World Academy of Science, Engineering and Technology
International Journal of Environmental and Ecological Engineering
Vol:9, No:1, 2015

A Study on Inference from Distance Variables in Hedonic Regression

Yan Wang, Yasushi Asami, Yukio Sadahiro

***Abstract*—**In urban area, several landmarks may affect housing price and rents, and hedonic analysis should employ distance variables corresponding to each landmarks. Unfortunately, the effects of distances to landmarks on housing prices are generally not consistent with the true price. These distance variables may cause magnitude error in regression, pointing a problem of spatial multicollinearity. In this paper, we provided some approaches for getting the samples with less bias and method on locating the specific sampling area to avoid the multicollinerity problem in two specific landmarks case.

***Keywords*—**Landmarks, hedonic regression, distance variables, collinearity, multicollinerity.

I. Introduction

THIS research addresses the problem that causes magnitude error such as unacceptable variance and mean value in regression when the analysis model adopted the distance variables in hedonic pricing method [1]. Hedonic pricing model is a classic regression method firstly proposed by Lancaster (1966) with consumer theory [2] and Rosen (1974) with the theoretical model [3]. The common form of hedonic pricing is:

$$P = a + b_1 d_1 \quad b_2 d_2 \quad \ldots \quad + e \qquad +$$

In the equation, a is a constant term, d_1, d_2 are the distances drawn from the landmarks, b_1, b_2 are the coefficient of distance variables, e is the error term.

This spatial multicollinearity issue presents when the regression includes more than two distance variables. It will cause high standard errors in the regression parameters. Most of the researchers focus on the influences from landmarks [4]. Hoerl and Kenard (1970) introduced ridge regression to identify the landmarks that didn't show influence on the regression [5]. Then Harrison and Rubinfeld (1978) used a weighted distance to employment centers to find out the demand for environmental quality [6]. Griffith (1981) proved that the third distance variable may be useless in a two nodes model, but not redundant in terms of their influence on rents. Eric Heikkila (1988) pointed out that it is possible to avoid or reduce potential collinearity problems by determining the geographic range. Also, a zero constant term is suggested in theoretical model [7]. Mahon (1996) has proved estimates for the standard errors of the parameters and has argued on the basis on simulation. Hood, Nix and Iles (1999) provided an analysis of the uncertainties in parameter estimates under a variety of constraints. Fik et al. (2003) used longitude and latitude distance from landmarks to properties to estimate the hedonic price in Tucson [8] - [9]. M. Tiefelsodorf (2003) used distance variables to address the spatial structure effects of general interaction models, multicollinearities in the estimated parameters make it extremely difficult to interpret the coefficient properly. Deaton and Hoehn (2004) adopted a second distance variable to mechanism to efficiently assess its value. Noonan et al. (2007) used the distance variables drawn from the historic city center to explain variation in median house value in the United States [10]. J. M. Ross et al. (2011) showed the method on confirming the landmarks affect the house price and the suggestion on the second landmark, also mentioned the multicollinearity will cause unstable error in analysis [11].

Here they pointed out the limitations of identifying the correlation relationship, which cannot be detected by the statistical specification tests and relative analysis. But in the case that in analysis region with the landmarks selected, selecting the valid data and identifying the samples which cause bias in regression is not mentioned. Two nodes in space, which considered as weighted vectors to some specific region will triangulated by geometry, the distance variables are measured on the same surface and the correlation between the two landmarks will cause an effect on the coefficients of regression. This paper will provide several solutions by avoiding the observations that cause bias and patterns of sampling method in hedonic pricing analysis to estimate the effect of the specific selected facilities to our observation location. In this study we will focus on a simple model with two landmarks.

To demonstrate the two landmarks sampling methods in hedonic pricing model, we built a set of simulations in theoretical model. This paper is arranged into 4 sections. After the introduction (Section I), the first simulation is in a wide region context, we will begin with the correlation analysis (Section II), in the second simulation is under the condition of a tinny region (Section III). By these simulations we will illustrate the parameters of distance variables which are changed by regression. Then we analyzed the influence of the sampling area on the distance variables. After that, we will illustrate the dominance of recommended sampling area by analyzing the coefficient and regressor change in simulations to locate the area that is with less bias. Finally, we will offer the conclusions and suggestions in hedonic analysis with two specific urban nodes (Section IV).

Yan Wang is Ph.D. Candidate of the Graduate School of Urban Engineering, the University of Tokyo, Japan (HP: 81-80-4926-9090; fax: 03-5841-8521; e-mail: wangyan@ua.t.u-tokyo.ac.jp / brucio99@gmail.com).

Yasushi Asami is Professor of Department of Urban Engineering, the University of Tokyo, Japan (e-mail: asami@csis.u-tokyo.ac.jp).

Yukio Sadahiro is Professor of Center for Spatial Information Science, the University of Tokyo, Japan (e-mail: sada@csis.u-tokyo.ac.jp).

II. Wide Region Simulation

Two Variables Wide Region Simulation

In two landmarks case, the basic Hedonic model can be written as:

$$P_i = a + b_1 d_{1i} + b_2 d_{2i} + e_i \quad (1)$$

$$\hat{P}_i = \hat{a} + \hat{b}_1 d_{1i} + \hat{b}_2 d_{2i} + \hat{e}_i \quad (2)$$

where i denotes each random sample point S, P_i denotes the price of the sample point S_i, a denotes the constant term in regression. d_{1i} and d_{2i} denotes the distances drawn from the two facilities Q_1 and Q_2 to point S_i. b_1, b_2 denote the coefficients of distance variables, and e_i denotes the error term.

