Assessing the Joint Influence of Ecological and Socioeconomic Determinants of Increases in the Built-Environment: A Study of Trends in Central North Carolina

Colin Vance & Rich Iovanna
German Aerospace Center, Institute of Transportation Research &
USEPA, National Center for Environmental Economics

Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Montreal, Canada, July 27-30, 2003

Assessing the Joint Influence of Ecological and Socioeconomic Determinants of Increases in the Built-Environment: A Study of Trends in Central North Carolina

Colin Vance and Rich Iovanna

For presentation at the AAEA meetings in Montreal, July 2003.

Introduction

The conversion of farm and forestlands on city-fringes throughout the United States has commanded attention from academics, policy makers, and the public since at least the early 1980s (Alig and Healy, 1987). Over the past two decades, the encroachment of rural lands by development has continued unabated, with the urbanized area expanding from approximately 51 to 76 million acres between 1982 and 1997 (Fulton et al, 2001). While partly reflecting growing prosperity and preferences for increased living space, this trend has raised concerns on several fronts. Through its strong association to the increase in impervious surfaces, expansion of the urban frontier degrades and fragments natural habitats, contributes to poor air quality through increased reliance on vehicle travel, and disrupts a multitude of ecosystem services such as aquifer recharge and nutrient cycling. Such disruptions can impose significant costs on municipalities, including damage from flooding, higher medical costs for air quality-related illnesses, and increased expenditures for the provision of public services and infrastructure. Social and aesthetic costs further compound these ecological and health impacts. The movement of populations away from central city areas not only contributes to urban blight, but also to a loss of cultural heritage as farmland and forest is replaced by what is often a pattern of helter-skelter development characterized by strip malls, office parks and disconnected residential communities.

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 of both why urbanization happens as well as where and when it happens. These questions are not only the concern of economists, whose focus is on the role of incentives, preferences and constraints in determining landscape patterns, but also to ecologists, whose analytical point of departure is the impact of these patterns on the physical and bio-geo-chemical flows that sustain ecosystem function. To the extent that present-day development decisions create landscape mosaics that constrain the choice set of future land use alternatives, a comprehensive understanding of urbanization requires integrating economic and ecological paradigms.

Recognition of the need for interdisciplinary approaches to the study of urban growth patterns has led to an increasing number of studies that combine principles from landscape ecology with spatial-econometric methods to account for how human decision-making, ecosystem function, and their interaction effect landscape changes across different spatial scales. In contrast to area-based approaches, which estimate the determinants of land use shares within aggregate geographic areas such as counties (e.g.

Lictenberg, 1989; Stavins and Jaffee, 1990; Parks and Kramer, 1995; Plantinga, Mauldin and Miller, 1999; Hardie et al., 2000), this literature 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 and Alig, 2001), and the causes of fragmented development patterns among residential land parcels on the rural-urban interface (Irwin and Bockstael, 2002).

While the increasing availability of high-resolution remotely sensed data and improvements in geographic information system (GIS) technologies have enabled major advancements in the literature concerned with spatially explicit assessments of land-use change, the role of temporal dynamics in the patterns that emerge remains poorly understood. This can be partly attributed to the fact that spatial data is typically available over a limited range of discrete dates, thereby limiting the extent to which the effect of time can be explicitly modeled. The present paper attempts to build on the existing spatial literature by advancing an empirical model assessing how, over both time and space, changes in land-use on the rural-urban interface respond to changing economic and ecological conditions. The estimated model is not only spatially explicit, but, unlike work to date, it also parameterizes the effect of time on the risk of conversion. Together, these features allow for the use of the coefficient estimates to simulate short- and medium term projections of future land use patterns.

The study examines the determinants of built-up area across a 10,000-mile square swath across central North Carolina, an area that has undergone extensive conversion of forest and agricultural land over the last two decades. For present purposes, built-up area is defined as impervious surface, which includes paved surfaces, structures, and medium to high-density residential areas. Between 1976 and 2001, the area covered by impervious surface in the region more than doubled, from 252 to 568 square miles, 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, 1998). 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.

