



**AgEcon** SEARCH  
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search  
<http://ageconsearch.umn.edu>  
[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

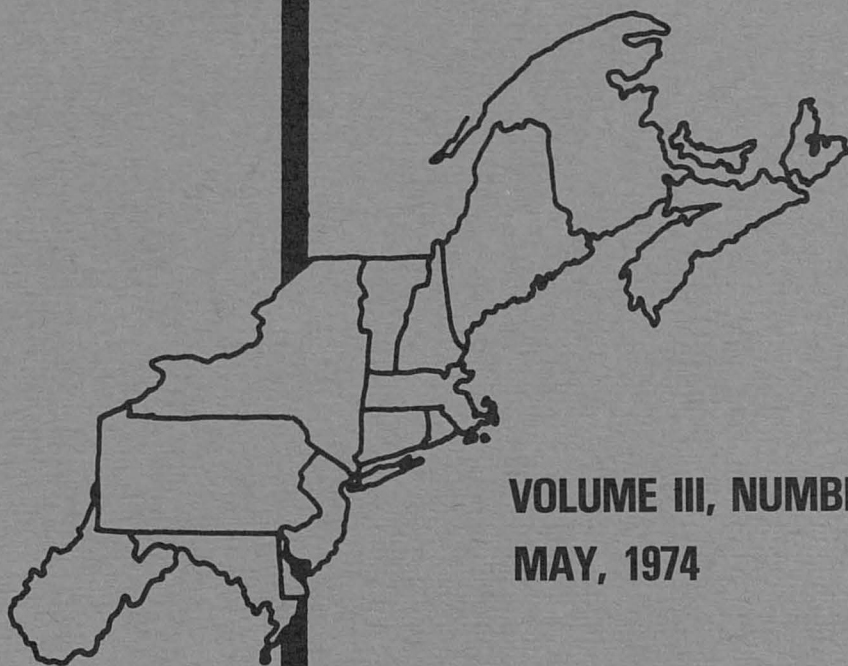
PER. SHELF

GIANNINI FOUNDATION OF  
AGRICULTURAL ECONOMICS  
LIBRARY

JUN 26 1974

# **JOURNAL OF THE**

## **Northeastern Agricultural Economics Council**



**VOLUME III, NUMBER I  
MAY, 1974**

FACTORS INFLUENCING LAND VALUES  
IN THE PRESENCE OF SUBURBAN SPRAWL

Bruce E. Lindsay  
Research Assistant  
and

Cleve E. Willis  
Assistant Professor of Resource Economics and Econometrics  
Department of Agricultural and Food Economics  
University of Massachusetts  
Amherst, Massachusetts

The spread of suburbs into previously rural areas has become commonplace in the United States. A rather striking aspect of this phenomenon has been the discontinuity which results. This aspect is often manifest in a haphazard mixture of unused and densely settled areas which has been described as "sprawl". A more useful definition of suburban sprawl, its causes, and its consequences, is provided below in order to introduce the econometric objectives of this paper.

Although sprawl is seldom adequately defined and often refers to different types of phenomena, it will be referred to here as comprising regions of essentially urban nature located in close juxtaposition to, or surrounded by, essentially undeveloped, rural, or agricultural lands. Sprawl may take one of three major forms:<sup>1/</sup> low density continuous development, ribbon development, and leapfrog development. The ribbon development refers to a continuous development in belts extending outward from the center with undeveloped areas occupying the zones between these belts. The leapfrog development is the type most often referred to as "sprawl", and is identified by its discontinuous (apparently haphazard) zones of urban development. This type of development generally requires the greatest expenditure for social services.

Perhaps the most obvious cause of sprawl is a physical terrain not suited to continuous development. Related to this factor, of course, is the nature and adequacy of transportation facilities. Another cause often mentioned is the independent decisions made by land purchasers (speculators). Such decisions with respect to future land demand are made individually, non-collusively, and with different expectations, resulting in discontinuous unrelated developments. Other causes include: speculation, imbalanced public regulation affecting the attractiveness of competing areas, and the real property tax (as it influences land settlement).<sup>2/</sup>

---

<sup>1/</sup> This taxonomy is made by [7].

<sup>2/</sup> See [7, pp. 2-5].



