

Article

Using Co-Ordinate Systems in Hedonic Housing Regressions

Steven B. Caudill ^{1,*}, Neela Manage ¹ and Franklin G. Mixon, Jr. ² ¹ Department of Economics, Florida Atlantic University, Boca Raton, FL 33431, USA; manage@fau.edu² Center for Economic Education, Columbus State University, Columbus, GA 31907, USA; mixon_franklin@columbusstate.edu

* Correspondence: scaudill@fau.edu

Abstract: Hedonic house price studies typically incorporate information about location by including either a set of dummy variables to represent individual locations called “neighborhoods” or by using a set of distance (or travel time) variables to characterize locations in terms of proximity to amenities and dis-amenities. As an alternative to these, relatively recent research advocates a latitude–longitude co-ordinate system for incorporating distance information into hedonic house price regressions. This study shows that many of the claims made in this research, particularly those referencing the elimination or diminution of “biases of coefficients of non-distance variables”, are given the particulars of the Monte Carlo experiments, not possible to investigate. We further show, both analytically and with our simulations, that there is no omitted variable bias present in their simulations because their randomly generated non-distance variable is uncorrelated with any of the other variables used in their regression models.

Keywords: house prices; hedonic regression model; distance; location; co-ordinate systems



Citation: Caudill, S.B.; Manage, N.; Mixon, F.G., Jr. Using Co-Ordinate Systems in Hedonic Housing Regressions. *Real Estate* **2024**, *1*, 41–64. <https://doi.org/10.3390/realestate1010004>

Academic Editor: Pierfrancesco De Paola

Received: 24 January 2024

Revised: 5 March 2024

Accepted: 5 March 2024

Published: 12 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Hedonic regression models [1] of house prices increasingly incorporate information about location into both the non-stochastic (the regression model) and stochastic (error structure) components of the empirical model (e.g., see the reviews in [2,3]). In hedonic house price studies, information about location has typically been incorporated into the non-stochastic portion of regression models in two ways: either including a set of dummy variables to represent individual locations called “neighborhoods” (e.g., see [4–8]) or using a set of distance (or travel time) variables to characterize house locations in terms of proximity to amenities and dis-amenities (e.g., see [9–13]). The main advantage of using dummy variables is that locations of amenities and dis-amenities need not be known, and the main disadvantage of this approach is that it requires the estimation of many additional parameters, although this is not usually a serious limitation as housing data sets are typically quite large. An alternative is the use of “distance” measures, either distance or travel time, to amenities and dis-amenities. The main advantage of this approach is that fewer parameters need to be estimated compared to the use of dummy variables. The main disadvantage is that the locations of the amenities and dis-amenities must be known.

Ross et al. [14] present an alternative to the use of distance variables in hedonic house price regressions that instead uses a latitude–longitude co-ordinate system that requires no information on the number of amenities (dis-amenities) or their locations. One virtue Ross et al. [14] claim for the use of a co-ordinate system is that it can eliminate omitted variable bias in the estimation of the coefficients of non-distance variables in the case where the locations of the amenities are unknown or incorrect. We conducted a very extensive set of Monte Carlo simulations in order to compare various model specifications, including the Ross et al. [14] co-ordinate system to examine this claim. These comparisons show that the claim made by Ross et al. [14] cannot be established because there is no omitted variable bias to eliminate in any of the specifications due to the nature of their non-distance variables.

The remainder of the paper is organized as follows: Section 2 contains a review of recent studies that account for the location of residences, with particular interest in how location is measured. In Section 3, we briefly describe the Ross et al. [14] approach, followed by a discussion of the peculiar features of the Ross et al. [14] simulations. In Section 4, we present the results of our Monte Carlo simulations and compare our results with those reported by Ross et al. [14]. In Section 5, we conduct an additional Monte Carlo experiment to directly assess the value of the co-ordinate system for eliminating omitted variable bias. Section 6 contains a summary and conclusions.

2. Literature Review

As indicated above, real estate pricing models have historically included location information either by including a set of dummy variables to represent individual locations called “neighborhoods” or by using a set of distance variables to characterize house locations in terms of proximity to (dis-)amenities. There are benefits to each approach. The former approach is useful when the location of the source of an externality is unknown, while the latter approach provides a more parsimonious specification. In the two subsections below, we describe some recent studies utilizing one or the other approach. The comprehensive bibliometric analysis by Li and Li [15] lists additional studies that fit one or the other category. Lastly, some of the studies reviewed in this section make use of spatial econometrics techniques. In addition to the real estate literature (e.g., [16–18]), these models are also used to examine environmental issues (e.g., [19,20]).

2.1. Studies Using Dummy Variables to Represent “Neighborhoods”

Lee et al. [5] examine the relationship between hedonically controlled housing price levels and subsequent changes in those prices across metropolitan areas in order to assess whether areas with a high price relative to an “imputed rent” are paying for higher appreciation. The study finds that identical houses in higher-priced ZIP codes appreciate faster than their counterparts elsewhere [5]. To reach this conclusion, Lee et al. [5] utilize repeat-sales housing price indices at the ZIP-code level that include a diverse set of demographic, geographic and housing market-related conditions.

A subsequent study by Sah et al. [6] utilizes data on 20,000 housing sales in San Diego County in order to investigate the presence (or absence) of a “school proximity premium” in hedonic residential pricing models. Their results indicate the presence of a school proximity “penalty”, with home prices rising by about 0.8% with every 1000 feet of distance between the home and a school, thus suggesting that schools represent a net nuisance to homeowners. To account for San Diego’s heterogeneous geospatial attributes (e.g., canyons, mountains, coastline and open spaces), Sah et al. [6] include robust spatial controls, particularly those identifying neighborhoods. These include ZIP codes, census tracts, elevation and distance to the closest library, shopping mall, open space and retail center, among others [6].

Wolf and Klaiber [7] explore the impact of harmful algae in lakes near homes located across six separate Ohio counties. In doing so, they report that reductions in lake water quality reduce the value of near-lake homes by 11% to 17%, while lower quality negatively impacts the price of lake-adjacent homes by 22% [7]. To measure proximity to a lake, Wolf and Klaiber [7] georeferenced each house sold to a spatial location using GIS data provided by county officials. This process linked spatial characteristics for each residential property, such as census block location and the distance to the nearest lake, park and school. Lastly, lakefront (near-lake) properties were classified as those within 20 m (300 and 600 m) of a lake [7]. In their study of food deserts near Memphis, which utilizes data from 3298 real estate transactions, Caudill et al. [8] report that location in a food desert (i.e., more than 10 miles from a grocery store) reduces the value of residences by 4% to 6%, *ceteris paribus*. This study uses location measures (e.g., ZIP codes) in a regression model that accounts for spatial correlation. The model employed to obtain the impact stated above is based on the three closest neighbors [8].

2.2. Studies Using Distance Variables to Characterize Location to (Dis-)Amenities

A study by Caudill et al. [9] uses data on 2026 residential real estate transactions in Memphis to examine the impact of the location of sex offenders' residences on house prices. The examination suggests that each additional sex offender residence reduces the value of homes situation within one mile of the domicile by 2% [9]. The regressors of interest in the models producing the above result are the distance to the nearest sex offender's residence and the concentration of such residences within a one-mile radius of the location of each property sold. Caudill et al. [9] also include ZIP-code fixed effects to capture the impact of neighborhoods. Liao et al. [10] investigate the negative impact of invasive watermilfoil in a large lake in northern Idaho on residential property values around the lake. They find that the presence of the invasive plant, which interferes with recreational use of the lake, reduces mean home values by 13% [10]. The property data utilized in the study were geocoded with shape files, and lakefront properties were defined as being within 500 feet of the lake.

A study by Affuso et al. [11] that examines distance to cellular towers from residential properties in southern Alabama reports that homes located within three-quarters of one kilometer of a cellular tower sell for about 2.5% less than those located farther away, while residences located within a visible range of a cellular tower suffer a 9.8% loss in value compared to those located outside of a visible range. The study uses latitude and longitude to plot the location of each residence, which is compared to a map of cellular tower locations in order to compute the distance between residences and cellular towers. A software tool known as Viewshed is used in combination with a digital elevation map to construct the cellular tower visibility variable [11].

Li and Li [13] examine the impact of the distance between residential properties and a landfill location on housing prices in Hong Kong in order to account for the negative externality (i.e., bad smell) produced by waste collection at a landfill. Regarding the impact on house prices from the location of a landfill, the study not only includes a measure of distance from the landfill, but it also includes the distance between a residence and the main road used by garbage trucks to transport waste from residential neighborhoods to the landfill [13]. In the case of the former, a one-unit increase in distance creates a five-fold increase in housing prices. In terms of the latter, a one-unit increase in distance is associated with between a two- and three-fold increase in housing prices [13]. Affuso et al. [12] assess the impact of noise on residential prices in Memphis, which is home to the busiest cargo airport in the U.S. Analysis of 9606 residential real estate transactions reveals that the noise surrounding Memphis International Airport is considered to be a dis-amenity, as it is associated with a potential per household average external cost of almost USD 5000 per decibel of noise [12]. This result is supported by a newer study by Tsao and Lu [21] that examines the impact of noise emanating from Taoyuan International Airport on house prices in Taoyuan City, Taiwan. Lastly, Affuso et al. [12] account for the impact of wind on noise travel, a georeferenced soundscape map to measure airport noise, and the curvature of the earth in plotting distances from residential locations to the closest major road, the Mississippi River and four major open space areas.

