Searching for the Urban Fringe: Exploring Spatio-Temporal Variations in the Effect of Distance versus Local Interactions on Residential Land Conversion Using a Conditionally-Parametric Discrete-Time Duration Model

Douglas H. Wrenn, Abdoul Sam, and Elena G. Irwin

Agricultural, Environmental, and Development Economics

The Ohio State University

Selected Paper prepared for presentation at the Agricultural & Applied Economics
Association's 2012 AAEA Annual Meeting, Seattle, Washington, August 12-14, 2012

Searching for the Urban Fringe: Exploring Spatio-Temporal Variations in the Effect of Distance versus Local Interactions on Residential Land Conversion Using a Conditionally-Parametric Discrete-Time Duration Model

Douglas H. Wrenn[1], Abdoul Sam and Elena G. Irwin

[1]Primary Author: Douglas Wrenn, Ohio State University, 2120 Fyffe Road Room 103, Columbus, OH 43210, wrenn.7@buckeyemail.osu.edu.

Abstract

The spatial configuration of land use can have a significant impact on both market and non-market outcomes. One type of land use configuration that has received considerable attention in recent decades - because of its impact on the efficiency of public service provisioning, on environmental quality and on the productivity of agricultural land - is fragmented development. Many theories have been presented to explain these fragmented development patterns including variable densities, competing land uses and heterogeneity in agents' expectations. More recently, a number of authors have hypothesized that it is the reduction in congestion and the increase in open space amenities in exurban areas that have attracted people thus creating an increase in low-density, scattered development. While many of these hypotheses have been addressed empirically, most of the work has looked at these factors separately. In addition, most of the previous work has been conducted using global parametric models, which make it difficult to capture the spatial heterogeneity of the impact of the different factors and the resulting response. To address these shortcomings it is necessary to have micro-level data that can capture the spatio-temporal nature of the land conversion process and can separately identify the most important factors, on both sides of the market, as well as have a modeling technique that can capture the potential for spatial heterogeneity and non-stationarity in the effects of the variables across space. In this paper, we fill this gap in the existing land use literature by combining several unique panel datasets of historical land use change for Carroll County, Maryland with an original nonparametric modeling technique to separately identify the heterogeneous effect of amenities, local congestion and real options prices effects on the exurban land conversion decision. The results from our model show that factors on both sides of the market matter and that these factors are not spatially stationary. We find that much of the fragmented growth pattern observed in our study region can be explained by the differential effect of local land use interactions across space. This result is particularly important for policy makers interested in protecting farmland, reducing public costs and protecting valuable ecosystem services and biodiversity.

Keywords: Land-use modeling; Nonparametrics; Spatial analysis

1 Introduction

Over the last half century suburban populations have accounted for the majority of the population growth in urbanized areas (Nechyba and Walsh, 2004). As these suburban areas have expanded, population spillovers have created fragmented exurban development further from traditional population centers with majority of this development taking place as single-family residential developments.[2] Some recent research (Brown et al., 2005) has shown that exurban land forms now make up close to 25% of the land use in the lower 48 U.S. states. While the extent of urban sprawl and the resultant suburban and exurban patterns have been quantified (Brown et al., 2005; Burchfield et al., 2006; Irwin and Bockstael, 2007), researchers are just beginning to understand the complex economic, political and regulatory relationships that are driving the processes leading to these seemingly-chaotic spatial outcomes.

All land-use patterns begin at the micro-level as agents make decisions on the optimal time and location of land conversion. Given that many of the factors influencing the conversion decision are uncertain, land is most aptly viewed as a option on a real asset. Conditional on the information available in each period, the agent makes an optimal stopping decision regarding the best time to develop. This decision is influenced by many factors including future costs and prices as well as the characteristics of the parcel, local policies and the interactions between parcels in space. Given the spatio-temporal nature of the conversion process and the heterogeneity of the parcels and agents, it should be clear that these factors will impact the spatial pattern of development.

Until now, the primary explanation for fragmented residential development patterns has been demand driven with the hypothesis being that consumers trade off accessibility and agglomeration near traditional urban centers with reduced congestion and open space amenities in more exurban areas. As a result of this process, local land use interactions arise from either negative development or positive open space spillovers which leads to more low-density scattered development. However, this explanation disregards land use regulations and how they influence the process. The decision of the agent to exercise her option is dependent upon uncertainties in both prices and costs. If land use regulations create more, rather than less, cost uncertainty or if they create uncertainty differences across space, then it is entirely plausible that they could produce patterns and results similar to those produced by local land use interactions.

Most areas in U.S. have enacted some form of land use regulation designed to control certain features of the housing and land markets in an effort to manage the development process. Such regulations, which

[2] Fragmentation is usually defined as discontinuous development patterns with larger parcels of land being separated into ever-smaller pieces with like land uses interspersed with unlike uses.

include constraints on housing starts, density and subdivision type and location, are typically motivated by desires to minimize land use externalities, reduce public service costs or protect agricultural and sensitive environmental resources. Regardless of the motivation, the upshot is a change in the incentive structure on the supply side of the market and an alternation of the optimal timing of investment decision. While the majority of regulations represent a *de jure* increase in the *explicit costs* of land development, it is the *de facto* increase in *implicit costs* or uncertainty that have the largest potential impact on housing and land market investments.

If there are spatial differences in these implicit costs, then we hypothesize that they would impact the spatial pattern of development. Specifically, it is the unequal application of regulations that results in differences in these implicit costs across space that is the key factor in explaining the spatial pattern of land development in exurban and rural areas. As a result, this provides an equivalently plausible and potentially observationally equivalent explanation for the fragmented development patterns observed given that exurban and rural areas also have a preponderance of open space and less stringent regulations that constrain residential development. However, in order to separately identify the impact of regulatory uncertainty from the impact of amenity-driven growth and determine how each of these affects the timing decision across space, it is necessary to have data on regulations that can proxy for uncertainty and separately identify its effect from the effect of local land use interaction as well as have a model that capture the decision process in spatially-flexible way.

In this paper, we remedy both of these issues by employing a unique dataset, identification strategy and modeling technique that allows us to separately identify (from local land use spillovers) the effect of implicit regulatory costs on the the conversion decision and on the spatial pattern of new residential development. Using an original panel dataset on historical subdivision development from an exurban county in Maryland, we estimate a conditionally-parametric discrete-time duration model of land developer decision making regarding the timing of land conversion. Our dataset allows us to capture the intertemporal decision-making process at the individual level and control for parcel-level factors affecting that decision. It also contains information on the length of time that it took the county planning department to approve each subdivision starting from the point at which the subdivision application was filed. Using these data on approval times, we create a time-varying measure of regulatory uncertainty for each parcel that captures the influence of variation in implicit costs on the expected completion times for developments. We use this measure as an explanatory variable in our duration model of development timing. By comparing the differences in the estimated coefficients between our local land use variables and our measure of regulatory

uncertainty we can answer our first research question regarding the magnitude of each factor in explaining the decision to convert. Further, given that our model produces a coefficient value for each observation, we can map the results for both the local land use variables and regulatory variable to answer our second question about how the factors are impacting the spatial pattern of development.

The results from our model reveal a number of interesting results. First, regulatory uncertainty has a larger impact on the development timing decision than the local land use variables. Second, the mapping of the coefficients for our regulation variable reveals that regulations reduce the probability of development more in urbanized areas than in the exurban. This is evidence that differences in costs across space may be contributing the scattered exurban development pattern we observe. The remainder of the paper is structured as follows. Section 2 provides a review of the literature of the related literature, section 3 provides the basic theoretical model, section 4 presents our empirical specification, section 5 presents the data used in the model and the construction of the measure of implicit costs, section 6 presents the results and discussion and section 7 concludes.

2 Related Literature

Investment Under Uncertainty

Over the last several decades, a number of authors have recognized that the irreversibility of the land conversion decision makes it equivalent to an option on a real asset (Titman, 1985; Williams, 1991; Dixit and Pindyck, 1994). The theoretical prediction from this literature is that any factor that increases uncertainty over future profits will raise the "hurdle" rate on the option and alter the optimal investment time. These papers, however, are aspatial and provide little insight into impact of uncertainty on the spatial pattern of development. In a seminal paper, Capozza and Helsley (1990) were the first to integrate the real options investment model into the urban growth model to produce a spatially explicit real investment model with uncertainty over future land rents. They show that as uncertainty over future rents increases it raises the option value on raw land and reduces development at the urban fringe. While their model provides insight into how uncertainty impacts the location of the urban boundary, it fails to explain the discontinuous pattern of urban growth often observed in the real world. The results from their model shows that this development pattern happens in a continuous manner and that uncertainty over rents is neither necessary nor sufficient for discontinuous or leapfrogged development patterns. To explain these patterns, several authors have shown that the model must have, in addition to future uncertainty, either variable density (Wheaton, 1982),

multiple land use types (Fujita, 1976; Mills, 1981), or heterogeneous expectations among the agents (**?**).

Subsequent critiques of these early real investment models of urban growth pointed out that they assume an instantaneous payoff at the optimal investment time and ignore the intensity of inputs following the decision to invest. It was recognized that a more realistic model would incorporate a delay between the start of the project and the time at which the first revenues are realized (Majd and Pindyck, 1987; Dixit and Pindyck, 1994; Bar-Ilan and Strange, 1996). Bar-Ilan and Strange (1996) were the first to include lags in an urban spatial model of real asset investment. In their model, landowners who wish to capitalize on an investment opportunity must take into account the time lag between the initial development decision and the time at which the project is completed. As a result, the foregone returns and the benefits are uncertain and the effect of uncertainty on investment is ambiguous. Bar-Ilan and Strange (1999) model both the timing and the intensity decision and show that uncertainty increases the optimal timing of the investment decisions and decreases the optimal amount of capital invested.

All of this previous work, however, has focused exclusively on uncertainty over future returns and not considered the impact of uncertainty over costs. In the context of this paper, it is the uncertainty over the regulatory environment and implicit costs that alter the timing and spatial outcome of the process. In one of the first real investment models to consider a stochastic cost environment, Pindyck (1993) models a real investment project that takes several periods to complete with uncertainty over costs. In this model, costs are divided between *technical* and *input* costs. Technical costs depend on the stage of the project, whereas input costs are independent of the investment decision and are assumed to be non-diversifiable and exogenous to the actions of the project owner. Input costs in his model are analogous to our statement of implicit costs above. He finds that, while an increase in uncertainty over technical costs does not necessarily delay investment, a similar change in input costs always increases the option value and decreases investment.

Mayo and Sheppard (2001) adapt the Pindyck (1993) model of uncertain costs to model the impact of *stochastic development controls* on the timing and intensity of housing supply. In their model, development regulations are stochastic because they are flexibly applied by regulators. As the difficulty in anticipating the human response increases so does the variance in the expected delay of approval by the land developer. Their theoretical analysis shows that it is not the increase in expected completion times that impacts development, instead it is the variance in the timing and the relative change in timing between development types and locations. They consider the market consequences of these uncertain regulatory effects by examining individuals' responses to these controls in the context of an aggregate model of housing demand and supply. They derive conditions based on the relative rates of growth of the housing price index, the value of vacant

land, and the profits from immediate development that correspond to cases in which an increase in regulatory risk results in either an increase or decrease in the cutoff price that causes developers to reduce their housing supply to zero. When conditions cause the cut-off price to decrease, this reduces housing supply in higher valued housing areas and leaves supply in lower valued housing areas unchanged. While their model is not spatial, they assign a spatial interpretation to the results by interpreting lower valued housing areas as the periphery and thus interpret conditions under which the cut-off price declines as those under which"anti-centric" development is favored. Based on this interpretation, they conclude that "land speculation and leapfrog development (both of which involve developers holding land in a vacant state waiting to develop in future periods) may be actually caused by the behavior of the planners (or planning system)...Efforts to reduce the uncertainty which developers face may be the most effective remedy available for speculation."[3] Despite these provocative conclusions, no one has empirically tested the hypothesized effects of uncertain regulatory costs on individual development decisions and the spatial pattern of development. Doing so is the central aim of this paper.

