575)
Exports	0.2559	0.212	0.2039	0.1567	0.1588	0.1568
	(0.0417)	(0.0378)	(0.0329)	(0.0214)	(0.0227)	(0.0217)
Imports	0.3806	0.3233	0.3108	0.1803	0.1840	0.1837
	(0.0942)	(0.0924)	(0.0867)	(0.0621)	(0.0599)	(0.0608)
Labour Pooling (province)		0.4235	0.4172	0.2567	0.2634	0.2639
		(0.0444)	(0.043)	(0.037)	(0.0351)	(0.0374)
Spillovers: γ_3		-0.0196	-0.0189			
		(0.0097)	(0.012)			
Distance, km/100 - δ_3		14.7985	15.8318			
		(24.2186)	(25.0457)			
Competition		0.9918	1.0085	0.5034	0.5137	0.5084
		(0.1487)	(0.1488)	(0.1288)	(0.1303)	(0.1291)
Jakarta	-0.0316	-0.0019	-0.0158	-0.0172	0.0577	0.0195
	(0.0312)	(0.0305)	(0.0329)	(0.0261)	(0.0407)	(0.028)
Firm size per 100				0.0058	0.0058	0.0058
				(0.0018)	(0.0018)	(0.0018)
Foreign ownership				0.3205	0.327	0.3283
				(0.0493)	(0.0492)	(0.0509)
Government ownership				0.3234	0.3265	0.3208
				(0.0503)	(0.0499)	(0.0495)
Female participation				-0.3266	-0.3257	-0.3257
				(0.0661)	(0.0651)	(0.0669)
High school educated				0.3827	0.3850	0.3876
				(0.0327)	(0.0321)	(0.0323)
Tertiary educated				1.7069	1.7054	1.709
				(0.1165)	(0.1156)	(0.117)
Kabupaten skill level				0.2960	0.4157	0.216
				(0.1151)	(0.2266)	(0.1064)
# Firms in 1986 per 100					0.0098	0.0105
					(0.0035)	(0.0027)
Coast					0.0395	0.0255
					(0.0179)	(0.0126)
Swamp					-0.2065	
					(0.5082)	
Land					0.2214	
					(0.1955)	
Distance to port, km					-0.0086	
					(0.0236)	
Industry	2 digit	2 digit	3 digit	3 digit	3 digit	3 digit
RSS	2927.8	2795.4	2745.2	2317.8	2308.3	2311.9
R-squared	0.332	0.362	0.373	0.471	0.473	0.472
N	11361	11361	11361	11361	11361	11361

Table 4: Sensitivity Tests

	Comparison Col (6) Table 3 (1)	Small GDP Kabupaten (2)	Drop own Kabupaten (3)	Lagging 5 years (4)	Dropping if Own Industry Input Use (5)
Supply Access - γ_1	0.093	0.0927	0.0742	0.1035	0.0938
	(0.0193)	(0.021)	(0.017)	(0.0165)	(0.0209)
Distance, km/100 - δ_1	0.8771	0.8107	0.9553	1.1053	0.8318
	(0.1703)	(0.1664)	(0.1979)	(0.2198)	(0.1665)
Market Access - γ_2:	0.1450	0.1658	0.1462	0.1284	0.1535
	(0.0329)	(0.0357)	(0.0317)	(0.0316)	(0.0349)
Distance, km/100 - δ_2	2.1368	2.3208	2.0183	2.0511	2.1193
	(0.5575)	(0.5553)	(0.4454)	(0.5937)	(0.5923)
Exports	0.1568	0.1468	0.1581	0.1643	0.1697
	(0.0217)	(0.0236)	(0.0215)	(0.0229)	(0.0214)
Imports	0.1837	0.138	0.1799	0.1758	0.1547
	(0.0608)	(0.0807)	(0.0593)	(0.0603)	(0.0685)
Labour Pooling (province)	0.2639	0.1996	0.2642	0.2676	0.2808
	(0.0374)	(0.0486)	(0.0368)	(0.0368)	(0.0398)
Competition	0.5084	0.4434	0.6419	0.7252	0.585
	(0.1291)	(0.1707)	(0.1363)	(0.1318)	(0.1477)
Industry	3 digit	3 digit	3 digit	3 digit	3 digit
RSS	2311.9	1501.7	2322.9	2299.0	2027.8
R-squared	0.472	0.499	0.469	0.472	0.469
N	11361	7317	11359	11310	10152

Table 5: Comparisons Across Years

	1996	1991	1983
Supply Access - γ_1	0.0985	0.1117	0.1129
	(0.0241)	(0.0315)	(0.0478)
Distance, km/100 - δ_1	0.9208	0.8453	0.8178
	(0.2022)	(0.3539)	(0.4211)
Market Access - γ_2:	0.1944	0.1927	0.0903
	(0.0384)	(0.045)	(0.0485)
Distance, km/100 - δ_2	2.6115	3.3872	7.2099
	(0.4782)	(1.009)	(5.535)
Exports	0.1527	0.0748	
	(0.0312)	(0.046)	
Imports	0.2409	0.1647	0.0892
	(0.0825)	(0.0481)	(0.0466)
Competition	0.7611	1.1025	0.4186
	(0.1445)	(0.1644)	(0.1286)
Jakarta	0.0166	0.0867	0.2007
	(0.0303)	(0.0334)	(0.051)
firm size per 100	0.0067	0.0071	0.026
	(0.0023)	(0.003)	(0.0066)
Foreign ownership	0.4308	0.669	1.2583
	(0.0652)	(0.1341)	(0.088)
Government ownership	0.4419	0.4828	0.5483
	(0.0611)	(0.064)	(0.0614)
# Firms lagged 10 years per 100*	0.011	0	0.0186
	(0.0037)	(0.0001)	(0.0071)
Coast	-0.0015	0.0116	-0.0061
	(0.0154)	(0.0258)	(0.0322)
Industry	3 digit	3 digit	3 digit
RSS	2708.6	2267.2	1192.5
R-squared	0.382	0.379	0.422
N	11361	7927	3857

* For 1983 we used the first available year of SI data which is 1976.

Table A1: Summary Statistics for 1991 and 1983

	Java - 1991				Java - 1983			
	Mean	Std	Min	Max	Mean	Std	Min	Max
wage	4339.46	4010.37	549.70	36368.39	1700.42	1527.10	167.81	10588.57
supplier access	0.05	0.08	0	1	0.05	0.08	0	1
market access	0.02	0.04	0	0.94	0.01	0.04	0	1
imports	0.15	0.30	0	1	0.24	0.35	0	1
exports	0.11	0.29	0	1				
jakarta	0.23	0.42	0	1	0.27	0.44	0	1
size	193.92	512.23	20	14830	129.71	281.30	10	5338
foreign ownership	0.03	0.15	0	1	0.03	0.14	0	1
govt ownership	0.02	0.14	0	1	0.03	0.17	0	1
firms86	264.77	267.24	0	869	303.75	274.48	4	869
coast	0.55	0.50	0	1	0.44	0.50	0	1
swamp	0.02	0.03	0	0.14	0.03	0.04	0	0.14
# industries	157				140			
# kabupaten	83				75			
N	7927				3857			