JBLD, VOL.25 NO.2, 2024, pp. 56-70

HEDONIC HOUSE PRICE INDEX FOR DAR ES SALAAM: EXAMINING THE EFFECTS OF DATA FROM INFORMAL AND FORMAL REAL ESTATE AGENTS

Frank Nyanda[1]

ABSTRACT

The Dar es Salaam housing market is among the nascent one in Sub-Saharan Africa with limited availability of housing transaction data. This has contributed to the absence of house price indices to reveal the house price dynamics. However, there are both formal and informal real estate agents with housing transaction data which could be useful in constructing a house price index. Nevertheless, no studies have examined the potential of data from both informal and formal real estate agents for developing house price indices. Using a pooled cross-sectional sample of data from both informal and formal agents, this study determines the effect of the two data sources on the house price index for Dar es Salaam city. The study employs OLS-based hedonic pricing and the spatial hedonic models (Spatial Durbin). Results from this study indicate that, adding data from formal real estate agents to the data from informal agents seems to marginally improve the hedonic model and produce a smoother house price index. However, the marginal improvement is probably due to the differences in the volumes of data rather than the data source. Findings suggest that, a house price index for Dar es Salaam could be developed using a combination of data from both formal and informal real estate agents.

Keywords: Property transaction, housing market, OLS, Spatial hedonic models

INTRODUCTION

Sources and volume of data for property price indices are important considerations for the quality of any index in question (Miller and Maguire, 2022). Since the indices are a result of some form of a model, the data quality will certainly determine the degree of precision of the model and the effectiveness of the index. In light of how sensitive the data would be for the resulting index, some authors have cautioned their research findings to the readers on account of data limitations (Sipan, Mar Iman, and Razali, 2018; Aliefendioğlu *et al.* 2022). In any single market, authors have also cautioned against the effects of different sources of data having the potential to bring about differences in the resulting constructed indices (Grover and Grover, 2013). Disregarding the different sources of data within any single market, experience has also shown that property markets in general are not similar in terms of data sources and quality of property indices. This implies that some indices in certain markets are generally superior to those in other markets because of the sources and quality of data.

Studies have shown that it is easier to get property transaction data in developed markets than it is in nascent markets, particularly for property indices (Owusu-Ansah *et al.*, 2017). Many property researchers in developed markets have

[1] Frank Nyanda is a lecturer Department of Business Studies, School of Earth Sciences, Real Estate, Business and Informatics, Ardhi University, P. O. Box 35176, Dar es Salaam
His email address is gfrank842003@yahoo.com

benefited from the availability of databases with property transaction data which are regularly updated. These databases are not well-established in many nascent property markets inclusive of the Tanzanian property market. However, this does not mean that property transaction data is completely inexistent. There are market operations such as property transactions, although there is also the absence of a system to appropriately capture such data for a consolidated database (Nyanda, 2024).

Using informal and formal property agents is among the alternatives to accessing property transaction data. These market players have data that may be useful inputs in constructing an indicative index. Previous research has demonstrated that it is possible to develop an index using data from informal property agents in Dar es Salaam (Nyanda, 2024). When the study was carried out, the only data that had been made available was from the informal property agents. The data from informal property agents could be subjected to several criticisms such as the question of representativeness of the resulting house price index. Henceforth, with a new set of data from the formal property agents, this study aims to examine the effect of data sources from the formal and informal real estate agents on the house price index for Dar es Salaam. In fulfilling the study objective, the study compares the index created with data from informal agents to the index created with data from both formal and informal agents to establish if the data from formal agents has a significant effect on the resulting index.

Studies on property price indices are highly scanty in Sub-Saharan Africa (SSA) and specifically in the Tanzanian context. To the knowledge of the author, no other author had attempted to carry out this kind of study in the Dar es Salaam housing market let alone the Tanzanian context. This may have contributed to investors' little interest, particularly foreigners, in considering investing in the Tanzanian property market due to the lack of indication of house price trends.

This study aims not only to fill the gap, given the scarcity of similar studies on property price indices, but also to provide valuable information on the usefulness of the available data as it examines the effect of the two data sources (formal and informal agents) on the Dar es Salaam house price index.

BRIEF LITERATURE REVIEW

For the past few decades and until recently, studies on house price indices have been more popular in developed property markets for several reasons, including the ease of data availability and their use in real estate appraisals (Knoll *et al.*, 2017). This has facilitated the availability of crucial market information to stakeholders such as planners, property developers and other policymakers in the property industry. Similar studies are highly scant in nascent markets and many of the reasons are centred on the limited data sources (Behr *et al.* 2023; Owusu-Manu *et al*. 2019). Because of the different levels of property market development in the globe, various sources of data have been used to develop the indices. Some of the popular data sources include municipal housing companies, government agencies, mortgage banks and real estate agents, to mention a few (Ismail *et al.*, 2021; Sipan *et al.*, 2018; Grover and Grover, 2013). It is important to also note that most of the studies on house price indices in the recent literature, do not directly focus on the different sources of data for indices but the issue of data sources is just part of other specific issues of focus.

Scholars in the developed markets have mostly used the collected and stored data for articles on housing markets and house price indices. Most of this is either transactional or appraisal data, which could have several problems (Deng *et al*., 2018; Silver, 2016). Since the focus is more on the prices than the values, transactional data is arguably more useful than appraisal data. In the study about house price inequalities between the UK and Germany, Blaseio and Jones (2020), explain that house transaction data in Germany is aggregated and sold by some private companies. This suggests that the data may not be freely publicly available in German but for sale by some companies. The authors point out that one challenge in their analysis was to do with data that was compiled differently between the two countries. The authors rely on house transaction data from private companies and federal government offices from the 1990s in the case of Germany and data from mortgage providers in the case of the UK.

