

UC Berkeley

International Conference on GIScience Short Paper Proceedings

Title

Local variation in hedonic house pricing in Hanoi, Vietnam: a spatial analysis of status quality trade-off (SQTO) theory

Permalink

<https://escholarship.org/uc/item/0qh4t98s>

Journal

International Conference on GIScience Short Paper Proceedings, 1(1)

Authors

Comber, Alexis
Harris, Paul
Nguyen, Quan
et al.

Publication Date

2016

DOI

10.21433/B3110qh4t98s

Peer reviewed

Local variation in hedonic house price, Hanoi: a spatial analysis of SQTO theory

A. J. Comber¹, P. Harris², T. Q. Nguyen³, P. K. Chi⁴, H. Tran⁴ and H. H. Phe⁵

¹ School of Geography, University of Leeds, LS2 9JT, UK
Email: a.comber@leeds.ac.uk

² Rothamsted Research, North Wyke, Okehampton, EX20 2SB, UK
Email: paul.harris@rothamsted.ac.uk

³ National University of Civil Engineering, Hanoi, Vietnam
Email: quannt@nuce.edu.vn

⁴ GeoViet Consulting Co. Ltd, Hanoi, Vietnam
Email: chi.geoviet@gmail.com

⁵ Vinaconex R&D, Pham Hung St, Hanoi City, Vietnam
Email: hoanghuuphe@googlemail.com

Abstract

This paper applies a local analysis to model and predict hedonic house price in Hanoi, Vietnam. It applies a locally compensated geographically weighted ridge regression to data survey data collected to support the Status Quality Trade Off theory proposed by Phe and Wakely (2000). This has an inherently local flavour is therefore suitable for local statistical approaches such as GWR (Brunsdon *et al.*, 1996). The locally compensated ridge regression accounts for the observed local collinearity. The results provide a spatially nuanced model of status poles associated with areas of desirable housing. Some key areas for future work are suggested.

1. Introduction

The SQTO (Status Quality Trade Off) theory (Phe and Wakely, 2000) seeks to explain urban development dynamics and links the temporally dynamic process of urban transformation (*housing status*) with the relatively permanent character of the physical environment (*dwelling quality*). In doing so, SQTO models changes the relationship between residential location choices, for example the social desirability of a particular area for different social groups, and housing physical characteristics such as floor area, number of bathrooms, number of storeys, etc.

SQTO has at its core the idea that desirable residential location patterns are explained by polar structures, where one or several *status poles* represent the highest points of certain kinds of social status held by the residential population relating to particular notions of wealth, political power, business, culture, ethnicity, education specific to particular social or geo-demographic groups. It is underpinned by 2 considerations: that status poles are local and that they vary for different social groups and as such SQTO theory has at its core the essence of Tobler's 1st law of Geography (Tobler, 1970), suggesting the suitability of explicitly local frameworks of analysis such geographically weighted (GW) approaches (Brunsdon *et al.*, 1996). This is in contrast to mainstream economic residential location theories that explain residential location patterns using classic utility maximization theories and budget constraints (Alonso, 1964; Fujita, 1989). In these, access to place of work to minimise commuting costs, for example, is one of the main factors explaining residential location choices for lower-income households and higher-income households choose larger properties in the suburbs

with choices moderated by factors such as school quality (Kim *et al.*, 2005), crime (Weisbrod *et al.*, 1980) and access to other amenities (Parkes *et al.*, 2002).

This research 1) applies a geographically weighted regression to model local variations factors predicting hedonic house price in Hanoi, and 2) applies a locally-compensated ridge regression to create a spatially distributed model of house prices in Hanoi.

2. Methods

A detailed house survey of nearly 1,000 households in Hanoi was undertaken in 2014 collecting data for variables, reflecting both physical (tangible) and social (intangible) factors related to housing choices. In this study 13 variables were selected as input to analyses as listed in Table 1.. The data were cleaned to remove NULL variables resulting in 633 data points whose locations are shown in Figure 1. The dependent variable in this case is *HPRICVND* - the house price in millions of Vietnamese Dong. The spatial distribution of this data is reasonable, providing good coverage of the study area.

Table 1. The SQTO variables. Type = 1 indicates a tangible variable related to dwelling quality, Type = 2 indicates an intangible variable related to housing status.

