The Geographic Concentration of Enterprise in Developing Countries

John S. Felkner and Robert M. Townsend[1]
University of Chicago, Massachusetts Institute of Technology

August 2011

A nation's economic geography can have an enormous impact on its development. In Thailand, we show that a high concentration of enterprise in an area predicts high subsequent growth in and around that area. We also find spatially contiguous convergence of enterprise with stagnant areas left behind. Exogenous physiographic conditions are correlated with enterprise location and growth. We fit a structural, micro-founded model of occupation transitions with fine-tuned geographic capabilities to village data and replicate these salient facts. Key elements of the model include costs, credit constraints on occupation choice, and spatially varying expansion of financial service providers.

JEL Codes: J240, 0120, R110, R120

Key Words: occupational choice; financial credit access; spatial analysis; enterprise growth; enterprise concentration; agglomeration; market access; geography; GIS; agriculture; environment

I. Introduction

Despite military coups, tsunamis, and the Asian financial crisis, over the long run Thailand's economic trajectory 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% and the number of households in non-farm enterprises

[1] We would like to thank Edward Glaeser and the anonymous reviewers for very helpful comments. We wish to thank Alex Karaivanov, Victor Zhorin, Yukio Koriyama, and Natalia Ramondo for their assistance in adopting and running the simulation programs for the various models. Many thanks to Mark Harris and Nery Perez 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 Consortium on Financial Systems and Poverty at the University of Chicago through a grant from the Bill & Melinda Gates Foundation. Authors' e-mails: john3@uchicago.edu and rtownsen@mit.edu.

increasing by 27% (Townsend 2009).[2] But this growth is not happening everywhere in Thailand. As in many developing countries, spatial inequalities and economic concentrations are increasing over time (Kanbur and Venables, 2005; Nordhaus 2006; Henderson and Wang 2005).[3] We use high-resolution, highly-accurate digital spatial data on Thailand to examine this heterogeneity of economic activity across space.

We provide evidence for seven key facts about spatial development in Thailand:

1. *There are high concentrations of enterprise.* On a national scale across Thailand, in the 1986 baseline year, high levels of industry and enterprise are clustered in and around Bangkok and in a central national corridor heading north along main transportation arteries. At the rural village level, concentrations of enterprise within each province are at or near large towns and intersections of major highways.

2. *Levels of enterprise are negatively correlated with distance to market and infrastructure.* Areas closest to Bangkok, larger cities, and main highways have higher fractions of village populations in enterprise. Within provinces, enterprise levels decrease significantly as travel-time and geographic distances to intersections of major highways and amphoe district centers increase.

3. *Over time, areas of high concentrations and those nearby show increasing levels of enterprise.* Our data show increasing spatial concentrations of enterprise emerge over the ten year period, from 1986 to 1996. Highest enterprise growth from l986 to l996 occurs in and adjacent to areas with high initial levels of enterprise (such as along the edges of the 1986 Thai central corridor). Areas of convergence with low initial levels of enterprise and high growth also exist, but these "catch up" areas are largely contiguous to the existing high enterprise concentrations.

4. *There is some attenuation of enterprise at the national level, but less so at the local*

[2] Industrialization and the rise of national income are not restricted to the creation of large firms and increasing profits. Even by 1996, after a decade of high growth, non-farm enterprise continued to account for 30% of national income, more than double that of corporate profits. Satellite images have shown that extensive deforestation and urbanization occurred during this time (Felkner 2000).

3 Empirical studies show that the spatial concentration of new economic activities occurs at much finer spatial levels than might be suggested by provincial or administrative data (Demurger et. al 2002).

level. At the national level, the large, central core concentrations dissipate a bit over time: some central areas of initially high levels are associated with later, below average growth in enterprise. However, at the rural village scale, the poorer, less developed provinces have mixed patterns and show little dissipation in existing areas of concentration.

5. *Numerous regions remain stagnant.* Convergence toward equality is not the dominant story. From the perspective of the national scale, some regions appear to be permanently lagging: the hinterland in the Northeast and other areas appear to have both low levels of enterprise and low growth. There are numerous stagnant areas at the local, within-province level, as well.

6. *Levels of enterprise are correlated with favorable environmental conditions.* High levels of enterprise and/or factory density correspond with high soil fertility, lower elevation, and, at the national level, less rainfall variability.

7. *Growth in enterprise is related to favorable, exogenous environmental conditions in rural areas, at the local scale.* High growth in enterprise at the rural village scale is highly correlated with favorable physiographic conditions: good soil, proximity to rivers, and flat topography.

In addition to documenting these patterns, this paper seeks to understand the process(es) that lie beneath enterprise concentration and these spatial patterns. Our goal here is not to try to prove causality through instrumental variable techniques; rather, we view the data through the lens of a structural model with fine-tuned geographic capabilities. The model explicitly considers the geography of village locations, takes a stand on causality, and allows us to see spatially how well it fits the facts. The key elements of this model are varying levels of initial wealth, mixed but expanding access to the financial system, and a fixed cost for a household to set up enterprise or move out of subsistence agriculture and/or wage work. Households can set up a firm in their village, at a fixed cost that varies with distance to initial agglomerations of enterprise, or they can choose to either stay in subsistence agriculture in the village or costlessly migrate to work for wages (hired by the firms that are run by others). An equilibrium economy-wide wage ensures that the demand and supply for labor are equated. In a financial autarky sector, both set-up costs and capital must come from initial household wealth, whereas in a financially intermediated sector, the supply of wealth and demand for capital (plus set-up costs) are equated at an equilibrium interest rate. The end of period wealth and earnings are either consumed or saved to become the initial wealth for next period. We estimate the key parameters of the model (preferences, technology, and spatial costs) by minimizing the village-level prediction errors, the squared error between model predictions for the final period, 1996, versus the actual data from 1996.

The model fits well on several dimensions. Set-up costs increase as distance of a village from the intersection of major highways increases, despite spatial and time varying wealth and financial intermediation that are also key ingredients of the model. In particular we are able to replicate the patterns of concentration and stagnation of enterprise that we see in the data, comparing maps of the occupation choice of villages in the real data with maps of villages from the model simulation.

However, there is nothing perfunctory about these results; in fact, we discover via quantitative experiments that each of the elements of the model seems to play a role. 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 services is also central to the results. This expansion is not endogenous, rather it is related to policy levers. To establish this, we simulate the model while keeping financial infrastructure constant at the base level or, alternatively, shutting down financial intermediation entirely. Under these conditions, the model then fails to deliver observed patterns.

We also use another structural model in which we replicate our procedures to evaluate the role of intermediation and its exogeneity. In this second structural model, households weigh the gains of better risk-sharing from financial intermediation, or access to financial services, with the fixed costs of entering the intermediated sector. This model, which thus features endogenous financial deepening, tends to over predict the spatial expansion of credit in and around towns and areas of initial concentration of wealth and education, while under predicting expansion in more distant and poorer rural areas. Put differently, when we estimate the fixed cost of entry into the financial system with this second model, we find that costs decrease as one moves away from areas of enterprise concentration, further evidence that what we see in the data is indicative of a policy distortion. We remain, however, somewhat agnostic about the welfare implications of this as the preference structure of the primary occupation choice model does not lend itself easily to comparisons.

The paper proceeds as follows: Section II provides an overview of the study area, data and GIS with detailed descriptions presented in Appendices 1 and 2. Section III describes in more detail the seven key stylized facts of the data. Section IV describes the underlying occupation choice model. Section V tests the model against the facts by simulating, estimating key parameters, and comparing the results to the actual data. Section VI performs robustness checks, including a discussion of endogenouos intermediation. Section VII concludes with a larger vision for this kind of research. Appendices provide detailed discussions of the data, spatial statistics, and identification of favorable areas.

II. Study Area, Data and GIS

Our data are extensive. They include the following sources:

- Data on enterprise from a number of Thai national household and village surveys and censuses, linked spatially to Thai tambon boundaries (third administrative level) and to more than a thousand Thai village geo-locations;
- High-resolution spatial data on major and minor road networks (allowing for calculation of accessibility and travel-costs) and provincial district centers;
- Extensive physiographic data (soil quality, rainfall variation, river networks, and topography);
- Data on wealth, education, and occupation and entrepreneurial income from the Household Population Census (HPC); a national rural census conducted by the Community Development Department (CDD); a national-scale household survey from the Thai Socio-Economic Survey (SES); and data on factory and industrial types and outputs from an industrial census.

Figures I and II, respectively, provide the overall context for the study at the national level and show the four provinces for which we have detailed village-level data. Table I presents descriptive statistics of the primary variables used in our analyses. We reserve a more complete description of the data sources and methods for the appendices. To emphasize, we conduct our analysis using high-quality digital spatial data, precisely rectified in a Geographic Information System (GIS).

III. Spatial Patterns of Enterprise Growth: The Facts

Fact 1: There Is a High Degree of Concentration of Levels of Enterprise

At the Thai national level, enterprise is regionally concentrated. The left graphic in Figure Ia displays entrepreneurial activity from the CDD at the national level for 1986, our baseline year, aggregated to the Thai amphoe (second administrative) level. The graphic has colored-coded variation in the percent of households engaged in enterprise per amphoe and shows a clear regional spatial clustering of entrepreneurial activity, notably surrounding the central Bangkok metropolitan region and along a large regional central corridor extending north from Bangkok (correlated with the location of wealth as in Figure II). These levels of enterprise are considerably higher than in the poorer areas outside this corridor, such as the Northeast region. To confirm this spatial clustering of enterprise, we used the Local Indicators of Spatial Association (LISA) (Anselin 1995) spatial statistical test, which detects significant spatial clusters. The results are shown in the right-hand graphic in Figure Ia. Clusters colored red in this graphic indicate statistically significant clusters of contiguous higher-than-average enterprise levels in the Central region, while the blue clusters identify statistically significant clusters of low enterprise levels in the Northeast, North and South. (See Appendix I for more discussion on the LISA and other spatial statistics we use.)

Similar regional spatial cluster patterns are also present in data from other sources, including enterprise income data from the Thai SES shown Figure Ib, again aggregated by amphoe, for 1990. In addition, similar regional patterns were found in data from the 1990 Household Population Census (HPC) (percent of respondents engaged in entrepreneurial occupations, not shown), and in the spatial density of factories across

Thailand, derived from the Department of Industrial Works (DIW) data linked to Thai tambons for 2005 (not shown).[4]

Overall, the considerable degree of regional concentration coincides with a cross-country comparison of the degree of spatial agglomeration in more than 200 countries, conducted by the World Bank (World Bank 2009), which found that Thailand's spatial economic agglomeration in 2000 was relatively high for a developing country, with an agglomeration index value of 35.6 on a scale of 0 (low) to 100 (high), compared to, for example, 13.5 for Vietnam or 11.9 for Ethiopia, but low relative to a value of 71.9 for the highly developed US.[5] Further, the central, monocentric regional agglomeration (centered around Bangkok) is typical of developing "primate city" countries, which tend to be more dominated by a single central city than are developed countries (Overman and Venables 2010).

At the rural village scale, we are able to map clear concentrations of enterprise near large towns and intersections of major roads. Figure IIIa displays a LISA map of statistically significant spatial clusters of fractions of households in enterprise in 1986 (each province analyzed separately).[6] Enterprise is clearly clustered in specific locations within each province, confirming considerable rural spatial heterogeneity, across villages. As anticipated, the central provinces of Lop Buri and Chachoengsao (adjacent to the Bangkok area) have larger regions of significant spatial clustering of enterprise than do the northeast provinces (Sisaket and Buriram).

[4] Regressions performed at the tambon and amphoe level indicate that early on, at least, credit was also concentrated in the central national corridor (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.

[5] The Agglomeration Index calculated for the 2009 World Bank World Development Report used international GIS databases and country statistical sources to calculate the percentage of a country that is agglomerated, with agglomerated defined as 1) a population density exceeding 150 persons per square kilometer; 2) having access to a sizable settlement of 50,000 inhabitants or greater; 3) and within 60 minutes travel time by road. The index followed processes outlined in Chomitz et al (2005) and Uchida and Nelson (2010).

[6] 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. 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 I and II, as those were evaluated at the Tambon level relative to the national sample.

Fact 2: Levels of Enterprise are Negatively Correlated with Distance to Market and Infrastructure

Market access in developing countries has been found to be significantly correlated with economic development and growth, in both cross-country regional studies (Davis and Weinstein 2003; Hanson 2005; Niebuhr 2006) and rural village studies (Zeller et. al 1998; Araujo 2003). Exogenous geographic conditions have also been shown to be significant in dictating the initial spatial location of economic activity and enterprise (Hoxby 2000; Burchfield et al. 2006; Roos 2005; Ellison and Glaeser 1999; Gallup et al. 1999). For Thailand, Table II displays robust regressions of CDD enterprise onto variables capturing distance to major highways, to Bangkok and to major cities at the national level in the base year, 1986, and for SES entrepreneurial income in 1990. The percent of households engaged in enterprise in the CDD falls off significantly in most cases. For example, the fraction in enterprise drops by 0.03% for every kilometer of increasing distance from major highways (or about a 1.8% reduction in enterprise incidence per kilometer of distance, calculated against an overall mean of 1.65% of the population in enterprise). For the SES data on entrepreneurial monthly profits, bivariate regressions are not as consistently significant, but enterprise income does fall significantly with geographic distance from major highways, by 11.5 Thai Baht (about half a US dollar at 1990 exchange rates) for every kilometer of additional distance from major highways.[7]

Table II also displays the same bivariate regressions at the provincial level, using rural village data, running the dependent variable in the column on each variable in the row, one at a time. [8] The patterns are clear, striking, and more significant than at the national level: enterprise levels in 1986 decrease significantly with travel-time to major highways and amphoe district centers and with geographic distances to major road intersections.

[7] There are two exceptions here in the analysis of spatial trends with distance to cities or infrastructure: SES National Growth in entrepreneurial Income and DIW factory spatial density. Both increase with distance from the 15 largest Thai cities. Part of this may be due to food processing and nearness to natural resources.

[8] 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).

Specifically, enterprise fraction decreases by 0.0473% with every minute of increase in travel-time to major highways (mean of 12.05 minutes travel-time for all villages in the sample), by 0.0493% for every minute of increased in travel-time to amphoe district centers (mean of 14.37 minutes for all villages), and by 0.0757% with every additional kilometer of distance from major road intersections.

Fact 3: Over Time, Areas of High Concentrations and Those Nearby Show Increasing Levels of Enterprise.

This paper focuses on whether or not the concentration of enterprise increases over time in Thailand, and if so, where it has occurred and with what spatial and environmental patterns. The bivariate Local Moran and LISA (Rey and Anselin 2010) tests can be used to detect significant spatial clusters of growth regions, comparing enterprise levels in our base year, 1986, with enterprise growth between1989 and 1996 in adjacent spatial areas or neighbors. We can identify clusters where enterprise is concentrating (high base year levels surrounded by high enterprise growth), spatially expanding (low base year levels surrounded by high growth), attenuating or losing enterprise (high base year levels surrounded by low growth areas), or remaining stagnant (low base year levels surrounded by low growth areas).

We apply the bivariate Local Moran and LISA techniques to Thailand's 6,500 tambons, mapping enterprise levels from the CDD in the base year with the value of enterprise growth from 1986 to1996 in their surrounding neighbors (adjacent tambons). The results are mapped in Figure IV, with the bivariate Local Moran on the left and the bivariate LISA on the right. The LISA map shows only those clusters that are statistically significant at a p-value of .05 or less. Areas colored red in Figure IV areas with high levels of base-year enterprise surrounded by areas of higher than average enterprise growth from 1986 to 1996. As is evident, many of these areas of growing concentration 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 in the North in and around Chiengmai / Chiengrai / Lampoon. There is also convergence to the mean from below: tambons with initially low levels of enterprise surrounded by areas of higher than average growth, colored in pink. These form large areas of catch-up buffer zones in and around the central core, including much of the South and North.

Similar but distinctive patterns are seen at the rural village level. Figure V shows the results for the bivariate LISA in four study provinces. Enterprise is clearly concentrating and growing in and around major local towns and cities. More specifically, villages with high initial levels of enterprise are surrounded by villages with higher than average growth in enterprise (colored in red). (These concentrations also occur in areas of the provinces with high initial levels of wealth and/or near major road intersections and with credit availability.) Areas of convergence from below, with low initial levels in the base year surrounded by areas of high growth (colored in pink), are largely interspaced in around the agglomerating "hot spot" areas.

Fact 4: There is Some Attenuation of Enterprise at the National Level, But Less So at the Local Level

Our spatial analysis of enterprise growth reveals that there are also areas that are losing enterprise concentration: originally high levels but subsequent low growth. These include large areas in the Bangkok metropolitan area and south and north of this area along the central developed corridor of Thailand (areas visible in Figure IV colored light blue). Thus, at the national scale, there is a dissipation of enterprise in the central Bangkok metropolis.

The local picture in Figure V is consistent, but only in part. The provincial maps for Lop Buri and Chachoengsao, in the Central region, do show evidence of attenuation of enterprise within rural village enterprise clusters. Specifically, we see areas with attenuating enterprise, colored light blue, in the initially concentrated centers in the western areas of both Provinces, closest to Bangkok.[9] But we do not see evidence of attenuation of rural village enterprise clusters in both provinces of the Northeast region. Areas losing their concentration of enterprise (light blues) are rare and tend to appear in other areas, if at all. It is as if the process were just beginning and has not settled down;

[9] Compare with the upper part of Figure III and the lower part of Figure VI, as these are consistent.

that is, there has not yet been enough time for any convergence from above, falling down, to occur.

Fact 5: Numerous Regions Remain Stagnant

Significant spatial clusters of enterprise stagnation persist over the period from 1986 to1996 at both the national scale (Figure IV) and at the within-province, village level (Figure V). Areas colored dark blue in both figures indicate low levels of enterprise in the base year surrounded by low levels of enterprise growth from 1986 to1996. These areas appear to have missed out on much of the dynamic 1986 to1996 growth. At the national scale, these regions include much of the Northeast, the lower North, and the far South (a largely Islamic area) of Thailand, as identified in the left Panel of Figure IV (without delineation of statistical significance). Statistically significant stagnate enterprise spatial clusters, shown in the right Panel of Figure IV, are detected in the same regions, although the significant clusters occupy much smaller areas, largely in the Northeast region (centered around Sisaket) and in the North. Likewise, at the rural village scale in our four study provinces (Figure V), there are lagging rural areas, those which remain stagnant, some but not all of which are away from large towns and major road networks.

