Economic Geography and Wages*

Mary Amiti
International Monetary Fund and CEPR

Lisa Cameron
University of Melbourne

July 30, 2003

Abstract

This paper estimates the effect of demand and cost linkages arising from vertical relations between manufacturing firms on wages in Indonesia. By utilizing unusually detailed manufacturing census data from Indonesia, we explicitly 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 analysis is firmly based on new economic geography theory developed by Krugman and Venables. The detailed location and input data allows us to employ a more comprehensive measure of spatial input/output linkages than in any previous study. This is also the first study of vertical relations for any developing country. The results show that demand and cost linkages have a significant impact on manufacturing wages. An understanding of the extent and strength of spatial linkages is crucial in shaping policies that seek to influence regional development - an issue that is especially pertinent in developing countries.

Comments welcome.

*We would like to thank Caroline Freund, Bill Griffiths, Russ Hilberry, David Hummels, Guy Lim, Stephen Redding and Tony Venables for their comments. This paper has been presented at the CEPR European Research Workshop in International Trade in Munich 2002, the Empirical Investigations in International Trade workshop in Atlanta 2002, 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 industries and regions within countries. In Indonesia in 1996, the average wage paid by firms in the chemical industry in the 95th percentile was 10 times that in the fifth percentile, with the highest paying firms located on the main island of Java and the lowest paying firms in Sumatra, one of the outer islands. Differences also exist between regions within Java. For example, in the textile industry firms in the 95th percentile in Indonesia paid 6 times higher wages than those in the fifth percentile - they were located on the main island of Java, 671 kilometers apart. This is rather surprising given that factors of production within a country are supposed to be freely mobile. Similarly, a standard neoclassical model would predict that manufacturing firms would locate in low wage regions and also bid away these wage differences. This has clearly not been the case.

The reasons why firms may not want to relocate to low wage regions could be due to the agglomeration benefits they enjoy from being close to other firms. Three main sources of geographic agglomeration have been identified by Marshall (1920) - they are (i) input/output linkages[1]; (ii) labour 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 'new 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 their 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 industry specific and region specific wages in Indonesia.[3]

[1] Also see Hirschman (1958) on input/output linkages.

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

[3] Benefits could also arise through higher total factor productivity and profits.We choose to estimate the effects on wages as this variable is likely to be the most accurately measured, and is more closely related to the theory.

We use an unpublished data set which provides detailed information on firms' vertical linkages with their suppliers - and the information on firms' geographic locations to model the demand and cost linkages proposed in Krugman and Venables (1995). Our study is the first that utilizes detailed connections between input suppliers and final producers across space to examine cost linkages, rather than relying on aggregate input/output tables.[4] It is also the first on any developing country that explicitly models vertical linkages between firms. We use three waves of Indonesia's Manufacturing census, which is a complete enumeration of large and medium scale manufacturing 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.

A further important distinguishing feature of this paper is that we do not assume that externalities are bound by district borders. For example, most locational studies examine the relationship between the variable of interest and the characteristics of the region in which the firm or industry resides. This ignores the characteristics of neighboring and other regions. Using the locational data, we explicitly model the spillovers into neighboring regions and estimate how far these linkages extend. It is the spatial linkages that determine the extent to which the fruits 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.

There are a small number of studies that are closely related to ours. Redding and Venables (2002) use the same theoretical structure to explain variation in average manufacturing wages at the country level. Their measures of supply and market access are based on import and export dummies coming out of a gravity equation. In contrast, we use detailed information on which inputs firms use to construct our measures of market and supplier access. Hanson (1998) estimates a spatial wages equation for US manufacturing firms, with the average wage in each county as the dependent variable. His theoretic framework is based on an earlier new economic geography model, Krugman (1991), where agglomeration forces arise due to the mobility of workers - vertical linkages are not modeled. Dumais, Ellison and Glaeser (1997) estimates the effect of Marshall's three sources of agglomeration on changes in employment using US firm level data at the metropol-

[4] Our dataset includes firm level information on intermediate inputs from firms surveyed in 1998. 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.

itan level and state level. They find evidence of all three of Marshall's 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 there are large changes at the plant level but small changes in agglomeration patterns at the aggregate level. [*other references to be added].* None of these studies have focused on developing countries.

Indonesia's geography, public policy and political history make it an especially interesting laboratory in which to test the predictions of new economic geography theories. Indonesia is a large developing country. Although its 200 million people are spread over 900 islands and an east-west distance of 5500 kms, manufacturing is very heavily concentrated on the island of Java. About three quarters of non-oil and gas manufacturing is located on Java, and within Java manufacturing is 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. Decentralization is currently a major political and public policy issue in Indonesia. 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 policies in an attempt to spread the benefits of industrialization to the other (outer) islands - with limited success. 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 (see Henderson and Kuncoro, 1996). Blalock (2001) estimates whether supply chains in Indonesia are the conduit for transferring technology from foreign direct investment using aggregate input/output tables. However he too implicitly assumes that the externalities are bound by district borders (at the provincial level) and hence does not model the spatial dimension that is the focus of our paper.

The results show that demand and cost linkages have a significant impact on manufacturing wages in Indonesia. We also test for the effects of other sources of agglomeration such as labour pooling and technological externalities; and we control for more standard labour explanations of wages such at the average size of firms and diferent skill levels. We find our results are robust to these additional controls. Firms do benefit from vertical linkages but these benefits are highly localized. These findings highlight why government policies often fail in trying to relocate individual industries to peripheral areas. Industries benefit from the availability of a large supply of inputs and good market access. We show that firms located in the outer islands are too far away to benefit from the agglomeration of industries on the main island of Java. The geography of Indonesia renders internal trade costs far too high for these firms to receive the benefits of agglomeration.

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

2. Theory

We derive our estimating equation from a new 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 aggregate consumption of varieties of differentiated industry i goods consumed in district k and s^i is the share of income spent on industry i goods. We denote all location specific variables with subscripts and industry specific variables with superscripts.

Aggregate demand for final manufactured goods in industry i can be represented by a quantity index or sub-utility function, C_k^i, 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 $t^i_{lk} \geq 1$.[5] In order to consume one unit of output, consumers must demand t^i_{lk} 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

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

The further a good must travel the more of it melts in transit. Dual to the quantity index is the price index, defined as follows:

$$P^i_k = \left[\sum_{l=1}^{K} \sum_{v=1}^{n^i_k} \left(p^{iv}_l t^i_{lk} \right)^{1-\sigma^i} \right]^{\frac{1}{1-\sigma^i}} \tag{2.4}$$

where p^{iv}_l 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 l for a variety v of an industry i good produced in district k is given by maximizing the sub-utility functions, equation 2.2, subject to the budget constraint $s^i Y_l$.

$$c^{iv}_{kl} = \left(p^{iv}_k \right)^{-\sigma^i} \left(t^{iv}_{kl} \right)^{1-\sigma^i} s^i Y_l \left(P^i_l \right)^{\sigma^i - 1} \tag{2.5}$$

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

$$\left(L^{iv}_k \right)^{\alpha^i} \left(K^{iv}_k \right)^{\beta^i} \prod_u \left(C^{ui}_k \right)^{\mu^u} = F + b^i x^{iv}_k, \qquad \alpha^i + \beta^i + \sum_u \mu^u = 1, \tag{2.6}$$

[5] We assume that $t^i_{kk} = 1$.

where L_k^{iv} and K_k^{iv} are the labour and capital[6] amounts required by each firm in industry i to produce output, x_k^{iv}. C_k^{ui} is a quantity index aggregated across varieties of intermediate inputs supplied by industry u to industry i, defined analogously to equation 2.2. Hence, industry $u's$ output of intermediate inputs enters the production function of each downstream firm in industry i through a CES aggregator as in Ethier (1982). Note that industry i purchases many varieties of inputs from multiple upstream industries. The cost function is given by

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

The total cost function comprises a fixed cost, F, a constant cost, b^i, and factor prices, where w_k^i is the wage of an industry i firm in district k,[7] r_k is the price of capital in district k (or any other factor of production), and P_k^{ui} is the intermediate input price index of each upstream industry u that supplies inputs to industry i. It is analogous to equation 2.4, with $i = u$.

