Colin Vance and Rich Iovanna

Cities and Satellites: Spatial Effects and Unobserved Heterogeneity in the Modeling of Urban Growth

#43

Ruhr Economic Papers

RWI
ESSEN

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstraße 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstraße 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI Essen)
Hohenzollernstrasse 1/3, 45128 Essen, Germany

Ruhr Economic Papers #43

Responsible Editor: Christoph M. Schmidt

ISSN 1864-4872 (online) – ISBN 978-3-86788-043-5

Ruhr Economic Papers

#43

Colin Vance and Rich Iovanna

Cities and Satellites: Spatial Effects and Unobserved Heterogeneity in the Modeling of Urban Growth

RWI

ESSEN

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

ISSN 1864-4872 (online)
ISBN 978-3-86788-043-5

Colin Vance and Rich Iovanna*

Cities and Satellites: Spatial Effects and Unobserved Heterogeneity in the Modeling of Urban Growth

Abstract

The confluence of factors driving urban growth is highly complex, resulting from a combination of ecological and social determinants that co-evolve over time and space. Identifying these factors and quantifying their impact necessitates models that capture both why urbanization happens as well as where and when it happens. Using a database that links five satellite images spanning 1976–2001 to a suite of socioeconomic, ecological and GIS created explanatory variables, this study develops a spatial-temporal model of the determinants of built-up area across a 25,900 square kilometer swath across central North Carolina. Extensive conversion of forest and agricultural land over the last decades is modeled using the complementary log-log derivation of the proportional hazards model, thereby affording a means for modeling continuous-time landscape change using discrete-time satellite data. To control for unobserved heterogeneity, the model specification includes an error component that is Gamma distributed. Results confirm the hypothesis that the landscape pattern surrounding a pixel has a major influence on the likelihood of its conversion and, moreover, that the omission of external spatial effects can lead to biased inferences regarding the influence of other covariates, such as proximity to road. Cartographic and nonparametric validation exercises illustrate the utility of the model for policy simulation.

JEL Classification: C41, Q15, R14

Keywords: Urban growth, landscape pattern, satellite imagery, hazard model, North Carolina

March 2008

---

* Colin Vance, RWI Essen; Rich Iovanna, Economic and Planning Analysis Staff, Farm Service Agency, U.S. Department of Agriculture, Washington, DC. – The authors wish to thank Manuel Frondel for his comments on an earlier draft. We also thank Andrew M. Jones for his helpful insights on the application of frailty models. – All correspondence to Colin Vance, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI Essen), Hohenzollernstraße 1-3, 45128 Essen, Germany, e-mail: vance@rwi-essen.de.

1. INTRODUCTION

Debate over the causes and consequences of urban growth in the United States has focused largely on the extent and rate of open-space conversion. Proponents of so-called smart growth strategies often cite the rapid expansion of urbanized area as a justification for land use controls, contrasting, for example, the fourfold increase in urban land area between 1945 and 2002 with the doubling in population (Lubowski et al. 2006). Those contesting interventionist land use policies counter that rapid expansion notwithstanding, less than 4% of the nation's land is currently urbanized, with most new development occurring within a kilometer of existing development (Duranton and Puga 2003). Open-space is thus argued to be a resource that is in plentiful supply, the preservation of which would, in any case, only increase development pressures elsewhere. Cross-cutting these broader themes are ancillary debates concerning the effect of anti-growth measures on congestion, air quality, housing affordability, and freedom of choice, among other issues.

Against this backdrop, transportation- and city planners are confronted with a complex set of considerations when evaluating alternative regulatory, infrastructural, and fiscal interventions and the urban shapes to which these give rise. In this regard, a critically important question – and one which has generally been overlooked in the debates on sprawl – concerns the spatial distribution of urban growth, and in particular, how existing landscape patterns determine future trajectories of change. While a road built in one location, for example, may offset development that would have otherwise occurred in another (Hartgen 2003), the resulting environmental and socioeconomic impacts are likely to be highly location-specific, irrespective of the total amount of development that transpires.

Recognition of the need for spatially-explicit approaches to the study of urbanization has led to an increasing number of studies that combine principles from landscape ecology with econometric methods to account for how human decision-making, ecosystem function, and their interaction effect landscape changes across different spatial scales. Models that are fine scale are particularly meaningful because ecologists and allied disciplines perceive an intimate connection between the provision of habitat and other services by ecosystems and the pattern of the landscape mosaic in which the ecosystems function. Unlike area-based approaches, which estimate the determinants of land use shares within aggregate geographic areas such as counties or parcels (Hardie et al. 2000; Stavins, Plantinga, and Lubowski 2003), much of the recent literature on urban growth draws on disaggregate point data derived from remotely sensed sources or ground surveys to estimate spatially explicit models of land use. An early example of such work in the U.S. context is Turner, Wear, and Flamm's (1996) multinomial logit analysis using a time series of satellite imagery to study the effect of socioeconomic, ecological, and locational factors on landscape changes in North Carolina and Washington. Other issues explored in this literature include the role of GIS-created spatial pattern metrics as determinants of property values (Geoghegan, Wainger, and Bockstael 1997), the joint influence of urban population growth and urban proximity on land use change (Kline, Moses, and Alig 2001), and the causes of fragmented development patterns among residential land parcels on the rural-urban interface (Irwin and Bockstael 2002).

The present study extends this line of inquiry by developing a spatial-temporal model of the determinants of built-up area across a 25,900 square kilometer swath in central North Carolina, an area that has undergone extensive conversion of forest and agricultural land over the last decades. Between 1976 and 2001, the area covered by impervious surface in the region more than doubled, from 625 to 1471 square kilometers, with the majority of the increase occurring in the metropolitan regions of Greensboro and Raleigh. In Raleigh, for example, the population increased by 32 percent between 1990 and 1996 while its urbanized land area increased nearly twofold (Sierra Club 2003). We model these landscape dynamics by exploiting a spatial database that links five satellite images spanning the years 1976-2001 to a suite of socioeconomic, ecological and GIS-created explanatory variables.

