

Kriging House Prices: A Predictive Model for Travis County

Problem Formulation

The goal of this study is to create a predictive house pricing model for Travis County, Texas through the use of Kriging. Much of the real estate industry bases their price valuing on collected historical and present data, however, many do not take into consideration the geospatial properties of house and pricing data (Montero and Larraz, 2006). The hypothesis of this study is that Kriging will create an accurate spatial interpolation of house prices in Austin based on the spatial autocorrelation properties of real estate. The study is founded on Tobler's First Law of Geography *that everything is related to everything else but nearer things are more related than distant things*.

Data Sources and Processing

Although there are plenty of house price information, there are no readily available house price data for Travis County that can easily be imported into ArcGIS. Therefore, I acquired the house price data for 218 homes from the real estate website, Zillow.com, which contains a database for the current price for homes in a particular area. Because only the addresses for the houses were available on the websites, I had to import the addresses with their prices into an excel table and format the addresses properly. The formatting would allow the addresses to be able to run through ArcGIS's address locator and convert them to geographic coordinates (refer to Figure 1). The Travis County and Tract shapefiles were obtained from the US Census Bureau's website. The year of the shapefile boundary was of the 2010 Census; the administrative boundaries do not usually change until the subsequent Census (refer to Figure 2). All of the data was projected into Texas State Plane Zone 3 coordinate system.

						Geocoded Adresses	
Street Address	City	State	ZIP	House_Price		Latitude	Longitude
4014 Rosedale Ave	Austin	Texas	78793	625000	>	30.3113	-97.7457
1212 Tartan Ln	Austin	Texas	78753	450000	>	30.35592	-97.6784
5707 Boulder Crk	Austin	Texas	78724	164000	>	30.30619	-97.6551

Fig. 1: Formated addresses ready to be imported into ESRI's Address Locator tool.

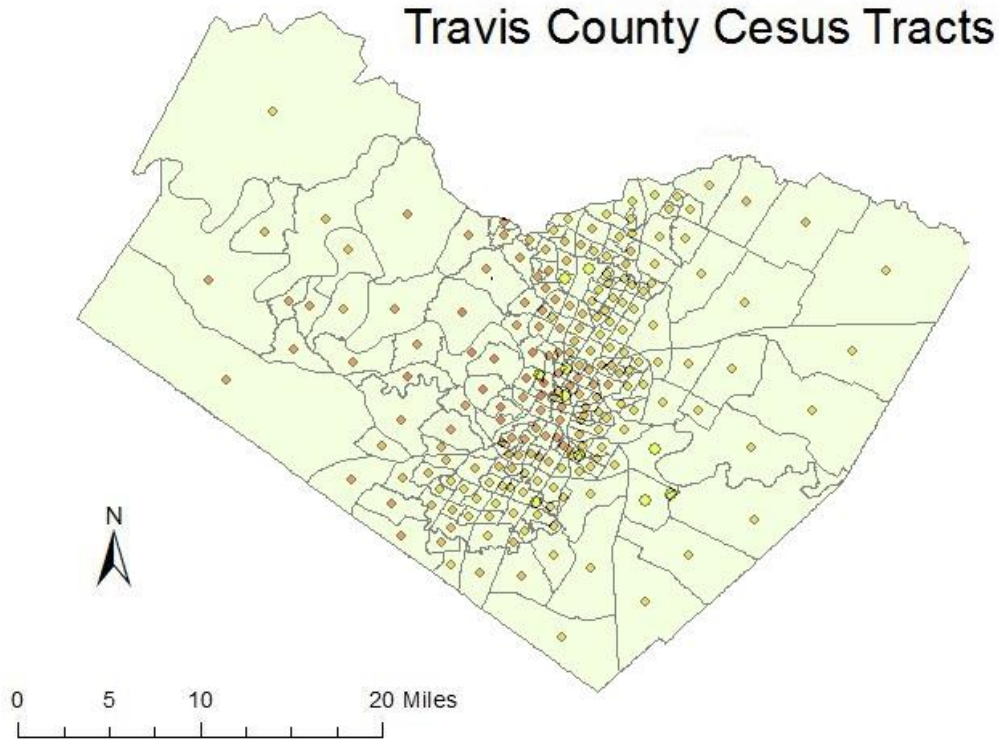


Fig 2 Travis County 2010 Census Tracts and the 218 sampled price homes

Methods

A sample of house prices from Travis County were collected from Zillow.com in order to create this predictive model, which was then compared to a different sample of house prices in Travis County from the same website and determine how similar or dissimilar the actual prices are to the estimated model based on location. Once the addresses and their corresponding house prices were converted into geographic point shapefiles, I proceeded to determine whether there were any ostensible patterns through the Data Exploration tools. A preliminary

