Global Research Unit Working Paper #2016-003

A varying coefficient approach to estimating hedonic housing price functions and their quantiles

Alan T.K. Wan, City University of Hong Kong
Shangyu Xie, University of International Business and Economics
Yong Zhou, Chinese Academy of Sciences and Shanghai University of Finance and Economics

A varying coefficient approach to estimating hedonic housing price functions and their quantiles*

Alan T.K. Wan[1,†], Shangyu Xie[2] and Yong Zhou[3,4]

[1] Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong

[2] School of Banking and Finance, University of International Business and Economics, Beijing, China

[3] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

[4] School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China

Abstract

The varying coefficient (VC) model introduced by Hastie and Tibshirani (1993) is arguably one of the most remarkable recent developments in nonparametric regression theory. The VC model is an extension of the ordinary regression model where the coefficients are allowed to vary as smooth functions of an effect modifier possibly different than the regressors. The VC model reduces the modeling bias with its unique structure while also avoiding the "curse of dimensionality" problem. While the VC model has been applied widely in a variety of disciplines, its application in economics has been minimal. The central goal of this paper is to apply VC modeling to the estimation of a hedonic house price function using data from Hong Kong, one of the world's most buoyant real estate markets. We demonstrate the advantages of the VC approach over traditional parametric and semi-parametric regressions in the face of a large number of regressors. We further combine VC modeling with quantile regression to examine the heterogeneity of the marginal effects of attributes across the distribution of housing prices.

Keywords: hedonic price function, heterogeneity, housing, kernel estimation, quantile regression, varying-coefficient

JEL classification: C14, C21, R21

[†]Corresponding author. E-mail: alan.wan@cityu.edu.hk; Phone: (852) 3442-7146

1 Introduction

Since Lancaster's (1966) and Rosen's (1974) seminal work, the hedonic pricing method has become the most employed revealed preference technique for non-market valuation of goods and amenities. Hedonic pricing, in particular, is widely utilized in housing studies and real estate appraisals. This method allows house prices to be deconstructed into values and quantities of trait variables such as living area, age of the dwelling, neighbourhood characteristics, etc., thus providing an important advantage in comparison to other methods. Traditionally, hedonic housing price models are estimated in a linear regression framework (e.g., Kain and Quigley, 1970; Witte, Sumka and Erekson, 1979), but some studies have allowed for non-linear relationships, mostly through the applications of Box-Cox transformations (e.g., Goodman, 1978; Halvorsen and Pollakowski, 1981).

Although there already exists a large body of literature on hedonic pricing, research continues on the development of new approaches to the formulation and estimation of hedonic relationships. Until the 1990s, virtually all published work on hedonic modeling of house prices was based on parametric models that involved assertions about the functional relations between house prices and the trait variables as well as their distributions. It has been argued (e.g., Mason and Quigley, 1996) that more often than not, economic theories offer researchers little guidance on specifying the functional form of the hedonic model; thus we should let the data determine an appropriate model, which is the essence of nonparametric methodology. Indeed, during the past twenty years, nonparametric and semi-parametric methods, which impose either no or minimal parametric assumptions about the data generating process, have undergone significant outgrowth, and are increasingly replacing parametric models for the latter's lack of sufficient flexibility. There is now a vast array of texts on these methods including several tailored to the needs of applied econometricians (Pagan and Ullah, 1999; Yatchew, 2003; Li and Racine, 2007; Henderson and Parmeter, 2015).

Most of the literature on nonparametric regression deals with the kernel method and its variants, primarily because of their mathematical and intuitive simplicity. Traditional kernel methods are based on local mean smoothing, although local polynomial smoothing (Fan and Gijbels, 1996) has recently gained prominence. Empirical applications of kernel regression methods to hedonic house price modeling can be found in the work of Pace (1993, 1995), Anglin and Gencay (1996), McMillen and Thorsnes (2000) and Parmeter, Henderson and Kumbhakar (2007). While kernel methods are by far the most popular, they are just one of the many approaches for constructing flexible models. Other well-known nonparametric approaches include spline smoothing, locally weighted regression, and nearest neighhbour. These methods have all been successfully applied to the estimation of hedonic house price models (Messe and Wallace, 1991; Pavlov, 2000; Bao and Wan, 2004; Sunding and Swoboda, 2010). One common feature of nearly all published studies is that they approach the nonparametric estimation problem by retreating it as a semi-parametric problem whereby the discrete housing attributes, such as view, enter the model in a linear parametric fashion, while the continuous attributes such as dwelling size enter the model nonparametrically. This is because conventional nonparametric methods generally do not handle discrete variables satisfactorily[1]. The development of a full-fledged nonparametric regression approach in the face of discrete variables is still an ongoing problem, but some progress has been made by Racine and Li (2004). An application of Racine and Li's (2004) method to hedonic house price modeling is given in Parmeter, Henderson and Kumbhakar (2007).

Although nonparametric methods have been extensively applied in a wide array of empirical domains, these methods do come with some costs. First, in order to yield any merit, nonparametric methods all require very large samples. Second, and more important, the rate of convergence of nonparametric estimators tends to decrease rapidly as the number of regressors grows. The latter is the well-known *curse of dimensionality* that afflicts virtually all standard nonparametric methods including those mentioned above, rendering these methods ineffectual when there is a large number of regressors in the model. This is clearly an issue with hedonic housing price modeling because many different characteristics can affect prices. Over the past two decades, econometricians and statisticians have spent a great deal of time and have made some progress in developing alternative approaches for alleviating the curse of dimensionality. The additive nonparametric model (Hastie and Tibshirani, 1990) is among the earliest approaches proposed to alleviate this problem. An application of the additive model to hedonic house price modeling was undertaken by Martins-Filho and Bin (2005), using data from Portland, Oregon, U.S.A. An alternative to the additive model is the varying coefficient (VC) model (sometimes referred to as the functional-coefficient model) popularised in the work of Hastie and Tibshirani (1993). The greatest appeal of the VC model is that it allows the unknown coefficients to vary as smooth functions of a small number of (non-discrete) variables, known as effect modifiers. Estimation of the VC model thus involves only low-dimensional smoothing, as opposed to the high-dimensional smoothing required for standard nonparametric procedures. Because only low-dimensional nonparametric functions are estimated, the curse of dimensionality can be circumvented even if there is a large number of regressors. Moreover, unlike many other nonparametric

