Hedonic Prices and Public Goods: An Argument for Weighting Locational Attributes by Lot Size

George R Parsons, University of Delaware
Hedonic Prices and Public Goods: An Argument for Weighting Locational Attributes in Hedonic Regressions by Lot Size

GEORGE R. PARSONS

College of Marine Studies and Department of Economics, University of Delaware, Newark, Delaware 19716

Received December 29, 1986; revised June 23, 1988

1. INTRODUCTION

The hedonic price technique is frequently used for valuing attributes of housing. In such an analysis a hedonic price function is specified: \( p = p(x) \). \( p \) is the price of a house and \( x \) is a vector of attributes of the house. The attributes include structural features such as size, age, garage, and fireplace and locational features such as neighborhood quality, crime, access to downtown, and air quality.

The function is interpreted as an equilibrium price schedule for houses (see Rosen [5]). Its first derivative with respect to an element of \( x \) is an implicit price for that attribute. The price function is estimated typically using cross-sectional data from a housing market.

I argue that the measure of each locational attribute in a hedonic price function should be weighted by lot size and that the measure of each structural attribute should not be weighted. The usual regression weights neither locational nor structural attributes. By my argument, if two houses are identical in every respect except that the measure of lot size of the first is twice that of the second, then the measure for air quality, access to parks, and so on for the first should be twice that of the second.

I also argue that failure to weight locational attributes in a hedonic regression that should be weighted gives biased implicit price estimates for structural and locational attributes. The direction and degree of bias depend on the correlations between the attributes and lot size in the housing market, on the variance of lot size in the market, and on the lot size of the observation for which implicit price is being estimated.

2. WEIGHTING LOCATIONAL ATTRIBUTES

A. The Argument in the Long Run

Locational and structural attributes differ in a fundamental way. If a locational attribute is provided to one household in a neighborhood it is provided to all households in that neighborhood. Everyone receives the same level of air quality, same access to downtown, and so on. This is not so for a structural attribute. If a structural attribute is provided to one household, it is not provided to others; only that household receives the services of its garage, fireplace, and so on. In this sense locational attributes are public goods and structural attributes are not.

The number of individuals that may consume a locational attribute of a given quality is limited by the area of land over which that quality is available. There is a fixed amount of land within 1 mile of downtown, a fixed amount of land with a certain air quality level, and so on. The more land taken by one household, the less available to others for consuming the attribute qualities at that location. The size of this lost opportunity for consumption by others is proportional to the amount of land, or lot size, occupied by the household. Doubling the lot size doubles the lost opportunity.

An equilibrium hedonic price function captures this lost opportunity only if its locational attributes are weighted by lot size. The following argument suggests that hedonic functions in competitive markets in long run equilibrium will have such a weighting.

Consider an owner of \( n \) acres of undeveloped land. Assume that the owner sells plots of land of a uniform size with a house of uniform design to achieve maximum profit. Assume that the owner takes an equilibrium hedonic price function for houses as fixed and that the level of all locational attributes is constant across the \( n \) acres. The landowner behaves so as to maximize

\[
\max_{m, x_1} \left\{ \sum_{i=1}^{n} p(x_{i1}, x_{2i}, x_{3i}) \cdot m - c(x_{1i}) \cdot m \right\},
\]

(1)

where \( p(\cdot) \) is the hedonic price function, \( x_{1i} \) is a \( 1 \times k_1 \) vector of structural attributes, \( x_{2i} \) is a \( 1 \times k_2 \) vector of locational attributes, \( x_{3i} \) is lot size, \( c(\cdot) \) is the owner's cost function for producing a house, and \( m \) is the number of lots sold. With \( n \) fixed and uniform size plots sold, choice of \( m \) determines \( x_{3i} \). Let \( x_{1i}^*, m^* \), and \( x_{2i}^* = n/m^* \) be the values that give maximum profit and \( x_{3i}^* \) be the constant level of locational attributes available in equilibrium.