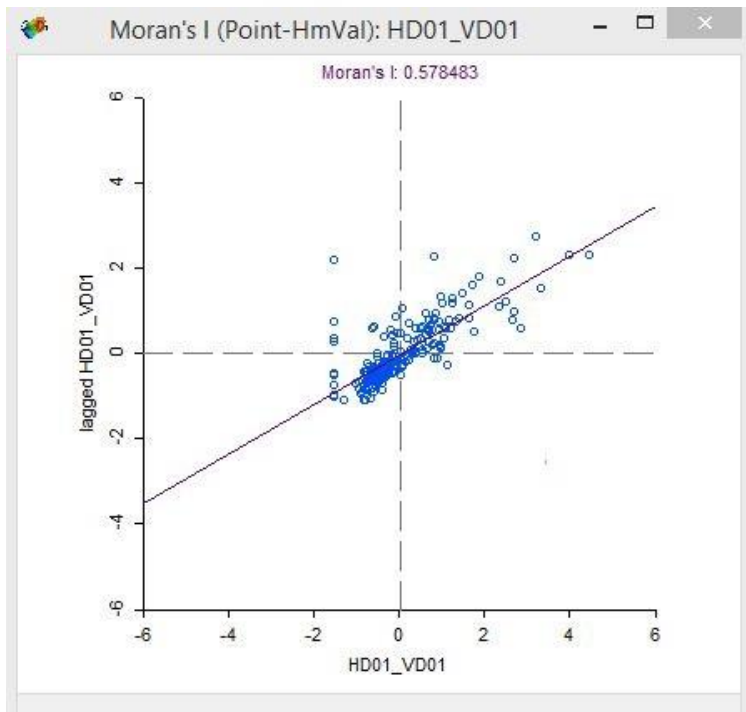


Fig. 3: Moran's I Scatterplot showing highly autocorrelated house prices in Travis County.

was to determine whether housing prices were autocorrelated using Moran's I spatial autocorrelation tool in GeoDa, a free software for spatial statistics. The value of homes in Travis County appear to be correlated, therefore, higher priced homes are near higher priced homes and lower priced homes are near other lower priced homes (refer to Figure 3).

