The Geographic Concentration of Enterprise in Developing Countries

John S. Felkner and Robert M. Townsend*
University of Chicago, Massachusetts Institute of Technology

July, 2009

Abstract

We find salient patterns of increased agglomeration of enterprise within a high growth Asian country, coupled with more subtle patterns of contiguous geographic convergence which leave stagnant areas behind. Exogenous favorable physiographic conditions are also an important factor, especially at the local level within provinces. These results are obtained with a variety of advanced spatial statistical methods using both national and village level socioeconomic and biophysical data, merged in a Geographic Information System (GIS). We fit a structural, micro-founded model of occupation transitions with fine-tuned geographic capabilities to these village data and replicate the salient spatial facts. Key ingredients appear to be costs which increase monotonically with distance to key infrastructure and markets, credit constraints on occupation choice which are mitigated with spatially varying and endogenously increasing wealth, and exogenous and spatially varying expansion of financial service providers. Furthermore, a structural model of endogenous financial deepening applied at the village level reveals costs which appear to decrease with distance to major infrastructure and markets, indicating policy distortions and the role of a key government development bank, which help drive the growth and inequality of local and regional economies.

I Introduction

Relatively little is known about increasing spatial concentration in enterprise in emerging market countries. One draws the impression from a debate on global inequality that, overall, less developed economies are characterized by convergence. That is, from low levels there has been substantial industrialization and reduced inequality. Certainly India and China have been playing this role in the world economy, reducing world inequality as poverty rates in these countries plummet. But this begs the issue of what is going on within countries such as these, individually. Are these countries

*We would like to thank Edward Glaeser for very helpful comments, as well as the perceptive and substantive comments of four anonymous reviewers. We wish to thank Alex Karaivanov and Victor Zhorin for their assistance in adapting and simulating the occupational choice model program. Thanks also to Yukio Koriyama and Natalia Ramondo for their assistance in adopting and running the financial deepening model simulation program, and to Juliano Assuncao for helpful discussions and assistance with econometrics and spatial statistics. Many thanks to Mark Harris for editorial assistance. Funding for the research was provided by the National Institute of Health, the National Science Foundation, the John Templeton Foundation, and the Bill & Melinda Gates Foundation through the University of Chicago Consortium on Financial Systems and Poverty. Authors' E-mails: john3@uchicago.edu and rtownsen@mit.edu.

in a phase of "catch-up" and convergence from within, with lagging regions converging toward more developed areas? Or rather are they repeating the earlier historical pattern of the world as a whole, with increasing concentration and agglomeration of economic activity spatially in key areas?

In this study, we use high resolution spatial data on enterprise to look closely at the spatial patterns of enterprise concentration over time in a rapidly developing country: Thailand from 1986 to 1996, a time of high economic growth. There we find striking and interesting patterns of increasing spatial concentration of enterprise, as well as evidence for a more nuanced convergence story. The first part of this paper is devoted to characterizing these within-country spatial patterns. The second part uses the lens of structural models to understand one possible logic of the patterns we see, using multiple robustness checks for each part.

Despite military coups, tsunamis and the Asian financial crisis, Thailand's economic trajectory, over the long run, typifies an Asian success story. Village wealth doubled between 1986 and 1996, the period we study in detail, and by 1992 the ratio of M2/GDP exceeded the level in the United States. At the same time, the country industrialized, with the fraction of GDP in manufacturing rising from 23 to 35 percent and the number of households in non-farm enterprises increasing by 27 percent (Townsend 2009). Indeed, it is important to stress that industrialization and the rise of national income are not restricted to the creation of large firms and increasing profits. In l986 in Thailand, non-farm proprietorships accounted for 40 percent of national income, with corporate profits at only 5 percent. Even by l996, after a decade of high growth, enterprise continued to account for 30 percent of national income, more than double that of corporate profits. Likewise, looking at income and inequality decompositions, movements of households from wage earning and agriculture to enterprise and industry account for 21 percent of overall income change and 40 percent of inequality change (Jeong and Townsend 2007). Satellite images have shown that extensive deforestation and urbanization occurred during this time (Felkner 2000). Poverty rates drop from 48% in 1976 and to a low of 9 percent in 2002, one of the largest drops for any country (Townsend 2009).

All of this may sound like a convergence story. But, mapping various available data for Thailand via a Geographic Information System (GIS) and using advanced spatial statistical techniques, reveals substantial spatial agglomeration of enterprise. Also, there is a more nuanced convergence story that has much to do with proximity and accessibility to existing high-enterprise areas and infrastructure, but which is also correlated with exogenous geography. To test these hypotheses, we utilize highly accurate digital spatial data on major and minor roads, provincial district centers, geopolitical boundaries, and more than a thousand geo-located Thai villages.

These spatial patterns can be summarized as follows. First, the baseline. Nationally across Thailand in the 1986 baseline year, high levels of income, industry and enterprise are clustered around the central corridor of Thailand: in and around Bangkok and in a backbone of highly developed economic activity heading north along main transportation arteries. A village level census and data from a socioeconomic survey that calculates entrepreneurial profits indicate that areas closest to Bangkok, larger cities and main highways have higher fractions of village populations in enterprise. Additionally, enterprise is also modestly correlated with the physiographic environment; high levels of these enterprise variables correspond with high soil fertility and low elevation. In contrast, the Northeast region and other areas that have low soil fertility and higher elevation also have low levels of enterprise, profits, and factories.

In terms of the key variable, the growth of enterprise, increasing spatial concentration of enterprise emerges over the next ten years, from 1986 to 1996. Specifically, we find areas located along the edges of the 1986 Thai central corridor that have high initial levels of enterprise, surrounded by areas with very high growth in enterprise 1986-1996. Related, some significant new hot spots emerge (in the north in Chiengmai, Lampoon and Chiengrai, and in parts of the south). There are also areas of convergence with low initial levels of enterprise and high growth, but these "catch up" areas are largely contiguous to the central core regions, extending the reach of these hot economic areas while leaving stagnant, more distant areas behind. At the same time, the central core areas do hollow out a bit over time: many central areas of high initial levels are associated with below average growth in enterprise. Finally, from the perspective of the national scale, the hinterland in the Northeast and other areas appear with both low levels of enterprise and low growth. In sum, at the national scale, while the central areas of Thailand flourish and expand, enterprise growth is relatively stagnant in more peripheral regions.

Making use of a unique dataset of the geo-locations of more than 1000 Thai villages in four provinces, we are able to zoom into the sub-provincial scale and look at spatial concentration in enterprise and other factors over time. This reveals similar trends to the national scale, but only in certain regions. Increasing concentration of enterprise is taking place in two representative provinces relatively near Bangkok, with some areas of adjacent new growth, and some hollowing out in areas with formerly high levels. However, in the two provinces in the Northeast where stagnation seems to be happening at the national scale, with low levels of enterprise and growth, we find that small but obvious areas of agglomeration are forming, as if the process of agglomeration and concentration were beginning anew. These less developed provinces show no hollowing out in existing areas of agglomeration, though, of course, stagnant provincial hinterlands remain.

Our analysis shows in fact that not only high levels but also high growth of enterprise across villages, within provinces, is significantly more likely to happen in areas with good infrastructure (near to major highways, government district centers and major highway intersections). Also at the village level within provinces, high growth in enterprise is correlated with favorable physiographic conditions (good soil, reliable rainfall, near rivers and flat topography).

In addition to documenting these patterns, this paper seeks to understand the processes, or better put, one possible process, that lies beneath enterprise agglomeration and the spatial patterns just noted. Our goal here is not to try to prove causality through instrumental variable techniques. Rather, we view the data though the lens of a structural model with fine-tuned, geographic capabilities. The model explicitly considers the geography of village locations, allowing us to test the model spatially, quantitatively assessing its ability to capture these spatial patterns. We also estimate how key enterprise costs parameters vary across geography in Thailand. The model takes a stand on causality, and we look closely to see how well it fits. The key ingredients of this model are varying levels of initial wealth (which we take as given though in the data wealth is much related to advantageous environment), mixed but expanding access to the financial system (with credit constraints for those without access alleviated by increased wealth), and a fixed cost of setting up enterprise and leaving subsistence agriculture and/or wage work. We estimate the key parameters of the model (preferences, technology) by minimizing the village level prediction errors, essentially the squared error between model predictions for the final period, l996, versus the actual data of l996.

The fixed cost of setting up enterprise and leaving subsistence agriculture/wage work is estimated to increase monotonically the further a village is to physical infrastructure and environmentally advantageous areas (although with a flatter slope). Further, because the model simulations are spatially explicit, we are able to map model simulation outputs with sharp spatial resolution. We find that, strikingly, the model delivers many of the same patterns of agglomeration and nuanced convergence that we see in the actual data. A quantitative comparison of the model spatial simulation with the actual spatial patterns indicates that we are doing quite well in understanding the growth of enterprise.

There is, however, nothing automatic about these results. In fact each of the ingredients of the model seems to play a role. In quantitative experiments, we find that if we do not allow enterprise set up costs to vary with distance, the structural model does worse: prediction errors are spatially concentrated, rather than random, and are correlated with wealth and distance to physical infrastructure. Further, the expansion of financial intermediation is also central to the results, and it is crucial that we use in the model what we see in the data. If we keep intermediation at its initial, low level, or rid the model of credit entirely, the simulated patterns of enterprise growth are again much poorer.

The expansion of financial infrastructure is not endogenous but rather seems related to policy levers. To establish this we use another structural model in which we replicate our procedures, to try to explain financial deepening in the same way that we previously explain the growth of enterprise. In this second structural model, households weigh off the gains of better risk sharing and project selection from intermediation with the fixed costs of entering the intermediated sector. Again, we repeat our steps of estimation and spatial prediction. The model with endogenous financial infrastructure tends to over predict actual expansion of credit in and around towns and areas of initial concentration of wealth and education and under predict actual expansion in more distant and poorer rural areas. Part of this can be attributed to the targeting of the government's Bank for Agriculture and Agricultural Cooperatives (BAAC), which tends to go to low wealth, distant areas. There is moreover an important implication to this exercise regarding the ingredients of agglomeration and enterprise growth. In contrast to the targeting and policy driven expansion we see in practice, the model under predicts the observed credit expansion in remote areas and its positive impact, as we have it in the occupational choice model. Thus, in searching for an alternative story and seeking to understand agglomeration forces, we would have imputed an enterprise setup cost that would have been more favorable and hence too low in remote areas, and we would have understated the clear monotonic spatial gradient increase in enterprise costs estimated in the enterprise model.

II Literature Review

Studies of enterprise and its spatial patterns at the regional or rural level in developing countries are not extensive, in many cases because high-resolution spatial data on enterprise growth is typically difficult to obtain. There are few, if any, micro-founded, macro-structured models. Some studies have noted that new economic development and enterprise tend to occur in close spatial proximity to existing centers of economic growth. For example, Burchfield et al. (2006) find that new urbanization in the US from 1976-1992 tends to occur adjacent to and on the fringe of existing major market and urban centers. Araujo (2003) finds that proximity to urban centers is strongly significant in predicting growth of rural non-farm employment in Mexico, while Conley, Flyer and Tsiang (2003) find that spillovers from local human capital are important in explaining the distribution of productivity in Malaysia.

Theoretical explanations include information spillovers and increasing returns to scale in the presence of large local markets (Becker, Murphy and Tamura 1988; Marshall 1890; Krugman 1991; Fujita, Krugman and Venables 1999). Some of these have been found to be significant in empirical testing (Conley, Flyer and Tsiang 2003; Grossman and Helpman 1992; Ellison and Glaeser 1997, 1999). Dijk and Sverrisson (2003) find that enterprise clusters and networks of small enterprises are highly important for technology diffusion in developing countries. Foster and Rosenzweig (1995, 2003) find that neighbor spillover effects are significant in new technology adoption and agricultural productivity growth among farmers in India from 1970-2000.

Exogenous geographic conditions have also been shown to be significant in dictating the initial spatial location of economic activity and enterprise. Hoxby (2000) finds that natural geographic boundaries can be used to define "communities", rather than political districts, to test the Tiebout hypothesis concerning provision of education in municipalities. Burchfield et al. (2006) find that, in addition to proximity to existing urban centers, physical geography (including groundwater availability and temperate climate) explains 25 percent of cross-city variation in new urbanization in the US. Roos (2005) finds that up to 36 percent of Germany's spatial GDP can be explained by direct and indirect effects of geography, while Ellison and Glaeser (1999) find that one-fifth or more of the agglomeration of US industries can be accounted for by natural geographic advantages, including natural resource location. More globally, Gallup, Sachs and Mellinger (1999) find that geographic factors including proximity to waterways, topography, and proximity to the equator (as a proxy for poor tropical soils and malarial incidence) are highly significantly correlated with per capita GDP output globally.

Research that use micro-founded structural models featuring occupation transitions include Gine and Townsend (2004), Jeong and Townsend (2007), based on Lloyd-Ellis and Bernhardt (2000) and Buera, Kaboski and Shin (2009), building on related work of Banerjee and Newman (1993) and Aghion and Bolton (1997), though these models are all aggregated to the macro level.

In this study, we take initial spatial locations of villages, roads and physiographic/environmental features as pre-determined, part of the initial specification. On the other hand, we do allow entrepreneurial activity, migration, interest rates, wages, interest rates and wealth to be endogenous, to evolve over time and space in interesting ways that can be compared to reality. That is, we do not dictate in which villages or areas firms are established, what occupations households choose, or where people work. In the endogenous financial access model we do not dictate how much is saved or invested or who has access.

The paper proceeds as follows: Section III provides a conceptual overview, describing the study area, data and GIS, and presents some stylized facts about entrepreneurship disaggregated spatially at both national and provincial levels. In Section IV, we describe a spatial occupation choice model and its estimation and simulation, comparing the spatiotemporal patterns of the model simulation and prediction errors. In Section V, we address the potential impact of credit on enterprise in a different way, by estimating and simulating a spatial model of endogenous financial intermediation.

III Study Area, Data and GIS

We consider the spatial distribution of certain key economic variables at both the national level and at the provincial level. These provinces were specifically chosen as representative of country-wide economic variation; they represent a gradient of decreasing development moving from west to east[1]. Chachoengsao and Lop Buri, are in the richer central region, relatively near Bangkok; and Buriram and Sisaket are in the poorer northeast (see Figure 1).

A Geographic Information Systems (GIS) database was constructed for the entire country that included Thai political boundaries, digitized road networks, and physiographic data. Further, a

[1] Per capita Gross Provincial Product in 1986 was, moving west to east, 27,300 Baht in Chachoengsao and 16,500 in Lop Buri, compared to 8,450 for Buriram and 7,950 for Sisaket, versus a national 1986 average of 23,944 (Thailand National Economic and Social Development Board- TNESDB).

high-resolution subset of this GIS was created for the four study provinces that included more than 1000 village geo-locations, high-resolution (primary/secondary and tertiary) road networks, Amphoe government district centers, and again physiographic data[2]. Highly detailed road network data for the four study provinces, displayed in Figure 2, was obtained from Royal Thai government maps from the early 1980s, to certify that all road networks were in place before the 1986 study base year.

New roads constructed after the 1986 baseline period could potentially interact with enterprise growth. To evaluate whether new roads were constructed after the baseline date we compared the Thai government road data with more recent maps obtained from American Digital Cartography (ADC) Inc.[3], as well as with current Thai road maps and Google Maps data. The comparisons showed that no new primary roads (highways and high-quality paved roads) had been constructed after 1986: all of them were present by the early 1980s in the Royal Thai map data. This allows us to avoid potential impacts of new roads and attests to the relatively high level of Thai infrastructure investment completed by the 1986 study base year period[4].

Some new secondary roads were constructed, however, as these are present in the ADC but not the earlier Thai maps. Still, these are very few[5], and we did not find any examples of clusters of new constructed secondary roads correlating with enterprise growth. Almost all of these new secondary roads were either relatively minor additions to market centers that already possessed extensive secondary road networks, or were constructed in rural areas that did not experience enterprise growth. A regression of enterprise growth locations onto distance from new secondary roads was not significant, with a p-value of .15, and further the coefficient was positive (indicating that new roads tended to correlate with increasing distance from enterprise growth clusters).

We utilize the GIS and much socioeconomic data to document trends and patterns relative to baseline infrastructure and to the environment. We include data from the Thai Socioeconomic Survey (SES), the Thai Household Population Census (HPC), the Thai Department of Industrial Works (DIW) factory census, the Thai Community Development Department (CDD) Rural Development Committee (RDC), and the Townsend Thai data[6]. For the physiographic elements, data on soil

[2]Political boundaries were obtained from the Thai government, allowing spatial aggregation of socioeconomic CDD village variables by administrative unit (as, for example, in Figure 1). Road networks for the four provinces were digitized from the Thai government maps, with further accuracy assessment from independent Landsat satellite images. Roads were assigned approximate speed categories using road category designations from the Thai government maps, ground-truth information collected on site, and comparison with satellite images. Local amphoe government District Center locations were obtained from the TNESDB. Both the CDD village and the amphoe District Center GIS point locations were linked to the road network in the GIS, allowing for the calculation of travel time along road networks. Thiessen polygons (or "proximal polygons") were used as a tool to make the village spatial variation in entrepreneurial activity, wealth and other key variables easier to appraise visually. Thiessen polygons are formed in the GIS by assigning every point in space to the nearest village, and then forming polygons to contain the set of that village's catchment area (Chou 1997). This made the spatial variation much easier to assess visually and had the advantage of not altering the original data (as would be required in any kind of aggregation to amphoe or tambon polygons, for example). All spatial data in the GIS was rectified to the Universal Transverse Mercator (UTM) map projection system.

[3]The ADC global road network data was rated as highest available quality for global roads databases by a World Bank study (Nelson, de Sherbinin and Pozzi 2006), including in comparison to declassified US CIA data.

[4]Primary roads are the most important infrastructure networks for movement of goods, harvest and people in the four provinces, and are thus most likely to determine relative village accessibility to market centers and agglomerations. Thus the fact that all these primary routes were in place before the advent of our study greatly reduces potential endogenous effects.

[5]Only 1 or 2 new secondary roads were constructed in Lop Buri, Sisaket and Chachoengsao.

[6]Townsend, R., principal investigator with Paulson, A., Sakuntasathien, S., Lee, T. J., and Binford, M., Questionnaire Design and Data Collection for NICHD Grant Risk, Insurance and the Family and NSF grants, The University

fertility were compiled from Thai government maps of soil types[7]. Data on annual rainfall variation was calculated from Thai meteorological stations collected by the Thai weather service 1951-1996, then interpolated to village locations using geostatistical techniques. Geographic distance to rivers was calculated for every village and every administrative district (including provinces, amphoes and tambons) using GIS data on river and stream networks compiled by the Thailand Environmental Institute (TEI). Slope was calculated from Digital Elevation Models (DEMs) extracted from vectorized contour lines on high quality Thai land cover maps. Basic statistical summaries of these key variables are reported in Table 1.

For the four study provinces we use primarily village socioeconomic data collected by the CDD every two years from 1986-1996[8]. The CDD is collected in every village, and responses are at the village, not the household, level. Thus all testing, analysis, estimation and simulation are done for villages. While the CDD survey does not specifically contain a variable for entrepreneurial activity, variables listing the number of households per village exclusively engaged in retail and cottage industries were summed per village and then divided by the number of households in a village to create a variable capturing the percentage of entrepreneurial activity in a village. This naturally raises the question of what these enterprises really are. Paulson and Townsend (2004) use the household level Townsend Thai data and report the most common enterprise is a shrimp pond, followed by trader, and then shop. In towns and urban areas the common enterprise is retail shops, followed by food and drink sales, then restaurants or noodle shops. As will be evident below, areas with household enterprise are also areas with industrialization. Indeed, the CDD data as a survey of villages excludes some urban neighborhoods but suburbs, new towns and other urban extensions are included in the data.

To measure variation in wealth levels, a wealth index variable was created as a function of four CDD survey asset variables. The four assets were combined using a principal components function, and the Eigen values of the first principal component were used to determine the wealth index values for each village[9]. Education was measured as a percent of the village population completing secondary education. The CDD also contains information on credit providers, both government and commercial. Travel times between villages, Thai government District Centers, major road intersections, city centers and physiographic features (e.g. rivers) were calculated using standard digital road network GIS travel-time algorithms including variation in road quality.

of Chicago, 1997

[7]More specifically, geo-located, cation exchange capacity and field capacity taken from national digital soil maps were used to construct a relative index of inherent soil fertility.

[8]The CDD village survey covers hundreds of socioeconomic variables including those pertaining to income, wealth levels, education, agricultural productivity, assets, and demographic factors. Village GIS point locations were linked to the CDD database by Thai administrative district identifier codes, taken from Thai government maps, and by village name. Village location information was not available for certain areas in southern Sisaket and Buriram near the Cambodian border.

[9]The four assets that were used for this procedure – per capita TV ownership per village, per capita motorcycles per village, per capita pick-up trucks per village, and the percentage of households having flush toilets per village – were chosen because each is arguably representative of a certain level of economic achievement in the various contexts that are present across the four provinces. The first principal components vector captures the axis of maximum variation across the multi-dimensional space of the four components and is therefore arguably a better indicator of wealth across the four provinces than any asset individually. However, any missing values for any of the four assets in the data made the calculation impossible. Unfortunately, this resulted in a number of villages with no wealth index values calculated: specifically, wealth index values could only be calculated for 76 percent of the total sample. However, this sample used is representative of overall wealth, as a comparison of means for each of the four input asset variables were within one-half a standard deviation in all cases.

