Multiple Linear Regression Model for Estimating the Price of a Housing Unit*

[1]P. Boye, [1]D. Mireku-Gyimah and [2]C. A. Okpoti
[1]University of Mines and Technology, P O Box 237, Tarkwa, Ghana
[2]University of Education, Winneba, P O Box 25, Winneba, Ghana

Boye, P., Mireku-Gyimah, D. and Okpoti, C. A. (2017), "Multiple Linear Regression Model for Estimating the Price of a Housing Unit", *Ghana Mining Journal,* Vol. 17, No. 2, pp. 66 - 77.

Abstract

This paper uses the respective unit costs, over fifteen (15) years, of selected Housing Unit Major Components (HUMC): cement, iron rods, aluzinc roofing sheets, coral paint, wood and sand, to develop Multiple Linear Regression Model (MLRM) for determining Housing Unit Price (HUP) for one-bedroom and two-bedroom housing units. In the modeling, the Ordinary Least Squares (OLS) normality assumption which could introduce errors in the statistical analyses was dealt with by log transformation of the data, ensuring the data is normally distributed and there is no correlation between them. Minimisation of Sum of Squares Error method was used to derive the model coefficients. The resultant MLRM is: $\hat{Y}_{i\,MLRM} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{Y}(x_i')$ where **X** is the sample data matrix. The specific model for one-bedroom housing unit is $\log_e (HUP_{MLRM})_{1\text{-Bed}} = 1.017 - 2.225 \times 10^{-5} \times CC + 2.512 \times 10^{-6} \times CS + 6.016 \times 10^{-4} \times CIR + 1.985 \times 10^{-4} \times CR + 5.694 \times 10^{-4} \times CP - 7.437 \times 10^{-4} \times CW$ and that for two-bedroom housing unit is $\log_e (HUP_{MLRM})_{2\text{-Bed}} = 5.760 - 7.501 \times 10^{-7} \times CC + 2.935 \times 10^{-6} \times CS + 1.898 \times 10^{-3} \times CIR + 6.695 \times 10^{-4} \times CR - 9.157 \times 10^{-3} \times CP + 6.136 \times 10^{-3} \times CW$, where CC, CS, CIR, CR, CP and CW are costs of the total quantity of cement, sand, iron rods, roofing, paint and wood respectively. The MLRM was validated by using it to estimate the known HUP in the 15.5th year. From the results, the percentage absolute deviations of the estimated HUP from the known HUP are 1.27% and 2.02% for one-bedroom and two-bedroom housing units respectively, which are satisfactory. The novel approach presented in this paper is a valuable contribution to the body of knowledge in modeling.

Keywords: Multiple Regression Analysis, Housing Unit Major Components, Housing Unit Price

1 Introduction

Apart from providing shelter, a house constitutes a major component of wealth. It is considered as an important form of savings that could serve as a hedge against inflation in the medium term. In other instances, it is utilised as collateral for borrowing, thereby generating funds for other investments and wealth creation. In addition, housing construction and renovation boost the economy through an increase in aggregate expenditures, employment and volume of house sales. Thus, the housing industry has the capacity to both cultivate and protect wealth (Anon., 2007a). Consequently, in Ghana, like anywhere in the world, it is the dream of individuals, institutions, companies and the government to own a housing property or housing properties.

Since Ghana's independence, provision of housing has remained central to the development agenda. Various policies have sought to address issues such as land tenure, land title regulation, and provision of affordable housing units to the working population. Nonetheless, the implementation of a number of these housing policies has been negatively affected by lack of funds, poor macroeconomic environment and lack of private sector participation. In recent times, however, improved macroeconomic environment in the country has attracted Real Estate Developers (RED) into the housing industry to facilitate the sale of both old and new housing properties (Anon., 2007b). The general concern of prospective buyers is the disparity in the price of a housing unit. RED usually would like to maximise their profit in pricing their housing units, while prospective buyers would like to get good value for realistic price.

To find an optimal price, researchers have tried some approaches such as Sales Comparison Method (SCM) and Multiple Linear Regression Method (MLRM) to determine the HUP. The SCM suggests that the value of the subject property equals sale prices of similar properties that have been sold recently and are in close proximity to the subject property with due consideration to adjustments for dissimilar characteristics (Isakson, 2002; Chaphalkar and Dhatunde, 2015). Since each housing unit is unique, the adjustments made by valuers may be inconsistent and speculative and thus cannot be relied upon to give realistic HUP. The MLRM can give better estimates of HUP (Chaphalkar and Dhatunde, 2015) but the possible multicollinearity issues in the independent variables and the assumption that they are normally distributed are sometimes not properly resolved by researchers. Again, researchers that have developed MLRMs to estimate HUP have often included intangible housing characteristics as the independent variables, such as quality of neighbourhood (King, 1976) and location (Ayan

**Manuscript received September 15, 2017*
Revised version accepted November 30, 2017
https://dx.doi.org/10.4314/gm.v17i2.9

and Erkin, 2014), all of which may not help in estimating the realistic HUP.

This study determines the HUP based on the monetary cost of the following HUMC: Cement, Sand, Iron Rods, Aluzinc Roofing Sheets, Coral Paints and Wood, all of which are tangible and the issue of multicollinearity in these independent variables and the assumption that they are normally distributed have been resolved by log transformation of the sample data.

2 Resources and Methods Used

2.1 Resources

This study used three main resources:

(i) Data comprising the quantities of Housing Unit Major Components (HUMC) obtained from Regimanuel Gray Estates Ltd., an estate development agency in Accra Metropolitan Area. The quantities and their units of measurement are: Cement (kg), Sand (m^3), Iron Rods (t), Aluzinc Roofing Sheets (m^2), Coral Paints (l), and Wood (m^3). See Table 1.

(ii) The respective unit costs of the HUMC over a 15 year period obtained through market survey (see Table 2).

(iii) Statistical software, R, and other computing facilities available at the University of Mines and Technology, Tarkwa and University of Cape Town, South Africa.

