V MODELING URBAN EXTENT AND EXPANSION

1. The role for analytic models of urban extent and expansion

The preceding chapter provided preliminary estimates of the global dimensions of urban areas in different world regions, income classes and city size classes and of the rate at which these areas are now expanding. While these numbers are no longer a mystery, they certainly pose great challenges, especially to the local, provincial and national public officials in developing countries and to international developing agencies that need to assist them. It is of the utmost importance to all stakeholders—be they ordinary citizens or planners and decision makers in the public, private or civic sector—to ensure that adequate quantities of *public goods* are put in place in a timely fashion, before it is too late: that there are adequate lands for absorbing the expected population growth; that there is an adequate capacity of trunk urban roads than can carry public transport; that there are adequate supplies of drinking water and effective means of sewerage disposal; that sensitive lands are protected from development; and that there is effective protection of open space. None of these are likely to be provided at adequate levels without concerted public action.

One possible reaction to the results presented so far might be to conclude that the magnitude of the problem has now been made manifest, and that the policy problems have been clearly delineated. If human societies are to accommodate the magnitude of urban growth that has been forecast, we must plan for and build adequate urban infrastructure for absorbing more than 1 million persons *every week* into cities for the next four decades. One might be forgiven for asserting that the time for analysis has passed, and the time for action is now upon us.

Still, some reflection and examination of existing trends and policies suggests that a deeper understanding of the forces that shape urban expansion in world cities is in order. In other words, there is still some serious thinking to be done before effective action can take place, because the phenomenon of urban expansion and the forces driving it are still not properly understood. An astonishing variety of factors have been put forward as contributing to urban expansion. In addition to the growing population of urban areas, these include the repulsion from central city problems[87] and the attraction of urban residents to rural amenities;[88] policy failures related to land use

---

[87] See Mills, E. and Lubuele, L. S., 1997, "Inner Cities", *Journal of Economic Literature*, 35, 727-756; Cullen, J. B. and Levitt, S., 1999, "Crime, Urban Flight, and the Consequences for Four Cities", *Review of Economics and Statistics*, 81, 159-169; or Brueckner, J., J. Thisse, and Y. Zenou, 1999, "Why is Central Paris Rich and Downtown Detroit Poor? An Amenity Based Theory", *European Economic Review*, 43, 91-107.

[88] See Irwin, E. and N. Bockstael, 2002, "Interacting Agents, Spatial Externalities and the Evolution of Residential Land Use Patterns" *Oxford Journal of Economic Geography,* 2, 31-54, or Wu, J., 2003, "Environmental Amenities and the Spatial Pattern of Sprawl" *American Journal of Agricultural Economics*, 83, 691-697.

regulation, housing finance[89], taxation and local public finance[90] or transportation;[91] increasing household incomes;[92] foreign direct investment and industrial structure;[93] access to drinking water through wells rather than through piped water;[94] and the increasing availability of automobiles and other alternatives means of transportation[95].

Even if each of these factors had some role to play in determining urban expansion, it is critical for policy makers to have some understanding of the *relative* contributions of different factors to urban expansion. This is particularly critical in the context of developing countries, where the sheer variety of economic circumstances, histories, levels of infrastructure, and modes of governance may pose situations far outside the sample of industrialized-country experiences that typically provide the setting for almost all of the empirical research on urban expansion.

It is towards this need that the present chapter is directed. We provide some initial analysis and modeling of the data that were described and presented in Chapters II through IV above. Our goal is to use the data collected so far to begin to address some basic questions concerning the relative importance of the factors that contribute to urban expansion. This provides an opportunity for developing a quantitative model of the urban extent and expansion that has been observed in our sample of cities—a model that will prove useful in planning and preparing for future urban growth. It also allows us

---

[89] See Voith, R. 1999, “Does the Federal Tax Treatment of Housing Affect the Pattern of Metropolitan Development?”, *Federal Reserve Bank of Philadelphia Review*, March, 3-16.

[90] See Brueckner, J. K. and Hyun-A Kim, 2003, “Urban Sprawl and the Property Tax”, *International Tax and Public Finance*, 10: 5–23, or Brueckner, J. K., 1997, “Infrastructure Financing and Urban Development: The Economics of Impact Fees”, *Journal of Public Economics,* 66**,** 383-407.

[91] See Hart, S. and A. Spivak, 1993, “Elephant in the Bedroom: Automobile Dependence and Denial Impacts on The Economy and Environment”; or, for a more serious analysis, see Hansen, M., D. Gillen, and M. Puvathingal, 1998, “Freeway Expansion and Land Development: An Empirical Analysis of Transportation Corridors”, Berkeley: Institute for Transportation Studies, University of California, Berkeley.

[92] Margo, R., 1992, “Explaining the Postwar Suburbanization of the Population in the United States: The Role of Income” *Journal of Urban Economics*, 31, 301-310.

[93] Seto, K. and Kaufmann, R., 2003, “Modeling the Drivers of Urban Land Use Change in the Pearl River Delta, China: Integrating Remote Sensing with Socioeconomic Data”, *Land Economics,* 79, 106-121, or Felsenstein, D., 2002, “Do High Technology Agglomerations Encourage Urban Sprawl?” *The Annals of Regional Science,* 36, 663-682.