Now, suppose that before the sale of the uniform plots of size \( n/m^* \) a buyer approaches the owner to purchase a lot that is \( t \) times as large as the uniform size. Assume \( t \) is an integer greater than one. Also, assume she
wants a structural design that is different than \( x_i^* \). The owner will sell the buyer the larger lot if

\[
\left[ p(x_1^*, x_2^*, x_3^*) - c(x_1^*) \right] \cdot m^* \\
\leq \left[ p(x_1^*, x_2^*, x_3^*) - c(x_1^*) \right] \cdot (m^* - t) + r. \tag{2}
\]

The left-hand side of this inequality is the maximum profit received if all plots and units are equivalent; the first piece on the right-hand side is the profit received on the \( m^* - t \) units with lots of size \( n/m^* \), and \( r \) is the profit from the larger lot. Assume that the change in housing density, presumably an element in \( x_2 \), is negligible if the larger lot is introduced. The profit on the larger lot is \( r = \bar{p} - c(\bar{x}_1) \), where \( \bar{p} \) is the selling price and \( \bar{x}_1 \) is the desired level of structural attributes. Substituting this expression into (2) and simplifying gives

\[
\bar{p} - c(\bar{x}_1) \geq t p(x_1^*, x_2^*, x_3^*) - tc(x_1^*). \tag{3}
\]

The logic of this condition is obvious. The owner knows she can sell \( t \) lots of uniform size for a profit of \( p(x_1^*, x_2^*, x_3^*) - c(x_1^*) \) each. Hence, if the new seller wishes to occupy \( t \) plots with a single house, she must be prepared to compensate the seller at least \( t \cdot (p(x_1^*, x_2^*, x_3^*) - c(x_1^*)) \).

If \( p(\cdot) \) and \( c(\cdot) \) are linear, (3) may be written

\[
\bar{p} \geq tx_1^* \beta_1 + tx_2^* \beta_2 + tx_3^* \beta_3 - tx_1^* \delta + \bar{x}_1 \delta, \tag{4}
\]

where \( \beta_1 \) is a \( k_1 \times 1 \) vector of parameters on the structural attributes in the hedonic price function, \( \beta_2 \) is a \( k_2 \times 1 \) vector of parameters on the locational attributes in the price function, \( \beta_3 \) is a parameter on lot size in the price function, and \( \delta \) is a \( k_1 \times 1 \) vector of parameters on the structural attributes in the cost function.

A competitive market for attributes implies that \( \beta_1 = \delta \) and that (4) holds with equality. If so, (4), the condition for sale of the large lot, may be written

\[
\bar{p} = \bar{x}_1 \beta_1 + tx_2^* \beta_2 + tx_3^* \beta_3, \tag{5}
\]

In equilibrium then, the sale of the house on the large lot must satisfy (5). Meanwhile, the \( m^* - t \) sales of the standard size lots must satisfy \( p = x_1^* \beta_1 + x_2^* \beta_2 + x_3^* \beta_3 \). In this setting it follows that

\[
p = x_1 \beta_1 + wx_2^* \beta_2 + wx_3^* \beta_3 \tag{6}
\]

is the equilibrium hedonic function, where \( w = t \) for the large lot, and \( w = 1 \) for the \( m^* - t \) standard lots.
Equation (6) holds for repeated calculations of \( \tilde{p} \) for other purchasers of nonstandard lots. An equation like (5) would be computed for each buyer, varying only by the values of \( \bar{x}_1 \) and \( t \). Collecting the set of equations across buyers and setting \( t = w \) in each, (6) remains the equilibrium price schedule.

The exercise may be repeated also for other owners under the same assumptions with the same results. After doing so, and after selecting units of measure for a standard lot such that \( x^*_1 = 1 \), (6) becomes

\[
p = x_1 \beta_1 + wx_2 \beta_2 + w \beta_3,
\]

where \( w \) is just lot size and the weighting applies for all levels of \( x_2 \). Note that it is the presence of many owners behaving as in my example that assures that (5) holds with equality and that Rosen's interpretation still applies. That interpretation is that each owner and buyer takes \( \beta_1 \), \( \beta_2 \), and \( \beta_3 \) in (7) as fixed, but as a group determines the value of these parameters.\(^1\)

The weighted price function in (7) has a somewhat different interpretation than the usual unweighted equation. Taking the derivative of (7) with respect to lot size, \( \partial p / \partial w = x_2 \beta_2 + \beta_3 \). The first term \( (x_2 \beta_2) \) is the implicit value of the foregone consumption of locational attributes by others if more land is occupied with a house. It is the "user cost" of occupying land. (Note that occupying more land does not increase the buyer's consumption of locational attributes. Air quality, distance to down-