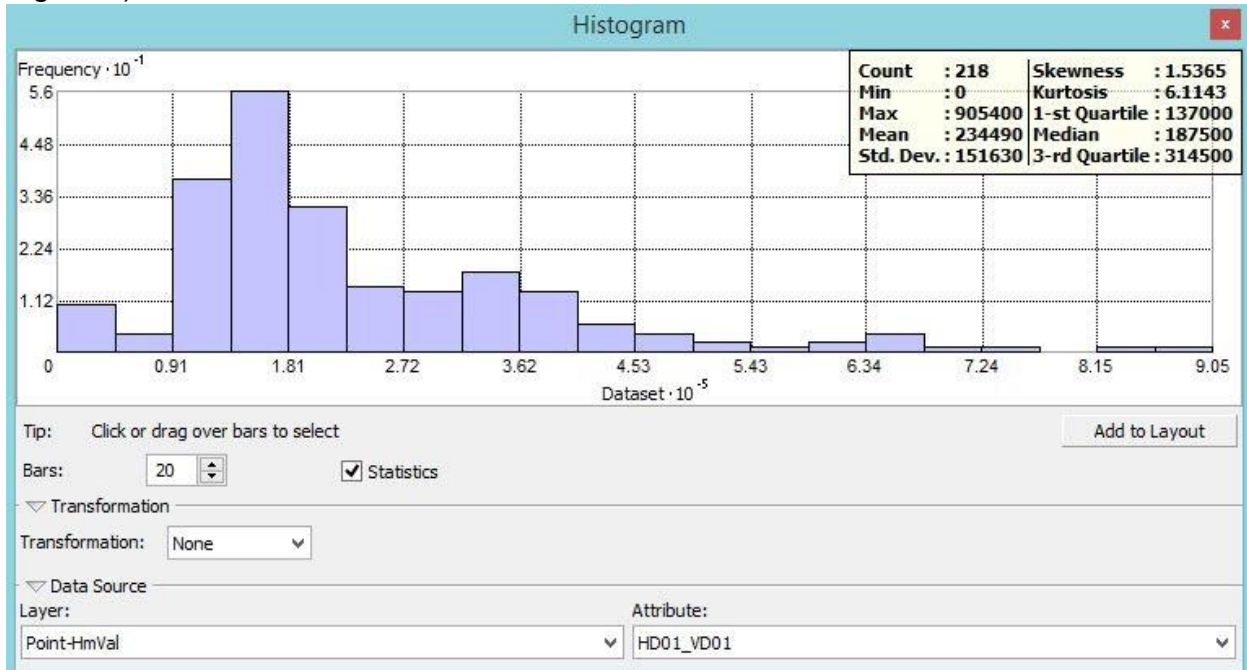


Fig. 4: Histogram of price value distribution

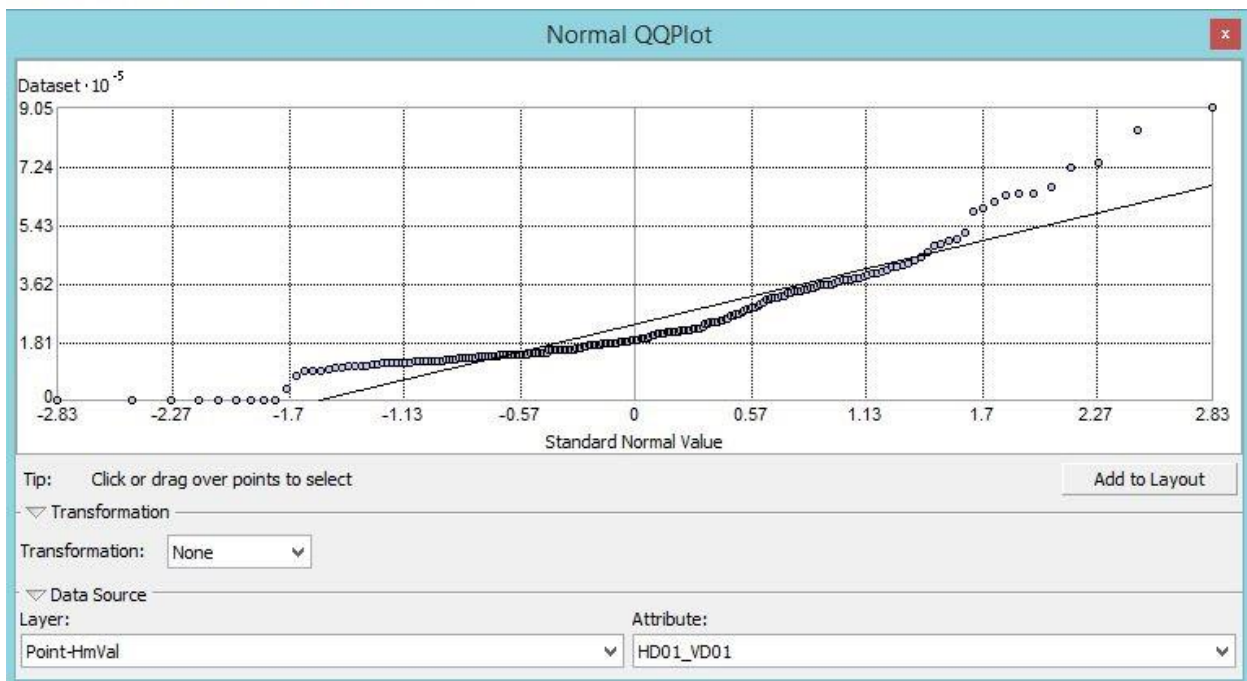


Fig 5: QQ Plot for distribution measure

The first step was to look at the distribution of the histogram. The distribution appears to be skewed towards the right or higher end of prices (refer to Figure 4). By looking at the QQ Plot, it can also be seen that the slope increases from left to right which indicates skewing to the right (refer to Figure 5). The trend analysis shows that the North to South direction is more correlated than the East-West direction (refer to Figure 6).