Profits of a single representative firm in district k are given by

$$\pi_k^{iv} = p_k^{iv} x_k^{iv} - \left(w_k^i\right)^{\alpha^i} r_k^{\beta^i} \prod_u \left(P_k^{ui}\right)^{\mu^u} \left[F + b^i x_k^{iv}\right]. \tag{2.8}$$

The relationship between firms is summarized in Figure 2. A firm in industry i (represented by the middle box), purchases a large variety of intermediate inputs from upstream firms (represented by the first box). This creates a cost linkage via the price index of intermediate inputs, P_k^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 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,

$$p_k^{iv} = \left(w_k^i\right)^{\alpha^i} r_k^{\beta^i} \prod_u \left(P_k^{ui}\right)^{\mu^u} b^i \theta^i, \qquad \theta^i = \frac{\sigma^i}{\sigma^i - 1}. \tag{2.9}$$

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

[7] Although free mobility of labour across industries within a region in theory leads to wage equality across industries, it is well-documented that empirically this is not the case. We allow wages to vary acrosss industries.

Given the symmetry of different varieties we drop the superscript v. 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.

$$x_k^i = \overline{x}^i = \frac{F\left(\sigma^i - 1\right)}{b^i}. \tag{2.10}$$

2.3. Aggregate demand

Demand for industry i goods not only comes from consumers but also from downstream firms, hence total expenditure on industry i, E_l^i, is

$$E_l^i = s^i Y_l + \mu^i n_l^d p_l^d x_l^d \tag{2.11}$$

Downstream firms spend a proportion μ of their total revenue, $p_l^d x_l^d$, on intermediate inputs produced by industry i (the second term in equation 2.11). 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 2.5 with the term $s^i Y_l$ replaced with $\mu^i n_l^d p_l^d x_l^d$. n_l^d is the number of downstream firms in district l.

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

$$c_k^i = \sum_{l=1}^{K} c_{kl}^i = \left(p_k^i\right)^{-\sigma^i} \sum_{l=1}^{K} \left(t_{kl}^i\right)^{1-\sigma^i} E_l^i \left(P_l^i\right)^{\sigma^i - 1}. \tag{2.12}$$

Substituting in for prices in the aggregate demand function, equation 2.12, gives

$$c_k^i = \left(\left(w_k^i\right)^{\alpha^i} r_k^{\beta^i} \prod_u \left(P_k^{ui}\right)^{\mu^u} b^i \theta^i \right)^{-\sigma^i} \sum_{l=1}^{K} \left(t_{kl}^i\right)^{1-\sigma^i} E_l^i \left(P_l^i\right)^{\sigma^i - 1}. \tag{2.13}$$

Substituting in for expenditure and transport costs (equations 2.11 and 2.3) into the aggregate demand function (equation 2.13) and setting demand equals supply in the product market gives

$$x_k^i = \left(w_k^{\alpha^i} r_k^{\beta^i} \prod_u \left(P_k^{ui}\right)^{\mu^u} b^i \theta^i \right)^{-\sigma^i} \tag{2.14}$$
$$\sum_{l=1}^{K} \left\{ e^{-\tau^d \left(\sigma^d - 1\right) d_{kl}} \left(s^i Y_l + \mu^i n_l^d p_l^d x_l^d \right) \left(P_l^i \right)^{\sigma^i - 1} \right\}.$$

This is the main equation we are interested in. The production of industry i goods, x_k^i, depends on supply side factors given by the first bracketed term, which include industry specific and location specific factors; and demand side factors given by the terms in the curly brackets; and the trade costs between location k and l. This equation embodies utility and profit maximization conditions, and product market equilibrium. Setting demand for industry i good equal to the zero profit level of output and rearranging gives the maximum wage industry i firms can afford to pay i.e. the zero profit wage,

$$\left(w_k^i\right)^{\alpha^i} = \left(\left(\overline{x}_k^i\right)^{\frac{1}{\sigma^i}} b^i \theta^i \right)^{-1} r_k^{-\beta^i} \prod_u \left(P_k^{ui}\right)^{-\mu^u} \tag{2.15}$$
$$\sum_{l=1}^{K} \left\{ e^{-\tau^d \left(\sigma^d - 1\right) d_{kl}} \left(s^i Y_l + \sum_d \mu^{di} n_l^d p_l^d x_l^d \right) \left(P_l^i \right)^{\sigma^i - 1} \right\}^{\frac{1}{\sigma^i}}.$$

So the theory posits that wages in location k are a function of industry specific effects D_i, location specific effects D_l, supplier access SA_k^i, and market access MA_k^i:

$$w_k^i = f\left(D_i, D_l, SA_k^i, MA_k^i\right). \tag{2.16}$$

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 or fixed capital. Our main focus is on the effects of supplier access and market access on wages.

Supplier Access The supplier access or cost linkage effect comes through the price indices of intermediate inputs, P_k^{ui}, defined as in equation 2.4 with $i = u$. Note that output does not usually enter the price index because in a Dixit-Stiglitz model it is constant across all firms within an industry. But since firm size does vary significantly in the data we allow for variation in outputs by modifying the price index as follows.

$$P_k^{ui} = \left[\sum_{l}^{K} \left(\frac{1}{X^u} \sum_{v=1}^{N_l} x_l^u \left(p_l^u e^{\tau^u d_{kl}} \right)^{1-\sigma^u} \right) \right]^{\frac{1}{1-\sigma^u}} \qquad (2.17)$$

where x_l^u is the output produced by a firm in upstream industry u in location l. The inner bracketed term sums across the number of upstream firms, N_l, that produce inputs used in the production of industry i goods. Our data do not provide individual input prices as required in equation 2.17 so we approximate it as follows:

$$P_k^{ui} = \left[\sum_{l}^{K} sh_l^{ui} e^{-\tau^u(\sigma^u-1)d_{kl}} \right]^{\frac{1}{1-\sigma^u}}, \text{ where } sh_l^{ui} = \frac{X_l^{ui}}{X^{ui}} = \frac{1}{X^{ui}} \sum_{v=1}^{N_l} x_l^{uv} p_l^{uv}. \qquad (2.18)$$

The term sh_l^{ui} gives the share of the value of intermediate inputs from industry u produced in district l that enter into the production of industry i goods. We have detailed information on the value of each input purchased. We also 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. Since we do not have individual prices, we cannot estimate the exponent on the price term, σ^u in equation 2.17. However, the cost linkages are still well-represented in equation 2.18 since this 'price index' is lower the higher the share of intermediate inputs that are produced in close proximity. Alternatively, we could calculate sh_l^{ui} as the share of the number of intermediate input firms rather than the share of output to see if it is the number of varieties in itself that is important rather than the value of the output. As well as benefiting from savings on transport costs, firms may also benefit from a large number of intermediate suppliers due to competition effects driving down the prices. We experiment with this option below. [*to be added*].