A Spatial Statistical Test for Spatial Clusters Simple visual appraisal of the spatial distributions of the primary variables revealed clear spatial patterns. However, visual appraisal can be subjective, as the human eye can erroneously identify apparent spatial patterns in randomly generated spatial data (Merrill and Selvin 1997).[10] Consequently, spatial statistical techniques were used to detect and confirm the presence of spatial concentrations, or "hot spots", in the primary variables. Specifically, we utilize the Local Moran statistic (following Anselin 1995)[11] The Local Moran calculates the Moran index for each observation i, as the cross-product of the standardized value of a variable at a location i with that of the average of neighboring (standardized) values. The statistic is expressed as:

$$I_i = \frac{Z_i}{m_2} \sum_j W_{ij} Z_j \tag{2}$$

where

$$m_2 = \frac{\sum_i Z_i^2}{N} \tag{3}$$

This statistic is a measure of the strength of the spatial correlation of an observation with its neighbors[12]. The results can be either positive (in the case that the standardized value Z_i of the observation is positive and the sum of its standardized neighbors Z_j is positive, or in the case that Z_i is negative and the sum of its standardized neighbors Z_j is negative - considered positive spatial association) or negative (in the case that Z_i is positive and the sum of its standardized neighbors is negative, or in the case that Z_i is negative and the sum of its standardized neighbors is positive - considered negative spatial association). The expected value is $-\frac{-1}{N-1}$ under complete spatial randomness (no positive or negative spatial autocorrelation/association).

[10] Also, some of the apparent patterns may be at least partly a function of GIS display choices regarding colors assigned to levels or the unequal spatial area taken up by certain village Thiessen polygons (the latter problem is referred to in the literature as the Modifiable Areal Unit Problem (MAUP), see (Fotheringham & Wong, 1991)).

[11] More specifically, the Moran's I (Moran, 1948) was developed as a global statistic of the degree of overall spatial autocorrelation in a spatial dataset. It measures the overall degree to which observations in that dataset tend to be correlated to their immediate spatial neighbors, relative to a random distribution. The statistic measures the degree of covariance between observations and their neighbors (or the "spatial lag"), and it can be used as an index of the global degree of spatial clustering or dispersion in data. Moran's I can be expressed as:

$$I = \frac{N}{\sum_i \sum_j W_{ij}} \frac{\sum_i \sum_j W_{ij} Z_i Z_j}{\sum_i Z_i^2} \tag{1}$$

where Z_i is the deviation of the variable of interest with respect to the mean, W_{ij} is the matrix of weights that determines the "neighbors" j for each observation i, and N is the number of observations. The numerator here, thus, is a measure of covariance between observations and their neighbors, summed across the whole data set, and the denominator is a measure of global variance, producing a normalized index. Moran's I returns a single global statistic for the full spatial dataset. However, it does not facilitate observation of smaller, or more "local", spatial clusters or hot-spots, which may exist and be significant even if the global statistic does not indicate significant spatial autocorrelation. For this, several "local" indicators exist which can be used to decompose the global statistic down to local levels, to identify hot-spot clusters.

[12] For our analysis, the W_{ij} neighbor weights used were, in the national level maps, the administrative district (either Tambon or Amphoe) directly bordering a given geopolitical unit (equivalent to "first-order" spatial lag), while at the village level neighborhoods were defined as all villages within 10 kilometers. These neighborhood weights were determined using sensitivity testing of a range of potential neighborhood weight specifications (including inverse distance weights up to larger and smaller distance ranges), with the Moran's I output evaluated to determine the optimal weight specification (following the accepted approach in the spatial statistical literature).

To detect significant spatial clusters, the Local Moran values are evaluated against a statistical hypothesis significance test. As general results on the distribution of this statistic may be hard to obtain, a conditional randomization or permutation approach is used to yield so-called pseudo significance levels (as in Hubert 1987).[13] To visualize the locations of (significant) spatial clusters, a widely-used approach is to map not the actual Local Moran's I values, but rather the corresponding values of the the Moran Scatterplot, binned into four groups depending on the type of spatial association that an observation has with its neighbors. The Moran Scatterplot technique plots the standardized response variable (that is, with the mean subtracted and divided by the standard deviation) against the standarized values of the spatial neighbors for that observation (the spatial lag) for all observations in the data (according to the definition of "neighborhood" used in the W_{ij} spatial weights matrix). Observations that are of higher than average standardized response value surrounded by neighbors whose spatial lag values are higher than the global average ("high-high" values, or hot-spots) are in the upper right quadrant of the Moran Scatterplot), while observations in the lower left quadrant are those with below average value surrounded by neighbors with below average mean values ("low-low", or low-spots). Observations in the lower right and upper left quadrants are the reverse: high values surrounded by low-value neighbors, and low values surrounded by high value neighbors. These bins can be colored and mapped, to visualize this local clustering across space. The left panel of Figure 4, for example, displays the Moran Scatterplot values for enterprise for all Tambons in Thailand. However, to visualize which of these clusters are statistically signficant, normally the Moran Scatterplot values are mapped only for the observations that have statistically significant Local Moran values. Such maps are known as LISA ("Local Indicators of Spatial Association") maps (Anselin, 1995).

Furthermore, both the LISA and the Moran Scatterplot can be applied across both space and time, by plotting the standardized values of an observation in one time period against the values of neighbors in another time period, or against the mean of the growth in that variable over time, using an approach known as the Bivariate LISA (Anselin, 1996). We have applied this method to enterprise in Thailand, with the Local Moran's I_i, calculated using enterprise in the base year at Z_i with the sum of enterprise growth 1986-1996 for the neighbors $\sum_j W_{ij} Z_j$, and then mapping corresponding values of the Bivariate Moran Scatterplot (for the same variables) for those observations that are significant.

III.1 Spatial Patterns in Enterprise at the National Level in the Base Year

The top left graphic in Figure 3 displays entrepreneurial activity from the CDD at the national level for 1986, our baseline year, aggregated to the amphoe level. The graphic clearly shows salient spatial patterns, with a clear clustering of entrepreneurial activity around the central Bangkok metropolitan region and the central "corridor" extending north and south from Bangkok, as in Figure 1. These levels are considerably higher than in the poorer areas outside this corridor, such as the northeast region. This clustering is confirmed by the top right-hand graphic in Figure 3, which displays a LISA map of enterprise levels. The red clusters in this graphic indicate amphoes that have higher enterprise activity than the national average, which are surrounded spatially by amphoes having

[13] That is, the randomization here is conditional in the sense that the value Z at a location i is held fixed (that is, is not used in the permutation) and the remaining values of the spatial data set are randomly permuted over the locations in the data. For each of these resampled data sets, the value of the Local Moran (for our Z_i) is computed (for this study , 999 permutations of the spatial data were used). The resulting empirical distribution provides the basis for a statement about the extremeness (or lack of extremeness) of the observed Local Moran at i, relative (and conditional on) the values computed under the null hypothesis (the randomly permuted values).

higher averages than the national average. Furthermore, the LISA map also identifies statistically significant blue areas that have low rates of enterprise in the northeast, north and south.

Quite similar spatial cluster patterns are also found in data from other sources, including entrepreneurial income data from the Socioeconomic Survey (SES) shown in Figure 3 (aggregated by amphoe for 1990), percent of respondents engaged in entrepreneurial occupations in the 1990 Household Population Census (HPC) (not shown), and in the spatial density of factories across Thailand, derived from the DIW data linked to Thai tambons for 2005.[14][15]

Table 2 displays regressions of enterprise onto infrastructure and environmental variables at the national level in the base year 1986, and in 1990 for SES entrepreneurial income. The percent in enterprise in the CDD falls off significantly with increasing geographic distance to major highways, to the 15 largest Thai cities, and to Bangkok. Specifically, the fraction in enterprise drops by 0.03 percent for every kilometer of increasing distance from major highways (or about a 1.8 percent reduction in enterprise incidence per kilometer of distance, calculated against an overall mean of 1.65 percent of the population in enterprise). For the SES data on entrepreneurial monthly profits, bivariate regressions were not as consistently significant, but income does fall significantly with geographic distance from major highways, falling by 11.5 Thai Baht (about half a US dollar at 1990 exchange rates) for every kilometer of additional distance from major highways[16].

In addition, the environment and geography matter, as CDD enterprise is also highly significant and varies positively with higher soil fertility (improving by 0.13 percent with every unit increase in the soil fertility index, which ranges from 0 to 9), and with decreasing elevation (declining by about .002 percent with every additional meter of elevation). DIW factory spatial density also varies significantly with proximity to major rivers (decreasing by .0145 factories per square kilometer with every additional kilometer of distance from rivers), to areas with lower elevation (declining by .005 factories per square kilometer for every additional meter of elevation) and lower annual rainfall variation (decreasing by .158 factories per square kilometer with every centimeter increase in annual rainfall variation, which ranges from 13 to 44 centimeters).

III.2 Spatial Patterns in Enterprise Growth at the National Level, 1986-1996

Our focus in this paper is whether or not an increasing concentration of enterprise develops over time and with what spatial and environmental patterns. Specifically, we use the Bivariate LISA technique to detect spatial agglomerations of increasing high enterprise activity. We do this by using the Bivariate LISA to compare enterprise levels for Thailand's 6500 Tambons in the base year with the value of enterprise growth 1986-1996 in their neighbors. The result identifies statistically significant clusters, hot-spots, where enterprise begins at high levels in 1986 and then is surrounded by areas with higher than average growth in enterprise 1986-1996. The technique also identifies areas that are stagnant (low initial enterprise surrounded by low enterprise growth). We plot the calculated LISA indices on the national map, showing the values for all Tambons in one map, and the statistically significant clusters in another (a p = .05 or less), shown in Figure 4. Areas

[14] Notably, analysis of the DIW data showed a high spatial density of factories in the poorer northeast and some areas south of Bangkok.

[15] Regressions performed at the tambon and amphoe level indicate that early on, at least, credit was also concentrated in the central backbone, high wealth areas (Assuncao, Mityakov and Townsend 2008). On the other hand, health clinics and schools do not correlate with distance so it is not obvious from this that more remote areas are inhabited by less talented or less healthy people.

[16] The only exception here in the analysis of spatial trends with distance to cities or infrastructure is DIW factory spatial density, which increases with distance from the 15 largest Thai cities.

colored red areas are areas with high levels of base-year enterprise, surrounded by areas of higher than average enterprise growth from 1986 to 1996, implying increasing agglomeration over time. As is evident, these areas of growing agglomeration lie on the edges of the highly developed central core areas in the base year, including areas in the south, on the eastern Seaboard and north in and around Chiengmai/ Chiengrai/ Lamboon. Areas colored light blue, by contrast, are tambons with higher than average enterprise in l986 surrounded by districts with lower than average growth. These areas display "convergence to the mean from above" and are largely in the central core areas. There is thus a hollowing out of the core areas, so to speak. There is also "convergence to the mean from below", or tambons with initially low levels of enterprise surrounded by areas of higher than average growth, colored in pinks. These form large areas of buffer zones in and around the central core, including much of the South and North. Less of this is statistically significant. Finally, the dark blues represent low levels of enterprise surrounded by low levels of growth, areas of stagnation. This includes much of the Northeast, the lower North (and the extreme Islamic south), though again significance is weak.

Table 2 also displays regressions of the growth of enterprise onto infrastructure and environmental variables at the national level, for growth in CDD percent in enterprise and 1990-1996 growth in entrepreneurial income. The results are weak and mixed: for the SES enterprise profit data, enterprise growth increases by .0649 percent with every unit increase in the soil fertility index (which ranges from 0 to 9), but elevation seems to have a perverse sign. Also, the positive signs on distance to Bangkok for CDD enterprise growth and on distance to the 15 largest Thai cities for growth in entrepreneurial income seem to indicate convergence. Evidently these regressions do not pick up the more striking geographic patterns in growth that are evident in the map.

III.3 Spatial Patterns in Entrepreneurship at the Sub-Provincial in the Base Year

We now look at spatial patterns at the village level, using village geo-locations and CDD data for more than a thousand villages in four provinces. The top panel of Figure 5 displays a LISA map of statistically significant clusters of high fraction in enterprise in 1986, while the bottom panel shows statistically significant spatial clusters in wealth for 1986, with each province analyzed separately[17].

As anticipated, the central provinces of Lop Buri and Chachoengsao (adjacent to the Bangkok area) have higher average levels of enterprise (0.02 and .0179 percent for Chachoengsao and Lop Buri, respectively, compared with 0.0065 and 0.0035 percent, respectively, for Buriram and Sisaket), consistent with the national agglomeration of enterprise in the Bangkok metropolitan area. Furthermore, the western parts of both of these central provinces stand out markedly for their solid clusters of high enterprise activity.

The central provinces also have considerably higher average wealth levels (average index values of 32.4 and 29.8 for Chachoengsao and Lop Buri, respectively, compared to 9.3 and 11.4 for Buriram and Sisaket). Statistically significant spatial clusters of wealth as confirmed by the Bivariate LISA technique largely mirror the patterns for enterprise in the central provinces: significant clusters of high wealth in western Chachoengsao and south-western Lop Buri. However, high-wealth clusters in the poor northeastern provinces vary somewhat from locations of high enterprise clusters, pointing

[17]For these Figures, the univariate LISA was calculated for each province individually, and thus clusters are detected and evaluated for statistical significance relative to other villages in that province only. Thus, the map reveals relatively large clusters of high enterprise and wealth in the northeastern provinces (Buriram and Sisaket) compared to the eastern provinces (Chachoengsao and Lop Buri), despite the overall higher levels of enterprise and wealth in the central provinces. Also, the patterns here look quite different than the patterns in Figures 3 and 4, as those were evaluated at the Tambon level relative to the national sample.

to some spatial independence between wealth and enterprise, although both wealth and enterprise high-value clusters tend to exist near the intersections of major highways. For both enterprise and wealth in all provinces, the statistically significant low-value clusters correlate with poorer, highly agricultural areas[18].

Table 2 displays the bivariate regressions at the provincial level[19]. The patterns are clear and striking: enterprise levels in 1986 increase highly significantly with improved travel-time to major highways and to Amphoe district centers, with distance to major intersections, and all geographic and environmental variables are significantly related to levels of enterprise. Notably, enterprise fraction decreases by .0473 percent with every minute of increase in travel-time to major highways (mean of 12.05 minutes travel-time for all villages in the sample), and by .0493 percent for every minute of reduction in travel-time to Amphoe district centers (mean of 14.37 minutes for all villages). Correlations between enterprise and exogenous geographic factors (distance to rivers, soil fertility, elevation and annual rainfall variation) are highly significant (p-values of less than one percent). Here, enterprise decreases by .132 percent with every kilometer of distance from major rivers, increases by 0.543 percent with every unit increase in the soil fertility index (as compared to 0.13 percent measured nationally, with the soil fertility index ranging from 0 to 9), and decreases by 0.105 percent with every additional meter of elevation (as compared to .002 percent measured nationally). Annual rainfall variation stands out here, as the significant positive correlation with enterprise seems surprising. Wealth (not shown) also correlates strongly with proximity to major road infrastructure and government centers, specifically distance to a major highway, distance to the intersection of major highways, or distance to a district center. This confirms intersections of major highways and government district centers as areas of initial concentration, and we come back to this below.

These correlations beg the issue of whether wealth is alleviating credit constraints and driving enterprise or rather profits from enterprise create wealth. The structural model takes the stand that it is the former, and we shall scrutinize prediction errors of the model below.

III.4 Spatial Patterns in Enterprise Growth at the Sub-Provincial Level

Taking these initial conditions as given, our interest here lies changes in the concentration of enterprise, shown in Figure 6. Areas of agglomeration of enterprise over time, with high initial levels surrounded by villages with higher than average growth in enterprise, are present in all four provinces (colored in reds). These are also in areas of the provinces with high initial levels of wealth and/or near major intersections. Areas of convergence, with low initial levels in the base year surrounded by areas of high growth, or vice-versa (colored in pinks and light blues, respectively) are interspaced in around the agglomerating "hot spot" areas (colored red). Large areas in both provinces are stagnant, with low initial village levels of enterprise surrounded by villages with low growth in enterprise (colored dark blue).

The northeastern provinces display areas of increasing concentration (red areas) in and around areas of convergence from below (colored in pinks). There is little pattern however with respect to areas of high initial growth surrounded by lower average growth (light blues), as if the process

[18] As for example in eastern Chachoengsao and central Sisaket, which are indeed heavily agricultural, as confirmed by satellite-derived land cover (Felkner 2000).

[19] Unequal numbers of villages between the 1986 baseline regressions and the 1986-1996 growth regressions at the province level were due to missing values in the CDD data (since the enterprise index includes percent of households in cottage and retail industries).

were just beginning and there has not yet been enough time for any hollowing out or convergence from above to occur. Most areas in each province remain painted with the dark blues of low levels of enterprise and low growth. That is, each province has its own hinterland.

As reported in Table 2, growth in enterprise is more likely closer to major highways and major intersections, decreasing by .0435 percent with every additional minute of travel-time from major highways, and by .0555 percent with every additional kilometer of distance from major intersections. With the environmental and physiographic factors, growth in enterprise is more likely when soil fertility is higher, in areas of lower, flatter elevation, and in areas closer to rivers and waterways, growing by 0.5 percent with every unit increase in the soil fertility index, decreasing by 0.192 percent with every additional kilometer of distance from major rivers, and decreasing by .0139 percent with every additional meter of elevation. Figure 7 displays locations that are exogenously "favorable" in all four of these physiographic dimensions (determined using a GIS "suitability" model that ignores road structures or other man-made objects, described in Appendix I. Not every one of these favorable areas turns into an economic hot spot, but many of them do. The process of identifying economic hot spots is described below.

III.5 Creation of Proxy Variables to Measure Market and Infrastructure Access

We created a series of spatial variables at both the national and village level that capture the existing variation in village and Tambon access to markets and infrastructure and thus act as proxies for access to agglomeration synergies that could be related to enterprise growth. Precedents exist in the literature for the creation of proxy measures of the spatial extent of agglomeration market benefits (Duffy 1987; Calem & Carlino 1991). These variables allowed us to test whether enterprise, wealth, industrial density and enterprise growth varied significantly with proximity to markets and infrastructure, and to explicitly test whether residuals in our spatial structural simulations, described below, correlate with market or infrastructure access. Summary statistics for the key continuous variables are show in Table 1, with regressions results for wealth, enterprise and enterprise growth onto the key travel-time and distance proxies given in Table 2, at the national and village levels. Regression of structural model simulation residuals onto the travel-time market and infrastructure access proxies for the occupational choice and financial deepening models are shown in Tables 4 and 5. In all cases, village access was measured by calculating travel-time through the road networks[20].

Market Access Variables The LISA spatial statistical tests were used to identify significant spatial clusters of high wealth, enterprise, education and private sector credit access (commercial banks). Spatial concentrations that were statistically significant (at a probability of five percent or less) in the same location across multiple variables were identified as spatial agglomerations, and these are displayed in the bottom panel of Figure 7. An additional variable sought to provide a

[20] A number of approaches for measuring market access have been used in the literature, from the computation of direct Euclidean geographic distance, to incorporating topography or the availability of transportation networks (Hanson 1998). Although our data did include a Digital Elevation Model (DEM) for the study provinces, topographic variation was minimal, ranging across only a few hundred meters within any province, and thus was not corrected for. However, sparse transport networks of widely varying quality result in widely varying accessibility for different locations. Villages relying on low-quality dirt roads could face considerably higher travel times and travel difficulty than those adjacent to paved roads or highways. Consequently access proxies were calculated through the digital road network, using the network information on approximate road speed as a function of road type (highway versus single-lane paved versus dirt road).

proxy for accessibility to major government, financial and regulatory institutions, by calculating the travel-time of each village to the nearest Thai government District Centers: a town where major public facilities, including hospitals and important government offices, as well as private commercial banks, are located (shown in Figure 2). Proximity to these important facilities could provide substantial economic benefits, increasing the likelihood of access to credit, for example.

Infrastructure Access Variables Infrastructure provides conduits for accessibility not just to local markets, but to regional or national markets as well. Thus, in addition to the variables measuring access to provincial markets, two additional variables were created measuring variation in access to primary infrastructure systems: travel-time to the nearest major road, and to the nearest major road intersection[21].

IV A Spatial Model of Enterprise

Our first and central notion is that households choose their occupation: some will become entrepreneurs and others will not. The alternative to setting up a firm is to work in an agricultural subsistence sector or work for equilibrium wages [22]. However, space can influence this choice, as space is a key ingredient in the technology of setup costs for firms. It is in this way that we capture what is known in the literature as agglomeration externalities, though for us, proximity to a hot spot simply lowers the set up cost. The estimated values for this freely varying parameter show robust patterns consistent with our expectations, as described below. Note that we do not model externalities and transport costs per se though these complications might we well be includes in subsequent research. The second crucial ingredient is limited though exogenously expanding intermediation. The third ingredient for those without intermediation is that wealth can help overcome credit constraints. High initial wealth makes limited intermediation less damaging and an evolving endogenous wealth distribution helps to determine the dynamic system. Again wages, and the interest rate in the intermediated sector are endogenously set to clear markets. There is free and costly migration for wage earners. Households are imagined to set up firms in the village in which they reside.[23]

As noted, some villages have access to credit and intermediation services and some do not. We impose onto the model what we know from the data: the location of villages which are in the intermediated sector. This part we take as exogenous (we shall reexamine this in the alternative endogenous financial deepening model below). So, if a household is in a village without access to credit in some period, then both set up costs and capital must be financed in that period out of

[21] The identification of "major" intersections was done subjectively, based on a consideration of a priori information collected on the ground, map sources, and the spatial statistical detection of economic concentrations described below.

[22] In the model at an equilibrium wage the latter two choices can be collapsed into one sector (we abstract away from heterogeneity and talent in wage work). We do, however, vary talent/human capital in enterprise.

[23] An analysis of the data indicated that entrepreneurs are less mobile overall than are both non-entrepreneurs and the population as a whole. From the SES household survey for 1990, the percent of the total population that has moved in the last 10 years is 23 percent, while the percent of entrepreneurs who have moved in the same time period is lower, at 19 percent. In municipal areas – covering the areas of major economic agglomerations and urbanization – the ratio is even more stark: again only 19 percent of entrepreneurs are mobile compared to 35 percent for the whole municipal population. The Townsend Thai survey as administered in towns and cities of the four study provinces in 2005 also provided some evidence to support this finding, indicating that a greater percentage of business owners have resided more than 5 years in urban areas than non-business owners, in three out of the four provinces. These results indicate that migration may not play as important a role in the movement to enterprise as might be supposed a priori, at both the national and provincial scales.

accumulated, beginning-of-period wealth. If that wealth is limited, this can be constraining. A firm may operate at a small scale, and indeed some households are constrained on the extensive margin, forced to be wage earners/ subsisters. On the other hand, if a household is in a village with intermediation, then the usual neoclassical separation theorem applies: it is as if all initial wealth were put on deposit in a bank at the equilibrium interest rate, and then the household decides whether to work for wages or set up a firm, borrowing to do so at the (same) equilibrium interest rate. In the intermediated sector there is a common cost threshold below which all households are firms, regardless of wealth.