[94] One of many contributing factors identified in Burchfield, N., H. Overman, D. Puga, and M. Turner, 2004, “The Determinants of Urban Sprawl: A Portrait from Space”, University of Toronto Working Paper.

[95] See Handy, S., 2005, “Smart Growth and the Transportation–Land Use Connection: What Does the Research Tell Us?” *International Regional Science Review,* 28, 146-167, or Glaeser, E. and M. Kahn, 2004, “Sprawl and Urban Growth” Chapter 56 in the *Handbook of Regional and Urban Economics,* Henderson, J. V., and J. F. Thisse, eds., Elsevier: Amsterdam.

to test our theoretical understanding of urban expansion by comparing our model outcomes with those predicted by the most accepted and widely used theories of urban spatial structure.

2. Theory and hypotheses explaining urban expansion

The expansion of urban areas is determined by the interaction of three broad types of phenomena: the physical constraints of geography and environment, the demand for land by the households and firms who inhabit the city, and the policy constraints that govern land use and spatial interactions in the city. The most useful models for informing public action on the management of urban expansion will be those models that incorporate each of these factors in some way, and that evaluate the relative contribution of each factor to urban expansion.

Unfortunately, we do not have the same level of theoretical understanding of the effects of the physical, economic, and policy environments on urban expansion. Very little work has been done on the effect of climate, ecological biomes or topography on the form of cities. And while some models of expected policy effects do exist, for the most part such analyses have been limited to an *ex post* evaluation of the extent to which a particular type of policy appears to have been effective, *ceteris paribus*, in influencing urban structure. This type of analysis remains a potentially useful exercise because it does provide important information to policy makers about where successes and failures have occurred. A serious constraint on its usefulness, however, has been the relatively limited variety of contexts within which such policy analysis has taken place. Those analyses that have been undertaken have usually focused on individual cities, and almost entirely on cities located in industrialized countries. Under what conditions can these limited results be extended to developing countries or to transition economies? An important long run goal of the present research is to investigate this question.

The economic model of urban spatial structure is, by contrast, relatively well developed, though not necessarily more accurate in predicting actual outcomes. Several authors[96] provide clear expositions of the by-now familiar theory, which proceed briefly as follows: We consider an urban area with exogenously given population of L households having income y and preferences that are represented by a common quasi-concave utility function $v(c,q)$ that depends on consumption of a composite good c and housing q. Each household has a worker who is employed in the city center and must commute to the center to earn income. The household's annual transportation costs for this commute are $t \times x$ if it resides in a house x units of distance from the center.

---

[96] See, *inter alia*, Mills, E., 1972, *Studies in the Structure of the Urban Economy,* Baltimore: Johns Hopkins University Press; Henderson, J. V., 1977, *Economic Theory and the Cities,* New York: Academic Press; or Brueckner, J., 1987, "The Structure of Urban Equilibria", Chapter 20 in *Handbook of Regional and Urban Economics*, E. Mills, ed., New York: Elsevier. We use their notation and basic approach in our discussion.

Equilibrium requires that a common utility level u be achieved by a household at any location within the built-up area of the city, so that the price per square meter of housing will vary with distance x. Households will allocate their income to select the most preferred combination of the composite good and housing, so that in equilibrium we must have:

$$\max_q v\left(y - t \cdot x - q \cdot p\left(x\right), q\right) = u \tag{1}$$

for all households.

Housing producers combine inputs of capital N and land l using a concave constant-returns production function $H(N, l)$ to produce square meters of housing. Housing production therefore exhibits diminishing marginal productivity of both capital and land. Constant returns to scale and free entry of housing producers is sufficient to determine an equilibrium land rent function $r(x)$ and a capital-land ratio (building density) $S(x)$ that depend upon distance x from the city center and satisfy:

$$\frac{\partial r\left(x\right)}{\partial x} < 0 \quad and \quad \frac{\partial S\left(x\right)}{\partial x}, \tag{2}$$

so that both land value and building density decline with distance from the city center. Combining the solution for building density $S(x)$ with the housing $q(x)$ demanded by a household at distance x provides a solution for the population density $D(x,t,y,u)$ at distance x, given the exogenous levels of transport costs and income and the achieved utility level u.

The maximum extent of the urban area $\bar{x}$ depends on the ability of housing producers to bid land away from its alternative uses. Let r_A represent the alternative use value of land (often explained heuristically as the market rent of land in agricultural use). The maximum extent of the urban area is then given implicitly by:

$$r\left(\bar{x}\right) = r_A \tag{3}$$

Finally, equilibrium requires that all households be accommodated in the urban area. If θ represents the share of land available for development at each distance, this is ensured by the following equilibrium condition:

$$\int_0^{\bar{x}} 2\pi \cdot \theta \cdot x \cdot D\left(x, t, y, u\right) dx = L\,. \tag{4}$$

This basic theory provides an endogenous solution for the maximum extent of urban land use, and relates this solution to several observable characteristics of the urban area. In particular, we can derive a number of comparative static results from this model that provide clear, testable hypotheses for our analysis.

