Parcel Size and Land Value: A Comparison of Approaches

Contact Author:

Karl L. Guntermann
College of business
Arizona State University
P. O. Box 874706
Tempe, AZ 85287
Phone: (480) 965-7206
Fax: (480) 965-8629
E-Mail: Karl.Guntermann@asu.edu

Alex R. Horenstein
Department of Economics
W. P. Carey School of Business
Arizona State University
Phone: 480.965.0329
Fax: 480.965.0748
E-mail: alex.horenstein@asu.edu

Gareth Thomas
Quantitative Micro Software
4521 Campus Drive, #336
Irvine, CA 92612-2621
Phone: 949-856-3368
Fax: 949-856-2044
E-mail: gareth.thomas@eviews.com

Submitted to Real Estate Economics

May 2007

The authors wish to thank Mark Ecker and Hans Isakson for making their data available and Seung Ahn and Crocker Liu for valuable comments. The views expressed in the paper are those of the authors and are not representative of the views of Quantitative Micro Software.

Abstract

Studies on non-linearity in the price-area relationship in urban land could be improved with a demonstration that the models being estimated have the flexibility to measure non-linearity if it exists in the data as well as through the use of larger data sets. The analysis presented here uses simulated data to test the ability of parametric and semi-parametric models to accurately reject a hypothesis about the co-existence of plottage and plattage when such a hypothesis is incorrect. The parametric estimator fails with simulated data, making it difficult to draw conclusions about the findings of plottage and plattage using real data. The semi-parametric estimator performs much better with the simulated data. It is 100 percent accurate in judging whether a hypothesis of the co-existence of plottage and plattage is correct or not when applied to the simulated data. However, the method fails when applied to real data because of the chaotic pattern of the estimated price-area relationship. Bayesian regression is proposed as an alternative to traditional approaches. While it may detect both plottage and plattage when they are present in data, it is more difficult and time consuming to use because Bayesian regression is not included in standard statistical packages. However, a combination of parametric and non-linear regression using Box-Tidwell produced results very similar to those from the Bayesian technique proposed in the literature.

Introduction

The appraisal literature has long included the concept of plottage value, which is defined in the Appraisal Institute's text as "The increment of value created when two or more sites are combined to produce greater utility". Because land assembly is not costless, it would occur only if the value of an assembled parcel were at least equal to the sum of the individual parcel values plus the costs associated with assembling them. These costs would include legal and other transaction related expenses, possible changes to infrastructure and higher prices extracted by holdout owners, etc. Land assembly costs would make the plottage portion of the parcel size-value curve convex, reflecting that land values are increasing at an increasing rate with size.

The term "plattage" was first used by Colwell and Sirmans (1978, 1980). It refers to the portion of the size-value curve where subdivision rather than land assembly is possible. Value can be created through the subdivision of land but that process also is not costless. For subdivision to occur, the total value of the smaller parcels would have to exceed their value in the original parcel by at least their subdivision costs, making the size-value curve concave over that range of parcel sizes. It is possible that both plottage and plattage could coexist in a market at the same time but they would be associated with different size parcels. If this is the case, then the relationship between price and parcel size would be "S" shaped, such as in Figure 1, rather than linear. Point A is the inflection point where land values stop increasing at an increasing rate and begin increasing at a decreasing rate. To the left of point A, plottage would occur, while to the right of point A, plattage would dominate.

The Literature

Rigorous analysis of the value-size relationship essentially began with the Colwell and Sirmans (1978, 1980) articles. These papers provided the initial explanation for the existence of an S shaped value-size curve, which ran counter to the prevailing view of a linear relationship between value and size. The models developed by Colwell and Sirmans found evidence for the existence of both plottage and plattage in residential lot data from Champaign, IL and Edinburgh, Scotland. More recent articles by Colwell and Munneke (1999), Thorsnes and McMillen (1998) and Ecker and Isakson (2005) specifically address the relationship between value and size. Tabuchi (1996) and Lin and Evans (2000) find evidence of plottage with data that includes very small parcels.

There is considerable evidence supporting the concave nature of the land value curve for large parcels but the evidence for convexity in the size range where plottage should occur is much weaker. Possible explanations for the lack of evidence include inappropriate model specification and data sets that do not contain sufficiently small parcels. It is possible that plottage exists but that the inflection point is beyond the range of the smallest parcels and, hence, plottage cannot be detected even if the appropriate functional form is used in the tests. Colwell and Munneke (1999) attempt to detect plottage by comparing the degree of concavity in the land price function between the Chicago Central Business District (CBD) and the rest of Cook County. They focus on the CBD because of the greater likelihood that land assembly would be associated with redevelopment projects. The coefficient on the size variable in the models is greater than one for the residential tests, which supports a convex relationship, while the commercial and industrial tests are consistent with a linear relationship. The authors note that the plottage/plattage tests are indirect in the sense that the transactions used have not been identified as being associated with either assembly or subdivision activity. Therefore, the results may reflect a large number of non-assembly sales, making it impossible to detect plottage even though it might be present.

Thorsnes and McMillen (1998) re-examine the relationship between land values and parcel size using a parametric estimator for distance and a semi-parametric estimator for size. The flexibility of the semi-parametric specification permits the function to reflect any shape (linear, convex or concave) in any portion of the size spectrum. Their model is estimated using data on 158 undeveloped residential parcels in the Portland, Oregon area that range in size from less than one-half acre to over 26 acres. Their results are consistent with studies that find the value-size relationship to be concave over the entire range of their data. However, they find no evidence that small parcels sell at a discount or that the value-size relationship is ever convex.

Ecker and Isakson (2005) develop a general model that reduces to the models previously discussed in special situations. It accommodates convexity for small parcels, concavity for large parcels and a non-deterministic change point while accounting for spatial correlation. In their model, all parameters are fit simultaneously from a Bayesian perspective by using Markov Chain Monte Carlo Techniques. Their model is estimated using data on 646 residential, arms-length sales of vacant land in two medium sized, Midwestern urban areas (Cedar Falls and Waterloo, Iowa). Their data contains parcels that range from 640 to 4.6 million square feet.

Location also would be expected to have a significant effect on value and is typically measured using distance variables to proxy for accessibility. Models testing for plottage and plattage include relatively few variables, typically size, distance and time, if the data might reflect changing market conditions. For example, Thornes and McMillen (1998) follow this format while using several distance variables (CBD, freeway interchange and arterial street) and several proxies for location including county dummy variables. It is generally assumed that land value is a negative exponential function of distance and most models specify distance in this way. Value would be expected to decline at a decreasing rate from a central point, which traditionally has been the CBD but a similar value-distance relationship might be associated with other distinct sub-peaks of value such as freeway interchanges, or commercial or industrial centers.

There is evidence that the decline of value with distance may not be as large as previously believed. Colwell and Munneke (1997) show that using price per acre as the dependent variable biases the distance coefficient upward since parcel size tends to increase with distance, while price per acre declines as parcel size increases. Models that allow for nonlinearity between price and size result in a price-distance relationship that declines at a slower rate but the relationship is still negative and significant. The irregularity of the land price gradient with respect to distance has been demonstrated by Colwell (1998) and Colwell and Munneke (2003) through the use of piecewise parabolic multiple regression. The complexity of the price surface supports the use of multiple accessibility measures (Thorsnes and McMillen, 1998) and the expectation that price should decline with distance.

The importance of model specification and functional form in testing for plottage and plattage is clear from previous studies. However, one aspect of this research has never been demonstrated, which is that the models used are capable of accurately rejecting a hypothesis about the coexistence of plottage and plattage when such a hypothesis is incorrect. It is possible that some studies produce inconclusive results not because of data limitations but because their models are incapable of consistently accepting or rejecting correct hypotheses about plottage and plattage. One contribution of this paper is to test the ability of parametric and non-parametric models to reject an incorrect hypothesis using simulated data.

A second issue with many published studies is that the data sets used are relatively small. For example, the original Colwell and Sirmans (1978 and 1980) articles had twenty-six and 102 observations respectively while the Brownstone and DeVany study that “rediscovered” plottage and plattage had 85 observations. Equally important is the size range covered by the data in the empirical tests. While data to test the plattage portion of the curve are readily available, it is more difficult to gather data on small parcels to test for plottage, especially in commercial parcels. The section containing the empirical tests uses a large data set of transactions covering a wide range of parcel sizes from two different metropolitan areas.

