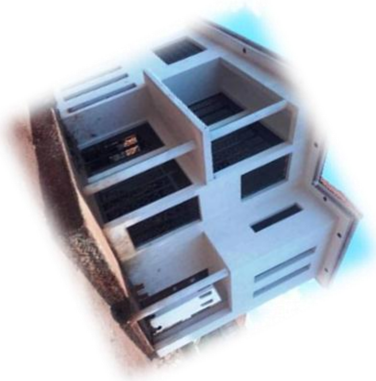
Uganda Bureau of Statistics
P.O. Box 7186, Kampala;
Tel: 0414 – 706000; Fax: 0414 – 237553; Email: ubos@ubos.org;
Website: www.ubos.org

RESIDENTIAL PROPERTY PRICE INDEX (RPPI) FOR GREATER KAMPALA METROPOLITAN AREA (GKMA) – UGANDA



First Quarter, 2018/19.

CONTENTS:

1. FOREWORD

This publication presents statistical information on real estate price developments on a quarterly basis. The information covers estimates of weighted residential Property Price Index (RPPI). The geographical coverage of the indices is the Greater Kampala Metropolitan Area, covering Kampala, and Wakiso urban areas.

The weights used in the development of RPPI are based on 2014 National Population and Housing Census results. The details on Stock of Houses by type occupied by the population were analyzed, considering the type of dwelling unit occupied by the household. Uganda Bureau of Statistics (UBOS) uses the method of Hedonic pricing approach to compute its Residential Property Price Index (RPPI). This is the recommended method internationally, as the best that deals with issues of quality changes. The hedonic method recognises that quality goods can be described as a function of their price determining characteristics.

The characteristics of residential properties considered include the following variables; House-type (bungalow, storied/flat, semidetached), Roofing (tiles, iron-sheet, non-if house is incomplete), Floor-type (tiles, cement, non-if house is incomplete), Number of garages, Number of bedrooms, Number of bathrooms, Number of toilets, fence (if available-yes/else-no), House-completeness, Availability of servant quarters /guest-house and Plot size(in square meters). The dependent variable is the quoted price of the residential property. The Laspeyres type index formula is used to compute the higher level indices while the lower level indices are computed by Jovens method of price relatives.

Uganda Bureau of Statistics appreciates the real estates managers, agents, brokers, developers and other stakeholders for providing the requisite data used to produce the information presented in this publication. Special gratitude is extended to the Bank of Uganda for their continued financial support, World Bank & International Monetary Fund (IMF) for technical assistance towards the Real Estate Price Index development.

Imelda Atai Musana
Ag. Executive Director

2. LIST OF ACRONYMS

BIS	Bank for International Settlements
BOU	Bank of Uganda
CPI	Consumer Price Index
EA	Elementary Aggregates
FSB	Financial Stability Board
GKMA	Great Kampala Metropolitan Area
HGMI	Hedonic Geometric Mean Imputation
IMF	International Monetary Fund
INSEE	National Institute of Statistics and Economic Studies
RPPI	Residential Property Price Indices
SNA	System of National Accounts
UBOS	Uganda Bureau of Statistics
WB	World Bank

3. INTRODUCTION

In collaboration with Bank of Uganda (BoU), the Uganda Bureau of Statistics embarked on the compilation of the Residential Property Price Index (RPPI) for Great Kampala Metropolitan Area (GKMA). The RPPI is a broad measure of the movement of residential property prices in the economy at a given time. It serves as a timely and accurate indicator of residential house price trends within GKMA.

Residential property prices are of significant interest to policy makers, market analysts and researchers for a range of economic and social reasons. This is because the housing market plays an important role in the Ugandan Economy. Housing is a major source of individual wealth in the country. Thus, changes in housing costs affect the disposable income of households, thereby influencing consumer spending and saving decisions. The RPPIs produced by the Bureau will serve the following purposes:

- As an input into Financial Stability models analyzed by Bank of Uganda
- A macroeconomic indicator of residential property price inflation
- Supporting the compilation of the non-financial assets component of the Household Balance Sheet in the Ugandan System of National Accounts (SNA).
- Analytical tool for estimating changes in the rates of mortgage defaults, prepayments and housing affordability in specific geographic areas within GKMA
- Used for international comparison of residential property price levels and changes.

This publication presents the technical methodology used to compute the RPPI and it is organized as follows: i) Section 4 provides an overview of the model selection; ii) Section 5 describes the data used to calculate the index, including sources and the data collection methodology, and provides information on limitations of this data; iii) Section 6 provides information on weights generation and computation; iv) Section 7 provides Hedonic models and index calculation; v) section 8 presents the results of the selected calculation methodology and model diagnostics and conclusion.

4. MODEL SELECTION AND JUSTFICATION

The RPPI measures the price change of residential dwellings over time. To capture these price changes, aggregated housing prices are needed over a period. However, accurate measures of aggregated housing prices are difficult to estimate. This is because dwellings are heterogeneous in nature and sales of a particular dwelling, especially in Uganda, are relatively infrequent. It is very rare to find two residential properties that are exactly the same: properties tend to differ according to various characteristics relating to the physical attributes of the properties or to their locations. Naturally, observed differences in characteristics between two houses will be reflected by differences in price.

The first requirement to calculate housing price indexes for a given period is to collect information on properties sold in that period. However, the quality (where quality refers to the characteristics of a residential property, both physical and locational) of the properties that are sold may differ from one period to another. For example, sales in one quarter could be disproportionately skewed towards low-quality properties, therefore producing a low and biased estimate of the average price. Therefore, it is necessary to 'standardize' this data to ensure that the varying mix of properties sold in each quarter does not give a false impression of the actual change in prices. In order to calculate the pure property price change over time, it is necessary to compare "like" with "like" and this comparison can only be achieved if the quality of properties is on average the same during the comparable periods.

There are a variety of methods that can be used to calculate an average house price and track trends in the property market, these include: Simple mean, repeated sales, and Hedonic price measurements.

4.1 Simple Mean/Median

The simple mean method calculates a simple average price (Mean or Median) of all sale prices observed within the period. With this method, it is assumed that with a very large number of transactions, the mix of properties in the sample will be sufficiently similar over time to give a reasonably accurate gauge of the change in property prices. The advantage of this method is that the calculation is simple and simple aggregate prices can be produced to summarize a large number of transactions on a timely basis.

However, using the mean can be problematic as the distribution of sale prices of properties can be positively skewed by some very high value properties sold and will therefore not give a true reflection of the average price. Though simple median is not affected by outliers to the same extent as the mean, this method ignores the problems of heterogeneity and infrequent sales. In addition, when using this method, no attempt is made to ensure that the sample of house sales used is comparable over time.

4.2 Repeated Sales Method

The repeated sales methodology analyses only residential properties that have been sold more than once. The idea behind this methodology is that the quality of a residential property remains approximately the same over time. If this is the case, then any observed change in sale price must relate to a change in aggregate prices or random “noise”. On the contrary, if the quality changed due to refurbish works or to depreciation, analyzing price change across a large number of properties filters out the “noise”.

The biggest problem with the repeated sales method is the constant quality requirement for properties included in the analysis. In reality, the quality of most properties changes significantly over time. In the cases where properties age and become run down, the methodology may underestimate the change in prices, and when owners improve their property, the repeat sales method may overestimate the change in prices. Also, due to infrequency of sales, another issue is that the repeated sales methodology is subject to transaction bias. Residential properties that are repeatedly sold may not be a representative sample of properties more generally.

4.3 Hedonic Price Measurement

The hedonic price method recognizes that heterogeneous goods can be described by their characteristics. A residential property can be viewed, mathematically speaking, as a collection of characteristics such as (Size, House Type, Housing Market Area, Roofing Material, etc). With enough data points, a regression model can be used to estimate prices for the characteristics that determine the ‘quality’ of the residential property. A hypothetical constant-quality residential property, i.e. one with the same attributes over time, can then be “constructed” and priced. The hedonic methodology does an excellent job of controlling for variations in the quality of residential property. The method’s main strength is that it can be used to estimate values based on actual sales. The hedonic modeling approach also allows many different characteristics to be considered and included in the analysis if they are determined to be significant contributors to house prices.

The main drawback of this approach is that it is data intensive. Large amounts of data must be gathered and manipulated. The method is relatively complex to implement and interpret, requiring a high degree of statistical expertise and the results depend heavily on model specification. The resource required to produce a hedonic model depends on the availability and accessibility of data and the statistical expertise available. In addition, the method assumes that buyers have the opportunity to select the combination of features they prefer, given their income. However, the housing market may be affected by outside influences, like taxes, interest rates, or other factors.

4.4 Justification for Hedonic Price Method (HPM)

UBOS uses the Hedonic price index approach (HPM) to compute its RPPI. HPM is the recommended method internationally, as it is considered to be the best way to deal with issues of quality changes. In addition, the decision to use this model was made following detailed research and comparisons of RPPI calculations internationally as well as recommendations from assessments performed by international experts on RPPIs. Key documents researched include the following: the Handbook on Residential Property Price Indices by Eurostat – 2013 Edition; the Housing Prices Indexes Methodology by the National Institute of Statistics and Economic Studies (INSEE); the Bank for International Settlements (BIS) Property Price publications and databank; the Financial Stability Board (FSB): Second Phase of the G-20 Data Gaps Initiative ; and Silver, 2016 . Collectively, these documents provide a sound and detailed rationale, including the economic theory, for the use of the hedonic model in RPPI calculations.