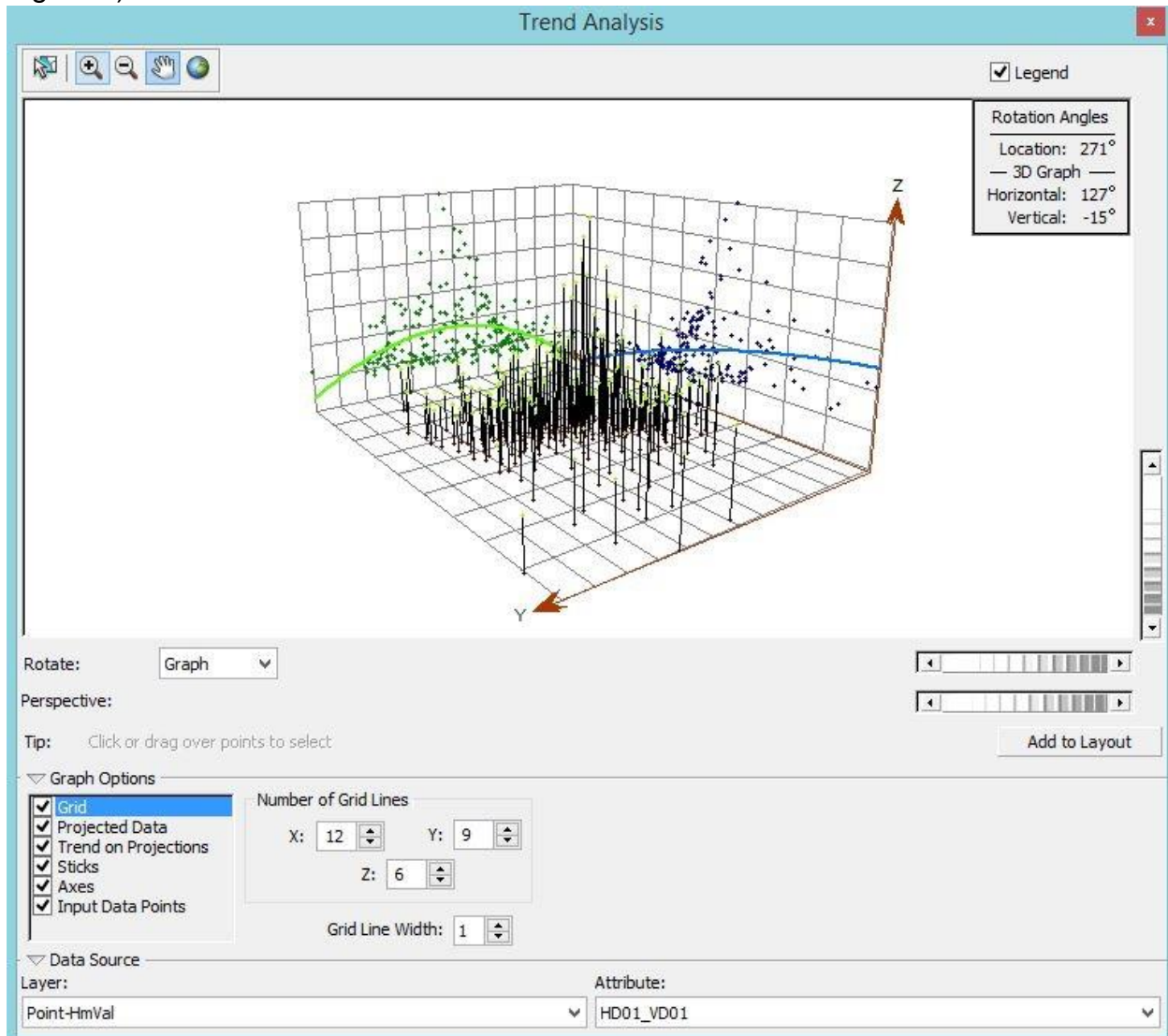


Fig. 6: Trend Analysis observing directional dependency of home values

The next step was to examine the semivariogram - the semivariogram exhibited a strong autocorrelation of home values, as well as a directional distribution; the distribution being slightly northwest to southeast (Refer to Figure 7a and 7b). The nugget which represents the precision of measurements is very small while the range

and sill are constant over a significant distance.

