A Hierarchical Regional Space Model for Contemporary China

- CONSTRUCTING AN URBAN-RURAL CONTINUUM

Wei Wang
Department of Statistics
University of California, Davis

G. William Skinner
Department of Anthropology
University of California, Davis

Paper prepared for the Geoinformatics '99 Conference
China Data Center, University of Michigan, Ann Arbor
20 June 1999

The objective of the second of our five analytical procedures is to classify county-level administrative units into finely graded classes along an empirical, spatially defined continuum from highly urban to overwhelmingly rural. Our urban-rural continuum index, or URC, is based on *structural* data describing the cities and towns in each county. As we describe in this paper, these structural features are captured in two indices, the Settlement Composition Index and the Combined Centrality Scores. We rely on *substantive* socioeconomic variables, summarized in a separate measure called the substantive urban index (SUI), only as a guide in integrating the two structural components of the URC.

Before detailing how we construct this urban-rural continuum, we should explain that such a continuum offers advantages over the approaches seen in previous studies. We reject as arbitrary and misleading the usual practice of dichotomizing Chinese socioeconomic data as urban or rural. In some studies, this involves comparing households with nonagricultural *hukou* status with agricultural households; in other studies, it means comparing towns of a specified size and larger with all other settlements. Many reports based on Chinese census data compare "cities" with "towns" with "counties" or "rural." This trichotomy is poorly specified. "Cities" in fact refers to municipalities, administrative units that may include sizable rural populations and in some cases whole *qu* (districts) that include no part of the central city. "Towns" actually refers to urban townships, which include villages as well as the central town. The third category, whether labeled "counties" or "rural," in fact refers to *xiang* (rural townships), which include market towns as well as villages. More defensible approaches develop an urbanization index by estimating the percentage of a region's population classified as urban, or by estimating the percentage of a region's land area that displays urban land use patterns. Here we adopt a more sophisticated approach that uses structural variables to place counties on an urban-rural continuum.

In partial compensation for arbitrary differences in the administrative geography of municipalities, we perform these procedures on the municipal compatible (MC) county-level units we have constructed, as described above. Further analysis of county-level units refers to these MC units. Noting that Tibetan society and its economy

is radically different from the remainder of areas under the jurisdiction of the People's Republic of China, we have excluded the Tibetan Plateau from this analysis. Indexes described in this paper were in all cases standardized for China minus that region.

Settlement Composition Index

The construction of the Urban-Rural Continuum index draws on two components, both categorical variables. The first of these, the Settlement Composition Index, is calculated from the residence data for county-level MC units in ChinaA.

Through manipulations of the census categories, we derived adjusted estimates of the number of residents in a) the central cities of muncipalities, 2) *zhen* towns, 3) villages in *zhen* qua townships, and 4) *xiang* (rural townships). In addition, interacting with the GIS maps, we discovered a fifth category of settlements that is not identified in the census reports: namely, the administrative centers of periurban *qu* (i.e., of urban districts that contain no part of the municipality's central city). Some of these towns have the status of *zhen* and hence are included in the census tables. But many others do not have such status and hence are invisible in the census; these towns form a fifth category of settlement that we designate "*qu* centers." The number of residents in *qu* centers can be inferred by determining the "excess" urban population in a given periurban *qu*. Thus, we were able to calculate the proportion of the population in each MC unit residing in each of five settlement types, and from these data construct a settlement composition index.

All units were classified according to the level of each unit's highest-order central place into one of six categories, for each of which we prepared a listing of units in descending order of index values. (We must postpone description of the final maneuver involving the settlement composition index until after introducing yet another index that plays a key role in the construction of the SUI.)

Combined Centrality Scores

The second component of the URC is derived from the central place analysis as recorded in ChinaT. Accordingly, we imported into the ChinaA county-level data file a number of variables from the ChinaT central place data file. Two of these variables describe the structure of urbanization and are relevant here. The first is the configuration by LUH (level in the urban hierarchy) of all central places within each county—for example, how many greater cities, local cities, central towns, and so forth. The second relevant variable specifies the LUH of the highest-level central place in the county. (Incidentally, in over forty cases the highest-level central place, according to our central place analysis, is *not* the seat of county government—a finding that draws attention to the differences between administrative and economic geographies.) The first of these variables takes the form of a string of digits, each digit standing for a central place at that level in the urban hierarchy. These were arrayed in logical order of centrality in lists controlled by the urbanization index. Figure 3.1 shows the listing for units whose urbanization index fell in the range 40-56%. Here we see the number of county-level units with unique configurations of central places at various levels. We then converted these strings of digits to numerical values indexing the combined centrality of all central places in the top seven levels. The value of a greater city (represented by the digit 5) was set at 100 and a multiplier of 3.33 was used between levels, so a regional city (digit 4) was valued 333, and a local city (digit 6) was valued 30, and so on. The values of all central places in a given unit were summed to yield the combined centrality score. We then grouped together units with exactly the same value and listed out these groups in descending order showing frequencies; this was done separately for each of 11 urbanization classes.