The sprawl phenomenon brings with it a number of problems. Perhaps the most important is the increasing per capita cost of public services as sprawl progresses. Local governmental officials generally respond by raising the local property tax. Population expansion into rural areas also contributes to an increasing demand for buildable open space, causing land values for such parcels to rise.<sup>3/</sup> Therefore, low income earners are often excluded from entering these areas because they are least able to afford such a location. Sprawl also contributes to a loss of agricultural land use. Increasing real property taxes, decline of supporting industry, and the attraction to "windfall" profits through land sales, has caused many farmers to discontinue operations.

Clawson suggests that establishment of an effective reporting of market transactions in suburban land would be helpful in providing information about the magnitude of sprawl.<sup>4/</sup> Such a market news reporting technique would keep more individuals advised regarding number of parcels, location of parcels, prices paid, and other terms of sale. Further suggestions include a more effective use of subdivision controls and zoning as a stabilizing force, better use of local real estate taxes to implement plans, and finally a more coordinated use of public services with respect to potential effects upon sprawl.<sup>5/</sup> Recently, research related to the problems of sprawl has centered upon economic models to provide a better understanding of the forces leading to various land valuation structures.<sup>6/</sup> This body of empirical research is subject to a number of limitations. The present analysis develops an econometric model (which is not subject to the same limitations) for the purpose of estimating the influence of various factors presumed to affect land values for buildable open space.<sup>7/</sup> The analysis provides local planners with a stronger basis upon which to place constraints on uneven growth so as to maintain the desirable amenities associated with open space (or for any set of community objectives).

The explanatory variables hypothesized to influence land value (selling price per acre) are parcel size, the distance of parcels to the nearest central city, the population of the town in which the parcel is situated, the number of yearly building permits issued by the community, employment of full-time workers in industries located within the town limits, and average annual income of those employed within these industries.

---

<sup>3/</sup> For present purposes, we conceive of land value in the context of selling prices.

<sup>4/</sup> See [3].

<sup>5/</sup> See [3, pp. 108-110].

<sup>6/</sup> Refer to [2].

<sup>7/</sup> Some limitations are: failure to include a combination of data across spatial and temporal units, a failure to seek the appropriate functional form, and failure to devote adequate attention to statistical problems (especially the existence of multicollinearity).

The econometric model, using the postulated exogenous variables above, is applied to an area typifying the "sprawl" concept mentioned previously.<sup>8/</sup>

The paper will be organized as follows. The second section contains the maintained hypothesis and econometric model used for estimation purposes. The estimation procedures employed and the empirical results are discussed in the third section. The final section consists of some concluding comments.

## II

In the previous section, a listing of the explanatory variables initially postulated to influence land value (selling price per acre) is provided. The *a priori* hypotheses regarding the direction of influence of each of these exogenous variables are treated individually below.

Following [4] it is hypothesized that parcel size and selling price of buildable open space land are inversely related. Traditionally, the land market faces a situation in which a greater number of buyers compete for smaller-sized parcels. Such a demand situation results in higher "per acre" selling prices for the smaller parcels than those for larger land lots. Similarly, the work of [1] suggests that land values decline as distances from the center of a city increase. Implicitly, this is based upon the assumption that parcels located further from the central city result in greater transportation costs, reflecting a decline in land values related to decreasing accessibility. Therefore, it is maintained that distance<sup>9/</sup> of a parcel from the nearest central city bears an inverse relationship with land values.<sup>10/</sup> It is generally thought that any population expansion into peripheral areas results in an increased demand for the essentially fixed supply of land. Therefore, the assumption is that community population increases in the interface lead to higher land prices.<sup>11/</sup> A direct relation is assumed for building permits, since this variable serves as a proxy for demand. The number of employees of industries located in a particular community

---

<sup>8/</sup> The subject area is contained in the Standard Metropolitan Statistical Area of Chicopee-Holyoke-Springfield, Massachusetts (Figure 1). For more in depth information, refer to [9].

