Confronting Price Endogeneity in a Duration Model of Residential Subdivision Development

Douglas H. Wrenn

Department of Agricultural Economics, Sociology, and Education,

Pennsylvania State University,

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

January 30, 2015

Selected Paper prepared for presentation for the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28.

This research is supported by a cooperative agreement with the U.S. Forest Service Northern Research Station, National Science Foundation DEB-0410336 and CBET-1160961.

Confronting Price Endogeneity in a Duration Model of Residential Subdivision Development

Abstract

Spatial equilibrium implies that distant factors are correlated with proximate locations through market mechanisms. Using this logic, we develop a novel approach for handling price endogeneity in reduced-form land use models. We combine a control function approach with a duration model of land development to shed new light on the role of price and supply-side factors that influence subdivision development at a micro level. We find that failure to control for endogeneity results in large differences in estimates of residential land supply price elasticities. Specifically, we find an elasticity of 2.06 compared to 0.67 in a model that ignores potential endogeneity.

Keywords: Endogeneity, Control Function, Duration Model, Land Supply Elasticity
JEL Codes: C26, R12, R14, R52

I. Introduction

Over the past several decades state and local planning agencies have become concerned about urban sprawl and its impact in terms of loss of farmland, congestion, and the degradation of urban ecosystems (Nechyba and Walsh, 2004; Glaeser and Kahn, 2004; Hansen et al. 2005).[1] Urban land use patterns are the result of supply and demand forces that determine urban development patterns and give rise to market clearing prices and influence individual landowner development decisions to alter the observed urban spatial structure (Irwin and Wrenn, 2014). Although a significant amount of empirical research has focused on the role of demand-side factors (schools, crime, and environmental amenities) in determining urban spatial structure, the recent literature on modeling the supply side of the housing and residential land market has been limited (Glaeser et al. 2006).

Addressing this deficiency is critically important from an academic as well as a policy perspective. Many smart growth policies are focused on limiting development in specific areas and redirecting it into designated development corridors. While there are many regulatory and market-based policy options designed to achieve this goal, price-based policies such as impact fees or green taxes are becoming increasingly popular as an option for growth management in practice and have been favored in lieu of regulatory approaches (Bruecker, 2000). The design and effectiveness of these policies, however, depends critically on knowing how responsive agents' land conversion decisions are to changes in prices.[2]

[1] By the 1990s, concerns about sprawl had translated into a number of statewide smart growth policies (Florida, Maryland, New Jersey, and Oregon) aimed at managing urban sprawl (Ingram et al., 2009).

[2] Of the recent research that has analyzed supply-side issues, most has focused on more aggregate measures of housing supply elasticity and the impact of regulation and geographic restrictions on the price and quantity of housing units supplied (Mayer and Somerville, 2000; Glaeser at al. 2006; Saks,

The aggregate supply of housing is a complex function of individual decisions made by landowners and developers.[3] This supply, however, must necessarily begin with the decision of an individual landowner to convert a previously undeveloped parcel to residential use, which effectively primes residential lots for future housing supply. The use of duration models to analyze land development decisions is rooted in the dynamic urban growth literature, where landowners with perfect foresight choose the optimal timing of land development (Anas, 1978; Arnott, 1980). Structuring the decision to develop land as an intertemporal problem allows these models to approximate the sequential development patterns observed in the real world while giving rise to the notion that changes in parcel and neighborhood-level factors over space and time influence the optimal timing of one's own development decisions. Reduced-form duration models have been used to examine the influence of land use externalities (Irwin and Bockstael, 2002), zoning and other growth management policies (McConnell et al. 2006; Newburn and Berck, 2006; Cunningham 2007), open-space conservation policies (Lewis et al. 2009; Towe et al. 2008), and regulatory costs (Wrenn and Irwin, 2014).

Duration models have also been used to test real options theory for land development with a focus largely on house price uncertainty (Cunningham, 2006, 2007; Bulan et al. 2009; Towe et al. 2008). These models are attractive as they operate at a micro level, easily incorporate temporal

---

2008; Saiz, 2010). See Murphy (2014) and Klaiber and Phaneuf (2014) for the two recent exceptions to the rule.

[3] While there is often a distinction between landowners and land developers, this distinction is not always clear as original landowners may sell the property to a land developer, develop their own land, or form a contract for a shared partnership with the developer. Because this distinction is not the focus of our research, here we assume the terms developer and landowner are synonymous throughout the remainder of the paper.

dimensions of choice, and avoid making explicit assumptions on underlying profit or cost functional forms. In the vast majority of the literature, however, covariates are typically included without instrumentation. One potential reason for this is that the nonlinear duration framework makes instrumentation inherently challenging.

Recently, a number of papers have attempted to address endogeneity concerns on the demand side of the housing market by applying the spatial equilibrium theory from Tiebout's (1956) model of urban sorting to estimate empirical structural models of household location choice (Bayer and Timmins, 2007; Klaiber and Phaneuf, 2010). Using the logic of long-run spatial equilibrium, these papers estimate the primal parameters from the household maximization problem and use these values along with the logic of the random utility model to conduct counterfactual policy analyses. One limitation of this demand-side approach to structural modeling is the difficulty in accounting for dynamics and the intertemporal optimization on the supply side of the market. While the estimation of dynamic, structural supply-side models is emerging in the literature (Murphy, 2014), these models largely operate at the level of the individual housing lot and have not considered the subdivision nature of the residential land conversion decision.

In this paper, we model the landowner conversion decision on whether to subdivide a developable land parcel to residential use while simultaneously instrumenting for housing price using the logic of spatial equilibrium underlying urban real estate markets. Our proposed method combines a control function (CF) approach to instrumentation (Rivers and Vuong, 1988; Papke and Wooldridge, 2008; Petrin and Train, 2010) with the spatial equilibrium logic used for the development of instruments in the demand-side literature of urban housing markets (Bayer and Timmins, 2007). The fundamental insight of our instrumentation approach is that variation in prices that reflects exogenous land supply characteristics of distant locations can serve as an instrument for price and can be captured easily using a control function approach suitable for use in a duration model. This approach builds on the instrumentation ideas present in structural models of housing markets where equilibrium levels of endogenous attributes at each location are influenced by the attributes of all other locations through the market equilibrium – i.e., distant factors are likely correlated with proximate locations through the market equilibrium (Epple and Sieg, 1999; Bayer and Timmins, 2007; Klaiber and Kuminoff, 2014). We apply our method to a unique data set where we have reconstructed the panel of residential subdivision events from 1994-2007 using historic plat archives from three counties in the Baltimore metropolitan region. Combining these data with quality-adjusted hedonic estimates for neighborhood housing prices, we estimate an instrumental variables (IV) duration model of residential subdivision development.

This paper makes several important contributions to the literature. First, by combining spatial equilibrium insights with a reduced-form duration model, we develop a novel methodology to control for price endogeneity in reduced-form land use models. As the results from our proposed IV model demonstrate, accounting for the endogeneity of price produces an estimate of the long-run price elasticity of residential land supply that differs significantly from the estimate in a model without instruments. Specifically, we recover an estimate of the price elasticity of residential land supply of 2.06 when using our preferred IV model compared to an estimate of 0.67 in a model without instrumentation. This contribution also has much broader appeal. Our results demonstrate that endogeneity is significant in reduced-form land use models, and our general modeling framework to control for it is flexible enough to be used to control for other endogenous variables that may arise in a variety of different contexts.