A similar source of data (private companies in the case of Germany) is documented in the article by Kajuth *et al.* (2016), which focused on apartment prices in comparison to the predicted prices in the previous years. The article compares the effectiveness of the time series and panel data in modelling house prices. Although the study is carried out in a developed property market, the authors point out that time series data is still limited for use in valuing property since the data representative for the whole of Germany is only available from 2004, making the sample period short and limiting the time-series model precision. As for the case of Germany in the article by Blaseio and Jones (2020), also Sipan *et al.* (2018), make use of the federal government's data to develop a spatial-temporal neighbourhood-level house price index in the Malaysian case. The authors point out the hurdle in getting data from other sources due to data collection process being expensive, and acknowledge the lower level of data accuracy.

Other scholars such as Kholodilin and Mense (2012), have also pointed out some specific companies as sources of data for Germany such as BulwienGesa AG and local expert communities for the case of the Destatis company. The data from BulwienGesa is criticized for the underrepresentation of the rural areas and focus on the cities while the dependence on local expert communities leads to low levels of data and hence a low sample for effective representation. There are cases in the European literature where some companies have worked with the country's statistical departments to avail data for house price indices. For instance, in the article by Aliefendioğlu *et al.* (2022), the authors use data from the DataStream database which collects data from the government's Department of Statistics. The authors do not highlight limitations of data for HPI but only report limitations on data for macro-economic variables which are relevant for their article. A similar limitation is also reported by Korkmaz (2020).

Besides the use of data from specific companies, some authors such as Kholodilin and Mense (2012), have constructed their hedonic-based house price indices using data from internet ads. Although the authors focused on the city of Berlin, they argue that it should be possible to replicate the practice in other cities in Germany. Other studies advocating for online data sources based in the developed property markets include Maguire *et al.* (2016); Anenberg and Laufer (2017). Wang, Li and Wu (2020), also point out that online data could be used for the case of China and that it provides a better trade-off between accuracy, reliability and feasibility. The authors sought to advance the adoption of the online data by Anenberg and

Laufer (2017), based on multiple listing services in the U.S. but in their case, in a nascent property market. A study by Mba *et al.* (2023) for the case of Nigeria also advocates the use of online ads for the richness of data for developing a house price index. The study is interesting since it focuses on a case within SSA. The authors developed the residential house price index with about 30,000 transactions obtained by way of web scraping. The study raises several questions. Most notably, the asking price may not be the actual transacted price. The other issue is the time difference between the listing of the property and the actual transaction date.

Data limitation for house price indices is more critical in smaller property markets such as those normally referred to as nascent and emerging markets (Rothacher, 2013). This highlights that the effects of data sources for house price indices in these markets where transactions are insufficiently captured could be much more pronounced (Owusu-Ansah and Talinbe Abdulai, 2014). Unlike the easy availability of data in developed property markets through government agencies, specialised private companies and online sources, such sources of data are not well prevalent in smaller property markets and more specifically the markets in SSA. There are lots of similarities in the challenges of data across the countries in SSA. The study by Baako (2019), in the case of Ghana, highlights that transactional data could be available in a government department which records transfers of ownership between property owners. However, the author points out the likely unreliability of data due to the possible underreporting of the transaction prices. A similar source of data was used in the article by Owusu-Ansah *et al.* (2017), for the same case of Ghana. The authors do not consider spatial dependence issues, probably due to the absence of coordinate data for each observation.

Limited studies such as Owusu-Manu *et al.* (2019), have focused on house price indices with data from real estate agents in nascent markets. The study focused on the relationship between housing attributes and the price in Ghana. The authors used a snowball approach in choosing the agents. Since the study only uses a sample of about 270 observations, this highlights critical data limitations. Also, the study does not account for spatial dependency. A similar kind of data is also used by Kieti and Ogolla (2020), in a paper on the hedonic valuation of apartments in Kenya. The authors used a sample size of 120 apartment sales from real estate firms practising agency. Despite a small sample size, they concluded that the hedonic model was accurate and reliable. Again, a small dataset points out severe data limitations, typical of nascent property markets (Behr *et al.* 2023). The study by Baffour Awuah *et al.* (2017), does not attempt to construct a house price index but highlights that property market information could easily be available from estate agents. While impliedly including transaction data as part of the market information, they point out the lack of substantial expertise especially among the informal agents in the data collection and management, thereby raising the data validity concern.

Very few studies have focused on the data from formal or informal real estate agents in nascent property markets. A recent study demonstrates the first house price index for Dar es Salaam city using data from informal agents by employing the hedonic approach (Nyanda, 2024). The study focused further on showing that such data could also provide meaningful results even with spatial considerations. Although the study demonstrates a meaningful index, the exclusion of data from the formal agents raises some concern since the assumption would be that the data from the formal agents would likely be more useful and would generate a more meaningful index. With property transaction data from the formal agents, the current study fills the gap by examining the effect of data from both formal and informal agents on the house price index for Dar es Salaam city.

MATERIAL AND METHODS

The Case Study Area

Dar es Salaam city is the major commercial hub of Tanzania, and according to the National Bureau of Statistics (NBS), it has a population of about 5.4 million residents (NBS, 2022). Given the 2013 census report (NBS, 2013), the population has increased by about 23% from 2012. The population density is estimated at 3,800 per km^2 (NBS, 2022). The city is the most populous in Tanzania followed by Mwanza, with a population of about 3.7 million residents and a population density of only about 391 per km^2. According to Andreasen and Agergaard, (2016), migration from other cities is believed to be the factor that caused the sharp increase in population in Dar es Salaam city. Compared to matured markets, the Dar es Salaam real estate market can only be described as nascent, due to factors such as limited transparency and low trading volumes of commercial real estate investments (Rothacher, 2013). There are both planned and unplanned settlements. Efforts are underway to formalise the unplanned settlements. The property market exists in both planned and unplanned settlements. The Dar es Salaam regional land office estimates about 3000 annual property transactions from planned settlements. According to CAHF (2022), the housing deficit is at 432,000 units. The government through the Tanzania Building Agency (TBA) and the National Housing Corporation (NHC), contributes to the supply of housing units. These, however, have not been able to saturate the market demand. Because of mortgage access difficulties, many have resorted to building own homes incrementally.