5. DATA SOURCES AND COLLECTION METHODS

5.1 Data sources

To calculate the RPPI, UBOS collects data on a monthly basis using a data questionnaire (see Annex 2) that is filled using information from a widely used real-estate, website, http://realestatedatabase.net/FindAHouse/Index.aspx#HomePage, and interviews with real estate agencies advertising these properties. The data questionnaire collects information on 23 variables on the characteristics of the property. The quality of the property is measured by variables collected on: i) the plot: size and tenure; ii) the building: status, structure, completion, and nature of the roofing material; iii) Building size, number of bedrooms, and flooring material; and iv) general description of the property: distance to tarmac, annexes and fencing material. As quality metrics change over time, UBOS intends to review this questionnaire on a regular basis to ensure that relevant quality variables are collected and analyzed.

The asking price for a property is used as a proxy for the transaction price of a property sold. To accurately estimate housing values, the most preferred data source for housing price indexes is the transaction price determined when the sale is signed and finalized. According to the BIS[1], most countries that publish RPPIs rely on transaction data, however four exclusively use asking prices – these are Macedonia, Malta, Slovakia and Switzerland. However, in Uganda, getting transaction price data is particularly difficult given the perceived tax implications of revealing this information. Further, for whatever final transaction price information exists, there is no central data repository from which this data could be feasibly collected.

[1] "Residential property price statistics across the globe", by Scatigna M., Szemere R., and Tsatsaronis K., BIS quarterly review, September 2014.

The biggest concern with estimating housing values by using asking price data is the potentially significant difference between the asking price and the agreed selling price. Lyons 2013[2] compares two large datasets with transaction and asking prices from Ireland's property market over the period 2001-2012. The paper finds that the two series are extremely closely correlated, both across space and time, suggesting that asking prices can be a very good proxy for transaction prices. In Uganda, interviews with key stakeholders in the real estate sector revealed that the spread between asking and transaction prices is not particularly large. Further, there is limited speculator activity in the housing market in Uganda, which can have an upward effect on the spreads.

At present, the database of all the information collected by UBOS covers 5 years, from mid-2014 to September-2018. The unit of observation is a single unit of housing and the total number of observations vary depending on available houses. Each observation is assigned a quarter variable (quarter's code) that denotes the quarter in which the data point was collected. The quarter's code is a continuous count from the quarter where data is available and begins from 1 (one). The tables in Annex 2 below provide descriptive statistics of the data.

5.2 Data Sampling

UBOS collects all the advertised data from the web and therefore no sampling method is employed to select the data. A census of the online selected database is taken as an input during the computation every quarter.

6. WEIGHTS SOURCES AND GENERATION

Weights used in the RPPI calculations can be derived from different sources. These are national accounts, national population and housing census data on the housing stock, and Commercial Banks on the loans taken out for house purchase. Other sources are construction sector statistics and official registers recording ownership among others. For this RPPI, the weights were developed from the 2014 National Population and Housing Census (NPHC) results. The details on the stock of housing by type, occupied by the population were analyzed.

The NPHC2014 data gives the type of dwelling unit occupied by a household in the following categories: (i) detached house (ii) semi-detached house (iii) flat in a block of flats (iv) room or rooms of a main house (v) servant quarters (vi) tenement (muzigo) (vii) garage (viii) go down/basement (ix) store (x) other unit. For computation of weights, the following assumptions were made:

- Three types of dwelling units were maintained, these are; detached house, semi-detached house and flat in a block of flats

[2] "Price signals and bid-ask spreads in an illiquid market: The case of residential property in Ireland, 2001-2012", Ronan C. Lyons, draft working paper, prepared for the EEA 2013 Congress. February 2013

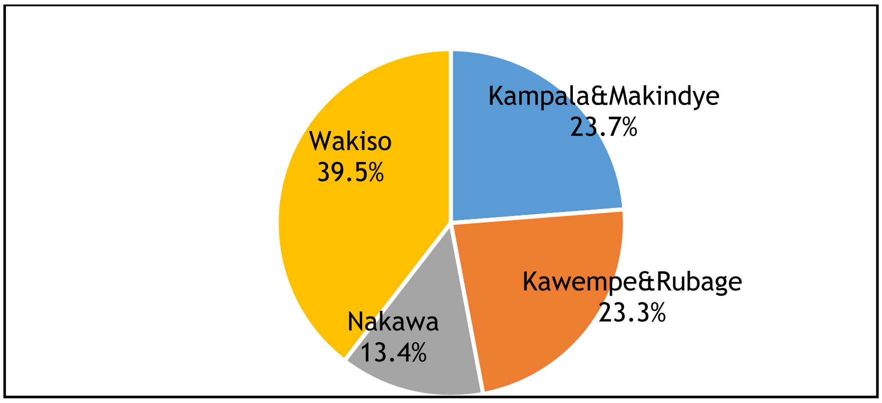
- Servant quarters, garage, go down/basement, store and other units were not included in the compilation of the stock of houses in order to avoid double counting, since these were considered to be part of the main houses indicated above.

Table 1: Stock of properties with corresponding weights for GKMA, 2nd Quarter 2014/15

Region	Kampala & Makindye	Kawempe & Rubaga	Nakawa	Wakiso
No. of Properties				
Detached Houses	25,283	63,620	19,878	165,040
Semi-Detached House	14,972	35,802	13,782	51,745
Condominium & Other Flats	6,271	7,523	3,931	4,562
Average Prices (M-UGX)				
Detached Houses	600	363	464	330
Semi-Detached House	1,338	530	1,069	426
Condominium & Other Flats	2,127	732	895	949
Values of Stock of houses(M-UGX)				
Detached Houses	15,159,714	23,109,173	9,215,422	54,425,199
Semi-Detached House	20,037,564	18,975,060	14,730,880	22,067,252
Condominium & Other Flats	13,335,375	5,510,447	3,517,603	4,331,521
TOTAL	**48,532,653**	**47,594,680**	**27,463,905**	**80,823,972**
Weights	**23.7%**	**23.3%**	**13.4%**	**39.5%**

The total number of units in each category (stock) obtained from the NPHC 2014 was multiplied by the average price of properties for the second Quarter (Q2) of 2014/15 as indicated in Table 1 to obtain the value of the stock. The Q2 is when the NPHC for Uganda was carried out. The weights were then computed using the value of the stock as the percentage of the total value of stock. The results show that Wakiso has a largest share of stock value of houses in GKMA of 39.5%, this was followed by Kampala Central combined with Makindye division at 23.7%. Kawempe combined with Rubaga division followed with 23.3% and lastly Nakawa division combined with Rubaga division at 13.4%, (Table 1 & Figure 1).

Figure 1: Residential Property weights per Geographical Area



7. HEDONIC MODEL AND INDEX CALCULATION

The hedonic method of pricing recognizes that quality goods can be described as a function of their price determining characteristics. The starting point for the modeling is the assumption that the price P_i^t of property i in period t, is a function of a fixed number of characteristic sZ_{ij}^t. The characteristics can be qualitative (such as the type of the property, location of the property etc.) and Quantitative (such as the size of the property, the number of bedrooms etc.). Each property i sold will differ in price due to differences in its characteristics. The characteristics are measured by "quantities" P_{ij}^t Therefore, the function for each property sold P_i^t can be written as:

$$P_i^t = f(Z_{i1}^t, \ldots, Z_{iK}^t, \varepsilon_i^t) \text{ .... (i)}$$

Where,ε_i^t are a group of unmeasured factors (assumed to be randomly distributed) which are specific to each property but for which data are not available.

The functional form used in this report is a log-linear—also referred to as a semi-logarithmic—form of the hedonic regression. This form arises from a hedonic relationship between P_i^t and Z_{ij}^t given by

$$P_i^t = \beta_0^t \left[\beta_1^{Z_{i1}^t}\right]\left[\beta_2^{Z_{i2}^t}\right]\left[\beta_3^{Z_{i3}^t}\right] \ldots\ldots \left[\beta_k^{Z_{ik}^t}\right] \varepsilon_i^t \text{.................... (ii)}$$

The log-linear form first allows for curvature in the relationships say between number of rooms and price, and second, for a multiplicative association between quality characteristics, i.e. that possession of a garage and a boys quarters may be worth more than the sum of the two. The estimation of ordinary least squares regression (OLS) equations requires a linear form; we transform the non-linear functional relationship in equation (ii) into a linear form by taking logarithms of both sides of the equation and use OLS, see Silver (2016)[3] .

$$\ln p_i^t = \ln \beta_0^t + \sum_{k=1}^{K} z_{k,i}^t \ln \beta_k^t + \ln \varepsilon_i^t \qquad \text{.....................(iii)}$$

An estimated OLS regression equation for equation (iii) is given as:

$$\ln \hat{p}_i^t = \ln \beta_0^t + \sum_{k=1}^{K} z_{k,i}^t \ln \beta_k^t \qquad \text{..................................(vi)}$$

where ($\hat{P}_i^t$ and P_i^t) are the predicted (and actual) price of property in period t; $Z_{k,i}^t$ are the values of each k=1,….,K price-determining characteristic for property i in period t; $\hat{\beta}_0^t$and β_k^t are the estimated coefficients for each characteristic Z_k^t; ε_i^t are independent and identically distributed (i.i.d) errors, using period t data and characteristics.

