Shakedown: Economic Geography Meets the Kobe Earthquake

Chad Steinberg*

International Monetary Fund

Abstract

This paper empirically investigates the claims of economic geography. This theory states that access to markets and access to suppliers creates benefits to firms: location and, therefore, local markets are a source of relative advantage. Yet while an abundance of evidence confirms the existence of a strong correlation between location parameters and production patterns, these empirical results often fail to account for the fundamental endogeneity of the explanatory regressors. In this paper, I use the exogenous variation generated by an earthquake in Kobe, Japan to solve this problem. Using a panel dataset of industrial production by regional center, I show that, yes, economic geography is a source of relative advantage, but only in markets and across regions where trade costs are relatively high. Specifically, because the relative costs of trade in intermediates is larger than trade in final goods, firms benefit from proximity to one another, but do not gain from relative proximity to consumers; and because trade costs are relatively larger when goods travel across borders than when goods travel within borders, economic geography can explain international, but not intranational patterns of production.

I. Introduction

Empirical evidence indicates that economic interactions between agents greatly decreases with distance [4]. Spatial economic geography[1] takes this one step further, and suggests that distance to potential markets and suppliers of inputs can help determine a region's relative advantage, in addition to differences in technology and factor endowments. This new geography framework–which is essentially a model of trade with imperfect competition combined with iceberg transport costs–can explain, amongst other things, wage differences across regions, the formation of an industrial North, and the failure of industrialization in developing countries.

Yet, interest in economic geography derives from the fact that the explanatory variables, the location of demand (access to markets) and the location of input supply (supplier access), are not exogenous. This generates the possibility of agglomeration and multiple equilibria. This, however, also creates a difficult task for econometricians because there is no clear mapping from the exogenous variables [14]. For example, take the case of transactions between firms: spatial geography predicts that a firm's success will depend on how close the firm is to markets to sell its output, and how close the firm is to suppliers to receive inputs. Thus, while output in an upstream firm may be affected by the abundance of local suppliers (supplier access), the abundance of local suppliers is equally affected by the presence of this upstream firm (market access). Causation runs in both directions, from dependent to independent variable, and back.

In this paper, I attempt to disentangle these forces using the exogenous variation created by an earthquake in Japan. The earthquake occurred on January 17, 1995 and directly affected two vital industrial districts, Kobe and Akashi. Damage was significant: approximately $100 billion in physical damage, 6,398 individuals killed, and 40,000 individuals injured[2]. Because the earthquake was exogenous and inflicted enough damage to adequately predict changes in measures of both market and supplier access, it is a potentially

good instrument.

I first estimate a reduced form equation of the spatial geography model using industry level data from 1993 and 1995 for 253 Japanese industrial districts. The estimated equation uses output indexed by industry and region as the dependent variable, with measures of access to suppliers of inputs, access to markets to sell intermediate inputs, and access to markets to sell final goods as the explanatory variables. The initial results are consistent with past research (e.g., Harris [8]): the relationship between location and output is robust. There exists a persistent correlation between the location of industry and the location of both intermediate and final demand; there is also a strong correlation between the location of suppliers and industry production.

However, after instrumenting for both market and supplier access using earthquake damage statistics, in addition to accounting for unobserved heterogeneity, the results are significantly weakened. The correlation between all variables is significantly reduced. In fact, the relative importance of access to markets for final goods disappears. That is, firms do not benefit from being close to consumers. On the other hand, access to local suppliers and markets to sell intermediate inputs remains important, although the variables exhibit a much weaker correlation. Combined with the previous result, this suggest that while firms benefit from proximity to one another, they do not gain from relative proximity to consumers. Quantitatively, the results show that a 10 percent increase in the index for intermediate market access is associated with a 9 percent increase in output. Alternatively, a single standard deviation increase in the index is associated with an approximate doubling of output.

After implementing the first difference equation there still remains two dimensions of variation in output: variation in output across regions in Japan and variation in output across industries in Japan. Surprisingly, neither the location of intermediate demand nor the location of suppliers can predict the distribution of output across regions within Japan. In contrast, both of the explanatory variables remain robust predictors of the variation in

output across industries.

This shows that access to markets and access to suppliers is not a source of relative advantage within borders. Thus, the distribution of output in Japan must be determined by other factors, and a producer of cd players is not necessarily better off locating in Kansai, near many producers of inputs for the electronic sector, as opposed to locating in Aomori, a city in the less-populated northern tip of Honshu.

The results also suggest, however, that access to markets and access to suppliers is a robust source of relative advantage across industries in Japan and therefore across borders. The difference between these two results is likely due to the difference between transaction costs within borders and transaction costs across borders; and because transaction costs greatly increase when goods cross an international border, geography is relevant for explaining international, but not intranational, patterns of production.

This is not the first empirical paper investigating the various hypotheses suggested by the economic geography models. In fact, indirect empirical work is not lacking[3]. As mentioned at the onset, the gravity models have consistently confirmed the importance of distance. Likewise, Sachs and his co-authors have documented the importance of physical geography measures like location in the tropics and landlockness [6]. Yet the empirical work directly testing these specific models is just beginning to emerge. This paper, perhaps, is closest to three recent empirical papers that implement structural geography models. Hanson [7] does this with counties in the U.S., Redding and Venables [15] do this for an international data set, and Combes and Lafourcade [2] implement a model using French employment areas. Each paper concludes economic geography is an important source of relative advantage. However, what my results suggest is that these results may be overstating the case, and geography is less relevant in markets and across regions with low transport costs.

The remainder of this paper is organized as follows. Section two describes the base case geography model. Section three describes the data. Section four presents the estimation technique. Section five presents the initial results. Section six deals with the endogene-

ity problem. Section seven disaggregates the results. Section eight investigates several robustness checks.

II. A Simple Geography Model

To illustrate the relationship between industrial output and a firm's proximity to markets and suppliers, this section presents a generic geography model. The building blocks include monopolistic competition, a two-level utility function with Dixit-Stiglitz preferences in the subutility function, a two-level production function also with Dixit-Stiglitz preferences in the subproduction function, and iceberg transport costs.

The model contains an unspecified number of industries and regions indexed by i and j, respectively. An apostrophe is utilized to distinguish multiple regions or industries, and the region or industry in the latter indexed position is the region or industry that purchases the good, as opposed to the origin of production. Thus, $x_{ijj'}$ is the quantity of an industry i good produced in region j and sold to region j', and $x_{ij'i'j}$ is the quantity of an industry i good produced in region j' and sold to industry i' in region j.

I begin with consumer preferences and the derivation of demand for finished products. Preferences in region j are given by the two-level utility function,

$$U_j = U\left(X_{1j}, ..., X_{Ij}\right) \quad s.t. \quad X_{ij} = \left(\sum_{j'} n_{ij'} x_{ij'j}^{\frac{\sigma_i - 1}{\sigma_i}}\right)^{\frac{\sigma_i}{\sigma_i - 1}},$$

where X_{ij} is the subutility derived from the consumption of product i in industry j and $U(\cdot)$ is the upper tier utility function that translates all sector subutility levels into an overall welfare level. In this representation, X_{ij} takes the form of a symmetrical constant elasticity of substitution function over all varieties within industry i, where σ_i is the elasticity of substitution between varieties in industry i, and $n_{ij'}$ is the number of varieties of industry i produced in region j'[4].

This particular form of preferences yields an easy representation of final demand for any

industry i good produced in region j and sold to region j'.

$$(1) \qquad x_{ijj'} = p_{ijj'}^{-\sigma_i} P_{ij'}^{\sigma_i - 1} E_{ij'(fin)} \quad s.t. \quad P_{ij'} = \left(\sum_j n_{ij} p_{ijj'}^{1-\sigma_i} \right)^{\frac{1}{1-\sigma_i}}$$

Where $p_{ijj'}$ is the price of any industry i good produced in region j and sold to region j', $P_{ij'}$ is the price index for industry i goods in region j' (in other words, the price of composite good $X_{ij'}$), and $E_{ij'(fin)}$ is the expenditure in region j' on finished goods in industry i. Thus, demand is decreasing in the individual price, and increasing in the overall price index (because the relative price of p decreases as P increases) and in expenditures on goods in industry i. Furthermore, if we assume the upper tier utility function exhibits Cobb-Douglas preferences, we can rewrite $E_{ij'(fin)}$ as $\alpha_i Y_{j'}$, where $Y_{j'}$ is income in industry j' and α_i is the share of income spent on industry i goods (or the exponent in the Cobb-Douglas function).

