Documentos CEDE
ISSN 1657-7191 edición electrónica

Using Frontier Models to Mitigate Omitted Variable Bias in Hedonic Pricing Models: A Case Study for Air Quality in Bogotá, Colombia

Fernando Carriazo
Richard Ready
James Shortle

11
MARZO DE 2011

Universidad de los Andes
Facultad de Economía

CEDE
Centro de Estudios sobre Desarrollo Económico

Serie Documentos Cede, 2011-11
ISSN 1657-7191

Marzo de 2011

Calle 19A No. 1 – 37, Bloque W.
Bogotá, D. C., Colombia
Teléfonos: 3394949- 3394999, extensiones 2400, 2049, 3233
infocede@uniandes.edu.co
http://economia.uniandes.edu.co

Ediciones Uniandes
Carrera 1ª Este No. 19 – 27, edificio Aulas 6, A. A. 4976
Bogotá, D. C., Colombia
Teléfonos: 3394949- 3394999, extensión 2133, Fax: extensión 2158
infeduni@uniandes.edu.co

Edición, diseño de cubierta, preprensa y prensa digital:
Proceditor ltda.
Calle 1ª C No. 27 A – 01
Bogotá, D. C., Colombia
Teléfonos: 2204275, 220 4276, Fax: extensión 102
proceditor@etb.net.co

Impreso en Colombia – *Printed in Colombia*

USING FRONTIER MODELS TO MITIGATE OMITTED VARIABLE BIAS IN HEDONIC PRICING MODELS: A CASE STUDY FOR AIR QUALITY IN BOGOTÁ, COLOMBIA

Fernando Carriazo*
Universidad de Los Andes

Richard Ready**
The Pennsylvania State University

James Shortle***
The Pennsylvania State University

Abstract

Hedonic pricing models use property value differentials to value changes in environmental quality. If unmeasured quality attributes of residential properties are correlated with an environmental quality measure of interest, conventional methods for estimating implicit prices will be biased. Because many unmeasured quality measures tend to be asymmetrically distributed across properties, it may be possible to mitigate this bias by estimating a heteroskedastic frontier regression model. This approach is demonstrated for a hedonic price function that values air quality in Bogotá, Colombia.

Key words: hedonic pricing model, omitted variables, air quality, frontier model.

JEL classification: Q51, Q53.

---

*Corresponding author
Assistant Professor-Researcher. Universidad de Los Andes. Department of Economics.
Email: *f.carriazo126@uniandes.edu.co*

**Associate Professor of Agricultural and Environmental Economics. The Pennsylvania State University. Department of Agricultural Economics and Rural Sociology. Email: *rready@psu.edu*

*** Distinguished Professor of Agricultural and Environmental Economics and Director of The Environment and Natural Resources Institute. The Pennsylvania State University. Department of Agricultural Economics and Rural Sociology. Email: *jshortle@psu.edu*

USANDO MODELOS DE FRONTERA PARA MITIGAR SESGO DE VARIABLES OMITIDAS EN MODELOS DE PRECIOS HEDÓNICOS: UN ESTUDIO DE CASO PARA CALIDAD DE AIRE EN BOGOTÁ, COLOMBIA

Resumen

Los modelos de precios hedónicos utilizan diferenciales en el precio de las propiedades para estimar cambios en la calidad ambiental. Si los atributos no cuantificados de las viviendas están correlacionados con una medida de calidad ambiental de interés, los métodos convencionales para estimar precios implícitos serían sesgados. Debido a que muchas medidas no cuantificadas de la calidad de la vivienda tienden a presentar una distribución asimétrica entre las propiedades, puede ser posible mitigar este sesgo estimando un modelo de regresión de frontera heteroscedástico. La implementación de este modelo se muestra para una función de precios hedónicos que valora calidad del aire en Bogotá, Colombia.

Clasificación JEL: Q51, Q53.

Palabras clave: modelo de precios hedónicos, variables omitidas, calidad del aire, modelo de frontera.

I. INTRODUCTION

The hedonic pricing model is commonly used to estimate the influence of environmental amenities and disamenities on both rural and urban property values. Recent examples include studies of the impacts of wetland amenities on urban property prices (Mahan et al 2000), the additional value on a property from being located nearby the coast (Conroy and Milosh, 2009), the impacts of wetland easements on agricultural land values (Shultz and Taff 2004), the impact of open space on non-residential land use (Shultz and King 2001), the effects of open space and the proximity to animal production facilities on residential values (Irwin 2002; Ready and Abdalla 2005), the effect of flood hazards on property values (Bin and Polasky 2004), the value of proximity to forest reserves (Thorsnes 2002), the impact of forest fires on house prices (Mueller et al 2009), and the externality effects of "supportive housing" on neighboring prices (Galster et al 2004).

Hedonic property models work by exploiting spatial variation in property values and environmental characteristics. This feature makes hedonic models of the impacts of environmental variables on property values vulnerable to omitted variable bias. When an omitted variable influences property values and is also spatially correlated with the environmental quality measure of interest, the estimated implicit price for that environmental measure will be biased.

In the case of residential housing, the structural characteristics of housing that are most important in determining housing values typically include structure size, age, lot size (for detached homes), and distance to employment and shopping centers. These characteristics are typically measured and included in applications of the hedonic price method. However, there are important measures of the quality of housing units that are often difficult or impossible for the hedonic analyst to measure. Let the unmeasured quality of a housing unit be represented by a single variable, UQ that incorporates many different unmeasured quality dimensions, including the quality of the construction and materials, exterior aesthetic (curb) appeal, and neighborhood quality measures such as school quality and quality of nearby parks and recreational facilities. We argue that UQ will tend to be positively correlated with environmental quality, leading to the potential for biased estimates of implicit prices for environmental quality. We present an econometric approach to mitigating this potential bias through the use of a frontier econometric model. While most hedonic applications are linked to the well known behavioral assumptions proposed by Lancaster (1966) and Rosen (1974), there is little theoretical guidance about the appropriate functional form for the hedonic price function or the appropriate error structure for the regression. Most studies assume a symmetric and homoskedastic (normal) random error in the hedonic regression. We argue that variation in UQ will tend to be asymmetric, and that a frontier model that includes two error components, one symmetric and the other asymmetric, can mitigate the bias from unmeasured UQ.

We demonstrate the approach using a dataset on rental properties in Bogotá, Colombia to explore the relationship between rental values and air quality. An important area of application of the hedonic pricing method is valuing improvements in urban air quality. Beginning with Ridker and Henning (1967), studies have consistently found a significant negative relationship between air pollution and property values (Smith and Huang 1995; Chattopadhyay 1999; Zabel and Kiel 2000; Kim et al 2003; Beron et al 2004; Chey and Greenstone 2005). We find the implicit price for air pollution to be negative, but to be smaller when estimated using a frontier model than when estimated using a conventional OLS model, consistent with the hypothesis that unmeasured variation in UQ biases conventional results.