It is important to note that the log-linear regression output from estimating equation (iii), that is $\ln(P_i^t)$ on z_{ik}^t provides us with the logarithms of the coefficients from the original log-linear formulation in equation (ii). As indicated by Silver (2016), exponents of the estimated coefficients from the output of the software have to be taken if the parameters of the original function, that is equation (ii), are to be recovered, that is exp(ln$\hat{\beta}_k^t$) = $\hat{\beta}_k^t$

Since many explanatory variables are dummy variables taking a value of zero or one—possession or otherwise of a characteristic—and since logarithms cannot be taken of zero values, the log-linear form is more convenient than a double-logarithmic transformation that would require logarithms be taken of the z_{ik}^ton the RHS. It should be noted that the interpretation of coefficients from a log-linear form differs from that of coefficients from a linear form. For a log-linear form our estimated coefficients are the logarithms of $\hat{\beta}_{1,}\ \hat{\beta}_{2,\ldots\ldots\ldots,}\ \hat{\beta}_k$. Thus, a unit change in the say square footage, z_{1i}, leads to a $\hat{\beta}_1$ percent change in price, while for a dummy explanatory variable, say "possession of a boys quarters, z_{2i} = 1, as opposed to z_{2i} = 0 otherwise" leads to an estimated $[\exp(\hat{\beta}_2\ ) - 1] * 100$ percent change in price as explained by Silver 2016.

7.1 Best base models generated for different regions in GKMA

When using hedonic regression techniques to adjust for quality changes, stratification is highly recommended. It is very unlikely that a single hedonic model holds true for all market segments, hence separate regressions should be run for different types of properties, different locations, etc., (Page 56, Para 5.32,

[3] Silver, Mick 2016. How to better measure hedonic residential property price indexes, IMF Working Paper WP/16/213, Washington DC, April 2016.

Handbook on RPPI). Stratification entails dividing the pool of transaction data into the desired subsets and running the hedonic regression method on each subset independently. The choice of strata is very much limited by the availability of transaction data. Too few transactions lead to very volatile or even unviable price indexes.

For this RPPI there are four separate strata, each producing their own elementary price indexes. These include, Wakiso, Nakawa, Kampala Central & Makindye, and Rubaga & Kawempe. A headline aggregate index is then created by combining the elementary price indexes. Therefore, UBOS calculates different regressions for different regions or strata. Once the best base models are determined, they are used to calculate the price indexes for each region or stratum as shown in Annex 2.

Two possible methodologies that can be used to compile the quality adjusted price indexes for each region: (i) the Time-Dummy Variable approach and (ii) the Imputation approach.

7.2 Time-Dummy Variable Approach

A single hedonic regression equation is estimated with observations across properties transacted over several time periods, including the reference period 0 and successive subsequent periods t. The (logarithm of) prices of individual properties are regressed on their characteristics and dummy variables δ^1 for different periods taking the values of 1if the house is sold in the period and zero otherwise. A log-linear specification is given by:

$$\ln \hat{p}_i^{0,t} = \hat{\beta}_0 + \sum_{k=1}^{K} z_{k,i}^{0,t} \ln \hat{\beta}_k + \sum_{t=1}^{T} \hat{\delta}^t D_i^t \quad \text{............... (vi)}$$

The $\hat{\delta}^t$ are estimates of the proportionate change in price arising from a change between the reference period t=0 (the period not specified as a dummy time variable) and successive periods t=1, having controlled for changes in the quality characteristics via the term,

$$\sum_{k=1}^{K} z_{k,i}^{0,t} \ln \hat{\beta}_k$$

The index derives from the time dummy parameters as follows,

$$I = 100 \exp(\hat{\delta}^t) \text{............................ (vii)}$$

I in equation (iv) requires an adjustment for it to be a consistent (and almost unbiased) approximation of the proportionate impact of the time dummy. The method implicitly restricts the coefficients on the quality characteristics to be constant over time: for example, for an adjacent period 0 and 1 regression, for K= 1,…,K, $\beta_k = \beta_k^0 = \beta_k^1$ and $100 \times \exp\left(\hat{\delta}^1\right)$ is an estimate of the RPPI for period 1 (period 0=100).

The extent of this restriction depends on the length of the time period over which the regression is run. If, for example, the regressions are run over quarterly data for a 10-year window, a property price comparison between say 2008Q1 and 2018Q1 with valuations of characteristics held constant may stretch credibility, though this can be alleviated by chained shorter and/or moving windows or adjacent period regressions[4].

7.3 Imputation Approach

The imputation approach works at the level of individual properties, rather than the average values of their characteristics. The rational for the imputation approach lies in the matched model method. Consider a set of properties transacted in period 0. We want to compare their period 0 prices with the prices of the same properties in period t. In this way there is no contamination of the measure of price change by changes in the quality-mix of properties transacted. However, the period 0 properties were not sold in period t (there is no corresponding period t price). The solution is to estimate the period t prices for the characteristics and then to calculate the estimated prices of each period 0 property. We use a period 0 regression to predict prices of properties sold in period t to answer the counterfactual question: what would a property with period t characteristics have sold at in period 0? A constant-quality Hedonic Geometric Mean Imputation (HGMI) price index from a log linear hedonic regression equation is a ratio of geometric means with characteristics held constant in the current period t, $\bar{z}_k^t$:

$$P_{HGMI:z_i^t}^{0\to t} = \frac{\prod_{i\in N^t}\left(\hat{p}_{i|z_i^t}^t\right)^{\frac{1}{N^t}}}{\prod_{i\in N^t}\left(\hat{p}_{i|z_i^t}^0\right)^{\frac{1}{N^t}}} = \frac{\exp\left(1/N^t \sum_{i\in N^t}\ln\hat{p}_{i|z_i^t}^t\right)}{\exp\left(1/N^t \sum_{i\in N^t}\ln\hat{p}_{i|z_i^t}^0\right)} \quad \text{.....................................(viii)}$$

From equation (viii), the value in the numerator of the above equation is the geometric mean of prices in period t with price-determining characteristics, $z_{i,k}^t$. This is compared, in the denominator, with the geometric mean of the period 0 predicted price of the same period t price-determining characteristics, $z_{i,k}^t$.

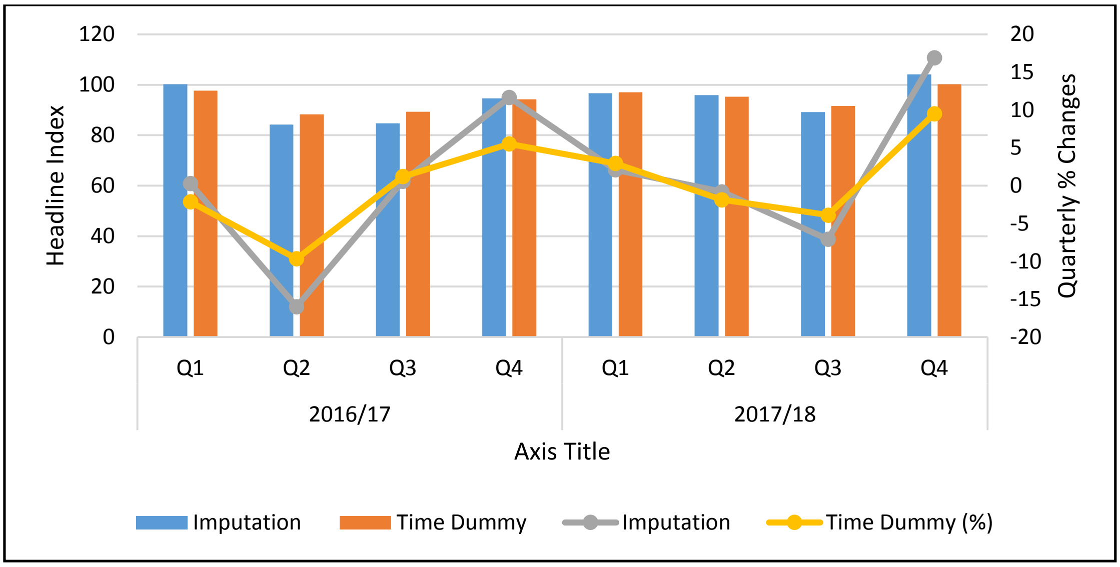
7.4 Index Computation Methodology Selection

The two methodologies described above have their advantages and disadvantages, *see Silver 2016*. UBOS decided to calculate the index for each region or stratum using the imputation approach especially as international practices indicate that

[4]Silver, Mick 2016. How to better measure hedonic residential property price indexes, IMF Working Paper WP/16/213, Washington DC, April.

the imputation method is generally preferred, particularly if the sample size is large enough. A Time Dummy Computation was also done and below is a comparison between the two methods.

Figure 2: Comparison of the Index using the Time Dummy and the Imputation method.



From figure 2, the data shows that Time Dummy Method is less volatile than the Imputation Method. However, the imputation method will be used. This method was selected for three main reasons: It is among the most sophisticated; is the simpler to explain to users; and easy to calculate. Its similarity with the rational of price index calculation makes it easy to understand. Its calculation is simpler than other methods such as the characteristics hedonic method, and since we are not using geospatial data it will provide the same results.

7.5 Aggregation of regional indexes to calculate the headline GKMA index

The GKMA index is computed from the indices at levels below the Elementary Aggregate (EA) indices. For the RPPI, the EA level is the selected regional index. Higher level indices are calculated as a weighted arithmetic means of the indices from all regions with in GKMA. The construction of these indices is based on a Laspeyres-type Fixed Base Weighted Index. The index at the EA level is calculated using Jevons Index (from equation v) as the geometric mean of the prices in the current period divided by the geometric mean of prices in the base period.

At GKMA Level, Laspeyres Price index is used to generate the headline index for the entire Kampala Metropolitan Area.