Fact 6: Levels of Enterprise Location are Correlated with Favorable Environmental Conditions

Environment and geography correlate with enterprise location in Thailand (again, see Table II). Enterprise percentage at the national level varies positively with higher soil fertility (improving by 0.13% with every unit increase in the soil fertility index, which ranges from 0 to 9. See Appendix I for more details), and negatively with elevation (declining by about 0.002% with every additional meter). Department of Industrial Work's factory spatial density also varies significantly with geographic distance to major rivers (decreasing by 0.0145 factories per square kilometer with every additional kilometer of distance from rivers), and to areas with lower elevation (declining by 0.005 factories per square kilometer for every additional meter of elevation) though increasing with higher annual rainfall variation (by 0.158 factories per square kilometer with every centimeter increase in annual standard deviation of rainfall, which ranges from 13 to 44 centimeters).

At the local village level, correlations between enterprise and exogenous geographic/environmental factors (geographic distance to rivers, soil fertility, elevation and annual rainfall variation) are even stronger than at the national level: coefficients are highly significant (p-values of less than one percent). Here, enterprise decreases by 0.132% with every kilometer of distance from major rivers, increases by 0.543% with every unit increase in the soil fertility index (as compared to 0.13% measured nationally), and decreases by 0.0105% with every additional meter of elevation (as compared to .002% measured nationally). (Annual rainfall variation stands out here as an exception, as the positive correlation with enterprise is surprising.

The impact of exogenous geographic or environmental factors on growth and development have been found to be significant in the US and in other countries, although direct comparisons with our findings here are not easy because the same factors were not always examined. For example, Burchfield et al. (2006) finds that, in addition to proximity to existing urban centers, physical geography (including groundwater availability and temperate climate) explains 25% of cross-city variation in new urbanization in the US. Roos (2005) finds that up to 36% of Germany's spatial GDP can be explained by direct and indirect effects of geography, while Ellison and Glaeser (1999) finds that one-fifth or more of the agglomeration of US industries can be accounted for by natural geographic advantages, including natural resource location.

Fact 7: Growth in Enterprise Is Related to Favorable Exogenous Environmental Conditions in Rural Areas at the Local Scale

At the rural village level within provinces, growth in enterprise is correlated with favorable exogenous environmental conditions. Specifically, growth in enterprise is more likely in areas of lower, flatter elevation and in areas closer to river sand waterways. As shown in the regression results in Table II, enterprise growth is decreasing by 0.192% with every additional kilometer of geographic distance from major rivers, and decreasing by 0.0139% with every additional meter of elevation.

The Figure VIa displays locations that are environmentally favorable in all four of these exogenous physiographic dimensions (determined using a GIS "suitability" model that ignores road structures or other man-made objects, described in Appendix I), while the lower Panel in the same figure displays, again, concentrations of wealth, enterprise, and credit in the baseline year (identified using a spatial analysis of wealth and enterprise data from Thai censuses, also described in Appendix I). Some of the areas of favorable exogenous physiographic/environmental conditions (Figure VIa) are also wealth/enterprise/credit clusters in the baseline year (Figure VIb). While not every one of these favorable areas turns into an economic hot spot, many of them do.

IV. A Spatial Model of Enterprise

We now turn to understanding these spatial patterns in the concentration of enterprise through the lens of a structural model. The model naturally emphasizes the occupation choices that households make; that is, whether they stay in subsistence agriculture or work for wages or, alternatively, set up a firm.[10] One constraint in setting up a firm is the talent of the owner, as well as other factors that can lower fixed costs, such as initial proximity to other firms (distance to existing initial concentrations of enterprise). This is a reduced form relationship that can be estimated in the base year cross-sectional data or as part of the overall estimation of parameters of the structural model via goodness of fit, end-of-period predictions. Actual and potential firms can also be constrained by initial wealth of the household/owner if access to credit markets is limited, as in a financial autarky sector. Initial wealth, which we measure in the data village-by-village and take as given, is, as already noted, spatially correlated in the data with exogenously physiographic conditions. Subsequent changes in wealth, which can alleviate constraints in the sector of firms without financial intermediation, are endogenous and a feature of the model dynamics (i.e., saving of end-of period income leads to an increase in wealth next period). This can then foster more business, on the extensive margin, or an expansion of existing business in size and employment, on the intensive margin. Otherwise, with perfect intermediation of savings to borrowers as in an intermediated sector, household wealth does not matter at all for enterprise,

[10] We are drawing, in fact, on a newly emerging literature which blends macro growth and development concerns with data from emerging market countries. This literature focuses on occupation and sectoral choice, talent and varying productivity, and financing constraints (Jeong and Townsend 2007, 2008; Townsend and Ueda 2006, 2010; Gine and Townsend 2004; and Buera, Kaboski, and Shin 2009; Buera and Kaboski 2007). It is motivated, in part, by ongoing measurement and analysis of the distribution of firm size and its evolution (Hseih and Klenow 2009; Pakes and Fernando(2008). But very little of this development literature sorts the data by space and location

neither for initial levels nor for growth. But a village gaining access to financial service providers can experience a shift both in the level of enterprise and in higher profits. Initial levels of financial service providers and their (exogenous) spread over time thus help to determine enterprise levels and growth. Finally, there are general equilibrium effects for all sectors, both those in the intermediated sector and those in financial autarky; the increasing demand for labor employed in firms eventually raises wages, suppressing the level of enterprise, but the increasing levels of wealth and savings decrease the interest rate, which is a boon for enterprise.

To elaborate and make the spatial assumption clear, households in each village in the model have three occupational choices: enterprise, subsistence agriculture, or wage work. To begin, each household in each village has one member who decides whether or not to become an entrepreneur (establish a firm). If a household does incur the fixed costs to run a business, it must do so in its given village location. Indeed, the fixed cost depends on the village location via proximity to infrastructure and other firms. In this sense migration of entrepreneurs is ruled out. One alternative to setting up a business is to work in subsistence agriculture, and this without loss of generality can be done in the village. As a third alternative or choice, the household can migrate and work for wages anywhere in the entire economy. For example, the household can travel to work for a firm in a neighboring or distant village (or the same village, for that matter). Migration per se is costless.[11] Location matters however for whether or not the villages have access to the services of financial sector providers.

To summarize, with regard to heterogeneity, households differ by initial, pre-determined beginning-of-period wealth and again by talent/costs in setting up businesses. The latter with the assumption that enterprise location is fixed is what makes location matter, and it influences the relative trade off for households in each village among the three occupations. This is an occupation choice model with spatial components rather than a model of location choice. At equilibrium wages, a household is indifferent about where it works for those wages, as the wage

[11] We ignore, in the present model, the location choice of wage workers and entrepreneurs . Behrens, Duranton, and Robert-Nicoud (2010) do allow some kinds of (limited) ex ante selection and replicate the stylized facts of city size in the U.S. Likewise, Desmet and Rossi-Hansberg (2009) study diffusion, concentration and the reduction of manufacturing and increase in services in the U.S. (and an increase in the value and dispersion of land rents). The inability to plot the model simulations at the level of cities or regions is symptomatic of the ever-present tension in modeling, as those authors make clear.

is common.[12] The presence or absence of the financial sector does respect geography; it is exogeneously imposed village by village, exactly as it is in the data.[13]

Of course, a village household compares returns across occupations in deciding what it chooses to do. Wages are paid in the unique consumption good of the model and are common throughout the country. Wages may equal subsistence agricultural returns early on in the development process except for a wage premium (a cost of leaving agriculture). Subsistence returns are also common throughout the country but are allowed to grow (at a common exogenous rate, mimicking technological progress). Increases in the number and scale of enterprise will, however, raise the wage over time. Profits for those running businesses subtract wages for hired labor. In addition, in the financially intermediated sector, initial wealth, in units of the consumption good, can be lent out by households in wage or subsistence agriculture and borrowed in by households running firms, all at a common interest rate, the same for everyone in this intermediated sector. Indeed, with neoclassical separation, all wealth can be saved in a financial institution and what is needed for production, if running an enterprise, then is borrowed back. Profits plus interest income thus vary across enterprises in the fully intermediated sector due (only to) variation in firm specific fixed costs and how much capital on net need be acquired

[12] In contrast, Donaldson (2010) is a prime example of a geographic analysis for India, specifically in considering the impact of railroads on spatial inequality. The Ricardian, Eaton Kortum trade model that he uses is static but tractable despite enormous geographic heterogeneity. Pawasutipaisit and Townsend (2011) is a trade model for Thailand that likewise respects differences across regions, though there are only two. The model they use is dynamic and is, in part, a generalization of the occupation/sectoral choice model of this paper. Regions can be considered to be integrated within the domestic economy up to (time varying) transportation costs and financial wedges. Villages and tambons in the Northeast are initially consuming a large fraction of their own product, though villages and tambons are much more open to domestic and international trade in the Central region. The wage is lower in the labor abundant Northeast, and the interest is lower in the capital abundant central region. As in Costinot and Komunjer (2007) study of countries, villages are shown to export products for which they either have high TFP or intensively use the factors with which they are relatively well endowed (an intensity that varies across agricultural and business/manufactured goods). Then, over time, as transportation time decreases, interest rates converge across regions. But unlike a Heckshire-Ohlin-Vanek model of factor price equalization, wages tend to rise in all regions, an apparent anomaly. Still, the dynamic structural model tells a story of increased demand for labor as financial markets improve and collateral constraints are lowered.

[13] Moll, Townsend, and Zhorin (2011) develops and computes solutions to a spatial model with an economic transition to a steady state of growth, with moral hazard constraints on the financing of enterprise in the Central region and limited commitment, and collateral constraints in the Northeast, as measured in the Townsend Thai data (Paulson, Townsend and Karaivanov 2006; Ahlin and Townsend 2007). We already know this type of analysis can be extended to much higher dimensions, with variation across villages, within regions, and as in the micro data used in this paper. Yet, tradeoffs remain, as we have highlighted in this concluding section. The incorporation of forward-looking dynamics with regional and village-level heterogeneity forces common, economy-wide wage and interest rates for tractability. Still, we are confident that tractable and increasingly realistic spatial models for developing countries are within reach.

on credit.

In a financial autarky sector, however, households in business must use their own wealth as capital (or store initial wealth in a backyard storage technology, to the end of the period, as inefficient informal savings). Thus, with no formal sector financial savings and no credit, neoclassical separation fails, and thus wealth also influences occupation choice. A low wealth household that chooses to run a business may be forced to do so at a small scale, and thus at a low level of profits, despite potentially high marginal returns.

Again, a comparison of end-of-period earnings across the three occupations determines which occupation a household will choose in each of the financial autarky and financially intermediated sectors in that period. So end-of-period wealth will vary across households and villages and differ from initial wealth by the amount of the chosen net earnings. Goods can be transported without cost in the model, i.e., there are no shipping costs, just as there are no migration costs, and so all returns are measured with consumption good (and wealth) as the *numéraire*.

The distribution of initial wealth by village location is taken from the data, but the subsequent evolution of wealth by location is endogenous, that is, determined by the model. The common wage of each period is endogenous, equating the demand for labor (as an entrepreneur in a firm, hires in firms as wage earners or employment in subsistence agriculture) to the supply of labor (each household has one unit and the population is fixed). The interest rate in the financially intermediated sector will balance the demand for wealth as capital and fixed costs with the supply of wealth.

In more precise notation, a household at a given village location is distinguished by a pair of beginning-of-period characteristics: initial wealth b_t and randomly drawn entrepreneurial set-up costs x_t (and the village is distinguished by whether it is in financial autarky or the fully intermediated sector). The latter cost x_t is a fixed cost independent of the scale of operation k_t and can be thought of as the purchase of a pickup truck, for those engaged in wholesale retail trade, the purchase of a building for storage, etc. Below, γ is the subsistence agriculture parameter and η is the cost for wage earners and entrepreneurs of switching out of agriculture.

Thus, given an equilibrium wage rate w_t, a household of type (b_t, x_t) in the financial autarky sector chooses its occupation to maximize his total end-of-period wealth, W_t:

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

For this last line, for entrepreneurs, firms in the enterprise sector choose capital input k and hired labor l to produce the consumption good, with profits:

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

The model implies that wage must be $\underline{w} = \gamma + \eta$ when the subsistence agriculture sector coexists with the modern sector firms doing production. We do generalize and allow subsistence income, γ, to grow exogenously at rate g_γ. All of η, γ, and g_γ are parameters to be estimated.

Again, there is a fixed cost x_t 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 agglomeration centers. These setup costs are assumed to be independent of total wealth b and randomly drawn each period from a time-invariant cumulative distribution with key parameter m determining the skewness of cost draws (and allowed to depend on distance d to economic centers). Specifically, the cumulative distribution of cost draws is[14]

$$H(x, m(d)) = m(d)x^2 + (1 - m(d))x.$$

The support of fixed cost x is unit interval $[0,1]$, and the range of possible values for parameter m is $[-1,1]$, where, again, parameter m is determined by d, written $m(d)$. This class

[14] 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.

of cumulative distributions subsumes the uniform distribution at $m(d)=0$. As $m(d)$ increases toward 1, the distribution of fixed cost x skews to the right, and hence potentially efficient, low-cost entrepreneurs become rare. We hypothesize that the $m(d)$ parameter will vary positively with geographic distance d to areas of initial concentration of enterprise. This is as a kind of reduced form relationship, meant to capture the hypothesis the farther a village is from an agglomeration center (in terms of travel-time), the less households in that village will be able to benefit from agglomeration economies.[15] We focus here on the estimation of parameter $m(d)$ using the spatial data, given the structure of the model, to see if our hypothesis is correct. Given limited data, we let d take on three values: near, medium, and far (see below). These are three more parameters to estimate.

In terms of additional parameters, the technology to produce the single consumption good is captured by production function f:

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

thus with five parameters $\alpha, \beta, \xi, \rho,$ and σ. Function f is intended as an approximation to any arbitrary production function. These five parameters must also be estimated.

Choices in the cross section, at any date t, are captured by the occupation map of Figure VII. For a given fixed wealth $b_t = b$, and set up cost x, a household's choice is pinned down, with the exception of a curve of indifference, where households are on the margin. Relatedly, with respect to the entire population, the fraction of households of wealth b who choose to run an enterprise is determined by their cost x value, that is, the number of households below the critical cost threshold of indifference separating the area marked in Figure VII as entrepreneurs (both constrained and unconstrained) from the area marked as wage earners. Intuitively, this fraction of households in enterprise is the distance from the abscissa axis to the occupation choice threshold indifference curve, exactly so if costs are uniformly distributed. Otherwise, for a

[15] Note that in the assumed cost technology x does not depend on d directly, but the greater is d the more likely x is higher.

household of wealth b at geographic distance d from an agglomeration center, one integrates under the cumulative distribution $H(x,m(d))$ of cost draws up to the threshold curve. In this, we are assuming a continuum of households, as an approximation.

Note from Figure VII that the higher 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 (plus subsistence) sector is influenced not only by geographic distance d through the distribution of set up costs $H(x,m(d))$ but also *by wealth* b.

Relatedly, if a household is in a village without access to a financial service provider in some period - i.e., without credit and savings in some period - then upon becoming a firm, both the set up costs and the working capital must be financed in that period out of accumulated, beginning-of-period wealth. If that wealth is limited, this can be constraining, as indicated in the same occupation map, Figure VII. A firm may operate at a small scale even with a high marginal return. If wealth is less than the fixed cost, then firms cannot operate at all, and enterprise frequency can fall off abruptly in the cross section, as wealth decreases.

The model features simplified household dynamic problems. A household with end-of-period wealth W_t at date t maximizes individual preferences over consumption c_t and saving (wealth tomorrow) b_{t+1}, as represented by the utility function:

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

subject to the end of period budget constraint, $c_t + b_{t+1} = W_t$. Here, ω is the savings rate for (all) households, another parameter which must be estimated. The optimal rules for consumption and saving will be linear functions of wealth, and preference maximization is equivalent to end-of-period wealth maximization, period by period. One could say that we have imposed a myopic savings rate. This helps make computation doable.

Again, there is also an (exogenously specified) intermediated sector that allows savings to flow to borrowers at an equilibrium interest rate. One can think of a household in a village in the intermediated sector as having access to a bank to either deposit funds, initial wealth, or borrow for running enterprise. Interest is received or paid at the end of the period. Within the period, funds (the consumption good) flow costless from savers to borrowers, within and across villages in the intermediated sector; there is no mark-up and intermediation per se is not costly.

In more precise notation, let $R_t = 1 + r_t$ denote the interest rate (again for simplicity both the deposit rate and the cost of funds). A household chooses to maximize end-of-period wealth W_t so that:

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

where the profits of an entrepreneur are unconstrained by wealth, and capital is obtained at the prevailing interest rate, a decision made jointly with hired labor:

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

The occupation map of Figure VII is quite simple for the intermediated sector. There is a constant threshold for set-up cost x*, the same for all wealth b, above which a household will be a wage earner (or in subsistence) and below which an entrepreneur. Population fractions in the intermediated are determined by this threshold (indifference) line, as before.

The weight of the intermediated sector overall in the entire economy depends on the fraction of villages having measured access to intermediation. The financial sector has expanded in Thailand during this period, so this too adds to the dynamics of the model. We know by name (and location) which villages have intermediation and which do not. This financial intermediation part of the model we take as exogenous (though we shall reexamine this in the alternative endogenous financial deepening model).

As described in the next section, all nine parameters of production $(\alpha, \beta, \rho, \sigma, \xi)$, savings rate ω, initial 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 last year of the CDD data and those predicted from the model, summing over villages in each province.

V. Testing the Spatial Occupational Choice Model With Agents in Villages

Now we use the model to try to make sense of the data. First, as anticipated, we simulate the model to get predictions at given guessed parameter values. Then we estimate the above list of parameters by a metric or criterion function, finding the simulation that fits the last period enterprise rates the best. Then we compare the model generated output at those estimated parameters to the data, contrasting maps of the actual data and the simulated data, also highlighting the estimated cost function. Finally, we conduct numerous robustness checks for parameters and assumptions of the model, to see what is driving the results.

V.A. Simulation

The occupational choice model, as exposited in the previous section, 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. For the initial base year, 1986, we impute to each household in a given village a wealth number that is the village average, and then let the model take over, allowing within village diversity in subsequent years. The model uses village level data and is run at the provincial level; here, we focus on enterprise growth patterns, one province at a time.

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 a government bank (the Bank for Agriculture and Agricultural Cooperatives, known as BAAC) and commercial bank credit access in a given year, and it was given a value of zero otherwise. This, then, is a conservative indicator, focusing on the two primary formal sector providers, jointly. That is, values of the index are imposed exogenously for 1986, 1988, 1990, 1992, 1994, and 1996, and 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.

An equilibrium is computed by finding a common market clearing wage overall and an interest rate in the intermediated credit sector, both computed in a numerical, bisection algorithm which varies factor prices up or down, depending on excess demand or supply. Essentially, if an initial wrong wage is used (at a given interest rate), then the population fractions assigned to various occupations do not add to unity, and the wage is adjusted, upwards, for example, if there is excess demand for labor. Likewise, demand for credit in the intermediated sector must equal initial aggregate wealth supplied, and this determines the interest rate.