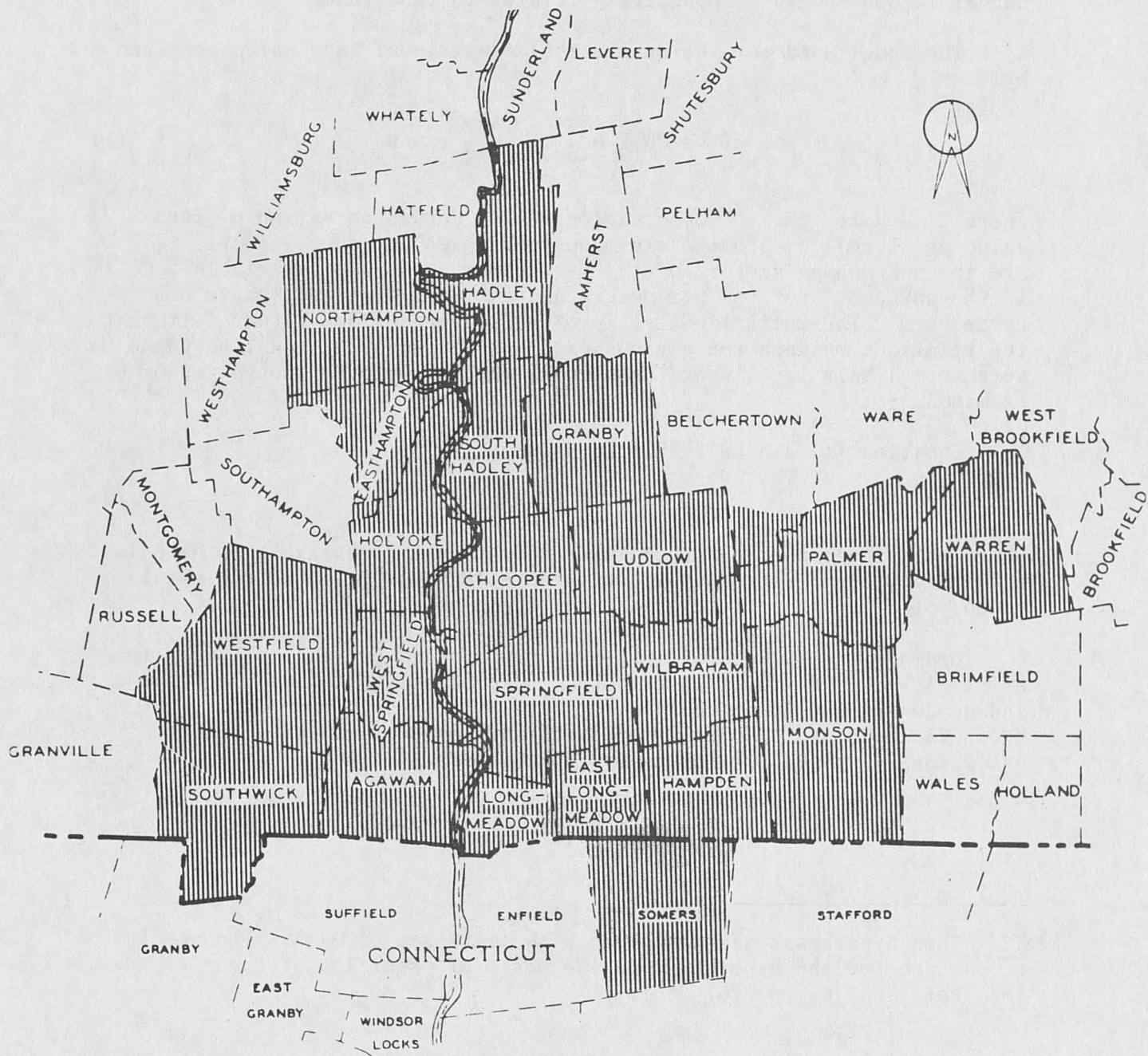
<sup>9/</sup> See Appendix for a treatment of the various distance proxies tested and the rationale for selecting the logarithmic form of distance to the nearest central city rather than the other weighted average distance proxies.

<sup>10/</sup> For further support of this hypothesis see [4]; also, [11] and [12].

<sup>11/</sup> For a study indirectly relating to this variable, see [12, p. 62]. Also, the hypothesis is given further support by [10, p. 248].



Figure 1  
Springfield-Chicopee-Holyoke, Massachusetts SMSA\*



\*Source: Department of Commerce and Development, Commonwealth of Massachusetts.

represents a proxy for those forces operating to attract population to the region, and therefore, employment and land values are presumed to be directly related. We also assume that average annual income reflects the relative purchasing power of the area residents, such that average annual income should be positively related to land value.