7.6 Treatment of the data

Data validation is done by checks to ensure that all inputs to the RPPI calculations are correct. Data is edited by adjusting or eliminating erroneous inputs. Once the required prices and corresponding characteristics have been surveyed, they are carefully examined for accuracy and validity before they are used for the RPPI calculations. This involves comparisons of individual house prices within and between the selected regions in GKMA. In addition, a comparison of prices is carried out between the current and the previous pricing period for the same region. Where outliers occur, prices are validated by contacting the outlet concerned and where necessary editing is done.

Data validation on prices is done by first setting up outlier boundaries. For all those prices that were identified as outliers5 or extreme outliers6, verifications are made to ensure that there are no mistakes during the data entry. Data entry mistakes that were found are corrected. In some cases, the other common errors that are causing the extreme outliers are due to poor recording or entering wrong unit of measurements. For instance, instead of reporting a price of a house in United States Dollars (USD) it is recorded in Uganda Shillings (UGX). Cases where the outliers are not due to data entry or unit of measurements errors, UBOS staff are sent back to the outlets to confirm the reliability of such prices and to make sure that the prices refer to exactly the same house with all its price determining characteristics. Through that process, some outlier and extreme outlier prices are confirmed as true prices and are left into the RPPI computations while those that are proved to be wrong prices are deleted.

When data is cleaned and ready to begin regression diagnostics two hedonic regressions are run for each elementary index of each reference quarter. The first or preliminary regression is run to identify outliers, transactions with unusual or extreme prices that exert undue influence. In the preliminary regression, Cooks Distance is computed for each transaction. Cooks Distance measures the leverage of each transaction on the overall regression fit (i.e. how influential a particular transaction is in the determination of the β explanatory price coefficients). Higher leverage is associated with extreme values (i.e. transactions where the house price appears exceptionally high or exceptionally low for its particular set of known characteristics). Transactions where the Cooks Distance exceeds (4/n), the conventional cut-off (where n is the number of transactions in the data pool), are considered outliers. These outliers are then examined further and are either excluded from the data pool or left and the regression is re-run.

8. RESULTS AND MODEL DIAGNOSTICS

8.1 Interpreting the index

The index numbers are interpreted following the two changes that are mostly calculated and published. These are;

[5] Outlier means a price or price change which falls outside a specified “outlier” boundary
[6] Extreme outlier refers to an outlier price or price change which falls outside a specified “extreme outlier” boundary

- The change in the RPPI between the current Quarter and the previous Quarter, referred to as Quarterly price change or the Quarterly Inflation.
- The change in the RPPI between the current Quarter and the same Quarter of the previous year is known as the annual price change or the Inflation.

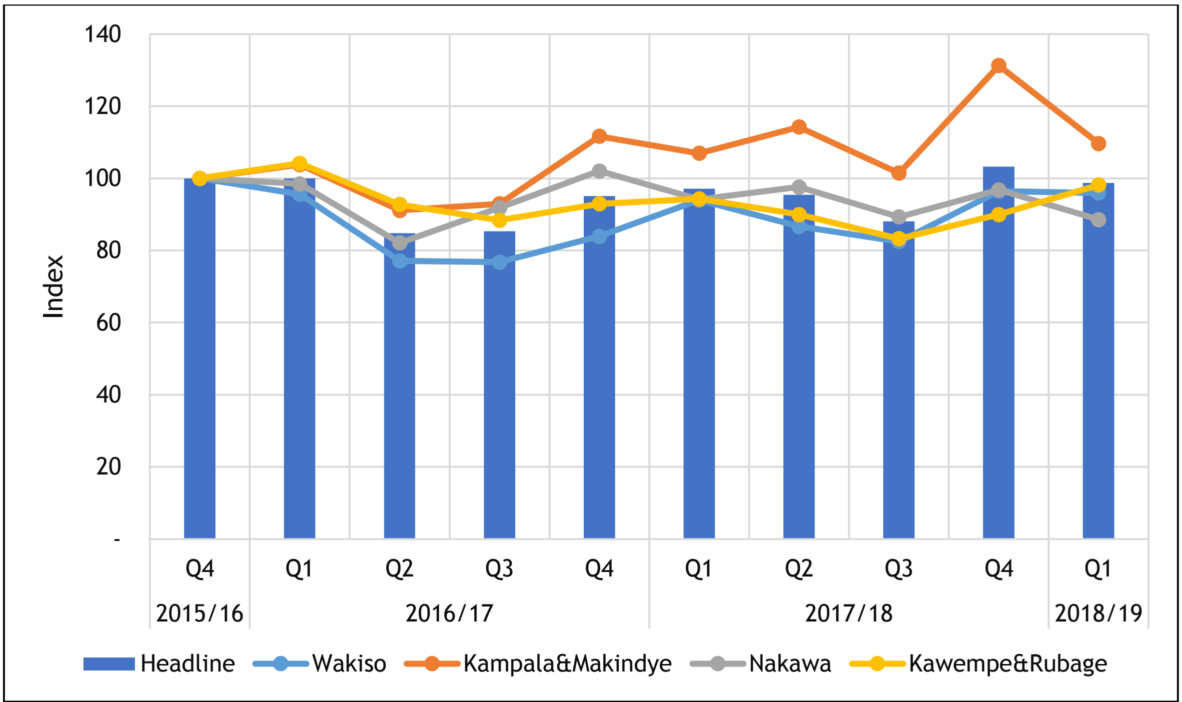
8.2 Results

Using the methodology detailed above, UBOS calculates the headline index for GKMA and the price indexes for the different regions or strata that comprise GKMA. The base period used is all quarters in Financial Years 2014/15 and 2015/16 so these quarters are set to 100. Table 3 and Figure 3 present the calculated RPPIs including the headline index and the strata indexes. Annex 1 provides detailed regression results for each stratum and the headline index.

Table 3: Headline Index and strata indexes for GKMA, Q1 2016/17 – Q1 2018/19

Financial Year		2016/17				2017/18				2018/19
Quarter	Weights	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
Wakiso	39.50	95.60	77.11	76.72	83.94	93.94	86.67	82.48	96.44	95.96
Kampala & Makindye	25.80	103.80	91.12	92.93	111.67	106.94	114.21	101.45	131.28	109.65
Nakawa	14.40	98.49	82.12	91.81	102.05	94.15	97.60	89.27	96.78	88.54
Kawempe &Rubaga	13.90	104.16	92.73	88.38	92.91	94.30	89.95	83.27	89.96	98.15
Headline	100.00	99.93	84.75	85.31	95.04	97.14	95.44	88.08	103.25	98.72
Quarterly % Changes		-0.1	-15.2	0.7	11.4	2.2	-1.7	-7.7	17.2	-4.4
Annual % Change					-5.0	-2.8	12.6	3.2	8.6	1.6

Figure 3: Headline RPPI and Strata Indexes, Q1 2016/17 – Q12018/19



The headline index shows that residential real estate prices decreased on average in Q2 2016/2017 and only recovered fully in Q4 2017/2018, where prices were finally higher on average than those in Q1 2016/2017. Results indicate that, on annual basis, the prices of properties in GKMA have been increasing since the second quarter of 2017/18 when they grew by 12.6%. This was followed by a growth of 8.6% in the fourth quarter of 2017/18 compared to 3.2% rise recorded in the third quarter. There was a decline in house prices inflation for the first quarter 2018/19 that registered a 1.6 percent, see table 3.

The results by region show that prices in Wakiso, Nakawa, and Kawempe and Rugaba followed a similar pattern to the headline index. However, although prices have shown some recovery, they have yet to surpass those in Q1 2016/2017. Prices in Nakawa seemed to recover rapidly in Q4 2016/2017, surpassing prices in Q1 2016/2017, only to drop again to levels below Q1 2016/2017. On the other hand, prices in Kampala and Makindye recovered fully in Q4 2016/2017 and have generally continued to increase higher than prices in Q1 2016/2017, see figure 3.

The model results also indicate a good fit of the data, R-squares for all the models are generally good enough implying that the selected price-determining characteristics used in the models such as; status of building (whether the house is new or used), Structure (dwelling type), Number of levels, Number of bedrooms, Nature of the floor, Nature of the roof, Nature of the fence and other house possessions, are sufficient to explain the price variation observed in the data. Table 4 indicates the total number of observations in the base period regressions and their corresponding R- squared for different strata, see table 4.

Table 4: Total Number of observations and corresponding R-squared for different Strata.

	Strata	No. of Observations of the base period	R- squared (%)
1	Kampala - Makindye	1097	71.48
2	Nakawa	1609	64.73
3	Wakiso	2483	66.96
4	Kawempe - Rubaga	344	89.15

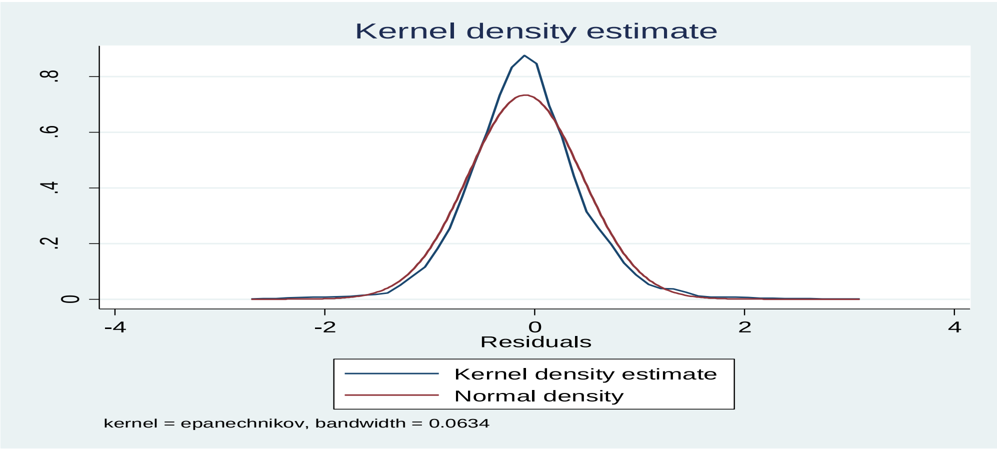
The sections that follow present various tests performed on the final models to ensure that they are robust, that assumptions made are appropriate, and that they correct for potential biases. The results of these tests validate the model specification and assumptions and ensure that the ensuing results do not suffer from potential biases.

