Advances in Economic Analysis & Policy

Volume 3, *Issue* 1 2003 *Article* 5

Spillovers from Local Market Human Capital and the Spatial Distribution of Productivity in Malaysia

Timothy G. Conley* Fredrick Flyer† Grace R. Tsiang‡

*University of Chicago, GSB, tconley1@gsb.uchicago.edu
†Lexecon, Inc., fflyer@lexecon.com
‡University of Chicago, gtsiang@midway.uchicago.edu

Spillovers from Local Market Human Capital and the Spatial Distribution of Productivity in Malaysia

Timothy G. Conley, Fredrick Flyer, and Grace R. Tsiang

Abstract

This paper examines whether spillovers from local market human capital are important in explaining the distribution of productivity across Malaysia. We develop an empirical method for describing local human capital distributions based on the idea that spillovers are limited in scope by costs of interaction or economic distance between agents. We use estimates of the economic distance between agents to construct measures of local market human capital based on schooling rates of the population within a given radius. These measures are then used in estimating equations obtained from a simple local public goods model. Our regressions are estimated using spatial GMM, allowing for general spatial correlation across observations as a function of economic distance. We find positive wage and rent differentials associated with local human capital, evidence consistent with productive human capital spillovers. Our results for rent differentials obtain with two distinct human capital measures; however, those for wage differentials depend on the human capital measure used.

KEYWORDS: Human Capital Spillovers, Local Public Goods, Spatial GMM

1. Introduction

Economists have long recognized that local market factors can affect productivity. An individual's neighbors, her local market, affect her productivity in several ways. They decrease her transaction costs and increase her returns to specialization by their proximity (Smith [32]). Neighbors can also be the source of external effects upon production and such agglomeration externalities are fundamental to theories of cities.[1] Among these agglomeration externalities, human capital spillovers are particularly interesting because of the prominent role they play in theories of economic growth (Romer [30], Lucas [20]). These spillovers can arise due to exchanges of ideas between agents and may be local— limited in scope by costs of interaction.[2] This presents the possibility that variation in local markets can be used to empirically identify the scope of these spillovers within an economy. In this paper, we investigate whether spillovers from local market human capital are important in explaining the distribution of productivity across Malaysia.

There are several empirical problems in estimating the effects of local market variables. Obviously, an empirical definition of what is local is required. Administrative regions are often used as local markets.[3] When the local market variables of interest are connected to governments, e.g. tax rates, these regions are the appropriate local markets.[4] However, we argue that administrative regions are often inappropriate for measuring human capital spillover effects. For example, external effects of human capital can arise from the sharing of knowledge between agents during formal and informal interaction. When costs of such interactions limit the extent to which agents can share knowledge, these spillovers will be local. Therefore, individuals' local markets will consist of those agents who are close enough for the costs of interaction with them to be sufficiently low. This set of agents will generally not correspond to an administrative region and may be different for each individual. Another difficulty arises from the endogeneity of location for firms and individuals. An endogenous local market variable like human capital

[1]See Anas, Arnott and Small [1] for a survey of this extensive literature.

[2]Examples of models with local human capital spillovers include Jovanovic and Rob [17], Benabou [4], and Lucas [21]. Relevant development applications about local learning include Besley and Case [5], Foster and Rosezweig[12], Munshi [26], Bandiera and Rasul [2] and Conley and Udry [11].

[3]Examples include Topel [35], Moulton [25], and Gyourko and Tracy [15]. Combinations of administrative districts and sectors have also been used, e.g. the city-industry pairs in Glaeser et.al.[13].

[4]As in, for example, Gyourko and Tracy [15].

may both drive and be a response to cross sectional productivity differentials. This simultaneity complicates the interpretation of observed correlations between local human capital and productivity. Finally, local human capital can provide amenity value as well as productive value. Therefore measures of productivity like wages will reflect compensating differentials for this amenity as well as productive effects of local human capital. The main contribution of this paper is the presentation of an empirical method designed to mitigate these difficulties.

The first step of our approach is to define and estimate a measure of the scope of an individual's local market. We assume that there is a metric that reflects transaction and information costs between agents, an *economic distance* metric. This economic distance is what limits the spillover effects of human capital and makes it a local public good. We use a primary component of transaction and information costs as our economic distance metric: travel time between locations. Measurements of travel times between a sample of locations are available from a household-level survey, allowing estimation of travel time cost between all locations. These costs are estimated as a function of population density and physical distance along the likely path connecting sample locations. Travel time estimates between locations throughout Malaysia are then constructed using a path network that approximates actual road networks. Economic distance between locations is taken to be the minimum time-cost connecting path on this network.

Local market human capital for an individual is defined using this economic distance. We can estimate the number of potential trading partners within, say, a two-hour trip for an agent located anywhere in Malaysia. Data on education and ethnic composition allow estimates of the characteristics of this population within two hours travel as well. By using our economic distance metric rather than administrative regions to define local human capital, we hope to more fully characterize the human capital of the set of neighbors interacting with each agent.

Defining local markets with an economic distance metric also allows us to vary the definition of local markets in order to examine the nature of spillover effects. We can estimate the relevant radius for a given effect: how close do you have to be to your neighbor to be influenced by spillovers? Furthermore, we can use measures of local market human capital within different radii to help isolate productivity or amenity effects. If the relevant radius for amenity effects is smaller than that for productive effects, we can identify a portion of the productive effect. For example, suppose you get amenity value from living within 30 minutes of people with high human capital but not those farther away; however, you get a productive benefit from being within 90 minutes of such people because you can occasionally interact and learn from them. In this case, local market human capital within 30 minutes will have both amenity and productive value while that between 30 minutes and 90 minutes away will have only a productive value. Therefore, this intermediate-range human capital can be viewed as a public good with only productive effects, yielding unambiguous predictions for wage and rent differentials.

Segmentation of regions can also be useful in handling some endogeneity problems due to unobserved productivity shocks attracting human capital. When productivity shocks have limited range, sorting in response to such shocks will tend to be reflected in human capital concentrations within a small radius of a location. Thus, human capital within say 30 minutes can be used as a control variable to mitigate this endogeneity problem in evaluating the impact of, e.g., human capital between 30 and 90 minutes away.

After constructing a measure of the local market human capital that is relevant for each agent, we follow Rauch [27] in using the local public goods model of Roback [28] and Rosen [29] to generate testable implications for wage and rent differentials. We estimate these differentials via wage and rent regressions containing our constructed measures of local market human capital along with individual and dwelling-specific controls. A primary concern for inference in these regressions is that there will be unobserved local market variables. This will result in unobservables being correlated for agents who are close to one another. Therefore, we use spatial GMM to estimate these regressions and conduct inference allowing for unobservable local market effects that are correlated across agents using a nonparametric covariance matrix estimation method (see Conley [10]). Thus, our regression unobservables are allowed to have a general correlation structure as a function of economic distance.

Our estimation method is a departure from those used in previous studies of local market variables. Some studies allow spatial correlation in individuals' unobservables with strict assumptions on the correlation structure, such as the equicorrelation-within-region error structure allowed by Moulton [25] and Rauch [27]. A few employ more elaborate parametric models of correlation structure such as the spatial autocorrelation model based on administrative districts used by Case [7]. These methods rely on exact measures of analogs of agents' economic distances, and are misspecified and inconsistent without them. In this application, it is certain that economic distances will not be perfectly measured and so it is important to conduct inference in a way that is robust to distance measurement errors. By using the nonparametric covariance estimation method in Conley [10], we obtain estimates that are consistent for general correlation structures and robust to measurement errors in economic distances.

The evidence from our investigation suggests local human capital provides positive productive spillovers. We find positive wage and rent differentials associated with local human capital, evidence consistent with productive human capital spillovers. While there is variation in the size of these estimated effects, their magnitude is consistently nontrivial. These productive effects have a range up to approximately 90 minutes away. These results display a modest amount of robustness to instruments and specific measures of human capital. Our results for rent differentials obtain when human capital is measured as either percentages or numbers of individuals with some secondary education within various economic distances. However, those for wage differentials obtain only for human capital measured using percentages of persons with some secondary education within a given radius.

The remainder of the paper is organized as follows. Section two contains a brief discussion of the suitability of Malaysia for a study of local market effects as well as a description of the data available. The third section describes our local public goods model and estimating equations. The fourth section describes the creation of an economic distance measure for Malaysia and construction of local market variables. The fifth section contains our regression results and a discussion of alternative specifications and interpretations. The sixth section briefly concludes.

2. Data

Several factors make peninsular Malaysia a good location for studying the effects of local market human capital. The greater reliance on face-to-face interactions in developing countries makes travel time a more important restriction on the extent of local markets. Characterizing an agent's local market would be more difficult in a industrialized country where better communication infrastructure enables transactions via a variety of channels. Among developing countries, Malaysia is a good choice due to its high level of migration across locations, implying that moving costs are not prohibitively high. Moreover, there has not been a massive rural to urban migration often present in developing countries. The spatial distribution of population plausibly reflects equilibrium conditions. Thus, a simple public goods model with costless migration can be useful in interpreting the data. In addition, individuals generally live near their workplace enabling us to get reasonable measures of the local human capital relevant for a given individual's workplace knowing only the location of their residences. Details of these advantages along with a description of the data are provided below.

The Second Malaysian Family Life Study (MFLS-2) and the Malaysian census provide good sources of data. In particular, the MLFS-2 provides data on travel times that enable us to construct economic distances between locations. The MFLS-2, is a population-weighted survey of households containing women under age 50, conducted in 1988-89.[5] Our wage, age, and schooling attainment information data are for men in these households. Households report rents paid and characteristics of residences. The MFLS-2 also provides information about the Census Enumeration Blocks (EB) containing each household. These data include travel times to nearest towns and anecdotal evidence on community characteristics. We also use data for 1100 mukims (subdistricts) on schooling and sectoral labor forces from the 1980 Malaysian Census. These data sources lead to measures of wages, rents, and local human capital that have substantial variation across the country. Summary statistics and precise variable definitions are contained in the data appendix.

There is a potential problem in specifying local human capital that is relevant for production. The MFLS-2 does not contain information on the location of individuals' workplaces. As a result we use human capital near an individual's residence as a proxy for that affecting his productivity. In a country where long commutes were common, this could lead to poor measures of local human capital. Fortunately, the community level MFLS-2 data contain substantial anecdotal evidence that the great majority of individuals do in fact reside very close to where they work, roughly within ten miles. Among this evidence is that agricultural capital-labor ratios are much higher for EBs within twelve miles of towns with large firms than for those immediately adjacent. These differences in the mix of farm inputs, for areas that are just a few miles apart, suggests that workers have limited access to employment opportunities that are more than twelve miles away.