Substantive Urban Index

The Urban-Rural Continuum index (URC) is generated through cross-tabulations and interdigitations of the two components just described. While the settlement composition index and the combined centrality scores are both strictly structural variables, we need an index with substantive "urbanness" content to serve as a guide in fixing cutpoints in the continuous structural components and in interdigitating

Figure 3.1 Level and Configuration of Central Places of county-level units whose urbanization Index is 40-56%.

(Raw data imported into ChinaA from ChinaT)

4 Regional city

5 Greater city

6 Local city

7 Central town

A901	_FREQ_
	429
46	1
47	1
55666	1
5666666666777	1
566666677777777	1
5666667	1
566666	1
5666677	1
566667	1
56666	1
566677777777	1
566677777	2
5666777	2
566677	2
56667	3
56677777777777	1
56677777	3
5667777	3
566777	6
56677	1
5667	3
566	7
5677777777	1
567777777	1
56777777	3
5677777	7
567777	2
56777	10
5677	11
567	12
56	19
57777777777777	1
5777777777	2
57777777	3
5777777	1
577777	5
57777	14
5777	5
577	25
57	30
5	37
666677	3
6666	1
666777777	1
666777	1
66677	1
666	3
6677777	1
667777	2
66777	2
6677	4
667	12
66	20
677777777	1
6777777	1
677777	3
67777	6
6777	7
677	18
67	30
6	73
7777777777	1
7	5

logically ordered categories of classes of county-level units. To this end, we identified through exploratory analysis a set of variables pertaining to higher education and particular urban occupations that best indexed "urbanness," namely:

a) Percentage of adults with advanced education;
b) Percentage of the labor force who are clerical and commercial employees.
c) Percentage of the female labor force in the utility services or health and welfare sectors (Industry census codes)
d) Percentage of the female labor force in professional-technical and office-managerial positions (Occupation census codes)

We determined that, of all of the more or less urban occupational categories, the three listed above are the most urban. The advanced education component is particularly apt because colleges and universities are strongly concentrated in cities, and because the level of institutions for higher learning is associated with level in the urban hierarchy, as are employment opportunities for the well educated.

From these variables we constructed a standardized index which we call the Substantive Urban Index (SUI). Since these same substantively urban variables also vary to some degree through the core-periphery structure at the scale of macroregional economies, we removed from each component that portion of the variance which could be ascribed to the regional systems index constructed for all of China Proper (described in the following paper). We then used the SUI as a guide in setting cutpoints in the continuous structural components of the URC and also in interdigitating logically ordered categories of classes of MC units.

Urban-Rural Continuum

In the case of the settlement composition index, we first established cutpoints in each of the 6 ordered listings (controlled by the urbanization index) to create discrete classes, and then interdigitated the 6 lists of ordered classes. This transformation yielded 44 ordered classes based on the settlement composition index with monotonically descending SUI means. A similar procedure was followed with the combined centrality scores: establishing cutpoints in each of the 11 ordered listings (controlled by level of the highest-level central place) to create discrete classes and then interdigitating the 11 lists of ordered classes. This transformation yielded a categorical variable of 36 classes based

on combined centrality score, which also exhibited monotonically descending SUI means.

These two components were then cross-tabulated. Cells containing one or more cases are, of course, concentrated along the diagonal, and we combined cells across this diagonal to create categories that maximized the homogeneity of SUI means while keeping the ordered progression across the diagonal. The end result was 48 classes. To avoid the assumption of equal intervals implied by the values 1 to 48, we label these classes not with those serial digits, but instead use the mean SUI value for each class.

These complicated procedures were designed to take full account of variations in the configuration and structure of county-level units that are a direct fall-out of the arbitrary boundaries of the administrative units for which we have data. Our procedures bring everything we know about the urban hierarchy to bear on the construction of an urbanization index applicable to units whose boundaries cross-cut the natural socioeconomic hierarchy. The success of the URC index in positioning administrative units along an urban-rural continuum will be apparent in the next paper.