Second, this research contributes to the urban economics literature on housing and land supply. Research on housing and land supply has been historically limited (Mayer and Somerville, 2000; Glaeser et al., 2006; Gyourko 2009; Saiz 2010). Of these papers, all have estimated models of housing supply that focus on the supply of individual housing units – i.e., building the structure after the subdivision conversion decision has already been made. However, the supply of new housing is a multistep process, which begins with the initial subdivision decision to create residential lots. Our paper makes a unique contribution to this literature by developing the first consistent estimate of the price elasticity of residential land supply. Our results on land supply are not directly comparable to the previous literature that focuses only on housing supply, though our estimated elasticity measure appears reasonable relative to previous estimates of the supply elasticity of housing. Saiz (2010), for example, uses aggregate data on geographic land restrictions and regulation from the Baltimore, Maryland MSA, and estimates a supply elasticity of housing of 1.23. This suggests our estimated elasticity of residential land supply of 2.06, which accounts for endogeneity, is at least consistent in that it shows a more elastic response of landowners to changes in price relative to the elasticity estimate of 0.67 for our non-IV model that ignores endogeneity. Finally, we make a contribution to the literature on land use policy analysis and design. Our results demonstrate that ignoring the endogeneity of housing prices results in large biases in price elasticity estimates. Analysis of price-based land use policies designed to manage residential development and urban spatial structure must account for landowner responsiveness to price and the potential endogeneity of price, which otherwise would provide misleading assessments for the effectiveness of these policies.

II. Econometric Model

Our econometric model extends previous reduced-form duration models and places it within the context of the urban spatial structure described by structural demand-side models (Walsh, 2007; Klaiber and Phaneuf, 2010). We assume that in each period t the landowner of an undeveloped parcel i located in neighborhood j decides whether or not to convert her parcel to a residential subdivision.[4] The decision to convert is based on factors that vary at the parcel level I_{it} as well as factors that vary at the neighborhood level X_{jt}, where neighborhoods in our application are defined at the census tract level. Parcel-level variables represent factors such as soil quality and slope, which affect the expected profitability of converting to residential use. Meanwhile, neighborhood-level variables represent broader-scale factors, which affect the average profitability of converting a given parcel in the neighborhood such as house and land prices or the amount of inventory in terms of previously approved residential lots. The intuition is similar to the urban sorting literature where parcels are individual observations nested within neighborhoods affected by neighborhood-level characteristics (Klaiber and Kuminoff, 2014).

Because we do not observe the actual profits and costs for an individual parcel, we use fine-scale data on the factors most likely to affect profitability on a given parcel and specify the following reduced-form profit model

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

where Π_{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 quality adjusted price of housing at the neighborhood level, and u_{it} is an idiosyncratic parcel-level error term.

Given the dynamic nature of the land development process, we model the optimal timing decision using a discrete-time duration model (Beck et al., 1998). Duration models take account of the fact that an action taken in period t implies the action was not taken in any previous period

[4] We only model the decision of landowners to subdivide their parcel to single-family residential use. While other land use types (e.g., commercial, industrial, apartments etc.) are important in determining the urban spatial structure, residential land use accounts for the majority of the developed land area in most urban and suburban areas.

($T < t$), which is the essence of an optimal stopping investment decision inherent in residential subdivision development (Capozza and Helsley, 1989; Dixit and Pindyck, 1994; Capozza and Li, 2002). The random variable, t, is the time until a subdivision event occurs, where we are interested in the effect of a set of covariates, including price, on an individual conversion decision.

The observations in our data are spells over time of the same parcel unit and realizations are characterized by the following density function

$$f(t) = P(t \leq T < t + dt). \tag{2}$$

with a corresponding cumulative density function of

$$F(t) = \int_0^t f(s)ds = P(T \leq t)\,, t \geq 0. \tag{3}$$

Combining equations (2) and (3) produces the following hazard function

$$h(t) = P(t \leq T < t + dt | T \geq t) = \frac{f(t)}{1 - F(t)}. \tag{4}$$

This function represents the instantaneous probability of a subdivision event occurring in the time interval dt given that it has not occurred prior to that time. Based on equation (4), the parametric proportional hazard we adopt is

$$h(t) = h_0(t)h(I'_{it}\beta + X'_{jt}\alpha + P'_{jt}\gamma), \tag{5}$$

where $h_0(t)$ is the baseline hazard, which is shifted proportionally by changes in the variables in the model. Given that our subdivision data are only available at a yearly time step, we use the discrete-time duration model proposed by Beck et al. (1998). This paper demonstrates that a simple binary probability specification with time fixed effects provides the same fit to the data as a piece-wise exponential duration model when the data is only observed at interval time steps.[5]

[5] Beck et al. (1998) show that, while a complementary log-log (cloglog) specification is the statistical equivalent to a continuous-time parametric survival model, any binomial probability model – i.e., a

By including a spatially and temporally varying housing price covariate in our duration model, we are able to estimate the impact of price on the timing of parcel-level land conversions. Our IV method follows from the CF approach developed by Rivers and Vuong (1988) and later extended by Papke and Wooldridge (2008) and Petrin and Train (2010). The CF approach uses a two-step estimation procedure to instrument for endogenous variables in the main equation using residual variation derived from a first-stage linear regression. Our model builds on this approach and extends it by providing a framework for controlling for price endogeneity in a parcel-level discrete-time duration model using the spatial equilibrium properties of the housing market (Bayer and Timmins 2007; Klaiber and Phaneuf 2010).

To make explicit the potential endogeneity of price in our duration model, we specify the neighborhood price vector P_{jt} as a function of our set of exogenous neighborhood variables X_{jt} and a set of excluded instrumental variables Z_{jt} that control for the correlation between price and the error term as

$$P_{jt} = X_{jt}^{'}\beta + Z_{jt}^{'}\delta + v_{jt}, \tag{6}$$

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

Endogeneity of house price arises if unobserved factors not accounted for in the latent profit equation are correlated with neighborhood house prices. For example, unobservable factors at the neighborhood level that positively affect the propensity to convert a parcel may include the

probit or logit – provides a reasonable fit to interval data. We use the probit specification here to maintain the joint normality assumption in the first and second stage of the control function model.

[6] As is the case in a standard 2SLS IV models, 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.

expected waiting time for subdivision approval or natural landscape features. If these factors influence the rate and number of parcels developed, it will result in a negative correlation between the unobservables and the price of housing at the neighborhood level yielding inconsistent estimates of the price coefficient γ in equation (1).

In the presence of endogeneity, the error term from equation (1) may be written as

$$u_{it} = v_{jt}^{'}\theta + e_{it}. \tag{7}$$

Assuming we have a properly specified first-stage regression in equation (6) and joint normality between u_{it} and v_{jt}, price is endogenous if u_{it} and v_{jt} are correlated, meaning if $\theta \neq 0$. To confront this endogeneity concern, we follow the two-step procedure in Papke and Wooldridge (2008). The first step is to estimate a reduced-form linear regression model for price with a set of excluded instruments added to control for endogeneity. The residual vector, $\hat{v}_{jt}$, from this first-stage regression model is added to the second-stage duration model as an additional covariate. Assuming that the instruments in the first stage are valid, a simple *t*-test of the coefficient on θ provides a valid test of the null hypothesis that P_{jt} is exogenous.

We now rewrite our original latent profit model as

$$\Pi_{it}^{*} = I_{it}^{\prime}\beta + X_{jt}^{\prime}\alpha + P_{jt}^{\prime}\gamma + v_{jt}^{\prime}\theta + e_{it}. \tag{8}$$

Assuming joint normality between the errors in both stages of the CF model, we can model equation (8) as a discrete-time duration model as follows

$$\begin{aligned} P\left(\Pi_{it}^{*} = 1 \middle| I_{it}, X_{jt}, P_{jt}, v_{jt}\right) &= h\left(t \middle| I_{it}, X_{jt}, P_{jt}, v_{jt}\right) \\ &= \Phi\left[\frac{I_{it}^{\prime}\beta + X_{jt}^{\prime}\alpha + P_{jt}^{\prime}\gamma + v_{jt}^{\prime}\theta + \tau_{t-t_0}}{\sqrt{1-\rho^2}}\right]. \end{aligned} \tag{9}$$

Equation (9) makes explicit the role that price plays and the potential endogeneity that arises by ignoring correlation between price and the error. Using a discrete-time duration model also allows us to nest our methodology of instrumenting for price in a parcel-level land use model in the context of the CF econometric model described above, while including a set of time fixed effects τ_{t-t_0} to model the baseline hazard and account for the censored nature of the data.