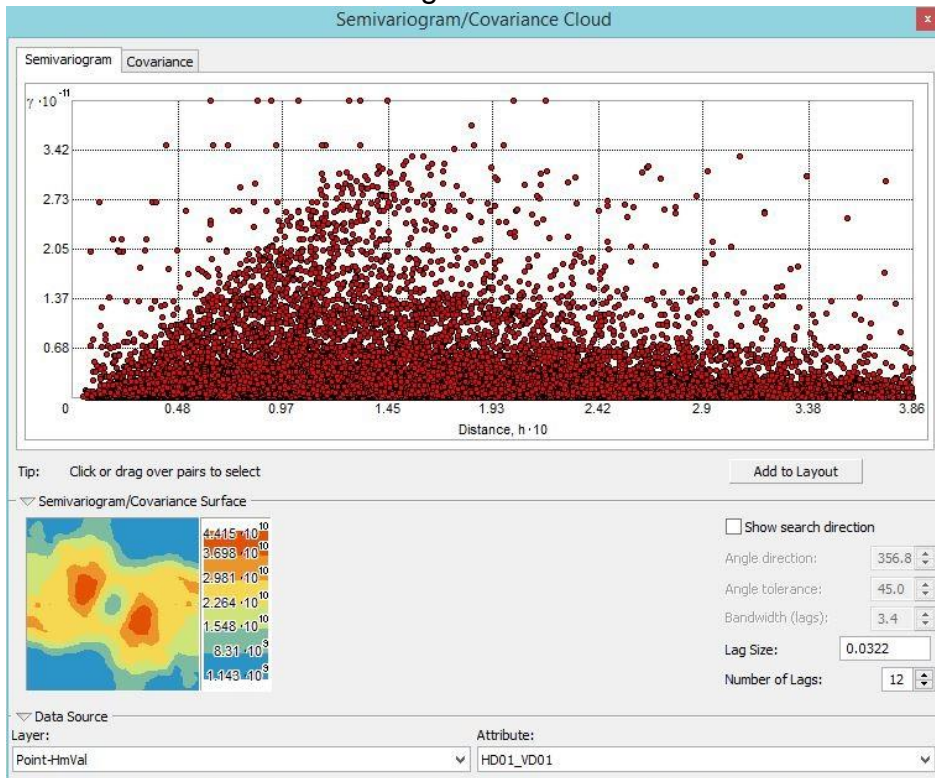


Fig 7: Semivariogram demonstrating autocorrelated values; small nugget and moderate range and sill.

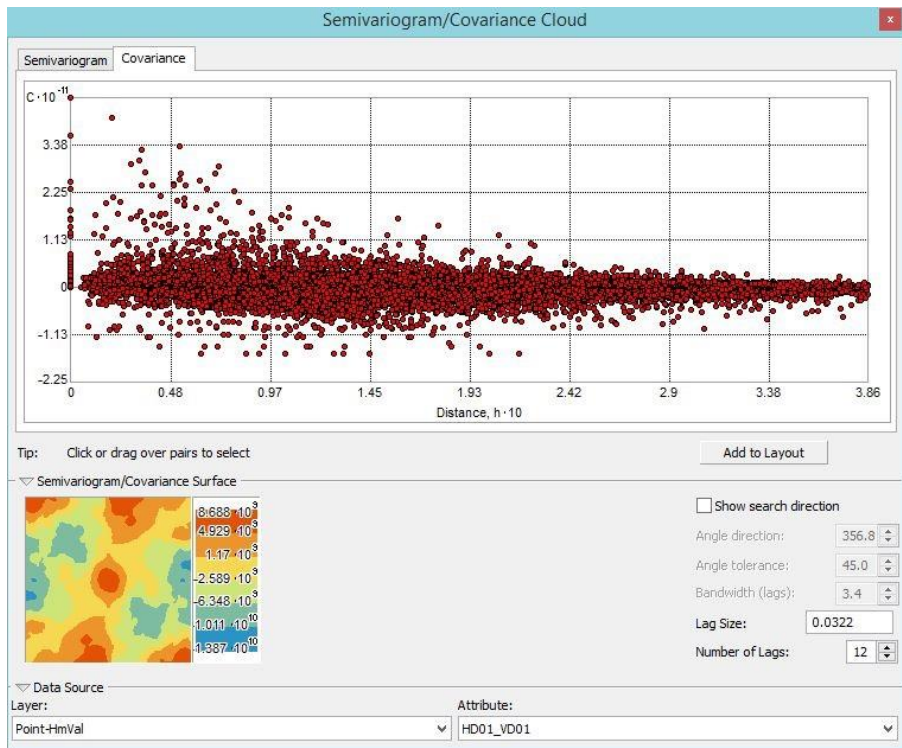


Fig 7b: Covariance Cloud

Simple Kriging was used to create a raster of a predictive model. As observed from data exploration tools there appeared to be a southwest to northwest trend. The resulting raster created from the Geostatistical Analyst depicts the same directional distribution (refer to Figure 8). Most of the higher income households are located in West Austin while the Lower Income households are located in East Austin. Therefore, it was expected that the value of homes in West Austin are much higher than anywhere else. There are two very distinct clustering of values or autocorrelation in Travis County, in the mid-west and mid-east with a neutral area in between the two.