We allow enterprise set-up costs to vary across village travel-time, as measured using the GIS through road networks, to the nearest economic agglomeration, as proxied by nearest intersection of major highways (displayed in the lower panel of Figure VI, for example). Of course, there are not enough villages across the continuum of travel-time distances d to estimate $m(d)$ reliably across all distances d. Again, the full sample was stratified into three discrete bins with equal number of villages in each bin, near, medium and far.

The criteria used to judge how well the model is fitting are the latter's point-by-point, village-by-village end-of-period prediction errors. That error is the final fraction of those running an enterprise in each village, for the end-period 1996, predicted from the model, relative to the actual fraction in each village in the data for 1996. Indeed, the squared prediction error summed over all the villages in a given province is the metric/criterion function, and this is minimized to deliver the best-fit parameter values. The list of parameters was enumerated earlier.

V.B. Model Predictions

To get to the heart of the matter, we feature the best fit prediction of the model, comparing it to the data. We do this not only for the final year but also for more dynamic patterns using the additional time varying data that the model generates, specifically growth rates from beginning to end of the sample. Using additional data, not used in estimation, is, of course, a form of model validation. A major criterion that we use *ex post* to judge how well this spatial structural model predicts is its ability to replicate the spatial patterns of increasing concentration that we have seen in the actual data, the maps.

Figure VIII displays the results of the best-fit structural simulation. Here, we have applied the Bivariate LISA technique to the final simulated results, just as we did for the actual data in Figure V, detecting statistically significant spatial regions where enterprise was concentrated and is growing (colored red), rising anew (pinks), attenuating (light blues), or remaining stagnant over time (dark blue) from 1986 to1996. As is evident from a comparison of Figure VIII and Figure V, the model does remarkably well. Areas of increasing concentration include the western corridors of both Chachoengsao and Lop Buri, both of which lie on the outskirts of the heavily developed Bangkok metropolitan area and were identified in the national-scale analysis (Figure IV) as experiencing significant initial concentration and subsequent growth of enterprise. In addition, the model does well in capturing areas where enterprise was concentrated and growing in the Northeast provinces, specifically in areas that were identified as environmentally advantageous (top Panel of Figure VI). The model also correctly simulates areas where new enterprise is developing (areas with initially low levels surrounded by high enterprise growth), colored in pinks, primarily in and adjacent to the areas of increasing concentration (reds) and along major road networks and major intersections in all provinces, such as in western Lop Buri and along the main central roads of Buriram. However, the model does simulate some areas of enterprise concentration that do not exist in the actual data, especially in Northeast.

V.C. Spatial Variation in Enterprise Set-Up Costs

Notably, in our primary simulation where we allow spatial set-up costs $m(d)$ to vary, the estimated $m(d)$ values increase monotonically with increasing travel-time/geographic distance from markets and agglomeration centers, though they are not required to do so in the estimation. See Table III. Again, a higher $m(d)$ value estimated by the model indicates a greater skewness towards higher enterprise set-up costs in the population distribution of villages. The estimated parameter value increases from -0.1013 for the sample of villages with lowest travel-time accessibility to infrastructure (bin 1), to 0.5986 for the sample of villages with intermediate travel time (bin2), and finally to 0.6202 for the sample with highest travel-time to markets (bin 3). This monotonic increase in costs across space was confirmed in multiple robustness tests.

VI. Robustness Checks and Extensions

One key to arriving at a superior fit to final-period data in our primary estimation, relative to a non-spatial simulation, is this spatially varying set-up cost function. Specifically, we perform a robustness check, where we shut down the model's ability to allow set-up costs to freely vary spatially, and impose common parameter m regardless of distance d. We then simulate through to the final period 1996. We find that the simulated results from this spatially restricted model are indeed inferior, both in terms of overall correlation with the actual end-period data and in terms of the ability of the model to capture the spatial patterns of enterprise growth. Allowing spatial variation in set-up costs results in a positive correlation coefficient of 0.1638 between the final-period predicted values and the actual 1996 data, and a p-value of 0.000, compared to 0.0746, with a p-value of 0.0444, when this variation is not allowed.

Further, Figure XI depicts a LISA map of the residuals between the best fit model simulation, allowing for spatial diversity in set-up costs, and the actual data, while the bottom Panel shows the results of LISA run on the residuals between the model, shutting down spatial diversity in costs, and the actual data. The model allowing for spatial diversity in costs has fewer significant spatial clusters in its residuals with the actual data. Likewise, Table IV reveals that the prediction errors for the non-spatial model are correlated with a set of variables capturing village travel-time accessibility to markets through the road network, as well as to base-year wealth and education.

There are further robustness checks. Rather than allowing all parameters to be estimated freely, as we do with our best fit simulation, we instead simulate the spatial model using the original parameter values taken from Jeong and Townsend (2008). Results are listed as Robustness Check 1 in Table III. Additionally, in Robustness Checks 2 and 3, we start by imposing the Jeong and Townsend (2007) parameters, but then allow a subset of them 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 an equal number of villages in each bin, as in the base line primary specification, and second, using five bins stratified equally by variation in travel-time agglomeration access (but not forcing an equal number of villages into each bin). Another Robustness Check 4 allows all parameters of the model (technology, preferences, etc.) to be estimated for each bin of differing market access separately. In another quite interesting variation, we re-estimate the model by binning villages across space by geographic travel-time to environmentally advantageous areas (identified using a GIS "suitability" model, described in detail in Appendix I, listed as Robustness Check 5). With only two exceptions, costs vary monotonically with distance.

Another set of robustness checks is conducted to gage the importance of the expansion of financial intermediation in generating accurate simulations. First, we eliminate the expansion, by freezing access to its initial 1986 level and then by getting rid of the intermediated sector entirely, so that no village is intermediated ever, with results listed as Robustness Checks 6 and 7. We display the application of the bivariate LISA technique to the predictions for increasing concentration in Figure X, holding intermediation fixed at 1986 levels. Though this simulation does capture increasing concentration in the Northeast, it appears to miss this concentration in the Central provinces, especially in Chachoengsao where this version of the model fails to predict any areas of increasing concentration (colored red). The model performs even worse at capturing the spatial patterns of enterprise concentration when we remove the intermediated sector entirely (the bivariate LISA applied to these results is shown in Figure XI). 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 two checks and the actual spatial patterns in enterprise growth, as shown in Table III, were 0.0634 (p-value of 0.049) for the intermediation fixed check and 0.0314 (p-value of 0.079) for the no intermediation check, again as compared to a correlation of 0.1638 when credit is included and as expanding in the data. We conclude that observed intermediation expansion is a key part of the model.

The final robustness check is related. We now allow the expansion of endogenous intermediation, though in a somewhat different structural model and then argue that had we used these endogenous predictions, rather than the observed actual expansion as above, we would have biased the gradient of the key spatial set-up cost function downward.

We use a spatial version of Greenwood and Jovanovic (1990). This is also the model used by Townsend and Ueda (2006) and Jeong and Townsend (2008). See earlier working version of Felkner and Townsend (2009). This model features the choice of joining the financial system, which we refer to hereafter as the financial deepening model. In the version here, there is a spatially varying fixed cost for entry into the financial system (capturing roads, bank infrastructure and/or household learning) and we treat households as villages; i.e., there is one representative household per village. This entry cost into the financial sector is nontrivial, so relatively low wealth villages choose to remain in autarky. But other villages, although possibly at the same given geographic distance from the branch of the bank (as proxied by distance to major intersections), may have wealth higher than a key threshold value and have thus joined the financial system, thus taking advantage of better information about macro shocks and improved allocation of risk for idiosyncratic shocks (full risk sharing). Key decisions for those villages in financial 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 aggregate and idiosyncratic shocks. The alternative is to put savings into relatively safe agricultural activity with low returns. A key decision for those villages in the financial sector is also how much to save out of income. Nonetheless, since for these villages investment is, in effect, under advice of a bank, with better information than those in autarky, and in a situation where all idiosyncratic shocks are pooled, project choice reduces to comparing expected returns. In sum, in this model, the higher the fixed cost of entering the financial system, the higher the critical value of wealth must be, and the larger the fraction of the villages that choose to stay in financial autarky.

The financial deepening simulation, with common entry costs that do not depend on distance, produces predictions of whether a village is financially intermediated or not, and its wealth, for all time periods. The simulation does an excellent job of capturing overall dynamic trends: starting with the imposed initial condition of 27% of villages in intermediation, the model accurately captures the growth rate of intermediation and closely matches the final period 1996 percent of all villages intermediated, at 46%. The overall end-of-period means for actual and simulated household wealth are also quite close (46.96 for actual versus 43.15 for simulated) as are the cross-sectional standard deviations of 38.83 for actual versus 36.23.

Nevertheless, a spatial analysis of the predictions from the model reveals a glaring anomaly. See Figure XII. Reds are areas where the model has over-predicted the fraction of villages in intermediation. The map reveals that the model tends to over-predict financial access in areas of economic agglomerations, including in western Chachoengsao (infrastructure corridor adjacent to Bangkok), southwestern Lop Buri, and clusters in Buriram and Sisaket. Likewise the greens are areas of underprediction, in areas distant from major roads.

This trend is confirmed quantitatively. There is indeed a significant correlation between model over-predictions of financial access and travel-time accessibility to markets, infrastructure, and government district centers In order to determine whether the model overestimates the influence of geographic conditions in the wealth and financial access per village, we computed the residuals subtracting the values yielded by the model from the actual values. Hence, we have four variables: Wealth Residuals, Credit Intermediation Residuals, BAAC Simulation Residuals and Commercial Banks Simulation Residuals.

Notice that for the case of Credit Intermediation Residuals, BAAC Simulation Residuals and Commercial Banks Simulation Residuals, observations could take three possible values: 0, if the predicted value is the same that the actual one; 1, if the model overestimates; and -1, if the model underestimates.

OLS regressions were run for every dependent variable using travel-time to major roads, travel-time to major highway intersections and travel-time to amphoe district centers as regressors, one at a time (major roads, major highway intersections, and amphoe district centers for the four study provinces are displayed in Figure XIII). In all the regressions, we controlled for wealth and education differences and included province fixed effects.

Table V presents the results for these regressions. It can be seen that all the three separate explanatory variables tend to overpredict wealth levels and BAAC presence. Regarding Credit Intermediation only travel-time to major roads was not statistically significant; while for Commercial Banks only distance to major intersections seems to be a source of overprediction.

When we allow costs of financial access to vary across space in the model simulation, the results thus imply that the costs of financial access would be higher in towns and economic agglomerations, or conversely lower in rural hinterland areas away from agglomerations. We do not believe this could be real. Rather, the data through the lens of this endogenous financial deepening model are telling us that there must be distortions. We suspect the primary one is targeting, that the government establishes branches of its agricultural bank, the BAAC, and seeks out clients, focusing on rural areas. This suggests that if we somehow had been able to put endogenous, rather than exogenous, access to financial intermediaries in the occupational choice model itself, we would likely have undercut the estimated firm set-up cost gradient, incorrectly raising costs near urban centers and lowering them in the hinterland to counteract the credit distortion. This thus validates, in a different way, the use of exogenous credit expansion in the occupation choice model.

VII. Conclusions

This paper looks at the complex process of development in Thailand by examining the heterogeneity of economic activity across space. To do so, we fit a structural, micro-founded model of occupation transitions with spatial heterogeneities and fine-tuned geographic capabilities to these village data and replicate the salient spatial facts. Key ingredients appear to be costs as functions of distance to infrastructure and markets, credit constraints on occupation choice, and variation in initial wealth.

Overall, we find salient patterns of increased concentration of enterprise, coupled with more subtle patterns of contiguous geographic convergence that leave stagnant areas behind. Exogenous favorable physiographic conditions are also an important factor, especially at the local level. Specifically, initially high levels of enterprise are clustered in and around Bangkok and along main transportation arteries, while at the rural village level, they are at or near large towns and major roads intersections. Fractions of population in enterprise decrease as travel-time and geographic distances to markets and infrastructure increase. Subsequently, the highest enterprise growth occurs in and adjacent to areas with high initial levels of enterprise. At the national level central core concentrations dissipate a bit over time, though at the rural village scale, this has not happened. More generally, convergence toward equality is not the dominant story as some regions are permanently lagging. The environment also plays a role. High initial levels and high growth in enterprise at the rural village scale are correlated with favorable physiographic conditions: good soil, proximity to rivers, and flat topography.

Our model embodies the agglomeration and externalities that urban economists have found to be important in the U.S., Europe, and other settings. We are also are drawing on a newly emerging literature which blends macro growth and development concerns with data from emerging market countries, here incorporating space explicitly.

However, we do not address and cannot distinguish potential spillovers, the role of technological diffusion, iceberg transportation costs, costly migration and the location choice of wage workers and entrepreneurs. The tradeoff is whether there are economy wide markets at common prices, as in this paper, or, if not, how these can vary spatially while retaining tractability. Yet, we are confident from ongoing efforts that these elements are increasingly within reach.

Interactions of the next generation of models with the analysis of available spatial data will allow us to better understand the forces for increasing inequality and geographic concentration. As this paper has established, the spatial aspects of enterprise are a crucial component of the larger process of development.

APPENDIX I: Data and the 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.[16] Chachoengsao and Lop Buri are in the richer Central region, relatively near Bangkok; Buriram and Sisaket are in the poorer Northeast (see Figure II).

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 high-resolution subset of this GIS was created for the four study provinces that included more than 1,000 village geo-locations, high-resolution (primary/secondary and tertiary) road networks, Amphoe government district centers, and again physiographic data.[17] Highly detailed road network data for the four study provinces, displayed in Figure XIII, 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. That is, new roads may cause development, unlike the assumptions of the model. 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

[16]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).

[17]Political boundaries were obtained from the Thai government, allowing spatial aggregation of socioeconomic CDD village variables by administrative unit. Road networks for the four provinces were digitized from the Thai government maps, with further accuracy assessment from independent lands at 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.

(ADC) Inc.,[18] 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.[19]

Some new secondary roads were constructed; however, these are present in the ADC but not in the earlier Thai maps. Still, these are very few in number,[20] 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 0.15, 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 extensive 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.[21] For the physiographic elements, data on soil fertility were compiled from Thai government maps of soil types.[22] Data on annual rainfall variation were calculated from Thai meteorological stations collected by the Thai weather service 1951-1996 then interpolated to village locations using

[18] 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.

[19] 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.

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

[21] 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 of Chicago, 1997.

[22] 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.

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) and extracted from vectorized contour lines on high quality Thai land cover maps. Basic statistical summaries of these key variables are reported in Table I.

For the four study provinces, we use primarily village socioeconomic data collected by the CDD every two years from 1986 to1996.[23] 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.[24] Education was measured as a percent of the

[23]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.

[24]The four assets that were used for this procedure -- per capita TV ownership per village, per capita motorcycles per village, per capita pickup 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

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.

A. Spatial Statistical Tests for Spatial Clusters, Concentration

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).[25] 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).[26] 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:

---

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% 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.

[25]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 and Wong, 1991).

[26]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:

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

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 that can be used to decompose the global statistic down to local levels, and identify hot-spot clusters.

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

where

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

This statistic is a measure of the strength of the spatial correlation of an observation with its neighbors.[27] 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).

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).[28] To visualize the locations of (significant) spatial clusters, a widely-used approach is to map not the actual Local Moran's I values, but

[27]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).

[28] 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 re sampled 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 to (and conditional on) the values computed under the null hypothesis (the randomly permuted values).

rather the corresponding values of the Moran Scatter plot, binned into four groups depending on the type of spatial association that an observation has with its neighbors. The Moran Scatter plot technique plots the standardized response variable (that is, with the mean subtracted and divided by the standard deviation) against the standardized 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 Scatter plot), 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 IV, for example, displays the Moran Scatter plot values for enterprise for all tambons in Thailand. However, to visualize which of these clusters are statistically significant, normally the Moran Scatter plot 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 Scatter plot 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 (Rey and Anselin, 2010). We have applied this method to enterprise in Thailand, with the Local Moran's , 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 Scatter plot (for the same variables) for those observations that are significant.

B. 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 and 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 test whether residuals in our spatial structural simulations correlate with market or infrastructure access. Summary statistics for the key continuous variables are show in Table I, with regressions results for enterprise and enterprise growth onto the key travel-time and distance proxies given in Table II 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 IV and V. In all cases, village access was measured by calculating travel-time through the road networks.[29]

C. 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 VI. An additional variable sought to provide a proxy for accessibility to major government, financial, and regulatory institutions, 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 XIII).

[29]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). 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). When traveling through road networks, routes for movement of people and goods are usually selected to minimize travel-time (and these routes may or may not be the shortest route in terms of distance, due to travel-time variation in road quality), thus we used a “least-cost path” network algorithm which sought to minimize travel-times, given the digital road network data. 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.

D. Infrastructure Access Variables

Infrastructure provides conduits for accessibility not only to local markets but also 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.[30]

[30] 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.

APPENDIX II: Modeling Areas of Favorable Geography Using a GIS Suitability Model

A. 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; Malczewski 2004) in a raster GIS context to identify locations in our study areas.

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

1. Areas of higher soil fertility are economically preferable, because of the potential for greater agricultural productivity;

2. 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 are more favorable for economic growth.

3. 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;

4. 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;

5. It is preferable (lower cost) to build in topographically flat areas, compared to steep slopes, and over time 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 were then added:

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.

The locations with the highest clustered values in these output layers were identified with points, and are displayed in the top Panel of Figure VI. Finally, the distance of each CDD village to the nearest of these points was calculated, with this value being a measure of relative distance to favorable geography for each village.

References

Ahlin, Christian, and Robert M. Townsend, "Using Repayment Data to Test Across Models of Joint Liability Lending," *Economic Journal*, 117 (2007), F11-51.

Anselin, Luc, "Local Indicators of Spatial Association-LISA," *Geographical analysis*, 27 (1995), 93-115.

---, "Interactive Techniques and Exploratory Spatial Data Analysis," in *Geographical Information Systems: Principles, Techniques, Management and Applications*, P. Longley, M. Goodchild, D. Maguire and D. Rhind, eds. (New York: Wiley, 1999).

---, "The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association," in *Spatial Analytical Perspectives on GIS*, Manfred M. Fischer, Henk J. Scholten, and David J. Unwin, eds. (London: Taylor and Francis, 1996).

Aghion, Philippe, and Patrick Bolton, "A Theory of Trickle-Down Growth and Development," *Review of Economic Studies*, 64 (1997), 151-172.

Araujo, Caridad, Alain de Janvry, and Elisabeth Sadolet, "Spatial Patterns of Non-agricultural Rural Employment Growth in Mexico During the 90's," Mimeograph, University of California at Berkeley, 2003.

Assuncao, Juliano, Sergey Mityakov, and Robert M. Townsend, "Commercial vs. Government Development Banks," Mimeograph, Clemson University, 2008.

Banerjee, Abhijit, and Andrew F. Newman, "Occupational Choice and the Process of Development," *Journal of Political Economy*, 101 (1993), 274-298.