The derivation of intermediate demand is equally straightforward. Starting again with a two-level function, the production technology for industry i in region j is given by,

$$x_{ij} = x_j \left(V_{ij}, X_{1'ij}, ..., X_{I'ij} \right) \quad s.t. \quad X_{i'ij} = \left(\sum_{j'} n_{(ij)'} x_{(ij)'ij}^{\frac{\sigma_{i'}-1}{\sigma_{i'}}} \right)^{\frac{\sigma_{i'}}{\sigma_{i'}-1}},$$

where x_{ij} is output of a good in industry i in a region j, V_{ij} is a vector of factors used in production, and $X_{i'ij}$ is the subproduction function derived from the use of industry i' inputs in the production of goods in industry i, region j. This technology is different across sectors, but does not vary across regions.

This particular production function also yields a familiar demand function for industry i intermediate goods produced in region j and used in the production of output in industry i' in region j'.

$$(2) \qquad x_{ij(ij)'} = p_{ijj'}^{-\sigma_i} P_{ij'}^{\sigma_i - 1} E_{i(ij)'(int)} \quad s.t. \quad P_{ij'} = \left(\sum_j n_{ij} p_{ijj'}^{1-\sigma_i} \right)^{\frac{1}{1-\sigma_i}},$$

where $x_{ij(ij)'}$ is the amount of an industry i good produced in region j and sold to industry i' in region j', and $E_{i(ij)'(int)}$ is the expenditure in region j', industry i', on intermediate inputs from industry i. If we assume the top-tier production function is also Cobb-Douglas we can rewrite $E_{i(ij)'(int)}$ as $a_{ii'}y_{(ij)'}$, where $y_{(ij)'}$ is nominal output and $a_{ii'}$ is the input-coefficient for intermediate inputs from industry i used in the production of industry i'.

Demand equations 1 and 2 are the building blocks for a more general expression of demand and eventually a statement about the relationship between geography and demand. First, notice that the only difference between equations 1 and 2 is the expression for expenditures. This allows us to easily sum first within a region (sum across industries to form intermediate demand and then add to the term for final demand), and then across all regions to obtain an expression for total demand for output from any industry i good produced in region j.

$$x_{ij} = \sum_{j'} \left(x_{ijj'} + \sum_{i'} x_{ij(ij)'} \right) = \sum_{j'} p_{ijj'}^{-\sigma_i} P_{ij'}^{\sigma_i - 1} E_{ij'}, \tag{3}$$

where $E_{ij'}$ is equal to $E_{ij'(fin)} + \sum_{i'} E_{i(ij)'(int)}$ which is simply $\alpha_i Y_{j'} + \sum_{i'} a_{ii'} y_{(ij)'}$ if we assume the upper-tier functions are Cobb-Douglas.

So far this expression says nothing about geography. Essentially, we have a demand function that is identical to that used in new trade models (e.g., Helpman and Krugman [9]) summed over a number of regions. What transforms this into a geography model, however, is the particular definition we choose for $p_{ijj'}$, the price of any industry i good produced in region j and sold in region j'. One can think of two components that determine a good's price in different regions. First, one has to account for transport costs between regions. The market price of a Japanese good sold in Germany is likely to be higher relative to the market price in Japan due to transaction costs incurred during transportation. Thus, location close to markets will reduce transport costs, and subsequently the market price, $p_{ijj'}$. Second, the technology employed and the price of factors and inputs in one's own region help determine

the cost of production. Because there are costs to transportation, the closer you are to markets for inputs, the more likely these inputs will be cheaper. That is, favorable supplier access can lower your costs of production and subsequently also lower the market price, $p_{ijj'}$.

I introduce transport costs using the following ubiquitous formulation:

$$p_{ijj'} = p_{ij} * t_{ijj'}, \tag{4}$$

where $t_{ijj'}$ is the iceberg cost factor on trading industry i good from region j to region j'. If we live in a "borderless world" then $t = 1$ between all regions and across all industries. Next, we can define the base price of industry i good produced in region j, p_{ij}, or the price of the good in its home market (because we assume $t_{ijj} = 1$). As mentioned above, the price is determined by both the firm's technology, the price of factors and intermediate inputs, and finally, the perceived demand elasticity with respect to own price, or in this case the industry elasticity of substitution between varieties.

$$p_{ij} = p_i\left(\mathbf{w}, \mathbf{P}_j, \sigma_i\right) \tag{5}$$

Technology is assumed to be different across industries but identical across regions, $\mathbf{w}$ is the vector of factor prices, and $\mathbf{P}_j$ is the vector of price indices for intermediate inputs in region j. If one substitutes equation 4 back into our definition of the price index, it is clear that proximity to suppliers lowers the cost of inputs. Finally, the function p_i is an increasing function of input prices, $\mathbf{w}$ and $\mathbf{P}_j$, and decreasing in the perceived demand elasticity of substitution.

Adding equations 4 and 5 to the main demand equation 3, results in the following geography-relevant expression for demand of industry i goods produced in region j.

$$x_{ij} = p_i\left(\mathbf{w}, \mathbf{P}_j, \sigma_i\right)^{-\sigma_i} \sum_{j'} \left(\frac{t_{ijj'}}{P_{ij'}}\right)^{1-\sigma_i} E_{ij'} \quad s.t. \quad P_{ij'} = \left(\sum_j n_{ij}\left(p_{ij} t_{ijj'}\right)^{1-\sigma_i}\right)^{\frac{1}{1-\sigma_i}} \tag{6}$$

Demand for industry i goods is now clearly a function of distance to and from markets. In the language of economic geography, this is supplier and market access. Where supplier access refers to the benefits a firm derives from proximity to producers of essential inputs. This effect enters the function p_i through the price index. Market access refers to the benefits a firm derives from being close to potential customers. This is apparent in the term after the summation sign. Distance discounts demand from each region, such that firms located closer to big markets will benefit from increased sales.

In comparison, the base case is when transport costs between regions is zero (i.e., $t_{ijj'} = 1$). Under this condition, demand for a good produced in industry i is invariant to location. Whether a producer locates in Hokkaido or Kyushu the demand for one's product will be unchanged because the price of the good is the same across varieties in a single industry. That is, $p_{ijj'} = p_{i'j'j'}$ and arises because technologies, input prices and market prices are equated across goods regardless of location.

An equilibrium in this economy can be defined with the addition of factor market clearing conditions which is beyond the purpose of this presentation. However, it is important to distinguish between a long- and a short-run equilibrium in this context. Two previous papers (Hanson [7] and Redding and Venables [15]) implemented a structural variant of the above model imposing the long-run monopolistic competition condition that profits will equal zero. Under this condition the number of firms is pinned down by the factor market clearing conditions, and the amount each firm produces, x_{ij}, is defined by the zero-profit condition. Thus, equation 6 implicitly defines regional factor prices as a function of supplier and market access.

In this paper, however, I want to model the reaction of firms directly following an earthquake in a new short-run equilibrium point; and in the short-run, it is unreasonable to assume that the zero-profit condition holds. Yet, it is likely that wages and the number of firms are rigid within the given time period. Under these conditions, supplier and market access explicitly defines output per firm in equation 6, rather than implicitly defining factor

prices.

Finally, to bring equation 6 closer to the data, I transform the demand equation into one for total nominal output for industry i in region j. Multiplying by the price and the number of firms in industry i within region j results in the following expression for nominal output:

$$y_{ij} = p_{ij} n_{ij} x_{ij} = n_{ij} p_i \left(\mathbf{w}, \mathbf{P}_j, \sigma_i\right)^{1-\sigma_i} \sum_{j'} \left(\frac{t_{ijj'}}{P_{ij'}}\right)^{1-\sigma_i} E_{ij'}. \tag{7}$$

To summarize, if the cost of transactions across distance is greater than zero, t is strictly greater than one, geography matters for the determination of production patterns across space and industries.

III. Data

Output data and industry characteristics are taken from the Japanese census of manufacturers. This census is implemented on an annual basis and is published in various report formats. To fully account for the geographical dispersion of economic activity, this paper uses the report by industrial district [16], which divides economic information based on the classification of 253 industrial districts. The report includes information on gross output, value added, salaries, employees, and fixed assets. Within each industrial district, data is additionally disaggregated between twenty-three industries; and to account for the large reclassification that occurred in 1994, I combine two of the industries to form a dataset with twenty-two distinct industries.