[1]Within the conventional nonparametric framework, categorical variables are handled by a frequency-based approach that splits the sample into cells. However, when the number of cells is large, each cell may have insufficient observations to estimate nonparametrically the relationship among the remaining continuous variables. For this reason most empirical studies involving discrete regressors use a semi-parametric approach.

models, the VC model assumes that there is a linear relationship between the dependent variable and regressors, albeit a changing one. The VC model thus has the flexibility of a nonparametric model and the easy interpretability of an ordinary linear regression. One further advantage of the VC model is that by allowing the coefficients to vary with the effect modifiers, it permits nonlinear interactions between the effect modifiers and other regressors. This is an important merit of the VC model relative to the additive model mentioned above. Moreover, the additive model utilized by Martins-Filho and Bin (2005) assumes all regressors are continuous.[2] In a typical hedonic house pricing analysis, discrete housing attributes are common and often outnumber continuous housing attributes. The VC model we consider here does not suffer from the same deficiency. The VC models have been exhaustively studied and applied to a variety of domains, including time series analysis (Cai, 2007), panel data analysis (Fan and Li, 2004), duration analysis (Fan, Lin and Zhou, 2006), and finance (Xie, Zhou and Wan, 2014), and an array of estimation and inference procedures have been developed. Fan and Zhang (2008) and Park, Mammen, Lee and Lee (2015) provide comprehensive surveys of this material. However, to the best of our knowledge, the VC approach has not been applied to hedonic price modeling, and the purpose of this paper is to take steps in this direction.[3]

Another frequent concern in the literature on hedonic house price modeling is the heterogeneity of the marginal effects of attributes across the distribution of housing prices. Indeed, it is not uncommon for a given housing characteristic to be priced differently for different dwellings; for example, buyers of lower-priced dwellings likely value proximity to transport higher than buyers of higher-priced dwellings, who probably consider attributes such as the presence of car parking spaces or recreational facilities as more important. Common hedonic models, parametric and nonparametric, typically estimate the marginal effects of the attributes as conditional mean functions of house prices, neglecting the possible heterogeneity of price behaviour. Zietz, Zietz and Sirmans (2008), Mak, Choy and Ho (2010), and Kim, Hung and Park (2015) have attempted to address this issue using quantile regression (QR), but their approaches are parametric, and as such they suffer from the potentially serious consequences of functional form mis-specification as discussed above. Recent advances in the statistics literature have developed methods of varying-coefficients to the estimation of regression quantiles (Honda, 2004; Kim, 2007; Cai and Xu, 2009; and Xie, Zhou and Wan, 2014). This approach has the potential of permitting heterogeneity in

[2] Refinements to the additive model allowing for discrete regressors have been considered by Fan, Härdle and Mammen (1998), and Camlong-Viot, Rodríguez-Póo and Vieu (2006).

[3] The VC model considered here must not be confused with a different model bearing the same name considered by Knight, Dombrow and Sirmans (1995). The latter is essentially a parametric linear regression model that allows the regression coefficients to vary.

the implicit price of attributes at different points in the distribution of housing prices, as well as being flexible in terms of functional form and computational feasibility even in the face of a large number of attributes. A second purpose of the present paper is to exploit the recent theoretical advances in varying coefficient quantile regression (VC-QR) in hedonic house price modeling.

As our empirical framework we choose a hedonic house price model with data from Hong Kong, one of the world's most active real estate investment markets. We extract our data from a database maintained by the Centaline Property Agency Ltd, the territory's largest real estate agency company. There have been several recent studies published on hedonic house pricing in Hong Kong using data from this same database (Bao and Wan, 2004, 2007; Bucchianeri, 2008; Magnus, Wan and Zhang, 2011; Kim, Hung and Park, 2015), which contains detailed information on transactions and dwelling characteristics of more than 550,000 condominium units across 118 private housing complexes (estates) in Hong Kong. A typical housing estate comprises some 10-30 high rise buildings, with each building containing 20-40 floors with 4-6 condos on each floor. Many of the newer estates are equipped with their own leisure and sports facilities. It is also not uncommon for a housing estate with more than one thousand condo units to have as few as four to six different floor plans, yet these floor plans can be so vastly different that there are huge price differences between the most expensive and cheapest units within the same estate. Nevertheless, given the relative uniformity of floor plans, many condos in the same estate are identical in terms of living area, i.e, kitchen size and number of bedrooms and bathrooms, but differ in terms of floor levels, views and direction. The housing estate being considered in this study is South Horizons, located on the southern side of Hong Kong Island near 'Ocean Park' and the industrial suburb of Aberdeen. Our observation period is December 3, 2007 - February 14, 2013, during which a total of 2683 transactions were recorded for South Horizons. Magnus, Wan and Zhang (2011) also considered the same housing estate in their study based on a parametric model averaging technique but a much shorter observation period. Kim, Hung and Park (2015) recently estimated a hedonic house price function for another major housing estate in Hong Kong by Box-Cox quantile regression, which is parametric and hence less flexible than the semi-parametric VC-QR approach considered in the present paper. Their study is also based on a sample substantially smaller than ours.

In Section 2, we introduce the VC model along with a description of a benchmark semi-parametric model for subsequent comparison purposes. Section 3 contains a description of the data. Section 4 reports the estimation results and compares the performance of the VC model to that of the benchmark parametric linear and semi-parametric models. In section 5, we discuss the VC-QR modeling framework and demonstrate how the response of housing prices to various housing traits can vary across quantiles. Our concluding remarks are in Section 6.