8.3 Normality Tests

Normality of residuals is only required for valid hypothesis testing, that is, the normality assumption assures that the p-values for the t-tests and F-test will be valid. Normality is not required in order to obtain unbiased estimates of the regression coefficients. OLS regression merely requires that the residuals (errors) be identically and independently distributed. Furthermore, there is no assumption or requirement that the predictor variables be normally distributed. If this were the case then we would not be able to use dummy coded variables in our models. It should be noted that most of our data obtained is qualitative and therefore most of our variables are dummies.

However, UBOS tries to test for the normality in the data. After running the regression, the residues were predicted using the predict r command in stata to create residuals (r) and then use commands such as kdensity to produce the kernel density plot with a normal option requesting that a normal density be overlaid on the plot.

For Wakiso, below is the data behavior after running the kdensity command.

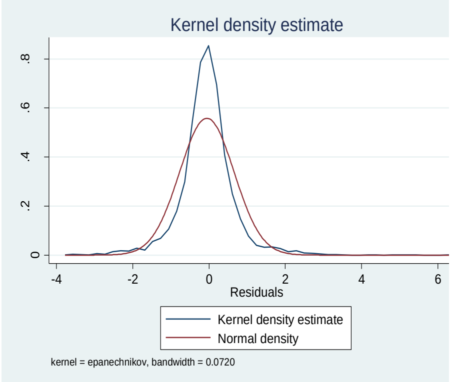


Another test available is the swilk test which performs the Shapiro-Wilk W test for normality. The p-value is based on the assumption that the distribution is normal. In our Wakiso example, it is very small (0.00), indicating that we do not accept that r is normally distributed.

```
. swilk r

                   Shapiro-Wilk W test for normal data

    Variable |       Obs       W           V         z       Prob>z
-------------+------------------------------------------------------
           r |     13412    0.98014     127.131    13.070    0.00000
```

The test result indicates that the residuals are not normally distributed, but from the kdensity test, it is clear that there are no severe outliers and the distribution seems fairly symmetric. The residuals have an approximately normal distribution. The table below shows how other regions behavior on normality.

1 Nakawa

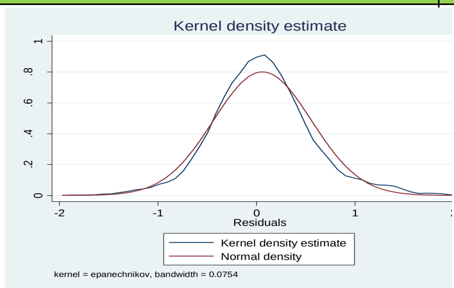



```
. swilk r

                   Shapiro-Wilk W test for normal data
```

Variable	Obs	W	V	z	Prob>z
r	6838	0.92031	284.892	14.971	0.00000

From Shapiro – W normality test, result of Nakawa indicates that the residuals are not normally distributed, but from the kdensity test, it is clear that we don't have any severe outliers and the distribution seems fairly symmetric. The residuals have an approximately normal distribution.

2 Kampala - Makindye

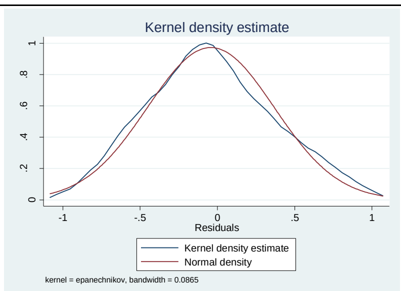



```
. swilk r

                   Shapiro-Wilk W test for normal data
```

Variable	Obs	W	V	z	Prob>z
r	4396	0.98771	29.692	8.861	0.00000

From Shapiro – W normality test, result of Kampala-Makindye indicates that the residuals are not normally distributed, but from the kdensity test, it is clear that we don't have any severe outliers and the distribution seems fairly symmetric. The residuals have an approximately normal distribution.

3 Kawempe - Lubaga



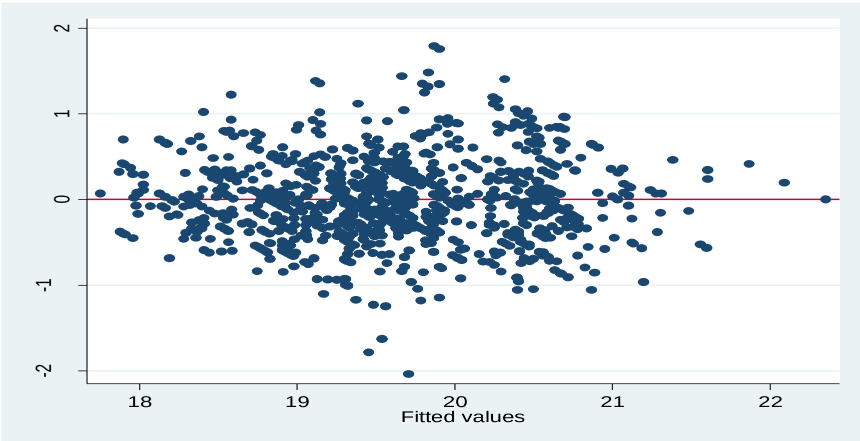

```
. swilk r

                   Shapiro-Wilk W test for normal data
```

Variable	Obs	W	V	z	Prob>z
r	1611	0.92529	72.769	10.818	0.00000

From Shapiro – W normality test, result of Kawempe–Lubaga indicates that the residuals are not normally distributed, but from the kdensity test, it is clear that we don't have any severe outliers and the distribution seems fairly symmetric. The residuals have an approximately normal distribution.

8.4 Checking Homoscedasticity of Residuals

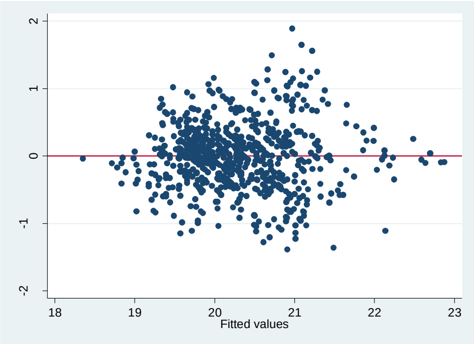
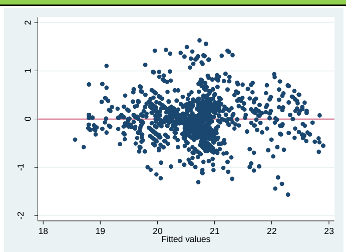
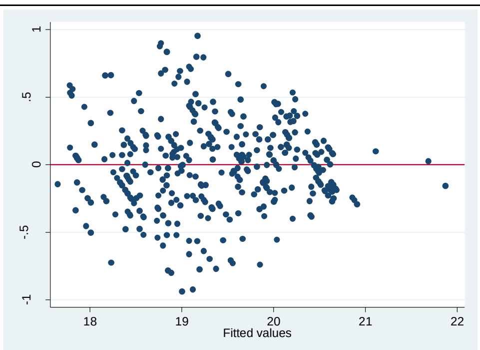
One of the main assumptions for the ordinary least squares regression is the homogeneity of variance of the residuals. If the model is well-fitted, there should be no pattern to the residuals plotted against the fitted values. If the variance of the residuals is non-constant then the residual variance is said to be "heteroscedastic." There are graphical and non-graphical methods for detecting heteroscedasticity. A commonly used graphical method is to plot the residuals versus fitted (predicted) values. We do this by issuing the *rvfplo*t command in stata. Below we use the *rvfplot* command with the *yline*(*0*) option to put a reference line at y=0 for Wakiso. We see that the pattern of the data points not following any pattern, which is an indication of no heteroscedasticity.



The non-graphical method used in stata was the hettest, which is better known as the Breusch-Pagan test. The test's null hypothesis is that the variance of the residuals is homogenous. Therefore, if the p-value is very small< (0.05), we would have to reject the hypothesis and accept the alternative hypothesis that the variance is not homogenous.