Chen et al. [22] investigate the relationship between rail accessibility, defined by the Euclidean distance from a residential community to the nearest rail transportation station, and house prices in Fuzhou, China. To reach a conclusion about this particular relationship, the authors model road accessibility using two Space Syntax indicators (i.e., connectivity and carrying capacity) in accordance with the spatial pattern of the road network. The results from a spatial interactive regression model suggest that road connectivity has a significant regulating effect on the significantly negative impact of the distance to the closest rail station on house prices [22].

Aziz et al. [23] investigate the impact of distance to neighborhood services, such as a hospital or market, and property attributes on land and property values in the congested urban area of Gujrat, Pakistan. In doing so, they find that although the distance from neighborhood services and land/property values are negatively related, as expected, these

relationships are marginally insignificant. The empirical approach used by Aziz et al. [23] includes the mapping of data collection points in Gujrat and their distance from neighborhood services in terms of spatial analysis. A contemporaneous study by Peng et al. [24] examines the impact of various types of “urban blue spaces” on house prices across eight megacities in China. The types of urban blue spaces considered in the study include metro stations, schools, hospitals, shopping centers, parks, lakes and rivers, some of which ameliorate air pollution and improve the thermal environment [24]. To conduct their analysis, the authors combine geographic coordinates of all communities through Amap API with point-of-interest data on the location of the urban blue spaces. The results indicate that distances from the types of urban blue spaces under consideration were generally negative and significant across all eight urban areas [24].

3. The Latitude–Longitude Model

Ross et al. [14] present an alternative to the use of distance variables in hedonic house price regressions. Their alternative replaces the distance variables with a latitude–longitude co-ordinate system that does not require information on the locations or number of amenities or dis-amenities. In its simplest form, the regression model they propose is

$$HP_i = Z_i + \theta_1 x_i + \theta_2 y_i + \theta_3 x_i^2 + \theta_4 y_i^2, \quad (1)$$

where HP is house price, and Z_i is a vector of non-distance house characteristics and their coefficients, including an intercept. The x and y variables refer to the latitude and longitude of the house location along with their squares, and the θ s represent the parameters to be estimated.

One important implication of this formulation can be seen by partially differentiating the regression model with respect to x and y . By doing so and solving yields,

$$\frac{\partial HP_i}{\partial x_i} = \theta_1 + 2\theta_3 x_i = 0 \Rightarrow x_i^* = \frac{-\theta_1}{2\theta_3}, \quad (2)$$

and

$$\frac{\partial HP_i}{\partial y_i} = \theta_2 + 2\theta_4 y_i = 0 \Rightarrow y_i^* = \frac{-\theta_2}{2\theta_4}. \quad (3)$$

These solutions indicate that underlying Ross et al.’s [14] co-ordinate system is the idea that the neighborhood is characterized by a single maximum property value. One obvious problem with this analytic solution is that it will not exist if either θ_3 or θ_4 , or both, equal zero. Ross et al. [14] illustrate some of the problems with the use of distance variables in regressions and investigate the relative merits of their latitude–longitude approach using three simulations: their *Illustrative Simulation*, *Simulation I* and *Simulation II*. We discuss each of these simulations below.

The Ross et al. [14] *Illustrative Simulation* is not really a Monte Carlo simulation in the usual sense, as the results are based on a single sample, and the true regression model contains no random error term. This model is used to illustrate the consequences of omitting and/or redefining independent variables on model performance measured in terms of slope estimates and model R^2 values. They estimate four models, all of which are variations of their true model. Two of the four models are straightforward applications of omitted variable results and do not require a simulation for analysis [14]. We discuss omitted variable bias and apply omitted variable results to obtain parameter estimates for these two cases in Appendix A.

The sample space used in Ross et al.’s [14] *Simulations I* and *II* is based, in part, on locations inside a 40×40 square centered at the origin. They also make use of a smaller square inside this larger square centered at the origin, which is of length 20. Ross et al.’s [14] *Simulations I* and *II* are based on the two models below, where A and B refer to the true

landmarks, and D^A and D^B are the distances to the true landmarks. A non-distance variable, x , is also included, yielding the following regression models:

$$HP_i^A = \beta_0 + \beta_1 x_i + \gamma_2 D_i^A + \varepsilon_i, \quad (4)$$

and

$$HP_i^{AB} = \beta_0 + \beta_1 x_i + \gamma_2 D_i^A + \gamma_3 D_i^B + \varepsilon_i. \quad (5)$$

Models (4) and (5) represent the true models or data-generating processes. Ross et al. [14] focus on the statistical properties of the estimators of slope parameters in these models, particularly those of the non-distance variable x , when the true landmark locations are unknown, and distances to the mis-measured landmarks, D^a and D^b , are used instead. These estimated models are calculated using

$$HP_i^a = \alpha_0 + \alpha_1 x_i + \delta_2 D_i^a + \varepsilon_i, \quad (6)$$

and

$$HP_i^{ab} = \alpha_0 + \alpha_1 x_i + \delta_2 D_i^a + \gamma_3 D_i^b + \varepsilon_i. \quad (7)$$

These incorrect landmark locations are generated randomly for each of Ross et al.'s [14] Monte Carlo experiments. As a result, there are always three possible models to be estimated: one model using distances to correct landmarks, one model using distances to incorrect landmarks, and the latitude–longitude version of the hedonic regression. Ross et al. [14] estimate separate distance models, with one based on incorrect landmarks, and another based on true landmarks *and* the latitude–longitude version of the hedonic regression. Ross et al.'s [14] simulations use 400 observations per trial and 1000 trials.

3.1. Simulation I

For Monte Carlo *Simulation I*, Ross et al. [14] choose random house locations inside the inner square, which we call the *neighborhood*. The locations outside the inner square but inside the larger square, we refer to as the *periphery*. The landscape features two true landmarks, A and B , both located in the periphery. Their co-ordinates are $A = (0, 15)$ and $B = (10, 15)$, both unknown to the researcher. Instead, random landmarks are chosen from the periphery, and distances to each are used in the estimation. All of the random house locations are chosen from the neighborhood. Ross et al.'s [14] *Simulation I* examines the consequences for the estimators in two cases. In the first case, the estimated model contains distances to one or two incorrect landmarks when the true model contains the distance to a single true landmark. In the second case, the estimated model contains distances to one or two incorrect landmarks when the true model contains distances to two true landmarks. In other words, Ross et al. [14] investigate parameter estimates if the true model is given by (4) above, but instead (6) and (7) are estimated, as well as what happens if the true model is (5) above, but models (6) and (7) are estimated. For these models, Ross et al. [14] obtain parameter estimates using several Monte Carlo experiments. Ross et al. [14] assume the coefficient of x is 2, x is an $N(0, 1)$ random variable, and the error term is an $N(0, 2)$ random variable. As previously noted, the true locations of A and B are $(0, 15)$ and $(10, 15)$, respectively. The true values of the model parameters are $\beta_0 = 1$, $\beta_1 = 2$, $\gamma_1 = -0.25$, and $\gamma_2 = -0.1$.

The regression models in Ross et al.'s [14] Monte Carlo *Simulation I* are based on distances to incorrect landmarks that also change from experiment to experiment. Based on their simulation results, they note that the explanatory power of the model is, on average, higher when the estimated model contains two incorrectly measured distance variables rather than one incorrectly measured distance variable, whether the data-generating process contains one or two true landmark distances (see Appendix B for a presentation of the Ross et al. [14] *Simulation I* results, along with our own re-estimation). From this simulation, Ross et al. [14] also conclude “that incorrectly identifying the landmark has consequences for producing unbiased estimates of the non-distance variables in the re-

gression". As we show below, this assertion is not true because the non-distance variable x in the Ross et al. [14] simulations is not correlated with any potentially included independent variable, so the slope coefficient of x is not subject to omitted variable bias in these simulations.

3.2. Simulation II

Their Monte Carlo *Simulation II* uses the same two true models as in the Monte Carlo *Simulation I* given in 4 and 5, but now, the true landmarks change from experiment to experiment [14]. Estimates from this model are compared to an alternative, which is the latitude–longitude version of the model preferred by Ross et al. [14]. The estimated distance models include their non-distance variable x and either one or two distances to true landmarks. The alternative model includes their non-distance variable but uses the Ross et al. [14] co-ordinate system in place of distances to the true landmarks (see Appendix C for details on the Ross et al. [14] *Simulation II* results, which we repair). The locations of the true landmarks and the housing units are randomly chosen over the larger square (i.e., the neighborhood and the periphery combined).

3.3. Conclusions and Conjectures

There are two main conclusions reached by Ross et al. [14] regarding the virtues of using the latitude–longitude system in place of distance variables. The first is that omitted variable bias can be eliminated or greatly reduced in the estimates of the parameters of non-distance variables, even if the true landmark locations are unknown, provided that the latitude–longitude version of the hedonic regression model is estimated. The second is that multicollinearity is reduced, and the possibility of extreme multicollinearity is eliminated as the use of sometimes highly correlated distance variables is avoided. We fully address these claims and more in the next section after presenting and discussing our simulation results. However, before doing so, we discuss several unusual features of the Ross et al. [14] simulations.