Set the observation area (Fig. 1) as 10×10 square region centered on $O(0,0)$, draw a radius R=1 circle centered on origin in the illustration. Take the sample points as uniform distribution interval defined by n in the region. Calculate the distance from any point $S_i(x_i, y_i)$ to $Q_1(-1,0)$ and $Q_2(1,0)$ in the square region, mark them as d_{1i} and d_{2i}. Then the price of S_i could be calculated according to (1). We take a set of observations and ran the regression with P_i and d_{1i}, d_{2i} for 1000 trails. Analyze the parameter $\hat{b}_1$ (coefficient of d_{1i}) and $\hat{b}_2$ (coefficient of d_{2i}), and estimated the model.

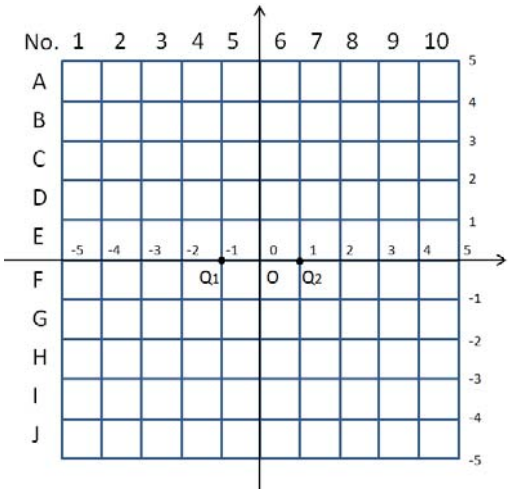

Fig. 1 Observe region of simulation I

Before performing the regression, we analyzed the correlation coefficient of the square region first (simulation I), for most of the multicollinearity problem was caused by spatial collinearity. More than two distance variables on one surface with longitude and latitude locations will form liner problems as proved by lots of other researchers.

To realize the performance of correlation analysis, the sampling area here is selected as a circle region C (radius r=0.5), interval step is set as t'=0.02, and 1000 samples will be taken randomly in this circle region. When C's center moving in the whole area as step t_1=0.05 (interval step) both on x and y axes, we measure the distance from sampling point S to node $Q_1(-1,0)$ and $Q_2(1,0)$, and calculate each distance d_{1i} and d_{2i} 's correlation coefficient. Plot the results of R_{12} (correlation of d_{1i} and d_{2i}), we can get the contour line map and the surface map of the entire observation region's correlation coefficient as below.

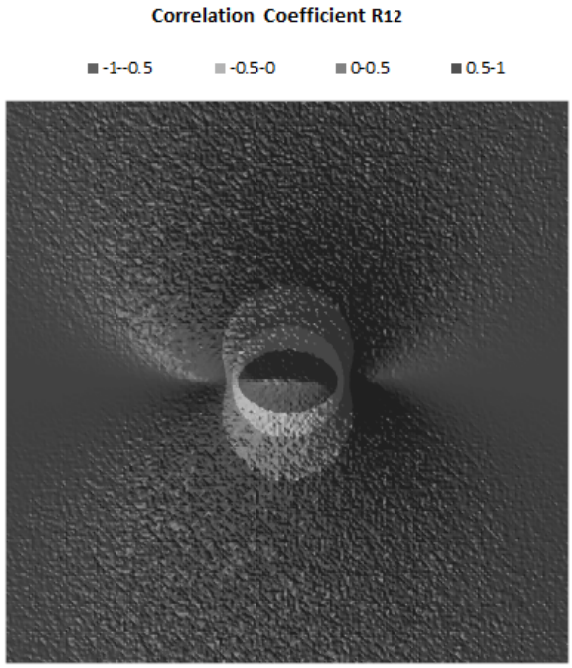

Fig. 2 Contour line map of correlation coefficient R_{12}

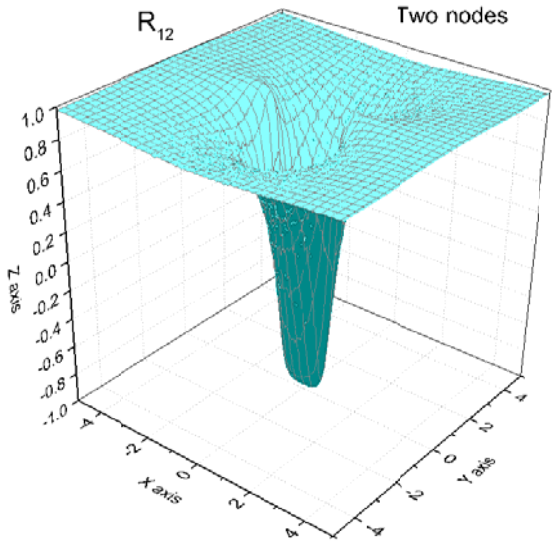

Fig. 3 Surface map of correlation coefficient R_{12}

To identify the valid samples, we should select the ones that in the area where $|R_{12}|$ (absolute value of correlation coefficient of d_{1i} and d_{2i}) is small. Obviously, it will not be the region inside or outside the circle depending on the results showed in Fig. 2. However, it is usually impossible to locate observations

exactly on a circle in the real world, so they should be scattered around or close to the circle. Furthermore, we should avoid the locations where R_{12} increasing rapidly such as the neighborhood around $Q_1(-1,0)$ and $Q_2(1,0)$ that was shown in Fig. 3.