The model used for the statistical analysis of land value is given by:

$$Z_i = C W_{i2}^{\beta_2} W_{i3}^{\beta_3} \dots W_{i7}^{\beta_7} \exp u_i \quad (1)$$

where  $Z_i$  denotes the  $i^{\text{th}}$  observation on the dependent variable (land value per acre),  $W_j$  are the exogenous variables introduced above,  $\beta_j$  are the unknown parameters of the respective explanatory variables,  $C$  is the unknown intercept parameter, and  $\exp u$  is the disturbance or error term. The multiplicative form was selected since it was felt that the relations between the various explanatory variables and land value were more likely log-linear than linear (additive) over the ranges of  $W_j$  examined.<sup>12/</sup>

Equation (1) can be rewritten as:

$$Y = X\beta + u, \quad (2)$$

where  $Y$  is the  $n \times 1$  vector of  $\ln Z$ ,  $X$  is an  $n \times 7$  matrix of  $\ln W$ ,  $\beta$  is a  $7 \times 1$  vector of regression coefficients ( $\beta_1 = \ln C$ ),  $u$  is the  $n \times 1$  vector of disturbance terms, and  $n$  is the number of observations.

Ordinary Least Squares (OLS) was employed to provide estimates (b) of  $\beta$ . Assuming  $X_j$  to be distributed independently of  $u$  and the  $u$  to be independently and identically distributed with a zero mean and an unknown variance  $\sigma^2$ , OLS yields unbiased and asymptotically efficient estimates<sup>13/</sup> of the parameters  $\beta_j$ , i.e.

$$\hat{Y}_i = b_1 + b_2 X_{i2} \dots + b_k X_{ik},$$

<sup>12/</sup> This hypothesis was tested on a separate set of data. The results supported the hypothesis on the basis of precision of the estimates.

<sup>13/</sup> Refer to [6, pp. 267-268].



where  $\hat{Y}$  is an unbiased estimate of  $Y$  given  $X$ ,<sup>14/</sup> and  $k$  is the number of parameters to be estimated.

The estimates (b) are based on cross-sectional data for two separate time periods. In order to include a measure of the influence of time, a dummy variable was introduced and the cross section data obtained for a 1965 period was combined with the 1970 data. For this formulation, equation (1) is replaced by (3),

$$Z_i = C W_{i2}^{\beta_2} \dots W_{i7}^{\beta_7} \exp \beta_8 X_{i8} \exp u_i, \quad (3)$$

where

$$X_8 = \begin{cases} 1, & \text{if the observation was taken in 1965,} \\ 0, & \text{otherwise} \end{cases}^{15/}$$

$\beta_j$  and  $C$  are the unknown parameters, and  $\exp u$  is the disturbance. Operationally, of course, the parameters  $C$  and  $\beta_j$  are estimated using the double logarithmic transformation of (3),

$$Y_i = \beta_1 + \beta_2 X_{i2} + \dots + \beta_8 X_{i8} + u_i \quad (4)$$

### III

Of the seven factors hypothesized to influence land values, the data regarding population and number of building permits was inadequate to suggest a significant influence on land values. These variables were

---

<sup>14/</sup> This relationship does not imply, however, that  $\hat{Z} = \exp(\hat{Y})$  is unbiased given  $W$ . Note also that the OLS estimation of models of the form above yields conditional medians rather than conditional means; i.e. under the model  $Z = C W^b e^u$ , where  $u \sim N(0, \sigma^2)$ ,  $E(\exp u) = \exp \frac{\sigma^2}{2}$  and median  $M(e^u) = 1$ . The implication is that  $E(Z|W) =$

$C W^b E(e^u) = C W^b \exp \frac{\sigma^2}{2}$ . Similarly the conditional median is given

by  $M(Z|W) = C W^b m(e^u) = C \exp \ln W^b$  and will of course lie below  $E(Z|W)$ . Since the difference is generally minor (roughly .00052 in our case) we have not made the corrections.

