**PRICE BASED POLICIES FOR MANAGING RESIDENTIAL LAND DEVELOPMENT: IMPACTS ON WATER QUALITY**


Douglas H. Wrenn

Department of Agricultural Economics, Sociology, and Education,

Pennsylvania State University,

112A Armsby Building

University Park, PA 16802

814-865-9216

814-865-3746

dhw121@psu.edu

H. Allen Klaiber

Department of Agricultural, Environmental, and Development Economics,

The Ohio State University,

klaiber.16@osu.edu

David A. Newburn

Department of Agriculture and Resource Economics,

University of Maryland

dnewburn@umd.edu


May 28, 2017

**PRICE BASED POLICIES FOR MANAGING RESIDENTIAL LAND DEVELOPMENT: IMPACTS ON WATER QUALITY**

**Abstract**: Urban land use plays a critical role in determining the health of urban ecosystems. In this paper, we link an instrumental variable land use model with a model of local nutrient pollution to analyze how price-based land use policies, designed to influence development outcomes, impact water quality. Our results demonstrate that these polices can be effective at limiting the impact growth on water quality outcomes. We also show that policymakers will be required to make tradeoffs between managing sprawl and limiting its impact on water quality, a counterintuitive result that stems from the interaction between the location of urban land conversion and baseline land cover.

*Keywords:* Land Use, Water Quality, Duration Models, Price Endogeneity

*JEL Codes:* Q24, Q53, Q56, Q58

# 1. INTRODUCTION

Land use plays a critical role in determining the health of urban ecosystems (Glaeser and Kahn, 2004; Hansen et al. 2005; Wu, 2008). This is particularly true for water-based systems where land use change has a significant impact on water quality. With the Environmental Protection Agency's (EPA) implementation of total maximum daily load (TMDL) restrictions in watersheds throughout the U.S., many localities have adopted policies designed to limit urban spatial expansion and encourage high-density development in an attempt to limit the impact of urban land use on water quality. To achieve these results, policymakers have increasingly turned to price and incentive-based policies (Bruecker, 2000). While these policies have been celebrated as market-based solutions for regulating development, research is still limited in assessing how specific policies, designed to influence the timing, density, and location of urban development, actually impact water quality outcomes. To explore this relationship, and evaluate the implications of different policies, it is necessary to couple economic land use models with models of water quality; it is also necessary to develop economic models that consistently estimate the price responsiveness of urban land conversion.

In this paper, we link an instrumental variable land use model with a watershed model of local nutrient pollution to examine how price-induced changes in residential development patterns impact nitrogen and sediment loading in the Baltimore metro region. The objectives of this study are to (1) econometrically identify the responsiveness of residential development to changes in housing prices; (2) combine the results from an econometric land use model with a model of water quality and develop a simulation framework suitable for analyzing price-based land use policy; and (3) use the simulation framework to analyze the most common real-world policies used to manage growth and water quality outcomes. A number of recent studies have combined econometric modeling and land use simulation to look at issues related to water use (Bigelow et al. 2017), open-space conservation (Newburn et al. 2006; Lewis et al. 2009; Lewis et al. 2011), forest conservation (Newburn and Ferris, 2017), property taxes and sprawl (Wrenn and Klaiber, 2017), and regulatory costs (Wrenn and Irwin, 2015). Our paper builds on this research and advances the literature by examining how incentive-based policies force policymakers to make tradeoffs between local water quality and the location and amount of residential development.

Urban land use in most areas is influenced by a combination of land use restrictions (e.g., urban growth boundaries and minimum-lot zoning) and incentive-based policy approaches (e.g., impact fees and taxes on residential development). In our Baltimore study region, land use regulations have been used extensively to manage growth. Historically, most of these regulations have been either minimum-lot zoning laws or urban growth boundaries. In 1997, Maryland established priority funding areas (PFAs), which combined urban-growth restrictions with an incentive-based tax policy – i.e., the state made infrastructure funding for each county contingent on the maintenance of PFAs (urban growth areas) designed to contain the urban footprint. More recently, the state implemented a price-based storm-water management fee designed to influence development and reduce the impact of land use change on water quality outcomes. Based this brief policy background it's clear that planners in Maryland are moving toward incentive-based (price-based) policies for managing both growth – a trend taking place in many other urban areas in the U.S. We contribute to this debate by examining several of the most common price-based policies used by planners in U.S. to determine their effectiveness at managing growth and limiting its impact on local water quality. Specifically, we are concerned with the types of tradeoffs that exist between policies designed to contain urban spatial expansion and those designed to limit development impacts on water quality outcomes.

To address these questions, we combine a unique data set on historical subdivision development with a parcel-level duration model capable of addressing price endogeneity. Our data are generated by combining detailed GIS shapefiles with subdivision plat maps to produce a complete history of residential development activity in the Baltimore region from 1994-2007. These data include information on the size and location of the original land parcel for each subdivision as well as information on the number of building lots created, zoning restrictions, land characteristics, and information on existing land cover. Our duration model accounts for changes in pre-existing land use following the creation of each residential development; it also handles price endogeneity to produce a consistent estimate of the price elasticity of land conversion. To instrument for price, we use a control function methodology appropriate for instrumentation in a nonlinear probability model with instruments based on demand-side drivers of housing consumption in distant neighborhoods (Petrin and Train, 2010; Klaiber and Kuminoff, 2014; Wrenn et al. 2017). Following estimation, we combine the parameters from our econometric model with our land use data and data on local loading rates for nitrogen and sediment taken from the Chesapeake Bay Program's (CBP) watershed model in a land use simulation to analyze three specific policies designed to limit growth and/or manage water quality.[1] While there are numerous land use policies that could be analyzed in the context of our study region, we focus on three policy scenarios which are both relevant to the Baltimore region and general enough to be used by planners in other urban areas.

In the first scenario (Scenario 1), we implement a uniform (one-size-fits-all) property tax on all parcels in the data. This scenario is designed to imitate the manner in which property taxes policies are implemented. In the second scenario (Scenario 2), we raise taxes on parcels in areas without public sewer and reduce taxes (implement a subsidy) on parcels in areas with public sewer.[2] This scenario is designed based on policies similar to Maryland's PFA legislation, which uses tax incentives and growth boundaries to concentrate growth in areas with public utility access. In our third scenario (Scenario 3), we implement a tax or subsidy based on a parcel's pre-existing land cover – i.e., based on the baseline loading-rate values for nitrogen and sediment from CBP model, we seek to limit the conversion of parcels with significant forest cover and encourage the conversion of parcels with significant cropland agriculture. Our specific policy is to implement a tax on parcels with more than 50% forest cover and provide a subsidy on parcels with greater than 50% cropland agriculture. This last scenario is focused on enhancing water quality outcomes and ecosystem services, as opposed to managing urban spatial expansion. We compare the results from this policy scenario, and the other two scenarios, to a baseline simulation where prices remain unchanged. In each scenario, we vary the tax and subsidy values from 2% to 10% in 2% increments.

The results from our land use simulations provide a number of important insights. First, we show that failure to account for price endogeneity produces biased and inconsistent elasticity values and leads to incorrect conclusions about the effectiveness of price-based policies. Using our preferred duration model, we find that the elasticity values in our IV model are 2.6 times larger than those in our non-IV model. This result suggests that not accounting for price endogeneity has important implications for determining the effectiveness of price-based policies. Second, we find that a uniform property tax (Scenario 1) leads to a large overall reduction in the number of residential developments created. Interestingly, however, this policy actually leads to an increase in average distance traveled to the center of Baltimore across all developments. Third, we find that a policy that attempts to limit urban spatial expansion and concentrate development in areas with

[1] The CBP's watershed model is the most policy relevant model in our study area as it is used by the EPA and all jurisdictions in the watershed to evaluate compliance with the Chesapeake Bay total maximum daily load (TMDL) requirements.
[2] We assume that the areas in our study region covered by public sewer and water are exogenously determined by land use planners and individual developers take them as given.

public sewer (Scenario 2) leads to a decrease in average travel distance and an increase in loading rates for nitrogen and sediment. This counterintuitive result stems from the fact that discouraging development in areas without public sewer, and pushing it toward areas with public sewer, leads to a reduction in the number of cropland farms parcels taken out of production and increases the number of high-density parcels with large amounts of impervious surface, both of which result in increased nitrogen and sediment loading. Finally, in Scenario 3, where we attempt to protect parcels with large amounts of forest cover, we find that nitrogen and sediment loading decrease, average travel distances increase, and the number of subdivisions created increases, albeit by only a small amount. This result stems from the fact that in Scenario 3 we replace cropland agriculture with low-density, low-impervious residential which has lower nitrogen and sediment loading per acre, compared to agriculture, and we protect forest cover, which reduces nitrogen and sediment loading. Hence, our simulations indicate that price-based policies which limit forest loss and remove cropland agriculture from production are the most effective for mitigating the water quality impacts from residential development in our study region. However, this policy appears to come at a cost as it increases the overall average travel distance from the city center for the development that is created which suggests an increase in urban spatial expansion.[3]

The rest of the paper is structured as follows. In the next section, we provide an overview of our study region and a description of our data and variables. In Section 3, we present our econometric model along with our control function approach to handling price endogeneity. Section 4 presents our main results, and Section 5 presents our land use simulations and policy analysis. Section 6 concludes.

## 2. DATA AND CONSTRUCTION OF VARIABLES

Our data cover the Baltimore, Maryland metro region for the years 1994 through 2007. Our study area, shown in Figure 1, covers the counties of Baltimore, Carroll and Hartford, which account for much of the residential development in this region, both in terms of land area and in terms of population during this time period.

To produce our subdivision data, we obtained GIS parcel data from each of the three counties, and manually matched individual housing units and other land parcels with scanned images

[3] A similar counterintuitive result was found in a recent paper analyzing urban land conversions and water usage in Oregon (Bigelow et al. 2017). In this paper, replacing cropland agriculture with low-density residential actually led to a reduction in overall consumptive water use under certain policy scenarios.

of the subdivision plat maps for the years 1994-2007.[4] The plat maps were obtained from the Maryland Department of Planning by direct download or by manual scanning. The matching process allowed us to group observations on individual housing units into their underlying subdivisions (parent parcels) as well as put them in the original land parcels and time periods (years) from which they were created. From this process, we also obtained information on the number of lots created and the original size of the raw land parcel.