Worker mobility is a key assumption in the local public goods model we use and the large degree of worker and job mobility in Malaysia has been well documented (e.g., Lee [19], Smith [33], and Smith and Thomas [34]). Although commuting costs seem to limit local job opportunities, moves are possible, as evidenced by the high level of both geographic and job changes. Patterns in long term moves more closely resemble those in developed, industrialized countries than developing countries. For example, there has not been a massive population flow from the

[5] The population weighting was accomplished by taking an independent random sample from the Statistics Department's sampling frame of 26,000 Census Enumeration Blocks (each made up of 100 living quarters.) Multiple households were then randomly sampled in each chosen EB.

rural to urban areas over recent decades.[6] The 1980 Census shows that among Malays, rural-to-rural migration accounts for the largest portion of moves in the 1970's and that rural-to-urban moves roughly equal the number of reported urban-to-rural moves. For the predominantly urban Chinese ethnic group, urban-to-urban migration is the largest category with some net migration from urban-to-rural areas (Cho [9]).

3. Model and Empirical Specifications

The impact of local market variables on the distribution of productivity is examined by estimating reduced-form wage and rent regressions based on the framework developed in the quality of life literature by Rosen [29] and Roback[28]. As in Rauch [27], wage and rent differentials are used to investigate the effect of local human capital on productivity. In the model, individuals consume local land l_{ij}, composite good x_i, and amenity A_j from a local public good, where i refers to the individual and j refers to location. For simplicity, the set of location choices is assumed to be finite and correspond to a set of non-overlapping markets where individuals live and work. We refer to these markets as input markets. The individual can only work and live in one input market, but faces zero cost in choosing this location. Firms can also costlessly locate anywhere but are restricted to hire only local labor and land inputs. As in Rauch [27], the local public good can also impact productivity via a Hicks neutral shift of the production function, denoted s_j.

The impact of the local public good on amenities and productivity is allowed to extend beyond input markets. The effects s_j and A_j may reflect the influence of public goods in locations that are a relatively small economic distance from input market j. We use the term spillover region to refer to the area containing local public goods that contribute to these effects. The size of the spillover region for local amenity and local productivity effect could be the same, but is not restricted to be in the model. For example, human capital endowments from input markets that are an hour away from j might impact productivity s_j, while the level of the

[6] One of the reasons for the relatively small flow of population from rural to urban areas is that government policies are not geographically neutral. Government spending encourages Malays to stay in rural areas. Meerman [24] has shown that public expenditures on agricultural assistance and rural resettlement programs have benefited rural Malays the most. The amount spent on rural development (by agencies such as the Federal Land Development Agency) is more than half of the total education expenditures.

local amenity A_j may depend on only human capital within a much smaller area.

Individuals' utility functions are assumed to be identical and homogeneous of degree one, denoted as:

$$U_{ij} = U(l_{ij}, x_i\,; A_j). \tag{3.1}$$

Ignoring initial wealth endowments and income from non-labor sources, an individual with one unit of human capital has budget constraint; $w_j = l_{ij}\,r_j + x_i$, where r_j and w_j are the rental price for a unit of land and wage per human capital unit in location j. Substituting the constraint into the objective function gives: $U_{ij} = U(l_{ij}\,, w_j - l_{ij}\,r_j\,; A_j)$ and first order condition $U_1 = r_j U_2$. These preferences imply: $l_i = \frac{(\alpha-1)w}{\alpha\, r}$ and $x_i = \frac{w}{\alpha}$, where $\frac{1}{\alpha}$ is the share of income allocated to the composite good, with $\infty > \alpha > 1$. The indirect utility function is:

$$V(r, w\,; A_j) = U(\frac{(\alpha-1)w}{\alpha\, r}, \frac{w}{\alpha}\,; A_j) = \frac{w}{\alpha}U(\frac{(\alpha-1)}{r}, 1\,; A_j). \tag{3.2}$$

The above utility structure implies that relative preferences for location are not affected by human capital endowments, since differences in endowments proportionally transform the indirect utility function. At equilibrium, all individuals must be indifferent to the various populated locations, giving:

$$w_j U(\frac{(\alpha-1)}{r_j}, 1\,; A_j) = w_k U(\frac{(\alpha-1)}{r_k}, 1\,; A_k)\ \ \forall\ j,\, k. \tag{3.3}$$

Note that conditional on the amenity, the above equation indicates that $w_j > w_k \Rightarrow \frac{w_j}{r_j} < \frac{w_k}{r_k}$, since $\frac{1}{\alpha} > 0$ and the price of the composite good is constant across locations. In other words, since individuals consume a mix of housing and the composite good in equilibrium, the disparity in rents across locations will be greater than that in wages.

Firms are assumed to have identical, constant returns to scale technology for producing the composite good x. Human capital and land are the direct inputs in production. Firms can costlessly locate anywhere, but are restricted to hire only local labor and land inputs. The unit cost function is:

$$C(r_j, w_j\,; s_j) = \min_{\{l_j\ ,\ h_j\}} w_j\,h_j + r_j\,l_j \qquad \text{s.t. } s_j\,f\,(l_j, h_j) = 1\ , \tag{3.4}$$

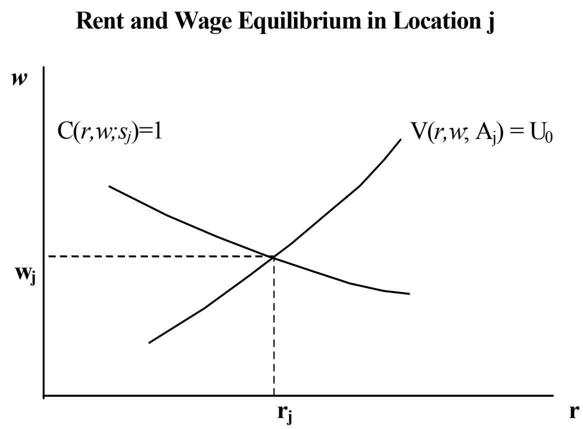

Figure 3.1:

where s_j is the shift parameter from the productive spillover, h is human capital. Given that firms can costlessly locate anywhere and that the price of the composite good is uniform across regions, at equilibrium, all firms must be indifferent to the various populated locations as unit costs are equated.

Location specific wages and rents are determined by the above conditions of spatial equilibrium for consumers and firms. This equilibrium is illustrated in Figure 3.1. The locus of points satisfying the condition $C(r_j, w_j\,; s_j) = 1$ define rent-wage combinations where firms are indifferent to producing in location j versus the alternative locations. It is necessarily downward sloping, as higher unit cost of one local input must be offset by lower unit cost for the other local input, if production costs are to be held constant. Similarly, spatial equilibrium for consumers requires that the location's rent-wage combination gives an indirect utility level per human capital unit, denoted U_0, that is common across locations (see equation (3.3)). The intersection between these two curves determines the location's rent and wage levels.

It is important to note that while the above spatial equilibrium depends on the assumption of costless migration, this assumption need not contradict our emphasis on using travel time costs to define economic distances in the empirical work. There is obviously a cost to relocation in the real world. However, relocation costs may be small enough relative to increases in discounted life cycle earnings from moves for this to be a useful model. There is a substantial amount of intra-urban and intra-rural migration in Malaysia as described above, so it is clear that moving costs are not prohibitively high. We argue that the use of travel time costs to define market and spillover region boundaries is consistent with negligible moving costs. Small single-trip costs could be paid infrequently by movers yet still accumulate to a significant amount when repeatedly paid by commuters.

In the regressions below we focus on local market human capital as our local public good, following Rauch [27]. We estimate reduced-form rent and wage regressions wherein we try to estimate equilibrium wage and rent differentials associated with human capital spillovers. The main measure of local market human capital we use is an estimate of the fraction of the population within a given region having some secondary education. Specifially, we use the fraction that have completed what is called lower secondary education, approximately 9 years of school (completion of form 3, see appendix for details).[7] Our primary motivation for using percentages of people with some secondary education is that we feel it is less likely to suffer endogeneity problems than measures of, say, the number of people with a given education level which is directly endogenous in our model. The use of education percentages instead of average educational levels is motivated by the difficulty of categorizing the several different types of education paths. That is, a meaningful average education is difficult to calculate. It also seems reasonably simple to interpret as a measure of the probability of meeting a 'high human capital' worker in a chance encounter. However, we acknowledge that our measure is *ad hoc* and, while a systematic study of alternate measures of human capital is beyond the scope of this paper, we investigate the robustness of our results to the use of one potential alternate: the level of people with some secondary education.

Reduced Form Equations Following Rauch [27] and the quality of life literature, e.g. Rosen [29], Roback[28], Blomquist et. al.[6], we characterize the variation in equlibrium wages and rents across locations by estimating reduced-form hedonic wage and rent equations. Our baseline log wage regression is:

$$\ln w_i = X_i'\beta_w + S_i'\delta_w + e_i \ . \tag{3.5}$$

The wage for individual i, denoted w_i, is the reported rate of payment for the primary employment activity, converted to a weekly rate. The vector X_i contains a

[7] We use completion of form 3 with or without completion of the lower secondary examination to define this regressor. See data appendix for details.

constant and variables used to control for differences in human capital endowments and certain local market characteristics. It includes years of schooling, an indicator for having some post-secondary education (beyond 11 years, see appendix for details), years of potential work experience (defined as age less years of school minus six) and its square, a Chinese ethnicity dummy variable[8], and a dummy variable for the ability to read English. In addition, X_i contains an estimate of the travel time to a major city/port[9] and an indicator variable for residence in the three largest urban areas.[10] These variables are meant to describe the infrastructure present in an individual's workplace. The error term is denoted e_i.

The vector S_i contains proxies for local market human capital. This includes human capital for both the input market inhabited by individual i and relevant spillover regions for this input market. Our proxy for the human capital within a given region is the percentage of individuals with some secondary education. This percentage is calculated with 1980 census data rather than MFLS2 data because of its superior coverage of areas outside major cities and advantages in terms of mitigating endogeneity concerns that are discussed below. S_i contains human capital measures for several regions that are defined using our measures of economic distance from an individual's residence. Human capital measures for the relevant local input market for individual i are captured by calculating percentages using the population within a small economic distance. Of course, the input market is a subset of the spillover region and so this region also serves as the first of the partitions of our spillover region. In the results reported below we use 30 units, roughly 30 minutes of travel time, for these input market measures.[11](See the following Section for a discussion of economic distance construction.) The remaining partitions of the spillover regions are defined using larger radii from the individual. We divide the area that is too far to be in the local input market but still close enough to be within the spillover region into rings. In the regressions presented below, we define these rings as the area between 30 units and various outer radii up to 120 units. We also include rings between 60 and 90 units and 90 to 120 units.

[8]There is also a large Indian minority group in Malaysia. We do not include an indicator for this group to conserve degrees of freedom as it does not appear to improve the fit of the wage regression. See footnote 20.

[9]The travel time to these big cities is, in view of our location resolution, equivalent with travel time to their associated ports.

[10]For the largest urban areas indicator, we used the three largest cities in terms of our sample's residences: Kuala Lumpur, Johor Bahru, and the Ipoh area.

[11]Although input markets do not overlap in the model, our proxies for them occasionally will.