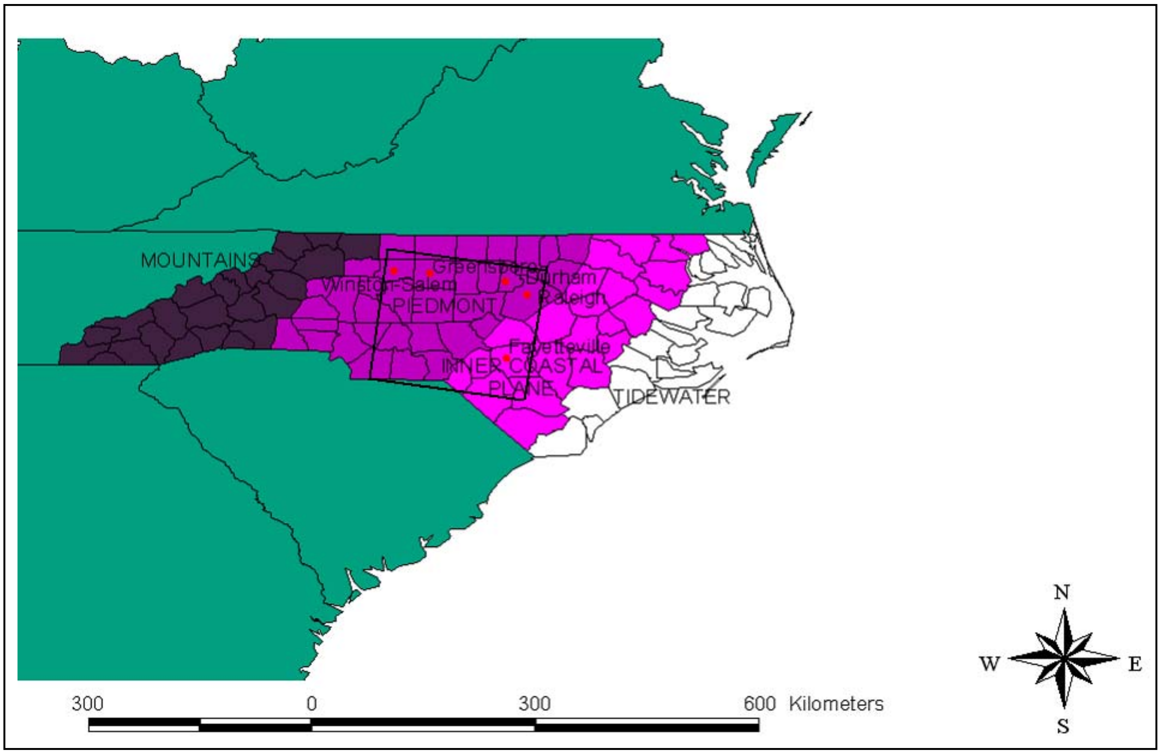
Our analysis takes as its point of departure a dynamic, profit-maximizing framework that suggests several possible determinants of land conversion from commodity-based to urban uses. Using 60 X 60-meter satellite pixels as the unit of observation, we subsequently test for the significance of these determinants 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. Because the data itself is observed at a very fine level of spatial resolution, we can 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 (2001), we are particularly interested in exploring the effects of pattern metrics, which are captured by three time-varying variables measuring fragmentation, the percent of impervious surface, and the percent of water in a two-by-two kilometer window surrounding the pixel. In addition to testing for the statistical significance of these variables and assessing their magnitude, we also gauge the extent to which their inclusion improves the predictive ability of the estimated model.

The Study Region

The study region straddles portions of the Piedmont and the Inner Coastal Plane of North Carolina, two distinct physiogeographic zones that cut diagonally north-south across the state (Figure 1). Elevations in the Piedmont range between 300 and 1500 hundred feet across a landscape primarily covered by deciduous and pine forests. The Inner Coastal Plane reaches maximum elevations of 500 feet and includes various wetland vegetation types such as gum cypress swamps and shrub bogs among its dominant land covers (Cooper, 2000; Bobyarchick and Diemer, 2000). Across the state as a whole, hardwoods cover more than half of the timberland acreage, while pine stands and oak-pine stands account for the remaining 33 and 14 percent, respectively (Brown, 1993; Brownlow et al., 2000). Centuries of human occupation have fragmented these forests by a patchwork of croplands, fields in varying stages of abandonment, and, increasingly, built-up areas. Virtually all of the woodland in North Carolina has been farmed at some point in the past and, excluding some swamplands and mountain slopes, there are no remaining virgin timber stands (Lilly, 1998). Among the state's most botanically diverse regions that has been threatened by human encroachment is the Carolina Sandhills, a longleaf pine/wiregrass ecosystem located in the in the southwestern corner of the Inner Coastal Plane and extending into South Carolina. Shaped by thousands of years of a natural fire disturbance regime, this ecosystem is currently in severe decline due to fire suppression efforts, clearing for agriculture, and logging. Today the ecosystem provides habitat for more than 30 plant and animal species, including the red-cockaded woodpecker, that are listed as threatened or endangered (U.S. Fish and Wildlife Service, 2003).

Figure 1: The study region boundaries and physiogeographic zones of North Carolina.



North Carolina is widely regarded as a state in which inefficiencies in land consumption are leading to excessively costly expansion of the built environment. A highly publicized report recently released from Smart Growth America (2003) ranked Greensboro and Raleigh-Durham as second and third among a listing of 83 U.S. cities in which the spread of development far outpaces population growth. Historical accounts suggest that the foundations for the sprawling patterns observed in these and other North Carolina cities can be traced back to the 1880s, when a low-density urban landscape emerged that was driven by the proliferation of tobacco factories and textile mills (Larsen, L. 1985; Orr and Stuart, 2000; Ingals, 2000). These employment centers spawned a dispersed network of small towns across the state that today serve as bedroom communities for regional metropolitan centers. By 1900 there were 177 mills in the state, with over 90% of them in the Piedmont (Ingals, 2000). To connect these emergent centers of economic activity, major investments in road infrastructure were undertaken with the result that by the early 1920s there were over 5,500 miles of roads were paved linking county seats ((Ingals, 2000). These developments ushered in a transition from an economy based largely on agriculture to one based on service sector and manufacturing industries, with heavy reliance on the forest-products sector.