3.4. Peculiar Features of Simulations I and II

Ross et al.'s [14] *Simulation I* is based on distances to two true fixed landmarks in the periphery, random house locations chosen from the neighborhood and distances to randomly chosen incorrect landmarks from the periphery. From these samples, Ross et al. [14] estimate models with one or two distance variables based on the inclusion of incorrect landmarks and compare estimation results for the two when the underlying true model contains either one true distance variable or two true distance variables. Their *Simulation II* is based on one or two randomly chosen landmarks from the neighborhood and periphery combined. Random house locations are chosen from the neighborhood and periphery combined. In each case, the true model is estimated, the latitude–longitude version of the model is estimated, and the results are compared [14]. Their *Simulation I* parameter estimates are obtained from regression models based on distances to incorrect landmarks, while their *Simulation II* includes only the true regression model and the associated latitude–longitude version. None of the Ross et al. [14] simulations estimate all three possible models, and the sample spaces are different for *Simulations I* and *II*. This setup greatly complicates comparisons.

The construction of the non-distance variable x is unusual. As previously mentioned, this variable is a normally distributed random variable with a mean of zero. Thus, even with a nonzero-associated coefficient, the variable's average contribution to the regression model is nil. In effect, this non-distance variable is akin to having a second error term in the model. Another peculiarity related to the first is found in the construction of the regression models used in the simulations; the problem is that variation in the generated house prices is largely dominated by the randomly generated non-distance variable and the randomly generated error term. Consider the Ross et al. [14] model with two known landmarks, 1 and 2, calculated using

$$HP_i = 1.0 + 2.0x_i - 0.25D_i^1 - 0.10D_i^2 + \varepsilon_i. \quad (8)$$

Recall that for the Ross et al. [14] Monte Carlo *Simulation I*, house locations are randomly chosen from a neighborhood square of length 20 centered at the origin. The locations of the true landmarks in their *Simulation I* are $A = (0, 15)$ and $B = (10, 15)$. The most distant location from points A and B inside the neighborhood is at the point $(-10, -10)$. From this location, the distance to landmark A is 27 (i.e., $\sqrt{10^2 + 25^2}$), and the distance to landmark B is 32. In the regression model above, the penalty subtracted for this maximum distance to A is one-fourth of 27 or 6.75, and the value subtracted for this maximum distance to B is one-tenth of 32 or 3.2. These penalties combined are likely to be dominated by the contributions of the non-distance variable and the error term. The coefficient of the non-distance variable is 2.00, and the values of the non-distance variable will typically fall between -2 and $+2$. This implies that the contribution to the hedonic regression from the distance variable will fall between -4 and $+4$ for a range of 8. The same is true for the random error term, which has a standard deviation of 2; the bulk of the values will fall between -4 and $+4$. Together, these two random components of the regression model can cause house prices to go up or down by approximately eight for a range of 16. Thus, the distance variables based on the most distant locations have about sixty-three percent of the impact of the non-distance variable and the random error combined. For less distant locations, the impact of the distance variables on house prices will diminish, but the impact of the non-distance variable and the error term will not. For the not-so-distant housing units, the random component is likely to be more dominant. This feature of the Ross et al. [14] simulations complicates pinpointing optimal house locations because of the enormous random component in house prices built into the simulation.

A third issue concerns the locations of the true landmarks A and B in Ross et al.'s [14] Monte Carlo *Simulation I*. The fact that they are both outside of the neighborhood square and on the same side of the neighborhood makes for an easier estimation problem because, on average, the highest-valued houses will be on the "north side" of the neighborhood. Compare this situation with a single centrally located landmark. From sample to sample, the estimated "optimal location" can be pulled in *any* direction. More curious, perhaps, is the fact that Landmark B does not matter for the determination of the optimal location as the central location and higher distance penalty (the coefficient) associated with Landmark A means that distances to B only impact the final house price but not the optimal location. In effect, the two-landmark model using distances to both A and B is essentially a one-landmark model consisting of Landmark A .

Below, we show that the highest value, on average, is the point in the neighborhood $(0, 10)$ closest to Landmark A . From this location, the distance to Landmark A is 5.0, and the distance to B is 11.18 (i.e., $\sqrt{125}$). Thus, the penalty for locating at $(0, 10)$, the center north position in the neighborhood, is 2.37 (i.e., $0.1(11.18) + 0.25(5) = 1.12 + 1.25$). Movements due east from this center north point will increase the distance to Landmark A and decrease the distance to Landmark B until the northeast corner of the neighborhood is reached at $(10, 10)$. The penalty at this corner is approximately 3.29 (i.e., $0.25(11.18) + 0.1(5) = 2.79 + 0.5$). Thus, the costs are lower at the center north point *even with two landmarks*. Although house prices will change in the two-landmark model because there is a second distance penalty to consider, the optimal location will not change; thus, Ross et al.'s [14] two-landmark model in *Simulation I* is, for the purposes of finding the optimal location, actually a one-landmark model.

Another unusual feature of the Ross et al. [14] simulations is the relationship between their sample space and their sample sizes. The landscape used by Ross et al. [14] is typically based on a square of length 40 centered at the origin. This large square contains 1600 unit squares. The Ross et al. [14] sample size is 400, which means that in each Monte Carlo experiment, at least 75% of these unit squares contain zero housing units, which is a problem that is mitigated when the inner square or neighborhood is used for the sample space. This is a very sparse landscape, which, along with the extra variation induced by

the construction of their non-distance variable, means there will be considerable parameter fluctuation from sample to sample. This fact prompted us to use a much larger number of Monte Carlo trials in our own simulations.

Ross et al.'s [14] *Simulation I* is based on distances to two known landmarks, but the true model is not estimated. Instead, Ross et al.'s [14] Table 1 compares estimates from *two different mis-specified models* based on incorrect landmarks for each of the two true underlying DGPs. In *Simulation II*, Ross et al. [14] estimate the true model and the latitude–longitude version of the model with the locations of the underlying true landmarks known but changing from sample to sample. The changing locations of the true landmarks add an additional layer of variability to the estimation results, which are already highly variable due to the unusual structure of the non-distance variable discussed earlier and the sparse coverage of the landscape afforded by the relatively small sample size in a relatively large neighborhood.

Finally, we note that there are two obvious omissions in the Ross et al. [14] work. These are issues mentioned frequently by Ross et al. [14] but not investigated in their paper. Despite their claim that the latitude–longitude approach is useful when the locations of the true landmarks are unknown or likely incorrect, Ross et al. [14] do not directly compare their latitude–longitude formulation to a distance regression model based on distances to incorrect landmarks. A second omission concerns the optimal location in space. Ross et al. [14] also frequently mention finding optimal locations in space as one of the virtues of their co-ordinate system, again, without any investigation of the issue. We address these omissions in our Monte Carlo simulations in the next section.

4. Our Monte Carlo Simulations

We conducted an extensive set of Monte Carlo experiments to examine omissions and extensions of the Ross et al. [14] analysis. We combined features of Ross et al.'s [14] Monte Carlo *Simulations I* and *II* into a single simulation. We estimated housing models containing from one to five known landmarks. The locations were fixed and given as follows: L1 = (−15, −15), L2 = (10, 10), L3 = (−5, −5), L4 = (0, 0) and L5 = (5, −13). The true model in each case is given by applying zero restrictions on the coefficients of the model below to change the number of distance variables included from one to five:

$$HP_i = 1.0 + 2.0x_i - 0.25D_i^1 - 0.10D_i^2 - 0.15D_i^3 - 0.10D_i^4 + 0.10D_i^5 + \varepsilon_i. \quad (9)$$

We also included the Ross et al. [14] non-distance variable x . For each version of the specification above, we estimated three models: the true model using distances to the true landmark(s), a model using distances to randomly chosen incorrect landmarks, and the Ross et al. [14] latitude–longitude version of the model. From each trial, we collected information on the parameter estimates, the model R^2 , the estimate of σ^2 and the implied optimal location, that is, the location of the highest property value. We used the same sample size (i.e., 400) as Ross et al. [14] but conducted one million Monte Carlo trials in an effort to reduce the sample variation we discussed previously. The sample space was a 40×40 square centered at the origin from which we randomly generated house locations.

The simulation results begin with Table 1. Panel A contains the average regression statistics for each model. Panel B contains the model estimates for the co-ordinates of the optimal locations in space. For each trial, the optimal location in space was determined by recording the co-ordinates of the highest predicted house price for each model. These values were averaged over the simulation to obtain means and standard deviations for each set of co-ordinates. We also calculated and tracked the co-ordinates of the optimal location obtained using the analytic solution given above in Equations (2) and (3). This solution is of considerable practical importance in applied work. The fact that this approach uses estimated model parameters allows one to obtain point estimates of the optimal location and the estimated variances of these co-ordinates from a single sample. Approximate variances can be obtained using the delta method (see Greene [25], pp. 172–173). A

standard error for the maximum predicted value from any hedonic regression model can be obtained using the bootstrap.

Table 1. Model with one true distance variable.

Panel A: Coefficients			
Variable	Correct Landmarks	Latitude–Longitude	Incorrect Landmarks
Intercept	1.000 (0.25)	−4.296 (0.19)	−4.645 (3.64)
Non-Distance	2.000 (0.10)	2.000 (0.10)	2.000 (0.14)
D1	−0.250 (0.01)		
d1			−0.025 (0.17)
X		−0.159 (0.01)	
Y		−0.159 (0.01)	
X ²		−0.007 (0.00)	
Y ²		−0.003 (0.00)	
R ²	0.744 (0.02)	0.725 (0.02)	0.443 (0.15)
σ ²	4.000 (0.28)	4.321 (0.31)	8.705 (2.42)
Panel B: Optimal Location			
Model	X Co-Ordinate	Y Co-Ordinate	House Price
Correct Landmarks	−13.965 (3.56)	13.963 (3.56)	4.058 (0.84)
Latitude–Longitude	−13.418 (4.99)	13.421 (4.98)	3.205 (0.92)
Analytic Solution	−27.981 (887.63)	29.030 (298.18)	
Incorrect Landmarks	−6.323 (10.97)	6.326 (10.97)	1.806 (1.34)

Note: Standard errors in parentheses.