In addition to the creation of our subdivision data, we obtained historical information on zoning boundaries and public sewer availability for each county. Using these zoning data, we derive the number of building lots permitted on each parcel of land, which in turn allow us to manually construct a micro-level land use data set that reflects the potential for development on each parcel and establishes a panel of parcels that are eligible for a residential subdivision development from 1994 through 2007. For our model, we define subdivision eligibility as any land parcel that can, according to zoning, accommodate a residential development of two or more building lots. Combining our subdivision data with our method of determining subdivision eligibility results in a data set of 15,015 parcels that were subdivision eligible as of 1994. Among these, 2,394 experienced a residential subdivision event during our study period. The balance remained undeveloped, or censored, at the end of 2007.

To develop an empirical strategy for handling price endogeneity, we begin by adopting the intuition underlying the urban location choice literature whereby households, or housing units, are nested within larger neighborhoods (Tiebout, 1956; Epple and Sieg, 1999; Bayer et al. 2007; Klaiber and Phaneuf, 2010). We assume that each of the 15,015 land parcels in our data are nested within one of 667 neighborhoods, where neighborhoods in our context are defined based on Census 2000 block-group boundaries. Figure 1 provides the basic intuition for this process. As with the urban location choice literature, we assume that factors varying at both the parcel and neighborhood level impact the probably of converting to a residential development in each time period.

Summary statistics for the variables used in our model are given in Table 1. The variables are separated based on their level of spatial variation. The top portion of the table lists the variables that control for parcel-level characteristics. First, to control for the locational attributes of the parcel, we include the distance, in kilometers, to the center of Baltimore City (Dist), which reflects accessibility to the largest employment center in the region. We also include the distance to the closest major

[4] Plat maps are a kind of architectural drawing, which reflect the location and boundaries of each subdivision as well as other important information about each development, such year of initiation, size, and number of building lots.

highway (DistMajRoad) as a measure of accessibility to transportation infrastructure. Both of these variables are expected to decrease the value of the parcel and its propensity to develop the farther the parcel is to the central business district (CBD) or major highway.

Zoning is also expected to play a role in determining the likelihood of conversion as the more densely zoned a parcel is the greater the number of building lots allowed. We obtained historic zoning maps for each of our counties from the Maryland archives and overlaid these maps with our parcel data. The variable ZndLots captures the zoned lot capacity for each parcel based on parcel size, zoning type, and maximum density regulations. While zoning did change slightly in Baltimore County during our study period, these changes were relatively small with most zoning in the region established in the mid-1970s. To account for the slight changes in zoning during out study period, we obtained historic zoning boundary maps for Baltimore County that enabled us to accurately calculate the zoned capacity for each parcel and each year in our data. We expect that parcels that have more development rights are likely more valuable and more likely to develop.

The final set of variables control for the physical features of the parcel. These include variables for the size of each parcel and soil quality characteristics derived from the SSURGO data provided by the Natural Resource Conservation Service (NRCS). We expect larger parcels are more likely to develop due to economies of scale. The NRCS soil classifications capture the hydrology, slope, percolation rate, and permeability of the soil. By combining these factors, we are able to determine variables for development suitability on each parcel. First, to proxy for the ability of a parcel to install residential septic systems and basements, we develop a septic suitability indicator (SepticSuit) based on soil permeability and percolation. We expect that parcels with a value of one will be more likely to develop as the soils on the parcel are more suitable for installation of septic systems and basements. Second, we use the slope classification for each parcel to develop an indicator (Slope) for whether the majority of the parcel has a slope of more than 15%. We also intersect our parcel data with maps for 100-year floodplains from the Federal Emergency Management Agency (FEMA), and create an indicator variable for whether or not the parcel is located in a floodplain zone (FloodPlain). We expect that parcels with steeper slopes or those located in floodplains are less likely to develop due to development limitations. Third, we use sewer boundary maps for each county and create an indicator variable (Sewer) for parcels with municipal sewer services. Finally, we include an indicator variable for whether the parcel has an existing structure (ExHouse) as well as variable for percentage of the parcel with forest cover (ForestPrcnt).

Our neighborhood variables comprise time-varying factors most likely to influence the probability of development in each neighborhood and year. First, we develop measures of the percentage of land in each neighborhood that is either preserved, or restricted from development (Preservation), and the percentage that remained undeveloped and developable. Each of these measures is obtained using detailed land use maps outlining the evolution of land development in our study region. Second, we capture competitive effects among developers using a one-year lag of subdivision activity (total lot approvals) at the neighborhood level (ApprvLots). And finally, we estimate neighborhood land (LandPrice) and housing (HousePrice) prices using auxiliary regressions. Both our land and house-price variables are created by estimating a series of yearly hedonic models on arms-length land and housing transactions obtained from a statewide database of sales provided in the Maryland Property View (MDPV) database. Each model includes, among other controls, a set of block-group fixed effects which serve as our quality adjusted neighborhood land and housing-price indices (Sieg et al. 2002). The details of the data and estimation procedures for our housing and land-price variables are given in the Appendix 1.[5] It is the housing-price variable produced from this process that is the focus of our model and our subsequent simulations.

## 3. ECONOMETRIC MODEL

Housing supply outcomes impacting ecosystem functioning and water quality are a result of many individual decisions made by landowners linked through market processes. As an individual landowner decides to convert previously undeveloped parcels to residential use, they consider not only attributes of their own parcel but also the likely prices and subsequent revenues this conversion with provide (Capozza and Helsley, 1989). Housing prices are spatially varying equilibrium outcomes arising from many individual transactions. Econometrically, the inclusion of prices in housing demand and supply estimation is challenging due to the potential for omitted variables to create endogeneity concerns. In nonlinear models, handling these types of endogenous attributes is a longstanding challenge.

Duration models are frequently used to capture the housing supply decision in a reduced form setting (Irwin and Bockstael, 2002; McConnell et al. 2006; Newburn and Berck, 2006; Cunningham, 2007; Lewis et al. 2009). Operating at parcel levels, these models easily accommodate micro-level land use features that vary spatially as well as temporally, such as prices. We extend the standard duration modeling framework by combining a nonlinear instrumental variable technique to

[5] This method of estimating prices is commonly used in the urban demand and supply literature (Walsh, 2007).

instrument for price using a control function methodology (Rivers and Vuong, 1988; Papke and Wooldridge, 2008; Petrin and Train, 2010; Wrenn et al. 2017). The instruments used in this control function model are formed from equilibrium relationships – a method commonly used in structural models of housing demand (Bayer et al. 2007; Walsh, 2007; Klaiber and Phaneuf, 2010).

The duration model we develop assumes that in each period $t$ the owner of an undeveloped parcel $i$ located in neighborhood $j$ decides whether or not to convert his parcel to a residential subdivision. Conversion decisions depend on parcel level attributes, $I_{it}$ and neighborhood level attributes, $X_{jt}$, where we define neighborhoods as Census 2000 block groups. We use discrete annual time-steps to define the time dimension of our duration model consistent with much of the existing housing supply literature. We can capture a reduced-form latent profit function underlying the duration model as

$$\Pi_{it}^{*} = I_{it}'\beta + X_{jt}'\alpha + P_{jt}'\gamma + u_{it} \tag{1}$$

where $\Pi_{it}^{*}$ is the latent profitability on parcel $i$, $I_{it}$ and $X_{jt}$ are parcel and neighborhood characteristics affecting profitability, respectively, $P_{jt}$ is the price of housing services at the neighborhood level, and $u_{it}$ is an idiosyncratic error term. Based on equation (1), the parametric proportional hazard we adopt is

$$h(t) = h_0(t)h(I_{it}'\beta + X_{jt}'\alpha + P_{jt}'\gamma) \tag{2}$$

where $h_0(t)$ is the baseline hazard, which is shifted proportionally by changes in the variables in the model. To empirically implement the model in equation (2), we use the discrete-time specification presented in Beck et al. (1998). We use the discrete-time duration framework in this paper as our subdivision events are only observed at yearly time steps.

We adopt a control function approach to address price endogeneity. This approach uses a two-step estimation procedure to instrument for endogenous variables using residual variation derived from a first-stage regression that includes exclusionary instrumental variables, $Z_{jt}$ that control for the correlation between the price and the error term. Based on this method, we can write our first-stage OLS regression as

$$P_{jt} = X_{jt}'\beta + Z_{jt}'\delta + v_{jt} \tag{3}$$

where the exogenous neighborhood variables, $Z_{jt}$ are a set of excluded variables that affect price, but not latent profit $\Pi^*_{it}$, and $v_{jt}$ is an idiosyncratic error term.[6]

In the presence of endogeneity, the error term from equation (1) is given as

$$u_{it} = v'_{jt}\theta + e_{it} \tag{4}$$

Assuming joint normality between $u_{it}$ and $v_{jt}$, the residual vector, $\hat{v}_{jt}$ from the first stage is added to the second-stage duration model as an additional covariate. Assuming that the instruments in the first stage are valid, we rewrite equation (1) as

$$\Pi^*_{it} = I'_{it}\beta + X'_{jt}\alpha + P'_{jt}\gamma + v'_{jt}\theta + e_{it} \tag{5}$$

where assuming joint normality between the errors in both stages, results in a discrete-time duration model as

$$P\left(\Pi^*_{it} = 1 \middle| I_{it}, X_{jt}, P_{jt}, v_{jt}\right) = \Phi\left[\frac{I'_{it}\beta + X'_{jt}\alpha + P'_{jt}\gamma + v'_{jt}\theta + \tau_{t-t_0}}{\sqrt{1-\rho^2}}\right] \tag{6}$$

where a set of time fixed effects $\tau_{t-t_0}$ are included to model the baseline hazard.

To estimate the two-stage control function approach, it is necessary to obtain instruments. The price instruments we construct borrow from the structural urban demand models of location choice. These models use the logic of Nash equilibrium to form instruments in a residential sorting context. The primary insight from this literature is that distant attributes impact prices in focal neighborhoods through spatial equilibrium. By isolating these distant, exogenous attributes it is possible to form instruments that are uncorrelated with unobservables in the focal neighborhood. We adopt this same methodology in the Baltimore metro region.

