

Time Is Money: An Empirical Examination of the Dynamic Effects of Regulatory Uncertainty on Residential Subdivision Development

Douglas H. Wrenn^{1,2}

Job Market Paper

October, 2012

Agricultural, Environmental and Development Economics

Ohio State University

¹Corresponding Author, Ohio State University, 2120 Fyffe Road Room 103, Columbus, OH 43210, wrenn.7@buckeyemail.osu.edu, www.douglaswrenn.org. Copyright 2012 by Douglas Wrenn. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

²Research support gratefully acknowledged from a cooperative agreement with the U.S. Forest Service Northern Research Station, National Science Foundation DEB-0410336 and Grant No. 0423476 and the James S. McDonnell Foundation.

Time Is Money: An Empirical Examination of the Dynamic Effects of Regulatory Uncertainty on Residential Subdivision Development

Abstract: Variations in regulatory costs over time and across different types of investment projects create non-diversifiable risk for developers who hold land. So-called implicit costs that arise from regulatory uncertainty are hypothesized to be potentially large, but empirical evidence is limited to a metropolitan-level study of aggregate housing markets by Mayer and Somerville (2000a). Using a unique micro dataset on parcel-level subdivision development, data on the timing of subdivision approvals by a local planning agency and a sample selection Poisson model, we test the effects of implicit costs that arise from uncertain subdivision approval on the timing, quantity and pattern of residential subdivision development. Consistent with theory, we find that these regulation-induced implicit costs reduce the probability and size of subdivision development on any given parcel. In addition, we find that systematic correlation between implicit costs and subdivision size has resulted in an implicit cost advantage for smaller subdivision projects that, in combination with zoning regulations that restrict the allowable size of exurban subdivisions, has resulted in a substantial increase in the likelihood of exurban development relative to larger-scale suburban development. We show that a hypothetical policy that reduces the differences in approval times across suburban and exurban areas generates a more concentrated pattern of development by increasing the average size of subdivision development and decreasing the incidence of exurban development. The results contribute to the growing supply-side literature on housing and land use, and provide a new explanation of scattered residential development as the outcome of heterogeneous regulatory costs and optimal land development.

Keywords: land conversion; land use policy; real options; sample selection; spatial analysis

1 Introduction

Land development is a risky business. Because development projects take time to complete, risks related to future prices or costs can significantly alter the development decision. Regulations influence these decisions by altering both the explicit costs of development, e.g., in the form of permitting and impact fees, and the so-called implicit costs that arise from uncertain regulation. Variation in regulatory costs over time and across different types of investment projects generate implicit costs associated with the non-diversifiable risk that this uncertainty imposes on the developer. While it is well-recognized that the majority of land use regulations represent a *de jure* increase in the explicit costs of land development, real options theory suggests that a *de facto* increase in these implicit costs can have an even larger impact on housing and land market investments (Pindyck, 1993; Mayer and Somerville, 2000a; Bar-Ilan and Strange, 1996; Quigley and Raphael, 2005; Gyourko and Saiz, 2006).

The theory of optimal investment with uncertain costs generates well-known predictions of the effect of implicit costs on the reduction in the likelihood and intensity of an investment project (Pindyck, 1993). Despite this theoretical interest in implicit costs, there is little empirical evidence of how these costs influence housing markets and none that examines their impact on individual development decisions. The majority of empirical studies of regulation and housing markets use aggregate data and focus on the overall impact of regulatory stringency on housing supply at a metropolitan scale (Green, Malpezzi, and Mayo, 2005; Glaeser, Gyourko, and Saks, 2006, 2005a,b; Ortalo-Magne and Prat, 2007; Paciorek, 2011). These studies largely confirm the relationship between increased regulatory stringency and a reduction in housing supply at this regional scale, but are unable to identify implicit costs from their aggregate measure of regulatory stringency. Other studies have considered the impact of specific land use regulations on the timing, density and spatial pattern of land development using parcel data on residential subdivisions or lots (Cunningham, 2006, 2007; Towe, Nickerson, and Bockstael, 2008; Newburn and Berck, 2006; McConnell, Walls, and Kopits, 2006). These studies provide a number of interesting insights about the relationship between spatially-heterogeneous regulations and the timing, density and location of development, but, like the aggregate studies, do not consider the role of implicit costs associated with regulation. An important exception is Mayer and Somerville (2000a), who use metro-level data on housing starts and a national survey of planners to estimate the impact of regulation-induced increases in expected approval times on the number of new houses being built. They divide regulations between those adding explicit costs and those that induce cost increases by extending approval times. They find that regions with increased approval times for subdivisions can have up to 45% fewer starts and elasticities that are more than 20% lower, and that those regulations

that lengthen the approval process serve to decrease the supply response the most. These results provide key empirical evidence of the potentially large impact that implicit costs can have on housing supply. However, the aggregate scale of analysis precludes consideration of how implicit costs influence individual development decisions and the spatial structure of land markets within metropolitan regions.

The purpose of this paper is to examine the impact of implicit costs on individual decisions regarding the timing and intensity of residential subdivision development and the influence of heterogeneous implicit costs on the spatial pattern of land development. We hypothesize that implicit costs arise from uncertainty over subdivision approval times and that, following Pindyck (1993), an increase in these implicit costs will both decrease the likelihood of development of a given parcel and reduce the intensity of development or project size chosen. We investigate whether the spatial pattern of development is impacted by heterogeneity in implicit costs arising from key differences in the approval process of differently-sized subdivisions. We hypothesize that less stringent regulations that reduce the implicit costs associated with smaller subdivision development projects have favored this type of development, and that has led to an increase in the amount of low density, scattered residential development in the exurban areas of our study region. We test these hypotheses using a unique panel dataset on residential subdivision development over a 13-year time period from 1995 to 2007 from a rapidly urbanizing county within the Baltimore, Maryland region. Our dataset includes information on the length of time the county planning department took to approve each subdivision from the point at which the subdivision application was first filed. Descriptive statistics reveal large variation in approval times over this time period, ranging from a minimum approval time of one month to a maximum of 105 months. To test whether this variation in approval times had a significant effect on development outcomes, we create a parcel-specific dynamic measure of regulatory uncertainty that captures the influence of implicit costs on the expected approval times for each developable parcel. A joint Probit-Poisson model of subdivision development timing and intensity is estimated using this parcel-specific measure of implicit regulatory cost (Lewis, Provencher, and Butsic, 2009). The richness of the data enables us to identify the effect of implicit regulatory costs on development decisions, while controlling for other parcel-level variables, including soil type, slope and other variables that influence explicit costs of development as well as key economic variables, including price drift and volatility, local competition and the opportunity cost of development. We use simulation methods to examine the implications of the model under benchmark and alternative scenarios for the spatial pattern of development.

Our findings confirm the hypothesis that variation in implicit costs resulting from regulatory uncertainty exert a significant influence on both the overall development process and the spatial pattern of development.

We find that a 1-month (10%) increase in the expected approval time has a significant, but small, effect on reducing the average probability of development (by 0.13%) and a significant and more substantial reduction in the average size of the development (by 0.65 lots or 5.6%). We also find significant spatial differences in average marginal effects and very large differences in the predicted probabilities of subdivision development across suburban and exurban areas. In comparing the distributions of the predicted probability of development of exurban versus suburban subdivisions, we find that the probability of exurban development is 60% more likely and that this difference can be fully attributed to the lower implicit costs associated with small-scale exurban development. We investigate the policy implications of our model using Monte Carlo simulation methods to predict land development patterns under benchmark and alternative policy scenarios. A reduction in the difference between implicit costs in the exurban versus suburban areas is found to be successful in fostering a more concentrated development pattern. Specifically, we find that by increasing the implicit costs in exurban areas by 10% and reducing them by 10% in suburban areas results in an increase in the predicted number of total lots by 6.5%, an increase in the predicted number of subdivisions in suburban areas by 21% and a decrease in the predicted number of subdivisions in exurban areas by 12%.

The paper makes several contributions to the literature on land use regulation and urban spatial structure. The role of heterogeneity in generating discontinuous development patterns has long been emphasized in the theoretical literature (Mills, 1981; Wheaton, 1982; Newburn and Berek, 2011), but empirical evidence thus far has been lacking. Instead, previous empirical studies have focused on the role of demand-side amenities and disamenities and the role of these local land use spillovers in generating scattered exurban land development (Irwin and Bockstael, 2002; Klaiber and Phaneuf, 2010; Walsh, 2007). Using a unique parcel-level panel dataset on subdivision development, we provide the first empirical evidence of the influence of implicit regulatory costs associated with the supply of residential land on individual development decisions and on the spatial structure of land markets. Our main results are that implicit costs due to uncertain subdivision approval significantly and substantially influence the size and distribution of subdivision development and in ways that generate unintended consequences for the spatial pattern of development. In particular, we find that the combined effects of implicit costs that reduce the optimal size of development and zoning regulations that restrict the size of exurban subdivisions have fostered a clear pattern of scattered subdivision development in the exurban areas of our study region. The results offer a new explanation of scattered, low-density residential development as the outcome of heterogeneous regulatory costs and optimal land development decision making.

The remainder of the paper is structured as follows. Section 2 provides the basic theoretical model.

Section 3 presents our empirical specification. Section 4 presents the data used in the model and the construction of the measure of implicit costs. Section 5 presents the results and discussion and section 6 concludes.

2 Model of Regulatory Uncertainty

Our theoretical framework follows from Pindyck (1993), and we model the timing-intensity decision of raw land conversion as equivalent to the exercise of a financial put option. In each period, t , a developer, $n \in \{1, \dots, N\}$, makes two decisions in order to maximize the profits on her parcel: the time at which to apply for permission to develop and the quantity or size, q , of the project proposed. We assume that the market for residential housing is competitive and that each developer derives a certain return, V , from the sale of the residential lots created from the parcel.³ The total cost of a project is given by $K(q)$, which is an increasing function of the quantity or size of the project ($K'(q) > 0$). The payoff function is given by:

$$\max[0, V - K]. \quad (2.1)$$

As is the case with any real option, an increase in uncertainty over the underlying values will increase the value of the option to wait (Dixit and Pindyck, 1994).

Each subdivision project takes time to complete, and the total amount of progress that is possible in each period is given by k . In the case of no uncertainty, the total time required for the project is given by $T = \frac{K}{k}$ and the value of the option to invest is given by:

$$F(K) = \max \left[Ve^{-r\frac{K}{k}} - \int_0^{\frac{K}{k}} ke^{-rt} dt, 0 \right], \quad (2.2)$$

where r is the real rate of interest. The investment rule for equation 2.2 is to invest as long as the return, V , is above the critical value, K^* , given by:

$$K^* = \left(\frac{k}{r} \right) \log \left(1 + r \frac{V}{k} \right). \quad (2.3)$$

When the real rate of interest, r , is equal to zero, the result is the same as the neoclassical NPV model. For values of r greater than zero, $F(K) < (V - K)$ as costs are spread over the entire project, but the discounted

³Given our interest in the impact of regulatory uncertainty on the subdivision development decision, we assume that returns from land development are known with certainty and focus on the role of cost uncertainty on the optimal timing-intensity decisions.

returns are only realized at the end of the project. As is clear from equation 2.2, $F(K)$ is a convex function of the total cost of the project. As a result, any increase in uncertainty over this value will increase the option value and delay the project.