Estimation Procedures

The three procedures that form the basis of our analysis are the parametric model used in Colwell and Sirmans (1978), the semi-parametric model given in Thorsnes and McMillen (1998) and the Bayesian approach by Eckerd and Isakson (2005). In the first two models we include an intercept term and a log-linear relationship between distance and land value. Thus the only way in which these two models differ is through their treatment of the relationship between land value and size. When testing the Bayesian model, we also include a term for time of sale since the data spans 20 years.

Colwell and Sirmans' parametric model gives an exact form for the relationship between land value and area. This function is as follows:

$$\ln\left[y_i\right] = \ln\left[\beta_1^p\right] + \beta_2^p D_i + \beta_3^p \left(A_i - s\right)^{1/3} + \varepsilon_i \quad (1)$$

Where:

y_i = The price of the property

D_i = The distance of the property from a central urban point

A_i = The area of the property

s = A shift parameter.

And the superscript p denotes a parameter in the parametric model.

This function (1) allows a relationship between land value and area that can exhibit both plottage and plattage. Land value is essentially an inverse cubic function of land area shifted by the shift parameter, s. This shift parameter allows for the inflection point in the curve and represents the point at which the first derivative of land value with respect to area is at its maximum. If s has a value less than the minimum value of A_i then the model would only exhibit plattage. If s has a value greater than the maximum value of A_i then the model would only exhibit plottage. For values of s that lie within the range of values of A_i both plottage and plattage exist.

This function is estimated with respect to $\{\beta_1^p, \beta_2^p, \beta_3^p\}$, for a given s. This estimation procedure is undertaken for a variety of different *s*, and the value of *s* that maximizes r^2 is the one used. The set of *s* used runs from slightly less the smallest value of A_i in the data set to slightly greater than the largest value of A_i. If the value of *s* that maximizes r^2 is either less than the smallest value of A_i or greater than the largest value of A_i then the theory that plottage/plattage co-exist can be rejected.

Colwell and Sirmans suggest that a further test of plottage/plattage can be undertaken by comparing this model to a model capable of detecting only plottage or plattage but not both simultaneously. This model is given by:

$$\ln[y_i] = \ln[\beta_1] + \beta_2 D_i + \beta_3 \ln[A_i] + \varepsilon_i \tag{2}$$

Here if β_3 takes a value less than one but greater than zero then the model includes only plattage, if β_3 takes a value greater than one then the model includes plottage and if β_3 is equal to one then land value is proportional to area. This model can be estimated through simple ordinary least squares. A J-test (see Davidson and Mackinnon (1981)) can then be undertaken to compare the two models, using the first regression as the null hypothesized true model. The fitted values from the second regression are included as an additional regressor in performing the J-test so that the estimated equation becomes:

$$\ln[y_i] = \ln[\beta_1] + \beta_2 D_i + \beta_3 (A_i - s)^{1/3} + \gamma Y^{est} + \varepsilon_i \tag{3}$$

where Y^{est} represents the fitted values of ln(*Price*) from the second regression. If the first model is correct, then the fitted values from the second regression should have no explanatory power on ln(*Price*) when they are included in model 1 and γ should be statistically insignificant (i.e. equal to zero). If γ is statistically significant, then the theory of co-existence of both plottage and plattage can be rejected.

The semi-parametric model of Thorsnes and McMillen estimates the following relationship:

$$y_i = g(A_i) + \beta_1^{sp} D_i + \varepsilon_i \tag{4}$$

Where the superscript *sp* stands for semi-parametric, $g(A_i)$ is some unknown function. This model is estimated using a kernel based non-parametric estimator for $g(A_i)$. Thorsnes and McMillen recommend a number of different kernels, although in our paper we restrict ourselves to the Gaussian Kernel[1]. This non-parametric estimation procedure attempts to estimate $g(A_i)$ on an observation-by-observation basis. Since there are no parameters involving A_i to be estimated, this procedure only provides point estimates and standard errors for the distance parameter, D_i. However it is still possible for this procedure to provide conclusions on the existence of plottage/plattage.

Thorsnes and McMillen suggest one possible process to test for plottage/plattage would be to look at the first and second derivatives of $g(A_i)$. They first take the average of the second derivatives, $\frac{1}{n}\sum_{i=1}^{n} g''(A_i)$, and test whether it is significantly different from zero. If this average is not significantly different from zero, then the hypothesis of a linear relationship cannot be rejected. Thorsnes and McMillen also take the average of the first derivatives, $\frac{1}{n}\sum_{i=1}^{n} g'(A_i)$ and observe whether this value is less than one (indicating a concave price-area relationship), greater than one (indicating a convex relationship), or equal to one (indicating a linear relationship).

By observing only the average of the first and second derivatives of $g(A_i)$ Thorsnes and McMillen are only able to make statements on the average shape of the function and are unable to test whether both plottage and plattage co-exist. They recommend use of averages since "the non-parametric estimator does not provide accurate point estimates of the second derivatives"[2]. However our simulations, admittedly generated with both well-behaved data and a well-behaved function, suggest otherwise. Thus, we propose that the existence of both plottage and plattage in the data could be observed by looking at the point estimates of the first derivative. If both plottage and plattage exist, the first derivative of price with respect to area should start at a value less then one, increase with area to a value greater than one, and then decrease back to a value less than one. Although we are unable to suggest a statistical test of whether this pattern holds,

---

[1] The Gaussian Kernel has the advantage of having continuous derivatives and is more easily implemented. Thorsnes and McMillen state that results do not change significantly with a change in kernel.

[2] Thorsnes and McMillen (1998) pp 237.

we can suggest that a simple observation of the point estimates of the first derivatives may give a good indication of whether both plottage and plattage co-exist in the data.

Ecker and Isakson (2005) use a Bayesian procedure to estimate at the same time 1) the regression parameters and the spatial correlation structure, 2) the optimal change point from convexity to concavity, and 3) the functional form for the convex part (small lots) and concave part (large lots).

With the Bayesian approach all unknown parameters are regarded as random variables. Using Bayes Rule, an a priori functional form of the distribution is updated using the observed data to estimate the ex-post distribution of the unknown parameters. This is the main difference with the classical econometric procedures where the random variable is the error between the real value of the unknown parameter and the estimated one, not the parameter itself.

Ecker and Isakson suggest that the large sample can be tested through the following model:

$$\ln\left[y_i^l\right] = \beta_0 + \beta_1 t_i + \beta_2 h_l(A_i) + \sum_{k=3}^{p} \beta_k x_{ik} + \varepsilon_i^l \tag{5}$$

Similarly, the small sample part should be tested by the following model:

$$\ln\left[y_i^s\right] = \alpha_0 + \alpha_1 t_i + \alpha_2 h_s(A_i) + \sum_{k=3}^{p} \alpha_k x_{ik} + \varepsilon_i^s \tag{6}$$

Where:

$h_l(A_i)$ is a function of the large sample parcel area like $h_l(A_i) = \log(A_i)$ in the standard plattage model.

p is the number of covariates

t_i refer to time of sale

x_i is the rest of covariates like zoning status and relative distances.

The subscript *s* refers to small parcels and the subscript *l* refers to large parcels

For the functions *h* they suggest a more general form. More precisely, they suggest using a Box-Tidwell function for the small area and large area part of the sample of the following form:

$$h_l(A_i, \lambda_l) = \begin{cases} \dfrac{A_i^{\lambda_l} - 1}{\lambda_l} & \text{if } \lambda_l \neq 0 \\ \log(A_i) & \text{if } \lambda_l = 0 \end{cases}$$ for the large area and,

$$h_s(A_i,\lambda_s)=\begin{cases}\dfrac{A_i^{\lambda_s}-1}{\lambda_s} \text{ if } \lambda_s\neq 0\\ \log(A_i) \text{ if } \lambda_s=0\end{cases} \text{ for the small area.} \quad (7)$$

The parameter λ is estimated from the data. The second derivative of $h(A_i,\lambda)$ with respect to A reveals convexity when $\lambda>1$ and concavity when $\lambda<1$. Note that when $\lambda=0$ the functional form of the model is identical to the "classical" logarithmic form of equation (2). It is claimed that "only the Bayesian technique allows one to assess and test for plottage (through a posterior probability) in the presence of both spatial correlation and a non-deterministic change point"[3].