Behrens, Kristian, Gilles Duranton, and Frederic Robert-Nicoud, "Productive Cities: Sorting, Selection, and Agglomeration," CEPR Discussion Paper No. 7922, 2010.

Buera, Francisco J., "A Dynamic Model of Entrepreneurship with Borrowing Constraints," Ph.D. Dissertation, (Northwestern University), 2003.

Buera, Francisco J., and Joseph Kaboski, "Scale and the Origins of Structural Change," *Journal of Economic Theory*, forthcoming, 2011.

Buera, Francisco J., Joseph Kaboski, and Yongseok Shin, "Finance and Development: A Tale of Two Sectors," NBER Working Paper No. 14914, 2009.

Burchfield, Marcy, Henry G. Overman, Diego Puga, and Matthew A. Turner, "Causes of Sprawl: A Portrait from Space," *Quarterly Journal of Economics*, 121 (2006), 587-633.

Calem, Paul S., and Gerald A. Carlino, "Urban Agglomeration Economies in the Presence of Technical Change," *Journal of Urban Economics*, 29 (1991), 82-95.

Chomitz, Kenneth M., Piet Buys, and Timothy S. Thomas, "Quantifying the Rural-Urban

Gradient in Latin America and the Caribbean," World Bank Policy Research Working Paper No. 3634, 2005.

Chou, Yue-Hong, *Exploring spatial analysis in geographic information systems*, (Santa Fe, NM: OnWord Press, 1997).

Costinot, Arnaud, and Ivana Komunjer, "What Goods Do Countries Trade? A Structural Ricardian Model," NBER Working Paper No. 13691 (2007).

Davis, Donald R., and David E. Weinstein, "Market Access, Economic Geography, and Comparative Advantage: An Empirical Assessment," *Journal of International Economics*, 59 (2003), 1-23.

Démurger, Sylvie, Jeffrey D. Sachs, Wing Thye Woo, Shuming Bao, Gene Chang, and Andrew Mellinger, "Geography, Economic Policy and Regional Development in China," *Asian Economic Papers*, 1 (Winter, 2002), 146-197.

Desmet, Klaus, and Esteban Rossi-Hansberg, "Spatial Growth and Industry Age," *Journal of Economic Theory*, 1446 (2009), 2477-2502.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," NBER Working Paper No. 16487, 2010.

Duffy, Neal, "Returns To Scale Behavior and Manufacturing Agglomeration Economies in US Urban Areas," *Regional Science Perspectives*, 17 (1987), 42-54.

Eaton, Jonathan, and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, 70 (2002), 1741-1779.

Ellison, Glenn, and Edward L. Glaeser, "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" *American Economic Review*, 89 (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 (Cambridge: Harvard University, 2000).

Felkner, John, and Robert M. Townsend, "The Geographic Concentration of Enterprise in Developing Countries," Working Paper, 2009. http://www.robertmtownsend.net/sites/default/files/files/papers/published/GeographicConcentration2009.pdf

Fernandes, Ana M., and Ariel Pakes, "Factor Utilization in Indian Manufacturing: A Look at the World Bank Investment Climate Surveys Data," NBER Working Paper No. 14178, 2008.

Fotheringham, A. Stewart, and David W. S. Wong, "The Modifiable Areal Unit Problem in Multivariate Statistical Analysis," *Environment and Planning*, A23 (1991), 1025-1044.

Gallup, John Luke, Jeffrey D. Sachs, and Andrew D. Mellinger, "Geography and Economic Development," *International Regional Science Review*, 22 (1999), 179-232.

Gine, Xavier, and Robert M. Townsend, "Evaluation of Financial Liberalization: A General Equilibrium Model with Constrained Occupation Choice," *Journal of Development Economics*, 74 (2004), 269-307.

Goodchild, Michael F., and Sucharita Gopal, *The Accuracy of Spatial Databases*, (London: CRC Press, 1989).

Greenwood, Jeremy, and Boyan Jovanovic, "Financial Development, Growth, and the Distribution of Income," *Journal of Political Economy*, 98 (1990), 1076-1107.

Hanson, Gordon H, "Regional Adjustment to Trade Liberalization," *Regional Science and Urban Economics*, 28 (1998), 419-444.

---, "Market Potential, Increasing Returns, and Geographic Concentration," *Journal of International Economies*, 67 (2005), 1-24.

Henderson, J.Vernon, and Hyoung Gun Wang, "Aspects of the Rural-Urban Transformation of Countries," *Journal of Economic Geography*, 5 (2005), 23-42.

Hoxby, Caroline M., "Does Competition Among Public Schools Benefit Students and Taxpayers?" *American Economic Review*, 90 (2000), 1209-1238.

Hsieh, Chang-Tai, and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124 (2009), 1403-1448.

Hubert, Lawrence J., *Assignment Methods in Combinatorial Data Analysis*, (New York, NY: Marcel Dekker, 1987).

Jeong, Hyeok, and Robert M. Townsend, "Sources of TFP Growth: Occupational Choice and Financial Deepening," *Economic Theory*, 32 (2007), 179-221.

Jeong, Hyeok, and Robert M. Townsend, "Growth and Inequality: Model Evaluation Based on an Estimation-Calibration Strategy," *Macroeconomic Dynamics Special Issue on Inequality*, 12 (2008), 231-284.

Joerin, Florent, Marius Thériault, and André Musy, "Using GIS and Outranking Multicriteria Analysis for Land-Use Suitability Assessment," *International Journal of Geographical Information Science*, 15 (2001), 153-174.

Kanbur, Ravi, and Anthony J. Venables, *Spatial Inequality and Development*, (New York, NY: Oxford University Press, 2005).

Lloyd-Ellis, Huw, and Dan Bernhardt, "Enterprise, Inequality and Economic Development," *Review of Economic Studies*, 67 (2000), 147-168.

Malczewski, Jacek, “GIS-Based Land-Use Suitability Analysis: A Critical Overview,” *Progress in Planning*, 62 (2004), 3-65.

Merrill, Deane W., and Steve Selvin, “Analyzing Spatial Patterns in Health Outcome Surveillance Data,” (paper presented at the Sixth Biennial CDC and ATSDR Symposium on Statistical Methods, Lawrence Berkeley National Laboratory, University of California at Berkeley, CA School of Public Health, January 28-30, 1997).

Moll, Benjamin, Robert M. Townsend, Victor Zhorin, “Financial Frictions and Development: Limited Commitment vs. Private Information,” Mimeograph, Massachusetts Institute of Technology, 2011.

Moran, Patrick Alfred Price, “The Interpretation of Statistical Maps,” *Journal of the Royal Statistical Society*, Series B (Methodological) (1948), 243-251.

Nelson, Andrew, Alexander de Sherbinin, and Francesca Pozzi, “Towards Development of a High Quality Public Domain Global Roads Database,” *Data Science Journal*, 5 (2006), 223-265.

Niebuhr, Annekatrin, “Market Access and Regional Disparities. New Economic Geography in Europe,” *Annals of Regional Science*, 40 (2006), 313-334.

Nordhaus, William D., “Geography and Macroeconomics: New Data and New Findings,” *Proceedings of the National Academy of Sciences*, 103 (2006), 3510-3517.

Overman, Henry G., and Anthony J. Venables, “Evolving City Systems,” UNU-WIDER Working Paper No. 2010/26, 2010.

Paulson, Anna L., and Robert M. Townsend, “Entrepreneurship and Financial Constraints in Thailand,” *Journal of Corporate Finance*, 10 (2004), 229-262.

Paulson, Anna L., Robert M. Townsend, and Alexander Karaivanov, “Distinguishing Limited Liability from Moral Hazard in a Model of Entrepreneurship,” *Journal of Political Economy*, 114 (2006), 100-144.

Pawasutipaisit, Anan, and Robert M. Townsend, “Wealth Accumulation and Factors Accounting for Success,” *Journal of Econometrics*, 161 (2011), 56-81.

Rey, Sergio J., and Luc Anselin, “PySAL: A Python library of spatial and analytical methods,” *Handbook of Applied Spatial Analysis*, (2010), 175-193.

Roos, Michael W. M., “How Important is Geography for Agglomeration?” *Journal of Economic Geography*, 5 (2005), 605-620.

Store, Ron, and Jyrki Kangas, “Integrating Spatial Multi-Criteria Evaluation and Expert Knowledge for GIS-Based Habitat Suitability Modeling,” *Landscape and Urban Planning*, 55 (2001), 79-93.

Townsend, Robert M., “Intermediation with Costly Bilateral Exchange,” *Review of Economic Studies*, 45 (1978), 417-425.

Townsend, Robert M., and Kenichi Ueda, “Financial Deepening, Inequality, and Growth: A Model-Based Quantitative Evaluation,” *Review of Economic Studies*, 73 (2006), 251-293.

Townsend, Robert M., and Kenichi Ueda, “Welfare Gains from Financial Liberalization,” *International Economic Review*, 51 (2010), 553-597.

Townsend, Robert M., *Financial Systems in Developing Economies: Growth, Inequality and Policy Evaluation in Thailand*, (New York, NY: Oxford University Press, 2010).

Townsend, Robert M., Anna Paulson, Sombat Sakuntasathien, Tae Jeong Lee, and Michael Binford, “Questionnaire Design and Data Collection for NICHD Grant Risk, Insurance and the Family and NSF grants,” Unpublished manuscript, University of Chicago, 1997.

Uchida, Hirotsugu, and Andrew Nelson, “Agglomeration Index: Towards a New Measure of Urban Concentration,” UNU-WIDER Working Paper No. WP2010/29, 2010.

Vanek, Jaroslav, “The Factor Proportions Theory: the *N*-Factor Case,” *Kyklos*, 21 (1968), 749–756.

World Bank. *World Development Report: Reshaping Economic Geography, 2009.* (Washington, D.C.: World Bank, 2009).

Zeller, Manfred, Aliou Diagne, and Chardles Mataya, “Market access by smallholder farmers in Malawi: Implications for technology adoption, agricultural productivity and crop income,” *Agricultural Economics*, 19 (1998), 219-229.

TABLE I
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
	-	(2.866)	(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
	-	(8.6823)	8.6823
Soil Fertility Index (Range: 1 to 9)	2.9472	2.6497	2.6497
	(1.7375)	(1.6956)	(1.6956)
Elevation (Meters)	217.527	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)	-	-
Tambon Distance to Major Highways (Kilometers)	6.2766	-	-
	(9.1575)	-	-
Tambon Distance to Bangkok (Kilometers)	381.3931	-	-
	(225.8363)	-	-
Tambon Distance to Nearest Major City (Kilometers)	84.4864	-	-
	(55.4940)	-	-

Notes:

1. Means with standard deviations in parentheses.

2. To estimate the 1986 Wealth Instrument we used a wealth index variable 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.

TABLE II:
Enterprise and Geography: Bivariate Regression Results
All regressions are bivariate, running the dependent variable in the column on each variable in the row, one at a time^.

At the National Level *At The Thai Amphoe District Level* *For all of Thailand* *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 SES Growth in Entrepreneurial Income	National 2005 DIW Factory Geographic Density (km ^ 2) (Tambon Districts)
	(1)	(2)	(3)	(4)	(5)
Distance to Major Highways (Kilometers)	-0.0303***	-9222	-11.53***	0.568	0.0109
	[0.00847]	[6938]	[2.581]	[0.848]	[0.0228]
Distance to Bangkok (Kilometers)	-0.00285***	1230**	-0.17	0.0129	-0.000934
	[0.000319]	[480.4]	[0.140]	[0.0267]	[0.00105]
Distance to 15 Largest Cities (Kilometers)	-0.00779***	3519*	-0.517	0.267**	0.0248*
	[0.00189]	[2090]	[0.672]	[0.131]	[0.0139]
Distance to Major Rivers (Kilometers)	-0.00574***	-4418**	-1.174	-0.0928	-0.0145**
	[0.00187]	[1717]	[1.055]	[0.171]	[0.00705]
Soil Fertility (Index, 0 to 9)	0.131***	36.93	6.072	0.0649*	-0.266
	[0.0498]	[474.7]	[16.62]	[0.0358]	[0.340]
Elevation (Meters)	-0.00211***	-3.25	-0.293	0.000985*	-0.00503***
	[0.000299]	[5.166]	[0.183]	[0.000595]	[0.00141]
Rainfall Variation (Centimeters)	0.00352	285.2*	2.272	-0.073	0.158***
	[0.00888]	[152.8]	[4.217]	[0.118]	[0.0398]
Observations	700	700	833	487	2588

At the Provincial Level *At the Village Level in Four Thai Provinces:* *Sisaket, Buriram, Lop Buri and Chachoengsao* *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*
	[0.00624]	[0.0253]
Travel-Time to Amphoe District Centers (Minutes)	-0.0493***	-0.023
	[0.00843]	[0.0347
Distance to Major Intersections (Kilometers)	-0.0757***	-0.0555**
	[0.00884]	[0.0280]
Distance to Major Rivers (Kilometers)	-0.132***	-0.192***
	[0.0164]	[0.0635]
Soil Fertility (Index, 0 to 9)	0.543***	0.508***
	[0.0549]	[0.165]
Elevation (Meters)	-0.0105***	-0.0139***
	[0.00177]	[0.00519]
Rainfall Variation (Centimeters)	0.175***	0.0421
	[0.0285]	[0.0627]
Observations	2987	1532

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

^ An example of how to interpret coefficients is the coefficient on *National 1986 CDD Percent In Enterprise* regressed on Distance to Major Highways, to be read as: for each additional kilometer of distance to major highways the percentage of households involved in enterprise activities falls by 0.03%.

Notes:
1. Thailand's Community Development Department (CDD) socio-economic data set is a census at the village level collected socio-economic data set is a census at the village level collected by the Rural Development Committee (RDC) within the Thai Ministry of the Interior.
2. CDD percent in enterprise defined as the percent of households in a village engaged in retail or cottage industries.
3. The Thai Socio-Economic Survey (SES) is a representative national survey of income and expenditures conducted at the household level by the Thai National Statistical Office (NSO).
4. SES entrepreneurial income is a distinct variable in the SES survey.
5. The DIW is a survey of Thai manufacturing companies conducted by the Thai Department of Industrial Works (DIW) under the Thai Ministry of Industry.
6. DIW factory geographic density is defined as the number of factories per k^2
7. Distance variables are geographic distance from Tambon or village centroid points.
8. Travel-time variables were calculated using a digital GIS road network digitized from high quality official Thai government road maps and satellite imagery.
9. Soil, elevation, river and rainfall data obtained from the Thai Meteorological Department and other Thai government sources.
10. Soil fertility index was constructed using data on soil cation exchange capacity and field capacity taken from national Thai soil maps, to construct a relative index of inherent soil fertility ranging from 1 to 9.
11. Rainfall variation was calculated as the standard deviation of total precipitation recorded from 1951-2000 using data from Thai meteorological stations supplied by the Thai Meteorological Department.

Table III:
Spatial Structural Model Parameter Values

SPATIAL OCCUPATIONAL CHOICE SIMULATION RESULTS

SPATIAL OCCUPATIONAL CHOICE SIMULATION ESTIMATED STRUCTURAL PARAMETERS:

	Occupational Choice Estimated Structural Parameters:	*All Parameters Allowed To Vary Freely Across Whole Sample*								
		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

	Occupational Choice: All Parameters Vary In All Bins *Increasing Geographic Distance*	*All Parameters Allowed To Vary Freely In Each Bin Across Space*								
		α	β	ω	η	ρ	σ	γ	g_γ	ξ
Robustness Check 4:	*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

	Estimation of the Spatially Varying *m(d)* Costs Parameter By Geographic Distance	*Increasing 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

	Estimation of the Spatially Varying *q(d)* Cost Parameter by Geographic Distance	*Increasing 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

Note: All nine parameters of production (α, β, ρ, σ, ξ), savings rate(ω), initial 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 last year of the CDD data and those predicted from the model, summing over villages in each province. Then we compare the model generated output at those estimated parameters to the data, contrasting the actual data and the simulated data. Finally, we conduct numerous robustness checks for parameters and assumptions of the model.

Table IV:
Occupational Choice 1996 Simulation Residuals Regressed Onto Agglomeration Proxies

	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)
Travel-Time to Major Roads	-0.0505**	-0.0178						
	(0.025)	(0.025)						
Travel-Time to Major Intersections			-0.00000138***	-0.000000168				
			(0.00000046)	(0.00000045)				
Travel-Time to District Centers					-0.0643**	-0.0252		
					(0.026)	(0.026)		
1986 Wealth							0.000344**	0.000181
							(0.00016)	(0.00016)
1986 Education							0.0213	0.00716
							(0.025)	(0.025)
Constant	0.00789	0.00216	0.0172**	0.00148	0.0141**	0.00490	-0.0244***	-0.0142
	(0.0056)	(0.0055)	(0.0069)	(0.0069)	(0.0070)	(0.0069)	(0.0088)	(0.0088)
Observations	731	731	731	731	731	731	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

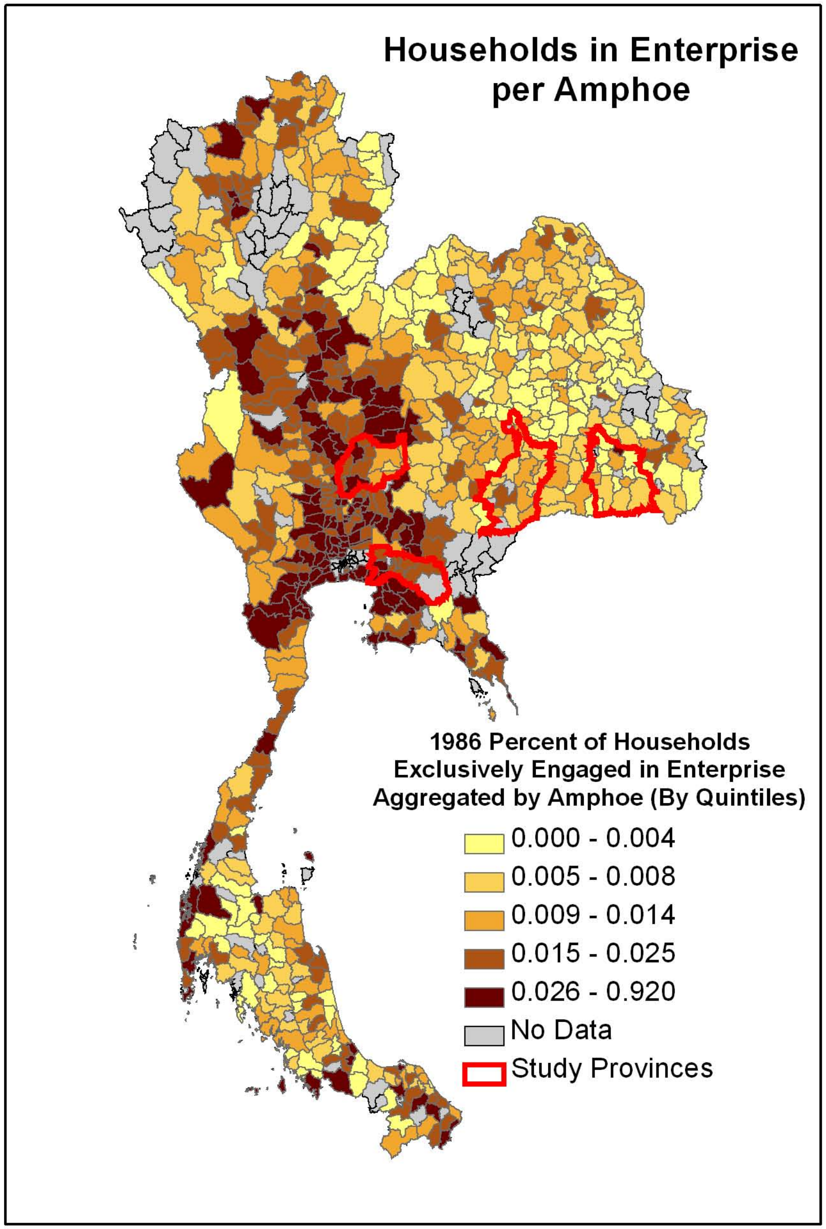
Note: In the table above, when regressing estimated residuals from non-spatial structural simulation models on variables that capture village travel-time accessibility to markets through road networks, as well as on base-year wealth, the explanatory variables appear to be statistically significant. However, when regressing estimated residuals from spatially-specified 3-bin structural simulation models the same explanatory variables do not appear to be statistically significant. This evidence supports the hypothesis that the spatially-specified structural simulation models significantly explain occupational decisions.