2 Model descriptions

2.1 A varying coefficient hedonic housing price model

A VC hedonic housing price model may be expressed as follows:

$$Y_i = a_1(U_i)X_{i1} + a_2(U_i)X_{i2} + \cdots + a_p(U_i)X_{ip} + \varepsilon_i, \quad i = 1, 2, \cdots, n, \tag{2.1}$$

where Y is the sale price of the dwelling (or its transformation such as the log transformation), $X_{.j}$ is the value of the j^{th} housing characteristic, U is a housing characteristic not belonging in the set containing $X_{.1}, \cdots, X_{.p}$ that modifies the effects of $X_{.j}$'s in a nonparametric way through the unspecified coefficient functions $a_j(U)$'s that vary smoothly over U, and ε is a random disturbance term such that $E(\varepsilon_i|X_{i1}, \cdots, X_{ip}, U_i) = 0$ for all $i's$. Model (2.1) says that Y is linearly dependent on $X_{.1}, X_{.2}, \cdots X_{.p}$, but U changes the coefficients of the $X_{.j}$'s through $a_1(U), a_2(U), \cdots a_p(U)$. Suppose that U and $X_{.1}$ are the floor level and size of the dwelling respectively. The VC model allows the marginal effect of an increase in dwelling size (say, a square foot) on the sale price Y to change smoothly with the floor level. Thus, the VC model simultaneously possesses the flexibility of a nonparametric model and the easy interpretability of a linear regression, while allowing a special kind of interaction between the effect modifier and each of the regressors. We typically estimate $a_j(U)'s$ by a kernel method, and because kernel smoothing acts only on the space of U, a single variable, the estimators have one-dimensional convergence rates (Stone, 1986), making them far more accurate than traditional kernel estimators of a p-dimensional function. Hence, the VC model can avoid the curse of dimensionality even in the face of a large number of $X_{.j}'s$. Although U is usually taken to be a one-dimensional variable, in general it can be a low-dimensional vector of variables.

The usual and simplest estimator of $a_j(U)$ is local least squares, although more complicated local polynomial fitting methods are also available. The local least squares method possesses a number of desirable properties, including those of optimal asymptotic minimax efficiency, design adaptation, and good boundary behaviour (Fan, 1993; Fan and Gijbels 1996; Ruppert and Wand 1994). Now, let $\mathbf{a} = (a_1, \cdots, a_p)^T$ and $\mathbf{b} = (b_1, \cdots, b_p)^T = (a_1^{'}(\cdot), \cdots, a_p^{'}(\cdot))^T$ be vectors of constants. Assume that $a_j(u)$ is twice continuously differentiable so that the function $a_j(\cdot)$ can be approximated locally by $a_j \approx a_j + b_j(u - u_0)$, where u is a point in the neighborhood of a given point u_0. Write $\boldsymbol{\beta} = (a^T, hb^T)^T$ and let

$\boldsymbol{\beta}_0 = (a_0^T(\cdot), ha_0'(\cdot)^T)^T$ be a vector comprising the unknown coefficient functions. The local least squares estimator $\widehat{\boldsymbol{\beta}} = (\widehat{\mathbf{a}}^T, h\widehat{\mathbf{b}}^T)^T$ of $\boldsymbol{\beta}_0$ is obtained by a minimization of

$$\frac{1}{n}\sum_{i=1}^{n}\big[Y_i - \sum_{j=1}^{p}(a_j + b_j(U_i - u_0)X_{ij})\big]^2 \times K_h(U_i - u_0), \tag{2.2}$$

where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function, and $h = h_n > 0$ is a bandwidth. In matrix notation, $\widehat{\boldsymbol{\beta}}$ may be written as

$$\widehat{\boldsymbol{\beta}} = (\widehat{\mathbf{a}}^T, h\widehat{\mathbf{b}}^T)^T = \{\mathbf{D}_{u_0}^T\mathbf{W}_{u_0}\mathbf{D}_{u_0}\}^{-1}\mathbf{D}_{u_0}^T\mathbf{W}_{u_0}\mathbf{Y}, \tag{2.3}$$

where

$$\mathbf{D}_{u_0} = \begin{pmatrix} \mathbf{X}_1^T & \frac{U_1-u_0}{h}\mathbf{X}_1^T \\ \vdots & \vdots \\ \mathbf{X}_n^T & \frac{U_n-u_0}{h}\mathbf{X}_n^T \end{pmatrix},$$

$\mathbf{X}_i = (X_{i1}, \cdots, X_{ip})^T$, $\mathbf{W}_{u_0} = diag(K_h(U_1-u_0), \cdots, K_h(U_n-u_0))$, and $\mathbf{Y} = (Y_1, \cdots, Y_n)^T$.

As in linear regression, a test that is always of interest is whether the housing attribute $X_{\cdot j}$ is significant or helps explain the variation in the sale price Y. For any function $g(u)$, define $||g||_\infty = \sup_{u\in[0,1]} |g(u)|$. Fan and Zhang (2000) shows that the hypothesis

$$H_0 : a_j(u) = 0 \quad \text{vs.} \quad H_1 : a_j(u) \neq 0$$

may be tested using the statistic

$$T_1 = (-2log_e(h))^{1/2}(\big|\big|\{\widehat{var}(\widehat{a}_j(u)|\mathcal{G})\}^{-1/2}(\widehat{a}_j(u) - \widehat{bias}(\widehat{a}_j(u)|\mathcal{G}))\big|\big|_\infty - d),$$

where $\widehat{a}_j(u)$ is the estimator of $a_j(u)$ at $U = u$, $\mathcal{G} = (U_1, \cdots, U_n, X_{11}, \cdots, X_{n1}, \cdots, X_{1p}, \cdots, X_{np})^T$ is the observed regressor vector, $\widehat{bias}(\widehat{a}_j(u)|\mathcal{G}))$ and $\widehat{var}(\widehat{a}_j(u)|\mathcal{G})$ are estimators of the conditional bias and variance of $\widehat{a}_j(u)$ respectively, and d is a constant that depends on the kernel function and bandwidth. We reject H_0 if T_1 exceeds the asymptotic critical value $c_\alpha = -\log_e(-0.5\log_e(\alpha))$ at a given significant level α.

More generally, T_1 may be modified to

$$T_2 = (-2log_e(h))^{1/2}(\big|\big|\{\widehat{var}(\widehat{a}_j(u)|\mathcal{G})\}^{-1/2}(\widehat{a}_j(u) - \widehat{\beta}_j - \widehat{bias}(\widehat{a}_j(u)|\mathcal{G}))\big|\big|_\infty - d),$$

for testing

$$H_2 : a_j(u) = \beta_j \quad \text{vs.} \quad H_3 : a_j(u) \neq \beta_j,$$

based on the same decision rule as for testing H_0, where $\widehat{\beta}_j$ is the estimator of β_j for the model $Y = \sum_{k=1,k\neq j}^{p} a_k(U)X_k + \beta_j X_j + \varepsilon$ under H_2 (see p.724 in Fan and Zhang, 2000). If H_2 is rejected then a_j does vary with U. On the other hand, if H_2 cannot be rejected at least for some $k < p$ coefficients, then the VC model becomes the semi-VC model studied by Zhang, Lee and Song (2002), where k of the p coefficient functions in (2.1) reduce to constants that are independent of U.