To focus on the role of implicit costs, we take explicit costs as given and introduce implicit costs with the following diffusion process:

$$dK = -Idt + \gamma Kdw, \quad (2.4)$$

where I is the rate of investment at time t with $0 \leq I(t) \leq k$, γ is the standard deviation or volatility parameter impacting the total cost of the project and dw is an increment of a Weiner process. This equation reflects uncertainty over both the effort needed to complete the project and time that it will take to complete, \tilde{T} , where time is now uncertain as result of the fact that K is uncertain. Note that the total cost of the project, K , is changing even in the case where $I(t) = 0$. The implicit costs of regulation are outside of the control of the developer, and they change regardless of whether the project has been started or not. This is evidenced by the uncertainty term, which is not a function of the current investment, $I(t)$.

Assuming that, because of the bang-bang nature of the problem, it is optimal to invest the maximum amount in each period ($I(t) = k$), the value of the investment opportunity is now given by the following equation:

$$F(K; V, k) = \max_{I(t)} E_0 \left[V e^{-r\tilde{T}} - \int_0^{\tilde{T}} I(t) e^{-rt} dt, 0 \right], \quad (2.5)$$

where E_0 signifies that the developer forms an expectation during the first period regarding the cost or time to complete the approval process, \tilde{T} , which is uncertain at the outset. Two conditions must hold for this equation to make sense in our particular context. First, $F(K; V, k)$ needs to be a decreasing function of the expected costs, which implies that the interaction between increased uncertainty, γ , and projects that take a longer time to complete, K , combine to increase the option value (to waiting) of a given investment project. And second, from equation 2.4, as the total cost of the project approaches zero so does the variance, which implies that smaller projects face less uncertainty.

The solution to the previous problem shows that $F(K; V, k)$ satisfies the following differential equation:

$$\frac{1}{2}\gamma^2 K^2 F_{KK} - IF_K - \gamma K F_K - I = rF, \quad (2.6)$$

which implies that the expected return on the asset portfolio - in this case the asset is the subdivision

development option - is set equal to the opportunity cost of capital or the risk-free rate of interest times the value of the investment.⁴

Equation 2.6 provides the theoretical basis for our empirical specification, and allows us to formulate the following testable hypotheses. First, if we take the total size of a particular project, $K(q)$, as given, then as the uncertainty, γ , increases, the value of the option to delay investment increases. This implies that as uncertainty increases only smaller projects are optimal and larger projects get delayed. Second, assuming that the developer also chooses the total size of the project, then as uncertainty increases the developer will reduce the optimal size of the project, $K(q)$. In the context of undeveloped land, this implies that as uncertainty increases the developer will hold some land vacant as a hedge against future uncertainty and reduce the number of lots created.

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 development for parcels located in more high-density suburban areas, while leaving exurban parcels unchanged, then a relative increase in expected future costs and reduction in expected future revenues on suburban parcels would increase the value of vacant suburban land, but not change the value of vacant exurban land. This should reduce the supply of suburban housing and increase the incentive to develop at or beyond the urban fringe. This result is consistent with the urban economic model in which intertemporal and spatial differences in costs lead to scattered development (Ohls and Pines, 1975; Mills, 1981; Peiser, 1989).

3 Econometric Model of Residential Subdivision Development

In our study region, subdivision approval is a two-step process. In the first stage, owners of raw land apply for conditional subdivision approval. The county approves the plan conditional on meeting certain regulatory requirements. Final approval, the second stage of the process, is granted once all of these regulatory demands have been met. The time required to meet these demands can take anywhere from a few months to many years, and it is at least partially uncertain, from the developer's perspective, at the beginning of the process. Given that the second stage of this process is necessarily endogenous to the developer, we model the decision beginning at the first stage where the optimal stopping decision (to begin the process at all) and the density decision (on how many buildable lots to create) are made conditional on expectations formed about the time

⁴A full solution to this problem can be found in Pindyck (1993).

needed to gain final approval.

We cast the landowner's decision problem as one of choosing the optimal number of lots, q , to create in time, t , in order to maximize profits, conditional on having reached her optimal development time. Given that we are modeling subdivision development and not a single developed lot, the smallest number of lots that can be created is two. Thus, the choice for each landowner in each period is $q_t = 2, 3 \dots$ with the lot quantity decision truncated at one lot. Given the decision to subdivide, each landowner chooses the optimal value of q to maximize the following profit equation:

$$\Pi(X_{nt}, \mu_{nt}) = \max_{(q_t=2,3,\dots)} [V_q(X_{nt}) + \mu_{nt}], \quad (3.1)$$

where X_{nt} is a set of parcel characteristics, including our measure of cost uncertainty, that affects the profit on the parcel and μ_{nt} is a decision-specific random effect that is unobserved by the researcher, but is observed by the landowner and impacts her timing decision. This specification makes explicit our empirical estimation approach to the two-step decision problem presented in section 2. As the developed value of a parcel rises, it eventually reaches its optimum, which translated into the optimal development time for that parcel. Given the panel nature of our data, the equation inside the brackets is the latent decision problem of the landowner of whether or not to develop in each period and it represents the optimal stopping decision.

Conditional on this optimal stopping decision, each choice of lot quantity, q , produces a different value of the function 3.1, and the maximization problem is over the optimal lot quantity conditional on having chosen to develop. This choice is equivalent to the intertemporal choice made in section 2 as landowners choose some amount of their parcel to develop and some amount to keep vacant. As a result, the optimal number of lots created on a parcel is given by the optimal land value function:

$$q^*(X_{nt}, \varphi_n) = \arg \max_q [V_q(X_{nt}) + \mu_{nt}], \quad (3.2)$$

where μ_{nt} is an i.i.d. standard normal random variable and φ_n represents a time-invariant random effect for each parcel that captures the unobserved individual heterogeneity and the fact that there are unobservable factors impacting both the value function, the equation inside of the brackets, as well as the optimal number of lots created.

Given the integer nature of our second-stage outcome, we model the expected number of lots created as a truncated count process represented by the following equation:

$$\text{Pois}(q_{nt}|X_{nt}, \varphi_n) = \frac{e^{-\exp(X'_{nt}\beta + \varphi_n)} (\exp(X'_{nt}\beta + \varphi_n))^q}{q!(1 - e^{-\exp(X'_{nt}\beta + \varphi_n)} - e^{(X'_{nt}\beta + \varphi_n)} e^{-\exp(X'_{nt}\beta + \varphi_n)})}, \quad (3.3)$$

where X_{nt} is a set of variables and φ_n is an individual-specific random term impacting the optimal quantity decision on each parcel, n , at time, t . If we specify the profit function, equation 3.1, by the latent variable model, $V^*(X_{nt}, u_n) = \alpha X_{nt} + u_{nt}$, then the unobservable factors in the equation for the decision to develop, μ_{nt} , and those for the decision on the number of lots to create, φ_n , are joint-normally distributed and are necessarily correlated as the unobserved random effects are correlated across both processes for the same decision maker. We specify the correlation between these random effects terms with the coefficient ρ . If this term is significantly different from zero, then the two decisions are not statistically independent and estimating them separately would produce inconsistent results. In the next section we introduce a sample selection model that accounts for the correlation as well as the truncated nature of our data.

3.1 Empirical Specification

To account for the two-step nature of the development decision and the potential correlation across the unobserved individual effects, we estimate a sample selection Poisson model with random effects. In each period, we observe first, whether a parcel converts, d_{nt} , with $d_{nt} = 1$ if the parcel develops, and second, conditional on the decision to develop, we observe the number of lots chosen on the parcel, q_{nt} . Both of these decisions are made conditional on the parcel characteristics, X_{nt} . This type of sample selection model was first developed by (Greene, 1995; Terza, 1998; Greene, 2005), and it has recently been applied in land use modeling (Lewis, Provencher, and Butsic, 2009) and previously to a travel-cost model (Englin and Cameron, 1996).

Let the decision to develop in period t be $d_{nt}^* = X'_{nt}\alpha + u_{nt}$, where $u_{nt} \sim N[0, 1]$. Under this assumption, the standard normal distribution function for the probability of development is given by:

$$\text{Prob}(d_{nt} = 1|X_{nt}) = \Phi(X'_{nt}\alpha). \quad (3.4)$$

This is a discrete choice Probit model that captures the binary timing decision. The expected value for the number of lots created is modeled as a truncated Poisson conditional on the covariates and the unobserved random effect is as follows (Cameron and Trivedi, 1998):

$$\text{Pois}(q_{nt}|X_{nt}, \varepsilon_n) = \frac{e^{-\exp(X'_{nt}\beta + \sigma\varepsilon_n)} (\exp(X'_{nt}\beta + \sigma\varepsilon_n))^q}{q!(1 - e^{-\exp(X'_{nt}\beta + \sigma\varepsilon_n)} - e^{(X'_{nt}\beta + \sigma\varepsilon_n)} e^{-\exp(X'_{nt}\beta + \sigma\varepsilon_n)})}, \quad (3.5)$$

where $\varepsilon_n \sim N[0, 1]$ and the term inside the parentheses in the denominator handles the truncation of the number of lots selected. The joint distribution of u_{nt} and ε_n follows a bivariate normal distribution, $[u_{nt}, \varepsilon_n] \sim [(0, 0), (\sigma, 1), \rho]$ and the selectivity is transmitted via the ρ term. By the joint normality of these terms, we can write the selection portion of the model as:

$$\text{Prob}(d_{nt}|X_{nt}, \varepsilon_n) = \Phi\left(\frac{X'_{nt}\alpha + \rho\varepsilon_n}{\sqrt{1 - \rho^2}}\right). \quad (3.6)$$

This result is obtained by the Cholesky decomposition of the Probit function using the joint normality assumption $f(u_{nt}|\varepsilon_n)$ (Greene, 2005). Combining equations 3.5 and 3.6, the unconditional joint density is obtained by integrating out the ε_n term from the density function:

$$\begin{aligned} \text{Prob}(d_{nt} = 1, q_{nt}|X_{nt}, \varepsilon_n) &= \int_{-\infty}^{\infty} \left[\frac{e^{-\exp(X'_{nt}\beta + \sigma\varepsilon_n)} (\exp(X'_{nt}\beta + \sigma\varepsilon_n))^q}{q!(1 - e^{-\exp(X'_{nt}\beta + \sigma\varepsilon_n)} - e^{(X'_{nt}\beta + \sigma\varepsilon_n)} e^{-\exp(X'_{nt}\beta + \sigma\varepsilon_n)})} \right] \\ &\quad * \left[\Phi\left(\frac{X'_{nt}\alpha + \rho\varepsilon_n}{\sqrt{1 - \rho^2}}\right) \right] \phi(\varepsilon_n) d\varepsilon_n. \end{aligned} \quad (3.7)$$

Given the symmetry of the normal cdf function, we can write equation 3.7 as:

$$\begin{aligned} \text{Prob}(d_{nt} = 1, q_{nt}|X_{nt}, \varepsilon_n) &= \int_{-\infty}^{\infty} \left[(1 - d_{nt}) + d_{nt} \frac{e^{-\exp(X'_{nt}\beta + \sigma\varepsilon_n)} (\exp(X'_{nt}\beta + \sigma\varepsilon_n))^q}{q!(1 - e^{-\exp(X'_{nt}\beta + \sigma\varepsilon_n)} - e^{(X'_{nt}\beta + \sigma\varepsilon_n)} e^{-\exp(X'_{nt}\beta + \sigma\varepsilon_n)})} \right] \\ &\quad * \left[\Phi\left((2d_{nt} - 1) \frac{X'_{nt}\alpha + \rho\varepsilon_n}{\sqrt{1 - \rho^2}}\right) \right] \phi(\varepsilon_n) d\varepsilon_n. \end{aligned} \quad (3.8)$$

When $d_{nt} = 0$, the Poisson problem disappears and the equation reduces to the lower portion of the cumulative Probit distribution. Note that this is a full information specification in the sense the decisions are modeled jointly.