To highlight the intuition of this process, Figure 1 shows a map of our study region displaying both county and neighborhood (2000 census block group) boundaries. This figure depicts a single focal neighborhood with a seven-mile distance ring drawn around a given focal neighborhood. Our instrumentation strategy uses variation in exogenous attributes from neighborhoods located outside the distance boundary defined by this ring as a means of controlling for price endogeneity. To establish the extent of "local" vs "distant" neighborhoods, we use a series of statistical tests to examine the validity of our distance thresholds. To do this, we exclude an increasing number of local neighborhoods around each focal neighborhood and create an IV matrix, $Z^n_{jt}$ that consists of area-weighted average values of exogenous attributes from neighborhoods

[6] As is the case in a standard 2SLS IV model, identification depends on having at least as many excluded variables in the first stage ($Z_{jt}$) as there are endogenous regressors in the main model.

outside of the boundaries defined by these local neighborhoods, where the superscript $n$ indexes the distance cutoff used in forming the $Z$ matrix. We add these area-weighted instrumental variables to the right-hand side of equation (3) and estimate

$$P_{jt} = X_{jt}'\beta + Z_{jt}^{n}\delta + v_{jt} \tag{7}$$

For each of these models, we use over identification tests and choose the optimal model (optimal distance cutoff used in forming our instruments) based on the Chi-squared values from these tests (Stock et al. 2002; Wooldridge, 2010). By increasing the distance used in forming our instruments, we are able to net out local sources of variation, while retaining the power of the instrument, a result that is predicted by urban spatial theory (Bayer and Timmins, 2007).

## 4. ECONOMETRIC RESULTS

### 4.1 Control Function Estimation

The control function residual in our duration model is obtained by estimating a pooled OLS regression in the first stage based on equation (7). This first-stage regression includes all exogenous neighborhood characteristics, time and county fixed effects, and a set of instrumental variables based on average values of exogenous attributes from distant neighborhoods. The excluded instruments used include the percentage of land in each neighborhood that is preserved or restricted from development (PreservationAvg), the percentage of land that remains undeveloped (UndevelopedAvg), and a measure of neighborhood income obtained from mortgage data collected through the Home Mortgage Disclosure Act (HMDA) (HHIncomeAvg). Each of these variables serves a role similar to a standard a demand-side exclusionary restriction – i.e., we expect each to impact the location choices of households, but not directly impact developer decision-making. Using equation (7), we generate neighborhood-level residuals based on different distance bands and include them as additional variables in our duration model. To account for any nonlinear impacts associated with our control function, we also include a quadratic term for the residual in the model (Papke and Wooldridge, 2008).

Intuition for the exogeneity of our instruments is derived from the urban sorting and IO literature (Bayer et al. 2007; Petrin and Train, 2010). This literature establishes that variation in the instruments derived from distant locations influence prices in a focal neighborhood due to spatial equilibrium linkages arising from household sorting. Identification uses both spatial and temporal variation in excluded variables as they change over time and space. For robustness, we estimate models based on omission of local neighborhoods falling within the distance cutoffs of five to seven miles.

We report results from our control function estimates in Table 2. Based on our IV strategy, using more distant neighborhoods to generate instrumental variables in equation (7) should isolate plausibly exogenous competitive effects on price, while reducing the potential for direct price impacts from local neighborhoods. The remaining residual variation in the first-stage models introduced by our instrumental variables, after removing local effects, should serve as a plausible control function for endogeneity in competing risks model.

The first column in Table 2 indicates the distance bands used to form the instrumental variables in the first-stage OLS models. In column (2), we present results for a set of joint hypothesis tests that the coefficients on our IVs are equal to zero in the first stage. Based on these $F$-statistics, it is clear that our instruments pass these exclusion tests for each of the distance bands in the duration models; $F$ statistics easily exceed the rule-of-thumb exclusion threshold of 10 needed for inference based on the 2SLS estimator (Stock et al. 2002).

We also conduct overidentification tests based on the methodology described in Wooldridge (2010) that is similar to a Hansen's $J$ overidentification statistic used in GMM models. Specifically, in addition to the residuals generated from the first-stage OLS model we add two of the three excluded IVs to the right-hand side of the competing risks model – we use the variable on percentage of preserved land area as our excluded instrument – and perform a series of Chi-squared joint hypothesis tests.

Column (3) in Table 2 presents the results for our overidentification tests. For each model, the results are based on nonparametric block bootstrapped standard errors (300 replications) with clustering (bootstrapped samples drawn) at the parcel level.[7] The overidentification tests show that we reject the null hypothesis that the excluded variables are correlated with the error terms in each of the models. Furthermore, as we continue to exclude additional neighborhoods in the first-stage regressions based on our distance cutoffs the $p$-values for these tests rise and become less significant. This result reflects the spatial equilibrium nature of our IV strategy – as more distant neighborhoods are used in constructing our instruments, the competitive impact of distant attributes

[7] Because the residuals used as controls in the competing risks model are generated regressors – i.e., they are generated within a separate auxiliary model and added to the second stage – it is necessary to bootstrap the standard errors in order to obtain a consistent estimate of the variance-covariance matrix and standard errors (Wooldridge, 2010). Because our data is structured as a panel, we bootstrap the standard errors by sampling based on the parcel IDs.

on housing prices in a focal neighborhood are purged of local sources of variation that may directly impact prices in a focal neighborhood.

Lastly, column (4) of Table 2 presents the control function residual, which has the expected negative sign. This indicates that failure to control for endogeneity of prices is likely to lead to an understated price elasticity of land conversion.[8]

4.2 Duration Estimation

Full results for both the non-IV and IV duration models are shown in Table 3. We separate the results into parcel and neighborhood characteristics. Standard errors are, once again, based on a nonparametric panel bootstrap procedure. Examining these results, we find similar signs and significance in models with and without instrumentation and the estimates have the expected signs. We now provide a brief review of the main findings in Table 3 followed by an analysis of the coefficients on price in each model.

For the parcel characteristics, the coefficient estimate on zoning (ZndLots), which accounts for how many lots can be developed on a particular parcel based on the zoning designation of that parcel, is positive and significant in both models, which suggests that zoned capacity play a key role in determining development. Further, we see that the coefficient on land area (Area) is also positive and significant in both model; the coefficient of distance is positive, but insignificant in both models; and the coefficient for sewer is negative and significant in both models. Together, these results suggest the availability of sewer and parcel size play a significant role development decisions.

For our neighborhood characteristics, we find that an increase in the number of prior approvals at the neighborhood level has a positive impact, but only in the non-IV model. The coefficient on preservation is negative and significant, but only for the IV model. Finally, the coefficient on land price is negative and significant in both models, which is as expected if land serves as an input in the production of housing.

Turning attention to the coefficients on the housing-price variable, we first see that they are of the expected sign, positive, and significant in both the non-IV and IV models. Next, we see that the coefficients for the price residuals are negative and significant in all models indicating a

[8] Table A3 in Appendix 2 reports results from our first-stage OLS models for each type of development. One indicator for the plausibility of our instruments is whether they have a differential, and potentially opposite effect, on price from those same variables in focal neighborhoods. Similar to Bayer et al. (2007), we find that the neighborhood level of preserved area in a focal neighborhood has a differential and opposite effect from the same characteristic based on values from distant neighborhoods, which suggests that our instrumental variables have a competitive effect on price, but do not directly affect latent profit in the focal neighborhood.

downward bias in price in models without instrumentation. Finally, we find that the price coefficients in the IV models are substantially larger in magnitude than those in the non-IV models, which suggests that not accounting for price endogeneity may significantly underestimate the impact of housing prices on the probably of conversion. To provide some intuition for both the sign and relative magnitude of our housing price coefficients, we convert these coefficients into elasticities in Table 4.[9] Examining these values, we find that the price elasticity in the IV model 2.6 times larger than in the non-IV model.

The results in Tables 3 and 4 demonstrate that accounting for price endogeneity is important. This result also has important policy implications in that it suggests that planners and policymakers interested in regulating development may be able to use targeted, price-based (market-based) policies in achieving their objectives – i.e., targeted taxes based on location may provide planners with ability develop more focused land use policies. We explore these ideas further in the next section using the results from our duration model in a series land use simulations using the bootstrapped parameters estimates from our IV duration model. Based on the result in this section, we use the IV duration results in the remainder of the paper.

## 5. POLICY SIMULATIONS AND WATER QUALITY OUTCOMES

In this section, we present results from a series of land use simulations where we combine the parameter estimates and data from our duration model with data on nutrient loading rates from the CBP's watershed model to examine how different price-based land use policies (taxes and subsidies), which alter development patterns, impact water quality outcomes. Specifically, we focus on how three policy scenarios, which differentially influence the location of residential development, impact loading rates for nitrogen and sediment. For each scenario, we compare simulation results based on five different tax levels with baseline results where we make no change to price. For each simulation, we employ a nested looping structure where the outer loop is over the parameter distribution from the duration model and the inner loop is over time (years). Each of the simulations proceeds as follows.

In the first step, we use the data from our duration model and replace all of the parcels in all periods that dropped out as a result of development – i.e., we effectively produce a balanced panel data set, where each parcel is assumed to be available for development in each period. Second, we

[9] The point estimates are the average elasticity values calculated at each point in the sample, and they represent the "price elasticity of residential land conversion" and reflect the long-run price elasticity of land supply in our three-county metro region for both low and high-density subdivision developments (Wooldridge, 2010).

begin the iteration of the outer loop by taking a draw from the parameter distribution from our econometric model and combine it with our land use data from the first step and a probit link function from the duration model to form a predicted probability of development for each parcel. The parameter draws for this steps are taken from the rows of the matrix of bootstrapped parameters – i.e., 1,000 $kx1$ parameter vectors from the bootstrap matrix. Next, we take a random draw from a uniform distribution for each parcel and compare it with the predicted probabilities produced in the previous step. We assume that a parcel is developed if the predicted probability of development for that parcel is larger than the draw from the uniform distribution. To replicate the dynamic nature of the development process in our original data, these probability comparisons take place sequentially – i.e., for each iteration of the outer loop (for each draw from the bootstrapped parameter matrix), we iterate (step) through time, make comparisons in each time period, and drop parcels from all subsequent time periods once they are assumed to be developed, which effectively replicates the terminal nature of the development process. Finally, we repeat this sequential inner-loop 1,000 times for each parameter vector (row) of our bootstrapped parameter matrix. Each set of 1,000 iterations, and the parcels predicted to develop within each iteration, represent a single land use simulation – i.e., 1,000 different sets of parcel IDs representing the predicted development outcomes for each iteration. We use the aforementioned simulation process for our baseline simulation – without a change in price – and for each policy scenario (tax and subsidy level).