Review of the Empirical Literature on Urban-Rural Land Conversion

Most empirical research on the interaction between regulation and the conversion decision has focused on the impact of regulatory stringency on aggregate housing supply decisions or house prices (Green, Malpezzi, and Mayo, 2005; Mayer and Somerville, 2000; Glaeser, Gyourko, and Saks, 2005a,b; Ortalo-Magne and Prat, 2007; Thorson, 1997; Malpezzi and Mayo, 1997). Glaeser, Gyourko, and Saks (2005a) look at long-run trends in the supply response of housing to demand shocks in metro regions across the U.S. They find evidence that the rise in house prices in many areas of the U.S. is the result of a change in regulatory regimes that have made large-scale development increasingly difficult in expensive regions of the country. They also show that cities with less regulation tend to be more sprawling. In a paper focused specifically on California cities, Quigley and Raphael (2005) develop a city-level index of regulatory stringency and relate this measure to a measure of local housing supply. They find that the supply response to a demand shock is significantly lower in more regulated cities. In a related paper, Mayer and Somerville (2000) use aggregate data on housing starts and a national survey of planners to estimate the impact of regulation-induced increases in development times on the number of new houses being built. They find that regions with increased approval times for subdivisions can have up to 45% fewer starts and elasticities more than 20% lower and that those

[3] A related theoretical literature not specifically focused on investment and uncertainty looks at the impact of numerous type of land use regulation such as development fees, minimum lot zoning, and growth boundaries (Brueckner, 1990, 1997; Lee and Fujita, 1997; **?**). Quigley and Rosenthal (2005) provides an excellent review of this and other literature related to land use controls.

regulations that lengthen the approval process serve to decrease the supply response most. Their results suggest that efforts to reduce development through increased approval times can be effective.[4]

While these previous papers find a relationship between increased regulatory stringency and a reduction in housing at the aggregate level, they are unable to evaluate the impact of uncertainty over regulations on individual development decisions. Consequently, they are unable to examine how heterogeneity in regulatory cost uncertainty (and the individual responses to it) differentially affects development timing and the spatial patterns and substitutions among locations. Given the theoretical evidence of the role of heterogeneity in explaining discontinuous development patterns (Mills, 1981; Wheaton, 1982), it is clear that a complete empirical analysis of the process producing these patterns requires analysis at the parcel level. Doing so requires micro-level data on land development, similar to that used in this paper, that can account for the impact of uncertainty on the behaviorial response of individual decision markers.

A number of recent papers have looked explicitly at the impact of land use regulations on the individual development decision. Cunningham (2006, 2007) uses a duration modeling framework to estimate the impact of the urban growth boundary around Seattle, WA. He finds that the expectation of the boundary increases development initially, but decreases it in long-run. Towe, Nickerson, and Bockstael (2008) uses a similar modeling technique, but estimates a real options model to look at the impact of the option to preserve a parcel on the timing of land conversion. They find that the preservation option has a significant and negative impact on the timing of development with large parcels which are eligible for development staying undeveloped longer than parcels which are not. In a model focused on the density of exurban and rural development, (Newburn and Berck, 2006) investigate how land use regulations differentially influence suburban versus rural development density beyond the urban fringe. They find that public service provisions impact suburban development, but that it does not significantly affect the likelihood of rural development. Similarly, McConnell, Walls, and Kopits (2006) examine the developer's decision about the density of development at the subdivision level, and analyze the influence of low density zoning regulations versus market forces. Their results show that factors affecting value and costs are important to density and that zoning may be leading to more scattered development pattern.

[4]See Fischel (2004) for a related literature that looks at the creation and impact of land use regulations from a political economy perspective and Sevelka (2004) for a comprehensive overview of the land use regulations from the appraisers' and planning perspective.

Empirical Models of Local Land Use Interactions

A number of recent papers have provided evidence of the role of local land use interactions in fostering lower density, more scattered development patterns particularly in exurban areas. Irwin and Bockstael (2002) estimate a spatially-explicit land use model of development timing and found evidence of negative spillover effects of neighboring development. They also find that the negative spatial spillovers predicted actual development patterns quite well. Walsh (2007) estimates a spatially-explicit vertical sorting model of land conversion that incorporates the endogeneity of the open space and land conversion decisions. One of his key findings is that location matters in determining the kind of open space policy applied. His model shows that spending on open space can lead to different results and spatial patterns depending on the location of parcel. Klaiber and Phaneuf (2010) estimate a horizontal sorting model and analyze how open space amenities affect residential location choices. They show that heterogeneity in preferences across households for open space are important in determining the welfare outcomes of conservation policies. In a model of joint development-density decision, Lewis, Provencher, and Butsic (2009) model the effect of open space conservation on the rate of growth in the density of land. They find that, contrary to previous work, if open space and parcel size are complementary in the land value function, then open space can actually reduce development timing and density.

These papers examine the land conversion decision and the spatial outcome of the process using micro-level data. This allows them to identify spatially heterogeneous policy effects, but they are limited in other ways. First, they do not explicitly model regulatory uncertainty and its impact on development. Many land use policies add explicit costs to development projects, but theory suggests that it is the unknown or implicit costs brought about by the flexible application of the policy and not the policy itself that matters most for the conversion decision. Second, they do not explicitly account for and compare the tradeoffs between the demand and supply-side factors in determining spatial patterns. And finally, while they use spatially explicit data, they model the development decision using a fully parametric model, which produces global estimates of the parameters that assume a stationary (i.e. stable) relationship between the variables across space. This assumption ignores the possibility of local variation (i.e. contextual effects) due to heterogeneity of the impact of the variables across space. Given that the premise of our second research question is that the spatially heterogeneous application of policy (by the county) and the heterogeneous response by agents may be leading to fragmented development, it seems inappropriate to make the assumption of stationarity up front as the goal of the paper is to test this very premise. The data and modeling techniques used in this paper are sufficient to address all of these shortcomings. In doing so, we demonstrate the important role

that spatially differentiated regulatory costs play in determining land use patterns and show that the role played by local land use interactions may have been overstated in previous research.

3 Model of Land Use Conversion

In our study region, the majority of the land conversion that takes place is from agriculture or forest land to residential. Moreover, the majority of this development is into some form of subdivision development. Thus, we model the decision of the developer to convert a raw parcel of land to a residential subdivision development. Specifically, we model the decision of the agent from the time in which she files her subdivision application. Because it necessarily takes time to complete any project and we are interested in how regulation impacts that time, we model the decision from the point in our data where the agent reached her optimal stopping point and filed her application.

In each period, t, a developer, $n \in \{1, \ldots, N\}$, makes a decision as to whether or not to apply for permission to develop. We assume that the market for residential housing is competitive and that each developer derives income only from the same of the developed lots. We call the optimal stopping or development time, t^*, and assume that it is chosen to solve the following maximization problem over the infinite horizon:

$$\pi_i = \max_{t^*} \left\{ \int_0^{t^*} A(X_i, \tau) e^{-r\tau} d\tau + (R(X_i, t^*) - C(X_i, t^*)) e^{-rt^*} \right\} \tag{3.1}$$

where A is the value of agriculture rent earned up to the optimal stopping time, R is one-time rent or sale value of the lots in their developed form, C is the cost of converting parcel to a developed state, r is the discount rate and X_i are characteristics specific to each parcel that influence the conversion decision.[5] In each period, a developer trades off the choice of developing in the current period with the choice of delaying development one period. The optimal development time is the point where the returns from the current period investment decision are equal to the second-period discounted returns plus the value of the agriculture rent earned between periods (Capozza and Helsley, 1990). In any fast-growing urban area, this condition will eventually be met and the parcel will development.

One issue with this specification, however, is that it assumes an immediate payoff at the time of development. However, as we stated previously, regulations can actually increase the implicit costs of development by extending the approval and completion times for development. To account for this uncertainty, we assume that the optimal stopping time is now uncertain and is composed of the baseline or instantaneous payoff

[5] We suppress the time subscript for the design matrix in this simply theoretical setup. However, it should clear that many of the factors affecting the optimal stopping decision vary with space as well as time.

case, t^*, and some additional time, α_a, where a stands for the approval and is the time it takes the county to approve a subdivision project. As a result, the maximization problem now becomes:

$$\pi_i = \max_{t^*+\alpha_a} \left\{ \int_0^{t^*+\alpha_a} A(X_i, \tau) e^{-r\tau} d\tau + (R(X_i, t^* + \alpha_a) - C(X_i, t^* + \alpha_a)) e^{-r(t^*+\alpha_a)} \right\} \tag{3.2}$$

From this equation we can see that the addition of the α_a alters the decision process by increasing the time until approval and reducing the profits by reducing the present discounted value of returns. As this value increases it alters the timing by making the previous project unprofitable.

To make explicit the optimal timing nature of the model, we appeal to the market conditions of our study region. This region has experienced significant population growth over the last several decades. This population pressure has translated into increased land values over time as per capita incomes have risen and the amount of developable land has declined. Consequently, the returns from development, π, in all states of the world have been rising as result of rising land rents, R. Given this continued rise in land value, we assume that the optimal *value* (or stopping time) for development on each parcel will eventually be reached. However, this optimal time will be altered or increased as the expectation and uncertainty about the approval time increases.

Although our theoretical model is not spatial, the results provide some idea of how spatial variation in regulatory uncertainty could be translated into a leapfrog pattern of development. If the planning authority increases the restrictions on the development of parcels that are closer to the urban center, while leaving exurban parcels unchanged, then a relative increase in expected future costs and reduction in expected future revenues on urban parcels would increase the value of vacant urban land, but not change the value of vacant exurban land. This should reduce the supply of urban housing and increase the incentive to develop at or beyond the urban fringe. While this result is not new - the fact that intertemporal and spatial differences in costs lead to scattered development (Ohls and Pines, 1975; Mills, 1981; Peiser, 1989) - this is the first paper to provide an explicit empirical test of how regulation-induced cost uncertainty contributes to development timing and the spatial pattern of development. Our results also separate out and provide an alternative interpretation to the amenity-drive growth explanation provided by the empirical land use literature presented in the previous section. We now turn to describing the econometric model used to implement this empirical test.

4 Empirical Model

Empirical Specification

Residential subdivision development is a multi-stage process. For example, in our study region the first stage consists of the submission of a preliminary plan; the second stage consists of gaining conditional approval; and the last stage is final approval. Given that the first stage is only a preliminary hearing with no official approval, we model the decision beginning at the second stage where landowners commit funds and would need to make predictions about the uncertainty in the price and cost environment during the period in which they are developing.[6]

We cast the developer's decision problem as one of choosing the optimal time, t^*, to file her application for development in order to maximize profits on her parcel. Given that we do not observe revenues and costs at the individual level, we cast the problem, in each period, as a latent decision by the developer over the value function for the parcel:

$$\max_{t^*} \mathrm{V}(\pi(t^* + \alpha_a, x_{it}) + \varepsilon(t^* + \alpha_a, \xi_i) \geq 0), \tag{4.1}$$

where x_{it} is a set of parcel characteristics in period t for parcel i, including our measure of cost uncertainty, that affects the value of the parcel and ξ_i is a vector of unobserved characteristics associated with parcel i and its developer. This function makes clear that the developer only chooses to start the development process when the value of the project in present value terms is greater than zero. To make the empirical specification of this value function explicit and specify the parameters to be estimated, we rewrite equation 4.1 as follows:

$$\mathrm{Prob}(\pi(x(t^* + \alpha_a, i)'\beta + \varepsilon(t^* + \alpha_a, \xi_i) \geq 0), \tag{4.2}$$

which is a per-period binary choice model that takes on a value of one at the optimal development time and zero otherwise. x_{it} is the same vector of covariates as before and β is set of parameters to be estimated. One possible estimation technique for this model would be to specify a distribution for the unobserved component and run a pooled binary choice model on the data. However, given our theoretical specification regarding the intertemporal decision-making process facing the developer and the fact that our data are time-series-cross-section on this intertemporal choice, there is clearly temporal dependence in the probability or optimal

[6] During the first stage, county planners determine if the landowner has permission to develop the parcel and whether it is located in a developable area. It is during the second stage that the official development plan is submitted and developers face uncertainty in gaining final approval.

stopping decision. Thus, must specify a model that can account for this dependence and model the optimal stopping decision in an empirically explicit way.

Discrete-Time Duration Model

To achieve both of these goals, we follow Beck, Katz, and Tucker (1998) and specify a binary time-series-cross-section model for discrete time or *grouped duration* (event history) data. Event history data accounts for the elapsed time until an "event" occurs or is no longer observed. Such a specification models the entire process for each observation and captures the cumulative impact of the process and variables on the decision-making process of the individual. Thus, an observation is at risk until it fails or an event occurs and the model captures the hazard rate or probability of failure in any particular time period.

The innovation of these authors was that for discrete-time or grouped duration data a continuous-time duration model could be estimated by specifying the binary stopping choice in each discrete time interval by a discrete-choice Logit model with the baseline hazard or "duration dependence" modeled by specifying time dummies for each period that an observation was still alive (or undeveloped in our case). The coefficients on these different time dummies provide for an explicit test of temporal dependence. The model also allows for the easy inclusion of time-varying covariates, which is particulary important in our case as our measures of local land use interactions and regulation vary with each time period and between observations.

The most common specification for duration data is the continuous-time proportional hazard model:

$$\mathrm{h}(t|x_{i,t}) = \mathrm{h}_0(t)e^{x_{i,t}\beta}, \tag{4.3}$$

where $x_{i,t}$ is the vector of independent variables at continuously-measured time steps. The hazard rate in the model depends on both the independent variables and the length of time that the observation has been at risk, $h_0(t)$. Depending on the type of model estimated, the baseline hazard can take any number of different types of time dependence. The continuous-time duration specification models the instantaneous probability of failure for each observation. In the case of grouped or discrete-time duration data, observations are only observed at discrete intervals. As result, more than one event is observed in each period of time and the model is simply the probability of a particular event occurring during a given time period. By letting $y_{i,t}$ be a binary indicator of an event occurring to observation i in period t, the discrete-time hazard is simply $P(y_{i,t} = 1)$. The discrete-time duration model corresponding to equation 4.3 is given by following the Logit model:

$$\mathrm{P}(y_{i,t} = 1) = \mathrm{h}(t|x_{i,t}) = \frac{1}{1 + e^{-r(x_{i,t}\beta + p_{t-t0})}}, \tag{4.4}$$

where $x_{i,t}$ now represents the value of the variable for the interval period t and p_{t-t0} is a dummy variable indicating the time period for which the observation is being observed. The first period is left out in each model to prevent multicollinearity. The maximum likelihood specification of this model is as follows:

$$\sum_{i=1}^{n} \left\{ \mathrm{y}_{i,t} \log(P_{i,t}) + (1 - \mathrm{y}_{i,t}) \log(1 - P_{i,t}) \right\}, \tag{4.5}$$

where $y_{i,t}$ is the binary decision in period t by developer i and the P term is the Logit link specification in equation 4.4.