Data collection

The author collected the data in person by moving with real estate agents across the observation of sold housing units from 2010 to 2019. This was done to inspect the properties, capture the coordinates and measure the distance for various proximity variables, such as arterial roads, hospitals, airports, and food markets. Figure 1 shows the observations in H-labelled balloon-shaped place marks. In total, 1000 observations were captured. The data clean-up process trimmed down the sample size to 954, including 430 observations from informal agents and 524 from the formal agents. The data collection process was quite challenging as the real estate agents were highly hesitant to share the data. However, the informal agents were more willing to share data than their formal counterparts. Therefore, data was first collected from the informal real estate agents and later on from the formal agents.

Compared to the formal real estate agents, there seemed to be a problem of inefficient data-keeping methods by the informal agents as they kept their data records in old notebooks. Data from the informal agents was verified by their peers to assure reliability. This was possible since at times the informal agents work in groups to maximise their chances of getting clients, although they would need to share the commission afterwards. Unlike the formal agents, the informal agents are not officially registered as real estate agents. Because of the sensitive nature of the data collection exercise, the houses were not internally inspected. This information only came from the agents. Coordinates were captured by moving close enough to each of the observations in a vehicle with the agents. This also made it possible to confirm some of the external features of the houses. The coordinates were later plotted on Google Earth and updated to the middle point of the houses. As the data from the formal property agents did not include coordinate points, Coordinates were captured with a GPS by moving across the observations. The annually aggregated data is of a pooled-cross-sectional structure.