\(^1\)The weighting argument applies with slight modification for nonlinear price functions as well. If the price function in (1) is nonlinear but additively separable in \( x_1 \) and \( (x_2, x_3) \), then \( p(x_1, x_2, x_3) = p_1(x_1) + p_2(x_2, x_3) \). Again, competitive markets for attributes imply \( p_1(x_1) = c(x_1) \). Following the same steps as in the linear case, I have \( p = p_1(x_1) + w p_2(x_2, x_3) \) as the counterpart to (7) for any form of \( p_2 \). For example, the weighted version of a semilog hedonic price function is \( p = (\log x_1) \beta_1 + (w \log x_2) \beta_2 + (w \log x_3) \beta_3 \), where \( p_1 = (\log x_1) \beta_1 \), \( p_2 = (\log x_2) \beta_2 + (\log x_3) \beta_3 \), and \( x^*_j \) is an arbitrary standard lot size that is constant. (Note that the rule calls for weighting the log of the attribute, not logging the weighted attribute).

If the price function is not separable in \( x_1 \) and \( (x_2, x_3) \), the steps to (4) and (5) cannot be taken as in my linear argument. But, (3) must hold with equality by assumption of competitive markets. This, in turn, implies that the price function must always satisfy \( \partial p / \partial x^*_j = (\partial p / \partial x_j) \), where \( x^*_j \) is any element in \( x^*_j \). For example, a double log form that satisfies this derivative property is \( \log p = (\log x_1) \beta_1 + (\log x_2) \beta_2 + (\log w) (1 + \beta_3) \). (It is equivalent to write \( \log (p/w) = (\log x_1) \alpha + (\log x_2) \beta + (\log w) \beta_3 \). A linear Box-Cox form that satisfies this derivative property is \( ((p/w)^\lambda - 1) / \lambda = x_1 \alpha + x_2 \beta \). Notice that the dependent variable in both cases simply enters in terms of price per unit of land occupied \( (p/w) \).

An additively separable weighted price function such as the semilog has a desirable feature that the nonadditively separable weighted price functions (double log and linear Box-Cox) do not have. In the separable functions, the prices for locational attributes increase proportionally with lot size but the structural attribute prices do not. In the nonseparable functions, prices of both structural and locational attributes increase proportionally with lot size. There is theoretical justification for this property for locational attributes. That there is theoretical justification for this property for structural attributes is doubtful.
town, and so on are the same on 1-, 2-, and 3-acre lots.) The second term is the implicit price of another unit of land for private use (perhaps more garden space, room for a garage, solitude, and so forth). An unweighted price function excludes the first term and thereby treats the value of a unit of land as invariant with respect to its locational attributes: an implausible assumption in most any housing market.

In the weighted equilibrium price equation (7), locational attribute prices \((w \beta_2)\) vary across houses in proportion to the amount of land occupied; structural attribute prices do not \((\beta_1)\). Much of the urban economics literature concerned with land markets has accepted this property of locational attribute prices for years. Following Alonso [1], researchers such as Diamond [2] and Diamond and Tolley [3] have formulated hedonic price regressions with the price per unit of land as a dependent variable and locational attributes of the land as independent variables. This is equivalent to a weighting rule (without structural attributes present) and gives locational attribute prices that vary proportionally with the amount of land occupied.

Unfortunately, most of the urban and environmental economics literature concerned with housing markets has specified price functions following Rosen’s model and has ignored this property of locational attributes in land markets. Prices per unit of land have not been used as dependent variables or weighted locational attributes by lot size. By doing so, the public good aspect of locational attributes has inadvertently been missed.

B. The Argument in the Short Run

Reconsider my argument in the case of an owner of developed land. When approached by a potential buyer for a larger lot, the owner of undeveloped land in my previous example could insist on a price of \(\bar{p}\) with profit of \(r\). Always present was the alternative of dividing the acres at no cost into smaller lots for a certain return of \(r\). An owner of developed land has no such alternative, or has an alternative with return of \(r\) less conversion cost. Hence, one might expect some short run (\(m\) fixed for owners of developed land) equilibrium sales for larger lots at a price \(p' < \bar{p}\). This would imply something less than a strict weighting by lot size for houses on developed lots.