3.2 Model Estimation

The empirical model has four sets of parameters to be estimated: $(\beta, \alpha, \sigma, \rho)$. The first set of coefficients, β , give the effects of the covariates on the number of lots created; the second set, α , give the effect of the covariates on the probability of the development; the third parameter, σ , gives the estimated standard

deviation of heterogeneity across parcels in the Poisson model;⁵ and the final parameter, ρ , captures the sample selection correlation between the unobservable factors across the two models. Given the panel data nature of our data, we let N be the total number of parcels available for development during our study period and T the total number of periods in our dataset. We estimate the parameters of the model using Full Information Maximum Likelihood (FIML).

The random effects specification of the log likelihood function is as follows:

$$\begin{aligned} \ln L = \sum_{n=1}^N \ln \left\{ \int_{-\infty}^{\infty} \left[\prod_{t=1}^T \left[\left[((1 - d_{nt}) + d_{nt}) \frac{e^{-\exp(X'_{nt}\beta + \sigma\varepsilon_n)} (\exp(X'_{nt}\beta + \sigma\varepsilon_n))^q}{q!(1 - e^{-\exp(X'_{nt}\beta + \sigma\varepsilon_n)} - e^{(X'_{nt}\beta + \sigma\varepsilon_n)} e^{-\exp(X'_{nt}\beta + \sigma\varepsilon_n)})} \right] \right. \right. \right. \\ \left. \left. \left. * \left[\Phi \left((2d_{nt} - 1) \frac{X'_{nt}\alpha + \rho\varepsilon_n}{\sqrt{1 - \rho^2}} \right) \right] \right] \right] \phi(\varepsilon_n) d\varepsilon_n \right\}, \end{aligned} \quad (3.9)$$

where the integral is taken across all n agents and the inner product handles the time-invariant random effect. As a result of the random effects, we are forced to take the expectation of a multidimensional integral, which has no closed form solution. This equation can be solved using either a quadrature method or Monte Carlo simulation (Greene, 2011). Given the large number of observations in our dataset, we use a simulation approach. In order to estimate parameters of the model, we replace equation 3.9 with simulated log-likelihood given by:

$$\begin{aligned} \ln L_{\text{sim}} = \sum_{n=1}^N \ln \left\{ \frac{1}{R} \sum_{r=1}^R \prod_{t=1}^T \left[\left[((1 - d_{nt}) + d_{nt}) \frac{e^{-\exp(X'_{nt}\beta + \sigma\varepsilon_n)} (\exp(X'_{nt}\beta + \sigma\varepsilon_n))^q}{q!(1 - e^{-\exp(X'_{nt}\beta + \sigma\varepsilon_n)} - e^{(X'_{nt}\beta + \sigma\varepsilon_n)} e^{-\exp(X'_{nt}\beta + \sigma\varepsilon_n)})} \right] \right. \right. \\ \left. \left. * \left[\Phi \left((2d_{nt} - 1) \frac{X'_{nt}\alpha + \rho\varepsilon_n}{\sqrt{1 - \rho^2}} \right) \right] \right] \phi(\varepsilon_n) d\varepsilon_n \right\}, \end{aligned} \quad (3.10)$$

and estimate the model via maximum simulated likelihood (Train, 2003; Greene, 2011). We simulate the random effects using a series of 200 Halton draws from a standard normal distribution.⁶

⁵This heterogeneity parameter serves the same role as the sigma in the lognormal Poisson model and relaxes the restrictive assumption in the general Poisson model that the conditional mean and variance of the data model must be equal.

⁶The entire model is estimated using maximum likelihood code written in Matlab.

4 Description of Data and Covariates

4.1 Study Region

Our study region is Carroll County, Maryland, an exurban county within the Baltimore metropolitan region that witnessed rapid population growth from the 1960's onward.⁷ In response to burgeoning growth pressures, the county passed its first comprehensive zoning plan in 1963, which restricted development density to one house per acre in all areas of the county without public sewer facilities. Increased growth in the 1970's led to the passage of a second comprehensive plan in 1978, which included a massive down zoning of 70% of the land in the county to low-density agricultural zoning. This zoning class has a stated density of one house per 20 acres, but its actual effective density is closer to one house per 15 acres due to some leeway that was built into the law. Specifically, each parcel located in an agriculture zoning area and having at least six acres of land is allowed to create two buildable lots; each additional lot requires an addition 20 acres. Apart from several small adjustments made in 1989, these same restrictions have been in place in the county since 1978.⁸

The 1963 comprehensive plan also provided the first formal procedure for the creation of large major and small minor residential subdivisions that is still in place today. Large subdivisions consist of any development with four or more buildable lots at the time of development and require the installation of streets, storm water management facilities and other infrastructure. Small developments are two or three lots and are not subject to any infrastructure requirements. In addition, while approval of large subdivisions requires a formal public hearing, small subdivisions can simply be approved by the chairmen of the planning board without a formal hearing. According to county planning officials, in most cases small developments can gain approval in less than two or three months; large developments, however, require an open public hearing as well as the approval of numerous county agencies, which can significantly increase the time until approval.

While the combination of the exurban zoning policy and the creation of the formal subdivision policy was intended to control exurban and rural development and reduce the fragmentation the rural landscape, our subdivision data indicate otherwise. Our data show that over 60% of all subdivisions created from 1995 through 2007 were platted in exurban areas and that of these, 82% were small minor developments. Table 4.1 reveals that, on average, large developments take as much as 13 months longer to get approved and that,

⁷A map of the county and our study region is shown in appendix 7.1.

⁸Given the amount of land in the county that falls within the agriculture zoning class and the drastic difference in density between this class and the other four zoning classes in the county, the estimation and analysis in the remainder of this paper will differentiate between these two areas by classifying all agricultural areas as low-density exurban development areas and the other four classes as high-density suburban development areas.

in general, there is a positive correlation between approval times and the size of the development. This table also shows that a substantial difference remains even between small and medium-sized developments suggesting it is not just a difference between large and small subdivisions. Table 4.2, which shows the differences in approval times between exurban and suburban density areas, reveals even larger differences in approval times between subdivisions. Consequently, if these differences in approval times between subdivision types and between different spatial units translate into differences in implicit costs faced by developers, then it is possible, given the theoretical predictions of our model, that they could have a substantial impact on the process and pattern of subdivision development as developers minimize costs between development types and between exurban and suburban areas of the county.

Table 4.1: Distribution of Approval Times by Subdivision Size

Time (Months)	Mean	25th	Med.	75th	N
Large Developments	22.24	8	13	28.5	136
Small Developments	8.95	2	4	7	258
Large Developments (Over 9 lots)	25.94	10	16	31	68
Medium Developments (4 to 9 lots)	17.25	4	9	18	67
Small Developments	8.95	2	4	7	258

Table 4.2: Distribution of Approval Times by Density Class

Time (Months)	Mean	25th	Med.	75th	N
Exurban Density	8.89	2	4	8	247
Suburban Density	23.25	6	14	28	150

4.2 Data Construction and Description

We constructed several micro panel datasets of residential subdivision at the parcel level to estimate the econometric model.⁹ First, we constructed a panel of historical residential subdivision development, which we assembled by combining a current GIS parcel boundary file with the tax assessor’s database and historical records of subdivision plats from the Maryland Archives. By matching the individual parcels in the parcel boundary shapefile with the plat maps, we determined all of the parcels in each development, assigned

⁹A full description of the data construction process for this paper is given in the appendix 7.2.

each development a unique ID number and a date when the subdivision first gained approval. Second, we constructed data on the historical evolution of land preservation and protected open space in the county. In 1980 Carroll County began its own purchase of development rights (PDR) program in an effort 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. Finally, we reconstructed the history of the subdivision approval process for each of the subdivisions in our dataset by collecting 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.

The final dataset consists of all undeveloped parcels that, as of 1995, 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 classes in the county: agriculture, conservation, and residential (specifically, R40, R20, or R10). This yields a total of 3,844 parcels of which a total 397 (or a little over 10% of the parcels in the county) gained final approval between 1995 and 2007. Another 343 were preserved during this period. We consider these parcels as undeveloped until the quarter they are preserved at which time they drop out of our dataset. We assume that once a parcel reaches its full development potential it leaves our dataset. Thus, some parcels that are in the dataset at the beginning are not there in the final periods.

4.3 Covariates Used in the Empirical Model

Table 4.3 defines all of the variables used in the estimation of our econometric land use model and gives their summary statistics. Given the structural nature of our model and the similarities between our micro-level model of land development and similar structural supply models estimated at the metropolitan level (Mayer and Somerville, 2000a,b), we frame the discussion of these variables in the context of a structural econometric model of the supply of buildable lots.

The first set of variables in the table pertain to explicit or variable costs hypothesized to impact the timing and intensity decision. Most macro-level structural models have accounted for variable costs using a metro-level real building or development cost index. This, however, is not possible in our context as we are estimating the model in a single metro region. Moreover, because of the richness of our dataset, we are able to capture a much more spatially-explicit measure of local parcel costs and proxy for the most important

factors that would impact this index in our region.¹⁰ The first variables, Sewer and ExUrbZone, are fixed effects for parcels located in areas with public sewer facilities and in low-density zoning areas, respectively. It is likely that parcels located in areas with sewer will be forced to pay impact fees or other explicit costs at the time of development. Since we do not have data on these costs, we proxy for them by using a fixed effect for this area. The sign on this variable could be positive or negative depending on how these costs enter the profit function and impact the value of the investment and the conversion decision. ExUrbZone captures a similar process for low-density areas that may reduce explicit costs and speed up development in these area for reasons discussed above. Both variables account for the direct costs as opposed to the indirect impact of being in these areas on implicit costs or expected approval times, which will be explained below.

Area is the size of the parcel in acres. Larger parcels are advantageous for both development as well as for agriculture. So, the sign on this coefficient can be taken as ambiguous. However, given that agriculture rarely out-competes urban development, we expect the sign to be positive. The final four variables, Soil1, Soil2, SteepSlp and FrstPrct, capture parcel-level costs that may impact development. These costs are similar to building materials in the case of building construction. As with capital inputs to building construction, the market for labor and capital inputs for land excavation and conversion is competitive and the costs associated with development on any parcel, while they vary spatially, should be directly proportional to the characteristics of the parcel and not be uncertain from the point of view of the developer. The soils variables (Soil1 and Soil2) are the percentage of each type of soil on the parcel with Soil1 being the best soil. There are three main types of soils so both of these are relative to the excluded category or worst type, Soil3. SteepSlp is the percentage of the parcel that is over 15% slope and FrstPrct in the percentage of the parcel covered in forest. We expect better soils to speed up development and Slope and Forest to reduce it if they raise costs. All of these variables are constant across time.

The second set of variables are what we term economic variables and are those variables hypothesized to impact revenues and proxy for local competition and the opportunity cost of development. The first variable, EaseElig, is a proxy for the opportunity cost of preservation. Given the size of the land preservation program in our study region, it is possible that the presence of the option to preserve could reduce willingness to develop on parcels that are both eligible for preservation and development. So, EaseElig is time-varying indicator variable for whether a parcel is eligible for easement in each time period. To be eligible for preservation in our study region a parcel must be greater than 50 acres and have more than 50% of type 1 and 2 soils combined or be between 25 and 50 acres and border a previously-preserved parcel. Thus,

¹⁰This method of accounting for variable costs is similar to other land use models (Towe, Nickerson, and Bockstael, 2008; Newburn and Berck, 2006; Irwin and Bockstael, 2002).

easement eligibility is based off of the size of the parcel, the percentage of certain soil types and the proximity of a parcel to other parcels that have preserved in the past. Because of this final clause, some parcels that are not eligible in one period may become eligible in latter periods as larger parcels around them preserve. Geltner, Riddiough, and Stojanovic (1996) have shown theoretically that the presence of multiple options can reduce development if they are correlated and Towe, Nickerson, and Bockstael (2008) have shown that it is empirically important in the context of land conversion. This variable is assumed to only impact the conversion or timing decision.

