THE DRIVERS OF HOUSING PRICES AND THE IMPACT OF SCHOOLS: EVIDENCE FROM GEORGIA

MITRA L. DEVKOTA
PHD, ASSOCIATE PROFESSOR, UNIVERSITY OF NORTH GEORGIA, DAHLONEGA, GA 30597, USA
e-mail:mldevkota@ung.edu

ERIC B. HOWINGTON
PHD, PROFESSOR, VALDOSTA STATE UNIVERSITY, VALDOSTA, GA 31698, USA
e-mail:ebhowington@valdosta.edu

Abstract

This paper attempts to describe the drivers of housing prices in Cumming, Georgia, a rapidly growing suburban area in the southeast of the US. Data from123 single family homes were collected and analyzed using multiple regression methodology. The findings from correlation matrix indicate that the price of the house is positively associated with the number of bedrooms, number of bathrooms, square footage of the house, the lot size, and the number of parking spaces available in the house, and negatively associated with the age of the house. The results from regression analysis suggest that number of bathrooms, square footage, parking spaces, and the dummy variables for Denmark High School, Forsyth Central High School, and North Forsyth High School are statistically significant predictors of the price of the house for Cumming, Georgia. Finally, about 80% of the variation in the prices of the houses is accounted for by our regression model. These findings may have important implications for decision-making by residents, real-estate agents, house buyers and sellers, financial institutions, policymakers, and scholars alike.

Keywords: *Multiple Regression, Correlation, Real Estate, Housing Prices, Dummy Variables*

JEL Classification: *R3*

1. Introduction

The intersection of education and real estate markets represents a dynamic and influential nexus, where the quality of local schools intricately intertwines with the trajectory of housing prices.

Forsyth county, GA, is a rapidly growing area located just north of Atlanta. With 43% growth over the decade from 2010-2020, the county was the fastest growing county in Georgia, and the 15th fastest growing county in the United States. The population of Cumming, GA, the sole incorporated area in the county, grew almost 35% between the 2010 and 2020 censuses. The housing market in the area has also experienced recent rapid growth. According to rockethomes.com, the median home price in Cumming, GA, from November 2022 to October 2023, has increased from $372,000 to $564,600, an increase of over 40% in a single year.

The Forsyth County school system has had to expand to accommodate rapid population growth. There are currently eight high schools in Forsyth County, with five high schools opening since 2007. As communities grow and urbanize, the perceived quality of local schools emerges as a critical factor influencing housing decisions. Homebuyers and investors alike often weigh the educational opportunities available to their families, recognizing the enduring impact of these choices on the overall quality of life. Consequently, an empirical understanding of how school districts affect housing prices assumes heightened significance for residents, policymakers, and scholars alike.

We are not aware of any research article that has analyzed the drivers of housing prices in the growing suburban south of the US after the COVID-19 Pandemic. The purpose of this paper is to fill this void.

The remainder of this study is organized as follows. A literature review of previous empirical studies on housing prices is carried out in section 2. The description of the data and the variables used in the study are presented in section 3. Section 4 contains the methodology used in the study, the empirical results, and their interpretations. Finally, the section 5 concludes the paper with concluding remarks and discussions.

2. Literature Review

Bin, O. (2004) estimated a hedonic price function using a semi-parametric regression and compared the price prediction performance using conventional parametric models utilizing a large data set from 2595 single-family residential home sales between July 2000 and June 2002 from Pitt County, North Carolina. They incorporated data from Geographic Information Systems (GIS) to account for locational attributes of the houses. The author found that the semi-parametric regression outperformed the parametric counterparts in both in-sample and out-of-sample price predictions. They concluded that the semi-parametric model could be useful for measurement and prediction of housing sales prices.

Zietz et al. (2008) studied the determinants of housing prices using a quantile regression approach. Their results show that purchasers of higher-priced homes value certain housing characteristics such as square footage and the number of bathrooms differently from buyers of lower-priced homes. They further document that other variables such as age were also shown to vary across the distribution of housing prices.

Nistor and Reianu (2018) presented a panel data econometric model of the drivers of house prices in the ten largest census metropolitan areas (CMA) in Ontario, Canada, for the years 2001, 2006 and 2011. The authors document that the main drivers of home prices over that time are interest rate, immigration, unemployment rate, household size and income.