Analogous reduced-form specifications are estimated for location-specific log rents:

$$\ln r_i = Z_i'\beta_r + S_i'\delta_r + \varepsilon_i \ . \tag{3.6}$$

Where r_i are rents paid on MFLS-2 household residences. The vector Z_i contains a constant and describes individual characteristics of dwelling i. It includes the number of sleeping rooms in the house (Nsleep), whether there is a flush toilet (Flush), and whether there is piped drinking water in the residence (Drink). The presence of any government development project in the EB (Anydev), the proportion of homes in the EB with piped drinking water (H_2O Proportion), the number of years since an EB was initially given electrical service (Elec. Years), and whether there was a flood disaster in the community within the last two years (Flood) are also in Z_i. In addition, Z_i contains time-to-city measures and an indicator for city residence. Only those who paid rent to non-relatives were used, this results in some EBs used in the wage regressions being omitted from the rent regressions. We also use some EBs for which we have rent but not wage information, so neither sample is a subset of the other.

We estimate these wage and rent regressions via instrumental variables because we are concerned that our human capital measures may be correlated with unobservables for several reasons. There is certainly measurement error in using this crude measure of local market human capital. We use population proportion of community service workers in 1980 for each of our regions as an instrument for secondary education percentage in that region. It should be correlated with human capital, since a large portion of community service workers are teachers. These instruments are in fact strongly correlated with the secondary education percentages.[12] We assume that measurement errors associated with using this measure of teachers in 1980 to proxy for local human capital in 1988 are not correlated with those arising from using 1980 secondary education percentages so this is a valid instrument. We investigated the sensitivity of our results to using IV rather than OLS and obtained almost equivalent results using OLS, as discussed in Section 5.2.

Potential endogeneity problems with local market human capital are of more concern. As our controls for local infrastructure are limited to time to city/port

[12]Regressions of each of our secondary education percentages upon our community service workers instruments have r-squareds above 76% and F-statistics (for the joint hypothesis of zero slopes) that are above 300.

measures and urban area dummies, there are likely to be unobserved productivity enhancing variables. These unobserved productive effects might have disproportionately affected high human capital workers, motivating their migration to these areas. To our knowledge, the leading candidate for unobserved infrastructure investments that might have motivated such selective migration is government-funded infrastructure projects implemented during the 1980s under the 4th and 5th 'Malaysia Plans.' Public investment under these plans was substantially more than that in the previous decade (total spending was approximately 4 times that in the 1970s, Cho[9]). In contrast to many countries, infrastructure expenditures did not tend to favor previously high human capital areas. Instead, expenditures on housing and roads targeted rural areas in efforts to balance employment and economic activity levels throughout Malaysia. Therefore, unobserved infrastructure implemented during this decade tends to disproportionately favor what were low human capital areas in 1980. The implicit measurement error from using 1980 human capital variables and instrument to proxy for those in 1988 is likely to be either uncorrelated or perhaps even negatively correlated with S_i in the cross section. Thus, if there is a bias due to these omitted 1980s infrastructure components, it will tend to reduce our estimated human capital effects. Therefore, we argue it is plausible that our results are not driven by endogeneous migration due to the leading suspects for omitted infrastructure components.

An additional argument against this type of endogeneity driving our results is that sorting in response to unobserved types of infrastructure that are limited in range will tend to be reflected primarily in input market human capital levels. Many candidate types of unobserved infrastructure like schools, housing developments, municipal services like water treatment are limited in range. Thus, we are able to use human capital percentages in input markets to control for sorting in response to such shocks and focus on estimated effects of human capital in spillover regions beyond the input markets. As for infrastructure of extensive range, like a highway network, it may be well-captured by our time-to-city/port and urban-area regressors and therefore not generate omitted variable bias. If some aspect of extensive infrastructure's effect on wages or rent does remain omitted, then as described above it ought not to be positively correlated with the 1980 local human capital measures due to the redistributive goals of Malaysian public expenditure of the 1980s that favored areas that had low human capital in 1980. We further discuss endogeneity issues in Section 5.3 where we discuss alternative explanations for our results.

Our method of dividing up spillover regions provides the potential to use both wage and rent differentials to identify productive effects of human capital even if it also has amenity effects. Equilibrium wage and rent differentials associated with local public goods are determined by wage rent indifference sets for both firms and workers. If productive and amenity effects operate within the same spillover region, then indifference sets for both workers and firms will be shifted to the right in Figure 3.1. Therefore, the sign of wage differentials is ambiguous, as is the case in typical local public goods models where local public goods with productive and amenity value are not consumable across the disjoint input markets. However, if the relevant spillover region is different for amenities and productivity enhancements, then we will be able to isolate the effect with the larger radius. Suppose human capital endowments of the region 30 to 90 units away provide productivity spillovers, but not utility-enhancing amenities, while human capital within 30 units provides both. Productivity spillovers can then be disentangled from amenities by looking at the impact upon wages and rents of human capital 30 to 90 units away. The model predicts positive wage and rent differentials will arise from a pure productive effect and this can be investigated in coefficient estimates on outer regions' human capital proxies.

When interpreting our empirical results, we assume that the radius of the spillover region for amenity effects of human capital is smaller than that for its productive effects. Our motivation for this assumption is that we think that the majority of amenity effects from human capital stem from repeated interactions of daily life near high human capital neighbors. In contrast, we think productive effects tend to arise from informal knowledge transfers that do not rely as heavily on very high frequency contacts. Therefore, we expect travel costs will be a larger component of the costs of interaction limiting amenity consumption than productivity consumption of human capital.

Estimation and Inference with Spatial Correlation It is important to use an estimation method that allows for spatial correlation. Local market variables should be highly correlated for observations with small economic distances between them and our construction method will result in observable measures of spillover region and input market variables that are highly correlated for nearby observations. There are likely to be local market variables that are left out of these sets of regressors and end up in the unobservable terms. So we expect that the error terms in the estimating equations will be correlated across individuals, according to economic distance. Therefore, we use the model of cross sectional dependence and spatial GMM estimation method detailed in Conley [10] to allow for residuals from our regressions to be correlated across observations in a general manner characterized by economic distance.

We estimate our wage and rent regressions with instrumental variables, a special case of GMM. For example, consider estimation of the wage regression in 3.5 via instrumental variables with instruments V_i. Our point estimator is a just-identified GMM estimator using the moment condition

$$E\{V_i e_i\} = 0.$$

Note that this expectation is with respect to the marginal distribution of the data and it is a just-identifed system since there the same number of instruments as parameters. Our estimator of the marginal distribution of the data is unchanged by the presence of spatial correlation, it is the empirical distribution of the data just as when the data are independent. Therefore, when we use the sample analog of this just-identified moment condition to obtain estimates, we get exactly the same estimates as if the data were independent. However, spatial correlation must be accounted for in conducting inference for these estimates.

To conduct inference, we use nonparametric covariance matrix estimators presented in Conley [10]. A consistent estimator of the asymptotic covariance matrix of the IV estimator of $[\beta_w, \delta_w]'$, Ω, is easily constructed using a weighted average of cross-products of instruments and residuals. Letting the regressors be denoted by $\phi_i \equiv [\,X_i', \; S_i]'$, $\hat{e}_i$ denote the residual for observation i, and D_{ik} denote the measured distance between agents i and k, a consistent estimator of Ω can be defined as:

$$\hat{\Omega} \equiv \left[\frac{1}{N}\sum_{i=1}^{N} V_i\phi_i'\right]^{-1} \left[\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N} W_N[D_{ik}]V_iV_k'\hat{e}_i\hat{e}_k\right] \left[\frac{1}{N}\sum_{i=1}^{N} V_i\phi_i'\right]^{-1\prime}$$

Where $W_N[D_{ik}]$ is a weighting function that will converge to one for all distances as the sample size grows. For simplicity, we use a weighting function that is one for $D_{ik} = 0$, declines linearly until D_{ik} equals a cutoff distance L, and is zero for distances greater than L. This estimator is consistent if L is allowed to grow with sample size at an appropriate rate.[13] In practice, we take the truncation point L to be the largest economic distance radius used to define a spillover region. Since our constructed local market variables are functions of spatial moving averages,

[13]We note that for this W_N, $\hat{\Omega}$ is not guaranteed to be positive definite; though it turned out to be positive definite for all our estimated specifications. W_N can be chosen to insure that $\hat{\Omega}$ is positive definite with only a slight increase in complexity, see Conley [10] for details.

we allow $V_i\hat{e}_i$ to have as much dependence as the broadest local market variable. This nonparametric covariance estimation method remains consistent even when used with imperfectly measured economic distances like the estimated travel times used here. These imperfectly measured distances would prohibit consistent estimation of parametric models of error correlation such those used by Moulton[25], Rauch[27], and Case [7] mentioned above.

4. Construction of Economic Distances

This section describes how we construct the estimates of economic distance used to define measures of local human capital. The basis for our estimates is data on travel times for a sample of 677 trips in the MFLS-2 survey.[14] Using this data, we estimate the time cost of travel as a function of physical distance and population density (a proxy for congestion and infrastructure quality). We use estimates of this function of physical distance and population density to estimate minimum travel times between locations. We use these minimum travel times to construct a map of Malaysia (illustrated in Figure 4.4) where towns' positions reflect the estimated minimum travel time between them. Our economic distance between any pair of locations corresponds to their distance on this map, reflecting a minimum travel time.

The first step in constructing a measure of economic distance is the selection of a network that is used to approximate paths of travel across Malaysia. The nodes of this network include EB locations and the 104 cities and towns for which travel times from EBs are known.[15] The path connections illustrated in Figure 4.1 approximate the actual road network of Malaysia and capture its principal features—coastal highways, extensive grids around cities, and limited cross-country routes.

The travel time data on trips from each EB to cities are then used to estimate travel time as a function of distance of the path traveled and a measure of population concentration along that path. This estimated relationship allows the construction of travel times between all network nodes. We take as our raw measure of economic distance between nodes the minimum travel time along any path connecting the two nodes. The statistical model we use assumes that the economic distances between agents can be represented in a Euclidean space, therefore, we

[14]The MFLS-2 contains data on travel times from each EB to the nearest small, medium, and large town as reported by community leaders in the EB.

[15]We use travel times to medium and large towns.

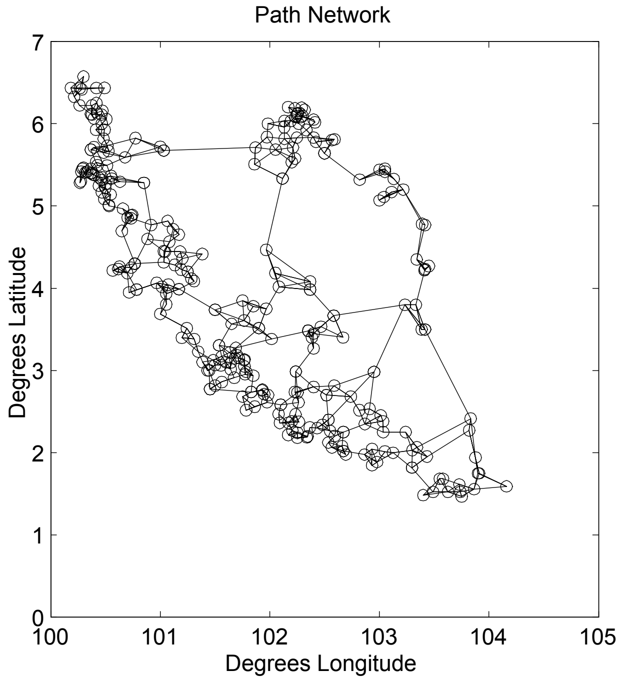

Figure 4.1: Path network approximating the actual road network in Malaysia.

use classical multidimensional scaling to obtain a configuration on the plane whose interpoint distances are close (in a sense made precise below) to the raw economic distances obtained by travel time minimization. We take this representation of travel times as interpoint distances of a configuration on a plane as our measure of economic distance. The use of this planar representation also allows us to create pictures that enhance understanding of the estimated structure of neighborhoods.