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of logprice

         chi2(1)      =     9.64
         Prob > chi2  =   0.0019
```

Thus, in this case, the evidence is against the null hypothesis that the variance is homogeneous (p=0.0019). It should be noted that these tests are very sensitive to model assumptions, such as the assumption of normality. Therefore, it is a common practice to combine the tests with diagnostic plots to make a judgment on the severity of the heteroscedasticity and to decide if any correction is needed for heteroscedasticity. In our case, the plot above does not show too strong evidence. But we can go ahead and correct for heteroscedasticity since there are methods available in STATA to correct it like *vce(robust)* command after a regression, which was done to come up with the final model. The table below shows how other region's diagnostics for the selected model behaved.

1	Nakawa	
		Breusch-Pagan / Cook-Weisberg test for heteroskedasticity Ho: Constant variance Variables: fitted values of logprice chi2(1) = 46.79 Prob > chi2 = 0.0000
	From the Breusch-Pagan test, null hypothesis is that the variance of the residuals is homogenous. So, in this case, the evidence is against the null hypothesis that the variance is homogeneous, since p is very small (0.000). But from the plot, the severity of the heteroscedasticity does not show too strong evidence. A correction was made by using vce (robust) command while the regression is ran.	
2	Kampala - Makindye	
		. hettest Breusch-Pagan / Cook-Weisberg test for heteroskedasticity Ho: Constant variance Variables: fitted values of logprice chi2(1) = 14.18 Prob > chi2 = 0.0002
	From the Breusch-Pagan test, null hypothesis is that the variance of the residuals is homogenous. So, in this case, the evidence is against the null hypothesis that the variance is homogeneous, since p is very small (0.0002). But from the plot, the severity of the heteroscedasticity does not show too strong evidence. A correction was made using vce(robust) command while the regression is ran.	
3	Kawempe – Lubaga	
		Breusch-Pagan / Cook-Weisberg test for heteroskedasticity Ho: Constant variance Variables: fitted values of logprice chi2(1) = 12.39 Prob > chi2 = 0.0004
	From the Breusch-Pagan test, null hypothesis is that the variance of the residuals is homogenous. So, in this case, the evidence is against the null hypothesis that the variance is homogeneous, since p is very small (0.0004). But from the plot, the severity of the heteroscedasticity does not show too strong evidence. A correction was made using vce(robust) command while the regression is ran.	

8.5 Checking Multicollinearity

When there is a perfect linear relationship among the predictors, the estimates for a regression model cannot be uniquely computed. The term collinearity implies that two variables are near perfect linear combinations of one another. When more than two variables are involved it is often called multicollinearity. The primary concern is that as the degree of multicollinearity increases, the regression model estimates of the coefficients become unstable and the standard errors for the coefficients can get wildly inflated. We can use the VIF (variance inflation factor) command after the regression to check for multicollinearity. As a rule of thumb, a variable whose VIF values are greater than 10 may merit further investigation. This was done and found that most of our variables had VIFs less than 10 apart from interactions which can be explained.

8.6 Conclusion

The RPPI is of significant interest to policy makers, market analysts and researchers. Given the importance of the housing market in the Ugandan Economy, changes in housing costs affect the disposable income of households, thereby influence consumer spending and saving decisions. The RPPI is therefore a key indicator for following:

i) as an input into Financial Stability models used by BOU to monitor the soundness of the financial sector;

ii) as a macroeconomic indicator of residential property price inflation that can be used in supporting the compilation of the non-financial assets component of the Household Balance Sheet in the Ugandan System of National Accounts (SNA);

iii) as an analytical tool used by Housing economists to estimate changes in the rates of mortgage defaults, prepayments and housing affordability in specific geographic areas.

The Uganda RPPI was computed for Greater Kampala Metropolitan Area (GKMA) and covers available houses that are up for sale in the market. GKMA covers areas of Kampala Central, Makindye, Nakawa, Rubaga, Wakiso, and Kawempe. UBOS collects data on a monthly basis using a data questionnaire filled using information from a widely used real-estate website, and also interviews with real estate agencies advertising these properties.

A constant-quality Hedonic Geometric Mean Imputation (HGMI) price index from a log linear hedonic regression equation is a ratio of geometric means with characteristics held constant in the current period is used by UBOS to calculate the indexes, international practices indicate that the imputation method is generally preferred, particularly if the sample size is large enough.

To produce separate price indices for subsets of the housing market, stratification is required. For this RPPI there are four separate strata, each producing their own elementary price indices. These include, Wakiso, Nakawa, Kampala Central & Makindye, and Rubaga & Kawempe. Aggregate indexes are created from combining the elementary price indexes into a headline.

The results for the headline index show that residential real estate prices decreased on average in Q2 2016/2017 and only recovered fully in Q4 2017/2018, where prices were finally higher on average higher than those in Q1 2016/2017. On an annual basis, the prices of properties in GKMA have been increasing since the second quarter of 2017/18. The results by region show that prices in Wakiso, Nakawa, and Kawempe and Rugaba followed a similar pattern to the headline index. On the other hand, prices in Kampala and Makindye recovered fully in Q4 2016/2017 and have generally continued to increase higher than prices in Q1 2016/2017. These findings also corroborate the trends described by key stakeholders in the real estate sector.

The model results also indicate a good fit of the data, R-squares for all the models are generally good enough implying that the selected price-determining characteristics used in the models are sufficient to explain the price variation observed in the data. Further, the residuals for all the models have an approximately normal distribution. Following other diagnostics, the models were improved by correcting for heteroscedasticity while multicollinearity was not an issue affecting the variables. Therefore, the model specification and assumptions are appropriate and do not bias the calculated results. In all, this exercise shows that the RPPI calculated by UBOS provides robust estimates of average prices and the trends of housing prices in GKMA.

ANNEX 1: DIFFERENT MODELS GENERATED FOR DIFFERENT STRATA

Annex 1.1: Generated Base Model for Nakawa

logprice	Coef.	Robust Std. Err.	t	P>t	[95% Conf.	Interval]
quartercode						
