CONFRONTING PRICE ENDOGENEITY IN A DURATION MODEL OF RESIDENTIAL SUBDIVISION DEVELOPMENT – APPENDIX

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

June 17, 2016

1. Map of Development Activity and Active Neighborhoods

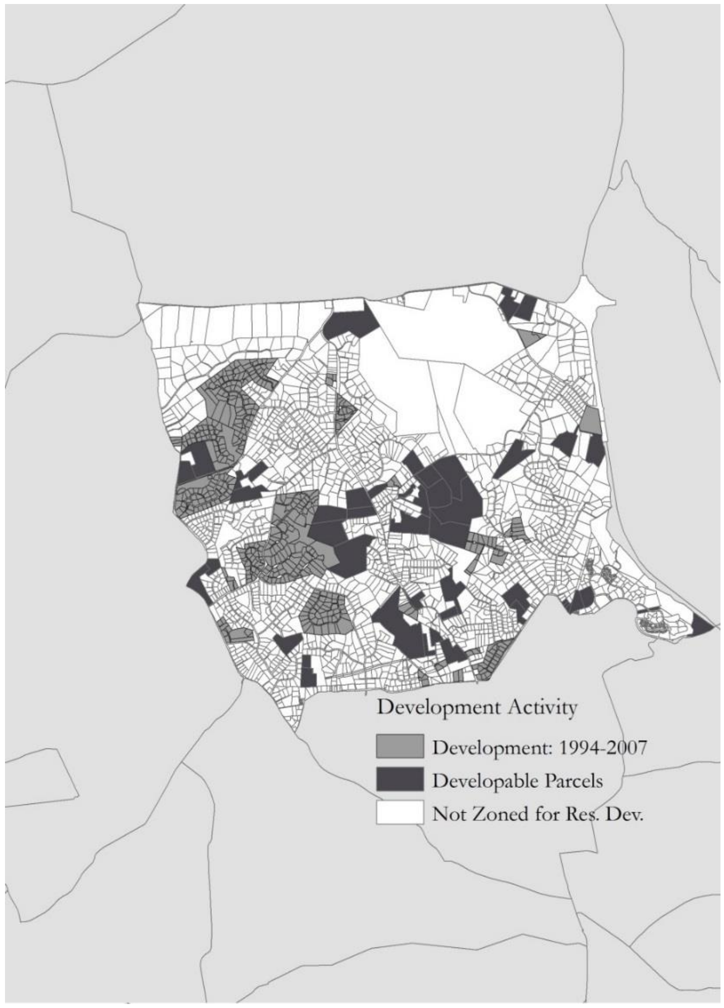

Figure A1. Map of subdivision development activity 1994-2007 and parcels remaining developable at the end of 2007 for the "focal" neighborhood shown in Figure 1. The category "Not Zoned for Res. Dev." includes two different sub categories: (1) development that occurred prior to 1994 and (2) parcels not zoned for residential development.

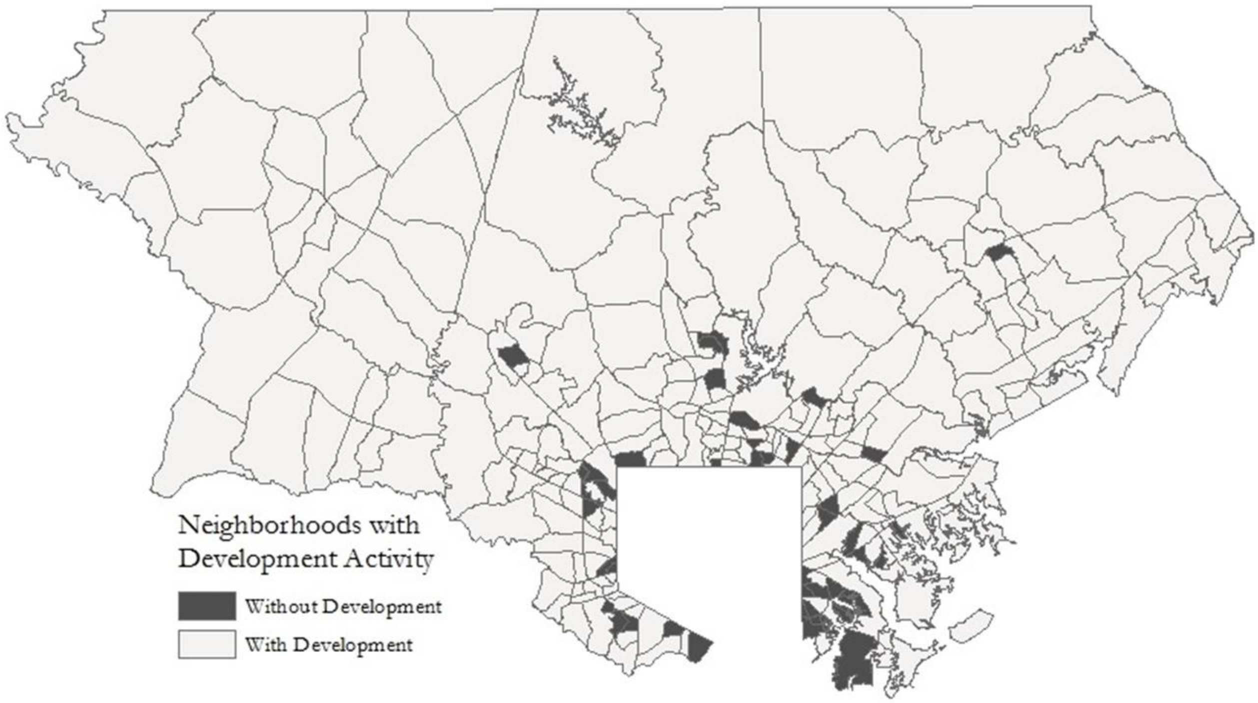

Figure A2. Map of neighborhoods with subdivision development activity during the period 1994-2007. The neighborhoods labeled "With Development" are those used to define the neighborhood boundaries in the duration models.

