Journal of Agricultural and Applied Economics, 39,3(December 2007):701–717

The Impact of an Urban Growth Boundary on Land Development in Knox County, Tennessee: A Comparison of Two-Stage Probit Least Squares and Multilayer Neural Network Models

Seong-Hoon Cho, Olufemi A. Omitaomu, Neelam C. Poudyal, and David B. Eastwood

The impact of an urban growth boundary (UGB) on land development in Knox County, TN is estimated via two-stage probit and neural-network models. The insignificance of UGB variable in the two-stage probit model and more visible development patterns in the western part of Knoxville and the neighboring town of Farragut during the post-UGB period in both models suggest that the UGB has not curtailed urban sprawl. Although the network model is found to be a viable alternative to more conventional discrete choice approach for improving the predictability of land development, it is at the cost of evaluating marginal effects.

Key Words: land development, multilayer neural network, two-stage probit least squares

JEL Classifications: C35, R14

During the 15-year period from 1982 to 1997, the developed area of Tennessee increased by 58%, from 1.5 to 2.4 million acres. This increase was the seventh largest among all 50 states and was much greater than the national rate of 34% for the time period. It was largely due to the conversion of land from agriculture and forests to residential use. Of the 870,000 acres developed between 1982 and 1997, about 340,000 acres, or 39%, was converted from prime farmland, including cropland and pastureland. Tennessee nearly doubled the rate of development of farmland, forests, and other open space during the 1990s, from 46,000 acres per year between 1982 and 1992 to 80,000 acres per year between 1992 and 1997 (NRCS).

One response to the rapid land development in Tennessee was the creation of the Growth Policy Act, implemented in 1998, requiring all counties and the cities within them to collaborate in defining urban growth boundaries (UGBs). Each county, and towns and the cities within the respective county, identified three classifications of land: rural areas, UGBs, and planned growth areas (PGAs). Rural areas included land to be preserved for farming, recreation, and other

Seong-Hoon Cho, Neelam C. Poudyal, and David B. Eastwood are, respectively, assistant professor, graduate research assistant, and professor emeritus at Department of Agricultural Economics, the University of Tennessee. Olufemi A. Omitaomu is with the Computational Sciences and Engineering Division at Oak Ridge National Laboratory.

nonurban uses. The land within UGB was expected to be reasonably compact but adequate to accommodate all of the city's expected growth for the next 20 years. PGAs were supposed to be large enough to accommodate growth expected to occur in unincorporated areas over the next 20 years (MPC 2001). The deadline for completing and approving all plans was July 1, 2001 (TACIR).

The City of Knoxville adopted a UGB in January, 2001. The UGB in Knoxville did not provide public facilities and subsidies inside the UGB to encourage development. Further, the UGB of Knoxville was drawn in response not only to concerns over growth management, but also to the fallout from local annexation battles. As a result of the local annexation battles, the Knoxville city government acquired the right to annex land parcels within the UGB boundary without the consent of landowners, which was another important provision of Knoxville's UGB. Cho et al. conducted the only published analysis of the impacts of UGBs in Tennessee. They estimated the effects of UGB on land development in Knox County, TN using a heteroscedastic probit model and concluded that with combined effects of increased land development within the city boundary and decreased development within the UGB and the neighboring town of Farragut after the implementation of UGB, the UGB of Knox County has been successful in urban revitalization within the city boundary and discouraging urban sprawl. However, three issues may have affected their results. They are addressed in the models described below and estimated with the same data set which lead to different inferences about the effects of UGB.

First, their study does not compare spatial distributions of predicted land development before and after the UGB. The change in spatial distribution for the predicted development between the pre-UGB and post-UGB periods is an important benchmark for the evaluation of the boundary.

Second, simultaneity needs to be addressed. Specifically, the appraised value of land parcel, treated as an exogenous variable in Cho et al., may be considered an endogenous variable. If the conversion rate increases in a certain area, the county assessor will likely adjust the appraised value of nearby land parcels, as any improvement on the parcel has spillover effects. Statistically, failure to accommodate the potential endogeneity between land development and land value produces biased and inconsistent parameter estimates, and thereby confounds the estimated effects of UGB and other socioeconomic, environmental, and jurisdictional characteristics on land development and land value.

Third, while the discrete choice model used in Cho et al. has been widely adopted for residential land–use change, the forecasting ability of the discrete choice model has been challenged. For instance, Landis pointed out that discrete choice models are constructed as a series of sequential and interrelated mathematical relationships. As a result, estimation and projection errors tend to propagate throughout the models. Missing data, outliers, heteroscedasticity of the error terms, and autocorrelation may lead to over- and under-prediction problems. Another major drawback of the discrete choice model is that a numerical or analytical solution to the system of equations must be obtained, limiting the level of complexity that may practically be built into the models (Parker et al.).

Artificial intelligence (AI) modeling has the potential for spatial data processing and forecasting (Smith). AI technologies, such as artificial neural networks, have been incorporated into land use change models to improve the forecasting ability (e.g., Balling et al.; Mann and Benwell). Yeh and Li used neural networks to simplify model structures and facilitate the estimation of parameters that can be used in land use planning for simulating alternative development patterns based on different planning objectives. Most recently, Lin et al. used a neural network to model urban sprawl and land use change. A common finding across these studies is that neural networks yield better predictions than traditional discrete choice models.

Estimated impacts of UGB on land development in Knox County, TN using two-stage probit and neural-network models are described and evaluated. We explore the application of neural networks in the modeling of land development and compare the predictability of the model with the discrete choice model. Land development at the parcel level is estimated as a functional form of individual parcel characteristics (distance and physical factors), neighborhood characteristics (socioeconomic factors at the census–block group level), and urban growth boundary as a land use policy. The endogeneity of land development and land value is accommodated with a simultaneous equations model for discrete–continuous endogenous variables that is modified for discrete choice. In addition, a multilayer perceptron (MLP) network is developed for the neural network model. Forecasts from the two models are compared. Mappings of predicted land development before and after the UGB from both models are used to illustrate the difference between the models and the impacts of UGB.

Modeling Techniques

In this section, we present an overview of the methods used in this paper. More in-depth descriptions are available in the references.