Under certain conditions a strict weighting rule should still apply. If existing houses on large lots have the same attributes as new houses on large lots, buyers always purchase the former instead of the latter if \(p' < \bar{p}\). Sellers would be unable to get their reservation price, \(\bar{p}\), and no new development on large lots would occur. The dynamics of this market are elementary. Until \(p'\) rises to \(p' = \bar{p}\), no new development occurs. If \(p'\) rises to \(p' = \bar{p}\), through growing demand for housing, new development occurs. In the latter market the strict weighting rule holds in short run equilibrium.
By this argument the necessary conditions for a strict weighting rule to apply are a growing housing market and existing houses being close substitutes for new houses. If these conditions are violated sales at \( p' < \hat{p} \) are likely for existing houses and a less than strict weighting rule would apply. Equation (7) would not hold in equilibrium for all houses.

This complicates the form of the equilibrium price function considerably. For example, it may take the form of two separate equilibrium price functions: one for existing houses without weighting and one for new houses with weighting. But, this ignores that some existing houses are good substitutes for new houses. There may be two separate equilibrium price equations, but there may be one for new houses and existing houses with attributes similar to new houses and another for existing houses with attributes dissimilar from new houses. The first is weighted, and the second is unweighted. This, however, ignores the degree of substitution between different types of existing and new houses. The better an existing house is a substitute for new houses the closer its locational attributes are to strict weighting, and the poorer an existing house is a substitute for new houses the closer its locational attribute is to no weighting.

Perhaps Graves et al.'s [4] partial weighting price function best approximates the equilibrium function for a set of existing houses with unknown degree of substitution for new houses. Their partial weighting function is \( p = x_i\alpha + w x_j\beta + x_i\gamma \). With the appropriate signs on \( \beta \) and \( \gamma \) the implicit price of locational attributes increases with lot size but less than proportionately.

3. IMPLICIT PRICE BIAS WITHOUT WEIGHTING

In this section I investigate the bias that occurs in estimating a hedonic price function that should be weighted but is not. The nature of the bias may be seen in a simple linear price function with one attribute as an argument. (Appendix section II presents the argument for the case of many attributes.)

Assume that the true price function is \( E(p_i) = w_i x_i \beta \), where the subscript \( i \) denotes an observation, \( x_i \) is a locational attribute, \( w_i \) is a lot size weight, and \( p_i \) is the price. For an unweighted function using ordinary least squares I estimate \( \beta = (x'x)^{-1}x'p \) as an implicit price for \( x_i \), where \( x \) and \( p \) are \( n \times 1 \) vectors of observations on \( x_i \) and \( p_i \). The true implicit price for \( x_i \) at observation \( i \) is \( w_i \beta \). Hence, the question of bias is: Is \( E(\beta) \approx w_i \beta \) for each \( i \)? Since the true implicit price of \( x_i \) varies across observations, the question is asked for each observation.

\[ E(p) = Wx\beta \], where \( W \) is an \( n \times n \) diagonal matrix with the \( w_i \) observations on the diagonal and zeros elsewhere. Hence, \( E(\beta) = (x'x)^{-1}x'Wx\beta \). But, \( (x'x)^{-1}x'Wx \) is simply a measure of correlation between lot size and
attribute quality in the sample, or a weighted average of the \( w_i \)'s across observations, which may be written

\[
\hat{w} = \left[ \left( \frac{x_1^2}{\sum x_i^2} \right) \cdot w_1 + \cdots + \left( \frac{x_n^2}{\sum x_i^2} \right) \cdot w_n \right]. \tag{8}
\]

This implies that \( E(b) = \hat{w}\beta \), and the question of bias becomes: Is \( \hat{w}\beta \approx w_i\beta \) for each \( i \)? The answer is conditional:

\[
\begin{align*}
\hat{w} = w_i, & \quad \text{then } E(b) = w_i\beta \\
\hat{w} > w_i, & \quad \text{then } E(b) > w_i\beta \\
\hat{w} < w_i, & \quad \text{then } E(b) < w_i\beta.
\end{align*}
\]

The direction and degree of bias depend on the relative sizes of \( w_i \) and \( \hat{w} \). Small \( w_i \) and large \( \hat{w} \) increase the likelihood that the estimate overstates the true price. (Large \( w_i \) and small \( \hat{w} \) increase the likelihood of understatement.)