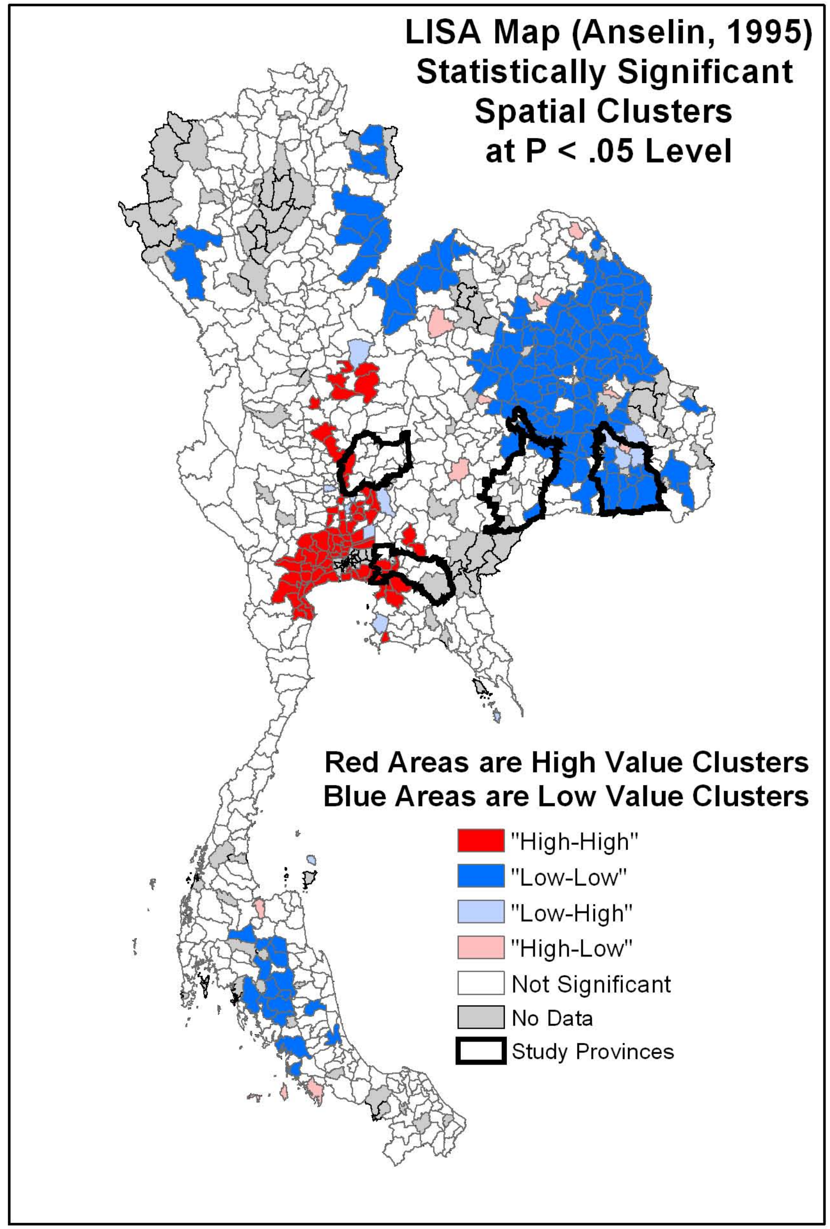
Table V:
Financial Deepening 1996 Simulation Residuals Regressed Onto Market and Infrastructure Access Proxies

	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 (Minutes	-0.297***			-0.000584			-0.00205**			0.000491		
	(0.0776)			(0.00180)			(0.00104)			(0.00183)		
Distance to Major Intersections (Km)		-0.759***			-0.0120***			-0.00725***			-0.0138***	
		(0.0893)			(0.00209)			(0.00122)			(0.00214)	
Travel-Time to District Centers (Minutes)			-0.533***			-0.00460**			-0.00229**			-0.00213
			(0.0839)			(0.00196)			(0.00113)			(0.00199)
1996 Wealth Index	-0.508***	-0.577***	-0.536***	0.00382***	0.00263***	0.00353***	0.00282***	0.00222***	0.00273***	0.00221***	0.000981	0.00207**
	(0.0343)	(0.0347)	(0.0344)	(0.000794)	(0.000811)	(0.000802)	(0.000469)	(0.000476)	(0.000473)	(0.000826)	(0.000834)	(0.000833)
1996 Educational Attainment	36.26***	32.37***	34.93***	0.109	0.0373	0.0921	-0.00669	-0.0459	-0.0120	0.0864	-0.0132	0.0696
	(5.210)	(5.134)	(5.170)	(0.121)	(0.121)	(0.121)	(0.0739)	(0.0734)	(0.0741)	(0.130)	(0.129)	(0.130)
R-Squared	0.212	0.243	0.226	0.036	0.067	0.043	0.080	0.100	0.080	0.065	0.092	0.066

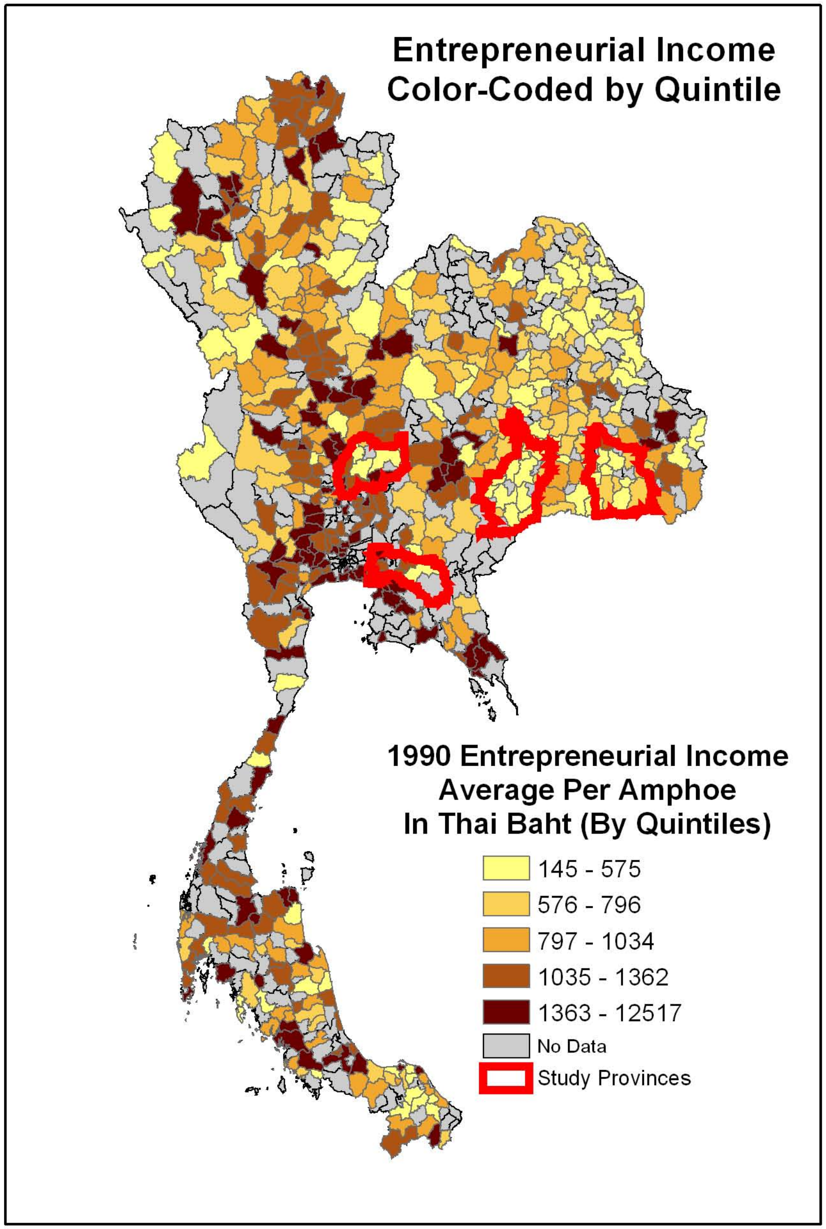
Notes :
1. All regressions include Provincial fixed effects dummy variables
2. Standard errors in parentheses
3. *** p<0.01, ** p<0.05, * p<0.1
4. Statistical significance in geographic variables implies that it is a source of over prediction on the model for the dependant variable.

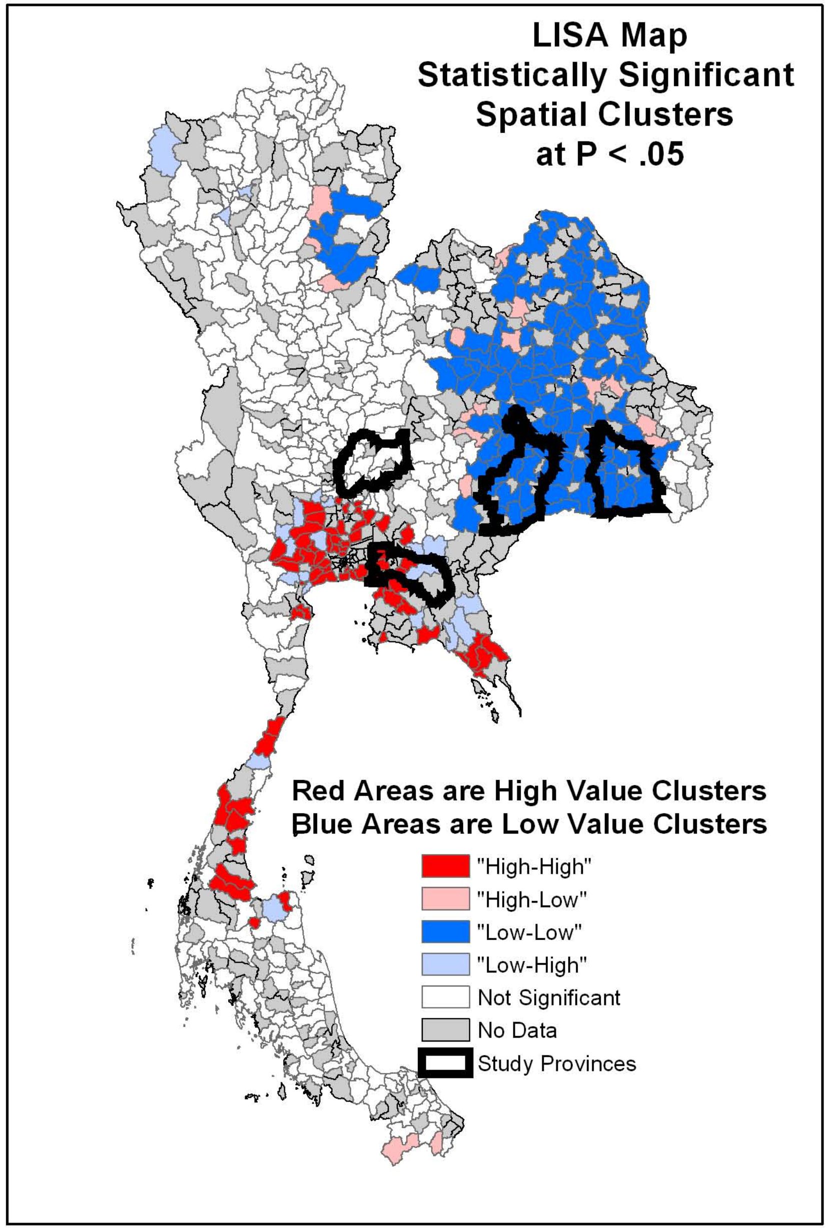
Enterprise at the National Level in 1986: CDD Survey Data





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





Variation in Thai Rural Wealth, 1988
Index of Wealth Based on Key Assets Aggregated to Thai Tambon Level

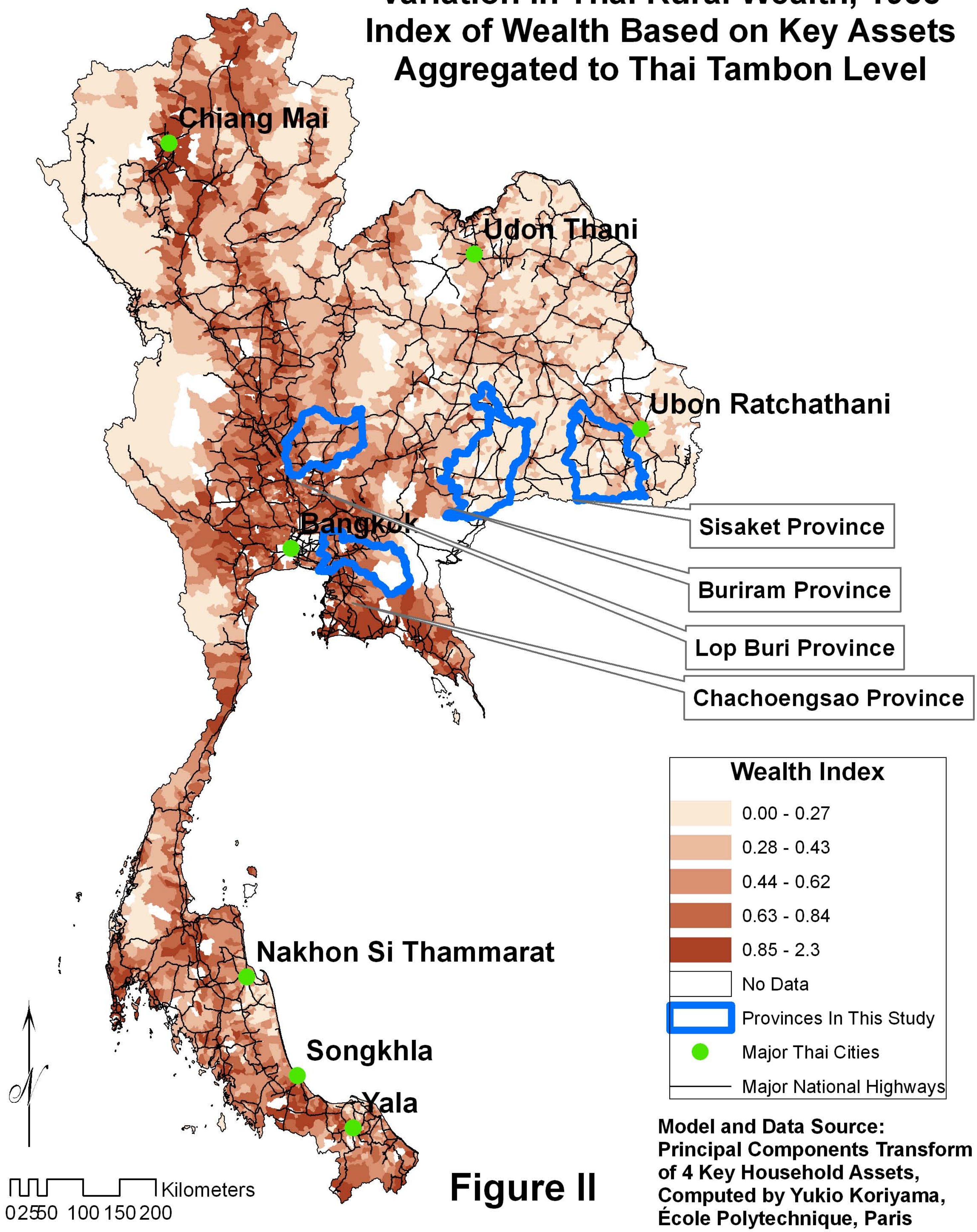


Figure II

Model and Data Source: Principal Components Transform of 4 Key Household Assets, Computed by Yukio Koriyama, École Polytechnique, Paris