The simulation results for models based on a single distance variable are given in Panel A of Table 1. We first examined the model R^2 values. The R^2 for the true model is 0.744, for the latitude–longitude model, 0.725, and for the model with incorrect landmarks, 0.443. We see that the true model and the longitude–latitude versions offer very similar explanatory power, with the model based on incorrect landmarks performing relatively poorly. Also worthy of note is that the estimated standard deviation of the model R^2 values for the true model and the latitude–longitude versions is 0.02, but the value for the model based on the distance to an incorrect landmark location is a much higher 0.15. We calculated the correlations between these R^2 values over the simulation. For the true model and the latitude–longitude version, the simple correlation is 0.94. For the true model and the latitude–longitude version, the correlations with the incorrect landmark model are 0.11 and 0.12, respectively. The high correlation between R^2 values from the true landmark model and the latitude–longitude version, along with the low standard deviations of the R^2 values,

indicates that both models explain the data well and that the difference in performance between the two is small in this case. The model based on incorrect landmarks explains the data relatively poorly compared to the other two models, and the explanatory power does not track the others very closely. Although the average R^2 values are lower for the incorrect landmark model, the maximum values of R^2 are nearly equal. This is not surprising as the probability of choosing a random landmark near the true landmark is high in one million samples. However, this is a reminder that the landmark must be at some location in the neighborhood, and, though unknown, searching for the landmark location with a computer is a possibility.

If we turn now to the parameter estimates, Column Two provides the results obtained from the estimation of the true model. The averages of all estimated model parameters, including the estimate of σ^2 , are equal to their true values. For the latitude–longitude version of the model, we find that the coefficient of the non-distance variable is also 2.000 and equal to its true value. The estimated value of σ^2 is somewhat high at 4.321 compared to the true value of 4.000. For the model based on incorrect landmarks, the coefficient of the non-distance variable is also 2.000. Thus, all cases support our claim about the absence of omitted variable bias. The estimate for σ^2 is much higher at 8.705 compared to the true value of 4.000. The reason for the absence of this bias is shown analytically in Appendix A.

Panel B of Table 1 contains co-ordinate estimates for optimal locations. The correct landmark model yields the average location of the maximum to be (−13.965, 13.963), with the average maximum house price of 4.058. The latitude–longitude version yields the optimum co-ordinates of (−13.418, 13.421), with an average maximum house price of 3.205. The analytic solution provides nothing useful at this point but will be of value for other models. This may be a result of division by zero as we indicated earlier. For the model with incorrect landmarks, the location of the highest price is (−6.323, 6.326), with an average maximum house value of 1.806. The first two models give somewhat similar results, but the model based on a single incorrect landmark is far from the results of the other two models.

The simulation results for models based on two true distance variables are given in Table 2. First, we note the average model R^2 values in Panel A. All three R^2 values increased over the single landmark results. For the true model, we obtained 0.761; for the latitude–longitude version, we obtained 0.749; and for the model based on incorrect landmarks, we obtained 0.590. The first two R^2 values are close, as before, and the model based on incorrect landmarks is closing the gap with the other two and enjoying a reduction in variance. The correlation between R^2 values for the true model and the latitude–longitude version is 0.96. The correlation between R^2 values for the incorrect landmark and either of the other two is about 0.14.

Column Two of Panel A in Table 2 shows the coefficient estimates for the true model. Not surprisingly, the average estimated values of 1.000, 2.000, −0.250 and −0.010 are all exactly equal to their true values, as is the value of σ^2 , which is equal to its true value of 4.000. For the latitude–longitude model, the R^2 is slightly lower than for the true model, as we saw previously, and, as before, the estimated value of σ^2 is larger than for the true model at 4.217. Again, the value of the non-distance coefficient is exactly 2.000. Column Four contains the estimation results for the model based on two incorrect landmarks. The coefficients are far different from the true value save for, once again, the coefficient of the non-distance variable, which is 2.000 on average. The estimated value for σ^2 is the highest in the group at 6.854.

Panel B of Table 2 shows the average optimal location estimates from each of our models. Using the correct landmarks, the co-ordinates of the average optimum location are (−11.426, 13.660). The average house price there is 1.715. For the latitude–longitude version, the average co-ordinates of the optimal location are (−8.975, 13.859), with an average house price of 1.143. The analytic solution gives the optimal co-ordinates as (−10.797, 20.780). Using two incorrect landmarks yields an average optimal location of (−6.024, 11.381), with an average house price of 0.990. The optimal co-ordinates of the true model and the

latitude–longitude version are similar, but the average predicted optimal house prices are closer for the latitude–longitude version and the incorrect landmark version.

Table 2. Model with two true distance variables.

Panel A: Coefficients			
Variable	Correct Landmarks	Latitude–Longitude	Incorrect Landmarks
Intercept	1.000 (0.32)	−5.692 (0.19)	−5.247 (4.72)
Non-Distance	2.000 (0.10)	2.000 (0.10)	2.000 (0.13)
D1	−0.250 (0.01)		
D2	−0.100 (0.01)		
d1			−0.050 (0.60)
d2			−0.049 (0.60)
X		−0.109 (0.01)	
Y		0.209 (0.01)	
X ²		−0.005 (0.00)	
Y ²		−0.005 (0.00)	
R ²	0.761 (0.02)	0.749 (0.02)	0.590 (0.14)
σ ²	4.000 (0.28)	4.217 (0.03)	6.854 (2.34)
Panel B: Optimal Location			
	X Co-Ordinate	Y Co-Ordinate	House Price
Correct Landmarks	−11.426 (4.85)	13.660 (3.57)	1.715 (0.78)
Latitude–Longitude	−8.975 (6.37)	13.859 (4.40)	1.143 (0.70)
Analytic Solution	−10.797 (2.21)	20.780 (4.04)	
Incorrect Landmarks	−6.024 (9.96)	11.381 (7.00)	0.990 (1.24)

Note: Standard errors in parentheses.

Panel A of Table 3 contains the estimation results for a true model containing three landmarks. The first row indicates the continued convergence of the model R² values for our three specifications. The R² for the model with true landmarks is 0.777; for the latitude–longitude version, it is 0.767; and for the model with three incorrect landmarks, the R² is 0.692. The standard deviations for each continue to decline, and the pairwise correlations continue to increase. For the true model and the latitude–longitude version, the correlation between R² values is 0.97. The correlation between R² values for the model using distances to incorrect landmarks is 0.20 with either of the other two.

Table 3. Model with three true distance variables.

Panel A: Coefficients			
Variable	Correct Landmarks	Latitude–Longitude	Incorrect Landmarks
Intercept	1.000 (0.44)	−6.971 (0.19)	−4.247 (5.05)
Non-Distance	2.000 (0.10)	2.000 (0.10)	2.000 (0.10)
D1	−0.250 (0.01)		
D2	−0.100 (0.01)		
D3	−0.150 (0.02)		
d1			−0.100 (1.12)
d2			−0.100 (1.32)
d3			−0.100 (1.03)
X		−0.150 (0.01)	
Y		0.167 (0.01)	
X ²		−0.010 (0.00)	
Y ²		−0.010 (0.00)	
R ²	0.777 (0.02)	0.767 (0.02)	0.692 (0.09)
σ ²	4.000 (0.28)	4.181 (0.30)	5.523 (1.68)
Panel B: Optimal Location			
	X Co-Ordinate	Y Co-Ordinate	House Price
Correct Landmarks	−9.342 (4.81)	9.316 (5.12)	−1.020 (0.69)
Latitude–Longitude	−7.751 (5.33)	8.544 (5.25)	−1.401 (0.65)
Analytic Solution	−7.886 (0.88)	8.779 (0.95)	
Incorrect Landmarks	−7.224 (6.93)	7.952 (6.65)	−0.947 (1.11)

Note: Standard errors in parentheses.

Column Two of Panel A of Table 3 indicates that the coefficients are well-estimated when the true landmarks are used. The coefficients of the intercept, the non-distance variable, the three true distance variables and the estimate of σ^2 are equal to their true values. For the latitude–longitude version, the coefficient of the non-distance variable equals its true value, and the estimate of σ^2 exceeds its true value by 0.181. In the version of the model using incorrect landmarks, the coefficient of the non-distance variable is also equal to its true value on average, and the estimate of σ^2 is highest at 5.523 but appears

to be getting smaller with additional landmarks. The other coefficients in this model bear little resemblance to their true values.

Panel B of Table 3 shows the estimates of the optimal house location based on these simulations. Using the correct landmarks, the highest valued house location is (−9.342, 9.316), with an estimated house value of −1.020. Using the latitude–longitude version of the model, the optimal location is (−7.751, 8.544), with an estimated house value of −1.401. The analytic solution using derivatives based on coefficients from the latitude–longitude regression yields the location of (−7.886, 8.544) for the highest value. For the model using incorrect landmarks, the optimal location is (−7.224, 7.952), with an optimal house value of −0.947, which is surprisingly *higher* than the values obtained from either of the other two specifications. This is possible because there are one million different incorrect landmark locations from which to choose and a large (as we explained previously) random component of house prices.