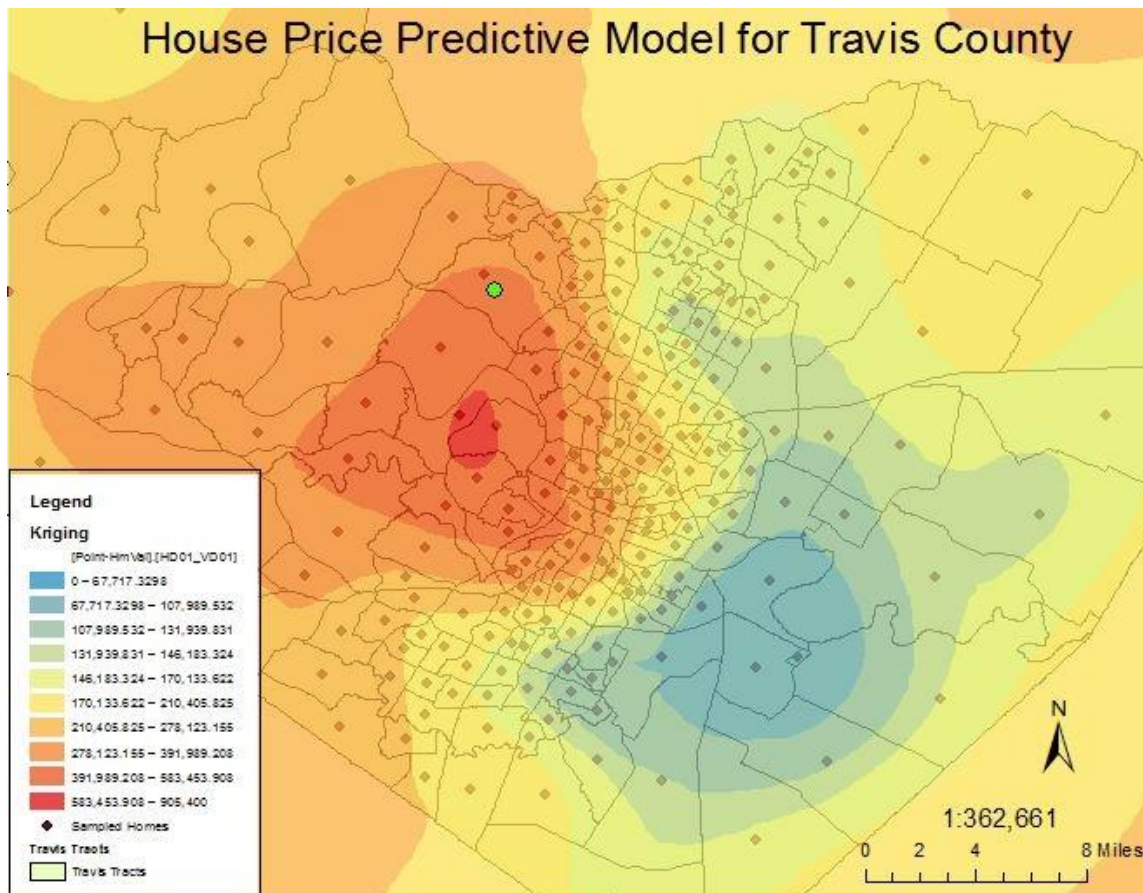


Fig 8. Predictive model of Travis County home prices derived from Simple Kriging

The following step was to determine whether the Kriging predictive model was accurate by collecting a small but different sample of house prices from Zillow.com and compare them. The comparison was fairly straight forward, the location of the homes and prices was collected and geolocated in Travis County.