7	-0.1380	0.0476	-2.9000	0.0040	-0.2313	-0.0446
8	-0.0646	0.0424	-1.5200	0.1280	-0.1479	0.0186
9	-0.0549	0.0422	-1.3000	0.1930	-0.1377	0.0279
10	0.0114	0.0456	0.2500	0.8020	-0.0780	0.1009
11	-0.0275	0.0544	-0.5100	0.6130	-0.1343	0.0792
12	-0.0709	0.0503	-1.4100	0.1590	-0.1696	0.0278
13	-0.0574	0.0445	-1.2900	0.1970	-0.1446	0.0298
Parishcode						
2	-1.5114	0.3357	-4.5000	0.0000	-2.1699	-0.8529
3	-0.5550	0.3659	-1.5200	0.1300	-1.2728	0.1628
4	-2.1427	0.2612	-8.2000	0.0000	-2.6550	-1.6305
5	-1.7028	0.2722	-6.2600	0.0000	-2.2368	-1.1689
6	-1.6636	0.3587	-4.6400	0.0000	-2.3672	-0.9600
7	-0.6504	0.2004	-3.2500	0.0010	-1.0434	-0.2573
8	-4.0173	1.2514	-3.2100	0.0010	-6.4719	-1.5627
9	-1.6897	0.3021	-5.5900	0.0000	-2.2823	-1.0972
10	-1.8216	0.2423	-7.5200	0.0000	-2.2969	-1.3462
11	-1.5613	0.2663	-5.8600	0.0000	-2.0837	-1.0389
statusofbuilding	0.0861	0.0303	2.8400	0.0050	0.0266	0.1456
structure						
2	0.8803	0.2013	4.3700	0.0000	0.4856	1.2751
3	2.0114	0.2506	8.0300	0.0000	1.5198	2.5031
4	0.8169	0.1949	4.1900	0.0000	0.4347	1.1991
numberoffloors	0.2122	0.0421	5.0400	0.0000	0.1297	0.2947
natureofroofingmaterial	0.4375	0.0475	9.2100	0.0000	0.3443	0.5306
numberofbedrooms	-0.0627	0.0554	-1.1300	0.2580	-0.1714	0.0459
natureoffloorfinish	0.2110	0.0663	3.1800	0.0010	0.0810	0.3410
natureoffencematerial	-0.0249	0.0954	-0.2600	0.7940	-0.2119	0.1622
structurepossession						
2	-0.1102	0.0608	-1.8100	0.0700	-0.2295	0.0092
3	0.0478	0.0395	1.2100	0.2270	-0.0298	0.1254
4	0.0341	0.0425	0.8000	0.4220	-0.0492	0.1174
structure#c.numberofbedrooms						
2	-0.2387	0.0537	-4.4500	0.0000	-0.3440	-0.1335
3	-0.4891	0.0762	-6.4200	0.0000	-0.6385	-0.3397
4	-0.1627	0.0364	-4.4700	0.0000	-0.2341	-0.0913
parishcode#c.numberofbedrooms						
2	0.5672	0.0839	6.7600	0.0000	0.4026	0.7318
3	0.2470	0.0685	3.6000	0.0000	0.1126	0.3813
4	0.3704	0.0664	5.5800	0.0000	0.2402	0.5007
5	0.2559	0.0689	3.7100	0.0000	0.1207	0.3912

logprice	Coef.	Robust Std. Err.	t	P>t	[95% Conf.	Interval]
6	0.5799	0.0780	7.4300	0.0000	0.4268	0.7329
7	0.0000		(omitted)			
8	0.7193	0.1861	3.8700	0.0000	0.3543	1.0843
9	0.5158	0.0690	7.4700	0.0000	0.3804	0.6512
10	0.2757	0.0599	4.6000	0.0000	0.1582	0.3931
11	0.3047	0.0664	4.5900	0.0000	0.1745	0.4349
structure#countycode						
2 2	-0.1279	0.1662	-0.7700	0.4420	-0.4539	0.1981
2 3	1.0267	0.2367	4.3400	0.0000	0.5623	1.4910
2 4	0.4469	0.1380	3.2400	0.0010	0.1762	0.7176
2 5	0.3685	0.2028	1.8200	0.0690	-0.0293	0.7663
2 6	-1.3007	0.2176	-5.9800	0.0000	-1.7275	-0.8738
2 7	1.0487	0.2081	5.0400	0.0000	0.6405	1.4568
2 9	0.7805	0.1970	3.9600	0.0000	0.3942	1.1669
2 10	0.1176	0.1515	0.7800	0.4370	-0.1795	0.4147
2 11	0.2875	0.1669	1.7200	0.0850	-0.0399	0.6149
3 2	-1.4944	0.1502	-9.9500	0.0000	-1.7890	-1.1999
3 3	0.1217	0.3401	0.3600	0.7200	-0.5454	0.7888
3 4	0.6762	0.1323	5.1100	0.0000	0.4166	0.9358
3 5	0.1816	0.1790	1.0100	0.3110	-0.1696	0.5327
3 6	-1.2115	0.2368	-5.1200	0.0000	-1.6759	-0.7471
3 9	-1.0335	0.1579	-6.5500	0.0000	-1.3432	-0.7239
3 10	0.9609	0.1754	5.4800	0.0000	0.6169	1.3049
3 11	0.4158	0.1469	2.8300	0.0050	0.1277	0.7039
4 2	-0.6577	0.2519	-2.6100	0.0090	-1.1519	-0.1636
4 3	-0.7533	0.2253	-3.3400	0.0010	-1.1952	-0.3113
4 4	0.0138	0.1445	0.1000	0.9240	-0.2697	0.2972
4 5	0.0273	0.1322	0.2100	0.8360	-0.2320	0.2867
4 6	-1.2803	0.2124	-6.0300	0.0000	-1.6969	-0.8636
4 9	-0.2870	0.1798	-1.6000	0.1110	-0.6396	0.0656
4 10	0.2361	0.1441	1.6400	0.1020	-0.0466	0.5189
4 11	0.0606	0.1317	0.4600	0.6450	-0.1976	0.3188
_cons	**19.9853**	**0.2486**	**80.3900**	**0.0000**	**19.4976**	**20.4729**

Annex 1.2: Generated Base Model for Kampala and Makindye

logprice	Robust Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
quartercode						
7	0.0110	0.0609	0.1800	0.8570	-0.1086	0.1306
8	-0.0499	0.0537	-0.9300	0.3520	-0.1552	0.0554
9	-0.0047	0.0507	-0.0900	0.9260	-0.1041	0.0947
10	0.0563	0.0563	1.0000	0.3170	-0.0541	0.1666
11	0.1299	0.0633	2.0500	0.0400	0.0058	0.2541
12	0.2343	0.0625	3.7500	0.0000	0.1116	0.3570
13	0.1612	0.0540	2.9800	0.0030	0.0552	0.2672
Parishcode						
2	0.0250	0.2162	0.1200	0.9080	-0.3993	0.4492
3	0.2356	0.5063	0.4700	0.6420	-0.7579	1.2292
4	0.0943	0.3294	0.2900	0.7750	-0.5522	0.7407
5	-1.0357	0.2909	-3.5600	0.0000	-1.6065	-0.4649
6	0.0253	0.2140	0.1200	0.9060	-0.3947	0.4453
7	-0.3547	0.1667	-2.1300	0.0340	-0.6817	-0.0277
8	1.0619	0.4283	2.4800	0.0130	0.2215	1.9023
9	0.6925	0.2372	2.9200	0.0040	0.2269	1.1580
10	1.0337	0.2820	3.6700	0.0000	0.4803	1.5872
11	-0.0606	0.2601	-0.2300	0.8160	-0.5710	0.4497
12	-0.1093	0.5268	-0.2100	0.8360	-1.1430	0.9245
statusofbuilding	0.1320	0.0382	3.4500	0.0010	0.0570	0.2069
structure						
2	0.5520	0.2951	1.8700	0.0620	-0.0271	1.1312
3	1.1526	0.4031	2.8600	0.0040	0.3616	1.9437
4	0.7912	0.1835	4.3100	0.0000	0.4312	1.1513
numberoffloors	0.4337	0.0527	8.2300	0.0000	0.3303	0.5370
natureofroofingmaterial	0.3071	0.0562	5.4600	0.0000	0.1967	0.4174
numberofbedrooms	0.2370	0.0420	5.6500	0.0000	0.1547	0.3194
natureoffloorfinish	0.3566	0.0609	5.8600	0.0000	0.2371	0.4761
natureoffencematerial	0.0260	0.0874	0.3000	0.7660	-0.1454	0.1975
structurepossession						
2	-0.1090	0.0672	-1.6200	0.1050	-0.2410	0.0229
3	-0.0640	0.0436	-1.4700	0.1430	-0.1496	0.0216
4	-0.0082	0.0436	-0.1900	0.8510	-0.0938	0.0773
structure#c.numberofbedrooms						
2	-0.0153	0.0754	-0.2000	0.8390	-0.1633	0.1327
3	0.0950	0.1105	0.8600	0.3900	-0.1219	0.3119
4	-0.2309	0.0421	-5.4900	0.0000	-0.3134	-0.1484
Parishcode#c.numberofbedrooms						
2	0.0371	0.0472	0.7800	0.4330	-0.0556	0.1298
3	-0.0391	0.1470	-0.2700	0.7900	-0.3276	0.2494
4	-0.1148	0.0976	-1.1800	0.2400	-0.3063	0.0767
5	0.1649	0.0700	2.3500	0.0190	0.0275	0.3023
6	-0.0668	0.0462	-1.4500	0.1480	-0.1575	0.0238

logprice	Robust Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
7	0.0347	0.0436	0.7900	0.4270	-0.0509	0.1202
8	-0.3819	0.1123	-3.4000	0.0010	-0.6023	-0.1615
9	-0.1322	0.0680	-1.9400	0.0520	-0.2656	0.0012
10	0.1551	0.0555	2.8000	0.0050	0.0463	0.2639
11	-0.0180	0.0547	-0.3300	0.7420	-0.1253	0.0893
12	0.1985	0.0808	2.4600	0.0140	0.0399	0.3570
structure#countycode						
2 2	0.6332	0.2355	2.6900	0.0070	0.1710	1.0955
2 4	-0.6310	0.2267	-2.7800	0.0050	-1.0758	-0.1862
2 7	0.0408	0.2015	0.2000	0.8400	-0.3547	0.4363
2 9	-0.1602	0.2461	-0.6500	0.5150	-0.6432	0.3228
2 10	-1.8850	0.2710	-6.9600	0.0000	-2.4169	-1.3532
2 11	-1.6051	0.3437	-4.6700	0.0000	-2.2795	-0.9306
2 12	-1.9370	0.3611	-5.3600	0.0000	-2.6456	-1.2284
3 2	0.8314	0.2038	4.0800	0.0000	0.4316	1.2313
3 3	0.5574	0.1967	2.8300	0.0050	0.1715	0.9433
3 6	-0.5157	0.2151	-2.4000	0.0170	-0.9377	-0.0937
3 7	1.0316	0.2158	4.7800	0.0000	0.6082	1.4550
3 10	-2.3855	0.2324	-10.2700	0.0000	-2.8415	-1.9295
3 11	-0.4612	0.2137	-2.1600	0.0310	-0.8806	-0.0418
3 12	-2.1638	0.3678	-5.8800	0.0000	-2.8855	-1.4421
4 2	0.0817	0.1227	0.6700	0.5060	-0.1592	0.3226
4 3	-0.2495	0.1673	-1.4900	0.1360	-0.5777	0.0787
4 4	0.0836	0.2425	0.3400	0.7300	-0.3922	0.5594
4 6	0.2482	0.1435	1.7300	0.0840	-0.0334	0.5298
4 7	0.4194	0.0899	4.6600	0.0000	0.2429	0.5958
4 9	-0.6639	0.2051	-3.2400	0.0010	-1.0664	-0.2615
4 10	-0.0692	0.1646	-0.4200	0.6740	-0.3921	0.2537
4 11	-0.0517	0.2269	-0.2300	0.8200	-0.4969	0.3935
_cons	**18.1394**	**0.1778**	**102.0300**	**0.0000**	**17.7905**	**18.4883**

Annex 1.3: Generated Base Model for Kawempe and Rubaga

logprice	Coef.	Robust Std. Err.	t	P>t	[95% Conf.	Interval]