The estimation procedure requires output data for both a year preceding and a year following the earthquake of January 17, 1995. Because data is reported on a calendar year (January through December), 1994 is a natural choice for the base year. However, the earthquake created significant problems for collection and accurate reporting of 1994 data. In general, data from Hyogo prefecture, which includes Kobe, is considered incomplete for

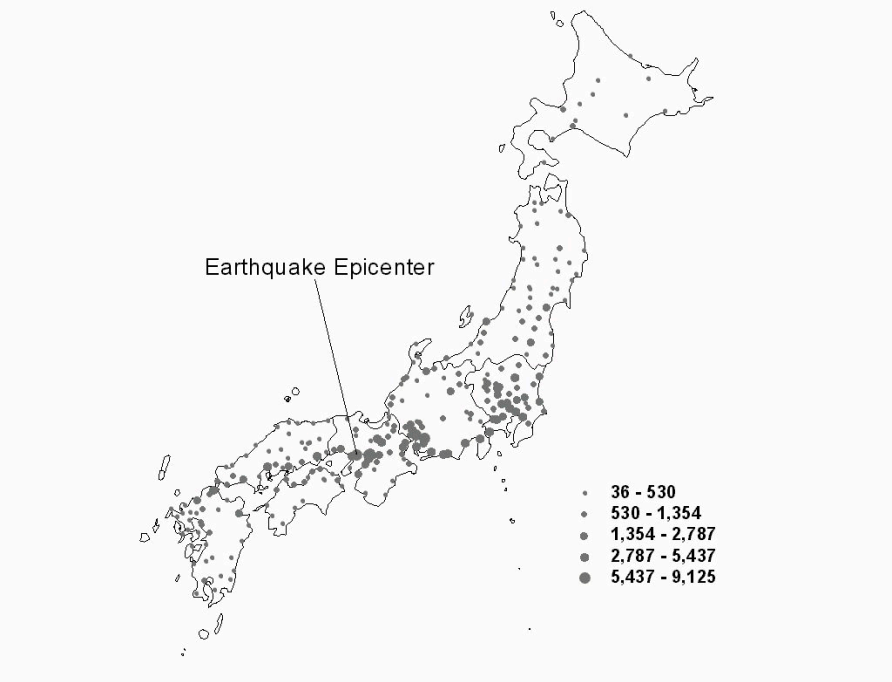


Figure 1: 1993 Gross Industrial Output by Industrial District (in billions of Yen)

1994. Instead, I use 1993 as the base year. For the year following the earthquake I use both 1995 and 1996.

Figure 1 spatially plots Japanese output in 1993. Each circle represents a district and the size indicates the extent of output. Output is clearly not evenly distributed over space. Production is heavily concentrated on the Pacific coast and clustered around three major industrial cities, Tokyo, Nagoya and Osaka. Kobe is part of the cluster of output surrounding Osaka. The economic districts are an improvement over arbitrary administrative definitions of regions commonly used in the literature (e.g., prefectures, states or counties). Yet the districts are often located in very close proximity, and thus, may be indistinguishable from one another in terms of access to markets, suppliers and factors used in production.

Technology and consumption preferences are taken from the 1995 Japanese input-output table [18]. At its most disaggregate level this table is 519X403. That is, there are 519 defined input industries, and 403 industries defined overall. Using the input-output manual [17], I transform the table into a technology matrix for the twenty-two aforementioned in-

dustries. Furthermore, the input-output table contains information on final demand that is used to infer consumer preferences over manufactured goods. Both technology and consumer preferences are assumed to be constant across districts in Japan and over the very short run.

Finally, a geography model needs some measure of transport costs between regions. I use the distance between districts as a proxy. This requires the calculation of a 253X253 distance matrix, such that the fifth element in the one-hundredth row is the distance between the fifth and the one-hundredth district. To calculate the distance between districts I first obtain the coordinates for each industrial district[5]. Next, I project these coordinates on to a flat service using a equidistant cylindrical projection with central median 138 degrees and reference latitude 36 degrees, and finally, using an extension in Arc View I am able to calculate the distance matrix between points[6].

IV. Empirical Framework

The empirical analysis is derived from the theoretical framework in section two. I proceed by taking the log of equation 7, adding a time dimension, and assuming the upper-tier production functions and consumption functions are Cobb-Douglas.

$$(8) \qquad \ln y_{ijt} = \alpha_{ijt} + (1-\sigma_i)\sum_{i'} a_{i'i} \ln P_{i'jt} + \ln \sum_{j'} \left(\frac{t_{ijj'}}{P_{ij't}}\right)^{1-\sigma_i} \left(\alpha_i Y_{j't} + \sum_{i'} a_{ii'} y_{(ij)'t}\right)$$

The first term on the RHS is a function of n_{ij}, factor prices, and the elasticity of substitution in industry i. Depending on our assumption about the rigidity of prices, this may or may not be constant in the short-run. The second term is a measure of supplier access, and the final term is a combination of market access for intermediate demand and final demand. Supplier access is a linear combination, based on technology coefficients, of composite input prices; market access measures proximity to demand.

Structural estimation of this equation requires the econometrician to first pin down

the 21X253 endogenous P's. Methods in other structurally estimated equations either use bilateral trade data to first estimate the region specific price indices (e.g., Redding and Venables [15]) or use methods that require one to assume a single aggregate manufacturing industry (e.g., Anderson and Van Wincoop [1] and Hanson [7]). In this paper neither method is reasonable given a lack of bilateral data between regions, and the interest in keeping some variation across industries.

I instead concentrate on estimating a reduced form version of equation 8 with appropriate proxies for supplier access (SA), market access to intermediate demand (IMA) and market access to final demand (FMA). While this method does not allow us to estimate industry trade costs or elasticities of substitutions, it does allow us to test the validity of the spatial geography hypothesis without imposing too many hard-to-believe assumptions.

The calculation of the proxy variables is straightforward. For both market access variables, IMA and FMA, the proxy is a distance-weighted average of regional demand expenditures,

$$IMA_{ijt} = \sum_{j' \neq j} d_{jj'}^{-1} \sum_{i'} a_{ii'} y_{(ij)'t} \qquad FMA_{ijt} = \sum_{j' \neq j} d_{jj'}^{-1} \alpha_i Y_{j't},$$

where $d_{jj'}$ is the distance between region j and region j'. These two terms closely approximate the third term in equation 8: essentially, the weight $d_{jj'}^{-1}$ replaces the more complicated term $\left(\frac{t_{ijj'}}{P_{ij't}}\right)^{1-\sigma_i}$. I also leave out a region's own demand to avoid any spurious correlation with the dependent variable.

A proxy variable for supplier access needs to capture the proximity of region j, industry i to suppliers, which in turn lowers the $\mathbf{P}$ vector for region j. The derived term also closely follows the definition of supplier access in equation 8. A distance-weighted measure of output, $\sum_{j' \neq j} d_{jj'}^{-1} y_{(ij)'}$, replaces $\ln P$ such that,

$$SA_{ijt} = \sum_{i'} a_{i'i} \sum_{j' \neq j} d_{jj'}^{-1} y_{(ij)'t},$$

and supplier access is a technology-weighted measure of composite input prices proxied by distance to suppliers.

The relationship between geography and regional output can then be estimated using the following stochastic transformation of equation 8:

$$\ln y_{ijt} = \alpha_t + \beta_1 \ln SA_{ijt} + \beta_2 \ln IMA_{ijt} + \beta_3 \ln FMA_{ijt} + \varepsilon_{ijt}, \tag{9}$$

where the error term enters additively in the logarithmic form and is comprised of components from the first term in equation 8, other omitted variables, and a white noise error term.

$$\varepsilon_{ijt} = \alpha_i + \alpha_j + \alpha_{ij} + \alpha_{it} + \alpha_{jt} + \alpha_{ijt} + u_{ijt}$$

The first six terms plus the constant, α_t, are simply a decomposition of α_{ijt} from equation 8. If factor prices and the number of firms are rigid in the short run, then the fourth through sixth terms are zero. Fixed over time, the first three terms represent fixed effects specific to industries, regions and industry-region. Stepping outside of the model, the α's can be interpreted to contain alternative explanations for the spatial distribution of industrial production. This may include industry specific items, such as the degree of increasing returns, or region specific items, such as preferable physical geography or access to a thick labor market. Finally, u_{ijt} is a normally distributed white noise component which may reflect productivity shocks inherent in manufacturing due to logistic problems.

Spatial geography predicts that relative advantage is partially determined by location to markets and suppliers. Thus, a result finding that β_1, β_2 and β_3 are statistically greater than zero would be consistent with the theory. More specifically, a result finding β_1 is strictly greater than zero, would suggest location close to suppliers of inputs is a source of relative advantage. Likewise, evidence that β_2 is greater than zero supports the theory that location close to firms buying intermediates is advantageous, and if β_3 is greater than zero, location close to consumers is a source of advantage.