<sup>15/</sup> To be sure, this formulation considers the change over time as occurring through the intercept only. At the sacrifice of degrees of freedom, of course, one can also use dummy variables to permit slope changes.



not significant, even at very low levels of confidence and after distortions stemming from conditions of multicollinearity were removed by the deletion of exogenous variables. Furthermore, their inclusion added to estimation problems by increasing the collinearity among the exogenous variables. Thus, zero restrictions were placed on these two variables. The results of the estimations with these zero restrictions are set out in Table I, where:

- $X_2$  denotes the log of parcel size in acres,
- $X_3$  denotes the log of distance to the nearest central city in miles,
- $X_4$  denotes the log of the number of employees for industries within the town limits,
- $X_5$  is the log of average annual income in dollars, and
- $X_6$  is the dummy variable.

Standard errors are in parentheses and one asterisk denotes significance at the 95% level of confidence, while two asterisks denote significance at the 90% level of confidence.

For the estimation using the 1965 data, all variables were significant at the 95% level except distance ( $X_3$ ). Income ( $X_5$ ), however, displayed a direction of influence contradictory to our prior thinking, perhaps reflecting a correlation between the demand for industrial land and locations of low income earners. The fit utilizing the 1970 observations was substantially weaker. Only distance was significant and the percent of the variation in  $Y$  which is explained by the linear influence of the combined exogenous variables is much lower (20 as compared with 54 for the 1965 formulation).<sup>16/</sup> For the combination of cross-sectional data, size, distance, and the influence of the time differential ( $X_6$ ) were significant. The sign of  $b_6$  reflects the fact that, *ceteris paribus*, land values have risen over the period 1965 to 1970.

The number of employees ( $X_4$ ) and average annual income ( $X_5$ ) were highly correlated,<sup>17/</sup> however, suggesting the existence of multicollinearity. The test suggested by [5] supports this conclusion for  $X_4$  and  $X_5$ , yielding values of the  $F$  statistic for  $X_4$  and  $X_5$  of 48.2 and 36.6, respectively, for the estimation of equation (2).

In order to reduce the disruptive effects<sup>18/</sup> of the multicollinear condition, a form of Restricted Least Squares<sup>19/</sup> (RLS) was employed.

---

<sup>16/</sup> Another example of a declining coefficient of determination during ensuing time periods is found in [12, pp. 68-69].

<sup>17/</sup> For example, the simple correlation coefficient of  $X_4$  and  $X_5$  was 0.96.

<sup>18/</sup> The effects are that the precision of estimation falls, investigators are led to delete variables incorrectly on the basis of conventional tests of significance, and estimates become particularly sensitive to particular sets of sample data.

<sup>19/</sup> For a more elaborate discussion, see [8].

Table I  
Estimated Parameters for Formulations (2) and (4)  
Under OLS and RLS Estimation

Estimation Method	Time Periods	Estimators of $\beta$						$R^2$	Observations
		$b_1 = \ln C$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$		
OLS	1965	45.6794	-.4157* (.0691)	-.0454 (.2009)	.9465* (.3999)	-5.0506* (1.8952)		.54	52
	1970	9.5267	-.2127 (.1963)	-.9437* (.4242)	-.0479 (.2818)	.2167 (1.4007)		.20	41
	1965 1970 combined	16.5042	-.3351* (.0896)	-.5521* (.2005)	.0613 (.1879)	-.6681 (.9108)	-.9512* (.4787)	.32	93
RLS	1965	22.8427	-.4120* (.0830)	-.3150* (.1220)	.2169** (.1475)	-1.6713* (.9130)		.49	52
	1970	15.7857	-.1860 (.1760)	-.7907* (.3690)	.2169** (.1475)	-.7415 (1.082)		.19	41
	1965 1970 combined	21.3010	-.3318* (.0945)	-.4882* (.1760)	.2169** (.1475)	-1.3486* (.3330)	-1.2482* (.2860)	.33	93



Using extraneous information in the form of a third set of independent cross-sectional data for another time period, an exact restriction was placed on the value of  $b_4$ , the coefficient on the number of employees. This extraneous information provided a coefficient of  $b_4$  equal to 0.2169 which was significant at the 90% confidence level, and this value is imposed in the estimation procedure<sup>20/</sup> for the sample data. The results of this estimation are, again, found in Table I.