Note again, however, that the imposed credit and initial wealth distributions are not the only variables which differentiate villages. Proximity to agglomerations centers is potentially a key to the occupation decision.

With all this heterogeneity in households and villages, choices are non-trivial equilibria phenomena. For example, wages and interest rates must be computed by integrating supply and demand over space, wealth, entrepreneurial talent and, for the wage, both intermediated and non-intermediated sectors. Thus the model features simplified dynamics. An agent with end-of-period wealth W_t at date t maximizes individual preferences over consumption c_t and saving b_{t+1} as represented by the utility function

$$u(c_t, b_{t+1}) = c_t^{1-\omega} b_{t+1}^{\omega} \tag{4}$$

subject to the budget constraint $c_t + b_{t+1} = W_t$.[24] Thus, ω is the savings rate for households, a parameter which must be estimated. The optimal rules for consumption and saving will be linear functions of wealth, and so preference maximization is equivalent to end-of-period wealth maximization, period by period. One could say that we have imposed a myopic savings rate. On the other hand, evolution of the distribution of wealth is highly nontrivial.

It is assumed that costs vary spatially, as follows. There are two kinds of production technologies. In the agricultural sector, everyone earns a safe subsistence return γ of a single consumption good. This is another parameter to be estimated, though if anyone works in the subsistence sector in equilibrium, we must equate it to the wage less a switching cost, η. In the enterprise sector, entrepreneurs use capital k_t and hired labor l_t at each date t to produce the single consumption good according to a production function

$$f(k_t, l_t) = \alpha k_t - \frac{\beta}{2} k_t^2 + \xi l_t - \frac{\rho}{2} l_t^2 + \sigma l_t k_t. \tag{5}$$

This function f, with five parameters α, β, ξ, ρ,and σ, is intended as an approximation to any arbitrary production function. These five parameters must be estimated. Each wage-worker is endowed with a single unit of time and is paid a (common) market clearing wage w_t at date t.

But only one occupation can be chosen. There is a fixed cost of entry into business: that is, the household pays an initial setup cost x to start up and run a business in the period. It is this setup cost which is explicitly allowed to be related to space through the agent's location in a village and measured proximity, in distance d, to economic centers. These setup costs are assumed to be independent of total wealth b and randomly drawn from a time invariant cumulative distribution[25]

$$H(x, m(d)) = m(d)x^2 + (1 - m(d))x. \tag{6}$$

[24] Preferences are given, and space is brought in through the technology.

[25] Extensions in an alternative model allow x and b to be correlated, but this does not alter their predictions substantially. See Buera (2003) for an endogenous relationship between talent and wealth.

The support of x is unit interval $[0, 1]$, and the range of possible values for parameter $m(d)$ is $[-1, 1]$, where, again, d represents geographic travel-time to economic agglomeration centers. This class of distributions subsumes the uniform distribution at $m(d) = 0$. As $m(d)$ increases toward 1, the distribution of x becomes more skewed to the right and hence potentially efficient, low cost entrepreneurs become rare. We thus hypothesize that the $m(d)$ parameter will vary positively with geographic distance d to economic agglomeration centers. That is, the farther a village from an agglomeration center (in terms of travel-time), the less households in that village will be able to benefit from agglomeration externalities and synergies.[26] We focus here on the estimation of parameter $m(d)$ using the spatial data given the structure of the model. Given limited data , we let d take on three values: near, medium and far (see below).

In sum, an agent is distinguished by a pair of beginning-of-period characteristics: initial wealth b_t and randomly drawn entrepreneurial costs x_t. Thus, given an equilibrium wage rate w_t, an agent of type (b_t, x_t) chooses his occupation to maximize his total end-of-period wealth, W_t, assuming as in the original model there is not access to intermediation:

$$\begin{aligned} W_t &= \gamma + b_t, \quad \text{for subsisters} && (7)\\ &= w_t + b_t - \eta, \quad \text{for wage earners} && (8)\\ &= \pi(b_t, x_t, w_t) + b_t - \eta, \quad \text{for entrepreneurs} && (9) \end{aligned}$$

where

$$\begin{aligned} \pi(b_t, x_t, w_t) &= \max_{k_t, l_t}\{f(k_t, l_t) - w_t l_t - k_t - x_t)\} \ s.t. && (10)\\ 0 &\leq k_t \leq b_t - x_t && (11) \end{aligned}$$

In the original Lloyd-Ellis and Bernhardt model (2000), parameter η is a cost-of-living parameter for those outside the traditional sector. Here it is interpreted more generally as an additional cost borne for those switching out of agriculture. Equations (7) and (8) suggest that there is a reservation wage level $\underline{w} = \gamma + \eta$ below which every potential wage-worker prefers to remain in subsistence sector. Likewise, if the wage rate exceeds that reservation wage, no one remains in subsistence sector. Therefore, the model implies that wage must be $\underline{w} = \gamma + \eta$ when the subsistence sector coexists with the modern sector. We do generalize and allow subsistence income, γ, to grow exogenously at rate g_γ. Both η and g_γ are, again, parameters to be estimated.

Choices in the cross section, at any date t, are captured in Figure 8. For a given fixed wealth $b_t = b$, the fraction of entrepreneurs is determined by the distance from the abscissa axis to the occupation choice threshold, exactly so if costs are uniformly distributed. Otherwise, for a household of wealth b at geographic distance d from an agglomeration center, one integrates under the distribution *H(x,m(d))* up to the threshold line, and this delivers a prediction of the fraction of entrepreneurs, or equivalently the likelihood that any given household of wealth b will be an entrepreneur. In this we are assuming a continuum of households. Note that the higher is the initial wealth, the more likely it is that a household will be an entrepreneur. On the other end, a potentially efficient low cost household, with $x_t = x$, may end up being a wage-worker, constrained by low initial wealth b. Thus, the likelihood of any given household being in either the entrepreneurial or wage-worker sector is influenced by geographic distance d, through the cost *H(x,m(d))*.

There is also an (exogenous) intermediated sector with a weight overall which depends on the fraction of villages having measured access to credit (and we know which particular villages are in

[26] Note that x does not depend on d directly, but the greater is d the more likely x is higher.

the intermediated sector). In this intermediated sector, $R_t = 1+r_t$ is the interest rate, for simplicity both the deposit rate and the cost of funds. A household chooses to maximize end-of-period wealth W_t so that:

$$\begin{aligned} W_t &= \gamma + R_t b_t, \quad \text{for subsisters} && (12)\\ &= w_t - \eta + R_t b_t, \quad \text{for wage earners} && (13)\\ &= \pi^{\mu}(w_t, R_t) - R_t x_t - \eta + R_t b_t, \quad \text{for entrepreneurs} && (14)\end{aligned}$$

where

$$\pi^u(w_t, R_t) = \max_{k_t, l_t}\{f(k_t, l_t) - w_t l_t - R_t k_t)\}^{27} \qquad (16)$$

Below, all nine parameters, of production $(\alpha, \beta, \rho, \sigma, \xi)$, savings rate ω, subsistence γ, cost of living η, and growth g_γ, are estimated via a mean square error metric, minimizing the squared distance between 1996 village enterprise rates in the CDD data and those predicted for the model, pooling both the intermediated and non-intermediated sectors, the latter varying exogenously as observed in the data. As a principal robustness check, we also used the parameters estimated in previous work by Jeong and Townsend (2007) using SES data, and also variation of these parameters.

V Dynamic Simulation of the Spatial Occupational Choice Model With Agents in Villages

The occupational choice model was simulated using computer codes based on Gine and Townsend (2004) and Jeong and Townsend (2007). The program was adapted to treat households in CDD villages as the decision makers. However, the data are at the village level and tell us only the fraction of households, on average, in various occupations. For the initial year, we impute to each household a wealth number that is the village average and then let the model take over, allowing within village diversity. The simulation begins with the base year of 1986 as given and produces results through 1996. The results are then compared with the actual data. [In this section, it is as if the set of parameters is taken as given, but these are estimated in the next section].

Participation history in the intermediation sector was taken from the biannual CDD data for each year of the survey. Because the CDD provides participation history for multiple types of financial credit providers, an intermediation index was given a value of one if the village had reported both BAAC and commercial bank access in a given year, and a value of zero otherwise. This, then, is a conservative indicator, focusing on the two primary formal sector providers. Values of the index are imposed exogenously for 1986, 1988, 1990, 1992, 1994 and 1996. That is, in each period, an entire village is treated as belonging to the intermediated sector if its intermediation index value is one and to the non-intermediated sector if its value is zero.

Again, the simulation begins with a distribution of 1986 wealth index, with each village treated as a single data point in the distribution of wealth, as is demonstrated on the abscissa in Figure 8. To convert the wealth index of the CDD data into model wealth units, the median CDD wealth value

[27] This yields the optimal choices:

$$k^u(w, R) = \frac{\rho(\alpha - R) + \sigma(\xi - w)}{\beta\rho - \alpha^2} \text{and } l^u(w, R) = \frac{\alpha k^u + (\xi - w)}{\rho} \qquad (16)$$

was matched to the median value in the SES data, as the latter was used in Jeong and Townsend (2007)[28]. Again, initially all households in a given village are imagined to have identical wealth (as we cannot measure initial within-village diversity), though quickly within-village distinctions emerge in the dynamic simulations. As in Figure 8, even with identical wealth, $b_t = b$, not all households in a village draw the same cost x_t. The initial wealth distribution is approximated by 200 equally spaced bins, and these are re-centered as the distribution evolves and shifts to the right over time. Thus the occupation fractions (subsistence, wage and entrepreneurs) are computed for each wealth bin for each village. The entire economy, with both an intermediated sector and a non-intermediated sector, is simulated with the relative weights of these two sectors computed for each period using the actual participation index values. For example, if in 1986 there were 20 villages reporting participation index values of one, having both BAAC and commercial bank access, and 80 reporting index values of zero, no access, then the credit economy is given a weight of 0.2. An equilibrium is computed giving a common market clearing wage overall, and an interest rate in the intermediated credit sector, both determined in a bisection algorithm. Essentially, if one guesses a wrong wage (at a given interest rate) then the population fractions assigned to various occupations do not add to unity and the wage is adjusted, e.g., upwards if there is excess demand for labor. Likewise, demand for credit in the intermediated sector must equal initial aggregate wealth, summing for k_t, x_t and b_t. We thus have the model prediction and endogenous dynamic paths for prices, w_t and R_t. The predictions are quite lively in the sense that entrepreneurship, wealth, inequality, and labor share all evolve with non trivial dynamics.[29]

Our focus here is on space and geography. To predict what would happen for each village at a given distance d from an agglomeration center, taking into account the distribution costs from parameter *m(d)*, and given the computed wage and interest rate path, the actual intermediation data are used to place each village in the appropriate sector at each date, intermediated or not. We thus attain from the evolving wealth histogram a (predicted) value for the fraction of entrepreneurs for each village for the initial 1986 period and onward. We then compound the transition probabilities appropriately (from conditional to non-conditional integrating from previous time periods) to obtain the predicted fraction of entrepreneurs for each village and each time period.

VI Estimation: Testing Entrepreneurial Choice Across Time and Space

The model was simulated repeatedly for various parameter values and then assessed for its ability to capture the differential spatial patterns of development. The end-of-period 1996 CDD entrepreneur rate data is used to iteratively modify the parameters in order to minimize the mean squared error (MSE) between the predicted rate and the actual data. Our goal is to estimate parameters, especially how m(d) varies across space (but see robustness check below). Unfortunately, there are not enough villages at every exact geographic distance d (on a continuum from the villages "closest" in terms of agglomeration access to those farthest away) to estimate m(d) reliably across all points d in space. Thus we approximate m(d) by putting villages into 3 geographic distance bins. Specifically, the full sample was stratified into three discrete bins with equal number of villages in each bin, partitioned across space by variation in travel-time to major intersections.

Travel-time to major intersections correlated most significantly with non-spatial simulation prediction residuals in a robustness check (as reported in Table 4 below), and thus appeared to capture

[28] Again, we have adopted the simulation codes for our use here, and our principal robustness check is to simulate at the original Jeong and Townsend (2007) parameter values.

[29] See Jeong & Townsend (2007) for an analysis of these variables but with no spatial component.

otherwise significant unexplained variation in spatial simulation predictions.

VII Spatial Model Best Fit Simulation Results

We arrive at a superior fit to final-period data in our primary estimation than in a non-spatial simulation. The estimated m(d) values increase monotonically with increasing geographic distance d, and decreasing accessibility to agglomeration centers, as shown in Table 3. Again, a higher m value indicates a greater skewness towards higher costs in the population distribution of villages. The estimated m parameter value increases from -0.1013 for villages in bin 1 (which have superior travel-time accessibility to agglomerations) to .0.6202 for bin 3 (the villages which have the highest travel-time and thus poorest access to agglomerations). This monotonic increase in costs across space was confirmed in multiple robustness tests, below (see Table 3).

Specifically, in terms of overall correlation with the actual final-period data, the spatial model had a positive correlation coefficient of .1638, with a p-value of 0.000, compared to .0746, with a p-value of .0444, for the non-spatial simulation. Also, the end-of-period prediction errors of the specified simulations relative to the data were regressed onto the market and infrastructure access variables. We normalize so that positive error indicates model overprediction. We also regressed these residuals onto wealth and education. Results are displayed in Table 4. For the non-spatial model (lacking spatially varying set up costs), all three agglomeration proxies are statistically significantly correlated with the end-of-period residuals. Likewise regressing the residuals onto wealth and education only, wealth is significant for the non-spatial residuals but is not with the spatial residuals.

Finally, to make the spatial comparison between the simulated and actual data more apparent, and to isolate statistically significant residual spatial concentrations, a LISA statistic was calculated and mapped for the end-of-period 1996 prediction errors. The results are shown in Figure 9. Spatially concentrated errors are dramatically reduced.

A major criterion that we use to judge how well this spatial structural model is predicting is its ability to replicate the patterns of increasing concentration that we have see in the data. Of course the model is run at the provincial level, only, not the national level, so here we focus on enterprise growth patterns within and across the four provinces. To examine this, a Bivariate LISA map of simulated enterprise in the base year surrounded by growth 1986-1996 is displayed in Figure 10, and can be compared to a Bivariate LISA map of actual enterprise growth 1986-1996 shown in Figure 6. As is evident from Figure 10, the model does remarkably well in replicating areas with initial high levels of enterprise for a village surrounded by villages with subsequent high growth in enterprise (colored in reds), compared to what we see in the actual data (Figure 6). These include the central corridor running though the western part of Lopburi, the increasing industrialization connecting Bangkok to the Eastern Seaboard and running though western Chachoengsao, and in environmentally advantageous areas in the two provinces of the Northeast. While there is not a one-to-one correspondence with the actual data, especially in Northeast, the prediction that increasing concentration should appear correlated with intersections of major highways is intact. Areas with initially low levels surrounded by high enterprise growth (colored in pink) lie in and around areas of increasing concentration in all provinces. Thus the model is consistent with enterprise "catching up" in areas contiguous to areas of initially high levels and not elsewhere. The correlation coefficient between the actual and simulated Bivariate Moran maps of enterprise growth was .1241 with a p-value of .023, confirming the ability of the model to simulate not only the levels but the spatial patterns of this growth.

VIII Robustness Checks

We have conducted various robustness checks on these results, displayed also in Table 3. To check on the monotonic increase in the enterprise setup-cost m(d) parameter with increasing geographic distance from agglomerations and infrastructure, we conduct four robustness checks. First, we simulate the spatial model using the original parameter values taken from Jeong and Townsend (2007), which were estimated from data at the national level, allowing only the m(d) cost parameter to vary across space (robustness check #1). We also allow a subset[30] of the Jeong and Townsend (2007) parameters to vary across the whole sample, while at the same time allowing m(d) to vary across space according to two different spatial specifications: first across three bins of decreasing accessibility to agglomeration/markets with equal number of villages in each, and second using five bins stratified equally by variation agglomeration access, robustness checks #2 and #3.[31]

Although time paths are sensitive to parameter values[32], the monotonic increase in m(d) with decreasing accessibility across space appears with robustness checks 1 and 3, while for robustness check 2 the m(d) estimated cost increases considerably from -0.21 in bin 1 to 0.36 and 0.27, respectively, in bins 2 and 3, with decreasing agglomeration accessibility. Thus, despite slight loss of the monotonic increase from bins 2 to 3 in robustness check 2, the monotonic increase of the m(d) parameter largely holds across these checks. In addition, spatial patterns in prediction errors - as measured both by the LISA map of residuals and by regression of prediction residuals onto the agglomeration proxies - were also robust: spatial modification resulted in both less residual significant correlation with agglomeration proxies and fewer significant residual spatial clusters (results not shown).

A fourth check on the monotonic increase of m(d) with poorer agglomeration access, robustness check #4, takes the opposite approach, rather than utilizing parameters estimated from other studies, it allows all parameters of the model (technology, preferences, etc.) to be estimated for each spatial bin separately. This produces the best fit against the data, but suffers in the interpretation. For example the savings rate decreases with distance, but the cost of transition out of subsistence agriculture decreases also. Likewise, exogenous growth in subsistence agriculture increases with distance. Related, perhaps, the part of the marginal product of capital captured by parameter α is increasing, while β is not monotonic (the parameters of the marginal production of labor, ξ and ρ, and the interactive term σ, are not monotonic either). In short, there seems to be too many free parameters moving to fit the data. We are, thus, not surprised m(d) is not monotonic in check #4.

As a final robustness check (see #5) on the estimated monotonic increase in m(d) across space, we re-estimate the model but this time binning villages across space by travel-time distance to environmentally advantageous areas (as identified using our digital spatial physiographic data using the GIS "suitability" model, described in detail in the Appendix). These endowments do not change

[30] Jeong & Townsend (2007) estimate well the parameters of the production technology using intial base year data and see how the model fits by comparing model predicted dynamics to what actually happens. Likewise, in this paper, we were tempted to use the parameters of Jeong & Townsend (2007) as best fits to micro SES data, so as to focus here on the village level data and spatial patterns, varying only the costs parameter m(d). However, dynamics in the LEB model are not so sensitive to parameters α, β and σ, so we allowed ourselves the ability to restimate these in the CDD data. On the other hand, if we go to the other extreme of varying all parameters, including ξ and ρ that have more influence on dynamics, then the goal is to allow the model to do as well as possible in fitting end of sample village level enterprise frequencies. Our preferred estimates do just this.

[31] Note that these latter two robustness checks were taken from earlier work and used a slightly different mean-squared error calculation.

[32] As, for example, with the robustness check # 1 (imposing values estimated from Jeong & Townsend [2007]): the m(d) parameter goes to extreme values quickly, reaching its maximum value of 1 for bins 2 and 3.

over time and so would not be sensitive to endogenous physical infrastructure, and thus this also serves as a check on our use of the distance from infrastructure variable to estimate these costs. Furthermore, this simulation allows us to check on the continuing role of natural endowments in enterprise location and growth, given that we establish above that the spatial location of enterprise and enterprise growth varies significantly with respect to natural endowments. The results are largely as we would now expect: m(d) increases with distance from natural endowments, from .17 in villages closest to endowments (bin 1) to .62 and .43 in bins 2 and 3. The increase is not uniformly monotonic, but there is a significant jump in costs from bin 1 to both bins 2 and 3. Bivariate LISA mapping of the simulated results are similar to those produced in our baseline simulation (Figure 10) but not as good a prediction of actual patterns: the coefficient of correlation with the actual data is .0985, compared to .1638 above.

There is nothing automatic about these spatial concentration patterns. Indeed to gage the importance of credit expansion, we eliminate it, first freezing credit use at its initial l986 level (robustness check #6), or then getting rid of the intermediated sector entirely (robustness check #7). The predictions for increasing concentration using the Bivariate LISA technique are shown in Figure 11 for credit fixed (robustness check #6). Though there is increasing concentration in the Northeast, we appear to miss this concentration in the Central provinces, especially in Chachoengsao where we have no areas of increasing concentration (colored red). When we get rid of the intermediated sector entirely (robustness check #7) as shown in Figure 12, the cross-province deviations from reality are enormous, with Lopburi and Sisaket showing only enterprise growth stagnation (in blues) and Chachoengsao showing only enterprise growth and concentration (in pinks and reds). Correlation coefficients between both of these checks and the spatial patterns in actual enterprise growth were .0634 (p-value of .049) for the credit fixed check and .0314 (p-value of .079) for the no credit check (as compared to .1638 when credit is included). We conclude that observed credit expansion is a key part of the model.

IX A Model of Financial Deepening

Again, as in the data, the strong results from the occupational choice structural model simulations take as given the spatial distribution and expansion of the intermediated sector[33]. If intermediation were endogenous, are our earlier results biased in some way? That is, selection into credit somehow driving the results more than selection into enterprise? In fact, the bottom line of this section is the opposite. Making credit endogenous goes against the data and would have given us misleading and underestimated agglomeration gains.

Thus, in this section we examine the endogeneity of credit through the lens of another structural model where credit is endogenous and that is well known in the theoretical literature. We estimate and test this model, again incorporating agglomeration effects allowing spatially varying transactions costs as a function of access to agglomeration benefits. Quite conveniently for us here, the model is already calibrated in macroeconomic work in Thailand (Townsend and Ueda 2006). We fix parameters based on that earlier work. As will become clear, if we had estimated parameters anew, these values would be influenced by what are in reality policy distortions, not spatial costs.