A key requirement of our CF approach and its application in the estimation of the duration model in equation (9) is the existence of a set of excluded variables Z_{jt} that are sufficient to control for the endogeneity of price. As in all IV models, sufficiency requires that: (1) the instruments have a direct and significant impact on the endogenous variable (i.e., the instruments must not be "weak"); and (2) they must not have a direct influence on the outcome variable in the main equation nor be correlated with the error term in that equation. Overcoming each of these requirements has been historically difficult in land use models as price is simultaneously determined as part of the spatial equilibrium outcome of the housing market.

Recent structural empirical models of residential location choice provide an approach to develop a credible set of instruments for price. Bayer and Timmins (2007) demonstrate that by exploiting the logic of the Nash equilibrium outcome of the residential housing market it is possible to form an optimal instrument for price. The intuition is that exogenous variables in distant neighborhoods should influence the price variable in a given focal neighborhood as a result of the spatial equilibrium in the market, but that those same exogenous variables are unlikely to be correlated with the error or influence the outcome variable directly in the focal neighborhood. [7]

---

[7] Recently, this method was extended by Klaiber and Phaneuf (2010) in a horizontal sorting model with multiple time periods. In their model, Klaiber and Phaneuf (2010) use a BLP (Berry et al., 1995) estimation technique and develop a per-period instrumental variable for use in the second stage of the model based on exogenous residual variation in the data outside a specified distance ring around each neighborhood.

We exploit a similar methodology and develop the price instrument in our model based on the spatial equilibrium of the housing market. We assume that in each time period the price in a focal neighborhood is influenced by the exogenous neighborhood characteristics in distant neighborhoods in space, but that those same exogenous attributes have no direct impact on the latent profitability of a given parcel in that focal neighborhood. Thus, by adding exogenous variables from proximate neighborhoods to the right-hand side of equation (6), we can effectively net out all local variation and use the residual from equation (6), which represents exogenous variation in distant neighborhoods, as an instrument for price in a manner similar to the urban sorting models.[8]

While spatial equilibrium theory does suggest a general method for developing an instrument for price, it does not reveal exactly what distant means in terms of developing the excluded variables in equation (6). We follow the previous literature (Klaiber and Phaneuf, 2010) and combine the theory of spatial equilibrium with a set of statistical tests for testing the validity of our instruments. For each focal neighborhood, we develop our set of excluded instruments based on the number of proximate neighbors in space. Specifically, for each model we use an increasing number of nearest neighbors in forming our IV matrix Z_{jt}^{n}, where the superscript n indexes the number of proximate tracts used in forming the Z matrix. Then, using each of these models we run overidentification tests and choose the optimal model based on the Chi-squared values from these tests (Stock et al, 2002; Wooldridge, 2010). As our results show, by adding more proximate neighbors to form the instruments in the first stage we retain the power of the instrument and pass all tests for IV validity; a result that is exactly predicted by urban spatial theory. The description of our data and the

[8] In developing our instrumentation method, we define the tract of interest in equation (6) as the "focal" tract, the nearest neighbor tracts used in estimating equation (6) as "proximate" tracts, and the residual variation for tracts outside of our nearest-neighbor cutoff as "distant" tracts.

development of our instruments are given in the next section, and the results of our statistical tests are given in the results section.

III. Data and Construction of Variables

The parcel-level data used in our duration model comes from three counties in the Baltimore metropolitan area – Baltimore, Carroll, and Harford counties. Each of these counties has experienced substantial population growth and residential development over recent decades. Figure 1 shows a map of our study region displaying both the county boundaries and census tract boundaries in 2000. We use the census tract boundaries to define neighborhoods. We selected this particular region because of the availability of micro-level data on residential subdivisions to model land conversion decisions. Data on residential land-use conversion in the three counties was derived from parcel data obtained from the Maryland Department of Planning. Using this data, we manually reconstructed the panel of residential subdivisions using historic archives for all recorded plats from 1994 through 2007. The year of subdivision approval from the historic plat maps is used for the timing of the residential conversion events.[9] By identifying all parcels in the same subdivision, we determine the original "parent" parcel and, thus, reconstruct the landscape for parcel boundaries in 1994.

We determined the baseline data set of developable parcels in 1994 as including those parcels that were eligible for residential development as of 1994 and could be subdivided into two or more buildable residential lots according to the parcel size and zoning. Because we are focused exclusively on modeling the subdivision conversion process for single-family residential

[9] While some development activity takes place outside of residential subdivisions (i.e., development of single lots), over 85% of the single-family houses in our data are located in a subdivision with two or more lots. As a result, we focus exclusively on residential subdivision conversion events in this paper.

development, we have screened out parcels that are zoned for commercial, industrial, multi-family dwellings (apartments), institutional, and protected areas. Parcels that are put into land preservation easements are considered developable from 1994 until the date of easement, after which they are not considered developable. The final data set consists of 14,576 parcels that were developable at the beginning of 1994. These parcels experienced 2,385 subdivision events during our study period of 1994 through 2007. As a demonstration of the unique nature of our data, Figure 2 shows development activity during 1994 to 2007 (light grey) and parcels remaining developable in 2007 (dark grey) for a single neighborhood or census tract. At the beginning of 1994, there were 75 developable parcels in this tract, with 27 residential subdivisions occurring between 1994 and 2007 and 48 parcels remaining undeveloped (or censored) at the end of the study period in 2007. The other parcels (white) in Figure 2 are either already developed prior to 1994 or zoned to not allow residential development.

For our model, we include a census tract in our data set if it is "active" such that it has at least one land conversion event occurring during our study period. There are 277 total census tracts in our three-county study region. However, only 229 experienced a subdivision event during our study period in 1994-2007. The majority of the tracts that did not have an event were: (1) located in very high-density residential areas with mostly apartment building development; (2) were completely developed with no remaining development potential; or (3) were located in areas zoned exclusively for commercial or industrial development. Figure 3 shows the data set of census tracts included in our model, which includes 229 census tracts over 14 years (or 3,206 tract-by-year observations). Hence, the final panel data set contains 183,580 parcel-by-year observations nested within 3,206 tract-by-year observations.

Summary statistics for variables used in our model are given in Table 1. 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 City of Baltimore (Dist), which reflects accessibility to the largest employment center in the region. We also include the distance to the closest major highway (DistMajRoad) as a local measure of accessibility to transportation infrastructure. Both of these variables are expected to increase the value of the parcel and its propensity to develop the closer 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 more individual lots that are allowable when the parcel is developed. We obtained historic zoning maps for each of the three counties from the Maryland archives and overlaid these maps with the parcel boundary data. The variable, ZndLots captures the zoned lot capacity for each parcel based on the parcel size, proportion of the parcel in each zoning type, and maximum density regulations by zoning type. While zoning has changed in the region during our study period, these changes were relatively small with the vast majority of the study area zoned in the mid-1970s. Since this time, zoning boundaries and rules have remained virtually unchanged for Carroll and Harford counties. Baltimore County had relatively minor changes to zoning boundaries between 1996 and 2008. To account for these changes, we obtained historic zoning boundary maps for Baltimore County that enabled us to accurately calculate the zoned capacity for each parcel and each year for our study period. We expect that parcels that have more development rights are likely more valuable and, thus, more likely to develop.

The final set of parcel-level 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 the permeability and percolation classification of each parcel. We expect that parcels with 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 development 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 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).