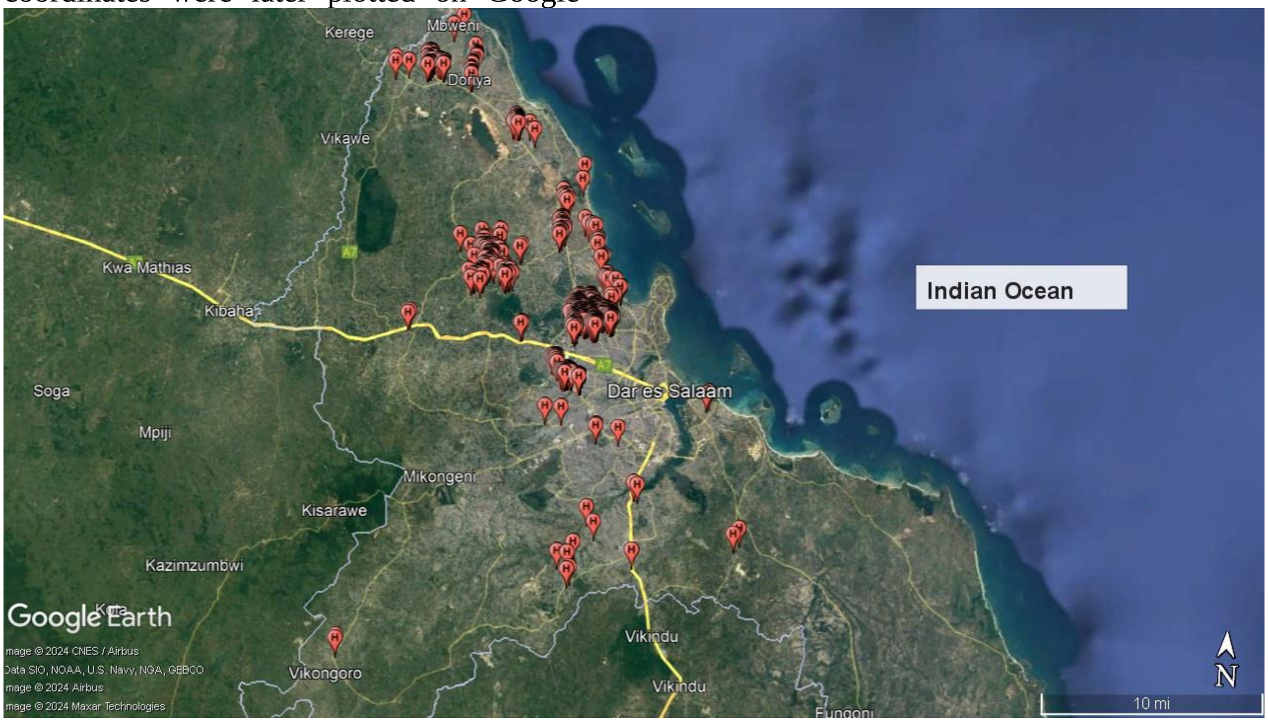

Figure 1: **Captured observations (Sold houses)**

Methods

The indices created adopted the hedonic method (Nishi *et al.*, 2021 and Lancaster, 1996). Given the data structure, the adoption of other index creation methods such as the repeat-sales method (Minne *et al.*, 2020), was not possible. This is because the repeat-sales method requires that the houses are transacted at least twice. This also makes it impractical to adopt the hybrid method (Leishman and Watkins, 2017), which combines both the hedonic and the repeat-sales methods. The hedonic regression model served as the basis for creating the indices with both data sets from the informal and formal real estate agents. The model includes transaction years as dummy variables, the coefficients of which, are used for the indices. The OLS model is used to compare the variable coefficients between data from the informal agents and data from both the informal and formal agents. The OLS models are created using STATA. Thereafter, the OLS model with data from both informal and formal agents is extended to take into consideration the spatial effects basing on the spatial autocorrelation tests. The spatial models include the Manski model (Burridge, 2012), the SARAR model (Kelejian and Prucha, 1998), and the Spatial Durbin Model (SDM) or the Spatial Durbin Error Model (SDEM) (LeSage and Pace, 2009). Several authors have discussed what would be the logical starting point in testing for spatial effects (LeSage and Pace, 2009). The analysis, therefore, adopts the Spatial Durbin Model (SDM) as the starting point, the benefit of which is the possibility to restrict the SDM to the other nested models using the likelihood ratio tests (LR-test) and the Lagrange Multiplier test (LM-test) (Michieka *et.al.*, 2022). Depending on whether the SDM is restricted or not, the coefficients of the year dummy variables are used to plot the house price indices: the index with data from only informal real estate agents and the index with data from both informal and formal agents. The plotted indices are then compared to establish if the addition of data from the formal real estate agents significantly alters the model with data from the informal real estate agents.

EMPIRICAL RESULTS AND DISCUSSION

Descriptive statics for the observations

Table 1 provides the descriptive statics for the observations' variables. The dependent variable is "logprice" which is the natural log of the prices of the transacted houses. The variable "neigh kimabu" represents the neighbourhoods of Kijitonyama, Makumbusho and Bunju. Because of the relative similarity, the neighbourhoods are combined into one variable as a solution to multicollinearity. The neighbourhood of Goba is represented by the variable "neigh goba". The neighbourhood of Tabata is represented by the variable "neigh tabata", and the neighbourhood of kawe is represented by the variable "neigh kawe".

Table 1. Descriptive Statistics of the variables for 954 sold houses from 2010 to 2019

Variable	Mean	Std. Dev.	Min	Max
logprice	8.114	.417	6.699	9.195
neigh kimabu	.481	.5	0	1
neigh goba	.123	.328	0	1
neigh tabata	.129	.335	0	1
neigh kawe	.071	.257	0	1
no storeys	1.06	.246	1	3
roof cas	.158	.365	0	1
roof asbestos	.021	.143	0	1
roof claytiles	.093	.291	0	1
ceil gypchtng	.773	.419	0	1
window wood	.637	.481	0	1
floor cerrtiles	.405	.491	0	1
floor terrazo	.006	.079	0	1
no bedrooms	3.351	.944	1	8
plotsize	303.32	268.455	24	2000
fence	.51	.5	0	1
year 2010	.072	.259	0	1
year 2011	.039	.193	0	1
year 2012	.048	.214	0	1
year 2013	.08	.271	0	1
year 2014	.078	.268	0	1
year 2016	.146	.353	0	1
year 2017	.151	.358	0	1
year 2018	.128	.334	0	1
year 2019	.101	.301	0	1
dist arteroad	.154	.191	.005	1.64
dist hospital	1.612	1.479	.045	10.544
dist airport	14.864	7.49	3.161	33.707
dist foodmkt	2.439	2.716	.131	12.53

The physical features of the observations are also abbreviated as follows: The number of storeys is denoted by "no storeys", the roof of corrugated aluminium sheets is represented by "roof cas" whereas the roofs of asbestos and clay tiles are represented by "roof asbestos" and "roof clay tiles" respectively. The ceilings of gypsum, chipboard and tongue and groove (TNG) are combined into the variable "ceil gypchtng". The window of wood is represented by "window wood". Floors of ceramic tiles and floors of terrazzo are represented by "floor cerrtiles" and "floor terrazzo" respectively. The variable "no bedrooms" represents the number of bedrooms. The minimum number of bedrooms is 1 while the maximum number is 8. On average, the surveyed homes have 3 bedrooms. The maximum plot size is 2000m^2 while the minimum size is 24m^2. Tiny houses are found in the neighbourhood of Goba. Plot size is represented by the variable "plotsize". The dummy variable "fence" stands for fence availability. Variables "year 2010" through "year 2019" stand for the years when the surveyed houses were sold.

The descriptive statistics in Table 1 also include the proximity variables. The variable "dist arteroad" represents the distance to the nearest arterial road. The maximum distance is 1.64 km and the minimum distance is 0.005 km. On average, houses are 0.154 km farther from the arterial roads. The distance from the nearest hospital is represented by the variable "dist hospital". The maximum distance is 10.544km and the minimum distance is 0.045 km, while the average distance is 1.612km. The variable "dist airport" represents the distance from the airport. The maximum distance is 33.707km and the minimum distance is 3.161km. The variable "dist foodmkt" represents the distance to the nearest local food market. The maximum distance is 12.53km, the minimum distance is 0.131km and the average distance is 2.439km.

Hedonic models

Table 2 provides the variables coefficients, the R^2 and adjusted R^2 for the two OLS models i.e. one with data from the informal property agents and the other with data from both the informal and formal agents[2]. The one general observation is that most of the significant variables with informal agents' data have also remained significant with both groups' data. There is also no change in the signs of the coefficients after adding the formal agents' data. This implies that the variables of the OLS hedonic models are not very different with the inclusion of data from the formal agents.