Table 1 Quantities of Housing Unit Major Components

Material	Unit	1 Bedroom	2 Bedroom
Cement	kg	35 640.00	40 200.00
Sand	m^3	86.00	99.00
16mm Iron Rods	t	2.00	3.00
Aluzinc Roofing	m^2	365.00	678.00
Coral Paint	l	287.50	322.00
Wood	m^3	4.81	6.36

Table 2 Unit Price of Housing Unit Major Components (US $), 2003 – 2017.5

	Year									
Material	2003	2003.5	2004	2004.50	2005	2005.50	2006	2006.5	2007	2007.50
Cement (kg)	0.0017	0.0025	0.003	0.0083	0.0147	0.017	0.0183	0.02	0.0222	0.0243
Sand (m^3)	17.06	22.17	27.28	32.42	37.50	40.60	43.75	58.83	73.75	84.30
16mm Iron Rods (t)	159.76	167.80	175.88	183.96	192	204.60	217.20	218.89	220.55	228.60
Aluzinc Roofing (m^2)	3.47	3.84	4.21	4.65	4.95	5.30	5.79	6.26	6.63	7.00
Coral Paint (l)	0.02	0.20	0.38	0.56	0.74	0.93	1.13	1.33	1.52	1.71
Wood (m^3)	58.23	67.78	76.76	85.71	95.20	105.34	115.27	125.31	135.61	145.92

Material	2008	2008.5	2009	2009.5	2010	2010.5	2011	2011.5	2012	2012.5
Cement (kg)	0.0258	0.0267	0.0438	0.0513	0.0567	0.0617	0.0697	0.0735	0.0742	0.0750
Sand (m^3)	95.00	106.14	117.19	118.40	119.69	120.50	121.25	58.83	123.00	0.0750
16mm Iron Rods (t)	236.73	245.98	255.15	266.60	278.05	282.78	287.50	295.00	302.50	311.25
Aluzinc Roofing (m^2)	7.47	7.92	8.31	8.70	9.15	9.58	9.99	10.40	10.83	11.34
Coral Paint (l)	1.91	2.11	2.30	2.50	2.69	2.89	3.08	3.27	3.47	3.68
Wood (m^3)	155.95	165.99	176.30	186.60	196.64	206.67	216.98	227.29	237.32	247.36

Material	2013	2013.5	2014	2014.5	2015	2015.5	2016	2016.5	2017	2017.5
Cement (kg)	0.0780	0.1090	0.1375	0.1383	0.1417	0.1488	0.1533	0.1583	0.1658	0.1717
Sand (m^3)	125.00	131.20	137.5	143.76	150.00	155.10	160.22	165.37	170.40	178.00
16mm Iron Rods (t)	320.00	328.70	337.50	353.43	369.25	377.30	385.37	393.44	401.49	409.55
Aluzinc Roofing (m^2)	11.67	12.00	12.51	12.82	13.13	13.50	13.87	14.25	14.60	14.97
Coral Paint (l)	3.86	4.05	4.25	4.46	4.65	4.83	5.01	5.19	5.37	5.54
Wood (m^3)	257.66	267.97	278.00	288.04	298.35	307.81	316.79	325.90	335.23	345.16

2.2 Methods

2.2.1 Development of Multiple Linear Regression Model (MLRM)

According to Montgomery *et al.* (2008), Verbeek (2004) and Brooks (2008), Multiple Linear Regression analysis is a statistical technique for modeling and investigating the relation between a response variable and one or more predictor variables.

Let the response variable the regression seeks to explain be the actual HUP denoted by Y and the (k-1) explanatory variables used to explain the variations in the response variable be the HUMC denoted by $x_2, x_3, ..., x_k$. The aim of this paper is to model the *best-fitting line* to the sample data that would minimise the Sum of Squared Errors (SSE).

Consider Equation (1) to be the *assumed* best-fitting model to the sample data

$$\hat{Y}_i = \beta_1 + \beta_2 x_{i2} + ... + \beta_k x_{ik} \quad (1)$$

where

$\hat{Y}_i$ is the i^{th} estimated HUP.

$\beta_1, \beta_2, ..., \beta_k$ are constants to be determined from the sample data.

The difference between the actual HUP, Y_i, and the estimated $H\hat{U}P$, $\hat{Y}_i$, is given as

$$e_i = Y_i - [\beta_1 + \beta_2 x_{i2} + ... + \beta_k x_{ik}] \quad (2)$$

where

e_i is the sample errors.

From vector algebra, Equation (2) can be written in a compact form as

$$\mathbf{e}_i = \mathbf{Y}_i - \boldsymbol{\beta}\mathbf{X}_i$$

where

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{1} & \mathbf{x}_{i2} & \mathbf{x}_{i3} & ... & \mathbf{x}_{ik} \end{pmatrix}' \ \forall \ i = 1, 2, ..., N$$

and

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 & \beta_2 & ... & \beta_k \end{pmatrix}'$$

The SSE objective function is therefore given as

$$SSE(\boldsymbol{\beta}) = \sum_{i=1}^{N} (\mathbf{Y}_i - \boldsymbol{\beta}\mathbf{X}_i)^2 \quad (3)$$

From Equation (3), the SSE can be rewritten in matrix notation as

$$\text{SSE}(\boldsymbol{\beta}) = \mathbf{e}'\mathbf{e} = \begin{bmatrix} e_1 & e_2 & \cdots & e_n \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

Let $\hat{\boldsymbol{\beta}}$ be an estimator of $\boldsymbol{\beta}$.

$$\text{SSE } \boldsymbol{\beta} = e_1^2 + e_2^2 + ... + e_n^2 = \sum_{i=1}^{n} e_i^2$$

$$= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

$$= \left[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right]'\left[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right]$$

$$= \left[(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\right]\left[(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right]$$

$$= (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$
$$+ (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

$$= (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$
$$+ (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + \mathbf{I}$$

$$\geq (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (4)$$

Equation (4) clearly shows that the minimisation of $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ is $(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$.

Hence

$$\text{SSE}(\hat{\boldsymbol{\beta}}) = (\mathbf{Y}' - \hat{\boldsymbol{\beta}}'\mathbf{X}')(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$
$$= \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$
$$= \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \quad (5)$$

where

$$\mathbf{X} = \begin{pmatrix} 1 & x_{12} & \cdots & x_{1K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N2} & \cdots & x_{NK} \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$$

In the $N \times K$ matrix $\mathbf{X}$, the i^{th} row refers to observation i, and the k^{th} column refers to the k^{th} explanatory variable.

From matrix algebra, Equation (5) can be partially differentiated with respect to $\hat{\boldsymbol{\beta}}$ to obtain Equation (6) as follows:

$$\frac{\partial\left(SSE\left(\hat{\boldsymbol{\beta}}\right)\right)}{\partial\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \qquad (6)$$

From Equation (6), a necessary condition for a minimum value to be attained is

$$-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0} \qquad (7)$$

The `*normal equations* are then obtained from Equation (7) as

$$\left(\mathbf{X}'\mathbf{X}\right)\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y} \qquad (8)$$

On the assumption that $\left(\mathbf{X}'\mathbf{X}\right)$ is non-singular, the *Ordinary Least Squared (OLS)* estimator of the betas in the Multiple Linear Regression model in Equation 1 is given in matrix form as

$$\hat{\boldsymbol{\beta}}_{OLS} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \qquad (9)$$

Hence, the Multiple Linear Regression Model (MLRM) is

$$\hat{Y}_i = \hat{\boldsymbol{\beta}}_{OLS}\mathbf{x}_i' = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'Y\left(\mathbf{x}_i'\right) \qquad (10)$$

Hypothesis Testing

The hypothesis governing the MLR theory is as follows:

H_0: HUMC do not contribute to HUP determination.