A Simultaneous Equations Model with Discrete–Continuous Endogenous Variables

We estimated a simultaneous equations model with an endogenous binary variable for development ($y_1 = 1$ for developed, 0 otherwise) and an endogenous continuous variable for land value (y_2). The model is characterized by the structural equations

$$(1) \quad y_1^* = \gamma_1 \log y_2 + \beta_{1J} X_{12} + \beta_{1k} X_{11} + u_1,$$

$$(2) \quad \log y_2 = \gamma_2 y_1^* + \beta_{2J} X_{12} + \beta_{2k} X_{22} + u_2,$$

where y_1^* is a latent variable corresponding to y_1, X_{12} is a vector of exogenous variables used in both equations, X_{11} and X_{22} are vectors of exogenous variables used exclusively in (1) and (2), respectively, β_{1J}, β_{2J}, β_{1k}, β_{2k} are conformable parameter vectors, γ_1 and γ_2 are scalar parameters, and the error terms u_1 and u_2 are assumed to be distributed as bivariate normal with zero means and a finite covariance matrix, with the variance of u_1 normalized at unity because land development is binary. The corresponding reduced-form equations are

$$(3) \quad y_1^* = \Pi'_1 X + v_1,$$

$$(4) \quad \log y_2 = \Pi'_2 X + v_2,$$

where, given the bivariate normality of u_1 and u_2, the error terms v_1 and v_2 are distributed as bivariate normal with zero means and covariance matrix

$$(5) \quad \Omega = \begin{bmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{bmatrix}.$$

Note that the elements $\omega^2{}_1$, $\omega^2{}_2$ and ω_{12} of the covariance matrix Ω in (5), as well as the reduced-form parameter vectors Π_1 and Π_2 in (3) and (4), are all functions of the structural parameters in (1) and (2). The observed value for development (y_1) corresponds to its latent counterpart ($y^*{}_1$) such that

$$(6) \quad \begin{aligned} y_1 &= 1 \quad \text{if} \quad y_1^* > 0, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Two-step estimation of the model is discussed by Maddala (Model 3, p. 244). Following the two-step estimation, the CDSIMEQ command in the Stata 9.0 was used to estimate the simultaneous equation system with corrected standard errors (Keshk). As the effects of exogenous variables are not trivial due to interdependence between the endogenous variables and the discrete nature of land development, more insight can be gained by calculating the marginal effects of exogenous variables and predicted land values at the second stage. Marginal effects refer to the effects of explanatory variables on the probability of land development. This is done by differentiating the probability of development with respect to each variable. Note that the UGB is treated as an exogenous variable in this model because the adoption of UGB by Knox County is not a result of political bargaining among the community residents and developers, but rather, a mandatory requirement by the Tennessee Growth Policy Act.

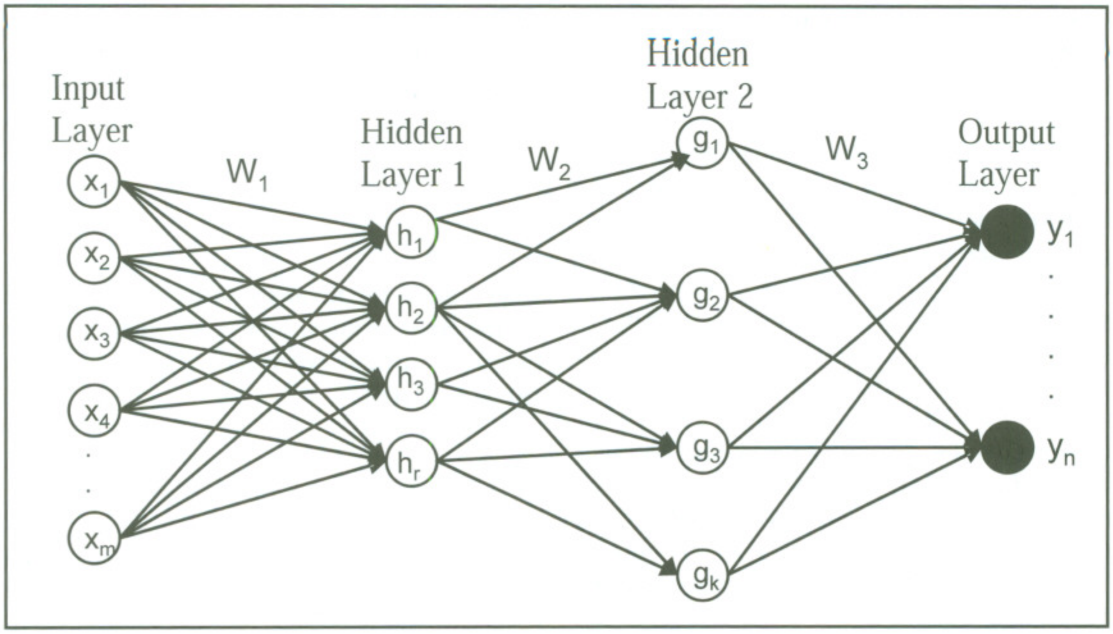

Figure 1. A Multilayer Perceptron Model

Neural Network Model

A neural network model is a data processing system that receives input variables, modifies the variables by a positive or negative weight, and combines these modified variables in order to produce an output. Therefore, this is a two-part process: the first part aggregates the weighted input, and the second part uses an activation function through which the combined input flows. There are several neural network model types depending on the method of aggregating the weights and how the combined input flows.

A multilayer perceptron (MLP) is a feedforward (one-direction) neural network that consists of an input layer, one or more hidden layer(s), and an output layer. The input layer does not perform any computation; but passes the input variables to the first hidden layer. The use of a hidden layer is one approach of approximating a nonlinear function. One or more hidden layers may be used depending on the complexity of the given problem. Theoretically, given enough hidden layers a neural network can approximate any function; hence, a neural network is also called a universal approximator. Implementation starts with a small number of hidden layers (for example, one hidden layer) and layers are added until a desired error goal is reached (e.g., minimizing the absolute error, meeting a number of iterations). An MLP is particularly attractive because the additional layers allow the results of one layer to be further processed, arranged, and put together in order to approximate a complex system, as shown in Figure 1.

We developed an MLP model with input neurons (x_i) of the explanatory variables and binary output neurons (y_i) of development land for parcel *i*. The model is made up of four layers: an input layer, two hidden layers, and an output layer. For this experiment, we used one, two, and three hidden layers but using two hidden layers was found to approximate best. The final model for one of the data sets (described below) uses 18 neurons in the first hidden layer and 9 neurons in the second hidden layer. The final model for the other data set (also described below) uses 22 and 11 neurons in the first and second layers, respectively. The equation that describes the model is defined as

$$(7) \quad \alpha_i^* = \sum_{i=1}^{n} f_i(x_i)w_i,$$

where w_i is the weight vector in each layer, f_i is the activation function (the function used to process the network) of the neurons in each layer, n is the number of neurons in each layer, and α_i^* is the output of the output layer. Note that α_i^* is different from y_i, the binary outcome of actual land development defined in the previous section. The output of the output layer, (α^*_i), is defined as

$$(8) \quad \alpha_i^* = \begin{cases} 1, & \alpha_i \geq \theta \\ 0, & \text{otherwise} \end{cases},$$

where θ is the threshold, and it is equal to 0.5 for this work in order to match the probability structure of probit. Following Nauck, Klawonn, and Kruse, we used a nonlinear (logarithmic sigmoid) activation function that constrains outputs to be between 0 and 1 for the hidden and the output layers. Variables free of measurement units were standardized by using a logarithmic sigmoid activation function, which is defined as

$$(9) \quad \text{logsig}(n) = \frac{1}{(1 + e^{-n})}.$$

The standardization gives equal importance to all inputs and prevents saturation of the sigmoid activation function (an activation function that uses exponential function). Our MLP model used backpropagation with the objective of making the sum of the errors as close to zero as possible. Using an iterative gradient descent process decreases the value of the error function on subsequent tests of the inputs (Haykin; Mehrotra, Mohan, and Ranka).