The informal agent's data suggest that the houses in the neighbourhoods of Goba—the variable "neigh_goba", would sell for 50.5% lower than the houses in the default neighbourhood of Sinza. However, with the addition of formal agents' data, they would sell for a 55.3% lower price compared to the houses in Sinza. Both coefficients appear to be highly significant. Likewise, a house in the neighbourhood of Tabata—the variable "neigh_tabata" would sell at 41.6% lower than a relatively similar house in Sinza as per the informal agents' data, while with both groups of agent's data, the house would sell for 37.3% lower compared to a comparable house in Sinza. Although houses in Kawe neighbourhood would sell for 15.4% higher than the houses in Sinza as per the data from informal agents, they would sell for 32.2% higher as per the data with both agent groups. However, the data with both groups show a much higher level of significance compared

[2] The data from informal agents was collected first since their formal counterparts were not ready to share their data. Therefore, hedonic models were first created with informal agents' data and later on with the additional data from the formal agents when they eventually agreed to share their data.

to the data with only informal agents. While the number of storeys does not appear to significantly contribute to house prices with informal agents' data, they appear to significantly contribute to house prices with data from both groups of agents by adding about 13.8% to the price, per each added storey. The ceilings of gypsum, chipboards, and Tongue and Groove significantly improve the house prices with data from both groups i.e. 7.9% higher price compared to the default scenario/variable of "no ceiling" but the ceilings would add about 17.5% as per the data from only the informal agents.

The floor of ceramic tiles significantly improves the house prices by about 9.33% compared to the default floor of sand and cement screed as per the data from the informal agents. However, the same floor type improves the house prices by about 11.5% as per the data from both groups of agents. The number of bedrooms significantly improves the house prices with data from informal agents and from both groups i.e. 2.69% and 1.92% respectively. Likewise, although the plot size improves the house prices by a very small percentage with data from informal agents and from both groups i.e. 0.039% and 0.032% respectively, the coefficients are significant with both datasets. The effect of fencing is also significant with all datasets. The year dummy variable coefficients have a slightly different picture compared to other variables.

Table 2. OLS regression results with data from informal, both formal and informal agents, and SDM with data from both agents

Variable	(*OLS coefficient*) Informal agents' data	(*OLS coefficient*) Both informal and formal agents' data	(Coefficient) Spatial Durbin – data from both formal and informal agents	
			Direct effect	Indirect effect
neigh_kimabu	-0.0549 (-1.87)	-0.0183 (-0.78)	-0.0868 (-1.58)	0.145 (0.94)
neigh_goba	-0.505*** (-7.56)	-0.553*** (-13.30)	-0.341*** (-3.54)	0.316 (1.18)
neigh_tabata	-0.416*** (-5.61)	-0.373*** (-10.33)	-0.400*** (-4.94)	0.546** (2.62)
neigh_kawe	0.154* (2.53)	0.322*** (6.48)	0.103 (1.22)	0.470 (1.60)
no_storeys	0.0892 (1.63)	0.138*** (3.62)	0.166*** (4.95)	0.0301 (0.07)
roof_cas	0.0325 (1.09)	0.00778 (0.31)	0.0254 (1.18)	-0.304 (-1.28)
roof_asbestos	0.0749(1.30)	0.0483 (0.95)	0.0594 (1.38)	0.220 (0.50)
roof_claytiles	0.0742* (2.18)	0.0127 (0.43)	0.0525* (2.06)	-0.483 (-1.58)
ceil_gypchtng	0.175** (3.22)	0.0794*** (3.52)	0.0829*** (3.76)	-0.0000162 (-0.00)
window_wood	0.0485 (1.14)	0.0184 (0.64)	-0.0143 (-0.58)	0.125 (0.54)
floor_cerrtiles	0.0933* (2.22)	0.115*** (4.29	0.103*** (4.28)	0.00820 (0.04)
floor_terrazo	0.00439 (-0.04)	0.0319 (0.36)	0.0380 (0.50)	0.0544 (0.05)
no_bedrooms	0.0269** (2.80)	0.0192* (2.34)	0.0206** (2.97)	-0.0773 (-1.32)
plotsize	0.000388** (2.74)	0.000324*** (10.14)	0.000220*** (7.61)	0.000436* (2.39)
fence	0.125*** (3.69)	0.0623** (3.20)	0.0276 (1.59)	0.129 (0.97)
year_2010	-0.176* (-2.34)	-0.157*** (-4.75)	-0.236*** (-7.71)	0.319 (1.48)
year_2011	-0.0317 (-0.25)	-0.0818* (-2.03)	-0.169*** (-4.65)	0.152 (0.64)
year_2012	-0.221 (-1.75)	-0.0535 (-1.45)	-0.139*** (-4.22)	0.181 (0.93)
year_2013	-0.0946 (-1.80)	-0.0725* (-2.39)	-0.111*** (-4.25)	-0.0129 (-0.07)
year_2014	-0.00634 (-0.15)	-0.0655* (-2.17)	-0.0637* (-2.52)	0.0525 (0.21)
year_2016	0.0381 (1.23)	0.0378 (1.52)	0.0344 (1.63)	0.299 (1.46)
year_2017	0.0143 (0.41)	0.0340 (1.36)	0.0613** (2.84)	-0.0556 (-0.34)
year_2018	0.0500 (1.32)	0.0548* (2.06)	0.0976*** (4.17)	-0.576** (-3.06)
year_2019	0.0552 (0.94)	0.102*** (3.39)	0.0989*** (3.71)	-0.0663 (-0.33)
dist_arteroad	-0.229*** (-3.68)	-0.207*** (-4.99)	-0.272*** (-5.13)	0.197 (0.84)
dist_hospital	-0.0108 (-0.90)	-0.00272 (-0.42)	-0.0200* (-2.39)	0.0597 (1.66)

dist_airport	0.0108** (3.04)	0.00457** (2.87)	0.0123* (2.05)	-0.0176 (-1.94)
dist_foodmkt	-0.0108 (-0.89)	-0.0333*** (-6.97)	-0.00710 (-0.93)	-0.0505* (-2.28)
R^2	0.774	0.754		
Adjusted R^2	0.758	0.746	Rho = 0.853*** (10.95)	
Observations	430	954	Sigma = 0.172*** (43.41)	