Exploring the role of landscape pattern in land-use change, we elect to decouple our exploration from the reliance on parcel-level data to focus on a consistent and finer unit of observation, a 60 by 60-meter satellite pixel. Although parcel analyses provide a direct link between the unit of observation and the land manager, the associated costs of data acquisition typically constrain the geographic coverage. Moreover, the pixel-level focus afforded by increasingly available satellite imagery recognizes that the conversion decision need not be an all or nothing proposition, an assumption that typically underpins parcel-level analyses (e.g. Stavins, Plantinga, and Lubowski 2003).

Our analysis takes as its point of departure a dynamic, profit-maximizing framework that suggests several possible determinants of land conversion from commodity-based to urban uses. We test for the significance of these determinants with a model derived from the proportional-hazards empirical specification. The model developed has several distinguishing features. By specifying the complementary log-log derivation of the proportional hazards model, we advance a methodology for modeling a continuous time process – the conversion of land to impervious surface – using discrete time satellite data. To control for unobserved heterogeneity, the model specification includes an error component that is Gamma distributed. Because the data itself is observed at a very fine level of spatial resolution, we additionally relax the assumption commonly invoked in land use shares models that all change occurs at the rural urban interface (Hardie et al. 2000). Finally, the model includes a broad array of covariates that measure the land allocation response to site, locational, and pattern attributes associated with each pixel. Following the works of Geoghegan, Waigner, and Bockstael (1997) and Irwin and Bockstael (2002), we are particularly interested in exploring the effects of spatial externalities, as captured by time-varying variables measuring the landscape pattern surrounding the pixel. Results confirm the hypothesis that what surrounds a pixel has a major influence on the likelihood of its conversion and, moreover, that the omission of external effects can lead to biased inferences regarding the influence of other covariates, such as proximity to road, that are commonly identified as important determinants of land use change.

2. LAND USE CHANGE MODEL

Understanding the timing, location, and causes of land use change is critical if policy makers wish to implement transportation projects in a cost efficient manner. Because almost all forms of land use change result from a comparative calculation of costs and benefits, we adopt an economic approach as the framework for constructing a predictive model of urban expansion in North Carolina. Land conversion decisions depend on a complex multiplicity of factors, including the market value of output from the land in alternative uses, expectations about the future use of neighboring lands, and the surrounding composition of land ownership. Following the work of Boscolo, Kerr, Pfaff and Sanchez (1998), the theoretical approach taken here attempts to structure this complexity by assuming that land will be converted if the net present discounted benefits of doing so are greater than the net present discounted benefits of leaving the land under its present use. In other words, the land manager converts pixel i in period T to maximize the following objective function:

$$Max_T \int_0^T A_{it}\left(X_{it}\right)e^{-rt}dt + \int_T^\infty D_{it}\left(X_{it}\right)e^{-rt}dt - C_T e^{-rT} \tag{1}$$

where

$A_{it}(X_{it})$ is the returns derived from a commodity-based use of the land in period t, i.e., the agricultural or forestry rent;

$D_{it}(X_{it})$ is the returns to development in period t, i.e., the development rent;

C_T is the cost associated with conversion at time T;

X_i is the vector of site and locational attributes of the pixel influencing returns, including environmental factors and accessibility costs; and

r is the discount rate.

Within X_i, the influence of the current pattern of land use surrounding a pixel that is yet at risk of conversion is of particular interest. The degree to which fragmented landscapes exhibit higher rates of conversion, for example, may help direct inquiry toward specific land-use change drivers, such as the loss of economies to traditional economic activities that may come from fragmentation, the relative appeal to homeowners of a proximate mix of land uses, or the amenities afforded by local landscape features such as water bodies.

Assuming irreversibility of the conversion process, there are two necessary conditions for conversion to take place. The first is that the discounted stream of returns derived from conversion are greater than that of leaving the plot in its present use, net of the one-time conversion costs:

$$\int_T^{\infty}(D_{it} - A_{it})e^{-rt}dt - C_T > 0 \tag{2}$$

The operative condition, however, is one that will be met well after that specified by equation (2): Conversion will occur when the development rent just equals the opportunity cost, *OC,* of developing that period as opposed to the next. Before time *T* and assuming development rents are rising over time and conversion costs are declining, it is more profitable to defer development for at least another period. After *T*, the landowner loses money every period that development is deferred. More formally, a developed pixel is one in which:

$$D_{it} \geq OC_{it} = A_{it} + rC_{it} - \frac{d}{dt}C_{it} \tag{3}$$

If the development rent in period t exceeds the sum of agricultural rent and the cost savings from deferring development, which relates to downward trend in costs as well as the fact that costs are discounted an additional period, the pixel has already been developed. With equality, time *T* is when conversion actually takes place.

The model of land-use conversion developed above is deterministic in assuming that the timing of development can be explained solely by variation in pixel attributes. To account for unobserved idiosyncratic factors associated with pixel *i*, such as an owner who gains satisfaction from maintaining open space, we add an error term to equation (3):

$$D_{it} \geq A_{it} + rC_{it} - \frac{d}{dt}C_{it} + \varepsilon_{it} \tag{4}$$

If we further define ε^* as the amount that makes (4) an equality, then we find the likelihood of conversion at time *t* to simply be the cumulative density of ε evaluated at ε^*. In other words, if the error for pixel *i* at time *t* is less than or equal to ε^*, conversion occurs.