The paper is organized as follows. Section II discusses how UQ might vary across properties. Section III discusses the frontier econometric model and how it can accommodate unmeasured UQ. Section IV describes the study area and the data used for estimations. Section V presents empirical results and Section VI includes a final discussion.

II. OMITTED VARIABLE BIAS AND ASYMETRIC ERROR

Omitted variables in hedonic studies are characteristics of a property that are observable to the buyer and seller, but not to the researcher. When an omitted variable influences property values and is also spatially correlated with the environmental quality measure of interest (e.g. air pollution, noise levels, proximity to landfills), the estimated implicit price for that environmental measure will be biased. Hedonic pricing models are typically estimated from secondary data sources, and there tend to be important determinants of house price that are not included in the data, such as school quality, view from the property, ambient noise, and physical condition of the property.

UQ will include many different unmeasured house characteristics. The argument why these characteristics could be asymmetrically distributed will be developed here in the context of one example. An important factor that determines housing values that is often unmeasured in hedonic pricing studies is construction quality. Some houses have marble floors and granite counter tops, while others have vinyl floors and laminate countertops. Because older properties may or may not have been remodeled, the age of structures is not a reliable indicator of quality of construction. Housing units with higher construction quality will tend to sell or rent for higher prices (Kain and Quigley 1970). While the researcher may know from assessment databases which properties are in poor physical condition, they typically do not know which houses have what standard of construction, but the buyers and sellers do.

How might construction quality vary across houses? First, there is the issue of the shape of the distribution of construction quality. Except in areas of severe poverty, there will tend to be a baseline standard of construction that no property drops below. This standard will be dictated by local building and rental codes and by cultural norms. While construction quality may be higher (in some cases quite a bit higher) than this standard, very few properties will be lower. This motivates the specification of an asymmetric error in the hedonic model, where the distribution of construction quality across properties has a long right hand tail, but no left hand tail. Because construction quality positively affects property values, the distribution of values will then also be asymmetric, where properties with higher than standard construction quality will have higher values than would otherwise be expected based on their measured characteristics.

The second issue regarding how construction quality varies across houses is the relationship between construction quality and environmental quality (in our case, air quality). Because environmental quality is a normal good, higher income households will tend to locate in neighborhoods with better environmental quality. Because construction quality is also a normal good, those same households are also more likely to demand higher construction quality. They are more likely to improve through remodeling the quality of their houses and to do so to a higher standard than lower income households in poorer quality neighborhoods. Builders will have to pay more for building sites in higher quality neighborhoods, and will spend more on construction quality on those sites than they will on cheaper sites in poorer quality neighborhoods. While there may still be properties in better neighborhoods that are built to the baseline standard, the proportion of properties that exceed the baseline standard, and the degree to which the baseline standard will be exceeded, will be higher in better air quality neighborhoods than in worse quality neighborhoods.

Other components of UQ will tend to show a similar pattern. For example, if higher income households demand and support better schools, school quality would tend to be correlated with environmental quality. If unobserved school quality has a baseline level below which no neighborhood schools fall, then its distribution would be asymmetric and the model developed here could mitigate the potential bias from its omission as well. Other examples of neighborhood measures that could be correlated with environmental quality and that have a baseline level such that their distribution would be asymmetric include quality of local parks and recreational facilities and quantity and quality of restaurants and cultural amenities. Other structure-level amenities that could vary with environmental quality and that would likely have an asymmetric distribution include aesthetics of the façade and quality of landscaping (curb appeal) and, for apartment buildings, quality of common areas.

The result is that the distribution of UQ is asymmetric with a variance that will be higher in better quality neighborhoods than in worse quality neighborhoods. This difference is illustrated in Figure 1. The baseline standard for unmeasured quality measures is shown as UQ_{BL}. The probability density function for the distribution of UQ is shown for a high environmental quality neighborhood and a low environmental quality neighborhood. No property in either neighborhood can have UQ less than UQ_{BL}, so both distributions are truncated at UQ_{BL}. The variance of UQ is higher in the better quality neighborhood, meaning that fewer properties will have UQ at or near the baseline standard. Because the variance of UQ differs between the two neighborhoods, the average level of UQ in the better neighborhood ($\overline{UQ}_H$) will be higher than the average level in the poorer quality neighborhood ($\overline{UQ}_L$).

Unless this heteroskedasticity in the unmeasured UQ is accounted for, it will bias conventional estimates of the implicit price of air quality. A simple comparison of properties located in better quality neighborhoods and properties located in poorer quality neighborhoods will capture the influence of both the difference in environmental quality and the difference in the average level of UQ. Ideally, we would directly measure all of the relevant quality variables for each property, but that may be prohibitively difficult. In the next section, we present an econometric model that allows for heteroskedastic asymmetric errors. This model will, at least to some degree, mitigate the bias from the unmeasured variation in housing quality.

III. A FRONTIER HEDONIC MODEL WITH ASYMMETRIC RANDOM ERROR

We extend the standard hedonic model proposed by Rosen (1974) to include two error components, a symmetric component representing the usual idiosyncratic error, and a non-negative asymmetric component which captures unobserved variation in UQ. This structure of the regression error term is identical to frontier models that have been estimated in production economics to measure efficiency of firms (See Kumbhakar and Lovell 2004). In production economics, the asymmetric component signals the distance from a sample point to a theoretical production frontier, and is often interpreted as a measure of technical efficiency. In our model, the asymmetric component captures omitted variables and does not have an efficiency interpretation.

The frontier hedonic model is given by

$$P_i = P(Z_i, A_i, N_i, \beta)\exp(u_i)\exp(v_i) \quad [1]$$

where P_i is the price (in our case rent) for property i, Z_i is a vector of measured structural characteristics of the property (size, age, etc.), A_i is a vector of environmental amenity measures for the property (air quality, noise, etc.), and N_i is a vector of neighborhood-specific amenities and disamenities (proximity to employment, crime, etc).

Given a parameter vector β, P(•) is a function that determines the price for a property with characteristics Z_i, A_i and N_i with UQ equal to the baseline level of quality, UQ_{BL}. There are two error terms. The usual, symmetric, mean zero random error term, v_i, captures all symmetrically-distributed idiosyncratic errors for property i, and is assumed to be distributed $N(0,\sigma_v^2)$. A second error term, u_i, represents unmeasured variation in UQ, and is assumed distributed half-normal, with u_i = $|z_i|$ where $z_i \sim N(0,\sigma_u^2)$. The distribution of u_i is therefore asymmetric, bounded below at 0 with a tail extending to the right. For a property with baseline UQ=UQ_{BL}, u_i would equal 0. For any property with UQ greater than the baseline, u_i would be positive. Finally, the variance of the asymmetric error term may vary across observations. The following heteroskedastic error model is explored

$$\sigma_v^2 = \exp(\delta_0 + \delta_1 Q_1 + \delta_2 Q_2 + ... \qquad [2]$$

where Q_1, Q_2, … are neighborhood characteristics that could be correlated with UQ, including environmental quality and δ_0 , δ_1 , δ_2,… are parameters to be estimated. If a neighborhood characteristic, Q_1, is measured as a bad (for example the concentration of particulate matter, PM10), then the argument we make regarding the relationship between unobserved quality (UQ) and environmental quality (Q_1) would imply a negative value for δ_1.