Conditionally-Parametric Discrete-Time Duration Model

The final step in our empirical specification and one of the main innovations of this paper is to extend the discrete-time duration model, (4.4), nonparametrically by allowing for flexible estimation of the coefficients across space. Nonparametric analysis has gained popularity in recent years, especially in the areas of land use, transportation modeling and housing. The main advantage these models is that they do not make any assumption about a stationary relationship across space thus allowing for the capture of spatially-varying and heterogeneous effects at a more localized scale.[7] While a number of studies have applied these models to cross-sectional land use and house price data, to our knowledge this is the first application of these models to time-series-cross-sectional data.

In this paper, we applied a conditionally parametric local likelihood model. The conditionally parametric model is a special case of locally-weighted regression, where a parametric regression model is applied locally to each observation in the dataset and the coefficients are estimated at each point based on the specification chosen. The intuition is that the variables are considered fixed at each point where the model is run, but by varying the weights applied to the observations in the models it is possible to optimize a smoothing statistic such that the model fits the as closely as possible to the true curvature of the parameter space. In our case, we are allowing the weights and coefficients to vary spatially in order to model the spatial heterogeneity of the land development process. The conditionally parametric specification of our global maximum likelihood model, equation 4.5, is as follows:

[7] McMillen and Redfearn (2010) provide an excellent overview of the different types of models being used and McMillen and McDonald (2004) provides a Monte Carlo study and application of nonparametric maximum likelihood.

$$\sum_{i=1}^{n} \mathrm{K}\left(\frac{Z_{1i}-Z}{h}\right) \mathrm{K}\left(\frac{Z_{2i}-Z}{h}\right) \left\{\mathrm{y}_{i,t}\log(P_{i,t}) + (1-\mathrm{y}_{i,t})\log(1-P_{i,t})\right\}. \tag{4.6}$$

In this model, we make the assumption that the binary choice at any location and time period is a parametric Logit model, but the marginal effects and constant term vary across space. In our model, we do not apply weight in the time dimension, but instead account for time by time dummies in the discrete-time duration model as explained above. Thus, the kernel weights in our model, K, are functions of the geographic or latitude and longitude coordinates, Z_1 and Z_2, at each location. These functions determine the weight that each observation receives in running the local model. In our model, we use a Gaussian kernel, $(2\pi)^{-.5}e^{\frac{-z^2}{2}}$. The choice of the this kernel has little impact on the results, but the choice of the window size or bandwidth, h, is more impact. The procedure for choosing this value will be discussed in the results section. Our final model consists of maximizing the weighted likelihood equation (4.6) at each location, i, in our dataset and using all of the cross sections, T, for that point as well as those for all other observations, j, included in the optimal window size.

5 Description of Data and Covariates

Study Region

The data used to estimate our model are micro-level land use data constructed from historical land use records from Carroll County, Maryland. Carroll is an urbanizing county located approximately 30 miles west of Baltimore, Maryland and 55 miles northwest of Washington, D.C. Figure 5.1 shows Carroll's location relative to the major metro areas of Baltimore and Washington, D.C. as well as the Chesapeake Bay, Potomac River, and other natural amenities. For much of its history Carroll was a predominantly rural county with most of its population and land dedicated to farming. But, in the last several decades, the county has become increasingly suburbanized with substantial population growth in many parts of the county. Figure 5.2 shows the cumulative population growth and population change over the last century. The county's population went from just over 34,000 in 1900 to fewer than 45,000 in 1950. However, from 1960 through 2010 the county's population grew by 215%. This growth has resulted in the county shifting away from an agriculture-based landscape to one with a large portion of the landscape in development. A significant portion of this new development is comprised of single-family housing. As of 2007, residential development made up 22.5% of the land area of the county (Carroll County GIS Department).

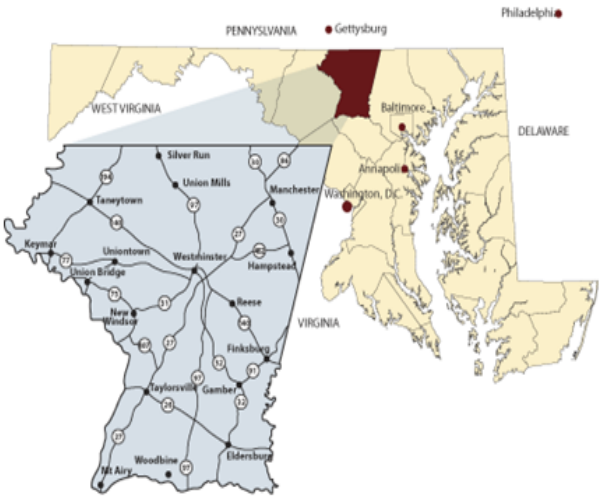

Figure 5.1: Baltimore/Washington, D.C. Metro Area

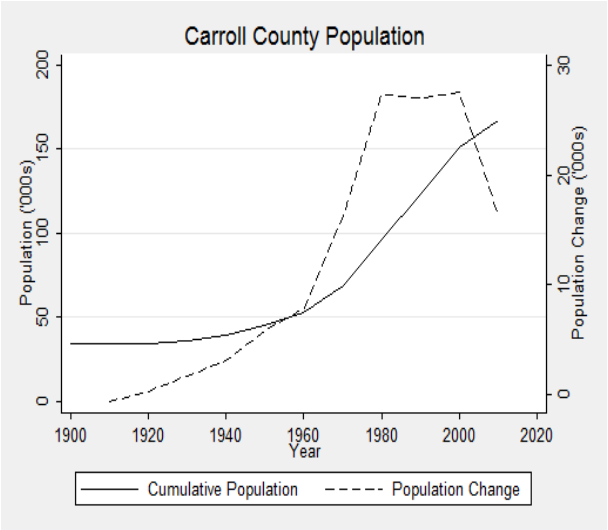

Figure 5.2: Carroll Population

Land Use Regulation History

In 1963, Carroll passed its first comprehensive zoning plan. The initial plan restricted development density to one house per acre in all areas of the county without public sewer facilities. However, in 1978, as a result of population pressures and the commensurate loss of productive agriculture land, the county passed its second major comprehensive plan. The intent of the plan was to significantly restrict development densities in much of the rural areas of the county and protect valuable agriculture land from further fragmentation. The plan included a massive down zoning of 70% of the land in the county. This new zoning class had a stated density

of one house per 20 acres, but given some weakness in the law its actual affective density was closer to one house per 15 acres.[8] Apart from several small adjustments made in 1989, these same restrictions have been in place in the county since 1978. Figure 5.3 shows the current zoning areas of the county for both the towns as well as the unincorporated areas. In addition to agriculture zoning, conservation zoning was set at one house per three acres; the densities in the rest of the zoning areas were based on derivatives of one acre.

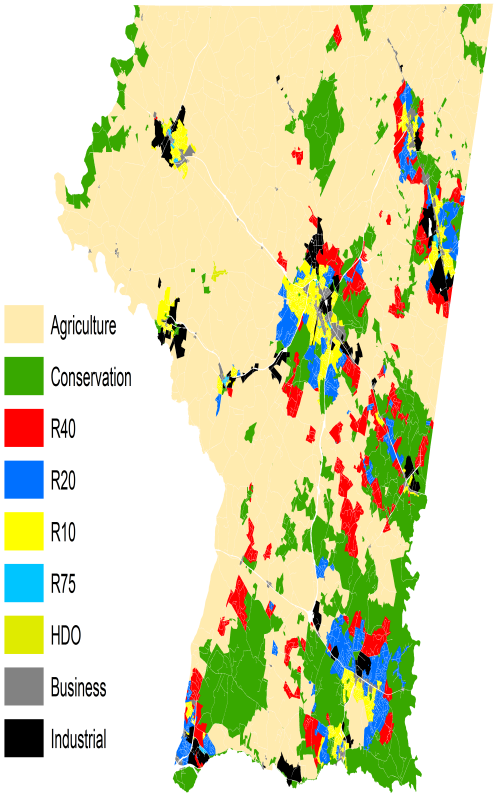

Figure 5.3: Carroll Zoning

In addition to placing density restrictions on residential development, the 1963 comprehensive plan provided a formal procedure for the creation of residential subdivisions. While there were subdivisions in the county prior to this point, a formal process of gaining approval was not in place. Following the 1963 plan, land owners could choose between either a major or minor subdivision. Major subdivisions consisted of any development with four or more buildable lots at the time of development and required the installation of formal infrastructure. Minor developments consisted of developments of two or three lots without any formal

[8] Large scale down zoning of this nature was common in the counties in this region during this period. Many of the counties had passed earlier zoning restrictions, but, because of weak application of the previous laws or because of insufficient restrictions to begin with, many counties recognized the need for more restrictive regulations on density and development in rural areas.

infrastructure. In addition to the differences in lot counts, the regulation also created significant regulatory differences between the two types of subdivisions. Minor developments did not require a formal hearing to gain approval and could be approved by the chairmen of the planning board. Major developments, however, do required a formal hearing, which, in many cases, significantly increased their approval times. Both of subdivision options has been in effect since that time.[9]

While the combination of the agricultural zoning policy and the creation of the formal subdivision policy was supposed to control exurban and rural development and reduce the fragmentation the rural landscape, several weaknesses in the agricultural zoning law (in combination with the minor subdivision policy) have made the outcome less than desirable. The enactment of agriculture zoning in 1978 was intended to down zone, to one house per 20 acres, over 70% of the the rural and exurban land in the county, to reduce development in these areas and to push development into designated development areas in the county. However, effective zoning is currently closer to one house per 15 acres, and, as our data and results indicate, it is not slowing development in these areas. This lower-than-expected result likely stems from a loophole in the agriculture zoning law. At the present, each parcel located in an agriculture district and having at least six acres of land is allowed to create two buildable lots or a two-lot minor development - each additional lot requires 20 acres. Thus, the zoning code creates a non-linearity in the effective zoning in agricultural districts and incentivizes the creation of minor developments. In our dataset, over 60% of all subdivisions created from 1995 through 2007 were platted in agricultural areas, and out of these 82% were minor developments. Figure 5.4 shows the spatial distribution of major and minor developments in the county in 2007 and the significant number of small developments in the county. Thus, the combination of weak zoning and the minor development policy appear to be interacting to create a environment more favorable for smaller developments in exurban and rural areas. This result will be made more explicit by the results from our empirical model.

Data Construction and Description

In addition to our contribution to the literature on suburban and exurban land conversion, the other primary contribution of this research is the construction of several micro panel datasets.[10] The first dataset we constructed was a panel of the historical subdivision development for the county. To construct these data,

[9] According to county planning officials, in most cases minor developments can gain approval in less than two or three months; major developments, however, require an open public hearing as well as the approval of up to 12 different county agencies, which can significantly increase the time until approval.

[10] This research is being conducted as part of the Baltimore Ecosystem Study (BES), a Long-Term Environmental Research (LTER) project funded by the National Science Foundation and charged with modeling the historical evolution of human-environmental interaction. The goal of the project is to build historical economic models of human decision making to be joined with existing environmental models to investigate the effect of human choice on the environment of the Baltimore Metro region.

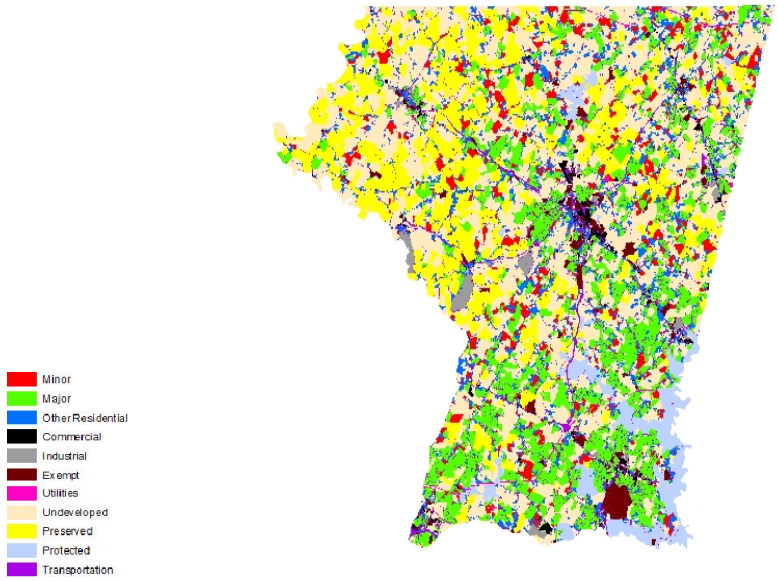

Figure 5.4: Carroll Land Use 2007

we joined the parcel boundary GIS shapefile of the county with the tax assessor's database using a tax assessment ID number. In addition to information on the attributes of the parcel, structure, purchase date and price, and information about the owner, the assessor's database contained information on the plat book and page number for the subdivision in which the parcel was located.[11] Using these numbers we were able to locate the original plats at the Maryland historical archives. By matching the individual parcels in the parcel boundary shapefile with the plat maps we were able to determine all of the parcels in each development, assign each development a unique ID number, and provide a date when the subdivision first gained approval.[12] By dating the subdivisions and dissolving the GIS layers on the years of interest, we were able to recreate the landscape in the county for any given time period.