The estimation results for hedonic regression models based on four landmarks are given in Panel A of Table 4. Row Two contains the model R^2 values, and the convergence in values continues. For the model using distances to the correct landmarks, the model R^2 values average 0.801. The R^2 values for the latitude–longitude version and the model using incorrect landmarks follow closely at 0.790 and 0.750, respectively. The average correlation between R^2 values for the true model and the latitude–longitude version is about 0.96, but the correlation between R^2 values for the model using incorrect distance variables and those of the other two models remains low but rises to about 0.30.

Table 4. Model with four true distance variables.

Panel A: Coefficients			
Variable	Correct Landmarks	Latitude–Longitude	Incorrect Landmarks
Intercept	0.999 (0.61)	−17.894 (1.80)	−15.470 (5.92)
Non-Distance	2.000 (0.98)	2.000 (0.10)	2.001 (0.11)
D1	−0.250 (0.01)		
D2	−0.100 (0.03)		
D3	−0.150 (0.06)		
D4	−0.100 (0.07)		
d1			−0.055 (0.70)
d2			−0.058 (0.66)
d3			−0.060 (0.72)
d4			−0.067 (0.57)
X		0.003 (0.21)	
Y		−0.001 (0.21)	

Table 4. Cont.

Panel A: Coefficients			
Variable	Correct Landmarks	Latitude–Longitude	Incorrect Landmarks
χ^2		−0.007 (0.00)	
γ^2		−0.007 (0.00)	
R^2	0.801 (0.02)	0.790 (0.02)	0.750 (0.06)
σ^2	4.000 (0.29)	4.223 (0.30)	5.033 (1.17)
Panel B: Optimal Location			
	X Co-Ordinate	Y Co-Ordinate	House Price
Correct Landmarks	−6.339 (4.69)	6.027 (5.09)	−2.209 (0.69)
Latitude–Longitude	−5.850 (4.82)	6.509 (4.80)	−2.556 (0.68)
Analytic Solution	−5.852 (0.54)	6.514 (0.57)	
Incorrect Landmarks	−5.868 (5.71)	6.333 (5.68)	−1.920 (1.04)

Note: Standard errors in parentheses.

As expected, the coefficients are well-estimated when the distances to the true landmarks are used, as the results in Column Two of Panel A indicate. All of the estimated coefficients, except the intercept, are exactly equal to their true values, while the intercept is low, on average, by 0.001. The average estimate of σ^2 is 4.000. With the latitude–longitude version, the estimate of the coefficient of the non-distance variable is unbiased, but the estimate of σ^2 is 0.222 higher than its true value. When the incorrect distance variables are used, the average bias in the estimate of the coefficient of the non-distance variable is zero, and the average estimate of σ^2 is 5.033.

Panel B of Table 4 shows the corresponding estimates of the optimal landmark locations. For the model using true landmarks, the average optimum location is (−6.339, 6.027), and the house value there is, on average, −2.209. For the latitude–longitude version of the model, the average co-ordinates of the optimum are (−5.850, 6.509), and the house value there is, on average, −2.556. The analytic solution using coefficients from the latitude–longitude version of the model is very nearly the same as when the locations are obtained from the predicted average house values from the latitude–longitude version of the model. The average analytic location of the optimum is (−5.852, 6.514). The average optimum location using the incorrect landmarks is (−5.868, 6.333), with a house value of −1.920. For the model with four distance variables, the optimal locations are nearly the same, but the estimated variances of the co-ordinates found analytically are much lower than the others and maybe only one-tenth as much as with the other methods. Again, the highest optimum house value is found using incorrect landmarks, with the latitude–longitude version giving the lowest maximum house price.

Table 5 contains the simulation results based on five landmarks. As indicated in Panel A of the table, the model R^2 values are consistent across specifications but slightly lower than the four-landmark models, as the result is dependent on the particular landmark locations chosen. For the regression model with the true distances, the R^2 is 0.786; for the latitude–longitude version, it is 0.770; and for the model including incorrect distance variables, it is 0.750. The two R^2 values for the true model and the latitude–longitude version fall slightly compared to the four-landmark values, but the R^2 for the model using

incorrect landmarks is unchanged. An interesting result is that the variance of the R^2 values for both the true model and the latitude–longitude version of the model increases slightly with the addition of the fifth landmark. In the model based on incorrect landmarks, the addition of the fifth landmark caused the variance to *fall* by about one-third, though it is still larger (about double) than that of the other two models. For the model using the correct landmarks or the latitude–longitude version of the model, the estimated standard deviations are 0.02. For the model with incorrect distances, the value is now 0.04. The average correlation between R^2 values for the true model and the latitude–longitude version of the model fell slightly to 0.95, but the correlation between R^2 values for the model based on incorrect distance variables and either of the other two R^2 s increased to 0.46.

Table 5. Model with five true distance variables.

Panel A: Coefficients			
Variable	Correct Landmarks	Latitude–Longitude	Incorrect Landmarks
Intercept	0.999 (1.20)	−8.994 (0.19)	−3.329 (5.20)
Non-Distance	2.000 (0.10)	2.000 (0.10)	2.000 (0.11)
D1	−0.250 (0.03)		
D2	−0.010 (0.04)		
D3	−0.150 (0.06)		
D4	−0.100 (0.09)		
D5	−0.010 (0.04)		
d1			−0.109 (0.99)
d2			−0.107 (1.33)
d3			−0.111 (0.97)
d4			−0.108 (0.89)
d5			−0.109 (0.84)
X		−0.125 (0.01)	
Y		0.103 (0.01)	
X ²		−0.016 (0.00)	
Y ²		−0.015 (0.00)	
R ²	0.786 (0.02)	0.770 (0.02)	0.750 (0.04)
σ ²	4.000 (0.29)	4.288 (0.31)	4.688 (0.73)

Table 5. Cont.

Panel B: Optimal Location			
	X Co-Ordinate	Y Co-Ordinate	House Price
Correct Landmarks	−4.015 (4.10)	2.408 (4.63)	−4.231 (0.75)
Latitude–Longitude	−4.021 (4.45)	3.556 (4.61)	−4.572 (0.71)
Analytic Solution	−4.019 (0.38)	3.553 (0.39)	
Incorrect Landmarks	−4.288 (4.89)	3.558 (5.06)	−4.005 (0.95)

Note: Standard errors in parentheses.

The simulation results for models using correct distance variables are shown in Column Two of Panel A in Table 5. The results show that the intercept is underestimated by 0.001, and all other estimated coefficients and the estimates of σ^2 have mean values equal to the true values (to three decimal places). The latitude–longitude version of the model is shown in Column Three of Table 5. The average estimated coefficient of the non-distance variable is 2.000, exactly equal to its true value. The average estimate of σ^2 is somewhat high, as usual, at 4.288. The results based on incorrect landmarks are given in Column Four of the table. Once more, the average value of the coefficient of the non-distance variable is 2.000. The average estimate of σ^2 is high at 4.688 but seems to fall as more landmarks are added.

Panel B of Table 5 presents the estimated optimal locations for each model and the analytic version. There is general agreement as to the optimal location. Using the correct distances, the optimal location is estimated to be (−4.015, 2.408), with an average house value of −4.231. Using the latitude–longitude version, the optimal location is (−4.021, 3.556), with an average house value of −4.572. The analytic solution gives (−4.019, 3.553) as the optimal location, and the optimal location using incorrect distance variables is (−4.288, 3.558), with an average house price of −4.005. These models all seem to give similar locations for the optimum, and there is good agreement between the two methods of regression and the analytic approach based on the latitude–longitude version of the model. Note that the model with incorrect distance variables produced the highest-valued house, but this may be misleading because that estimate is based on one million different sets of distance variables.

5. Omitted Variable Bias and Correlated Regressors

In this section, we take a different approach and present the results of a simulation that is unlike our previous simulations but goes directly to the heart of the two main issues in Ross et al. [14]. These are (1) the extent to which the latitude–longitude system is a substitute for distance measures and (2) the effectiveness of the latitude–longitude system at reducing omitted variable bias. Unlike prior simulations, the true model in the upcoming simulations contains only two independent variables of two true distance variables. We examined the impact of omitting one of the distance variables on the coefficient estimates for the other distance variable when the omitted distance variable is replaced with the latitude–longitude system.

In these simulations, we fixed the location of one landmark and changed the location of the second landmark to alter the correlation between the resulting two distance measures, as this correlation is central to an examination of omitted variable bias. Unlike the previous simulations, there actually *is* omitted variable bias in the slope estimator of the remaining distance variable. We sought to determine the effectiveness of the latitude–longitude system in reducing or eliminating this bias. In each Monte Carlo experiment, we replaced the omitted distance variable with the Ross et al. [14] co-ordinate system to determine whether the resulting omitted variable bias is reduced by the inclusion of the co-ordinate

system. We examined the impact of this replacement on the estimated coefficient of the included distance variable and noted how this impact is affected by the correlation between the included and excluded distance variables.