Since \( w_i \) is just the lot size for an observation, the smaller the lot where the estimate is being made the greater the likelihood of an overestimate. (The larger the lot the greater the likelihood of an underestimate.) The interpretation is that the unweighted regression assumes implicitly that all lots are of the same size. In effect, it treats small lots as larger than their true size and large lots as smaller than their true size. By assuming that small lots occupy more land than they do, the unweighted regression attributes too large a user cost to these units and thereby overestimates the price of the attribute. It does the opposite for large lots.

From (9) it is easy to see what makes \( \hat{w} \) large and small; its possible range is the same as the range of \( w_i \). (Note that if \( w_i \) is a constant, \( \hat{w} \) equals that constant.) \( \hat{w} \) gets larger (closer to the largest \( w_i \) in distribution of \( w_i \)'s) as the correlation between \( w_i \) and \( x_i^2 \) increases. It gets smaller as the correlation decreases. Also, the greater the variance of \( w_i \) the more sensitive the size of \( \hat{w} \) is to the correlation.

For example, assume that \( x_i \) is a measure of air quality as a locational attribute of housing. If air quality is good in areas where lots tend to be large and bad where lots tend to be small, I expect an unweighted regression to overestimate the implicit price of air quality. And, the larger the variance of lot size the larger I expect the overestimate to be. If air quality is good in areas where lots tend to be small, my expectation is reversed.
The interpretation is similar to omitted variable bias. If an attribute in an unweighted hedonic regression is positively correlated with lot size, it serves in part as a proxy for "omitted" weighting. That is, the estimated implicit price for that attribute will pick up some of the user cost of occupying more land. The reverse occurs for attributes negatively correlated with lot size.²

4. CONCLUSIONS

I argue that locational attributes of houses in hedonic price functions will be weighted by lot size in long run equilibrium. The weighting results because there is a user cost associated with the land purchased with a house. Although purchasing more land does not increase the buyer's use of locational attributes, it does reduce the amount of land available for others to enjoy the locational attributes at that location. The reduced availability is the user cost. The price of land, which is part of the price of a house, will account for this user cost and result in weighted locational attributes in long run equilibrium hedonic price functions. It follows that the implicit price of locational attributes will vary across houses in direct proportion to the amount of land occupied by the house.

I also argue that in the short run if existing houses are perfect substitutes for new houses, the weighting rule will apply for the entire housing market. If existing houses are imperfect substitutes for new houses, a strict weighting rule need no longer apply. Separate price functions are likely to exist for old houses that are close substitutes for new houses and new houses and for existing houses that are not close substitutes for new houses. The former would follow a strict weighting rule, but the latter would not.

Estimation of a hedonic price function without weighting where strict weighting should hold in equilibrium will give bias implicit prices for locational and structural attributes. The direction of bias for a given price depends on the direction of correlation between that attribute and lot size and on the lot size of the observation where the estimate is being made. Positive correlation increases the probability of overestimation, and negative correlation increases the probability of underestimation. The degree of overestimation or underestimation increases with the degree of this correlation and the size of the variance of lot size in the housing market. Also, prices for locational attributes are more likely to be underestimated for houses on large lots and overestimated for houses on small lots.

²In the case of a hedonic price function with many attributes, these findings still hold. Locational and structural attributes that are positively correlated with lot size tend to be overestimated. Those that are highly negatively correlated with lot size tend to be underestimated. The Appendix shows these results and a simple diagrammatic explain for the single attribute case.
I. A Graphic Explanation of Implicit Price Bias in a Single Attribute Hedonic Price Function

Consider Fig. 1a. I have two observations on houses: \(a\) and \(b\). The lot sizes, prices, and attribute quantities are \(w_a, w_b, p_a, p_b\) and \(x_a, x_b\). Lot size and attribute quantity are positively correlated, so \(x_b > x_a\) and \(w_b > w_a\). Assume \(w_a = 1\) and \(w_b > 1\). The unweighted and true hedonic regressions are shown.

Both regressions go through point \(A\). The true regression goes through \(B'\), the point \((p_b, w_b x_b)\), where \(x\) is weighted by lot size. The unweighted regression goes through \(B\), the point \((p_a, x_b)\), where \(x\) is not weighted by lot size. The true slope is less than the slope of the unweighted regression, and the true attribute price, \((p_b - p_a)/(w_b x_b - x_a)\), is less than the price for the unweighted regression, \((p_b - p_a)/(x_b - x_a)\). Hence, I overestimate with the unweighted regression.