Following each simulation, we combine the results, which are the parcels predicted to develop within each iteration, with data on existing land cover, loading rates for nitrogen and sediment, and development density to generate measures for the change in loading rates following the implementation of a given land use policy. We assume our data on loading rates, taken from the CBP's model, proxy for changes in water quality following this policy-induced change in land use. The details of this process are as follows.

First, we take the GIS data used to estimate our duration model – the 15,015 parcels that were developable as of 1994 – and overlay them with 1992 USGS land cover data for the Baltimore region. This process allows us to determine the acreage of forest and cropland agriculture cover on each parcel at the beginning of our study period. Next, we use the GBP's data on loading rates to determine the baseline loads on each parcel for nitrogen and sediment based on each parcel's existing land cover. Specifically, we combine the loading rates in Table 5 for each county with the total acreage of each type of land cover (forest and agriculture) on each parcel to determine baseline loads for nitrogen and sediment. Lastly, we use the parcel IDs for the predicted subdivision developments (produced within each iteration of the simulation) and generate another a set of loading rates based on the predicted density of development for each subdivision, where the loading rates for different densities are based on the values shown in the bottom of Table 5. Our duration model, while effective at modeling the optimal timing decision, does not explicitly model density. So, to assign a density to each simulated development, we use density values from our actual subdivision data. Specifically, we use our data on actual subdivision activity and generate average density values for in each year and county and for areas with and without public sewer; we use these average density values to assign development densities to each predicted subdivision development based the timing and location of the development. It is the comparison of the baseline loading rates on each parcel, based on forest and agriculture land cover, with the loading rates based on development density that we use in assessing the water-quality implications of each of our three policy scenarios. We now describe these policy scenarios in more detail.

There are an endless number of price-based land use policies that we could analyze in our study region. However, our main interest is in running simulations that allow us to draw generalizable conclusions about the impacts of policies that have, or are most likely to be, implemented in the real world. Thus, we focus on three policy scenarios which we believe are both relevant to the Baltimore region and are general enough to be used in land use planning in other areas of the U.S.

In Scenario 1, we implement a uniform property tax on all parcels in the data. This scenario is patterned after the manner in which most residential property taxes are implemented in the U.S. In Scenario 2, we increase property taxes on all parcels located in regions without public sewer and subsidize development on all parcels located in areas with public sewer. This scenario is designed to mimic the methods used by land use planners to control urban spatial expansion and sprawl – i.e., to concentrate growth in areas with public utility access. In Scenario 3, we implement a tax and subsidy policy based on the pre-existing land cover on each parcel and its ability to impact loading rates for nitrogen and sediment and thus local water quality – i.e., based on the loading rate values for nitrogen and sediment from CBP model (Table 5), we seek to limit the conversion of parcels with large amounts of forest cover and encourage the conversion of parcels with large amounts of cropland agriculture given the benefit that forest cover provides in managing runoff. Our specific policy is to implement a tax on parcels with greater than 50% forest cover and provide a subsidy to parcels with greater than 50% cropland agriculture. This last scenario is focused on enhancing water quality outcomes and ecosystem services, as opposed to managing urban spatial expansion. In all three scenarios, we make comparisons between the simulation results following our taxed-induced price changes and results for a baseline simulation where we do not alter prices. In each scenario, we vary tax and subsidy values from 2% to 10% in 2% increments.

Before turning to the results from our policy simulations, it is important to assess how well our simulation results match our actual data. To make this comparison, we run our baseline simulation – without a price change – and compare the distribution for the number of developments created with mean amount from the actual data (2,394 residential subdivision created from 1994 through 2007). The distribution of developments created from this baseline simulation is shown in Table 6. As is apparent from these results, the mean value for the number of development created across all 1,000 iterations of our baseline simulation almost perfectly matches the actual mean from the data. Thus, we can are confident that our model and simulation fit the actual data very well. We now present the results from our policy simulations.

Our simulation results are presented in Table 7 and Figures 2 through 4. For each scenario, we present results for the number of developments created, the travel distance, in kilometers, from each development to the center of Baltimore City, and loading rates for nitrogen and sediment. For the development counts, we compare baseline totals for each iteration of the simulation with totals for each iteration after a policy change. For distance, we compare averages across predicted developments for each iteration of the baseline simulation with similar values following a policy change. Finally, for nitrogen and sediment, we generate average for changes in loading rates for each iteration of the baseline simulation and compare those values to each iteration of our policy simulations. The results in Table 7 are for the mean differences between the baseline simulation and each of the three policy simulations. Figures 2 through 4 present box-and-whisker plots for the distribution of these differences across all simulation iterations. Table 7 and three figures present results for each of the five different tax and subsidy levels listed above.

Our policy simulations produce a number of interesting results. First, we find that the number of developments created is reduced in both Scenarios 1 and 2 and this reduction increases tax and subsidy rates increases. Conversely, in Scenario 3 we find a very limited impact from the policy on the number of developments created with developments actually increasing ever so slightly as taxes and subsidies increase. In terms of distance, we observe average increases for Scenarios 1 and 3, but an average decrease for Scenario 2. This result is intuitive when we consider that both Scenario 1 and 3 favor lower-density, rural parcels, and Scenario 2 focuses on concentrating development in areas with public sewer and water with much of this land located closer to Baltimore

City. Finally, we observe a decrease in average loading rates for nitrogen and sediment for both Scenarios 1 and 3 and an increase in Scenario 2 with both increasing in absolute value as tax and subsidy rates increases. In addition, we observe the largest decrease in loading rates across all three scenarios is for Scenario 3, where, for example, a 2% tax-and-subsidy policy leads to an average reduction of 24 pounds per acre per year for nitrogen and an average reduction of 132 pounds per acre per year for sediment.

Based on the results Table 7, it appears the policy in Scenario 3, which targets the preservation of forest cover and conversion of agriculture to residential development, produces the greatest benefit, in terms of reductions in nitrogen and sediment loading rates; it also leads to the smallest impact in terms lost development. Alternatively, we observe that in Scenario 2, which implements a tax and subsidy policy designed to limit urban spatial expansion and concentrate growth in areas with public sewer, that we actually have an increase in average loading rates for both nitrogen and sediment. This counterintuitive result stems from the fact that discouraging development in areas without sewer, and pushing it toward areas with sewer, leads to a reduction in the number of cropland farm parcels taken out of production and increases the number of high-density parcels with large amounts of impervious surface, both of which result in increased nitrogen and sediment loading. Conversely, in Scenario 3, where we are replace cropland agriculture with low-density, low-impervious residential development and protect forest cover, we get a significant reduction in nitrogen and sediment loading given the benefit provided by maintaining and protecting forest cover. Hence, our simulations indicate that price-based policies, which limit forest loss and remove cropland agriculture from production, are the most effective for mitigating the water quality impacts from residential development, at least in our simulations and study region. However, this benefit come at a cost as this policy (Scenario 3) leads to an increase the average travel distance, which suggests that it may also lead to an increase in urban spatial expansion.

## 6. DISCUSSION AND CONCLUSIONS

In this paper, we estimate an instrumental variable duration model of residential subdivision development in large metro region and link the econometric results with a model of water quality to analyze how different price-based land use policies, which differentially impact development patterns, impact loading rates for nitrogen and sediment. The linking process is achieved by running a series of land use simulations where we compare results from a baseline simulation in which we leave price unchanged with results from simulations based on three specific policy scenarios which tax or subsidize land parcels based on location and land cover. The results from this process produce another of important results.

First, we find that traditional, uniform property taxes lead to significant reductions in development and very little reduction in nitrogen and sediment loading rates. Thus, to the extent that uniform taxes may be proposed as a method to limit the water quality impacts of urban development, our results suggest otherwise. Second, we find that standard urban-growth policy, which attempts to limit urban sprawl and concentrate development in areas with public sewer and water, actually lead to increases in nitrogen and sediment loading relative to policies that specifically focus on reducing land use impacts on water quality outcomes. This result stems from the fact that they limit the amount of cropland agriculture taken out of production – a land use with significant impacts in terms of nutrient loading rates – and replace it with high-density, high-impervious development which also contributes relatively more to nutrient loading. Finally, we find that price-based policies which limit forest loss and encourage the conversion of cropland agricultural parcels to low-density residential actually lead to significant reductions in nitrogen and sediment loading. The downside of this policy, however, is that may lead to an increase in the urban footprint and an increase in sprawl.

The main result in this paper – that land use policies designed to manage or limit urban growth may lead to conflicting environmental outcomes – is not entirely original. A number of recent papers have examined the impact of land use restrictions on environmental outcomes in different contexts and found corresponding results. Glaeser and Kahn (2010) explore the impact that inter-urban heterogeneity in the restrictiveness of land use regulations have on emissions. Their results indicate that increased land use regulations, which lead to high housing prices in those cities, may lead to increased population growth in less-regulated, lower-priced cities. If these less-restrictive cities are also relatively warmer and produce energy using dirtier methods, then it is possible for land use regulations designed to reduce emissions in one urban area to lead to an overall increase in emissions.

In a more recent paper similar in spirit to ours, Bigelow et al. (2017) use micro-scale econometric simulation methods to analyze how different land use policies impact consumptive water in three cities in Oregon. In one simulation scenario in their paper, increased urban expansion and residential development actually leads to a reduction in overall consumptive water use. That result stems from the fact that the increased residential development removes irrigated agricultural land from production, which uses relatively more water. This results is largely similar to ours – i.e., we demonstrate that removing agricultural land from production, and replacing it with residential development, leads to a relative reduction in nitrogen and sediment loading albeit on land developed farther from the urban center.