The bottom portion of Table 1 describes the neighborhood-level variables that represent characteristics for each census tract. The first tract-level variable in our model controls for the amount of the census tract that is covered by farmland preservation (Preservation). Maryland has an extensive farmland preservation program, and we were able to obtain data on these preservation events. Using these data, we develop time varying variable of the total percentage of land area in each tract that is preserved in each time period. We also control for the percentage of land area in each tract that remains developable in each period (UDArea).

The next two variables control for the tract-level impact of zoning and local competition. The first variable, ZndLots, represents the total zoned capacity for number of allowable residential lots on undeveloped parcels in each census tract for each time period. This variable is similar to the parcel-level zoned capacity variable but aggregated to the census tract. While we expect zoned capacity to have a positive impact at the parcel level, the sign of this variable at the neighborhood level is ambiguous. Higher zoned capacity may increase the rate of development if it represents more development potential. However, after controlling for price and other factors, increased zoned capacity may also signal lower potential profit potential, which would decrease the likelihood of conversion. The variable, ApprvLots, controls for local competition using a one-year lag on recent subdivision activity at the tract level. Using our historical subdivision data, we create a lagged measure of the number of approved lots in residential subdivisions in each period. For example, in 1994 the model includes the number of approved lots in 1993 for each tract. This variable is updated in each time period and census tract based on recent development activity in the prior year.

The last two variables control for land (LandPrice) and housing (HousePrice) prices at the neighborhood level. Both variables were created by applying hedonic land and housing-price models to arms-length transactions we obtained from Maryland Property View (MDPV). MDPV is a statewide GIS database which has yearly snapshots of all parcels and land and housing transactions in Maryland. In addition to the sale price for each transaction, the data sets also include detailed information of the characteristics of the houses and land parcels including lot area, structure size in square feet, structural quality, numbers of bathrooms, age of the structure, garage, and other attributes. We follow Sieg et al. (2002) and estimate a series of hedonic models that permit us to separate out the price of housing services at the neighborhood level from the quantity index of housing that is determined by structural and lot-specific characteristics of the house. Similarly, we estimate a hedonic regression to obtain a tract level measure of land prices included as an additional control variable. Details of this procedure are in the appendix.

To develop our instrumental variables for housing price, we create a set of instruments Z_{jt} used in equation (6) based on the exogenous tract-level variables located in proximate census tracts in a given time period. The intuition is that for each focal tract in each time period, we take the area-weighted average of the exogenous tract-level variables in a given number of proximate census tracts surrounding the focal tract in each time period and add these variables to the right-hand side of equation (6) as follows

$$P_{jt} = X_{jt}'\beta + Z_{jt}^{n}\delta + v_{jt}, \tag{10}$$

where superscript n on the Z matrix specifies the number of proximate tracts used to form the matrix of excluded instruments. We estimate equation (10) as a pooled OLS regression and include both time and county-level fixed effects in the model.[10]

By controlling for the effect of proximate and focal tract-level determinants of price, the error term in equation (10) effectively accounts for residual variation in the exogenous attributes located outside of the nearest-neighbor cutoff. This exogenous variation in distant tracts serves as a sufficient instrument for price.[11] To examine the robustness of this method and choose the optimal second-stage model, we use a varying number of proximate tracts, ranging from seven to eleven neighbors, in each time period and take the average value for a set of exogenous tract-level variables in each of the proximate tracts.

Summary statistics for the IVs used in the first-stage OLS model are shown in Table 2. We use the average values of preservation, zoned lot capacity, and undeveloped area as our excluded instruments. Figure 1 shows an example focal tract as well as the specification of the census tracts

[10] County fixed effects are included in both stages of the model to account for county-level unobservables. In Maryland, most land use policies are established and applied at the county level, so it is important to proxy for any time-invariant county-level effects not accounted for by the covariates in the model.

[11] This method is also related to the Hausman-style instrument used to control for price endogeneity in structural empirical demand and IO models (Hausman, 1997; Petrin and Train, 2010).

used to create the IV matrix for the model with eleven nearest neighbors as proximate tracts. Because each of the IVs changes over time, there is both temporal and spatial variation in the excluded variables. To account for any nonlinear impacts of our instrument, we also include a quadratic term for the residual in the main model (Papke and Wooldridge, 2008).

IV. Results

Before discussing our primary results, we begin by describing the process to determine the optimal IV model. Based on the nature of our IV strategy, adding more neighbors to generate the excluded variables in equation (10) will net out more of the local variation in the spatially lagged exogenous variables, with the remaining variation contained in the residual representing exogenous variation outside the nearest-neighbor cutoff. In order for this method to work, and the first-stage residuals to serve as an instrument for price in the duration model, the excluded variables must both be significant in the first-stage OLS model and be reasonably excluded from the main duration model.

Table 3 reports the results for the first-stage OLS model in equation (10) for five different nearest-neighbors specifications, ranging from seven to eleven neighbors. Based on these results, it is clear that our instruments pass the first-stage exclusion test in all five models. Previous work (Stock et al. 2002) on 2SLS IV estimation has suggested that, in models with one endogenous regressor, the F statistics should exceed 10 for inference based on the 2SLS estimator. Our F statistics are well above this level suggesting that the residuals from these models serve as a valid instrument for price in the duration model.

To gauge the appropriateness of the residuals as instruments in our duration model, we perform overidentification tests for each of the nearest-neighbor specifications. Following the methodology described in Wooldridge (2010), in addition to the residuals from the first-stage OLS model we add two of the three excluded IVs from the first-stage OLS model to the right-hand side of the duration model and perform Chi-squared tests of the joint significance of these variables.

Wooldridge (2010) shows that this procedure provides a consistent overidentification test in cases where the number of excluded variables in the first stage is larger than the number of endogenous variables. We use the variable on undeveloped area as our excluded IV to perform overidentification tests.

Table 4 presents the results of the overidentification tests for each of the nearest-neighbor model specifications as well as the coefficient values for the house price variable and the IV residuals in the duration model. For each of these models, results are based on nonparametric bootstrapped standard errors (300 replications) with errors clustered at the parcel level.[12] Table 4 shows that for models (1) and (2) we reject the hypothesis that the excluded variables are uncorrelated with the error term in the duration model. In model (3), however, after adding the ninth neighbor to generate the first-stage IVs, the Chi-squared test is no longer significant at the 10% level. Furthermore, as we continue to add more neighbors, the coefficient on price falls and becomes stable in models (4) and (5).[13]

This result is intuitive when we consider the spatial equilibrium nature of our IV strategy. As more neighbors are added, more of the local variation is removed that may be correlated with the error term; however, the residual variation outside the nearest-neighbor cutoff is still correlated with price via spatial equilibrium of the housing market. These results suggest that our method of instrumentation is both consistent with spatial equilibrium theory and statistically valid as an IV

[12] Because the residuals used as IVs in the duration model are estimated values from the first-stage model, it is necessary to bootstrap the standard errors in order to obtain consistent estimates of the standard errors and the covariance matrix in the duration model.

[13] We also estimated our duration model for specifications with 12 or more nearest neighbor, and the estimates were similar to those in the model with 11 nearest neighbors.

strategy in providing a general framework to control for price endogeneity in reduced-form land use models.

The results from both the standard non-IV discrete-time duration model and our preferred IV discrete-time duration model are shown in Table 5. We separate the results into parcel and neighborhood characteristics. Standard errors are, once again, based on a nonparametric panel data bootstrap method sampled at the parcel level. Examining the results from both model specifications, we find similar signs and significance across models with and without instruments and the estimates have the expected signs.