While the state remains a major producer of tobacco, sweet potatoes, and hog products, the area under agriculture has declined drastically since its peak in the early 1900s (Lilly, 1998). The area under commercial timberland, by contrast, has remained relatively stable, peaking in the early 1970s to 20.13 million acres and then dropping back down to approach the 1938 level of 18.1 million acres by 1990 (Brown, 1993). Nevertheless, a recent U.S. Forest Service report projected that North Carolina will lose 30% of its privately owned, natural forest by 2040, with the Interstate 85 corridor extending southward from Raleigh-Durham designated as a "hotspot" of forest loss due to continuing urbanization (Prestemon and Abt, 2002; Wear and Greis, 2002).

Theoretical Considerations

Whether a land manager decides to convert a given tract of land depends 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 Parks (1995), Boscolo, Kerr, Pfaff, and Sanchez (1998), and Irwin and Bockstael (2002), the theoretical approach taken here attempts to reduce this complexity by assuming that land will be converted if the net present discounted benefits of doing so is greater than the net present discounted benefits of leaving the land under its present use. This approach considers that there is a continuous latent random variable reflecting net returns from pixel *i* in land use *m* (commodity) at time *t*, where returns are influenced by a vector of site and locational attributes of the pixel. Important site attributes include such factors as soil quality, slope, elevation, the cost of conversion and its value in the alternative use. Locational attributes include accessibility costs of the pixel to both roads and centers of economic activity as well as the existing pattern of land use surrounding the pixel. The land manager will choose the time of conversion, T, to maximize discounted net benefits from land at location *i*:

$$(1) \qquad Max_T = \int_0^T A_{it}(X_{it})e^{-rt}dt + \int_T^\infty D_{it}(X_{it})e^{-rt}dt - C_T e^{-rt}$$

where:

$A_{it}(X_{it})$ is the potential returns derived from a commodity-based use of the land at time t

$D_{it}(X_{it})$ is the potential returns derived from a developed use of the land at time t

C_T is the cost associated with conversion

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

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

The second condition considers that although conversion may yield net positive returns at time t, there may still be benefits to waiting because of the potential for even higher returns at some future date. Such a circumstance could arise, for example, in anticipation of improved technologies that reduce conversion costs. The second condition to be satisfied is thus:

$$(3) \qquad D_{it}(X_{it}) - A_{it}(X_{it}) - rC_{it} + \frac{d}{dt}C_{it} > 0$$

Because equation (2) is likely to be met well before equation (3), we subsequently focus on equation (3) to derive the empirical specification.

The Empirical Model

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 unobservable heterogeneity, we add an error term to equation 5:

$$(4) \qquad D_{it}(X_{it}) - A_{it}(X_{it}) - rC_{it} + \frac{d}{dt}C_{it} - \varepsilon_{it} \geq 0$$

The hazard rate – or instantaneous risk that pixel *i* is cleared in period *t* conditional on not having been converted before *t* – can then be expressed as

$$(5) \qquad h_{it} = \frac{f(\varepsilon^*(X_{it}, t)}{1 - F(\varepsilon^*(X_{it}, t))}$$

where *f(.)* and *F(.)* are the probability and cumulative distribution functions for ε, respectively, and ε^* is the ε that makes (6) an equality.

It bears pointing out that the hazard rate itself is not a probability, but rather a measurement of the number of events per unit interval of time, where an event is defined as some discrete transition across states. To estimate the determinants of h_{it} we draw upon a class of statistical methods collectively referred to as duration – or survival – analysis. Such methods have been applied widely in the biomedical, engineering, and social sciences to measure phenomena in which timing is the critical aspect of interest. Examples of such phenomena include the time until death after diagnosis of a terminal disease, the time until component failure, and the duration of unemployment spells. While conventional methods such as linear or logistic regression can sometimes be applied to these issues, these methods are generally ill-equipped to handle the features that often characterize duration data, including time-varying explanatory variables and censoring or truncation of the dependent variable.[1]

The data used in this study is interval censored, meaning that each observation's survival time is known only to fall somewhere between two dates. If a conversion occurs between the dates, the dependent variable assumes a value of 1; otherwise it assumes a value of 0. To reconcile the temporal continuity of the conversion process being modeled

[1] Truncation and censoring are pervasive features of duration data, resulting respectively from the data selection process inherent in the study design or from observation-specific random features that make observations on survival time incomplete (Hosmer and Lemeshow, 1999).

with this coarseness in the measurement of timing, we specify a complementary log-log model, written as:

$$(6) \qquad \log[-\log(1-P_{it})] = \beta' X_{it}$$

where P_{it} is the probability that pixel i is converted in interval t given that the pixel was not converted in any earlier intervals, β is a vector of estimated parameters, and the X_{it} are exogenous covariates. This model is a discrete analogue Cox's proportional hazards model, a highly flexible specification that is estimated using partial likelihood methods. Two major advantages of the Cox model is that it readily accommodates time-varying covariates and that it requires no assumptions on the functional form of the baseline hazard rate or on the factors that may change this rate over time. This enables attention to be focused specifically on the effect of the covariates on the relative risk of a transition.