The next variable, *Compete*, is a proxy for the stock of existing developed lots in a local area around each parcel. We do not explicitly account for population growth in this model but, given the richness of the our subdivision and approval data, we can determine, in each time period, the number buildable lots that have been approved based off of the dates from the plat maps. So, for each parcel in our data set, we calculate the number of approved buildable lots in a 10% region around each parcel in the previous two years. This variable proxies for both local competition and the local stock or existing supply of buildable lots for each parcel, in each period. The last two variables in this section are our real-options variables for price drift and volatility (*Drift* and *Volatile*). Mayer and Somerville (2000b) showed that in the context of a structural supply model it is the change in prices, and not the level, that determines supply. So, to account for the change or drift in price level in our model, we use historical house price data and a hedonic estimation strategy to develop a census-tract-level measure of price drift for each time period. We then use the residuals from each of the tract-level hedonic models and develop a measure of price volatility based off of the coefficient of variation.¹¹ We also include a full set of time fixed effects to account for other time-constant unobservables that may impact development such as interest rates, returns to the stock market or macro-level income shocks. The construction of our proxy for implicit regulatory costs, (*Regulate*), is explained in the next section.

4.4 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.¹² We then use the estimates from each model and in each period to predict approval times for all undeveloped parcels in the same period. Duration modeling has been applied extensively in land use modeling to capture

¹¹A full description of the construction of the price variables is given in the appendix 7.3.

¹²This is a multi-event duration model of sequential events or development stages and not a model of competing risks.

Table 4.3: Covariates Used in Sample Selection Model

Variables	Description	Mean	Std. Dev.	Min.	Max.
Explicit Costs					
Sewer	Public Services	0.15	0.36	0.00	1.00
ExUrbZone	ExUrban Density	0.51	0.50	0.00	1.00
Area	Area (Acres)	32.63	41.02	0.46	365.56
Soil1	Type 1 Soils (%)	40.04	43.21	0.00	100.00
Soil2	Type 2 Soils (%)	52.67	43.19	0.00	100.00
SteepSlp	Greater than 15%	17.27	29.21	0.00	100.00
FrstPrct	Forest Cover (%)	33.50	32.21	0.00	100.00
Economic Variables					
EaseElig	Easement Eligibility	0.24	0.42	0.00	1.00
Compete	Competition	13.95	21.19	1.00	135.00
Drift	Price Drift	0.97	1.77	-6.09	7.74
Volatility	Price Volatility	0.006	0.005	0.001	0.045
Implicit Costs					
Regulate	Regulatory Costs	10.64	5.15	1.98	95.39

the optimal stopping and dynamic nature of the land conversion decision (Irwin and Bockstael, 2002; Towe, Nickerson, and Bockstael, 2008; Cunningham, 2006, 2007). In this essay, we extend the traditional single-event duration approach to multiple events in order to generate expected approval times based on previous approval events.

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:

$$F(t) = \text{Prob}(T \leq t) = \int_0^t f(u)du, \quad (4.1)$$

and the survival function:

$$S(t) = 1 - \text{Prob}(T \leq t) = \int_0^t f(u)du = \text{Prob}(T > t), \quad (4.2)$$

where equation 4.1 is the probability that a particular event happens before time t and equation 4.2 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:

$$h(t) = \lim_{\Delta t \rightarrow \infty} \frac{P(t \leq T < t + \Delta t)}{\Delta t} = \frac{f(t)}{S(t)}. \quad (4.3)$$

This is the unconditional 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 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:

$$h_n(t; X_{nt}) = \lambda_0 \exp(X'_{nt} \beta), \quad (4.4)$$

where λ_0 is the baseline hazard and X_{nt} is a K -dimensional set of covariates for parcel n . Thus, each observation is classified by the following set, $(t_0, t_n, d_{nt}, x_{nt})$.

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, $k \in \{1, 2\}$, in each period, and the model is conditional in the sense that a parcel must experience the first event to be eligible for the second. That is, the landowner must have gained second-stage approval to be eligible to gain final approval. Thus, each parcel is classified by the following two sets, $(t_{n01}, t_{n11}, d_{n1}, x_{n1})$ and $(t_{n02}, t_{n12}, d_{n2}, x_{n2})$, which are sets describing the start, stop and events or censoring for each stage of the two-stage model. 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 $Y_{n2} = d(t_{n12} \geq t > t_{n02})$.

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

$$h_{nk}(t; X_{nk}) = \lambda_{0k}(t - t_{k-1}) \exp(X_{nk}(t)' \beta_k), \quad (4.5)$$

where the baseline specifies that we are modeling the time from the previous event and $\beta_k = (\beta_{1k}, \beta_{2k}, \dots, \beta_{pk})$ is a $p \times 1$ 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 standard 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 in the piece-wise exponential model. While the piece-wise exponential model is preferable, it suffers from a number of issues that make its use difficult in our particular context. First, while it provides a 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 estimate 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 simpler parametric model was used (Cleves et al., 2008). Given these difficulties and the fact that we are using the model to generate predictions for expected approval times to be used in the sample selection model, we estimate the model in each period using just 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. The final likelihood function for the multi-event duration model is given by:

$$L(\beta_k) = \prod_{n=1}^N \prod_{k=1}^2 h_{nk}((t_{nk1} - t_{nk0}), \beta_{nk})^{d_{nk}} S_{nk}((t_{nk1} - t_{nk0}), \beta_{nk})^{1-d_{nk}}, \quad (4.6)$$

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

Given the spatial nature of our data and the unobserved differences between the parcels, 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 parcel owners 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 parcel 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 parcels, we use a robust sandwich Huber/White variance-covariance estimator.

To produce our estimates of expected approval time for each parcel in each time period of our main model, we estimate 13 separate models starting in 1994 and use the lagged value of the estimates in our model starting in 1995 to control for contemporaneous correlation between the estimates of expected approval time from the duration model and the observed covariates in Probit-Poisson model. 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 from the model and estimate of the second-stage baseline hazard. The coefficient values 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 chooses to subdivide. It is this process of modeling the first-stage decision and using past activity to create predictions about expected approval times that allows us to model the impact of regulatory uncertainty on the joint timing-intensity decision and control for the inherent endogeneity in second stage of the process.

Table 4.4 gives the summary statistics for those variables used in each duration model. We include those variables that are most likely to impact the timing of subdivision approval on a parcel conditional on the decision to develop. The first two variables, travel time to Baltimore (DisttoBalt) and the distance to primary roads (DistPrimeRoads), proxy for accessibility. It is assumed that if a parcel is located closer to the CBD in a more densely-populated area or if it is closer to a major road, then it could potentially increase the approval time on the parcel as the zoning board or local residents delay the approval process as they debate the increased congestion and traffic as result of the proposed development. The variable ExHouse is a dummy for whether the parcel had a house on it in each period. In many cases, small developments result from the further subdivision of an existing parcels that already have an existing structure (Lewis, Provencher, and Butsic, 2009). If that is indeed the case and it expedites the approval process, then we would expect this variable to decrease expected approval time. The next variable, FloodPlne, is our proxy for environmental sensitivity and is an indicator variable for whether the parcel is located in a flood zone, wetland area or riparian area. If these areas require more time to get approved as the appropriate impact

studies are conducted, then we expect this to increase expected approval time. The remaining variables, ZonedLots-R10Zone, all relate to zoning. The first variable is the zoned number or allowable number of lots on each parcel and the last four are fixed effects for zoning with R10Zone being the highest density area and CnsvZone the lowest. All of these zoning dummies are in terms of the remaining dummy for low-density, exurban zoning. Given our previous discussion about the the location of development in the county, we expect that approval times will be higher in each of the areas relative to low-density, agricultural areas.

Table 4.4: Multi-Event Duration Model: Summary Statistics

Variables	Mean	Std. Dev.	Min.	Max.
DisttoBalt	39.55	8.09	23.17	65.50
DisttoPrimeroads	1.39	1.48	0.01	9.98
ExistHouse	.51	.49	0	1
FloodPlne	.53	.49	0	1
ZonedLots	9.56	28.10	0	653
CnsvZone	0.17	0.35	0	1
R40Zone	0.09	0.28	0	1
R20Zone	0.06	0.24	0	1
R10Zone	0.09	0.29	0	1

Given the number of models run in forming our implicit cost predictions and the amount of output produced, it is not possible to present all of the results here. So, to summarize the results from these predictions and show that they follow the assumptions presented above, we first plot our predictions for the entire dataset versus the actual development times for all subdivisions with approval-time data. Figure 4.1 shows both of these kernel density curves for the predicted and actual times. While these density curves do not overlap perfectly, they do reveal a similar shape and the predictions appear to cover the entire domain of the actual approval times. Given the the out-of-sample nature of our cost prediction these results seem reasonable.

As a second means of evaluating our measure of implicit costs, we regress the cost predictions for the entire dataset, which includes the predictions from all years, on the variables used to form those predictions. We also include a full set of time dummies to soak up the time-varying unobservable effects in the predictions, but do not display them. This process allows us to get an indication of the relationship between the coefficients in the duration models and the variables used to form the predictions and summarize our cost predictions over our entire observation period. The results from this process are shown in Table 4.5 and generally conform to the expectations stated above.

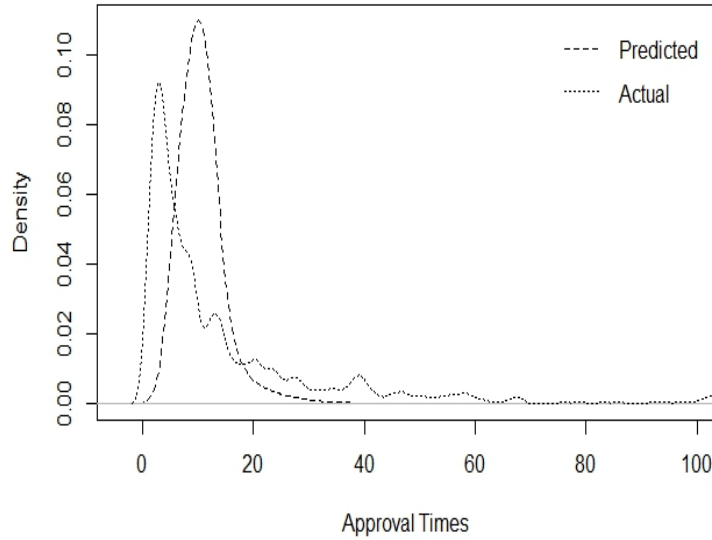


Figure 4.1: Expected Approval Times: 1994-2006

Table 4.5: Cost Variable Regression

Variables	Coeff.	Std. Err.
DisttoBalt	-1.588	0.246
DisttoBaltSqrd	0.025	0.002
DisttoPrimeroads	-0.295	0.026
ExistHouse	-0.758	0.089
FloodPIne	1.417	0.106
ZonedLots	-0.023	0.002
CnsvZone	6.079	0.242
R40Zone	6.196	0.191
R20Zone	6.943	0.025
R10Zone	6.729	0.206
Constant	30.117	5.005

4.5 Endogeneity of Cost Variable

One important issue that must be addressed is the potential endogeneity that may still exist between our prediction of expected approval time and the error structure in the sample selection model. First, there is the potential for temporal endogeneity between the data used to form the prediction and the errors in the sample selection model. However, this issue has been accounted for by our strategy of modeling the initial decision

to start the development project and not the final approval date. Thus, by using only previously-developed subdivisions to form the predictions and not past data on the the same parcel, we can control for temporal correlations between the error structures in the two models. Second, even after accounting for the temporal effects it is still possible that unobserved time-invariant effects in the error structure of the prediction model could cause our cost variable to be correlated with those same unobservables in the sample selection model, which would produce inconsistent and inefficient parameters estimates on our measure of cost.