Notes: t statistics in parentheses; $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Compared to the base year 2015, only the prices in the year 2010 were significantly different i.e. prices less by 17.6%. As per the data from both informal and formal agents, only the prices in the years 2016 and 2017 are not significantly different from the prices in the base year 2015. The rest are all significantly different from the base year 2015: houses sold in years 2010, 2011, 2012, 2013 and 2014 sold at prices less by 15.7%, 8.18%, 5.35%, 7.25%, and 6.55% respectively, whereas houses in years 2018 and 2019 sold at prices higher by 5.48% and 10.2% respectively. Most of the proximity variables are also significant as per the data from both groups (informal and formal agents). Each kilometre further from the arterial road reduced house prices by 20.7%, implying that buyers are willing to pay more for houses near the arterial roads, regardless of the associated disamenities. Each kilometre further from the airport improved the house prices by 1.08%, suggesting that buyers are willing to pay less for houses located near the airport, probably due to disamenities such as noises and the areas near the airport being used for industrial activities and warehouses. Each kilometre further from the local food market reduced the house prices by 3.33%, implying that buyers are willing to pay more for houses near the local food markets.

Given the likelihood of the presence of spatial effects on house prices, the testing of the existence of such effects starts with the Spatial Durbin Model. Columns 4 and 5 of Table 2 provide the direct and indirect effects as generated by the Spatial Durbin Model. Below the columns, the Rho and Sigma coefficients are highly significant with t statistics of 10.95 and 43.41, respectively. The significance of the coefficients of both Rho and Sigma coefficients imply that there exist spatial influences which should not be ignored in the interest of the accuracy of the hedonic model which would be the basis of the plotted house price index. Because there are other nested models in the Spatial Durbin Model (SDM) such as the Spatial Lag model (SLM), and the Spatial Lagged X Model (SLX), it is best to test if the SDM should be restricted to any of the nested models. The likelihood ratios are the basis of this test. For the sake of comparison of the ratios between data from the informal agents and data from both the informal and formal agents, the SDM likelihood ratios were generated from data from informal agents and thereafter with data from both agents.

With data from only the informal agents, the assumed zero value for Rho returned the p-value of 0.0007, suggesting against the restriction of the model to SLX. However, the assumed zero value to lagged X's returned the p-value of 0.0661, suggesting that the model should be restricted to SLM. The picture is a little different with data from both informal and formal agents. The assumed zero value for Rho returned the p-value of 0.0000 and the assumed zero value for lagged X's also returned the p-values of 0.0000, both of which suggest that the SDM should neither be restricted to SLX nor SLM. It turns out that there are more spatial influences which were not picked up by the data from only the informal agents. Nevertheless, it is still possible to compare the resulting house price index with the SDM as per the data from both informal and formal agents and the resulting index with the SLM as per the data from both groups of real estate agents.

LeSage and Pace (2009) highlight the complexity of parameter estimation in SDM due to spatial spillover effects or spatial diffusion, emphasizing caution in coefficient interpretation. The SDM model provides the twin effects i.e. Direct and indirect. The first effect: the direct effect is concerned with the impact of the variable linked to the corresponding house. In Table 2, most of the direct effects in the SDM appear more significant as compared to the indirect effects. Given the direct effects, the houses in Goba and Tabata sold for 34.1% and 40% respectively, lower than the houses in Sinza. Each added storey improves the house price by 16.6%. A roof of clay tiles directly improves the house price by 5.25% compared to the default roof of corrugated iron sheets. Ceilings of gypsum, Tongue and Groove, and chipboard directly improve the house prices by 8.29% over the default variable of "no roof". A floor of ceramic tiles directly improves the house price by 10.3% over the default floor of sand and cement screed. As also expected, each added bedroom directly improves the house price by 2%. For each added square metre of plot size, the house price directly improves by 0.02%.

The prices according to the year dummy variables in the SDM direct effects are all significantly different from the default year 2015, except for the year 2016. In 2010, 2011, 2012, 2013, and 2014, houses sold at 23.6%, 16.9%, 13.9%, 11.1%, and 6.37% lower prices than in 2015, respectively. The proximity variables are also significant according to the direct effects in the SDM model. For every kilometre further from the arterial road, the house prices directly decrease by 27.2%. For every kilometre further from the hospital, house prices directly decrease by 2%. For every kilometre further from the airport, house prices directly improve by 1.23%. The direct effects of the proximity variables on house prices are still very much logical even with the SDM model.

The Indices

With parameter estimates using data from both the informal and formal agents and data from only the informal agents, the coefficients of the year dummy variables are used to plot the two indices in Figure 2. Compared to the index with data from only the informal agents, the index with data from both informal and formal agents is relatively smoother. The index with data from informal agents is represented by the red line whereas the index with data from both informal and formal agents is represented by the blue line. It is best to keep noting that the index with data from only the informal agents was created with 430 observations while the blue index with data from both informal and formal agents was created with a total of 954 observations.