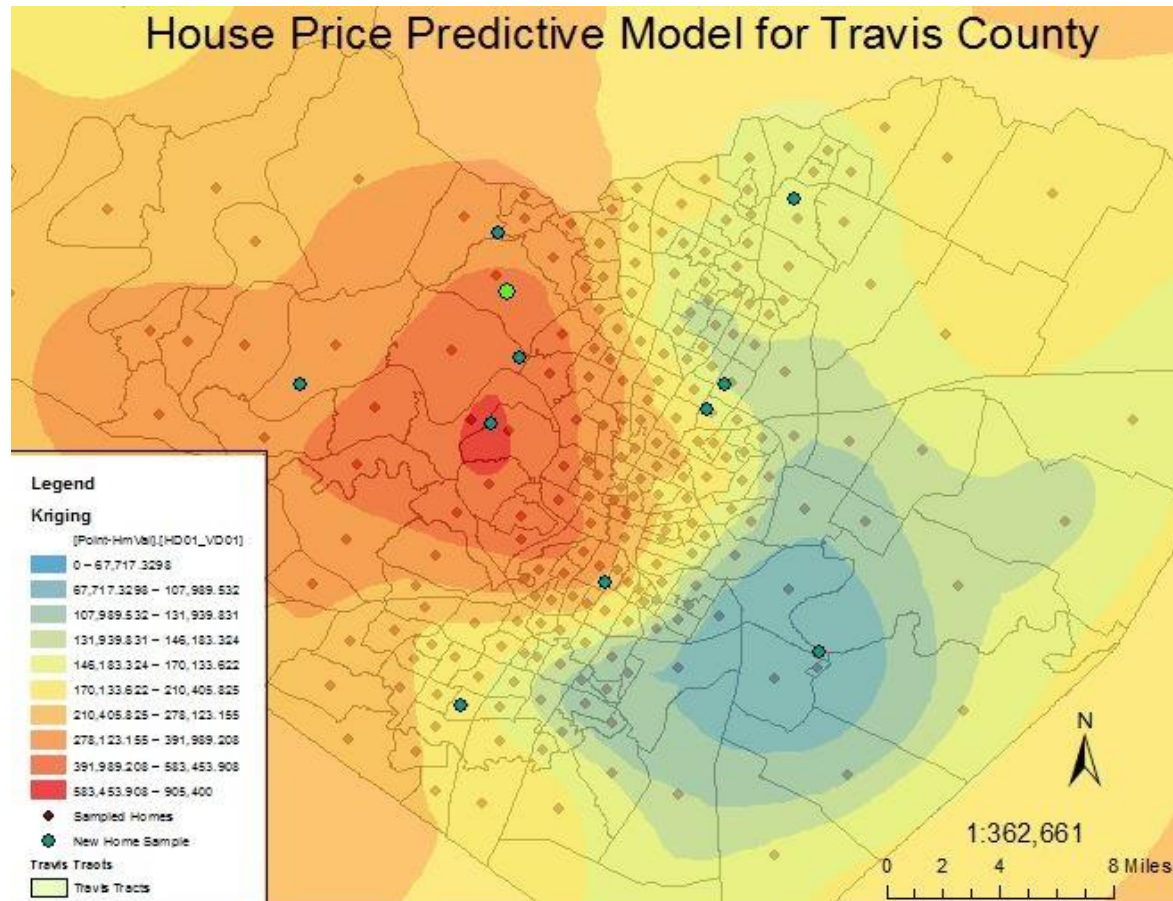


Figure 9a: Sample of new home prices overlaid the price prediction raster created by Kriging. The actual price of the recently sampled homes will be compared with the underlying value of the raster created through Kriging (refer to Figure 9a). Figure 9 shows 10 sampled homes and their corresponding price values, the estimated price value created by Kriging, the individual percentage error, and mean percentage error. There was a combination of overestimation and underestimation of home prices as well as a dispersion of percentage error. The majority of the predictions were underestimations. Perhaps a bigger primary sample should have been selected as well as the comparison sample. Another possible reason for this underestimation could be amended by creating local zone in which to perform Kriging. The mean percentage error was 46 which is somewhat higher than expected. Nevertheless, this type of study is a good start to further expand the use of Kriging for urban development issues such as house pricing.

and estimation.

Address	City	Long	Lat	Estimated Price	Actual Price	Percentage Error	Absolute Error
7208 Bennet Ave	Austin	-97.70028	30.33195	148528	155500	4.483601286	0.044836013
6000 Shepard Mo	Austin	-97.79566	30.35786	503967	200000	151.9835	1.519835
8305 Loralinda Dr	Austin	-97.69138	30.34445	148287	310000	52.16548387	0.521654839
2020 S. Congress A	Austin	-97.75223	30.24339	248627	140000	77.59071429	0.775907143
9618 Old Lampasa	Austin	-97.80714	30.42174	349991	598000	41.47307692	0.414730769
9410 Castlewood I	Austin	-97.82637	30.18062	180267	375000	51.9288	0.519288
2906 Ellon Rd	Del Valle	-97.64359	30.20796	76338	60000	27.23	0.2723
15133 Donna Jane	Pflugervil	-97.65634	30.43928	156271	188000	16.87712766	0.168771277
12217 Palisades Pl	Austin	-97.90791	30.34418	350851	550000	36.20890909	0.362089091
2800 Waymaker W	Austin	-97.811	30.32469	601553	585000	2.82957265	0.028295726
						Mean Percentage Error:	46.27707858

Fig 9: Error Calculation of the of 10 additional houses and their corresponding prices

Conclusion

The application of geostatistics such as Kriging to the field of real estate is becoming more crucial than ever after the housing and mortgage crisis. Spatial phenomenon although different in context is similar in characteristic over space, which is the reason we can apply geostatistics like Kriging, intended for geology, to urban practices (Martinez, Lorenzo, and Rubio, 2000). In principle, the behavior of house values is an appropriate candidate for the use of Kriging to create predictive models due to its spatial autocorrelation. However, my hypothesis only answered half of the question that real estate is a phenomenon that can be spatially quantified, the other half regarding the accuracy of the predictive model has yet to be improved. Although there are several procedures regarding the data collection and methods that can be improved, once further developed, Kriging is a practices that can great improve housing policy and valuation. Because the characteristics of real estate is almost always universal, this type of predictive model not only applied to Travis County but to any other location in the US and the world, of course allowing for the variances in space due to the particular housing market.

Sources

Larraz, B. (2006). Estimating Housing Price: Kriging the Mean. *International Advances in Economic Research*, 12(419), 419-419.