There were 1,910 subdivisions developed from 1924-2007. Of these, 1,098 were major developments and 812 were minor developments. Figure 5.5 shows the cumulative subdivision activity in the county from 1924 forward. As was stated above, formal subdivision approvals for minor developments did not begin until after the 1963 comprehensive plan was passed. Since that time the number of minor developments has grown at

[11] After a subdivision gains final approval from the county zoning commission, the plat of that development becomes public record, and is recorded and stored at the Maryland historical archives. These plats and the information contained on them are available to the public online at the following address: `www.plats.net`.

[12] In 20% of the developments the subdivision was completed in more than one phase. In the case of these multi-phase developments we dated and assigned unique ID numbers to each section. We also gathered information about open space requirements, sewer, zoning, developer information, and whether the development was a major or minor subdivision.

the same rate as majors and in recent years has actually grown faster than major developments.[13]

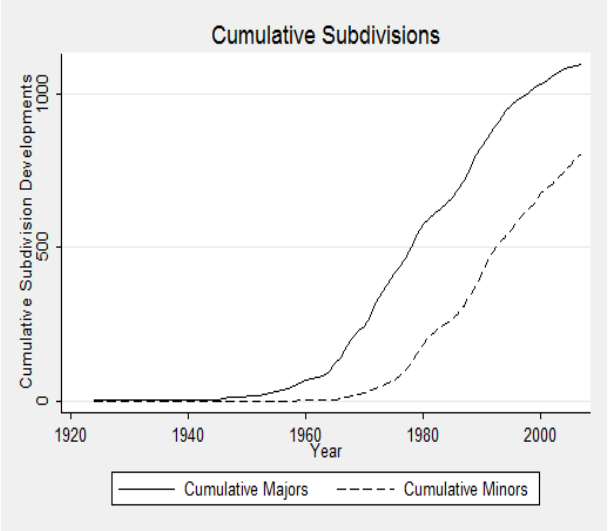

Figure 5.5: Cumulative Subdivisions

To provide a better perspective on the extent of residential development, since 1990 Carroll has had approximately 17,600 parcels of land developed. Out of these, 95% where in some form of residential development, and 85% were located inside of one of the two types of development. Figures 5.6 and 5.7 show, respectively, the marginal and cumulative growth of buildable lots in the county from the first quarter of 1980 through the fourth quarter of 2007. The graph of marginal lot counts shows a large amount of variability from one quarter to the next, and the graph on cumulative lots shows a steep upward growth trend until the late-1990's at which time it begins level off. Much of this leveling off may be due to the peak of the housing boom in mid-2006, but it is also likely a result of the counties crack down on larger subdivision developments beginning in 2002.[14]

The second dataset we created was for the historical evolution of land preservation and protected open space in the county. Over the past several decades many state and local governments throughout the U.S. have developed and used voluntary incentive-based programs as a mechanism to prevent sprawl, limit growth, and protect agriculture land. Within these programs landowners receive actual payment or equivalent tax deductions in exchange for voluntarily foregoing development on their property in perpetuity. In addition to the down zoning that took place in 1978, in 1980 Carroll began its own purchase of development rights

[13]In many ways the minor subdivision policy was internalizing a process that was already underway. Many small developments and single family homes were being built in the period preceding 1963. Part of the impetus for the plan was to help document and control the amount this type of development and protect vital farm land from develop-lead fragmentation. Thus, as is the case with many land use policies, the subdivision policy for Carroll formalized an existing trend.

[14]In 2002, two-thirds of the county commissioners were replaced in a election centered on increased land use regulation, the reduction in large subdivision, and the reduction in rural development.

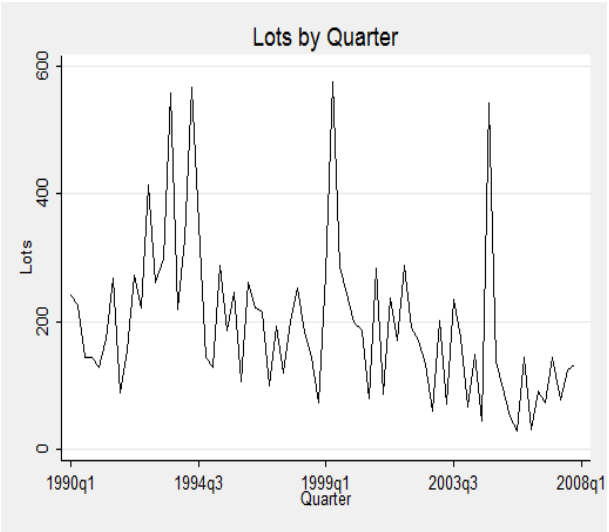

Figure 5.6: Lots by Quarter

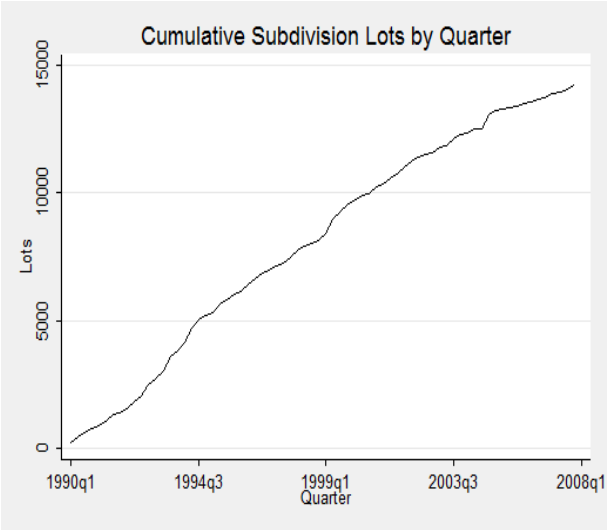

Figure 5.7: Cumulative Subdivision Lots by Quarter

(PDR) program as an additional measure to protect farmland. Using state and county funding sources, the county has preserved over 54,000 acres of land in four different programs since 1980. We created the data for the history of these programs by matching data received from the county officials with the parcel boundary file using names and tax ID numbers. While we do explicitly model this decision, these data give us the ability to control for this decision in our analysis by removing preserved parcels from the dataset in each time period as well as model the impact of local interactions between preserved land and the development decision.

The final dataset we created was for the historical subdivision approval process. As was noted above, when landowners wish to subdivide a parcel they must follow the rules in the county subdivision development guide. One of the most uncertain aspects of the development process is the necessary time to gain final approval and the regulatory hurdles that delay the process. To reconstruct the history of this process for each of our subdivisions, we collected the official minutes from the planning commission's monthly meetings. Using these data, we matched subdivision names with the information from the commission's database to provide dates for the stages of the development process for each of the developments. Given that the county only had electronic data starting in 1989, we only have data on the process from 1989 through 2010.

Before we describe the process of converting these data into into our proxy for regulatory uncertainty, it seems appropriate to get a sense for how the development times differ across different types of development and how they vary by lot quantity for the subdivisions in the county. Clearly, the most appropriate divide in development types is between major and minor subdivisions. In table 5.1, we show the summary statistics for the subdivisions developed in the county from 1995 through 2007. In the top part we compare the development times for major versus minor subdivisions. It is clear that there exists a significant difference in the development times between these two types of development, and that this difference is significant across all quantiles of the distribution. These results reveal that, on average, major developments take as much as 10 months longer to get approved. In the middle part of the table we divide major developments into two groups based on median lot quantity. While the values drop some, there is still a significant difference between smaller major developments and minors. Finally, in the bottom section, we show the unconditional correlation between the logged value of development times (in months) and the total lot quantity for each of the of subdivisions. This value shows a positive relationship between approval times and the size of the development. It should be noted that these results are just for approval times and do not include the time it takes to complete the project which could be even longer for larger developments.

Table 5.1: Major and Minor Development Times

Subdivision Type	**Mean**	**25th**	**Med.**	**75th**	**N**
A: Two Types					
Major Developments	15.34	7	12	20	118
Minor Developments	5.95	2	4	6.5	244
B: Three Types					
Major Developments (Over 9 lots)	17.05	10	14	22	58
Major Developments (4 to 9 lots)	13.70	3.5	9	15.5	60
Minor Developments	5.95	2	4	6.5	244
Correlation Coefficient	0.2566				

Our datasets as well as our economic analysis are at the parcel level. Given limited availability of historical price data, we estimate the model on all subdivisions that were granted final approval from 1995 through 2007. The data consists of all undeveloped parcels as of 1995 that were eligible to be subdivided into at least two buildable lots according to the zoning regulations for the parcel. We use all parcels located in one of five zoning districts in the county: agriculture, conservation, R40, R20, or R10. These parcels also include 604 parcels that were preserved during the observation period. We consider these parcels as undeveloped until the year they are preserved at which time they drop out of the dataset. The final dataset consists of 3,852 parcels. During our study period a total 400 or a little over 10% of the parcels in the county gained final approval. Our final time-series-cross-section dataset, when expanded by parcel-years, contained 46,138 observations.

Covariates Used in the Empirical Model

Table 5.2 defines all of the variables used in the estimation of our model and gives their summary statistics. The first set of variables proxy for the accessibility effects hypothesized by the classic monocentric model to impact the decision to convert land. If those predictions hold, then we would expect distance, (Dist_to_balt), to decrease the likelihood of development and accessibility, (Slu_roads), to increase it.[15] In addition to the predictions of the monocentric model, evidence of local land use interactions have also been shown to affect the speed and placement of development (Irwin and Bockstael, 2002, 2004). For example, if there are positive agglomerative effects of surrounding development, then we would expect surrounding residential development to speed up the pace of development. However, if surrounding residential has a congestion effect, then it may have a negative impact on other development. To account for the effect of local interactions among land uses, we create a 200 meter buffer around each parcel in each period using ArcGIS software and calculate the percentage of land uses of each type in that buffer for each period. The land use types we include in our model are surrounding subdivision development, (Slu_subdiv), other residential, (Slu_resident), undeveloped land with an existing structure, (Slu_undev_exhouse), private open space, (Slu_open_preserve), protected open space (parks and reservoirs), (Slu_open_public), commercial and public facilities, (Slu_commerce), and utilities and industrial, (Slu_indust). All of these are in terms of the omitted category, undeveloped land.

The next set of variables are parcel characteristic variables that proxy for the physical costs of development on the parcel. Certain characteristics of a parcel - size in area, soil type, slope, zoning restrictions, existing structures, and the percentage of forest cover - may make development more or less likely. For example, a

[15] In a long-run model it is likely that roads or some improvement to them may be endogenous. However, given the short time period of our model we take them to be exogenous.

larger parcel is ideal for a larger development, but it may also be more suitable for farming as there may be increasing returns to scale from farming larger parcels. Also, better soils may make it more likely to get approved for septic permits in areas without sewer, but this same soil is better for farming. Thus, many of these effects are ambiguous, and their sign is an empirical question. The variables we include as proxies for costs are the area of the parcel in acres, (Area), the quantity of lots allowed on the parcel, (Zoned_lots), an indicator variable for whether the parcel is located in an agricultural zoning area, (Ag_zoning), the percentage of type 1 and type 2 soils, (Soil1) and (Soil2) (these variables are in terms of the omitted soil category, (Soil3), which is the worst type of soil), the percentage of the parcel with slope over 15%, (Steep_slope), and the percentage of forest cover on the parcel, Forest.

In addition to the factors affecting the physical costs of development, we also include the variables that proxy for local competition, local price level, and the opportunity cost of investment capital. The first variable, (Local_competition), is a time-varying measure of the number of lots approved by the zoning commission in county. For each year that a parcel is eligible for development, we calculate the number lots approved over the previous two years in an area accounting for the closest 10% of all parcels. This variable is a proxy for the amount of buildable stock in a geographic region around each parcel, a measure of local competition and designed to separate out the impact of local supply-side competition effects from the local land use spillovers captured by our surrounding land use variables. Our prediction for this variable is that increased local competition will create a local congestion effect in the market and as the number of locally-approved lots increases the probability of development will decrease. The last three variables in this section are our real options variables for cost and price. Our measure of price drift, (Price_drift), and volatility, (Price_volatile), are constructed using a locally-weighted nonparametric regression technique. This process is explained in the appendix A. The construction of our proxy for implicit regulatory costs, (Regulation), is explained in the next section.

Proxy for Regulatory Costs

To capture the effect of regulation-induced implicit costs on the decision of landowners to subdivide, we use the data on completion times for previously developed subdivisions to predict the expected time to until approval for each eligible parcel, in each period of time. Starting in 1995, we use approval time information for all previous developments and estimate a parametric *multi-event* duration model in each period. We then use the estimates from each model, in each period to predict expected approval times for all undeveloped parcels in the same period.