In taking the above framework to the empirics, we focus on the critical role that timing plays in the risk of conversion. Given that conversion may occur at any point in time during the period under observation and that the factors influencing conversion are often continuous processes, survival modeling is uniquely suited to the task of estimating the parameters of interest. Rather than modeling the direct influence of a covariate on conversion probabilities, survival models are concerned with the hazard rate underlying the probabilities, i.e., the instantaneous risk that pixel *i* is cleared in period *t* conditional on not having been converted before *t*. While conventional methods such as linear or logistic regression have been applied in these contexts, they are ill-equipped to handle the features that often characterize survival data, including time-varying explanatory variables and censoring or truncation of the dependent variable.

Derived from satellite imagery, our data are interval censored. We know simply whether or not an observation's survival time falls somewhere between two dates. Accordingly, the dependent variable assumes a value of one if conversion occurs over an interval between the dates and zero otherwise. To reconcile the temporal continuity of the conversion process being modeled with this coarseness in the measurement of timing, we specify a complementary log-log survival model. By doing so, the relationship between the *X* covariates and the probability that opportunity costs (OC) are low enough for conversion to occur (i.e., that ε is less than or equal to ε^*) is assumed to be:

$$P_{itOC} = 1 - \exp\left[-\exp(X_{it}\beta + \theta_t(t) + \gamma_i\right] \qquad (5)$$

As a proportional hazards model and a discrete analogue to that developed by Cox (1972), the complementary log-log model requires no assumptions regarding the functional form of the baseline hazard rate, $\theta(t)$. In the present application, $\theta(t)$ is modeled as a step function by using dummy variables to represent each interval at risk, thereby enabling attention to be focused specifically on the effect of the covariates on the relative risk of a transition. The inclusion of these dummies additionally serves to control for the unequal interval durations as given by the dates of the satellite imagery (Allison, 1995).

Our specification of γ in Equation (5) takes two alternate forms, one in which it is set equal to zero and one in which it is specified as having a predetermined mixing distribution to capture unobserved heterogeneity, sometimes referred to as *frailty*. The frailty specification can alternatively be thought of as a random intercept model, with the intercept equal to $\beta_0 + \gamma$ (Jenkins 2005). A commonly used mixing distribution – and the one applied here – is the Gamma distribution, with unit mean and variance equal to σ^2. By comparing results with a standard non-frailty model in which γ=0, it is possible to gauge the extent to which unobserved individual effects lead to biases in the estimated coefficients. As proven by

Lancaster (1990) for the case of Gamma distributed frailty, the direction of this bias is downward, resulting in underestimated positive parameters and overestimated negative parameters.

3. DATA

A time series of five classified satellite images spanning central North Carolina for the years 1976, 1980, 1986, 1993 and 2001 comprise the core data used in this analysis. Data for the years 1976 and 1980 were derived from the Landsat Multispectral Scanner (MSS) imaging system, while the Thematic Mapper (TM) imaging system was the data source for the years 1986, 1993, and 2001. Because TM and MSS data have different spatial resolutions – 58 X 79 meters for MSS and 30 X 30 meters for TM – the data was spatially degraded to a 60 X 60 meter resolution for consistency. The data set and code used to estimate the models, available either in Stata or SAS format, can be obtained from the authors upon request.

The process of imagery classification was preceded by the standard pre-processing activities, including geometric correction, spectral-spatial clustering, and radiometric normalization. Classification then proceeded according to a hybrid change detection methodology combining radiometric and categorical change techniques on a pixel-by-pixel basis. This procedure produced four land cover classes: forest, non-forest vegetation, impervious surface, and water. From these classes, we generated a binary dependent variable equaling 1 if a conversion from forest or non-forest vegetation to impervious surface occurred between two dates and 0 otherwise. Conversions to water were treated as censored, while pixels whose classification in the first year (1976) was either water or impervious surface were eliminated from the data. Transitions between forest and non-forest vegetation were also treated as censored as these may be attributable more to forest rotations than permanent conversion from one land cover to another. After overlaying two GIS layers of tenure data from ESRI (2000 a,b) and the North Carolina Department of Parks and Recreation (2006), those pixels falling under public ownership (e.g. national, state, and municipal parks) were also eliminated.

Upon classifying the imagery, a systematic sample of pixels was drawn that provided 65,991 pixels for model estimation. The grid pattern across the satellite scene was such that roughly 1.2 kilometers separated each pixel from their nearest neighbors. Systematic sampling is a commonly applied technique to handle spatial correlation of unobserved variables that affect the probability of conversion (Turner, Wear, and Flamm 1996; Kline, Moses, and Alig 2001; Cropper, Puri, and Griffiths 2001). A major source of spatial autocorrelation arises from multiple observations falling under common landowners (Kline, Moses, and Alig 2001). In North Carolina, the average size of private forest ownership is 9.7 hectares (Powell et. al. 1992), while the average farm size is approximately 75 hectares

(USDA 1997). Consequently, we assume that 1.2 kilometer pixel separation in our sample is an adequate distance to exclude the possibility of spatial autocorrelation arising from common ownership.

Several static and time-varying covariates are included in the model, the values for which correspond to the start year of the interval given by the dates of the satellite imagery. The suite of variables specified captures both site and locational attributes that are hypothesized to affect the returns to land in developed and undeveloped uses. To capture the influence of what Alig and Healy (1987) have termed "spatially bounded externalities that affect adjoining or nearby land," we derived three time-varying spatially lagged variables from the imagery that measure the landscape configuration surrounding a pixel. The first is the percent of the area within a window of approximately two square kilometers that is classified as impervious (*inner impervious surface*). The size of the window is admittedly arbitrary, yet also based both on a typical developer's spatial frame of reference and on previous studies that have found window-sizes of similar magnitude to capture spatial externalities (Geoghegan, Wainger, and Bockstael 1997; Irwin and Bockstael 2002). Given the likely predominance of agglomeration effects associated with impervious surface in the immediate vicinity of the pixel, we hypothesize the sign of this variable to be positive. To allow for nonlinearities in the effect, we also include the variable's square.