quartercode						
7	-0.0003	0.0712	0.0000	0.9960	-0.1405	0.1399
8	0.1077	0.0660	1.6300	0.1040	-0.0222	0.2375
9	-0.0704	0.0661	-1.0600	0.2880	-0.2006	0.0598
10	-0.0552	0.0717	-0.7700	0.4420	-0.1962	0.0859
11	-0.0740	0.0710	-1.0400	0.2990	-0.2138	0.0659
12	-0.1021	0.0871	-1.1700	0.2420	-0.2735	0.0693
13	-0.1677	0.0778	-2.1500	0.0320	-0.3208	-0.0145
Parishcode						
2	0.4212	0.5332	0.7900	0.4300	-0.6282	1.4707
3	1.5409	0.0960	16.0500	0.0000	1.3520	1.7299
4	-1.0511	0.4602	-2.2800	0.0230	-1.9569	-0.1452
6	-1.6897	0.8952	-1.8900	0.0600	-3.4518	0.0724
7	2.4246	0.3499	6.9300	0.0000	1.7359	3.1133
8	0.2922	0.3133	0.9300	0.3520	-0.3245	0.9088
9	-0.2195	0.6158	-0.3600	0.7220	-1.4317	0.9927
10	-3.0531	1.2555	-2.4300	0.0160	-5.5243	-0.5819
11	-0.3027	0.6122	-0.4900	0.6210	-1.5077	0.9023
13	2.0652	0.4442	4.6500	0.0000	1.1908	2.9395
14	-0.5536	0.8662	-0.6400	0.5230	-2.2586	1.1513
15	-0.0271	0.4775	-0.0600	0.9550	-0.9670	0.9128
16	1.2130	0.3064	3.9600	0.0000	0.6099	1.8161
statusofbuilding	0.1470	0.0495	2.9700	0.0030	0.0495	0.2445
structure						
2	2.4677	0.5623	4.3900	0.0000	1.3608	3.5745
3	0.2067	0.1210	1.7100	0.0890	-0.0315	0.4449
4	3.5018	0.6878	5.0900	0.0000	2.1479	4.8557
numberoffloors	0.5673	0.0944	6.0100	0.0000	0.3814	0.7532
natureofroofingmaterial	0.0958	0.0583	1.6400	0.1010	-0.0188	0.2105
numberofbedrooms	0.5717	0.0809	7.0700	0.0000	0.4124	0.7310
natureoffloorfinish	0.3101	0.0632	4.9100	0.0000	0.1857	0.4345
natureoffencematerial	0.1229	0.0873	1.4100	0.1600	-0.0489	0.2947
structurepossession						
2	0.0975	0.0678	1.4400	0.1520	-0.0360	0.2310
3	0.3779	0.0864	4.3800	0.0000	0.2079	0.5479
4	0.3593	0.0803	4.4800	0.0000	0.2013	0.5173
structure#c.numberofbedrooms						
2	-0.5796	0.1883	-3.0800	0.0020	-0.9504	-0.2089
3	0.0000	(omitted)				
4	-0.8191	0.1359	-6.0300	0.0000	-1.0867	-0.5515
Parishcode#c.numberofbedrooms						
2	-0.2020	0.1895	-1.0700	0.2870	-0.5750	0.1710
4	0.3599	0.1291	2.7900	0.0060	0.1058	0.6141

logprice	Coef.	Robust Std. Err.	t	P>t	[95% Conf.	Interval]
6	0.2031	0.2300	0.8800	0.3780	-0.2496	0.6558
7	-0.8963	0.1261	-7.1100	0.0000	-1.1444	-0.6481
8	-0.1100	0.0933	-1.1800	0.2400	-0.2937	0.0738
9	0.2772	0.2027	1.3700	0.1720	-0.1217	0.6761
10	1.3404	0.4347	3.0800	0.0020	0.4847	2.1961
11	0.0996	0.1804	0.5500	0.5810	-0.2554	0.4546
13	-0.4025	0.1246	-3.2300	0.0010	-0.6478	-0.1572
14	-0.0259	0.2224	-0.1200	0.9070	-0.4637	0.4119
15	-0.1393	0.1131	-1.2300	0.2190	-0.3620	0.0834
16	-0.2658	0.0812	-3.2700	0.0010	-0.4256	-0.1060
structure#countycode						
2 6	1.2266	0.4837	2.5400	0.0120	0.2744	2.1788
2 7	-1.0482	0.2703	-3.8800	0.0000	-1.5803	-0.5160
2 8	0.1675	0.3901	0.4300	0.6680	-0.6003	0.9353
2 9	-0.5784	0.1617	-3.5800	0.0000	-0.8966	-0.2602
2 10	-0.4433	0.4693	-0.9400	0.3460	-1.3670	0.4805
2 11	0.4421	0.5442	0.8100	0.4170	-0.6291	1.5132
2 13	-0.7425	0.1405	-5.2800	0.0000	-1.0192	-0.4659
2 16	-0.1913	0.2361	-0.8100	0.4180	-0.6560	0.2734
4 4	-0.9158	0.3016	-3.0400	0.0030	-1.5095	-0.3221
4 7	-0.1656	0.3203	-0.5200	0.6060	-0.7961	0.4649
4 9	-1.0626	0.2722	-3.9000	0.0000	-1.5983	-0.5268
4 11	-0.9477	0.2395	-3.9600	0.0000	-1.4192	-0.4761
4 13	-1.5431	0.4447	-3.4700	0.0010	-2.4184	-0.6678
_cons	**16.0480**	**0.2645**	**60.6700**	**0.0000**	**15.5273**	**16.5686**

Annex 1.4: Generated Base Model for Wakiso

logprice	Coef.	Robust Std. Err.	t	P>t	[95% Conf.	Interval]
quartercode						
7	-0.02907	0.038566	-0.75	0.451	-0.1046919	0.0465585
8	0.004438	0.033667	0.13	0.895	-0.0615809	0.0704562
9	0.031931	0.03241	0.99	0.325	-0.031623	0.0954839
10	0.0139	0.03162	0.44	0.66	-0.0481059	0.0759053
11	-0.02884	0.04003	-0.72	0.471	-0.1073394	0.0496545
12	-0.03637	0.042672	-0.85	0.394	-0.1200493	0.0473059
13	-0.00363	0.032609	-0.11	0.911	-0.0675709	0.0603166
Parishcode						
2	0.531427	0.120987	4.39	0	0.2941791	0.768674
3	0.027779	0.101308	0.27	0.784	-0.1708792	0.226438
4	-0.88523	0.169733	-5.22	0	-1.218069	-0.5523966
5	0.015728	0.121958	0.13	0.897	-0.2234245	0.2548797
6	-0.60109	0.161862	-3.71	0	-0.9184937	-0.2836919
statusofbuilding	0.053495	0.022375	2.39	0.017	0.0096198	0.0973708
structure						
2	0.680748	0.103198	6.6	0	0.4783839	0.8831128
4	0.840813	0.123821	6.79	0	0.5980082	1.083617
numberoffloors	0.327059	0.038948	8.4	0	0.250684	0.4034347
natureofroofingmaterial	0.437099	0.025346	17.25	0	0.3873975	0.4868011
numberofbedrooms	0.167385	0.022215	7.53	0	0.1238225	0.2109469
natureoffloorfinish	0.3606	0.033009	10.92	0	0.2958717	0.4253274
natureoffencematerial	0.157656	0.038484	4.1	0	0.082192	0.2331208
structurepossession						
2	-0.05844	0.034073	-1.72	0.086	-0.1252587	0.0083702
3	0.134934	0.029814	4.53	0	0.076472	0.1933966
4	0.249695	0.033328	7.49	0	0.1843421	0.3150485
structure#c.numberofbedrooms						
2	-0.03825	0.035583	-1.08	0.282	-0.10803	0.0315206
4	-0.10465	0.024284	-4.31	0	-0.1522696	-0.0570326
Parishcode#c.numberofbedrooms						
2	-0.00967	0.028562	-0.34	0.735	-0.0656764	0.0463382
3	-0.01306	0.028461	-0.46	0.646	-0.0688741	0.0427463
4	0.266958	0.050947	5.24	0	0.1670536	0.3668616
5	0.000653	0.034446	0.02	0.985	-0.0668934	0.0681998
6	0.102339	0.046986	2.18	0.029	0.010203	0.1944741
structure#countycode						
2 2	-0.06073	0.13632	-0.45	0.656	-0.328044	0.2065847
2 3	-0.01462	0.087521	-0.17	0.867	-0.186246	0.1569991

logprice	Coef.	Robust Std. Err.	t	P>t	[95% Conf.	Interval]
2 4	-0.33688	0.164546	-2.05	0.041	-0.6595419	-0.0142126
2 5	-0.15977	0.124589	-1.28	0.2	-0.4040773	0.0845449
2 6	0.286125	0.174744	1.64	0.102	-0.0565356	0.6287862
4 2	-0.09195	0.099486	-0.92	0.355	-0.287036	0.1031346
4 3	-0.18687	0.0733	-2.55	0.011	-0.330604	-0.0431305
4 4	-0.69539	0.161918	-4.29	0	-1.012904	-0.3778812
4 5	-0.05139	0.089405	-0.57	0.565	-0.2267079	0.1239274
_cons	**17.65082**	**0.09122**	**193.5**	**0**	**17.47194**	**17.82969**

ANNEX 2: QUESTIONNAIRE USED IN DATA COLLECTION

SERIAL No	
Village-where BUILDING(S)on sale is located	
PARISH Where Building on sale is built.	
DIVISION/SUBCOUNTY where BUILDING(S)on sale is located	
DIVISION/COUNTY where BUILDING(S)on sale is located	
DISTRICT - where BUILDING on sale is built	
PLOT SIZE-where BUILDING on sale is built	
PLOT SIZE-Unit of Measurement 1= Hectares , 2= Decimals, 3= Acres, 4=feet, 5=Others(Specify)	
LAND TENURE of the plot: 1. Mailo 2.Freehold 3.Communal 4.Leasehold 5.Others(Specify)	
BUILDING SIZE in Square meters(Dimensions)	
STATUS OF THE BUILDING ON SALE: 1=New 2=Used	
STRUCTURE: 1=Bungalow. 2=Semi-Detached. 3=Condominium. 4=Storied	
NUMBER OF levels (If Storied/Flat in column 9)	
DISTANCE TO TARMAC (KM) MAIN ROAD	
STATE THE NATURE OF THE ROOFING MATERIAL: 1=Tiles. 2=Ironsheets-Tiles Shape. 3=Iron sheets-Ordinary. 4=Others(specify)	
NUMBER OF BEDROOMS.	
NUMBER OF BATHROOMS	
STATE THE NATURE OF THE MATERIAL USED TO FINISH THE FLOOR. 1.Tiles. 2.Cement. 3.Others(Specify).	
STATE THE NATURE OF THE MATERIAL USED TO BUILD THE FENCE. 1.Permanent. 2. Not-Permanent or None	
DOES HOUSE POSSESS: 1=None. 2=Garage. 3=Boy's Quarters. 4=Both 2 &3	
NUMBER OF BUILDINGS/ UNITS ON SALE (Of same characteristics)	
ADVERTISED PRICE FOR THE BUILDING (Asking Price)	

Price Unit 1=UGX 2=USD	
AMENITIES: 1.None 2.Swimming pool. 3.Electronic gates. 4.CCTV. 5.Parking Space. 6.Others.	

Note: It should be noted that some of the variables captured now are not used in the regression and will be used after rebasing the index.