Unlike the Cox model, the complementary log-log model is estimated using maximum likelihood methods. Drawing from the discussion in Hosmer and Lemeshow (1999), let I_j denote the jth time interval, and a_i and b_i denote two known values that bound the observed time for the ith pixel. Then define y_{ij} as a binary variable indicating the specific time interval observed for the ith pixel, where

$$(7) \qquad \begin{aligned} y_{ij} &= 1 \text{ if } (a_i, b_i) = I_j \text{ and} \\ &= 0 \text{ otherwise.} \end{aligned}$$

Now define a "pseudo" binary outcome variable as $z_{ij} = y_{ij} * c_i$, where $c_i = 1$ if the event occurred and $c_i = 0$ otherwise. Under the assumption that the underlying survival model is distributed as a type I extreme value (Irwin, 1998; Hosmer and Lemeshow, 1999), the model parameters can be estimated by maximizing the following likelihood function:

$$(8) \qquad \ell(\beta) = \prod_{i=1}^{n} \prod_{j}^{k_i - 1 + c_i} (1 - P_{ij})^{1 - z_{ij}} P_{ij}^{z_{ij}}$$

where n denotes the number of pixels and k_i denotes the observed interval for the ith pixel. The estimated coefficients from maximizing this function have a relative risk interpretation, just as with the Cox model. Because the model is estimated using maximum likelihood, however, it is also possible to readily generate estimates for the effect of time on the odds of a transition, an effect that cancels out when using the partial likelihood approach (see Allison, 1995 for further discussion).

Data and Methods

The Dependent Variable

The econometric model presented in this paper is estimated using a time series of five classified satellite images across a 25-year time span that includes the years 1976, 1980, 1986, 1993 and 2001. The images are taken from the northern half of path 16, row 36 and the southern half of path 16, row 35 of the Landsat satellite orbit, a region covering roughly 10, 000 square miles across central North Carolina. Data for the years 1976 and 1980 was derived from the Landsat Multispectral Scanner (MSS) imaging system, while the Landsat 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 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 across 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. After overlaying two GIS layers of tenure data from ESRI and the North Carolina Department of Parks and Recreation, those pixels falling under public ownership (e.g. national, state, and municipal parks) were also eliminated. The final data set used for model estimation comprised 65,999 observations.

Upon classifying the imagery, pixels were systematically sampled along a grid pattern across the satellite scene such that roughly twelve kilometers separated each pixel on a side. 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; Cropper, Puri and Griffiths, 2001; Kline and Alig, 2001). The consequences of spatial autocorrelation include inefficient but asymptotically unbiased estimates. However, in cases in which the unobservable variables are spatially correlated with the included explanatory variables, the coefficient estimates on the included variables will additionally be biased (Irwin and Bockstael, 2001). A major source of spatial autocorrelation arises from multiple observations falling under common landowners (Kline and Alig, 2001). Given that the average size of private forest ownership in North Carolina is 24 acres (Powell et al., 1992) while the average farm size is approximately 184 acres (U.S. Census of Agriculture, 1997), twelve kilometer separation was deemed an adequate distance to sufficiently reduce the likelihood of this occurring.

The Explanatory Variables

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 likelihood of land use conversion. Table 1 presents descriptive statistics and the units of measurement for each variable; an appendix lists web sites from which data was downloaded, if applicable.

Five variables are included in the model that do not change with time: elevation (elev), slope, and dummy variables indicating forested pixels (forest), poor soil quality (poorsoil), and wetlands. The measures of elevation, slope and the forest dummy were derived directly from the satellite imagery. Soil quality data was taken from the Land Capability Classes of the USDA Soil Conservation Service, which indicates the soil's suitability for agriculture. The wetland category was derived from the 1993 land use and land cover data from the EROS Data Center.

To capture the influence of what Healy (1985) has termed *juxtaposition effects* – or "spatially bounded externalties that affect adjoining or nearby land" (Alig and Healy, 1987: 225) – we derived three time-varying window-based metrics from the imagery that measure the landscape configuration surrounding a pixel. The window size was set at approximately two square kilometers, an admittedly arbitrary area, but one which was based both on a judgment call of 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, Waigner, and Bockstael, 1997; Flemming, 1999; Irwin and Bockstael, 2002). The calculated metrics are: the percent of area classified as impervious surface (p_imper), the percent of area classified as water (p_water), and a fragmentation metric (ppu). The latter of these, ppu, was developed by Frohn (1998) and is defined as:[2]

$$PPU_{it} = \frac{m_{it}}{n * \lambda} \tag{9}$$

where *i* denotes the pixel, *t* denotes the date of the image, *m* is the total number of patches, *n* is the total number of pixels in the window, and λ is a scaling constant equal to the area of the pixel. Because *n* and λ are constants in our data, the metric essentially reduces to a count of patches.[3] Hence, as the landscape becomes more fragmented, ppu increases. In the model estimation, we additionally include the interaction of ppu with p_imperv (ppu_imperv) to further capture the ecological integrity of the landscape. We would expect, for example, that a landscape with a preponderance of vegetated area would provide a more favorable habitat for plant and animal species, but that this effect would by attenuated by the extent to which that vegetation is fragmented.