Assuming a Cobb-Douglas form for P(•) with k1, k2 and k3 elements in the vector of structural, environmental and neighborhood characteristics (Z, A, and N) respectively, the price equation can be specified as

$$\ln(P_i) = \beta_0 + \sum_{k=1}^{k_1} \beta_k \ln Z_{ki} + \sum_{j=1}^{k_2} \beta_j \ln A_{ji} + \sum_{l=1}^{k_3} \beta_l \ln N_{li} + u_i + v_i \qquad [3]$$

Equations (2) and (3) can be estimated simultaneously using maximum likelihood techniques. We estimate this model twice. First, a homoskedastic model (with δ_1, δ_2,… set equal to 0) is estimated, to test whether σ_u^2 is significantly different from zero, i.e. to test whether there is a positive asymmetric error term. To carry out this test we use a likelihood ratio test. This test measures if the difference between the log likelihood function of the homoskedastic (restricted) model and the heteroscedastic (unrestricted) model is significantly different from zero. Second, if an asymmetric error structure is detected, then the more general model is estimated, and the null hypothesis of $\delta_1=0$ is tested. An estimated value of δ_1 not equal to zero is consistent with the proposed correlation between UQ and environmental quality, Q_1.

IV. STUDY AREA AND DATA

Study Area

Information from the rental housing market in Bogotá, Colombia was used to examine the relationship between the asymmetric error and air pollution. Bogotá is one of the most polluted cities in Latin America, with particulate matter less than 10 micrometers (PM10) concentrations comparable to the levels of Santiago and Mexico City (World Bank, 2006, pp 147). During the period 2001-2003, annual PM10 levels averaged 98 µg/m3 at four monitoring stations, and at one monitoring station a Total Suspended Particles (TSP) annual average of 288 µm/m3 was registered. In contrast, the U.S National Air Quality Standard is 50 $\mu g/m^3$ for PM10 while local annual standards for PM10 and TSP are 80 and 100 ug/m3 respectively. Air quality in Bogotá varies spatially in consistent ways. According to the local environmental agency (DAMA), the PM10 readings from monitoring stations located in the northwestern and west central parts of the city are consistently higher than those in the north, south and center areas of the city.

Data

The data set includes 6544 apartment rentals. Proprietary data on the rents and property characteristics were obtained from Metrocuadrado, a company that publishes prices for property rentals and sales in Bogotá, under a confidentiality agreement. The data includes structural characteristics for apartments that were listed for rent during the period 2001-2006, the rental price (including administrative fees), and the latitude and longitude for each housing unit. Because the dataset was limited to apartment rentals, the resulting hedonic price model will reflect only the preferences of households who rent apartments. Households who live in owner-occupied units may have different preferences. Rental prices were inflated or deflated using the Colombian Consumer Price Index with 2005 as the base year. Information on neighborhood amenities were obtained from digital maps from Bogota's planning office. Distances were calculated for each housing unit to the closest drainage ditch (open and concrete ditches that carry storm and waste waters), the closest main road, the closest metropolitan (large) park, and the closest zonal (small) park. Based on average public transportation, walking and driving times in Bogotá, commuting times were calculated for each apartment to the closer of two Central Business Districts (the Historic Center and 72nd Street).

A measure of neighborhood well-being, called "socioeconomic stratum," was obtained from the Bogotá local planning office. This variable classifies neighborhoods into homogeneous residential areas based on zoning, the quality of the neighborhood (e.g road conditions) and specific housing development's characteristics aggregated at the census block level. In addition to signaling neighborhood quality, socioeconomic stratum is also used to determine the amount that is charged for public utilities such as water, electricity, and telephone. Even though income levels were not available in our data, socioeconomic stratum is likely to be highly correlated with income. The socioeconomic stratum variable takes discrete values from 1 (low quality neighborhoods) to 6 (high quality neighborhoods). Very few rental properties are located in neighborhoods in strata 1 or 2, which typically are areas of extreme poverty or areas dominated by nonresidential uses, and these are not included in this analysis. Information from Bogotá's Observatory of Delinquency and Violence was used to estimate each neighborhood's crime rate, expressed as the number of homicides per 100,000 inhabitants. This data was incomplete. For neighborhoods with missing information on crime, the five nearest observed neighboring crime rates were averaged.

Data on air quality was obtained from the local environmental agency for 11 monitoring stations. For each station, average annual PM10 concentration was calculated for the period 2001-2006. A map of air quality for the entire city was constructed using Inverse Distance Weighted (IDW) spatial interpolation. Using IDW, the imputed pollution measurement at any point within the urban area is given by the weighted average of the 11 monitoring station readings, where the weights are inversely proportional to the distance to each of the measurement points. The resulting map (Figure 2) was shown to experts at the local environmental agency, who stated that it was a good representation of spatial variability in air quality throughout the city.[1]Broadly, the northern area of the city presents average pollution levels for the 2001-2005 period close to 38 ug/m^3 whereas pollution levels in the western area are much higher, ranging from 74 to 86 ug/m^3 . Elevation data comes from USGS (2004). Table 1 defines and provides summary statistics for the variables used in the hedonic model.

V. RESULTS

An initial (homoskedastic) frontier model was estimated to determine whether an asymmetric error component does in fact exist. Full results are not presented here, but are available from the authors. A likelihood ratio test strongly rejected the null hypothesis of zero variance for the asymmetric error (χ_1^2 =15.75). This finding provides support for our argument

[1]Ideally, a dispersion model that captures both wind direction and weather conditions would be used. The local environmental agency in Bogota is developing a dispersion model based on the TAPOM pollution model by Zarate et al (2004), but it is not yet available. The raster developed based on Inverse Distance Interpolation (IDW) of monitoring stations' PM10 readings describes general local air quality conditions and accurately captures differences of PM10 within the urban area

that asymmetrically-distributed unmeasured quality variables have an important impact on housing values.

Estimation results for the second (heteroskedastic) frontier model (HFM) are presented in the first three columns of Table 2. [2] Coefficient estimates for structural characteristics are of expected signs. Apartments located in neighborhoods with higher socioeconomic stratum rent for more. Apartments located at higher elevations rent for more. Apartments far away from main roads rent for more. The estimated coefficient for distance to drainage ditches is negative, but not statistically significant. Drainage ditches in Bogotá are generally unattractive, and may be viewed as a disamenity. Apartments near metropolitan (large) parks rent for more than apartments located far away. Zonal (small local) parks did not have a significant impact on rents. Apartments located closer to the central business districts (as measured by commute time) have higher rents. Apartments located in areas with higher crime rates have lower rents. Of particular interest is the estimated coefficient on PM10. This was negative and significant at all conventional levels, implying that higher PM10 concentrations are a disamenity.