Table 5.2: Covariates Used in Sample Selection Model

Variable	Descriptions	Average	Std. Deviation	Minimum	Maximum
Monocentric Model					
Dist_to_balt	Baltimore (Minutes)	41.07	8.07	23.18	65.51
Slu_roads	Primary Roads (%)	5.12	5.41	0.00	63.41
Local Interactions					
Slu_subdiv	Subdivision (%)	15.79	16.96	0.00	87.71
Slu_resident	Other Residential (%)	12.03	10.90	0.00	73.50
Slu_undev_house	Undev. with House (%)	22.67	20.22	0.00	89.17
Slu_open_preserve	Private Open Space (%)	5.39	11.92	0.00	98.05
Slu_open_public	Public Open Space (%)	0.75	4.58	0.00	81.34
Slu_commerce	Commercial (%)	3.51	8.30	0.00	69.89
Slu_indust	Industrial (%)	0.84	3.65	0.00	61.60
Costs					
Area	Area (Acres)	32.63	41.02	0.46	365.56
Zoned_lots	Zoned Lots	8.57	25.47	0.00	653.00
Existing_house	Has House	0.50	0.50	0.00	1.00
Sewer	Public Services	0.15	0.36	0.00	1.00
Ag_zoning	Min. Lot Zoning	0.51	0.50	0.00	1.00
Soil1	Type 1 Soils (%)	40.04	43.21	0.00	100.00
Soil2	Type 2 Soils (%)	52.67	43.19	0.00	100.00
Steep_slope	Greater than 15%	17.27	29.21	0.00	100.00
Forest_cover	Forest Cover (%)	33.32	32.69	0.00	100.00
Opportunity Costs					
Local_competition	Competition	13.32	21.01	1.00	135.00
Price_drift	Price Drift	0.27	1.03	0.00	5.11
Price_volatile	Price Volatility	0.55	0.07	0.38	0.84
Regulation	Regulatory Costs	13.61	3.04	5.29	33.33

To produce our estimates of regulatory uncertainty we estimate 13 separate duration models. Each of the models is estimated using the undeveloped parcels and subdivision events that occurred in all periods preceding the one of interest. For example, to produce the predicted approval time for undeveloped parcels in 1995, we use the estimates of the second-stage duration model, which are for all previous subdivisions that had completed both stages of the development process before 1995. Thus, in each period, only those developments that had finished the second stage and gained final approval were used in the predicting the development time for undeveloped parcels. The implication is that each landowner uses previous approval timing information and her parcel's characteristics to produce an estimate of the likely time until approval for her own parcel if she chose to subdivide.

Given the amount of information generated by these models, it is not feasible nor necessary to show all

of the results. Instead, figure 5.8 shows the kernel density curves for the predicted results from the models for the years 1994 and 2002.[16] The figures show two separate curves - one curve for parcels that remain undeveloped throughout the model and a second for those parcels that ended up developing in future periods. As is clear from the figures, there are definite differences between the two, but they still retain the same shape.

The predicted values from each of the duration models is matched with the parcels in each period. The intuition is that in each period of a sample that a parcel remains undeveloped the landowner uses past approval times to form her own expected approval time on her parcel for that period. These data serve as our measure of expected future regulatory costs for subdivision development on each parcel and the value of this parameter identifies the effect of regulatory uncertainty on conversion timing nonparametric model. A more complete explanation of this process is given in Appendix B.

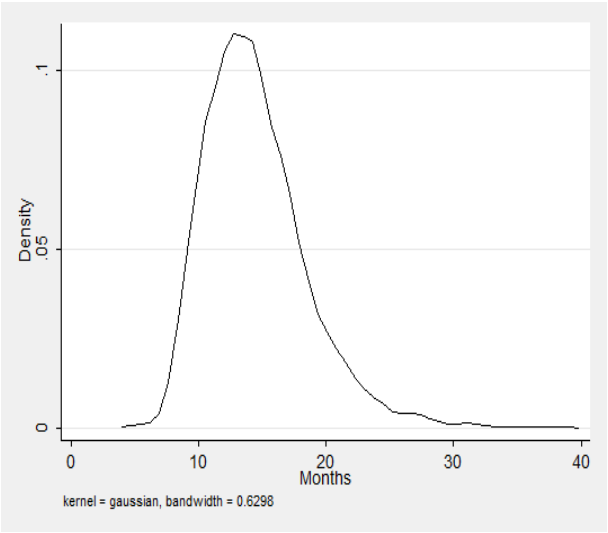

Figure 5.8: Predicted Development Times in 1994

[16]We chose 1994 as it is the set of predictions used in the first year of the sample selection model; we chose 2002 as it was the year that the county increased regulation on larger developments.

6 Results and Discussion

Model Results

The results from our models are shown in table 6.1. The first two columns show the results from running the global discrete-time time-series-cross-sectional model, and the last four columns show the results from our conditionally parametric version of the duration model. The global model was run using a standard GLM Logit model, and all of the nonparametric procedures were running using original code written R version 2.12.1. [17] We report the coefficients and standard errors for the global model, but report the mean, standard deviation, and the minimum and maximum of the coefficients for the conditionally parametric model, which is standard in nonparametric modeling. As is apparent from this table, most of the results for the nonparametric procedure are similar to those from the global model, but the log-likelihood values show that the nonparametric model provides for slightly better fit. Before turning to a discussion of these results, we discuss the selection procedure for the optimal bandwidth.

The choice of bandwidth, h, is of critical important for nonparametric modeling. The bandwidth determines how many observations receive weight and how fast the weights decline with distance. By placing more weight on more distant observations, high values of h imply local regressions that produce more smoothing. The bandwidth can be one of two kinds. First, it can be a "fixed window" bandwidth, where the window size determining which observations receive weight is fixed and the number of observations used in each estimation of the model is variable. The second type of bandwidth is "adaptive window", which allows for a flexible window size, but the same number of observations to used in each model. While there is little difference in the two in when the data are uniformly distributed. However, as we will see in the data section, our data are far from uniform and can we quite sparse in some areas. So, we use a the adaptive window specification to guarantee sufficient observations for each run of the model.

There are a number of methods for determining the optimal bandwidth, h, for local-likelihood models including the Akaike Information Criterion (AIC) and the Cross-Validation method (CV) or leave-out method.(Loader, 1999) In this paper, we use the latter. The cross-validation method is carried out by running model at each point, i, but leaving out the observation for that space, x_{-i}, in process similar to out-of-sample predictions. This process is repeated numerous times for different specifications for the window size and the estimated values at each point, β_{-i}, are used to produce the local *deviance residuals* for the binary choice model.

[17] I would like to thank Dan McMillen for this helpful R package McSpatial. It was from his cross-sectional conditionally parametric procedure that I created the code and procedures used in this paper.

The test statistic for the cross-validation method is then calculated using these deviance residuals as follows:

$$\frac{-2}{J}\sum_{j=1}^{J} \mathrm{D}(y_j, \beta_{-j}(x_j)), \tag{6.1}$$

where the summation in this statistic is over J, which is the total number of observation in all each periods, NxT. After this statistics for each window size, we determine the optimal window size by choosing the minimum value of the statistic. In our model, the optimal window size turns out to be 65%. The results from this procedure are shown in figure 6.1. The distance calculations between points are based on the Mahalanobois technique.

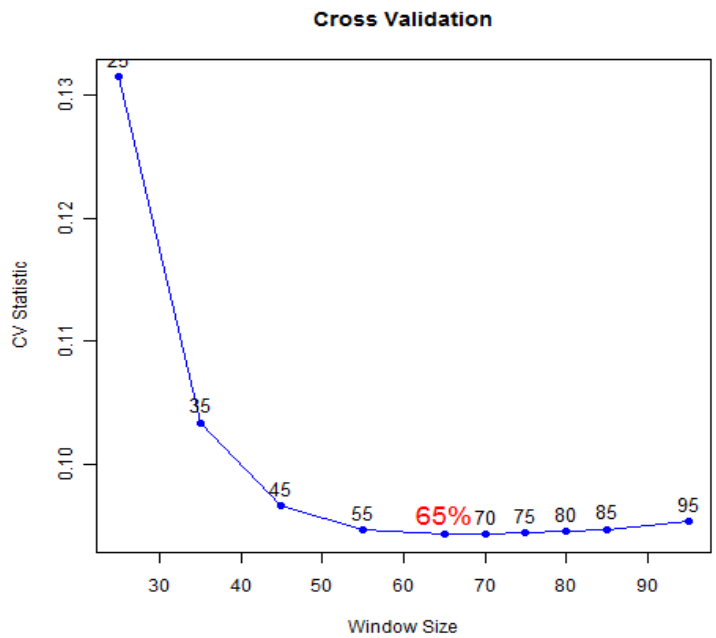

Figure 6.1: Cross Validation Statistic

Our monocentric variables show that an increase in travel time from Baltimore city reduces both the probability of development, which conforms to the predictions of the urban monocentric model. However, accessibility, measured as the amount of surrounding primary road access, has no impact. From our local interactions variables, we find that amenities have no impact on development in either model. The local interaction variables show that neither public open space nor preservation has any impact on the probability of the development, which is contradictory to what we would expect if permanently preserved land is viewed

as an amenity. The variables for surrounding development indicate that local agglomeration and congestion effects matter. Both the surrounding residential and surrounding subdivision development variables increase the likelihood of development in both models. In addition, our variable for developable land with an existing house, Slu_undev_house, indicates a reduction in both the probability of development in both models.

Table 6.1: Model Results

	Global Discrete Survival		Conditionally Parametric 65% Window			
	Coeff.	Std. Err.	Coeff.	Std. Dev.	Min.	Max.
Monocentric Model						
Dist_to_balt	-0.016*	0.009	-0.012**	0.006	-0.026	0.003
Slu_roads	0.011	0.014	0.013	0.012	-0.015	0.041
Local Interactions						
Slu_subdiv	0.029**	0.003	0.028**	0.007	0.015	0.039
Slu_resident	0.026**	0.005	0.025**	0.008	0.010	0.044
Slu_undev_house	-0.046**	0.004	-0.044**	0.005	-0.056	-0.032
Slu_open_preserve	0.006	0.005	0.007	0.007	-0.009	0.017
Slu_open_public	0.002	0.014	-0.029	0.039	-0.159	0.017
Slu_commerce	-0.018**	0.010	-0.016	0.018	-0.056	0.015
Slu_indust	0.001	0.016	-0.007	0.014	-0.045	0.021
Costs						
Area	0.017**	0.004	0.019**	0.003	0.013	0.028
Area_sqrd	0.000**	0.000	0.000**	0.000	0.000	0.000
Zoned_lots	0.004**	0.001	0.003**	0.001	0.000	0.006
Existing_house	0.502**	0.117	0.505**	0.262	0.065	1.098
Sewer	0.420	0.284	0.581	0.311	-0.292	1.193
Ag_zoning	0.774**	0.156	0.741**	0.346	0.174	1.313
Soil1	-0.001	0.004	-0.001	0.006	-0.015	0.012
Soil2	-0.002	0.004	-0.001	0.006	-0.016	0.010
Steep_slope	-0.003	0.003	-0.003**	0.001	-0.006	0.001
Forest	0.007*	0.004	0.006	0.005	-0.003	0.016
Opportunity Costs						
Local_competition	-0.004**	0.002	-0.005**	0.001	-0.011	-0.003
Price_drift	-0.169	0.130	0.012	0.092	-0.226	0.256
Price_volatile	2.333**	0.947	0.127	0.085	-0.083	0.334
Regulation	-0.216**	0.056	-0.222**	0.051	-0.359	-0.131
Constant	-4.746**	0.976	-4.411**	0.952	-6.700	-2.091
Log Likelihood	-2085.6591		-2072.7224			

* $p < 0.10$, ** $p < 0.05$

N=46143

Our cost variables indicate that as the size of the parcel and the number of buildable lots allowed increase

the probability of development. This result also implies that larger parcels, which are more profitable for agriculture as a result of economies of scale, are also more profitable in their residential form and out compete agriculture, which is as we would expect. The variable for sewer and public facilities has no impact on the probability of development. Our variable for agricultural zoning indicates that parcels located within these areas are more likely to develop. This increase in the probability of development is contradictory to the intension of the agricultural zoning regulation (which is to reduce the likelihood of development in agricultural areas). This result likely stems from regulatory differences between zoning areas and subdivision types and will be discussed further below. We also find that forest cover increase the probability of development in the global model, but has no impact in the nonparametric model with opposite effect occurring for the slope variable.