The true regression is accounting for the user cost of occupying land at \(b\); the unweighted regression is not. This is seen in the unweighted regression.
attributing the change in price \((p_b - p_a)\) to a smaller change in the locational attribute \((x_b - x_a)\) than the true regression \((w_b x_b - w_a x_a)\). As such, the increments of \(x\) are assigned greater values in the unweighted regression than in the true regression.

Figure 1b shows why an increase in the variance of lot size increases this bias. Two true regressions are drawn. One goes through the point B', where \(w = w_b\), and another through the point C, where \(w = w_c\). I assume \(w_c > w_b\). The lot size variance is greater for the true regression that goes through C.

The smaller slope on the second true regression implies a lower true price and hence a larger overestimate by using an unweighted regression. With more land occupied there is simply a greater user cost accounted for in the true regression that is missed in the unweighted regression.

II. Implicit Price Bias in a Many Attribute Hedonic Price Function

I accept that a weighted hedonic is the true model and consider the bias of estimating prices if I believe that an unweighted hedonic is the correct model. I assume that I have \(n\) observations on \(x_1, x_2, x_3,\) and \(p\), and I let \(i\) denote an observation, \(X_1\) is an \(n \times (k_1 + 1)\) matrix of observations on \(x_1, x_3\); \(X_2\) is an \(n \times k_2\) matrix of observations on \(x_2\); \(k = k_1 + k_2 + 1\); \(X = [X_1, X_2]\); \(Y = [X_1, W X_2]\); \(W\) is an \(n \times n\) diagonal matrix with the \(w_i\) weights on the diagonal; and \(p = n \times 1\) vector of prices.

The true model, written as a conditional expectation function satisfying all the properties of the classical regression model, is

\[
E(p|Y) = Y \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \quad V(p|Y) = \sigma_y I, \quad \text{and} \quad \text{rank}(Y) = k. \quad (A1)
\]

The model I accept in error is

\[
E(p|X) = X \begin{bmatrix} \alpha^* \\ \beta^* \end{bmatrix}, \quad V(p|X) = \sigma_y I, \quad \text{and} \quad \text{rank}(X) = k. \quad (A2)
\]

I take \(\alpha^*\) as the prices for structural attributes and \(\beta^*\) as the prices for locational attributes. The ordinary least-squares estimates of \(\alpha^*\) and \(\beta^*\) are my estimates of implicit prices:

\[
\begin{bmatrix} a \\ b \end{bmatrix} = [X'X]^{-1} X' p. \quad (A3)
\]
True prices are found by differentiating (A1) with respect to $X$. For a given observation, these prices are $\alpha$ and $w_i\beta$. Then, the question of bias is

$$\text{Is } E\left[\begin{array}{c}
\alpha \\
w_i\beta
\end{array}\right] \stackrel{\text{MV}}{\sim} \left[\begin{array}{c}
\alpha \\
w_i\beta
\end{array}\right] \text{ for each } i?\hspace{1cm} (A4)$$

Because the true prices of locational attributes vary with lot size, the bias is evaluated at each observation. Expression (A4) is $k + 1$ separate questions. The first $k_1$ rows pertain to the bias of the price of the structural attributes, and the last $k_2$ rows to the bias of the price of the locational attributes.

If I take the expected value of (A3) conditioned on $Y$, substitute (A1) into that expression, and partition matrices—all shown in Subappendix A—I may write (A4) as

$$\text{Is } F\beta \stackrel{\text{MV}}{\sim} V\beta \text{ for each } i?\hspace{1cm} (A5)$$

$F = (X'X)^{-1}X'WX_2$ and $V = \begin{bmatrix} 0 & 0 \\
w_iI_2 & I_2 \end{bmatrix}$. $F$ is a $k \times k_2$ matrix, $0$ is a $(k_1 + 1) \times k_2$ matrix of zeros, and $I_2$ is a $k_2 \times k_2$ identity matrix. The values in $V$ are different for each observation; the values in $F$ are constant. Like question (A4) there are $k$ rows in question (A5); the first $k_1$ pertain to the bias of the structural attributes and the last $k_2$ to the bias of locational attributes.