In conclusion, the results in our paper, and the papers listed above, indicate that there are a clear tradeoffs to be made by policymakers. Specifically, policymakers must recognize that reducing sprawl and urban spatial expansion may reduce certain environmental impacts – leapfrogging and fragmentation of the landscape – while increasing others – consumptive water use and local water pollution – based on interactions between existing land uses and new residential land use. Whether planners should target one type of environmental objective over the other is open question, which requires a more thorough welfare analysis – which we leave for future work.

## ACKNOWLEDGEMENTS

This research is supported by a cooperative agreement with the U.S. Forest Service Northern Research Station and National Science Foundation grants CBET 1160961, DEB 0410336, and GSS 1127044 and from support from the James S. McDonnell Foundation.

REFERENCES


Bayer, P., Timmins, C. 2007. Estimating equilibrium models of sorting across locations. *The Economic Journal* 117: 353-374.

Bayer, P., Ferreira, F., McMillan, R. 2007. A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy* 115: 588-638.

Beck, N., Katz, J.N., Tucker, R. 1998. Taking time seriously: time-series-cross-section analysis with a binary dependent variable. *American Journal of Political Scienc*e 42: 1260-1288.

Bigelow, D.P., Plantinga, A.J., Lewis, D.J., Langpap, C. 2017. How does urbanization affect water withdrawls? Insights from an econometric-based landscape simulation. Forthcoming: *Land Economics.*

Brueckner, J.K. 2000. Urban sprawl: diagnosis and remedies. *International Regional Science Review* 23: 160-171.

Capozza, D.R., Helsley, R.W. 1989. The fundamentals of land prices and urban growth. *Journal of Urban Economics* 26: 295-306.

Cunningham, C.R. 2007. Growth controls, real options, and land development. *The Review of Economics and Statistiscs* 89: 343-358.

Epple, D., Sieg, H. 1999. Estimating equilibrium models of local jurisdictions. *The Journal of Political Economy* 107: 645-681.

Glaeser, E.L., Kahn, M.E. 2004. Sprawl and urban growth. In Henderson, V., J. Thisse, J. Edition, vol. 4, *Handbook of Regional and Urban Economics*. 2481-2527. North-Holland.

Gyourko, J. 2009. Housing supply. *Annual Review of Economics* 1: 295-318.

Hansen, A.J., Knight, R.L., Marzluff, J.M., Powell, S., Brown, K., Gude, P.H., Jones, K. 2005. Effects of exurban development on biodiversity: patterns, mechanisms, and research needs. *Ecological Applications* 15: 1893-1905.

Irwin, E.G., Bockstael, N. 2002. Interacting agents, spatial externalities and the endogenous evolution of residential land use patterns. *Journal of Economic Geography* 2: 31-54.

Klaiber, H.A., Kuminoff, N. 2014. Equilibrium sorting models of land use and residential choice. In J.M. Duke and J.J. Wu Edition, *Oxford Handbook of Land Economics*. Oxford University Press.

Klaiber, H.A., Phaneuf, D.J. 2010. Valuing open space in a residential sorting model of the twin cities. *Journal of Environmental Economics and Management* 60: 57-77.

Lewis, D.J., Provencher, B., Butsic,V. 2009. The dynamic effects of open-space conservation policies on residential development density. *Journal of Environmental Economics and Management* 57: 239-252.

Lewis, D.J., Plantinga, A.J., Nelson, E., Polasky, S. 2011. The efficiency of voluntary incentive policies for preventing biodiversity loss. *Resource and Energy Economics*. 33: 192-211

McConnell, V., Walls, M., Kopits, E. 2006. Zoning, TDRs, and the density of development. *Journal of Urban Economics* 59: 440-457.

Newburn, D.A., Berck, P. 2006. Modeling suburban and rural-residential development beyond the urban fringe. *Land Economics* 82: 481-499

Newburn, D.A., Ferris, J. 2017. Additionality and forest conservation regulation for residential development. (Manuscript)

Papke, L., Wooldridge, J.M. 2008. Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics* 145: 121-133.

Petrin, A., Train, K. 2010. A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research* 47: 3-13.

Rivers, D.,Vuong, Q. 1988. Limited information estimators and exogeneity test for simultaneous probit models. *Journal of Econometrics* 39: 347-366.

Sieg, H., Smith, V.K., Banzhaf, H.S., Walsh, R.P. 2002. Interjurisdictional housing prices in locational equilibrium. *Journal of Urban Economics* 52: 131-153.

Stock, J.H., Wright, J., Yogo, M. 2002. A survey of weak Instruments and weak identification in generalized method of moments. *Journal of Business Economics and Statistics* 20: 518-529.

Tiebout, C. 1956. A pure theory of local expenditures. *The Journal of Political Economy* 64: 416-424.

Towe, C.A., Nickerson, C.J., Bockstael, N. 2008. An empirical examination of the timing of land conversions in the presence of farmland preservation programs. *American Journal of Agricultural Economics* 90: 613-626.

Walsh, R.P. 2007. Endogenous open space amenities in a locational equilibrium. *Journal of Urban Economics* 61: 319-344.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd Edition. The MIT Press.

Wrenn, D.H., Irwin E.G. 2015. Time is money: an empirical examination of the effects of regulatory delay on residential subdivision development. *Regional Science and Urban Economics* 51: 25-36.

Wrenn, D.H., Klaiber H.A. 2017. Price based land use policy and development density: impacts on urban land use. (Manuscript)

Wrenn. D.H., Klaiber, H.A., Newburn, D.A. 2017. Confronting price endogeneity in a duration model of residential subdivision development. *Journal of Applied Econometrics.* 32: 661-682.

Wu, J.J. 2008. Land use changes: Economic, social, and environmental impacts. *Choices*. 23: 6-10.

## Figures

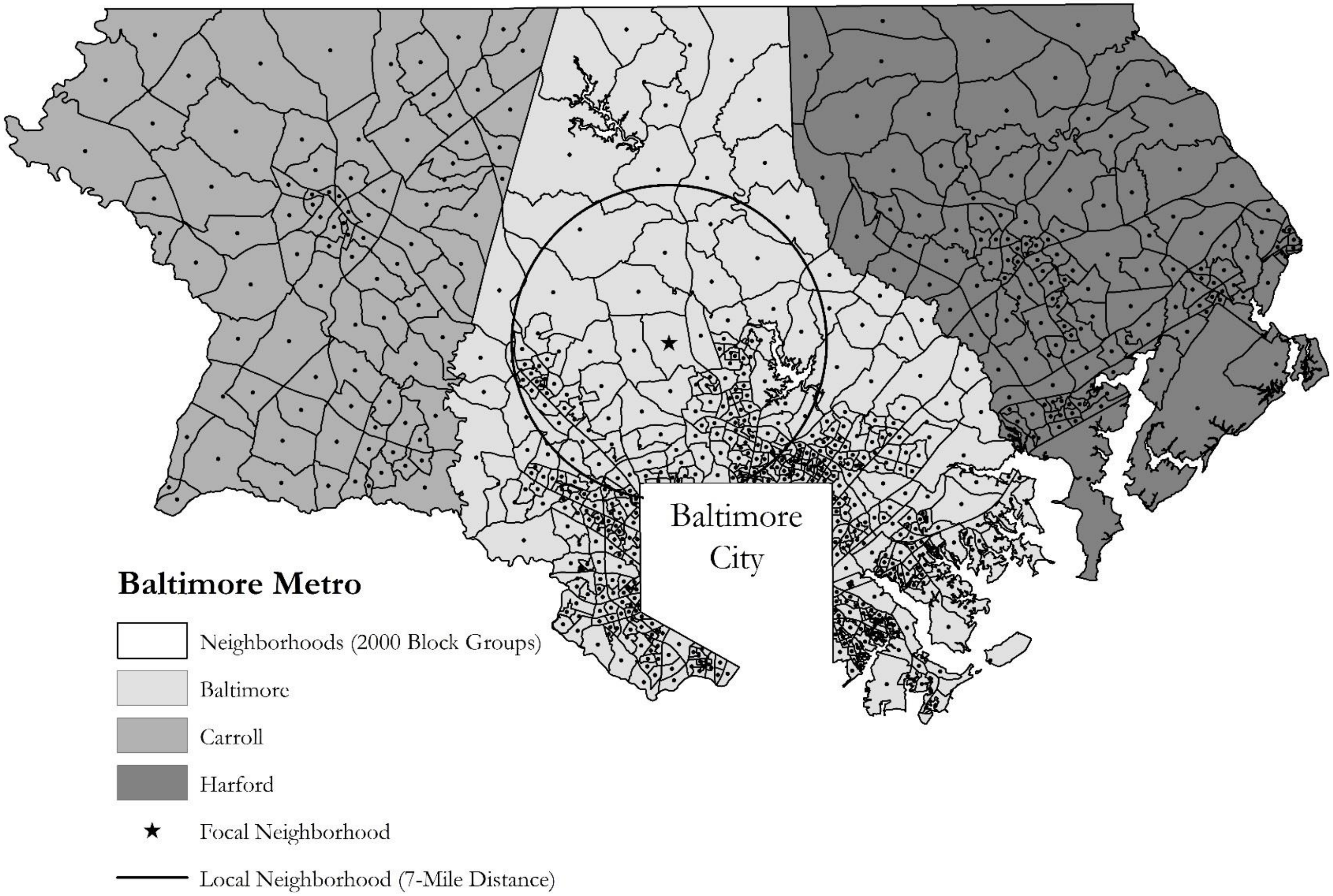


Figure 1. Map of the Baltimore metro study region. Neighborhoods are defined based on 2000 block group boundaries. The "local" neighborhoods in the figure are those that fall within the 7-mile distance cutoff drawn around the centroid of the "focal" neighborhood.