H_1: At least one HUMC contribute(s) to HUP determination.

2.2.2 Analysis of Data

In formulating the MLRM, the quantities of HUMC data in Table 1 was converted to monetary values in US Dollars by using their respective unit costs of the total quantity of each HUMC in Table 2 to obtain secondary data for one-bedroom and two-bedroom housing units. See Tables 3 and 4. The monetary values were then used to plot scatter diagrams to assess the correlation among the HUMC. See Figs. 1 and 2. Distributions of observed prices for one-bedroom and two-bedroom housing units were plotted to verify the Ordinary Least Squares (OLS) normality assumption. See Figs. 3 and 6. It can be seen that the OLS normality assumption was not met since the distributions were 'skewed' to the right. Consequently, log transformed costs for one-bedroom and two-bedroom housing unit distributions were plotted as shown by Figs. 4 and 7 . Here again, the distribution did not appear to be normal but their normal quantile-quantile plots (see Figs. 5 and 8) showed the characteristic straight line of a normal distribution, indicating that the distributions of the log-transformed prices for one-bedroom and two-bedroom housing units are approximately normal. Since the normality assumption for the sample data has been achieved, the MLRM was formulated by using log-log linear model. That is, log HUP was constructed as well as log HUMC since the sample data are of different scale. Statistical analyses were then performed on these log variables to obtain the model coefficients, betas, as shown in Equation 9. Consequently, the MLRM was obtained using Equation 10. Finally, the OLS residual normality assumption for one-bedroom and two-bedroom housing units were checked by looking at the distributions of the least squares residuals (see Figs. 9 and 10) and their normal quantile- quantile plots (see Fig. 11 and 12). Since the plotted points in the Figs. 11 and 12 lie close to the 45^{o} diagonal line, the least squares residuals are approximately normally distributed.

Table 3 Price of Housing Unit and Cost of Total Quantity of HUMC (US $) 2003 – 2017
(One-Bedroom)

Year	Housing Unit Price (HUP)	Cost of Cement (CC)	Cost of Sand (CS)	Cost of Iron Rods (CIR)	Cost of Roofing (CR)	Cost of Paint (CP)	Cost of Wood (WP)
2003	31 455.00	60.59	1 467.16	319.52	1 266.55	5.75	280.33
2003.5	33 260.00	89.10	1 906.62	335.60	1 401.60	57.50	326.30
2004	35 065.00	106.92	2 346.08	351.76	1 536.65	109.25	369.53
2004.5	36 870.00	295.81	2 788.12	367.92	1 697.25	161.00	412.62
2005	38 675.00	523.91	3 225.00	384.16	1 806.75	212.75	458.30
2005.5	40 587.50	605.88	3 491.60	409.20	1 934.50	267.38	507.12
2006	42 500.00	652.21	3 762.50	434.40	2 113.35	324.88	554.92
2006.5	42 500.00	712.80	5 059.38	439.78	2 284.90	382.38	603.26
2007	42 500.00	791.21	6 342.50	441.10	2 419.95	437.00	652.84
2007.5	44 604.00	866.05	7 249.80	457.20	2 555.00	491.63	702.47
2008	46 708.00	919.51	8 170.00	473.46	2 726.55	549.13	750.76
2008.5	49 020.00	951.56	9 128.04	491.96	2 890.80	606.63	799.09
2009	51 332.00	1 561.03	10 078.34	510.30	3 033.15	661.25	848.73
2009.5	51 332.00	1 828.33	10 182.40	533.20	3 175.50	718.75	898.31
2010	51 332.00	2 020.79	10 293.34	556.10	3 339.75	773.38	946.65
2010.5	53 873.00	2 198.99	10 363.00	565.56	3 496.70	830.88	994.93
2011	56 414.00	2 484.11	10 427.50	575.00	3 646.35	885.50	1 044.56
2011.5	59 206.50	2 619.54	10 500.60	590.00	3 796..00	940.13	1 094.20
2012	61 999.00	2 644.49	10 578.00	605.00	3 952.95	997.63	1 142.48
2012.5	61 999.00	2 673.00	10 668.30	622.50	4 139.10	1 058.00	1 190.82
2013	61 999.00	2 779.92	10 750.00	640.00	4 259.55	1 109.75	1 240.40
2013.5	65 068.00	3 884.76	11 283.20	657.40	4 380.00	1 164.38	1 290.03
2014	68 137.00	4 900.50	11 825.00	675.00	4 566.15	1 221.88	1 338.32
2014.5	69 942.00	4 929.01	12 363.36	706.86	4 679.30	1 282.25	1 386.65
2015	71 747.50	5 050.19	12 900.00	738.50	4 792.45	1 336.88	1 436.29
2015.5	73 552.00	5 303.23	13 338.60	754.60	4 927.50	1 388.63	1 481.83
2016	75 357.50	5 463.61	13 778.92	770.74	5 062.55	1 440.38	1 525.06
2016.5	77 162.00	5 641.81	14 221.82	786.88	5 201.25	1 492.13	1 568.92
2017	78 967.50	5 909.11	14 654.40	802.98	5 329.00	1 543.88	1 613.83
2017.5		6 119.39	15 339.66	819.10	5 464.05	1592.75	1 661.63

Table 4 Price of Housing Unit and Cost of Total Quantity of HUMC (US $) 2003 – 2017
(Two-Bedroom)