It would be computationally difficult to include a separate intermediate input price index for each industry u that supplies inputs to industry i. Instead, our supply access variable is a weighted sum of each industry u price index, as follows:

$$SA_k^i = \left[\sum_u^U a^{ui} \left(\sum_l^K sh_l^{ui} e^{-\tau^u (\sigma^u - 1) d_{kl}} \right) \right]^{\frac{1}{1-\sigma^u}} . \qquad (2.19)$$

The weights, a^{ui}, are given by the share of industry u in the total cost of industry i inputs.

The supplier access variable is only summed across locations within Indonesia whereas firms also purchase inputs from the rest of the world. Our data set provides information on total inputs imported which we include separately as a proportion of total inputs purchased rather than trying to model the supply of inputs from the whole world, which would clearly be unmanageable.

Market Access The Market Access or demand linkage effect arises due to the benefits of being in close proximity to final demand. Firms sell their output to consumers and to downstream firms as can be seen from the second term in equation 2.15.

$$DemandLinkage = \sum_{l=1}^K \left\{ e^{-\tau^d \left(\sigma^d - 1\right) d_{kl}} \left(s^i Y_l + \mu^i n_l^d p_l^d x_l^d \right) \left(P_l^i \right)^{\sigma^i - 1} \right\}^{\frac{1}{\sigma^i}} . \qquad (2.20)$$

Consumers spend a proportion s^i of their income on industry i goods; and downstream firms spend a proportion μ^i of their revenue on industry i output. This is summed across all locations in Indonesia to get total demand. The price index P_l^i is meant to reflect the price of substitute goods in industry i, however this price data is unavailable. Hence, we will construct a separate measure of this price index analogous to the supplier access variable to capture the competition effect of being close to firms within the same industry.

Our market access variable to proxy for demand linkages is

$$MA_k^i = \frac{1}{TD^i} \sum_{l=1}^K \left\{ e^{-\tau^i \left(\sigma^i - 1\right) d_{kl}} \left(s^i Y_l + \sum_d^D \mu^{di} n_l^d p_l^d x_l^d \right) \right\}^{\frac{1}{\sigma^i}} . \qquad (2.21)$$

The inner bracketed term sums demand across all downstream firms and consumers in location l that demand industry i goods. This, scaled by total demand in Indonesia by firms and consumers TD^i, is distance adjusted so that demand from consumers within the same kabupaten receives a higher weighting than demand from locations further away. The size of the distance adjustment will depend on the estimation results. We include actual exports by each industry in location k as a proportion of total output separately instead of trying to model demand from every location in the rest of the world.

Our estimating equation, after taking logs of equation 2.15, becomes

$$\ln w_k^i = \gamma_0 + D_l + D_i + \gamma_1 * \ln SA_k^i\left(\delta_1\right) + \gamma_2 * \ln MA_k^i\left(\delta_2\right) + \gamma_3 * EXP_k^i + \gamma_4 * IMP_k^i + \varepsilon_{ik}. \quad (2.22)$$

The coefficients to be estimated are γ_0, γ_1, γ_2, γ_3, γ_4, δ_1, and δ_2 with

$$\delta_1 = \tau^u\left(\sigma^u - 1\right), \ \delta_2 = \tau^d\left(\sigma^d - 1\right), \quad (2.23)$$

We estimate equation 2.22 using non-linear estimation. This enables us to estimate a distance adjusted supplier and market access rather than imposing the distance effect.[8] By estimating δ_1 and δ_2 we can recover τ^u and τ^i for given values of $\sigma's$.[9] Once we have τ^u, and τ^d we can calculate a measure of transport costs i.e. t_{kl}^i as in equation 2.3, which gives us an indication of the proportion of output that melts in transit

We hypothesize that all the parameters are positive. An increase in the supplier access term increases wages. This is essentially an inverse proxy of the price index in equation 2.17 - the closer a firm is to its input suppliers the lower its total cost and the higher the wages it can afford to pay. An increase in the market access also leads to an increase in 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 the wages it can afford to pay. The larger are δ_1 and δ_2 the more localized are the cost and demand pecuniary externalities and hence the higher the transport costs.

Extensions and modifications to the theory Before going to the data with this theory we need to ask how realistic the assumption of the theory are and whether there are any other important variables omitted from the theory that

[8]Other studies usually divide market access proxies, such as GDP, by distance as originally done in Harris (1954).

[9]If we had price data we would be able to also estimate $\sigma's$ using the functional form in equation 2.15. Instead, we will use the range of values of $\sigma's$ from other studies. See Hummels (1999.)

affect wages. First, consider the zero profit assumption. Although firms may not earn zero profits in practice, the relationship in equation 2.22 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 cannot ignore that there are significant industry variations in wages. The industry differences may be explained by standard labour 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 kabupaten. Third, the theory assumes free mobility between industries within the same location and no labour mobility between locations. Clearly, that is not the case in Indonesia - there is some mobility between locations and there are some frictions in mobility between industries. Provided that there are also some frictions in labour mobility between locations then the relationship in 2.22 should hold. Fourth, other sources of agglomeration such as technological spillovers and labour pooling could give rise to higher wages. We construct variables to capture these effects and include them as additional regressors.

3. Data

Our analysis is conducted at the *kabupaten* (district) level. Indonesia has a five-tiered geographic system – national, provinces, districts, sub-districts (kecamatan) and villages (desa).[10] A map of the nation showing the total value-added for each kabupaten in 1983 is presented in Figure 1. It shows that value-added is concentrated largely around Java's urban centers, with some activity in Sumatra, and to a lesser extent Kalimantan. There is little formal sector manufacturing in the Eastern Islands - consisting of Nusa Tenggara Timur, East Timor, Maluku and Irian Jaya. For this reason we drop these regions from our initial sample. Sulawesi has slightly more in the way of manufacturing and we leave it in because

[10] 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. This was most commonly to separate out the urban centres of economic activity (kotamadya) to become their own districts (for administrative purposes). We avoid problems associated with changing kabupaten borders by using the kabupaten borders from the earliest year (1983). We also merge all kotamadya that existed in 1983 back into their neighbouring kabupaten. This avoids any bias that might arise from economic activity being concentrated in kotamadya which are much smaller in total area than the average kabupaten.

it is a large, important land mass.[11] Our sample consists of 210 kabupaten, 88 of which are on the island of Java. These cover an area of 1,375,369 square kilometers.[12] 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.[13]

Our main data source is the Manufacturing Survey of Large and Medium-sized firms (Survei Industri, SI). This is a census of all manufacturing firms in Indonesia with 20 or more employees. It is an ongoing survey and was initiated in 1975. The SI data can be thought of as capturing the formal manufacturing sector. It does not capture the informal sector. Hence, our estimates pertain only to the impact of potential demand on formal sector wages and employment. Informal sector employment plays an important role in Indonesia (as in most developing countries), however it generates only a small percentage of the nation's total value-added. Our attention on the formal sector is justified by its importance in the development process and role in raising the population's standard of living. The SI collects an unusually rich array of firm level data. We use this data set to construct our measure of kabupaten manufacturing wages by 5-digit industry and the value of output. See Table 1 for the summary statistics.