<sup>20/</sup> The RLS approach adopted originates from [8, pp. 164-165]. It assumes the usual model,

$$Y = X\alpha + u,$$

where  $X$ ,  $\alpha$ , and  $a$  (the estimates of  $\alpha$ ) are partitioned as:

$$X = \begin{bmatrix} X_r & X_{k-r} \end{bmatrix}, \quad \alpha = \begin{bmatrix} \alpha_r \\ \alpha_{k-r} \end{bmatrix}, \quad \text{and } a = \begin{bmatrix} a_r \\ a_{k-r} \end{bmatrix}$$

and where  $X_r$  is the  $n \times r$  sub-matrix consisting of the first  $r$  columns of  $X$ ,  $X_{k-r}$  is the sub-matrix of the  $k-r$  columns of  $X$ ,  $\alpha_r$  and  $a_r$  are similarly partitioned, and  $k$  is the number of parameters to be estimated. In our case, of course,  $k-r$  is one. OLS estimators,

$$a_r = (X_r' X_r)^{-1} X_r' Y^*,$$

are found, where,

$$Y^* = Y - X_{k-r} a_{k-r},$$

whose covariance matrix is given by

$$\text{var}(a_r) = \sigma_u^2 (X_r' X_r)^{-1} + (X_r' X_r)^{-1} X_r' X_{k-r} V X_{k-r}' X_r (X_r' X_r)^{-1}.$$

In general,  $V$  is the covariance matrix,

$$V = E(a_{k-r} - \alpha_{k-r})(a_{k-r} - \alpha_{k-r})'.$$

In our case, of course, it is a scalar which is provided by the regression on the extraneous information. The covariance matrix,  $\text{var}(a_r)$  is adjusted to recognize the uncertainty attached to the coefficients ( $b_4$  in our case) used in the adjustment of  $Y$ .

Most of the  $t$  statistics<sup>21/</sup> for testing significance of parameters are greater for the RLS estimation than for the OLS approach, suggesting that the former has improved the precision of the estimators. Similarly, the RLS estimation has reduced standard errors relative to regression coefficients to the extent that estimates for the 1965 period and  $a_5$  for the combination of observations become significant at the 95% level. Indeed, all estimators were significant in the RLS estimation where the data were aggregated, with only the sign associated with average annual income ( $a_5$ ) differing from the prior hypotheses of the investigators. One possible explanation is that the demand for industrial land is often higher for areas in which low income earners are located.

Since the parameters were estimated in double-logarithmic form, the estimates can be interpreted as elasticities. For example, focusing on the lower row of Table I, the results suggest the expectation that for a 1% increase in parcel size, land values will decline by .33% ( $b_2$ ); for a 1% increase in distance to the nearest central city, land values will decline by .49%; for a 1% increase in number of employees, land values will increase by .22%; for a 1% increase in average annual income, land values will fall by 1.35%. The coefficient  $b_6$ , again, simply reflects the intercept shift attributable to the higher land values in 1970 as compared to those in the 1965 sample.

#### IV

This analysis has sought to develop an econometric model to estimate the influence of various factors on land values for buildable open space for an area displaying the characteristics and problems associated with suburban sprawl. This was accomplished by applying a RLS procedure to the double-logarithmic transformation of a multiplicative model which incorporates two cross-sectional sets of data in a dummy variable context. The restrictions were formed by extraneous information procedures. Empirical investigations of land values in which the inherent multicollinearity (which surely exists in most cases) is recognized and its adverse effects are adjusted for do not appear to be commonplace. The search for appropriate functional form and proper distance proxy is also a rarity in this type of empirical research.

The coefficients for parcel size ( $b_2$ ) and distance ( $b_3$ ), both negatively signed, suggest a possible basis for zoning restrictions, e.g. special stipulations on minimum and/or maximum size and locational classifications. Employment was found to be significant and positive. This suggests that rising employment levels have an upward pressure upon land values. Town officials normally strive to increase the amount of

---

<sup>21/</sup> These, of course, are calculated by dividing the regression coefficients of Table I by their respective standard errors.



employment in their jurisdictions. However, evidence suggests that such a policy contributes to increasing land values, with possible adverse effects given the particular set of community goals and priorities, possibly causing a detrimental impact to a community. Further,  $b_5$  suggests a direct relation between low income workers and industrial demand for land. Finally,  $b_6$  reflects changes in the level of economic activity, during past and recent years, that affect the usefulness of land and the ability of land procurers.