Population concentration along paths We use mukim level census data to create a measure of population concentration along paths between network nodes. Mukim population density is identified with the center of each mukim. This allows estimation of a population density number for points along paths as a weighted average of mukim centers' population density. Since Malaysia is a small country near the equator we use latitude and longitude as approximate Cartesian coordinates. Our weights decline linearly with physical distance with a maximum distance of 1/2 degree (about 56 km) and are zero thereafter. The estimated population density surface is depicted in Figure 4.2. The concentration of population in port cities and expanses of uninhabited jungle in the interior of the peninsula are evident in this figure.

The major cites lying in areas corresponding to the largest concentrations of population are labelled in Figure 4.3 for later comparison. The city of Johor Bharu (labelled J in Figure 4.3) is the eastern most spike of the surface in Figure 4.2. The highest concentration of people can be seen in the sharp spike about the city of Melaka (M in Figure 4.3). Kuala Lumpur is the major city associated with the bump to the northwest of Melaka (K in Figure 4.3). The city of Georgetown (G in Figure 4.3) and, further inland, the city of Ipoh (I in Figure 4.3) are in the large concentration of population in the northwest corner of the country. The eastern coast has large population concentrations around the city of Kota Bharu in the extreme northeast and Kuala Terengganu in east central region (respectively B and T in Figure 4.3).

Paths between nodes are broken down into segments approximately 5 km in length.[16] The population density along each segment was estimated by the average of the heights of the population surface in Figure 4.2 above each of the segment's endpoints. This results in a measure of population concentration along each segment of all the network paths. These segments were grouped into seven categories based upon this measure of population concentration. The cutoff points for categories are depicted as the vertical axis tick marks in Figure 4.2, namely (100,250,325,400,475,550).

Estimation of travel time function The data on travel time from EBs to the nearest cities and towns allow estimation of the time it takes to travel across each density category segment. The total time cost of each trip is assumed to be the sum of a fixed transactions cost, the number of segments in each density category multiplied by the time cost per segment, and an IID measurement error:

$$c_i = \tau_0 + \tau_1 N_i^1 + \tau_2 N_i^2 + ... + \tau_7 N_i^7 + \eta_i \tag{4.1}$$

where τ_j is the time cost of traversing a segment in density category j and N_i^j

[16]This division was into .051 degree segments with of course some remainder segments of shorter length. We treat remainder segments in the same fashion as the full-length ones. The sample paths used are long enough that we feel this source of measurement error is not a major concern in the estimation of equation 4.1. As detailed below, the imputed time costs of shorter paths are dominated by fixed costs so this source of error has a small impact there as well.

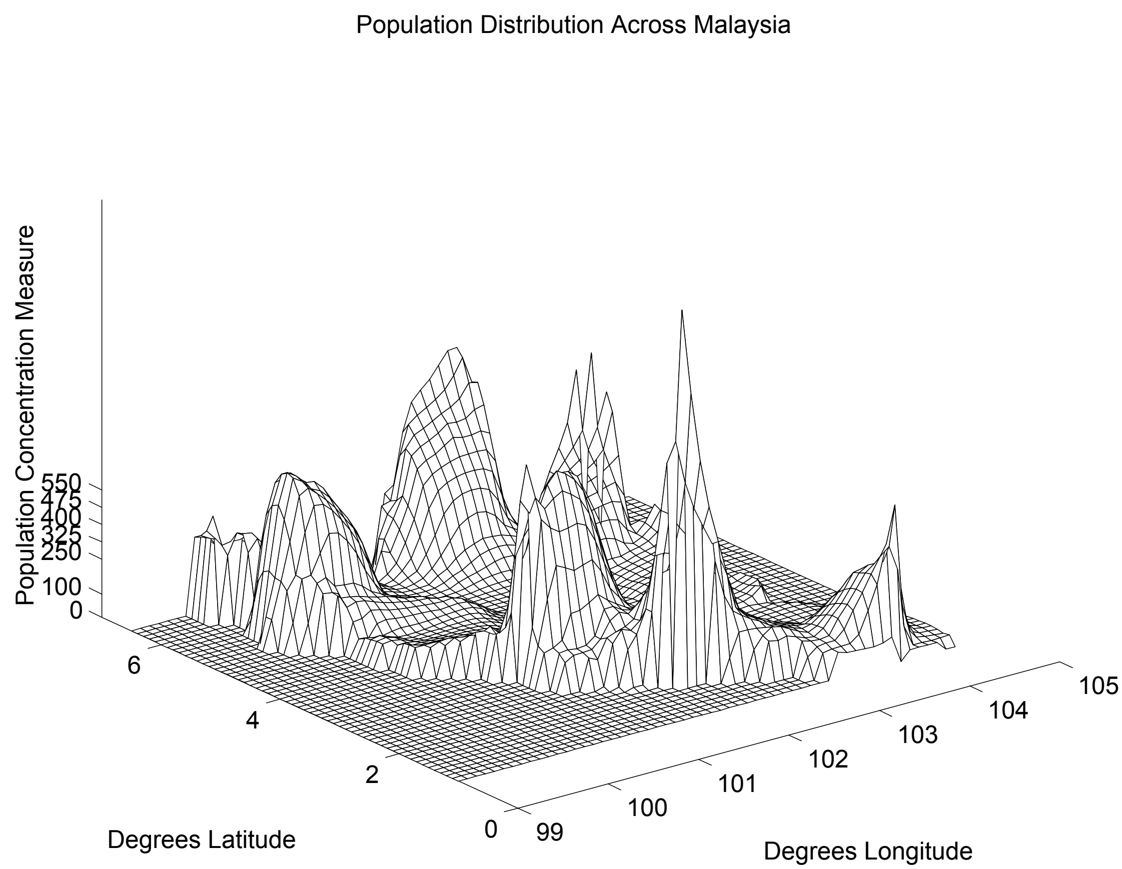

Figure 4.2: Population concentrations across Malaysia. The vertical axis is graded according to population categories used in segment classification.

is the number of density category j segments in trip i. We used a road map of Malaysia, to pick the path connecting each EB and nearest city for each trip i by hand, yielding imputed measurements of $\{N_i^1, ..., N_i^7\}$.

We estimate the time cost parameters τ_j in 4.1 by OLS. The estimated time costs for traversing segments in each population density category are presented in Table 4.1. The coefficients $\tau_1, ..., \tau_7$ are ordered in increasing population concentration. These estimates indicate that the most expensive segments to cross are those located just outside of heavily concentrated urban areas. The centers of metropolitan areas may well be equipped with better infrastructure than the surrounding 'suburban' sprawl. Thus congestion may be decreasing with population concentration at high levels. The decline in travel costs with lowering levels of population is suggestive of lessening congestion as population decreases. The very sparsely inhabited segments are estimated to be somewhat more expensive to travel than slightly more populous segments, which is not surprising because the least inhabited areas in Malaysia are often jungle.

The parameter estimates from equation 4.1, $\{\hat{\tau}_j\}$, enable us to estimate the time cost of travel between all grid nodes along all possible paths by $\hat{\tau}_0 + \sum_j \hat{\tau}_j N^j$ for the $\{N^j\}$ associated with each path. The raw economic distance measurement between any two grid points is then taken to be the time cost along the minimum estimated cost path between them.[17] Define the matrix R whose elements contain this raw economic distance between nodes.

Creating an economic distance map The raw economic distances between points in R will not necessarily be Euclidean and our statistical model (detailed in Conley [10]) assumes that economic distances can be represented in a Euclidean space. Furthermore, a graphical representation of relative locations will greatly aid our understanding of how this economic metric differs from physical distance. A map with distances defined using our economic metric will also be valuable in understanding how neighborhoods overlap. Therefore, we take as our final estimate of economic distance the distance matrix of a configuration of points in $\Re^2$ that has interpoint distances that best approximate our minimum estimated travel time costs between nodes. In order to calculate this configuration on the plane, we use classical multidimensional scaling (CMDS). A discussion of the CMDS procedure and a measure of the quality of fit of the constructed configuration is

[17] To calculate this minimum time path we use a shortest path algorithm generously provided by Boris Cherkassky, described in B.V. Cherkassky, Goldberg, and Radzik [8].

presented in the appendix.[18]

Economic distance estimates The fitted distances D from applying CMDS to R provide our measure of economic distance. The configuration of nodes based upon the economic distance D yields an economic configuration of points that differs from their physical configuration. The physical locations of 104 cities and towns are depicted in Figure 4.3. Contrast this with the fitted economic configuration of these cities and towns depicted in Figure 4.4. Fitted economic distances D are expressed on the map in the same scale as the raw time cost distances R because we want to retain the interpretation of these fitted distances as approximately reflecting travel time costs in minutes. So we can interpret 120 economic distance units as roughly 2 hours. These fitted distances capture the salient medium and large-scale features of the distances in R. However, there is distance information in R, particularly for short distances, that is not captured in the fitted distances.[19] This is largely because there are many close nodes connected by a small number of segments where the relatively large fixed cost estimate $\hat{\tau}_0$ dominates the economic distance estimate between them. These close nodes are effectively estimated to be equidistant and thus not easily representable in a low-dimensional Euclidean space.

The relative distances between major population centers are substantially altered when the metric of economic distance, reflected in D, is used. The most important alteration is that Kuala Lumpur is isolated with an economic distance metric due to the high cost of travel across its outskirts. With the economic distance metric, the areas around Georgetown and Ipoh appear closer to being one metropolitan area. The economic distance map also suggests that northwest coast population center around Kota Bharu may be as close to Melaka in economic distance as to Georgetown, contrary to their physical locations.

The estimates of economic distances in D will play two roles in our empirical work. First, we will define local market variables for a given individual as a weighted average of his neighbors' characteristics, defining neighbors using the metric in D. Thus our local market variables can be thought of as spatial moving averages in the space depicted in Figure 4.4. Figure 4.4 can be consulted to get an idea of what large and small travel-time neighborhoods look like in Malaysia. In addition, we will allow general spatial dependence in regression unobservables

[18] See Mardia et. al. [23] for a more complete description.

[19] The goodness of fit measures described in the appendix are 72% for the measure comparing squares of the first two eigenvalues to the rest and 30% for that using absolute values.