Monte Carlo Simulations for the Parametric and Semi-parametric Procedures

In order to compare the ability of both the parametric procedure of Colwell and Sirmans (1978) and the semi-parametric procedure of Thorsnes and McMillen to detect the presence of both plottage and plattage, a number of Monte Carlo simulations were undertaken.

The basis of the simulations is the following model:

$$\ln[y_i]=\ln[\alpha_1]+\beta X_i+\alpha_2(A_i-s)^{1/p}+\varepsilon_i$$

This model closely resembles the parametric model used by Colwell and Sirmans in their estimation procedure. However we allow ρ to take differing values rather than restricting it to 3. With ρ equal to an odd integer, this model allows for plattage (if s is less than the minimum value of A_i), plottage (if s is greater than the maximum value of A_i) or both (if s is between the minimum and maximum values of A_i). Increasing the value of ρ creates a more dramatic point of inflection as indicated in Figures 2 and 3. Although this model allows for multiple X variables, our simulations indicate that our results do not change as we vary the number of variables, and so we only report results for the inclusion of one X variable, namely Distance.

In our simulations the exogenous variable X_i is generated from a (0-1) uniform distribution and A_i from a (5-25) uniform distribution. ε_i is generated from a standard normal distribution. We chose parameter values of $\alpha_1=0.5,\ \beta=0.8,\ \alpha_2=0.15$, and non-reported simulations indicate that changing these values does not affect the results. Therefore, having generated X, A_i, and ε_i, y_i is generated from the following:

[3] Eckerd, M. and Isakson, H. (2005), pp. 267

$$\ln[y_i] = \ln[0.5] + 0.8X_i + 0.5(A_i - s)^{1/p} + \varepsilon_i$$

Previous studies testing for the presence of a non-linear price-area relationship have tended to use small to medium sized samples. For example, Colwell and Sirmans (1978) use 26, Colwell and Sirmans (1980) use 102, and Thornses and McMillen use 158, and so in our simulations we choose three different sample sizes, one small (20), one medium sized (100) and one large (1000). The large sample is used as a method to compare procedures asymptotically, while the medium and small samples are used to judge how the procedures may differ in a real data situation.

In order to create data with just plottage, just plattage and both, we choose three values for *s:* 1, 15 and 30. Note that 1 is less than the minimum possible value of A_i (5) and 30 is greater than the maximum possible value of A_i (25). The value 15 is used since it is the midpoint of the range of possible values of A_i. We also choose three values for *ρ:* 3, 5 and 9. With each simulation we perform the two estimation procedures outlined in Section 3. The parametric procedure estimates Model (1):

$$\ln[y_i] = \ln[\beta_1^p] + \beta_2^p D_i + \beta_3^p (A_i - s)^{1/3} + \varepsilon_i$$

and the point estimate of the parameters (β_1^p, β_2^p, β_3^p and the value of s that maximized r^2), their bias (i.e. the value of the estimate minus the true value) and the mean square error (MSE), their standard errors (not available for s), the r^2, the result of the J-test (simple reject or non-reject of the model incorporating both plottage and plattage at a 5% significance level), and whether the estimated s lay within the minimal and maximal values of A_i are recorded for each simulation. In place of the mean of the standard errors for s we report (in parenthesis) the standard deviation of the point estimates of s through the simulations.

The semi-parametric procedure estimates model (4). For each simulation we record the point estimate of β_1^{sp}, its bias and MSE and its standard error, the r^2, and if the semi-parametric estimator of the first derivatives follows a pattern of starting at a value less than one, increasing to a value greater than one, then decreasing to a value less than one.

For each set of parameter values and sample sizes we undertake 1,000 simulations. The results of the simulations are reported in Tables 1-3. The first conclusion that can be drawn from Tables 1-3 is the clear accuracy of the parametric estimator <u>for all the sample sizes</u> when the data are generated from a function that matches the function the parametric model estimates, i.e. when s=15 and ρ=3. Under this data generating process, the point estimates of all three parameters are very close to the true values, even in the small sample size. Both the J-test and the Min/Max (which rejects the co-existence of both plottage and plattage if the estimate of s is less than the minimum value or greater than the maximum value of A_i) test are able to non-reject the true hypothesis the vast majority of the time, although in the small sample size neither are able to non-reject 100% of the time, and in the large sample size only the Min/Max test can non-reject 100% of the time. The second conclusion, perhaps more interesting, is that as ρ increases to 5 or 9 the performance of the parametric estimator does not drop significantly. Even with ρ=9 the point estimates are close to the true values and both the J-test and the Min/Max test are able to non-reject the majority of the time.

Where the parametric estimator does begin to fail, however, is when the s parameter is outside the range of possible values of A_i. When this occurs, the point estimates of β_1^p and β_3^p start to become inaccurate, even in the large sample and the estimates have a large bias. The estimate of β_2^p, which is log-linear in price and unrelated to the relationship between area and price, remains unaffected. Also of note is the fact that the point estimate of s is drastically inaccurate. The parametric estimator, unsurprisingly, seems incapable of estimating s if the true value lies outside of the data range. The estimates are distributed widely and take a wide range of values, as shown by the increase in the standard deviation of the estimates of s. Given this, it is unsurprising that both the J-test and the Min/Max test are unable to reject the incorrect hypothesis of the co-existence of both plottage and plattage the majority of the time.

The semi-parametric estimator fairs much better. The point estimate of β_1^{sp} from the semi-parametric estimator remains accurate for all parameter values and sample sizes. More importantly, it is able to correctly non-reject a true hypothesis 100% of the time, and to reject an incorrect hypothesis 100% of the time. It should be noted, however, that both the data generating function, and the data generated were both “smooth” and “clean”. The function is of a perfect *s* shape, and there were no measurement errors or other irregularities in the data. When using real world data this is not the case, and the method used in the simulations to judge the co-existence of both plottage and plattage from the semi-parametric estimator may not be useable.

Empirical Analysis and Comparison

While the simulation analysis suggests that the semi-parametric technique is more accurate than the parametric model used by Colwell and Sirmans, it is more challenging to test the models on actual data. In addition to the two models used in the simulation tests, the Bayesian model used by Ecker and Isakson will also be estimated using land transactions data. Testing the models with data from both Arizona and Iowa will allow for a comparison of model properties that would not be possible with just one database.

The Arizona database contains information for 444 residential and 144 industrial land parcels from 1998 and 1999 which were obtained from Costar/Comps, Inc. An additional 134 improved single-family lot transactions were provided by Marketron, Inc. The lot data was included to provide additional sales of smaller parcels and to extend the range of the smallest parcels in the Costar/Comps, Inc. database. Residential parcel sizes ranged from 18,865 to 41,327,948 sq. ft. while the additional improved lots ranged from 3,851 to 65,993 sq. ft. Part of the value of the databases is the detailed information that each company adds to the transaction. For example, in this study the longitude and latitude of each parcel was used to measure distance to various centers of industrial and commercial employment and retail activity (a proxy for residential location) as well as to a central location in downtown Phoenix. Distance was calculated from each parcel to the various central points appropriate to its zoning category. The value assigned to the Distance variable for each sale was the minimum calculated distance. The data used includes only the sale price, size of the parcel in square feet and distance in miles. Table 4 contains summary statistics for the Arizona data used in the regressions.

The Iowa database used by Ecker and Isakson (2005) contained 646 residential, arms-length sales of vacant land in and around Cedar Falls and Waterloo Iowa obtained from a public source. The sales span the period from 1980 to 2000 during which land values appreciated at a slow but steady rate. The location of the parcels is known because the state plane (x-y) coordinates of the centroid of each parcel were collected and geo-coded. Distances to the two CBDs are used as covariates. Table 5 contains summary statistics for the data used in the regressions.