In addition to the window-based metrics, several time-varying proximity-based metrics are also included in the specification: the Euclidean distance to the nearest primary road (dis_road), the Euclidean distance to the nearest woodchip mill (dis_mill), and a gravity index (gravity). The variable dis_road is based primarily on the road network available from ESRI, but was modified using image interpretation of Landsat

---

[2] Frohn (1998) suggests that unlike conventional measures of fragmentation, ppu allows comparisons of landscape fragmentation across images having different spatial resolutions, raster orientations, and numbers of land cover classes.

[3] The metric does not, however, assume only integer values in our data because of the GIS algorithm used to calculate it.

data to reflect the conditions existing at the beginning of each interval. Distance to the nearest woodchip mill, a potentially important cost attribute of forestry operations, was obtained by overlaying a GIS layer of woodchip mill locations and their establishment dates that is available from the Economic Research Service of the USDA's Forest Service. To isolate the effect of this variable for forested pixels, we interact it with the forest dummy. The final proximity metric, termed a gravity index, is introduced to capture the joint influence of urban proximity and urban population on land use change. While several specifications exist for capturing this influence (Haynes and Fotherham, 1984; Shi, Phipps, and Coyler, 1997), we follow a specification similar to that used by Kline and Alig (2001), defined as:

$$(10) \qquad gravity_{it} = \sum_{m=1}^{3} \frac{\sqrt{population_{mt}}}{distance_{im}}$$

where m represents the three nearest cities having a population of greater than 25,000. The index thus allows multiple cities to exert an effect on land use change while giving greater influence to larger cities of closer proximity, thereby capturing the combined effects of population growth and its spatial distribution. Data on metropolitan population counts used in the index were obtained from the U.S. Department of Commerce, Bureau of the Census.

We include two indicators of county-level economic conditions – per capita income (pc_inc) and median house values (hous_prc) – that were also obtained from the Department of Commerce for the years 1982, 1987, and 1992. Per capita income values were linearly interpolated for years in which published data and the satellite imagery did not correspond. Because of volatility in the housing market, linear interpolation was not deemed appropriate for constructing the series for median house prices. We instead used a home price index from Freddie Mac to extrapolate median housing prices for years in which Census data is not available. The price index is available for Metropolitan Statistical Areas, thereby allowing us to maintain intra-regional variability.

Two variables are included in the model to control for the effect of time. The first is a trend variable (trend) that measures the years elapsed since the start date of the time series in 1976. The inclusion of this trend as a covariate is a distinguishing feature of the model, as it enables us to parameterize the direction of duration dependence. That is, by controlling for the effects unobservable inter-temporal factors that affect land use change, we can answer the question of whether the conditional risk of conversion is an increasing or decreasing function of time. The second variable is derived by taking the natural log of the interval length (ln_span), and is introduced to control for the fact that the intervals are of differing lengths.

Finally, dummy variables representing the 31 counties in the region are included in the model to limit omitted variable effects arising from county-level differences governance, zoning, and expenditures that may affect land use.

Results

Table 2 presents results of two complementary log-log models of the determinants of increases in impervious surface. The first model is distinguished from the second by its exclusion of the window-based metrics. As interpretation of the coefficient estimates from the complementary log-log model is complicated by the log-odds transformation of the dependent variable, we derive a more intuitive interpretation through calculation of the "risk ratio". In the case of the continuous covariates, the risk ratio is interpreted as the percent change in the hazard rate from a unit increase in the covariate. These values are obtained by subtracting one from e^{β} and multiplying the resulting value by 100. In the case of the dichotomous variables, the risk ratio is simply equal to e^{β}, and can be interpreted as the ratio of the estimated hazard for observations with a value of one to the estimated hazard for those with a value of zero (Allison, 1995).

While Models 1 and 2 are both highly significant, with chi square values of 1626.10 and 2439.08, respectively, a likelihood ratio test of the null-restrictions imposed by Model 1 on the effects of the window based metrics suggests that it be rejected in favor of Model 2. The chi square value of the test is 285.04 with 27 degrees of freedom, providing clear-cut evidence that the metrics improve the fit of the model. As an additional gauge of the predictive performance of the two models, we calculated Goodman and Kruskal's gamma, a non-parametric, symmetric metric that is based on the difference between concordant (C) and discordant (D) pairs of predicted and actual values of the dependent variable as a percentage of all pairs ignoring ties. Gamma is computed as (C – D)/(C + D), and can be interpreted as the contribution of the independent variables in reducing the errors of predicting the rank of the dependent variable.[4] The value of gamma calculated from the constrained model is 0.809, while that of the unconstrained model is 0.909. The improvement in the predictive ability of the model with the inclusion of the window metrics is thus considerable, reducing the fraction of uncertainty remaining in the constrained model by 52 percent.