The HFM models the unmeasured variation in UQ by allowing the variance of the asymmetric error component to differ by neighborhood. Some effort was made to determine which neighborhood characteristics were most important for explaining variation in the variance of the asymmetric error component. In particular, we searched for neighborhood characteristics that were correlated with air quality. We found a strong negative correlation between elevation and environmental quality (-0.5447). When elevation and its interaction with PM10 are included as variables that explain the variance of the asymmetric error component, these variables were not significant in the HFM specification. However, when socioeconomic stratum and ln(PM10) were included, they were found to be important determinants of the variance of the asymmetric

[2] For variables in natural logs, such as PM10, the estimated coefficient is interpreted as a price elasticity.

error, as was their interaction. The specific form used to model the variance of the asymmetric error component was then

$$\sigma_u^2 = \exp[\delta_0 + \delta_1 * \ln PM10 + \delta_2 * stratum + \delta_3 * \ln PM10 * stratum] \qquad [4]$$

Figure 3 shows how the variance of the asymmetric error component varies with PM10 and socioeconomic stratum. If both air quality and neighborhood quality are poor, then the variance of the asymmetric error is low. This suggests that in areas with both poor air quality and low socioeconomic stratum, houses tend to be built close to the UQ baseline. However, if either air quality or socioeconomic stratum increases, the variance of the asymmetric error term increases, suggesting increased variation in UQ.[3]

If it is the case that variation in UQ is greater in areas with better environmental quality and/or better socioeconomic stratum, and consequently that UQ is on average higher in those areas, then OLS estimation will tend to overestimate the implicit price of both air quality and socioeconomic stratum. OLS regression estimates are presented in the last three columns of Table 2. The OLS estimated price elasticity for PM10 is 59% larger than the estimate from the HFM. Similarly, the OLS estimate of the marginal impact of socioeconomic stratum on rents was 29% higher than the HFM estimate. In both cases, failure to account for unobserved variability in UQ biased the OLS parameter, so that it overestimated the true impact of air quality on apartment rents.

As a robustness check, we examined if spatial dependence was driving our results. Methods to estimate an HFM model with spatial dependence are not available. However, it is possible to compare spatial models to our OLS results, to determine whether parameter estimates are sensitive to the assumed spatial structure. Two spatial econometric models were estimated, a

[3] When the interaction term between lnPM10 and stratum was included in the deterministic part of the implicit price function, its estimated coefficient was not statistically significant.

Spatial Autoregressive Model (SEM) and a Spatial Lag (S-Lag) model. Parameter estimates from both spatial models were close to OLS estimates, suggesting that parameter estimates are robust to spatial structure, and that our HFM estimates are not influenced by spatial dependence. Complete results from SEM and S-Lag models are available upon request.[4]

The price elasticity estimates for PM10 from Table 2 can be expressed as marginal implicit prices by multiplying the estimated elasticity by the ratio of price (rent) divided by the pollution level (PM10). Table 3 shows the calculated marginal implicit price for an increase in PM10 using the average rental price and PM10 level for each socioeconomic stratum. The marginal implicit price is higher for higher socioeconomic strata both because average rents are higher and because average PM10 levels are lower. Table 3 again shows the bias from using OLS estimation.[5]

VI. DISCUSSION

Our estimation results suggest that failure to account for unobserved quality differences in house characteristics that vary with pollution levels will bias parameter estimates from the hedonic regression. We found that the asymmetric variance increases with lower PM10 and higher stratum, but these effects are greatest for the lowest strata and the worst PM10. As a consequence, the price elasticity for air quality was 59% higher in the OLS specification than in a frontier model with asymmetric random errors. Published estimates of the marginal implicit price for air quality improvements may suffer from similar biases.

We might expect the same pattern of results for other environmental amenities and disamenities such as noise, exposure to environmental toxics, and proximity to undesirable land

[4] Please see table A1 in reviewers appendix

[5] See appendix for a discussion of simulation results showing how the HFM mitigates omitted variable bias in this model.

uses. This can be seen anecdotally by observing that high end homes (with high construction quality) tend to be built in neighborhoods with better environmental quality. In fact, this issue will arise for any environmental amenity that is a normal good. From a policy perspective, marginal implicit prices for environmental amenities and disamenities will tend to be overstated if the issue of differences in UQ is ignored. Use of OLS estimates of marginal implicit prices could result in overinvestment in environmental quality improvement. In a city like Bogota with many social priorities, overinvestment in one area means underinvestment in another area and losses in social welfare. Fortunately, the approach suggested here is relatively easy to implement, and does not require any extra data. Software for estimation of frontier models is widely available.[6]

While the results obtained here are suggestive, they are not conclusive. Ideally, to test the thesis that unobserved quality was driving these results, information on selected dimensions of UQ (for example construction quality, curb appeal, quality of local restaurants) would be obtained for a sample of houses, along with information on household income, to see if UQ is indeed correlated with both environmental quality and income. Frontier and OLS models could then be estimated that included and excluded information on these selected UQ measures, to see if UQ is indeed leading to omitted variable bias, and whether the use of a frontier model mitigates that bias. Measurement of variables such as construction quality and curb appeal may require site visits to each housing unit by the researchers (see Morris et al. 1972).

Another area for future research would be to develop a spatially explicit frontier model that allows both the idiosyncratic and asymmetric error components to be spatially correlated. We would expect that UQ would tend to be more similar for closely neighboring properties, if only

[6] We use Stata for the HPM estimation. Other commercial packages such as Limdep, Frontier, Gauss, TSP and Shazam can be readily used for the estimation of frontier models.

because closely neighboring properties are often built at the same time and often by the same builder. It would be necessary to specify the likelihood function for this type of model and develop methods for its estimation.

Of course the best approach to any situation with potential omitted variable bias is to collect information on that variable. That approach has practical limitations, and even the most aggressive data collection efforts will miss some potentially-important variables that are known to the buyers and sellers. The estimation approach suggested in this paper is relatively easy to implement. If the the HFM model gives similar results to the OLS model, then the analyst can be assured that asymmetrically-distributed omitted quality variables are not biasing estimated implicit prices.

References

Beron K.J., Hanson Y., Murdoch J. C., Thayer Mark A. (2004). Hedonic Price Functions and Spatial Dependence: Implications for the Demand for Urban Air Quality. *In Advances in Spatial Econometrics.Methodology Tools and Applications*. Anselin L., Florax R.J.G.M, Rey S.J. (eds). Springer, Germany. 267-281

Bin, O. and Polasky, S. (2004) "Effects of Flood Hazards on Property Values: Evidence before and after hurricane Floyd." *Land Economics*, 80, (4).490-500

Chattopadhyay, S. (1999) Estimating the Demand for Air Quality: New Evidence Based on the Chicago Housing Market." *Land Economics* 75(1):22-38.