Table 1: Geographic Explanations for Output Dispersion

	(1)	(2)	(3)	(4)
Model	OLS	OLS	FE	RE
Dependent variable	$\ln y_{ijt}$	$\ln y_{ijt}$	$\ln y_{ijt}$	$\ln y_{ijt}$
$\ln SA_{ijt}$	0.2087	1.3641	0.9605	1.2061
	(0.0603)**	(0.2705)**	(0.2006)**	(0.1406)**
$\ln IMA_{ijt}$	0.4858	1.5107	0.5497	0.7976
	(0.0245)**	(0.2333)**	(0.1810)**	(0.1089)**
$\ln FMA_{ijt}$	0.1800	-0.7469	-0.3044	-0.5620
	(0.0131)**	(0.4502)+	(0.3523)	(0.4645)
Year dummy	0.0220	0.0799	0.0272	0.0439
	(0.0075)+	(0.0110)**	(0.0080)**	(0.0089)**
Industry Dummies	No	Yes	No	Yes
Region Dummies	No	Yes	No	Yes
R-squared	0.25	0.65	0.03	
Observations	9043	9043	9043	9043
Number of unique			4601	4601

Robust standard errors allowing for regional clusters in parentheses
Constant not reported
+ significant at 10%; * significant at 5%; ** significant at 1%

V. Results

Results from the estimation of equation 9 using output from Japanese industrial districts categorized by industry are reported in the first column of table 1. The results are consistent with the theory, namely, that there exists a strong and persistent correlation between industry-specific regional output and all three explanatory variables, SA, IMA, and FMA. Proximity to suppliers, proximity to intermediate goods markets, and proximity to final goods markets are all important. All three proxies are positive and significant at the one-percent level and the R-squared is 0.25. This is not surprising given the long line of papers dating back to Harris [8] and recently including Hummels [11] and Leamer [13] that have found related results. This evidence is also similar to the general observation that production tends to agglomerate (e.g., Silicon Valley) and locate near large markets (see figure 1).

The scatter plots in figure 2 convey the same result. There is a clear positive correlation

between all proxies and the log of industry-specific regional output. The correlation coefficient ranges between 0.26 for final market access and 0.44 for intermediate market access, supplier access falls inbetween.

Because this data set has variation across three dimensions, across time, across industries and across regions, it is straighforward to begin controlling for omitted variables that are constant over the short-run. First, without implementing a fixed effect model, we can control for both differences across industries and differences across regions that are constant over time, or, in other words, the first two terms in the definition of the error term, $\alpha_i + \alpha_j$.

Omitted industry effects can affect the estimated results through two channels: national comparative advantage and industry scale effects. First, any Japanese comparative advantage may not only disproportionately affect specific industries, but can also affect a set of industries linked through the input-output table. Linked industries with a common technological edge over world competitors, or linked industries drawing from the same skill-specific labor pool may be equally influenced. Thus, omitted industry effects can positively bias the estimated coefficients on supplier access and intermediate market access.

Second, production scale economies help determine the optimal plant size and in turn can affect the regional distribution of output within Japan. At one end of the spectrum, industries with no scale economies or constant returns to scale, can spread production evenly across space. Industries with a high degree of scale economies, in contrast, will tend to concentrate production in fewer locations. Thus, industry scale economies which are embodied in industry specific effects can help determine the distribution of production over space and may also be a source of omitted variable bias. See, for example, Kim's [12] evidence to this respect using two centuries of U.S. regional data.

Regional effects, like location fundamentals and regional labor pooling, may also bias the estimated results. The physical characteristics of a region may naturally explain why

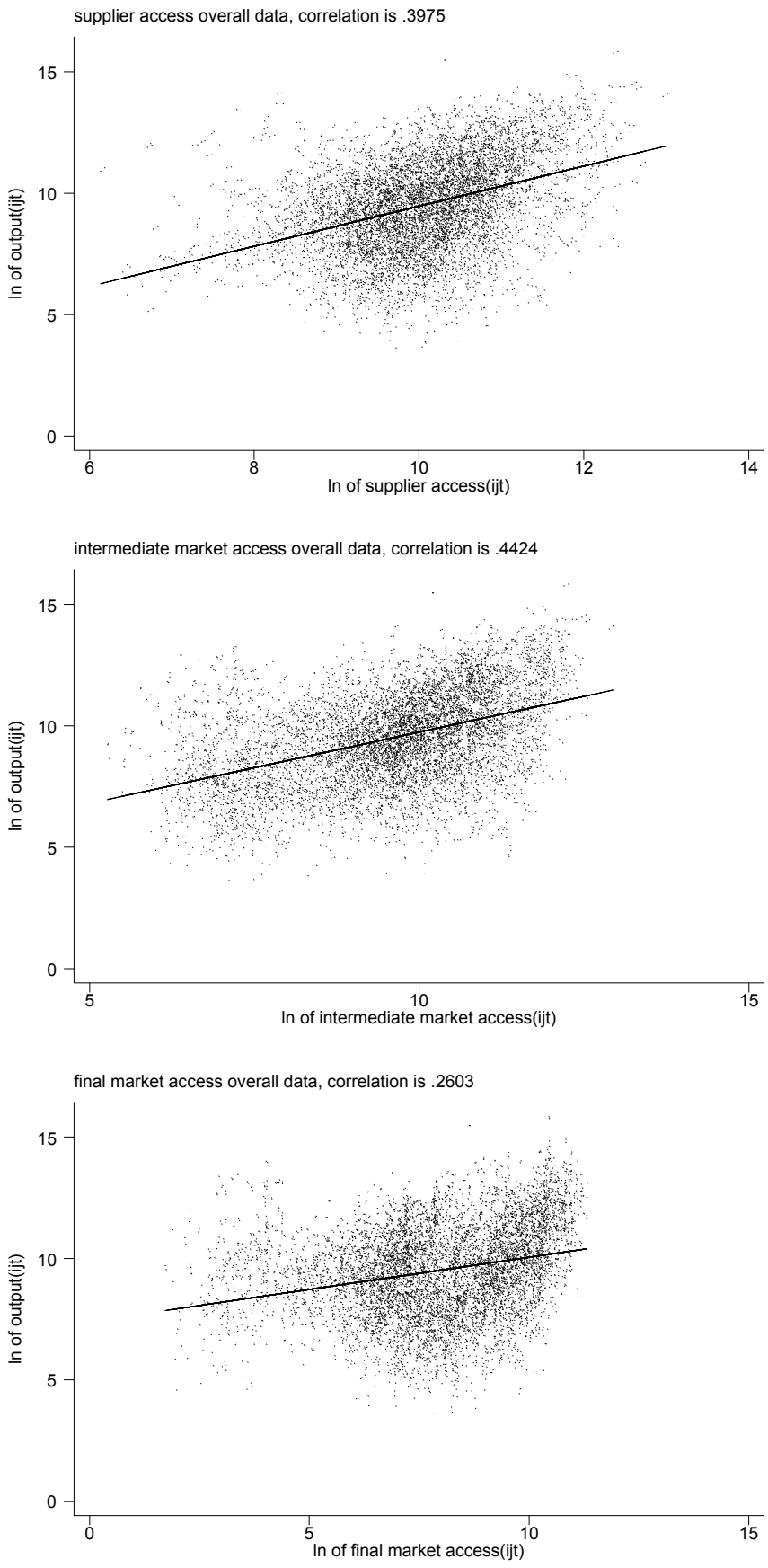


Figure 2: Overall Relationship Between Output and SA(top), IMA(middle), and FMA(bottom)

economic activity agglomerates. The quality of a region's land, favorable access to the sea, and climate, are all location fundamentals that are fixed over time and are natural explanations for the spatial distribution of production. In Davis and Weinstein [3], the persistence of cities in Japan to exogenous shocks and over a 8,000 year time period suggests that location fundamentals are an important explanatory variable for the pattern of relative regional densities. And because location fundamentals will not only be correlated with regional output, but also with regional measures of SA, IMA and FMA, omitting regional effects can positively bias the estimates.

The positive impact of regional labor markets might also explain the concentration of industry and will be embodied in any region fixed effect over the short run. Firms benefit from a thick labor market which allow workers to specialize their skills, and will therefore have an incentive to locate near such markets. This, in turn, can create incentives for the agglomeration of economic activity. In this dataset the relevant labor market may not be restricted to any specific district but might apply to a larger region encompassing several districts. This, too, can positively bias the estimated results.

I proceed to control for both regional and industry specific effects by estimating an equation with both 21 industry and 252 regional dummy variables. The results are reported in the second column of table 1. Contrary to expectations, the coefficients on both supplier access and intermediate market access greatly increase and remain significant at the one percent level.

The main finding, however, is that the coefficient on final market access is no longer significantly greater than zero. That is, regional and industry effects, once accounted for, can explain the previous strong correlation between final market access and industry level output. Combined with the previous result, this suggests that while firms benefit from proximity to one another, they do not gain from relative proximity to consumers.