Year	Housing Unit Price (HUP)	Cost of Cement (CC)	Cost of Sand (CS)	Cost of Iron Rods (CIR)	Cost of Roofing (CR)	Cost of Paint (CP)	Cost of Wood (CW)
2003	34 500.00	68.34	1 688.94	479.28	2 352.66	6.44	370.49
2003.5	37 280.00	100.50	2 194.83	503.40	2 603.52	64.40	431.26
2004	40 070.00	120.60	2 700.72	527.64	2 854.38	122.36	488.39
2004.5	41 880.00	333.66	3 209.58	551.88	3 152.70	180.32	545.34
2005	43 680.00	590.94	3 712.50	576.00	3 356.10	238.28	605.72
2005.5	45 841.00	683.40	4 019.40	613.80	3 593.40	299.46	670.24
2006	48 000.00	735.66	4 331.25	651.60	3 925.62	363.86	733.42
2006.5	48 000.00	804.00	5 825.17	656.67	4 244.28	428.26	797.30
2007	48 000.00	892.44	7 301.25	661.65	4 495.14	489.44	862.83
2007.5	53 579.00	976.86	8 345.70	685.80	4 746.00	550.62	928.43
2008	59 160.00	1 037.16	9 405.00	710.19	5 064.66	615.02	992.25
2008.5	64 001.00	1 073.34	10 507.86	737.94	5 369.76	679.42	1 056.13
2009	68 840.00	1 760.76	11 601.81	765.45	5 634.18	740.60	1 121.73
2009.5	74 419.00	2 062.26	11 721.60	799.8	5 898.60	805.00	1 187.26
2010	80 000.00	2 279.34	11 849.31	834.15	6 203.70	866.18	1 251.14
2010.5	80 000.00	2 480.34	11 929.50	848.34	6 495.24	930.58	1 314.96
2011	80 000.00	2 801.94	12 003.75	862.50	6 773.22	991.76	1 380.56
2011.5	89 001.00	2 954.70	12 087.90	885.00	7 051.20	1 052.94	1 446.16
2012	98 000.00	2 982.84	12 177.00	907.50	7 342.74	1 117.34	1 509.97
2012.5	103 671.00	3 015.00	12 280.95	933.75	7 688.52	1 184.96	1 573.85
2013	109 340.00	3 135.60	12 375.00	960.00	7 912.26	1 242.92	1 639.39
2013.5	114 919.00	4 381.80	12 988.80	986.10	8 136.00	1 304.10	1 704.99
2014	120 500.00	5 527.50	13 612.50	1 012.50	8 481.78	1 368.50	1 768.80
2014.5	126 727.00	5 559.66	14 232.24	1 060.29	8 691.96	1 436.12	1 832.68
2015	132 955.00	5 696.34	14 850.00	1 107.75	8 902.14	1 497.30	1 898.28
2015.5	139 183.00	5 981.76	15 354.90	1 131.90	9 153.00	1 555.26	1 958.47
2016	145 410.00	6 162.66	15 861.78	1 156.10	9 403.86	1 613.22	2 015.61
2016.5	151 637.00	6 363.66	16 371.63	1 180.32	9 661.50	1 671.18	2 073.57
2017	157 865.00	6 665.16	16 869.60	1 204.47	9 898.80	1 729.14	2 132.93
2017.5		6 902.34	17 424.27	1 228.65	10 149.66	1 783.88	2 193.53