The second metric complements the first, and is the percent of impervious surface in a region between the aforementioned window and a larger one with sides twice as large (*outer impervious surface*). The inclusion of this variable, which is non-overlapping with *inner impervious surface*, recognizes the possibility of varying parameters with increased distance from the pixel. Such variation may arise, for example, from spatial externalities associated with neighboring parcels. While the effect of such externalities is expected to vary depending on the surrounding configuration of land uses, evidence obtained from Irwin and Bockstael (2002) suggests the net effect to be negative, a finding that they attribute to "repelling effects" associated with low-density residential development. It bears noting that Irwin and Bockstael use parcel-level data, allowing them to readily identify the effects of neighboring land parcels. The absence of parcel boundary information here makes it difficult to distinguish true externalities associated with a neighboring parcel from spatial effects within the parcel itself. With respect to the variable *outer impervious surface*, however, there are two reasons why any difference between a window-based calculation of imperviousness centered on the parcel and that on an undeveloped pixel within the parcel is likely to be negligible. First, the calculation covers a surface area that is a kilometer in width and a kilometer removed from the pixel on all sides, so that overlap with the parcel is difficult to conceive. Moreover, to the extent that the undeveloped pixels comprising the sample are within parcels that are themselves undeveloped, whatever impervious surface entering in the calculation of the metric will be primarily – if not exclusively – associated with neighboring parcels.

The remaining metric, based on the smaller window, is the percent of area classified as water (*percent water*). The effect of percent water on the likelihood of development is ambiguous. While developers may covet increased water surface area as a residential amenity, this feature could also confer benefits to agricultural activities.

In addition to the window-based metrics, two time-varying proximity-based metrics are also included in the specification, the first of which is the Euclidean distance to the nearest woodchip mill (*distance to chipmill*) (Prestemon et al. 2000). A recent study concludes that while mills in North Carolina are generally located in areas where wood supply is plentiful, they usually do not harvest at rates that exceed growth levels within a 50 mile radius (Schaberg et al. 2005). The study goes on to note that the cumulative effect of overlapping mill procurement zones remains an open question. We therefore leave the sign of this variable to the empirical results. The second proximity metric is the Euclidean distance to the nearest major road (*distance to road*), which includes interstate highways, U.S. and state highways, and other major thoroughfares. This variable is expected to have a negative effect on the conversion hazard given higher access costs.

Three time-invariant variables are included in the model that capture the effects of proximity to other landscape features. The first of these, *distance to city,* measures the Euclidean distance to the nearest city with a population of over 50,000, which is expected to exert a negative effect on the conversion hazard. The second, *near public lands,* is binary and indicates whether public lands are nearby the pixel (within the outer window mentioned, above) (ESRI 2000 a and b; NCS 2006). The third proximity metric, *distance to hazardous waste site*, measures the Euclidean distance to the nearest hazardous waste site (CGIA 2006). These latter two variables are hypothesized to have positive and negative coefficients, respectively, through their effects on the amenity value of the pixel.

Additional pixel-level variables are included in the model that also do not change with time, including *elevation*, *slope,* and dummy variables indicating forest cover (*forest*) or wetlands (*wetland*) (USGS 1992). All of these variables are expected to have negative effects given higher conversion costs as well as higher opportunity costs associated with pixels under mature or ecologically important vegetation.

Three time-varying, county-level variables are also modeled, the first of which is a measure of returns-to-agriculture to capture the opportunity costs of commodity uses (*agricultural returns*). This metric, which is expected to negatively affect the conversion hazard, is calculated as county total farm receipts less costs, divided by farm acreage in the county (USDA 1997). Two additional time-varying indicators of county-level socioeconomic conditions included in the model are the deflated *per capita income* and *population density* (BEA 2001). As proxies for increased demand for developed land, both variables are expected to increase the conversion hazard.

Finally, we include a set of county dummies and a set of year dummies indicating the beginning of each interval. The former serve to limit omitted variable effects arising from county-level differences in governance, zoning, and other factors that may be fixed over time while the latter control for the effects of autonomous shifts in the policy and economic environment that occur over time in the region as a whole.

5. RESULTS

Table 1 presents results of four complementary log-log models of the determinants of the hazard of conversion. Because interpretation of the coefficient estimates is complicated by the log-odds transformation of the dependent variable, the figures presented in the table are the transformed coefficients in terms of risk ratios, which are interpreted as the percent change in the hazard rate from a unit increase in the covariate. This value is obtained by subtracting one from e^{β} and multiplying the resulting value by 100.

Models 1 and 2 (the constrained models) both exclude the spatially lagged variables and are distinguished by whether they control for unobserved heterogeneity via the inclusion of a Gamma-distributed frailty term. Models 3 and 4 (the unconstrained models) are likewise distinguished by the inclusion of a frailty term but additionally include the spatially lagged variables. A likelihood ratio (LR) test of Model 2 versus Model 1 yields a chi square statistic of 6.57 with one degree of freedom, suggesting statistically significant frailty. The same conclusion is reached in a comparison of Model 3 versus Model 4, which yields a chi square of 29.23. As expected, the estimated coefficients from the frailty models, when significant, are uniformly higher than in the non-frailty models, in some cases substantially so.

Interestingly, the estimated Gamma variance in Model 4, 4.70, is over four times the magnitude of the corresponding estimate of 1.10 in Model 2, despite the fact that the inclusion of the spatially lagged variables in Model 4 significantly reduces the log-likelihood relative to Model 2. While it might be expected that additional covariates reduce the unobserved heterogeneity, Fielding (2004) notes that interpretation of changes in intercept variances is difficult for the case of generalized linear models. The addition of covariates induces an implicit scale change as the model is re-standardized to maintain the error variance of $\pi^2/6$, thereby rendering it impossible to draw direct comparisons.