Finally, I implement a full fixed effect model and report the results in column three. This model eliminates all unobserved heterogeneity in the model, thus, the first three terms

in the definition of the error term are swept away, $\alpha_i + \alpha_j + \alpha_{ij}$. This differs from the regression in column two by also sweeping away the constant industry specific region effect, α_{ij}. The results are not dissimilar from the estimated coefficients in column two. The impact of supplier access and intermediate market access decreases to a number between the original estimate and the estimate in column two which includes regional and industry dummy variables, but the results remain strongly significant and robust. The estimate on final market access is now negative and insignificant from zero.

An alternative random effects model with dummy variables for both industry and region effects is also tested (column 4). Although the estimated coefficients are remarkably similar to the coefficients in the fixed effect model, a Hausman test of the hypothesis that the correlation between α_{ij} and the regressors is zero results in a $\chi^2(4)$ greater than fifty and can be easily rejected. Thus, region and industry specific effects are not sufficient to control for all unobserved heterogeneity, and a full fixed effect model is maintained through the remainder of this paper.

VI. Instrumented Equation

This section continues to deal with the correlation between the error term and the regressors. In particular I am concerned with the endogeneity of all three variables, SA, IMA and FMA. The direction of causality is not clear, and for example, while output may be influenced by supplier access, supplier access is equally influenced by downstream output levels. Thus, the earthquake provides us with a unique exogenous shock to help disentangle these forces and map the exogenous regressors to the dependent variable.

A. Earthquake Instruments

Instruments for both supplier and market access can be constructed using proxies for earthquake damage. The earthquake occurred on January 17, 1995, and is a possible strong

predictor of changes in supplier and market access during 1995. I develop proxies to predict both variation across space and variation across industries.

To predict changes across space, a straightforward instrument for earthquake damage is the distance to the epicenter, or d_j. The closer an industrial district is to the epicenter, the more likely market access and supplier access were affected by output fluctuations and transportation costs in the vicinity of the earthquake.

To predict changes across industries I develop separate instruments for intermediate market access and supplier access[7]. To predict changes in intermediate market access, I measure the percentage of total Japanese intermediate demand for industry i located in either Akashi or Kobe for 1993, the two districts most heavily affected by the earthquake.

$$\text{demand share}_i = dem(y)_i = \frac{\sum_{i'} a_{ii'} y_{i'kobe,93}}{\sum_{i'} a_{ii'} y_{i'japan,93}}$$

This instrument can also be constructed using either employment, $dem(l)_i$, or the number of establishments, $dem(n)_i$, in place of output.

The constructed instrument for supplier access is also a measured share of activity in the earthquake affected region. Using the transpose of the input-output table, this is a technology-weighted average of relative supplier activity in Kobe and Akashi,

$$\text{supply share}_i = \sup(y)_i = \frac{\sum_{i'} a_{i'i} y_{i'kobe,93}}{\sum_{i'} a_{i'i} y_{i'japan,93}}.$$

This instrument can similarly be constructed using employment, $\sup(l)_i$, and establishment figures, $\sup(n)_i$. *A priori*, these instruments should be negatively correlated with the proxies for supplier and market access.

The impact of the constructed instruments is also likely to diminish with distance from the earthquake epicenter. To account for this effect interactive terms with the log of distance to Kobe are included as additional instruments.

Table 2: First Stage Regressions

	(1)	(2)	(3)
Model	OLS	OLS	OLS
Dependent variable	$\Delta \ln SA_{ij}$	$\Delta \ln IMA_{ij}$	$\Delta \ln FMA_{ij}$
$\ln d_j$	0.0029	0.0092	0.0055
	(0.0008)**	(0.0019)**	(0.0010)**
$dem(n)_i$	-2.6990	-7.2497	-0.0575
	(0.2431)**	(0.7526)**	(0.0531)
$dem(l)_i$	-0.7962	0.2165	0.0359
	(0.3393)*	(1.3696)	(0.1141)
$dem(y)_i$	1.1150	-1.3552	-0.0396
	(0.4201)**	(1.0993)	(0.1669)
$\sup(n)_i$	5.6221	10.2860	0.0645
	(0.3387)**	(1.3542)**	(0.0622)
$\sup(l)_i$	-0.3064	-0.0273	-0.0396
	(0.3212)	(0.5957)	(0.1365)
$\sup(y)_i$	-1.3898	1.5720	0.0463
	(0.4756)**	(0.8431)+	(0.2163)
$\ln d_j * dem(n)_i$	0.2009	0.5866	0.0112
	(0.0454)**	(0.1419)**	(0.0103)
$\ln d_j * dem(l)_i$	0.1163	0.9838	-0.0035
	(0.0592)+	(0.2360)**	(0.0197)
$\ln d_j * dem(y)_i$	-0.0913	-0.6568	0.0013
	(0.0763)	(0.1973)**	(0.0281)
$\ln d_j * \sup(n)_i$	-0.3957	-1.1820	-0.0140
	(0.0635)**	(0.2636)**	(0.0117)
$\ln d_j * \sup(l)_i$	-0.0347	-0.4581	0.0030
	(0.0583)	(0.1098)**	(0.0233)
$\ln d_j * \sup(y)_i$	0.1698	0.5112	-0.0004
	(0.0878)+	(0.1582)**	(0.0367)
Constant	-0.0672	-0.1447	-0.0159
	(0.0043)**	(0.0083)**	(0.0060)**
Observations	4442	4442	4442
R-squared	0.08	0.32	0.30
F	461.64	672.99	6.20

Robust standard errors allowing for regional clusters in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%

B. First Stage

The first stage regressions are estimated and reported in table 2. The earthquake damage proxies are strong predictors of changes in supplier access, intermediate market access and final market access. The instruments are most robust for the prediction of intermediate and final market access. The R-squared statistics are 0.32 and 0.30, respectively. The log of distance to the epicenter of the earthquake alone explains thirty percent of the variation in the change in final market access. The prediction of supplier access is weaker, but is still robust with an F-statistic of 461 and an R-squared of 0.08.

In all three equations the log of distance to the earthquake epicenter is the expected sign, positive. Proximity to the earthquake clearly negatively affected the change in supplier and market access between 1993 and 1995. The estimates, however, are only strongly significant in the prediction of changes in intermediate and final market access.

The instruments for the percentage of demand located in the region of the earthquake, $dem(x)_i$, are generally negative, and the estimate is strongly significant and negative for the proxy built with the number of establishments, n. Although this is an instrument specifically used for the prediction of the change in intermediate market access it also has predictive power for changes in supplier access. The negative sign confirms our priors that the change in intermediate market access for industries with a greater concentration of demand located in the earthquake region were disproportionately affected. The interactive term with distance and $dem(n)$ and $dem(l)$ are both positive and significant, also suggesting that the impact on the change in intermediate market access weakened with distance to the epicenter.

Finally, the predictive power of the instrument for the percentage of suppliers located near the epicenter is significant in both equations for the change in supplier access and the change in intermediate market access. Several of the interactive terms with distance are also significant. The signs on the estimates, however, are opposite of the expected sign. This may arise due to a strong correlation with the $dem(x)$ instruments. In an equation

Table 3: Instrumented Estimation

	(1)	(2)	(3)	(4)	(5)
Model	FD	2SLS-FD	2SLS-FD	2SLS-FD	2SLS-FD
Dep. Variable	$\Delta \ln y_{ij}$	$\Delta \ln y_{ij}$	$\Delta \ln y_{ij}$	$\Delta \ln y_{ij}$	$\Delta \ln y_{ij}$
$\Delta \ln SA_{ij}$	0.9605	0.6641	1.8885	2.8474	0.7040
	(0.2006)**	(0.5905)	(0.5712)**	(2.7844)	(0.6052)
$\Delta \ln IMA_{ij}$	0.5497	0.8750		0.4384	0.9006
	(0.1810)**	(0.2329)**		(1.5084)	(0.2386)**
$\Delta \ln FMA_{ij}$	-0.3044	0.1868		-0.8172	
	(0.3523)	(0.9314)		(1.9719)	
Constant	0.0272	0.0235	0.0247	-0.7110	-0.0853
	(0.0080)**	(0.0221)	(0.0128)+	(0.6833)	(0.0173)**
Industry dummies	No	No	No	Yes	No
Region dummies	No	No	No	No	Yes
Observations	4442	4442	4442	4442	4442

Robust standard errors allowing for regional clusters in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%

with only the $\sup(x)$ instruments, the signs switch in the direction of theory.

Overall, the instruments are successful predictors of the change in all three explanatory variables over the very short run. Problems with weak instruments is not a major concern.