Results with Arizona Data

The results from the estimation procedures are presented in Table 6. The residential regression includes a dummy variable to control for differences between the databases (improved lots versus raw land) with Lot=1 for improved lots and 0 for raw land parcels. The explanatory power of the models is quite good and is consistent with recent studies on urban land prices such as Colwell and Munneke (1997, 1999) but slightly lower than the R^2 reported by Thornes and McMillen (1998). The coefficients on the distance variables cannot be directly compared to the distance variables in similar studies, which measure distance from the CBD, because of the way distance is defined here. However, the coefficients have the correct sign and their magnitude is plausible. They are also fairly consistent between the parametric (β_2^p) and the semi-parametric models (β_1^{sp}). The coefficients indicate a decline in value of 8.9 percent per mile for the industrial land.

The coefficients for size (β_3^p) in the parametric models are statistically significant. The s values (in thousands of square feet) that produced the maximum R^2 are reported for each land use category. For industrial land, s=20,000 square feet, which is below the smallest industrial parcel (Min/Max range in 10,000s is 2.54/1,624). This means that there is no evidence of plottage over the range of our data, ie. the inflection point, if there is one, would be associated with parcel sizes below 20,000 square feet. Plattage or the benefits of subdivision occurs over the entire range of our data. The fitted values of ln(Price), Y^{est}, are used in estimating model 3, the J-test for plottage and plattage. If plottage and plattage exist, γ should equal zero. For the industrial data, the J-test rejects the existence of plottage and plattage.

The results are different for the residential data where s=13,056 square feet, which is above the minimum parcel size of 3,851 square feet, suggesting that the data demonstrates both plottage and plattage. However, the J-test also rejects the coexistence of plottage and plattage in the residential data and we cannot conclude that concavity and convexity are both present. Instead the results support the presence of plattage throughout the range of the data, which may mean that only plattage exists in the Arizona data or that plottage and plattage coexist in the data but the parametric model is unable to correctly identify the inflection point.

Unfortunately, there is no comparable J-test that can be used for the semi-parametric results in Table 6. The method we suggest earlier of checking the estimates of the first derivatives of $g(A_i)$ fails to lead to any conclusions when used with the real data since visual inspection of the estimates yields no obvious patterns in the derivatives, as was the case with the simulated data. However, the semi-parametric estimates of $g(A_i)$ for the industrial and residential land are graphed in Figures 4-5 where they are subject to visual interpretation. The curves are irregular as the Gaussian kernel attempts to fit a function to the data. The industrial graph (Figure 4) appears to be convex, particularly over the larger parcel sizes. In contrast, the residential graph of $g(A_i)$ (Figure 5) appears to reflect a more linear relationship between value and size, except for the smallest and largest parcels.

Results with Iowa Data

Ecker and Isakson (2005) use a Bayesian approach with data from Cedar Falls and Waterloo Iowa to simultaneously estimate the inflection point *s* and the shape of the value-size functions to the right and left of the inflection point. Since their data covered twenty years, a time variable was included in the models along with distance variables to the CBD of each city. The following two models are estimated simultaneously:

$$\ln\left[y_i^l\right] = \beta_0 + \beta_1 t_i + \beta_2 DistCF + \beta_3 DistW + \beta_4 \frac{A_i^{\lambda_l} - 1}{\lambda_l} + \varepsilon_i^l \quad (8)$$

$$\ln\left[y_i^s\right] = \alpha_0 + \alpha_1 t_i + \alpha_2 DistCF + \alpha_3 DistW + \alpha_4 \frac{A_i^{\lambda_s} - 1}{\lambda_s} + \varepsilon_i^s \quad (9)$$

where the subscript *l* is for large parcels and *s* is for small parcels. The Box-Tidwell function for the small and large size portions of the sample estimates λ_l and λ_s along with *s* using the Bayesian technique. The data revealed a convex value-size relation for parcels smaller than 3,516 square feet (λ_s = 1.543), which became concave for larger parcels (λ_l = 0.2588). The signs of the coefficients for the distance variables are negative as expected. Their results are presented in Table 7 only for the larger parcels, which represented all but 22 of the 646 Iowa sales. Ecker and Isakson also estimated the (posterior) probability that convexity is present in the data at 86.9%.

In order to compare techniques, the parametric and semi-parametric models have been estimated with the Iowa data and those results are also presented in Table 7. With the Iowa data the parametric model detects an inflection point at 2,618 feet, which is only slightly smaller than the 3,516 estimated with the Bayesian approach. The inflection point is robust to any size of the shift parameter (s) included in the regression and the estimate is within the 95% Bayesian confidence interval. Unlike the inflection point estimated with the Arizona data, the J-test (p-value=0.3034 does not reject the null hypothesis that plattage and plottage coexist in the Iowa data. Since the Iowa data contains parcels considerably smaller than the Arizona data (640 versus 3,851 square feet), and well below the inflection points estimated with either model, it is likely that the inability of the parametric model to detect plottage in the Arizona data was because the minimum parcel size was too large.

The results from the semi-parametric regression using the Iowa data (Table 7) are better than those obtained with for Arizona but they are still inconclusive. Specifically, the first derivative behaves chaotically. The graph of $g(A_i)$ in Figure 6 appears to indicate a basically linear relationship between price and size but there is some evidence of convexity and concavity for the smallest and largest parcels.

Use of the Bayesian technique for regression analysis can be more cumbersome and time consuming because it is unavailable in standard econometric packages. As a simplification, we use a two-step process combining the parametric model with non-linear least squares and compare those results to the more general form proposed by Ecker and Isakson. As previously reported, the parametric model produced an estimate of s=2,618 square feet with the Iowa data. However, since there are only 9 observations smaller than *s*, the Box-Tidwell functional form could be used to obtain parameter values only for the plattage portion of the curve. The non-linear results are presented in Table 7 next to the Bayesian results for large parcels. The non-linear λ is not significantly different from zero, which denotes a logarithmic functional form for the relationship between size and price. For the Bayesian model, $\lambda = 0.2588$ and this is consistent with the original Colwell and Sirmans inverse cubic function (equation 1) for the parametric portion of the comparison.

Conclusions

The results from the simulation part of the study raise doubts about the ability of the parametric estimator to accurately estimate the inflection point (*s*) if its true value is outside of the data range. The point estimates of s vary widely if the true value is greater than the maximum value or less than the minimum value of the Area variable. This occurs consistently no matter what value the curvature for the size relationship (ρ) takes, or what size sample is generated. With such inaccuracy of the estimate of s neither the Min/Max approach to judging whether plottage and plattage co-exist in the data, nor the J-test approach are able to successfully reject a hypothesis of co-existence when such a hypothesis is incorrect. Given this, it is difficult to draw a conclusion about the existence of plottage and plattage from a data set using the parametric estimator when the inflection point is outside the data range. However, an inflection point that lies within the data range may be accurate, suggesting there is an inflection point in the price-area relationship. We found a possible inflection point for the Arizona residential data within the data range but the J-test was insignificant. However, the failure of the parametric model to detect an inflection point probably is the result of problems with the data.

The semi-parametric estimator performs much better in the simulations. By observing the pattern of the estimates of the first derivatives it is 100% accurate in judging whether a hypothesis of co-existence of plottage and plattage is correct or not. However, when applied to the real data sets this method fails, due to the chaotic pattern of the first derivatives. Thus, we are only able to graph the estimate of the size-function of the semi-parametric model "$g(A_i)$". With the residential and industrial datasets, this graph would seem to suggest a convex price-area relationship.

The Bayesian technique detected plottage and plattage in the Iowa data. However, Bayesian regression is difficult to use at this time since it is not included in standard econometric packages. Since the parametric model is generally reliable at estimating an inflection point if it is within the range of the data, it can be combined with non-linear least squares using the Box-Tidwell functional form. When applied to the Iowa data, the combined approach produced a similar inflection point and parameter estimates for the plattage portion of the curve, offering reliable yet much faster and easier techniques to use than the Bayesian methodology.

