NOTES

THE ECONOMICS OF URBAN SPRAWL: THEORY AND EVIDENCE ON THE SPATIAL SIZES OF CITIES

Jan K. Brueckner and David A. Fansler*

I. Introduction

Many commentators believe that the phenomenon of urban sprawl, which is characterized by vigorous spatial expansion of urban areas, is a symptom of an economic system gone awry. By transforming pastoral farmland into often-unattractive suburbs, sprawl is thought to disrupt a natural balance between urban and non-urban land uses, leading to a deplorable degradation of the landscape.[1] This sentiment is often translated into policy through zoning restrictions designed to inhibit the conversion of land from agricultural to urban use (see Bryant and Conklin (1975)).

The economist's view of urban expansion stands in stark contrast to this emotionally-charged indictment of sprawl. Economists believe that urban spatial size is determined by an orderly market process which correctly allocates land between urban and agricultural uses. The model underlying this view, which was originally developed by Muth (1969) and Mills (1972) and more completely analyzed by Wheaton (1974), suggests that urban spatial size is determined in a straightforward way by a number of exogenous variables. By showing empirically that urban size is related to the given variables (population, income, agricultural rent, and commuting cost) in the manner predicted by the model, the present paper achieves two goals. First, the empirical results suggest that the economist's view of urban sprawl is justified: rather than being determined by a process which indiscriminately consumes agricultural land, urban sizes are the result of an orderly market equilibrium where competing claims to the land are appropriately balanced.[2] Second, by confirming the urban size predictions of the underlying model, the empirical results constitute yet another piece of evidence validating the basic framework of urban economic analysis.[3]

The plan of the paper is as follows. Section II sketches the structure of the Muth-Mills model and presents the main comparative static results relevant to urban sprawl. With the model's predictions in focus, section III discusses the sample and the data, and section IV presents the empirical results. Section V offers conclusions.

II. Theory

Fundamental assumptions of the basic Muth-Mills model are that all consumers earn the same income y at the central business district (CBD) and that tastes over housing q and a composite non-housing good c are identical. Housing, which is measured in square feet of floor space, is available for rent at price per square foot p, with p depending on distance x to the CBD. At each location, consumers maximize utility $v(c, q)$ subject to the budget constraint $c + pq = y - tx$, where t is commuting cost per round-trip mile. Spatial variation in p guarantees that all consumers reach the same utility level u regardless of location. Together, the uniform-utility condition and the first-order condition for utility maximization determine p and q as functions of x, t, y, and the utility level u.

Housing is produced with capital and land under constant returns. Producers maximize profit per acre of land, which equals $ph(S) - iS - r$, where r is land rent, i is the rental price of capital, S is structural density (capital per acre of land), and h gives square feet of housing per acre of land. Since p depends on x, t, y, and u from the consumer problem, the housing producer's first-order and zero-profit conditions determine S and r as functions of these same variables.[4] Population density in the model is $h(S)/q \equiv D$ (floor space per acre divided by floor space per dwelling), which is also a function of x, t, y, and u.

Explicitly denoting the functional dependencies of r and D, the equilibrium conditions for the city are

$$r(\bar{x}, t, y, u) = r_a \tag{1}$$

$$\int_0^{\bar{x}} 2\pi x D(x, t, y, u)\, dx = L, \tag{2}$$

Received for publication March 8, 1982. Revision accepted for publication November 9, 1982.

* University of Illinois at Urbana-Champaign.

[1] See Mills (1972, Ch. 6) for a critical summary of this position.

[2] While the Muth-Mills model portrays a world without market failure, it is, strictly speaking, illegitimate to use a favorable test of the model as proof of the efficiency of real-world urban equilibria (inefficient equilibria could possess the same comparative static properties as Muth-Mills equilibria). However, by showing that urban areas are not indiscriminate consumers of agricultural land, the empirical results go a long way toward deflating criticism of urban sprawl.

[3] Muth (1969) reports the results of regressions relating urban land areas in 1950 to a number of explanatory variables. Muth's regressions, however, omitted agricultural land rent and included some variables not explicitly suggested by a theoretical model.

[4] The capital cost parameter i is ignored as a determinant of S and r on the belief that intercity variation in the cost of capital is of minor importance compared to variation in the other determinants of the urban equilibrium. The assumption that the non-housing good c is numeraire similarly rules out empirical consideration of intercity variation in the prices of non-housing goods. This omission is again justified on the belief that such variation is of minor importance.

where r_a is agricultural land rent, L is the urban population size, and $\bar{x}$ is the distance to the urban-rural boundary.[5] Equation (1) says that urban and agricultural rents are equal at $\bar{x}$, while (2) says that the urban population exactly fits inside $\bar{x}$.