This algorithm required two passes: a forward pass of the function signal and a backward pass of the error signal to modify the weights. The model was implemented using MATLAB software. There were four steps: initialization, presentation, error computation, and learning. Initialization involved specifying the weights and biases using MATLAB "INITFF" function (MathWorks). The INITFF function used information about the input variables, the activation function, and the size of the layers to compute the weights and biases for use in training the network. The error computation step compared the predicted output to the actual values to obtain errors for the backward pass. The final step, the learning process, consisted of both the backward computation and iteration. In the backward computation, errors were backpropagated through the network and used to adjust the weights to minimize the errors.

The MATLAB "TRAINBPX" function was used to identify data characteristics. This function used adaptive learning and momentum for backpropagation, and its performance depended on the value of the initial learning rate and momentum set *a priori*. The learning rate controlled the step size of the gradient approximation. Momentum, on the other hand, decreased backpropagation sensitivity to minor details in the error surface. It incorporated pseudo-second order information by adding part of the last weight change to the current weight change. The learning rate determined the extent of the change in the weights at each step.[1] After several experiments with different learning rates and momentum combinations, a starting learning rate of 0.01 and initial momentum of 0.5 were used. The criterion to stop the training was either an error goal of 0.5 (the difference between actual output and the predicted output) or a maximum number of 2,000 training iterations. In all experiments the error goal was never reached, so the backward computation step stopped after 2,000 iterations. The model maximum was set to avoid overfitting the training data (Hines; Haykin).[2]

Data

Information about housing parcels within Knox County was gathered from the Knoxville/Knox

[1] If the learning rate is too small (large), the algorithm will take a long time to converge (diverges). The use of momentum, on the other hand, helps avoid local minima. It allows a fraction of the previous weight update to the current one. Therefore, it is often necessary to reduce the learning rate when using a lot of momentum (Haykin).

[2] Overfitting is when a network learns peculiarities in a training (estimation) data set that are not general to the data as a whole. In other words, it learns the noise in the data. Therefore, it does well on the training set but poorly on out-of-sample (validation) data; that is, overfitting causes poor generalization.

County/Knoxville Utilities Board's Geographic Information System (KGIS) and 2000 long form data of census-block groups. The county adopted the UGB system in January 2001. In order to see the impact of the UGB on development, the sample data were divided into two periods, four years before the UGB (1997–2000) and four years after the UGB (2001–2004). The pre-UGB data for the period of 1997–2000 included all the vacant parcels at the start of 1997. Similarly, the post-UGB data for the period of 2001–2004 included all the vacant parcels at the start of 2001. The developed parcels pertained to residential, commercial, and industrial properties that were built during the subperiods. The 2000 long form data at census-block group level were merged with the parcel data. The timing cycle of the census and development of parcel did not match except in 2000. However, given the periodic nature of census taking, census data for 2000 were considered proxies for real-time data for 1997, 1998, 1999, 2001, 2002, 2003, and 2004.

Median housing values reflecting real estate market conditions and average travel times to work reflecting spatial measures of the distance to the employment hub as a neighborhood location at the level of census-block group were included as exogenous variables to capture spatial dependence in both land development and land value. Average per-capita income, unemployment rate, and vacancy rate reflecting socioeconomic characteristics at the census-block group level were included to capture spatial dependence in land development (Downs; Phillips, and Goodstein). Housing density was used to measure the level of development surrounding each lot (Yen and Cho).

Another set of neighborhood variables were dummy variables for high school districts to capture the effects of school quality on land value. School quality has been found to be a determining factor in real estate market values (e.g., Bogart and Cromwell; Hayes and Taylor) and to capture neighborhood properties that have similar characteristics (Basu and Thibodeau). There are 12 high school districts in Knox County, and the town of Farragut has one high school district. Farragut is used as a reference district. High school dummy variables were not included in the land development equation because the high school districts did not seem to have a strong conceptual relation to the land conversion decisions.

In addition, a dummy variable measuring whether the parcel is located within the city of Knoxville was included to capture neighborhood effects of locally provided public service in both land value and land development because each jurisdiction provides a different bundle of public services. A dummy variable indicating UGB area was included to capture the effects of growth plan boundaries in land development and land value.

Distance and physical variables included distance to downtown Knoxville, nearest water body, nearest greenway, nearest railroad, nearest park, nearest interstate highway, and slope of each lot. These distance and physical variables were to capture the effects on land development decisions to the proximity of various amenities and disamenities.

Parcel data included assessed land value, size of the parcel, and construction year. All the assessed values were determined by the Knox County tax assessors' office at the end of 2004. There were 234 census block groups in Knox County. The 2000 census-block group data provided housing density, average travel time to work, unemployment rate, per capita income, and home vacancy rate measures. The boundary data (high school districts, growth plan boundaries including the UGB, and jurisdiction boundaries) were provided by the Knoxville–Knox County Metropolitan Planning Commission (MPC 2006).[3] Environmental feature data, such as shape files of water bodies, parks, and interstate highways, were acquired from 2004 Environmental Systems Research Institute (ESRI) data and maps. Other environmental feature data, such as shape files for railroads and greenways, were acquired from KGIS. Distance calculations for various location vari-

[3] There may be some grounds for arguing that real estate markets in Knoxville would anticipate the imposition of the UGB between 1998 and 2001. Based on a series of interviews with local realtors and planners about the issue, we concluded that the anticipation factor may be reasonably trivial.

Table 1. Variables and Descriptive Statistics

Variables	Pre-UGB (n = 38,045)		Post-UGB (n = 27,745)	
	Mean	SD	Mean	SD
Endogenous variables				
Land value per square foot ($)	4.31	8.64	1.00	2.25
Develop dummy	0.27	0.44	0.21	0.40
Parcel variables				
Lot size (square feet)	104,074.90	3,69,402.70	123,509.40	411,879.87
Census-block group variables				
House density	0.90	1.04	0.96	1.18
Travel time	23.29	4.15	23.16	4.31
Per capita income ($)	22,870.00	10,800.00	22,320.00	11,170.00
Unemployment rate	0.04	0.04	0.05	0.04
Vacancy rate	0.07	0.03	0.07	0.04
Jurisdiction and high school district variables				
Knoxville City	0.32	0.46	0.37	0.48
Austin High	0.05	0.22	0.06	0.25
Bearden High	0.09	0.28	0.07	0.26
Carter High	0.08	0.27	0.09	0.28
Central High	0.07	0.26	0.08	0.27
Doyle High	0.11	0.31	0.13	0.33
Fulton High	0.05	0.23	0.07	0.25
Gibbs High	0.08	0.28	0.08	0.27
Halls High	0.07	0.25	0.06	0.24
Karns High	0.10	0.30	0.08	0.28
Powell High	0.06	0.24	0.05	0.23
West High	0.09	0.28	0.09	0.29
Distance and physical variables				
Downtown (feet)	43,281.34	23,093.30	40,884.89	24,005.75
Water (feet)	8,619.95	6,489.25	8,153.36	6,333.24
Green (feet)	9,310.55	7,133.14	9,116.82	7,265.08
Rail (feet)	7,562.18	6,973.95	7,253.85	7,026.72
Park (feet)	9,609.61	6,556.77	9,230.29	6,501.53
Interstate (feet)	14,033.35	11,240.47	13,714.57	11,430.10
Slope (degree)	1.77	1.30	1.84	1.37
Urban growth boundary variables				
UGB	—	—	0.10	0.30

ables were made using the shape files and ArcGIS 9.1 (ESRI).