The statistical significance and magnitude of the coefficient estimates on the window-metrics provide further evidence of their importance as determinants of land use change. Increases in fragmentation, as measured by ppu, increase the hazard of conversion, as do increases in impervious surface. The former result likely reflects the deleterious impacts of fragmented land on commodity-based or recreational uses that rely on ecosystem viability, while the positive coefficient on p_imper may reflect the effect of agglomeration economies from positive spillovers of existing development. Evaluated at the mean of ppu, we find that a percent increase in p_imper induces a 9.97 percent

---

[4] As an illustration, consider the following hypothetical list of predicted and actual values: a. (0.9, 1); b. (0.3, 0); c. (0.8, 1); d. (0.2, 1). From this list, *ab* and *bc* are concordant pairs as the value of both elements is higher (or lower) in one set then in the other. *bd* is a discordant pair because the value of one element is higher in one set while the value of the other element is higher in the second set. Pairs *ac, ad,* and *cd* are ties and are therefore ignored in the calculation of gamma. Gamma in this example equals 0.33.

increase in the hazard of conversion. It is notable that Irwin and Bockstael (2001) obtain a contrary finding on a similarly constructed variable measuring the percent of developed area. Their study focuses primarily on explaining leap-frog development of land parcels in exurban areas, and they interpret the negative coefficient as representing "repelling effects" that result from negative externalities among residential land parcels. We attempted to replicate their result by limiting the sample to pixels located beyond 10, 15, and 20 kilometer gradients of the nearest city of greater the 25,000, but found the positive and significant parameter estimate to be robust.[5] As pointed out by Irwin and Bockstael and confirmed empirically by Geoghegan, Waigner and Bockstael, however, the direction of landscape pattern effects may vary over different window sizes, a possibility that data constraints precluded us from pursuing. With respect to the interaction of ppu and p_imper, we find evidence for different pressures on highly fragmented land with a preponderance of impervious surface as compared with highly fragmented land with a preponderance of vegetation. The latter has a higher risk of conversion, which is consistent with the idea that the marginal returns to development on such land is higher than on land already predominated by development.

Beyond improving the fit of the model, the inclusion of the window metrics produces several noteworthy discrepancies with respect to the proximity and ecological covariates. The statistical significance of these variables is either absent in Model 2 or their magnitude is substantially reduced. Model 1 confirms evidence found elsewhere in the literature for what has been dubbed "urban influence potential" (Shi, Phipps, and Coyler, 1997; Hardie, Narayan, and Gardner, 2001; Kline and Alig, 2001), whereby the positive and highly significant coefficient estimate on the gravity index suggests an increased hazard of conversion on those pixels in close proximity to urban areas expanding in population. By contrast, in Model 2 we fail to reject the null that the coefficient estimate on gravity is significantly different from zero. Likewise, while Model 1 supports the hypothesis that higher elevations and poorer soil quality significantly increase the hazard of conversion to developed uses, in Model 2 both variables are insignificant. With respect to the variable dis_road, both models are consistent with the hypothesis that decreasing primary road proximity discourages peripheral location through increases in the money and time cost of travel per kilometer, but the coefficient estimate in Model 2 is nearly half that of Model 1. A discrepancy is also evident with respect to the variable dis_mill, though the negative effect of the variable in Model 1 is just significant at the 10% level and its economic significance is questionable given the small magnitude of the coefficient estimate.

The remaining statistically significant variables across the two models are largely in agreement. Model 2 indicates that the hazards of conversion for forested and wetlands

---

[5] A possible explanation for this result traces back to the dispersed pattern of urban development, organized around mill towns, that emerged in North Carolina at the turn of the century. To the extent that a leap frog pattern of development was already established at this time, subsequent development occurring at the end of the century may have been driven largely by urbanization economies arising from city size itself. Indeed, the notion that a contemporary leap-frog development pattern prevails in U.S. cities may be open to scrutiny. As noted by Duranton and Duga (2003), 1.9% of the land area in the U.S. was built-up or paved by 1992, with almost all recent development being less than 1 kilometer away from earlier development despite the wide availability of open space.

pixels are 48 and 61 percent of the hazard for those pixels not having these attributes, with similar magnitudes seen in Model 1. These findings are consistent with the hypothesis of higher conversion and opportunity costs associated with pixels under mature or ecologically important vegetation. Per capita income and median housing prices are also seen to be significant determinants of conversion. Based on the results from Model 2, every 1000 dollar increase in per capita income increases the hazard by roughly 28.34% while every 1000 dollar increase in the median price of a house increases the hazard by 3.26%. The former result is consistent with the hypothesis that land is a normal good, so that increases in real incomes cause households to demand larger plots. The latter result confirms the intuition that as prices for houses increase, the hazard of conversion will increase as developers seek more land to profit from the sale of housing services (Deaton and Laroque, 1998). The negative coefficient on the trend variable in Model 2 suggests that the land use conversion process is characterized by negative duration dependence, with each year decreasing the hazard by 33.59%.