2.2 A benchmark semi-parametric model

We consider the following partially linear semi-parametric model as a benchmark model for comparison:

$$Y_i = Z_{i1}\beta_1 + \cdots + Z_{ip_1}\beta_{p_1} + g(W_{i1}, \cdots, W_{ip_2}) + \varepsilon_i, \quad i = 1, 2, \cdots, n, \tag{2.4}$$

where Y is dwelling price, $Z_{.1}, \cdots, Z_{.p_1}$ are dummy variables representing the qualitative characteristics of housing, $\beta_1, \cdots, \beta_{p_1}$ are unknown coefficients, $W_{.1}, \cdots, W_{.p_2}$ are (non-dummy) variables representing quantitative housing characteristics, $g(.)$ is an unknown functional form and ε, the disturbance term, is assumed to be i.i.d. This model allows the qualitative variables to enter the model linearly and the quantitative variables to enter the model nonparametrically.

One method that is frequently used to estimate $\beta_1, \cdots, \beta_{p_1}$ and $g(\cdot)$ is profile least squares (PLS) developed originally by Speckman (1988). Here, we briefly describe this method. Let us rewrite (2.4) as

$$Y_i^* = g(W_{i1}, \cdots, W_{ip_2}) + \varepsilon_i, \quad i = 1, 2, \cdots, n, \tag{2.5}$$

where $Y_i^* = Y_i - Z_{i1}\beta_1 - \cdots - Z_{ip_1}\beta_{p_1}$. This transforms the partially linear semi-parametric model into a nonparametric regression. Denote $W_i = (W_{i1}, \cdots, W_{ip_2})^T$ and let $w_0 = (w_{01}, \cdots, w_{0p_2})^T$ be a vector of fixed points in the neighbourhood of W_i. Then the kernel estimator of the coefficient function g at w_0 is

$$\widehat{g}(w_0) = \frac{\sum_{i=1}^n K_h((W_i - w_0)/h)Y_i^*}{\sum_{i=1}^n K_h((W_i - w_0)/h)}. \tag{2.6}$$

This leads to, in matrix notations,

$$\widehat{g} = (\widehat{g}(W_1), \cdots, \widehat{g}(W_n))^\top = \Omega\mathbf{Y}^* = \Omega(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta}), \tag{2.7}$$

where the (l,s)-th element of the matrix Ω is $[\Omega]_{l,s} = \frac{K_h((W_s - W_l)/h)}{\sum_{i=1}^n K_h((W_i - W_l)/h)}$, $Y^* = (Y_1^*, \cdots, Y_n^*)^T$, $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_n)^T$, $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \cdots, \mathbf{Z}_{ip_1})^T$ and $\beta = (\beta_1, \cdots, \beta_{p_1})^T$. Now, substituting $\widehat{g}(\cdot)$ into (2.4) yields

$$(I - \Omega)Y = (I - \Omega)\mathbf{Z}\beta + \varepsilon. \tag{2.8}$$

Applying OLS to (2.8), we obtain

$$\widehat{\boldsymbol{\beta}} = \{\mathbf{Z}^T(I - \Omega)^T(I - \Omega)\mathbf{Z}\}^{-1}\mathbf{Z}^T(I - \Omega)^T(I - \Omega)Y, \tag{2.9}$$

which in turn yields

$$\widehat{g} = \Omega(\mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\beta}}). \tag{2.10}$$

3 Data

As described in Section 1, we estimate our hedonic price function based on sales data for the estate South Horizons between December 3, 2007 and February 14, 2013, during which the Hong Kong Land Registry recorded $n = 2683$ transactions. South Horizons comprises a total of 9812 condos across 34 high-rise blocks, each containing 25 - 42 floors. The sizes of these condos range between 632 and 1633 square feet, and the majority of them have 2-3 bedrooms and 1-2 bathrooms. The age of these condos also varies as South Horizons was completed in four stages between 1991 and 1995. The geographical location of South Horizons is shown in Figure 1. The estate is connected to the northern part of Hong Kong Island through the Aberdeen Tunnel, and is on the Mass Transit Railway's South Island Line that is currently nearing completion. The new railway line will reduce traveling time from South Horizons to the central business district to 10 minutes as compared to 25-30 minutes on the road. This improved travel link is expected to boost property prices in South Horizons and its vicinity.

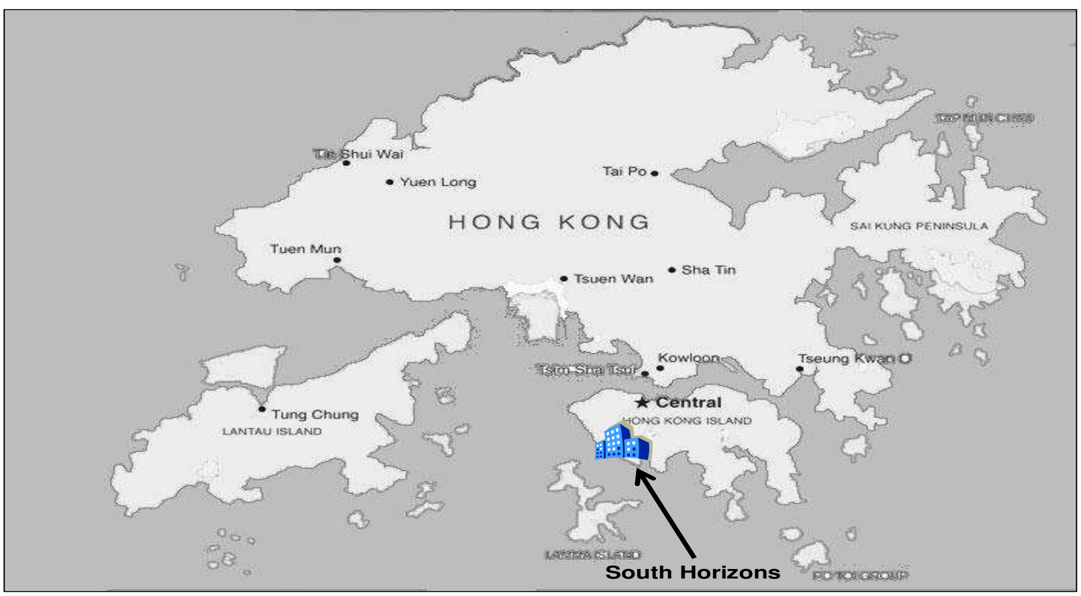

Figure 1: Geographical location of South Horizons

The transaction prices available at the Centaline database are in nominal terms. As the objective of our study is to examine the impacts of housing attributes on housing prices,