The model discussed above has housing producers (and agricultural producers outside of the urban area whose demand for land generates the rents r_A) as the only direct consumers of land. It is easy to generalize this model so that firms who trade in the city center are included as well, combining inputs of capital and land according to *f(N,l)* to produce an export good for external markets sold at price *w*. These firms provide a separate commercial demand for land. Assuming that the cost (in terms of reduced profitability) of moving production away from the urban center is greater than the aggregate commuting cost of the households who would occupy an equal amount of land, the firms will be more centrally located than the households. In this case we can derive two additional hypotheses concerning the impact of changes in the productivity of land in export-good production, and the impact of an increase in the world demand for the export good. All of the hypotheses derived from this model of urban spatial structure are summarized and described in table V-1 below.

Table V-1: Hypotheses concerning urban spatial structure derived from the standard economic model

No.	Comparative Static Result	Description of prediction and hypothesis
1.	$\frac{\partial \bar{x}}{\partial L} > 0$	An increase in population will increase urban extent and urban expansion.
2.	$\frac{\partial \bar{x}}{\partial y} > 0$	An increase in household income will increase urban extent and urban expansion.
3.	$\frac{\partial \bar{x}}{\partial t} < 0$	An increase in transportation costs will reduce urban extent and limit urban expansion.
4.	$\frac{\partial \bar{x}}{\partial r_A} < 0$	An increase in the opportunity cost of non-urban land will reduce urban extent and limit urban expansion.
5.	$\frac{\partial \bar{x}}{\partial H_l} > 0$	An increase in the marginal productivity of land in housing production will increase urban extent and urban expansion.
6.	$\frac{\partial \bar{x}}{\partial \theta} > 0$	An increase in the share of land available for housing development will increase urban extent and urban expansion.
7.	$\frac{\partial \bar{x}}{\partial f_l} > 0$	An increase in marginal productivity of land in production of the export good will increase urban extent and urban expansion.
8.	$\frac{\partial \bar{x}}{\partial w} > 0$	An increase in the world price of the export good will increase urban extent and urban expansion.

One of the primary objectives of this study was to test these hypotheses with the data from our sample of cities, so as to provide some confirmation of the traditional economic theory of urban extent and urban land use. While successful predictions derived from our empirical data cannot establish the ‘truth’ of the theory, they can surely help increase our confidence in applying this theory, especially in anticipating the impacts on urban expansion that might result from alternative policies or from exogenous changes in the economic environment.

In a full test of this model of urban spatial structure, we would account not only for the factors explicitly identified in the theory as important, but also for those external factors such as policy, environmental, and geographic constraints. A more complete test awaits the completion of ongoing data collection in the second phase of this study, but we are already able to provide some initial results to examine the predictive value of this model at the present time. The following section discusses the data we have collected for analysis so far. Following that is a section that presents our estimates of models of urban land use. We then proceed to engage in a very preliminary exploration of the predictive value of economic variables in explaining some measures of the structure of urban areas, such as their *compactness* and *contiguity* (as defined in Chapter IV). We conclude this chapter with an evaluation of the predictive value of the models and an outline of the future directions of our ongoing econometric research into the causes and consequences of urban expansion.

3. The collection of data for the analysis of urban expansion

The data used for our analysis begins with the measures of urban land use in the sample of 90 cities discussed extensively in Chapters II through IV above. These provide key measures—for each city in the sample—of both population and urban land cover (in square kilometers) at two points in time. We can use these data as either 180 observations of urban population and land use, collected at various points in time, or, alternatively, as 90 observations of change in urban land use.

Measures of the built-up area and the maximum slope in the built-up area for each city were derived by the research team as described in Chapter IV above (see also the individual city images and their associated measures presented in Chapter VII). The total population of the cities in the sample was derived from small area population estimates provided by CIESIN.[97] These were collected by CIESIN from national census offices in each country, as part of its project for providing population estimates for the entire world at a 1km-grid resolution. The CIESIN population estimates for the administrative districts that contained each of the cities in the sample were extrapolated to match the dates T_1 and T_2 of each city's satellite images.

To test the hypotheses listed in Section 2 above, we combined the data generated by our study with available information on population, income, and other relevant variables. These additional data were gathered from a variety of sources. The following three tables, V-2, V-3 and V-4, list the data used in estimating the three groups of econometric models that will be presented below. The final line of each table provides descriptive statistics for the sampling weights based on urban populations. These were used in calculating the descriptive statistics and in estimation of all models discussed below.

---

[97] Documented more fully at http://beta.sedac.ciesin.columbia.edu/gpw/ and http://sedac.ciesin.columbia.edu/plue/gpw/index.html?main.html&2.

Descriptive statistics for variables used in estimating models 1–3 (discussed in Section 4 below) are presented in Table V-2. The final three variables in table V-2 are dichotomous variables indicating location within one of three biomes[98] that turned out to be significant determinants of the level of urban land cover. These three biome indicators and the Ground Water indicator were used in all models.