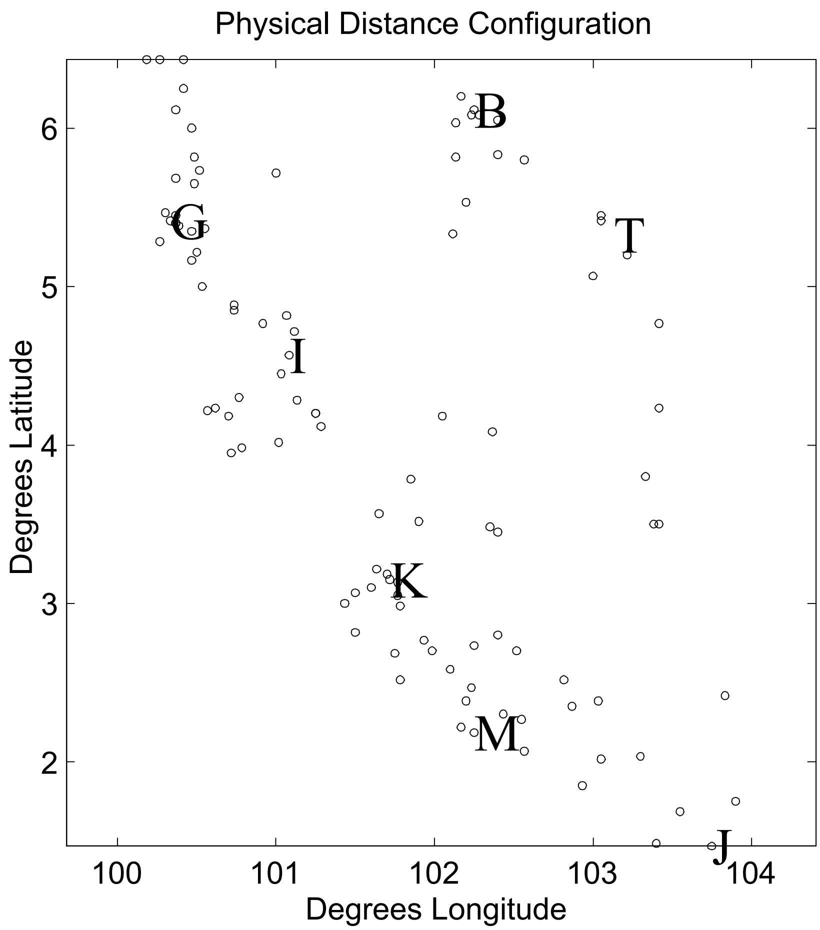

Figure 4.3: City and town physical configuration. Major cities are labelled: Georgetown—G, Ipoh—I, Johor Bharu—J, Kota Bharu—B, Kuala Lumpur—K, Kuala Terengganu—T, and Melaka—M.

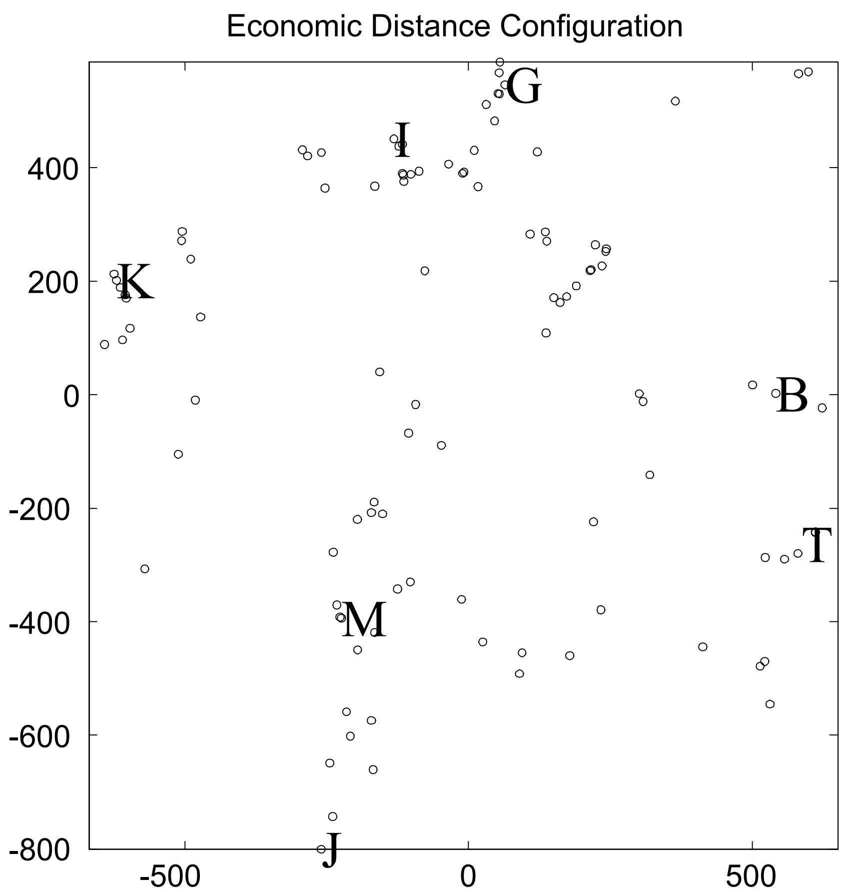

Figure 4.4: City and town economic distance configuration. Major cities are labelled: Georgetown—G, Ipoh—I, Johor Bharu—J, Kota Bharu—B, Kuala Lumpur—K, Kuala Terengganu—T, and Melaka—M

as a function of the economic distances in D, using the spatial GMM method in Conley [10].

5. Regression Results

This section presents our estimates of the regressions described in Section 3. We present estimates of our baseline specification in Subsection 5.1. Subsection 5.2 contains results designed to investigate the robustness of our baseline results to the use of alternate instruments, specifications, and measures of human capital. We discuss caveats and alternative explanations of our results in Subsection 5.3.

5.1. Baseline Regressions

Tables 5.1 and 5.2 contain the results from our wage and rent regressions, respectively. Each column of these tables corresponds to a different description of the spillover regions. In the specification reported in the first column, the spillover region is a single disk that ranges from zero to 30 units (roughly zero to 30 minutes travel time). The percentage of people within zero to 30 units that have some secondary education is labelled SEC% 0-30. The specification reported in the second column has a disk that is zero to 30 units and a ring that is between 30 and 90 units away from the individual. The percentage in the 30 to 90 ring is labelled SEC% 30-90. The third column has results from a specification that divides the spillover region into a zero to 30 disk and a 30 to 120 ring, with secondary education percentage for the latter region labelled SEC% 30-90. Finally, the fourth column's specification has a zero to 30 disk and rings from 30-60, 60-90, and 90-120, labelled correspondingly. Each regression also contains either individual or dwelling-specific controls a city indicator variable for residence in the three largest cities and an estimate of the travel time in minutes to the nearest of these cities. Standard errors that allow for spatial correlation are in parentheses. The weighting function used to obtain the covariance matrix $\hat{\Omega}$ declines linearly until the largest distance used in the local market human capital.

The wage regression coefficients on individual level variables in Table 5.1 are stable across specifications and have the expected signs and magnitudes. The estimated return to a year of education is about 6 percent across specifications. This is similar to returns estimated with U.S. data. The estimated coefficient on the some post-secondary education indicator implies a 17 to 18 percent wage premium in the different equations. A year of potential work experience is associated with roughly a five percent wage increase. Although the coefficient on the quadratic experience term is negative, its small magnitude implies a rather linear life-cycle wage profile. There is a large positive effect for Chinese ethnicity. It amounts to nearly an 80 percent wage premium.[20] Ability to read English has a 14 to 15 percent premium associated with it, depending on the specification.

There is little evidence that wages systematically vary by physical location, controlling for our other covariates. The point estimate on the city dummy is negative but imprecisely measured. The coefficient on the time-to-city variable has the expected negative sign, being farther away reduces predicted wages. The coefficents are not precisely estimated and point estimates are modest in size, being 100 units (roughly 100 minutes) away from the nearest city corresponds with predicted wages being 1% lower.

The wage regression coefficients on human capital measures are consistent with productive spillovers. Secondary schooling percentages have a positive relationship with wages across specifications. This effect is clearly significant in specifications {3} and {4}; and borderline significant in specification {2}. The strongest effects are for human capital measures in regions beyond 30 units and up to 120 units and in regions 90 to 120 units distance.

There is less evidence of a positive wage differential associated with human capital measures in input markets. Point estimates for 0 to 30 rings for all specifications and the 30 to 60 ring are positive but have large standard errors. This is not surprising if, as we argue above, there is an amenity value to nearby human capital that results in a negative differential that offsets a positive wage differential due to a productive spillover.

The rent regression coefficients on dwelling characteristics in Table 5.2 are fairly stable across specifications, although several coefficients are statistically insignificant. The presence of government developments and the number of years electrical service has been available in the area and the presence of a flush toilet in the dwelling have insignificant impacts on rents. Access to drinking water and number of sleeping rooms have large, significant impacts. Having had a flood has a positive and significant effect, perhaps reflecting age of rebuilt homes.

There is evidence that rents vary by physical location. The larger urban area

[20]In seperate specifications (available upon request) we also included dummy variables for Indian ethnicity, the other large minority ethnic group in Malaysia. The Indian effect was small and insignificant in all specifications and there were no qualitative changes in estimates of other coefficients. Therefore, in an effort to conserve degrees of freedom, our baseline specifications omit the Indian indicator.

indicator coefficients are significant though rather imprecisely measured across specifications but point estimates are always large and positive. Taken as a group these estimates imply that rents are systematically higher in these cities, but it is difficult to pin down how much higher. The time-to-city variable does not have the expected sign but it is not statistically significant in any specification.

There is evidence of a positive human capital effect in the rent regressions in Table 5.2, although it is more local than that for wages. All specifications find evidence of a significant, positive relationship between rents and secondary schooling percentages in the zero to 30 region. Specifications {2} and {4}also provide evidence that human capital effects for the 30-60 and 30-90 regions are significant. The impact of secondary education percentages for areas that are 90 to 120 units away are insignificant. Thus, the range of these human capital effects appears to be more limited than those in the wage regression.

The range of human capital effects in both wage and rent regressions appears limited to about 120 units. We found no evidence of significant effects in wage or rent regressions for distances beyond this distance. Coefficients for rings beyond 120 units in alternate specifications were insignificant and so they are not presented to conserve space.

Overall, our most persuasive evidence of productive spillovers is the siginificant, positive wage effects found for human capital percentages in the 30 to 120 units range and significant, positive rent effects for the 30 to 90 units range (areas with substantial overlap). These findings are consistent with human capital having amenity effects concentrated within 30 units and productive effects extending farther. In this scenario, our model predicts positive wage and rent differentials for human capital in regions beyond 30 units away, a prediction that is consistent with our empirical results.

5.2. Robustness Checks

We estimated several alternative specifications to examine the robustness of our baseline results to changes in instruments and measures of human capital. Sensitivity of our results to these instruments is examined by comparing our baseline results to those using the human capital regressors themselves as instruments. We use an alternative measure of local human capital in levels rather than percentages of people with some secondary schooling.