Chey, K.J. and Greenstone, M. (2005) "Does Air Quality Matter? Evidence from the Housing Market". *Journal of Political Economy*, Vol 13, No2.

Conroy, S.J. and Milosch, J.L. (2009) "An Estimation of the coastal Premium for Residential Housing Prices in San Diego County." *Journal of Real Estate and Finance Economics* / Open access-Springer

Galster, G., Tatian, P. and Petit, K. (2004) "Supportive Housing and Neighborhood Property Value Externalities. " *Land Economics*,80, (1). 33-54

Irwin E.G. (2002) "The Effects of Open Space on Residential Property Values." *Land Economics*, 78, (4).465-480

Kain, J.F and Quigley, J.M. "Measuring the Value of Housing Quality". *Journal of the American Statistical Association.* Vol. 65, No. 330, (Jun., 1970), pp. 532-548.

Kiel,K.A and Zabel,J.E. (1999). "The Accuracy of Owner-Provided House Values: The 1978-1991 American Housing Survey" . *Real Estate Economics*. Vol27. No 2,pp,263-298.

Kim, C.W., Phips, T.T. and Anselin, L. (2003) "Measuring the Benefits of Air Quality Improvements: A Spatial Hedonic Approach". *Journal of Environmental Economics and Management* 45. 24-39

Kumbhakar,S.C and Lovell C.A K. (2004). Stochastic Frontier Analysis. Cambridge, Cambridge University Press.

Lancaster, K.J.(1966). "A New Approach to Consumer Theory." *The Journal of Political Economy,* 74 (2), 132-157.

Mahan,B.L., Polasky, S. and Adams R.M. (2000) "Valuing Urban Wetlands: A Property Price Approach." *Land Economics*, 76, (1). 100-113.

Morris, E.W., Woods, M.E and Jacobson A.L. (1972) "The Measurement of Housing Quality". *Land Economics,* Vol 48. No 4. pp. 383-387

Mueller, J., Loomis, J. and González-Cabán, A. (2009) "Do repeated wildfires Change Homebuyers' Demand for Homes in High-risk Areas? A Hedonic Analysis of the Short and Long-Term effects of Repeated Wildfires on House Prices in Southern California." *Journal of Real Estate and Finance Economics*, 38: 155-172.

Ready, R.C. and Abdalla, C. (2005) "The Amenity and Disamenity Impacts of Agriculture: Estimates from a Hedonic Pricing Model." *American Journal of Agricultural Economics*. 87 : 314-326

Ridker R.G. and Henning J.A. (1967) "The Determinants of Residential Property Values with Special Reference to Air Pollution." *Review of Economics and Statistics* . 49: 246-25

Rosen, S (1974) "Hedonic prices and implicit markets: product differentiation in pure competition." *Journal of Political Economy* 82, 34-55.

Shultz,S.D., and Taff, S.J. (2004) "Implicit Prices of Wetland Easements in Areas of Production Agriculture." *Land Economics*, 80, (4). 501-512

Shultz, S.D., and King,D.A. (2001) “The Use of Census Data for Hedonic Price Estimates of Open-Space Amenities and Land Use.” *Journal of Real Estate Finance and Economics* 22:2/3, 239-252

Smith V.K. and Huang J.C. (1995) “Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models.” *Journal of Political Economy* 103 : 209-227

Thorsnes, P.(2002) “The Value of a Suburban Forest Preserve: Estimates from Sales of Vacant Residential Building Lots.” *Land Economics*, 78, (3).426-441.

Worldbank (2006) , “Republic of Colombia. Mitigating Environmental Degradation to Foster Growth and Reduce Inequality”. World Bank Document. Report No 36345-CO , February 25. pp1-345

Zabel, J.E, and Kiel, K.A. (2000) “Estimating the demand for Air Quality in Four U.S Cities”. *Land Economics*. 76 (2): 174-194

Zarate,E.,Belalcazar,L.C. and Echeverry, D. (2004) "Air Quality Modelling Over Bogotá City" Paper presented in the 9th International Conference on Harmonization within Atmospheric Dispertion Modelling for Regulatory Purposes". Accessed online:http://www.harmo.org/conferences/proceedings/_Garmisch/publishedSections/6.31.pdf. June 2005.

Table 1. Summary of statistics for property characteristics.

Variable	Definition	Mean	SD
Rent	Apartment rent in 2005 pesos	2150682	1806038
Area	Constructed area in squared meters	126.72	74.79
Bedrooms	Number of bedrooms	2.56	0.77
Bathrooms	Number of bathrooms	2.44	0.97
Carpet	=1 if apartment is carpeted.	0.87	0.34
Dinning-room	=1 if apartment has an independent dinning room.	0.35	0.48
Electric Kitchen	=1 if apartment has electric kitchen.	0.16	0.37
Door Keeper	=1 if building has a doorkeeper 24 hours.	0.94	0.24
Floor number	Floor number of apartment in the building	3.99	2.63
Elevator 1	=1 if building has 1 elevator.	0.61	0.49
Elevator 2_3	=1 if building has 2 or 3 elevators,	0.18	0.39
Elevator 4_5	=1 if building has 4 or 5 elevators,	0.01	0.08
Age0_5	=1 if apartment is between 0 and 5 years old.	0.16	0.37
Age5_10	=1 if apartment is between 5 and 10 years old.	0.34	0.47
Age10_20	=1 if apartment is between 10 and 20 years old.	0.33	0.47
Age20more	=1 if apartment is more than 20 years old.	0.07	0.26
Y2002	=1 if apartment was listed in year 2002.	0.07	0.25
Y2003	=1 if apartment was listed in year 2003.	0.10	0.31
Y2004	=1 if apartment was listed in year 2004.	0.26	0.44
Y2005	=1 if apartment was listed in year 2005.	0.39	0.49
Y2006	=1 if apartment was listed in year 2006.	0.09	0.29
Stratum	Socio economic sratum of the neighborhood	5.15	1.03
Air quality	Air pollution level of PM10. (mg/m^3)	48.87	12.62
Crime	Neighborhood Crime Index (# homicides per 100000)	2.57	4.10
Elevation	Elevation at the property (meters)	28.96	23.13
Dist_Canal	Distance to the closest canal (meters)	512.68	421.37
Dist_Main_Road	Distance to the closest main road (meters)	132.23	105.54
Dist_Met_Park	Distance to the closest metropolitan park (meters)	1325.71	786.83
Dist_Zon_Park	Distance to the closest zonal park (meters)	1788.53	935.25
Accesibility	Commute time to closest employment center (minutes)	8.94	5.55