One solution to this problem would be to run a fixed effects model and difference out the unobservables. However, given the nonlinear nature of our data this is not feasible. Thus, we employ the technique for handling endogeneity in nonlinear models suggested by Zabel (1992). Using this technique, we build correlation into the model by specifying each individual-specific random effect, ε_n , as a function of the initial value of the cost variable in the first period: $\varepsilon_n = \delta \text{RegulateFixed} + \gamma_n$, where γ_n is distributed standard normal, ρ and σ still capture the correlation and heterogeneity, respectively and the variable `RegulateFixed` serves as a proxy for individual-specific fixed effects and controls for time-invariant effects for each person that may cause correlation between our measure of cost and the standard error in the sample selection model. We only include this measure in the first-stage Probit model as the second-stage Poisson does not have enough variation over time and produces collinearity problems during estimation.

5 Results and Discussion

5.1 Model Results

The results from our empirical model are shown in Table 5.1. First, we see that the correlation coefficient is negative, but not statistically different from zero, which indicates that the unobservable factors impacting the timing decision are statistically-independent from the factors affecting the density decision. Thus, our analysis of the models and their marginal effects can proceed separately as we do not have to take account of the indirect effect in analyzing the marginal effects for the Poisson model. Second, we see that the heterogeneity term, σ , which is the parcel-level standard deviation from the Poisson model, is significant. This implies that the conditional mean and variance of the standard Poisson model are not equal and that the log-normal specification provides a better fit to the data. This result is not surprising given the large number of small developments during our study period, which necessarily produces a non-standard count distribution. Finally, we see that our individual fixed effect, `RegulateFixed`, is not significant indicating that endogeneity between our cost measure and the errors in the sample selection model do not appear to be an

Table 5.1: Model of Regulatory Uncertainty

	Timing Decision		Intensity Decision	
	Probit Model		Poisson Model	
	Coeff.	Std. Err.	Coeff.	Std. Err.
Explicit Costs				
Sewer	0.078	0.069	0.750**	0.217
ExUrbZone	0.038	0.056	-2.338**	0.199
Area	0.002**	0.000	0.017**	0.001
Soil1	0.002*	0.001	-0.011**	0.004
Soil2	0.002*	0.001	-0.007*	0.004
SteepSlp	-0.001	0.001	-0.005	0.003
FrstPrnt	-0.002**	0.000	-0.009**	0.003
Economic Variables				
EaseElig	-0.167**	0.064		
Compete	-0.001*	0.000	-0.001	0.003
Drift	0.040*	0.022	-0.007	0.091
Volatile	-2.206	5.295	-7.901	26.538
Implicit Costs				
Regulate	-0.067**	0.018	-0.084**	0.034
RegulateFixed	0.016	0.018		
Constant	-1.911**	0.188	3.555**	0.587
Sample Selection				
Correlation (ρ)	-0.043	0.073		
Heterogeneity (σ_n)	1.031**	0.064		
Log Likelihood	-3007.388			
* $p < 0.10$, ** $p < 0.05$				
N=45556				
Note: A full set of time fixed effects was included in the model.				

issue. In the remainder of this section, we present a brief overview of the results from the model. In the next section, we analyze the marginal effects for our measure of cost uncertainty, present our analysis of the broad spatial distribution of these marginal effects and present a counterfactual exercise using the marginal effects from the timing model that shows how our findings can be used to impact the spatial pattern of development. Finally, we use the coefficients from our model in a Monte Carlo policy simulation to compare the in-sample predictions of our model with the actual quantity of developments and lots created during our study period.

Our measures of explicit costs generally conform to expectations. First, we see that large parcels both increase the probability of development as well as the number of lots created upon development. This

result implies that in our study region development is out-competing agriculture. Second, the parcel-level characteristics for cost in the Probit model have the appropriate signs with better soils increasing the likelihood of development and parcels with a larger percentage of forest cover reducing it. Third, we see that the presence of public facilities has no impact on the probability of development, but that it does increase the number of lots created. This is as we would expect given that areas with public facilities allow for higher-density development. It also indicates that any additional explicit costs associated with being in an area with public facilities must be offset by the returns to investing in more lots. Fourth, contrary to expectations, parcels located in agriculturally-zoned areas are no less likely to develop than parcels in other areas, although they do experience a reduction in the number of lots created when they do develop. The reduction in the number of lots created is as expected, but the fact that agricultural zoning has no impact on the probability of development is contradictory to economic theory and to the intention of the agricultural zoning regulation. Finally, we see that better soils actually reduce the number of lots created. This result is opposite of what we would expect if better soils reduce the variable costs associated with development. The main explanation for this result is that within our study region the highest-quality soils are located in the areas closest to the urban center. This is also the same area that has experienced the greatest development pressure. As result, if this increased development has translated into higher land prices and increased implicit costs, then it is easy to see that better soils, in this context, could lead to a reduction in the number of lots created if these other factors offset the benefit of better soils.

The results from our economic variables show that real options matter, but only for the timing decision. Our measure for the opportunity cost of preservation is significant and indicates that for parcels that are eligible for preservation, the probability of development is reduced compared to those that are not. This conforms to earlier results in the land use literature (Towe, Nickerson, and Bockstael, 2008). Our measure of local competition is also significant and shows that as the stock of locally-approved lots increases it leads to a reduction in the probability of development. We also see that our measure of price drift is positive and significant indicating that a rise in house price levels leads to an increase in the probability of development. However, our measure of volatility, while it has the appropriate sign, is not significant in either equation. One explanation for this result and for the fact that our real option's variables are not significant in the density model is that we also included a full set of time fixed effects. It is likely that all of the uncertainty impacting the timing and density decision is being soaked up by these variables. These variables also capture systematic macro-level trends in uncertainty and interest rate changes that vary over time, but not across parcels.

In the final section, we observe that our measure of implicit cost has the expected sign and is significant in both models. The Probit model indicates that an increase in the expected time until approval reduces the overall probability of development. The Poisson model indicates that a similar increase reduces the expected number of lots created. We now present the marginal effects for our implicit cost variable and analyze the spatial distribution and policy implications of changes in this variable.

5.2 Marginal Effects

The results reported in Table 5.1 provide an indication of the significance of the covariates on the timing and intensity decisions, but, because of the highly nonlinear form of the model, are not equal to the marginal effects. In order to evaluate the magnitude of our measure of implicit costs in both the Probit and Poisson models, we calculate the corresponding marginal effects using a discrete change method and use the Delta method to calculate the standard errors.¹³ Because of the log-normality and truncation in the Poisson model, the marginal effects for this equation must be simulated. Simulation is carried out using the same set of Halton draws used to estimate the FIML model. The results from this process are given in Table 5.2.¹⁴

The values in these tables can be interpreted as the average change in the absolute value of the probability of development and the average change in the number of lots created across all parcels for a 10% increase in the expected approval time on each parcel, in each time period. We chose to use a percentage change (as opposed to an absolute change in the approval time) to account for the fact that absolute changes can have different impacts depending on the initial starting value of costs on the the parcel. A percentage change, however, adds proportionally to each parcels cost variable and allows for a better comparison.

The marginal effects for the Probit model indicate that a 10% increase in expected approval time, which is equivalent to 1.06 months, leads to a 0.131% reduction in the average probability of development. For the Poisson model, we find that the same increase in approval time decreases the average number of lots created by 0.646. This latter result is equivalent to a full one-lot reduction for each 1.3 month increase in approval time. These results are consistent with the basic theoretical predictions of our model: an increase in uncertainty over implicit costs implies that it is optimal to delay development, thus reducing the likelihood of development in a given period, and to reduce the optimal size of the project conditional of deciding to

¹³See appendix 7.4 for a full derivation of these marginal effects.

¹⁴It is important to point out that, while the estimates and marginal effects for the Poisson model are normal in size, the probabilities and marginal effects for the Probit model are very small (in comparison to what is normally presented in other articles). The reason for these small probability values is that during our study period we have a lot of potentially-developable parcels (45,556), but only 397 subdivisions actually occur. Thus, the overall probability of any parcel developing at any point in time is small. This is a standard result in land use models looking at subdivision conversions and not the development of individual lots - especially when the number of cross sections or time periods increase.

develop.

Table 5.2: Marginal Effects for Implicit Cost Variable

	Avg. dydx	Z-Score
Probit Model (%)		
Regulate (Exp. Approval Time)	-0.131**	-2.948
Poisson Model (# of Lots)		
Regulate (Exp. Approval Time)	-0.646*	-1.894
Note: The marginal effects are for a 10% increase in expected approval times.		
Note: Standard errors calculated using the Delta Method.		

In examining the change in the average effects across suburban versus exurban areas (Table 5.3), we find that a 10% increase in expected approval time decreases the average predicted probability of development in suburban areas by 0.122% and in exurban areas by 0.14%. This latter result makes sense in the context of the spatial optimization problem faced by developers. Exurban land development costs, including the implicit costs of subdivision approval, are low, which generates profitable opportunities for developers. However, returns are also low and buildable land is plentiful, implying that an increase in developer costs will not be fully capitalized into the price of a finished lot. Thus any increase in development cost reduces the relative advantage of developing in this exurban zone and it becomes more profitable to build in suburban areas where buildable land is more limited and the increase in cost is reflected in higher prices for subdivision lots.

Table 5.3: Marginal Effects by Density Class: Probit

	Avg. dydx	Z-score
Exurban		
Regulate (Exp. Approval Time)	-0.140**	-2.961
Suburban		
Regulate (Exp. Approval Time)	-0.122**	-2.913
Note: Marginal effects are for a 10% increase in expected approval times.		
Note: Standard errors calculated using the Delta Method.		

5.3 Spatial Pattern of Development

Although we find that the effects of an increase in implicit costs are slightly larger in magnitude in the exurban relative to suburban areas at the margin, the fact remains that implicit costs in total are substantially lower

in exurban versus suburban areas due to the differences in subdivision size. As was reported in Table 4.2, the average time, in months, for approval of subdivision projects in the exurban versus suburban areas over our study period was 8.89 versus 23.25. Given large differences in implicit costs between the two regions, we hypothesize that the greatest impact of heterogeneous implicit costs may be on the spatial pattern of development.