Table 1: Complementary log-log model of the hazard of conversion to impervious surface

	Constrained		Unconstrained	
	1. Non-frailty	2. Frailty	3. Non-frailty	4. Frailty
Forest (1,0)	-49.449	-53.989	-49.061	-67.656
	(0.000)	(0.000)	(0.000)	(0.000)
Wetland (1,0)	-56.926	-59.250	-50.046	-52.236
	(0.000)	(0.000)	(0.000)	(0.007)
Slope (degrees)	1.686	1.308	1.761	1.958
	(0.688)	(0.775)	(0.686)	(0.755)
Elevation (meters)	0.574	0.597	0.173	-0.043
	(0.004)	(0.006)	(0.378)	(0.878)
Distance to city (km)	-3.229	-3.558	0.415	0.073
	(0.000)	(0.000)	(0.490)	(0.928)
Distance to road (km)	-60.979	-63.181	-41.500	-42.961
	(0.000)	(0.000)	(0.000)	(0.000)
Distance to chipmill (km)	-0.393	-0.329	0.152	0.300
	(0.120)	(0.207)	(0.560)	(0.302)
Near public lands (1,0)	80.990	112.691	20.340	45.974
	(0.000)	(0.000)	(0.099)	(0.039)
Distance to hazardous waste site (km)	-16.060	-17.320	-8.098	-8.957
	(0.000)	(0.000)	(0.000)	(0.000)
Per capita income ($1000/person)	17.356	16.501	12.389	9.001
	(0.057)	(0.069)	(0.167)	(0.343)
Population density (people/km^2)	1.159	1.457	1.091	1.923
	(0.030)	(0.009)	(0.049)	(0.002)
Agricultural returns ($1/acre)	-0.267	-0.283	-0.297	-0.399
	(0.003)	(0.002)	(0.001)	(0.000)
Inner impervious surface (%)			16.940	28.086
			(0.000)	(0.000)
Inner impervious surface squared (%)			-0.157	-0.221
			(0.000)	(0.000)
Outer impervious surface (%)			-1.742	-3.166
			(0.009)	(0.017)
Percent water (%)			3.894	4.366
			(0.010)	(0.015)
intercept	-6.828	-6.848	-7.765	-7.866
	(0.000)	(0.000)	(0.000)	(0.000)
Chi^2 county dummies	113.560	110.40	58.040	54.300
	(0.000)	(0.000)	(0.000)	(0.001)
Chi^2 time dummies	148.490	145.841	142.760	139.400
	(0.000)	(0.000)	(0.000)	(0.000)
Gamma variance		1.098		4.696
		(0.035)		(0.000)
Log-likelihood	-2452	-2449	-2229	-2190

p-values in parentheses; number of observations=65991

Turning to the coefficients of the spatially lagged variables in Models 3 and 4, all are seen to be highly significant, with the inner ring variable having the strongest positive effect on the conversion hazard. Its magnitude, however, decreases with increases in impervious surface, as evidenced by the negative coefficient of the squared term. Increased water surface also has a positive but somewhat weaker effect, with the estimate from Model 4 suggesting that a 1% increase of water surface in the window increases the hazard by 4.4%. The only spatially lagged variables metric having a negative effect is that measuring the outer band of impervious surface, pointing to the presence of varying parameters across adjacent bands surrounding the pixel. This finding confirms results obtained by Irwin and Bockstael (2002), who employ similarly constructed variables derived from parcel-level data in Maryland to test for spatial externalities. The negative coefficient in theirs and the present study suggests that existing development in the vicinity of an undeveloped pixel reduces the hazard of conversion, a likely reflection of preferences for open space.

Beyond improving the fit of the model, the inclusion of the spatially lagged variables produces several noteworthy discrepancies with respect to the significance and magnitude of the remaining covariates. The coefficient on *elevation* is significant but unexpectedly positive in Models 1 and 2, a counterintuitive result that is insignificant in Models 3 and 4. Another discrepancy is seen with respect to the effect of distance to the nearest city. The estimates from Models 1 and 2 indicate this variable to be a negative and highly significant determinant of the conversion hazard, an effect that fades away with the inclusion of the spatially lagged variables in Models 3 and 4. Returns to land conversion evidently bear little relation to the proximity to the urban core when controlling for the influence of the immediately surrounding landscape pattern.

The effect of distance to the nearest road is, as expected, negative, but differs substantially depending on whether the spatially lagged variables are included. The estimate from Model 4 suggests that a one kilometer increase in distance produces a 43% decrease in the hazard of conversion, an effect that is roughly a third lower than the corresponding frailty estimate from Model 2. Even larger differences across the models are seen for the dummy indicating proximity to public lands and the distance to the nearest hazardous waste site, both of which are reduced by at least half in Model 4 compared with Model 2. The positive sign on the latter is counterintuitive at first glance, but may reflect, among other things, the potential for larger tracts to be at risk of development in less desirable neighborhoods because land is less expensive (Alonso 1964).

The variables measuring the return to agricultural land uses and population density have the hypothesized negative and positive effects, respectively, on the hazard of conversion. Likewise, the forest and wetlands dummy both have the expected negative coefficients. Based on the results from Model 4, forested pixels have a roughly 68% lower hazard of conversion than non-forest pixels, with the corresponding magnitude for wetlands at 52%.

While the coefficients of the 27 county dummies in the model are not shown in the table, using a chi-square test of their joint significance we reject the hypothesis at the 1% level that all of these coefficients are zero in all models. Finally, joint tests of the year dummies are also found to be statistically significant at the 1% level.

5. VALIDATION

To explore the validity of the models, we developed maps indicating the pattern of development between 1976 and 2001 (i.e., across all four intervals) that is observed and that is predicted by the models. For this visual inspection, we depict a 'surface' of estimated conversion probabilities, generated using a nearest-neighbor algorithm (Childs 2004). The performance of the constrained models in Figure 1is not impressive, even without the others for comparison. Many of the observed conversions are found outside areas of high predicted probability. Although not necessarily problematic, the overwhelming influence of particular variables is readily apparent: Due to *distance to city,* areas of highest predicted probability are tightly concentrated around city centers. The county dummies work to keep most of the high probability areas within particular county boundaries. *distance to road* contributes to the irregular pattern within each county.