To construct the dependent variable, w_k^i- industry $i's$ wage in kabupaten k - we divide each firms' annual wage bill by the average number of workers they employ over that 12 month period. We then take the median of this variable within 5-digit industry/kabupaten categories. The output measure that is used to calculate the supplier access and market access variables is the total within 5-digit industry/kabupaten categories of firms' self-reported output in rupiah.

The SI questionnaire asks each firm to list all of their individual intermediate

[11] In addition to having little formal sector manufacturing, NTT, East Timor and Maluku do not constitute large land masses but are clusters of small islands. One would expect spatial linkages to be different from those experiences on the larger land masses. Although Irian Jaya is a large land mass, it is very sparsely populated and much less developed than the other large islands.

[12] There is considerable variation in the size of kabupaten. The average size of a kabupaten in the Outer islands is 10,145 square kilometres compared to 1,465 in Java. This may introduce heteroscedasticity to the error terms in our equations. All reported standard errors are heteroscedasticity-robust. Kabupaten size is more uniform within Java and so presents much less of a problem for the Java only estimates.

[13] For example, at the provincial level. The kabupaten is the smallest geographic unit for which there is gross regional product data. Even if more disaggregated regional income data were available, working at a more disaggregated level would be computationally difficult.

inputs and the amount spent on each in rupiah.[14] The standard SI data files purchased from BPS only provide the total spent on intermediate inputs. We however have access to the unpublished individual input data for 1998. Thus we know for each firm how much they spent on each input. We aggregate this up within 5 digit industry categories. This is far more disaggregated than standard I/O tables - we have disaggregated input data information from 300 manufacturing industries whereas the most disaggregated I/O tables in Indonesia include 90 manufacturing industries . The input codes allow us to link these inputs to the firms that produce them (although not the actual source of the input for each firm) and construct the supplier access variable. Similarly in reverse, we can identify the location of firms that are potential purchasers of an industry's output and so construct the marker access variables. We only have the input data for 1998 and so assume that technology does not change between 1998 and our sample period. Published data at the kabupaten level on raw material production is not readily available.[15] 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). A list of industries that are not included as a dependent variable is included in the Appendix. For example, the industry "threads" is not included in the dependent variable but it may still be an input into another industry so would be included in the calculation of the explanatory variable SA. Dropping raw materials industries reduces the number of industries covered from 263 to 172.

In addition to the SI data, we use data on non-oil gross regional domestic product (GRDP) at the kabupaten level to construct the regional income data needed in the calculation of the final demand component of the market access variable. These data are published by the Indonesian Central Statistical Agency, Badan Pusat Statistik (BPS). BPS (1995a), BPS (1997) and BPS (2000) together cover the period 1983 to 1998. 1983 is the earliest year for which such data are available. 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 product. Non-oil GRDP figures are published from 1993. For years prior to 1993 we predict kabupaten oil revenues from concurrent provincial figures which are available and subtract this from the GRDP (including oil) data. See the appendix for details. Final demand shares from Input-Output tables published in

[14] Those constituting less than 10% of total inputs could be grouped as "other".

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

BPS (1995b) and BPS(1990) are applied to the income to construct final consumer demand at the 5 digit industry level.[16]

The other important data source is a kabupaten-level map of Indonesia. We use ArcView's GIS technology to calculate the geographic center (centroid) of each kabupaten and then construct pairwise measures of "as the crow flies" distance between the centroids.[17] 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 3304 km from Aceh Besar in the north-western tip of Sumatra to Sangihe Talaud in the far north-east of North Sulawesi. Other minor data sources are detailed in the Appendix.

4. Results

Our initial sample covers 210 kabupaten and 172 industries. The total number of observations is 3075. Of these, 2377 are on the island of Java and 698 in the Outer Islands. Table 1 presents summary statistics of the data. One concern that has arisen in previous studies (for example, Redding and Venables, 2000) is that the cost and demand linkage variables are likely to be highly correlated. This is a problem that the detailed location and input data we use has enabled us to overcome. As a result of being able to accurately pinpoint the location of suppliers and also to identify suppliers at the 5-digit level, the correlation between our cost linkage and demand linkage variables is only 0.43. This allows us to separately and precisely estimate the two different - and sometimes competing- vertical linkages.

Equation 2.22 is estimated using non-linear least squares. All standard errors have been corrected for heteroscedasticity using the White method. Table 2 presents the results for the whole of Indonesia, and Java and the Outer Islands separately. In addition to the variables suggested by the equation, dummy variables for the islands of Sumatra, Kalimantan and Sulawesi are included. We also include a Jakarta dummy. Our industry controls are at the two digit level and are relative to the textiles, clothing, footwear and leather industry. We in-

[16] The input-output tables have 161 sectors in 1990 and 172 in 1995. These are more aggregated than the 5-digit ISIC industry categoroies. 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.

[17] The map used shows the 1993 kabupaten boundaries. As explained above we merge the kotamadya into the kabupaten. It was also necessary to merge a small number of other kabupaten whose boundaries had changed between 1983 and 1993. See Appendix 2.

clude more disaggregated industry controls in further specifications below. For the whole of Indonesia, all of the coefficients associated with the cost and demand linkage terms are consistent with the theory and strongly statistically significant. Both the coefficients on distance (δ) and the coefficients on the distance-adjusted demand linkage (γ) are significant. The estimates for Indonesia suggest that a distance-adjusted increase in the supplier access of 10 percent allows a zero profit firm to increase its wages by 0.65%. The analogous estimate for the demand linkage is 0.062, hence a 10 percent increase in market access allows firms to increase their wages by 0.62%. For Indonesia as a whole supplier access and market access appear to be of about equal importance to industries. Both the coefficients in the cost linkage term are significant. The parameters on distance, δ, indicate how quickly the market and supply access spillovers decay with distance. If $\delta = 0$, then any increase in the externality in any kabupaten in Indonesia has the same effect on wages in all locations. If $\delta = \infty$ then an increase in the externality in location l will have no effect on wages in kabupaten k – all effects are completely localized.

Using $\delta = \tau(\sigma - 1)$ and equation 2.3, we can calculate the iceberg transport cost t_{kl} between any two kabupatens, k and l, for different values of σ. As discussed above, our data does not enable us to estimate σ, however by assuming a range for sigma - as identified by estimates in the literature - we can calculate the transport costs. Reliable estimates of σ lie in the range 4 to 7 (see Hummels, 1999). For two kabupatens that are the closest in distance of 6.2 kilometers (and for $\sigma = 5.5$), the iceberg transport cost for upstream goods $t^u = 1.016$ and for downstream goods $t^d = 1.032$, which implies that 1.6% of intermediate goods and 3.1% of final goods would melt in transit. To transport intermediate inputs 100 kilometers t^u lies between 1.22 and 1.49 and t^d lies between 1.46 and 2.14. This suggests that for final goods, somewhere between 32% and 53% of the good "melts" in transit over a distance of 100 kms. In contrast, somewhere between 18% and 33% of intermediate goods melt over the same distance. Hence, these estimates suggest that transport costs are higher for final goods and demand linkages are more localized than cost linkages. For the purposes of comparison, below we will compare estimates of transport costs assuming a σ^u in the middle of this range at 5.5.[18] The point estimates, overall, indicate that the effects of activity are highly

[18] Another way to examine the localisation of the linkages is to calculate their half lives, that is, at what distance from the kabupaten are the effects of the kabupaten's externality on wages halved. This involves finding the D* that satisfies $0.5 = e^{-\delta D^*}$ (Keller, 2002). For example, the demand linkage has a half life of 30.5 kilometres, whereas as determined above, the cost linkage

localized. The equation explains 31% of the variation in log average wages in 1996. The coefficients on the percentage of output exported and the percentage of inputs imported are positively signed confirming that the more internationally focused firms pay higher wages.