From Fig. 3 we can see that the correlation coefficient changed sharply from nearly 1 to -1, but tend to be 0 when the location is near to the circle passed through the two facilities, which was marked as circle F. The points located far from circle F produced large positive correlations, also the points around the circle F's center produced magnitude negative correlations too. The R_{12} started to fall when the sample region C is getting close to the circle F. Sample points on or close to circle F showed correlation value close to 0, they are more effective than others in the observe region.

Here we built a regression model to confirm our hypothesis in wide region context. Before running the regression, we also calculated the correlation of the observation region as we introduced previously.

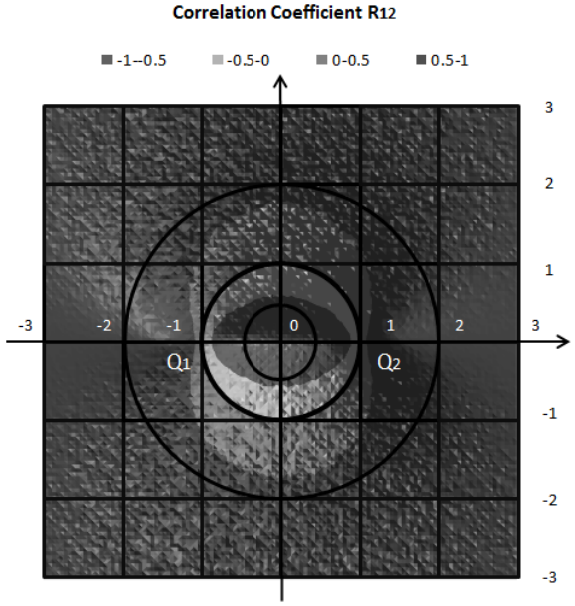

Fig. 4 Contour line map of correlation coefficient R_{12} in simulation II

This time we take 6×6 square region as the study region of simulation II. In the performance of correlation analysis, the sampling area here is selected as a circle region C_2 (radius r=0.5) set as uniform distribution, the interval step is t"=0.02, and 1000 samples are taken randomly in this circle. Then, C_2's center moving in observation area as interval step t_2=0.05 on both on x and y axes, the distance from sampling point S to the two landmarks can be measured, then the correlation of d_{1i} and d_{2i} can be calculated. The result was plotted into contour map and surface map. They are showed in Figs. 4 and 5.

We could see the value of correlation R_{12} decayed when the location is getting close to the circle that passed through two landmarks in Fig. 4. The value of R_{12} is falling sharply from 1 to -1, reach the maximum, zero, at the location on the specific radius $R=1$ circle F. From Fig. 5, we can see the $2R$ circle located at the high correlation area. The correlation value is between 0.5 and 1. The $R/2$ circle also located at the area with high correlation, the correlation floated from -0.5 to -1. Usually, samples with high correlation value will produce more bias in regression. So in the next phase we want to confirm it as the following steps.

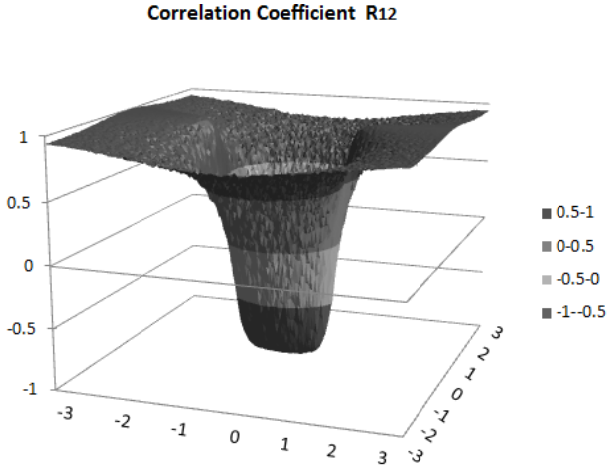

Fig. 5 Surface map of correlation coefficient R_{12} in simulation II

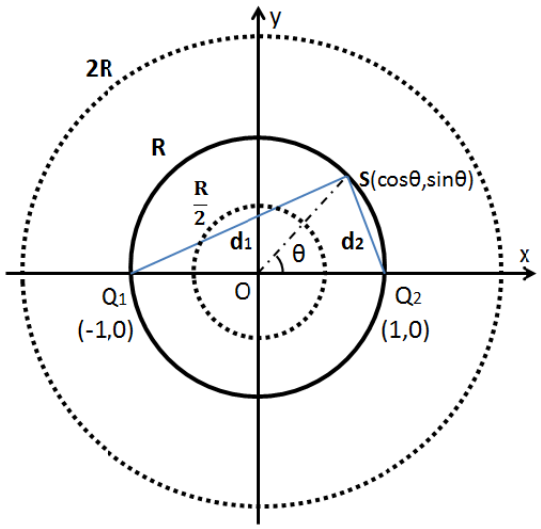

Fig. 6 Observation region of simulation III

In simulation III, three circles centered on origin were drawn in Cartesian coordinate as Fig. 6, their radius were taken as $R/2$, R and $2R$, O is the center of the 6×6 uniform distributed square region, Q_1, Q_2 are the first and second landmarks (nodes) located at points $(-1,0)$ and $(1,0)$, R is the half of segment Q_1Q_2. In this simulation R was taken equal to 1. S is the sample point which was taken on the radius $R=1$ circle. If the angle of $\angle SOQ_2$ is taken as θ, the coordinate of our uniform distributed sample point S could be written as $S(\cos\theta,\sin\theta)$.