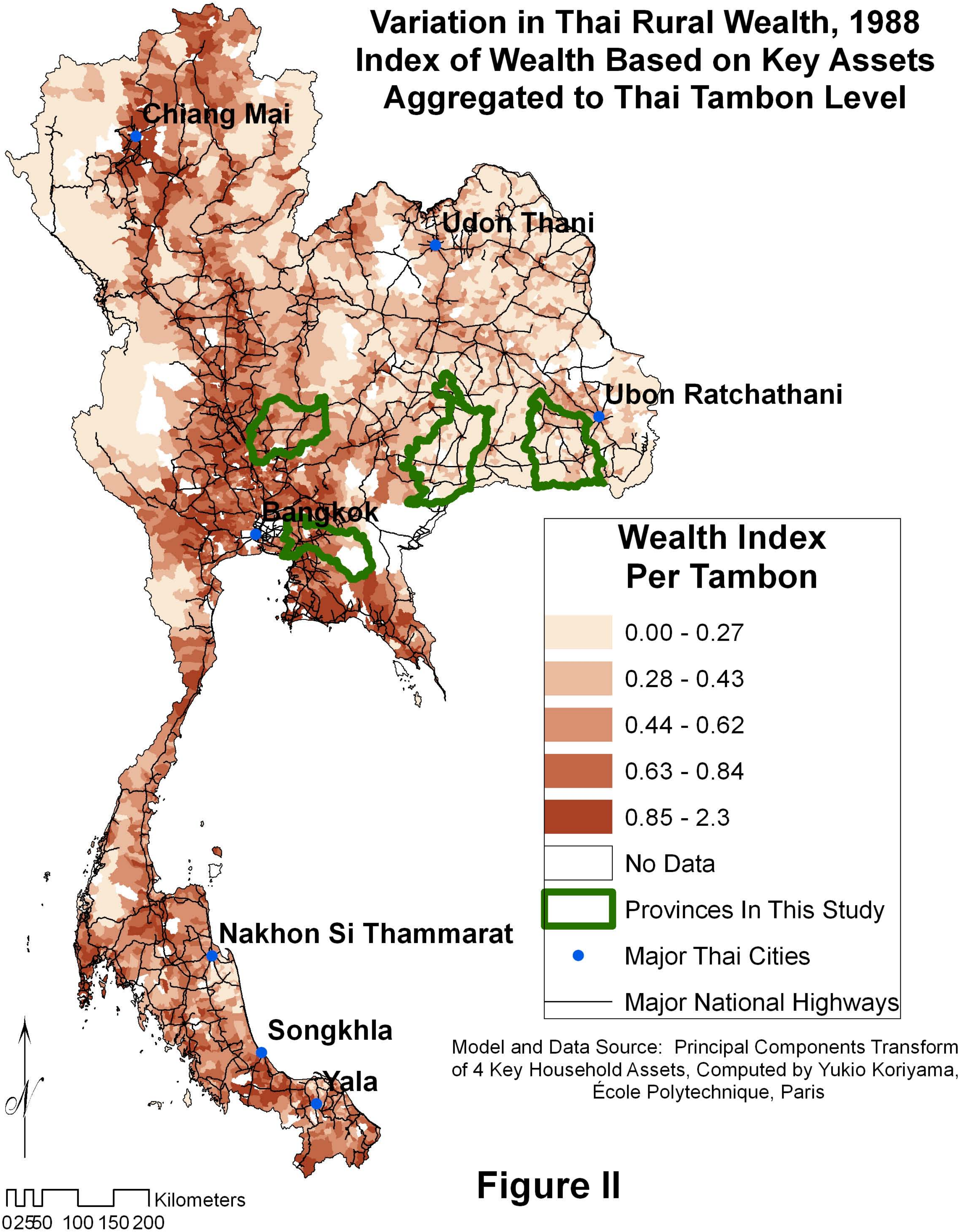
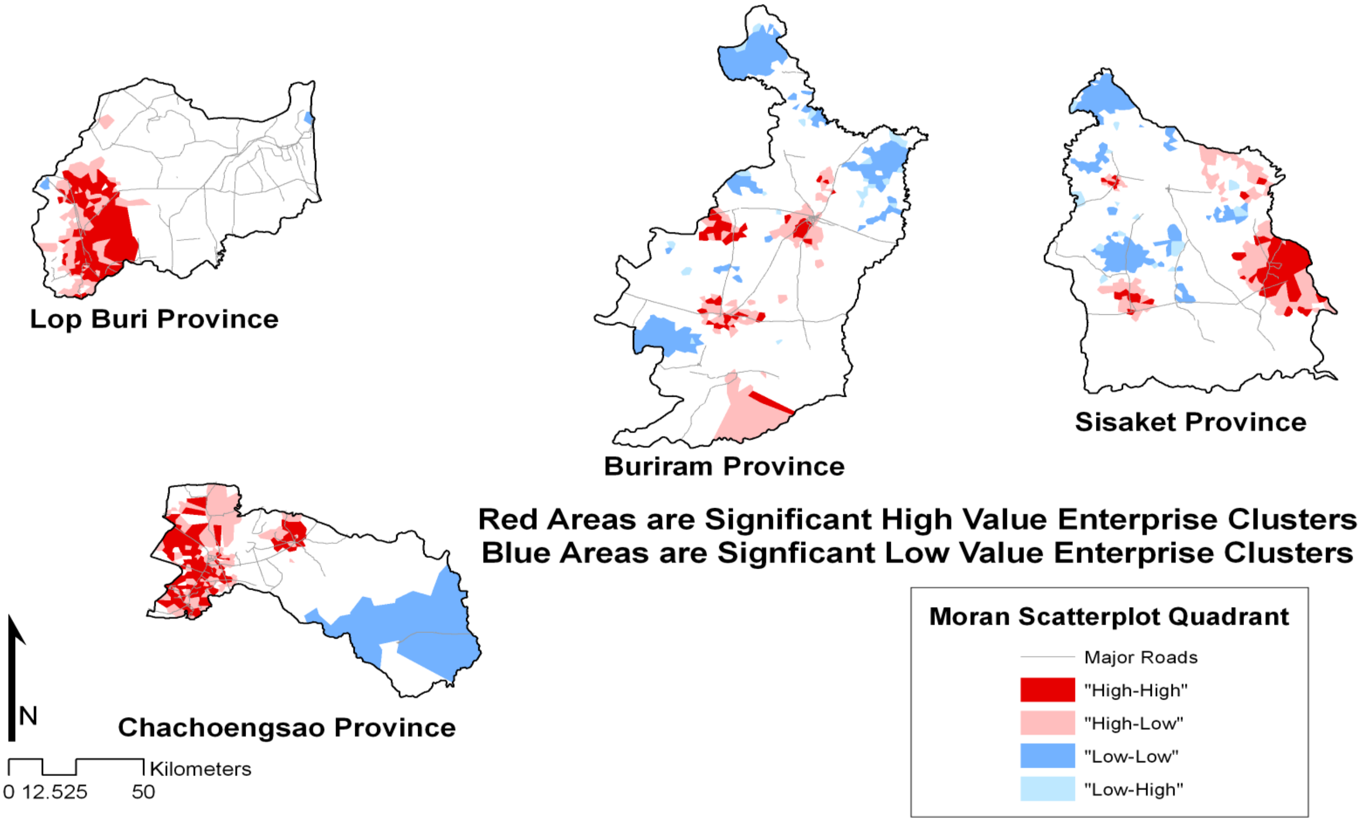

Variation in Thai Rural Wealth, 1988
Index of Wealth Based on Key Assets
Aggregated to Thai Tambon Level
Chiang Mai
Udon Thani
Ubon Ratchathani
Bangkok
Nakhon Si Thammarat
Songkhla
Yala
Wealth Index
Per Tambon
0.00 - 0.27
0.28 - 0.43
0.44 - 0.62
0.63 - 0.84
0.85 - 2.3
No Data
Provinces In This Study
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
N
Kilometers
0 25 50 100 150 200

Figure II

Enterprise Concentration in 1986

Univariate LISA Map, Values Significant at P < .05 ,
Distance Weights 10 Kilometer Binary



Wealth Concentration in 1986

Univariate LISA Map, Values Significant at P < .05 ,
Distance Weights 10 Kilometer Binary

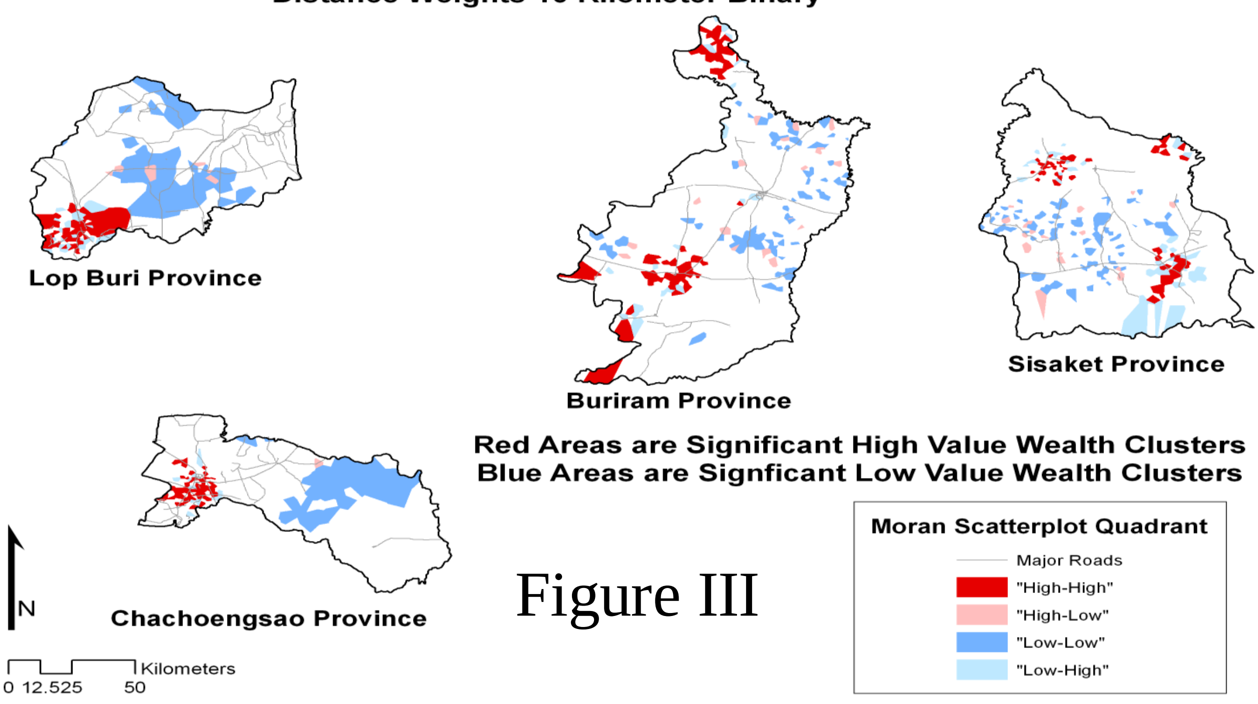
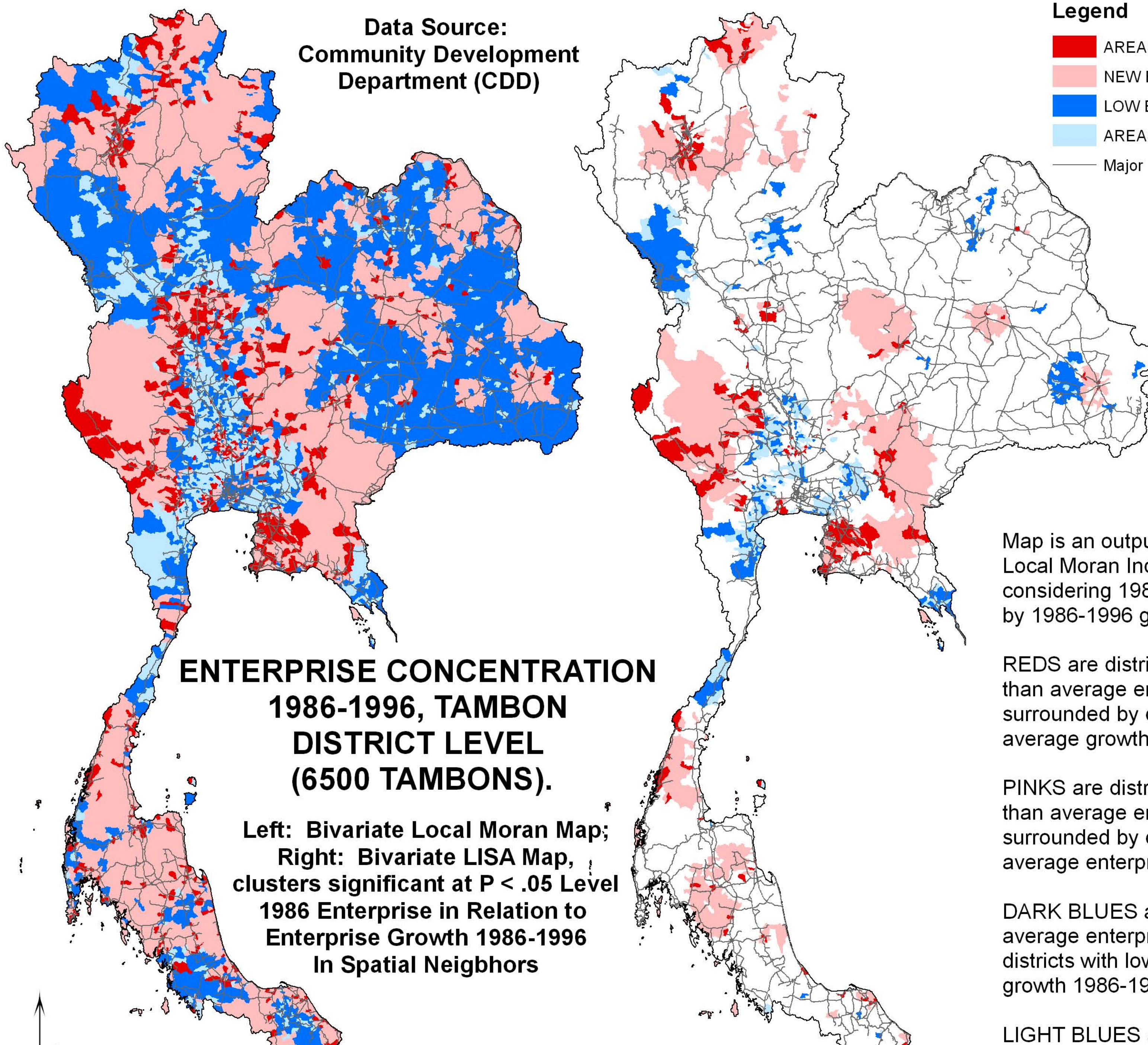


Figure III



Map is an output from a bivariate
Local Moran Index for 6500 districts,
considering 1986 fraction in enterprise surrounded
by 1986-1996 growth in fraction in enterprise