To further investigate the spatial implications of our model, we use the model estimates to calculate the predicted probabilities of development for all developable parcels in our sample and plot the kernel density of parcels located within the areas zoned for higher-density suburban development versus parcels located in areas that are zoned for lower-density exurban development. We compare the means and distributions of these two plots under this baseline case versus an alternative case in which the heterogeneity in implicit costs is eliminated by setting the expected approval time equal to the observed mean approval time (10 months) for all parcels. By examining how the elimination of heterogeneity in implicit costs alters the relative means and distributions, we can infer the extent to which differences in development outcomes are due to the large differences in implicit costs across these regions.¹⁵

The results from this exercise are shown in Figures 5.1(a) and 5.1(b). It is clear from the baseline density plots that the predicted probability of development between the two areas is significantly different. The average predicted probability of development of parcels in exurban areas is almost 60% larger than of those parcels located outside these areas. After setting the expected approval times equal across all parcels, the densities and average probabilities between the two areas are virtually indistinguishable. These results illustrate the important role that implicit costs and zoning combine to play in altering the distribution of subdivision development across space. Implicit costs, arising from uncertain subdivision approval time, reduce the optimal size of subdivision development. The zoning policy greatly constrains the size of subdivision development in these exurban areas. In combination, these policies result in a substantial implicit cost advantage associated with small subdivision development, which increases the likelihood of development in exurban relative to suburban areas. When the implicit cost advantage is removed by equalizing the approval times of all subdivision projects, regardless of size, the systematic difference in the predicted probability of development between parcels located in exurban versus suburban areas disappears.

¹⁵In generating these predictions here, and for the policy simulations in the next section, we assume that a change in the implicit costs generates a direct impact on development decisions and that the general equilibrium feedbacks of these changes are negligible. Accounting for these general equilibrium effects would require a structural model of the spatial price equilibrium and demand and supply interactions between the suburban and exurban areas. These considerations are beyond the scope of this paper.

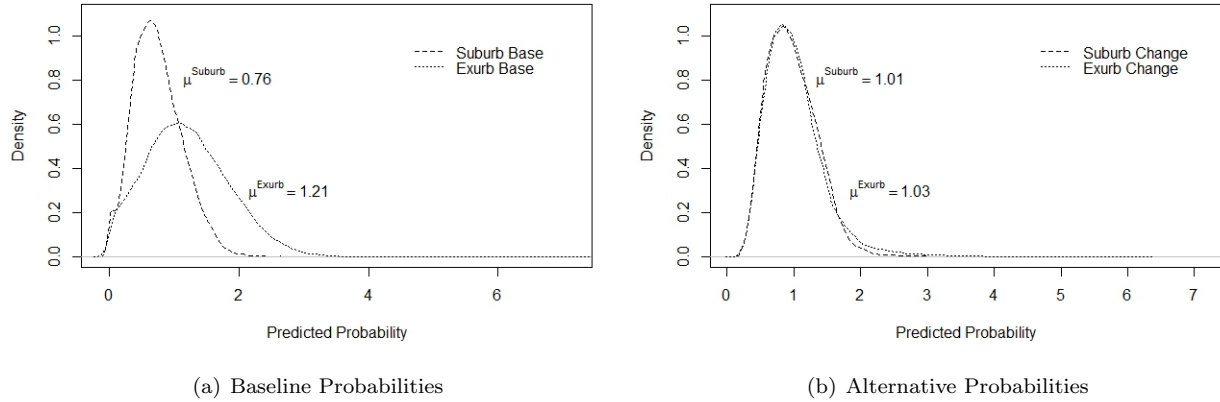


Figure 5.1: Spatial Analysis of Development Timing

5.4 Policy Simulations

The estimated marginal effects suggest a policy lever that could be purposefully manipulated to reduce exurban development and encourage more suburban development. By either lengthening the approval time needed for exurban subdivisions or streamlining the review process for suburban subdivisions or doing both, policy makers can alter the relative implicit costs of development across these two regions. We investigate the potential implications of such a counterfactual policy by comparing predicted development outcomes under a baseline scenario, in which the implicit costs are unchanged, versus a series of alternative policy scenarios in which the relative lengths of the suburban and exurban subdivision review processes are purposefully altered.¹⁶

To examine the efficiency of our model in matching in-sample development outcomes, we apply a similar simulation technique to Lewis, Provencher, and Butsic (2009). We begin with the entire sample of developable parcels at the beginning of our study period. Then, we use the estimated parameters and the covariance matrix from the model and a set random draws from a standard normal distribution to produce a random draw from our estimated parameter distribution in a method similar to Krinsky and Robb (1986).¹⁷ Using each random draw of the parameters, the predicted probability of development is calculated for each parcel in the first period using the first-stage Probit model. Then, a random uniform draw, $U \sim [0, 1]$, equal to the number of parcels in the dataset in the first period is taken. The random probabilities on each parcel

¹⁶Throughout this section we take parcel to mean the original raw land parcel and lots to mean the number of buildable, subdivided lots created after development occurs.

¹⁷Each simulated parameter is a combination of the original parameter, a Cholesky decomposition of the covariance matrix and a standard normal draw, $\hat{\theta} = \theta + C d_r$, where θ is the entire parameter vector from our model, C is the Cholesky decomposition of the variance-covariance matrix and d_r is a random draw from a standard normal distribution.

are compared to this draw with parcels whose predicted probabilities are greater than or equal to this draw considered as developed. Finally, conditional on being developed, we use the second-stage Poisson model to determine the conditional mean number of lots for that parcel. The number of new lots created is either the prediction from the model or the number allowed by zoning on that parcel if the prediction is larger than the amount allowed by zoning.¹⁸ Once a parcel develops it is removed from the dataset and the process is repeated for each of our 13 time periods with the updated dataset used in each subsequent period. The time fixed effects from both stages of the model account for changes in unobserved macro-level changes. This process is repeated 200 times for each simulation exercise.

While our model does an excellent job of predicting lot quantity outcomes for most small, medium and large subdivisions, we find that it does not do as well at predicting the number of buildable lots for very large subdivisions, i.e., for subdivisions with buildable lots over 100. Given that these subdivisions account for a larger portion of the lots created (on a per development basis), under-predicting them using the aforementioned simulation technique causes the total number of in-sample lots predicted by our model (3,392) to fall significantly below the actual number of lots created during our observation period (4,577). The reason for this deficiency is that there is a limited number of very large subdivisions in the sample and, as result, there is a limited amount information on these types of subdivisions to estimate that portion of the Poisson distribution function.¹⁹

To remedy this deficiency and increase the predictive efficiency of our model for use in doing policy analysis, we use the aforementioned simulation technique to examine the ability of our model to match the actual number of lots created under three different scenarios for the second-stage Poisson (count) predictions. In the first scenario, which we term the “baseline”, we use the simulation technique described above and form the predicts (for the number of lots created in each iteration of the simulation) based only on the predictions from the Poisson model and the zoned lot capacity for the parcel, i.e., if the prediction from the Poisson model is above the maximum number of zoned lots allowed by zoning, then we use the zoned lot quantity instead of our prediction. The results for the baseline scenario are shown in Figure 5.2²⁰, which shows that the mean predicted value for our model falls over 1,000 lots below the actual number created

¹⁸In many areas of the U.S., variances are allowed that make zoning flexible. However, inspection of our data reveals that this is rare for our study region so, for the purposes of our simulations, we take zoning as given when the predicted value is larger than the zoned maximum number of lots for the parcel, which implies that zoning holds on all parcels during our simulations.

¹⁹During our observation period there were only 12 subdivisions (out of 397) that had buildable lots over 100. However, these 12 subdivisions accounted for 1,815 of the 4,577 lots developed during this period.

²⁰For comparison, each of the figures also shows the “actual” number of lots created in the county over our 13-year study period (True) and the number of lots that would be created in each simulation if we allowed each parcel to develop to its maximum zoned capacity (Zoned). That is, conditional on the development predictions from the Probit model, Zoned is created in each simulation by using the maximum zoned capacity on each parcel and not the predictions from the Poisson model.

(True), most of which can be attributed to our models inability to capture very large subdivisions.²¹ In the second scenario, we use the same simulation approach as in baseline scenario, but in this scenario, for parcels with a zoned lot capacity above the 95th percentile for all developable parcels in the dataset (29 lots available for development), we use the maximum number of lots allowed by zoning in all cases where the prediction from the model is less than or equal to the zoned number of lots. That is, if the prediction is below the maximum capacity, but capacity is above 29 lots, then we use the maximum capacity in place of the prediction. This scenario allows the predictions of the model to hold for the bottom 95% of the parcels, but forces predictions on larger parcels (in terms of allowable lots) to match the zoned quantity and accounts for very large subdivisions. The results from this scenario are shown in Figure 5.3. In this scenario we over-predict the number of lots created, but it is closer, in absolute value, to the actual amount created than in the baseline case. In the final scenario we increase the percentile cutoff to 99%, which is equivalent to zoning for 104 allowable lots in the data. The results from this simulation are shown in Figure 5.4 and produce the closest approximation to the data of each of the three scenarios. We use the final scenario as our “benchmark” case in conducting the policy simulations.

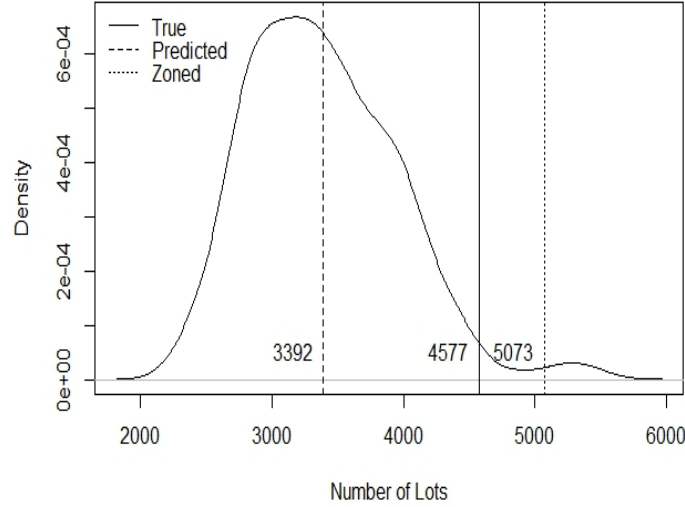


Figure 5.2: Prediction Scenario 1 (Baseline)

To analyze the impact of a purposeful change in implicit costs on development outcomes, we run four counterfactual policy simulations. The results for all of our policy simulations are shown in Table 5.4 along

²¹This deficiency is very similar to any models inability to predict extreme outliers.