FIGURE 1A Predictions from constrained model, non-frailty:

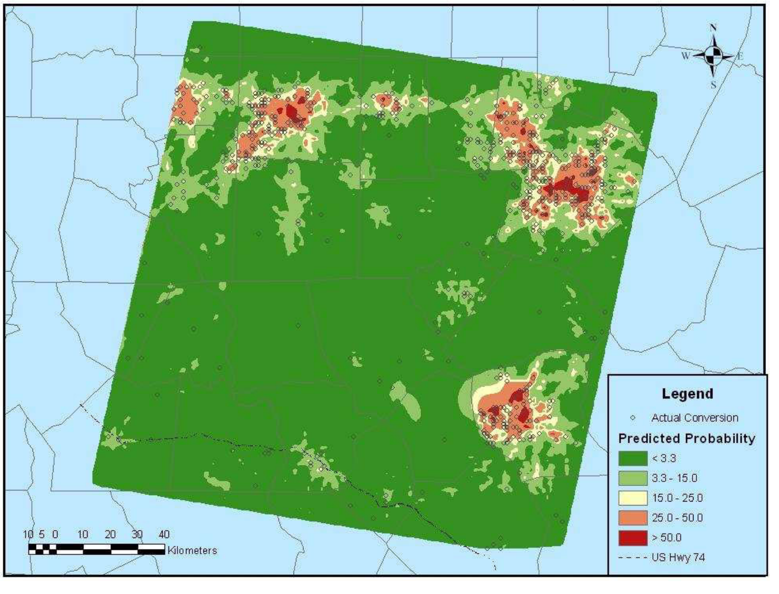


FIGURE 1B Predictions from constrained model, frailty:

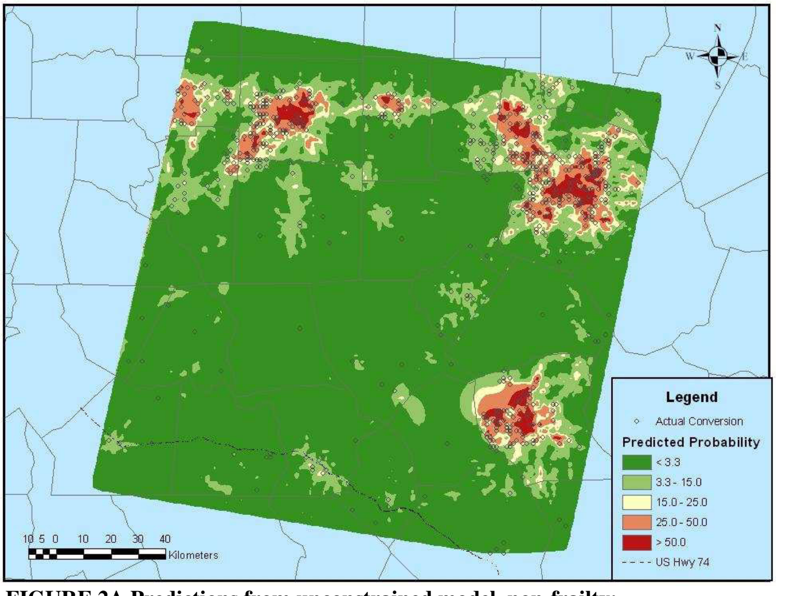


FIGURE 2A Predictions from unconstrained model, non-frailty:

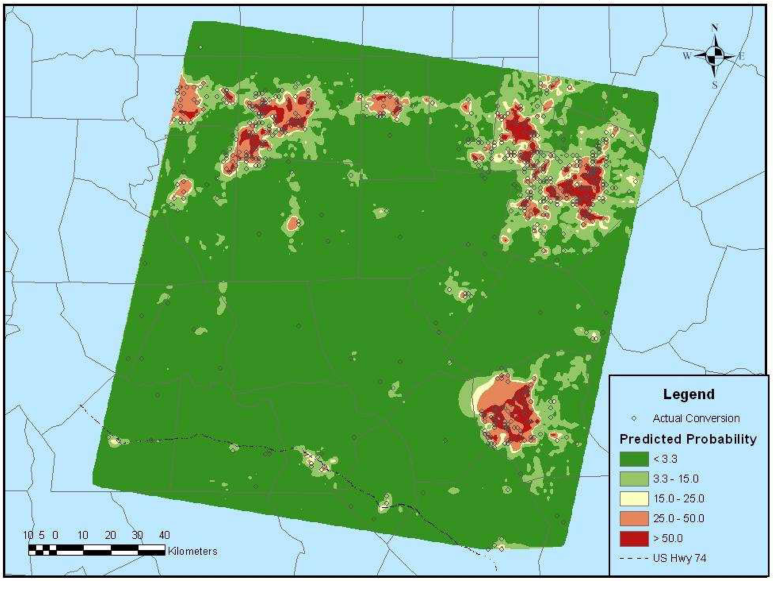
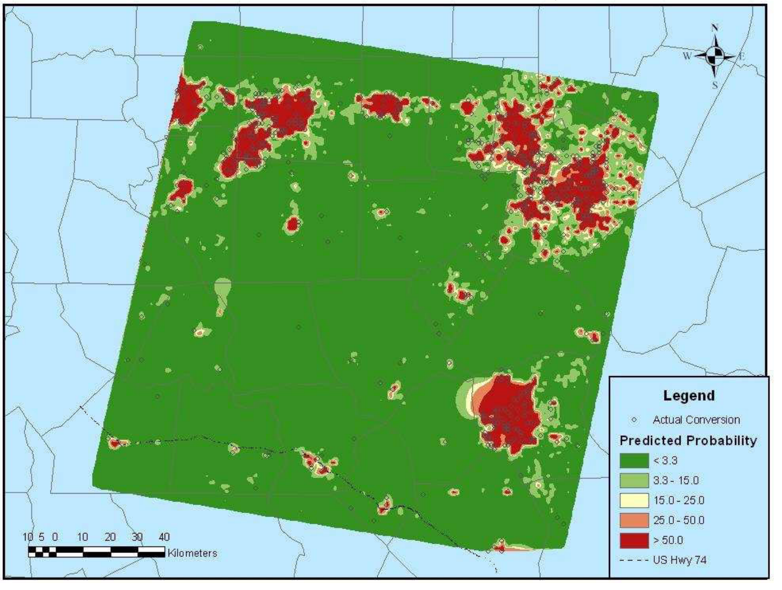