Table 1 Small Sample Simulation Results

	Parametric							Semi-Parametric		
	β_1^p	β_2^p	β_3^p	S	r^2	Rejections		β_1^{sp}	r^2	Rejections
	0.5	0.8	0.15			J-test	Min/Max	0.8		
N=20										
s = 15, ρ=3	0.5018	0.8149	0.2105	14.17	85.17	11.0	6.9	0.8152	83.33	0.00
SE	0.0406	0.1117	0.0256	(2.21)				0.1693		
Bias	0.0018	0.0149	0.0605					0.0152		
MSE	0.0905	0.0207	0.0277					0.0218		
s = 15, ρ=5	0.4849	0.8130	0.1511	14.30	83.02	11.3	6.3	0.8143	80.08	0.00
SE	0.0337	0.1109	0.0226	(2.42)				0.1690		
Bias	-0.0151	0.0130	0.0011					0.0143		
MSE	0.0220	0.0196	0.0091					0.0211		
s = 15, ρ=9	0.4838	0.8123	0.1273	14.39	81.25	14.7	5.3	0.8138	78.1	0.00
SE	0.0326	0.1113	0.0221	(1.92)				0.1689		
Bias	-0.0172	0.0123	-0.0227					0.0138		
MSE	0.0109	0.0191	0.0059					0.0210		
s = 1, ρ=3	0.6932	0.8057	0.0553	13.31	87.51	3.3	6.7	0.8064	77.87	100
SE	0.0339	0.0713	0.0175	(6.36)				0.0713		
Bias	0.1932	0.0057	-0.0947					0.0057		
MSE	0.0459	0.0080	0.0105					0.0080		
s = 1, ρ=5	0.6304	0.8070	0.0251	14.34	84.18	9.4	1.90	0.8081	75.37	100
SE	0.0333	0.0806	0.0192	(5.27)				0.1646		
Bias	0.1304	0.0070	-0.1249					0.0081		
MSE	0.0204	0.0099	0.0168					0.0113		
s = 1, ρ=9	0.6001	0.8111	0.0097	14.80	82.52	8.2	1.0	0.8087	74.13	100
SE	0.0339	0.0856	0.0205	(5.82)				0.1654		
Bias	0.1001	0.0111	-0.1403					0.0087		
MSE	0.0131	0.0110	0.0211					0.0130		
s = 30, ρ=3	0.3351	0.8394	0.0556	14.89	62.71	5.20	2.00	0.8340	61.91	100
SE	0.0352	0.1581	0.0376	(6.01)				0.1812		
Bias	-0.1649	0.0394	-0.0946					0.0340		
MSE	0.0327	0.0422	0.0140					0.0486		
s = 30, ρ=5	0.3776	0.8252	0.0243	14.94	66.73	6.10	0.70	0.8259	64.34	100
SE	0.0336	0.1386	0.0323	(5.96)				0.1758		
Bias	-0.1224	0.0252	-0.1257					0.0259		
MSE	0.0350	0.0301	0.0193					0.0350		
s = 30, ρ=9	0.3951	0.8292	0.0098	14.70	67.91	0.00	0.00	0.8194	63.56	100
SE	0.0337	0.1332	0.0309	(5.82)				0.1740		
Bias	-0.1049	0.0292	-0.1402					0.0194		
MSE	0.0133	0.0243	0.0225					0.0293		

Table 2 Medium Sample Simulation Results

	Parametric							Semi-Parametric		
	β_1^p	β_2^p	β_3^p	*S*	R^2	Rejections		β_1^{sp}	R^2	Rejections
	0.5	0.8	0.15	15		J-test	Min/Max	0.8		
N=100										
s = 15, ρ=3	0.4869	0.8225	0.1554	14.92	83.43	9.60	0.40	0.8223	82.80	0.00
SE	0.0151	0.0539	0.0095	(0.89)				0.0760		
Bias	-0.0131	0.0225	0.0054					0.0223		
MSE	0.0011	0.0040	0.0012					0.0042		
s = 15, ρ=5	0.4884	0.8214	0.1231	14.95	80.81	18.00	0.20	0.8213	79.95	0.00
SE	0.0148	0.0539	0.0092	(0.60)				0.0758		
Bias	-0.0116	0.0214	-0.0269					0.0213		
MSE	0.0007	0.0038	0.0010					0.0040		
s = 15, ρ=9	0.4901	0.8146	0.1069	14.93	79.13	30.80	0.40	0.8159	78.36	0.00
SE	0.0147	0.0523	0.0092	(0.71)				0.0752		
Bias	-0.0099	0.0146	-0.0431					0.0159		
MSE	0.0007	0.0032	0.0022					0.0035		
s = 1, ρ=3	0.6676	0.8067	0.0762	10.95	85.62	0.04	33.20	0.8072	83.72	100
SE	0.0223	0.0341	0.0130	(7.38)				0.0720		
Bias	0.1676	0.0067	-0.0738					0.0072		
MSE	0.0467	0.0013	0.0088					0.0014		
s = 1, ρ=5	0.6239	0.8045	0.0357	11.48	83.56	0.06	18.33	0.8103	80.56	100
SE	0.0345	0.0123	0.0112	(8.23)				0.0358		
Bias	0.1239	0.0045	-0.1143					0.0103		
MSE	0.0112	0.0006	0.0186					0.0009		
s = 1, ρ=9	0.6337	0.8098	0.0734	8.43	78.56	1.65	5.67	0.8223	81.34	100
SE	0.0321	0.0885	0.0192	(8.35)				0.0398		
Bias	0.1337	0.0098	-0.0766					0.0223		
MSE	0.0243	0.0105	0.0102					0.0013		
s = 30, ρ=3	0.3054	0.8467	0.0704	10.22	62.49	3.68	28.33	0.8195	83.44	100
SE	0.0205	0.0384	0.0156	(9.33)				0.0345		
Bias	-0.1946	0.0467	-0.0796					0.0195		
MSE	0.0262	0.0034	0.0103					0.0007		
s = 30, ρ=5	0.3642	0.8315	0.0245	8.22	59.45	2.95	8.46	0.8213	80.99	100
SE	0.0095	0.0345	0.0107	(5.54)				0.0456		
Bias	-0.1358	0.0315	-0.1255					0.0213		
MSE	0.0164	0.0024	0.0183					0.0012		
s = 30, ρ=9	0.3875	0.8298	0.0119	7.39	60.26	0.08	0.34	0.8023	83.84	100
SE	0.0086	0.0299	0.0084	(8.54)				0.0234		
Bias	-0.1125	0.0298	-0.1381					0.0023		
MSE	0.0123	0.0117	0.0208					0.0003		

Table 3 Large Sample Simulation Results

	Parametric							Semi-Parametric		
	β_1^p	β_2^p	β_3^p	*s*	r^2	Rejections		β_1^{sp}	r^2	Rejections
	0.5	0.8	0.15			J-test	Min/Max	0.8		
N=1,000										
s = 15, ρ=3	0.4895	0.8193	0.1534	15.00	83.13	11.60	0.00	0.8195	82.47	0.00
SE	0.0069	0.0244	0.0042	(0.00)				0.0339		
Bias	-0.0105	0.0193	0.0034					0.0195		
MSE	0.0002	0.0010	0.0000					0.0010		
s = 15, ρ=5	0.4900	0.8183	0.1226	15.00	80.54	48.60	0.00	0.8184	79.71	0.00
SE	0.0067	0.0238	0.0041	(0.00)				0.0338		
Bias	-0.0100	0.0183	-0.0274					0.0184		
MSE	0.0002	0.0009	0.0008					0.0010		
s = 15, ρ=9	0.4903	0.8480	0.1060	15.00	78.77	79.40	0.00	0.8180	77.98	0.00
SE	0.0067	0.0238	0.0041	(0.00)				0.0338		
Bias	0.0097	0.0180	-0.0440					0.0180		
MSE	0.0009	0.0009	0.0019					0.0009		
s = 1, ρ=3	0.5513	0.8079	0.1346	4.15	85.02	0.80	91.20	0.8079	84.61	100
SE	0.0148	0.0156	0.0111	(9.36)				0.0323		
Bias	0.0513	0.0079	-0.0154					0.0079		
MSE	0.0192	0.0003	0.0017					0.0003		
s = 1, ρ=5	0.6179	0.8095	0.0327	11.11	81.63	0.00	30.60	0.8097	81.21	100
SE	0.0102	0.0173	0.0064	(9.21)				0.0326		
Bias	0.1179	0.0095	-0.1173					0.0097		
MSE	0.0165	0.0004	0.0143					0.0004		
s = 1, ρ=9	0.6304	0.8070	0.0251	14.34	84.18	9.40	1.90	0.8081	75.37	100
SE	0.0333	0.0806	0.0192	(9.12)				0.1646		
Bias	0.1304	0.0070	-0.1249					0.0081		
MSE	0.0204	0.0099	0.0168					0.0113		
s = 30, ρ=3	0.3766	0.8381	0.0779	19.81	55.87	0.00	38.60	0.8382	55.96	100
SE	0.0144	0.0343	0.0136	(6.82)				0.0368		
Bias	-0.1234	0.0381	-0.0721					0.0382		
MSE	0.0204	0.0028	0.0081					0.0028		
s = 30, ρ=5	0.3808	0.8290	0.0245	16.95	60.98	0.00	7.80	0.8291	60.86	100
SE	0.0084	0.0299	0.0079	(6.16)				0.0354		
Bias	-0.1192	0.0290	-0.1255					0.0291		
MSE	0.0147	0.0018	0.0160					0.0018		
s = 30, ρ=9	0.3991	0.8256	0.0116	16.14	63.46	0.00	1.80	0.8257	63.32	100
SE	0.0078	0.0281	0.0070	(6.12)				0.0350		
Bias	-0.1009	0.0256	-0.1384					0.0257		
MSE	0.0103	0.0115	0.0193					0.0015		