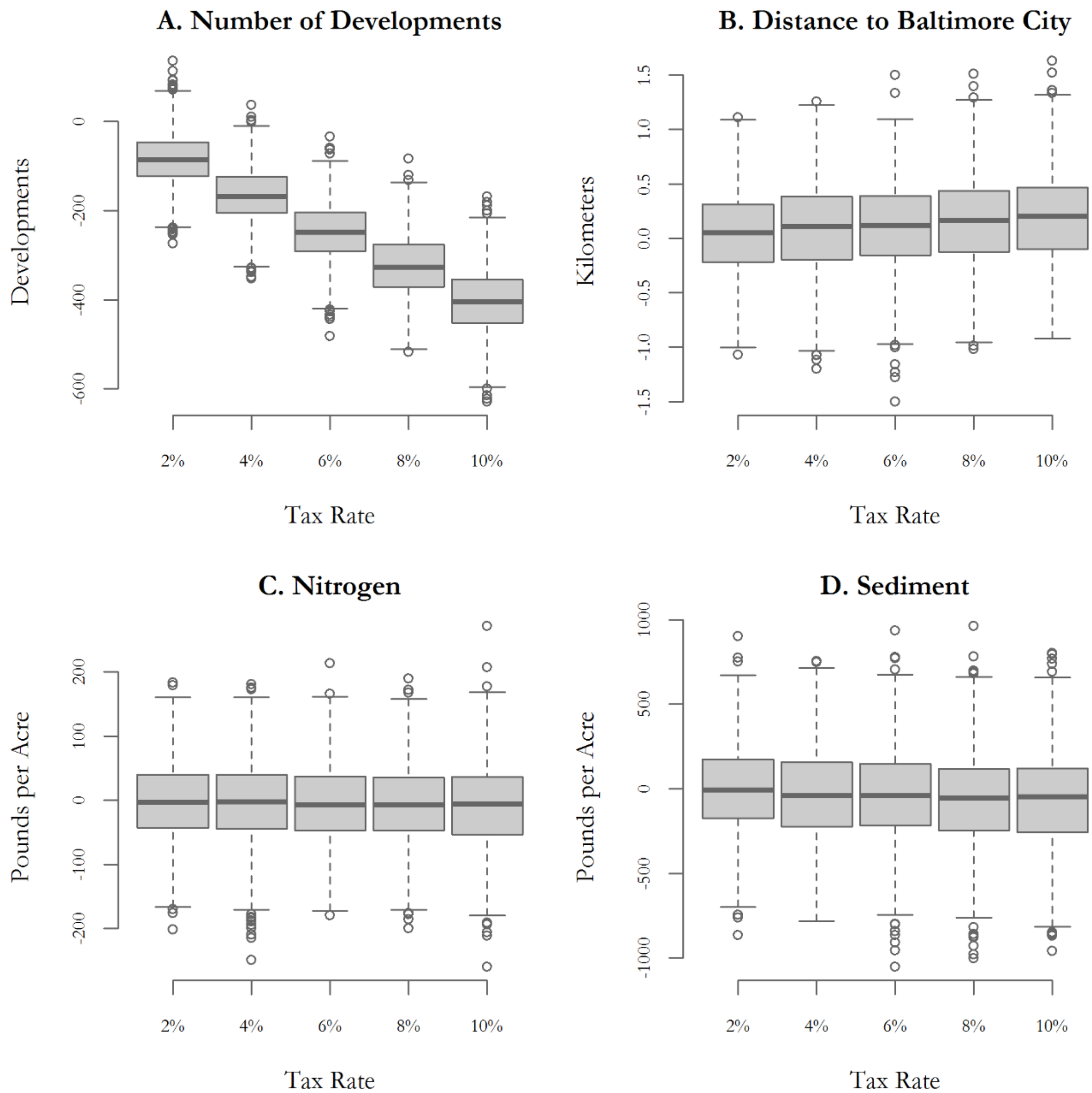


Figure 2. This figure displays results from a series of land use simulations under tax Scenario 1. In these simulations, a uniform tax is imposed on all parcels in the data with the different tax levels defined along the horizontal axis. Each simulation consists of 1,000 iterations, where each iteration combines land use and price data with a draw from the parameter distribution from the duration model and a probit link function to determine the likelihood of development for each parcel in each time period. The results for each tax simulation are compared with a baseline simulation, where prices are held at their original values and development outcomes are predicted. Part A. displays results for the change in the number of subdivision developments created; Part B. displays results for the change in the average commuting distance (in kilometers) for each of the developments in each iteration of the simulation; and Parts C. and D. display results for the change in average loading rates (lbs/acre) for nitrogen and sediment.

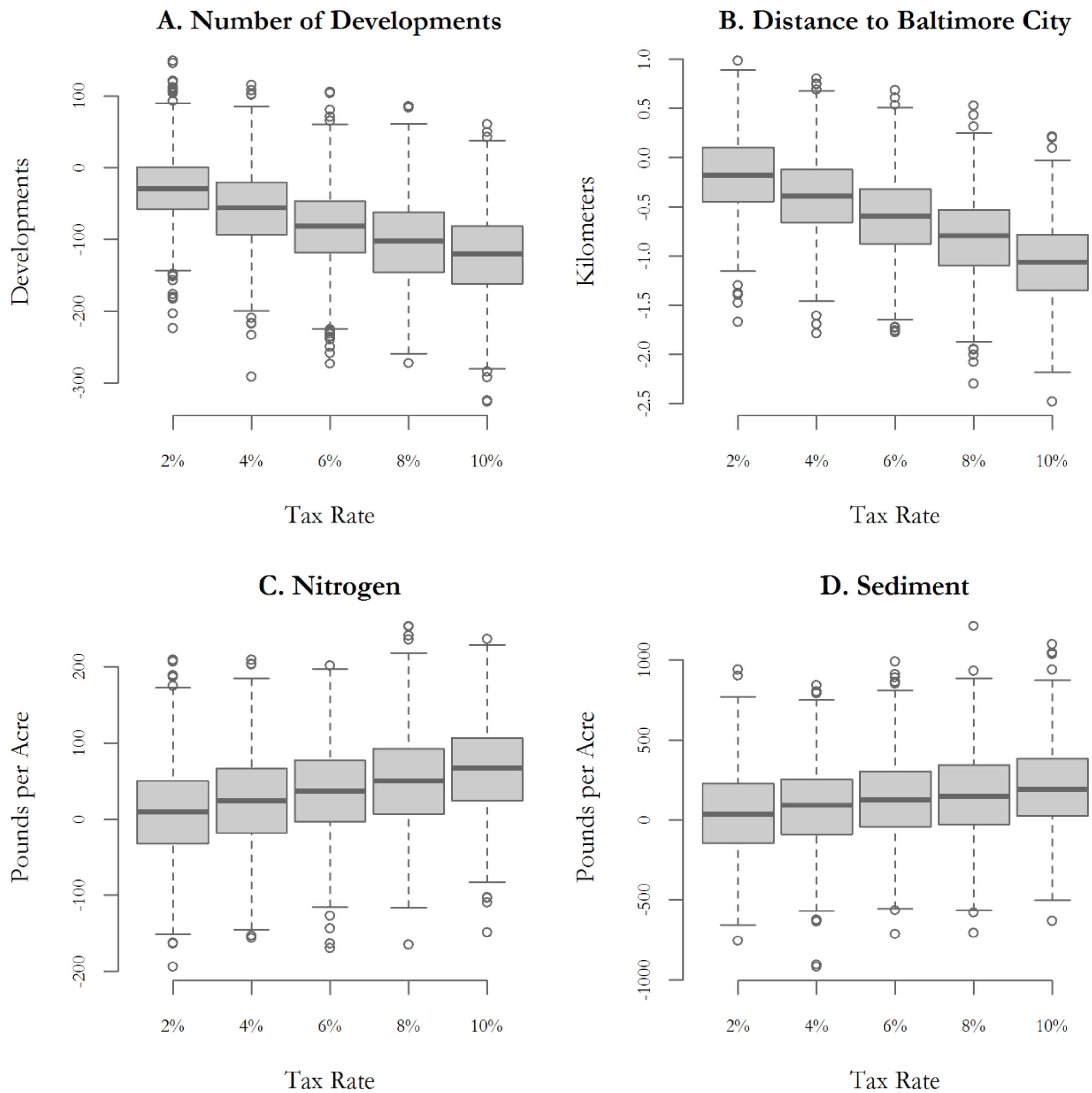


Figure 3. This figure displays results from a series of land use simulations under tax Scenario 2. In these simulations, a uniform tax is imposed on all parcels located in areas without public sewer and a uniform subsidy is imposed on all parcels with public sewer. The different tax/subsidy levels are defined along the horizontal axis. Each simulation consists of 1,000 iterations, where each iteration combines land use and price data with a draw from the parameter distribution from the duration model and a probit link function to determine the likelihood of development for each parcel in each time period. The results for each tax simulation are compared with a baseline simulation, where prices are held at their original values and development outcomes are predicted. Part A. displays results for the change in the number of subdivision developments created; Part B. displays results for the change in the average commuting distance (in kilometers) for each of the developments in each iteration of the simulation; and Parts C. and D. display results for the change in the average loading rates (lbs/acre) for nitrogen and sediment.