In passing we note something important about our methods: adding spatially varying transactions costs into a structural model does not guarantee the researcher success in explaining spatiotemporal data. This serves to further reinforce the positive results we achieved in the enterprise

[33] Private commercial banks correlate positively with wealth in the standard OLS regression, while the public BAAC correlated negatively.

model above.

More specifically, we use in this section a spatially specified version of Greenwood and Jovanovic (1990) that models the choice of whether agents join the financial system (referred hereafter as the financial deepening model). In this model there is a spatially varying fixed cost for entry into the financial system (capturing roads, bank infrastructure and/or household learning). Again we combine all these things to keep the model traceable. This cost is nontrivial, so relatively low wealth households choose to remain in autarky. But other households at the same given geographic distance from the bank have wealth higher than a key threshold and so have joined the financial system, taking advantage of better information about macro shocks and improved allocation of risk for idiosyncratic shocks (full risk sharing). Key decisions for those in autarky are how much to save out of income, and, given savings, how much to put into a risky business. Returns in enterprise are subject to an aggregate and idiosyncratic shocks. The alternative is to put savings into relatively safe agricultural activity, with low return. A key decision for those in the financial sector is how much to save, but investment is in effect under advise of the bank with better information, and all idiosyncratic shocks are pooled. In sum, in this model, the higher is the fixed cost of entering the financial system, the higher must be the critical value of wealth, and so the larger is the fraction of the villages who choose to stay in financial autarky.

Thus consider an economy with a continuum of villages on the unit interval $[0,1]$. Each village persists for an infinite, discrete number of periods $t = 0, 1, 2\cdots,$. All people in a village are considered to be identical and we restrict attention here, in this model, to one "representative" agent per village in all periods. For every village j, there are two technologies available that can convert the capital investment i_{jt} at date t into a yield $y_{j,t+1}$ at date $t+1$. One technology yields a safe but relatively lower rate of return δ per unit capital, and the other gives a risky rate of return $\left(\zeta_{t+1} + \epsilon_{j,t+1}\right)$ with higher expected value, where ζ_{t+1} represents a common aggregate shock and $\epsilon_{j,t+1}$ an idiosyncratic shock specific to village j. The aggregate shock ζ_{t+1} is governed by a time-invariant uniform distribution with support $[\underline{\zeta}, \overline{\zeta}]$, and the i.i.d. idiosyncratic shock $\epsilon_{j,t+1}$ is governed by a time-invariant uniform distribution with support $[-\overline{\epsilon}, \overline{\epsilon}]$ with $E\left(\epsilon_{j,t+1}\right) = 0$. Let $\eta_{j,t+1} = \zeta_{t+1} + \epsilon_{j,t+1}$ be the composite shock, and Ψ^{η} be its distribution function. The lower bound for the composite shock is positive, i.e., $\underline{\zeta} - \overline{\epsilon} > 0$.

Each village j can run either of two technologies, with portfolio share ϕ_{jt} for the risky one, so that the beginning-of-period wealth $k_{j,t+1}$ at $t+1$ can be written as

$$k_{j,t+1} = \left[\phi_{jt}\eta_{j,t+1} + (1-\phi_{jt})\delta\right] i_{jt}. \tag{17}$$

At the beginning of period t, village j allocates its current disposable wealth k_{jt} into current consumption c_{jt} and capital investment i_{jt}, namely $k_{jt} = c_{jt} + i_{jt}$. The objective is then to maximize the discounted utility stream, with contemporary utility *u(c)*. That is, the objective function for village j is:

$$E\sum_{t=0}^{\infty}\beta^t \frac{c_{jt}^{1-\sigma}}{1-\sigma} \tag{18}$$

subject to the sequence of resource constraints of $k_{jt} = c_{jt} + i_{jt}$ and law of motion for those in autarky, (17).[34] Villages differ in their wealth level in each period t for two reasons: first, the initial endowment k_{j0} at date 0 may be different across villages j, distributed as in the cross-section by

[34] In their original model, Greewood and Jovanovic (1990) consider a log utility function, a special case of the CRRA preferences with $\sigma \to 1$.

a cumulative distribution function H_0, measured in the data. This is a given initial condition. Second, the history of realizations of random shock $\{\epsilon_{j,s}\}_{s=0}^{t}$ differs across villages.

In contrast a formal financial intermediary can run a countably large number of trials for the risky technology and get advanced information on next period's return to the risky project. Then, the intermediary invests in the risky project only if this return exceeds the safe return δ. Furthermore, the intermediary can diversify the village-specific idiosyncratic shocks by pooling participants' resources. It can pay back a promised return $r(\zeta_{t+1})$ per unit of capital invested at time t contingent on the realized aggregate shock ζ_{t+1}. Therefore, every village has a gross gain to joining the financial system. The model does not seem to distinguish among various possible financial sector providers, but we shall do this below.[35]

But, intermediary trading arrangements are costly, as in Townsend (1978). There is an initial fixed cost $q(d)$ of incorporating each village j into the formal financial sector and a variable cost of $(1-\gamma)$ in proportion to the amount of funds each village invests in the coalition. Here, using data on the geolocation of villages and the GIS, we allow fixed entry costs q(d) to vary with geographic distance d to the district centers.

Given this entry fee, $q(d)$, not everyone immediately joins the financial system. Only villages whose wealth levels exceed some critical level k*(d) are willing to join, hence the choice of financial sector participation is again constrained by wealth.

The decision making of villages can be characterized by the pair of value functions: v^0, the value function of non-participant villages, and v^1, the value function of participant villages. A village j at date t with initial wealth k_{jt} which currently is not in the formal intermediated sector chooses the total investment and the portfolio share between safe and risky projects according to a functional equation:

$$(P1) : v^0(k_{jt}) = \max_{i_{jt},\phi_{jt}} \left\{ u\left(k_{jt} - i_{jt}\right) + \beta E_{\eta_{j,t+1}} \max\left[v^0(k_{j,t+1}), v^1(k_{j,t+1} - q(d))\right] \right\} \qquad (22)$$

$$\textit{subject to } (17),$$

where $E_{\eta_{j,t+1}}$ is the expectation with respect to the composite shock $\eta_{j,t+1}$. Note the decision to join is incorporated next period. Likewise, a village j at date t with initial wealth k_{jt} which currently is in the formal intermediated sector chooses the total investment and the portfolio share between safe and risky projects according to the functional equation:

$$(P2) : v^1(k_{jt}) = \max_{i_{jt},\phi_{jt}} \left\{ u\left(k_{jt} - i_{jt}\right) + \beta E_{\zeta_{j,t+1}} v^1(k_{j,t+1}) \right\} \qquad (23)$$

Financial participation constrained by wealth and geographic distance is thus the key micro foundation of the financial deepening model. Let D_{jt} denote the participation decision of village j

[35] There are key restrictions on the parameter space to make the above economy work properly. In order to assure the benefits of intermediation and the incentive to invest positive amounts in production every period, we need to assume the following condition:

$$E\left\{r(\zeta_t)\right\} > E\left\{\zeta_t\right\} > \delta. \qquad (19)$$

To avoid the economy shrinking to negative infinity, we need:

$$\delta > 1/\beta. \qquad (20)$$

Due to the linear production technology, unbounded growth is possible in this model, and in order to make the economy not explode in utility terms, we also need:

$$\beta E\left\{r(\zeta_t)^{1-\sigma}\right\} < 1. \qquad (21)$$

at date t, which assigns 1 if village j decides to participate in the financial sector, and 0 otherwise:

$$\begin{aligned} D_{jt} &= 1, \quad if\ v^1(k_{jt} - q(d)) \geq v^0(k_{jt}) \\ &= 0, \quad if\ v^1(k_{jt} - q(d)) < v^0(k_{jt}). \end{aligned} \tag{24}$$

Again, there exists a unique critical value $k{*}(d)$ such that the participation decision in (24) is equivalent to

$$\begin{aligned} D_{jt} &= 1, \quad if\ k_{jt} \geq k^*(d) \\ &= 0, \quad if\ k_{jt} < k^*(d) \end{aligned} \tag{25}$$

There is no closed form solution for $k^*(d)$ because there are no analytic solutions to the dynamic program in (22). However, $k^*(d)$ is a function of the underlying parameters of the financial deepening model, $(q(d), \delta, \beta, \sigma, \gamma, \underline{\zeta}, \overline{\zeta}, \overline{\epsilon})$.

Initial model parameters used are those calibrated in Townsend and Ueda (2006) (see Table 3, bottom panel). These are in turn strikingly close to the maximum likelihood estimates of Jeong and Townsend (2007). These were taken as initial starting values because, as with the occupational choice structural model, they fit the Thai macroeconomic aggregate trends well, and also fit the aggregate dynamic mean trends well at the provincial scale. Again, as with the occupational choice model, our primary goal is to estimate the key fixed setup cost $q(d)$ across space at the provincial scale.[36]

Specifically, the support of the idiosyncratic shocks ε was given a lower bound of –0.6 and an upper bound of 0.6 to represent the difference between the top and bottom one percent returns of income/capital ratios for those in non agricultural businesses, obtained from the Townsend Thai survey that was conducted in Thailand (Townsend, Paulson et al. 1997, or see cier.uchicago.edu). The discount rate β was set at 0.96, following the business cycle literature. The fixed cost free parameter q was set initially at 5 in model units of capital but see below for its estimation over space. The upper and lower bounds values for the aggregate shock were derived from data on Thai real per capita growth from 1976 to 1996, under the assumption that the underlying variation of shocks would be larger than the difference between the minimum and maximum of this growth rate. The actual difference was about 8.7 percent, and thus the range ζ was set at 10 percent. The mean of ζ was determined by picking one that minimized the mean squared errors between the actual Thai growth rate and predicted analytic path of the model. The resulting mean was set at 1.097, giving the upper and lower bounds for ζ of 1.047 and 1.147, respectively. The safe return δ (value of 1.054), risk aversion σ (value of 1), and transaction costs 1-γ (value of 0) parameters were also taken from Townsend and Ueda (2006).

[36] It becomes immediately apparent that the key cost parmeters q(d) are counter to what we anticipate. We might have allowed ourselves to re-estimate all parameters. However, estimation (rather than calibration) of all parameters is beyond computability at this point in time (see discussion in Townsend and Ueda 2007). Economically, if we had allowed other parameters to vary across space, in addition to q(d), other parameter values would have to move in strange ways to reconcile these anomalies. That is, productivity of enterprise would have to decrease in towns, agents would have to be more risk averse, and idiosyncratic shocks would have to be larger. So in the end we use the Jeong & Townsend (2007) original parameter estimates and accentuate the anomalous cost parameters that, we believe, have to do with policy distortions.

X Simulation of the Financial Deepening Model

Our results will be clearer if we begin with the model in which spatial access costs $q(d) = q$ are the same for all villages regardless of geographic distance, and the same for both financial sector providers (BAAC and commercial banks). Then, each village has the same critical value of the wealth k*(d)=k*. The 1986 CDD empirical distribution of the village wealth index was converted into a distribution (a Cumulative Distribution Function – CDF) of 1000 bins. Data from the CDD survey are used to determine the percentage of village participation in the base year, namely 27 percent. For intermediation here we use, for now, participation in both commercial banks *and* BAAC, which is the same intermediation index used in this paper, above, for the occupational choice simulation. (Note, however, in this section below we make an important further distinction between private and public credit providers.) We then obtain the corresponding critical level of wealth $\hat{k}$ in the data, i.e., the wealth level at which 27 percent of the villages have higher wealth values in the CDF. This is where the uniformity in costs comes into play and also suggests the modification we perform momentarily. This wealth value is matched to the critical value of capital k^* obtained from the model at its given parameter values. The scalar to convert model units to CDD wealth index units (to make predictions) is thus equal to $\hat{k}/k^*$.

Beginning with the initial 1986 wealth distribution, and assigning initial observed wealth to each village, the model economy is simulated for 11 periods (one for each year, 1986-1996), treating villages as agents. Each simulation produces a participation index and wealth index prediction for each village and time period as a function of realized (simulated) idiosyncratic and aggregate shocks. 500 simulations were run, and then averaged. Model units were converted back into CDD wealth index units.[37]

XI Endogenous Credit Tested Dynamically

The financial deepening simulation with common entry costs produced predicted probabilities of credit intermediation and wealth for all time periods. The simulation does an excellent job of capturing overall dynamic trends. The simulation closely matches the 1996 actual mean for intermediation with between 46 and 47 percent of all villages having reported intermediation, increased from 27 percent in 1986. The overall end-of-period means for actual and simulated wealth are close (46.96 for actual versus 43.15 for simulated) as are the cross-sectional standard deviations (not shown) of 38.83 for actual versus 36.23.

On the other hand, the simulated intermediation residuals, displayed visually in Figure 13[38], reveal a correlation between model *overprediction* and proximity to base-year economic clusters. The overprediction areas include western Chachoengsao (infrastructure corridor adjacent to Bangkok),

[37] See Townsend and Ueda (2006) for a discussion of best fit path versus the Monte Carlo average path. We use the average of the Monte Carlo simualtions as the best overall prediction. Another possibility is to pick the particular simulation that is the best fit relative to the time series, here we do not make much use of the temporal predictions and so use the average value.

[38] To make the spatial residual trends more apparent visually in Figure 13, a spatial statistical smoother was run on the residual values in the villages, namely a window-average smoother with the "window" defined at the 10 nearest neighboring villages. Using this method, residual values for the 10 nearest neighbors for a given village are averaged, and then applied to the village in question. Window-averaged values for villages are then displayed using Thiessen polygons rather than village geo-points to make the spatial trends more visually apparent.

southwestern Lop Buri, and clusters in Buriram and Sisaket. The apparent spatial pattern in the residuals - visible in Figure 13 - was confirmed with a LISA map.

Evidently, if we had taken the endogenous access to credit predicted by this financial deepening model and placed it into the occupation choice model, we would have put in too much credit in these urbanized areas and would have over predicted enterprise frequency, or otherwise we would have had to raise the business entry cost for those near agglomeration centers.

At the same time, there is some indication that under-prediction clusters in the endogenous deepening model tended to occur outside the areas of major urbanization, with lower wealth levels (areas as we have seen with fewer businesses). These included eastern Chachoengsao and Lop Buri, areas in central Lop Buri and much of Buriram and Sisaket.

If we had used endogenous access to credit predicted by this financial deepening model in the occupation choice model, we would have put in too little and under predicted the frequency of business there, or otherwise would have had to lower the business entry costs. In sum if we had not used observed credit access as exogenous credit, we would have undercut the estimated cost gradient of the previous occupation choice model, raising costs near urban centers and lowering them in the hinterland.

More specifically, the financial deepening intermediation index end-of-period simulation residuals were regressed onto the agglomeration proxies, wealth and education, with provincial fixed effects. Results are shown in the upper panel of Table 5, which also displays residual regression results for the financial deepening end-of-period wealth simulation (as well as the further simulations described below contrasting results for BAAC and commercial banks). The results confirms significant and negative results for all three proxy variables measuring village variation in access to infrastructure as well as significant and positive results for both wealth and education. That is, the model over-predicts for those with high wealth and education[39].

We now explore the possibility that public and private credit providers exhibit varying behavior with respect to economic agglomerations and that access cost q(d) varies systematically with geographic distance from agglomeration centers. To keep dimensions simpler, the full sample was stratified into three bins by equal number of villages along the axis of the time-travel to major intersections. In addition, the model was simulated separately for commercial banks only, and then the BAAC only. This allowed for the estimation across space of the variation in costs of using each major financial provider, as captured by the $q(d)$ parameters, allowing direct tracking of financial system access costs as a function of spatial access to economic agglomerations.

The model tends to underpredict the BAAC expansion over time for each bin. The BAAC is a chartered institution with a mission of getting credit to farmers and those in rural areas and thus almost certainly benefits from policy distortions for this purpose. At the same time, the model overpredicts the expansion of commercial banks in bin 1, near infrastructure, almost as if there were policy barriers to entry. This may be the flip side of the same policy distortion. The credit market in Thailand is segmented, featuring a role for the BAAC in agricultural and rural areas. With increases in wealth, the initial financial plan with its target segmentation may have left the educated, young, middle class near towns and cities underserved.

[39] The end-of-period residuals were also regressed onto the variables measuring base-year spatial clusters, with controls for education, shown in the lower panel of Table 5. The spatial cluster variables have significant positive and negative coefficients that follow the same remarkably consistent pattern: positive coefficients for the binary variables (with a value of one for a village that is inside existing agglomeration areas, and zero if it is outside) and negative coefficients with the continuous agglomeration proxies (which have increasing values with increasing geographic distance from agglomeration centers). In short the model is overpredicting in and near agglomeration centers.

Again, we estimated cost $q(d)$ by provider and geographic distance d from agglomerations using the structure of the model. Initially, all parameters were fixed (as reported in Table 3) and the k^* value computed. We verified that k^* and q are proportional to each other. For a particular initial spatial bin, credit provider, and the base year, we got from the CDD data the fraction of villages f reporting access to credit in that year. Considering the observed empirical distribution of wealth, the critical value $\hat{k}$ was calculated such that f percent of villages have higher wealth values. Conversion of empirical wealth units to model units was then fixed by calculating the ratio of $\hat{k}/k^*$ for that baseline spatial bin. Then, for all other bins b, credit provider p and indeed for each year t, we used f_{bpt} from the CDD data and computed the corresponding $\hat{k}_{bpt}$. We then convert each $\hat{k}_{bpt}$ to k^*_{bpt}, hence to q_{bpt} via proportionality. This allows us to see the ratio of costs for access to credit for each bin (and provider and year, as well).

Calculated relative costs are reported in Table 3, which reveals that q was 2.33 times higher closer to roads. This of course runs contrary to the presumed hypothesis that nearness to district centers should lower costs. Likewise, and related, BAAC costs are systematically lower than for commercial banks. In sum, the estimated $q(d)$ parameter seems more able to pick up the impact of policy distortion than the actual costs.[40]

XII Conclusions

We find salient patterns of increased agglomeration of enterprise within a high growth Asian country, coupled with more subtle patterns of contiguous geographic convergence which leave stagnant areas behind. Exogenous favorable physiographic conditions are also an important factor, especially at the local level within provinces,. These results are obtained with a variety of advanced spatial statistical methods using both national and village level socioeconomic and biophysical data, merged in a Geographic Information System (GIS). We fit a structural, micro-founded model of occupation transitions with fine-tuned geographic capabilities to these village data and replicate the salient spatial facts. Key ingredients appear to be costs which increase monotonically with distance to key infrastructure and markets, credit constraints on occupation choice which are mitigated with spatially varying and endogenously increasing wealth, and exogenous and spatially varying expansion of financial service providers. Furthermore, a structural model of endogenous financial deepening applied at the village level reveals costs which appear to decrease with distance to major infrastructure and markets, indicating policy distortions and the role of a key government development bank, which help drive the growth and inequality of local and regional economies.

We are encouraged by our fact finding mission, our spatial disaggregation and analysis of the data, and use of structural models. The next steps on both dimensions are compelling. The first, as regards measurement, is to try to secure data from other countries at high spatial resolution spanning growth episodes. The second is to extend the models so that they are more realistic on several dimensions: potential spill-overs, inclusion of transport costs, forward looking, variety in financial services, and endogenous industrial organization of financial markets with public and private providers that, evidently, have divergent objectives.

XIII Appendix: Modeling Areas of Favorable Geography Using a GIS

[40] If we had also allowed other parameters to vary across space in addition to $q(d)$ (say δ, the safe return parameter, or $(1-\gamma)$, the level of transaction costs), their values would likely have been also obscured by this apparent policy distortion.

Suitability Model

When Adam Smith noted that economic activity tended to cluster along rivers and waterways, he was acknowledging the fact that geography and environmental conditions have always played a large role in determining the location of agglomerations of economic activity. Building in flat elevations is less costly than on steep slopes. Planting crops in areas with rich soils and consistent rainfall is preferable to planting in areas with poorer soil and rainfall conditions. And, as Smith described, distance to rivers and waterways provides a means for lower cost (and reliable) shipping of goods.

For this study, we are interested in the factors driving the growth of enterprise in Thailand, including the influence of geography and the environment. Consequently, we are interested in identifying which geographic locations are relatively optimal – in terms of geographic and environmental conditions – for the development of economic activity. Further, this allows us to consider whether areas that are geographically favorable to agglomerations actually develop them.

Suitability Modeling in Thailand to Identify Locations of "Favorable Geography" Because we have access to extensive, high-quality GIS physiographic data describing of our study area, we used a GIS "suitability" modeling process (Goodchild and Gopal 1989; Joerin, Thériault, and Musy 2001; Store and Kangas 2001; Jankowski 2006; Malczewski 2004) in a raster GIS context to identify locations in our study areas that are relatively optimal for the growth of economic agglomerations, based on a set of logical assumptions. A "suitability" modeling approach seeks to identify optimal locations as a function of a set of weighted, discrete input criteria that are combined algebraically to identify spatial locations meeting that criteria.

We begin with a set of simple assumptions about physiographic conditions and economic activity in Thailand:

a) we assume that areas of higher soil fertility will be economically preferable, because of the potential for greater agricultural productivity;

b) since consistent rainfall supply is a crucial factor for farmer's long-term economic security, and because frequent variation in rainfall intensity makes it more difficult for farmer's to plan and make planting decisions, we assume that areas of lower rainfall variability will be more favorable for economic growth.

c) we assume that distance to rivers and waterways in general provides an economic advantage, because of the better access to a reliable and low cost method of transport;

d) furthermore, we assume that spatial distance to intersections of rivers and waterways are even more favorable, because they provide the option of sending transporting goods or people in more than one direction using a low cost transport system;

e) we assume that it is preferable (lower cost) to build in topographically flat areas, compared to steep slopes, and that over time it will be cheaper to conduct economic activity in flatter areas. For these reasons, economic agglomerations are more likely to arise in topographically flat areas.

Input GIS data layers to perform this analysis included:

- Data on variation in soil quality across Thailand, including a variety of soil quality indicators;
- Spatial data on rainfall variation over time;
- River and stream networks, in GIS vector format;
- A digital elevation model for Thailand giving elevation variation.