TableV-2: Variables Used in Models 1-3

Variable	Mean	σ	Min	Max
Urban Land Use (km^2)	400.6871	533.7343	8.91769	2328.87
Total Population	3,287,357	4,179,050	105,468	1.70E+07
Per Capita GDP (PPP 1995 $)	9,550.217	9,916.317	562.982	32,636.5
National share of IP addresses	0.085741	0.193696	3.50E-06	0.593672
Air Linkages	88.78808	117.6716	0	659
Maximum Slope (percent)	25.34515	14.55289	4.16	72.78
Agricultural Rent ($/Hectare)	1,641.608	3,140.596	68.8372	19,442.1
Cost of fuel ($/liter)	0.581498	0.328673	0.02	1.56
Cars per 1000 persons	144.7495	191.4476	0.39	558.5
Ground Water (1=shallow aquifer)	0.281518	0.451022	0	1
Temperate Humid Climate	0.077395	0.267979	0	1
Mediterranean Warm Climate	0.005109	0.071499	0	1
Mediterranean Cold Climate	0.017234	0.130515	0	1
Sampling Weight	0.011168	0.010542	0.000834	0.068174

Descriptive statistics for additional variables used in estimating models 4–6 (also discussed in Section 4) are presented in Table V-3 below.

Table V-3: Additional Variables Used in Models 4-6

Variable	Mean	σ	Min	Max
Change in Built-Up Area	125.8202	163.3169	-322.559	527.368
Change in Total Population	751827.3	1474634	-470586	5.40E+06
Change in Per Capita GDP	1566.28	2156.812	-4552.33	6722.88
Air Links in 1990	88.03663	124.1801	0	659
Maximum Slope in 1990	25.03812	14.3309	4.16	70.63
Agricultural Rent in 1990	1589.797	3396.454	84.9003	19442.1
Fuel Cost in 1990	0.436883	0.247924	0.02	1.18
Cars per 1000 in 1990	130.7622	182.7599	0.39	489.2
Sampling Weight	0.011168	0.010573	0.000834	0.068174

The cost of fuel at the pump, the numbers of cars per 1,000 persons, and per capita GDP (in constant 1995 US dollars converted as PPP exchange rates)—all available only at the national level—were obtained from the *World Development Indicators* (WDI) website of the World Bank. The 'Agricultural Rent' variable was also available only at

---

[98] Obtained from U.S. Department of Agriculture, Natural Resources Conservation Service, Soil Survey Division, World Soil Resources, Washington, D.C.

the national level. It was computed from WDI data by calculating the total value-added in agriculture (in constant US dollars) per hectare of arable land. The Share of IP Addresses was calculated from data available from *MaxMind*[99]. It provided, for each country, the share of Internet Protocol (IP) addresses assigned to that country relative to the total of all IP addresses assigned throughout the world. The available data is current as of 2005, but we used it as a general indicator of available business infrastructure in the country, and hence of the marginal productivity of land for business production.

Table V-3 above provides descriptive statistics for those variables used in the models of *change* in urban land cover (rather than absolute levels of urban land cover). Several of the variables shown in the table were available at the city level. The Air Links variable provided the number of airports[100] connected via incoming flights (both non-stop and connecting) to the cities in our sample from mid-March through mid-June for 1990 (T_1) and 2000 (T_2). This measure and similar measures have been shown[101] to be highly correlated with general global connectedness and the volume of international trade. We used this measure, therefore, as an indicator of global connectedness and the price that domestically produced export goods are likely to receive in global markets. The dichotomous variable 'Ground Water' indicates that the city is located in an area with localized and shallow aquifers[102], so that while drilling a well to obtain water may not be possible at all locations, it is possible at many locations and the required depth is not great.

Descriptive statistics for variables used in estimating models 7–10 (discussed in Section 4) are presented in Table V-4 below.

Table V-4: Variables Used in Models 7-10

Variable	Mean	σ	Min	Max
Ln(Urban Area)	5.217764	1.302409	2.18804	7.75314
Ln(Total Population)	14.26064	1.243901	11.5662	16.6682
Ln(Per Capita GDP)	8.596582	1.099758	6.33325	10.3932
Ln(Share IP Addresses)	-5.249607	3.012159	-12.5592	-0.52143
Ln(Air Links+1)	2.923513	2.21341	0	6.49224
Ln(Maximum Slope)	3.065746	0.595572	1.42552	4.28744
Ln(Agricultural Rent)	6.757474	0.980555	4.23174	9.8752
Ln(Fuel Cost)	-0.71369	0.640135	-3.91202	0.444686
Ln(Cars Per 1,000)	3.399618	2.1609	-0.941609	6.32525
Sampling Weight	0.011168	0.010542	0.000834	0.068174

---

[99] Available at http://www.maxmind.com/app/geoip_country.

[100] Measured from the *OAG Data* database.

[101] See Gugler, J., 2004, *World Cities beyond the West: Globalization, Development and Inequality*, Cambridge: Cambridge University Press.

[102] The data are obtained from *Groundwater Resources of The World* (world map at the scale of 1:50 Million meant to provide a global overview), BGR Hannover /UNESCO, Paris 2004.

Table V-4 provides the information for the data used in the final four models that relate the natural logarithm of urban land cover to the natural logarithm of non-dichotomous variables.

4. Models of urban extent and urban expansion

The economic model of urban extent and urban land use outlined in section 2 above does not specify a particular functional relationship between urban extent (the built-up area of the city) or urban expansion (change in total built-up area) and the variables discussed in the hypotheses listed in table V-1. The functional form would depend on the exact functional forms of the utility and production functions in the model.