Summary statistics for the variables created from the data set are listed and defined in Table 1. After cleaning up the parcel data (deleting missing observations, unreasonably low appraised value of land parcels below $1,000, and unreasonably small parcel sizes below 100 square feet according to the land planner and data provider who have been working on the parcel data), there were 38,045 parcels for the pre-UGB data and 27,745 parcels for the post-UGB data that were used for the estimations.[4]

[4] Alternatively, we tried Grubbs' test for the detection of outliers and found more outliers compared with what land planner and data provider indicated. Our judgment was that the test may detect correct data as outliers because the test is structured to detect outliers and errors in the data. Our intention with the elimination of the data was not to delete outliers but to eliminate errors in the data.

For the cross validation, a holdout method sometimes called test sample estimation was used. The data were partitioned into two mutually exclusive subsets of estimation and validation samples. The estimation subset was used for the parameter estimations, and the validation subset was used for the testing of the validation of the parameter estimates. It is common to designate two thirds of the data as the estimation set and the remaining one third as the validation set (Kohavi). Accordingly, the pre-UGB and post-UGB samples were divided into 60% estimation and 40% validation samples. Out of the 38,045 parcels of the pre-UGB sample, 22,739 were randomly selected as an estimation set and 15,306 were selected as a validation set, while out of the 27,745 parcels of the post-UGB sample, 16,560 were randomly selected as an estimation set and 11,185 were selected as a validation set.

Results

The first empirical task was to test for differences in the time frame, before and after the UGB, and determine whether the empirical model should be estimated with separate pre-UGB and post-UGB samples or a pooled sample. This was accomplished with a likelihood ratio (LR) test. Specifically, denote the maximum log-likelihoods for the pre-UGB sample, post-UGB sample, and pooled sample (with a dummy variable indicating before and after the UGB, time–UGB variable) as f_{pr}, f_{po}, and f_P, with corresponding numbers of parameters k_{pr}, k_{po}, and k_P. Then, the LR statistic $2(f_{pr} + f_{po} - f_P)$ is Chi-square distributed with $(k_{pr} + k_{po} - k_P)$ degrees of freedom. The hypothesis that all slope parameters (i.e., except the constants) are equal was rejected (LR $=$ 360.38, df $=$ 26, $p < 0.001$), suggesting that inclusion of a time–UGB variable in the pooled sample did not fully capture differences of pre-UGB and post-UGB. The second empirical task was to test for endogeneity of land value (i.e., whether the two-stage probit least square model or simple probit model should be used). The hypothesis of no endogeneity of land value in the development equation was rejected for both pre-UGB (LR $=$ 26956.14, df $=$ 1, $p < 0.001$) and post-UGB (LR $=$ 1167.9, df $=$ 1, $p < 0.001$) samples.

Parameter Estimates and Marginal Effects of the Two-Stage Probit Least Squares

Parameter estimates and marginal effects of the development equation of the probit model (henceforth, two-stage probit least square model or probit model) with an endogenous binary variable for development and an endogenous continuous variable for land value using the full sample of 38,045 for the pre-UGB sample and the full samples of 27,745 for the post-UGB sample are presented in Tables 2 and 3, respectively. Similar results for the land value equations for both samples are available upon request. We limit our discussion to the results from the development equation because the focus is on the impact of UGB on development and predictive accuracy of the development models.

Results for both samples suggest land value and land development are simultaneously determined, as evidenced by significance of the endogenous variables in the corresponding equations. Further, parcels with higher land values are more likely to be developed conditional upon the exogenous variables. This positive effect of land value on development is explained by the fact that development typically occurs in the neighborhoods of developed areas where development pressure increases land value.

There is an interesting difference between pre-UGB and post-UGB samples in the effect of development. Notable differences, in terms of both statistical significance and magnitudes, between the two samples are found. Travel time to work is significant at the 1% level for the development prior to the UGB, whereas the corresponding estimate is not significant, even at the 10% level, for the post-UGB development. While proximity to railroad is a significant negative factor for development prior to the UGB at the 1% level, the variable is not a significant factor at the 5% for the post-UGB development. Quantitative differences also exist between the two samples. For instance, whereas the estimate of lot size is

Table 2. Estimates for the Development Equation of the Two-Stage Probit Model (Pre-UGB)

Variables	Coefficient	Marginal Effect
Intercept	−5.198***	—
	−0.259	
Endogenous variables		
Ln(Land value/square feet)	0.655***	20.610***
	−0.025	−0.007
Develop	—	—
Parcel variables		
ln(lot size)†	0.270***	0.830***
	−0.02	−0.006
Census-block group variable		
House density	−0.073***	−2.300***
	−0.012	−0.003
Travel time	0.007***	0.235***
	−0.002	0
ln(per capita income)†	−0.022***	−0.700*
	−0.001	0
Unemployment rate	−1.329***	−41.786***
	−0.264	−0.083
Vacancy rate	−0.705***	−22.19***
	−0.263	−0.082
City jurisdiction variables		
Knoxville‡	−0.070***	−2.188***
	−0.025	−0.007
Distance and physical variables		
ln(water)†	0.067***	0.242***
	−0.008	−0.002
ln(greenway)†	0.045***	0.162***
	−0.009	−0.003
ln(rail)†	0.029***	0.108***
	−0.007	−0.002
ln(park)†	0.029	0.103***
	−0.01	−0.003
ln(interstate) †	0.082***	0.283***
	−0.008	−0.002
slope	0.009	0.002
	−0.007	−0.002
Urban growth boundary variables		
UGB	—	—
Adjusted R^2	0.1	

Notes: Dagger (†) indicates marginal effect of the natural log variables, which were divided by mean values of the respective variables × 10^{-3}, and ‡ indicates marginal effect of discrete changes in dummy variables. Only significant marginal effects are shown in the table. ***, **, and * denote statistical significance at 1%, 5%, and 10% levels, respectively.

0.270 for the pre-UGB development equation, the corresponding estimate is 0.156 for the post-UGB development equation.