2. Creation of Housing and Land Price Variables

2.1 *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

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

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 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. 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 tract fixed effects are used to construct the estimate of mean land price per acre in each neighborhood. For tracts and years without a sale we use a distanced weighted average of the values of the tract-level fixed effects for the closest five tracts 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

2.2 *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

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

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 provides a value for the price of housing services for each census tract and year in our model.[1] 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.

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 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. Although as mentioned above, we did a pooled regression for the land price hedonic model due to the smaller number of land transactions over time. This is an important consideration in applying our method in other settings.

[1] 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).

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

3. Additional Robustness Checks

3.1 *Results for IV Models with Instruments Based on Distance Cutoffs*

The results presented in the Table A2 are for a model run with distance-based cutoffs – 3, 4, and 5 miles – used in defining the "local" neighborhoods. These results are in comparison to those in the paper, which are based on a count-of-nearest-neighbor cutoff.

	(1) 3 Miles			(2) 4 Miles			(3) 5 Miles		
	First-Stage Coefficients								
	Coef.		St. Err.	Coef.		St. Err.	Coef.		St. Err.
Neighborhood Characteristics									
Preservation (%)	0.383	***	0.101	0.439	***	0.0927	0.499	***	0.090
UDArea (%)	0.545	***	0.063	0.448	***	0.061	0.479	***	0.059
ZndLots	-0.003	**	0.001	-0.002	*	0.0013	-0.004	***	0.001
Excluded Instruments									
PreservationAvg	-23.319	***	2.961	-7.906	***	2.0014	-2.739	*	1.545
UDAreaAvg	4.239	*	2.228	-9.308	***	1.813	-12.600	***	1.578
ZndLotsAvg	0.364	***	0.048	0.501	***	0.0363	0.497	***	0.034
	Exclusionary Restrictions								
	F-Stat		*p*-value	*F*-Stat		*p*-value	*F*-Stat		*p*-value
	75.20	***	0.000	118.84	***	0.000	121.95	***	0.000
	Overidentification Tests								
	$\chi 2$		*p*-value	$\chi 2$		*p*-value	$\chi 2$		*p*-value
	13.42	**	0.0012	5.89	*	0.0527	1.19		0.5506
	Price Elasticity Values								
	Coef.		St. Err.	Coef.		St. Err.	Coef.		St. Err.
	2.2754	***	0.3376	2.0980	***	0.2776	2.0045	***	0.2753

Note: The instrumental variables are based on the area-weighted average values of the variables specified in the table, and are based on values in "distant" census tracts, or in tracts that are greater than or equal to distances specified in the table. The average values are calculated for each tract in each time period and use tracts in the same year to created the average values. All models include county and time fixed effects. The standard errors are based on a block bootstrap procedure with 500 replications and clustered at the parcel level.

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

Table A3. Results using a distance-based cutoff in generating the instrumental variables

3.2 *Results for non-IV and IV Truncated Poisson Models*

The results presented in the Table A4 are for the price elasticities produced from our truncated Poisson model. The dependent variable in each model is the count of buildable lots for each subdivision. The buildable lots count for each development is generated by counting up the number of buildable lots produced by each subdivision event in our data. Using these data, which are truncated at one given that subdivision events are defined as events that create two or more buildable lots, we estimate a truncated Poisson model using the same covariates as our duration model and instrument for price using the same spatial equilibrium procedure. The results in Table A4 are for the non-IV and IV truncated Poisson models, where the IV model uses the 12-nearest-neighbor specification.

	Non-IV Model		IV Model	
	Coef.	St. Err.	Coef.	St. Err.
Price Elasticity	0.2815	0.2082	1.2958 *	0.7135

Note: The price elasticity values are calculated using a standard marginal effects formula for a truncated Poisson model, and represent the average percentage change in the number of lots created for a small change in price. The standard errors for the price elasticity estimates are calculated using the delta method, and are based on the bootstrapped variance-covariance matrix from the second-stage truncated Poisson model. The instrumental variable results are based on the 12-nearest-neighbor specification

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

Table A4. Price elasticity values from a non-IV and IV truncated Poisson model estimated using lot quantity

3.3 *Standard Errors Clustered at the Neighborhood Level*

The results presented in the Table A4 provide price elasticities using a variance-covariance matrix estimated using a block bootstrap procedure with standard errors clustered at the neighborhood level. These results are analogous to those in Table V, which used block bootstrap standard errors clustered at the parcel level.

	Non-IV		IV 8 Neighbors		IV 9 Neighbors		IV 10 Neighbors		IV 11 Neighbors		IV 12 Neighbors	
	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.
Price Elasticity	0.6747 ***	0.1144	2.6625 ***	0.6611	2.5437 ***	0.5879	2.2461 ***	0.5491	2.0767 ***	0.5175	2.0587 ***	0.4967

Note: This table presents price elasticity values calculated using a standard marginal effects formula for a binary probit model and represent the percentage change in the probability of development for a 1% change in price. The standard errors for the price elasticity estimates are calculated using the delta method and are based on the bootstrapped variance-covariance matrix from the second-stage duration model. The standard errors are clustered at the neighborhood level.
* Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level

Table A5. Price elasticity estimates with standard errors clustered at the neighborhood level