The Java-only results have the same signs, with the linkage effects much stronger - a 10% increase in supplier access increases wages by 1.1% and a 10% increase in the market access increases wages by 1.5%; and the demand linkage is more localized. The R-squared is also higher for Java with 36% of variation being explained. The results for the Outer Islands are in strong contrast. Both the demand and cost linkage for the outer islands are insignificant. The Outer Islands are much more sparsely populated and much less industrialized. Much of the industry in this region involves the processing of natural products like wood and rubber. The results are consistent with linkages between firms only being important once a critical mass of industry has been reached. In 1996 there were only 4339 formal sector firms (or 0.003 firms per square kilometers) in the Outer regions compared to 18506 (0.145 per square kilometer) in Java. Further, Table 2 presents results for Java when we exclude linkages to the other islands. The results show that linkages to the outer islands are of no importance to firms on Java - the coefficients are almost identical. These results are consistent with the difficulty the Indonesian government has experienced in trying to move industry to the outer islands. Not only is the very small number of firms in the Outer Islands a problem, the outer regions are so far from Java so as to not benefit from the existence of the Javanese markets and suppliers.[19] Below we restrict our attention to more closely characterizing the linkages on Java (excluding linkages to the Outer islands).

Equation 2.22 is derived from economic geography theory and does not control for other variables that vary across industries and are well-known to affect wages. For example, differences in human capital requirements across industries. Larger firms and foreign firms are also well-known to pay higher wages. The 2-digit industry dummies will capture these differences across industries at the two-digit level but differences may persist within these categories. We deal with this in two ways. First, Column 2 in Table 3 presents results with 3 digit dummies. The coefficients only change slightly. Column 3 then adds further control variables which we calculate from the SI data. Specifically, we include variables that reflect the av-

---

is less localised the cost linkage half life is 58.3 kms.

[19] The insignificance of the linkage variables persists with the inclusion of further controls.

erage percentage of workers that are tertiary educated, and high school educated and the percentage of workers who are female within each 5-digit industry and kabupaten. Female workers are normally paid less owing to discrimination and/or occupational segregation. In addition we control for the education attainment of the population within each kabupaten. The variable *skill* is calculated from the 1995 Intercensal Survey (SUPAS) which is a household survey and represents the percentage of a kabupaten's population that has at least a high school education. Adding these controls increases the adjusted R^2 from 0.38 (with the 3 digit dummies) to 0.52. All of the additional controls are strongly statistically significant and are signed as expected. For example, a 1 percentage point increase in the percentage of workers who are female decreases average wages by half a percent. The coefficients on the demand and cost linkages remain statistically significant and are now slightly smaller. They are now similar in magnitude (0.10 and 0.09 respectively). The estimates of the deltas are also slightly smaller.

4.1. Alternative Forces of Agglomeration

4.1.1. Labour Pooling

In Column 4 of Table 3 we add variables that attempt to capture other forces of agglomeration. As mentioned in the introduction, three main forces of agglomeration were identified by Marshall (1920). The first being the distance to suppliers and customers which is the focus of this study. The second was labour market pooling and the third information spillovers. To examine labour pooling we follow Dumais, Ellison and Glaeser (1997) and construct an index that captures the similarity of industry i in kabupaten k's labour requirements to the requirements of other firms in the kabupaten. The index is calculated as:

$$LP_k^i = -\sum_s (L^{is} - \sum_{j \neq i} + \frac{E_k^j}{E_k - E_k^i} L^{js})^2, \tag{4.1}$$

where $L^{is}$is the fraction of industry i's labour force that has education level s, E_k^i is the number of workers in industry i in kabupaten k, and E_k is the total number of workers in kabupaten k. The index thus compares the educational composition used by industry i with the education composition used by the other firms in the kabupaten. 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.[20] We hypothesize that firms will benefit from the presence of other firms that use a similar mix of skills and as a result will be more profitable. The results in column 4 are consistent with this hypothesis. The labour pooling index is strongly significant and positive.

4.1.2. Technological and Knowledge Spillovers

The regression results in column 4 also include a variable that is constructed to capture technological or knowledge spillovers amongst like firms. We calculate the number of firms in the same 5 digit industry in every kabupaten and then distance weight this variable in the same way as we do the linkage variables. It is not clear from which industries technology spillovers flow without a technology flow table. So we calculated this spillover variable at the 5,4,3 and 2 digit levels but the 5 digit level gave the lowest residual sum of squares. [*to be added*]. [21] In addition to capturing spillovers (which would allow firms to pay higher wages), or instead of capturing spillovers, this variable could capture the "competition effect" ie it could be seen as an inverse proxy of the price index of substitute goods in equation 2.15 hence putting downward pressure on firms' profits and hence their ability to pay high wages. Thus, a priori the direction of this variable's impact is ambiguous. The results show this variable to be statistically insignificant. It may be that the two effects offset one another or that neither force affects wages. We are unable to distinguish between these two scenarios.

4.1.3. Number of firms vs output

[*to be added*].

4.1.4. Simultaneity Issues

One 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 our

[20] We calculated this measure at the provincial and kabupaten level. The kabupaten level variable gave a better fit. Four observations were dropped because there was only one industry in the kabupaten.

[21] We also constructed an analogous variable using same industry output instead of the number of firms in each kabupaten. The number of firms was a better fit.

linkage variables. For example, it may be that firms are attracted to kabupaten 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 (19??) average temperature, humidity and wind speed are typically used. These variables don't adequately capture differences in exogenous amenity in Indonesia or Java which are almost invariably hot and humid. Instead, to capture exogenous amenity we have included a dummy variable for whether the kabupaten is on the coast and another measuring the percentage of the kabupaten's area that is swamp land.

Another way to control for the possible endogeneity is to include the total number of manufacturing firms in each kabupaten as an explanatory variable.[22] This variable reflects the attractiveness of a kabupaten to firms (including pre-existing infrastructure). The drawback of using this variable is that it is likely to be endogenous. To reduce the extent of the correlation between this variable and the error term we lag the number of firms 10 years. This takes us back to the very start of the rapid industrialization in Java. The number of formal sector firms almost doubled in Java between 1986 and 1996 (10159 in 1986 compared to 18506 in 1996).

Further, we include provincial level dummy variables. Java's 88 kabupaten are divided into 5 provinces -West Java, Jakarta, Central Java, Yogyakarta and East Java. The fixed provincial effects capture non-time varying characteristics of these regions. Thus they capture pre-existing (prior to 1991) infrastructure, and any other non-time variant characteristics that affect the attractiveness of geographic location to firms. The provincial dummies control for geographic location which is what we are trying to explore through the linkage terms, thus greatly reduce the information that is available for the identification of the linkage terms. We nevertheless present these results in column (6) of Table 3 for completeness. The linkage terms remain significant even when the information that they are identified off has been severely restricted. The distance parameter on the demand linkage parameter estimates are largely unchanged but the estimate on the distance term for the cost linkage rises significantly.[23]

[22] Note that this is the total number of firms with more than 20 employees (as collected in the Survei Industri).