The coefficients of the second-stage probit model are estimated effects on a cumulative normal function of the probabilities that the response variable equals 1, and therefore they have no directly intuitive interpretation. For this reason, marginal effects are calculated and more closely investigated than the coefficients

Table 3. Estimates for the Development Equation of the Two-Stage Probit Model (Post-UGB)

Variables	Coefficient	Marginal Effect
Intercept	−4.647***	—
	−0.415	
Endogenous variables		
ln(land value/square feet)	0.786***	20.354***
	−0.053	−0.013
Develop	—	—
Parcel variables		
ln(lot size)†	0.156***	0.394***
	−0.031	−0.008
Census-block group variable		
House density	−0.110***	−2.863***
	−0.015	−0.004
Travel time	0	0.022
	−0.002	0
ln(per capita income)†	−0.013***	−0.330***
	−0.002	0
Unemployment rate	−1.248***	−32.310***
	−0.291	−0.075
Vacancy rate	−0.614**	−15.890**
	−0.309	−0.08
City jurisdiction variables		
Knoxville‡	−0.331***	−8.259***
	−0.035	−0.008
Distance and physical variables		
ln(water)†	0.128***	0.387***
	−0.01	−0.002
ln(greenway)†	0.054***	0.161***
	−0.012	−0.003
ln(rail)†	0.016***	0.050*
	−0.009	−0.002
ln(park)†	0.054***	0.160***
	−0.013	−0.003
ln(interstate)†	0.122***	0.340***
	−0.011	−0.003
slope	−0.003	−0.001
	−0.008	−0.002
Urban growth boundary variables		
UGB	0.026	0.007
	−0.031	−0.008
Adjusted R^2	0.13	

Notes: Dagger (†) indicates marginal effect of the natural log variables, which were divided by mean values of the respective variables × 10^{-3}, and ‡ indicates marginal effect of discrete changes in dummy variables. Only significant marginal effects are shown in the table.

***, **, and * denote statistical significance at 1%, 5%, and 10% levels, respectively.

Table 4. Comparison of Prediction Accuracy Using Two Methods

	Pre-UGB (1997–2000)				
	Classification (*n*)				Accuracy (%)
Data set	Undeveloped to Undeveloped	Undeveloped to Developed	Developed to Undeveloped	Developed to Developed	Overall
	Result from the Simultaneous Probit Model				
Estimation	15,639	839	4,931	1,330	74.63
Validation	10,594	584	3,322	806	74.48
Full	26,343	1,313	8,387	2,002	74.50
	Post-UGB (2001–2004)				
Estimation	12,717	321	3,023	499	79.81
Validation	8,615	209	2,095	266	79.40
Full	21,342	490	5,152	731	79.64
	Result from the Neural Network Model				
	Pre-UGB (1997–2000)				
Estimation	15,852	626	660	5,601	94.34
Validation	10,722	456	404	3,724	94.38
Full	26,574	1,082	1,064	9,325	94.36
	Post-UGB (2001–2004)				
Estimation	11,821	1,217	1,555	1,967	83.26
Validation	7,991	833	1,114	1,247	82.59
Full	19,812	2,050	2,669	3,214	82.99

to evaluate the relationships in the model. Marginal effects of continuous variables are evaluated at the sample means, and marginal effects of dummy variables are evaluated for discrete changes from 0 to 1.

Both land value variables are found to be significant at the 1% level. A $1 per square foot increase in land value increases the probability of development by about 21% and 20% for pre-UGB and post-UGB samples respectively. An increase in lot size by 10 square feet increases the probability of pre-UGB development by about 8%, while the same change increases the post-UGB development by about 4%. This discrepancy in marginal effects for the lot size variable before and after the UGB indicates that the size of lot mattered more for development prior to the UGB imposition than after.

A comparison of marginal effects of census-block group variables shows notable differences, in terms of both statistical significance and magnitudes, between the pre-UGB and post-UGB data. Travel is significant at the 1% level in the pre-UGB data, and the corresponding estimate is not significant at the 10% level for the post-UGB data. The insignificant effect of Travel for the post-UGB development may reflect the geographically sprawled employment opportunities regardless of distance to the employment center. Quantitative differences also exist between the estimates. Whereas a 1% increase of unemployment decreases the probability of development by about 42% in the pre-UGB data, the corresponding change decreases the probability of development by about 32% in the post-UGB data. An increase of one house per square mile decreases the probability of development by about 2% in the pre-UGB sample, and the change decreases by about 3% of the probability of development in the post-UGB. This is an indication that more development occurred in lower density areas before and after the UGB. This may be evidence of sprawl development in the areas as the sprawl can be defined as development in less-dense areas.

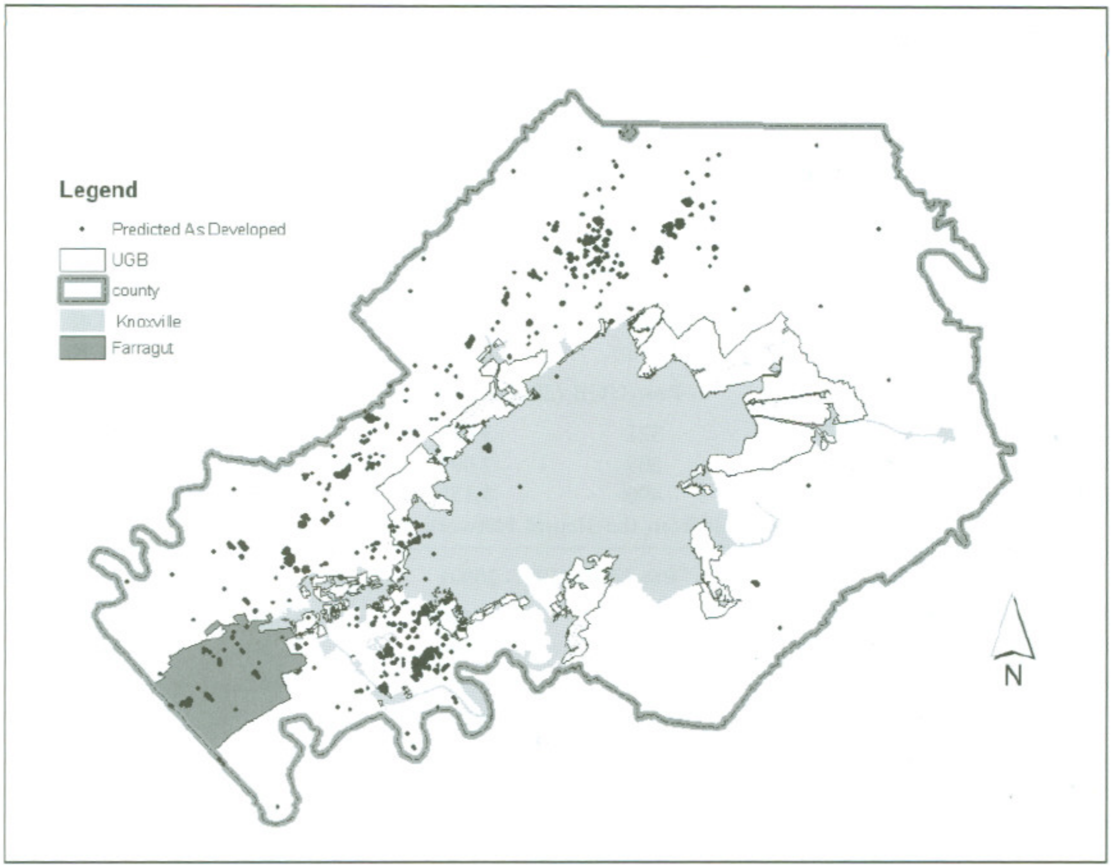

Figure 2. Predicted Development of Pre-UGB Period Using Two-Stage Probit Least Squares

According to the estimated marginal effects of the Knoxville variable, a lot inside the Knoxville city boundary is about 2% less likely to be developed for the pre-UGB sample, while the corresponding lot is about 8% less likely to be developed for the post-UGB sample. This higher probability of development outside of the city boundary for the post-UGB sample compared with pre-UGB sample indicates that the development outside of the city boundary occurred more frequently in the post-UGB period than in the pre-UGB period.