Jafari, A., & Akhavian, R. (2019) studied the characteristics that determine the driving forces of housing prices in the USA. Using a data set of 13,771 houses extracted from the 2013 American Housing Survey, they employed a Hedonic Pricing Method (HPM). They also used another data set of 22 houses in the city of San Francisco, CA extracted from Redfin real estate brokerage database to test and validate the model. They found that the main driving force for housing price in the USA is the square footage of the unit, its location, and its number of bathrooms and bedrooms.

Yuan (2019) collected the house sale price in seven major cities in Los Angeles County and constructed a multiple regression model to estimate the factors affect the house price in the real estate market. The author concluded that lot size, number of bathrooms, local school quality, median household income, city population, and school district ratings are significant predictors of the house price.

Tripathi (2019) investigated the macroeconomic determinants of housing prices from a cross country perspective. Using the analysis from random-effect models, the author documented that the rent, price-to-income ratio, price-to-rent ratio, urbanization, per-capita GDP, inflation, the share of population aged 15-64, GDP growth rate, broad money, and real exchange rate had a positive and statistically significant effect on housing prices, while the percentage share of employment in services had a negative effect on housing prices.

Xu and Tang (2014) examined the determinants of housing prices in UK by applying a cointegration approach and its error correction model. Using the quarterly data from 1971 to 2012, the results from the cointegration test showed that construction cost, credit, GDP, interest rate and unemployment rate had a positive impact, while disposable income and money supply had negative impact on housing prices. Furthermore, the error correction model showed that the growth of house prices was affected by the growth of construction cost, credit, interest rate, and disposable income in the short run.

Nghiep and AI (2001) compared the predicted performance of artificial neural networks (ANN) and multiple regression analysis (MRA) for single family houses. They made multiple comparisons between the two data models in which the sample size, the functional specification, and the temporal prediction were varied. The results indicated that the ANN performed better than MRA when a moderate to large sample size was used.

Yusof and Ismall (2012) used multiple regression analysis and its extension, Hedonic regression analysis to study the explained price variation for houses in Malaysia. The authors documented that the locality and the area of the house were the most important drivers of housing prices.

Randeniya et al. (2017) analyzed the housing price data for Sri Lanka using 50 single house sales in Maharagama urban neighborhood area to illustrate the applicability of the hedonic pricing model. Using correlation analysis, the authors found that the design of the house, distance to the local road, infrastructure quality, garden size, number of the bedrooms, and age of the house were the drivers of housing prices.

Bui (2020) analyzed the determinants of apartment prices in Ho Chi Minh city of Vietnam by using multiple regression methodology. They used data of 124 apartments traded during the first six months of 2019 in their analysis. The results from the analysis showed a positive relation of the apartment price to the apartment size, presence of balcony, presence of swimming pool, presence of shopping malls, and periodic rental income, and negative relation of the apartment price to the distance from the city center.

3. Data and variables

In this empirical research, we have used data from 123 single family homes from Forsyth Cumming, Georgia for the year 2023. This data set was collected from real estate research website www.zillow.com. The variables in the data set are the number of bedrooms, number of bathrooms, square footage, lot size, age, number of parking spaces, and the high school assigned to the specific house. These variables are described in Table 1 below. The price ranges from $195,000 to $1,785,000; the number of bedrooms ranges from 1 to 8; the number of bathrooms ranges from 1 to 7; the square footage ranges from 720 to 10,454; the lot size ranges from 0.07 acres to 3 acres; the age of the house ranges from 1 to 66; and the number of parking spaces ranges from 0 to 6. There are eight high schools within the sampled region.

Table 1: Variables used in the study

Variable	Description	Details
Price	Listing price of the houses	Ranges from $195,000 to $1,785,000
Bedroom	Number of bedrooms	Ranges from 1 to 8
Bathroom	Number of bathrooms	Ranges from 1 to 7
Sqft	Floor size in square feet	Ranges from 720 sq. ft. to 10,454 sq. ft.
Lot size	Lot size in square feet	Ranges from 0.07 acres to 3 acres
Age	Age in years	Ranges from 1 to 66
Parking	Number of parking spaces	Ranges from 0 to 6
High School	The name of the high school assigned	Name of High Schools (Eight categories)

The eight high schools used in the study are Alliance Academy for Innovation, East Forsyth High School, Forsyth Central High School, Denmark High School, Lambert High School, North Forsyth High School, South Forsyth High School, and West Forsyth High School.