Table 5.3 presents estimates of wage and rent regressions using the human capital regressors themselves as instruments. The point estimates in this table are the same as OLS estimates but of course the standard errors are different as they allow for spatial correlation (and heteroskedasticity). The regressors in these specifications are the same as those in specification {3} and {2} from Tables 5.1 and 5.2, respectively. When all specifications in Tables 5.1 and 5.2 are estimated using OLS (with adjusted standard errors), our results are essentially unchanged. Local market human capital beyond 30 units appears useful in predicting wages and rents and that within 30 units helps predict rents. We do not report all specifications to conserve space and choose to present these two specifications illustrating our that our strongest results are robust: human capital between 30 and 120 units (the outer ring in specification {3}) seems most relevant for wage prediction and for rent prediction that between 30 and 90 units (the outer ring in specifcation {2}) appears most relevant.

Table 5.4 presents results from wage and rent regressions analogous to specification {4}, using an alternative measure of local market human capital. The alternative measure is the number of secondary graduates in a region. The number of secondary graduates, measured in 10,000s, for various regions are calculated just as we do above for percentages. Instruments were also constructed analogously as the number of community service workers. These results are representative of (unreported) estimates using the analogs of all specifications from Tables 5.1 and 5.2.

The results using levels of human capital are qualitatively similar in some respects to the previous findings. Human capital in the zero to 30 range and in a range beyond that (30 to 60) clearly has a positive, siginificant effect in the rent regression. This is essentially the same story as in our baseline specifications and thus our rent results appear robust to this change in human capital measure. In contrast, our wage regression results are not robust to this change. Point estimates for human capital in the zero to 30, 60 to 90, and 90 to 120 ranges are positive. However, these effects are not precisely estimated and not siginificant at conventional levels. Thus our overall conclusions about differentials are sensitive to the measure of human capital that we use.

5.3. Alternative Explanations

This section discusses alternative specifications and caveats in interpreting our results. There are likely to be many local public goods that influence wages and rents in our sample. An important shortcoming of our data is the lack of controls for these types of effects. We are limited to measurements of the travel time to the nearest big city/port and an indicator for residence in the largest urban areas. These controls may do an adequate job of controlling for local public goods like population congestion disamenities, port infrastructure advantages, or highway locations; however, there will certainly be other relevant local public goods they do not capture well. Likely examples include school districts, sanitation facilities, and climate or pollution (dis)amenities. Our use of spatial GMM allows general correlation across observations as a function of economic distance. Therefore, our inferences do account for regression error correlations induced by the impact of unobserved local public goods upon nearby observations in our data. However, the possibility exists that these unobserved local public goods are correlated with local human capital levels up to two hours away and the resulting omitted variable bias could be an alternative explanation for our results.

A prime alternative explanation for our results combines unobserved local public goods with a breakdown of our model's assumption that workers and firms can relocate with negligible cost. If workers face high costs of migration, then high human capital workers would have disproportionately larger incentive to move in response to positive productivity shocks. Thus, high human capital levels in productive areas might simply be a response to unobserved productivity-enhancing infrastructure rather than the cause of increased productivity.

The substantial literature on migration in Malaysia provides evidence that is relevant for this explanation. First, there is evidence of substantial mobility across demographic groups, favoring our argument that effective migration costs are small. However, there is also evidence that more highly educated workers have greater migration propensities, all else equal (see Smith and Thomas [34] and Lee [19]). This evidence of correlation between education and migration is consistent with the premise of this alternative explanation that high human capital workers are disproportionately able to move to take advantage of productivity-enhancing local public goods. Such selective migration along with an unobserved local public good could generate empirical results akin to ours. For example, consider an unobserved local public good with productivity-enhancing effects that extended beyond input markets into spillover regions coupled with amenity value that was more limited in range. The limited-range amenity would avoid generating obvious, large wage premia along with distinct rent differentials at short radii but both wage and rent differentials at longer radii. If in addition, the selective migration of high human capital workers into input markets containing this good was such that their equilibrium levels would not perfectly predict wage and especially rent differentials, outer spillover market educated worker rates would retain predictive power even with conditioning on input market rates.

As we have argued in Section 3, our results are unlikely to be generated by this alternative process if the unobserved local public goods motivating selective migration were implemented during the 1980s under the 4th and 5th Malaysia Plans or are limited in range to input markets. Our measures of local human capital dated from 1980 obviously do not directly include the effect of migrations due to these goods and are plausibly not correlated with migrations due to the 1980s investments because of these investments redistributive nature. Indeed, it is reasonable to expect that the selective targeting of these investments to areas where our measure indicates low human capital could result in a downward bias in our estimated wage and rent differentials associated with local human capital. Thus, it is unlikely that our positive partial correlations are spuriously driven by this cause. In addition, our decomposition of spillover regions mitigates the impact of selective migration in response to unobserved local public goods that are limited in range. Many productivity-enhancing unobservable are plausibly limited in scope to the input market, e.g. a new drinking water system benefiting only the town in which it exists. For such limited-range goods, selective migration would tend to result in the distribution of human capital across input markets reflecting accumulated productivity shocks in each input market. Controlling for input market human capital, that in outer spillover regions (beyond the input market) would not be significant predictor of wage and rent differentials. Thus, it is plausible that our finding of significant effects of human capital beyond input markets (that in regions 30 to 120 units away), while controlling for input market human capital, is not driven by limited-range omitted local public goods. However, it remains clear that we cannot definitely rule out alternative explanations based on selective migration and unobserved infrastructure with the data we have available and it must remain an important caveat.

Our empirical findings may also be consistent with stories involving complementarity between inputs in the production process that could be internalized by firms. If outputs for some firms are inputs for other firms, then proximity can increase productivity as it may reduce costs associated with transportation and delays in acquiring inputs. In general, if these industries cluster and are also human capital intensive, then the correlation pattern observed in these data could be generated. This might be the case if, for example, disk drive manufacturers locate near computer assembly factories. Similar patterns could also be generated with other forms of complementarity between highly educated workers and various inputs. In order to gain insight regarding the source of positive productive effects we included an interaction term between the individual's education and our local human capital measures in wage regressions, analogous to specification {3} in Table 5.1. The interaction term was the product of SEC% for 30-120 units and a dummy variable for the individual having some post-secondary education. This breakdown addresses the question of whether local human capital levels primarily influence the productivity of highly educated workers, as is suggested by some explanations that focus on complementarity. Estimates in Table 5.5 show that the interaction term is positive but insignificant with a large standard error and the SEC%30-120 variable itself remains positive and significant. Thus, there is some evidence that higher local human capital has positive effects on wages for low as well as high-skilled workers, and if some type of complementarity is driving our results, it does not appear confined to high-skilled workers.

6. Conclusion

This paper develops an empirical method for detecting human capital spillovers. Our approach uses an economic distance metric to define boundaries of the local public good. This economic distance metric enables us to segment regions in such a way as to elicit strong testable restrictions from a simple local public goods model. This segmentation of regions also mitigates some of the endogeneity concerns in our regressions. Our econometric approach allows for unobservables in our regressions to be correlated across observations as a function of economic distance, a major concern due to unobserved local public goods. This estimation method remains consistent even when economic distances are imperfectly measured.

Our main empirical finding is that local human capital as measured by secondary schooling rates is associated with significant, positive wage and rent differentials. These differentials are consistent with positive productive spillovers from local human capital. While their magnitudes do vary across specifications and are somewhat imprecisely measured, in general they are nontrivial in size. In specification {2} for example, a one standard deviation increase in the secondary education percentage from 30 to 90 units away would result in an increase in predicted wages of about 5% and 10% in rents. The estimated effects of human capital have an appreciable range. Human capital in a given location appears to impact productivity in areas that are as far as ninety minutes away. Finally, these qualitative findings for rents are robust when human capital is measured using schooling levels rather than percentages, however wage differentials appear only when it is measured using secondary schooling rates.

7. Colophon

The authors thank Gary Becker, Boris Cherkassky, Adeline Delavande, Paul Gertler, Lars Hansen, James Heckman, Judy Hellerstein, Ethan Ligon, James Rauch, Sherwin Rosen, Anne Royalty, Chris Taber, Robert Topel, Chris Udry, and seminar participants at the University of Chicago, UC Berkeley, Carnegie Mellon, Columbia, Federal Reserve Bank of New York, Northwestern, NYU, Rochester, UCSD, and Virginia for many helpful suggestions. Tsiang gratefully acknowledges support from the Institute for Policy Reform. Conley acknowledges support from the Margaret Reid Fund and the National Science Foundation (SES-9905720). Excellent research assistance was provided by A. Curtis. We are responsible for any errors. Contact information for the corresponding author is: Tim Conley, University of Chicago Graduate School of Business, 1101 E. 58th Street, Chicago IL 60637; tim.conley@gsb.uchicago.edu.

8. Appendix

8.1. Classical Multidimensional Scaling

This brief description of multidimensional scaling borrows heavily from Mardia et.al. [23] who should be consulted for a much more complete treatment of this technique.

The goal of CMDS is to obtain a configuration of T points in $\Re^2$ with distance matrix D that is close to the raw economic distances R. Of course the configuration is unique only up to translations and rotations. For the sake of exposition, consider a general distance matrix R that has zero elements on the main diagonal and whose other elements R_{ij} are positive. Define a centering matrix $H \equiv I - \frac{1}{T}\mathbf{11}'$, a matrix A whose elements $A_{ij} \equiv -\frac{1}{2}R_{ij}^2$, and $B \equiv HAH$.

The basic approach of CMDS can be made clear by considering what the matrix B would look like for a Euclidean distance matrix. As was first proved by Schoenberg [31] R can be a distance matrix for a configuration of points in a k-dimensional Euclidean space not lying in a proper subspace if and only if B is positive semi-definite and rank k. Consider a sketch of the proof of this proposition for $k = 2$.

First take the "if" part, if R were a matrix of Euclidean interpoint distances for a configuration $Z = (z_1, z_2, ..., z_T)'$ in $\Re^2$ then

$$R_{ij}^2 = -2A_{ij} = (z_i - z_j)'(z_i - z_j) \tag{8.1}$$

The matrix B can be written as:

$$B = HAH = A - \frac{1}{T}A\left(\mathbf{11'}\right) - \frac{1}{T}\left(\mathbf{11'}\right)A + \frac{1}{T^2}\left(\mathbf{11'}\right)A\left(\mathbf{11'}\right)$$

Define:

$$\bar{A}_{i\cdot} = \frac{1}{T}\sum_{j=1}^{T} A_{ij},\ \ \bar{A}_{\cdot j} = \frac{1}{T}\sum_{i=1}^{T} A_{ij},\ \bar{A}_{\cdot\cdot} = \frac{1}{T^2}\sum_{i=1}^{T}\sum_{j=1}^{T} A_{ij} \tag{8.2}$$

Then the elements of B can be expressed as:

$$B_{ij} = A_{ij} - \bar{A}_{i\cdot} - \ \bar{A}_{\cdot j} + \ \bar{A}_{\cdot\cdot} \tag{8.3}$$

Using equations 8.1 and 8.2 to rewrite equation 8.3 yields an expression for the elements of B:

$$B_{ij} = \left(z_i - \bar{z}\right)'\left(z_j - \bar{z}\right)$$
$$\text{or in matrix form: } B = (HZ)(HZ)'$$

Thus the matrix B, referred to as a centered inner product matrix, is positive semi-definite and rank two (assuming the points in Z are not on a line).