In simulation III, we took θ from 0 to 360 degree (circle), set $a = 2500$, $b1 = -130$, $b2 = -100$, $e = 0.01$. If we take the coordinate of the sample point as (xi , yi), then each distance drawn from our observation to the landmarks' distance $d1i$ and $d2i$ could be calculated by (3) and (4) as below.

$$d1i = \left[(xi + 1)^2 + yi^2 \right]^{1/2} \quad (3)$$

$$d2i = \left[(xi - 1)^2 + yi^2 \right]^{1/2} \quad (4)$$

Since Pi could be calculated by (1), we take 500 observations randomly on the three circles and run the regression with $d1i$, $d2i$ and Pi for 1000 trails. We can get the results of regressors' parameters, variance, mean value and standard deviation divided by the total numbers of samples ($n = 50$) in Table I.

TABLE I
RESULTS OF SIMULATION III

Area	Value	$\hat{b}1 - b1$	$\hat{b}2 - b2$
OUT (2R)	Variance	6.95E-05	1.59E-05
	Mean	-0.00012	6.11E-05
	SD/N	1.67E-05	7.96E-06
ON(R)	Variance	2.56E-06	9.04E-07
	Mean	5.46E-05	-5.1E-05
	SD/N	3.2E-06	1.9E-06
IN (1/2R)	Variance	0.008661	0.000175
	Mean	0.002384	-0.00023
	SD/N	0.000186	2.65E-05

Simulation based on 1,000 trials on sample size of 10,201. House price has two distance variables in data producing process. SD/N means Standard Deviation divided by N (N=10,201).

Based on the results, we can figure out that on the $R = 1$ circle, variance and mean value of $\hat{b}1$ are 2.56×10^{-6} and 5.46×10^{-5}, variance and mean value of $\hat{b}2$ are 9.04×10^{-7} and -5.1×10^{-5}. They are much smaller than the value of variance and mean on the $2R$ circle, also smaller than the value of variance and mean on the $R/2$ circle's situation. This result confirmed the hypothesis that when observations were taken randomly on the circle passed through the two facilities, the regression results are more persuasive than the results of observations that were taken on the inside and outside the circles.

In Eric Heikkila (1988)'s analysis method, observations should be taken randomly in the region between the area that some facilities are located in the center, and others are on the edge of the observe region [7]. Following this method, we can locate the sampling area in our observation region. It will be the region inside the dotted circle's area centered $Q1$ described in Fig. 7 when considering $Q1$ $Q2$ as the landmarks. However, the correlation coefficient analysis we did in simulation I had already pointed out that the location in region close to $Q1$ $Q2$ and the origin will produce unacceptable $R12$ (correlation) value, which means that these biased data will cause unstable coefficients and error in regression. This method was confirmed in the wide region simulation. The solid circle and the dotted circle showed the difference of the sampling region in Fig. 7.

Based on the correlation analysis of simulation II, sample region should be located on the circle that passed through the two nodes $Q1$ and $Q2$. This area (without the region close to the two nodes) will be the best sampling region as we certified. However, in real database, it will be very difficult to take samples right on the circle. There will be a lot of reasons such as the spatial and other limitations when we gathering the data.

Basically, the observation location should be close to the specific circle. And avoiding the high correlation value area, such as the region just around $Q1$ and $Q2$, the region close to the origin (midpoint of the two nodes) and the region too far away from the specific solid circle in Fig. 7 (the area between the dotted circle and the solid circle on the left of the Fig. 7 was also suggested by Eric Heikkila [7]).

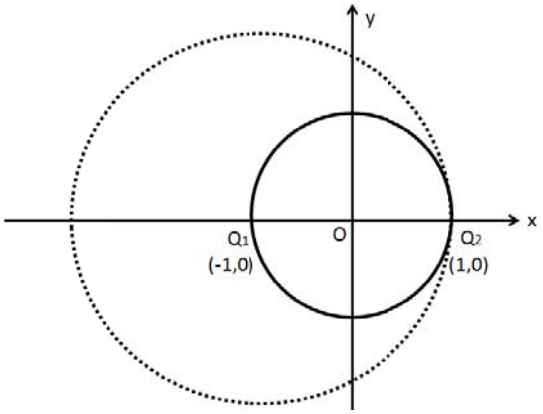

Fig. 7 Comparison on the observation region with other method

III. TINNY REGION SIMULATION

Two Variables Tiny Region Simulation

In real database, it will be hard to gathering the samples of the whole observation area. When researchers are facing the limited observation area, we have to estimate the house price with data in some specific region. This phase we will build another tiny region regression model to demonstrate the sampling area's location.

Before running the regression, the correlation $R12$ will be considered in simulation IV. To realize this performance, the correlation sampling region is selected as a tiny circle region $C3$' (radius r=0.05) which is set as uniform distribution, the interval step is $t3' = 0.02$. 1000 samples were taken randomly in this circle region. When the center of $C3$' is moving in the whole 6×6 square region as step $t3 = 0.005$ both on x and y axes, measure the distance from each sample point S' to node $Q1(-1,0)$ and $Q2(1,0)$, calculate each distance of $d1i$ and $d2i$'s correlation coefficient. Plot the result of $R12$, we could get the contour line map of the square observation region described as Fig. 8.