Rubio, N. G. (2000). Kriging methodology for regional economic analysis: Estimating the housing price in Albacete. *International Advances in Economic Research*, 6(3), 438-450.

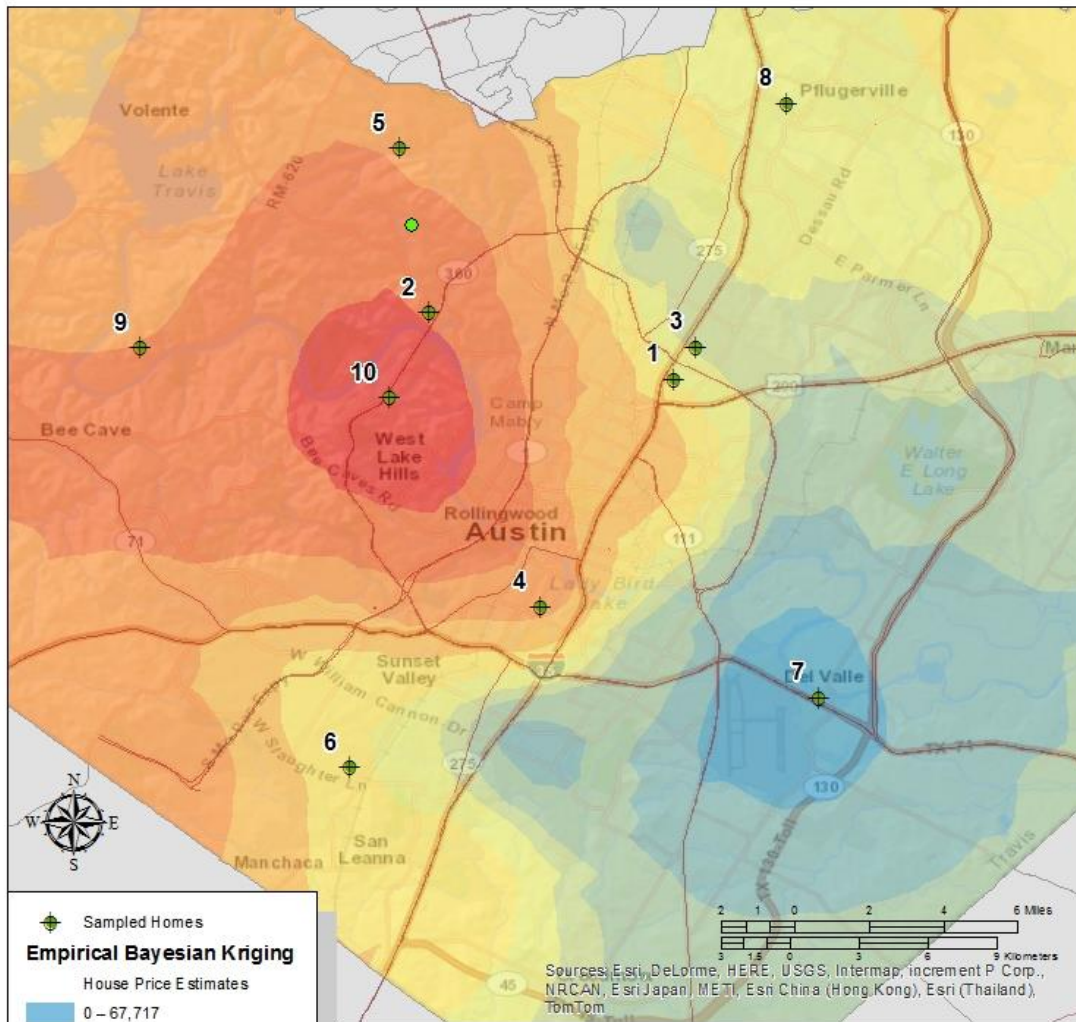
United States Census Bureau; 2010 TIGER/Line Shapefiles
<http://www.census.gov/cgi-bin/geo/shapefiles2010/main>

Zillow Real Estate Network; Travis County homes for sale
http://www.zillow.com/homes/travis-county_rb/

A Geostatistical Approach to Real Estate

Predictive Model for House Prices in Travis County, TX

Spatial interpolating using the Kriging method



Kriging is an interpolation technique in which the surrounding measured values are weighted to derive a predicted value for an unmeasured location. Weights are based on the distance between the measured points, the prediction locations, and the overall spatial arrangement among the measured points.

State Plane Texas Central Datum: North American 1983
Lambert Conformal Conic Unit of Measurement: U.S. Foot
Source: US Census Bureau, Real Estate, Zillow.com, ESRI
Created by Edgar Estrada, April 19, 2014