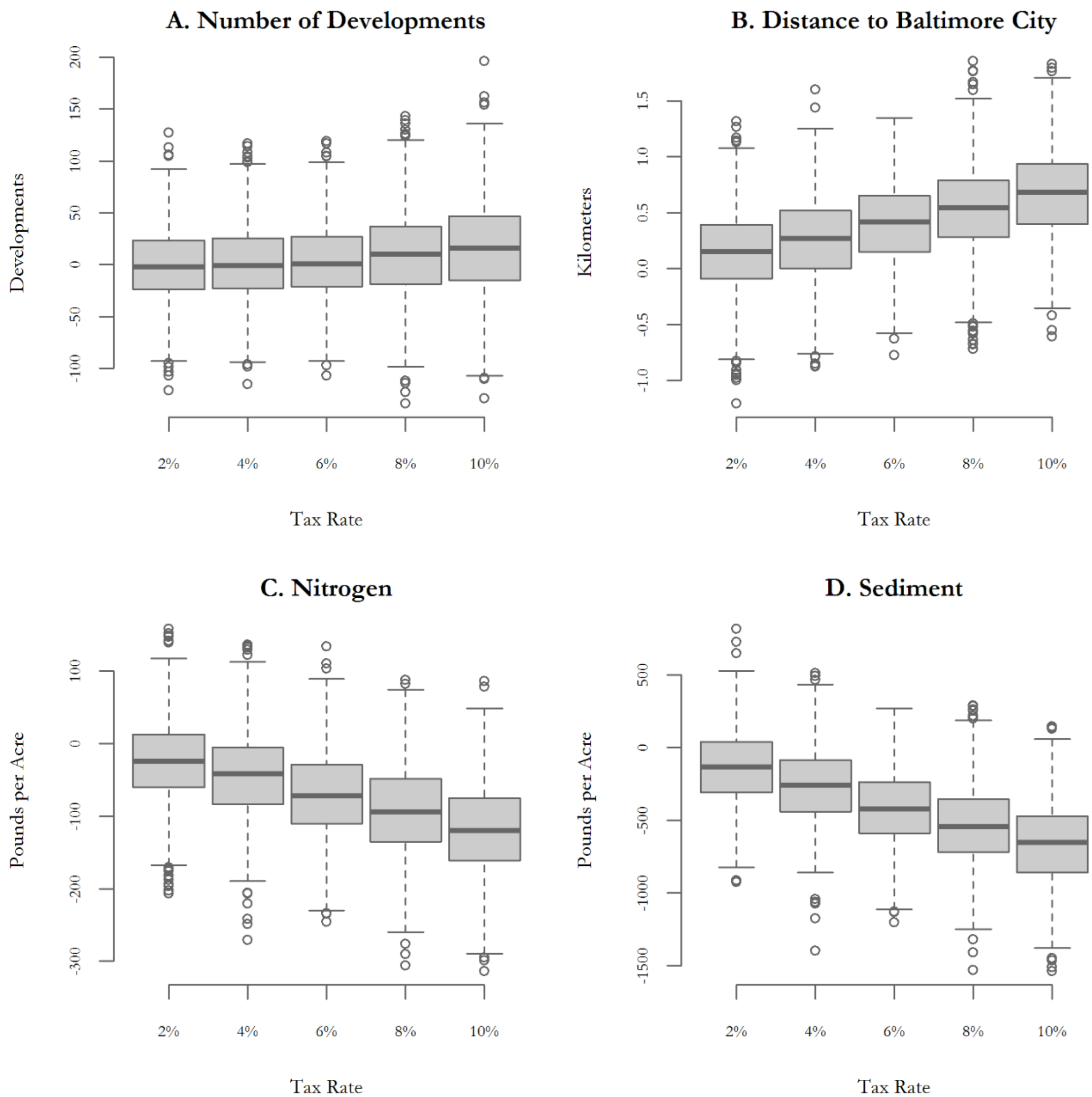


Figure 4. This figure displays results from a series of land use simulations under tax Scenario 3. In these simulations, a uniform tax is imposed on all parcels with greater than 50% forest cover and a uniform subsidy is imposed on all parcels with greater than 50% agricultural cover. The different tax/subsidy levels are defined along the horizontal axis. Each simulation consists of 1,000 iterations, where each iteration combines land use and price data with a draw from the parameter distribution from the duration model and a probit link function to determine the likelihood of development for each parcel in each time period. The results for each tax simulation are compared with a baseline simulation, where prices are held at their original values and development outcomes are predicted. Part A. displays results for the change in the number of subdivision developments created; Part B. displays results for the change in the average commuting distance (in kilometers) for each of the developments in each iteration of the simulation; and Parts C. and D. display results for the change in the average loading rates (lbs/acre) for nitrogen and sediment.

## Tables

Table 1. Descriptive statistics for parcel and neighborhood variables

| Variables | | Mean | St. Dev. | Min. | Max. |
|---|---|---|---|---|---|
| Parcel | | | | | |
| Dist | Kilometers to Baltimore City | 30.54 | 16.74 | 0.00 | 76.31 |
| DistMajRoad | Kilometers to closest major highway | 0.71 | 0.72 | 0.00 | 5.34 |
| Area | Parcel area in acres | 19.38 | 36.74 | 0.07 | 946.95 |
| ZndLots | Count of zoned lots allowed | 9.71 | 32.20 | 0.00 | 1378.00 |
| Sewer | Indicator for municipal sewer service | 0.46 | 0.50 | 0.00 | 1.00 |
| FloodPlain | Located in 100-year flood plain | 0.18 | 0.38 | 0.00 | 1.00 |
| SepticSuit | Indicator for septic suitability | 0.54 | 0.50 | 0.00 | 1.00 |
| Slope | % of parcel with slope > 15% | 9.81 | 23.22 | 0.00 | 101.93 |
| ExHouse | Has an existing structure | 0.54 | 0.50 | 0.00 | 1.00 |
| ForestPrcnt | % of parcel with forest cover | 38.77 | 34.98 | 0.00 | 100.00 |
| Neighborhood | | | | | |
| Preservation | % of neighborhood in preservation | 5.76 | 10.23 | 0.00 | 88.18 |
| UDArea | % of neighborhood undeveloped | 26.31 | 15.94 | 0.03 | 120.82 |
| ApprvLots | Count of lots approved - 1-year lag | 25.96 | 40.86 | 0.00 | 519.00 |
| LandPrice | In $1,000s per acre | 77.21 | 66.02 | 4.22 | 718.84 |
| HousePrice | In $1,000s | 128.42 | 44.68 | 20.25 | 493.73 |
| Baltimore | Located in Baltimore County | 0.58 | 0.47 | 0.00 | 1.00 |
| Carroll | Located in Carroll County | 0.24 | 0.43 | 0.00 | 1.00 |
| Harford | Located in Harford County | 0.19 | 0.39 | 0.00 | 1.00 |

Note - The statistics for the parcel variables are based on the 15,015 land parcels that were developable or developed during our study period (1994-2007). The statistics for the neighborhood variables are based on the 667 block groups (neighborhood boundaries) over the same time period.

Table 2. Diagnostic tests of IV duration model

| (1) IV Distance | (2) First-Stage $F$-Statistic | (3) OverID Statistic | (4) Control Function |
|---|---|---|---|
| 5 Miles | 1436.21 | 0.5530 | -0.0027 *** |
| | | | -2.3E-05 *** |
| 6 Miles | 1247.19 | 0.6429 | -0.0030 *** |
| | | | -2.3E-05 *** |
| 7 Miles | 1060.79 | 0.916 | -0.0029 *** |
| | | | -2.3E-05 *** |

Note - This table presents a series statistical tests of the IV durationmodel. Column (2) is a first-stage weak instrument test (F-stat); column (3) is an overidentification test (p-value); and column (4) presents the coefficient values and significance levels for the control function variables in the main duration model - a direct test of price endogeneity. The results on each line are based on the distance cutoff (column (1)) used in forming the instruments in the first-stage OLS model. The residuals from each first-stage model are included as controls (Wooldridge, 2010) in the duration model. Al lresults are based of a bootstrapped models with 300 reps.

* Significant at 10% level

** Significant at 5% level

*** Significant at 1% level

Table 3. Results for the non-IV and IV duration models

| | Non-IV Model | | | IV Model | | |
|---|---|---|---|---|---|---|
| | Coef. | | St. Err. | Coef. | | St. Err. |
| Parcel Characteristics | | | | | | |
| Dist (km) | 0.0017 | | 0.0011 | 0.0011 | | 0.0011 |
| DistMajRoad (km) | 0.0450 | *** | 0.0114 | 0.0484 | *** | 0.0114 |
| Area (acres) | 0.0007 | *** | 0.0002 | 0.0007 | *** | 0.0002 |
| ZndLots | 0.0021 | *** | 0.0002 | 0.0021 | *** | 0.0002 |
| Sewer | -0.0911 | *** | 0.0240 | -0.0583 | ** | 0.0245 |
| FloodPlain | -0.0731 | *** | 0.0230 | -0.0655 | *** | 0.0232 |
| SepticSuit | 0.0807 | *** | 0.0183 | 0.0499 | ** | 0.0192 |
| Slope | 0.0000 | | 0.0004 | -0.0001 | | 0.0004 |
| ExHouse | -0.0122 | | 0.0176 | -0.0088 | | 0.0177 |
| ForestPrcnt | -0.0018 | *** | 0.0002 | -0.0019 | *** | 0.0002 |
| Constant | -2.4335 | *** | 0.0577 | -2.7880 | *** | 0.0781 |
| Neighborhood Characteristics | | | | | | |
| Preservation (%) | -0.0013 | | 0.0009 | -0.0029 | *** | 0.0009 |
| UDArea (%) | -0.0027 | *** | 0.0008 | -0.0029 | *** | 0.0008 |
| ApprvLots | 0.0007 | *** | 1.7E-04 | 0.0003 | | 0.0002 |
| LandPrice (1K) | -0.0011 | *** | 0.0002 | -0.0011 | *** | 0.0002 |
| HousePrice (1K) | 0.0022 | *** | 0.0002 | 0.0058 | *** | 0.0006 |
| Residual | | | | -0.0029 | *** | 0.0008 |
| ResidualSqrd | | | | -2.3E-05 | *** | 6.1E-06 |
| Log-Likelihood | -12401.548 | | | -12374.785 | | |

Note - The table presents results from the Non-IV and IV duration models. The IV results are produced using a control function methodology (Wooldridge, 2010). The residuals are generated using a 7-mile distance cutoff in forming the instruments. All models include time and county fixed effects. The standard errors are based on a block bootstrap procedure with 300 replications and clustered at the
* Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level

Table 4. Price elasticity estimates

| | Coef. | | St. Err. |
|---|---|---|---|
| Non-IV | 0.7514 | *** | 0.0782 |
| IV | 1.9652 | *** | 0.2185 |

Note - The elasticity values are calculated using a marginal effects formula for probit model and represent the average percentage change in the probability of development (conversion) for a small change in price. The standard errors are calculated using the delta method and are based on the bootstrapped variance-covariance matrices from the duration model.