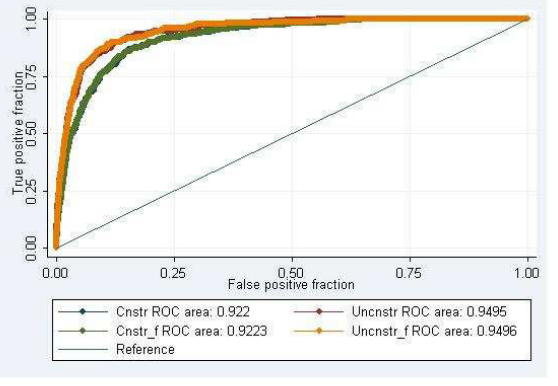
FIGURE 2B Predictions from unconstrained model, frailty:



The maps from the unconstrained models in Figure 2 present a clearly superior picture: The models not only appear to capture the pattern of conversion around metropolitan areas better, but high predicted conversion probabilities tend to be coincident with the more dispersed actual occurrences. This is especially evident in the frailty model, which tends to ascribe higher probabilities to regions that converted. Development well outside North Carolina's largest cities, such as that around Burlington in the north-center of the scene or along Route 74, is also predicted by the unconstrained models to be in the vicinity of where it actually occurs.

As a final diagnostic check on the performance of the models, receiver-operating characteristic (ROC) curves are presented in Figure 3, which plot the percentage of converted pixels correctly forecast (on the y-axis) against the percentage of non-converted incorrectly forecast (on the x-axis) for each possible prediction threshold. The area under the ROC curve, which ranges from zero to one and is non-parametric, can be interpreted as the proportion of correct forecasts across all possible thresholds. The closer the ROC curve is to the diagonal, the less useful is the model for discriminating between open space and converted pixels. Comparing the four curves, we see that the incorporation of frailty has, by this measure, a negligible impact on the model performance for both the constrained and unconstrained sets of models. The inclusion of the spatially lagged variables has a more notable impact on the ROC curves, increasing the area by roughly 0.03. Moreover, a chi-square test of equality in the areas generated by the constrained and unconstrained frailty models is, at 40.5, clearly rejected (P<0.001), providing further evidence that the spatially lagged variables significantly improve classification accuracy.

FIGURE 3: Receiver operating curves



6. CONCLUDING REMARKS

This paper has presented an application of a discrete-time hazard model as a means of analyzing the effects of time-varying socioeconomic and ecological covariates on the conditional risk that land is converted for developed use. Alternative variants of the model were estimated that were distinguished by controls for unobserved heterogeneity via the inclusion of a Gamma- distributed error component. In addition, alternative specifications were estimated to explore the influence of spatially lagged effects via the inclusion of explanatory variables derived from the satellite imagery. Both features are found to improve the model's ability to predict where the relative probability of change is high and where it is low. Moreover, the results suggest that controlling for landscape pattern and unobserved heterogeneity can have a substantial bearing on the conclusions drawn with respect to the influence of landscape features and socioeconomic factors, many of which have immediate relevance for policy planning.

From an urban planning perspective, the impact of major roads is of particular interest given the centrality of this variable to debate about the causes of sprawl. As noted by Hartgen (2003), isolating the relationship between roads and growth is difficult due to the confluence of factors that simultaneously determine each, including a region's economic health, prior growth, site suitability and demographics. The evidence presented here suggests that even after controlling for these determinants, proximity to roads has a statistically significant impact on the likelihood of conversion to developed use. Such information, particularly when linked to ecological impact models, can play a critical role in informing decisions relating not only to the siting of roads, but also to other landscape development features such as protected areas and hazardous waste facilities.

Nevertheless, in assessing the utility of econometric modeling of satellite imagery for planning purposes, it is also important to understand the implications of what this data source does not deliver. An important caveat in this regard is the inability to link the dependent and independent variables at the level of the decision unit, the land manager. The resulting spatial mismatch in measurement can obscure the relationship between the modeled determinants and the conversion hazard, particularly for variables measured at the county level. By reducing potentially important local heterogeneity to average values, variables such as per capita income – found here to be insignificant – must be interpreted with greater care since information at the county level may not necessarily correlate with localized circumstances.

A related shortcoming is the absence of information on the parcel boundaries within which the pixels are situated, as well as the potentially confounding role of zoning regulations. As an example of why this may be important, consider randomly selecting what appears to be a developable pixel using satellite data. If this pixel belongs to a lot in which the maximum allowable density has been reached, then it will in fact not be subject to development pressures. A large share of such pixels in the data could, in turn, impart a downward bias on the estimated effects of the variables included in the model. To the extent that variability in zoning plays out at the county level, this bias will be attenuated by the inclusion of the county dummy variables. Nevertheless, the modeling results should be interpreted in light of this issue.