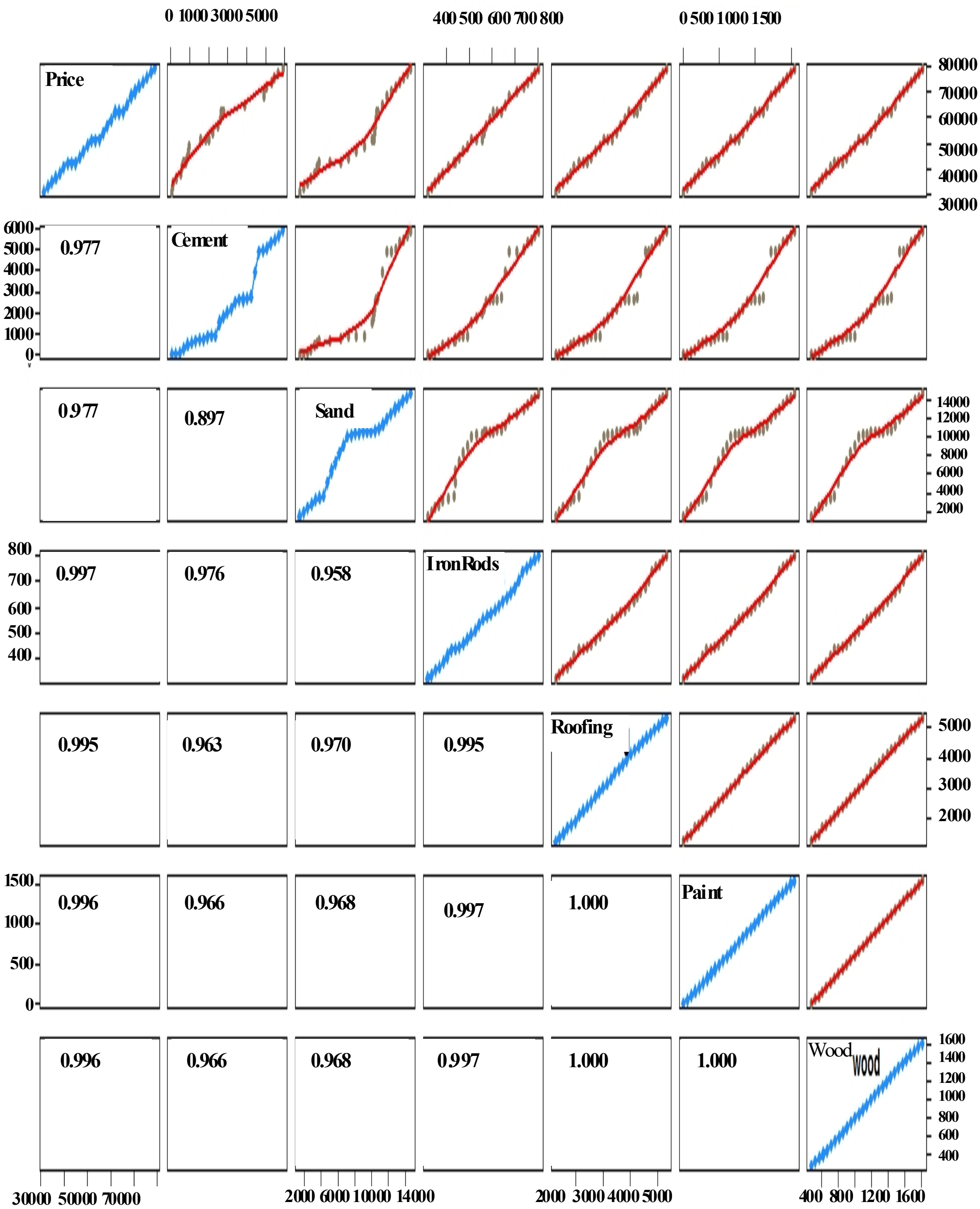

Fig. 1 Scatterplot of Housing Unit Major Components for One-Bedroom Unit

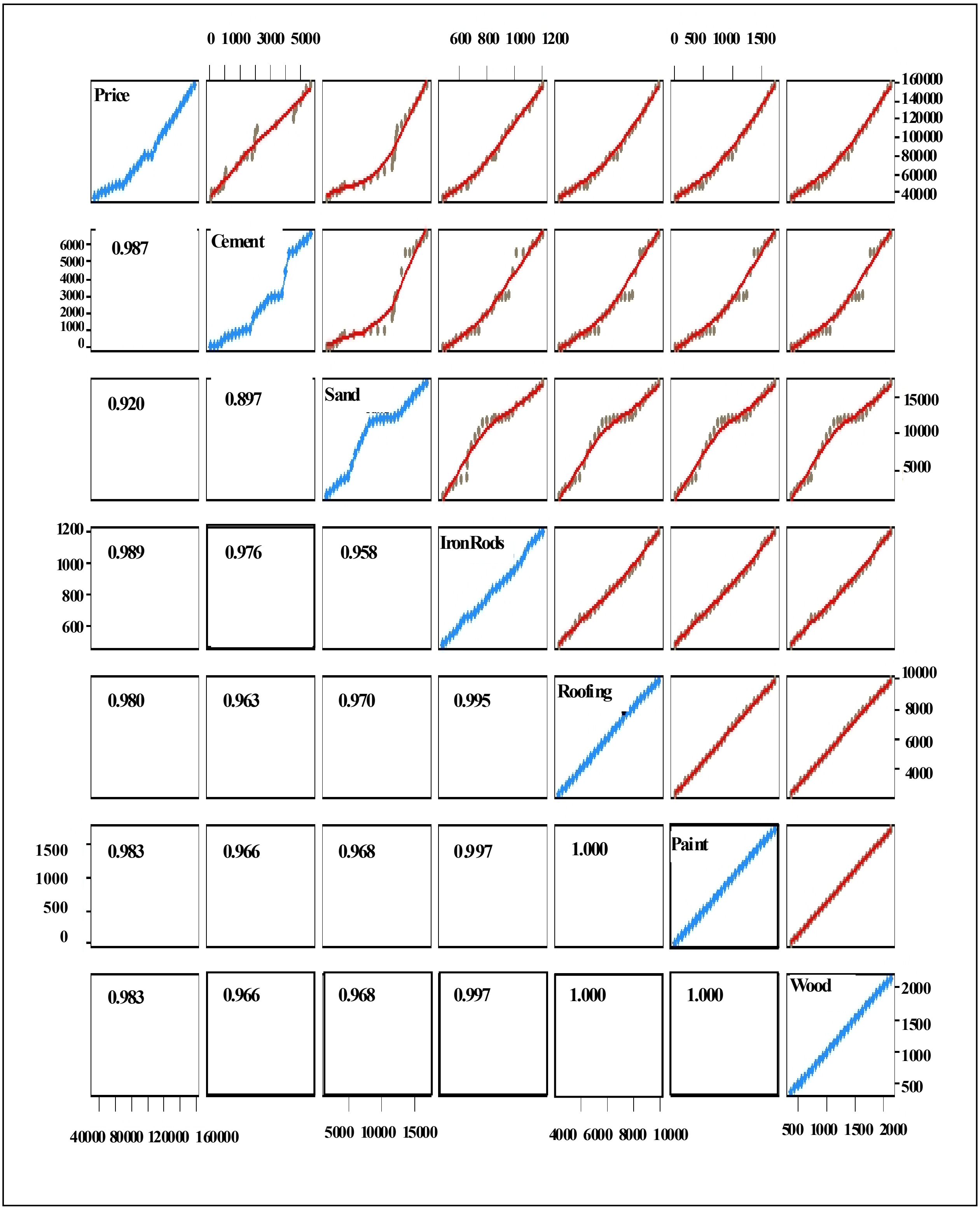

Fig. 2 Scatterplot of Housing Unit Major Components for Two-Bedroom Unit

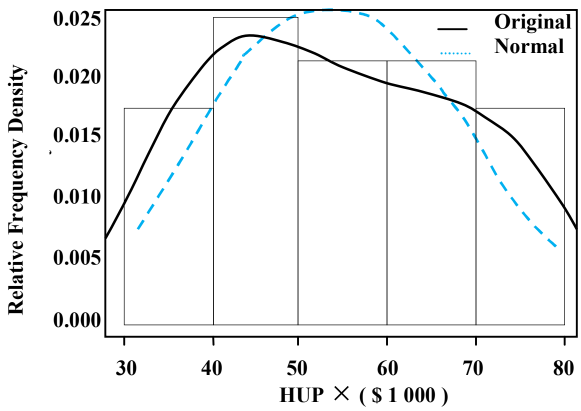

Fig. 3 Distribution of Observed Prices for One-Bedroom Housing Unit

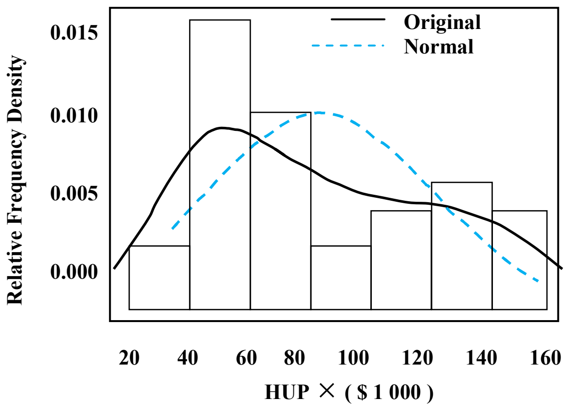

Fig. 6 Distribution of Observed Prices for Two-Bedroom Housing Unit

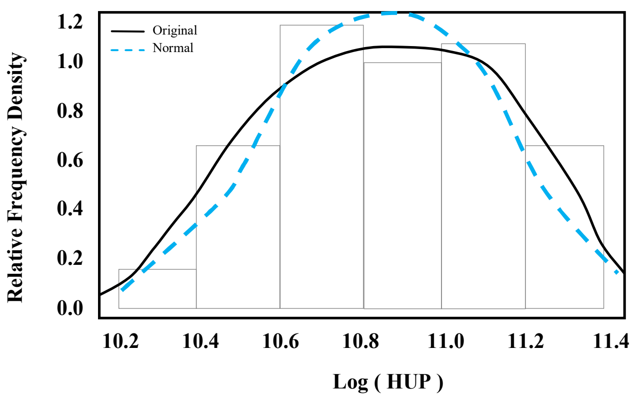

Fig. 4 Distribution of Log Transformed of Prices of One-Bedroom House

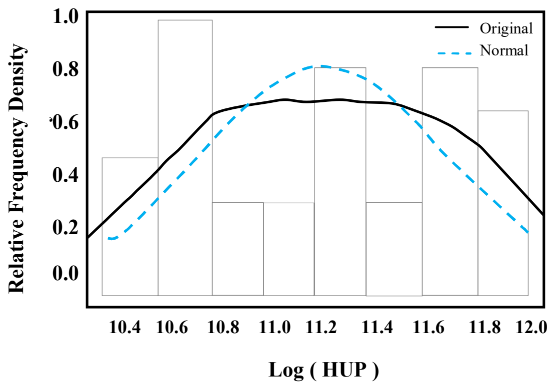

Fig. 7 Distribution of Log Transformed Prices of Two-Bedroom House

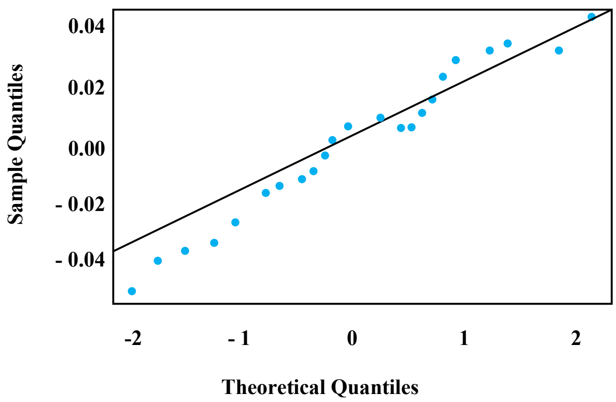

Fig. 5 Normal Quantile-Quantile Plot for One- Bedroom Unit

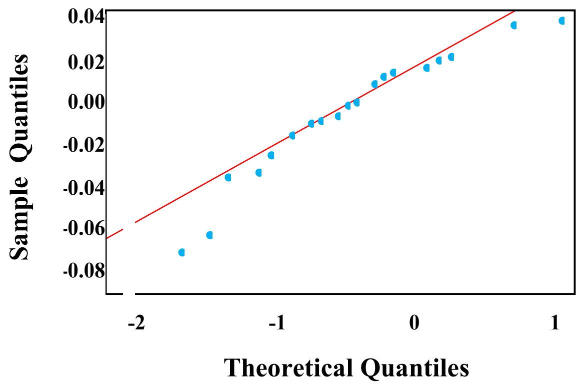

Fig. 8 Normal Quantile-Quantile Plot for Two-Bedroom Unit

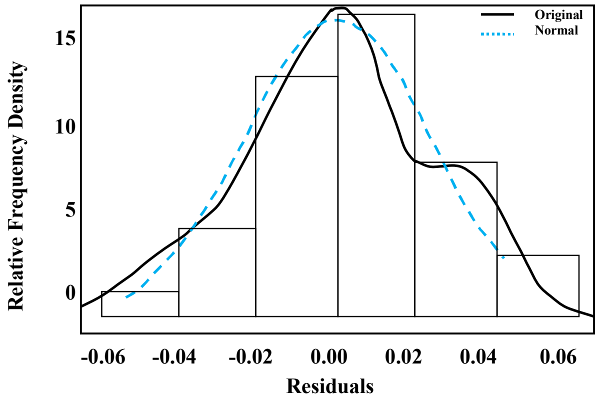

Fig. 9 Ordinary Least Squares Residuals Plot (One-Bedroom House)

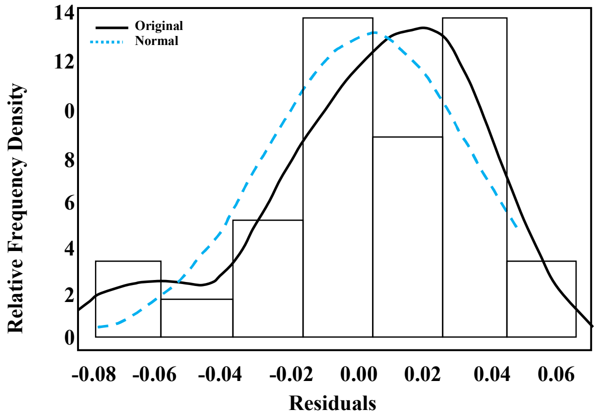

Fig. 10 Ordinary Least Squares Residuals (Two-Bedroom House)

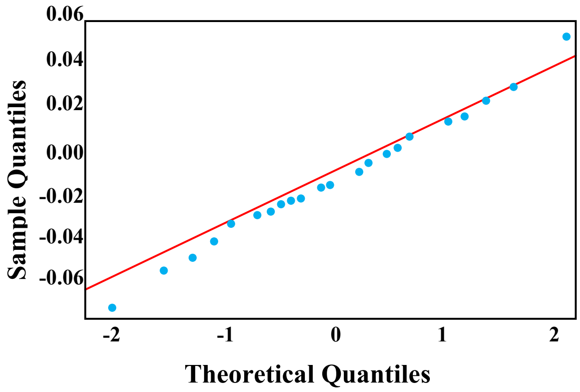

Fig. 11 Normal Quantile-Quantile for Residuals (One-Bedroom Unit)

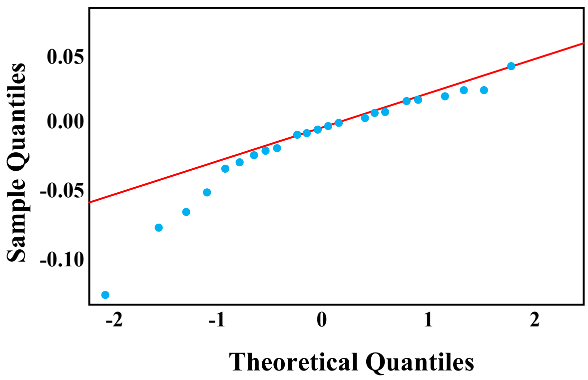

Fig. 12 Normal Quantile-Quantile Plot for Residuals (Two-Bedroom Unit)

3 Results and Discussion

The developed MLRM for one-bedroom and two-bedroom housing units is as shown in Equation 10. After derivation of the coefficients, β, the respective equations for one-bedroom and two-bedroom housing units are as follows:

$$\log_e (HUP_{MLRM})_{1\text{-}Bed} = 1.017 - 2.225 \times 10^{-5} \times CC + 2.512 \times 10^{-6} \times CS + 6.016 \times 10^{-4} \times CIR + 1.985 \times 10^{-4} \times CR + 5.694 \times 10^{-4} \times CP - 7.437 \times 10^{-4} \times CW \quad (11)$$

$$\log_e (HUP_{MLRM})_{2\text{-}Bed} = 5.760 - 7.501 \times 10^{-7} \times CC + 2.935 \times 10^{-6} \times CS + 1.898 \times 10^{-3} \times CIR + 6.695 \times 10^{-4} \times CR - 9.157 \times 10^{-3} \times CP + 6.136 \times 10^{-3} \times CW \quad (12)$$

where

CC = Cost of total quantity of cement
CS = Cost of total quantity of sand
CIR = cost of total quantity of iron rods
CR = cost of total quantity of roofing
CP = cost of total quantity of paint and
CW = cost of total quantity of wood

The standard error of prediction using simple linear regression is the residual standard deviation on the basis that it was an estimate of the standard deviation of the 'error process' which produced deviations of individual points around the linear function and it is an estimate of the accuracy of the dependent variable being measured. The summary results under Tables 5 and 6 show the results of the residual standard error values for one-bedroom and two-bedroom housing units as: 0.02688 and 0.03486 respectively. Since the standard errors are within acceptable limits, it can be concluded that the developed MLRM fitted the sample data very well.