Table 4 Descriptive Statistics for the Arizona data

	Industrial			Residential		
	Price	Size (Square feet)	Distance (Miles)	Price	Size (Square feet)	Distance (Miles)
N	144	144	144	578	578	578
Mean	$1,106,148	611,875	3	$968,269	1,434,796	10
Standard Deviation	$1,449,411	1,696,617	3	$1,527,834	3,092,161	7
Maximum	$12,000,000	16,247,880	16	$15,809,589	41,327,985	69
Minimum	$250,000	25,400	0	$5,000	3,851	0

Table 5 Descriptive Statistics for the Iowa data

	Price	Time (Years)	Size (Square feet)	Distance Cedar Falls*	Distance Waterloo*
N	646	646	646	646	646
Mean:	$27,574	14.33	32,223	26,464	19,298
Standard Deviation.	$34,011	5.65	190,000	14,434	11,201
Maximum	$500,000	20.56	4,623,023	58,870	43,983
Minimum	$1,000	0	640	298	1,474

*Feet

Table 6 Empirical Results with the Arizona Data

	Parametric		Semi-parametric	
	Industrial	Residential	Industrial	Residential
Intercept	287,983 (27996)	337223 (1695)	___	___
Distance	-0.089 (0.016)	-0.063 (0.0039)	-0.078 (0.011)	-0.0576 (0.004)
Size	0.403 (0.029)	0.015 (0.0006)	___	___
λ	___	___	___	___
Lot	___	-1.07 (0.079)	___	-1.0487 (0.1)
Shift (s)	20,000	13,056	___	___
R^2	.5789	.7811	.6410	.6972
J-test	0.003	0.0284	___	___
Min/Max	25,400/ 16,247,880	3,851/ 41,327,985	25,400/ 16,247,880	3,851/ 41,327,985

Standard errors are in parentheses

Table 7 Empirical Results with the Iowa Data

	Parametric	Semi- parametric	Non-Linear Least Squares	Bayesian*
Intercept	8.610 (0.21)	___	7.965851 (0.467634)	8.356 (7.24,9.57)
Time	0.080 (0.007)	0.0790 (0.0072)	0.081132 (0.007594)	0.068 (0.058,0.081)
Distance Cedar Falls	-0.0000219 (0.000003849)	-0.000002247 (0.000000260)	-0.0000223 (0.00000389)	-0.00002 (-0.0004,-0.0001)
Distance Waterloo	0.00000218 (0.000005096)	0.000000094 (0.000000377)	0.00000342 (0.00000528)	0.000006 (-0.00002,0.00004)
Size	0.016 (0.0034771599)	___	0.105 (0.05)	0.024 (0.008,0.058)
λ	___	___	0	0.2588 (0.176,0.354)
Shift (s)	2,618	___	___	3,516 (1965,6766)
R^2	.271	.3005	0.257	___
J-test	.3034	___	___	___
Min/Max	640/ 4,623,023	640/ 4,623,023	640/ 4,623,023	640/ 4,623,023

*Results from Ecker and Isakson, 2005, large parcels only. Standard errors are in parentheses except for the Bayesian results, which contain the 95% confidence intervals.