The marginal effects of distance variables are different between the two samples quantitatively and qualitatively. Railroad is significant at the 1% level for the pre-UGB data, but the corresponding estimate is not significant at the 5% level using the post-UGB data. This may be explained by the fact that Knoxville does not have use of rail mass transit, and proximity to a railroad track is likely to be associated with disamenities. Although the negative marginal effect of proximity to railroad exists before and after the UGB, the effect becomes less significant for the development in the post-UGB period.

Quantitative differences also exist in the "Interstate" variable between the two samples. An increase in distance to the nearest interstate highway by 100 feet increases the probability of development by about 28% for the pre-UGB sample and around 34% for the post-UGB sample, suggesting that probability of development increases with increasing distance from an interstate highway. Reluctance to develop away from an interstate highway is decreased during the post-UGB period relative to the pre-UGB period. We find that the "Slope" variable is not significant at the 10% level in the both pre-UGB and post-UGB samples. The insignificance of the variable may be explained by the fact that the Knox County still has low-slope undeveloped land.

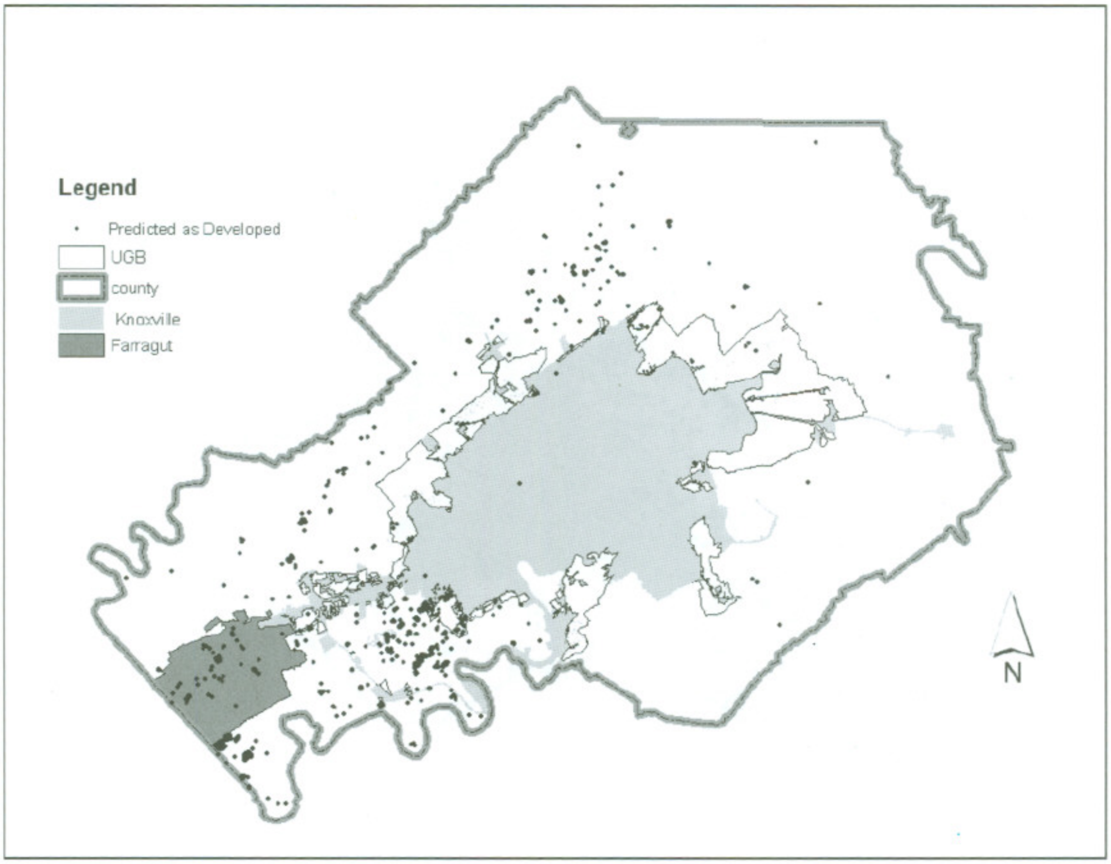

Figure 3. Predicted Development of Post-UGB Period Using Two-Stage Probit Least Squares

The UGB variable during the post-UGB period is found to be not significant at the 10% level. It stands in sharp contrast to the prediction using simple binary approach by Cho et al. This is empirical support for endogenous land value in the land development equation to account for the endogeneity of land value in the development equation, and also supports the result of significant differentiated development patterns between pre-UGB and post-UGB periods.

Prediction Comparison

Summaries of the predictive accuracy of full sample, estimation, and validation data for the two-stage probit least square and neural network models are presented in Table 4. Simulations using various neural network parameters led to using a learning rate of 1.0×10^{-2} and a momentum of 0.5 for one of the data sets and a learning rate of 1.0×10^{-6}, a momentum of 0.9 for the other data set. We used a *logsig* activation function to give the smallest classification errors for all models. Summaries of the prediction accuracies of full sample and estimation and validation data for the neural network model using two hidden layers with seven neurons in the first layer and five neurons in the second layer are given in Table 6. A combination of the input variables is used in the first hidden layer and a combination of the combined neurons in the first hidden layer is used in the second hidden layer.[5]

Overall percentages of correct predictions (Wooldridge, p. 465) show that the neural network model is preferred to the probit

[5] These combinations, however, are not user defined, and this is one reason the neural network has been described as a black box.