Conclusions

This paper has presented an application of a hazard model as a means of analyzing the effects of static and time-varying socioeconomic and ecological covariates on the conditional risk that land is converted for developed use. Our analysis confirmed several findings uncovered elsewhere in the literature, including significant impacts of per capita income, housing prices, and road proximity on the likelihood of conversion. In contrast to previous studies, however, we did not find unambiguous support for the hypothesis that the combined effects of proximity to and population of multiple metropolitan centers, as measured by the gravity index, is a significant determinant of land use. Moreover, the magnitude of our estimate of the impact of road proximity was substantially reduced when the model specification included the window metrics. Taken together, these results point to the risk of incorrect inferences with respect to the role of the spatial relations described by proximity metrics when the spatial relations described by pattern are not included in the model. Moreover, given that the gravity index includes an argument for population, land-use change scenarios based on projected values of population may lead to faulty conclusions that do not provide reliable information for policy analysis.

There are several possible extensions for using the empirical model estimated in this paper to explore the issue of urbanization. One extension would assess the significance of ownership patterns on the hazard of conversion, and could include tests for parameter consistency across federal and private tenures. Another possibility is to explore the effects of different specifications of the role of time by including squared and higher order terms in the model. Finally, the dependent variable could be expanded to include multiple land use classes, thereby allowing for the estimation of competing risks models of land use change.

References

Alig, R. and R. Healy. 1987. "Urban and Built-Up Area Changes in the United States: An Empirical Investigation of Determinants," *Land Economics* 63(3): 215-226.

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

Anas, A. 1999. "The Costs and Benefits of Fragmented metropolitan Governance and the New Regionalist Policies," Paper prepared for presentation at "Regionalism: Promise and Problems," a symposium held at the State University of New York at Buffalo School of Law, March 6, 1999.

Bobyarchick, A.R. and J.A. Diener. 2000. "Land Regions and Geology," in: Orr, D.M. Jr. And A.W. Stuart (eds.), *The North Carolina Atlas: Portrait for a New Century.* The University of North Carolina Press, Chapel Hill, N.C.

Boscolo, M., S. Kerr, A. Pfaff, and A. Sanchez. 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.

Brown, M.J. 1993. *North Carolina's Forests, 1990*. USDA Forest Service Technical Report. Asheville, N.C.: USFS.

Brownlow, B., N. Lineback, J. DeHart. 2000. "Forest Sector Strategies," in Lineback, N. (ed.), *North Carolina's Sensible Greenhouse Gas Reduction Strategies*. Report of the Department of Geography and Planning, Appalachian State University.

Cooper, A.W. 2000. "Vegetation," in: Orr, D.M. Jr. And A.W. Stuart (eds.), *The North Carolina Atlas: Portrait for a New Century.* The University of North Carolina Press, Chapel Hill, N.C.

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

Deaton, A. and G. Laroque. 1998. "Housing, Land prices, and the Link between Growth and Saving," http://www.worldbank.org/research/projects/savings/pdffiles/deaton.pdf, retrieved 5/12/03.

Duranton, G. and D. Puga. 2003. "Micro-foundations of Urban-Agglomeration Economies," (preliminary draft) in J. Vernon Henderson and Jacques-Francois Thisse (eds.), *Handbook of Regional and Urban Economics*, Volume 4. North-Holland, Forthcoming.

Flemming, M. 1999. "Growth Controls and Fragment Suburban Development: The Effect on Land Values," *Geographic Information Sciences*, 10.

Frohn, R.C. 1998. *Remote Sensing for landscape Ecology: New Metric Indicators for Monitoring, Modeling, and Assessment of Ecosystems*. CRC Press LLC: Boca Raton, FL.

Fulton, W., R. Pendall, M. Nguyen and A. Harrison. 2001. "Who Sprawls Most? How Growth Patterns Differ Across the U.S." Center on Urban & Metropolitan Policy, The Brookings Institution.

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

Goodman, L. and W.H. Kruskal. 1954, 1959, 1963. Measures for association for cross-classification, I, II, III. *Journal of the American Statistical Association* 49: 732-764; 54: 123-163; and 58: 310-364, respectively.

Hardie, I.W., T.A. Narayan, and B.G. Grardner. 2001. "The Joint Influence of Agricultural and Nonfarm Factors on Real Estate Values: An Application to the Mid-Atlantic Region," American Journal of Agricultural Economics 83(1): 120-132.

Hardie, I., P. Parks, P. Gottlieb, D. Wear. 2000. "Responsiveness of Rural and Urban Land Uses to Land Rent Determinants in the U.S. South," *Land Economics* 76(4): 659-673.

Hosmer, D. and S. Lemeshow. 1999. *Applied Survival Analysis: Regression Modeling of Time to Event Data*. John Wiley and Sons, Inc., New York, N.Y.

Ingalls, G.L. 2000. "Urbanization," in: Orr, D.M. Jr. And A.W. Stuart (eds.), *The North Carolina Atlas: Portrait for a New Century.* The University of North Carolina Press, Chapel Hill, N.C.

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

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

Lichtenberg, E. 1989. "Land quality, irrigation development, and cropping patterns in the northern High Plains," *American Journal of Agricultural Economics* 71: 187-201.