Figure 1 Urban Land Price-Size Relationship

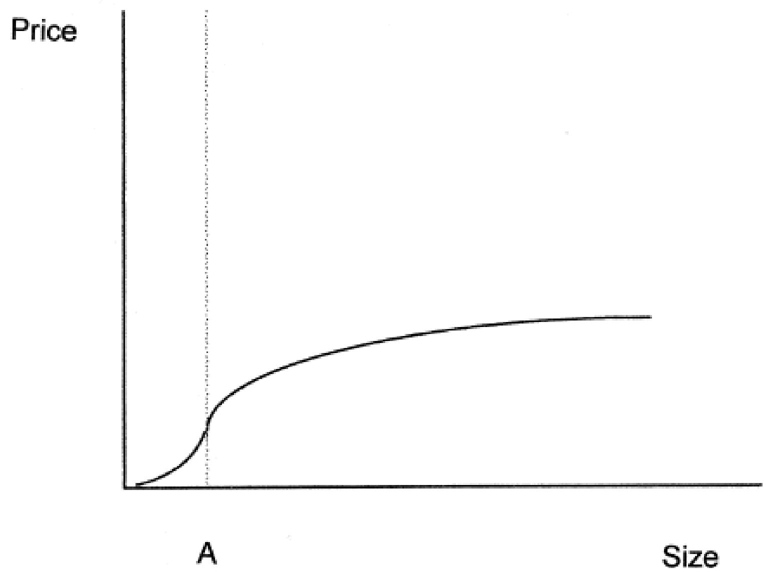


Figure 2: Land Value and Area for ρ=3

$$\ln[y_i] = \alpha_1 + \alpha_2 (A_i - s)^{1/3}$$

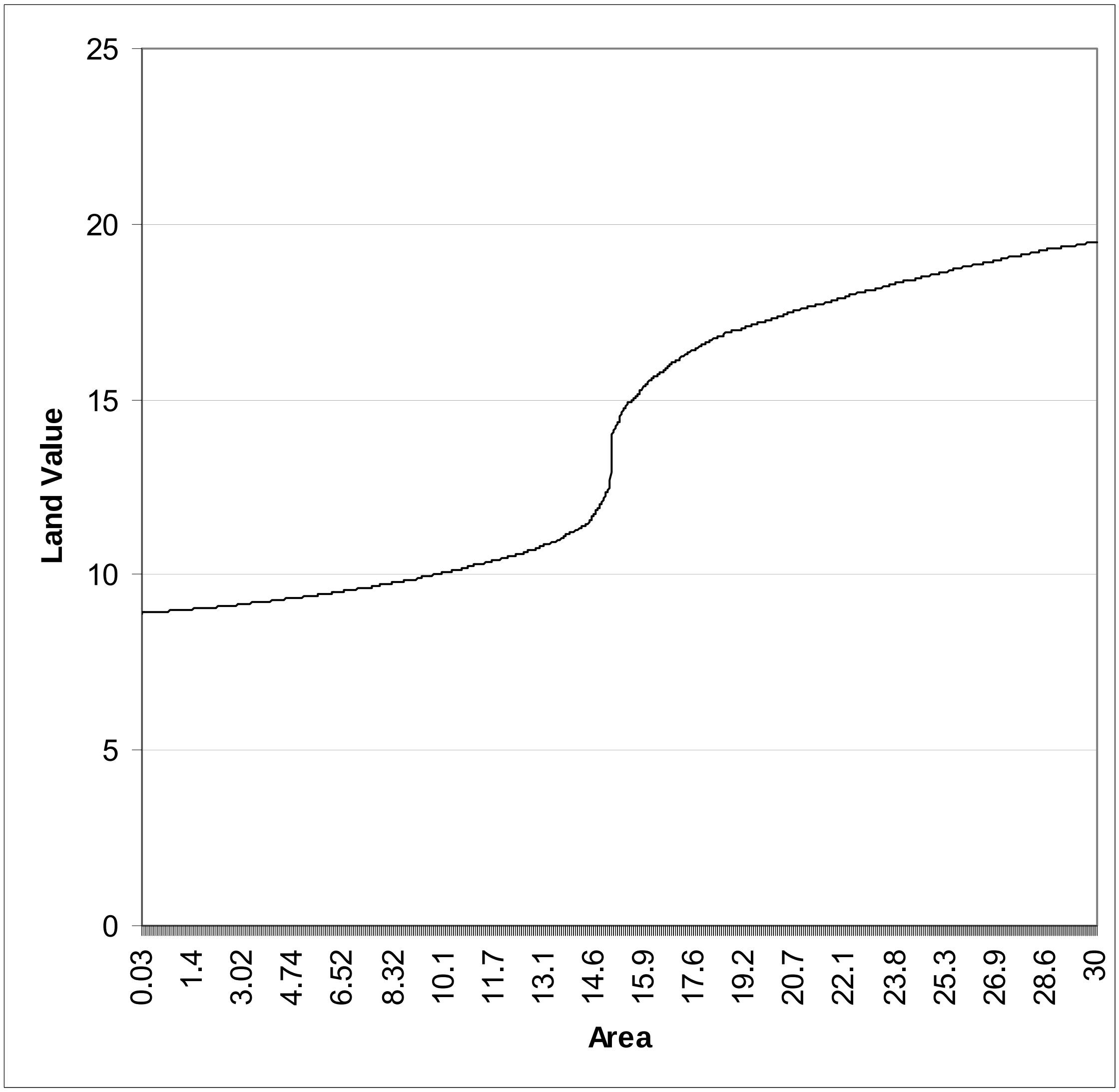


Figure 3: Land Value and Area for ρ =5

$$\ln[y_i] = \alpha_1 + \alpha_2 (A_i - s)^{1/5}$$

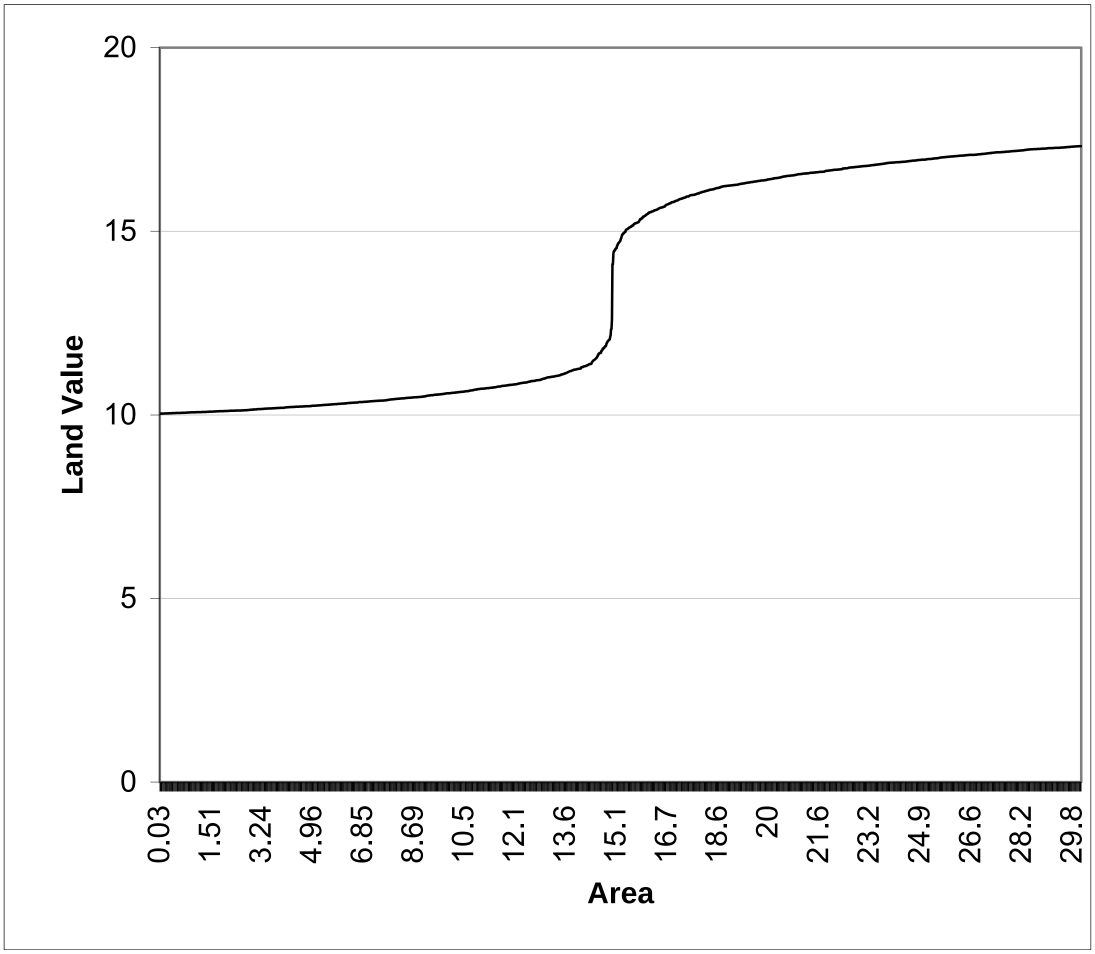


Figure 4 Semi-Parametric Estimate of $g(A_i)$ - Arizona Industrial Data

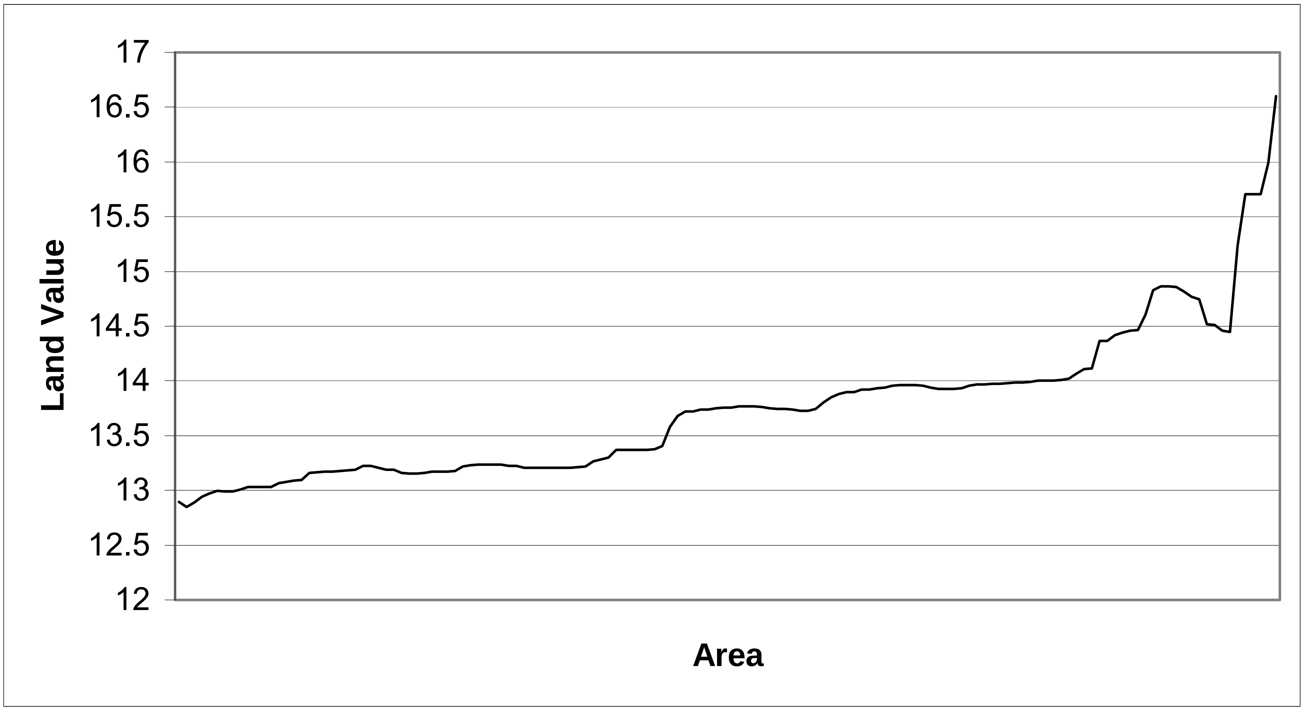


Figure 5 Semi-Parametric Estimate of $g(A_i)$ - Arizona Residential Data

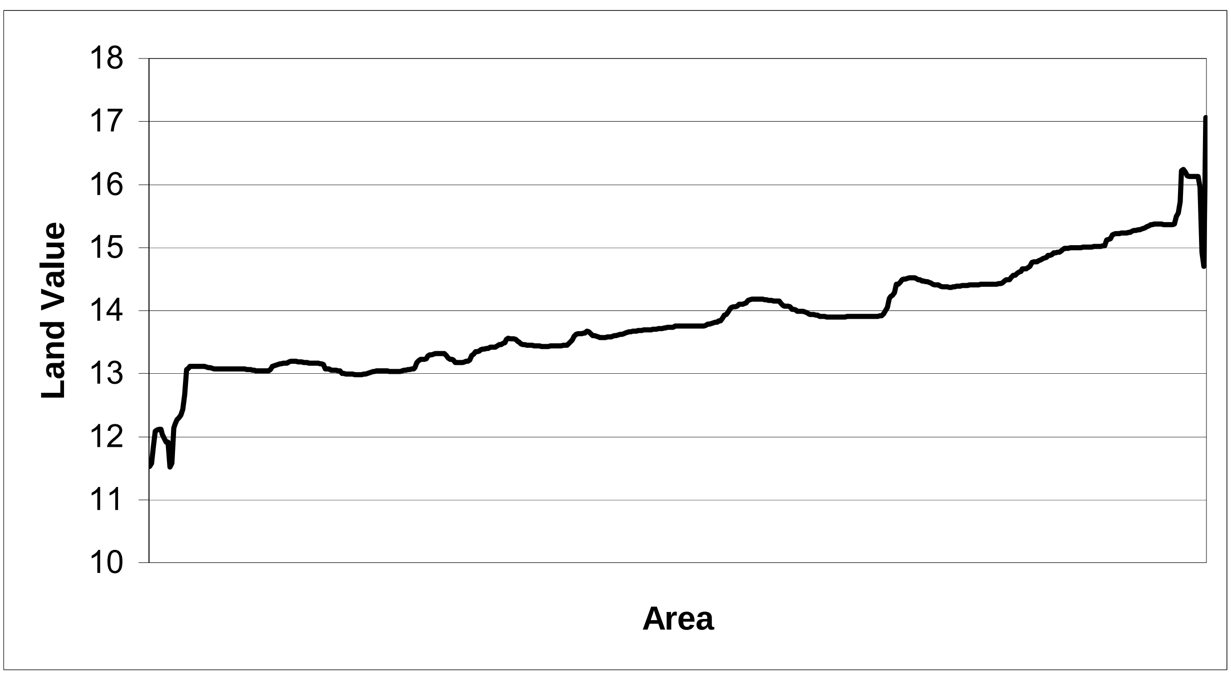
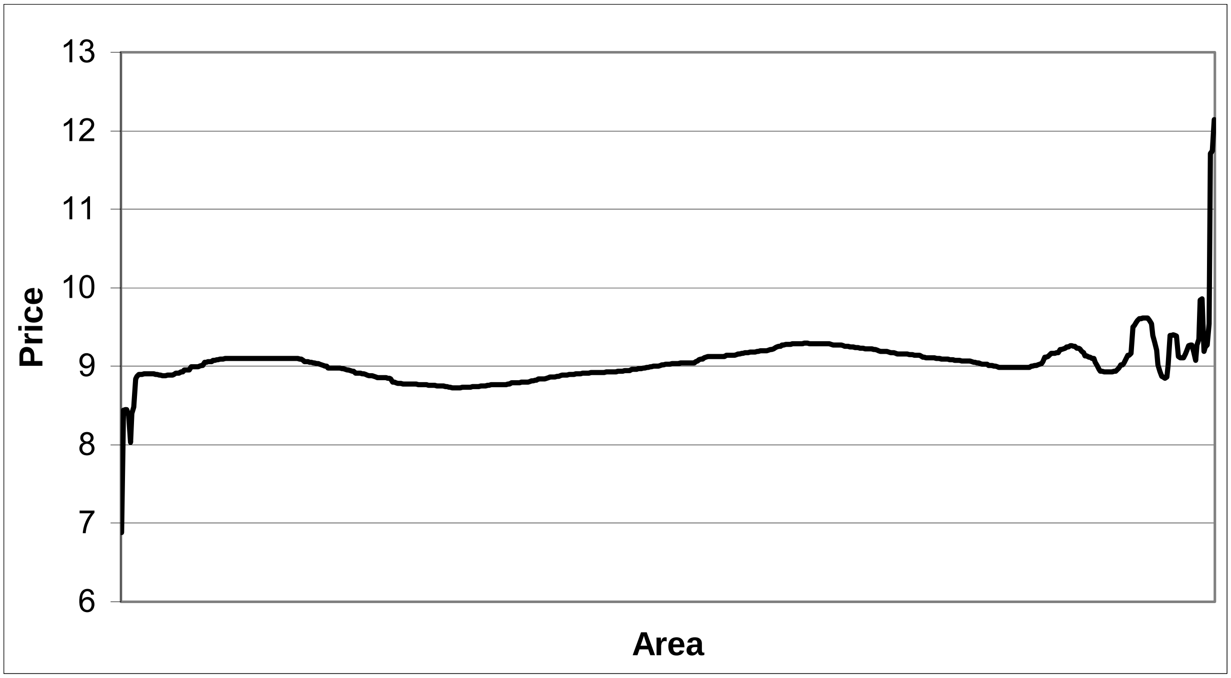


Figure 6 Semi-Parametric Estimate of $g(A_i)$ - Iowa Data



References

Appraisal Institute. The Appraisal of Real Estate, 12th edition, 2001.

Brownstone, D. and A. De Vany, "Zoning, Returns to Scale, and the Value of Undeveloped Land", Review of Economics and Statistics, 1991 (Nov.), 73:4, 699-704.

Colwell, P. "A Primer on Piecewise Parabolic Multiple Regression Analysis via Estimations of Chicago CBD Land Prices", Journal of Real Estate Finance and Economics, Special Issue on Spatial Econometrics, 1998, 17:1, 87-97.

Colwell, P. and H. Munneke, "The Structure of Urban Land Prices," Journal of Urban Economics, 1997, 41 (May), 321-336.

Colwell, P. and H. Munneke, "Land Prices and Land Assembly in the CBD", Journal of Real Estate Finance and Economics, 1999, 18:2, 163-180.