When θ equals to 30 degree, the tiny regions located on the $2R$, R and $R/2$ circle (here we also take the radius equal R=1), are marked as: X-1 ,which is centered on point $x1$ (1.73, 1) located on the $2R$ circle; X-2, which is centered on point $x2$ (0.865, 0.5) located on the R circle; and X-3, which is centered on point $x3$ (0.43, 0.25) located on the $R/2$ circle. When θ equals to 30 degree, the tiny observation regions located on the three circles are marked as: Y-1, which is centered on point $y1$ (1, 1.73), located on the $2R$ circle; Y-2, which is centered on point $y2$ (0.5, 0.865) located on the R circle; and Y-3, which is centered on point $y3$ (0.25, 0.43) located on the $R/2$ circle. In 90 degree's situation they are marked as: Z-1, which is centered on point $z1$ (0, 2) located on the $2R$ circle; Z-2, which is centered on point $z2$ (0, 1) located on the R circle; and Z-3, which is centered on point $z3$ (0, 0.5) located on the $R/2$ circle.

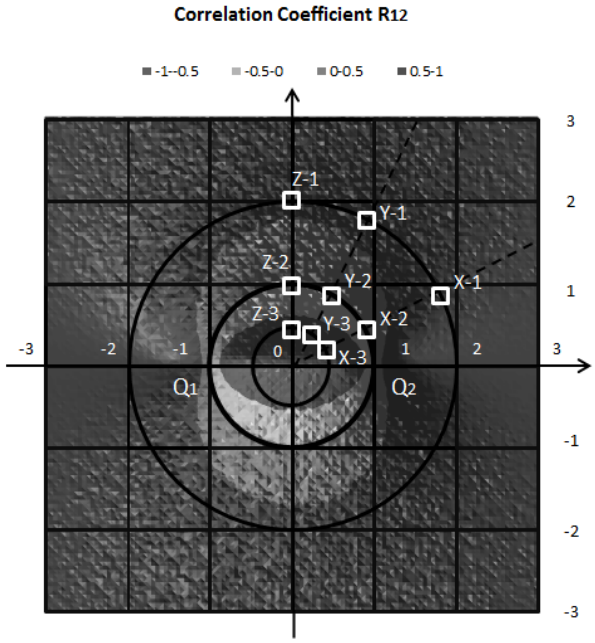

Fig. 8 Contour line map of correlation coefficient R_{12} in simulation IV

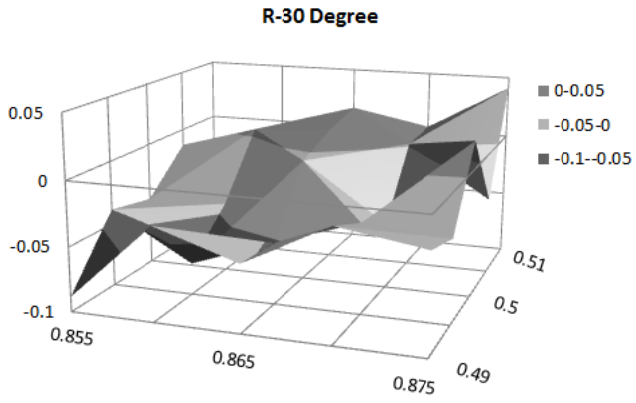

Fig. 9 Surface map of correlation coefficient R_{12} of X-2

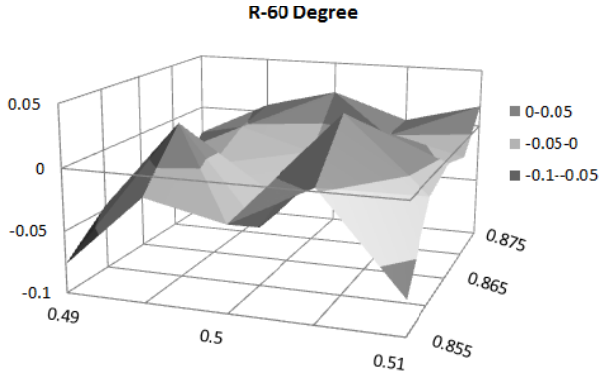

Fig. 10 Surface map of correlation coefficient R_{12} of Y-2

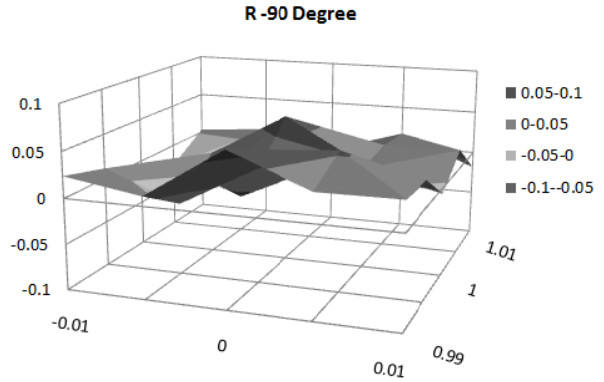

Fig. 11 Surface map of correlation coefficient R_{12} of Z-2

From the contour map of the correlation in Fig. 8, we could figure out that the tiny observation locations: Region X-2, Y-2 and Z-2 on the radius circle passed through landmarks sharing the minimum correlation value in this correlation analysis. The value changed from -0.05 to 0.05, which is quite close to zero. The changes are very tinny and smoothly. We could see the details in the surface maps of correlation of X-2, Y-2 and Z-2 in Figs. 9-11.