We tested these hypotheses by estimating three general types of models. We first examined linear models of urban extent, estimating three slightly different relationships between the total built-up area of the city and the variety of explanatory variables discussed and described in the preceding section. These three models—labeled Models 1, 2 and 3—are presented in table V-5. Each model has the total built-up area of the city (measured in square kilometers) as the dependent variable.

In this table (and in all subsequent tables presenting model estimates), the estimated model is presented in a column, with a blank spot indicating that the variable associated with that row is not included as part of the specification of that model. Each parameter estimate is presented in larger type, with the standard error of the estimate presented in italics below the estimated parameter. Those parameters whose estimates are statistically significant at the 10 percent level or better are printed in bold face. Robust standard errors σ are used and reported for all estimates. All model estimates were obtained using the *STATA 9 Special Edition* statistics software.

The primary difference between models 1, 2 and 3 is the inclusion of a variable to measure the impact of transportation costs t, so as to provide a test of hypothesis 3 presented in Section 2 above. The data available to us are of limited use, since they are available only at the national level (rather than the individual city), and since they are often highly correlated with income. While it can be argued that the use of national level data provides an exogenous measure of transport costs whose variation is not simultaneously determined by the level of urban land cover, this measure is certainly noisier than a direct observation of local fuel costs and transport mode choices at the city level.

The available measure of automobile use—cars per 1,000 persons—is difficult for two additional reasons. First, it is not clear whether increasing the number of automobiles per capita decreases transport costs (since automobiles are faster than public transportation) or increases it (due to increased congestion costs). Furthermore, the numbers of cars per 1,000 persons are very highly correlated with national income, so that problems of colinearity arise when both are included in the model. This is clear from examination of table V-6, which provides the Variance Inflation Factors for variables in Model 3 in table V-5. The table clearly demonstrates that the inclusion of both income and cars per 1,000 people in the model has a considerable impact on the variance of parameter estimates.

Table V-5: Linear Models of Total Urban Land Cover

	Model 1	Model 2	Model 3
Total Population	**0.000046**	**0.000046**	**0.000045**
σ	*0.000012*	*0.000012*	*0.000012*
Income	**0.007656**	**0.007204**	0.012503
σ	*0.0030*	*0.0028*	*0.0116*
Share of IP Addresses	**1035.0870**	**1059.1800**	**1003.3360**
σ	*279.5773*	*297.3413*	*282.5092*
Air Links	**1.6540**	**1.6467**	**1.6908**
σ	*0.2803*	*0.2953*	*0.2880*
Maximum Slope	-1.3593	-1.3574	-1.3620
σ	*1.7102*	*1.7061*	*1.7285*
Agricultural Rent	**-0.0111**	**-0.0114**	**-0.0122**
σ	*0.0043*	*0.0048*	*0.0054*
Fuel Cost		17.2982	
σ		*70.4921*	
Cars/1000			-0.2458
σ			*0.4976*
Shallow Ground Water	**97.2364**	**98.1368**	**95.3943**
σ	*51.6306*	*52.9738*	*52.8588*
Temperate Humid	**-225.0211**	**-224.1264**	**-217.8070**
σ	*106.3447*	*107.1763*	*105.1351*
Mediterranean Warm	**275.4711**	**274.0859**	**271.0372**
σ	*42.4932*	*40.5861*	*44.2845*
Mediterranean Cold	63.9141	61.7096	66.0667
σ	*36.8820*	*34.0951*	*36.9881*
Constant	-19.2077	-26.3652	**-26.4953**
σ	*44.8569*	*62.1321*	*50.1161*
Number of observations	176	176	176
R-squared	0.7858	0.7858	0.7862
Root MSE	254.43	255.16	254.91

The parameter estimates for fuel costs and cars per 1,000 are also disconcerting. While neither is statistically significant (owing to the large standard errors with which the parameters are estimated) they also have the wrong sign, with higher fuel costs apparently leading to increased urban land use and higher cars per 1,000 persons leading to reduced urban land use.

Despite the difficulties in obtaining accurate estimates of the impacts of changes in transport costs, the models reported in table V-5 seem to perform surprisingly well. Each model explains nearly 79 percent of the total sample variation in urban land cover.

Population, income, and agricultural rent are statistically significant in each model, and the signs of each parameter estimate are consistent with hypotheses 1, 2 and 3 respectively.

Taking the Share of IP Share as a measure of the marginal productivity of land in export production and Air Links as a measure of the demand for the cities export product, we note that parameter estimates for both are statistically significant and signed in a way to be supportive of hypotheses 7 and 8. The Maximum Slope variable may be taken as indicative of both limitations in the amount of land available for residential development (an index of θ), and as an indicator of the marginal productivity of land in the production of housing. While the parameter estimate is not statistically significant in any of the three models, it is signed in a way that is consistent with hypotheses 5 and 6.