The model examined in this simulation is like those previously estimated. In particular, the true model is

$$HP_i = \gamma_0 + \gamma_1 D_i^A + \gamma_2 D_i^B + \varepsilon_i, \quad (10)$$

where, as before, A and B are the true and known locations of Landmarks A and B , and D^A and D^B represent their respective distances. The coefficient of interest is γ_1 , which is equal to -0.1 ; this is also the true value of γ_2 . The true value of the intercept is 1.00 and that of σ^2 is 4.00 . The problem for estimation is that D^B is unknown and replaced by the Ross et al. [5] co-ordinate system to yield the estimated model:

$$HP_i = \gamma_0 + \gamma_1 D_i^A + \theta_1 x_i + \theta_2 y_i + \theta_3 x_i^2 + \theta_4 y_i^2 + \varepsilon_i. \quad (11)$$

By allowing the location of B to change relative to A , we varied the correlation between the distances to A and B , which changes the impact of the omitted variable, D^B , on the size of the bias in the estimated coefficient of the included distance variable, D^A .

For this simulation, we used random samples from a square of length 20 centered at the origin. The location of Landmark A is fixed at $(10, 0)$. The locations for B vary with each Monte Carlo experiment from $(-10, 0)$, $(-10, 5)$, $(-10, 10)$, $(-5, 10)$, $(0, 10)$, $(5, 10)$, $(10, 10)$ and $(10, 5)$. These changing B locations allow the correlations between distances to change from large negative values to large positive values, respectively. For each pair of true landmarks, we conduct 10,000 Monte Carlo experiments. For each Monte Carlo experiment, three models were estimated: the *true* model, including distances to both A and B ; the model with distance to B omitted and thus subject to omitted variable bias; and the model with distance to B omitted but replaced with the Ross et al. [14] latitude–longitude system or LL. The results of this simulation are shown in Table 6.

As the first column indicates, the average correlations between the distances to the true landmarks are -0.821 , -0.752 , -0.620 , -0.379 , 0.044 , 0.452 , 0.680 and 0.876 . For the *true* model with both distance variables included, the estimated average slope of the distance to Landmark A is -0.100 regardless of the landmark distance correlation. Note that in some of these cases, the correlation between distance measures has a large magnitude (-0.821 and 0.876), and the slope estimator remains unbiased just as the theory suggests; multicollinearity does not invalidate the statistical properties of the OLS estimator and is not a problem that needs a solution.

A more interesting comparison can be made between the estimates from the model using the Ross et al. [14] co-ordinate system (LL) and estimates from the model with distance to B omitted with no replacement (*omitted*). Of the eight cases presented in Table 6, the mean slope estimate with the distance to B replaced by the LL system is closer to the truth in seven of the eight cases compared to the *omitted* variable model, and, on average, the latitude–longitude estimates are much closer. These results appear to suggest that the use of the Ross et al. [14] co-ordinate system provides some relief for omitted variable bias. However, this is misleading as the overall performance of the co-ordinate system as a replacement for the distance variable is poor.

This conclusion is apparent if one examines the root mean square errors for the estimates. As the table shows, the RMSEs for the *true* model are the smallest in every case, as the range for these RMSEs is 0.020 to 0.041 . The RMSEs for the omitted variable model range from 0.021 to 0.094 and are not nearly as good as the true model but are close when the correlation between distance variables is zero. However, the RMSEs for the latitude–longitude version of the model range from 0.146 to 0.166 . There is not much variation. The omitted variable model gives RMSEs about twice as large as the omitted variable model regardless of the correlation.

Table 6. Omitted variable bias and correlated regressors.

P (A, B)	Model	Mean	SD	Min	Max	RMSE
−0.821	True ^{a/}	−0.100	0.034	−0.225	0.043	0.034
	Omitted ^{b/}	−0.017	0.020	−0.088	0.063	0.085
	LL ^{c/}	−0.024	0.148	−0.531	0.523	0.166
−0.752	True	−0.100	0.030	−0.212	0.009	0.030
	Omitted	−0.021	0.020	−0.093	0.049	0.082
	LL	−0.060	0.148	−0.597	0.524	0.153
−0.620	True	−0.100	0.025	−0.200	−0.002	0.025
	Omitted	−0.031	0.020	−0.109	0.055	0.072
	LL	−0.109	0.150	−0.725	0.422	0.150
−0.379	True	−0.100	0.021	−0.184	−0.008	0.021
	Omitted	−0.060	0.020	−0.138	0.017	0.045
	LL	−0.112	0.149	−0.706	0.433	0.150
0.044	True	−0.100	0.020	−0.183	−0.026	0.020
	Omitted	−0.105	0.020	−0.187	−0.028	0.021
	LL	−0.103	0.146	−0.693	0.457	0.146
0.452	True	−0.100	0.022	−0.186	−0.020	0.022
	Omitted	−0.148	0.020	−0.222	−0.073	0.053
	LL	−0.089	0.149	−0.622	0.508	0.149
0.680	True	−0.100	0.027	−0.204	−0.004	0.027
	Omitted	−0.176	0.020	−0.244	−0.105	0.079
	LL	−0.081	0.147	−0.661	0.491	0.149
0.876	True	−0.100	0.041	−0.280	0.060	0.041
	Omitted	−0.192	0.020	−0.269	−0.120	0.094
	LL	−0.139	0.147	−0.639	0.468	0.152

^{a/} The true model including distances to both landmarks, A and B. ^{b/} In this model, the distance to Landmark B is omitted and thus subject to omitted variable bias. ^{c/} In this model, the distance to Landmark B is omitted but replaced with the Ross et al. (2011) [14] latitude–longitude system.

As expected, the RMSEs for the true model and the omitted variable model are nearly equal when the correlation is near zero, which is exactly what the theory of omitted variable bias presented earlier predicts (and is exactly the situation with the Ross et al. [14] non-distance variable). On the other hand, the LL RMSEs are, by far, the largest and are *independent of the correlations*. The LL RMSEs hover around 0.150 regardless of the correlation. This value is about five to seven times the average RMSEs from the true model and about *two to three times larger* than the average RMSEs from the omitted variable model. Thus, we conclude that using the LL model in this case is demonstrably worse than doing nothing; the omitted variable model is better, often much better. Furthermore, this simulation indicates that the use of the latitude–longitude system does nothing to reduce omitted variable bias in models actually subjected to omitted variable bias and could make the situation worse.

6. Summary and Conclusions

As pointed out above, hedonic house price studies typically incorporate information about location into the non-stochastic portion of regression models by including either a set of dummy variables to represent individual locations called “neighborhoods” or by using a set of distance (or travel time) variables to characterize locations in terms of proximity

to amenities and dis-amenities. The latter approach requires relatively few additional parameters, although, unlike the dummy variable approach, the locations of the amenities and dis-amenities must be known. Ross et al. [14] present an alternative to the use of distance variables in hedonic house price regressions that replaces the distance variables with a latitude–longitude co-ordinate system that requires no information on the locations or the number of amenities or dis-amenities and, based on Monte Carlo simulation results, claim that the use of the latitude–longitude system eliminates omitted variable bias in the estimation of the coefficients of non-distance variables in the case where the locations of the amenities are unknown or incorrect.

Despite the claims of Ross et al. [14], we show that their models are not subject to omitted variable bias due to their construction and inclusion of a non-distance variable. Thus, their true model is not subject to omitted variable bias as there is none to eliminate. The random nature of their non-distance variable removes the possibility of omitted variable bias, as their non-distance variable is uncorrelated with other regressors. In addition to showing how the theory predicts no bias in this case, we also illustrate this fact with several simulations. The results of 15 simulations presented above in this study reveal that the average value of the coefficient estimates attached to their non-distance variable is equal to its true value of 2.000 in 14 of the simulations, while in the 1 remaining simulation, the average value is 2.001. Moreover, we failed to encounter any problems with extreme multicollinearity, even though we estimated 15 different models, including from one to five distance variables (both correct and incorrect), one million times. We report that the model R^2 values for the model using correct distances and the Ross et al. [14] latitude–longitude version were nearly equal and highly correlated over different specifications. The model using the incorrect distance variables yielded a lower R^2 than the other two models initially, but the gap closed as more incorrect distance variables were added. After three or four distance variables were added, the average R^2 values were similar for the three models, although the R^2 for the incorrect distance model had a larger variance that decreased with the addition of more distance variables.

We extended Ross et al. [14] Monte Carlo experiments to examine how well their co-ordinate system can replace biases in the coefficients of distance variables when the regressors actually are correlated. The results of this simulation are based on a true regression model containing two independent variables—both true distance variables—but one is omitted. We altered the omitted variables to change the correlation with the included distance variable, as this relationship underlies any omitted variable bias. We investigated the parameter estimates for the retained distance variable from the estimation of the true model, the model with the omitted variable, and a model including the retained distance variable and the Ross et al. [14] co-ordinate system. We allowed the correlations between the true distance variables to change and found that in terms of the root-mean-squared error, the *omitted variables model provides better estimates* than the model that replaces the missing distance variable with the Ross et al. [14] co-ordinate system.

What benefits, if any, does the Ross et al. [14] co-ordinate system provide? In the absence of any distance variables, the use of the system can lead to a higher model R^2 , although a model with a few distance variables to incorrect locations can also improve the explanatory power (i.e., choosing locations for incorrect landmarks to use is easy to implement and does appear to provide some benefits). Our results seem to suggest that distances to incorrect landmarks are better than no distance variables at all. Is increasing the R^2 a desirable goal? If the goal is the prediction of the dependent variable, the answer is yes. However, given that R^2 is a non-decreasing function of the number of explanatory variables, the inclusion of *any* additional variables will cause an increase or at least no decrease in R^2 . If prediction of the dependent variable is not the objective, but rather explaining the dependent variable is, then one should consider specification issues based on some underlying theory rather than adding variables in order to maximize R^2 .