[23] We also experimented with stacking the data for 1991 and 1996 and including kabupaten effects. This limits identification of the linkage terms to variation withing 3 digit industry

4.2. Sensitivity Tests

We further explore the possibility of simultaneity bias by conducting some sensitivity tests. First, following the approach of Hanson (2000) and Keller (2002) we re-estimate the equation with the full set of controls but dropping kabupatens that individually constitute more than 2% of Indonesia's GDP. This drops the main industrial centers of Jakarta, Surabaya and Bandung. Conditions in these large centers of economic activity are the most likely to affect conditions in neighboring kabupaten. Hence, we would expect the linkage terms for these centers to be the most problematic in terms of reverse causality. That is, the share of output in own and neighboring kabupaten is likely to be more highly correlated with wages in these centers than wages elsewhere. Hence the sensitivity of the results to dropping these observations gives us an indication of the extent of simultaneity bias in our results. The results are presented in Column (2) of Table 5. A comparison with Column (1) shows us that the estimates of the parameters in the linkage terms and the coefficients on these terms are largely unaffected. The point estimates are very close to the original results and the linkage terms remain significant.

In a similar vein, Column (3) presents the results when we drop the own kabupaten component of the market and supplier access variables. If we are concerned that the linkage terms are a function of wages (reverse causality) then this is more likely to be the case for own kabupaten effects. That is, kabupaten k's share of supplier output is more likely to be a function of kabupaten k's wages than the share of supplier output in another kabupaten. Similarly for the market access variable. The coefficients on the linkage terms are largely unaffected by this change. The distance parameter in the market access variable falls from around 2 to 1.3 and the distance parameter in the supplier access variable rises from around 1 to 1.37. Both of these estimate however still lie in the 5% confidence interval of the original estimates.

Column (4) presents results where we drop observations on industries that receive more than 50% of their inputs from within their own 5 digit industry.

Column (5) presents results when we lag both the linkage variables and the spillover variable 5 years. This reduces the possible correlation between the error term and these variables. However, to the extent that these variables are correlated

categories and within kabupaten. The point estimates and standard errors were almost identical to those obtained when provincial dummies were used. The linkage terms were signed correctly and remained statistically significant.

over time some simultaneity bias will persist. The point estimates are largely unaffected by the lagging and remain significant (although the distance parameter in the market access term at the 10% level).

Finally, Column (8) presents instrumental variable results where we instrument for GDP by kabupaten using the % of arable land in each kabupaten. [*to be added*]..

Having established that the results are robust to concerns about simultaneity bias we now examine changes over time.

4.3. Changes Over Time

Table 4 presents results for 1983 and 1991. 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 results are remarkably stable across time. The point estimates on the distance parameter within the market access variable are the most affected, being larger in the earlier years. This suggest that being close to one's market has become less important over time - likely due to improving transport infrastructure. However, all of the coefficients for 1991 and 1983 lie within 2 standard deviations of the 1996 results, suggesting that the differences are unlikely to be statistically significant. This is a significant result in two senses. First in terms of the robustness of our results. The variables for 1991 and 1983 were constructed from a completely separate set of data. Indonesia has experienced dramatic change over the past twenty years. Manufacturing has boomed since the lifting of trade restrictions in the mid-1980's and there has been associated dramatic improvements in infrastructure - both roads and telecommunications. Thus, the result is significant in substantive terms because given the extent of change over this period, it would not have been surprising if linkages between firms had changed over this period. In particular one might have expected the extent of localization to have diminished with better communication and lower transport costs (in terms of time and money).

5. Conclusions

This study shows that demand and cost linkages have a significant impact on spatial manufacturing wages in Indonesia, and both are of similiar magnitude. These results were robust after controlling for alternative sources of agglomeration such as labor pooling; more standard explanations of wage variation such as skill levels and firm size; and infrastructure variables. The results were robust to all of these additional controls. We found that the benefits of linkages are quite localized, which one would expect to be likely in a developing country that does not have good internal transport infrastructure. Firms do benefit from vertical linkages but not if the suppliers are too far away. The spread of the linkages over space is remarkably stable across time. This is particularly noteworthy given the dramatic changes experienced in Indonesia during the period of study. Our results for the outer islands, showing that linkages are insignificant, underscore the difficulty governments have in igniting economic growth in far flung regions - where the citizens are often the poorest and benefiting the least from economic growth. Linkages between vibrant Java and the quieter Outer Islands were found to be non-existent. The geography of Indonesia renders internal trade costs far too high for these firms to receive the benefits of agglomeration.

6. Appendix

1. ISIC Code lists
    2. Details on netting out oil GRDP
    3. Apportioning final demand shares (? - maybe not necessary)
    4. List of raw materials industries dropped.

References

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

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

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

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

[5] Combes, Pierre-Philippe and Miren Lafourcade (2001). Transport Costs Decline and Regional Inequalities: Evidence from France.

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

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

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

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

[10] Henderson, J.Vernon (1999). The Effects of Urban Concentration on Economic Growth. NBER 7503.

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

[12] Hummels, David (1999). *Towards a Geography of Trade Costs.* mimeo.

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

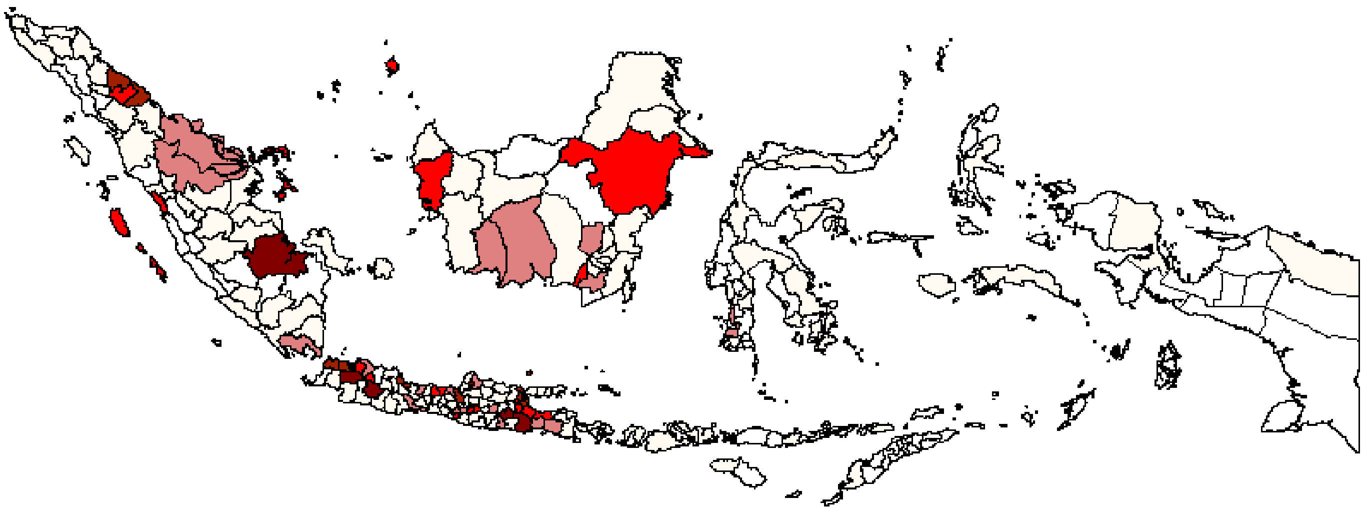
[14] Krugman, Paul (1991). Increasing Returns to Scale and Economic Geography. *Journal of Political Economy* 99(31), 483-499.