C. Instrumented Results

The results from the two-stage least squares procedure using first differences is reported in the second column of table 3. Column one reports again the results from the full fixed effect model or equivalently the first difference model for reference. Notice that the estimated coefficients between column one and column two are not that drastically different. A Hausman test between these two regressions cannot reject the hypothesis that the estimated coefficient are identical ($\chi^2(4) = 5.44, p = .1426$). In other words, once we implement a fixed effect model, the hypothesis that the regressors are exogenous cannot be rejected. This was surprising, as the major goal of this paper was to deal with the endogeneity of the explanatory variables. Although, failure to reject this hypothesis may be due to the short period in which a large component of the variation in the regressors was due to the exogenous earthquake. This is partially evident by the high R-squares in the first stage regressions. Regardless,

theory tells us these regressors before instrumenting are endogenous and I continue with the instrumented results.

In the instrumented equation, the significance of both final and intermediate market access are unchanged; final market access is insignificant from zero and intermediate market access is significant at the one-percent level. The coefficient on supplier access decreases, as expected, but is now insignificant from zero. A higher standard error due to both strong collinearity with IMA and a weaker predicted value from the first stage regression may account for the insignificant result. However, estimating the equation with only supplier access results in a significant coefficient (column three). Thus, as concluded in the earlier sections, firms benefit from proximity to one another, but do not gain from relative proximity to consumers.

How much do they benefit? I really cannot say, but it is quite clear that the explanatory power of the theory is greatly decreased. The R-squared from the first regression in table 1 of 0.25 is now a small 0.03 in the within instrumented regression in the second column of table 3. This is also demonstrated in figure 3 which plots the instrumented variables against the change in industry output. In fact, if one juxtaposes figure 2, the before shot, against figure 3, the after shot, the reduction in correlation for each variable is remarkable. Yet if one squints hard enough, it is still apparent that there exists a positive correlation between intermediate market access and output (middle), as well as supplier access and output (top). The coefficient in the former case has actually increased.

A final issue is whether or not the instruments are truly exogenous. In particular, distance to the epicenter might be affecting the regressors independently through other channels such as transportation costs[8] or direct damage. To test the exogeneity assumption I implement an overidentification test, which is essentially another Hausman test between the regression in column two and a similar regression where I leave out distance and the interactive terms as instruments. This can only be done for SA and IMA since the change in FMA only varies across regions. The $\chi^2(2) = 0.32$ with a p-value equal to .8520. Thus,

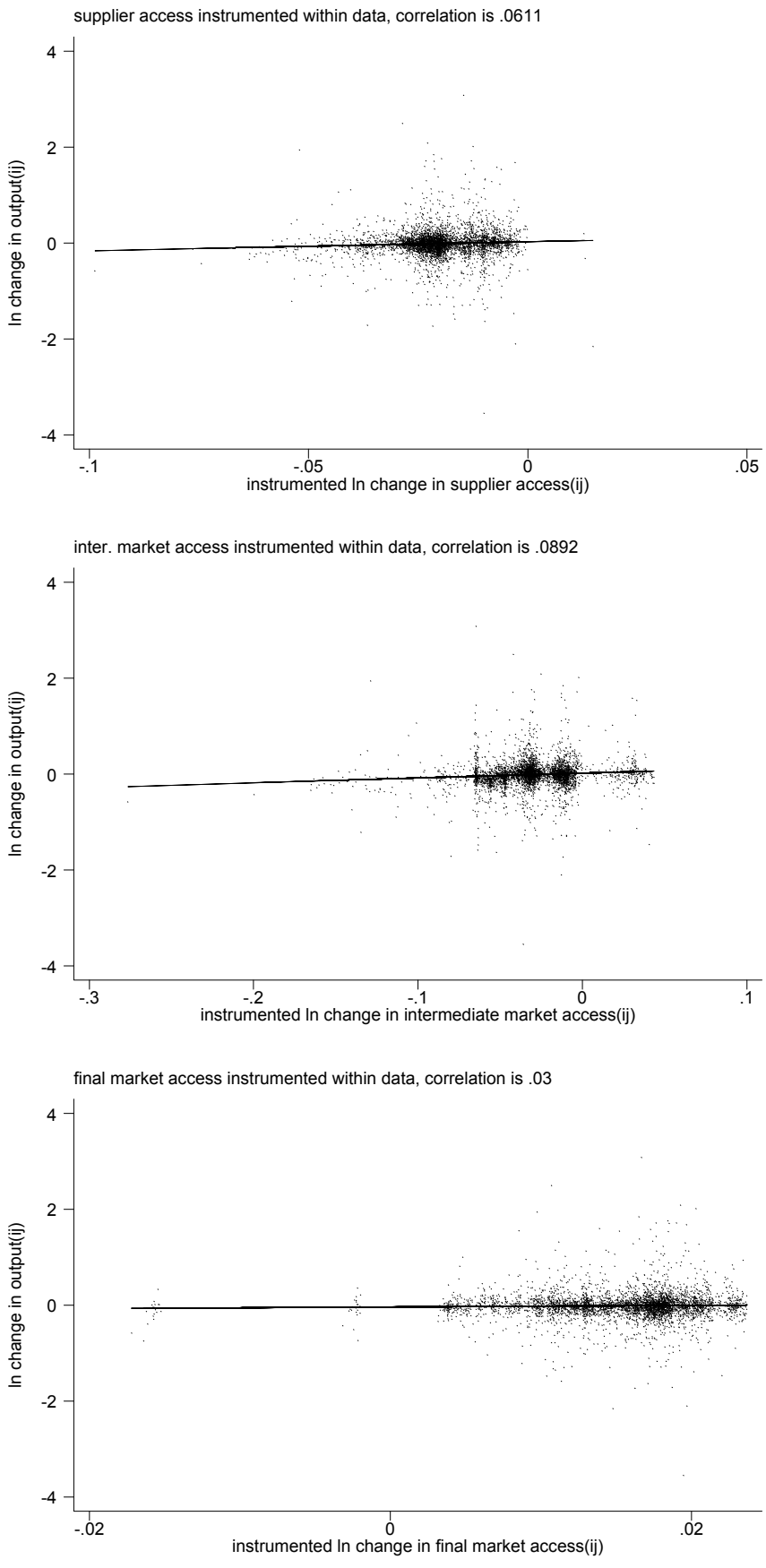


Figure 3: Instrumented Within Relationship Between Output and SA(top), IMA(middle), and FMA(bottom)

(assuming the other instruments are exogenous) we cannot reject the hypothesis that distance is an appropriate instrument. I discuss and check this assumption more thoroughly in section A.. Likewise, a similar overidentification test on the validity of the instruments, $dem(y)$ and $sup(y)$, also fails to be rejected.

D. Intra- or International Geography?

With two dimensions of variation remaining, across industry and across region variation, it is possible to ask one final question: Does geography matter for intranational trade, international trade, or both? Stated differently, do the geography relevant regressors explain the within-variation across region, the within-variation across industries, or both?

If geography matters for trade and the location of production within national borders then the correlation between the regressors and output should persist after controlling for common industry changes over the time period, α_{it}. This is equivalent to subtracting the industry mean from output, and the industry mean from the regressors, and looking at the within industry variation across regions.

I implement this strategy by including industry dummies in the estimation and the results are reported in column four of table 3. A scatter-plot of the correlation for intermediate market access is also presented in the top portion of figure 4. Any significant variation is wiped away in the new results. All variables are insignificant and the scatter plot is indicative of something close to zero correlation. This suggests that location within Japan does not give advantage to producers near consumers or to producers located near suppliers.

If within industry across region variation cannot account for the significant correlation in column two, then within region across industry variation must be the answer by default. I test this with region dummies, α_{jt}, and the results are reported in the last column of table 3[9]. A cropped scatter-plot of the correlation for intermediate market access is also presented in the bottom of figure 4.

The results are nearly identical to the original instrumented equation in column two.

And to add statistical oomph to the previous result that regional variation is not important, the region dummies cannot be jointly statistically distinguished from zero. The scatter-plot also indicates an upward trend for intermediate market access. A scatter plot for supplier access (not shown) indicates the same upward trend.

What does this all mean? The results in column four suggests that access to markets and access to suppliers is not a source of relative advantage within borders. Thus, the distribution of output in Japan must be determined by other factors, as is indicated by the previously mentioned papers of Kim [12] and Davis and Weinstein [3]. This, however, is contradictory to the papers by Hanson [7] and Combes and LaFourcade [2] that find positive evidence for geography in the US and France, respectively.

In contrast, the estimated coefficients in column five suggest that access to markets and access to suppliers is a source of relative advantage across industries in Japan and therefore across borders. The difference between these two results is likely due to the difference between transaction costs within borders, and transaction costs across borders. Transaction costs greatly increase when goods cross an international border, therefore, geography is relevant in explaining international, but not intranational, patterns of production.

VII. Robustness Checks

To check the robustness of the results presented thus far, it is straightforward to carry out two robustness checks. First, the validity of distance to the epicenter of the earthquake as an instrument may be compromised by events specific to 1995. Second, the 1995 data may largely be capturing the impact of transitory shocks. I use 1996 data to check the robustness of both the distance instrument and to confirm the results over a longer period. Finally, the data is significantly censored at zero. This may downwardly bias the results and account for the insignificance of both final market access and all variables once we control for industry variation. I start with the timing of the data.