Variable	Description	Type
HPRICVND	Price of house in Millions of Vietnamese Dong	Target
AIRCON	Air-Conditioner (Yes, No)	1
GFA	Total floor area (incl. mezzanine) (m ²)	1
PLOTAREA	Total plot area (m ²)	1
SHOPFRNT	Shop Front (Yes, No)	2
PLUMBING	Plumbing Quality (Good, Other)	1
HOUSEGRADE	Permanent, Other	1
CAR	Car ownership (Yes, No)	2
CENTDISR	Measured distance to Centre District	2
DISCENDI	Perceived travel time to the Centre District	2
EDYEARS	Time in education of the interviewee (years)	2
OCCUP_PRIVBIZ	Private Business owner (1=Yes, 0=No)	2
SCHOOQLT	School Quality (Good, Other)	2
STRTYPE	Type of street (Business, Residential)	2

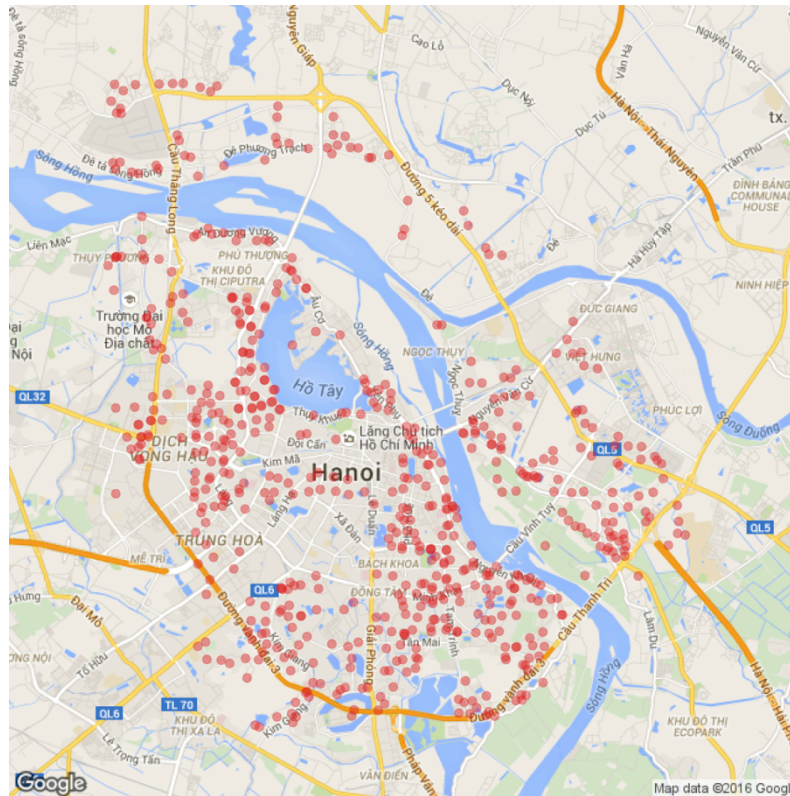


Figure 1. The study area and the 633 survey data points, shaded with a small transparency term and a Google Maps background.

Geographically weighted approaches uses a moving kernel and data under the kernel are used to make a *local* calculation of some kind, such as a regression. The data are weighted by their distance to the kernel centre and local variable collinearity was tested for. The analysis consisted of the following stages:

1. Collation of the optimum bandwidth for the geographically weighted model;
2. An exploratory GW analysis using GW correlations, a GWR fit and various GWR collinearity diagnostics.
3. The calibration of a locally-compensated ridge regression model;

A full description of robust geographically weighted analysis can be found in Gollini *et al.* (2013). In this study a bi-square function was applied and result of the weighting means that data nearer to the kernel centre make a greater contribution to the estimation of local regression coefficients and a weighted least squares regression model is constructed at each regression point i . Here, an optimum kernel bandwidth for GWR can be found by minimising a model fit diagnostic and this case a leave-one-out cross-validation (CV) score (Bowman 1984; Brunson *et al.* 1996) was used under a bi-square kernel. The standard GWR model is:

$$y_i = \beta_{i0} + \sum_{k=1}^m \beta_{ik} x_{ik} + \epsilon_i \quad (\text{Eqn 1})$$

where y_i is the dependent variable at location i , x_{ik} is the value of the k^{th} independent variable at location i , m is the number of independent variables, β_{i0} is the intercept term at location i , β_{ik} is the local regression coefficient for the k^{th} independent variable at location i and ϵ_i is the random error at location i .