Table 2. Estimation Results: HFM and OLS

Variable	Heteroskedastic Frontier Model Coeff	S.E.	t-stat	OLS Model** Coeff	S.E.	t-stat
Intercept	9.3210	.14805	62.96	9.3387	.1200921	77.76
Ln(Area)	0.8966	0.0128	70.00	0.9135	0.0127	71.77
Bedrooms	-0.0627	0.0064	-9.83	-0.0678	0.0065	-10.51
Bathrooms	0.0391	0.0053	7.44	0.0375	0.0052	7.15
Carpet	0.0859	0.0087	9.87	0.0811	0.0087	9.31
Dinning-room	-0.0439	0.0092	-4.79	-0.0399	0.0094	-4.24
Electric Kitchen	0.0039	0.0095	0.41	0.0001	0.0097	0.01
Door Keeper	0.1879	0.0141	13.31	0.1773	0.0147	12.02
Floor number	0.0078	0.0013	6.04	0.0074	0.0013	5.66
Elevator 1	0.0840	0.0095	8.86	0.0869	0.0099	8.75
Elevator 2_3	0.1461	0.0127	11.52	0.1457	0.0130	11.23
Elevator 4_5	0.2131	0.0420	5.07	0.2121	0.0422	5.03
Age0_5	-0.0754	0.0128	-5.90	-0.0742	0.0130	-5.71
Age5_10	-0.1692	0.0115	-14.74	-0.1720	0.0117	-14.75
Age10_20	-0.2507	0.0117	-21.46	-0.2545	0.0119	-21.47
Age20more	-0.3863	0.0160	-24.14	-0.3899	0.0162	-24.02
Y2002	0.0581	0.0163	3.56	0.0565	0.0165	3.42
Y2003	0.1252	0.0146	8.59	0.1294	0.0148	8.74
Y2004	0.1269	0.0126	10.09	0.1301	0.0128	10.13
Y2005	0.1671	0.0121	13.76	0.1676	0.0124	13.56
Y2006	0.2027	0.0154	13.18	0.2026	0.0154	13.13
Stratum	0.1472	0.0072	20.41	0.1905	0.0054	35.35
Ln(PM10)	-0.0908	0.0258	-3.52	-0.1448	0.0186	-7.77
Crime	-0.0039	0.0008	-4.62	-0.0037	0.0008	-4.42
Ln (Elevation)	0.0604	0.0065	9.29	0.0648	0.0066	9.85
Ln(Dist drainage ditch)	-0.0018	0.0041	-0.44	-0.0044	0.0042	-1.04
Ln(Dist main road)	0.0232	0.0048	4.86	0.0260	0.0048	5.37
Ln(Dist zon park)	-0.0045	0.0050	-0.90	0.0023	0.0052	0.44
Ln(Dist met park)	-0.0194	0.0045	-4.33	-0.0188	0.0045	-4.19
Ln(Commute Time)	-0.0643	0.0068	-9.41	-0.0680	0.0069	-9.82
$\ln \sigma_v$	-2.9732					
σ_v	0.2261					
Variance Function ($\ln \sigma_u^2$)						
Intercept	43.6137	7.4174	5.88			
Ln(pm10)	-13.3523	1.9915	-6.70			
stratum	-8.2407	1.2999	-6.34			
Ln(pm10)*stratum	2.3551	0.3489	6.75			

** F(29, 6514)= 2142.77

R-squared 0.9047

Table 3. Marginal implicit price for an increase in PM10

Socioeconomic stratum	Average Rent (2005 US$)*	Average PM10 (mg/m^3)	Marginal Implicit Price (2005 US$ per mg/m^3)	
			HFM	OLS
3	217.7	64.6	-0.31	-0.49
4	357.9	56.3	-0.58	-0.92
5	648.3	50.2	-1.17	-1.87
6	1358.9	43.0	-2.87	-4.58