Now consider the "only if" part. Since B is positive semi-definite and rank 2, let its nonzero eigenvalues be denoted λ and μ. Take the eigenvectors ℓ and m associated with these nonzero eigenvalues and normalize them so that $\ell'\ell = \lambda$ and $m'm = \mu$. Let a configuration of points $Z = (z_1, z_2, ..., z_T)'$ in $\Re^2$ have as first coordinates ℓ_i and second coordinates given by m_i, i.e. $Z = \begin{bmatrix} \ell & m \end{bmatrix}$. The spectral decomposition theorem implies that:

$$B = \begin{bmatrix} \gamma_1 & \gamma_2 \end{bmatrix}\begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}\begin{bmatrix} \gamma_1 & \gamma_2 \end{bmatrix}'$$

for $\gamma_1'\gamma_1 = \gamma_2'\gamma_2 = 1$. The definitions of ℓ and m imply that $\ell = \lambda^{-1/2}\gamma_1$ and $m = \mu^{-1/2}\gamma_2$, therefore:

$$B = \begin{bmatrix} \ell & m \end{bmatrix}\begin{bmatrix} \ell & m \end{bmatrix}' = ZZ'$$

So $B_{ij} = z_i'z_j$. Now 8.3 can be used to show that the interpoint distances of this configuration equal R.

$$\begin{aligned}\left(z_i - z_j\right)'\left(z_i - z_j\right) &= z_i'z_i - 2z_i'z_j + z_j'z_j \\ &= B_{ii} - 2B_{ij} + B_{jj}\end{aligned}$$

Equation 8.3 implies:

$$B_{ii} - 2B_{ij} + B_{jj} = A_{ii} - 2A_{ij} + A_{jj}$$

$A_{ii} = 0$ so:

$$A_{ii} - 2A_{ij} + A_{jj} = -2A_{ij} = R_{ij}^2$$

Therefore the configuration Z has interpoint distances given by R.

When R is not Euclidean or is a Euclidean distance matrix of a configuration in greater than k dimensions, the above result offers the basis of the CMDS fitted configuration (Torgerson [36]). Since the normalized eigenvectors of the constructed matrix B produce a configuration with distance matrix equal to R when it is a Euclidean distance matrix in k dimensions, it is reasonable to suspect that the k eigenvectors with the largest eigenvalues might, when normalized as above, give a configuration in k dimensions whose distances are close to R. This is in fact the CMDS solution to finding a representation in k-dimensional Euclidean space of points with interpoint distances R : take the eigenvectors (normalized as above) corresponding to the k largest eigenvalues of the constructed matrix B as the points' coordinates. Let the CMDS solution in k dimensions be summarized by a distance matrix D. The CMDS solution is optimal in the sense that the configuration summarized by D minimizes the quantity: $tr\left(B - \hat{B}\right)^2$ over all $\hat{B}$ where $\hat{B}$ is the centered inner product matrix of a configuration in $\Re^k$ (Mardia [22]). Furthermore, if the matrix R corresponds to a configuration in $\Re^n$ where $n > k$, the CMDS solution is an optimal projection that minimizes $\sum_i \sum_j \left(R_{ij}^2 - \hat{D}_{ij}^2\right)$for any projection of the configuration R onto k dimensional subspaces of $\Re^n$.[21]

A measure of the goodness of fit for a configuration in k dimensions to the distances in the matrix R can be obtained by looking at the magnitude of the first k eigenvalues of B relative to the whole set of eigenvalues. For example, the ratio of squares of the first k eigenvalues to the sum of squares of all of its eigenvalues or the corresponding ratio of absolute values of eigenvalues provide measures of the distance information in the matrix R captured by the CMDS configuration.

8.2. Data Appendix

This study uses country-wide information from the 1980 Malaysian Census and micro-level information from the Second Malaysian Family Life Survey (collected

[21] Proofs of both of these optimal properties can be found in Mardia et.al. [23] Sec. 14.4.

in 1988 through 1989). Specifically, population and population by education data were taken from the Census for 1100 mukims (sub-districts) in the 78 administrative districts which comprise the 11 states of Peninsular Malaysia. The 1100 mukim aggregates were matched with the physically nearest nodes in our economic distance network depicted in Figure 4.1 to allow construction of local market variables using economic distance radii.

We use the male respondents in the New and Senior samples of the Second Malaysian Family Life Survey (MFLS2) to obtain wage, age, and schooling attainment information. In addition, a subset of households report rents paid and characteristics of residences. The male respondents were selected as husbands of a population-weighted survey of women under age 50. The weighting was accomplished by taking a random sample from the Statistics Department's sampling frame of 26,000 Census Enumeration Blocks (each EB is made up of 100 living quarters.) The key ingredient in our construction of economic distance is community information obtained from MFLS2 that includes travel times from EBs to nearest towns and cities.

A brief description of the Malaysian education system is useful in describing our individual and local market variables. The Malaysian education system parallels that of Great Britain. The first 6 years are primary education and begin at age 6. At age 12, children enter lower secondary (Forms 1-3) after which they receive an examination for the Lower Certificate of Education (LCE). If they pass, they enter secondary school (Forms 4-5) and prepare for the Malaysia Certificate of Education (SPM). Postsecondary school (Forms 6 lower and upper) corresponds to the 12th and 13th year of schooling and grant the Higher School Certificate (STP). University entrance requires this STP. Polytechnics and teacher and nurse colleges require only the SPM. Our indicator of some postsecondary education for an individual corresponds to their completing either form 6 or a year of training requiring the SPM.

Local market human capital measures are created from mukim-level educational attainment information from the 1980 census contained in concentric rings of the MFLS-2 respondent's EB defined using the economic distance metric detailed in Section 3. We use completion of Form 3 (the 9th year of schooling) with or without obtaining the LCE for passing the exit examination as the cutoff for this measure. The name for this variable takes the form SEC%n-m, corresponding the percentage of population with some secondary school within economic distances n to m.

The rental information is restricted to the subset of respondents who pay rent to non-relatives and non-employers to avoid the complication of transfer payments implicit in reported rental rates. The variables from the rent regressions are taken primarily from the respondent replies in the MFLS2 survey. Community leaders provided the EB-level information on government programs, piped drinking water, an indicator for whether major flooding took place within the past decade, and the date of electrification.

Wage regression variables are described in Table 7.1 and the rent variables not described therein are described in Table 7.2.

References

[1] Anas, Alex Richard Arnott, and Kenneth Small [1998] "Urban Spatial Structure" *Journal of Economic Literature* 36:1426-1464.

[2] Bandiera, O. and Rasul I.[2002] "Social Networks and Technology Adoption in Northern Mozambique" Working paper LSE.

[3] Becker, Gary S. and Kevin M. Murphy [1992] "Division of Labor, Coordination Costs, and Knowledge," *Quarterly Journal of Economics* CVII: 1138-60.

[4] Benabou, Roland. [1996] "Equity and Efficiency in Human Capital Investment: The Local Connection" *Review of Economic Studies* 63(2):237-64.

[5] Besley, T. and A. Case (1994). "Diffusion as a Learning Process:Evidence from HYV Cotton." Discussion Paper 174, Princeton University.

[6] Blomquist, Glenn C. , Mark C. Berger and John P. Hoehn [1988] "New Estimates of the Quality of Life in Urban Areas" *American Economic Review* 78:89-107.

[7] Case, Anne C. [1991] "Spatial Patterns in Household Demand" *Econometrica* 59:953-965.

[8] Cherkassky, BV, A. Goldberg, and T. Radznik [1993] "Shortest Paths Algorithms: Theory and Experimental Evaluation" Stanford Computer Science Dept. Technical Report STAN-CS-93-1480.

[9] Cho, George [1990] *The Malaysian Economy; Spatial Perspectives.* Routledge, Chapman, and Hall: New York.

[10] Conley, Timothy G. [1999] "GMM Estimation with Cross Sectional Dependence" *Journal of Econometrics* 92:1-45..

[11] Conley, Timothy G. and C. Udry [2002] "Learning About a New Technology: Pineapple in Ghana" Working Paper.

[12] Foster, A. and M. Rosenzweig [1995] "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture" *Journal of Political Economy* Vol. 103(6) p 1176-1209.

[13] Glaeser, Edward L. Hedi D. Kallal, José A. Scheinkman, and Andrei Schleifer [1992] "Growth in Cities" *Journal of Political Economy* 100:1126-1152.

[14] Government of Malaysia [1991] *The Second Outline Perspective Plan 1991-2000,*Kuala Lumpur: National Printing Department.

[15] Gyourko, Joseph and Joseph Tracy [1991] "The Structure of Local Finance and the Quality of Life" *Journal of Political Economy* 99(4):774-806.

[16] Haaga, John, J. DaVanzo, C. Peterson, T. N. Peng, T.B. Annzz ./1993*] The Second Malaysian Family Life Survey: Overview and Technical Report.* MR-106-NICHD/NIA. Rand Corporation. Santa Monica.

[17] Jovanovic, Boyan and Rafael Rob [1990] "Long Waves, Short Waves: Growth Through Intensive and Extensive Search" *Econometrica* 58 (6):1391-1409.

[18] Khoo Teik Huat [1987] *Population and Housing Census of Malaysia, 1980; Population Report for Mukims.* Department of Statistics Malaysia, Kuala Lumpur.

[19] Lee, Kye Sik [1989] "Migration, Income and Fertility in Malayasia: A Simultaneous Equations Model with Limited Dependent Variables" in *Applied Economics*, vol 21 pp. 1589-1610.

[20] Lucas, Robert E. Jr. [1988] "On the Mechanics of Economic Development." *Journal of Monetary Economics.* 22:3-42.

[21] Lucas, Robert E. Jr. [1998] "Externalities and Cities" Working paper University of Chicago.

[22] Mardia, K.V. [1978] "Some Properties of Classical Multi-dimensional Scaling" *Communications in Statistics; A: Theory and Methods.* 7:13, 1233-1241.

[23] Mardia, K.V., Kent, J.T., and Bibby, J.M.[1979] *Multivariate Analysis.* Academic Press, London.

[24] Meerman, Jacob. [1979] *Public Expenditure in Malaysia; Who Benefits and Why.* World Bank Research Publication. Oxford University Press.

[25] Moulton, Brent R. [1990] "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units" *Review of Economics and Statistics.* 334-338.

[26] Munshi, Kaivan (2001) "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution" Working Paper Univeristy of Pennsylvania.

[27] Rauch, James E. [1993] "Productivity Gains from Geographic Concentration of Human Capital: Evidence from Cities" *Journal of Urban Economics*, 34:380-400.

[28] Roback, Jennifer [1982] "Wages,Rent, and the Quality of Life." *Journal of Political Economy*, 92(6):1257-1278.

[29] Rosen, Sherwin. [1979] "Wage-based Indexes of Urban Quality of Life" In Current Issues in Urban Economics, P. Mieszkowski and M. Straszheim eds. Johns Hopkins Univ. Press.