In a closed city, where population is fixed, the equilibrium conditions (1) and (2) equate the supply and demand for housing, determining u and $\bar{x}$ as functions of L, r_a, y, and t. Although housing is a final consumption good in the Muth-Mills model, it may be shown that the equilibrium given by (1) and (2) has the same comparative static properties as the one analysed by Wheaton (1974) (land was directly consumed in his model). The signs of Wheaton's comparative static derivatives for $\bar{x}$, which form the basis for the empirical work on urban sprawl presented below, are as follows:

$$\frac{\partial \bar{x}}{\partial L} > 0, \frac{\partial \bar{x}}{\partial r_a} < 0, \frac{\partial \bar{x}}{\partial y} > 0, \frac{\partial \bar{x}}{\partial t} < 0. \quad (3)$$

The intuition behind the results in (3) is easily stated. First, an increase in the urban population clearly must increase the distance to the edge of the city since more people must be housed. On the other hand, an increase in r_a raises the opportunity cost of urban land and makes the city more compact. Raising the level of income increases housing demand and leads to a larger city, while an increase in commuting cost per mile lowers disposable income at all locations, reducing housing demand and leading to a smaller city.[6]

To appreciate the complicated adjustments which underlie the results in (3) it is helpful to consider in detail the simplest type of change, that of a population increase, using a heuristic approach. If the city starts in equilibrium and population increases, excess demand for housing is created at the old prices: the urban population no longer fits inside the old $\bar{x}$. As a result, housing prices are bid up throughout the city, which means that the utility level u falls.[7] On the consumption side of the market, the housing price increase leads to a decrease in dwelling sizes at all locations.[8] On the production side of the market, the price increase causes land rents to be bid up everywhere, and higher land rents in turn lead producers to substitute away from land, resulting in higher structural densities. Since buildings are taller and dwellings smaller, population density rises everywhere, so that more people fit inside any given $\bar{x}$. Finally, the increase in the level of the land rent function leads to an increase in the value of $\bar{x}$ which satisfies (1). The resulting spatial expansion of the city, together with the increase in population densities, tends to eliminate the excess demand for housing. In the new equilibrium, the city has a lower utility level, a larger area, smaller dwellings, and higher housing prices, land rents, and structural and population densities. Although the effects of an increase in r_a are similar to the above (the internal changes are identical, though $\bar{x}$ falls), an increase in y or t leads to more complex internal adjustments.

While the preceding discussion concerns the closed-city urban model, an alternative assumption is that the city is open to (costless) migration, with urban residents enjoying whatever utility level prevails in the rest of the economy (u is exogenous and L endogenous in this case).[9] Use of the closed-city rather than the open-city model to derive comparative static predictions reflects a belief that job immobility, family ties, and other frictions impede movement among urban areas, so that urban utility levels are largely determined without regard to external considerations.[10]

III. Sample and Data

The model's predictions were tested using data from a sample of "urbanized areas," as defined in the 1970 Census. By closely approximating the actual built-up portion of the city, the urbanized area is the Census statistical unit which best corresponds to the requirements of the theory. The sample consists of 40 urbanized areas with 1970 populations ranging from 52,000 to 257,000 (a list is available on request). Only those

[5] Equation (2) reflects the assumption that all the land at any given x is available for housing. Since cities with topographical irregularities will not conform to this requirement, none were included in the sample.

[6] The comparative static derivatives of u have the following signs: $\partial u/\partial L < 0$, $\partial u/\partial r_a < 0$, $\partial u/\partial y > 0$, $\partial u/\partial t < 0$.

[7] This fact is a consequence of $\partial p/\partial u < 0$. The remaining results in this paragraph use the inequalities $\partial q/\partial u > 0$, $\partial r/\partial u < 0$, and $\partial S/\partial u < 0$.

[8] Since the model is static, ignoring the time dimension of the urban economy, the adjustment process to the new equilibrium is not specified. It is possible to view the model as implicitly assuming perfect malleability for housing capital, so that adjustment in effect takes place immediately. On the other hand, the comparative static results can be viewed as characterizing the outcome of a long-run adjustment process during which aging buildings are torn down and replaced. A logical problem with the latter view is that the model is being interpreted in a dynamic sense even though producer decision-making has been modeled in a static context. For analysis of a truly dynamic urban model where the longevity of buildings is explicitly recognized, see Brueckner and von Rabenau (1981).