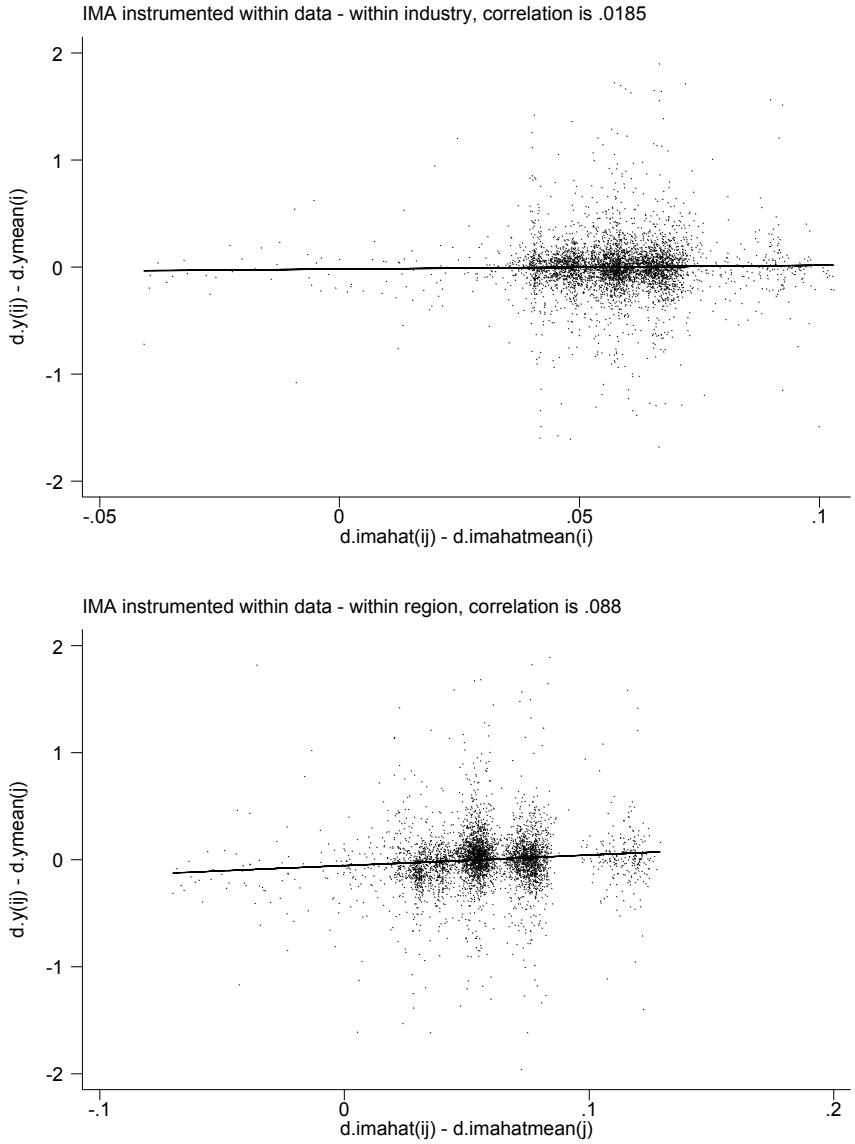


Figure 4: Variation Across Regions(top) and Across Industries(bottom)

A. 1996 Data

A robustness check with the end year of 1996 rather than the end year of 1995 solves two problems. First, it helps check the exogeneity assumption of the distance to epicenter instrument. Second, it helps confirm these results are not just a reaction to transitory shocks and do contain some evidence for the impact of permanent shocks and permanent production patterns.

The earthquake impacted output in the Kinki region through two, and possibly three channels. The first was direct damage to industrial clusters near the epicenter. The second was through additional transportation costs created by congestion and disruption in the regional infrastructure system. The third possible channel was through a general malaise that inflicted the workforce and may well have impacted productivity levels.

Distance will directly impact both output in Kobe and Akashi through the first channel. The remaining districts, however, were equally unaffected by direct damage and will not be affected by distance through the same channel. Furthermore, excluding both Kobe and Akashi from the sample does not alter the results.

The second channel, increased transportation costs, will be greater for districts closer to the earthquake and weaker for districts far away. Yet there is no *a priori* reason to think that these costs will work through anything other than a firms ability to reach markets and receive inputs on time. Still, if one believes there is a direct relationship between transportation costs, distance and output, the 1996 data helps alleviate this problem. Because congestion affected output in 1995 and not in 1996, any bias that was related to the distance instrument should have less or no impact in the implementation of the same test using the 1996 data.

The final channel, a reduction in workforce productivity, may certainly dissipate with distance to the earthquake. This affect, however, should also be less apparent in the output level of 1996. So an implementation of the same test, with 1996 data should also rid the data of any problems associated with the correlation between worker productivity, distance and output.

Table 4: Robustness Check with 1996 Data

	(1)	(2)	(3)	(4)	(5)	(6)
Model	OLS	OLS	FD	2SLS-FD	2SLS-FD	2SLS-FD
Dep. variable	$\ln y_{ijt}$	$\ln y_{ijt}$	$\Delta \ln y_{ij}$	$\Delta \ln y_{ij}$	$\Delta \ln y_{ij}$	$\Delta \ln y_{ij}$
$\ln SA_{ijt}$ or $\Delta \ln SA_{ij}$	0.2093	1.2839	0.7248	0.9187	1.9620	0.8892
	(0.0599)**	(0.2608)**	(0.1908)**	(0.3144)**	(2.7106)	(0.3235)**
$\ln IMA_{ijt}$ or $\Delta \ln IMA_{ij}$	0.4870	1.5518	0.6349	0.9095	0.5996	0.9233
	(0.0246)**	(0.2282)**	(0.1584)**	(0.2275)**	(1.3468)	(0.2324)**
$\ln FMA_{ijt}$ or $\Delta \ln FMA_{ij}$	0.1794	-1.3411	-0.2367	-1.0105	-2.1236	
	(0.0132)**	(0.4287)**	(0.3255)	(1.1277)	(3.9667)	
yeardummy or constant	0.0175	0.0770	0.0189	0.0515	-0.7104	-0.1612
	(0.0090)+	(0.0171)**	(0.0128)	(0.0426)	(0.7110)	(0.0024)**
Industry dummies	No	Yes	No	No	Yes	No
Regional dummies	No	Yes	No	No	No	Yes
Observations	9049	9049	4417	4417	4417	4417
R-squared	0.25	0.65	0.04	0.03	0.05	0.09

Robust standard errors allowing for regional clusters in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%

Finally, the 1996 data helps sort out the impact of transitory and permanent shocks. The variation in the RHS variables using 1995 data includes reductions in supplier, intermediate market and final market access that were both transitory and permanent. A transitory shock to a supplier may interrupt production of downstream users, but because this is a transitory shock, firms may not look for new suppliers and will only be temporarily affected by the shortage of inputs. A permanent shock to a supplier may also have a transitory affect on downstream users as firms incur search costs and identify new suppliers. Both of these transitory affects should be negligible in the 1996 data. Therefore, the 1996 data measures the impact of the remaining permanent shock on downstream output levels in an intermediate equilibrium, and not just the impact of transitory inconveniences.

Results using the 1996 data are included in table [4]. The first three columns replicate the first three regressions in table 1 using the 1996 data in place of the 1995 data, and the last three columns replicate the second, fourth and fifth regressions in table 3 also using the 1996 data. The coefficients and standard errors are surprisingly similar. The impact of final market access turns negative after controlling for industry and region effects, and becomes

insignificant from zero after the full-fixed effect model is implemented. Thus, linkages between firms, and not between firms and consumers, continue to be the main source relative advantage in this framework. Furthermore, after instrumenting, in addition to intermediate market access, supplier access also is positive and significant at the ninety-nine percent level. Thus, the results appear to be stronger when using the 1996 data. This is partially due to a lower standard error for supplier access resulting from a stronger prediction in the first stage regression. Finally, the results in columns five and six continue to suggest that geography is important in explaining international rather than the intranational distribution of output.

B. Censored Regression

The estimates presented up to this point exclude 2,089 of the 11,132 possible industry and region combinations that report zero output. This may downwardly bias the results and account for the insignificance of both final market access and all variables once we control for industry variation. In fact, if one looks at either the scatterplot for intermediate market access or final market access in figure 2, the estimated line appears slightly flatter than the actual relationship.