Soil Fertility Soil fertility data was obtained from the Thailand Environmental Institute (TEI) and included numerous indicators of soil conditions. Data on variation in cation exchange capacity of the upper portion of the topsoil was binned and used to product an index of soil fertility. This index ranked all areas of Thailand on a scale of 1 to 9, with 9 having the highest potential fertility for agricultural productivity. This data was converted to a regular grid of raster pixels, with each pixel having a value between 1 and 9.

Rainfall Variation Rainfall variation was derived from an analysis of variation over time using monthly Thai rainfall data from 1951-2000 obtained from more than Thai meteorological stations. For each calendar month, the standard deviation of total monthly rainfall from 1951-2000 was calculated for each meteorological station, and then an Inverse Distance Weighted geo-statistical interpolation was performed to create a continuous surface of rainfall variation, with higher values indicating higher rainfall variability (higher standard deviation) over time. The hypothesis here was, of course, that greater variability in rainfall would produce conditions less favorable for agricultural productivity over time. As with the soil fertility raster layer, this layer was also in raster format, with each grid cell's value reflecting the degree of rainfall variation for that location.

Distance to Rivers and Streams, and to River Intersections Highly detailed Thai GIS data on river and stream network locations nation-wide was obtained from TEI, and the two highest orders of rivers/streams were extracted. For the river/stream linear networks, a gridded raster surface was generated with the value of each grid cell indicating the spatial distance (in meters) from that grid cell's centroid point to the nearest river. A separate raster layer was created describing distance from river intersection points.

Elevation/Slope Beginning with a digital elevation model for Thailand obtained from TEI, the layer was rasterized into a regular grid with each cell's value reflecting the elevation above sea level (in meters) of that location. An kernel algorithm was then applied to this layer which calculated the relative slope of each pixel as a function of the elevation values of its neighbors (relative to its elevation value). These values were then binned into an index, with slopes higher than 40 percent given null values, and values lower than 40 percent a value of 1.

Combining the Input Layers All the above input layers were normalized to indices of equal range and then algebraically summed in a raster context, with all layers given equal weights, thus:

soil fertility index + (1/rainfall variability) + (1/distance to rivers) + (1/distance to river intersections) + slopes less than 40% (binary variable, 1/null) = favorable locations raster, as follows:

The locations with the highest clustered values in this output layers were identified with points, and these are displayed in the top panel of Figure 7.

Finally, the distance of each CDD village to the nearest of these points was calculated, with this value a measure of relative distance to favorable geography for each village.

XIV References

Anselin, L. "Interactive techniques and exploratory spatial data analysis." Geographical Information Systems: principles, techniques, management and applications (1996).

—. “Local indicators of spatial association-LISA.” Geographical analysis 27, no. 2 (1995): 93-115.

—. “The Moran scatterplot as an ESDA tool to assess local instability in spatial association.” Spatial analytical perspectives on GIS (1996): 111-126.

Aghion, Phillippe, and Patrick Bolton. 1997. “A Theory of Trickle-Down Growth and Development.” Rev. Econ. Studies 64 (April): 151–72.

Araujo, C. “Spatial patterns of non-agricultural rural employment growth in Mexico during the 90s.” University of California, Berkeley (2003).

Assuncao, J., S. Mityakov, and R. Townsend. Commercial vs. Government Development Banks. working paper, University of Chicago, 2008.

Banerjee, A., and A. Newman. "Occupational Choice and the Process of Development." Journal of Political Economy 101, No. 2 (1992): 274-298.

Becker, G. S., K. M. Murphy, and R. Tamura. “Human capital, fertility, and economic growth.” Journal of Political Economy (1990): 12-37.

Buera, F. “A dynamic model of entrepreneurship with borrowing constraints.” Northwestern University (2003).

Burchfield, M., H. G. Overman, D. Puga, and M. A. Turner. “Causes of Sprawl: A Portrait from Space*.” Quarterly Journal of Economics 121, no. 2 (2006): 587-633.

Calem, P. S., and G. A. Carlino. “Urban agglomeration economies in the presence of technical change.” Journal of Urban Economics 29, no. 1 (1991): 82-95.

Chou, Y. H. “Exploring spatial analysis in geographic information systems” (1997).

Conley, T. G., F. Flyer, and G. R. Tsiang. “Spillovers from local market human capital and the spatial distribution of productivity in Malaysia.” The BE Journal of Economic Analysis & Policy 3, no. 1 (2003): 5.

Duffy, N. “Returns to scale behavior and manufacturing agglomeration economies in US urban areas.” Regional Science Perspectives 17, no. 1 (1987): 42-54.

Ellison, G., and E. L. Glaeser. “Geographic concentration in US manufacturing industries: a dartboard approach.” Journal of political economy 105, no. 5 (1997): 889-927.

—. “The geographic concentration of industry: Does natural advantage explain agglomeration?.” American Economic Review (1999): 311-316.

Felkner, John. “A Statistical Predictive Model of Land Use Change in Thailand, 1989-1999, Using a Tree Classification Approach.” Harvard Graduate School of Design (2000): 400.

Foster, A. D., and M. R. Rosenzweig. “Agricultural Productivity Growth, Rural Economic Diversity, and Economic Reforms: India, 1970-2000*.” Economic Development and Cultural Change 52, no. 3 (2004): 509-542.

—. “Learning by doing and learning from others: Human capital and technical change in agriculture.” Journal of political economy (1995): 1176-1209.

Fotheringham, A. S., and D. W. S. Wong. “The modifiable areal unit problem in multivariate statistical analysis.” Environment and planning A 23, no. 7 (1991): 1025-1044.

Fujita, M., P. R. Krugman, and A. Venables. The spatial economy. Mit press, 1999.

Gallup, J. L., J. D. Sachs, and A. D. Mellinger. “Geography and economic development.” International Regional Science Review 22, no. 2 (1999): 179.

Gine, X., and R. M. Townsend. “Evaluation of Financial Liberalization: A general equilibrium model with constrained occupation choice.” Journal of Development Economics 74, no. 2 (2004): 269-307.

Goodchild, M. F., and S. Gopal. The accuracy of spatial databases. CRC Press, 1989.

Greenwood, J., and B. Jovanovic. "Financial Development, Growth, and the Distribution of Income." Journal of Political Economy (1990): 1076-1107.

Hanson, G. H. "Regional adjustment to trade liberalization." Regional Science and Urban Economics 28, no. 4 (1998): 419-444.

Hoxby, C. M. "Does competition among public schools benefit students and taxpayers?." American Economic Review (2000): 1209-1238.

Hubert, L. J. Assignment Methods in Combinatorial Data Analysis. New York: Marcel Dekker, 1987.

Jeong, H., and R. M. Townsend. "Growth and inequality: Model evaluation based on an estimation-calibration strategy." Macroeconomic Dynamics 12, no. S2 (2008): 231-284.

—. "Sources of TFP growth: Occupational choice and financial deepening." Economic Theory 32, no. 1 (2007): 179-221.

Joerin, F., M. Thériault, and A. Musy. "Using GIS and outranking multicriteria analysis for land-use suitability assessment." International Journal of Geographical Information Science 15, no. 2 (2001): 153-174.

Krugman, P. "Increasing returns and economic geography." Journal of political economy 99, no. 3 (1991): 483.

Lloyd-Ellis, H., and D. Bernhardt. "Enterprise, inequality and economic development." Review of Economic Studies (2000): 147-168.

Malczewski, J. "GIS-based land-use suitability analysis: a critical overview." Progress in Planning 62, no. 1 (2004): 3-65.

—. "Integrating multicriteria analysis and geographic information systems: the ordered weighted averaging (OWA) approach." International Journal of Environmental Technology and Management 6, no. 1 (2006): 7-19.

Marshall, A. Principles of economics. Macmillan and co., 1890.

Merrill, D. W., and S. Selvin. "Analyzing Spatial Patterns in Health Outcome Surveillance Data." In Sixth Biennial CDC and ATSDR Symposium on Statistical Methods, Lawrence Berkeley National Laboratory, University of California, Berkeley School of Public Health, 1997.

Moran, P. A. P. "The interpretation of statistical maps." Journal of the Royal Statistical Society. Series B (Methodological) (1948): 243-251.

Nelson, A., A. de Sherbinin, and F. Pozzi. "Towards development of a high quality public domain global roads database." Data Science Journal 5, no. 0 (2006): 223-265.

Paulson, A. L., and R. Townsend. "Entrepreneurship and financial constraints in Thailand." Journal of Corporate Finance 10, no. 2 (2004): 229-262.

Paulson, A. L., R. M. Townsend, and A. Karaivanov. "Distinguishing Limited Liability from Moral hazard in a model of entrepreneurship." Journal of Political Economy 114, no. 1 (2006): 100-144.

Roos, M. W. M. "How important is geography for agglomeration?." Journal of Economic Geography 5, no. 5 (2005): 605-620.

Store, R., and J. Kangas. "Integrating spatial multi-criteria evaluation and expert knowledge for GIS-based habitat suitability modelling." Landscape and Urban Planning 55, no. 2 (2001): 79-93.

Townsend, R. M. "Intermediation with costly bilateral exchange." The Review of Economic Studies (1978): 417-425.

Townsend, R. M., and K. Ueda. "Financial deepening, inequality, and growth: A model-based quantitative evaluation." Review of Economic Studies 73, no. 1 (2006): 251-293.

Townsend, Robert M. Financial Systems in Developing Economies: Growth, Inequality and Policy Evaluation in Thailand. Oxford/New York: Oxford University Press, 2009.

Van Dijk, M. P., and Á Sverrisson. "Enterprise clusters in developing countries: mechanisms of transition and stagnation." Entrepreneurship & Regional Development 15, no. 3 (2003): 183-206.

TABLE 1: Statistical Parameters of Primary Variables

	National Level 1986	Four Provinces 1986	Four Provinces 1996
	(1)	(2)	(3)
Percent in Enterprise (CDD)	**1.6537**	**1.2294**	**1.8259**
	3.7573	*4.3332*	*7.0660*
1986 Wealth Instrument (CDD)	-	**25.0659**	**46.9642**
	-	*19.1844*	*38.8291*
Percent Completing Secondary Education	-	**1.6926**	**10.5234**
	-	*2.8660*	*12.9296*
Percent of Villages Having Financial Credit Access (Credit Index)	-	**18.2247**	**36.3610**
	-	*38.6105*	*48.1109*
Percent of Villages Having BAAC Credit Access	-	**74.6721**	**95.6682**
	-	*43.4963*	*20.3603*
Percent of Villages Having Commercial Bank Credit Access	-	**24.6622**	**38.8545**
	-	*43.1117*	*48.7495*
Travel-Time to Major Roads (Minutes)	-	**12.0468**	**12.0468**
	-	*10.6406*	*10.6406*
Travel-Time to Amphoe District Centers (Minutes)	-	**14.3741**	**14.3741**
	-	*9.0629*	*9.0629*
Distance to Major Highway Intersections (Kilometers)	-	**14.7458**	**14.7458**
	-	*8.682.333*	*8.682.333*
Soil Fertility Index (Range: 1 to 9)	**2.9472**	**2.6497**	**2.6497**
	1.7375	*1.6956*	*1.6956*
Elevation (Meters)	**217.5270**	**185.9592**	**185.9592**
	184.2478	*67.8361*	*67.8361*
Distance to Major Rivers (Kilometers)	**31.8648**	**3.7998**	**3.7998**
	34.1651	*3.6866*	*3.6866*
Annual Rainfall Variation (Centimeters)	**22.9625**	**23.6391**	**23.6391**
	6.7473	*3.2558*	*3.2558*
SES Entrepreneurial Monthly Income (in Baht): 1990	**910.0200**	-	-
	120.3423	-	-
DIW Factory Spatial Density (per square mile): 2005	**1.6413**	-	-
	14.4786	-	-
Amphoe Distance to Major Highways (Kilometers)	**6.2766**	-	-
	9.1575	-	-
Amphoe Distance to Bangkok (Kilometers)	**381.3931**	-	-
	225.8363	-	-
Amphoe Distance to Nearest Major City (Kilometers)	**84.4864**	-	-
	55.4940	-	-

Mean in bold, standard deviation italics

TABLE 2: ENTERPRISE AND GEOGRAPHY: BIVARIATE REGRESSION RESULTS

At the National Level ***Bivariate Regression Results***	National 1986 CDD Percent In Enterprise	National 1986-1996 CDD Growth in Percent in Enterprise	National 1990 SES Entrepreneurial Income	National 1990-1996 Growth in Entrepreneurial Income	National 2005 DIW Factory Geographic Density
	(1)	(2)	(3)	(4)	(5)
Distance to Major Highways (Kilometers)	-0.0303*	-9222	-11.53***	0.568	0.011
	(0.016)	(13594)	(3.149)	(0.849)	(0.000)
Distance to Bangkok (Kilometers)	-0.00285***	1230**	-0.17	0.0129	-0.00093
	(0.001)	(541.8)	(0.129)	(0.030)	(0.000)
Distance to 15 Largest Cities (Kilometers)	-0.00779***	3519	-0.517	0.267**	0.0248*
	(0.003)	(2157)	(0.523)	(0.120)	(0.000)
Distance to Major Rivers (Kilometers)	-0.00574	-4418	-1.174	-0.0928	-0.0145**
	(0.004)	(3520)	(0.850)	(0.202)	(0.0001)
Soil Fertility (Index, 0 to 9)	0.131*	36.93	6.072	0.0649*	-0.27
	(0.080)	(673.1)	(16.730)	(0.036)	(0.340)
Elevation (Meters)	-0.00211***	-3250	-0.293*	0.985***	-0.005***
	(0.001)	(6680)	(0.157)	(0.366)	(0.001)
Rainfall Variation (Centimeters)	0.00352	3707	2.272	-0.073	-0.158***
	(0.021)	(2253)	(4.307)	(0.123)	(0..0397)
Observations	700	700	700	700	2588

At the Provincial Level ***Bivariate Regression Results***	The Four Provinces 1986 CDD Percent In Enterprise	The Four Provinces 1986-1996 CDD Growth in Percent in Enterprise
	(1)	(2)
Travel-Time to Major Highways (Minutes)	-0.0473***	-0.0435*
	(0.007)	(0.025)
Travel-Time to Amphoe District Centers (Minutes)	-0.0493***	-0.023
	(0.009)	(0.029)
Distance to Major Intersections (Kilometers)	-0.0757***	-0.0555*
	(0.009)	(0.029)
Distance to Major Rivers (Kilometers)	-0.132***	-0.192**
	(0.022)	(0.083)
Soil Fertility (Index, 0 to 9)	0.543***	0.508***
	(0.044)	(0.134)
Elevation (Meters)	-0.105***	-0.0139***
	(0.012)	(0.005)
Rainfall Variation (Centimeters)	0.175***	0.000
	(0.024)	(0.080)
Observations	2987	1532

Robust standard errors in parentheses
***** p<0.01, ** p<0.05, * p<0.1**

Table 3: Spatial Structural Model Parameter Values

SPATIAL OCCUPATIONAL CHOICE SIMULATION RESULTS

SPATIAL OCCUPATIONAL CHOICE SIMULATION ESTIMATED STRUCTURAL PARAMETERS:

All Parameters Allowed To Vary Freely Across Whole Sample

	Occupational Choice Estimated Structural Parameters:	α	β	ω	η	ρ	σ	γ	g_γ	ξ
Primary Estimation:	*Estimated Values:*	1.0519	0.0536	0.5791	0.0009	0.0056	0.0001	0.0346	0.0035	0.0663

All Parameters Allowed To Vary Freely In Each Bin Across Space

Robustness Check 4:	**Occupational Choice: All Parameters Vary In All Bins** ***Increasing Geographic Distance***	α	β	ω	η	ρ	σ	γ	g_γ	ξ
	Bin 1	1.0127	0.0491	0.6391	0.0010	0.0049	0.0020	0.0305	0.0010	0.0599
	Bin 2	1.5095	0.0470	0.5601	0.0010	0.0020	0.0021	0.0096	0.0010	0.0207
	Bin 3	1.5889	0.1802	0.2410	0.0009	0.0050	0.0019	0.0305	0.0031	0.0601

VARIATION IN ESTIMATED M(d) PARAMETER ACROSS SPACE

Increasing Geographic Distance

	Estimation of the Spatially Varying *m(d)* Costs Parameter By Geographic Distance	Bin 1	Bin 2	Bin 3
Primary Estimation:	*All Parameters Vary Freely Initially; M(d) Varies by 3 Bins of Increasing Distance from Major Intersections*	-0.1013	0.5986	0.6202

	m(d) Costs Parameter By Geographic (Travel-Time) Distance	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Robustness Check 1:	*Jeong &Townsend (2003) Initial Parameters; M(d) Varies by 3 Bins of Increasing Distance from Major Intersections*	0.8400	1.0000	1.0000	-	-
Robustness Check 2:	*Jeong &Townsend (2003) Initial Parameters; M(d) Varies by 3 Bins of Increasing Distance from Major Intersections*	-0.2143	0.3612	0.2659	-	-
Robustness Check 3:	*Jeong &Townsend (2003) Initial Parameters; M(d) Varies by 5 Bins of Increasing Distance from Major Intersections*	-1.0000	-1.0000	-0.1587	0.3978	0.4256
Robustness Check 4:	*All Parameters Allowed to Vary Freely in Each Bin of Increasing Distance from Major Intersections*	-0.0898	-0.4831	-0.0048	-	-
Robustness Check 5:	*All Parameters Vary Freely Initially; M(d) Varies by 3 Bins of Increasing Distance from Favorable Geography*	0.1700	0.6200	0.4300	-	-

VARIATION IN FRACTION IN ENTERPRISE, ACTUAL AND SIMULATED

	Actual and Simulated Fraction in Enterprise, By Geographic Distance	Bin 1	Bin 2	Bin 3
	Actual Data	0.0572	0.0249	0.0221
Primary Estimation:	*All Parameters Vary Freely, 3 Bins by Equal Villages*	0.0575	0.0250	0.0237

	Fraction in Enterprise by Distance Bin	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Robustness Check 1:	*Jeong &Townsend (2003) Initial Parameters, 3 Bins by Equal Villages:*	0.0298	0.0192	0.0193	-	-
Robustness Check 2:	*Jeong &Townsend (2003) Initial Parameters, Allowing Subset to Vary, 3 Bins by Equal Villages:*	0.0294	0.0146	0.0198	-	-
Robustness Check 3:	*Jeong &Townsend (2003) Initial Parameters, Allowing Subset to Vary, 5 Bins By Equal Distance:*	0.0436	0.0435	0.0278	0.0151	0.0144
Robustness Check 4:	*All Parameters Allowed to Vary Freely, in All Bins Across Space:*	0.0572	0.0249	0.0221	-	-

CORRELATION COEFFICIENTS WITH ACTUAL DATA

		Correlation Coefficient	Probability Value
	Occupational Choice Non-Spatial Model: Predicted Spatial Patterns of End-Of-Period Simulation	0.0746	0.0444
Primary Estimation:	*Occupational Choice Spatial Model: Predicted Spatial Patterns of End-Of-Period Simulation*	0.1638	0.0010
	Occupational Choice Spatial Model: Predicted Spatial Patterns of Enterprise Growth (Bivariate LISA)	0.1241	0.0230
Robustness Check 5:	*Occupational Choice Spatial Model, Villages Binned by Distance from Favorable Geography: Predicted Spatial Patterns of Enterprise Growth (Bivariate LISA)*	0.0985	0.0510
Robustness Check 6:	*Occupational Choice Spatial Model CREDIT FIXED AT INITIAL LEVELS: Predicted Spatial Patterns of Enterprise Growth (Bivariate LISA)*	0.0634	0.0490
Robustness Check 7:	*Occupational Choice Spatial Model CREDIT ELIMINATED: Predicted Spatial Patterns of Enterprise Growth (Bivariate LISA)*	0.0314	0.0790

SPATIAL FINANCIAL DEEPENING SIMULATION RESULTS

SPATIAL FINANCIAL DEEPENING SIMULATION STRUCTURAL PARAMETERS:

	Financial Deepening Initial Structural Parameters:	σ	δ	ζ	ε	β	γ
Primary Estimation:	*Initial Values (from Townsend and Ueda, 2003):*	1.0000	1.0540	[1.047,1.147]	[-0.600,0.600]	0.9600	0.0000

Increasing Geographic Distance

	Estimation of the Spatially Varying *q(d)* Cost Parameter by Geographic Distance	Bin 1	Bin 2	Bin 3
Primary Estimation:	*Sample Divided into Three Bins by Equal Geographic (Travel-Time) Distance From Major Intersections*	43.4763	24.6950	18.6233

Table 4: Occupational Choice 1996 Simulation Residuals Regressed Onto Market and Infrastructure Access Proxies

	Non-Spatial Residuals^	Spatial Residuals^^	Non-Spatial Residuals	Spatial Residuals	Non-Spatial Residuals	Spatial Residuals	Non-Spatial Residuals	Spatial Residuals	Non-Spatial Residuals	Spatial Residuals
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Travel-Time to Major Roads	-0.0505**	-0.0178								
	(0.025)	(0.025)								
Travel-Time to Major Intersections			-0.00000138***	-0.000000168					-0.000000719	0.000000428
			(0.00000046)	(0.00000045)					(0.00000045)	(0.00000045)
Travel-Time to District Centers					-0.0643**	-0.0252				
					(0.026)	(0.026)				
1986 Wealth							0.000344**	0.000181	0.000265	0.000228
							(0.00016)	(0.00016)	(0.00016)	(0.00016)
1986 Education							0.0213	0.00716	0.0146	0.0112
							(0.025)	(0.025)	(0.025)	(0.025)
Constant	0.00789	0.00216	0.0172**	0.00148	0.0141**	0.00490	-0.0244***	-0.0142	-0.0103	-0.0226*
	(0.0056)	(0.0055)	(0.0069)	(0.0069)	(0.0070)	(0.0069)	(0.0088)	(0.0088)	(0.012)	(0.012)
Observations	731	731	731	731	731	731	704	704	704	704