The coefficient of determination R^2 is a measure of how well the developed MLRM performs as a predictor of the HUP (measure of fit). The higher the R^2 value, the more useful the model is. Since R^2 can take on any value between zero and one, with a value closer to 1 indicating that a greater proportion of variance in the data about the mean was accounted for by the model.

It must be note that, if the number of fitted coefficients in a model is increased, R^2 value will increase although the fit may not improve in a practical sense. To avoid this situation, the adjusted R^2 statistics is recommended for use. The summary results under Tables 5 and 6 show the coefficient of determination for one-bedroom and two-bedroom housing units respectively as follows: 99.23% and 99.58%. This implies that 99.23% and 99.58 of the variabilities observed in the HUPs for one-bedroom and two-bedroom housing units respectively were explained by their respective HUMC. These R^2 values show that the HUMC contributed a lot of information about the observed HUP.

In model adequacy analysis, the F-statistic must be used in combination with the p-value when deciding if the overall model coefficients are significant or not. If the p-value is less than the significance (alpha) level, the hypothesis test is statistically significant. From the summary results under Table 5, the F-statistic = 470.2 and p-value = $2.2 \times 10^{-16} < \alpha = 0.05$ level of significance for the one-bedroom housing unit. This is an indication that the model coefficients contributed significantly to the prediction of the housing unit price. From Table 6, the F-statistic = 873.9 and p-value = $2.2 \times 10^{-16} < \alpha = 0.05$ level of significance for the two-bedroom housing unit. This is also an indication that the model coefficients contributed significantly to the prediction of the housing unit price.

Explanation to One-Bedroom Housing Unit Model Coefficients

The coefficient of the intercept is 1.017e+01. It shows the value of beta zero when all the HUMC values are held constant.
The coefficient of cement = -2.225e-05. It shows a negative contribution effect of cement on the estimated housing unit price, adjusting for all the other HUMC. Thus, one US Dollar increase in the price of cement will cause the estimated HUP to decrease by 2.225e-05 US Dollars.

Table 5 Ordinary Least Squares Regression Results (One-Bedroom)

	Coefficients of Fitted Model	
	Estimate	**Standard Error**
Intercept	1.017e+01	1.375e+00
Cement	-2.225e-05	1.630e-05
Sand	2.512e-06	6.509e-06
Iron Rods	6.016e-04	1.017e-03
Roofing	1.985e-04	3.594e-04
Paint	5.694e-04	3.527e-03
Wood	7.437e-04	3.445e-03

Summary Result:

Residual Standard Error: 0.02688 on 22degrees of freedom

Multiple R-squared: 0.9923, Adjusted R-squared: 0.9902

F-statistic: 470.2 on 6 and 22 DF, p-value: 2.2e-16

Table 6 Ordinary Least Squares Regression Results (Two-Bedrooms)

	Coefficients of Fitted Model	
	Estimate	**Standard Error**
Intercept	5.760e+00	1.77e+00
Cement	-7.501e-07	1.875e-05
Sand	2.935e-06	7.319e-06
Iron Rods	1.898e-03	8.866e-04
Roofing	6.695e-04	2.530e-04
Paint	-9.157e-03	4.043e-03
Wood	6.136e-03	3.344e-03

Summary Results:

Residual Standard Error: 0.03486 on 22 degrees of freedom

Multiple R-squared: 0.9958, Adjusted R-squared: 0.9947

F-Statistic: 873.9 on 6 and 22 DF, p-value: 2.2e-16

The coefficient of sand = 2.512e-06. It shows a positive contribution effect of sand on the estimated HUP, adjusting for all the other HUMC. Thus, one US Dollar increase in the price of sand will cause the estimated HUP to increase by 2.512e-06 US Dollars.

The coefficient of iron rods = 6.016e-04. It shows a positive contribution effect of iron rods on the estimated HUP, adjusting for all the other HUMC. Thus, one US Dollar increase in the price of iron rods will cause the estimated HUP to increase by 6.016e-04 US Dollars.

The coefficient of roofing = 1.985e-04. It shows a positive contribution effect of roofing on the estimated HUP, adjusting for all the other HUMC. Thus, one US Dollar increase in the price of roofing will cause the estimated HUP to increase by 1.985e-04 US Dollars.

The coefficient of paint = 5.694e-04. It shows a positive contribution effect of paint on the estimated HUP, adjusting for all the other HUMC. Thus, one US Dollar increase in the price of paint will cause the estimated HUP to increase by 5.694e-04.

The coefficient of wood = -7.437e-04. It shows a negative contribution effect of wood on the estimated HUP, adjusting for all the other HUMC. Thus, one US Dollar increase in the price of wood will cause the estimated HUP to decrease by -7.437e-04.

Explanation to Two-Bedroom Housing Unit Model Coefficients

The coefficient of the intercept = 5.760. It shows the value of beta zero when all the HUMC values are held constant.

The coefficient of cement = -7.501e-07. It shows a negative contribution effect of cement on the estimated HUP, adjusting for all the other HUMC. Thus, one US Dollar increase in the price of cement will cause the estimated HUP to decrease by 7.501e-07 US Dollars.

The coefficient of sand = 2.935e-06. It shows a positive contribution effect of sand on the estimated HUP, adjusting for all the other HUMC. Thus, one US Dollar increase in the price of sand will cause the estimated HUP to increase by 2.935e-06 US Dollars.

The coefficient of iron rods = 1.898e-03. It shows a positive contribution effect of iron rods on the estimated HUP, adjusting for all the other HUMC. Thus, one US Dollar increase in the price of iron rods will cause the estimated HUP to increase by 1.898e-03 US Dollars.

The coefficient of roofing = 6.695e-04. It shows a positive contribution effect of roofing on the estimated HUP, adjusting for all the other HUMC. Thus, one US Dollar increase in the price of roofing will cause the estimated HUP to increase by 6.695e-04 US Dollars.