Estimates for the full sample using first OLS and then Tobit[10] are reported in the first two columns of table 5. The Tobit estimates when compared with both the results in column one of this table and the OLS estimates in column one of table 1 are, as expected, significantly larger. The original estimates for both intermediate market access and final market access were, in particular, downwardly biased.

After controlling for both regional and industry fixed effects, column three, the coefficient on final market access turns negative and is no longer insignificant from zero. Thus, the result that linkages between consumers and firms are not important is robust to this estimation. The next step, the full fixed effect model, is impossible to implement using this MLE estimator because of the potentially large bias created by the incidental parameters problem[11]. The instrumented equation is also impossible to implement because the

Table 5: Robustness Check with Censoring Model

	(1)	(2)	(3)	(4)	(5)
Model	OLS	Tobit	Tobit	Tobit	Tobit
Dep. Variable	$\ln y_{ijt}$	$\ln y_{ijt}$	$\ln y_{ijt}$	$\ln y_{ijt}$	$\ln y_{ijt}$
$\ln SA_{ijt}$	0.5601	0.3346	1.2373	1.2306	1.2325
	(0.0379)**	(0.0488)**	(0.3813)**	(0.3891)**	(0.3810)**
$\ln IMA_{ijt}$	0.7721	1.1014	3.7026	3.7886	3.7102
	(0.0187)**	(0.0261)**	(0.2991)**	(0.3067)**	(0.2990)**
$\ln FMA_{ijt}$	0.4215	0.5993	-2.9623	-2.7874	-1.3477
	(0.0168)**	(0.0214)**	(6.1420)	(6.1427)	(0.3868)**
yeardummy	-0.0160	-0.0374	0.1126	0.1331	0.5143
	(0.0632)	(0.0771)	(0.1104)	(0.2659)	(0.8781)
Industry dummies	No	No	Yes	Yes	Yes
Regional dummies	No	No	Yes	Yes	Yes
Industry*Year	No	No	No	Yes	No
Region*Year	No	No	No	No	Yes
Observations	11132	11132	11132	11132	11132

Standard errors in parentheses, constant not reported

+ significant at 10%; * significant at 5%; ** significant at 1%

instruments predict changes in variables rather than levels.

However, to check on the validity of the final conclusion, that geography explains international rather than intranational specialization patterns, I attempt to control for the same components of the error term as was carried out in section D.. These are the fourth and fifth components of the equation for the error term, α_{it} and α_{jt}. I include dummies for α_{it} in column four and dummies for α_{jt} in column five. Because these estimates do not account for the full fixed effect, and the RHS variables have not been purged of endogenous variation, these results can only confirm, but not overturn, the conclusions of section D.. Column five confirms the result that industry variation is an important component of the observed correlation. Column four, however, cannot confirm the observation that regional variation in production cannot be explained by geography.

Overall, these results indicate that the impact of supplier access and intermediate market access may in fact be larger than the reported estimates in the previous sections.

VIII. Conclusion

In this paper I present an empirical test of the geography model using both the variation across industries and industrial clusters in Japan, in addition to using the exogenous variation created by the Kobe earthquake. Three results deserve special attention.

First, although the theory is quite robust in the overall data, explaining as much as twenty-five percent in the variation in output across industries and regions, the theory explains very little of the within-industry-region variation in output. In fact, after controlling for all unobserved heterogeneity that is fixed over time and isolating the exogenous component in the explanatory variables, the three variables, final market access, intermediate market access and supplier access, explain only three percent of the within-variation in output. Whether the explanatory power of the theory for the overall variation in output is similarly small once we isolate the true impact of these variables, or whether this is a peculiarity of this exercise which concentrates on short to intermediate equilibrium points and not long run outcomes, is left unresolved. If the explanatory power is truly only three percent in the overall data, we would have to conclude that economic geography is only a small source of long-run relative advantage.

Second, the evidence strongly suggests that linkages between firms and not linkages between firms and consumers are a source of relative advantage. Firms producing final goods do not need to locate near large markets but do need to locate near suppliers. Likewise, firms producing intermediate manufactured goods do gain from locating near both downstream users and upstream suppliers. The relative size of trade costs in intermediates relative to final goods is the likely source of this difference.

Finally, the evidence suggests that economic geography cannot explain the within-variation regional dispersion in output. In other words, economic geography is not relevant for models of intranational trade. The geography variables are, however, extremely robust in explaining the within-variation across industry variation in output. This suggests that economic geography is a much more robust model for explaining international patterns of

production. Once again, the relative size of trade costs across borders relative to within borders is the likely source of this difference. Thus, as a source of relative advantage, economic geography models are relevant in markets and across regions where trade costs are significant.

REFERENCES

James E. Anderson and Eric Van Wincoop. Gravity with gravitas: A solution to the border puzzle. NBER Working Paper No. 8079, January 2001.

Pierre-Philippe Combes and Miren LaFourcade. Transport costs, geography, and regional inequalities. mimeo, 2002.

Donald Davis and David Weinstein. Bones, bombs and break points: The geography of economic activity. NBER Working Paper No. 8517, October 2001.

Jeffrey Frankel, Ernesto Stein, and Shang-Jin Wei. *Regional Trading Blocs in the World Economic System*. Institute for International Economics, Washington, DC, 1997.

Masahisa Fujita, Paul Krugman, and Anthony Venables. *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Cambridge MA, 1999.

John Gallup and Jeffrey Sachs with Andrew Mellinger. Geography and economic development. *Annual World Bank Conference on Development Economics*, 1998.

Gordon H. Hanson. Market potential, increasing returns, and geographic concentration. mimeo, University of Michigan, November 2000.

C Harris. The market as a factor in the localization of industry in the united states. *Annals of the Association of American Geographers*, 44:315–348, 1954.

Elhanan Helpman and Paul Krugman. *Market Structure and Foreign Trade*. MIT Press, Cambridge MA, 1985.

Bo E. Honore. Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica*, 60(3):533–565, May 1992.

David Hummels. Global income patterns: Does geography play a role? Chapter 2 of PhD Thesis, University of Michigan, 1995.

Sukkoo Kim. Expansion of markets and the geographic distribution of economic activities: The trends in u.s. regional manufacturing structure, 1860-1987. *Quarterly Journal of Economics*, 110(4):881–908, November 1995.

Edward Leamer. Access to western markets and eastern effort. In S Zecchini, editor, *Lessons from the Economic Transition, Central and Eastern Europe in the 1990s*, pages 503–526. dordrecht: Kluwer Academic Publishers, 1997.

Henry Overman, Stephen Redding, and Anthony J. Venables. Trade and geography: A survey of empirics. mimeo, LSE, 2001.

Stephen Redding and Anthony J. Venables. Economic geography and international inequality. mimeo, London School of Economics, April 2001.

Tsusho Sangyo Daijin Kanbo Chosa Tokeibu (Research, Ministry of Industrial Trade Statistics Department, Minister's Secretariat, and Industry). *Kogyo Tokeihyo Kogyo Chiku Hen (Census of Manufactures. Report by Industrial District)*. Tsusho Sangyo Chosakai, Tokyo, 1993, 1995, 1996.

Sangyo Renkan Bukyoku Chokaigi. *Heisei Nana-Nen (1995) Sangyo Renkanhyo Sakusei Kihon Yozumi*. sangyo renkan bukyoku kaigicho, Tokyo, 1997.

Kyodo Henshu Somucho. *Heisei 7-Nen Sangyo Renkanhyo (1995 Input-Output Tables)*. Somucho, Tokyo, 1999.

Notes

*The views expressed in this working paper are those of the author and do not necessarily represent those of the IMF or IMF policy.

[1]This refers to the set of model summarized in Fujita, Krugman, and Venables [5].

[2]Source: http://web.pref.hyogo.jp/syoubou/english/dmh1.html

[3]See Overman, Redding and Venables for an overview [14].

[4]This assumes that a consumer consumes equal amounts of different varieties from the same region, which implicitly requires the price to be the same for each variety from that region.

[5]These are paired down from an extensive list of 50,000 points in Japan made available by NIMA at http://164.214.2.59/gns/html/cntry_files.html.

[6]The extension is "Distance Matrix of Point Features" by Hanna Maoh.

[7]Final market access does not vary across industries over time, $\Delta \ln FMA_{ijt} = \Delta \ln \alpha_i + \Delta \ln \sum_{j' \neq j} d_{jj'}^{-1} Y_{j't} = \Delta \ln \sum_{j' \neq j} d_{jj'}^{-1} Y_{j't}$.

[8]Although the impact of transportation costs should be working through SA, IMA and FMA.

[9]This regression excludes the change in FMA because it is perfectly collinear with the region dummies since the change in FMA does not vary across industries.

[10]In using the Tobit model I am assuming the underlying process that describes a firms decision to produce or not is the same as deciding how much to produce.

[11]One solution would be to implement a semiparametric method similar to Honore 10.