Colwell, P. and H. Munneke, "Estimating a Price Surface for Vacant Land in an Urban Area", Land Economics, 1 March 2003, 79:1, 15-28(14).

Colwell, P. and C.F. Sirmans, "Area, Time, Centrality and the Value of Urban Land", Land Economics, November 1978, 54:4, 514-519.

Colwell, P. and C.F. Sirmans, "Nonlinear Urban Land Prices", Urban Geography, 1980, 1:2, 141-152.

Colwell, P. and C. F. Sirmans, "A Comment on Zoning, Returns to Scale, and the Value of Undeveloped Land", Review of Economics and Statistics,1993, 75:4, 783-786.

Davidson and MacKinnon, Estimation and Inference in Econometrics. Oxford University Press, New York 1981.

DiPasquale, D. and W. Wheaton, Urban Economics and Real Estate Markets, Prentice Hall,1996.

Ecker, M. and H. Isakson, "A Unified Convex-Concave Model of Urban Land Values", Regional Science and Urban Economics, May 2005, 35:3, 265-277.

Lin, Tzu-Chin and A. W. Evans, "The Relationship between the Price of Land and Size of Plot when Plots are Small", Land Economics, 2000, 76(3): 386-394.

Mills, E. S., "The Value of Urban Land" in The Quality fo the Urban Environment, (H. S. Perloff, editor), Johns Hopkins University Press, Baltimore, 1971.

Tabuchi, T. “Quantity Premia in Real Property Markets”, Land Economics, 1996, 72(2): 206-217.

Thornes, P., and D. P. McMillen, “Land Value and Parcel Size: A Semi-parametric Analysis,” Journal of Real Estate Finance and Economics, 1998, 17:3, 233-244.