Here, "High school" is a categorical independent variable with eight categories-Alliance Academy for Innovation (Alliance), East Forsyth High School (East Forsyth), Forsyth Central High School (Forsyth Central), Denmark High School (Denmark), Lambert High School (Lambert), North Forsyth High School (North Forsyth), South Forsyth High School (South Forsyth), and West Forsyth High School (West Forsyth). Alliance Academy for Innovation (Alliance) is used as the reference category of High School. Dummy variables represent the other seven high schools. Thus, the coefficient of the dummy variable "East Forsyth" represents the difference in price between the home in East Forsyth High School region and Alliance Academy for Innovation region, the dummy variable "Denmark" represents the difference in price between the home in Denmark High School region and Alliance Academy for Innovation region, the dummy variable "Lambert" represents the difference in price between the home in Lambert High School region and Alliance Academy for Innovation region, and so on.

4. Methodology and results

In this study, we create the correlation matrix to study the strength of linear relationships between the variables. We then fit a multiple linear regression model to estimate the price of the house (response) using the number of bedrooms, the number of bathrooms, square footage, lot size, the number of parking spaces, and the High School assigned to the home as the independent variables.

The multiple linear regression model is

$$price = \beta_0 + \beta_1 Bedroom + \beta_2 Bathroom + \beta_3 sqft + \beta_4 Lotsize + \beta_5 Age + \beta_6 Parkings + \beta_8 Denmark + \beta_9 East\ Forsyth + \beta_{10} Forsyth\ Central + \beta_{11} Lambert + \beta_{12} North + \beta_{13} South + + \beta_{14} West + \epsilon \quad (1)$$

The correlation between the variables is summarized in the correlation matrix given below in Table 2. The figures in the correlation matrix suggest that, as one would expect, there are positive linear associations between the price of the house and number of bedrooms, price of the house and number of bathrooms, price of the house and square footage of the house, price of the house and its lot size, and price of the house and number of parking spaces available in it, and a negative linear association between the price of the house and its age. The correlation matrix also suggests a very strong and positive linear associations ($r > 0.70$) between number of bedrooms and number of bathrooms and number of bedrooms and square footage. The correlation coefficients between the remaining pairs of the independent variables are within the acceptable range ($r < 0.70$). The strong correlations between the independent variables indicates a possibility of a multicollinearity issue in the regression model (Anderson et al., 2014; Bowerman et al., 2014).

Table 2. Correlation Matrix

	Price	Bedroom	Bathroom	Sqft	LotSize	Age	Parkings
Price	1.00	0.69	0.84	0.69	0.16	-0.31	0.25
Bedroom	0.69	1.00	0.75	0.72	0.19	-0.29	0.18
Bathroom	0.84	0.75	1.00	0.66	0.15	-0.36	0.18
Sqft	0.69	0.72	0.66	1.00	0.16	-0.18	0.15
LotSize	0.16	0.19	0.15	0.16	1.00	0.17	0.0035
Age	-0.31	-0.29	-0.36	-0.18	0.17	1.00	-0.24
Garages	0.25	0.18	0.18	0.15	0.0035	-0.24	1.00

Table 3. Parameter Estimates Table for multiple regression model

Parameter Estimates					
Term	Estimate	SE	t-Ratio	Prob>|t|	VIF
Intercept	-36314.98	78869.59	-0.46	0.6461	.
Bedroom	-8066.345	20828.88	-0.39	0.6993	3.2036
Bathroom	165771.78	18184.31	9.12	0.0001	2.8480
Sqft	42.491818	13.52752	3.14	0.0022	2.4380
LotSize	42277.609	39124.8	1.08	0.2823	1.2479
Age	-766.0479	1168.92	-0.66	0.5136	1.6829
Parkings	31385.012	17952.94	1.75	0.0832	1.1348
Denmark	107858	48378.65	2.23	0.0278	1.7422
East Forsyth	-41610	35925.41	-1.16	0.2493	1.4092
Forsyth Central	-82466.6	31123.25	-2.65	0.0093	1.3098
Lambert	-29909.18	61947.17	-0.48	0.6302	2.0196
North Forsyth	-82503.52	36636.94	-2.25	0.0263	1.5126
South Forsyth	-6428.313	32047.15	-0.20	0.8414	1.2609
West Forsyth	28258.901	42004.81	0.67	0.5025	1.4584
RSquare=0.7989, RSquare Adjusted=0.7650, F(13, 109)=31.5, P(F>31.5) <2.2e-16					