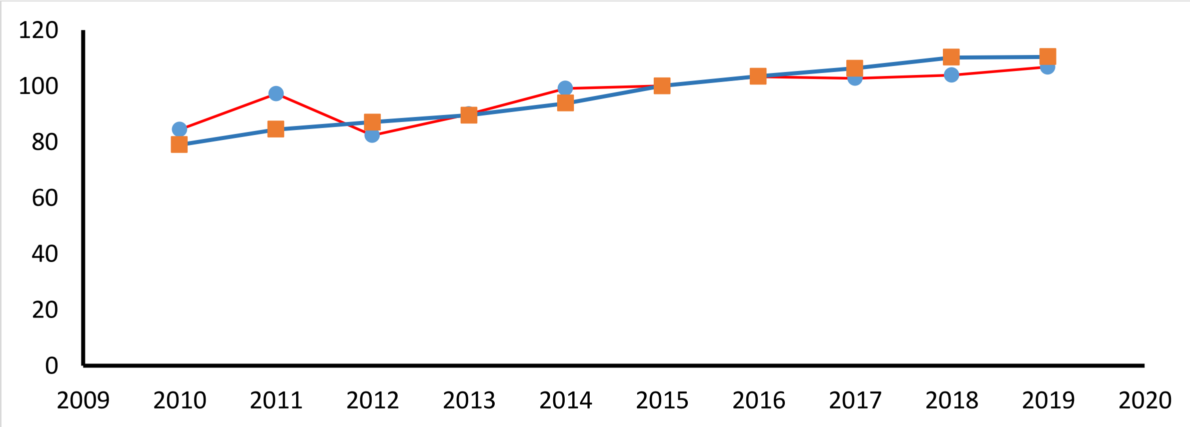

Figure 2. Hedonic house price indices—SDM: Informal agents' data vs both informal and formal agents' data

Both indices show a relatively similar trend: an upward trend, highlighting that prices have generally been increasing from 2010 to 2019. While the index with data from the informal agents shows that prices decreased from the year 2011 to the year 2012, the index in blue shows a slightly different picture: prices increased from the year 2011 to 2012, although both prices were significantly lower than the prices of the default year 2015. In almost all the other years, prices seem to have increased, however, the index in red crossed the index in blue going upwards in 2013 and going downwards in 2016. With fewer data, the index in red seems to exaggerate some points. This is not surprising since, with more data, the distribution of number of the observations corresponding to the respective years improved. Comparing the two indices, the index in blue shows the highest point in year 2019 and it also shows the lowest point in year 2010. Overall, the index in blue with data from both informal and formal agents seems to portray the price trend in a steadier increase as compared to the index in red.

CONCLUSION

The two indices: one with only data from the informal agents and the other with data from both informal and formal agents show an upward trend. The addition of data from the formal agents appears to enhance the trend by smoothing it. The earlier index with data from the informal agents is limited to only 430 observations. With fewer data, it is likely possible for the index to exaggerate certain points and thereby demonstrate some bias. The author noted that real estate agency by formal agents is also done collaboratively with informal agents since some formal agents admitted using their informal counterparts to get the buyers. Some of the formal agents seemed to have information about the houses sold by their informal counterparts and could confirm the prices. The resulting indices could suggest that the quality of the index is not necessarily due to the source of data between the informal and the formal agents but rather, the volume. The index in blue improves simply due to the inclusion of more data but not necessarily due to the inclusion of data from the formal agents. Therefore, it may not matter whether the data comes from formal or informal agents but it matters more if the dataset is large enough for an index of better quality. The key issue regarding data is that the informal agents have poor records keeping of data and may need more time to retrieve the data for past transactions, albeit, when available, the data seems to be reliable. With better data-keeping skills by the formal agents, the data-collection process is much faster. The problem with the formal agents is getting their acceptance to share the data. The hedonic model with fewer data does not show some of the variables as significant whereas some of these variables are significant in the SDM with the direct effects. The model with fewer data may potentially underestimate the significance of certain variables.

Informal real estate agents in nascent markets like Dar es Salaam can provide valuable data for house price indices and thereby contribute in improving the housing market transparency. To capture data, a database should be created for transactions. Regulations should demand agents report transactions, motivating them to share data for research and index creation. This could among the criteria to formalize their role in the housing market.

REFERENCES

Aliefendioğlu, Y., Tanrivermis, H. and Salami, M. A. (2022), "House price index (HPI) and Covid-19 pandemic shocks: evidence from Turkey and Kazakhstan". International Journal of Housing Markets and Analysis, Vol. 15 No. 1, pp. 108-125.

Andreasen, M. H. and Agergaard, J. (2016), "Residential mobility and homeownership: tracing intraurban settlement trajectories of residents in the periphery of Dar es Salaam", Population and Development Review, Vol. 42 No. 1, pp. 95-110.

Anenberg, E. and Laufer, S. (2017), "A more timely house price index". The Review of Economics and Statistics, Vol. 99 No. 4, pp. 722-734.

Baako, K. T. (2019), "Determining House Prices in Data-Poor Countries: Evidence from Ghana.". International Real Estate Review, Vol. 22 No. 4, pp. 571-595.

Baffour Awuah, K. G., Gyamfi-Yeboah, F., Proverbs, D. and Lamond, J. E. (2017), "Sources and reliability of property market information for property valuation practice in Ghana". Property Management, Vol. 35 No. 4, pp. 448-466.

Behr, D. M., Chen, L., Goel, A., Haider, K. T., Singh, S. and Zaman, A. (2023). "Estimating House Prices in Emerging Markets and Developing Economies". World Bank, Policy Research Working Paper Series.

Blaseio, B. and Jones, C. (2020), "Regional economic divergence and house prices: a comparison of Germany and the UK". International Journal of Housing Markets and Analysis, Vol. 12 No. 4, pp. 722-735.

Burridge, P. (2012), "Improving the J test in the SARAR model by likelihood-based estimation". Spatial Economic Analysis, Vol. 7 No. 1, pp. 75-107.

CAHF. (2022), Housing finance in Africa yearbook: 13th edition – 2022.

Deng, Y., Girardin, E. and Joyeux, R. (2018), "Fundamentals and the volatility of real estate prices in China: a sequential modelling strategy", China Economic Review, Vol. 48, pp. 205-222.

Grover, R. and Grover, C. (2013), "Residential property price indices: fit for purpose?". International Journal of Housing Markets and Analysis, Vol. 6 No. 1, pp. 6-25.