Author Contributions: Conceptualization, S.B.C. and N.M.; methodology, S.B.C. and N.M.; investigation, S.B.C. and F.G.M.J.; writing—original draft preparation, S.B.C. and F.G.M.J.; writing—review and editing, S.B.C. and F.G.M.J.; project administration, F.G.M.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Omitted Variable Bias and Ross et al.'s [14] Illustrative Simulation

Omitted Variable Bias. We begin with a simple two-variable regression model where all variables are mean-scaled, so the intercept is zero. The true model is

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i. \quad (\text{A1})$$

With the usual assumptions about the error term, estimation by OLS produces best linear unbiased estimates of the two slope coefficients, β_1 and β_2 . Now, suppose a variable, for example, x_2 , is omitted, and the model estimated includes only x_1 . In this case the OLS expression for the estimator of β_1 based on the omitted-variable model is

$$\tilde{\beta}_1 = \frac{\sum x_{1i} y_i}{\sum x_{1i}^2} = \beta_1 + \beta_2 \frac{\sum x_{1i} x_{2i}}{\sum x_{1i}^2} + \frac{\sum x_{1i} \varepsilon_i}{\sum x_{1i}^2}. \quad (\text{A2})$$

Thus, the expected value of $\tilde{\beta}_1$ from the omitted-variable model is

$$E[\tilde{\beta}_1] = \beta_1 + \beta_2 \rho(x_{1i}, x_{2i}) \sum x_{2i}^2. \quad (\text{A3})$$

Equation (A3) provides the well-known result that the estimator for β_1 in the model omitting x_2 will be biased unless either $\beta_2 = 0$, meaning x_2 did not truly belong in the model in the first place, or x_1 and x_2 are uncorrelated. The second condition is extremely important. The implication is that there is no omitted variable bias if the omitted variable or variables are uncorrelated with those remaining in the regression model. This is exactly the case with the non-distance variable, x , used in all the Ross et al. [14] simulations; it is white noise. Thus, no alternative estimator or procedure can eliminate or reduce omitted variable bias because the absence of a correlation between x and the other independent variables in the Ross et al. [14] simulations means there is no omitted variable bias. Next, we examined aspects of the Ross et al. [14] *Illustrative Simulation* in light of these results on omitted variable bias.

Ross et al.'s [14] *Illustrative Simulation*. In this simulation, Ross et al. [14] begin with a grid of twenty-five points at integer values on a square of length 4 centered at the origin. To these twenty-five points, Ross et al. [14] add three points: $(-0.75, 0.75)$, $(-0.5, 0.5)$ and $(-0.25, 0.25)$. Ross et al. [14] use this data set to investigate the estimation of several variations of the true model given below. House price is a function of distances to three landmarks: $A(-1, -1)$, $B(-1, 1)$, and $C(1, 1)$. The true model is given by

$$\text{HP}_i = \delta_0 + \delta_1 A_i + \delta_2 B_i + \delta_3 C_i, \quad (\text{A4})$$

where A , B and C refer to distances to each of these landmarks. Ross et al. [14] assume the true intercept is 14, and all slope coefficients equal -1 . This model is used to investigate the estimation of alternative specifications based on and related to this relationship in the Ross et al. [14] *Illustrative Simulation*.

A peculiarity of this simulation is that the true model contains no error term, so in every case Ross et al. [14] examine, the error term consists of whatever has been omitted from the true model in (A4). Unfortunately, this means that the error term will also change from alternative model to alternative model making comparisons between models

complicated as both the regression model *and* the error term are changing. We investigate these issues below.

In their *Illustrative Simulation*, Ross et al. [14] present parameter estimates for four variants of the model in (A4) above. These models are, in order of presentation in their Table 1. The first model uses two new distance variables constructed from distance variables to use in place of distances to A , B and C . This model includes distances to $X = (A + B)/2$ and $U = (B + C)/2$. The second model omits distances to C from the original regression and uses only distances to A and B . The third model omits distances to B from the original regression and uses only distances to A and C , and the fourth model includes only distances to point b , which is $(-0.5, 0.5)$ in Figure 1 of their results [14].

Below, we focus only on estimates from Ross et al. [14] Models 2 and 3, as they are both classic cases of omitted variables. The estimated coefficients are easily obtained using the analytical results on omitted variable bias presented earlier. In particular, we examine the two models below

$$HP_i = \delta_0 + \delta_1 A_i + \delta_2 B_i + [\delta_3 C_i] \quad (\text{Model 2}) \quad (\text{A5})$$

and

$$HP_i = \alpha_0 + \alpha_1 A_i + \alpha_2 C_i + [\alpha_3 B_i], \quad (\text{Model 3}) \quad (\text{A6})$$

where the terms in brackets are the omitted variables/error terms.

Ross et al.'s [14] simulation results for their models 2 and 3 are easily obtained using the simple correlations between pairs of distance measures: $\rho(A, B) = -0.257$, $\rho(A, C) = -0.440$ and $\rho(B, C) = 0.257$. Note that all are nonzero. These nonzero correlations, along with nonzero slope coefficients in the full model, portend a problem with omitted variable bias. In addition to these correlations, our analytic solution requires the slope coefficients from the auxiliary regressions: $C = f(A, B)$ and $B = f(A, C)$. These are two simple regression models, the first with C as the dependent variable and A and B as independent variables and the second with B as the dependent variable and A and C as independent variables. With variables written in deviation form, these simple regressions are given by

$$\hat{C} = -0.54A + 0.36B \quad \text{and} \quad \hat{B} = 0.50A + 0.50C. \quad (\text{A7})$$

Econometric theory informs us that the bias in the estimate of δ_1 , the coefficient of A , will be γ_3 times the coefficient of A in the first simple regression, or $(-1)(-0.54) = 0.54$. We note that in Ross et al.'s [14] Table 1, the estimated coefficient of $\delta_1 = -0.46$ (i.e., $-1 + 0.54$). Using the same guidance, we found that the bias for δ_2 is -0.36 (i.e., $(-1)(0.36)$), and the estimated coefficient given in Ross et al.'s [14] Table 1 is -1.37 . These results can be established without resorting to simulations.

Appendix B. Ross et al.'s [14] Simulation I and Table 2

Ross et al.'s [14] *Simulation I* is based on two data-generating processes. The first is based on distance to a known single landmark in the periphery, and the second is based on distances to two known landmarks in the periphery. The estimated models contain distances to one or two incorrect landmarks. Both one and two incorrect landmark models are estimated for both one and two true landmark models. The results of the Ross et al. [14] simulations are shown in Table 2 of their study, and many of the values seem to be implausible. In particular, the estimated values of the intercept and σ^2 seem to be suspiciously large. See, for example, Column Four of Table A1 below, labeled RFL2, which reproduces simulation results from Ross et al.'s [14] Table 2. These results are based on a true model containing one true distance variable but estimated with a single, incorrect, distance variable. In Column Five, labeled NEW2, we re-estimated Ross et al.'s [14] Table 2. Our results differ substantially from those in Ross et al.'s [14] Table 2. Our estimated mean intercept is -2.969 rather than Ross et al.'s [14] Table 2 estimate of -69.78 . In fact, in our one million random trials, the *lowest* intercept value we obtained is -9.589 , which

is far above Ross et al.’s [14] mean value. For the average estimate of σ^2 , Ross et al. [14] report 928.27. Our average for σ^2 is much lower at 4.99, with a range of 2.92 to 7.93. The differences between our estimates and those of Ross et al. [14] are substantial.

Table A1. Ross et al.’s [14] simulations and reproduction simulations.

	RFL1	NEW1	RFL2	NEW2	RFL3	NEW3	RFL4	NEW4
Intercept	−54.27 (88.76)	−2.676 (4.30)	−69.78 (59.90)	−2.969 (3.18)	−86.16 (126.55)	−4.483 (5.93)	−108.85 (85.83)	−4.841 (4.35)
x	2.06 (0.71)	2.000 (0.10)	2.20 (1.26)	2.000 (0.11)	2.10 (1.01)	2.000 (0.11)	2.27 (1.78)	2.000 (0.12)
d ^A	−0.04 (0.44)	−0.013 (1.22)	0.00 (0.17)	−0.004 (0.17)	−0.06 (0.77)	−0.016 (1.60)	0.00 (0.24)	−0.006 (0.23)
d ^B	0.00 (0.44)	−0.011 (1.22)			0.01 (0.76)	−0.013 (1.60)		
σ^2	303.11	4.196	928.27	4.99	617.69	4.349	1853.91	5.837
R ²	0.85	0.58	0.54	0.49	0.85	0.62	0.54	0.49

Note: Standard errors in parentheses.

Our results from the re-estimation of the model in Ross et al.’s [14] Table 3 are shown in Column Six of Table A1, which is labeled RFL3. The underlying model contains distances to two true landmarks, but the estimated model uses distances to two incorrect landmarks. Once more, the mean Ross et al. [14] intercept estimate of −86.16 is far below our mean estimate of −4.48, as shown in the Table A1 column titled NEW3. In one million Monte Carlo trials, we found the range of intercept values to be −41.55 to 35.51. The Ross et al. [14] estimate is well below our lowest value. Ross et al.’s [14] estimated mean value of σ^2 is 617.69, which far exceeds our mean estimate of 4.35. Our estimated range for σ^2 is 2.89 to 9.42, very far from the Ross et al. [14] average value.