The observation locations X-1, Y-1, Z-1 located on the $2R$ produced high correlation value from 0.5 to 0.8. X-1 region showed the $R12$ value from 0.79 to 0.85, which is the largest among the three observation region on 30 degree.

We could see the correlation coefficient value detail changes in the surface maps of correlation of X-1, Y-1 and Z-1 in Figs. 12-14.

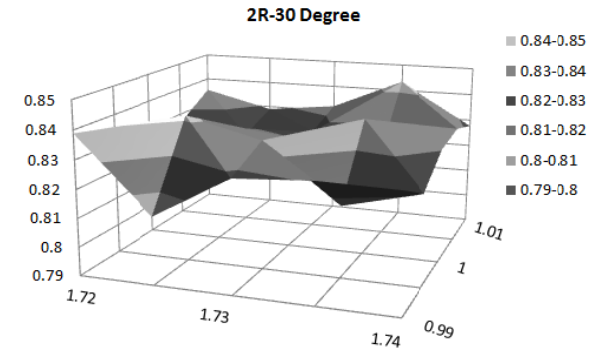

Fig. 12 Surface map of correlation coefficient R_{12} of X-1

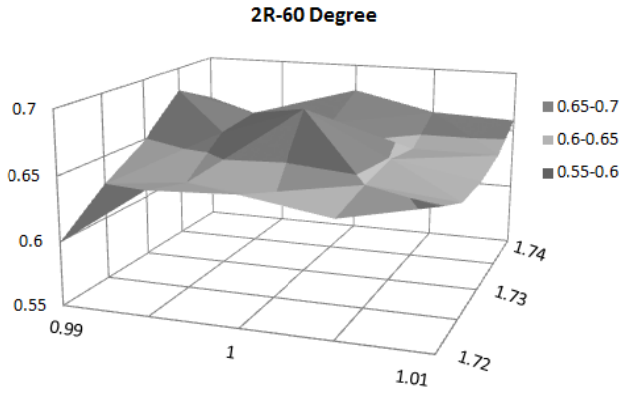

Fig. 13 Surface map of correlation coefficient R_{12} of Y-1

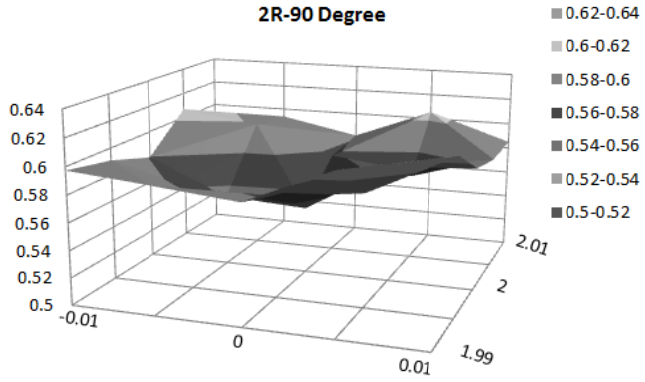

Fig. 14 Surface map of correlation coefficient R_{12} of Z-1

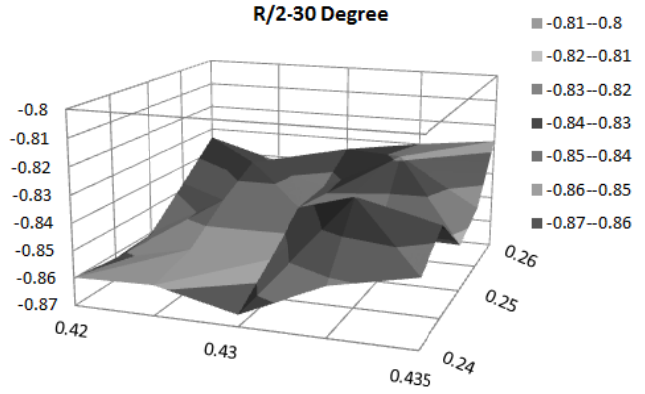

Fig. 15 Surface map of correlation coefficient R_{12} of X-3

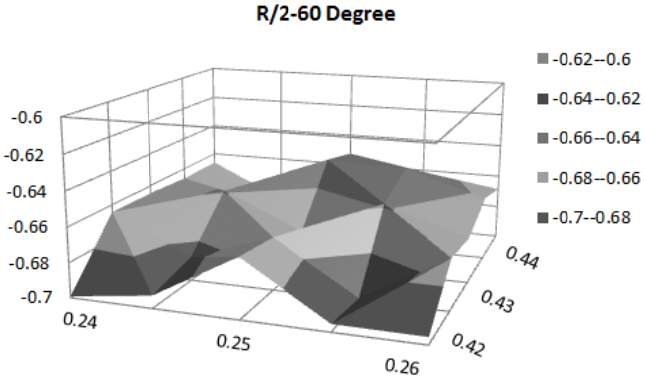

Fig. 16 Surface map of correlation coefficient R_{12} of Y-3

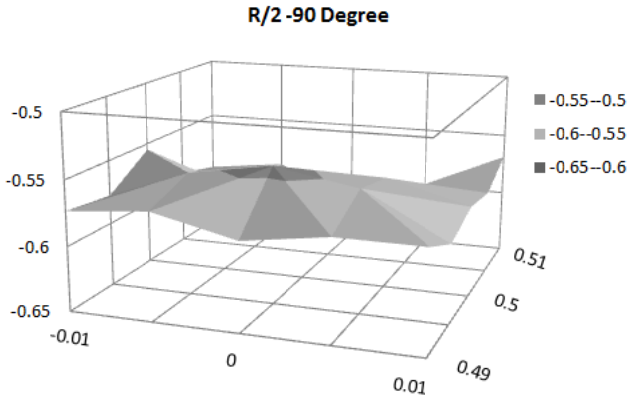

Fig. 17 Surface map of correlation coefficient R_{12} of Z-3

And observation locations X-3, Y-3, Z-3 located on the $R/2$ circle are facing the data with high correlation value from -0.5 to -0.87. Region X-3 showed the largest $R12$ value from -0.8 to -0.87.