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

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

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

Figure 1: Geographic Distribution of Value-Added



Ratio of Kabupaten Value-Added to Mean Value-Added

0 -0.001	(lightest)
0.001-0.5	
0.5 -1	
1 -3	
3-8	
8-30	(darkest)

Figure 2: Vertical links

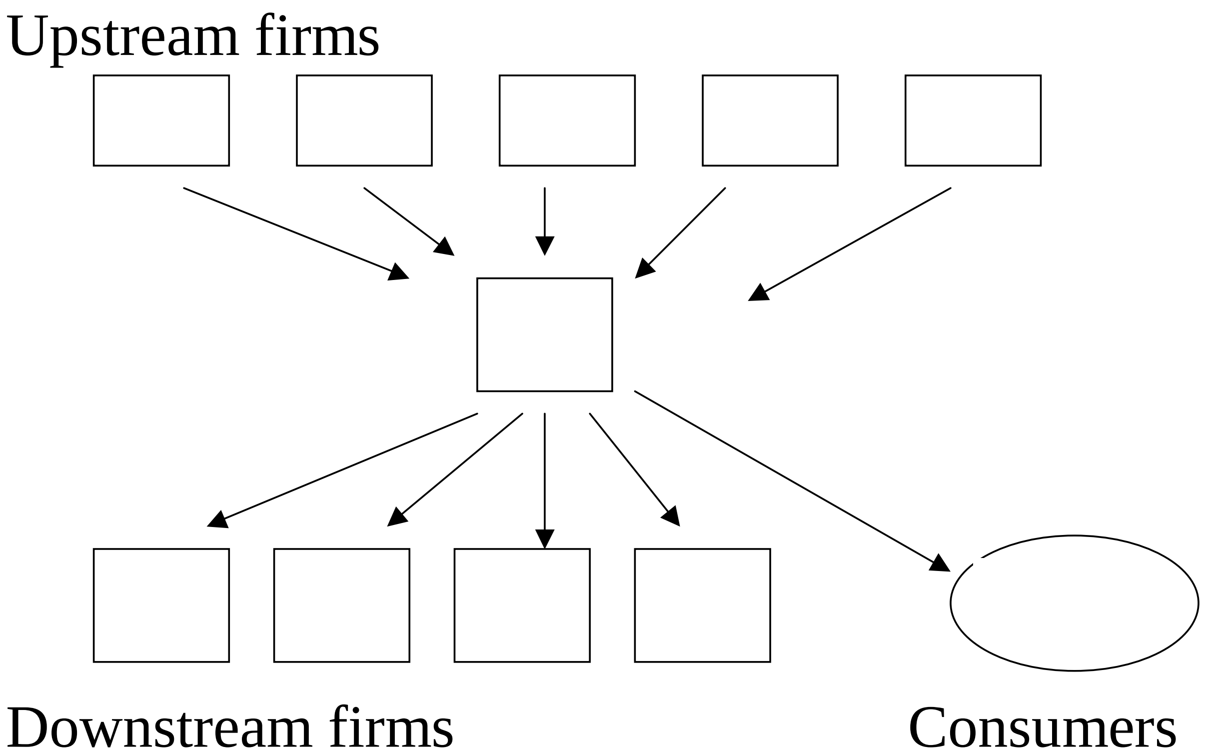


Table 1: Summary Statistics

	Indonesia				Java				Outer Islands			
	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
ln(wage)	8.79	0.63	5.71	10.75	8.80	0.63	5.71	10.75	8.76	0.65	6.20	10.72
wage	8083.63	6081.43	300.66	46598.53	8145.44	6182.20	300.66	46598.53	7873.13	5724.22	490.36	45261.98
market access	0.00	0.02	0	0.83	0.00	0.02	0	0.83	0.00	0.02	0	0.46
supplier access	0.00	0.03	0	1	0.00	0.03	0	1	0.00	0.03	0	0.71
exports	0.19	0.32	0	1	0.17	0.30	0	1	0.26	0.38	0	1
imports	0.15	0.28	0	1	0.17	0.28	0	1	0.09	0.24	0	1
foreign ownership	0.07	0.19	0	1	0.07	0.18	0	1	0.06	0.19	0	1
govt ownership	0.03	0.15	0	1	0.03	0.14	0	1	0.04	0.18	0	1
average size	214.34	544.30	14	15651	217.14	564.56	17	15651.00	204.78	469.05	14	5184
female participation	0.32	0.28	0	1	0.32	0.27	0	1	0.33	0.29	0	1
high school education	0.38	0.24	0	1	0.37	0.24	0	1	0.42	0.25	0	1
tertiary education	0.04	0.06	0	0.78	0.04	0.07	0	0.78	0.03	0.05	0	0.60
population skill level	0.23	0.10	0.07	0.62	0.22	0.13	0.07	0.62	0.24	0.08	0.11	0.56
labour pooling	-0.06	0.81	-0.67	0.00	-0.07	0.08	-0.62	0.00	-0.06	0.08	-0.67	0.00
spillovers	0.38	4.35	0	393	0.36	4.30	0.00	393.00	0.48	4.55	0.00	348.00
coast	0.69	0.47	0	1.00	0.60	0.49	0.00	1.00	0.77	0.43	0.00	1.00
swamp	0.03	0.07	0	0.60	0.01	0.03	0	0.14187	0.05	0.08	0	0.60
firms86	60.39	121.74	0	1143	115.44	165	2	1143	20.67	47.20	0	450
N	3071				2733				694			
kabupatens	210				88				122			

TABLE 2: BASIC SPECIFICATION

	INDONESIA	JAVA	OUTER ISLANDS	JAVA
Supply Access (gamma1)	0.0654	0.111	0.017	0.109
	(0.0121)	(0.0181)	(0.0143)	(0.0179)
distance (delta1)	1.1916	1.050	1.625	1.090
	(0.2827)	(0.2007)	(1.415)	(0.2105)
Market Access (gamma2):	0.0618	0.149	0.003	0.150
	(0.0166)	(0.0209)	(0.0191)	(0.021)
distance, kms/100 (delata2)	2.2909	3.218	3.818	3.117
	(0.9416)	(0.5561)	(24.7126)	(0.539)
Exports	0.2100	0.087	0.424	0.087
	(0.0331)	(0.0375)	(0.0567)	(0.0374)
Imports	0.5599	0.491	0.408	0.491
	(0.0467)	(0.0499)	(0.1056)	(0.0498)
Region Dummies:				
Sumatra	0.3713		0.138	
	(0.0416)		(0.0598)	
Kalimantan	0.6401		0.345	
	(0.0675)		(0.077)	
Sulawesi	0.2584		-0.139	
	(0.0853)		(0.0915)	
Jakarta	0.2104	0.021		0.022
	(0.0357)	(0.0353)		(0.0353)
Industry Dummies:				
Wood/Furniture	0.1266	0.197	0.241	0.197
	(0.0303)	(0.034)	(0.07)	(0.0339)
Paper/Printing	0.3171	0.270	0.525	0.270
	(0.0439)	(0.0486)	(0.1007)	(0.0486)
Chemicals/Plastics	0.2889	0.288	0.449	0.288
	(0.0322)	(0.0346)	(0.084)	(0.0346)
Non-metallic Minerals	0.1812	0.188	0.310	0.188
	(0.0430)	(0.0459)	(0.1064)	(0.0459)
Metals	0.5467	0.463	0.634	0.463
	(0.0632)	(0.0671)	(0.1613)	(0.0671)
Machinery and Components	0.3831	0.316	0.684	0.316
	(0.0280)	(0.0288)	(0.0744)	(0.0288)
Other	-0.0284	-0.056	0.289	-0.055
	(0.0485)	(0.0512)	(0.1005)	(0.0512)
constant	8.7429	9.195	8.262	9.192
	(0.0514)	(0.0526)	(0.1289)	(0.0525)
Linkage Variables Coverage:	Indonesia	Indonesia	Indonesia	Java
RSS	847.7	598.774	194.5	598.6
R-squared	0.31	0.360	0.34	0.36
N	3071	2377.000	694	2377