For parcel-level characteristics, the coefficient estimate on distance to a major road is negative and significant, implying that being closer to the highway transportation network increases the probability of development. We find that while the coefficient on distance to the CBD is positive it is statistically insignificant. As expected, we also find that parcels with higher zoned capacity and larger parcel area are significantly more likely to develop, suggesting that economies of scale exist for larger development projects. Parcels located in floodplains are less likely to develop, presumably due to limitations on development in sensitive floodplain areas. Parcels that have soil conditions suitable for septic systems and basements are more likely to develop. Meanwhile, parcels located within municipal sewer service areas are less likely to develop, suggesting an increase in the likelihood of exurban development on septic systems which is a common form of development in our study region.

For time-varying census tract characteristics, we find that an increase in prior approved lots has a positive impact on the likelihood of development while an increase in zoned capacity at the neighborhood level has a negative impact on development. For the lot approvals, the positive sign suggests that increased approvals in the previous period may serve as a signal to other landowners that potential profits are high in a given neighborhood, which increases the probability of conversion. This result is also similar to previous work in the Baltimore region (Towe et al., 2008). For the zoned capacity variable, our results suggest that, after controlling for zoning at the parcel level, increased capacity may signal a reduction in overall neighborhood profitability, which reduces development propensity. The coefficient on preservation is negative in both models but is not significant for the non-IV model and only significant at the 10% level for the IV model. 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.

Finally, our estimated coefficients on the housing price variables are positive and significant, as expected, with the estimate being of larger magnitude in the IV model. Specifically, we find that the estimate in the IV model is over three times larger than the estimate in the non-IV model (Table 5). Moreover, the coefficients on the price residuals are negative and significant indicating a downward bias in the model without instruments, which corresponds with what would be expected in a supply-side model where unobservable factors are negatively correlated with price.

To provide additional context on the role of our IV strategy in identifying the residential land price elasticity, Tables 6 and 7 report implied price elasticity estimates for all of our nearest-neighbor models and the duration model without instrumentation. Table 6 simply converts the parameter estimates from Table 4 to elasticity values. As was the case in Table 4, the values decrease with each additional neighbor and settle to value of 2.06 in the model with 11 nearest neighbors.

Table 7 reports estimates of the implied price elasticity for both duration models with and without instruments. These elasticity estimates reflect a long-run price elasticity of land supply given the long time dimension of our model. For the non-IV model, the implied price elasticity is 0.6747, which is low compared to the long-run price elasticities on housing supply (Gyourko, 2009) and near the bottom range relative to the more recent estimates for price elasticities of housing supply that take account of both land-use regulations and geographic restrictions (Saiz, 2010). This previous literature has placed the long-run elasticity of housing supply in range of 0.6 to 5.45. For our IV model, we find an estimate of 2.0587 for the price elasticity of land supply, which is within the range of estimates in this previous work focused on housing supply. That said, our elasticity measure is not directly comparable to these estimates as we aim to provide a unique estimate for the elasticity of land supply, and not the supply of individual housing units. Moreover, the prior literature on estimates of housing supply elasticity relies on aggregate data, whereas our proposed method has the advantage of being estimated using micro-level data for the actual parcel-level conversion decision of landowners.

V. Discussion and Conclusions

In this paper, we estimate a parcel-level duration model of subdivision development and apply an innovative method of instrumentation based on theory from the urban sorting literature to control for price endogeneity (Bayer and Timmins, 2007). Many previous papers have included some form of price as a variable in parcel-level duration models of land development (e.g., Cunningham, 2007; Towe et al. 2008; Bulan et al. 2009; Wrenn and Irwin, 2014), but have not addressed the potential endogeneity of price and its impact on the model estimation. Our results demonstrate that the potential endogeneity bias in much of this existing work could be substantial.

The results from our model show that, by controlling for price endogeneity, we recover a price elasticity estimate that is more than three times larger than the value estimated in a model without instrumentation. Hence, our results demonstrate that not accounting for the endogeneity of price is likely to provide misleading results when duration models of land development are used for the policy analysis. Given the growing availability of fine-scale parcel and house price data, our method is particularly useful both as a method for controlling for endogeneity in reduced-form land use models and as means of analyzing current and future price-based policies aimed at managing urban growth. Beyond our context, we demonstrate a novel approach to confront endogeneity concerns in the growing body of literature on land supply using nonlinear duration models. Our price instrumentation strategy is generic in the sense that other instruments could be obtained from either traditional exclusionary assumptions or through additional exploitation of the logic of spatial equilibria.

References

Anas, Alex. (1978), "Dynamics of Urban Residential Growth," *Journal of Urban Economics*, 5 (1978), 66-87.

Arnott, Richard, "A Simple Urban Growth Model with Durable Housing", *Regional Science and Urban Economics*, 10 (1980), 53-76.

Bayer, Patrick and Christopher Timmins, "Estimating Equilibrium Models of Sorting Across Locations", *The Economic Journal*, 117 (2007), 353-374.

Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker, "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable", *American Journal of Political Science*, 42 (1998), 1260-1288.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium", *Econometrica*, 63 (1995), 280-296.

Brueckner, Jan K., "Urban Sprawl: Diagnosis and Remedies", *International Regional Science Review,* 23 (2000), 160-171.

Bulan, Laarni, Christopher Mayer, and C. Tsuriel Somerville, "Irreversible Investment, Real Options, and Competition: Evidence from Real Estate Development", *Journal of Urban Economics*, 65 (2009), 237-251.

Capozza, Dennis R. and Robert W. Helsley, "The Fundamentals of Land Prices and Urban Growth", *Journal of Urban Economics,* 26 (1989), 295-306.

Capozza, Dennis R. and Yuming Li, "Optimal Land Development Decisions", *Journal of Urban Economics,* 51 (2002), 123-142.

Cunningham, Christopher R., "House Price Uncertainty, Timing of Development, and Vacant Land Prices: Evidence for Real Options in Seattle", *Journal of Urban Economics*, 59 (2006), 1-31.

Cunningham, Christopher R., "Growth Controls, Real Options, and Land Development", *The Review of Economics and Statistiscs*, 89 (2007), 343-358.

Dixit, Avinash K. and Robert S. Pindyck, "Investment Under Uncertainty", (1994), Princeton University Press.

Epple, Dennis and Holger Sieg, "Estimating Equilibrium Models of Local Jurisdictions", *The Journal of Political Economy,* 107 (1999), 645-681.

Glaeser Edward L. and Matthew E. Kahn, "Sprawl and Urban Growth", In V. Henderson, J. Thisse Edition, *The Handbook of Urban Economics*, vol. IV (2004), North-Holland.

Edward L. Glaeser, Joseph Gyourko, and Raven E. Saks, "Urban Growth and Housing Supply", *Journal of Economic Geography,* 6 (2006), 71-89.

Gyourko, Joseph, "Housing Supply", *Annual Review of Economics*, 1 (2009), 295-318.

Hansen, Andrew J., Richard L. Knight, John M. Marzluff, Scott Powell, Kathryn Brown, Patricia H. Gude, and Kingsford Jones, "Effects of Exurban Development on Biodiversity: Patterns, Mechanisms, and Research Needs", *Ecological Applications,* 15 (2005), 1893-1905.

Hausman, Jerry, "Valuation of New Goods Under Perfect and Imperfect Competition", In R. G. Bresnahan Edition, *The Economics of New Goods*, (1997), University of Chicago Press.

Ingram, Gregory K., Armando Carbonell, Yu-Hung Hong, Anthony Flint, "Smart Growth Policies: An Evaluation of Programs and Outcomes", (2009), Lincoln Institute of Land Policy.

Irwin, Elena G. and Nancy Bockstael, "Interacting Agents, Spatial Externalities and the Endogenous Evolution of Residential Land Use Patterns". *Journal of Economic Geography*, 2 (2002), 31-54.

Irwin Elena G. and Douglas H. Wrenn. "An Assessment of Empirical Methods for Modeling Land Use", In J.M. Duke and J.J. Wu Edition, *Oxford Handbook of Land Economics*, (2014), Oxford University Press.