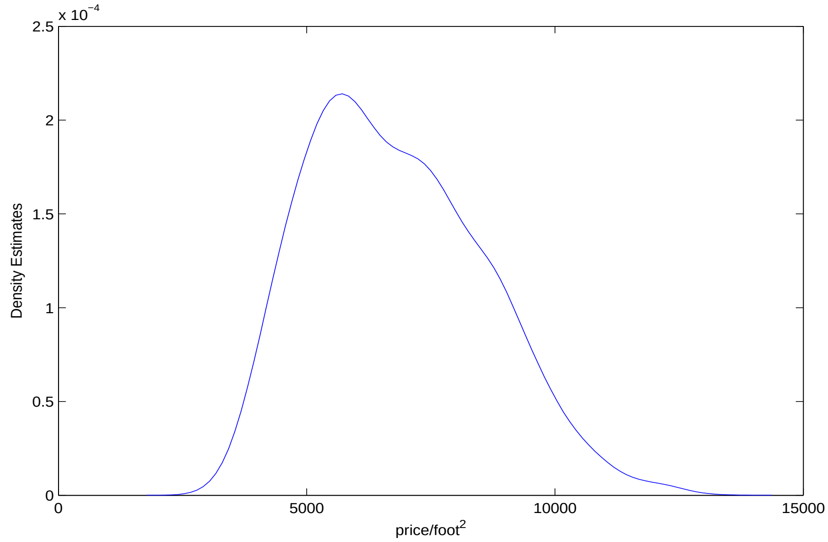

Figure 2: Estimated density of house prices

we convert the nominal transaction prices into real prices using the sub-index pertaining to South Horizons in the Centa-City Index (CCI) as a deflator.[4] The CCI, based on 118 constituent estates, is the most widely quoted property price index for Hong Kong. This index reflects real estate price movements after removing the effects of attributes.[5] In Hong Kong, property prices are commonly reported in terms of price per square foot. We denote the (CCI-deflated) real price/foot^2 as P. Figure 2 gives the plot of the estimated density of P corresponding to the 2683 observations in our sample based on kernel method with bandwidth set to the default value in MATLAB. The plot shows that there is high level of price dispersion with the bulk of the properties sold between HKD3,000/foot^2 and HKD10,000/foot^2 (1USD $\approx$ 7.75HKD). The average real sales price is HKD6340/foot^2. The density has a skewness of 0.5330, indicating that price distribution is mildly right-skewed, and a kurtosis 3.3695, indicating that the density is leptokurtic with a peak higher than that of a normal distribution.

Most studies in the hedonic house pricing literature use the natural log of house prices as the dependent variable in order to alleviate heteroscedasticity associated with skewed sales price data. We follow this practice in our study and denote the natural log of real sales price/foot^2 as LP. Twelve other variables containing a wealth of information on the condos' characteristics obtained from the Centaline database are included in our hedonic model. These include the natural logarithm of gross area in square feet ($LGAREA$), floor

[4]The sub-indices are referred to as "adjusted unit prices" within the CCI system.

[5]See $http://hk.centadata.com/cci/cci_e.htm$, Bao and Wan (2004) and Bucchianeri (2008) for a description of the index.

level ($FLOOR$), age (AGE), direction (DIR) and various characteristics on views. AGE is coded as 4,5,6 or 7, and the higher the value of AGE, the older is the building. DIR is a binary variable that takes on 1 if the condo faces east, south, south east or south west, and 0 otherwise. From a "feng-shui" perspective (which plays a significant role in the culture of Hong Kong), condos with a east-facing exposure are normally preferred as they face the direction of the sunrise; south-facing condos, which tend to be warmer in winter and more breezy in summer, are also expected to command a premium price over condos that face north. Among the various views that might be visible from condos in South Horizons, a view of the gardens within the estate ($GARDV$) generally adds value to the price of the condo, as does a sea view of the nearby East Lamma Channel and Shek Pai Wan. The presence of a sea view, especially an unobstructed one, generally has a pronounced positive influence on property prices. In the case of South Horizons, the impact of a sea view can differ depending on whether it is full unobstructed sea view ($SEAVF$), a semi-sea view ($SEAVS$) or a minor-sea view ($SEAVM$), and we distinguish between them accordingly. On the other hand, buyers usually shy away from properties from which the Aberdeen cemetery ($CEMV$) is visible. Some condos have a distant view of Brick Hill ($MONV$) where Ocean Park is located. Although a view of greenery is generally considered good feng-shui, in the case of South Horizons, the view of Brick Hill does not yield the expected added value because it also includes the view of some public housing estates located nearby. Street views within the estate ($STRI$) and surrounding areas tend to negatively affect sales prices due to traffic noise. It is also not uncommon to find condos in South Horizons that have no street view at all ($STRN$), a factor we also consider.

4 Estimation results

The marginal prices of housing attributes are rarely constants. Typically in a multi-storey condo complex, floor level strong impacts the marginal prices of attributes, such as the direction the condo faces and the views. One distinct merit of the VC approach, as described in Section 1, is that it allows some attributes to modify the effects of other attributes. We take advantage of this feature to investigate the modifying effects of floor levels on the effects of other attributes on property prices. We define $FLOOR = \min(FLOOR, 40)$, and let $U = (FLOOR - \min(FLOOR))/(\max(FLOOR) - \min(FLOOR))$ be the effect modifier in our VC model, where $\max(FLOOR)$ and $\min(FLOOR)$ are the highest and lowest ordinal floor levels in the building where the condo is housed. The variable U thus converts $FLOOR$ to standard units. We use U instead of $FLOOR$ because housing blocks in South Horizons do not all have the same number of floors. As well, floor levels are sometimes labelled differently across the different blocks because in Cantonese culture the number 4 is commonly associated with bad luck; for example, in some blocks, levels 4, 14 and 24 are labelled as levels 5, 15 and 25 respectively. We use the first 2383 transactions for model estimation and the remaining 300 observations for out-of-sample forecast evaluation.