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

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

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

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

Figure IV

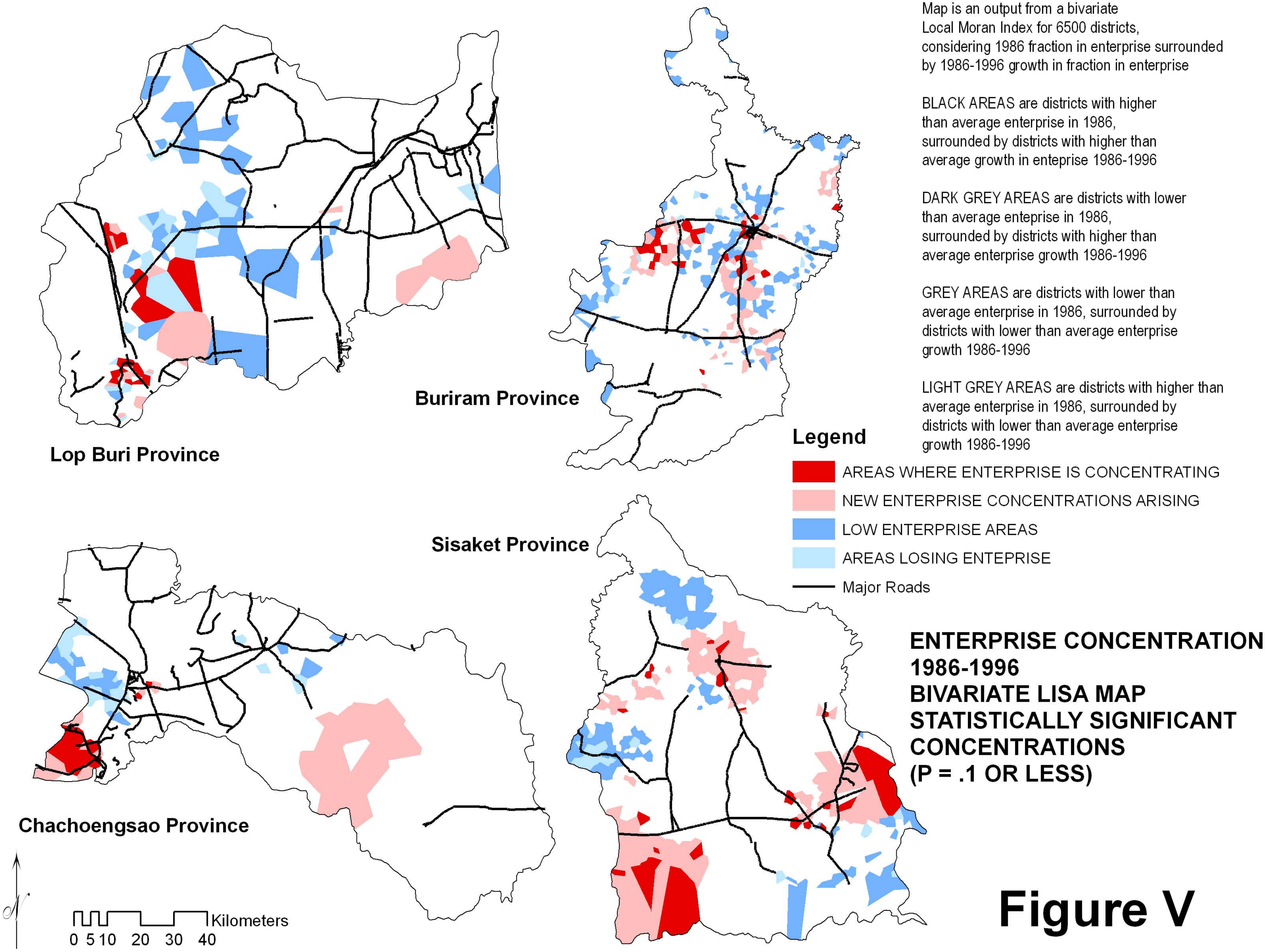
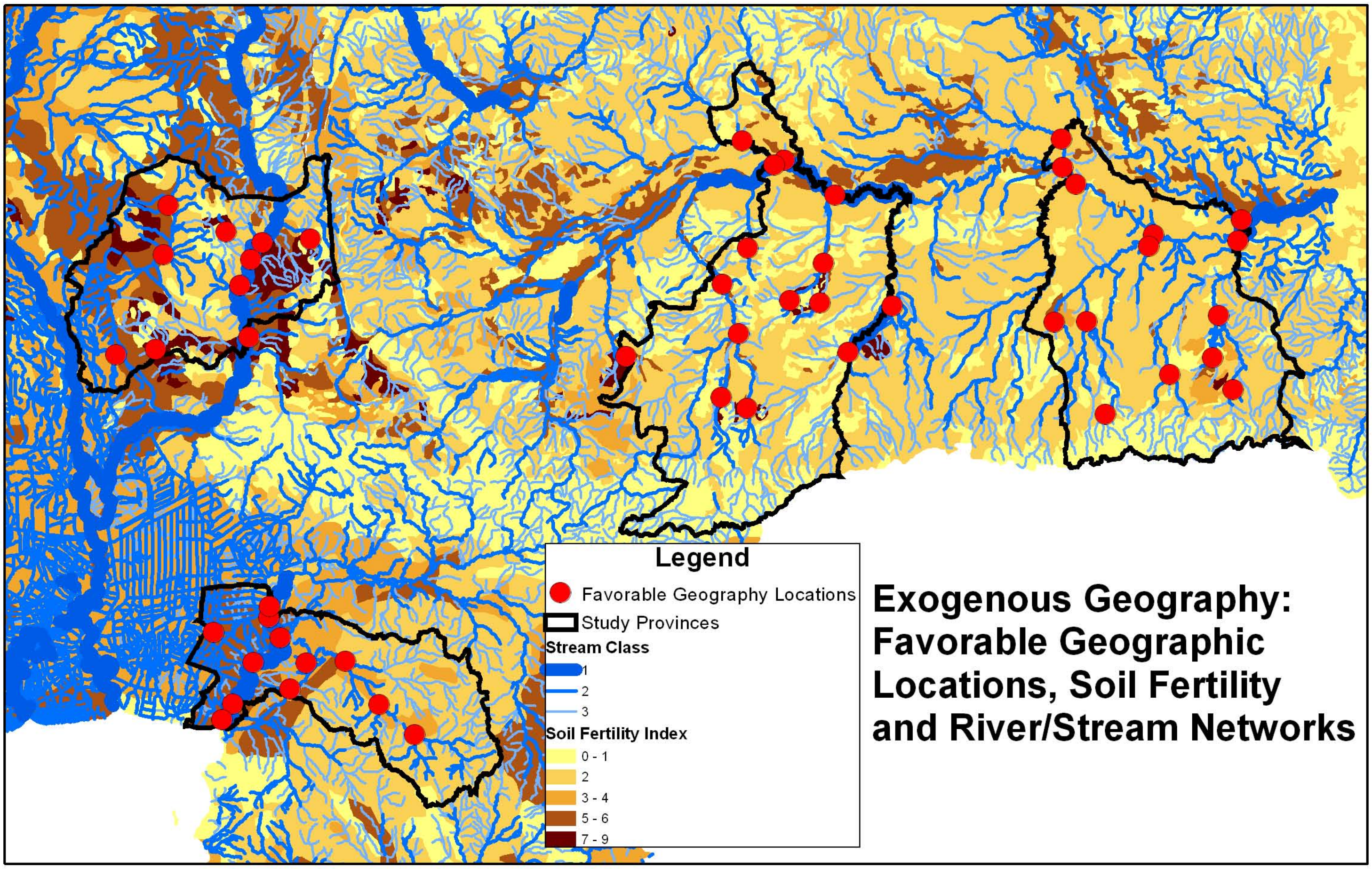
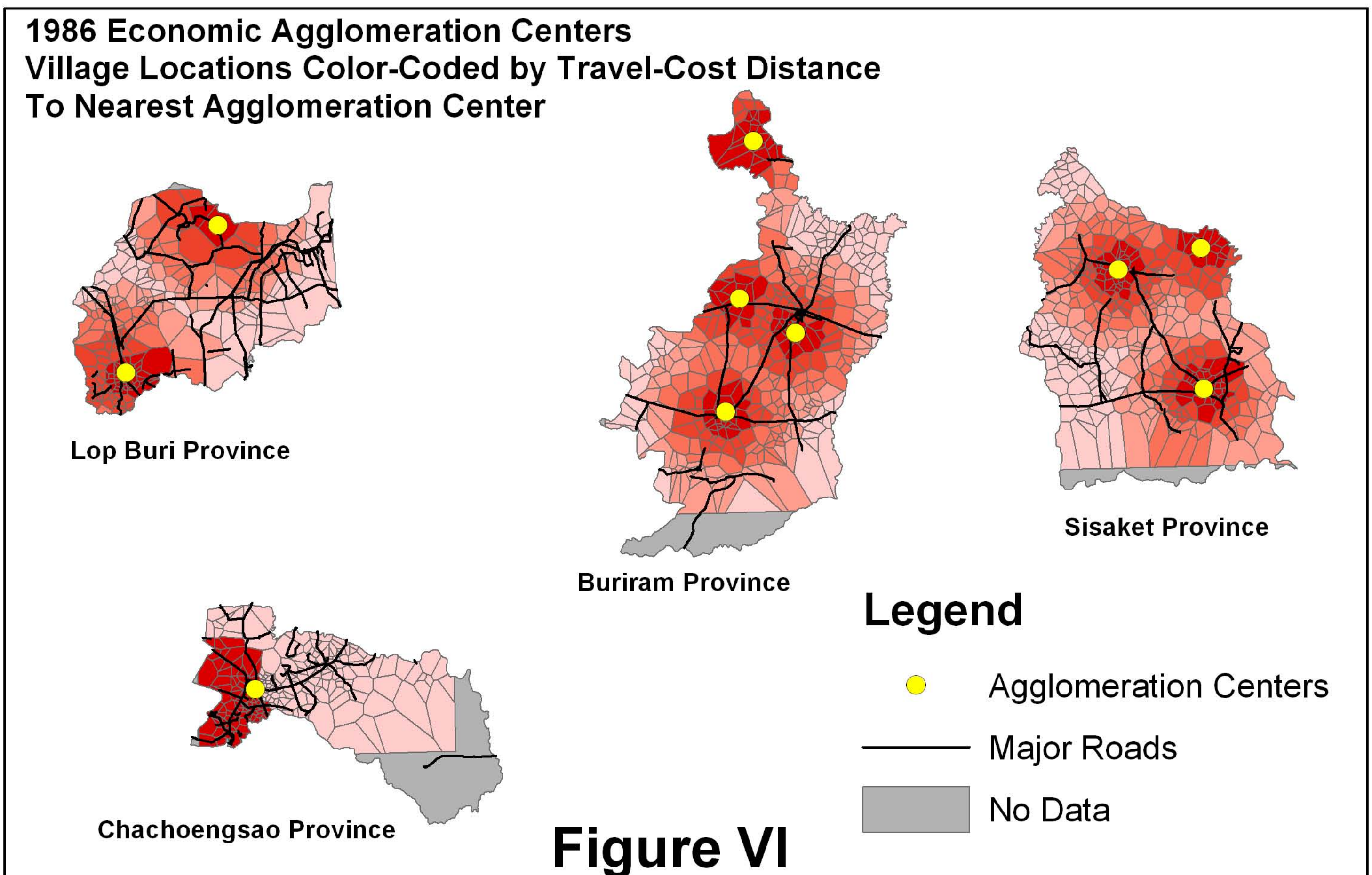

Map is an output from a bivariate Local Moran Index for 6500 districts, considering 1986 fraction in enterprise surrounded by 1986-1996 growth in fraction in enterprise
BLACK AREAS are districts with higher than average enterprise in 1986, surrounded by districts with higher than average growth in enteprise 1986-1996
DARK GREY AREAS are districts with lower than average enteprise in 1986, surrounded by districts with higher than average enterprise growth 1986-1996
GREY AREAS are districts with lower than average enterprise in 1986, surrounded by districts with lower than average enterprise growth 1986-1996
LIGHT GREY AREAS are districts with higher than average enterprise in 1986, surrounded by districts with lower than average enterprise growth 1986-1996
Lop Buri Province
Buriram Province
Sisaket Province
Chachoengsao Province
Legend
AREAS WHERE ENTERPRISE IS CONCENTRATING
NEW ENTERPRISE CONCENTRATIONS ARISING
LOW ENTERPRISE AREAS
AREAS LOSING ENTEPRISE
Major Roads
ENTERPRISE CONCENTRATION 1986-1996
BIVARIATE LISA MAP
STATISTICALLY SIGNIFICANT CONCENTRATIONS
(P = .1 OR LESS)
N
0 5 10 20 30 40 Kilometers

Figure V



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
Village Locations Color-Coded by Travel-Cost Distance
To Nearest Agglomeration Center
Lop Buri Province
Buriram Province
Sisaket Province
Chachoengsao Province
Legend
Agglomeration Centers
Major Roads
No Data

Figure VI

Occupational Choice Model -
Choices in Cross-Section at a Given Date *t*:

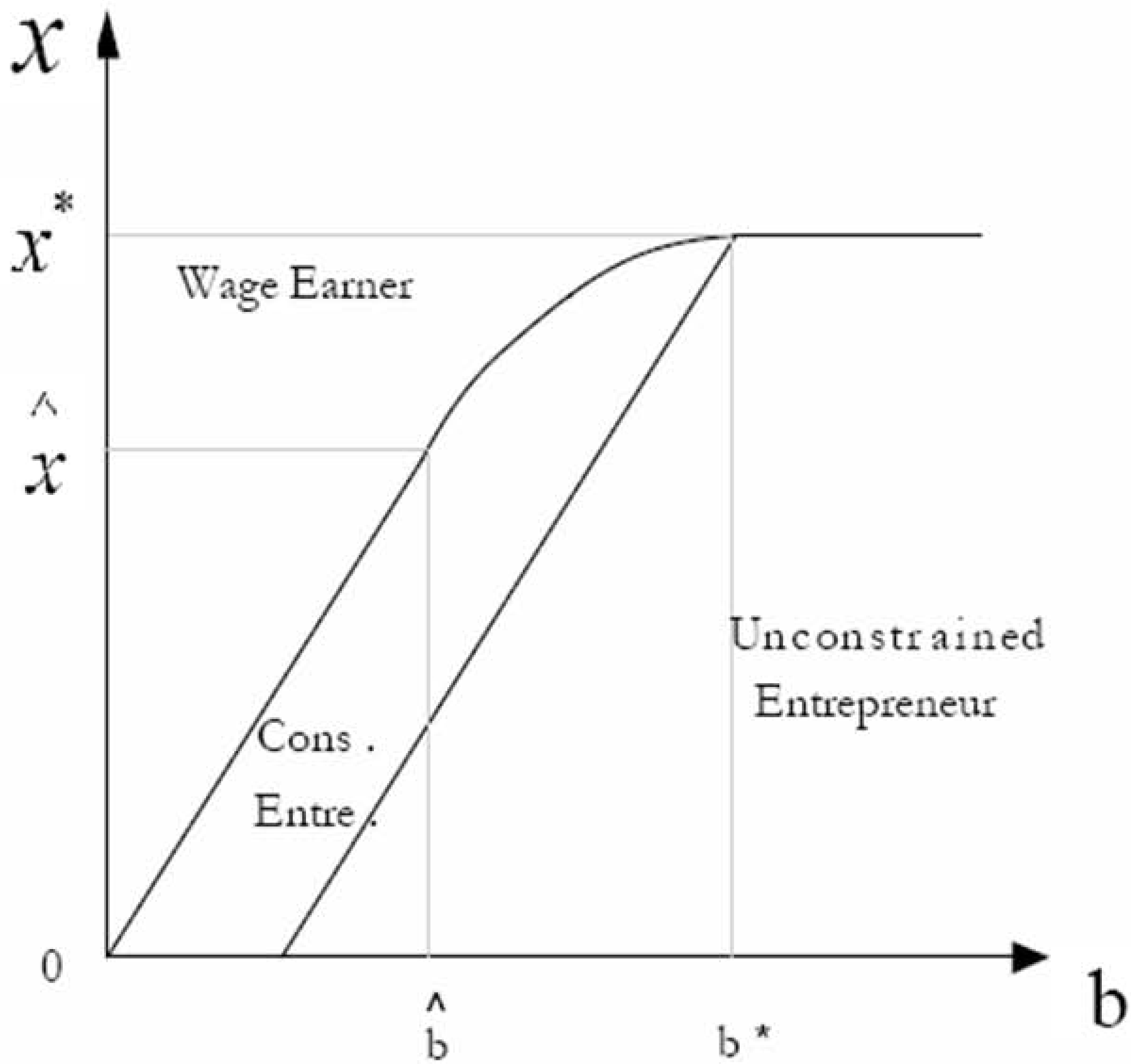
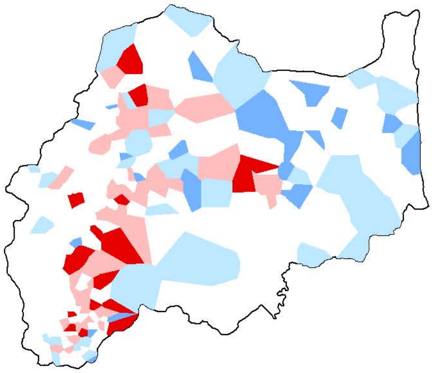
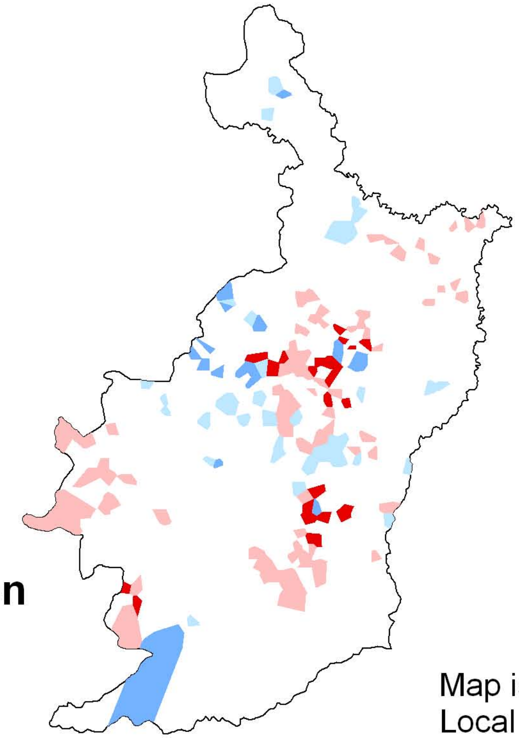


Figure VII

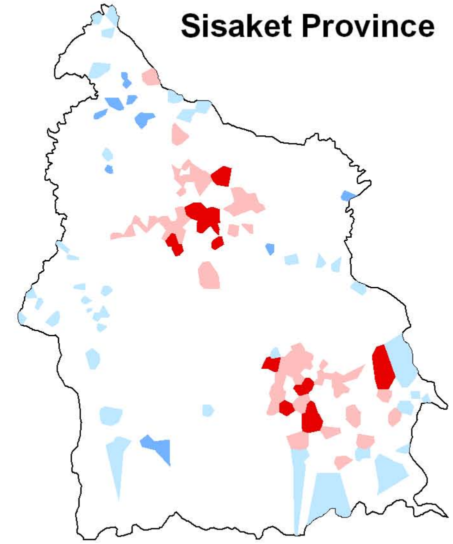
Lop Buri Province



Buriram Province

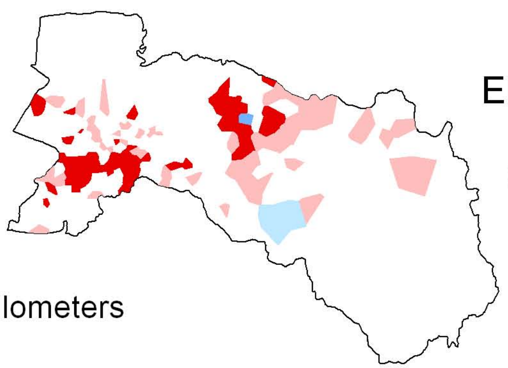


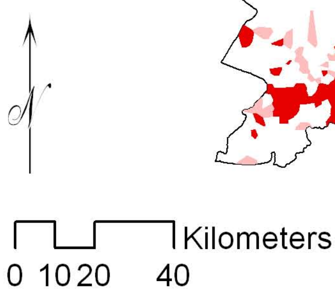
Sisaket Province



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





Chachoengsao Province

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

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

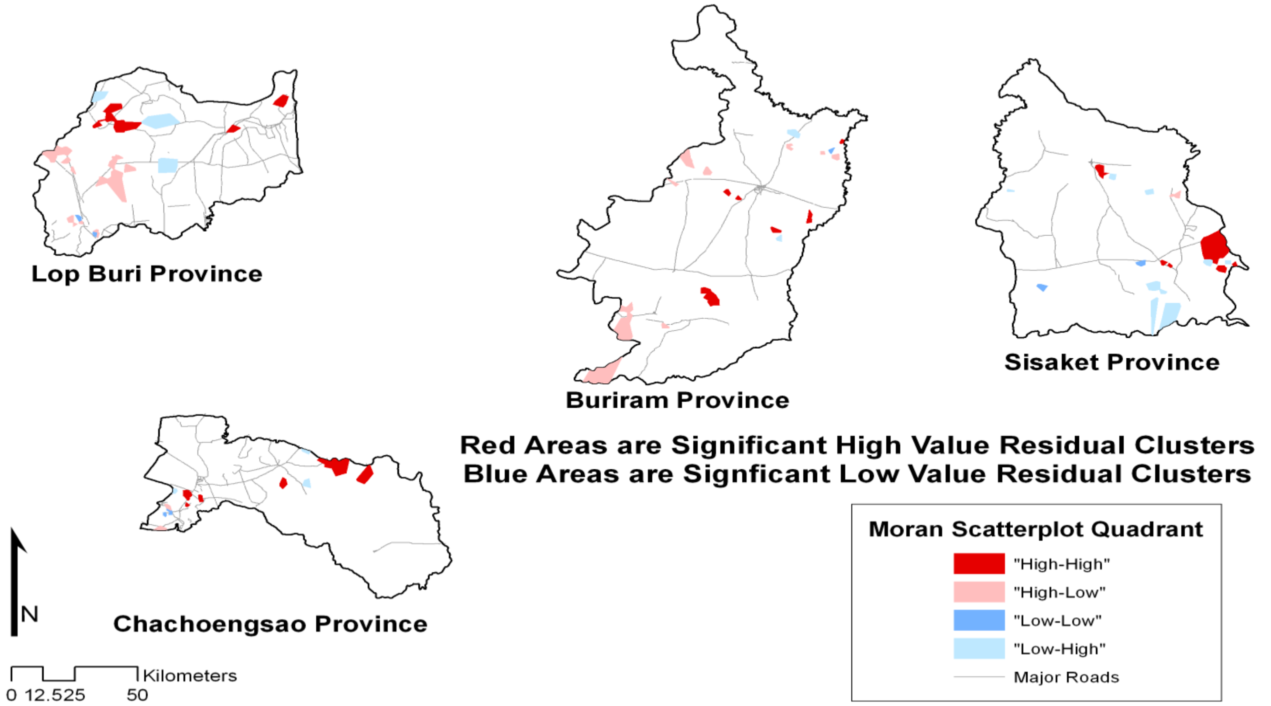
Legend

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

Figure VIII

Occupational Choice Structural Simulation

Spatial Specification (3 Bin) Simulation Residuals: M Parameter Varies With Space; LISA Map at P = .05 Cut-Off Value; Distance Weights Are 10 Nearest Neighbors.