Table 3: Estimates for Java

	Basic	+ 3-digit controls	+human capital controls	+spillovers +labour pooling	+exog. Amenity + firms	+provincial dummies
	(1)	(2)	(3)	(4)	(5)	(6)
Supply Access (gamma1)	0.1085	0.1275	0.0929	0.0888	0.0832	0.0184
	(0.0179)	(0.019)	(0.0165)	(0.0178)	(0.0186)	(0.0064)
distance (delta1)	1.0904	0.87	0.9761	1.0212	0.9676	13.1828
	(0.2105)	(0.1594)	(0.2045)	(0.2445)	(0.2583)	(4.811)
Market Access (gamma2):	0.1503	0.1356	0.1006	0.099	0.1012	0.0407
	(0.021)	(0.0183)	(0.0181)	(0.0199)	(0.0207)	(0.0147)
distance (delta2)	3.12	3.92	2.8066	2.3426	2.0889	4.5509
	(0.539)	(0.6916)	(0.7144)	(0.6862)	(0.6481)	(2.2985)
Exports	0.087	0.0834	0.0304	0.0268	0.0259	0.0134
	(0.0374)	(0.0374)	(0.0385)	(0.0385)	(0.0385)	(0.0373)
Imports	0.4914	0.4793	0.2412	0.238	0.2356	0.2118
	(0.0498)	(0.0478)	(0.047)	(0.0469)	(0.0469)	(0.0459)
Human Capital Controls: firm size			0.0001	0.0001	0.0078	0.0079
			(0)	(0)	(0.0033)	(0.0029)
foreign ownership			0.4723	0.4626	0.4769	0.4295
			(0.0634)	(0.0637)	(0.0645)	(0.0627)
government ownership			0.3218	0.3318	0.3268	0.3302
			(0.0938)	(0.0903)	(0.0902)	(0.0865)
female participation			-0.5075	-0.499	-0.5018	-0.475
			(0.0505)	(0.0504)	(0.0501)	(0.0493)
high school educated			0.3494	0.3621	0.3697	0.3719
			(0.0505)	(0.0505)	(0.0505)	(0.0504)
tertiary educated			1.6232	1.6254	1.6091	1.6526
			(0.2068)	(0.2031)	(0.2021)	(0.2013)
kabupaten skill level			0.2801	0.2439	0.149	0.2609
			(0.1021)	(0.1023)	(0.1141)	(0.1128)
Labour Pooling				0.5367	0.5318	0.4166
				(0.1538)	(0.1538)	(0.151)
Spillovers: (gamma3)				-0.0025	-0.0009	0.0004
				(0.0096)	(0.0097)	(0.0097)
distance (delta3)				0.0974	0.0638	-50.2143
				(4.8838)	(13.321)	(1168.65)
# firms in 1986					0.0116	-0.0011
					(0.0048)	(0.0052)
coast					0.019	0.0032
					(0.0206)	(0.0208)
swamp					0.3207	0.3009
					(0.2913)	(0.2811)
industry	2 digit	3 digit	3 digit	3 digit	3 digit	3 digit
region dummies	Jakarta	Jakarta	Jakarta	Jakarta	Jakarta	provincial
RSS	598.6	567.8	447.6	445.1	443.2	424.89106
Adjusted-R-squared	0.356	0.384	0.515	0.516	0.517	
N	2377	2377	2377	2377	2377	2377

Table 4: Comparisons Across Years

	1996	1991	1983
Supply Access (gamma1)	0.1170	0.1111	0.1044
	(0.0245)	(0.0297)	(0.0375)
distance (delta1)	0.7008	0.6998	0.6912
	(0.2029)	(0.2532)	(0.4026)
Market Access (gamma2):	0.1270	0.1635	0.0985
	(0.0187)	(0.0245)	(0.03)
distance (delta2)	3.4116	5.3162	5.5413
	(0.6887)	(1.1071)	(2.1229)
Exports	-0.0255	-0.0532	
	(0.0383)	(0.0473)	
Imports	0.3628	0.2366	0.1377
	(0.0492)	(0.0539)	(0.0582)
Indy/Kab Controls: firm size	0.0123	0.0086	0.0121
	(0.0049)	(0.0029)	(0.0056)
foreign ownership	0.5711	1.0301	1.4384
	(0.0723)	(0.1185)	(0.1324)
government ownership	0.4845	0.6870	0.5766
	(0.1043)	(0.0914)	(0.0827)
Spillovers: (gamma3)	-0.0033	-0.0298	-0.0249
	(0.0104)	(0.0134)	(0.0178)
distance (delta3)	-0.1715	6.8145	15.6681
	(3.5166)	(6.7651)	(52.7422)
# firms lagged 10 years	0.0079	0.0010	0.0001
	(0.0031)	(0.0089)	(0.0001)
coast	0.0250	0.0554	0.0187
	(0.0226)	(0.0305)	(0.0422)
swamp	0.4066	-0.7962	-0.0953
	(0.3172)	(0.4941)	(0.5018)
1991 year dummy			
industry	3 digit	3 digit	3 digit
region dummies	Jakarta	Jakarta	Jakarta
RSS	530.51096	492.87426	298.57204
R-squared	0.43297312	0.4439798	0.45856285
N	2377	1834	1067

Table 5: Sensitivity Tests

	Comparison	Small GDP Kabupaten	Drop own Kabupaten	Lagging 5 years	Dropping 50% Own Input Use	Instrumental Variables
	(1)	(2)	(3)	(4)	(5)	(6)
1996						
Supply Access (gamma1)	0.0832	0.0734	0.0684	0.0822		
	(0.0186)	(0.0199)	(0.015)	(0.0173)		
distance (delta1)	0.9676	0.9028	1.3786	1.2286		
	(0.2583)	(0.2802)	(0.4231)	(0.3045)		
Market Access (gamma2):	0.1012	0.1039	0.1012	0.0602		
	(0.0207)	(0.0236)	(0.0251)	(0.0206)		
distance (delta2)	2.0889	1.9559	1.3326	2.176		
	(0.6481)	(0.6832)	(0.5385)	(1.2496)		
RSS	443.2	345.7	446.2	436.3		
R-squared	0.526	0.524	0.523	0.523		
N	2377	1828	2375	2335		