[9] For an open city, the system (1)–(2) is recursive, so that $\bar{x}$ is given directly by (1). The regression equation appropriate for a system of open cities would relate $\bar{x}$ to r_a, y, and t (u would be constant in cross section).

[10] Excessive utility differentials across cities may in reality be eliminated by interurban income variation. The variation in income required to cancel utility differences is found by equating $u(L, r_a, y, t)$ (the utility solution from (1) and (2)) to a constant. This equation implicitly determines y as a function of L, r_a, and t (its partial derivatives are all positive). Note that the resulting compensating variation in y yields correlation between income and the other determinants of $\bar{x}$.

urbanized areas contained within a single, relatively small county were eligible for inclusion in the sample. This restriction, which explains the relatively small populations of the sample areas, arose from the desire to accurately measure the value of agricultural land immediately adjacent to the built-up part of the city (farmland value data are available at the county level).

Although the comparative static results in (3) are in terms of $\bar{x}$, the sign predictions are identical for the total land area variable $A \equiv \pi\bar{x}^2$, which is used as the dependent variable in the regressions. A is measured by the 1970 spatial size of the urbanized area in square miles. The population variable L is represented by 1970 urbanized area population, while r_a is represented by the 1969 median agricultural land value per acre for the county containing the urbanized area. The income variable y is represented by a measure of average household income similar to median family income.[11] Since no direct measure of commuting cost per mile exists, the estimation was repeated using two different proxies for t, both of which reflect the belief that high auto usage indicates a low commuting cost per mile. The first of the variables, *TRANSIT*, equals the percentage of commuters using public transit in 1970. Since bus transit, the predominant public transit mode in small and medium-size cities, has a high time cost per mile, a high value of the variable *TRANSIT* would indicate a high value of t.[12] The other commuting cost proxy, *AUTOS*, equals the percentage of 1970 households owning one or more automobiles. Other things (especially income) being equal, a high value of *AUTOS* would be associated with ease of automobile usage (reflecting low congestion levels) and would indicate a low value of t.

IV. Empirical Results

Since the theory provides no guidance as to the functional form of the estimating equation, the empirical work makes use of the flexible non-linear specification of Box and Cox (1964). Table 1 shows the maximum likelihood parameter estimates for the two equations based on the different commuting cost proxies (t-statistics are in parentheses). The optimal value of the functional form parameter λ for each equation equals 0.53, indicating that a square-root transformation of the variables is approximately correct. The signs of the other estimated coefficients in table 1 largely confirm the predictions of the theory. In both equations, the population coefficient is significantly positive, the agricultural rent coefficient is significantly negative, and the income coefficient is significantly positive, as predicted. Thus, as predicted by the theory, city area is an increasing function of population and income, and a decreasing function of agricultural rent. Although the estimated coefficients of the commuting cost proxies do have the correct signs, neither coefficient is significantly different from zero (recall that cities with high proportions of public transit users are presumed to have high values of t and thus small areas while cities with a high incidence of auto ownership presumably have low t's and large areas). The poor performance of the commuting cost variables may reflect either of two possibilities: first, the proxies may in fact do a poor job of capturing actual commuting cost differences; second, the proxies may be correlated with commuting costs, but their small range of variation within the sample may prevent the emergence of precise estimates.[13] Note finally the high R^2 values for both equations.

The 95% confidence intervals for λ in the *TRANSIT* and *AUTOS* equations are, respectively, [0.04, 0.99] and [0.04, 1.00]. Since the linear specification ($\lambda = 1$) is covered by the latter interval and lies just outside the former, it is of interest to consider this specification, which cannot be decisively rejected on the basis of the data. The linear results are presented in table 2, and inspection shows that the estimates are qualitatively identical to those in table 1. To grasp the implications of the estimated magnitudes of the coefficients, consider the elasticities of area A with respect to the significant exogenous variables, which are presented in table 3 (the elasticities are evaluated at sample means). The elasticities show that a 1% increase in L leads to a slightly greater than 1% increase in A, that a 1% increase in r_a reduces A by approximately 1/4%, and that a 1% increase in y increases A by about 1-1/2%. It is interesting to note that while the theory predicts that the elasticity of $\bar{x}$ with respect to L is less than unity, the elasticity of A with respect to L (which is exactly twice as large) may in fact be greater than one, in conformance with the empirical results.