Details on how such a global ridge regression model is adapted to a GW form to provide the locally-compensated ridge GWR model can be found in Brunson *et al.* (2012); Lu *et al.* (2014); Gollini *et al.* (2013). These studies also describe associated local collinearity diagnostics for GWR, such as localised correlations amongst pairs of predictors, local VIFs

for each predictor, local variance decomposition proportions (VDPs); and cross-product matrix Condition Numbers (CNs).

3. Initial Results and Discussion Points

A GWR collinearity diagnostic procedure was run which returns CNs describing local correlations amongst pairs of predictors. CNs greater than 30 suggest evidence of local collinearity. In this case the CNs were found to range from 12.78 to 941.76, with a mean CN of 38.78, suggesting that there are areas with very high levels of collinearity present in the study area when the data are considered in the context of local analyses such as GWR. The CN values for each data point are shown in Figure 2 and show that regions of strong collinearity exists in areas around the central portion of the study area.

In order to take account of the observed local collinearity, a locally compensated ridge GWR was calibrated and the analysis was conducted over a 200m grid of points covering the study area. The spatially distributed variables were combined using the spatially distributed coefficient estimates arising from the locally compensated GWR in order to construct the model of house price in Figure 3. This shows distinct patterns of predicted house prices reflecting not only the downtown poles that one would expect but also distinct poles of different levels of house price in areas around the centre.

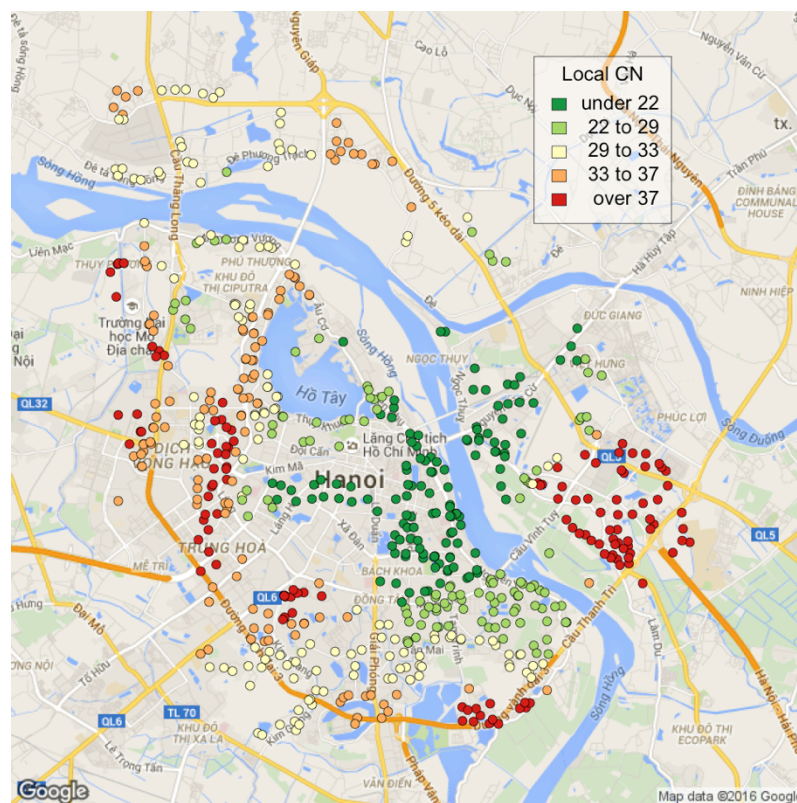


Figure 2. The spatial variation in condition number (CN) in the survey data, with GoogleMaps backdrop.