We can see the detail changes in the surface maps of correlation of them in Figs. 15-17.

Base on the analysis above, Figs. 9-11 showed the three dimensional image of the correlation coefficient R_{12}'s changes. These area are right on the specific circle that passed through urban nodes Q_1 Q_2.

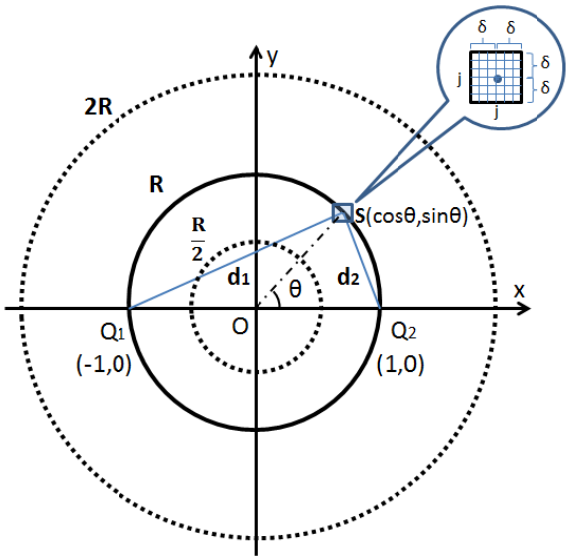

Fig. 18 Observe region of simulation V

In this step we want to check whether the regression show the positive result. Here, in tiny region simulation V, there are three circles centered on origin in Cartesian coordinate as Fig. 18, their radius were taken as $R/2$, R and $2R$, O is the center of the 6×6 uniform distributed square region, $Q_1(-1,0)$ and $Q_2(1,0)$ are the landmarks (urban nodes), the radius of the circle, R, was taken as the distance of the half of segment Q_1 Q_2. In this simulation radius was also taken as equal to 1. S is our sample point which was taken on the $R=1$ circle.

Take the angle of $\angle SOQ_2$ as θ .The coordinate of this sample area centered on S , the coordinate will be set as: $x \in (\cos\theta - \delta, \cos\theta + \delta)$, $y \in (\sin\theta - \delta, \sin\theta + \delta)$, each observation's coordinate S_i could be calculated following (5) and (6).

$$xi = \cos\theta - \delta + \delta\frac{j}{n} \tag{5}$$

$$yi = \sin\theta - \delta + \delta\frac{j}{n} \tag{6}$$

where δ denotes the range of the square area centered on S , δ is divided into n , j denotes the length and width of the tiny region. When $n = 50$, the numbers of samples in the tiny observation region N can be calculated by (7).

$$N = (2n+1) \times (2n+1) \tag{7}$$

We will check the regression result is positive or not. Setting θ as 30, 60 and 90 degree, adjusting the parameters to $a = 2500$, $b_1 = -130$, $b_2 = -100$, $e = 0.01$. Take 500 samples randomly in each of the tiny observation zones and run the regression for 1000 trails in each simulation. (These zones are located on the radius $R/2$, R and $2R$ circles.)

We can get the results of variance, mean, standard deviation divided by the total numbers of samples ($n = 50$) in Tables II-IV.

TABLE II
RESULTS OF SIMULATIONS V ON 30 DEGREE

Area	Value	$\hat{b}_1 - b_1$	$\hat{b}_2 - b_2$
OUT (2R)	Variance	179.1488	181.2094
	Mean	0.143752	0.10219
	SD/N	0.026769	0.026923
ON(R)	Variance	55.21421	60.25449
	Mean	0.174904	0.181987
	SD/N	0.014861	0.015525
IN (1/2R)	Variance	181.8755	186.4329
	Mean	-0.5521	0.377261
	SD/N	0.026972	0.027308

Simulation based on 1,000 trials on sample size of 10,201.SD/N=Standard Deviation/ N

TABLE III
RESULTS OF SIMULATIONS V ON 60 DEGREE

Area	Value	$\hat{b}_1 - b_1$	$\hat{b}_2 - b_2$
OUT (2R)	Variance	102.9973	101.2691
	Mean	0.193642	0.026698
	SD/N	0.020298	0.020127
ON(R)	Variance	62.48154	66.10719
	Mean	0.064727	0.083577
	SD/N	0.015809	0.016261
IN (1/2R)	Variance	105.4606	99.44662
	Mean	0.492117	-0.31337
	SD/N	0.020539	0.019945

Simulation based on 1,000 trials on sample size of 10,201.SD/N=Standard Deviation/ N

TABLE IV
RESULTS OF SIMULATIONS V ON 90 DEGREE

Area	Value	$\hat{b}_1 - b_1$	$\hat{b}_2 - b_2$
OUT (2R)	Variance	98.10827	89.30927
	Mean	-0.39122	-0.35039
	SD/N	0.01981	0.018901
ON(R)	Variance	59.8667	59.06657
	Mean	-0.24307	0.004901
	SD/N	0.015475	0.015371
IN (1/2R)	Variance	85.20443	94.40464
	Mean	-0.01009	-0.3511
	SD/N	0.018461	0.019432

Simulation based on 1,000 trials on sample size of 10,201.SD/N=Standard Deviation/ N

Based on these results, we can figure out that in tiny region simulations:

When the observations realized on the condition that θ =30 degree, variance value of $\hat{b}_1$ and $\hat{b}_2$ on the circle passed through landmarks are 55.21421 and 60.25449. Mean values are 0.174904 and 0.181987. They are much smaller than the variance and mean values than the results of $2R$ and $R/2$ circle.