As the availability of spatially-explicit land use data increases, it will become more feasible to merge satellite imagery with GIS layers denoting administrative and regulatory features and thereby purge the data of pixels that fall outside of the modeling framework. In the meantime, future research should explore approaches for making the human-environment linkage more explicit via the use of remotely sensed data. In this regard, one particularly promising area of inquiry comprises studies that combine georeferenced field surveys of land managers with satellite imagery. Such data would not only support testing of hypotheses derived from richer theoretical models, but could also significantly improve the predictive ability of the empirical models. Another extension for using the empirical model estimated in this paper to explore the issue of urbanization would relax the assumption of irreversibility of the conversion process. This would involve expanding the dependent variable to include multiple land use classes, which would allow for the estimation of competing risks models of land use change, including transitions from development to open-space.

REFERENCES

Alig, R., Healy, R., 1987. Urban and Built-Up Area Changes in the United States: An Empirical Investigation of Determinants. *Land Economics* 63: 215-226.

Allison, P. D., 1995. *Survival Analysis Using the SAS System: A Practical Guide*. SAS Institute Inc., Carey.

Alonso, W., 1964. *Location and Land Use: Toward a General Theory of Land Rent*. Harvard UniversityPress, Cambridge MA.

BEA (U.S. Department of Commerce), 2001. Local Area Personal Income Tables, www.bea.gov/regional/reis/ Accessed November 2006.

Boscolo, M., S. Kerr, A. Pfaff, Sanchez, A., 1998. What Role for Tropical Forests in Climate Change Mitigation? The Case of Costa Rica. Paper prepared for the HIID/INCAE/BCIE Central America Project.

CGIA (North Carolina Corporate Geographic Database Data Layers), 2006. http://cgia.cgia.state.nc.us/cgdb/catalog.html

Childs, C., 2004. Interpolating Surfaces in ArcGIS Spatial Analyst, *ArcUser*, July-September: 32-35.

Cox, D.R., 1972. Regression Models and Life Tables. *Journal of the Royal Statistical Society*, B34, 187-220.

Cropper, M., J. Puri, Griffiths, C., 2001. Predicting the Location of Deforestation. *Land Economics* 77: 172-186.

Duranton, G., and D. Puga, 2003. Microfoundations of Urban Agglomeration Economies, CEPR Discussion Papers 4062, C.E.P.R. Discussion Papers.

ESRI (Environmental Systems Research Institute, Inc.) 2000a. U.S. National Atlas Federal and Indian Land Areas, ESRI Data & Maps (CD 2).

ESRI (Environmental Systems Research Institute, Inc.) 2000b. U.S. GDT Park Landmarks, ESRI Data & Maps (CD 3).

Fielding, A., 2004. Scaling for Residual Variance Components of Ordered Category Responses in Generalised Linear Mixed Multilevel Models. *Quality and Quantity* 38: 425-433.

Geoghegan, J., Wainger, L.A., Bockstael, N.E., 1997. Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS. *Ecological Economics* 23: 251-264.

Hardie, I.W., P. Parks, P. Gottlieb, and D. Wear, 2000. Responsiveness of Rural and Urban Land Uses to Land Rent Determinants in the U.S. South. *Land Economics* 76: 659-673.

Hartgen, D. T., 2003. Highways and Sprawl in North Carolina. A report prepared for the John Locke Foundation. http://www.johnlocke.org.

Irwin, E.G.,, Bockstael, N.E., 2002. Interacting Agents, Spatial Externalities, and the Endogenous Evolution of Residential Land Use Pattern. *Journal of Economic Geography* 2: 31-54.

Jenkins, S. P. Survival Analysis, July 2005. Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK. www.iser.essex.ac.uk/teaching/degree/stephenj/ec968/pdfs/ec968lnotesv6.pdf. Accessed July 1, 2007.

Kline, J.D., Moses, A., Alig, R.J., 2001. Integrating Urbanization into Landscape-level Ecological Assessments. *Ecosystems* 4: 3-18.

Lancaster, T., 1990. *The Econometric Analysis of Transition Data.* Cambridge University Press, Cambridge.

Lubowski, R. N., Plantinga, A. J., Stavins R. N., 2003. Determinants of Land-Use Change in the UnitedStates 1982-1997. Resources for the Future, Washington D.C., Discussion Paper 03-47.

NCS (North Carolina Division of Parks & Recreation, Resource Management Program), 2003. http://www.ncsparks.net.

Powell, D.S., Taulkner, J.L., Darr, D.R., Zhu, Z., MacCleery, D., 1992. Forest Resources of the United States. USDA Forest Service, Gen. Technical Report RM-234.

Prestemon, J., Pye, J., Butry, D., Stratton, D., 2000. Economic Research Unit, USDA Forest Service. "Locations of Southern Wood Chip Mills for 2000," http://www.srs.fs.usda.gov/econ/data/mills/chip2000.htm, retrieved April 2003.

Schaberg, R. H., P. B. Aruna, F. W. Cubbage, G. R. Hess, R. C. Abt, D. D. Richter, S. T. Warren, J. D. Gregory, A. G. Snider, S. Sherling, and W. Flournoy. Economic and Ecological Impacts of Wood Chip Production in North Carolina: An Integrated Assessment and Subsequent Applications. *Journal of Forest Policy and Economics* 7: 157-174.

Sierra Club, 1998. Sierra Club Sprawl Report: 30 Most Sprawl-Threatened Cities, http://www.sierraclub.org/sprawl/report98/raleigh.asp, retrieved January 2003.

Stavins, R., A. Plantinga, and R. Lubowski, 2003. Determinants of Land-Use Change in the United States 1982-1997, Discussion Papers dp-03-47, Resources For the Future.

Turner, M.G., Wear, D.N., Flamm, R.O., 1996. Land Ownership and Land-Cover Change in the Southern Appalachian Highlands and the Olympic Peninsula. *Ecological Applications* 6: 1150-1172.

USDA (U.S. Department of Agriculture), 1997. Census of Agriculture.

USGS. EROS Data Center. National Land Cover Data 1992, http://edc.usgs.gov/products/landcover/nlcd.html, retrieved January 2003.