Our opportunity cost variables indicate that local competition decreases the probability of development, which is an indication that local market competition effects matter in the decision to development and the quantity of lots chosen. This result also shows that supply-side congestion effects, while not as large as the local land use interactions, still play a significant role in determining development. While the variable for price drift has not impact in either model, our measure of price volatility is significant in the global model and actually positive, which opposite of what we would expect for an increase in price uncertainty. While this result for price volatility is contradictory to the findings of several other authors (Cunningham, 2006, 2007), it likely stems from both the type of data we used - the other authors used data on individual house and lot transactions, while we are using data on the subdivision decisions - and the inclusion in our model of other factors of uncertainty such as regulation. While uncertainty over returns matters for short durations, Bar-Ilan and Strange (1996) showed that for longer term projects, with a lag between the start and completion, the downward effect of uncertainty on development disappears and can actually be reversed. Towe, Nickerson, and Bockstael (2008) also found opposite and insignificant signs for volatility using a similar modeling technique and dataset. Finally, we observe that our regulation variable has the expected sign and is significant in both models. We also notice, in both the selection and outcome equations, that the impact of this increase in implicit costs has a larger impact than any of our surrounding land use variables, which is an indication that regulation, more than spatial spillovers, may be impacting the development timing and quantity decision.

Discussion of Results

Previous research has suggested the open space amenities may be one of the driving factors behind the scattered and fragmented growth often observed in the exurban areas. However, in our model, after account for numerous other factors including zoning and regulation we find not indication that these factors have an effect on the development timing decision. A plausible explanation for this result stems from the spatial relationship between the preservation program and the land rent gradient in the county. The preservation program is run as an auction. Landowners submit bids to the program each quarter indicating the minimum amount (per acre) they would be willing to accept to preserve their parcel. Following these bids, the preservation program repeatedly selects the lowest bids until the funding for that quarter is exhausted. In Carroll, the most expensive land is located in the South and consequently, most of the preservation takes place in the North where the bids are the lowest. The preservation program also allows newly-preserved parcels to set aside up to two buildable lots for future residential use. While these are not officially platted at the time of preservation, when and if the landowner chooses to sell them she must create a minor development. As result of this process, it is easy to see why preservation has no impact on the probability of development (the bidding process has effectively separated preservation parcels from the land most likely to develop) and why the number of lots decreases with an increase in preservation (the surrounding development is potentially linked to the preservation process).[18]

Our local interaction variables for surrounding development do indicate, however, that both local agglomeration effects matter The results for the surrounding residential variables indicate that an increase surrounding residential development increases the probability of development. The fact that surrounding residential development increases the probability of development is consistent with local agglomeration effects. For example, a local school, shopping center, or park (not sufficiently captured by our current land use measurements) may attract development and thus increase the probability of development of all parcels. On the other hand, the negative effect of the variable for undeveloped, but developable land. These results indicate that neighboring properties with existing housing and remaining building rights create a congestion effect as the potential for further development increases potential future density and congestion. The overall conclusion is that local land use interactions matter, but that it is a trade-off between local agglomeration and congestion effects, rather than the positive spillovers of open space amenities, that have influenced land development outcomes.

[18]Visual inspection of the landowner names on the plat maps and those in our preservation data revealed a number on cases where the same person was active in both processes, which is a significant results in itself and one worth further research in the future.

The results for the indicator variables for agricultural zoning reveal that parcels located in an agricultural zoning areas have a higher probability of developing. This result is counterintuitive given that county wishes to reduce development in this area. This result likely follows from the confluence of the minor subdivision process, which reduces the approval times for minor subdivision developments versus majors, and the weak agricultural zoning law, which incentivizes minor development. The minor subdivision policy reduces subdivision approval times and uncertainty and the weak zoning law incentivizes small-scale development, which speeds up development times and create an abnormally large number of minor subdivisions in agricultural areas. Given that most of the agriculturally-zoned land is also located in more exurban and rural areas, these results also provide a regulatory explanation for the discontinues development patterns observed in these areas. This result is also consistent with previous work in this region that found that low density zoning may exacerbate low-density residential sprawl (McConnell, Walls, and Kopits, 2006).

We should also note the potential relationship that exists between this zoning variable and our measure of regulatory uncertainty - especially as it relates the spatial pattern of the development in the county. While the agricultural zoning variables are much larger than our regulatory variables, it should be clear from our previous discussion that the two effects are not unrelated. Since the county practically subsidizes minor developments in these agricultural areas through its lenient approval process for minor developments and its allowance of two-lot developments on parcels of only six acres, it is likely that this indicator variable in picking some of the regulation effect we hypothesized to be impacting development timing and type. For example, our theory suggested that if a parcel can expect a more expeditious approval process for one type of development over another or for a more rural parcel over a more urban one, then it will have an impact on the type and location of development. Since we know that most of the land zoned agriculture is either exurban or rural and that the incentive for minor developments (over major development) exists in these same areas, then the coefficients on the agricultural zoning variables are producing the exact results that our theory model predicts - increased development in agricultural versus non-agricultural and a reduction lots as landowners substitute away from major developments to minors. Thus, both the zoning variable and our regulation variable are providing evidence of the same set of effects - that current land use regulations are leading to the increase in low-density exurban and rural development in the county. Our analysis in the next section will make explicit exactly how this impacts the spatial pattern of development.

Spatial Pattern of Development

To answer our second research question, we need to get some idea of how the coefficients for our regulation variable and surrounding land use variable vary across space. To accomplish this, we plot these coefficients, using their latitude and longitude coordinates, along the major roads in the county and several of the largest towns. Its is important to remember that most of the most developed areas of the county are in the South and in and around the county seat, Westminster, which is located in the center of the county. Figures 6.2-6.5 plot these coefficient values for regulation, surrounding subdivision, surrounding residential and surrounding developable land, respectively. There are several important findings from these figures.

First, from figure 6.2 we can that, while an increase in regulation has a negative impact throughout the county, we see evidence that the downward pressure is more pronounced around the Westminster and in the South part of the county providing some evidence that differences in regulatory stringency may be impacting the timing and location of development. Second, we find that there is an agglomeration effect of surrounding residential throughout the county, but that the impact is reduced in greater in the further one goes from the urbanized area in the South. Finally, figure 6.5 shows that congestion effect matter throughout the county, but the most significant impact of this effect is in the South. This is likely a result of the fact that this areas is more developed and open space is scarce and commands a premium. This is also an explanation for why agglomeration has less of an impact in this region. The overall conclusion from these figures is that both local land use interactions and regulation may be contributing to the fragmented development patterns in the county, but that it is the regulatory uncertain (in terms of subdivision approval times) that appear to be having the largest impact on the spatial development patterns in the county.

7 Conclusions

The process of converting raw land to a residential subdivision development is time-consuming and subject to many uncertainties. It is influenced by many factors including the physical characteristics of the parcel, local land use interactions, local market conditions and the local regulatory policies impacting the type and location of development. It is this latter regulatory factor that often creates the most uncertainty for developers as inconsistent or opaque regulations extend the necessary approval time of development and thereby increase the implicit costs for developers. Based on theory, as uncertainty over the expected approval time for a development increases, it increases the option value of the real investment asset. This will in turn increase the value of vacant land (relative to developed land) and increase the optimal development time

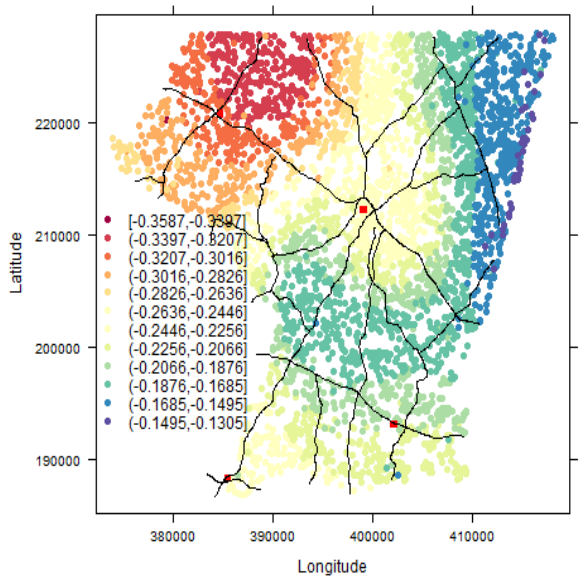

Figure 6.2: Regulatory Costs

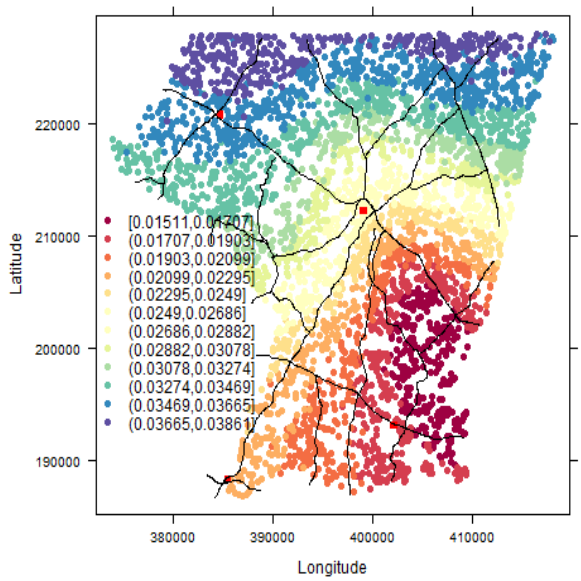

Figure 6.3: Surrounding Subdivision

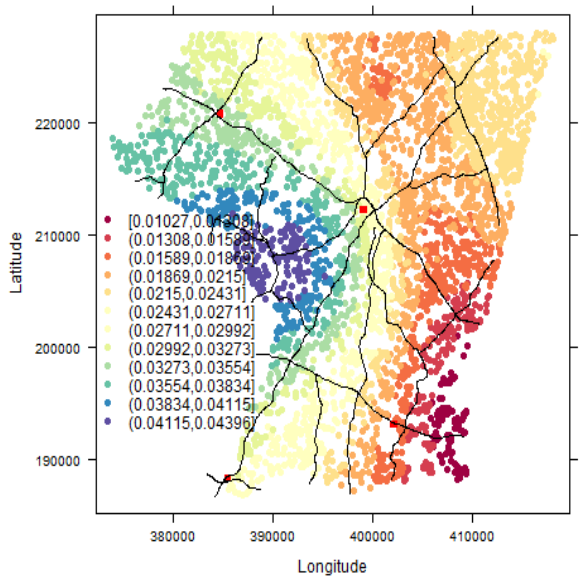

Figure 6.4: Surrounding Residential

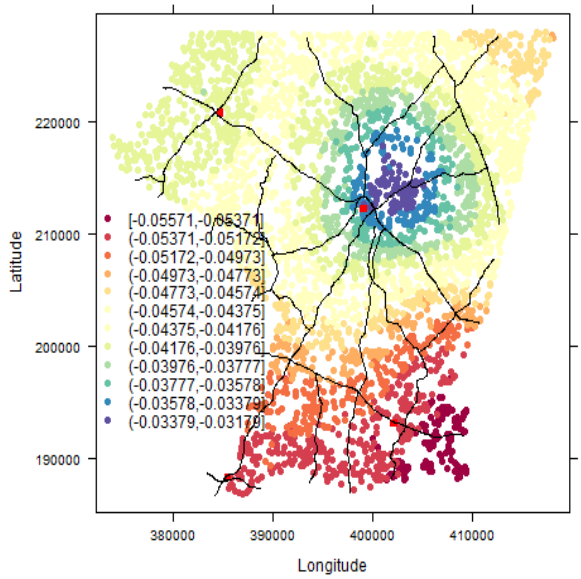

Figure 6.5: Surrounding Developable with Structure

thereby reducing the probability of development. Assuming that vacant land near the urban boundary is more valuable, then an increase in regulatory uncertainty will necessarily reduce the probability development on more urban parcels and increase or leave constant development on more exurban and rural parcels. As a result, development will occur in a discontinuous manner as costs are reduced and profits raised on more remote parcels. This latter result provides an alternative economic explanation, separate from the amenity-driven growth explanation, for the leapfrogged development patterns observed in many exurban areas.

In this paper, we examine the impact of increased regulatory uncertainty on the development timing decision of the developer and the spatial pattern of development. We also separately identify and compare this regulatory effect from the effects of local supply-side and demand-side factors. Using a unique panel dataset of subdivision approval times from an exurban county in Maryland, we develop a dynamic measure for the impact of increased regulation on the development process. We then use this measure as proxy variable for regulation in a nonparametric duration model. The results from our model show that regulation-induced implicit costs reduce the probability, that they have a larger impact on the development decision than local land use interactions and that spatial heterogeneity in the implicit costs are contributing to the leapfrogged development patterns observed in our study region. Our results show that the impact of regulation on the joint development decisions is three times larger than the impact of local land use interactions. This result is particularly important given the emphasis that previous studies have placed on land use interactions. Our results suggest that, when the spatial effects of regulatory uncertainty and local land use interactions are separately accounted for, that the implicit costs of regulation play a greater role in impacting the location, timing and pattern of land development than do local open space or congestion effects. Finally, spatial mapping of the parameter estimates of the model show that spatial heterogeneity in the application of subdivision regulations and in the resulting regulatory uncertainty have contributed to a substitution away from major subdivision development located closer to urban areas towards minor subdivisions located in more rural areas. Given these results, we conclude that these unequal regulatory effects have had clear unintended consequences in terms of the spatial distribution of development. Rather than preserving large areas of agriculturally zoned areas, as perhaps was the original intention of the county, these zoning policies have fostered a notable increase in land use fragmentation and patterns of leapfrog development.

References

Bar-Ilan, A., and W. Strange. 1999. "The Timing and Intensity of Investment." *Journal of Macroeconomics* 21:57–77.