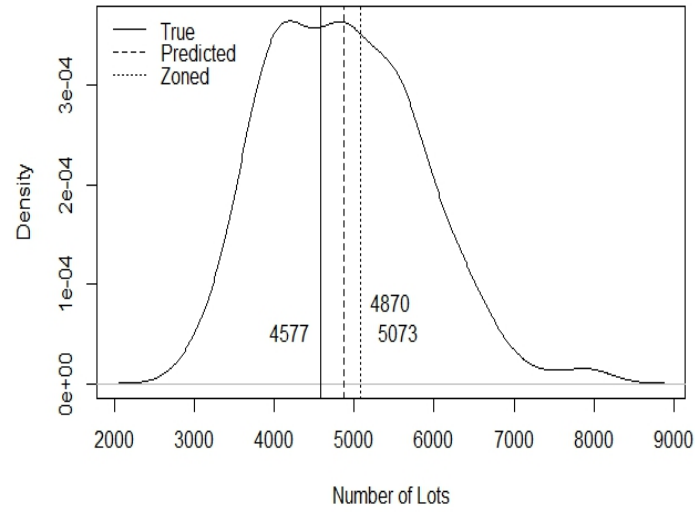


Figure 5.3: Prediction Scenario 2 (95th Percentile)

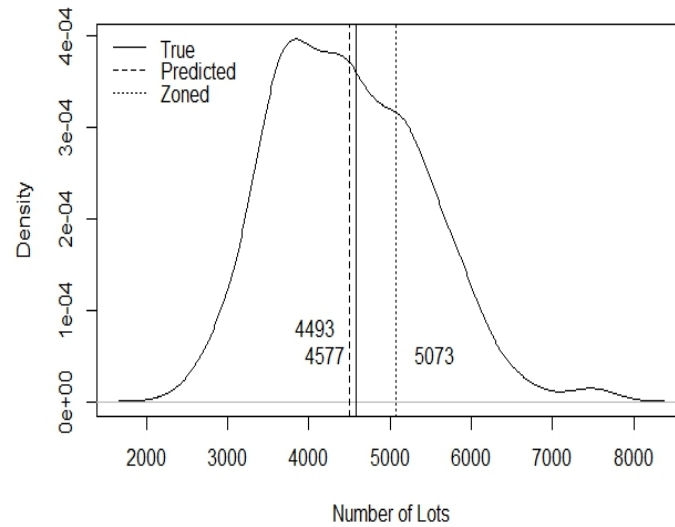


Figure 5.4: Prediction Scenario 3 (99th Percentile)

with the results from the benchmark scenario and the actual outcomes from the data during our study period. Each of these policy simulations provides analyzes of a different strategy that could potentially be taken by county planners in controlling the number (timing), size (density) and/or location (spatial) of the

developments in the county. The first policy simulation is a naive policy in which no distinction is made between the exurban versus suburban location of parcels and the expected subdivision approval time of *all* parcels is increased by 10%. The predicted number of subdivisions and number of lots created both decrease under this scenario. However, the predicted number of subdivisions falls by a smaller percentage for exurban (12.5%) versus suburban developments (18.2%). In our second simulation, the implicit costs are increased by 10% for only those parcels located in the exurban areas. In this case, the predicted number of subdivisions declines for only the parcels located in these exurban areas, but the number of lots created drops by less than in the first scenario in which both areas experienced an increase in implicit costs. In the third simulation, we increase by 10% the implicit costs associated with only the parcels located in suburban areas. Here, we find that the number of subdivisions drops by less than in the previous case, but that the reduction in the number of lots is almost double. In terms of lots per subdivision, these reductions translate into six lots per subdivision in exurban areas and almost 15 lots per subdivision in suburban areas. In the final simulation, implicit costs are increased by 10% for parcels located in the exurban areas and decreased by a similar amount in the suburban areas. This policy simulation produces a net loss of one development in total. However, by decreasing implicit costs in suburban areas we realize a net gain in the number of lots created by 291. Most notably, the result of this policy change is to substantially decrease the number of subdivisions predicted in the exurban areas and increase this number in the suburban areas thereby fostering a more concentrated development. Specifically, the predicted number of total lots increases by 6.5%, the predicted number of subdivisions in higher-density suburban areas increases by 21% and the predicted number of subdivisions in exurban areas decreases by 12%.

6 Conclusion

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. As uncertainty over the expected approval time for a development increases, the theory of optimal investment under uncertainty predicts that the option value of the real investment asset increases. Leapfrogging can result if the planning authority increases the restrictions

Table 5.4: Policy Simulation Results

	Predicted Subdivs	Predicted Lots	Exurban Subdivs	Suburban Subdivs
True Outcomes from the Data	397	4577	247	150
Benchmark Scenario (99th Percentile)	412	4493	264	148
(1) Cost Increase on All Parcels	353	3802	231	121
	(14.3)	(15.4)	(12.5)	(18.2)
(2) Cost Increase for Exurban Only	379	4293	231	147
	(8.0)	(4.5)	(12.5)	(0.0)
(3) Cost Increase for Suburban Only	386	4113	265	121
	(6.3)	(8.5)	(0.0)	(18.2)
(4) Cost Increase for Exurban/Decrease for Suburban	411	4784	232	179
	(0.0)	(6.5)	(12.1)	(20.9)

Note: The cost increases are for a 10% increase in implicit costs on each parcel.

Note: The true values are the actual results from the original data.

Note: The percent increase or decrease from the benchmark is shown parentheses.

on the development of parcels that are closer to the urban center, but leaves exurban parcels unchanged. In this case, the expected future net returns to suburban development is decreased, which increases the value of holding vacant suburban land while leaving unchanged the value of vacant exurban land. This increases the incentive to develop in exurban areas and leads to greater development in the exurban region. This latter result provides an alternative economic explanation of the leapfrog development patterns observed in many exurban areas.

Despite the theoretical implications of uncertain regulation, empirical evidence of how these effects influence residential land supply and development patterns is lacking. Using a unique panel dataset of subdivision approval times from an exurban county in Maryland, we provide new empirical evidence of the hypothesized effect of implicit costs on the timing, intensity and pattern of residential land development. The empirical results from our model show that regulation-induced implicit costs reduce both the probability and the quantity of development and that systematic correlation between implicit costs and subdivision size has resulted in an implicit cost advantage for smaller subdivision projects. We find that this small subdivision advantage, combined with zoning regulations that restrict the allowable size of subdivisions in the exurban areas of our study region, has resulted in a substantial increase in the likelihood of rural residential development relative to higher-density suburban development. Land development simulations under benchmark and alternative policy scenarios illustrate the potential for influencing residential development patterns by purposefully altering the implicit costs associated with the subdivision review process. We show that by changing the relative magnitudes of implicit costs across exurban and suburban areas, planners can maintain

the same number of developments, but foster greater concentration by increasing the number of lots created and reducing the amount of exurban subdivision development.

These results provide the first empirical evidence of the important role that heterogeneous costs and uncertain regulation play in influencing land development outcomes and leapfrog development patterns. The role of heterogeneity in generating discontinuous development patterns has long been emphasized in the theoretical literature (Mills, 1981; Wheaton, 1982; Newburn and Berck, 2011), but empirical evidence thus far has been lacking. Instead previous empirical studies have focused on the role of demand-side amenities and disamenities and the role of these local land use spillovers in generating scattered exurban land development (Irwin and Bockstael, 2002; Klaiber and Phaneuf, 2010; Walsh, 2007). Our results offer a new explanation of scattered, low-density residential development as the outcome of heterogeneous regulatory costs and optimal land development decision making and underscore the importance of supply-side factors that, to-date, have not been empirically tested. Given the importance of implicit costs to land developers, then any policy change that increases the implicit costs associated with development projects in higher-density areas can have the unintended effect of fueling this type of development. Conversely, the fact that spatially heterogeneous regulatory costs can influence leapfrog development suggests an explicit policy instrument that can be used to better manage land development. This final policy prescription is relevant not just for our study region, but for any urbanizing area that is experiencing increased exurban growth and scattered development as a result of systematic spatial differences in land use regulations.

References

- Bar-Ilan, A., and W. Strange. 1996. "Urban Development with Lags." *Journal of Urban Economics* 39:87–113.
- 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.
- Cameron, A., and P. Trivedi. 1998. *Regression Analysis of Count Data*. Cambridge University Press, Cambridge.
- 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.
- Englin, J., and T. Cameron. 1996. "Augmenting Travel Cost Models with Contingent Behavior Data." *Environmental and Resource Economics* 7:133–147.
- Geltner, D., T. Riddiough, and S. Stojanovic. 1996. "Insights on the Effect of Land Use Choice: The Perpetual Option on the Best of Two Underlying Assets." *Journal of Urban Economics* 39:20–50.
- Glaeser, E., J. Gyourko, and R. Saks. 2006. "Urban Growth and Housing Supply." *Journal of Economic Geography* 6:71–89, Cambridge, MA: National Bureau of Economic Research.
- . 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.
- Greene, W. 2011. *Econometric Analysis, 7th Edition*. Prentice-Hall, Inc., New York, New York.
- . 2005. "Functional Form and Heterogeneity in Models for Count Data." *Foundations and Trends in Econometrics* 1:113–218.
- . 1995. "Sample Selection in the Poisson Regression Model." Unpublished, Working Paper EC-95-06, Department of Economics, Stern School of Business, New York University.
- Gyourko, J., and A. Saiz. 2006. "Construction Costs and the Supply of Housing Structure." *Journal of Regional Science* 46:661–680.
- Irwin, E., and N. Bockstael. 2002. "Interacting Agents, Spatial Externalities and the Endogenous Evolution of Residential Land Use Patterns." *Journal of Economic Geography* 2:31–54.

- 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.
- Krinsky, I., and L. Robb. 1986. "On Approximating the Statistical Properties of Elasticities." *The Review of Economics and Statistics* 68:715–719.
- 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.
- Mayer, C., and C. Somerville. 2000a. "Land Use Regulation and New Construction." *Regional Science and Urban Economics* 30:639–662.
- . 2000b. "Residential Construction: Using the Urban Growth Model to Estimate Housing Supply." *Journal of Urban Economics* 48:85–109.
- McConnell, V., M. Walls, and E. Kopits. 2006. "Zoning, TDRs, and the Density of Development." *Journal of Urban Economics* 59:440–457.
- Mills, D. 1981. "Growth, Speculation, and Sprawl in a Monocentric City." *Journal of Urban Economics* 10:201–226.
- Newburn, D., and P. Berck. 2011. "Exurban Development." *Journal of Environmental Economics and Management*, pp. .
- . 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 Supply: Homeowners, Workers, and Voters." *Discussion Paper: London School of Economics*, pp. .
- Paciorek, A. 2011. "Essays on Housing Supply and House Price Volatility." Ph.d. dissertation, University of Pennsylvania.
- 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.
- Terza, J. 1998. "Estimating Count Data Models with Endogenous Switching: Sample Selection and Endogenous Treatment Effects." *Journal of Econometrics* 84:129–154.
- 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.
- Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge.
- 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.
- Zabel, J. 1992. "Estimating Fixed and Random Effects Models with Selectivity." *Economics Letters* 40:269–272.

7 Appendix

7.1 Study Region

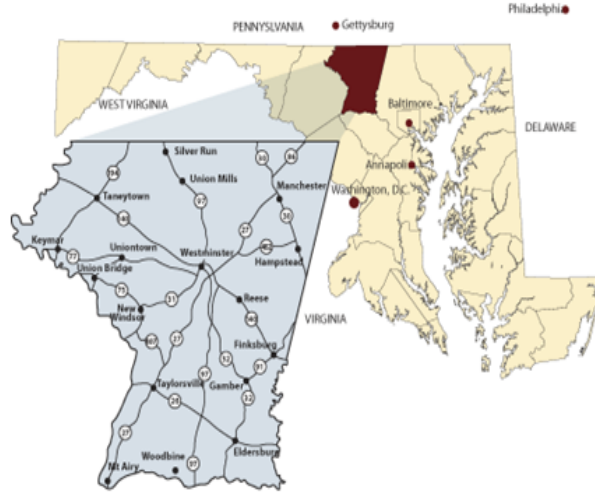


Figure 7.1: Baltimore/Washington, D.C. Metro Area

7.2 Data Construction

To estimate our econometric model, we constructed several micro panel datasets. The first dataset we constructed was a panel of the historical subdivision development for the county. To construct these data, 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.²² 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 could 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.²³ There were 1,910 subdivisions developed from 1924-2007. Of these, 1,098 were major developments and 812

²²After 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.

²³In 12% of the cases 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.

were minor developments.²⁴

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 (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 that they are preserved as opposed to removing all of these parcels at the beginning of the study period. In theory, we could model this process simultaneously with the development decision, but that is outside of the scope of this paper.

The final dataset we created, and one of the main contributions of this paper, was a dataset of the historical subdivision approval process for county. As was noted in the paper, 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 and pattern matching algorithm written in Python, 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 2010.²⁵

Figure 7.2 shows an example plat map and Figure 7.3 shows the land use pattern for the county at the end of 2007.

²⁴In 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.

²⁵We also collected a dataset of historical house price transactions over our study period. These data and their use in constructing our local measure of house price drift and volatility are described in appendix 7.3.