Consider two locational attributes in the hedonic price function, $x_1$ and $x_2$. The question of bias is

$$\text{Is } E\left[\begin{array}{c}
b_1 \\
b_2
\end{array}\right] \stackrel{\text{MV}}{\sim} w_i\left[\begin{array}{c}
\beta_1 \\
\beta_2
\end{array}\right] \text{ for all } i?\hspace{1cm} (A6)$$

Or,

$$\text{Is } E(b_1) = f_1\beta_1 + g_1\beta_2 \stackrel{\text{MV}}{\sim} w_i\beta_1 \text{ for all } i?\hspace{1cm} (A7)$$

And,

$$\text{Is } E(b_2) = f_2\beta_1 + g_2\beta_2 \stackrel{\text{MV}}{\sim} w_i\beta_2 \text{ for all } i?$$

From the expressions for $F$ and $V$, $f_1$, $g_1$, $f_2$, and $g_2$ are the coefficients in the ordinary least-squares regressions:

$$w_ix_1 = f_1x_1 + f_2x_2$$
$$w_ix_2 = g_1x_1 + g_2x_2.\hspace{1cm} (A8)$$
(Subappendix B shows expressions (A6)-(A8) for the case of \( n \) attributes with structural attributes included.) Consider the bias of \( b_1 \). Since both attributes are measured as goods, \( \beta_1 \) and \( \beta_2 \) are positive. The term \( f_1 \) is the partial correlation between \( w_{i}x_{1} \) and \( x_{1} \), controlling for \( x_{2} \). It is analogous to \( f \) in my previous example. The term \( g_1 \) is the partial correlation between \( w_{i}x_{2} \) and \( x_{1} \), controlling for \( x_{2} \).

The interpretation of the \( f \) and \( g \) coefficients is akin to omitted variable bias. The weighted locational attributes are “omitted” from the unweighted hedonic regression. The unweighted locational attributes are included. The latter, therefore, serve in some fashion as proxies for the former. This is captured in the \( f \) and \( g \) terms in the auxiliary regressions (A8). \( f_1 \) measures the extent to which \( x_{1} \) is serving as a proxy for \( w_{i}x_{1} \) in the unweighted regression, and \( g_1 \) measures the extent to which \( x_{1} \) is serving as a proxy for \( w_{i}x_{2} \).

The larger \( f_1 \) and \( g_1 \) are, the greater the probability that \( b_1 \) will overestimate the price of the first locational attribute. \( f_2 \) and \( g_2 \) are large when \( x_{1} \) explains a large share of the variation in \( w_{i}x_{1} \) and \( w_{i}x_{2} \), that is, when \( x_{1} \) is a proxy for the omitted weighting information. A large \( \beta_2 \) would increase the probability of overestimation more; \( x_{1} \) would be serving as a proxy for an attribute that accounted for a larger portion of the price of a house.

If \( w_{i} \), \( x_{1} \), and \( x_{2} \) are highly positively correlated, then values of \( f_1 \), \( f_2 \), \( g_1 \), and \( g_2 \) are likely to be large and the probability of overestimating both attribute prices is large. If \( w_{i} \) is highly positively correlated with one of the attributes, say \( x_{1} \), but not with \( x_{2} \), then \( f_1 \) and \( g_1 \) are large and \( f_2 \) and \( g_2 \) are small. If so, it is likely that \( b_1 \) overestimates the true price of the first attribute, and \( b_2 \) underestimates the true price of the second. If \( w_{i} \), \( x_{1} \), and \( x_{2} \) are positively correlated, the attribute with the greater partial correlation with \( w_{i} \) is more likely to overestimate its true price. In this case the more highly correlated attribute serves as the primary proxy for the omitted weighting information on both attributes. The less highly correlated attribute serves as a secondary proxy. Opposite results follow with negative correlation between an attribute and lot size.

The analysis resembles the case of a hedonic price function with one locational attribute. Prices for attributes with greater partial correlation with lot size are more likely to be overestimated. Prices for attributes with lesser partial correlation with lot size are more likely to be underestimated. The degree of each depends on the degree of correlation and the variance of lot size. The analysis applies for the prices of structural attributes also. If a structural attribute is highly correlated with lot size controlling for other attributes, the price of that attribute is likely to be overestimated.