Standard errors in parentheses
***** p<0.01, ** p<0.05, * p<0.1**
Residuals are simulated minus actual values
^"Non-Spatial Residuals" are residuals from non-spatial structural simulation
^^"Spatial Residuals" are residuals from spatially-specified 3-bin structural simulation

Table 5: Financial Deepening 1996 Simulation Residuals Regressed Onto Market and Infrastructure Access Proxies

	Wealth Simulation 1996 Residuals	Wealth Simulation 1996 Residuals	Wealth Simulation 1996 Residuals	Credit Intermediation 1996 Residuals	Credit Intermediation 1996 Residuals	Credit Intermediation 1996 Residuals	BAAC Simulation 1996 Residuals	BAAC Simulation 1996 Residuals	BAAC Simulation 1996 Residuals	Commercial Banks Simulation 1996 Residuals	Commercial Banks Simulation 1996 Residuals	Commercial Banks Simulation 1996 Residuals
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Travel-Time to Major Roads	**-17.4512**			**-0.7028**			**-0.1249**			**0.0281**		
	*0.0002****			*0.0001****			*0.0457***			*0.7985*		
Travel-Time to Major Intersections		**-0.0008**			**-0.0001**			**0.0001**			**-0.0001**	
		*0.0001****			*0.0001****			*0.0001****			*0.0001****	
Travel-Time to District Centers			**-31.5170**			**-0.6625**			**-0.1390**			**-0.1282**
			*0.0001****			*0.0002****			*0.0398***			*0.2821*
1996 Wealth Index	**2.1993**	**2.8967**	**2.1405**	**0.0027**	**0.0024**	**0.0024**	**0.0028**	**0.0022**	**0.0027**	**0.0022**	**0.0010**	**0.0021**
	0.1376	*0.0462***	*0.1448*	*0.0002****	*0.0009****	*0.0006****	*0.0001****	*0.0001****	*0.0001****	*0.0077****	*0.2364*	*0.0128***
1996 Educational Attainment	**34.6477**	**30.5663**	**33.3364**	**1.1009*****	**1.0195**	**1.1015**	**-0.0057**	**-0.0451**	**-0.0110**	**0.0866**	**-0.0130**	**0.0700**
	*0.0001****	*0.0001****	*0.0001****	*0.0001****	*0.0001****	*0.0001****	*0.9382*	*0.5394*	*0.8816*	*0.5056*	*0.9198*	*0.5912*
R-Squared	**0.3051**	**0.3320**	**0.3171**	**0.5853**	**0.5863**	**0.5844**	**0.1058**	**0.1252**	**0.1060**	**0.0652**	**0.0922**	**0.0660**

Residuals are simulated minus actual values
***** p<0.01, ** p<0.05, * p<0.1**
Coefficient values in bold, probability values in italics

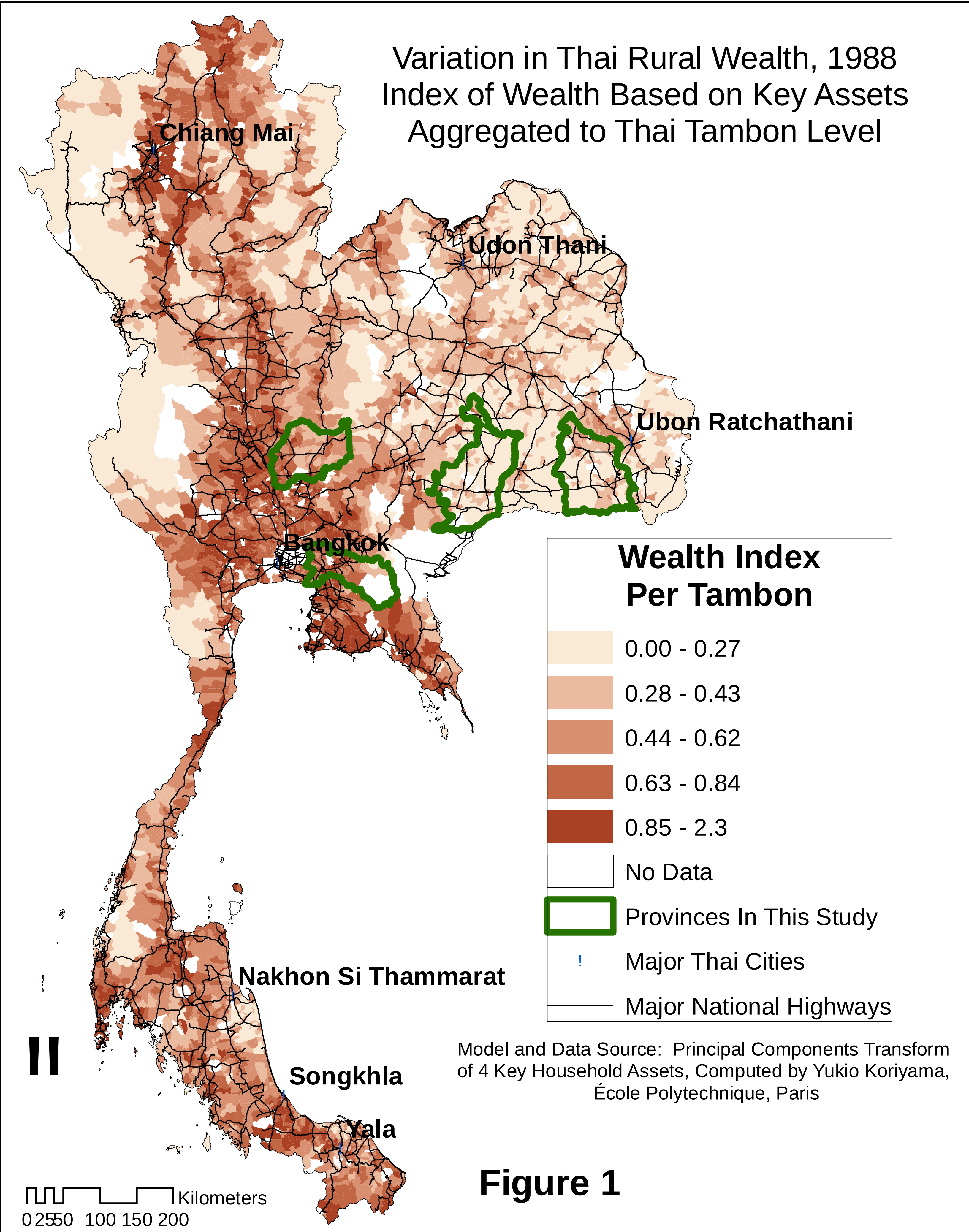

Variation in Thai Rural Wealth, 1988
Index of Wealth Based on Key Assets
Aggregated to Thai Tambon Level
Chiang Mai
Udon Thani
Ubon Ratchathani
Bangkok
Nakhon Si Thammarat
Songkhla
Yala
Wealth Index
Per Tambon
0.00 - 0.27
0.28 - 0.43
0.44 - 0.62
0.63 - 0.84
0.85 - 2.3
No Data
Provinces In This Study
Major Thai Cities
Major National Highways
Model and Data Source: Principal Components Transform of 4 Key Household Assets, Computed by Yukio Koriyama, École Polytechnique, Paris
Kilometers
0 25 50 100 150 200

Figure 1

Village, Roads, Major Intersections and Amphoe District Center Locations

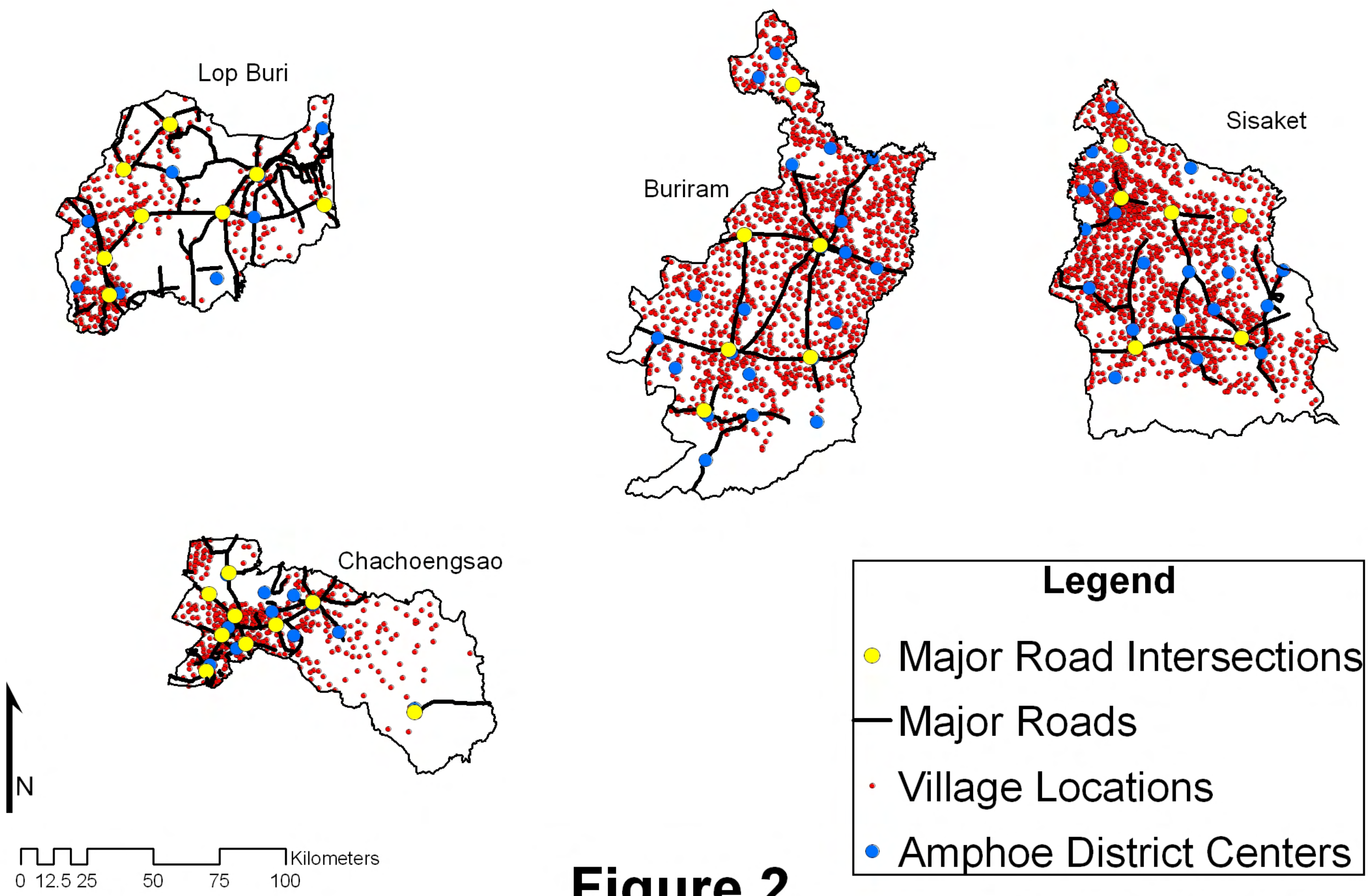
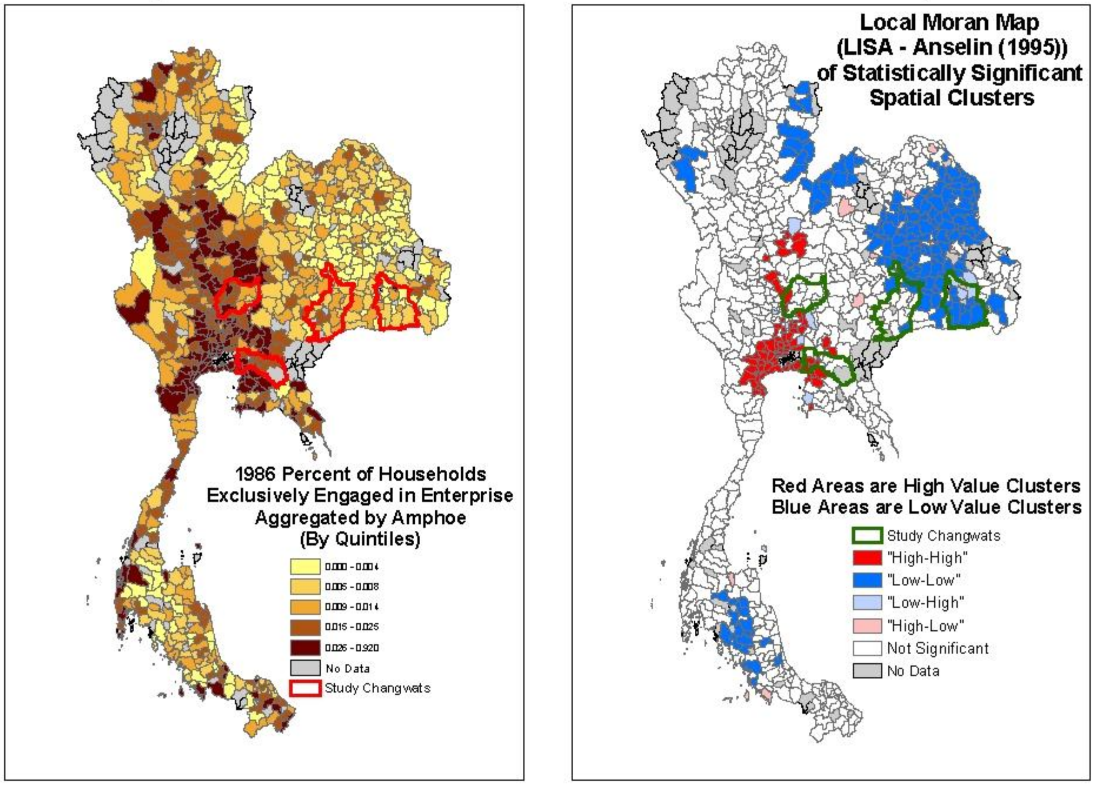