#### References

1. Alonso, W., "A Theory of the Urban Land Market", Papers and Proceedings of the Regional Science Association, Volume 6, 1960.
2. Bahl, Roy W., "A Land Speculation Model: The Role of the Property Tax as a Constraint to Urban Sprawl", Journal of Regional Science, Winter 1968.
3. Clawson, Marion, "Urban Sprawl and Speculation in Suburban Land", Land Economics, May 1962.
4. Clonts, Howard A., Jr., "Influence of Urbanization on Land Values at the Urban Periphery", Land Economics, November 1970.
5. Farrar, D. E. and R. R. Glauber, "Multicollinearity in Regression Analysis: The Problem Re-visited", Review of Economics and Statistics, 49:1967.
6. Goldberger, Arthur S., Econometric Theory, New York: John Wiley and Sons, 1964.
7. Harvey, R. O. and W. A. V. Clark, "The Nature and Economics of Urban Sprawl", Land Economics, February 1965.
8. Johnston, J., Econometric Methods, New York: McGraw-Hill, 1972.
9. Lindsay, B. E., "The Influence of Selected Variables on Land Values in the Rural-Urban Interface", Unpublished M.S. Thesis, Department of Agricultural and Food Economics, University of Massachusetts, Amherst, Massachusetts, 1973.
10. Mohring, Herbert, "Land Values and the Measurement of Highway Benefits", Journal of Political Economy, June 1961.
11. Pendleton, W. C., "The Valuation of Accessibility", Ph.D. Dissertation, University of Chicago, 1962.
12. Yeates, Maurice H., "Some Factors Affecting the Spatial Distribution of Chicago Land Values, 1910-1960", Economic Geography, January 1965.

# Appendix: Weighted Distance Proxy

In establishing a weighted distance proxy for the study, various models were tested to establish the best fit, including the following, where:

Y = land value per acre  
 $X_i$  = distance to city i;  $i = 1, \dots, 5$   
 $X_j$  = distance to nearest city  
 $N_j$  = population of city j  
a = Y-intercept

Model I:

$$Y = a + b_i \sum_i X_i$$

Model II:

$$Y = a + b_j X_j$$

Model III:

$$Y = a + b_j X_j + b_i \sum_{i \neq j} X_i$$

Model IV:

$$Y = a + b_j N_j X_j$$

Model V:

$$Y = a + b_j N_j X_j + b_i \sum_{i \neq j} N_i X_i$$

Model VI:

$$Y = a + b_j X_j^2$$

Model VII:

$$Y = a + b_j X_j^2 + b_i \sum_{i \neq j} X_i^2$$

Model VIII:

$$Y = a X_j^{b_j} \text{ or } \ln Y = \ln a + b_j \ln X_j$$



Model IX:

$$Y = a \pi_i X_i^{b_i} \text{ or } \ln Y = \ln a + \sum_i b_i \ln X_i$$

Data for 1952 was used for the testing since it was not used in the estimation of (2) or (4). Least-squares regression was used, where Models I through VII are of the linear additive case, while Models VIII and IX are multiplicative cases. Results of the testing are presented in Table II.

Model VIII, the logarithmic distance to the nearest city, gives the best fit. Accordingly, this proxy was selected to express distance as a variable for the analysis.

Table II  
Empirical Results for Weighted Distance Proxies

Models	Y-intercept	b		R <sup>2</sup>	Number of Observations
I	511.81	-19.6219 (13.4879)		.06	38
II	2506.62	-143.4999* (67.2936)		.12	38
III	264.97	-138.5906* (65.8892)	.5237 (.4154)	.16	38
IV	3133.92	-.4653 (.3268)		.06	38
V	243.81	-.4812 (.3341)	-.0999 (.4426)	.06	38
VI	1321.61	-8.0624 (4.4553)		.09	38
VII	529.19	-18.9410 (9.7541)	.4265 (.3421)	.13	38
VIII	9.15	-1.1852* (.4078)		.20	38
IX	11.20	-2.0053* (.7778)		.16	38

(1) \* denotes significance at the 95% level of confidence.

(2) Standard error of estimated parameter in ( ).