The same principle applies for hedonic price functions with \( n \) attributes. Subappendix B presents the relevant question of bias and auxiliary regressions for the \( n \) attribute case.
SUBAPPENDIX A

Derivation of Eq. (A5)

I take the expected value of (A3) conditioned on Y and substitute (A1) into that expression and I have

\[ E(\alpha | \beta) = (X'X)^{-1}X'(X_1:WX_2)(\alpha | \beta), \]  

(SA1)

\[ E(\alpha | \beta) = \begin{bmatrix} I_1 & \cdots & 0 \\ (m \times m) & \cdots & (X'X)^{-1}X'WX_2 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (k \times k_2) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \]  

(SA2)

\(I_1\) is an identity matrix, \(0\) is a matrix of zeros, and \(m = k_1 + 1\). So, \(k = m + k_2\). I rewrite the right-hand side of Eq. (A4) as

\[ \begin{bmatrix} \alpha \\ w_2 \beta \end{bmatrix} = \begin{bmatrix} I_1 & \cdots & 0 \\ (m \times m) & \cdots & (m \times k_2) \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_2I_2 \\ \vdots & \ddots & \vdots \\ (k \times m) & \cdots & (k \times k_2) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \]  

(SA3)

I substitute (SA3) and (SA2) into question (A4) and simplify to get

\[ \text{Is } (X'X)^{-1}X'WX_2\beta \overset{\text{se}}{\approx} \begin{bmatrix} 0 \\ w_2I_2 \end{bmatrix} \beta? \]  

(SA4)

\[ (k \times k)(k_2 \times 1) \]

I let \(F = (X'X)^{-1}X'WX_2\) and \(V = \begin{bmatrix} 0 \\ w_2I_2 \end{bmatrix}\), and I have Eq. (A5). Note that the value of \(\alpha\) is irrelevant in assessing the degree of bias for locational and structural attributes.

SUBAPPENDIX B

The Question of Bias with n-Attributes in the Hedonic Function

The question of bias is

\[ \text{Is } F\beta \overset{\text{se}}{\approx} V\beta \text{ for all } i? \]  

(SB1)

\(F = (X'X)^{-1}X'WX_2\) and \(V = \begin{bmatrix} 0 \\ w_2I_2 \end{bmatrix}\). \(F\) is a matrix of coefficient estimates for \(k_2\) regressions. In each regression the regressors are the complete set of unweighted attributes. Successively, each of \(k_2\) weighted locational at-
tributes is regressed on this set. The first column of $F$ is the set of coefficient estimates from the regression of the first weighted locational attribute on the unweighted set of attributes, the second column is the same for the second weighted locational attribute, and so on.

For a single structural attribute price, say the $l$th price, the question of bias is

$$\text{Is } f_{1l} \beta_1 + f_{2l} \beta_2 + \cdots + f_{kl} \beta_k \hat{z}_l \leq 0? \quad (SB2)$$

This is the $l$th row of question (SB1). The $\beta$'s are the elements of $\beta$. The $f$'s are the elements from the $l$th row of $F$. The corresponding $f$'s are shown below in the auxiliary regressions (SB4). The bias on structural attributes, unlike locational attributes, does not vary with lot size.

The question for a single locational attribute price, say the $j$th price, is

$$\text{Is } g_{1j} \beta_{1j} + \cdots + g_{2j} \beta_{2j} + \cdots + g_{kj} \beta_{kj} \hat{z}_j \leq w_j \beta_j \text{ for all } l? \quad (SB3)$$

The $g$'s are elements of the $j$th row of $F$ and are shown below in the auxiliary regressions (SB4).

The matrix $F$ is the set of $f$ and $g$ vectors from the regressions

$$w_i \hat{z}_j = f_{1j} x_1 + \cdots + f_{1k} x_k + g_{1j} \hat{z}_1$$
$$+ \cdots + g_{1j} \hat{z}_k$$

$$w_i \hat{z}_j = f_{2j} x_1 + \cdots + f_{2k} x_k + g_{2j} \hat{z}_1$$
$$+ \cdots + g_{2j} \hat{z}_k$$

$$\vdots$$

$$w_i \hat{z}_j = f_{kj} x_1 + \cdots + f_{kj} x_k + g_{kj} \hat{z}_1$$
$$+ \cdots + g_{kj} \hat{z}_k.$$

(SB4)

The structural attributes are defined as $x_1 = (x_1, x_2, \ldots, x_k)$, and the locational attributes are defined as $x_2 = (\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_k)$. The coefficients from the first regression form the first column of $F$, the second regression the second column of $F$, and so on.

REFERENCES