Table V-6: Variance Inflation Factors for Full Linear Model

Variable	VIF
Income	20.35
Cars per 1,000	15.85
Share of IP Addresses	3.31
Air Links	2.35
Total Population	1.90
Agricultural Rent	1.47
Maximum Slope	1.23
Temperate Humid	1.21
Ground Water	1.17
Mediterranean Cold	1.03
Mediterranean Warm	1.02
Mean	4.63

The dichotomous indicator for shallow aquifers provides a more direct test of hypothesis 5, since such aquifers reduce the amount of capital required to provide housing with water (in contrast with the higher capital cost of water that is obtained through deep wells or through the extension of water pipes from municipal sources). Thus land in areas with shallow aquifers has higher marginal productivity in housing production, and should be associated with increased urban land cover. The parameter estimates for this variable are statistically significant and directly support this hypothesis.

Models 4, 5 and 6 reported in table V-7 below take a different approach to estimating these relationships. Rather than use the total built-up area of cities as a measure of urban extent, these models use the change in built-up area as a direct measure of urban expansion.

Estimation of these models poses two types of data problems. First, the time period over which the change in urban land cover takes place is not the same for each city. The time periods depend on the availability of usable cloud-free satellite images and therefore range from periods of approximately 8 years to more than 12. These changes would ideally be matched with changes in the independent variables. For some data (for example, the Share IP Addresses or other measures of business infrastructure) we lack measures of change over this period. In other cases we have measures of change that fail to correspond to the periods of change in our land cover data.

For the two most important determinants—population and income—we do have relatively complete data and we can interpolate the data to get change in these variables that corresponds exactly to the time period over which our land cover change is measured. For our other variables, we use either the one measure we have for each city, or the measure available at the beginning of the time period, so as to provide an indication of the economic conditions that produced the observed urban expansion.

Again, examining table V-7, we see that all three models perform quite well, explaining roughly 82 percent of the total sample variance in urban expansion. Population change and income change are statistically significant in every model, and their signs are consistent with hypotheses 1 and 2. The Share of IP Addresses and the Shallow Ground Water variables are also statistically significant and supportive of hypotheses 7 and 5 respectively.

Table V-7: Linear Models of Urban Expansion

	Model 4	Model 5	Model 6
Population Change	**0.000083**	**0.000085**	**0.000084**
σ	*0.000007*	*0.000007*	*0.000006*
Income Change	**0.02169**	**0.01813**	**0.020129**
σ	*0.0077*	*0.0075*	*0.0073*
IP Share	**237.1614**	**279.7229**	**270.6102**
σ	*96.9735*	*101.1712*	*110.2253*
T_1 Airlink	0.1383	0.1154	0.1301
σ	*0.1003*	*0.1044*	*0.1105*
T_1 Maximum Slope	**-1.2954**	**-1.1688**	**-1.2267**
σ	*0.7050*	*0.7033*	*0.7111*
T_1 Agricultural Rent	-0.0011		
σ	*0.0011*		
T_1 Fuel Cost		21.0234	
σ		*34.4916*	
T_1 Cars/1,000			-0.0199
σ			*0.0902*
Shallow Ground Water	**36.0570**	**35.8025**	**36.5591**
σ	*19.6228*	*19.1266*	*19.1902*
Temperate Humid	-54.7146	-49.8376	-47.4455
σ	*40.3215*	*40.3067*	*41.9332*
Mediterranean Warm	**148.9260**	143.7444	**147.4802**
σ	*27.0154*	*26.9233*	*26.6382*
Mediterranean Cold	9.9700	13.9181	12.9924
σ	*15.6299*	*15.5493*	*14.8440*
Constant	24.2468	10.6364	20.8378
σ	*17.0721*	*21.2756*	*16.3076*
Number of observations	88	90	90
R-squared	0.8207	0.816	0.8154
Root MSE	73.515	74.035	74.154

In Models 4, 5 and 6 the Maximum Slope parameter is statistically significant and signed as would be suggested by hypotheses 5 and 6. While neither the Air Links variable nor the Agricultural Rent variable are statistically significant, both are signed as would be expected by hypotheses 8 and 4 respectively. The Fuel Cost and Cars per 1,000 variables continue to be insignificant and incorrectly signed.

Our final set of models returns to examination of the relationship between total urban land cover and the explanatory variables. Models 7 through 10, reported in table V-8 below, relate the *logarithm* of total urban land cover to the *logarithm* of all the non-dichotomous variables in table V-4 as well as to the binary variables in table V-2. This logarithmic functional form evaluates a non-linear relationship between the variables, and permits easy interpretation since estimated parameters provide elasticity measures of the dependent variable with respect to the independent variables.