[30] Romer, Paul [1990] "Endogenous Technical Change." *Journal of Political Economy.98:5 pt.2 S71.*

[31] Schoenberg, I.J. [1935] "Remarks to Maurice Frechet's Article 'Sur la Definition Axiomatique d'Une Class d'Espace Distances Vectoriellement applicable sur l'espace Hilbert"' Annals of Mathematics 36:724-732.

[32] Smith, Adam [1965] *The Wealth of Nations.* New York: Modern Library.

[33] Smith, James P. [1991] "Labor Markets and Economic Development in Malayasia" in Research in Population Economics, edited by T. Paul Schultz vol. 7 pp.131-156.

[34] Smith, James P. and Duncan Thomas [1992] "On The Road: Marriage and Mobility in Malaysia" Working Paper.

[35] Topel, Robert. [1986] "Local Labor Markets."*Journal of Political Economy.* 94:3 S111-S143.

[36] Torgerson, Warren S. [1958] *Theory and Methods of Scaling.* Wiley, New York.

[37] von Thünen, H. [1921] *Der isolierte Staat.* Waentig, Jena.

Table 4.1

Estimated Segment Travel Times by Population Density Category

	Travel Time Cost Estimates	
	Pt. Estimate	Std. Error
τ_0 (Fixed cost)	20.07	1.62
τ_1 (Least populated segment cost)	2.41	0.18
τ_2	1.87	0.17
τ_3	1.91	0.74
τ_4	4.88	0.92
τ_5	4.31	1.31
τ_6	8.84	1.55
τ_7 (Most populated segment cost)	1.54	0.68

OLS estimates, sample size 677, r-squared 48%.

Table 5.1
Instrumental Variables Regression Estimates, Log Wage as Dependent Variable

	Wage Regression Estimates			
	{1}	{2}	{3}	{4}
Years of School	0.0569 (0.0140)	0.0587 (0.0121)	0.0571 (.0126)	0.0582 (.0130)
Some Postsec. Ed.	0.1693 (0.1039)	0.1562 (0.0790)	0.1692 (0.0644)	0.1615 (0.0979)
Reads English	0.1429 (0.1175)	0.1340 (0.0910)	0.1356 (0.0782)	0.1313 (0.1066)
Experience	0.0491 (0.0140)	0.0500 (0.0147)	0.0506 (0.0121)	0.0511 (0.0139)
Experience^2	-0.0008 (0.0003)	-0.0008 (0.0003)	-0.0008 (0.0002)	-0.0008 (0.0003)
Chinese	0.5788 (0.0721)	0.5718 (0.0906)	0.5716 (0.0918)	0.5786 (0.0803)
Large Urban Areas	-0.0835 (0.0832)	-0.0834 (0.0692)	-0.0854 (0.0658)	-0.0890 (0.0765)
Time to City/Port	-0.0002 (0.0001)	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0001 (0.0001)
SEC % 0-30	0.0077 (0.0054)	0.0073 (0.0052)	0.0077 (0.0052)	0.0050 (0.0068)
SEC % 30-60				0.0040 (0.0042)
SEC % 30-90		0.0092 (0.0061)		
SEC % 30-120			0.0133 (0.0066)	
SEC % 60-90				-0.000006 (0.0037)
SEC % 90-120				0.0067 (0.0035)
Constant	3.4803 (0.2051)	3.2835 (0.2729)	3.1668 (0.3107)	3.3087 (0.2915)

Spatial standard errors in parentheses, sample size is 432.

Table 5.2
Instrumental Variables Regression Estimates, Log Rent as Dependent Variable

	Rent Regression Estimates			
	{1}	{2}	{3}	{4}
Flush	-0.2467 (0.2068)	-0.2914 (0.2022)	-0.2392 (0.2313)	-0.2502 (0.1938)
Drink	1.2335 (0.2852)	1.2975 (0.2526)	1.2234 (0.2490)	1.2526 (0.2396)
Nsleep	0.2972 (0.0846)	0.3079 (0.0899)	0.2911 (0.0853)	0.3091 (0.0943)
Anydev	0.1650 (0.1919)	0.1945 (0.1826)	0.1481 (0.1976)	0.1774 (0.2153)
H_2O Proportion	1.9161 (0.5789)	1.9628 (0.5908)	1.9160 (0.6015)	2.0002 (0.6134)
Flood	0.4490 (0.1697)	0.3916 (0.1559)	0.4271 (0.1618)	0.3196 (0.1703)
Elec. Years	0.0037 (0.0051)	0.0016 (0.0049)	0.0044 (0.0051)	0.0005 (0.0053)
Large Urban Areas	0.4940 (0.2413)	0.4819 (0.2502)	0.5077 (0.2489)	0.5471 (0.2609)
Time to City/Port	0.0006 (0.0006)	0.0009 (0.0007)	0.0007 (0.0008)	0.0007 (0.0009)
SEC % 0-30	0.0532 (0.0129)	0.0578 (0.0106)	0.0537 (0.0108)	0.0433 (0.0187)
SEC % 30-60				0.0206 (0.0118)
SEC % 30-90		0.0228 (0.0114)		
SEC % 30-120			0.0090 (0.0214)	
SEC % 60-90				0.00032 (0.0246)
SEC % 90-120				0.0005 (0.0171)
Constant	-0.7023 (0.8503)	-1.3232 (0.7600)	-0.9202 (0.8873)	-0.8939 (1.130)

Spatial standard errors are in parentheses. Sample size is 167.

Table 5.3

OLS Regressions, Log Wage and Log Rent as Dependent Variables

Wage Regression Specification {3}		**Rent Regression** Specification {2}	
Years of School	0.0572 (0.0126)	Flush	-0.3176 (0.2008)
Some Postsec. Ed.	0.1696 (0.0645)	Drink	1.3107 (0.2593)
Reads English	0.1374 (0.0781)	Nsleep	0.3067 (0.0889)
Experience	0.0507 (0.0122)	Anydev	0.2053 (0.1852)
Experience^2	-0.0008 (0.0002)	H_2O Proportion	2.0282 (0.5871)
Chinese	0.5744 (0.0905)	Flood	0.4015 (0.1599)
Large Urban Areas	-0.0878 (0.0647)	Elec. Years	0.0014 (0.0050)
Time to City/Port	-0.0001 (0.0001)	Large Urban Areas	0.5093 (0.2443)
SEC % 0-30	0.0067 (0.0046)	Time to City/Port	0.0009 (0.0007)
SEC % 30-120	0.0119 (0.0059)	SEC % 0-30	0.0642 (0.0113)
		SEC % 30-90	0.0227 (0.0105)
Constant	3.2161 (0.2966)	Constant	-1.5452 (0.7605)

OLS point estimates and spatial standard errors. Sample sizes 432 and 167 for the log wage and log rent regressions, respectively.

Table 5.4
Instrumental Variables Regressions Measuring Local Human Capital in Levels

Wage Regression Estimates		**Rent Regression Estimates**	
Years of School	0.0576 (0.0136)	Flush	-0.0888 (0.1937)
Some Postsec. Ed.	0.1601 (0.0989)	Drink	1.1735 (0.2281)
Reads English	0.1513 (0.1042)	Nsleep	0.3386 (0.0991)
Experience	0.0521 (0.0141)	Anydev	0.1459 (0.2702)
Experience^2	-0.0008 (0.0003)	H_2O Proportion	1.6788 (0.6989)
Chinese	0.5902 (0.0769)	Flood	0.4104 (0.2058)
Large Urban Areas	-0.0590 (0.0940)	Elec. Years	0.0029 (0.0053)
Time to City/Port	-0.0001 (0.0001)	Large Urban Areas	0.3742 (0.3298)
		Time to City/Port	0.0007 (0.0008)
SEC Level 0-30	0.0046 (0.0035)	SEC Level 0-30	0.0156 (0.0051)
SEC Level 30-60	-0.0009 (0.0066)	SEC Level 30-60	0.0528 (0.0185)
SEC Level 60-90	0.0050 (0.0041)	SEC Level 60-90	0.0029 (0.0190)
SEC Level 90-120	0.0016 (0.0039)	SEC Level 90-120	0.0014 (0.0129)
Constant	3.5562 (0.2008)	Constant	0.3660 (0.7967)

Instrumental variables regression estimates with log wages and log rents as dependent variables, with local human capital measured in levels. Spatial standard errors in parentheses, sample sizes 432 and 167, respectively.

Table 5.5
Log Wage Regression with SEC % and Individual Post-Secondary Education Interaction

	Wage Regression Estimates
Years of School	0.0568 (0.0127)
Some Postsec. Ed.	0.0224 (0.2460)
Reads English	0.1361 (0.0780)
Experience	0.0505 (0.0122)
Experience^2	-0.0008 (0.0002)
Chinese	0.5738 (0.0908)
Large Urban Areas	0.0869 (0.0660)
Time to City/Port	-0.0001 (0.0002)
SEC % 0-30	0.0078 (0.0053)
SEC % 30-120	0.0118 (0.0065)
SEC % 30-120*Some Postsec. Ed.	0.0070 (0.0112)
Constant	3.1977 (0.3052)

Spatial standard errors in parentheses, sample size 432.

Table 7.1 Variables used in Wage regressions

Variables	Mean	Std. Dev.	Description
Log Wage	5.00	0.74	Log average weekly earnings
Years of School	8.40	3.60	Maximum schooling attainment
Secondary Educ.	0.27	.	Any schooling > level 6 Yes=1, No=0
Reads English	0.35	.	Yes=1, No=0
Experience	24.2	9.4	Potential work experience =(age-schooling-6)
Chinese	0.28	.	Chinese ethnicity indicator
Time to City/Port	258	291	Travel time to big city / port in minutes
SEC % 0-30	29.4	16.2	% Pop. w/ secondary school within distance d, 0<d<30
SEC % 30-60	16.6	8.4	% Pop. w/ secondary school within distance d, 30<d<60
SEC % 30-90	19.4	5.0	% Pop. w/ secondary school within distance d, 30<d<90
SEC % 30-120	21.9	4.8	% Pop. w/ secondary school within distance d, 30<d<120
SEC % 60-90	16.6	7.1	% Pop. w/ secondary school within distance d, 60<d<90
SEC % 90-120	15.4	11.3	% Pop. w/ secondary school within distance d, 90<d<120

Table 7.2
Additional Variables used in Rent Regressions

Variables	Mean	Std Dev	Description
Log Rent	4.37	1.24	Log monthly rental expenditure (paid to owners not related, nor employer)
Flush	0.63	.	Residence has flush toilet Yes=1, No=0
Drink	0.95	.	Residence has piped water Yes=1, No=0
Nsleep	2.28	1.06	Number of rooms used for sleeping in respondent's house
Anydev	0.13	.	Has EB received development aid? (FELDA, etc) Yes=1, No=0
H_2O Proportion	0.92	0.22	Percentage of living quarters w/ piped water supply in EB
Flood	0.04	.	Any disastrous floods within 10 years in this EB? Yes=1, No=0
Elec. Years	18.5	14.8	How many years since EB first supplied with electricity?