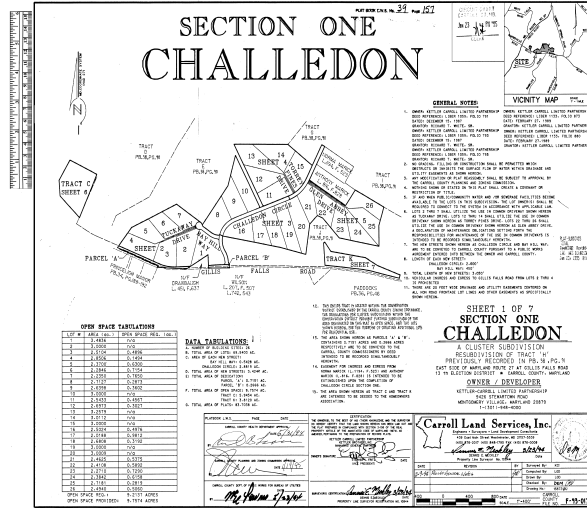


Figure 7.2: Plat Map Example

7.3 Proxy for Price Drift and Volatility

To construct our measure of house price drift and volatility, we use a hedonic technique. In a fully structural model, we would use estimates for the evolution of the asset value for each parcel. However, we lack adequate revenue data to construct these values. So, we proxy for the potential growth and volatility in 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 housing, this is insufficient to capture the quality-related differences over time and space. In this paper we adopt a similar approach to measuring these values as previous real options 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 an irreversible investment.

To construct these indices, we used historical data on house prices and property characteristics collected from the state tax assessor's 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

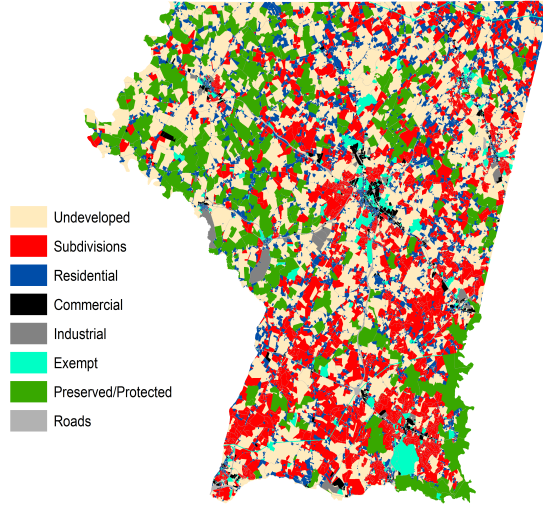


Figure 7.3: Carroll County Land Use 2007

800 or over 10000 to remove outliers. The final dataset contained 34,311 arms-length transactions. The descriptive statistics for the covariates used in each of the price index models are shown in Table 7.1.

Table 7.1: Summary Statistics: Price Index Data

Variables	Mean	Std. Dev.	Min.	Max.
Log Real Price	12.20	0.43	10.94	13.82
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
Travel Time to Baltimore City	37.01	6.85	22.72	64.49
Area	3.65	2.30	0.00	131.59
Square Footage	1799.21	736.66	800.00	9600.00
Age of Structure	18.67	24.77	1.00	9.99
Structure Quality	3.54	0.76	1.00	8.00
Townhouse	0.14	0.35	0.00	1.00
N=34,311				

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 our context as the entire county is under a single school district. Thus, we produced our estimates of drift and volatility at the census tract level and matched those values to each parcel in the dataset based on their census tract and for the years they are in the dataset.

Then, using our data on real house-price transactions, we estimate the following hedonic equation for each of the tracts and for each of our 13 time periods:

$$\ln(P_{it}) = \alpha_{jt} + \gamma_{jt}\tau_i + \beta'_{jt}X_i + \varepsilon_{it}, \quad (7.1)$$

where the dependent variable is the natural log of real house prices (in 2000 dollars) and the explanatory variables are the logs of the parcel and location characteristics of the parcel. 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 7.1. 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:

$$CV_{jt} = 100 * \frac{RMSE_{jt}}{\sum_{i=1}^n RP_{jit}}. \quad (7.2)$$

Given the number of time periods and census tracts in the county, this process produces too much information to display here. So, to give the reader some idea of how these values change over space and time we plot the mean values in each time period for both drift and volatility based on whether the parcel is located in an low-density exurban zoning area or not. These plots are shown in Figures 7.4 and 7.5. While the drift value is fairly similar for the two areas, the volatility measure is clearly higher for the exurban areas. This result stems from the fact that there are fewer sales in each period in these areas (thinner markets) producing a higher RMSE and CV statistic.

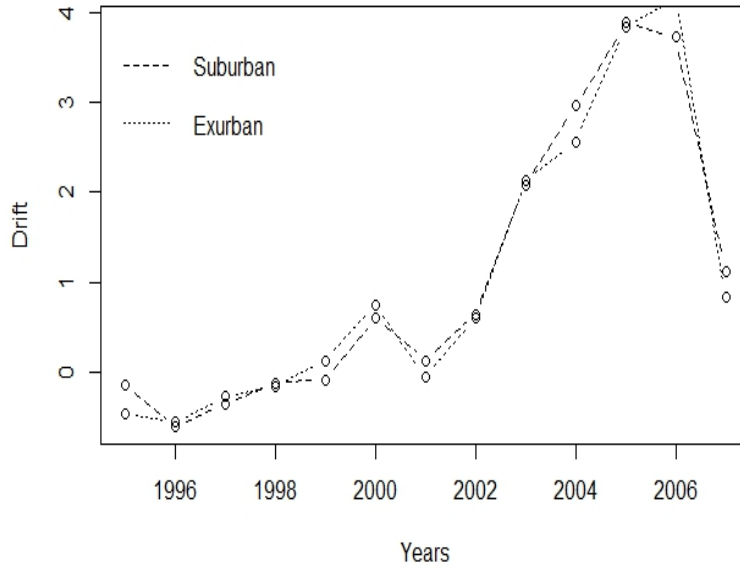


Figure 7.4: House Price Drift

7.4 Derivation of the Marginal Effects

To calculate the marginal effects for the Probit model, we begin with the predicted probability of development:

$$\Phi[d_s = 1|x_{nt}, \rho] = \Phi\left(\frac{(x'_s \alpha)}{\sqrt{1 - \rho^2}}\right), \quad (7.3)$$

where x_s is the value of the covariate for parcel n in period t ($S=n*t$) and α and ρ are the parameter estimates from the model. Setting equation 7.3 equal to $\Phi(x'_s \alpha, \rho) = \hat{\Phi}_s$, the average predicted probability is equal to:

$$\bar{\Phi} = \frac{1}{S} \sum_{s=1}^S \hat{\Phi}_s. \quad (7.4)$$

In order to evaluate the marginal change in the predicted probability for a change in one of the variables, we use equation 7.4 and estimate the discrete change in the predicted probability for a change in that variable. The marginal effects and standard errors for any nonlinear transformation model are not as simple as in the linear case. This is especially true in our particular model given the complexity of the likelihood function

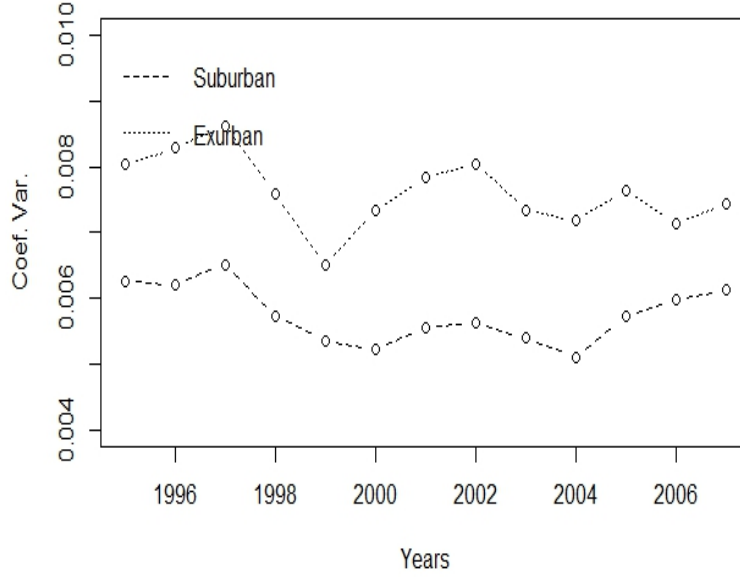


Figure 7.5: House Price Volatility

and introduction of truncation in Poisson outcome equation. Applying a discrete change methodology greatly simplifies the calculations needed for hypothesis testing and produces almost identical results and interpretations for small changes in the variables (Greene, 2011).²⁶

Again, setting 7.3 equal to $\hat{\Phi}_s$, the marginal effect for the selection equation is produced by taking the difference in this equation before and after the change as follows:

$$\Delta \hat{\Phi}_s = \hat{\Phi}_{s1} - \hat{\Phi}_{s0}, \quad (7.5)$$

where the 0 and 1 indicate the before and after time periods, respectively. In order to produce the standard errors, we take the derivatives of this equation with respect to the coefficients, θ , $\frac{\hat{\Phi}_s}{\partial \theta}$, and use the Delta method to produce the standard errors:

$$\text{Asy.Var.}(\Delta \hat{\Phi}_s) = \left(\frac{\hat{\Phi}_s}{\partial \theta} \right)^T V(\hat{\theta}) \left(\frac{\hat{\Phi}_s}{\partial \theta} \right), \quad (7.6)$$

where the first and last terms are the derivatives with respect the coefficients in the selection equation and

²⁶The alternative would be to calculation the continuous change effects, which require first and second derivatives.

the middle term is the robust covariance matrix of the parameters from the selection model. These standard errors are averaged in same way as the change in the predicted probabilities.

To calculate the marginal effect for the Poisson model for the same discrete change in one of the variables in the outcome equation, we begin with the following conditional mean equation:

$$E[q_s|x_s] = \int_{-\infty}^{\infty} \Phi\left(\frac{x'_s\alpha + \rho\varepsilon_n}{\sqrt{1-\rho^2}}\right) \frac{e^{x'_s\beta + \sigma\varepsilon_n}}{\text{Pois}(1)}, \quad (7.7)$$

where the first term in this equation is the adjustment controlling for the selection effect and the second term is the conditional mean for the truncated Poisson regression model. For variables present in both stages of the model, the first term captures the indirect effect of those variables and the second term captures the direct effect. In the case of a positive correlation coefficient, the marginal effects for both stages are summed to get the total effect. However, in our model we did not find a significant estimate for the correlation coefficient and thus can evaluate just the direct marginal effect of a change in the cost variable on the number of lots created.

Setting the second term in equation 7.7 equal to $F_s(\mathbf{x}'_s\boldsymbol{\beta}, \sigma) = \hat{F}_s$, the average value for the conditional mean is equal to:

$$\bar{F} = \frac{1}{S} \sum_{s=1}^N \hat{F}_s. \quad (7.8)$$

In a similar manner to the predicted probabilities, the marginal effect for a discrete change in one of the variables is produced by taking the difference in \hat{F}_s before and after the change:

$$\Delta\hat{F}_s = \hat{F}_{s1} - \hat{F}_{s0}, \quad (7.9)$$

and averaging across this difference using equation 7.8.

In order to produce the standard errors for these changes, we take the derivatives of this equation with respect to the coefficients in outcome equation, $\frac{\partial \hat{F}_s}{\partial(\boldsymbol{\theta})}$, and use the Delta method to generate the standard errors using equation 7.6.