Ismail, M., Warsame, A. and Wilhelmsson, M. (2021), "Do segregated housing markets have a spillover effect on housing prices in nearby residential areas?", Journal of European Real Estate Research, Vol. 14 No. 2, pp. 171-188.

Kajuth, F., Knetsch, T.A. and Pinkwart, N. (2016), "Assessing house prices in Germany: evidence from a regional data set", Journal of European Real Estate Research, Vol. 9 No. 3, pp. 286-307.

Kelejian, H. H. and Prucha, I. R. (1998), "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances". Journal of Real Estate Finance and Economics, Vol. 17, pp. 99-121.

Kholodilin, K. A. and Mense, A. (2012). Internet-Based Hedonic Indices of Rents and Prices for Flats: Example of Berlin, Discussion Papers of the German Institute for Economic Research DIW, Berlin.

Kieti, R. M. and Ogolla, W. (2020), "Hedonic valuation of apartments in Kenya", Property Management, Vol. 39 No. 1, pp. 34-53.

Knoll, K., Schularick, M. and Steger, T. (2017), "No Price Like Home: Global House Prices 1870-2012". American Economic Review, Vol. 107 No. 2, pp. 331-353.

Korkmaz, Ö. (2020), "The relationship between housing prices and inflation rate in Turkey evidence from panel Konya causality test". International Journal of Housing Markets and Analysis, Vol. 13 No. 3, pp. 427-452.

Korkmaz, Ö. (2020), "The relationship between housing prices and inflation rate in Turkey evidence from panel Konya causality test", International Journal of Housing Markets and Analysis, Vol. 13 No. 3, pp. 427-452.

Lancaster, K. J. (1966), "A new approach to consumer theory". Journal of Political Economy, Vol. 74, pp. 132-57.

Leishman, C. and Watkins, C. (2017), "Assessing the spatial impact of policy interventions on real estate values: an exemplar of the use of the hybrid hedonic/repeat-sales method". Regional Studies, Regional Science, Vol. 4 No. 1, pp. 202-217.

LeSage, J. P. and Pace, R. K. (2009), Introduction to Spatial Econometrics, CRC Press, Boca Raton, FL.

Maguire, P., Miller, R., Moser, P. and Maguire, R. (2016), "A robust house price index using sparse and frugal data". Journal of Property Research, Vol. 33 No. 4, pp. 293–308.

Mba, M. K., Olorunfemi, O. O., Adeboye, A. A., Takaya, V. A., Mohammed, M. G., Igweze, A. H., Gumbi, L. S. and Onumonu, O. G. (2023), "Residential Property Price Index in Nigeria: A Data Mining Approach", International Journal of Business and Economics Research, Vol. 12 No. 1, pp. 27-33.

Michieka, N. M., Lacombe, D. J. and Ampatzidis, Y. (2022), "Golf courses and home values in California's Central valley". International Journal of Housing Markets and Analysis, Vol. 15 No. 1, pp. 247-272.

Miller, R. and Maguire, P. (2022), "A real-time mix-adjusted median property price index enabled by an efficient nearest neighbour approximation data structure". Journal of Banking and Financial Technology, Vol. 6 No. 1-2.

Minne, A., Francke, M., Geltner, D. and Geltner, R. (2020), "Using revisions as a measure of price index quality in repeat-sales models", The Journal of Real Estate Finance and Economics, Vol. 60 No. 4, pp. 514-553

NBS. (2013), "2012 Population and housing census: population distribution by administrative areas", available at: www.nbs.go.tz/index.php/en/census-surveys/populationand-housing-census/162-2012-phcpopulation-distribution-by-administrative-areas (accessed 27 April 2024).

NBS. (2022), "Administrative Units Population Distribution Report", available at:https://www.nbs.go.tz/ nbs/takwimu/Census2022/Administrative_units_Population_Distribution_Report_Tanzania_volume1a.pdf (accessed 27 April 2024).

Nishi, H.,l Asami, Y. and Shimizu, C. (2021), "The illusion of a hedonic price function: nonparametric interpretable segmentation for hedonic inference". Journal of Housing Economics, Vol. 52, 101764.

Nyanda, F. (2024), "The effect of proximity and spatial dependence on the house price index for Dar es Salaam". International Journal of Housing Markets and Analysis, Vol. 17 No. 4, pp. 945-963.

Owusu-Ansah, A. and Talinbe Abdulai, R. (2014), "Producing hedonic price indices for developing markets: Explicit time variable versus strictly cross-sectional models". International Journal of Housing Markets and Analysis, Vol. 7 No. 4, pp. 444-458.

Owusu-Ansah, A., Adolwine, W. M. and Yeboah, E. (2017), "Construction of real estate price indices for developing housing markets: Does temporal aggregation matter?". International Journal of Housing Markets and Analysis, Vol. 10 No. 3, pp. 371-383.

Owusu-Manu, D. G., Edwards, D. J., Donkor-Hyiaman, K. A., Asiedu, R. O., Hosseini, M. R. and Obiri-Yeboah, E. (2019), "Housing attributes and relative house prices in Ghana". International Journal of Building Pathology and Adaptation, Vol. 37 No. 5, pp. 733-746.

Rothacher, N. (2013), "Maturity of real estate market in Sub-Saharan Africa: a comparison between Tanzania and South Africa". European Real Estate Society (ERES), Vol. 1 No. 1, pp. 1-21.

Silver, M. (2016), "How to better measure hedonic residential property price indexes", working paper WP/16/213, International Monetary Fund (IMF), Washington, DC, November 2016.

Sipan, I., Mar Iman, A. H. and Razali, M. N. (2018), "Spatial–temporal neighbourhood-level house price index". International Journal of Housing Markets and Analysis, Vol. 11 No. 2, pp. 386-411.