Klaiber, H. Allen and Nicolai Kuminoff, "Equilibrium Sorting Models of Land Use and Residential Choice", In J.M. Duke and J.J. Wu Edition, *Oxford Handbook of Land Economics*, (2014), Oxford University Press.

Klaiber, H. Allen and Daniel J. Phaneuf, "Valuing Open Space in a Residential Sorting Model of the Twin Cities", *Journal of Environmental Economics and Management*, 60 (2010), 57-77.

Klaiber, H. Allen and Daniel J. Phaneuf, “Linking Developers and Homeowners in a Horizontal Sorting Framework: Implications for General Equilibrium Welfare Measurement”, (2014), Manuscript.

Lewis, David J., Bill Provencher, and Van Butsic, “The Dynamic Effects of Open-Space Conservation Policies on Residential Development Density”, *Journal of Environmental Economics and Management*, 57 (2009), 239-252.

Mayer, Christopher J. and C. Tsuriel Somerville, “Residential Construction: Using the Urban Growth Model to Estimate Housing Supply”, *Journal of Urban Economics,* 48 (2000), 85-109.

McConnell, Virginian, Margaret Walls, and Elizabeth Kopits, “Zoning, TDRs, and the Density of Development”, *Journal of Urban Economics*, 59 (2006), 440-457.

Murphy, Alvin, "A Dynamic Model of Housing Supply", (2014), Manuscript.

Nechyba Thomas J. and Randall P. Walsh, "Urban Sprawl", *The Journal of Economic Perspective*s, 18 (2004), 177-200.

Newburn, David A. and Peter Berck, "Modeling Suburban and Rural-Residential Development Beyond the Urban Fringe", *Land Economics*, 82 (2006), 481-499.

Papke, Leslie and Jeffrey M. Wooldridge, “Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates”, *Journal of Econometrics*, 145 (2008), 121-133.

Petrin, Amil and Kenneth Train, “A Control Function Approach to Endogeneity in Consumer Choice Models”, *Journal of Marketing Research*, 47 (2010), 3-13.

Rivers, Douglas and Quang Vuong, “Limited Information Estimators and Exogeneity Test for Simultaneous Probit Models”, *Journal of Econometrics*, 39 (1988), 347-366.

Saiz, Albert, "The Geographic Determinants of Housing Supply", *The Quarterly Journal of Economics*, 125 (2010), 1253-1296.

Saks, Raven E., “Job Creation and Housing Construction: Constraints on Metropolitan Area Employment Growth”, *Journal of Urban Economics,* 64 (2008), 17-195.

Sieg, Holger, V. Kerry Smith, H. Spencer Banzhaf, and Randy Walsh, “Interjurisdictional Housing Prices in Locational Equilibrium”, *Journal of Urban Economics,* 52 (2002), 131-153.

Stock, James H., Jonathan Wright, Motohiro Yogo, “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments”, *Journal of Business Economics and Statistics*, 20 (2002), 518-529.

Tiebout, Charles, "A Pure Theory of Local Expenditures", *The Journal of Political Economy*, 64 (1956), 416-424.

Towe, Charles A., Cynthia J. Nickerson, and Nancy Bockstael, "An Empirical Examination of the Timing of Land Conversions in the Presence of Farmland Preservation Programs", *American Journal of Agricultural Economics*, 90 (2008), 613-626.

Walsh, Randy, "Endogenous Open Space Amenities in a Locational Equilibrium", *Journal of Urban Economics*, 61 (2007), 319-344.

Wooldridge, Jeffrey M., Econometric Analysis of Cross Section and Panel Data, 2nd Edition, (2010), The MIT Press.

Wrenn, Douglas H. and Elena G. Irwin, Time Is Money: An Empirical Examination of the Effects of Regulatory Delay on Residential Subdivision Development, (2014), Forthcoming: Regional Science and Urban Economics.