Lilly, J.P. 1998. "North Carolina Agricultural History," retrieved 5/12/03 from http://www.ncagr.com/stats/history/history.htm of the North Carolina Department of Agricultural and Consumer Services.

Orr, D.M. Jr. And A.W. Stuart (eds.). 2000. *The North Carolina Atlas: Portrait for a New Century.* The University of North Carolina Press, Chapel Hill, N.C.

Plantinga, A.J., T. Mauldin, and D.J. Miller. 1999. "An Econometric Analysis of the Costs of Sequestering Carbon in Forests," *American Journal of Agricultural Economics* 81: 812-824.

Prestemon, J.P. and R.C. Abt. 2002. "Chapter 13: Timber Products Supply and Demand," *The Southern Forest Resource Assessment*. USDA Forest Service.

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

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

Stavins, R.N. and A.B. Jaffe. 1990. "Unintended impacts of public investments on private decisions: the depletion of forested wetlands," *American Economic Review* 80: 337-352.

Wear, D.N. and J.G. Greis. 2002. *The Southern Forest Resource Assessment Summary Report*. USDA Forest Service Southern Research Station.

Table 1: Descriptive statistics of variables used in the analysis

Variable name	Units	Mean	Standard deviation
dep. var, (1=conversion)	0,1	0.01	0.10
forest	0,1	0.63	0.48
Poorsoil	0,1	0.11	0.31
elev	meters	137.18	65.81
slope	degrees	0.60	1.19
wetland	0,1	0.12	0.32
dis_road	kilometers	1.41	1.32
dis_mill	kilometers	41.53	48.26
discity1	kilometers	34.61	19.88
discity2	kilometers	50.31	25.21
discity3	kilometers	58.70	26.83
popcity1	persons	77699.93	69999.52
popcity2	persons	73023.82	75239.52
popcity3	persons	79125.52	75177.71
gravity	index	25.26	23.77
hous_prc	1000s/dollars	77.23	25.01
pc_inc	1000s/dollars	15.78	4.10
ppu	index	9.60	6.40
pwater	percent	0.51	3.05
pimper	percent	1.69	6.26
ppu_imp	percent	46.56	205.94
ln_span	ln of time	1.80	0.26
trend	years	13.92	7.89

Table 2:Complementary log-log model of the hazard of conversion to impervious surface

	Model I		Model II	
	Coef. Est.	Risk ratio	Coef. Est.	Risk ratio
forest	–0.401	0.670	–0.734	0.480
	(0.004)		(0.000)	
poorsoil	0.371	1.449	0.147	1.158
	(0.003)		(0.265)	
elev	0.009	1.009	–0.001	0.999
	(0.000)		(0.682)	
slope	0.026	1.026	0.061	1.063
	(0.517)		(0.157)	
wetland	–0.739	0.478	–0.492	0.611
	(0.000)		(0.013)	
dis_road	–1.247	–71.266	–0.473	–37.706
	(0.000)		(0.000)	
dis_mill	–0.002	–0.235	–0.001	–0.144
	(0.101)		(0.328)	
gravity	0.130	13.919	–0.025	–2.490
	(0.000)		(0.405)	
hous_prc	0.032	3.206	0.032	3.258
	(0.000)		(0.000)	
pc_inc	0.298	34.654	0.250	28.339
	(0.000)		(0.001)	
ppu			0.116	12.316
			(0.000)	
pwater			0.021	2.165
			(0.191)	
pimper			0.128	13.623
			(0.000)	
ppu_imp			–0.003	-0.343
			(0.000)	
trend	–0.424	–34.590	–0.409	–33.586
	(0.000)		(0.000)	
ln_span	8.384	282.481	8.150	268.396
	(0.000)		(0.000)	
cons	–23.258	–100.000	–21.535	–100.000
	(0.000)		(0.000)	
chi^2 county dummies	285.040		73.500	
	(0.000)		(0.000)	
n_obs	65991		65991	
LR chi^2 (39, 43)	1626.1		2439.08	
	(0.000)		(0.000)	
log-likelihood	–2646.575		–2240.082	

p-values in parentheses

Appendix: Data sources available on the internet.

Variable	Data source	Website
badsoil	USDA National Resources Conservation Service	http://soils.usda.gov/soil_survey/main.htm
elevation	North American Land Characterization	http://eosims.cr.usgs.gov:5725/CAMPAIGN_DOCS/nalc_proj_camp.html
wetland	National Land Cover Data 1992	http://landcover.usgs.gov/natllandcover.html
d_mill	USDA Economics Research Unit	http://www.srs.fs.usda.gov/econ/data/mills/chip2000.htm
citypop	U.S. Census Bureau, County and City Data Books	http://fisher.lib.virginia.edu/ccdb/city94.html, http://eire.census.gov/popest/data/cities.php http://www.census.gov/prod/cen1990/cph2/cph-2-35.pdf
ppi for housing	Freddie Mac	http://www.freddiemac.com/finance/cmhpi/#new
pop_den	U.S. Census Bureau, County and City Data Books	http://fisher.lib.virginia.edu/ccdb/
pc_inc	U.S. Census Bureau, County and City Data Books	http://fisher.lib.virginia.edu/ccdb/