* 1 dolar=2320.77 pesos

LIST OF FIGURES

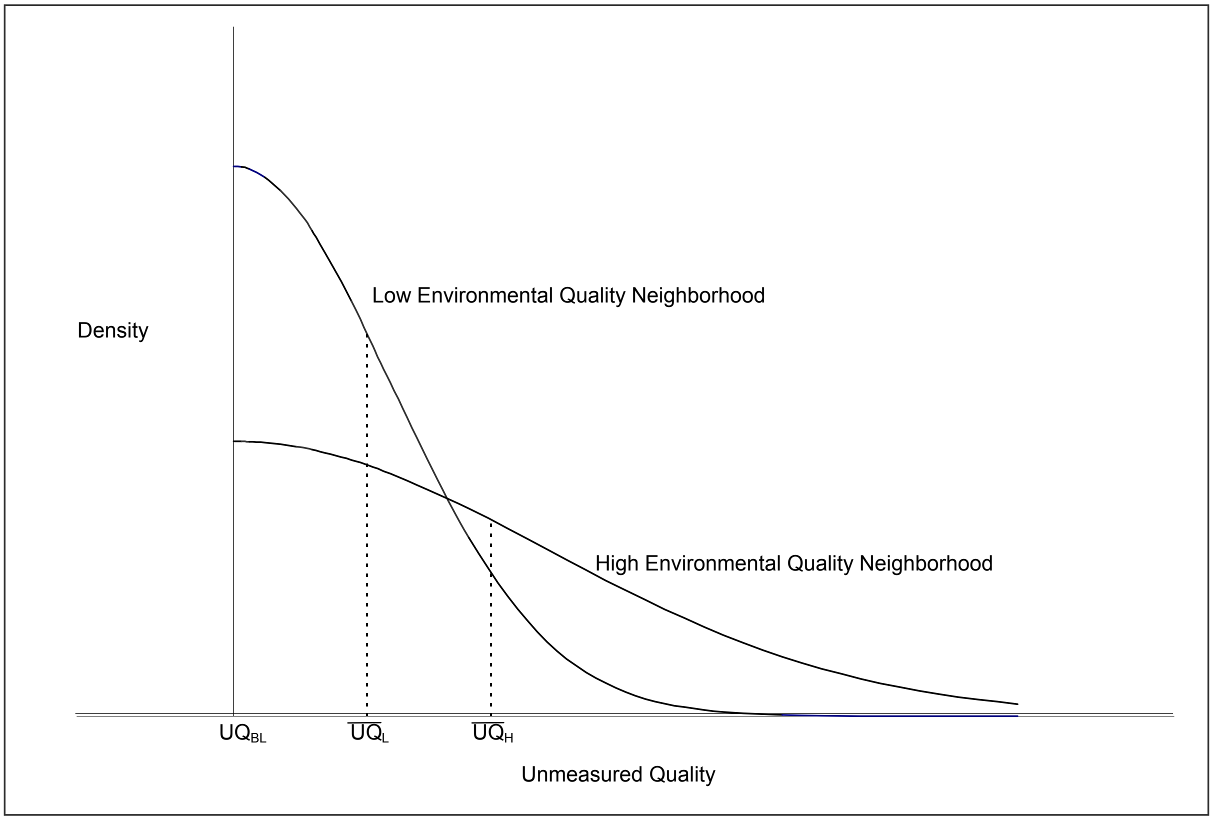

Figure 1. Distribution of Unmeasured Quality

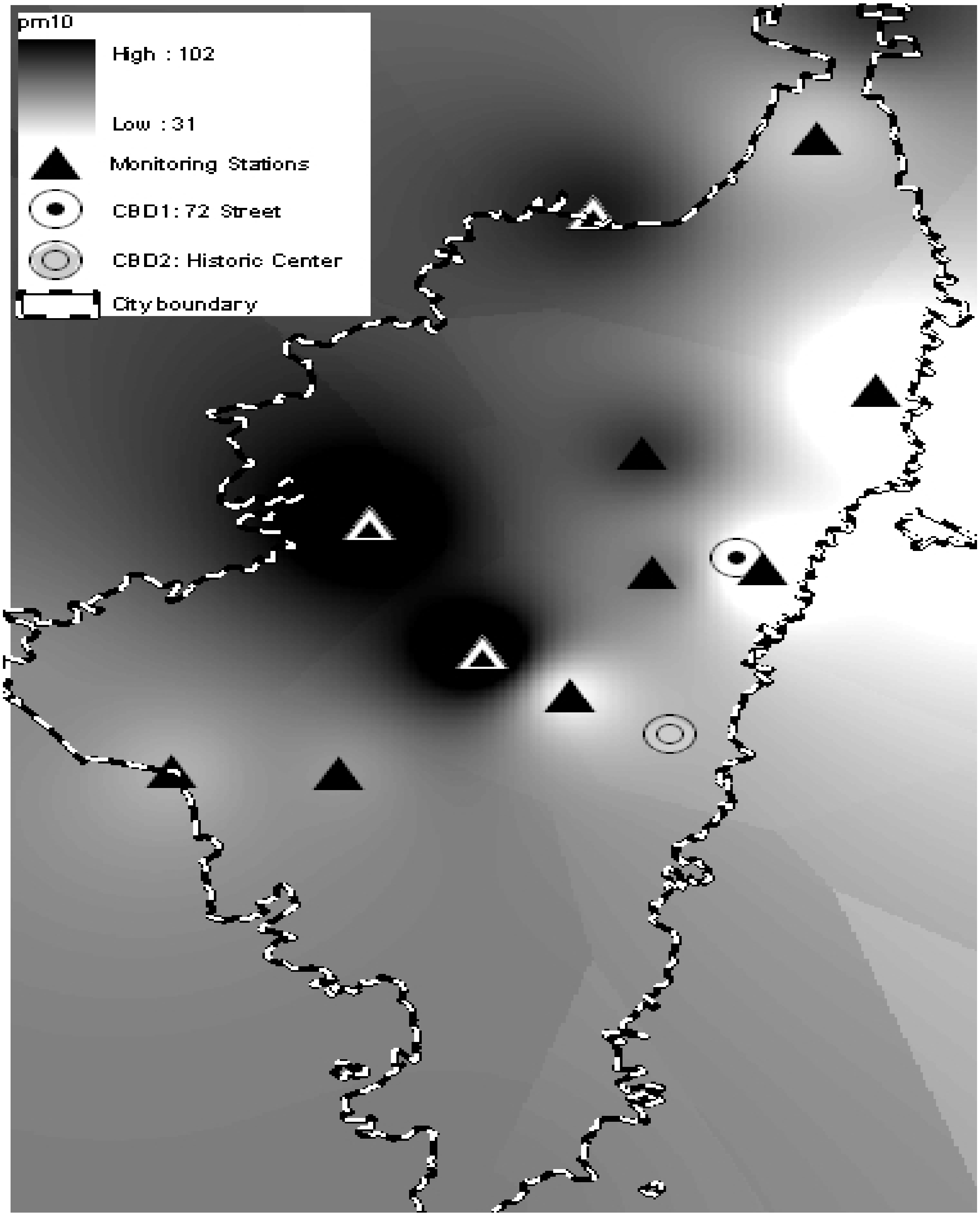

Figure 2. Spatial Distribution of PM10 (IDW Interpolation)

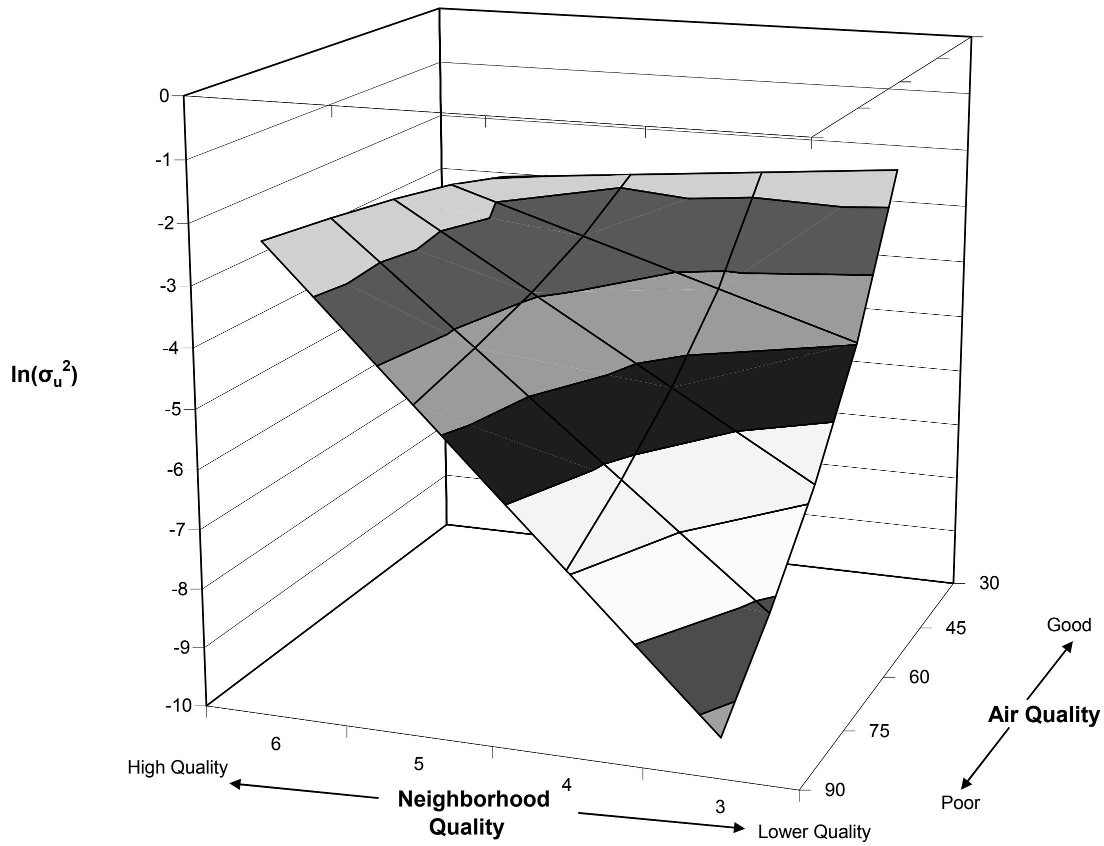

Figure 3. Impact of PM10 and Neighborhood Quality on the variance of the asymmetric error component.