The following VC model, which includes all of the twelve housing attributes, forms the basis of our investigation at the initial stage:

$$\begin{aligned} LP_i &= a_0(U_i) + a_1(U_i)AGE_i + a_2(U_i)DIR_i + a_3(U_i)SEAVF_i + a_4(U_i)SEAVS_i + a_5(U_i)SEAVM_i \\ &+ a_6(U_i)GARDV_i + a_7(U_i)MONV_i + a_8(U_i)STRN_i + a_9(U_i)STRI_i + a_{10}(U_i)INDUSTV_i \\ &+ a_{11}(U_i)CEMV_i + a_{12}(U_i)LGAREA_i + \varepsilon_i, \quad i = 1, \cdots, 2383. \end{aligned} \tag{4.1}$$

Our estimation of $a_j(U)$'s in (4.1) is based on the local least squares method described in Section 2.1. We use the Epanechnikov kernel $K(u) = 3(1-u^2)(|u| \le 1)/4$ and set the bandwidth to $h = 0.20 \times (\max(U) - \min(U))$ using Silverman's rule of thumb. Table 1 shows that the $p-$values of the hypothesis tests of $H_0 : a_j(u) = 0$ and $H_2 : a_j(u) = \beta_j$. It is observed that at the 5% level of significance, $H_0 : a_j(u) = 0$ is rejected except for $j = 5, 7, 8$ and 9, indicating that $SEAVM$ (minor-sea view), $MONV$ (mountain view), $STRN$ (no street view) and $STRI$ (internal street view) have no significant effect on the the log of house prices. Results of the tests of $H_2 : a_j(u) = \beta_j$ show that for the remaining nine coefficient functions that differ significantly from zero, $a_1(U)$ (AGE (dwelling age)), $a_6(U)$ ($GARDV$ (garden view)) and $a_{11}(U)$ ($CEMV$ (cemetery view) are invariant with respect to U, but $a_2(U)$ (DIR (direction)), $a_3(U)$ ($SEAVF$ (full-sea view)), $a_4(U)$ ($SEAVS$ (semi-sea view)), $a_{10}(U)$ ($INDUSTV$ (industrial view)) and $a_{12}(U)$ ($LGAREA$ (log gross area)) are varying functions of U. With respect to $a_0(U)$, the test results also suggest that standardised floor level has a significant and varying effect on the prices of dwellings.

Table 1: p-value for tests of $H_0 : a_j(u) = 0$ and $H_2 : a_j(u) = \beta_j$

$j =$	1	2	3	4	5	6
$H_0 : a_j(u) = 0$	0.0233	0.0001	0.0002	0.0018	0.0766	0.0303
$H_2 : a_j(u) = \beta_j$	0.1033	0.0010	0.0288	0.0258	0.2065	0.1371
$j =$	7	8	9	10	11	12
$H_0 : a_j(u) = 0$	0.1206	0.2085	0.0881	0.0374	0.0046	0.0000
$H_2 : a_j(u) = \beta_j$	0.1923	0.2469	0.0747	0.0391	0.2393	0.0056

We then re-estimate the VC model by removing $SEAVM$, $MONV$, $STRN$ and $STRI$ from the full model, and treat the coefficients of AGE, $GARDV$ and $CEMV$ as constants.

This results in the following estimated semi-VC model:

$$\widehat{LP}_i = \widehat{a}_0(U_i) + \widehat{a}_2(U_i)DIR_i + \widehat{a}_3(U_i)SEAVF_i + \widehat{a}_4(U_i)SEAVS_i + \widehat{a}_{10}(U_i)INDUSTV_i$$
$$+ \widehat{a}_{12}(U_i)LGAREA_i - \underset{(0.0062)}{0.0136} \times AGE_i + \underset{(0.0057)}{0.0355} \times GARDV_i - \underset{(0.0074)}{0.0315} \times CEMV_i, \quad i = 1, \cdots, 2383, \tag{4.2}$$

where the figures inside the parentheses are standard errors of the estimates. The estimates of the (non-varying) coefficients suggest that the view of garden has a large positive impact on house prices. On the other hand, a view of the Aberdeen cemetery and the age of the building have a negative impact on condo prices. Plots of the estimated VC functions and their corresponding 95% confidence bands versus $FLOOR$ are shown in Figures 3(a)-(f). The strong desire to live as high as possible is illustrated in Figure 3(a), where $\widehat{a}_0(U)$ is invariably positive and large, although as the floor level goes beyond 25, there is a noticeable decline in the additional premium one is willing to pay for higher level living. Figure 3(b) shows that condos which are not north- or directly west-facing are generally preferred, but it is also observed that $\widehat{a}_2(U)$ first decreases as $FLOOR$ increases, bottoming-out near zero at $FLOOR = 15$ before increasing again for $FLOOR > 15$. This suggests that the differentials in directions have little impact on prices for condos in the middle levels. A partial sea-view ($SEAS$) is value-adding, and its impact becomes more pronounced as the standardised floor level rises beyond 30, as shown in Figure 3(d). In comparing the values of the estimated coefficient functions for $SEAVF$ in 3(c) and $SEAVS$ in 3(d), it is apparent that a full sea-view is usually preferred over a semi sea-view. The impact of a view of the nearby industrial plants in Aberdeen is not strong, but is more often negative than positive, as shown in Figure 3(e). The invariably positive values of $\widehat{a}_{12}(U)$ revealed in Figure 3(f) indicate that consumers are willing to pay a higher price per square foot in order to live in a larger condo.