The estimated multiple linear regression equation is

$$\widehat{price} = -36314.98 - 8066.345Bedroom + 165771.78Bathroom + 42.491818sq.ft + 42277.609Lotsize - 766.0479Age + 31385.012Parkings + 107858Denmark - 41610.41East - 82466.6Forsyth - 29909.18Lambert - 82503.52North - 6428.313South + 28258.901West$$

Our model estimates that coefficient of slope for bedrooms is -8066.345. It means that, on average, we'd expect the price to be lower by $8,066.345 for each bedroom in the house, keeping the other variables fixed. This interpretation is counterintuitive to the common notion that price of a home increases with the increase in the number of bedrooms. The coefficient of slope for bathroom means that, on average, we'd expect the price to be higher by $165,771.78 for each bathroom in the house, keeping the other variables fixed. The coefficient of slope for square footage tells us that on average, we'd expect the price to be higher by $42.49 for each additional square footage in the house, keeping the other variables fixed. The coefficient of slope for lotsize tells us that, on average, we'd expect the price to be higher by $42,277.61 for each additional acre of lotsize, keeping the other variables fixed. The coefficient of slope for age tells us that on average, we'd expect the price to decrease by $706.05 for one year increase in the age of the house, keeping the other variables fixed. The coefficient of slope for number of parking spaces tells us that, on average, we'd expect the price to be higher by $31,385.01 for each additional parking space in the house, keeping the other variables fixed.

The coefficients of slopes for the dummy variables for High School [Denmark] tells us that, on average, we expect the price of house in Denmark Highschool region is $107,858 more than a house in Alliance School region, keeping the other variables fixed. The coefficient of slope for High School [East] tells us that, on average, we expect the price of house in East Forsyth region is $41,610 lower than a house in Alliance School region, keeping the other variables fixed. The coefficient of slope for High School [Forsyth] tells us that, on average, we expect the price of house in Forsyth Central

School High School region is $82,466.6 lower than a house in Alliance School region, keeping the other variables fixed. The coefficient of slope for High School [Lambert] tells us that, on average, we expect the price of house in Lambert High School region is $29,909.18 lower than a house in Alliance School region, keeping the other variables fixed. The coefficient of slope for High School [North] tells us that, on average, we expect the price of house in North Forsyth High School region is $82,503.52 lower than a house in Alliance School region, keeping the other variables fixed. The coefficient of slope for High School [South] tells us that, on average, we expect the price of house in South Forsyth High School region is $6,428.31 lower than a house in Alliance School region, keeping the other variables fixed. The coefficient of slope for High School [West] tells us that, on average, we expect the price of house in West Forsyth High School region is $28,258.90 higher than a house in Alliance School region, keeping the other variables fixed. The coefficient of determination of our regression model is 0.7989. This means that about 80% of the variation in the listing prices of the houses is accounted for by our regression model.

As Wooldridge (2011) notes, multicollinearity would be of serious concern if the variance inflation factor (VIF) for the independent variables exceed 10. To ensure that our model is free of multicollinearity issues, we have calculated the variance inflation factor (VIF) for each of the independent variables. These are reported in the far-right column of the parameter estimates table of the regression model in table 3. The figures there indicate that all the VIFs for the independent variables are well below the acceptable range. This shows that our model is free of multicollinearity.

5. Conclusions and Discussion

In this paper, we have analyzed the housing prices data on 123 single family houses for the city of Cumming, Georgia, a rapidly growing city located just north of Atlanta, USA. The data were collected from the realtor's website www.zillow.com. We have used the listing price of the house as the dependent variable and the number of bedrooms, the number of bathrooms, square footage, lot size, the number of parking spaces, and the High School assigned to the home as the independent variables.

Some of the results of the regression model are surprising. Several of the regression coefficients are not statistically significant: number of bedrooms, lot size, age, and number of parking spaces. In addition, the coefficients for the schools East Forsyth, South Forsyth, West Forsyth, and Lambert are not statistically significant. The statistically significant drivers of house price were square footage, number of bathrooms, and dummy variables for the schools Denmark (positive), Forsyth Central (negative), and North Forsyth (negative). Forsyth Central, built in 1955, and North Forsyth, which opened in 1994, are two of the older schools in the area. Denmark, on the other hand, is relatively new, having opened in 2016. Newer schools are often more popular than older schools, possibly explaining the signs of the coefficients of these three dummy variables.