V. Conclusion

The results of this paper justify a dispassionate view of urban sprawl. By showing that urban spatial area is related to population, income, and agricultural rent in

[11] The population of the city not living in group quarters (e.g., prisons and fraternities) was multiplied by per capita income, and the resulting figure was divided by the number of households in the urban area (shown in the Census as the number of occupied housing units).

[12] Note that a high value of *TRANSIT* may indicate that commuters are relying on bus transit to escape auto traffic congestion. In other words, *TRANSIT* reflects a mode choice decision whose outcome may depend on the prevailing level of congestion.

[13] On the belief that old cities have road networks less suited to automobile commuting, a city age variable equal to the number of decades since the central city of the urbanized area attained a population of 50,000 was substituted as a proxy for t. The estimated coefficient of this variable was also insignificant.

TABLE 1.—MAXIMUM LIKELIHOOD ESTIMATES

Constant	L	r_a	y	*TRANSIT*	*AUTOS*	λ	R^2
− 16.71150[a]	0.01554[a]	−0.07150[a]	0.07908[a]	−0.04670	—	.53	.7760
(−3.046)	(9.043)	(−2.856)	(3.233)	(−0.1979)			
−18.71665	0.01539[a]	−0.07054[a]	0.07905[a]	—	0.11168	.53	.7760
(−1.309)	(9.158)	(−2.736)	(3.230)		(0.1573)		

Sources: *A, L, y, TRANSIT, AUTOS—1972 County and City Data Book* (U.S. Census Bureau); r_a—*U.S. Census of Agriculture*, 1969.
Note: Variable definitions and units of measurement:
Dependent variable (A)—land area of urbanized area in square miles
L—population in actual numbers
r_a—median agricultural land value in dollars for county containing urbanized area
y—household income in dollars (for variable definition, see footnote 4)
TRANSIT—percentage of commuters using public transit
AUTOS—percentage of households owning one or more automobiles.
[a]Coefficient significant at the 5% level, two-tailed test.

TABLE 2.—LINEAR ESTIMATES

Constant	L	r_a	y	*TRANSIT*	*AUTOS*	R^2
− 41.07232[a]	0.00041[a]	−0.03028[a]	0.00620[a]	−0.24440	—	.7982
(−2.277)	(10.030)	(−3.090)	(3.033)	(−0.4056)		
−63.46913	0.00040[a]	−0.02888[a]	0.00624[a]	—	0.24746	.7985
(−1.244)	(9.876)	(−2.888)	(3.050)		(0.4604)	

[a]Coefficient significant at the 5% level, two-tailed test.

TABLE 3.—ELASTICITIES FROM LINEAR EQUATIONS[a]

	L	r_a	y
TRANSIT equation	1.097	−0.234	1.497
AUTOS equation	1.086	−0.231	1.496

[a]Evaluated at sample means.

the manner predicted by the model, the empirical results suggest that sprawl is the result of an orderly market process rather than a symptom of an economic system out of control. In this context, it is interesting to note that by demonstrating the negative impact of agricultural rent on urban size, the empirical results undermine the sprawl critic's claim that the transfer of farmland to urban uses represents the "waste" of a valuable resource. By showing that high-quality, high-priced farmland is more resistant to urban expansion than poor-quality land, the empirical results establish that the land market balances the gains and losses from urban sprawl, restricting spatial growth when the process consumes a valuable resource.

As well as providing a measured view of urban sprawl, the results of this paper constitute further evidence of the empirical robustness of the Muth-Mills urban model. Although the model's excellent performance in predicting the internal structure of cities is widely recognized, the present results show its ability to successfully explain intercity differences in urban spatial structure. This finding should further increase our confidence in the model as a tool for policy analysis.

REFERENCES

Box, G. E. P., and D. R. Cox, "An Analysis of Transformations," *Journal of the Royal Statistical Society*, Series B 26 (Apr. 1964), 211–243.

Brueckner, Jan K., and Burkhard von Rabenau, "Dynamics of Land-Use for a Closed City," *Regional Science and Urban Economics* 11 (Feb. 1981), 1–17.

Bryant, William R., and Howard E. Conklin, "New Farmland Preservation Programs in New York," *Journal of the American Institute of Planners* 41 (Nov. 1975), 390–396.

Mills, Edwin S., *Urban Economics* (Glenview, Ill.: Scott Foresman, 1972).

Muth, Richard F., *Cities and Housing* (Chicago: University of Chicago Press, 1969).

Wheaton, William C., "A Comparative Static Analysis of Urban Spatial Structure," *Journal of Economic Theory* 9 (Oct. 1974), 223–237.