APPENDIX

A1. Table A1. Estimation Results: Spatial Lag and Spatial Error Models

Variable	Spatial Lag Model Coeff	S.E.	t-stat	Spatial Error Model Coeff	S.E.	t-stat
Intercept	8.6063	0.0281	305.64	9.5780	0.0251	381.28
Ln(Area)	0.9226	0.0126	72.78	0.9169	0.0125	72.80
Bedrooms	-0.07236	0.0064	-11.22	-0.0692	0.0064	-10.77
Bathrooms	0.0375	0.0052	7.15	0.0375	0.0052	7.21
Carpet	0.0017	0.0094	0.18	0.0009	0.0090	0.10
Dinning-room	0.0829	0.0087	9.52	0.0816	0.0086	9.46
Electric Kitchen	-0.0366	0.0093	-3.90	-0.0398	0.0092	-4.29
Door Keeper	0.1696	0.0146	11.59	0.1594	0.0146	10.91
Floor number	0.0066	0.0001	5.04	0.0068	0.0013	5.28
Elevator 1	0.0913	0.0099	9.19	0.0914	0.0100	9.18
Elevator 2_3	0.1497	0.0129	11.55	0.1483	0.0129	11.51
Elevator 4_5	0.2145	0.0420	5.07	0.2136	0.0419	5.10
Age0_5	-0.0764	0.0130	-5.87	-0.0750	0.0128	-5.86
|Age5_10	-0.1725	0.0116	-14.79	-0.1690	0.0116	-14.62
Age10_20	-0.2548	0.0118	-21.58	-0.2470	0.0118	-20.98
Age20more	-0.3758	0.0161	-23.28	-0.3650	0.0161	-22.69
Y2002	0.0519	0.0165	3.13	0.0482	0.0164	2.94
Y2003	0.1256	0.0148	8.46	0.1203	0.0147	8.17
Y2004	0.1283	0.0128	9.98	0.1242	0.0127	9.75
Y2005	0.1670	0.0123	13.48	0.1600	0.0123	13.03
Y2006	0.2009	0.0154	12.99	0.1943	0.0153	12.66
Stratum	0.1691	0.0052	31.97	0.1702	0.0052	32.53
Ln(PM10)	-0.1348	0.0139	-9.65	-0.1840	0.0134	-13.70
Crime	-0.0025	0.0006	-3.80	-0.0020	0.0005	-3.97
Ln (Elevation)	0.0681	0.0060	11.23	0.0656	0.0064	10.19
Ln(Dist drainage ditch)	-0.0062	0.0042	-1.48	-0.0080	0.0043	-1.85
Ln(Dist main road)	0.0272	0.0047	5.67	0.0267	0.0048	5.58
Ln(Dist zon park)	0.0063	0.0050	1.26	-0.0200	0.0048	-4.18
Ln(Dist met park)	-0.0173	0.0043	-3.95	0.0098	0.0055	1.77
Ln(Commute Time)	-0.0689	0.0067	-10.19	-0.0690	0.0077	-8.97
Rho	0.0517	0.0056	9.23			
Lambda				0.5899	0.0036	161.70

R-squared for S-lag= 0.9047
R-squared for S-Error=0.9066

A2. To explore the potential bias that can occur when UQ is not considered in a hedonic pricing model, we simulated a house price dataset and estimated OLS and frontier models. The variables in the dataset were living area, ln(PM10) and two error terms, an asymmetrically-distributed term representing UQ and a symmetrically distributed term, ε. *Living area* and ln(*PM10)* came from the actual data of 6544 observations used in the hedonic estimations in the paper. UQ was drawn for each observation from a half-normal distribution, where the variance of the half normal is correlated with Ln(*PM10)* as follows

$$z_i \sim N(0, 43.61-13.35*Ln(PM10))$$

$$UQ_i = | z_i |$$

The symmetric random error component is drawn from a normal distribution, $\varepsilon \sim N(0,\sigma^2)$, with σ=0.2 . Based on this simulated data, ln(price) was simulated for each observation according to

$$Lnprice = 10 + 0.9 * Lnarea - 0.1 * Ln(PM10) + 1.0 * UQ + \varepsilon$$

The parameters of the hedonic price function, the variance function for the half-normal distribution, and the variance of the symmetric error component were chosen to be similar to those estimated in the frontier model presented in the paper in Table 2.

In order to show the potential bias that may arise from omitting construction quality and the suitability of the frontier model to correct this bias, three regressions were estimated:

Model 1 - an OLS regression that included ln(Area), ln(PM10) and UQ as explanatory variables

Model 2 - an OLS regression that included ln(Area) and ln(PM10), but that did not include UQ as an explanatory variable

Model 3 - a frontier model that included ln(Area) and ln(PM10) as explanatory variables, and that models UQ as an unobserved, asymmetric heteroskedastic error.

Estimation results from models (1), (2) and (3) are shown in Table A2.

Table A2. Estimation results from simulation

	Model (1)			Model (2)			Model (3)		
Variable	**Coeff**	**S.E.**	**t-stat**	**Coeff**	**S.E**	**t-stat**	**Coeff**	**S.E**	**t-stat**
Hedonic Price Function									
Intercept	10.0660	0.0661	166.07	10.8016	0.0559	193.14	10.05	0.0770	130.41
LnArea	0.8872	0.0046	191.03	0.8844	0.0048	181.65	0.8865	0.0048	185.14
LnPM10	-0.1026	0.0127	-8.08	-0.2788	0.0111	-24.94	-0.0990	0.0168	-5.87
CQ	1.0456	0.0409	25.55						
Symmetric Error Variance									
σ_v^2	0.0382			0.0420			0.0381		
Asymmetric Error Variance									
Intercept							43.693	4.678	9.34
LnPM10							-13.333	1.3447	-9.92

As would be expected, Model (1) results matched the original parameter values. The estimated implicit price for ln(PM10) differed by only 2.6% from the known value. A comparison of Models (1) and (2) reveals the potential bias of the parameter estimate for the air quality variable (LnPM10) when omitting unmeasured quality variables. Not accounting for UQ in this simulation inflates the pollution variable coefficient by 171.73%. However, the heteroskedastic frontier model did a good job identifying the correct parameters for the hedonic price function and for the variance function. The estimated implicit price for ln(PM10) from the HFM model differed by only 1% from the known value. These results suggest that implementing the frontier approach in the hedonic model estimation does mitigate the potential bias that arises from omitting unmeasured quality variables that are asymmetrically distributed and correlated with environmental quality levels.