—. 1996. "Urban Development with Lags." *Journal of Urban Economics* 39:87–113.

Bayer, P., R. McMillan, A. Murphy, and C. Timmins. 2010. "A Dynamic Model of Demand for Houses and Neighborhoods." Working paper.

Beck, N., J. Katz, and R. Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Po* 42:1260–1288.

Brown, D., K. Johnson, T. Loveland, and D. Theobald. 2005. "Rural Land Change in the Conterminous U.S., 1950-2000." *Ecological Applications,* 15:1851–1863.

Brueckner, J. 1990. "Growth Controls and Land Values in an Open City." *Land Economics* 66:237–248.

—. 1997. "Infrastructure Financing and Urban Development: The Economics of Impact Fees." *Journal of Public Economics* 66:383–407.

Bulan, L., C. Mayer, and C. Somerville. 2009. "Irreversible Investment, Real Options, and Competition: Evidence from Real Estate Development." *Journal of Urban Economics* 65:237–251.

Burchfield, M., H. Overman, D. Puga, and M. Turner. 2006. "Causes of Sprawl: A Portrait from Space." *Quarterly Journal of Economics* 121:587–633.

Capozza, D., and R. Helsley. 1990. "The Stochastic City." *Journal of Urban Economics* 28:187–203.

Cleves, M., R. Gutierrez, W. Gould, and Y. Marchenko. 2008. *An Introduction to Survival Analysis Using Stata*. Stata Press, College Station, Texas.

Cunningham, C. 2007. "Growth Controls, Real Options, and Land Development." *The Review of Economics and Statistics* 89:343–358.

—. 2006. "House Price Uncertainty, Timing of Development, and Vacant Land Prices: Evidence for Real Options in Seattle." *Journal of Urban Economics* 59:1–31.

Dixit, A., and R. Pindyck. 1994. *Investment Under Uncertainty*. Princeton University Press, Princeton, New Jersey.

Fischel, W. 2004. "An Economic History of Zoning and a Cure for it Exclusionary Effects." *Urban Studies* 41:317–340.

Fujita, M. 1976. "Spatial Patterns of Urban Growth: Optimum and Market." *Journal of Urban Economics* 4:428–447.

Glaeser, E., J. Gyourko, and R. Saks. 2005a. "Why Have Housing Prices Gone Up?" *American Economic Review* 95:329–333.

—. 2005b. "Why Is Manhattan So Expensive?" *Journal of Law and Economics* 48:331–370.

Green, R., S. Malpezzi, and S. Mayo. 2005. "Metropolitan-Specific Estimates of the Price Elasticity of Supply of Housing, and Their Sources." *American Economic Review* 2:334–339.

Irwin, E., and N. Bockstael. 2007. "The Evolution of Urban Sprawl: Evidence of Spatial Heterogeneity and Increasing Land Fragmentation." *Proceedings of the National Academy of Sciences* 104:20672.

—. 2002. "Interacting Agents, Spatial Externalities and the Endogenous Evolution of Residential Land Use Patterns." *Journal of Economic Geography* 2:31–54.

—. 2004. "Land Use Externalities, Open Space Preservation, and Urban Sprawl." *Regional Science and Urban Economics* 34:705–725.

Klaiber, H.A., and D. Phaneuf. 2010. "Valuing Open Space in a Residential Sorting Model of the Twin Cities." *Journal of Environmental Economics and Management* 60:57–77.

Lee, C., and M. Fujita. 1997. "Efficient Configuration of a Greenbelt: Theoretical Modeling of Greenbelt Amenity." *Environment and Planning A* 29:1999–2017.

Lewis, D., B. Provencher, and V. Butsic. 2009. "The Dynamic Effects of Open-Space Conservation Policies On Residential Development Density." *Journal of Environmental Economics and Management* 57:239–252.

Loader, C. 1999. *Local Regression and Likelihood*. New York: Springer.

Majd, S., and R. Pindyck. 1987. "Time to Build, Option Value, and Investment Decisions." *Journal of Financial Economics* 18:7–27.

Malpezzi, S., and S. Mayo. 1997. "Getting the Housing Incentives Right: A Case Study of the Effects of Regulation, Taxes, and Subsidies on Housing Supply in Malaysia." *Land Economics* 73:372–391.

Mayer, C., and C. Somerville. 2000. "Land Use Regulation and New Construction." *Regional Science and Urban Economics* 30:639–662.

Mayo, S., and S. Sheppard. 2001. "Housing Supply and the Effects of Stochastic Development Control." *Journ of Housing Economics* 10:109–128.

McConnell, V., M. Walls, and E. Kopits. 2006. "Zoning, TDRs, and the Density of Development." *Journal of Urban Economics* 59:440–457.

McMillen, D., and J. McDonald. 2004. *L. Anselin and J.G.M. Florax and Sergio J. Rey eds. of Advances in Spatial Econometrics*, Springer-Verlag, chap. Locally Weighted Maximum Likelihood Estimation: Monte Carlo Evidence and Application. pp. 225–239.

McMillen, D., and C. Redfearn. 2010. "Estimation and Hypothesis Testing for Non-Parametric Hedonic Price Functions." *Journal of Regional Science* 50:712–733.

Mills, D. 1981. "Growth, Speculation, and Sprawl in a Monocentric City." *Journal of Urban Economics* 10:201–226.

Murphy, A. 2010. "A Dynamic Model of Housing Supply." Mimeo.

Nechyba, T., and R. Walsh. 2004. "Urban Sprawl." *Journal of Economic Perspectives* 18:177–200.

Newburn, D., and P. Berck. 2006. "Modeling Suburban and Rural-Residential Development Beyond the Urban Fringe." *Land Economics* 82:481–499.

Ohls, J., and D. Pines. 1975. "Discontinuous Urban Development and Economic Efficiency." *Land Economics* 3:224–234.

Ortalo-Magne, F., and A. Prat. 2007. "The Political Economy of Housing Suppy: Homeowners, Workers, and Voters." *Discussion Paper: London School of Economics*, pp. .

Peiser, R. 1989. "Density and Urban Sprawl." *Land Economics* 65:193–204.

Pindyck, R. 1993. "Investments of Uncertain Cost." *The Journal of Financial Economics* 34:53–76.

Prentice, R., B. Williams, and A. Peterson. 1981. "On the Regression Analysis of Multivariate Failure Time Data." *Biometrika* 68.

Quigley, J., and S. Raphael. 2005. "Regulation and the High Cost of Housing in California." *America* 95:323–328.

Quigley, J., and L. Rosenthal. 2005. "The Effects of Land Use Regulation on the Price of Housing: What Do We Know? What Can We Learn?" *Cityscape* 8:69–137.

Sevelka, T. 2004. "Subdivision Development: Risk, Profit, and Developer Surveys." *The Appraisal Journal* Summer:242–252.

Thorson, J. 1997. "The Effect of Zoning on Housing Construction." *Journal of Housing Economics* 6:81–91.

Titman, S. 1985. "Urban Land Prices Under Uncertainty." *American Economic Review* 75:505–514.

Towe, C., C. Nickerson, and N. Bockstael. 2008. "An Empirical Examination of the Timing of Land Conversions in the Presence of Farmland Preservation Programs." *American Journal of Agricultural Economics* 90:613–626.

Walsh, R. 2007. "Endogenous Open Space Amenities in a Locational Equilibrium." *Journal of Urban Economics* 61:319–344.

Wheaton, W. 1982. "Urban Residential Development Under Perfect Foresight." *Journal of Urban Economics* 12:1–21.

Williams, J. 1991. "Real Estate Development as an Option." *Journal of Real Estate Finance and Economics* 4:191–208.

8 Appendix

Appendix A: Proxy for Price Drift and Volatility

While the focus of this research is on the impact of increased regulatory costs on the joint timing-density decision and its ability to explain exurban and rural development patterns, previous work in modeling real options in the context of land use has shown that prices also plays a role in determining the timing of development. Thus, we include measures of price drift and volatility as controls for these real options variables in effort to separate the impact of prices from our variable of interest regulatory costs.

To construct our measure of house price drift and volatility, we use a locally-weighted regression technique. In a fully structural model, we would use estimates for the evolution of the asset value. However, we lack adequate revenue data on these values. So, we proxy for the potential returns to housing by constructing local measures of price drift and volatility for each parcel, in each period. The best proxies for the real options variables are those that can capture the effects of price growth and variability as it evolves over time. One possibility is to use median home prices or some other measure of the central tendency. However, given the differentiated product nature of houses, this is insufficient to capture the quality-related differences among them. In this paper, we adopt a similar approach to measuring these values as previous literature and use data on previous house sales and market activity to construct local quality-adjusted price indices for each observation in our dataset, for each period the parcel was eligible for development (Towe, Nickerson, and Bockstael, 2008; Bulan, Mayer, and Somerville, 2009; Cunningham, 2006). The prediction of the real options model is that both drift and volatility in returns to investment affect development because they affect the option value of the irreversible investment. As they increase it should delay investment.

To construct these indices we used historical data on house prices and property characteristics collected from the state tax assessors office. We matched the property transactions from 1993 through the last quarter of 2007 with another dataset containing the characteristics of the structure and parcel. These characteristics were used to control for quality differences between the houses. We kept only observations with price values within three standard deviations of the mean. We also threw out any houses with a square footage under 800 or over 10000 to remove outliers. The final dataset contained 18,158 arms-length transactions. The descriptive statistics for the covariates used in each of the price index models are shown in table 8.1.

In previous real options research the values for the drift and volatility were calculated for predetermined geographic regions such as school districts or census tracts. While this makes intuitive sense in the case of school districts (schools often have large explanatory power in house price indices), this is not possible in

Table 8.1: Summary Statistics: Price Index Data

Variables	**Mean**	**SD**	**Min**	**Max**
Log Real Price	12.36	0.44	11.32	13.97
Quarter 1	0.21	0.41	0.00	1.00
Quarter 2	0.29	0.45	0.00	1.00
Quarter 3	0.26	0.44	0.00	1.00
Quarter 4	0.24	0.43	0.00	1.00
Log Dist. Balt.	3.57	0.17	3.14	4.15
Log Dist. West.	2.49	0.78	-1.48	3.45
Public Services	0.73	0.44	0.00	1.00
Ag. Zoning	0.07	0.26	0.00	1.00
Log Area (Acres)	-1.00	1.16	-3.57	3.93
Log Sqrft.	7.55	0.37	6.68	8.87
Struc. Qual.	1.65	0.56	1.00	4.00
Garage	0.73	0.44	0.00	1.00
Full Baths (#)	2.03	0.61	1.00	6.00
New House	0.39	0.49	0.00	1.00
Age Struc.	4.22	4.69	0.00	14.00
Town House	0.17	0.37	0.00	1.00
N=18,158				

our context as the entire county is under a single school district. Thus, the only alternative is to divide the county between census tracts or to divide it along some other random geographic line. However, both of these are arbitrary and there is no reason to believe that either of them defines the boundaries of local markets; it is much more likely that the coefficients and predicted values vary across space.

To keep from making an arbitrary judgment on the geographic cutoff, we use a nonparametric or locally weighted regression (LWR) technique to estimate values for drift and volatility in price. This approach to estimating local prices indices is similar in nature to two recent dynamic structural models of housing demand and supply (Bayer et al., 2010; Murphy, 2010).[19] We estimate the following model for each price observation, in each period (quarter) for our house price dataset:

$$\sum_{i=1}^{n}(ln(RP_{it}) - \alpha_{jt} - \gamma_{jt}\tau_i - \beta'_{jt}X_i)^2 K\left(\frac{(X_i - X_j)}{h}\right), \tag{8.1}$$

where the dependent variable is the natural log of real house prices (in 2000 dollars), and the explanatory variables are parcel and location characteristics of the parcel. The final term, $K(\bullet)$, is the kernel density function, which determines the weight that each observation, i, receives in estimating the values of the target observation, j. In this paper we use a tri-cubic kernel. The intuition is that in each year, t, we estimate

[19]The application of LWR and other nonparametric techniques has gained ground in research in recent years. The main advantage of using these models is the functional form flexibility and ability to capture spatial heterogeneity in the coefficient effects that is smoothed over in fully parametric models that only return the global average for each variable. For a recent review and application of these techniques in the areas of hedonic modeling see McMillen and Redfearn (2010).

this equation at each point, j, using all of the observations, i, that occurred in the previous two years and that fall within our observation window for point j. Thus, in period t the sample set for the entire county consists of all house sales in the years $((t-1)-(t-2))$. The number of observations used to estimate the coefficients at each point and in each period in time is obtained by selecting all those parcels that fall within a 40% window, which determines the number of observations that receive positive weight in each regression. While the choice of kernel has been shown to have little impact on the coefficient values, the choice of the window size, h, is much more important.