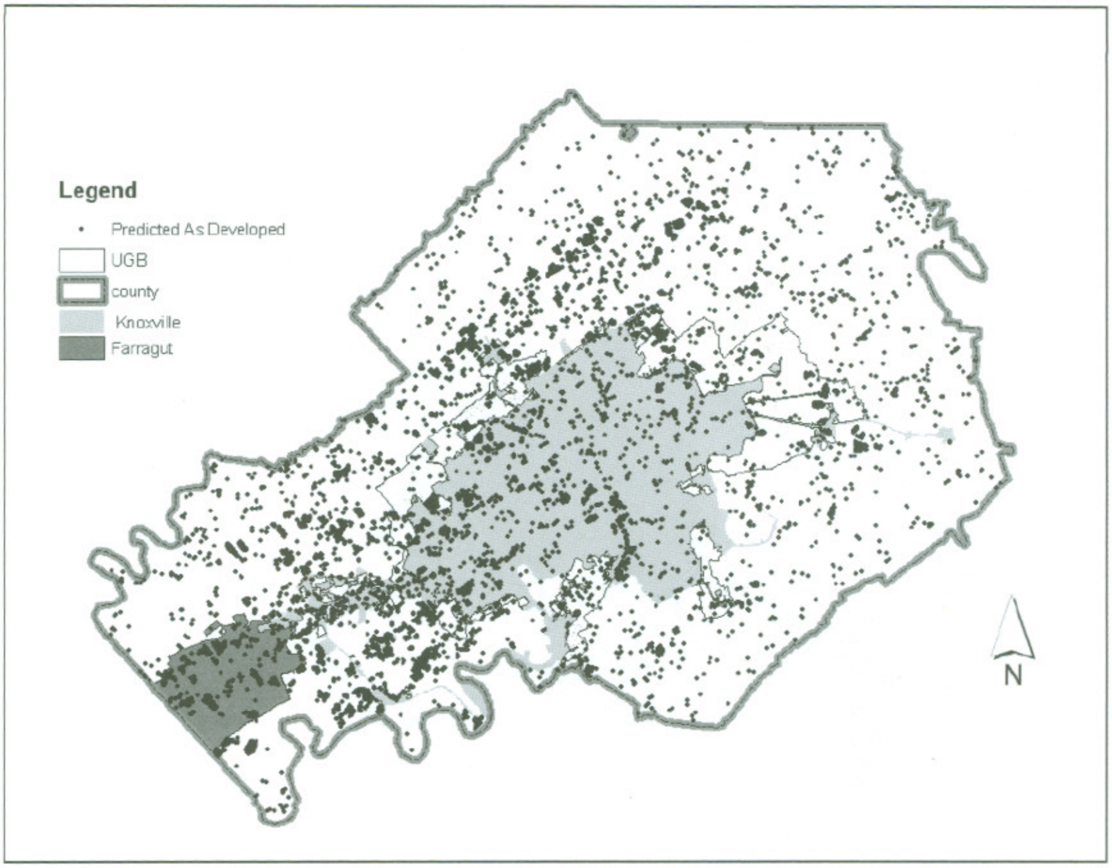

Figure 4. Predicted Development of Pre-UGB Period Using Neural Network Model

model. The overall improvement of the neural network over the probit model for the entire pre-UGB sample is 19.9%. The improvement in the prediction for the entire post-UGB sample declined significantly (3.4%). The cross validations of the prediction accuracies of validation samples in both models for the both pre-UGB and post-UGB data are consistent with prediction accuracies of corresponding estimation samples suggesting the reliability of the both models.

The differences in the prediction accuracy may be attributed to the difference in the structure of the two models. A neural network selects the parameters that best approximate the underlying functional form of the data, while in the probit model some restrictive assumptions must be made about the function form of the relationship between the explanatory variables and the probabilities of each of the alternative outputs. Furthermore, the error terms for the probit model are assumed to be normally distributed while no assumption about error terms is made in the neural network.

Based on the results of the second stage development equations, maps of predicted development probabilities greater than 0.5 are drawn in Figures 2 and 3 for the pre-UGB and post-UGB samples, respectively. Predicted development probabilities greater than 0.5 using the neural network model are drawn in Figures 4 and 5 for the pre-UGB and post-UGB samples, respectively. A comparison of the development patterns during the pre-UGB period between the probit and neural network models shows some notable differences. Whereas the probit model shows a clear west and north development pattern outside of the city boundary in Figure 2, the neural network model shows a similar but less clear development pattern due to more wide-

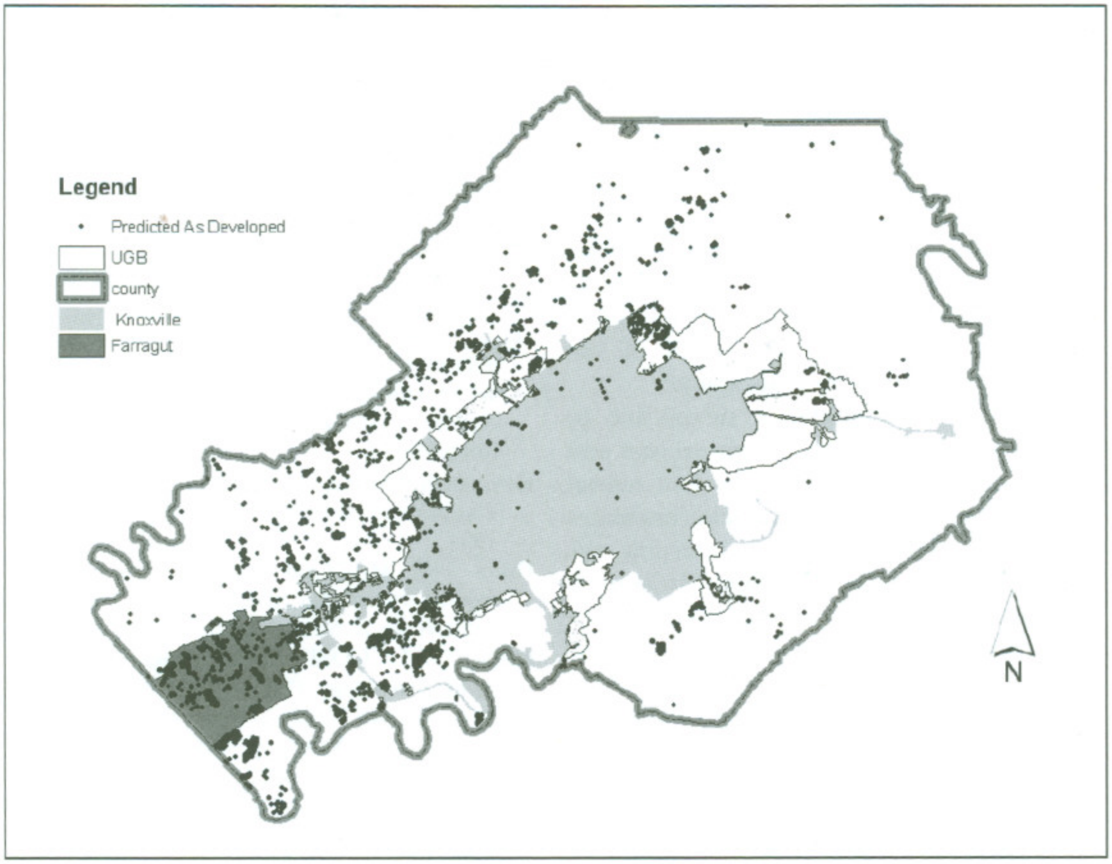

Figure 5. Predicted Development of Post-UGB Period Using Neural Network Model

spread predicted development across the county in the Figure 4. Comparison of the development patterns during the post-UGB between the probit and neural network models in Figures 3 and 5, respectively, reveals fewer differences. Both models predict that most development occurred in the west and north outside of the city boundary. Both models predict more development occurred in the town of Farragut during the post-UGB period than during the pre-UGB period. They also show that the development to the west of the city boundary gained development momentum, while the development in the north declined during the post-UGB period compared with pre-UGB period.