On the condition that θ =60 degree, the result on the specific circle, variance of $\hat{b}_1$ and $\hat{b}_2$ are 62.48154 and 66.10719, mean values are 0.064727 and 0.083577. They are smaller than the outside and inside circles' too.

On the condition θ =90 degree, the regression results on the radius equal to 1 circle, the variance of $\hat{b}_1$ and $\hat{b}_2$ are 59.8667 and 59.06657, the mean value are -0.24307 and 0.004901. They also proved our hypothesis.

Smaller variance, mean value and standard deviation denote the regressor coefficients' changes are quite tiny, and acceptable. Since the simulations showed positive result when it was realized on the specific circle passed through the two landmarks, if researchers adopt the data with limitation to some specific area, the data on or close to region on the circle that passed through the nodes are our suggestion.

IV. CONCLUSION AND DISCUSSION

In hedonic pricing regression usually employ a variable such as the distance drawn from one or some important locations that we considered as a station or an amenity in the urban context.

When more than two variables are taken, terrible collinearity may cause fluctuations in regression. Unfortunately these distance variables are not as effective as we excepted. When the observing location is in some specific area such as the region close to one landmark or share the same line with the two landmarks, multicollinearity problem will appear because of these invalid data. They will cause the instability and error in regression too.

In the estimation of two specific landmarks model, both wide region and tiny region simulation achieved acceptable variance and mean value. Regression results showed that the Hedonic price model is comparatively stable under the condition of two distance variables. The method considering the correlation

coefficients of the distance variables can lower down the bias in regression. It is possible to reduce the potential collinearity problem, and researchers should avoid gathering the samples as:

1) The samples located at the area are too far away from the circle that passed through the facilities which are considered as specific landmarks or amenity/hazard (Q_1 , Q_2).
2) The samples located at the area are just right on or close to the landmarks (Q_1 , Q_2).
3) The samples from the area are right on or close to the origin (the midpoint of the two landmarks).

We suggest gathering the data as the following methods:

1) Sampling the data which are located on or close to the circumcircle of the landmarks.
2) When the sampling area is limited to some specific location, we should take the observations which are located on or close to the circumcircle of the landmarks. And setting θ (the angle of $\angle SOQ2$, which is the angle of the line that passed through the center point of the tiny observation region, the midpoint and the line that passed through the two landmarks) equal to 30, 60 and 90 degree. Since these areas are easy to be identified in real world database. They are more effective than other data in regression.

However, when the third facility is considered in hedonic analysis, the multicollinearity problem will cause unacceptable variance and mean value. Then estimating the effect of three landmarks on specific observation region and the method to lower down the bias caused by the spatial collinearity problems will be our new task in the future.

REFERENCES

[1] Silvey, F. *"Multicollinearity and Imprecise Estimation", Journal of the Royal Statistical Society, Series B, 35,* 1985, pp.99-115
[2] Lancaster, K.J.*"A new approach to consumer theory", Journal of Political Economy 74,* 1966 , pp.132–57.
[3] Sherwin Rosen, *"Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", Journal of Political Economy, Vol. 82, No. 1,* Jan. - Feb., 1974, pp. 34-55
[4] Richard J. Howarth, *"A History of Regression and Related Model-Fitting in the Earth Sciences (1636?-2000)", Natural Resources Research,* Vol.10, No.4, 2001.
[5] Herl, Arthur E. and Robert W. Kennard. *"Ridge Regression: Biased Estimation of Nonorthogonal Problems", Technimetrics, 12,* 1970a. pp.55-67.
[6] Harrison, D., & Rubinfeld, D, *"Hedonic housing prices and demand for clean air.", Journal of Environmental Economics and Management, 5,* 1978, pp.81-102.
[7] Eric Heikkila, *"Multicollinearity in Regression Models with Multiple Distance Measures", Journal of Regional Science, vol.28. No.3,* 1998.
[8] Fik, T. J., Ling, D. C., & Mulligan, G. F. *"Modeling spatial variation in housing prices a variable interaction approach". Real Estate Economics, 31(4),* 2003, pp.623-646.
[9] Fik, T.J.,Ling, D.C.,& Mulligan, G.F. *"Modeling spatial variation in housing prices: a variable interaction approach.", Real Estate Economics, 31(4),* 2003,pp.623-646
[10] Noonan, D.S., Krupka, D.J.,& Baden, B. M, *"Neighborhood dynamics and price effects of superfund site clean-up", Journal of Regional Science, 47(4),* 2007,pp.665-692.
[11] Justin M. Ross, Michael C. Farmer, Clifford A. Lipscomb, *"Inconsistency in Welfare Inferences from Distance Variables in Hedonic Regressions", Journal of Real Estate Finance and Economics, 43,* 2011, pp.385-400.