We consider two other models for comparisons with the VC model. The first is the conventional parametric linear model. Initially, we include all attributes and the interaction terms between U and the other attributes in the model. We find that none of the interaction terms is significant. The following is our final model after eliminating the insignificant attributes:

$$\widehat{LP}_i = \underset{(0.0804)}{7.8407} + \underset{(0.0102)}{0.0535} \times DIR_i + \underset{(0.0134)}{0.0975} \times SEAVF_i + \underset{(0.0117)}{0.0572} \times SEAVS_i + \underset{(0.0089)}{0.0293} \times GARDV_i$$
$$- \underset{(0.0103)}{0.0427} \times STRI_i - \underset{(0.0141)}{0.0435} \times INDUSTV_i - \underset{(0.0094)}{0.0422} \times CEMV_i + \underset{(0.0402)}{0.3536} \times LGAREA_i$$
$$+ \underset{(0.0140)}{0.1486} \times U_i, \qquad i = 1, \cdots, 2383. \tag{4.3}$$

The signs of the estimates produced by the linear model are all consistent with prior

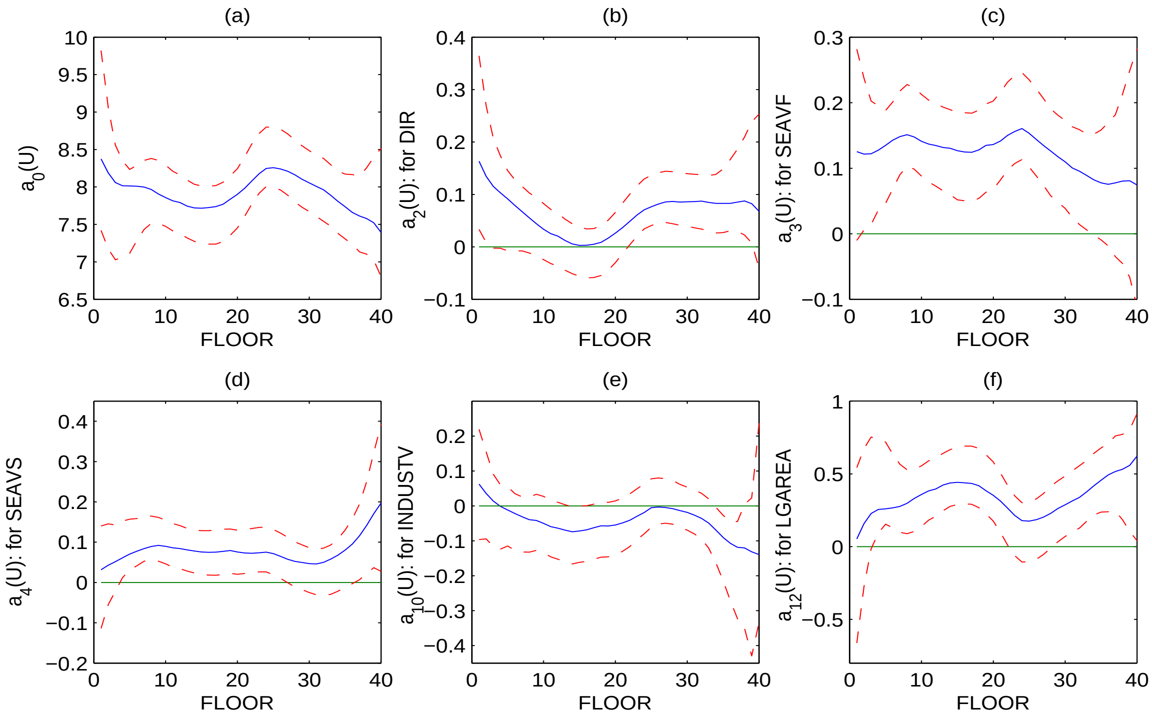

Figure 3: Plots of the estimated coefficient functions (solid curve) and the 95% simultaneous confidence bands (dashed curves) for estimated coefficient functions for the semi-VC model (4.2).

expectations. Several of the attributes found to be significant under the semi-VC model are also significant under the linear model; in particular, the coefficient estimates for $GARDV$ and $CEMV$ are very close to the corresponding (non-varying) estimates produced under the semi-VC model. The estimates of DIR, $SEAVF$, $SEAVS$, $INDUSTV$ and $LGAREA$ obtained under (4.3) are about the same as the means of the corresponding semi-VC coefficient functions produced by (4.2).

Our second benchmark model is a partially linear semi-parametric model, of which the nonparametric and parametric components are represented by the unspecified smooth function $g(LGAREA, U)$, and a linear combination of the dummy attributes and interaction terms of each pair of these attributes respectively. Our estimation of this model follows the two-step profile estimation procedure described in Section 2.2, with kernel smoothing based on the Gaussian kernel $K_2(x_1, x_2) = (2\pi)^{-1}\exp(-(x_1^2 + x_2^2 + 2x_1^2x_2^2)/2)$ applied to the estimation of the $g(.,.)$. After eliminating the insignificant attributes, our final model is

$$\widehat{LP}_i = \underset{(0.0101)}{0.0565} \times DIR_i + \underset{(0.0165)}{0.1153} \times SEAVF_i + \underset{(0.0126)}{0.0800} \times SEAVS_i + \underset{(0.0116)}{0.0260} \times SEAVM_i$$
$$+ \underset{(0.0096)}{0.0277} \times GARDV_i - \underset{(0.0144)}{0.0310} \times STRI_i - \underset{(0.0098)}{0.0424} \times CEMV_i + \widehat{g}(LGAREA_i, U_i), \quad i = 1, \cdots, 2383. \tag{4.4}$$

The coefficient estimates in the linear part of the semi-parametric model all have the expected signs and are very close to those obtained under the semi-VC and linear models. The plot of $\widehat{g}(LGAREA, U)$ in Figure 4 shows that $\widehat{g}(LGAREA, U)$ is always positive, and large condos located on high floors are the most expensive in terms of price per square foot. Larger condos are generally priced higher (in price per square foot terms) than smaller condos located on the same floor, except for condos located on very low levels, where there is a tendency for the condo price/foot^2 to first increase then decrease as the condo size increases. Generally speaking, the impact of $LGAREA$ on house prices is larger for condos located on higher floors than for those located on lower floors. This is consistent with the findings obtained under the semi-VC model (Figure 3(f)).