Conclusion

The insignificance of UGB variable in the probit model for the post-UGB sample and the more visible development patterns of the west and the town of Farragut during the post-UGB period in both the probit and neural network models suggest the UGB adopted for Knox County has not slowed down urban sprawl. In fact, the sprawl toward west Knox County has been more intense during the post-UGB period than the pre-UGB period. Although the cause of more intense sprawl in the west outside of the city boundary during the post-UGB period is not clear, this pattern of development confirms the insignificant role of UGB found in the probit model. While this result may not be surprising given the UGB has been adopted in Knox County for a relatively short time period and UGBs are usually considered long-term growth management tools, this is in sharp contrast to the finding by Cho et al. that the UGB in Knox County has had significant impacts on land development pattern. However, the role of UGB boundary in their model may be masked by the use of a pooled sample (both pre-UGB and post-UGB samples) and exhibit a bias caused by the endogeneity of land value.

A neural network model is found to be a viable alternative to the more conventional discrete choice probability approach of land use modeling for the purpose of improving the predictability of land development. The ability of neural networks to "learn" relationships from a set of variables and a flexible functional form are the likely causes of the improvement in prediction accuracy.

The regression model is susceptible to outliers affecting the estimation process and subsequent predictions. However, a neural network does not yield parameter estimates for the marginal effects to interpret the relationships between development and its determinants. Thus, the performance improvement of neural networks is rather costly. In our case of land use modeling, the gain in prediction accuracy with the neural network model over the probit model was at the cost of estimating the effects of the UGB on development. An implication is that neural network models should not be used as a sole investigation tool. Rather, it is a complementary tool for the use of land development projection. In addition, the performance of the network model depends on the number of hidden layers, the number of neurons, the value of the learning rate, and the value of the momentum. There is no one best answer on how to handle these parameters.

[Received April 2006; Accepted December 2006.]

References

Balling, R.J., J.T. Taber, M. Brown, and K. Day. "Multiobjective Urban Planning Using a Genetic Algorithm." *ASCE Journal of Urban Planning and Development* 125(June 1999):86–99.

Basu, S., and T.G. Thibodeau. "Analysis of Spatial Autocorrelation in House Prices." *Journal of Real Estate Finance and Economics* 17(July 1988):61–85.

Bogart, W.T., and B.A. Cromwell. "How Much More is a Good School District Worth?" *National Tax Journal* 50(June 1997):215–32.

Cho, S., Z. Chen, S.T. Yen, and D.B. Eastwood. "Estimating Effects of an Urban Growth Boundary on Land Development." *Journal of Agricultural and Applied Economics* 38(August 2006):287–98.

Downs, A. "Have Housing Prices Risen Faster in Portland than Elsewhere?" *Housing Policy Debate* 13(2002):7–31.

ESRI. 2006. Environmental Sensitivities Research Institute. Internet site: www.esri.com/ (Accessed July, 2006).

Hayes, K.J., and L.L. Taylor. "Neighborhood School Characteristics: What Signals Quality to Homebuyers?" *Federal Reserve Bank of Dallas Economic Review* Fourth Quarter(1996):2–9.

Haykin S. *Neural Networks—A Comprehensive Foundation*. New York: Macmillan Company, 1994.

Hines J.W. *MATLAB Supplement to Fuzzy and Neural Approaches in Engineering*. New York: John Wiley & Sons, 1997.

Keshk, O.M. "CDSIMEQ: A Program to Implement Two-Stage Probit Least Squares." *The Stata Journal* 3(June 2003):1–11.

Kohavi R. "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." International Joint Conference on Artificial Intelligence Conference Proceeding, 1995. Internet site: http://robotics.stanford.edu/~ronnyk/accEst.ps (Accessed October, 2006).

Landis J. "The California Urban Futures Model: A New Generation of Metropolitan Simulation Models." *Environmental and Planning B, Planning and Design* 21(July 1994):399–420.

Lin H., K. Lu, M. Espey, and J. Allen. "Modeling Urban Sprawl and Land Use Change in a Coastal Area." Paper presented at the annual meetings at American Agricultural Economics Association, Providence, RI. 2005.

Maddala, G.S. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge, UK: Cambridge University Press, 1983.

Mann, S., and G. Benwell. "The Integration of Ecological, Neural, and Spatial Modeling for Monitoring and Prediction for Semi-arid Landscapes." *Computers and Geosciences* 22(November 1996):1003–1012.

MathWorks. *Neural Network Toolbox—User's Guide Version 4*. Massachusetts: The MathWorks Inc., 2002.

Mehrotra K., C. Mohan, and S. Ranka. *Elements of Artificial Neural Networks*. Cambridge, MA: MIT Press, 1997.

MPC, Knoxville/Knox County Metropolitan Planning Commission 2006. "Directory of Neighborhood Organizations." Internet site:

www.knoxmpc.org/director/orgs/ndhome.htm (Accessed August, 2006).

MPC, Metropolitan Planning Commission, Tennessee Public Chapter 1101: Growth plan for Knoxville, Knox County, and Farragut, Tennessee, 2001. Internet site: www.knoxmpc.org/plans/growthpl.htm (Accessed December, 2006).

Nauck, D., F. Klawonn, and R. Kruse. *Foundations of Neuro-Fuzzy Systems*. New York: John Wiley & Sons, 1997.

NRCS, *Summary Report 1997 National Resources Inventory* (revised December 2000). Washington, DC: Natural Resources Conservation Service, U. S. Department of Agriculture, 2000.

Parker, D.C., S.M. Manson, M.A. Janssen, M.J. Hoffman, and P.J. Deadman. "Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review." *Annals of the Association of American Geographers* 93(June 2003):316–40.

Phillips, J., and E. Goodstein. "Growth Management and Housing Prices: The Case of Portland, Oregon." *Contemporary Economic Policy* 18(July 2000):334–44.

Smith, T.R. "Artificial Intelligence and its Applicability to Geographical Problem Solving." *Professional Geographer* 36(May 1984):147–58.

TACIR, Tennessee Advisory Commissionon Intergovernmental Relations. Growth Policy, Annexation, and Incorporation Under Public Act 1101 of 1998: A Guide for Community Leaders, 1999. Internet site: www.state.tn.us/tacir/PDF_FILES/Growth_Policy/Annexation98.pdf (Accessed September, 2006).

Wooldridge, J.M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press, 2002.

Yeh, A., and X. Li. "Urban Simulation Using Neural Networks and Cellular Automata for Land Use Planning." *Advances in Spatial Data Handling*. D. Richardson and P. Van Oosterom, eds. Berlin: Springer, 2002.

Yen, S.T., and S. Cho. "Residential Development and Land Value: The Differentiated Roles of Socioeconomic and Growth Policy Variables between Urban and Rural–Urban Interface Areas." Working paper, Dept. of Agr. Econ., University of Tennessee, Knoxville, 2006.