Occupational Choice Structural Simulation

Non-Spatial Specification Residuals: Global M Parameter; LISA Map at P = .05 Cut-Off Value; Distance Weights Are 10 Nearest Neighbors

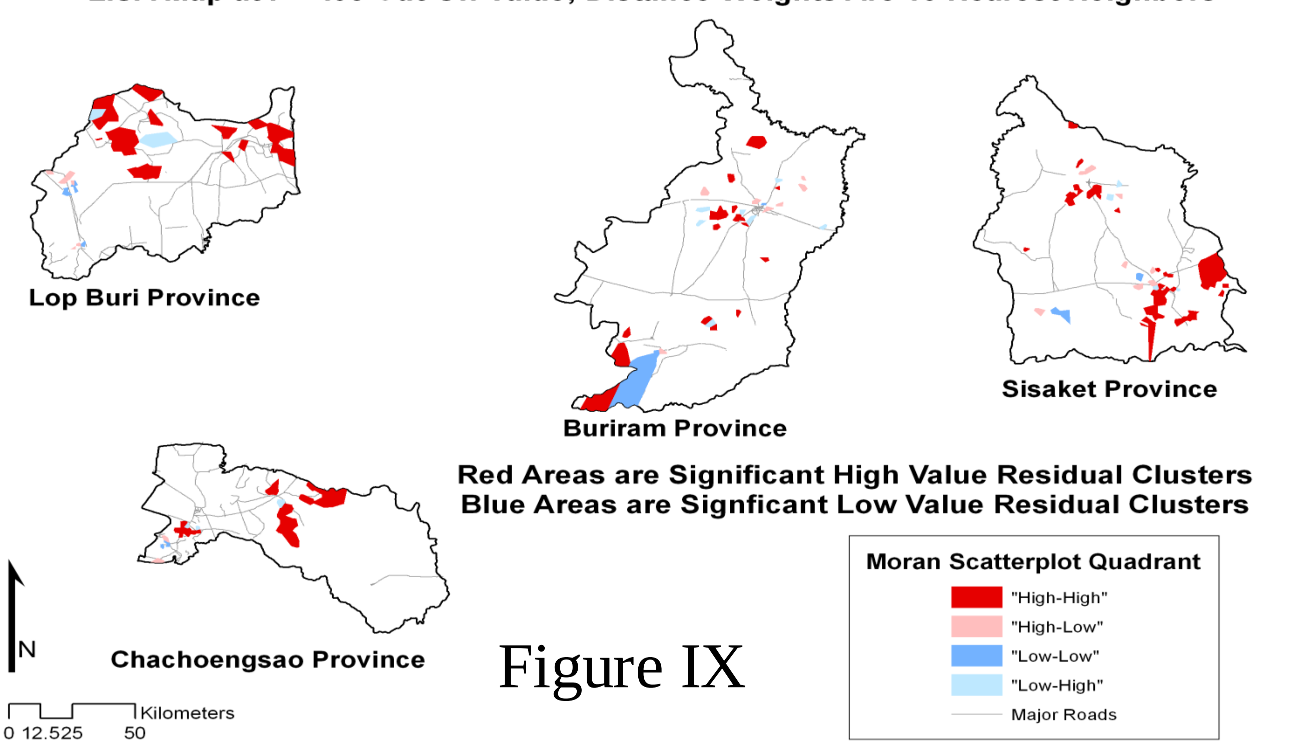
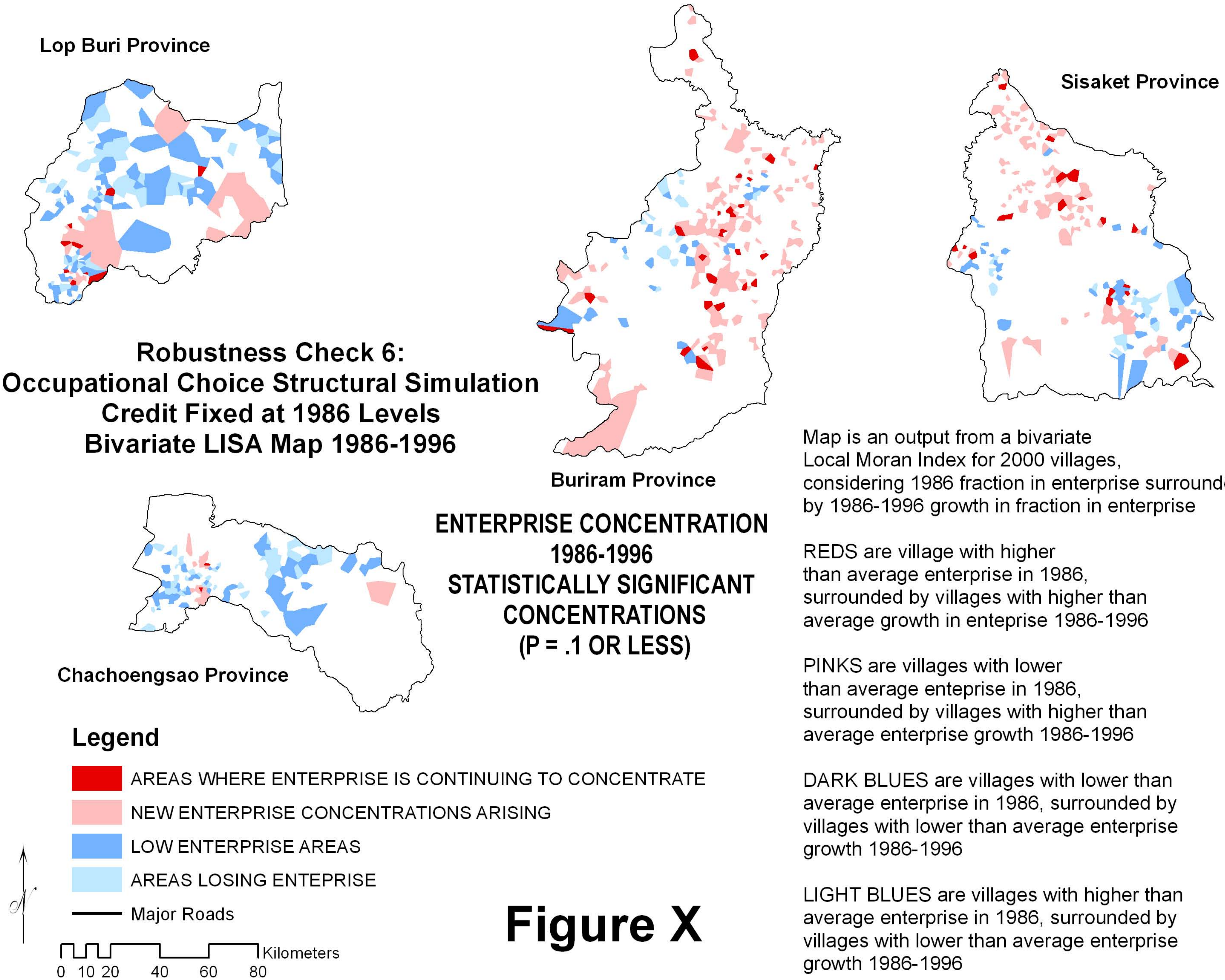


Figure IX



Robustness Check 6:
Occupational Choice Structural Simulation
Credit Fixed at 1986 Levels
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

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

Figure X

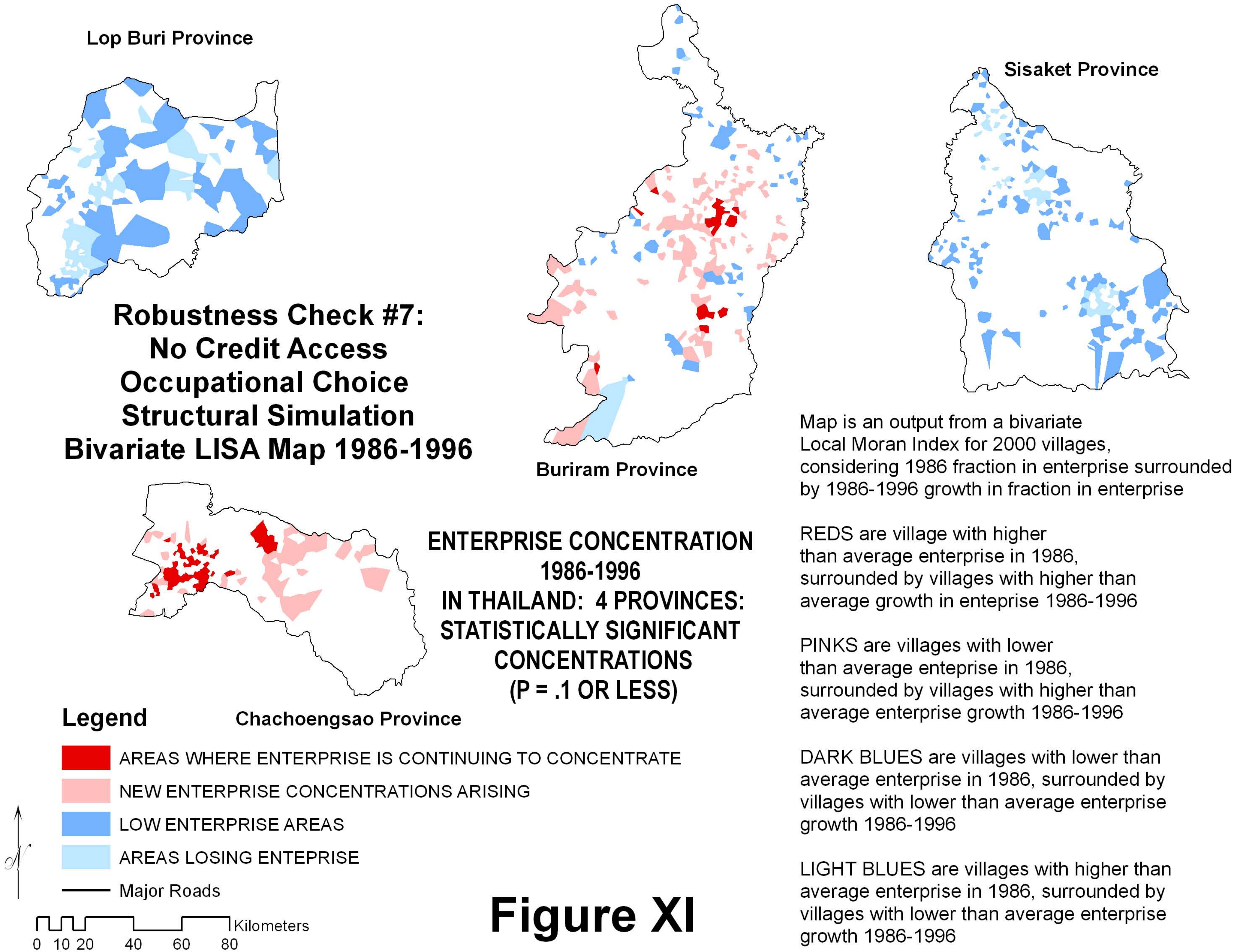

Lop Buri Province
Sisaket Province
Buriram Province
Chachoengsao Province
Robustness Check #7:
No Credit Access
Occupational Choice
Structural Simulation
Bivariate LISA Map 1986-1996
ENTERPRISE CONCENTRATION
1986-1996
IN THAILAND: 4 PROVINCES:
STATISTICALLY SIGNIFICANT
CONCENTRATIONS
(P = .1 OR LESS)
Legend
AREAS WHERE ENTERPRISE IS CONTINUING TO CONCENTRATE
NEW ENTERPRISE CONCENTRATIONS ARISING
LOW ENTERPRISE AREAS
AREAS LOSING ENTEPRISE
Major Roads
Kilometers
0 10 20 40 60 80
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

Figure XI

Financial Deepening Simulation
1996 Credit Access Spatial Residuals

Residuals are Simulated Minus Actual
Window-Average Smoothed Values, Using 10 Nearest Neighbors

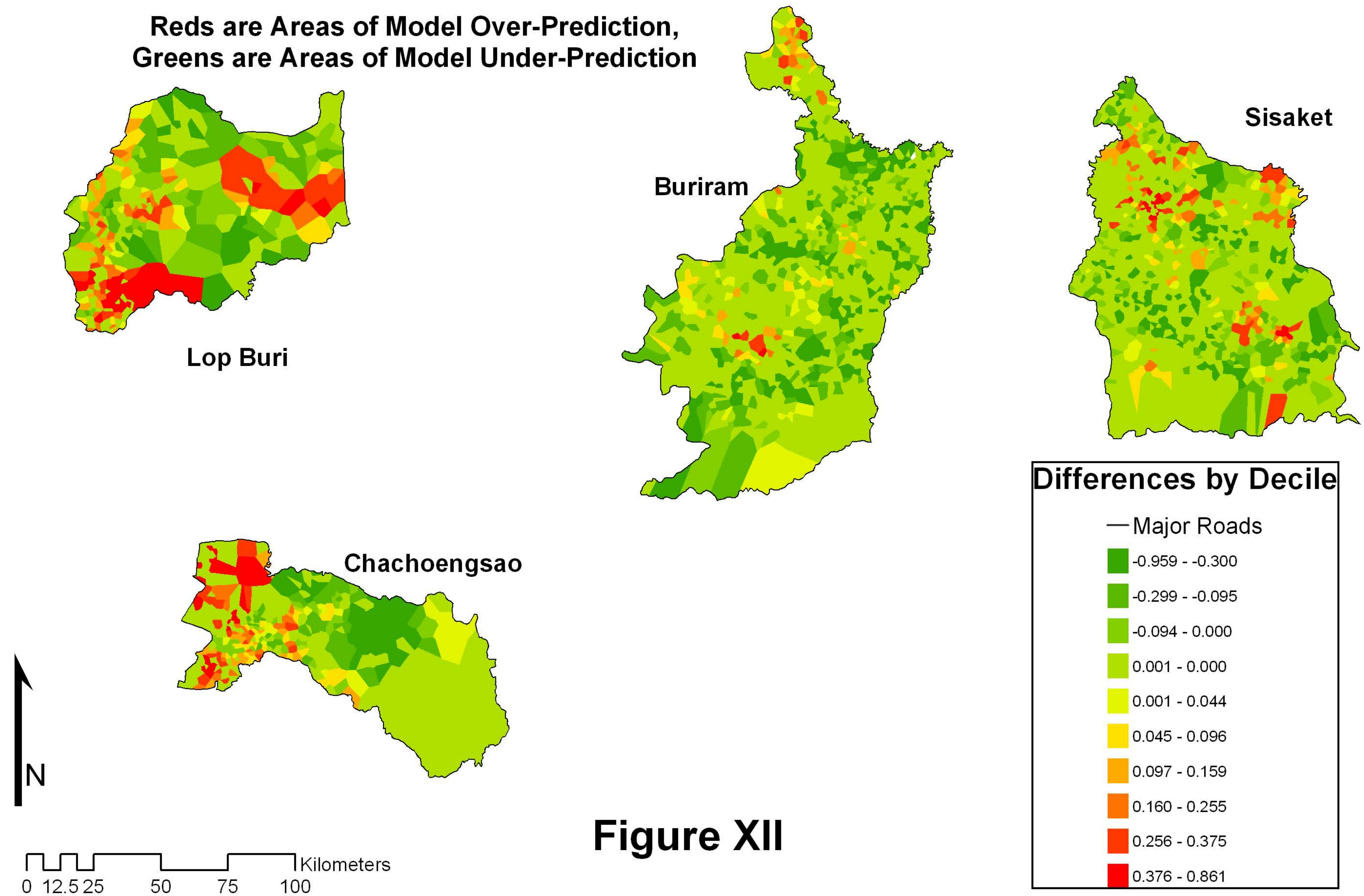


Figure XII

Village, Roads, Major Intersections and Amphoe District Center Locations

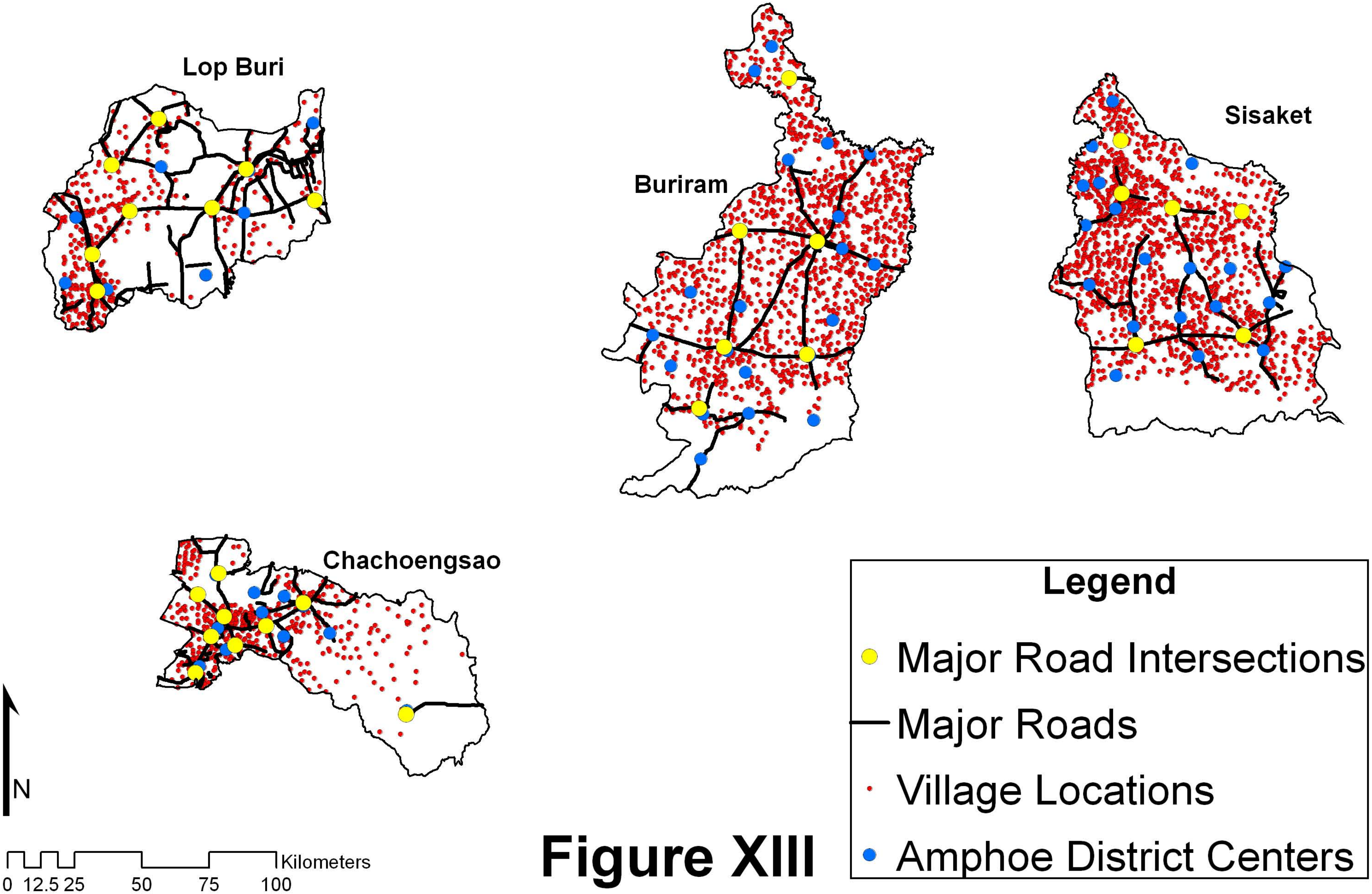


Figure XIII