Consistent with Boarnet and Chalermpong (2003), our research shows that the coefficient of the number of bedrooms of the houses is negative. A possible explanation could be that the higher-priced, large luxury homes have relatively few bedrooms. The price of such houses could be more influenced by dwelling size rather than by the number of bedrooms. Therefore, the negative coefficient on the number of bedrooms is not surprising. Further research may reveal the possible explanations. However, this research contradicts the common notion and the results from the studies carried out by other researchers. We also find that an additional bathroom increases the price of the house by $165,772 on average, keeping the other variables fixed. About 80% of the variation in the prices of the houses is accounted for by our regression model. These findings may have important implications for decision-making by residents, real-estate agents, house buyers and sellers, financial institutions, policymakers, and scholars alike.

6.References

[1]. **Anderson, D., Shoesmith, E., Sweeney, D., Anderson, D. and Williams, T.A.,** *Statistics for Business and Economics 3e*. Cengage, 2014;

[2]. **Bin, O.,** A prediction comparison of housing sales prices by parametric versus semi-parametric regressions. *Journal of Housing Economics*, *13*(1), 68-84, 2004;

[3] **Boarnet, M. G., & Chalermpong, S.**, New highways, house prices, and urban development: A case study of toll roads in Orange County, CA. *Housing policy debate*, *12*(3), 575-605, 2001;

[4]. **Bourassa, S., Cantoni, E., & Hoesli, M.**, Predicting house prices with spatial dependence: a comparison of alternative methods. *Journal of Real Estate Research*, *32*(2), 139-160, 2010;

[5]. **Bowerman, B.L., O'Connell, R.T. and Murphree, E.**, *Business statistics in practice*. McGraw-Hill/Irwin, 2014;

[6]. **Bui, T.**, A study of factors influencing the price of apartments: Evidence from Vietnam. *Management Science Letters*, *10*(10), pp.2287-2292, 2020;

[7]. **Jafari, A., & Akhavian, R.**, Driving forces for the US residential housing price: a predictive analysis. *Built Environment Project and Asset Management*, *9*(4), 515-529, 2019;

[8]. **Newsome, BA and J. Zeitz**, "Adjusting Comparable Sales Using Multiple Regression Analysis-The Need for Segmentation," The Appraisal Journal, January, 1992, pp.129-135, 1992;

[9]. **Nghiep, N. and Al, C.**, Predicting housing value: A comparison of multiple regression analysis and artificial neural networks. *Journal of real estate research*, *22*(3), pp.313-336, 2001;

[10]. **Nistor, A., & Reianu, D**., Determinants of housing prices: evidence from Ontario cities, 2001-2011. *International Journal of Housing Markets and Analysis*, *11*(3), 541-556, 2018;

[11]. **Randeniya, T. D., Ranasinghe, G. and Amarawickrama, S.**, A Model to Estimate the Implicit Values of Housing Attributes by Applying the Hedonic Pricing Method. International Journal of Built Environment and Sustainability, IJBES 4(2) 2017, 113-120, 2017;

[12]. **Sharpe, N.R., De Veaux, R.D. and Velleman, P.F.,** *Business statistics*. Pearson Education, 2012;

[13]. **Tripathi, S.**, Macroeconomic determinants of housing prices: a cross country level analysis, 2019;

[14] **Wooldridge, M.**, Modern econometrics, 2011;

[15]. **Xu, L., & Tang, B.,** On the determinants of UK house prices. *International Journal of Economics and Research*, *5*(2), 57-64, 2014;

[16]. **Yuan, L.**, A Regression Model of Single House Price In LA, 2019;

[17]. **Yusof, A. and Ismail, S.**, Multiple Regressions in Analyzing House Price Variations. Communications of the IBIMA, 1–9, 2012;

[18]. **Zietz, J., Zietz, E. N., & Sirmans, G. S**., Determinants of house prices: a quantile regression approach. *The Journal of Real Estate Finance and Economics*, *37*, 317-333, 2008;

[19]. ******https://www.rockethomes.com/real-estate-trends/ga/cumming

[20]. ******https://www.zillow.com