Table V-8: Logarithmic Models of Urban Expansion

	Model 7	Model 8	Model 9	Model 10
LN Total Population	**0.662338**	**0.664504**	**0.662429**	**0.664468**
σ	*0.0502*	*0.0487*	*0.0505*	*0.0488*
LN Income	**0.495863**	0.024581	**0.498571**	0.014567
σ	*0.0700*	*0.1195*	*0.0704*	*0.1296*
LN Share of IP Addresses	**0.0513**	**0.0901**	**0.0500**	**0.0917**
σ	*0.0237*	*0.0228*	*0.0240*	*0.0241*
LN Air Links	**0.1222**	**0.1057**	**0.1210**	**0.1065**
σ	*0.0256*	*0.0246*	*0.0258*	*0.0245*
LN Maximum Slope			0.0300	-0.0238
σ			*0.0791*	*0.0788*
LN Agricultural Rent	**-0.2601**	**-0.2142**	**-0.2675**	**-0.2076**
σ	*0.0350*	*0.0358*	*0.0361*	*0.0376*
LN Fuel Cost	-0.0870	-0.1168	-0.0843	-0.1194
σ	*0.0743*	*0.0793*	*0.0741*	*0.0779*
LN Cars/1000		**0.2228**		**0.2265**
σ		*0.0508*		*0.0539*
Shallow Ground Water	**0.2665**	**0.2057**	**0.2530**	**0.2154**
σ	*0.0913*	*0.0917*	*0.0853*	*0.0835*
Temperate Humid	**-0.3141**	**-0.3456**	**-0.3267**	**-0.3362**
σ	*0.1301*	*0.1274*	*0.1347*	*0.1303*
Mediterranean Warm	**1.1353**	**1.8896**	**1.1080**	**1.9238**
σ	*0.1642*	*0.2079*	*0.1757*	*0.2418*
Mediterranean Cold	**0.6808**	**0.5289**	**0.6883**	**0.5204**
σ	*0.2375*	*0.2699*	*0.2455*	*0.2724*
Constant	**-6.9516**	-3.7494	**-7.0141**	**-3.6463**
σ	*1.2198*	*1.3542*	*1.2575*	*1.4917*
Number of observations	176	176	176	176
R-squared	0.8858	0.8967	0.8859	0.8968
Root MSE	0.45327	0.43237	0.45439	0.43353

Again, the models appear to perform very well, explaining between 88 and 90 percent of the total variation in the log of total urban land cover. The parameters associated with variables measuring Total Population, the Share of IP Addresses, Air Links, Agricultural Rent and Shallow Ground Water are statistically significant in every model and signed correctly, providing support for hypotheses 1, 4, 5, 7, and 8.

The income variable is correctly signed in all models and statistically significant in Models 7 and 9, providing general support for hypothesis 2. The estimation in Models 8 and 10 is again plagued by colinearity problems that increase the variance of parameter estimates, as suggested by the Variance Inflation Factors presented in table V-9.

Table V-9: Variance Inflation Factors for Full Logarithmic Model

Variable	VIF
Income	18.71
Cars / 1,000	12.99
Share of IP Addresses	3.84
Population	3.26
Air Links	3.11
Agricultural Rent	1.54
Maximum Slope	1.5
Shallow Ground Water	1.25
Mediterranean Warm	1.23
Fuel Cost	1.17
Temp Humid	1.16
Mediterranean Cold	1.09
Mean	4.24

Models 7 through 10 provide the first support for hypothesis 3, with the impact of higher fuel costs being to reduce urban extent, although the estimated parameters remain statistically insignificant. The estimated parameter associated with Cars per 1,000 Persons is statistically significant in the two models where it appears. Assuming that increasing this value indicates lower transportation costs, this result also provides support for hypothesis 3.

5. Models of urban compactness and contiguity

Other characteristics of urban structure besides urban extent and urban expansion have drawn attention of some policy makers, urban planners and scholars. In Chapter IV above we introduced and provided estimates for two of these measures of urban form—compactness and contiguity—for regions, income groups and city size classes. The *Compactness Index* was defined as 'the share of the *buildable* area (area with no bodies of water or excessive slope) in a circle of minimum radius containing the main built-up area of the city that is actually built-up'. The *Contiguity Index* was defined as 'the share of the main, contiguously built-up area in the total built-up area of the city'. Other measures of urban form—such as the *Openness Index*—are presently being explored and will be reported on in later publications of the present study. Not much has been written concerning the determinants of these aspects of urban form[103]. The considerable

---

[103] Although Mayo, S. and S. Sheppard, 2001, "Housing Supply and the Effects of Stochastic Development Control" *Journal of Housing Economics,* 10, 109-128, present a model that identifies increasing holdings of vacant land by housing producers as a natural response to

interest in these variables, however, leads us to undertake an initial exploration of the extent to which our economic variables can explain the variations in these indices.

Table V-10 reports the results of our analysis of Contiguity and Compactness. We have estimated linear and logarithmic versions of models for both the Contiguity Index and the Compactness Index. These models are not derived from a theoretical model as was done for the models presented in the section above, so we must view these results as very preliminary.