Figure 2

Enterprise at the National Level in 1986: CDD Survey Data



Entrepreneurial Income at the National Level in 1990: SES Survey Data

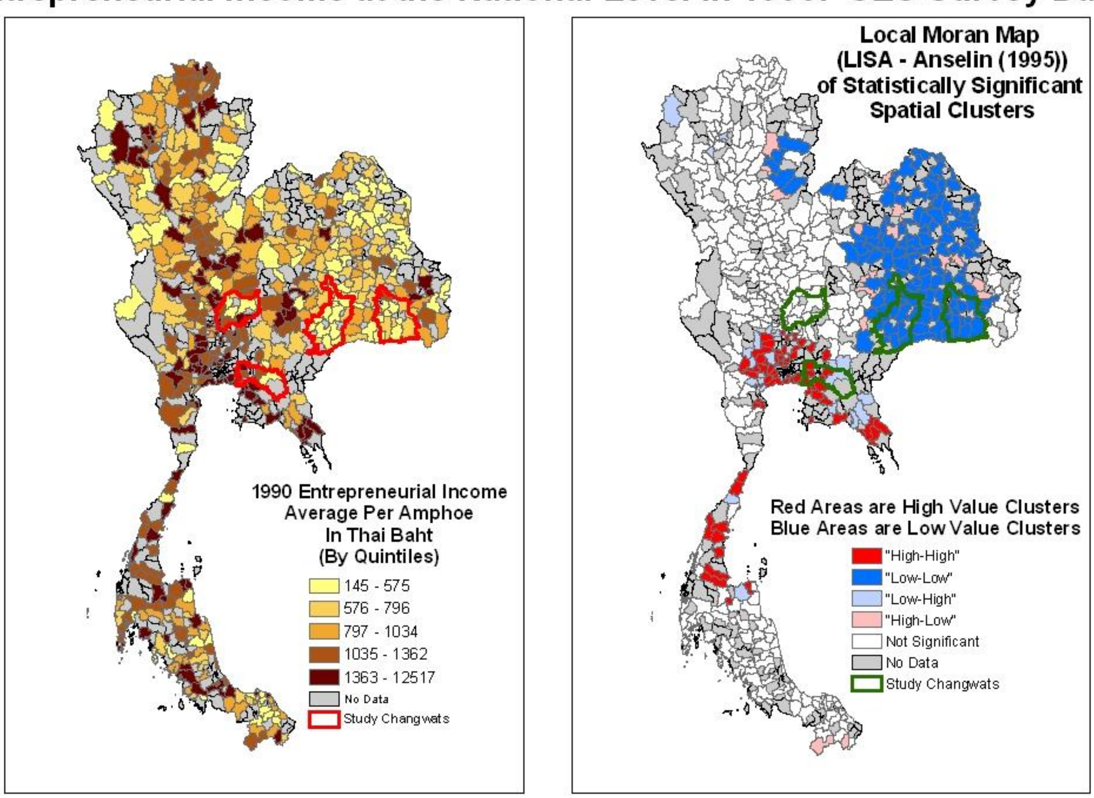


Figure 3

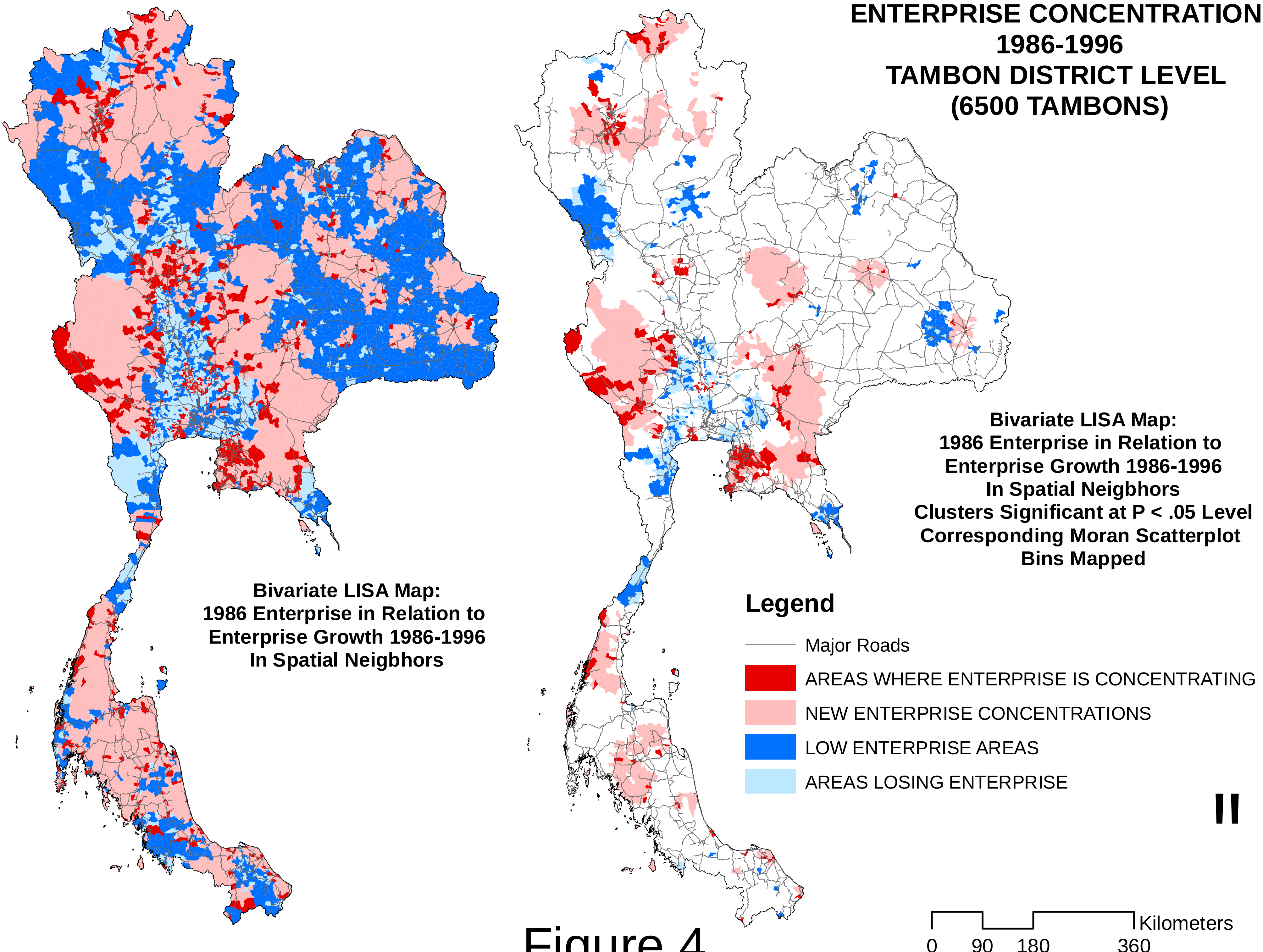
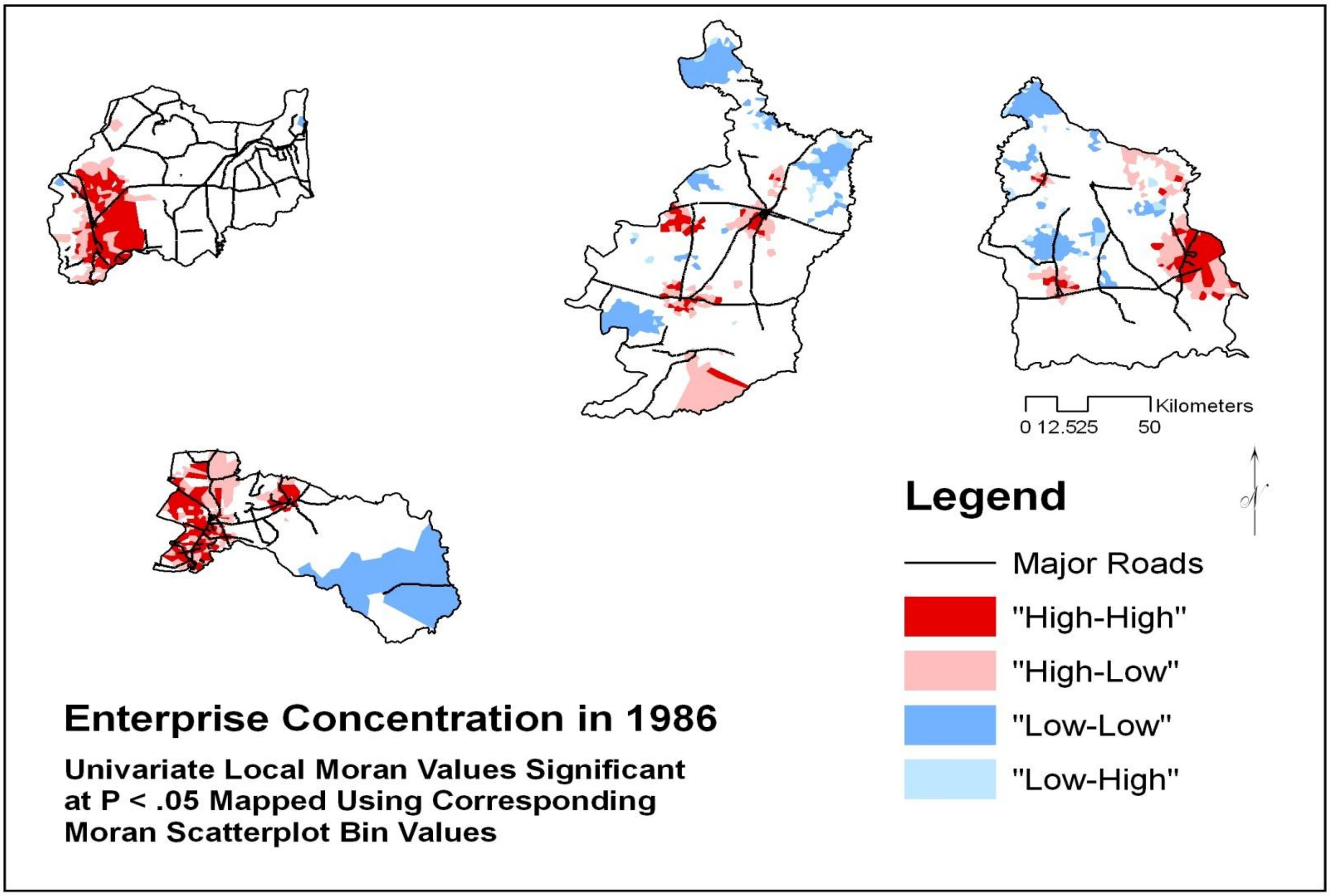
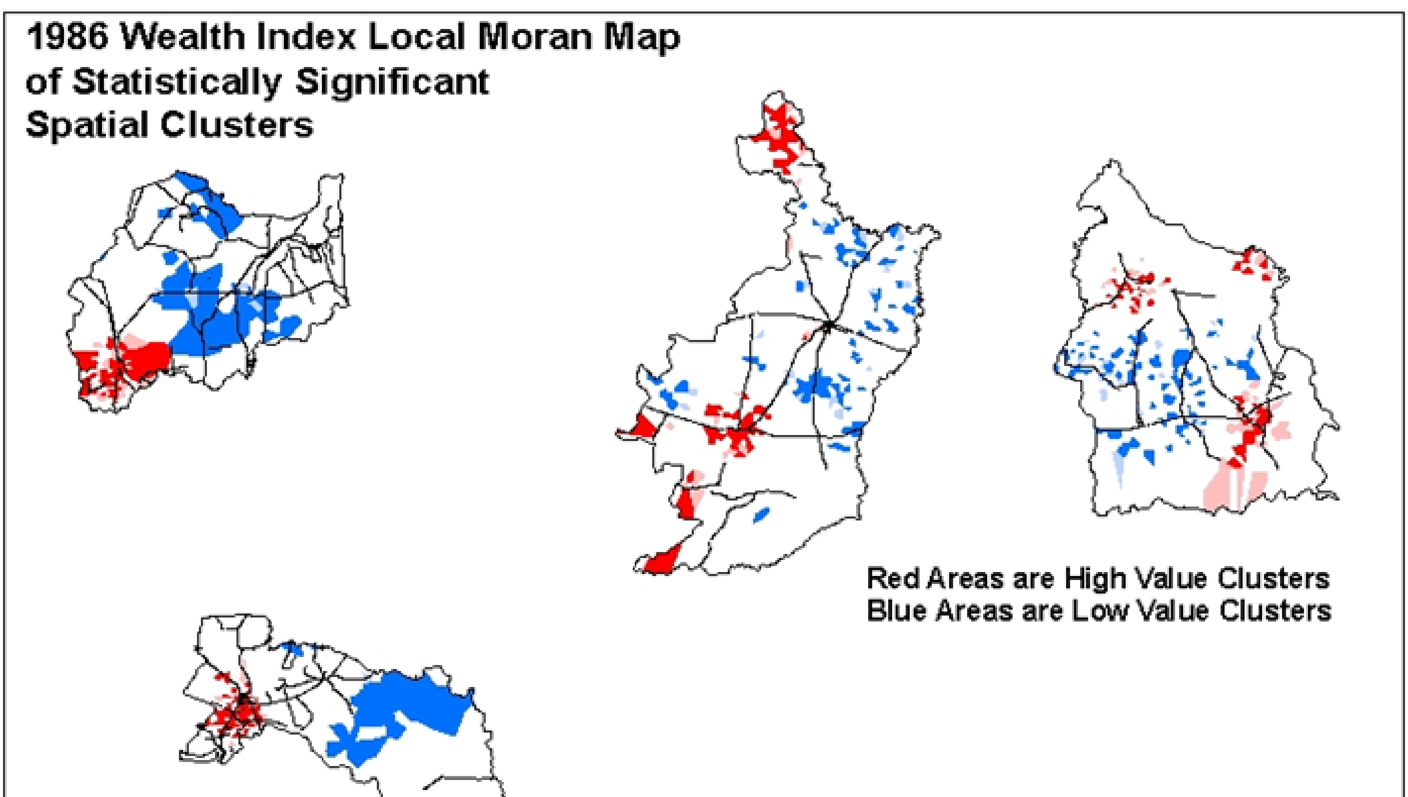

ENTERPRISE CONCENTRATION
1986-1996
TAMBON DISTRICT LEVEL
(6500 TAMBONS)
Bivariate LISA Map:
1986 Enterprise in Relation to
Enterprise Growth 1986-1996
In Spatial Neigbhors
Bivariate LISA Map:
1986 Enterprise in Relation to
Enterprise Growth 1986-1996
In Spatial Neigbhors
Clusters Significant at P < .05 Level
Corresponding Moran Scatterplot
Bins Mapped
Legend
Major Roads
AREAS WHERE ENTERPRISE IS CONCENTRATING
NEW ENTERPRISE CONCENTRATIONS
LOW ENTERPRISE AREAS
AREAS LOSING ENTERPRISE
II
Kilometers
0
90
180
360

Figure 4

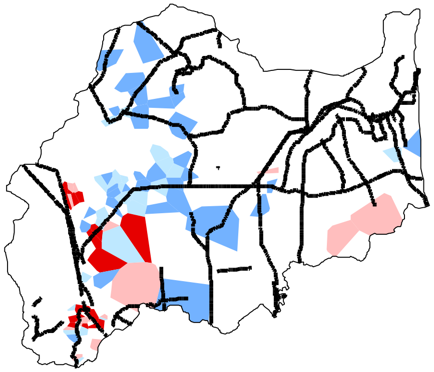


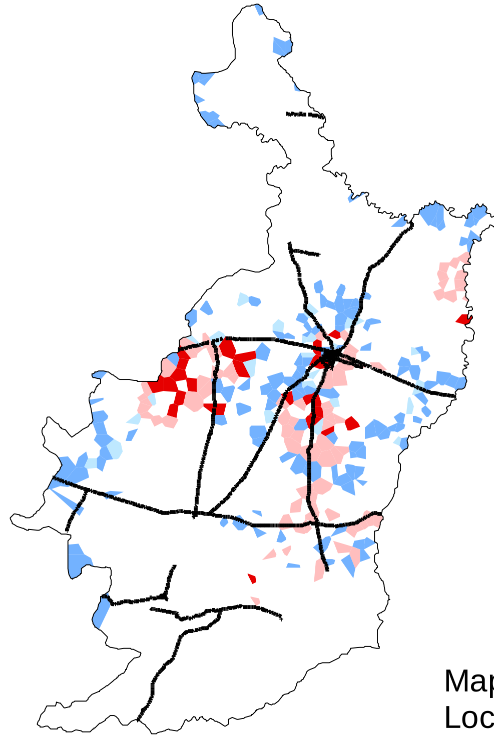
Kilometers
0 12.525 50
Legend
Major Roads
"High-High"
"High-Low"
"Low-Low"
"Low-High"
Enterprise Concentration in 1986
Univariate Local Moran Values Significant at P < .05 Mapped Using Corresponding Moran Scatterplot Bin Values

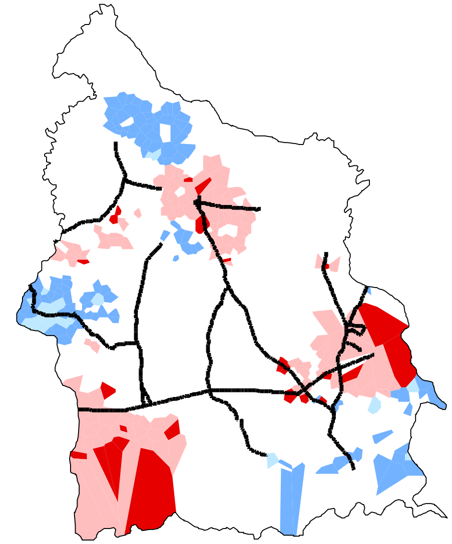


1986 Wealth Index Local Moran Map of Statistically Significant Spatial Clusters
Red Areas are High Value Clusters
Blue Areas are Low Value Clusters

Figure 5







ENTERPRISE CONCENTRATION 1986-1996
BIVARIATE LISA MAP
STATISTICALLY SIGNIFICANT
CONCENTRATIONS (P < .1)

Map is an output from a bivariate Local Moran Index for 1532 villages, considering 1986 fraction in enterprise surrounded by 1986-1996 growth in fraction in enterprise

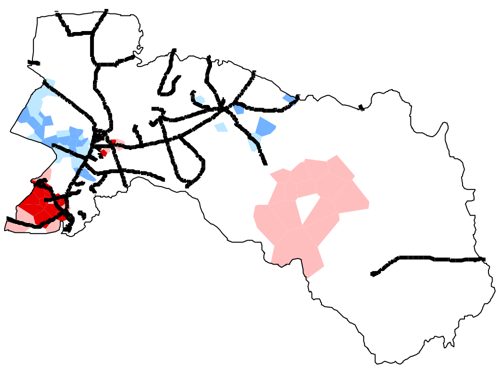
REDS are village with higher than average enterprise in 1986, surrounded by villages with higher than average growth in enteprise 1986-1996

PINKS are villages with lower than average enteprise in 1986, surrounded by villages with higher than average enterprise growth 1986-1996

DARK BLUES are villages with lower than average enterprise in 1986, surrounded by villages with lower than average enterprise growth 1986-1996

LIGHT BLUES are villages with higher than average enterprise in 1986, surrounded by villages with lower than average enterprise growth 1986-1996

II



Legend

- AREAS WHERE ENTERPRISE IS CONTINUING TO CONCENTRATE
- NEW ENTERPRISE CONCENTRATIONS ARISING
- LOW ENTERPRISE AREAS
- AREAS LOSING ENTEPRISE
- Major Roads

Figure 6

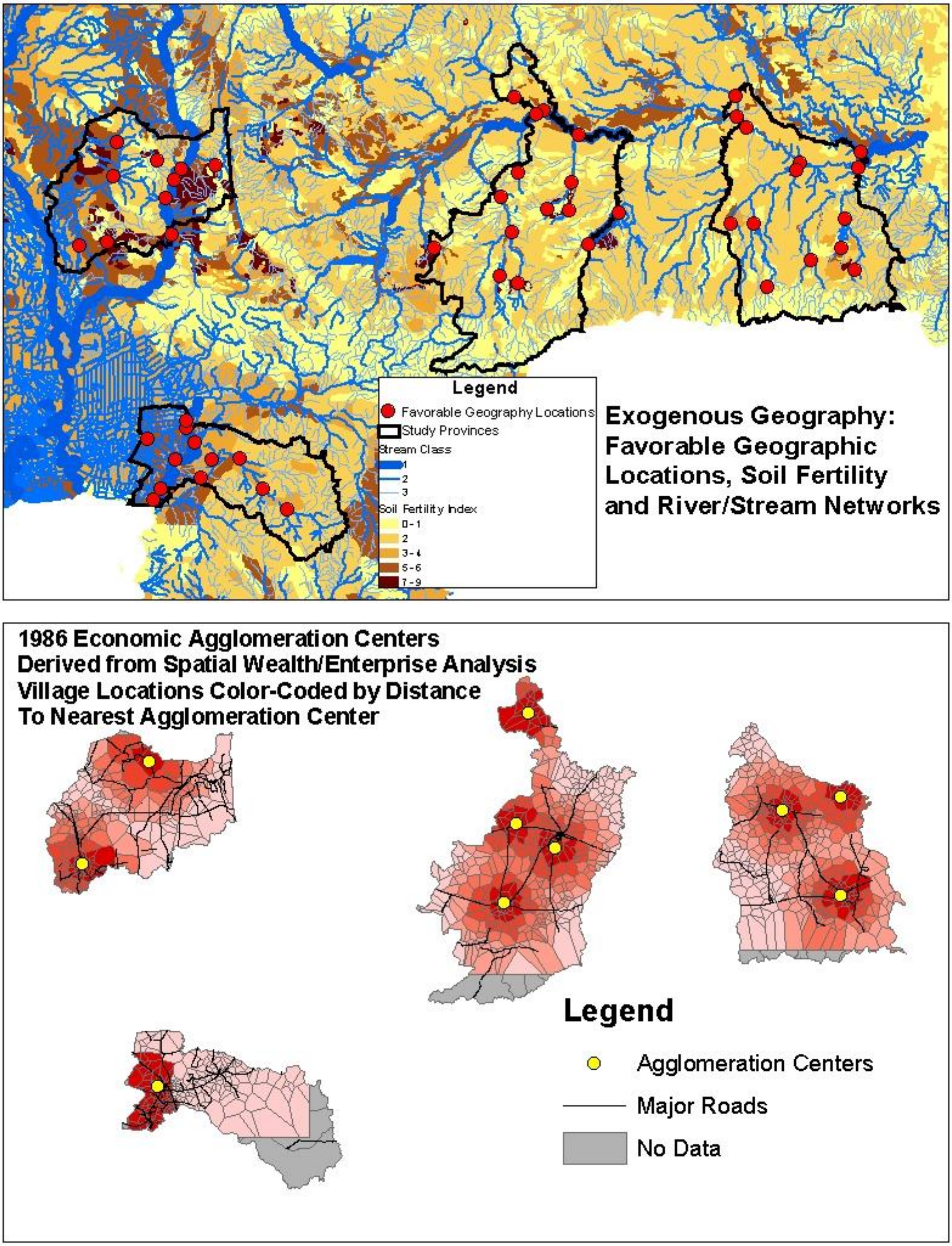
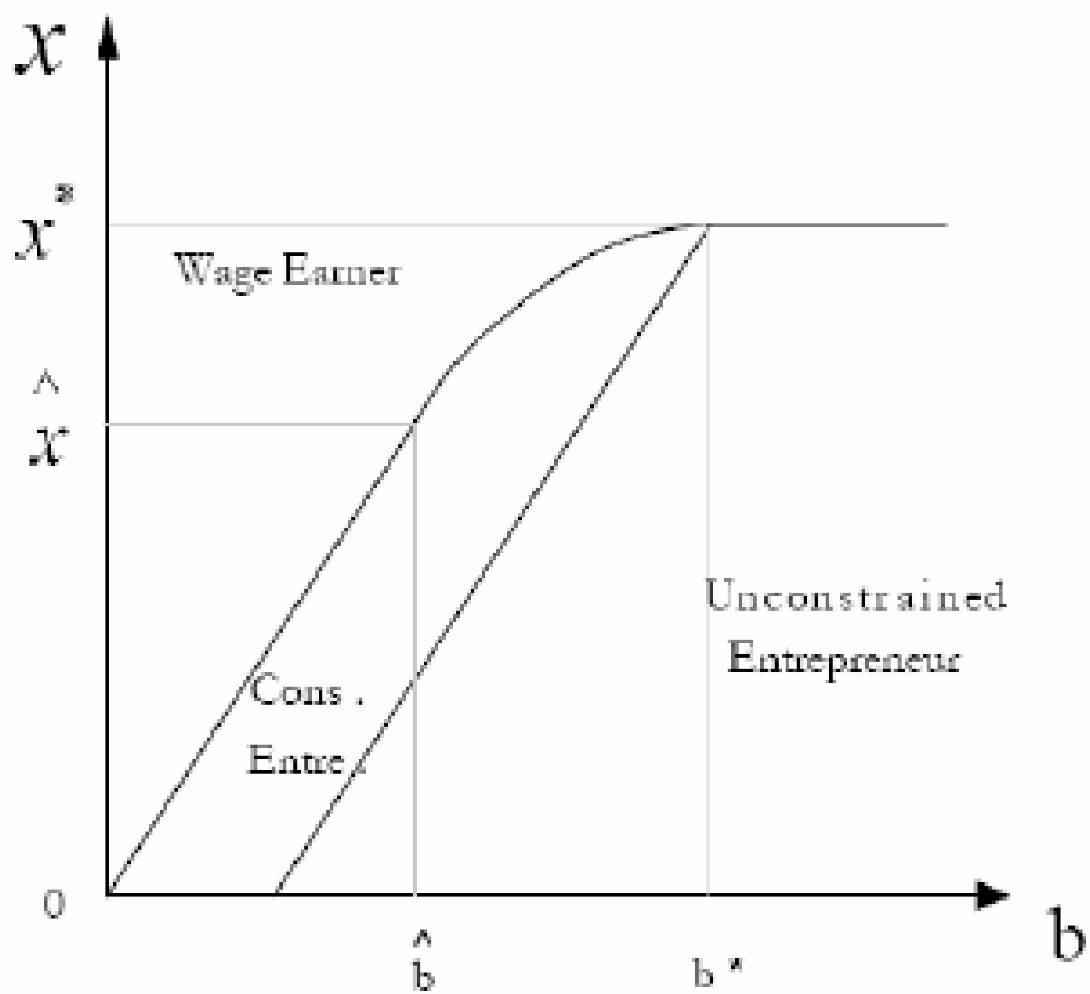
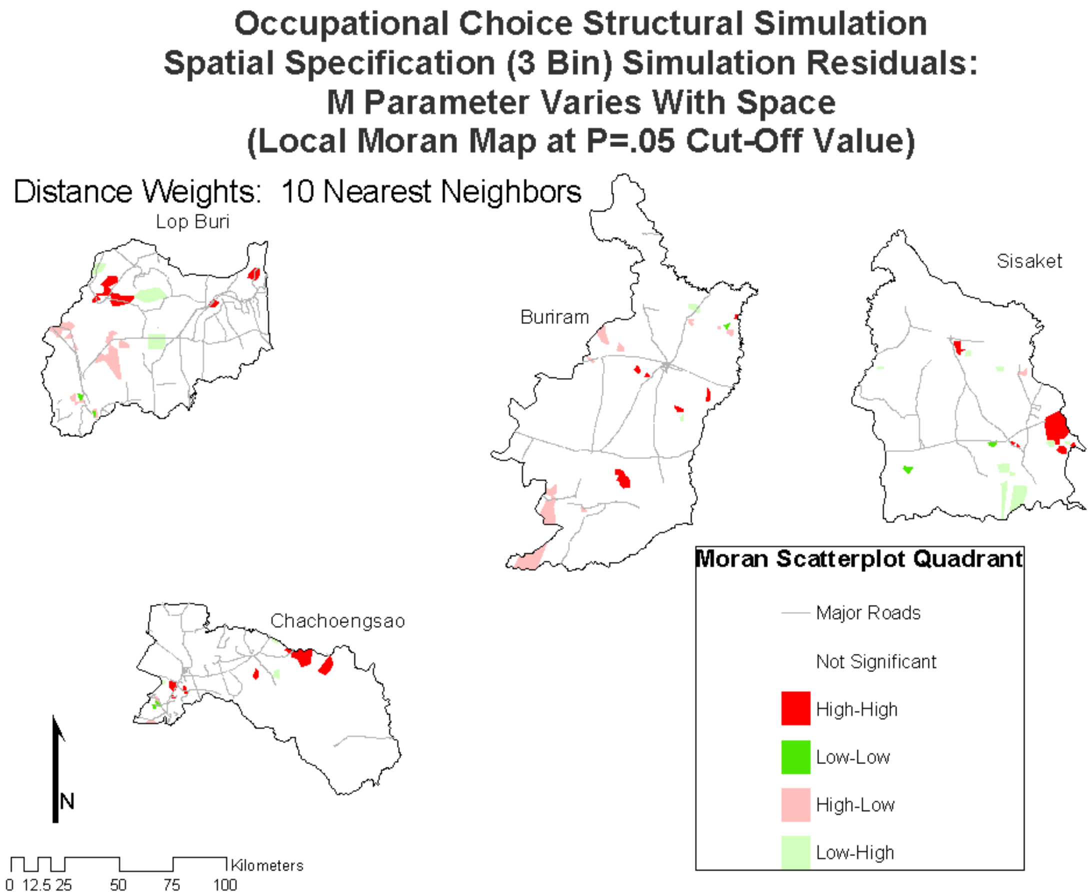

Legend
Favorable Geography Locations
Study Provinces
Stream Class
1
2
3
Soil Fertility Index
0 - 1
2
3 - 4
5 - 6
7 - 9
Exogenous Geography:
Favorable Geographic
Locations, Soil Fertility
and River/Stream Networks
1986 Economic Agglomeration Centers
Derived from Spatial Wealth/Enterprise Analysis
Village Locations Color-Coded by Distance
To Nearest Agglomeration Center
Legend
Agglomeration Centers
Major Roads
No Data