Our discussion has focused on the re-estimation of Ross et al.’s [14] Tables 2 and 3, but as Table A1 shows, we re-estimated all four of their models and found results that substantially differ from those shown in Table 2 of their study. A similar conclusion can be reached analytically. For the single true landmark $A = (0, 15)$, the most distant house in the neighborhood is at either lower corner of $(-10, -10)$ or $(10, -10)$. The approximate distance between A and either is about 27, which, given the true model parameters, yields an average minimum price of about one minus seven or −6 (with the impact of the non-distance variable and the random error term averaged out). Now consider a house location and an incorrect landmark location with the same co-ordinates on the border between the neighborhood and the periphery. At such a point, the distance between the two is zero, and the predicted average price, using Ross et al.’s [14] Table 2, is −69.78. This value is the *highest* possible average price prediction using an incorrect landmark. The predictions for other locations will be lower because the distance between the house and landmark will be positive, and the distance coefficients below are either negative or zero or, if positive, are too small to change the conclusion. The highest value from the Ross et al. [14] model is far below the lowest value in the data set.

Appendix C. Problems with Ross et al.’s [14] Table 3

There are problems with some of the reported results in Ross et al.’s [14] Table 3 as well. Several of the models in Ross et al.’s [14] Table 3 are estimated as part of our larger Monte Carlo simulation results shown in Tables 1 and 2. The most troubling are the parameter estimates for the latitude–longitude versions of the models in Ross et al.’s [14] Table 3. In fact, the results in Columns Two and Four of Table 3 in their study correspond to our results in Table 1, Panel A, Column Three, and Table 2, Panel A, Column Three. The results differ

greatly. Particularly conspicuous are the very large negative intercepts in the estimated latitude–longitude models given in Ross et al.’s [14] Table 3.

The very large negative RFL intercepts are suspicious. Below, we show that the average prediction from the Ross et al. [14] latitude–longitude model given in Column Two of Table 3 is impossibly low. Consider a true distance model in the Ross et al. [14] neighborhood with a true single landmark at the origin. In the Ross et al. [14] landscape, the greatest distance between two points will occur if the points lie at opposite corners of the neighborhood. The distance between these opposite corners is about 56. Using the true parameter values, the average minimum house price when the house and the landmark are the maximum distance apart is about -16 (i.e., $1 - 0.25(56)$).

Next, consider the latitude–longitude estimation results in Column Two of Ross et al.’s [14] Table 3. In repeated samples, the mean value of the non-distance variable, x , and the mean values of the latitude and longitude variables will equal zero. The mean values of the squares of latitude and longitude are both positive, but consider them to be zero as well for the time being. Using the Ross et al. [14] estimation results, the predicted average house price at the origin consists of the intercept of -66.64 . This value is much lower than the average value of -16 based on the true distance regression and the maximum possible distance between a landmark and a house. This value is based on the assumption that the average value of the non-distance variable, latitude, longitude and their squares all equal zero. This is true for the non-distance variable, latitude and longitude but not the squares, which both must have positive mean values. However, the squared values of latitude and longitude both have negative coefficients in Ross et al.’s [14] Table 3, meaning the predicted average house price at the origin using the latitude–longitude version of the model is less than -66.64 , which is already impossibly low.

References

- Rosen, S. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *J. Political Econ.* **1974**, *82*, 34–55. [[CrossRef](#)]
- Yiu, C.; Tam, C. A Review of Recent Empirical Studies on Property Price Gradients. *J. Real Estate Lit.* **2004**, *12*, 305–322. [[CrossRef](#)]
- Sirmans, S.; Macpherson, D.; Zietz, E. The Composition of Hedonic Pricing Models. *J. Real Estate Lit.* **2005**, *13*, 1–44. [[CrossRef](#)]
- Goodman, A.C.; Thibodeau, T.G. Housing Market Segmentation and Hedonic Prediction Accuracy. *J. Hous. Econ.* **2003**, *12*, 181–201. [[CrossRef](#)]
- Lee, N.J.; Seslen, T.N.; Wheaton, W.C. Do House Price Levels Anticipate Subsequent Price Changes within Metropolitan Areas? *Real Estate Econ.* **2015**, *43*, 782–806. [[CrossRef](#)]
- Sah, V.; Conroy, S.J.; Narwold, A. Estimating School Proximity Effects on Housing Prices: The Importance of Robust Spatial Controls in Hedonic Estimations. *J. Real Estate Financ. Econ.* **2016**, *53*, 50–76. [[CrossRef](#)]
- Wolf, D.; Klaiber, H.A. Bloom and Bust: Toxic Algae’s Impact on Nearby Property Values. *Ecol. Econ.* **2017**, *135*, 209–221. [[CrossRef](#)]
- Caudill, S.B.; Costello, M.; Mixon, F.G., Jr.; Affuso, E. Food Deserts and Residential Real Estate Prices. *J. Hous. Res.* **2021**, *30*, 98–106. [[CrossRef](#)]
- Caudill, S.B.; Affuso, E.; Yang, M. Registered Sex Offenders and House Prices: An Hedonic Analysis. *Urban Stud.* **2015**, *52*, 2425–2440. [[CrossRef](#)]
- Liao, F.H.; Wilhelm, F.M.; Solomon, M. The Effects of Ambient Water Quality and Eurasian Watermilfoil on Lakefront Property Values on the Coeur d’Alene Area of Northern Idaho, USA. *Sustainability* **2016**, *8*, 44. [[CrossRef](#)]
- Affuso, E.; Cummings, J.R.; Le, H. Wireless Towers and Home Values: An Alternative Valuation Approach using Spatial Econometric Analysis. *J. Real Estate Financ. Econ.* **2018**, *56*, 653–676. [[CrossRef](#)]
- Affuso, E.; Caudill, S.B.; Mixon, F.G., Jr.; Starnes, K.W. Is Airport Proximity an Amenity or Disamenity? An Empirical Investigation based on House Prices. *Land Econ.* **2019**, *95*, 391–408. [[CrossRef](#)]
- Li, R.Y.M.; Li, H.C.Y. Have Housing Prices Gone with the Smelly Wind? Big Data Analysis on Landfill in Hong Kong. *Sustainability* **2018**, *10*, 341. [[CrossRef](#)]
- Ross, J.M.; Farmer, M.C.; Lipscomb, C.A. Inconsistency in Welfare Inferences from Distance Variables in Hedonic Regressions. *J. Real Estate Financ. Econ.* **2011**, *43*, 385–400. [[CrossRef](#)]
- Li, N.; Li, R.Y.M. A Bibliometric Analysis of Six Decades of Academic Research on Housing Prices. *Int. J. Hous. Mark. Anal.* **2024**, *17*, 307–328. [[CrossRef](#)]
- Cameron, T.A. Directional Heterogeneity in Distance Profiles in Hedonic Property Value Models. *J. Environ. Econ. Manag.* **2006**, *51*, 26–45. [[CrossRef](#)]

17. Fik, T.J.; Ling, D.C.; Mulligan, G.F. Modeling Spatial Variation in Housing Prices: A Variable Interaction Approach. *Real Estate Econ.* **2003**, *31*, 623–646. [[CrossRef](#)]
18. Pavlov, A.D. Space-Varying Regression Coefficients: A Semi-parametric Approach Applied to Real Estate Markets. *Real Estate Econ.* **2000**, *28*, 249–283. [[CrossRef](#)]
19. Du, H.; Chen, Z.; Mao, G.; Li, R.Y.M.; Chai, L. A Spatio-Temporal Analysis of Low Carbon Development in China's 30 Provinces: A Perspective on the Maximum Flux Principle. *Ecol. Indic.* **2018**, *90*, 54–64. [[CrossRef](#)]
20. Shan, M.; Wang, Y.; Lu, Y.; Liang, C.; Wang, T.; Li, L.; Li, R.Y.M. Uncovering PM2.5 Transport Trajectories and Sources at District within City Scale. *J. Clean. Prod.* **2023**, *423*, 138608. [[CrossRef](#)]
21. Tsao, H.C.; Lu, C.J. Assessing the Impact of Aviation Noise on Housing Prices using New Estimated Noise Value: The Case of Taiwan Taoyuan International Airport. *Sustainability* **2022**, *14*, 1713. [[CrossRef](#)]
22. Chen, K.; Lin, H.; Liao, L.; Lu, Y.; Chen, Y.J.; Lin, Z.; Teng, L.; Weng, A.; Fu, T. Nonlinear Rail Accessibility and Road Spatial Pattern Effects on House Prices. *Sustainability* **2022**, *14*, 4700. [[CrossRef](#)]
23. Aziz, A.; Anwar, M.M.; Abdo, H.G.; Almohamad, H.; Al Dughairi, A.A.; Al-Mutiry, M. Proximity to Neighborhood Services and Property Values in Urban Area: An Evaluation through the Hedonic Pricing Model. *Land* **2023**, *12*, 859. [[CrossRef](#)]
24. Peng, C.; Xiang, Y.; Chen, L.; Zhang, Y.; Zhou, Z. The Impact of the Type of and Abundance of Urban Blue Space on House Prices: A Case Study of Eight Megacities in China. *Land* **2023**, *12*, 865. [[CrossRef](#)]
25. Greene, W.H. *Econometric Analysis*, 5th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2003.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.