Table V-10: Linear and Logarithmic Models of Contiguity and Compactness

	Contiguity	Compactness	Ln (Contiguity)	Ln(Compactness)
Total Population	-3.4400000E-09	3.5300000E-09	**0.0611118**	0.0129526
σ	*7.4400000E-09*	*3.2300000E-09*	*0.0335*	*0.0358*
Income	-0.0000016	-0.0000008	**0.1737462**	-0.015874
σ	*0.0000033*	*0.0000017*	*0.0692*	*0.0685*
Share of IP Addresses	0.0657763	-0.0864179	**-0.0683419**	-0.0241164
σ	*0.1462*	*0.0706*	*0.0233*	*0.0225*
Air Links	**0.0004387**	0.0000154	0.0080391	0.0284618
σ	*0.0002*	*0.0000971*	*0.0200*	*0.0191*
Maximum Slope	**-0.0033939**	-0.0006475		
σ	*0.0013*	*0.0007*		
Agricultural Rent	**0.0000115**	-0.0000008	-0.0193364	0.0009924
σ	*0.0000048*	*0.0000024*	*0.0252*	*0.0257*
Fuel Cost			0.0118521	0.0426241
σ			*0.0390*	*0.0469*
Shallow Ground Water	0.0646657	-0.0120385	0.0238876	-0.0410067
σ	*0.0458*	*0.0252*	*0.0821*	*0.0691*
Temperate Humid	**0.1717063**	0.0067534	0.1288521	-0.0075718
σ	*0.0556*	*0.0290*	*0.0907*	*0.0887*
Mediterranean Warm	**-0.267969**	0.0423031	**-0.948665**	-0.0052976
σ	*0.0449*	*0.0249*	*0.0717*	*0.0829*
Mediterranean Cold	0.0861863	-0.0573825	**0.2809879**	-0.1641055
σ	*0.0727*	*0.0295*	*0.0879*	*0.1454*
Constant	**0.7036003**	**0.365947**	**-3.059962**	-1.338613
σ	*0.0444*	*0.0233*	*0.9433*	*0.9990*
Number of observations	176	176	176	176
R-squared	0.1485	0.0827	0.238	0.1275
Root MSE	0.19683	0.10494	0.33187	0.30761

Overall the models explain far less of the variance in the Contiguity Index or the Compactness Index than the models of urban extent or urban expansion managed using the same data. In only one of the models (for the logarithm of Contiguity) is either population or income a significant determinant of the dependent variable.

---

the risk associated with land use regulations. This analysis would suggest that decreased contiguity is a natural response to increasing regulatory risk.

The biome within which the city is located and the maximum slope in its built-up area are both more frequently significant than economic variables. This suggests a conclusion that we advance tentatively, with the understanding that it requires further investigation and analysis: it seems as if the contiguity and compactness of urban form are determined, to an important extent, by the physical and climatic constraints that affect the urban area. Were this conclusion supported by subsequent investigation, it would indicate that there is little scope for *compact city* policies to have significant influence on the contiguity or compactness of urban form. Compactness might well be important in determining the welfare of residents or the environmental impacts of urbanization, but there may be little that policy makers—in many, if not in most cities—can do to influence their overall compactness in any significant way.

6. Conclusions and directions for further research

This chapter opened with an observation about the important role for analytic models of urban expansion in policy making. We proceeded to use a standard neoclassical urban economics model to derive eight testable hypotheses about factors that influence urban extent and urban expansion. We then assembled data from a variety of sources and combined it with the measures of urban land cover and changes in urban land cover that have been presented in chapters II through IV above.

These hypotheses were tested using these data through the estimation of ten models. Each of the hypotheses was directly supported in at least two model estimates, and none of the hypotheses were directly contradicted with statistically significant estimates in any model. In general the estimated models performed well, explaining over 80 percent of the total variation in urban extent and urban expansion.

Our logarithmic models provide for easy interpretation and offer a striking and important observation. These models suggest that a doubling of urban population is generally associated with an increase in urban land area of about 66 percent. This means that, holding other factors constant, the process of urbanization should result in denser cities.

The difficulty lies in the fact that other factors are not held constant. Income growth also results in urban expansion, and these models suggest that a doubling of income is associated with a 49 to 50 percent increase in urban land area. Thus if global efforts to encourage economic development are successful, the policy challenges generated by urban expansion naturally follow—as urban residents improve their economic circumstances, they consume more land.

We find that globalization and interconnectedness tend to increase the rate of urban expansion. Doubling the linkages via air transport, for example, is associated with an 11 to 12 percent increase in urban land use. On the other hand, increasing agricultural productivity can provide a countervailing force, with a doubling of the value added per hectare resulting in a 26 percent decline in urban land use. This confirms the result reported on by Brueckner and Fansler[104] who found that US cities were less expansive in regions where the surrounding farmland was more valuable.

There are several directions in which we plan to move forward as the research progresses. These relate both to data collection and to econometric analysis. Beyond the measures of total urban land use and slope, the data available for this research was drawn from published or publicly available sources. In many cases these provided only a rough indication of the conditions in the urban area itself. No data whatsoever were available to provide a measure of the level of planning constraints, the type of development policy in the city, or the price of land and housing in the city. Local consultants hired by the project are now in the process of collecting these data in our global sample of 120 cities, and we plan to analyze the returns once data collection is completed.

The second important direction for our analysis is to evaluate the possible endogeneity associated with some of our independent variables. This is most notably relevant in variables that determine the cost of transportation, but may affect other variables used in this first analysis as well. These concerns will surely be relevant for considering the impact of planning policies that are devised in direct response to growth pressures in cities.

In summary, the results presented in this chapter have been encouraging in that they provide support to varying degrees for all of our derived hypotheses. The traditional neoclassical theory of urban spatial structure that is supported by our findings should prove useful in devising policy responses to the problems associated with preparing for urban expansion. There still remain important policy issues for which little analytical support is now available. Future analytical research, using the global data set generated by this study, should shed some light on these issues in the near future.

* * *

---

[104] Brueckner, J. K., and D. Fansler, 1983, "The Economics of Urban Sprawl: Theory and Evidence on the Spatial Sizes of Cities", *Review of Economics and Statistics,* 55**,** 479-82.