Figure 7

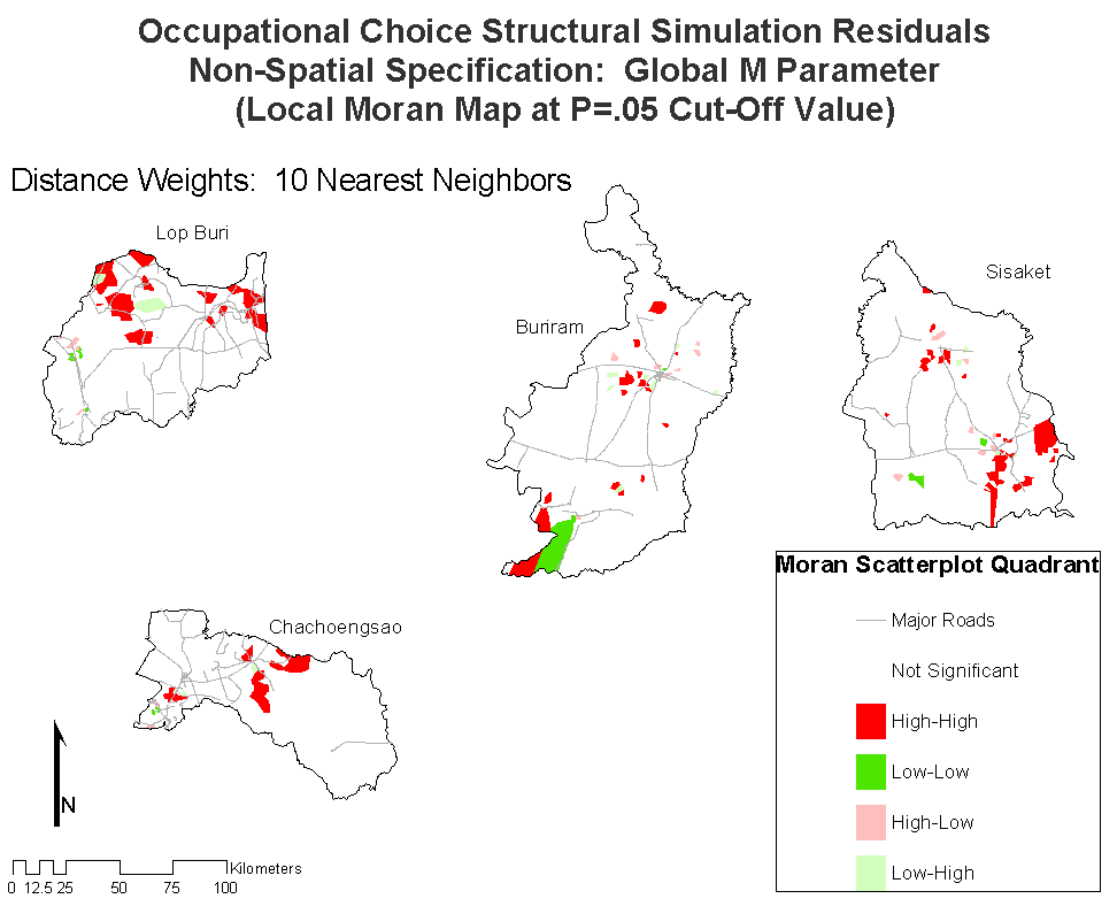


x
x*
Wage Earner
x̂
Unconstrained
Entrepreneur
Cons.
Entre.
0
b̂
b*
b

Figure 8

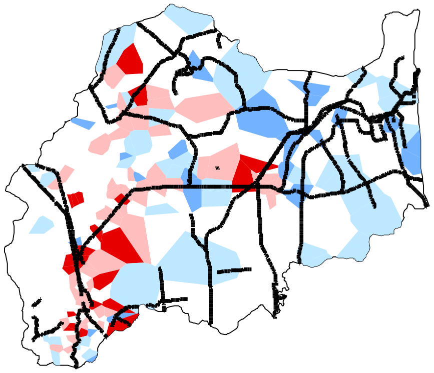


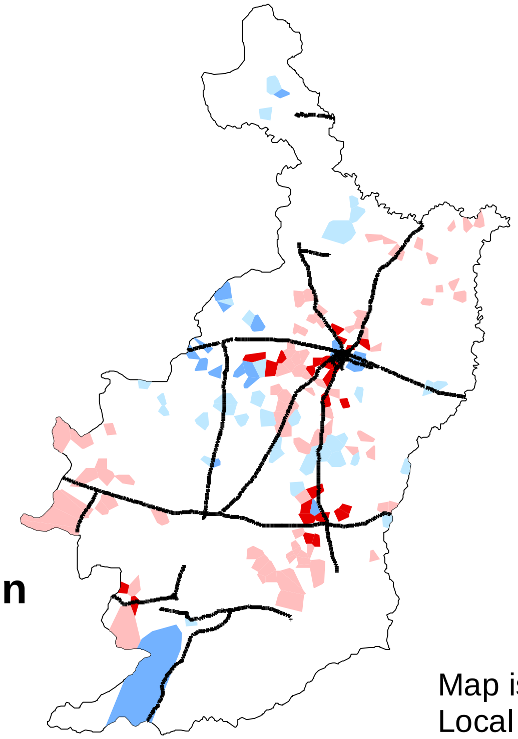
Occupational Choice Structural Simulation
Spatial Specification (3 Bin) Simulation Residuals:
M Parameter Varies With Space
(Local Moran Map at P=.05 Cut-Off Value)
Distance Weights: 10 Nearest Neighbors
Lop Buri
Buriram
Sisaket
Chachoengsao
Moran Scatterplot Quadrant
Major Roads
Not Significant
High-High
Low-Low
High-Low
Low-High
N
Kilometers
0 12.5 25 50 75 100

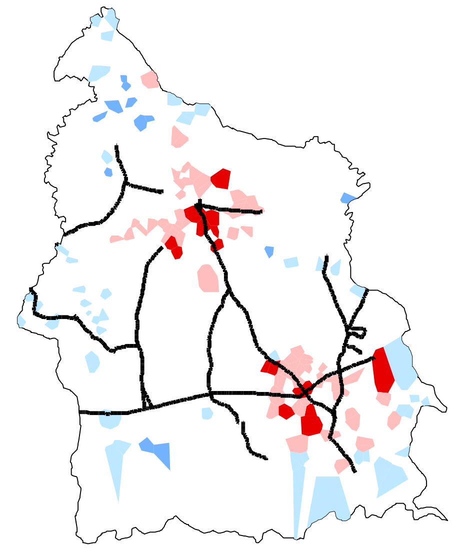


Occupational Choice Structural Simulation Residuals
Non-Spatial Specification: Global M Parameter
(Local Moran Map at P=.05 Cut-Off Value)
Distance Weights: 10 Nearest Neighbors
Lop Buri
Buriram
Sisaket
Chachoengsao
Moran Scatterplot Quadrant
Major Roads
Not Significant
High-High
Low-Low
High-Low
Low-High
N
Kilometers
0 12.5 25 50 75 100

Figure 9



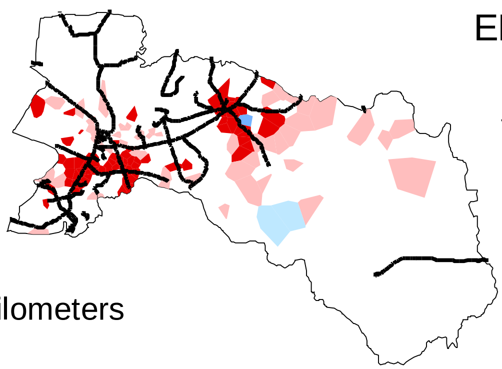




Primary Estimation:
Occupational Choice Structural Simulation
Spatial Model:
M Parameter Varies Across Space
Bivariate LISA Map 1986-1996

Map is an output from a bivariate
Local Moran Index for 1532 villages,
considering 1986 fraction in enterprise surrounded
by 1986-1996 growth in fraction in enterprise

II



ENTERPRISE CONCENTRATION
1986-1996
STATISTICALLY SIGNIFICANT
CONCENTRATIONS
(P = .1 OR LESS)

REDS are village with higher
than average enterprise in 1986,
surrounded by villages with higher than
average growth in enteprise 1986-1996

PINKS are villages with lower
than average enteprise in 1986,
surrounded by villages with higher than
average enterprise growth 1986-1996

Kilometers
0 10 20 40

Legend

AREAS WHERE ENTERPRISE IS CONTINUING TO CONCENTRATE

NEW ENTERPRISE CONCENTRATIONS ARISING

LOW ENTERPRISE AREAS

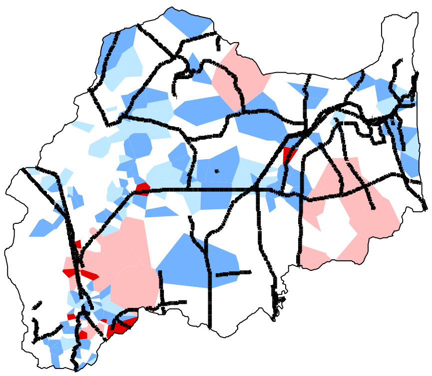
AREAS LOSING ENTERPRISE

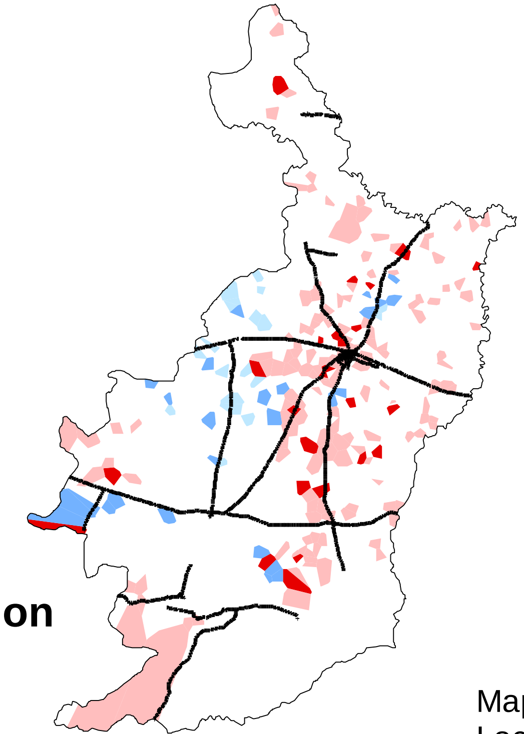
Major Roads

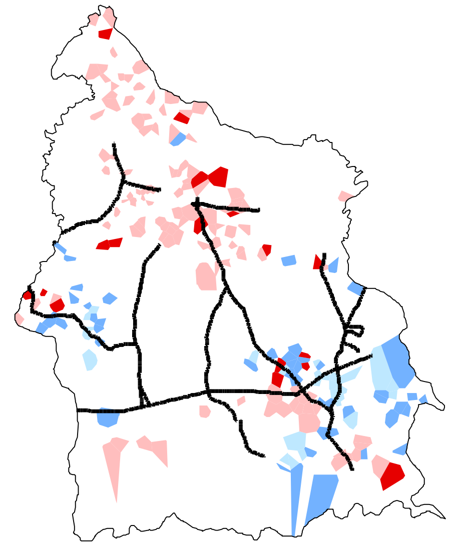
DARK BLUES are villages with lower than
average enterprise in 1986, surrounded by
villages with lower than average enterprise
growth 1986-1996

LIGHT BLUES are villages with higher than
average enterprise in 1986, surrounded by
villages with lower than average enterprise
growth 1986-1996

Figure 10

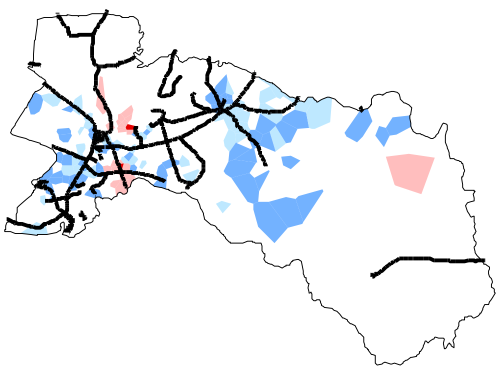






Robustness Check 6:
Occupational Choice Structural Simulation
Credit Fixed at 1986 Levels
Bivariate LISA Map 1986-1996

II



ENTERPRISE CONCENTRATION
1986-1996
STATISTICALLY SIGNIFICANT
CONCENTRATIONS
(P = .1 OR LESS)

Map is an output from a bivariate Local Moran Index for 2000 villages, considering 1986 fraction in enterprise surrounded by 1986-1996 growth in fraction in enterprise

REDS are village with higher than average enterprise in 1986, surrounded by villages with higher than average growth in enteprise 1986-1996

PINKS are villages with lower than average enteprise in 1986, surrounded by villages with higher than average enterprise growth 1986-1996

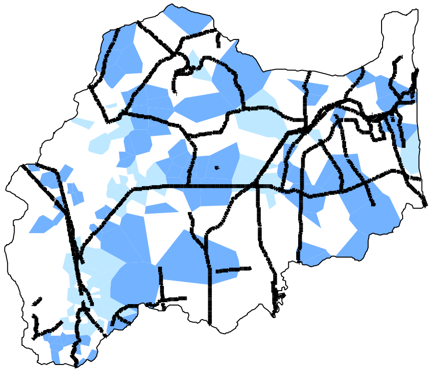
DARK BLUES are villages with lower than average enterprise in 1986, surrounded by villages with lower than average enterprise growth 1986-1996

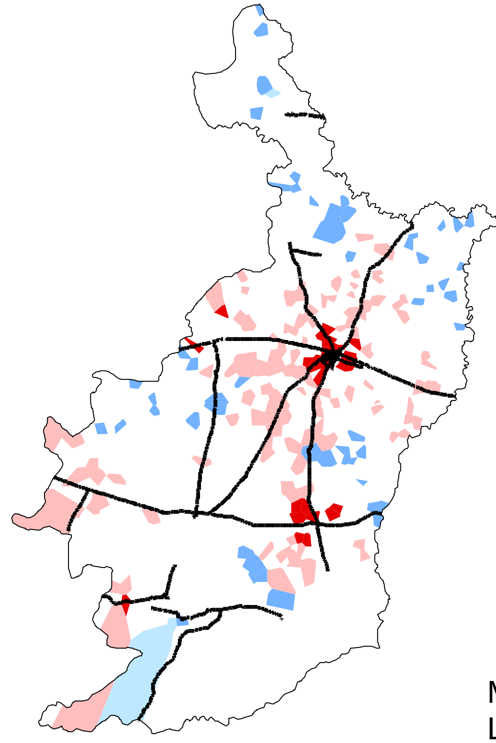
LIGHT BLUES are villages with higher than average enterprise in 1986, surrounded by villages with lower than average enterprise growth 1986-1996

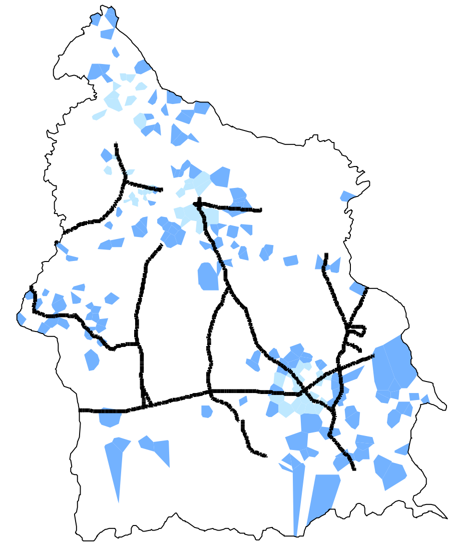
Legend

- AREAS WHERE ENTERPRISE IS CONTINUING TO CONCENTRATE
- NEW ENTERPRISE CONCENTRATIONS ARISING
- LOW ENTERPRISE AREAS
- AREAS LOSING ENTEPRISE
- Major Roads

Figure 11



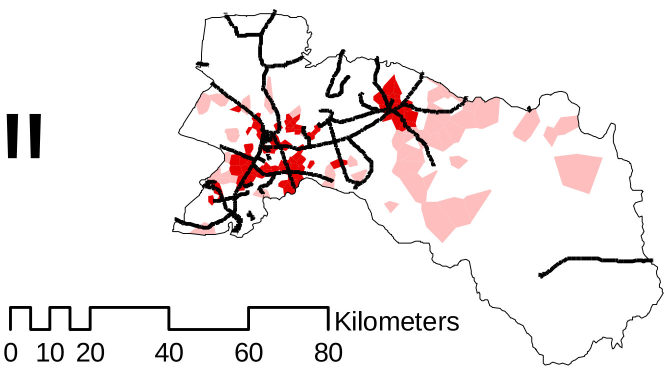




Robustness Check #7: No Credit Access Occupational Choice Structural Simulation Bivariate LISA Map 1986-1996

Map is an output from a bivariate Local Moran Index for 2000 villages, considering 1986 fraction in enterprise surrounded by 1986-1996 growth in fraction in enterprise

II



ENTERPRISE CONCENTRATION
1986-1996
IN THAILAND: 4 PROVINCES:
STATISTICALLY SIGNIFICANT
CONCENTRATIONS
(P = .1 OR LESS)

REDS are village with higher than average enterprise in 1986, surrounded by villages with higher than average growth in enteprise 1986-1996

PINKS are villages with lower than average enteprise in 1986, surrounded by villages with higher than average enterprise growth 1986-1996

DARK BLUES are villages with lower than average enterprise in 1986, surrounded by villages with lower than average enterprise growth 1986-1996

LIGHT BLUES are villages with higher than average enterprise in 1986, surrounded by villages with lower than average enterprise growth 1986-1996

Legend

- AREAS WHERE ENTERPRISE IS CONTINUING TO CONCENTRATE
- NEW ENTERPRISE CONCENTRATIONS ARISING
- LOW ENTERPRISE AREAS
- AREAS LOSING ENTEPRISE
- Major Roads

Figure 12

Figure 18: Financial Deepening Simulation
1996 Credit Access Spatial Residuals

Differences are Between Actual and Simulated
Window-Average Smoothed Values, Using 10 Nearest Neighbors

Reds are Areas of Model Over-Prediction,
Greens are Areas of Model Under-Prediction

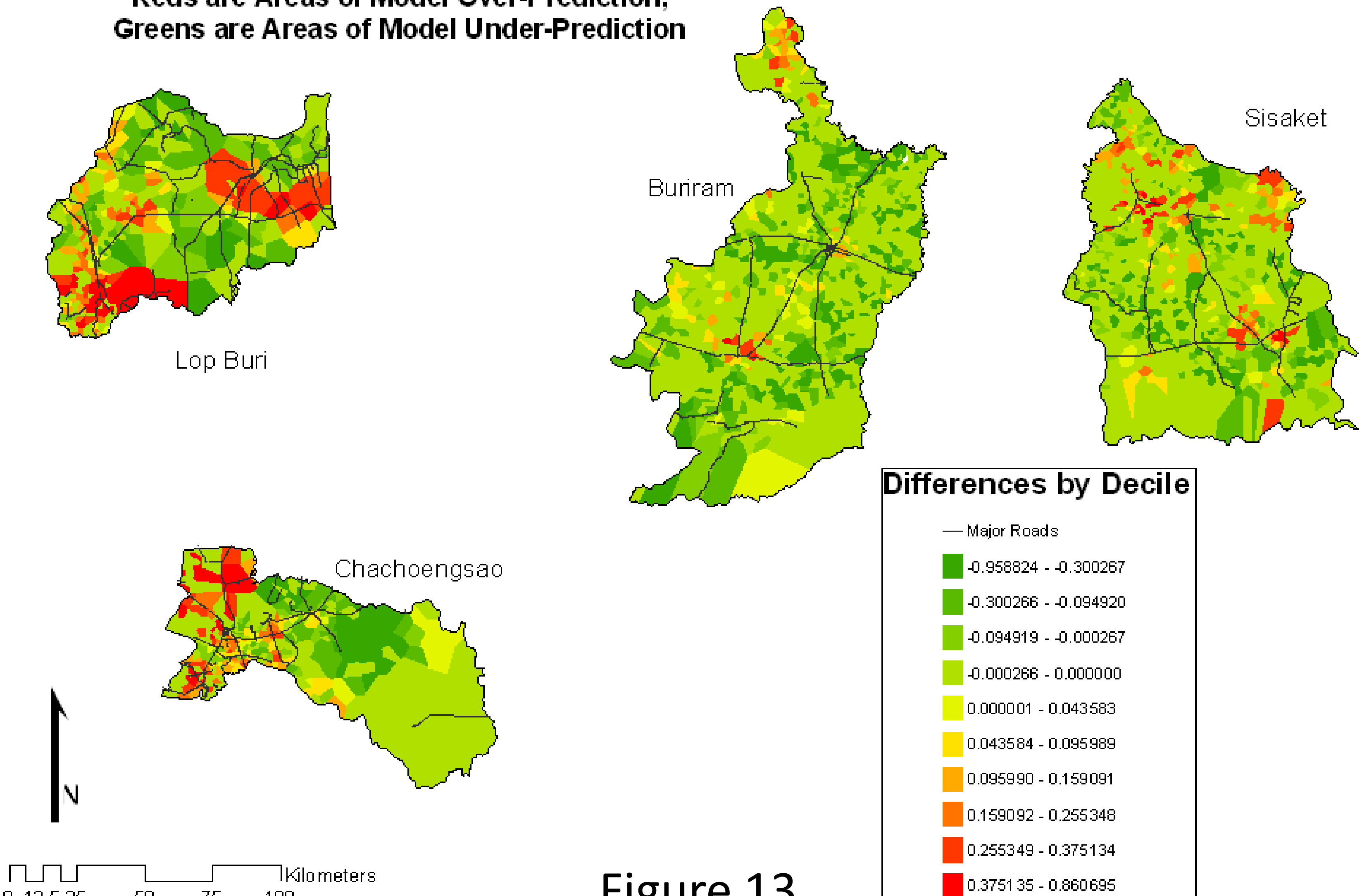


Figure 13