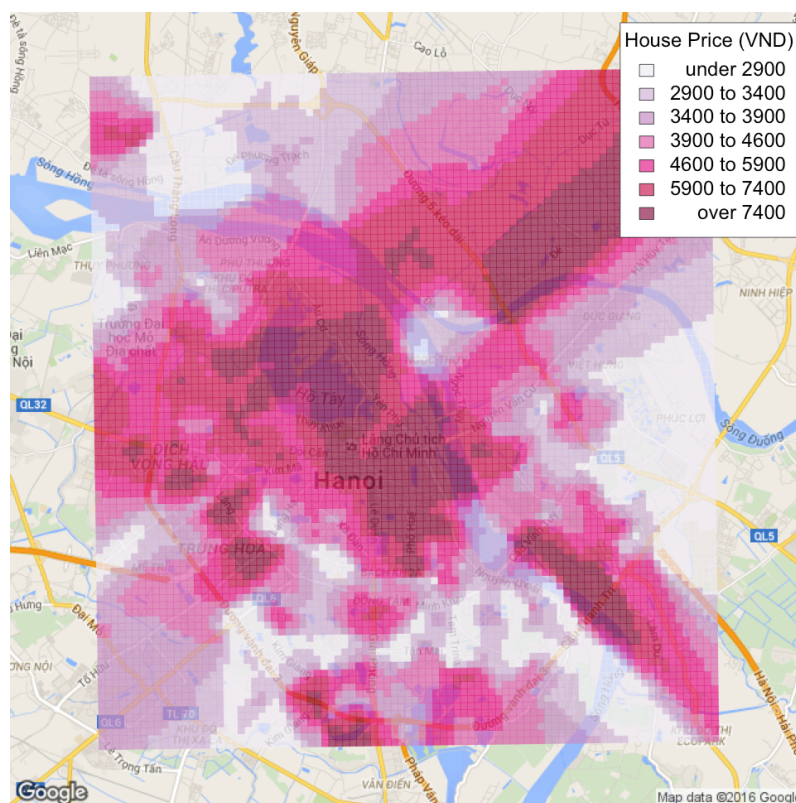


Figure 3. The spatial distribution of house prices in Hanoi derived from a locally compensated GWR model.

The results of this analysis demonstrate that it is important to consider and test for local collinearity even where none is found to exist globally. When local collinearity is found, local models should be applied that are able to handle it. Wheeler (2007; 2009) proposed penalized GWR models with the ridge GWR model and the GW lasso (Wheeler 2009). This study also applied a ridge GWR, but only applied a ridge term where necessary, when the local CN was found to be greater than 30. Thus local ridges were found (Brunsdon *et al.* 2012) and not a global one as applied in the Wheeler (2007) study.

Future work will extend this analysis in a number of directions: network analyses to account for the river running through Hanoi; the application of mixed GWR approaches that include both globally-fixed with locally-varying model coefficients (Nakaya *et al.* 2005) and will disaggregate the poles to identify geo-demographic specific areas of high status. The approach will also be extended to model house price bubbles.

Acknowledgements

This collaborative research was funded by British Academy Newton Advanced Fellowships NG150097, *Developing a housing model based on the status-quality trade off theory*.

References

- Alonso W. (1964). *Location and Land Use*. Harvard University Press, Cambridge, Massachusetts.
- Bowman A (1984). An Alternative Method of Cross-Validation for the Smoothing of Density Estimates. *Biometrika*, 71, 353–360.
- Brunsdon C, Charlton M, Harris P (2012). Living with Collinearity in Local Regression Models. In *Proceedings of the 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*. Brasil.
- Brunsdon, C., Fotheringham, A.S., Charlton, M., (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28(4), 281-298.

- Fujita, M. (1989). *Urban Economic Theory, Land Use and City Size*, Cambridge: Cambridge University Press.
- Gollini, I., Lu, B., Charlton, M., Brunson, C., & Harris, P. (2013). GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software*, 63.
- Kim, J-H., Pagliara, F. and Preston, J., 2005, The intention to move and residential location choice behaviour. *Urban Studies*, Vol. 42, 1621–1636.
- Lu, B., Harris, P., Charlton, M., & Brunson, C. (2014). The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-spatial Information Science*, 17(2), 85-101.
- Nakaya, T., Fotheringham, A. S., Brunson, C., & Charlton, M. (2005). Geographically weighted Poisson regression for disease association mapping. *Statistics in medicine*, 24(17), 2695-2717.
- Parkes, A., Kearns, A. and Atkinson, R., 2002, What makes people dissatisfied with their neighbourhoods? *Urban Studies*, Vol. 39, 2413–2438.
- Phe HH and Wakely P (2000). Status, quality and the other trade-off: Towards a new theory of urban residential location. *Urban Studies*, 37(1), 7-35.
- Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46, 234-240.
- Weisbrod, G. E., Lerman, S. R., and Ben-Akiva, M., 1980, Tradeoffs in residential location decisions: Transportation versus other factors. *Transport Policy and Decision Making*, Vol. 1, 13–26.
- Wheeler D (2007). Diagnostic Tools and a Remedial Method for Collinearity in Geographically Weighted Regression. *Environment and Planning A*, 39(10), 2464–2481.
- Wheeler D (2009). Simultaneous Coefficient Penalization and Model Selection in Geographically Weighted Regression: the Geographically Weighted Lasso. *Environment and Planning A*, 41(3), 722–742.