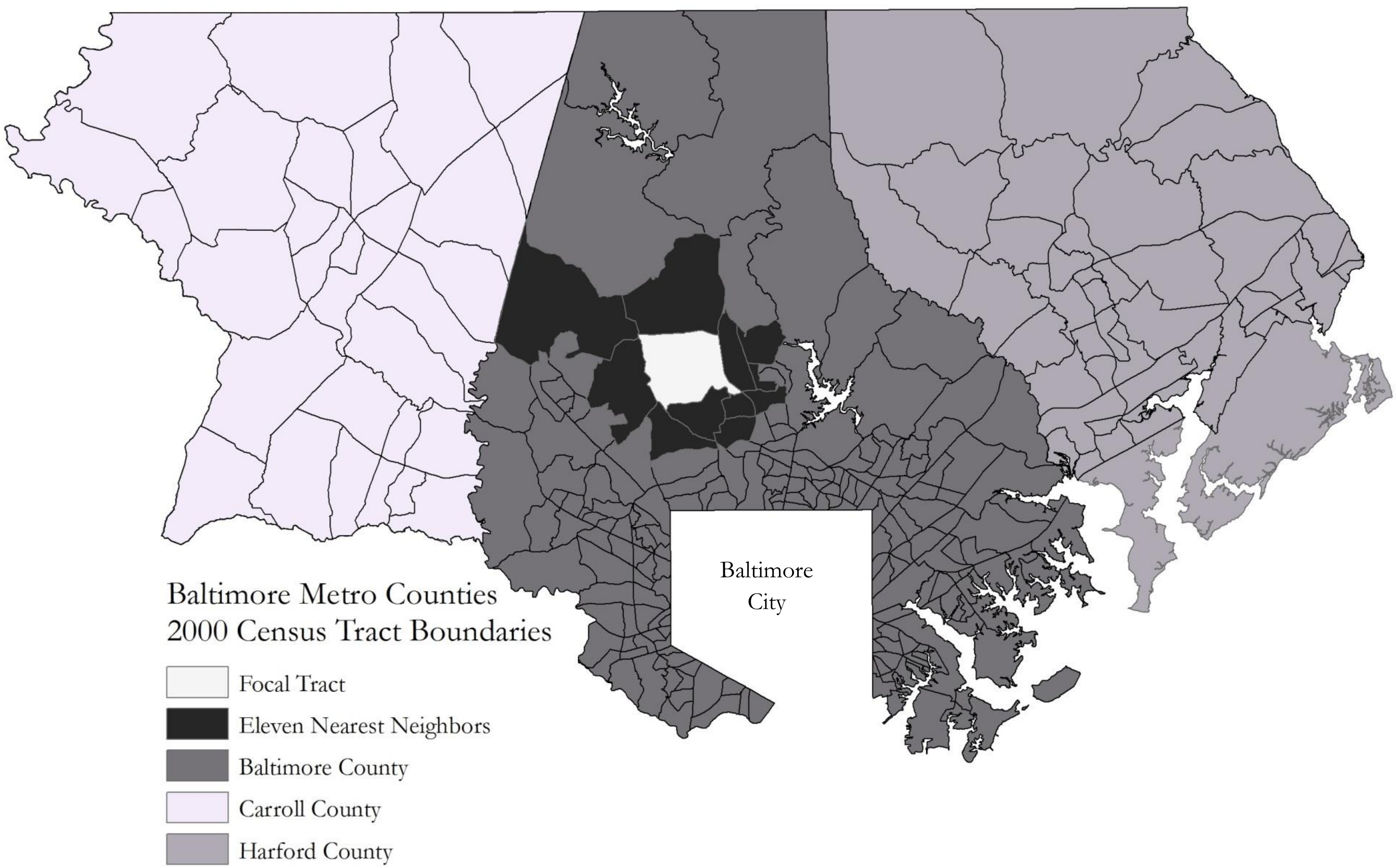

Figure 1: Map of Baltimore Metro Region

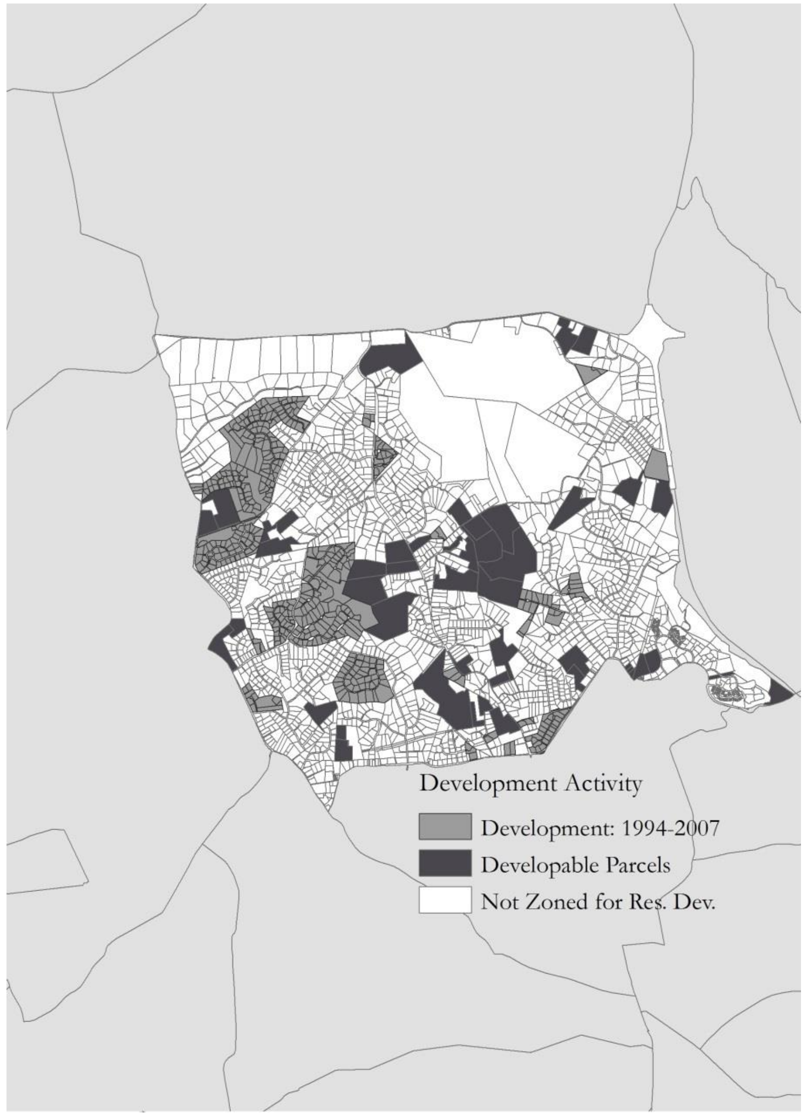

Figure 2: Subdivision Development Activity in 1994-2007 and Parcels Remaining Developable in 2007 in Given Focal Tract

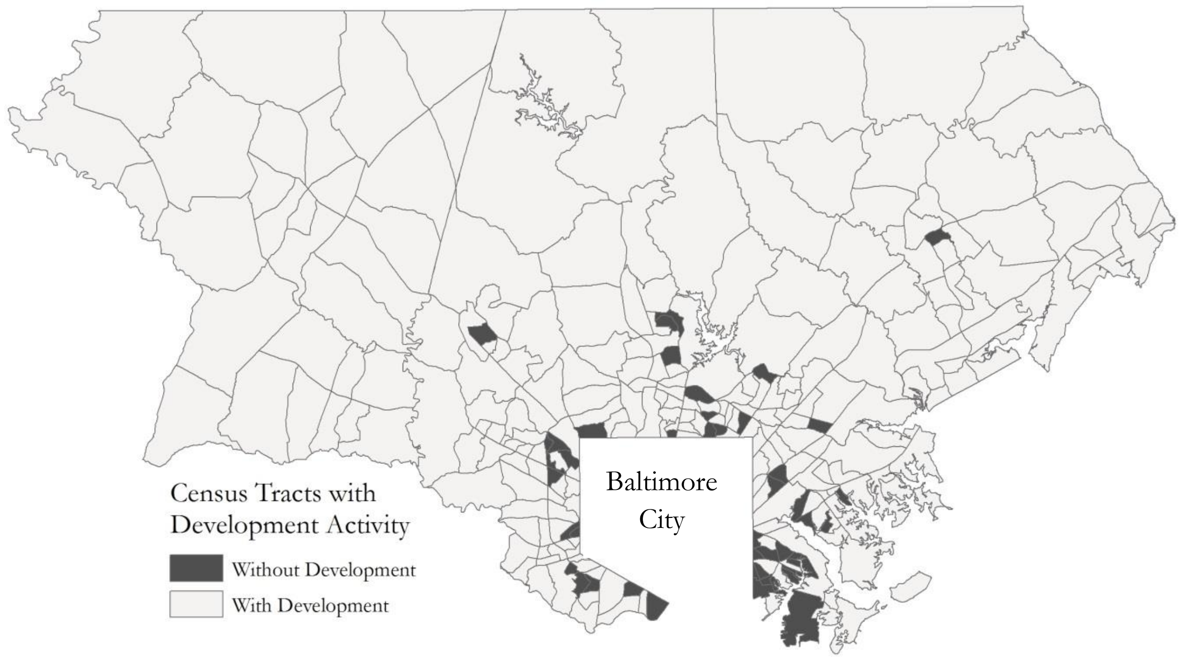

Figure 3: Map of Census Tracts with Subdivision Development Activity between 1994 and 2007

Table 1: Summary Statistics

Variable		Mean	S.D.	Min.	Max.
Parcel Characteristics					
Dist	Kilometers to Baltimore City	31.13	16.67	0	76.31
DistMajRoad	Kilometers to closest major highway	0.37	0.54	0	4.37
Area	Parcel area in acres	19.93	37.12	0.07	946.95
ZndLots	Count of zoned lots allowed	9.65	32.28	2.00	1378.00
Sewer	Indicator for municipal sewer service	0.45	0.50	0.00	1.00
FloodPlain	Located in 100-year flood plain	0.17	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%	10.10	23.47	0.00	100.00
ExHouse	Has an existing structure	0.54	0.50	0.00	1.00
Neighborhood Characteristics					
Preservation	% of neighborhood in preservation	6.22	9.02	0.00	47.99
UDArea	% of neighborhood undeveloped	25.72	14.23	0.20	67.78
ZndLots	Total zoned lot capacity	903.31	737.63	3.00	4808.00
ApprvLots	Count of lots approved - 1-year lag	74.80	83.41	0.00	800.00
LandPrice	In $1,000s per acre	71.59	59.89	7.47	708.93
HousePrice	In $1,000s	128.04	41.95	41.08	370.17
PriceResids	IV regression residuals	1.43	28.83	-111.31	172.19
PriceResidsSqrd	Residuals squared	833.14	1902.81	0.00	29649.19
Baltimore	Located in Baltimore County	0.57	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 parcel characteristcs are from 14,576 parcels that are developable during the 14-year study period; the neighborhood characteristics are from 229 active census tracts. The IV residuals are from the first-stage OLS model (11-nearest-neighbors specification).

Table 2: Summary Statistics for Instrumental Variables

Variable	Mean	S.D.	Min.	Max.
Excluded Instruments				
PreservationAvg	4.47	4.73	0.00	22.02
UDAreaAvg	22.08	10.34	4.60	47.48
ZndLotsAvg	662.56	319.35	104.18	1651.18

Note: Values are for the first-stage OLS (11-nearest-neighbors specification).