* Significant at 10% level;

** Significant at 5% level;

*** Significant at 1% level

Table 5. Nutrient and sediment load rates for Baltimore region

| | Nitrogen | | Sediment | |
|---|---|---|---|---|
| | Farmland | Forest | Farmland | Forest |
| Baseline Load (lbs/acre/year) | | | | |
| Baltimore | 17.2 | 2.5 | 835.2 | 61.3 |
| Carroll | 26.7 | 5 | 464.0 | 129.2 |
| Harford | 18.1 | 3.7 | 816.5 | 124.3 |
| Residential Loading Rates (lot size in acres) | | | | |
| 2 | 9.45 | | 235.2 | |
| 1 | 9.60 | | 268.2 | |
| 0.5 | 9.89 | | 330.6 | |
| 0.25 | 10.17 | | 388.5 | |
| 0.125 | 10.37 | | 431.2 | |
| 0.0625 | 10.86 | | 526.4 | |

Note - This table presents information on loading rates, in pounds per acre per year, for nitrogen and sediment across all three counties in the model. The top part of the table shows baseline loading rates for agriculture and forest land. The bottom part of the table shows loading rates for different residential density classes. All of these values come from the Chesapeake Bay Program's land use and water quality model.

Table 6. Predicted number of subdivisions created in baseline simulation using IV duration model

| Number of Developments | | | |
|---|---|---|---|
| Mean | St. Dev. | 95% CI | |
| 2396 | 60 | 2283 | 2518 |

Note - This table presents summary statistics for the predicted number of developments created across a series of simulations using the results from IV duration model. Each simulation (1000) is based on a draw from the matrix of bootstrapped parameter estimates from the model. The actual number of developments created in the data is 2,394.

Table 7. Results from land use simulations

| | Price Change | | | | |
|---|---|---|---|---|---|
| | 2% | 4% | 6% | 8% | 10% |
| Scenario 1 | | | | | |
| Number of Developments | -85 | -167 | -248 | -326 | -402 |
| Distance to Baltimore (km) | 0.05 | 0.09 | 0.11 | 0.15 | 0.20 |
| Nitrogen (lbs/acre) | -2.82 | -4.33 | -6.16 | -7.09 | -7.14 |
| Sediment (lbs/acre) | -6.16 | -32.18 | -47.25 | -60.39 | -57.51 |
| Scenario 2 | | | | | |
| Number of Developments | -29 | -56 | -82 | -102 | -121 |
| Distance to Baltimore (km) | -0.18 | -0.39 | -0.61 | -0.82 | -1.06 |
| Nitrogen (lbs/acre) | 8.03 | 23.39 | 35.89 | 49.19 | 65.38 |
| Sediment (lbs/acre) | 40.81 | 79.55 | 126.71 | 155.68 | 196.58 |
| Scenario 3 | | | | | |
| Number of Developments | -1 | 1 | 3 | 10 | 16 |
| Distance to Baltimore (km) | 0.14 | 0.26 | 0.40 | 0.53 | 0.67 |
| Nitrogen (lbs/acre) | -24.77 | -43.99 | -69.24 | -92.96 | -118.32 |
| Sediment (lbs/acre) | -132.66 | -259.31 | -415.97 | -528.19 | -660.52 |

Note - This table presents results from a series of land use simulations using the results from the IV duration model. The columns represent the sequence of price changes imposed in each policy scenario ranging from a 2% tax/subsidy to a 10% tax/subsidy. The table presents results for three different policy scenarios: Scenario 1 - a uniform tax imposed on all parcels; Scenario 2 - a tax imposed on parcels in areas without public sewer and a subsidy imposed on parcels in areas with public sewer; and Scenario 3 - a tax imposed on parcels with more than 50% forest cover and subsidy imposed on parcels with more than 50% agricultural land cover. The baseline land cover values are taken from the USGS land cover data for the Baltimore region. For each scenario and price change, we present results for the change in the average number of developments created, the change in the average distance of those developments from Baltimore City (in kilometers), and the change in average reduction (pounds per acre) of nitrogen and sediment for those developments. All changes (comparisons) are made relative to a baseline land use simulation without a change in price.

**Appendix**

**1. Creation of Housing and Land Price Variables**

1.1 *Land Price Indices*

To create our land-price variable we select all arms-length land transactions from the MDPV databases that occur between 1994 and 2007. We further refine these data by excluding any parcels that already contained a farmland preservation easement on the property, which precludes it from being sold for development at full market value. We further exclude observations that were clearly not land sales based on the improvement value of the parcel. Finally, we exclude the top and bottom 1% of the sample based on the sale price per acre of the parcel to reduce the potential influence of outliers. The final data set on land transactions includes 10,669 arms-length land sales from 1994 to 2007.

To create our land price variable we estimate the following hedonic regression

$$ln(rlppacre_{lt}) = Par_{lt}'\beta + \delta_j + \tau + e_{lt} \tag{A1}$$

where $rlppacre$ is the real land price per acre in year 2000 for land parcel $l$, $Par_{lt}$ is a set of parcel-level controls, and $\delta$ and $\tau$ are block group and year fixed effects, respectively. The set of parcel-level controls includes the size of the parcel in acres as well as an indicator for whether the sale was for a previously subdivided lot, which controls for any differences in price between subdivided and unsubdivided parcels. We estimate the land price hedonic model using the pooled data set due to the limited number of land sales during our study period (Table A1 lists the number of arms-length land transactions for each year in our data). After controlling for land parcel characteristics, the year and block group fixed effects are used to construct an estimate of mean land price per acre in each neighborhood. For block groups and years without a sale we use a distanced weighted average of the values of the block group fixed effects for the closest five block groups in space in each year. Since land is an input in the production of housing we expect land prices to negatively affect latent profitability.

| Year | Observations |
|---|---|
| 1994 | 851 |
| 1995 | 754 |
| 1996 | 972 |
| 1997 | 922 |
| 1998 | 1047 |
| 1999 | 1035 |
| 2000 | 909 |
| 2001 | 795 |
| 2002 | 831 |
| 2003 | 617 |
| 2004 | 664 |
| 2005 | 572 |
| 2006 | 376 |
| 2007 | 354 |

Table A1. Number of arms-length land transactions by year

1.2 *House Price Indices*

The data used to generate our house-price variable also comes from MDPV. Similar to the approach for the land price data, we use only arm's-length single-family housing transactions between 1994 and 2007. After excluding the top and bottom 1% of the sample to remove outliers and removing any transactions that do not appear to be of single-family dwellings, such as multi-family dwellings and commercial structures, the final sample for 1994-2007 has 187,497 individual transactions. We convert the nominal sale price of each house to year 2000 dollars using the consumer price index (CPI) for the Baltimore metropolitan area.

To construct our housing price indices we follow Sieg et al. (2002) and estimate a series of hedonic models that separate out the price of housing services at the neighborhood level from the quantity index of housing determined by structural and lot-specific characteristics of the house. To do this we estimate the following house-price hedonic for each year

$$ln(rlhspr_h) = P_j + H_h'\beta + \epsilon_h \tag{A2}$$

where $rlhspr_h$ is the real transaction price for house $h$ in census tract $j$, $P_j$ is a fixed effect for the block group in which the house is located, and $H_h'$ and $\epsilon_h$ are the observable and unobservable attributes for house $h$, respectively. We control for structure and lot-specific attributes of each house by combining our house price data with the tax assessor's data for each house. As shown in Sieg et al. (2002), $P_j$ represents the price of housing services for each block group. Repeating the estimation process in equation (A2) for each of the 14 years in our data provides a value for the price of housing services for each block group and year in our model.[10] This block group and year house price value is used in both our competing risks model and the first-stage regression as our measure of neighborhood house price.

One concern with the hedonic estimation strategy above is that by estimating yearly, instead of pooled, house price hedonics we implicitly assume that the quantity index is changing from year to year. However, estimating the quantity index (the coefficient on the housing characteristics) in each year comes at a price as sampling error is likely to increase statistical noise in the neighborhood fixed effects estimates. This extra noise is not likely to be a major issue in large samples, though it may affect the fixed effect estimates as yearly sample sizes decrease. Given that the sample sizes of our housing transactions data in each year are quite large (Table A2 lists the number of yearly arms-length housing transactions in our data), we are able to run separate hedonic models for each year to generate our neighborhood-level house price indices. We did, however, run a pooled regression for

[10] A similar method for estimating the price of housing services has been applied in other structural models (see Klaiber and Phaneuf (2010) and Walsh (2007), among others).

the land price hedonic model in the previous section due to the smaller number of land transactions over time. This is an important consideration in applying our method in other settings.

| Year | Observations |
|---|---|
| 1994 | 11127 |
| 1995 | 11032 |
| 1996 | 12708 |
| 1997 | 12254 |
| 1998 | 13130 |
| 1999 | 14523 |
| 2000 | 12940 |
| 2001 | 13706 |
| 2002 | 14487 |
| 2003 | 14974 |
| 2004 | 16125 |
| 2005 | 14970 |
| 2006 | 14147 |
| 2007 | 11374 |

Table A2. Number of arms-length housing transactions by year

## 2. Additional Results

| | Coef. | | St. Err. |
|---|---|---|---|
| Neighborhood Variable | | | |
| PreservationArea (%) | 0.6078 | *** | 0.0422 |
| Undeveloped Area (%) | 0.2291 | *** | 0.0231 |
| Instrumental Variable | | | |
| HHIncomeAvg | -11.0758 | *** | 0.2242 |
| PreservationAvg | 18.7935 | *** | 1.9077 |
| UndevelopedAvg | -10.6843 | *** | 0.7569 |

Note - This table presents results from the first-stage OLS price regressions. The dependent variable is the quality-adjusted hedonic price index for each "focal" neighborhood in each time period. The instrumental variables are area-weighted average values of the variables specified in the table and are based on the values of those variables in "distant" neighborhoods, or in neighborhoods that are greater than the distance cutoff for contiguity to the "focal" neighborhood. The distance cutoff used here is 7 miles. All models include time and county fixed effects.
N = 9,213

* Significant at 10% level; ** Significant at 5% level; *** Significant at 1% leve

Table A3. First-stage OLS price regression