The literature on the choice of optimal window size is still growing, but the most common method for determining the window size is the method of *cross validation*. This technique runs LWRs over a series of window sizes, and finds the optimal window size by minimizing the sum of the squared residuals over all window sizes. For our model, the average optimal window size was 40%. In actuality, the window size changes size at each point and the number of observations stays the same. At each point in the dataset all distances that fall below the 40th percentile are selected and used in estimating the local model. Thus, for areas with clustered sales data the window is quite small as are the weights; for areas with very sparse data the window size can be quite large, but weights decline quickly and decrease the effect of distant observations.[20]

Our measure of price *drift* is defined as the regression average growth rate in real house prices over the previous two years, and it is defined by γ_{jt} in equation 8.1. In each regression, at each location, and for a given time variable, τ_i takes on the value of the quarter in which the sale occurred. Thus, the coefficient on this variable picks up the quality adjusted price trend over the previous two years for a given period estimate.

Our measure of *volatility* is constructed from the residuals of each model. The Root Mean Squared Error (RMSE) of a regression model represents the error of the model in properly predicting the outcome variable. In many real estate and land development settings the best predictive measure of price volatility that a potential developer has is the uncertainty inherent in the hedonic price regression of the market forecast. The better the fit of the model, the lower will be the RMSE. One problem, however, with using the RMSE as a measure of volatility is that the value is not unit free, and it is not possible to compare different models, i.e. between time periods and regions. So, to control for this, we use the Coefficient of Variation (CV), which divides the model's RMSE by the weighted value of the dependent variable so that the units cancel out. The equation for this statistic is given by:

[20]This technique implies that in rural areas where the most recent sales may be relatively far away the local model still uses the closest 40%, but, given their distance, they are significantly down-weighted. This technique makes intuitive sense in modeling how actual landowners make their conversion decisions.

$$CV_{jt} = 100 * \frac{RMSE_{jt}}{\sum_{i=1}^{n} ln(RP_{it}) K\left(\frac{(X_i - X_j)}{h}\right)}. \tag{8.2}$$

To attach our estimated values of drift and volatility to each parcel we first determine the closest sales observation to each parcel using a Mahalanobis distance calculation. Then, we assign the value of the drift parameter and calculated coefficient of variation of that point to the undeveloped parcel. The intuition is that in each period that a parcel is undeveloped the owner of that parcel will determine the value of housing in that period and in her geographic market by using some subset of the most recent house sales in the county and weighting those parcels that are closer more than those further away. Our application of the LWR technique and subsequent matching to undeveloped parcels accomplishes that goal.

Given that we estimate the hedonic model at each point and across time, our model for price drift and volatility produces a lot of data and it would be impossible to report all of it. So, we report the average values for both measures across the county and separated out by sales inside and outside of the agriculture zoning areas. Figures 8.1 and 8.2 report these for the entire county. From these figures we can see that, while the drift is fairly similar for the two regions, the volatility measure is somewhat higher in agriculture districts.

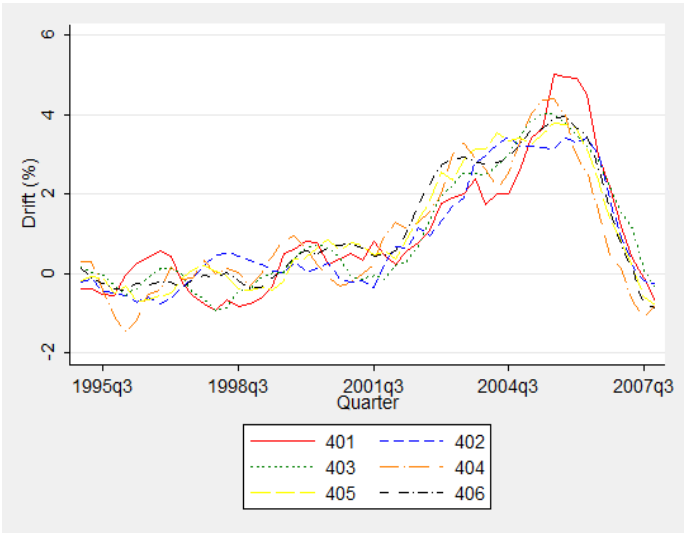

Figure 8.1: Carroll County House Price Drift

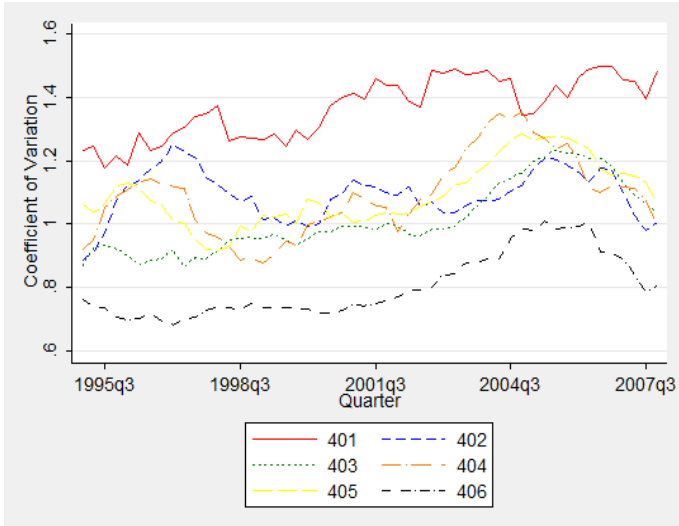

Figure 8.2: Carroll County House Price Volatility

Appendix B: Proxy for Regulatory Costs

Duration analysis models the time, t, until the occurrence of a specific event, d_n, while controlling for censoring in the event that an observation leaves the dataset early or reaches the end of the observation period without an event occurring. The entire event history for each observation is captured by the cumulative distribution function:

$$F(t) = Prob(T \leq t) = \int_0^t f(u)du, \tag{8.3}$$

and the survival function:

$$S(t) = 1 - Prob(T \leq t) = \int_0^t f(u)du = Prob(T > t), \tag{8.4}$$

where equation 8.3 is the probability that a particular event happens before time t and equation 8.4 is the probability of survival past that point or censoring. Taking the derivative of the first equation and then taking the limit of the ratio of the two gives the hazard rate:

$$h(t) = \lim_{\Delta t \to \infty} \frac{P(t \leq T < t + \Delta t)}{\Delta t} = \frac{f(t)}{S(t)}. \tag{8.5}$$

This is the conditional probability of an event occurring in the period $[t, t + \Delta t]$.

This rate is referred to as the "baseline" hazard rate and it is the rate of occurrence without controlling for

any other factors influencing the event. In the case of land development, this is the rate at which landowners choose to develop. In the full proportional hazard model, a series of covariates, x_n, are included, which provides an explanation of the differences in the hazard rates among the observations. The full model is specified as follows:

$$h_n(t; x_n) = \lambda_0 exp(x_n' \beta), \tag{8.6}$$

where λ_0 is the baseline hazard and x_n is a K-dimensional set of covariates for individual n. Thus, each observation is classified by the following set, (t_0, t_n, d_n, x_n).

Most duration models of land use have considered the occurrence of a single event and modeled only final approval of the land conversion. However, in our particular case, we are interested in modeling the time until approval from the date of the initial submission of the plan as this is the time period that is most uncertain to the developer and the event most likely to affect the choice of development timing and lot quantity. Consequently, we model the entire subdivision timing decision using a multi-event duration model. To do this, we apply the inter-event (or gap time) conditional risk set model of Prentice, Williams, and Peterson (1981). In each period, we consider all N^t parcels that were undeveloped as of 1989 and were eligible for subdivision into at least two lots, where the superscript t indicates the period of estimation. Each parcel, n, is eligible for one of two events, $k \epsilon \{1, 2\}$, in each period, and the model is conditional in the sense that a person must experience the first event to eligible for the second. That is, the landowner must have gained second-stage approval to be eligible to gain final approval. Each parcel is classified by the following two sets, $(t_{n01}, t_{n11}, d_{n1}, x_{n1})$ and $(t_{n02}, t_{n12}, d_{n2}, x_{n2})$. In this regard, we model the time for each event as the time from the previous event. For example, for the second event we have $Y_{i2} = d(t_{i2} \geq t > t_{i1})$.

The hazard function for the multi-event duration model is given by:

$$h_{nk}(t; x_{nk}) = \lambda_{0k}(t - t_{k-1}) exp(x_{nk}(t)' \beta_k), \tag{8.7}$$

where the baseline specifies that we are modeling the time from the previous event and $\beta_k = (\beta_{1k}, \beta_{2k}, \ldots, \beta_{pk})$ is a $p\ x\ 1$ vector of event-specific coefficients with k specifying the particular event.

Up until this point we have not specified a functional form for the baseline hazard other than stating that it is measuring the time from the previous event and it is indexed by the event type k. One option is to leave the baseline hazard unspecified and estimate a flexible Cox proportional hazard model. However, this is not possible in our particular context as we are interested in using the results of the model to predict completion

times. The Cox model is nonparametric in the baseline hazard and cannot be used for prediction. Thus, we must choose a parametric functional form for the baseline hazard.

The baseline hazard can be considered constant, as in the exponential model, it can increase or decrease monotonically as in the Weibull distribution, or it can take on a more flexible specification that captures both positive and negative duration dependence as is the case with the piece-wise exponential model. While this latter model is preferable, it suffers from a number of issues that make its use in our particular context difficult. First, while it provides solution to the issues of prediction in the Cox model, it does so at the expense of efficiency as it adds time fixed effects for each period of observation. This is particularly important in the early years of our model when we have limited subdivision activity with which to identify the parameters. And second, the addition of the time-varying baseline hazard makes prediction difficult and in many cases worse than would be the case if a simply parametric model was used (Cleves et al., 2008). Given these difficulties and the fact that we are using the model to generate predictions for regulatory uncertainty to be used in the main model, we estimate the model in each period using the exponential ($h_0(t) = exp(a)$) and the Weibull model ($pt^{p-1}exp(a)$) and use a likelihood ratio test to compare the two. In all cases, we reject the null hypothesis of the Weibull parameter being equal to zero. Thus, all of our prediction models use this latter specification. The final likelihood function for the multi-event duration model is given by:

$$L(\beta_k) = \prod_{i=1}^{N} \prod_{k=1}^{2} h_{ik}((t_{ik1} - t_{ik0}), \beta_{ik})^{d_{ik}} S_{ik}((t_{ik1} - t_{ik0}), \beta_{ik})^{1-d_{ik}}, \tag{8.8}$$

where the first term is over the individuals and the events that fail during the first or second observation period and the second term is over the individuals and observations that are censored during either the first or second event.

Given the spatial nature of our data and the unobserved differences between the agents, it seems likely that there will be many unobserved factors that influence the decision to submit the first application as well as the rate at which the process is completed. Some agents may have a better legal aptitude in being able to complete the regulatory process, they may be better able to forecast future demand, or have different financial needs and desires for their parcel. Because we cannot observe all of these factors, it is likely that individual heterogeneity and event dependence between the two stages exists. By modeling the choices as conditional on previous events and allowing the baseline hazard to vary between events, we can account for the event dependence in the data. To account for the heterogeneity across individuals, we use a robust sandwich Huber/White variance-covariance estimator.

To produce our estimates of regulatory uncertainty we estimate 13 separate duration models. Each of

the models is estimated using the undeveloped parcels and subdivision events that occurred in all periods preceding the one of interest. For example, to produce the predicted approval time for undeveloped parcels in 1995, we use the estimates of the second-stage duration model, which are for all previous subdivisions that had completed both stages of the development process before 1995. Thus, in each period, only those developments that had finished the second stage and gained final approval were used in the predicting the development time for undeveloped parcels. The implication is that each landowner uses previous approval timing information and her parcel's characteristics to produce an estimate of the likely time until approval for her own parcel if she chose to subdivide. For each model, we include those covariates most likely to affect the decision to convert. Table 8.2 gives the summary statistics for those variables used in each Weibull model.

Table 8.2: Multi-Event Duration Model: Summary Statistics

Variables	**Mean**	**SD**	**Min**	**Max**
Dist to Balt	41.22	8.09	23.17	65.50
Dist to West	14.73	6.42	.08	34.90
Trans Access	4.97	5.34	0	63.41
Area (acres)	35.47	45.48	.32	591.16
Zoned Lots	9.56	28.10	0	653
Existing House	.51	.49	0	1
Ag. Zoning	.53	.49	0	1
Public Services	.14	.35	0	1
Soil 1 (%)	39.90	42.90	0	100
Soil 2 (%)	52.61	42.90	0	100
Slope (15%)	16.88	28.72	0	100
Forest Cover (%)	32.07	31.91	0	100.00

Given the amount of information generated by these models, it is not feasible nor necessary to show all of the results. Instead, figures 5.8 and **??** (shown in the data section of the paper) show the kernel density curves for the predicted results from the models for the years 1994 and 2002.[21] The figures show two separate curves - one curve for parcels that remain undeveloped throughout the model and a second for those parcels that ended up developing in future periods. As is clear from the figures, there are definite differences between the two, but they still retain the same shape.

The predicted values from each of the duration models is matched with the parcels in the sample selection model in each period. The intuition is that in each period of a sample that a parcel remains undeveloped the landowner uses past approval times to form her own expected approval time on her parcel for that period.

[21] We chose 1994 as it is the set of predictions used in the first year of the sample selection model; we chose 2002 as it was the year that the county increased regulation on larger developments.

These data serve as our measure of expected future regulatory costs for subdivision development on each parcel and the value of this parameter identifies the effect of regulatory uncertainty on conversion timing and the number of lots created in our Probit and Poisson models.