Table 3: First-Stage IV Tests

	(1) 7 Neighbors		(2) 8 Neighbors		(3) 9 Neighbors		(4) 10 Neighbors		(5) 11 Neighbors	
HousePrice	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.
Excluded Instruments										
PreservationAvg	1.218 ***	0.203	1.386 ***	0.220	1.718 ***	0.231	1.701 ***	0.238	1.928 ***	0.249
UDAreaAvg	-0.255 *	0.143	-0.178	0.154	-0.253	0.163	-0.035	0.169	-0.322 *	0.176
ZndLotsAvg	-0.011 ***	0.003	-0.017 ***	0.003	-0.020 ***	0.004	-0.028 ***	0.004	-0.027 ***	0.004
	F-Stat	p-value	F-Stat	p-value	F-Stat	p-value	F-Stat	p-value	F-Stat	p-value
F-Statistics	21.13 ***	0.000	28.91 ***	0.000	38.65 ***	0.000	44.92 ***	0.000	53.19 ***	0.000

Note: Coefficients are for the first-stage pooled OLS price regression. All models include tract-level exogenous variables and time and county fixed effects. The excluded instruments are based on weighted average values of the exogenous regressors in neighboring census tracts in each time period. The F-statistics are for the joint hypothesis test that the excluded instruments are significant in the first-stage OLS regression model.

* Significant at 10% level.

** Significant at 5% level.

*** Significant at 1% level.

Table 4: Overidentification Tests

	(1) 7 Neighbors			(2) 8 Neighbors			(3) 9 Neighbors			(4) 10 Neighbors			(5) 11 Neighbors		
Variable	Coef.		St. Err.	Coef.		St. Err.	Coef.		St. Err.	Coef.		St. Err.	Coef.		St. Err.
HousePrice (1K)	0.0080	***	0.0016	0.0076	***	0.0014	0.0067	***	0.0013	0.0062	***	0.0012	0.0062	***	0.0012
Residual	-0.0049	***	0.0016	-0.0046	***	0.0014	-0.0037	***	0.0013	-0.0032	***	0.0012	-0.0032	***	0.0012
ResidualSqrd	-2.9E-05	***	5.8E-06	-2.8E-05	***	5.8E-06	-2.8E-05	***	5.8E-06	-2.8E-05	***	5.9E-06	-2.9E-05	***	6.0E-06
	Chi2		p-value	Chi2		p-value	Chi2		p-value	Chi2		p-value	Chi2		p-value
$\chi 2$ – Statistic for Overidentification	6.53	**	0.0383	5.44	*	0.0660	4.18		0.1238	3.56		0.1688	3.62		0.1665

Note: This table presents the tests of endogeneity based on the second-stage probit model and the Chi-squared overidentification tests (Wooldridge, 2010). The IV for average undeveloped area is the excluded instrument in the overidentification test (2 df). All models include time and county fixed effects. The standard errors are bootstrapped with 300 replications and clustered at the parcel level.

* Significant at 10% level.

** Significant at 5% level.

*** Significant at 1% level.

Table 5: Estimation Results for Duration Models

Variable	Non-IV Model Coef.		Non-IV Model St. Err.	IV Model Coef.		IV Model St. Err.
Parcel Characteristics						
Dist (km)	0.0007		0.0012	0.0012		0.8000
DistMajRoad (km)	-0.0382	*	0.0207	-0.0426	**	0.0208
Area (acres)	0.0005	**	0.0002	0.0005	**	0.0002
ZndLots	0.0021	***	0.0002	0.0021	***	0.0002
Sewer	-0.0666	***	0.0257	-0.0577	**	0.0263
FloodPlain	-0.0545	**	0.0239	-0.0517	**	0.0240
SepticSuit	0.0614	***	0.0186	0.0453	**	0.0186
Slope	-0.0005		0.0003	-0.0006	*	0.0003
ExHouse	-0.0072		0.0185	-0.0039		0.0185
Constant	-2.4457	***	0.0623	-2.8072	**	0.1131
Neighborhood Characteristics						
Preservation (%)	-0.0006		0.0012	-0.0029	*	0.0016
UDArea (%)	0.0004		0.0009	-0.0012		0.0012
ZndLots	-6.2E-05	***	1.7E-05	-4.1E-05	**	1.9E-05
ApprvLots	0.0007	***	9.5E-05	0.0004	***	0.0001
LandPrice (1K)	-0.0008	***	0.0002	-0.0008	***	0.0002
HousePrice (1K)	0.0020	***	0.0003	0.0062	***	0.0012
Residual				-0.0032	***	0.0012
ResidualSqrd				-2.9E-05	***	6E-06
Log-Likelihood	-12340.186			-12322.676		

Note: All models include time and county fixed effects. The IV results are for model with 11 nearest neighbors. The standard errors are bootstrapped with 300 replications and clustered at the parcel level.
N = 183,580
* Significant at 10% level.
** Significant at 5% level.
*** Significant at 1% level.

Table 6: Price Elasticity Estimates

	(1) 7 Neighbors		(2) 8 Neighbors		(3) 9 Neighbors		(4) 10 Neighbors		(5) 11 Neighbors	
	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.
Price Elasticity	2.6625 ***	0.5290	2.5437 ***	0.4715	2.2461 ***	0.4297	2.0767 ***	0.4101	2.0587 ***	0.3917

The standard errors for the price elasticity estimates were calculated using the Delta Method.
* Significant at 10% level.
** Significant at 5% level.
*** Significant at 1% level.

Table 7: Price Elasticity Estimates

	Non-IV Model		IV Model	
	Coef.	St. Err.	Coef.	St. Err.
Price Elasticity	0.6747 ***	0.0880	2.0587 ***	0.3917

The standard errors for the price elasticity estimates were calculated using the Delta Method.

* Significant at 10% level.

** Significant at 5% level.

*** Significant at 1% level.

Appendix: Creation of Housing and Land Price Variables

To create our variable for land price, we select all arms-length land transactions occurring between 1994 and 2007. We further refine these data by excluding any parcels that already had a farmland preservation easement on the property, which may preclude it from being sold for development at the market value or 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 run the following hedonic regression

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

where $rlppacre$ is the real price per acre of land in \$2000 for land parcel l, Par_{lt} is a set of parcel-level controls, and δ and τ are tract 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. Ideally we would run the model in equation (A1) for the transactions in each year and use the tract-level fixed effects as a measure of quality-adjusted land price per acre. However, because of the limited number of individual land sales over our study period, we estimate the model using the full data set and, after controlling for land parcel characteristics, use the time and tract-level fixed effects to predict the mean price per acre of land in each census tract and year. For tracts and years without a sale, we use a distanced weighted average of the predicted values of the closest five tracts in space. Since land is an input in the production of housing, we expect land prices to negatively affect latent profitability.

The data we use to generate our house-price variable also comes from MDPV. As was the case with the land price data, we use only arm's-length single-family housing transactions between

1994 and 2007 and create a set of yearly housing transactions data sets. The entire sample for 1994-2007 has 187,497 individual transactions, after excluding the top and bottom 1% of the sample to remove potential outliers and any transactions that do not appear to be of single-family dwellings, such as multi-family dwellings and commercial structures. We convert the nominal sale price of each house to 2000 dollars using the consumer price index (CPI) for the Baltimore metropolitan region.

Given the sample size of our housing transactions data, we are able to run separate hedonic models for each year to generate our neighborhood-level house price indices. To construct our housing prices indices, we follow Sieg et al. (2002) and estimate a series of hedonic models that permit us to separate out the price of housing services at the neighborhood level from the quantity index of housing that is 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 census tract in which the house is located, and $H_h^{'}$ and ϵ_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 census tract. Repeating the estimation process in equation (A2) for each of the 14 years in our data gives a value for the price of housing services for each census tract and year in our model.[14] This tract and year house price value is used in both our duration model and the first-stage regression as our measure of neighborhood house price.

[14] 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).