The coefficient of paint = -9.157e-03. It shows a negative contribution effect of paint on the estimated HUP, adjusting for all the other HUMC. Thus, one US Dollar increase in the price of paint will cause the estimated HUP to decrease by 9.157e-03.

The coefficient of wood = 6.136e-03. It shows a positive contribution effect of wood on the estimated HUP, adjusting for all the other HUMC. Thus, one US Dollar increase in the price of wood will cause the estimated HUP to increase by 6.136e-03.

Model Validation

In order to find the efficiency of the developed MLRM, Equations 11 and 12 were used to estimate the known HUP in the 15.5 year for one-bedroom house and two-bedroom house. Table 7 is a summary of the results. From the results, the percentage absolute deviations, (Δ%), of the estimates of the HUP from the known HUP are between 1.27 and 2.03%, which are considered to be satisfactory.

Table 7 Estimated HUP and Respective Percentage Absolute Deviation (Δ%) from the Known HUP

Housing Unit	Known HUP ($)	Estimated HUP ($) from MLRM	Δ %
1-Bedroom	83 600.00	82 530.24	1.27
2-Bedroom	169 000.00	172 413.10	2.02

4 Conclusions and Recommendation

In this paper, MLRM has been developed from the unit costs of HUMC which are cement, iron rods, aluzinc, roofing sheets, coral paint, wood and sand over a period of 15 consecutive years, obtained from an estate development agency, to determine realistic HUP for one-bedroom and two-bedroom housing units. In developing the MLRM, multicollinearity which existed among the sample data and could have caused wrong statistical inferences was resolved by log transformation of the data to ensure that the data is normally distributed and there is no correlation between them. Subsequently, minimisation of Sum of Squares Error method was used to derive the model coefficients, betas. The resultant MLRM which accounted for 99% and 99% of the total variation in the sample data for the one-bedroom and two-bedroom housing units respectively is $\hat{Y}_{i\ MLRM} = (\mathbf{X}'\mathbf{X})^{-1}\ \mathbf{X}'\mathbf{Y}(x_i')$ where **X** is the sample data matrix.

The model which determined the HUP for one-bedroom housing unit is $\log_e (HUP_{MLRM})_{1\text{-}Bed} = 1.017 - 2.225 \times 10^{-5} \times CC + 2.512 \times 10^{-6} \times CS + 6.016 \times 10^{-4} \times CIR + 1.985 \times 10^{-4} \times CR + 5.694 \times 10^{-4} \times CP - 7.437 \times 10^{-4} \times CW$ and that which determined the HUP for two-bedroom housing unit is $\log_e (HUP_{MLRM})_{2\text{-}Bed} = 5.760 - 7.501 \times 10^{-7} \times CC + 2.935 \times 10^{-6} \times CS + 1.898 \times 10^{-3} \times CIR + 6.695 \times 10^{-4} \times CR - 9.157 \times 10^{-3} \times CP + 6.136 \times 10^{-3} \times CW$, where CC; CS; CIR; CR; CP and CW are the costs of the total quantity of cement, sand, iron rods, roofing, paint and wood respectively.

The absolute deviation (Δ %) using Equations (11) and (12) to estimate the HUP from the known HUP in the 15.5 year is 1.27% for one-bedroom housing unit and 2.02% for two-bedroom housing unit, meaning that the developed MLRMs are good.

The novel approach presented in this study for deriving the MLRM is a valuable contribution to the body of knowledge in modeling. The MLRM should give prospective house owners a timely, good idea of the price of a house they intend to purchase.

Acknowledgements

We are thankful to Regimanuel Gray Estate Ltd., an estate development agency in Accra Metropolitan Area for providing the primary data to make this study a successful one.

References

Anon. (2007a), 'The Housing Industry in Ghana: Prospects and Challenges', Research Department, Bank of Ghana, pp. 1 - 6.

Anon. (2007b), 'The Housing Market in Ghana', Research Department, Bank of Ghana, pp.1- 46.

Ayan, E. and Erkin, H. C. (2014), "Hedonic Modeling for a Growing Housing Market: Valuation of Apartments in Complexes", *International Journal of Economics and Finance*, Vol. 6, No. 3, pp. 1- 12.

Brooks, C. (2008), *Introductory Econometrics for Finance*, 2nd edition, Cambridge University Press, pp. 88 - 258.

Chaphalkar, N. B. and Dhatunde, M. (2015), 'Real Property Valuation Using Sales Comparison Method and Multiple Regression Analysis', *International Journal of Modern Trends in Engineering and Research*, Vol. 2, No. 8, pp. 304 – 315.

Isakson, H. R. (2002), 'The Linear Algebra of the Sales Comparison Approach', *Journal of Real Estate Research*, Vol. 24, No. 2, pp. 117 – 128.

King, A. T. (1976) 'The Demand for Housing: A Lancastrian Approach', *Southern Economic Journal*, Vol. 43, No. 2, pp. 1 - 13.

Montgomery, D. C., Jennings, C. L. and Kulahci, M. (2008), *Introduction to Time Series Analysis and Forecasting*, Hoboken, N.J: Wiley-Interscience (Wiley Series in Probability and Statistics), pp. 73 - 275.

Verbeek, M. (2004), *A Guide to Modern Econometrics*, 2nd edition, Southern Gate, Chichester, West Sussex, England Hoboken, NJ: John Wiley and Sons, pp. 7 - 14.

Authors

P. Boye is a Lecturer at the Department of Mathematical Sciences of the University of Mines and Technology (UMaT). He obtained his MSc and BSc degrees from Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana. He has completed his PhD programme at UMaT. His research interest is in applied Mathematics and Statistics.

D. Mireku-Gyimah is a Professor of Mining Engineering and a Consulting Engineer currently working at the University of Mines and Technology, Tarkwa, Ghana. He holds the degrees of MSc from the Moscow Mining Institute, Moscow, Russia, and PhD and DIC from the Imperial College of Science, Technology and Medicine, London, UK. He is a member of Institute of Materials, Minerals and Mining of UK and New York Academy of Sciences and also a fellow of Ghana Institution of Engineers and the Ghana Academy of Arts and Science. His research and consultancy works cover Mine Design and Planning, Mine Feasibility Study, Operations Research, Environmental Protection and Corporate Social Responsibility Management.

C. A. Okpoti is an Associate Professor of Mathematics at the University of Education, Winneba, Ghana. He holds the degree of PhD from Lulea University of Technology, Sweden; MPhil from the Institute of Mathematical Sciences, Ghana; MEd from the University of Leeds, UK; and BSc from the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. His research areas include Hardy's Inequalities and Functional Analysis and Quantitative Techniques and Mathematical Applications in Business