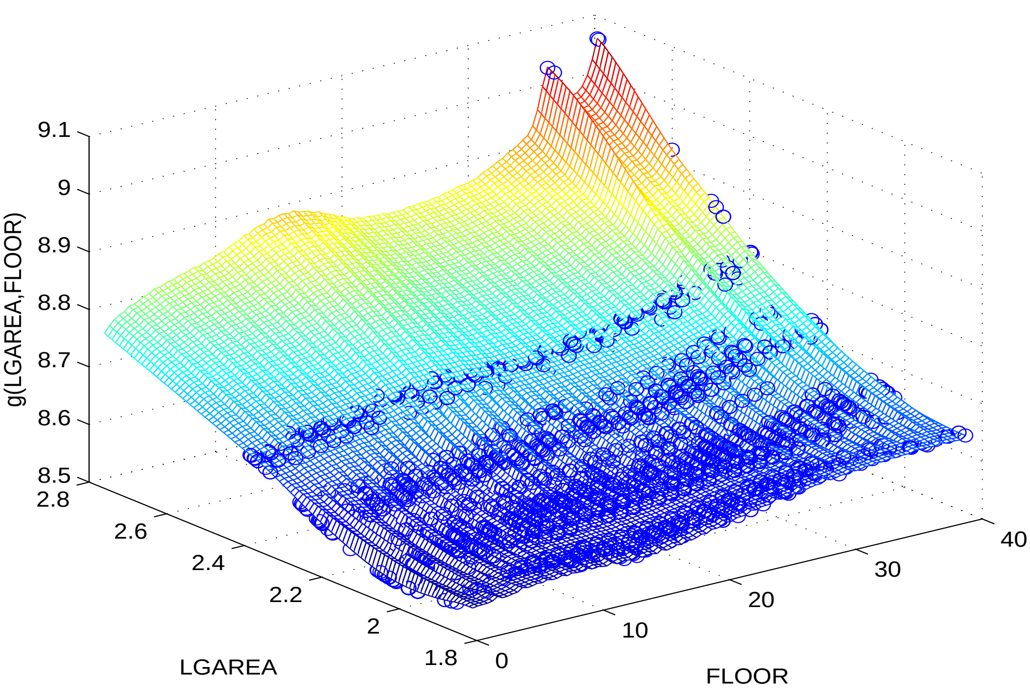

Figure 4: Surface plot of the estimated $g(\cdot, \cdot)$ as a varying function of LGAREA and FLOOR

We also evaluate the in-sample and out-of-sample predictive accuracy of P based on the three models in terms of the mean predictive squared errors (MPSE). Our out-of-sample evaluation is based on the last 300 observations in the sample not included in the estimation of the models. It is found that the semi-VC, linear, and the semi-parametric models yield in-sample MPSE of 0.0283, 0.0339 and 0.0333 respectively. The corresponding out-of-sample MPSE values are 0.1274, 0.1330 and 0.1352. The semi-VC model thus outperforms both benchmark models under both in-sample and out-of-sample evaluations. We have also tested the equality of the MPSE values using the sign test and Wilcoxon's signed-rank test (Diebold and Mariano, 1995). We are able to reject the hypotheses of the linear model and the semi-parametric model yielding the same predictive accuracy as the semi-VC model using both of these tests. Table 3 reports the test results.

Table 2: P-values of statistical tests of predictive accuracy

in-sample	linear vs. semi-VC model	semi-parametric vs. semi-VC model
Sign Test	<0.0001	<0.0001
Wilcoxon's Signed-rank Test	<0.0001	<0.0001
out-of-sample	linear vs. semi-VC model	semi-parametric vs. semi-VC model
Sign Test	0.0921	0.0744
Wilcoxon's Signed-rank Test	0.0939	<0.0001

5 A varying coefficient quantile regression model for hedonic house pricing

Our hedonic pricing analysis based on the VC model has revealed that many housing characteristics are not priced the same across different floor levels. This notwithstanding, the VC model assumes that for a given standardised floor level, a housing characteristic has the same effect on housing prices across the distribution of prices. As argued in Section 1, the response of housing prices to various housing traits often vary across quantiles, and a housing trait can have strong impact on prices at certain quantiles but little or no impact at other quantiles. Floor level adjustments cannot fully account for all of the heterogeneity in marginal effects, e.g., two condos located on the same floor of the same estate block can have other, vastly different characteristics which contribute to large price differences.

To address the issue of possible heterogeneity in the marginal effects of a housing attribute at different points in the distribution of prices, we apply the previously described VC approach to quantile regression. Building on the VC model formulated in the last section, at a given quantile $\tau(0 < \tau < 1)$ of LP, assuming that $Pr(\varepsilon_i < 0|\mathbf{X}_i, U_i) = \tau$ and ε_i follows a continuous, but otherwise unspecified, distribution, the VC-QR of LP conditional on the housing traits may be expressed as

$$Q_\tau(LP_i|\mathbf{X}_i, U_i) = c_{0,\tau}(U_i) + c_{1,\tau}(U_i)AGE_i + c_{2,\tau}(U_i)DIR_i + c_{3,\tau}(U_i)SEAVF_i + c_{4,\tau}(U_i)SEAVS_i$$
$$+ c_{5,\tau}(U_i)SEAVM_i + c_{6,\tau}(U_i)GARDV_i + c_{7,\tau}(U_i)MONV_i + c_{8,\tau}(U_i)STRN_i + c_{9,\tau}(U_i)STRI_i$$
$$+ c_{10,\tau}(U_i)INDUSTV_i + c_{11,\tau}(U_i)CEMV_i + c_{12,\tau}(U_i)LGAREA_i, \quad i = 1, \cdots, 2383, \tag{5.1}$$

where $c_{j,\tau}(U)$ is a smooth VC coefficient function of U that captures the marginal effects of the j^{th} housing trait on the τ-quantile of LP as U varies; $c_{j,\tau}(U)$ is thus different from the corresponding $a_j(U)$ in (4.1), which represents the effects of the same housing trait on the mean of LP at varying values of U. When τ=0.5, (5.1) becomes a median regression; the VC functions accordingly show the responses of the median of LP to a one-unit change in the housing traits across different values of U.

To estimate $c_{j,\tau}(U)$'s, let $\mathbf{c} = (c_1, \cdots, c_p)^T$, and $\mathbf{b} = (b_1, \cdots, b_p)^T = (c_1'(\cdot), \cdots, c_p'(\cdot))^T$ be vectors of constants. Assume that $c_j(u)$ is twice continuously differentiable so that the function $c_j(\cdot)$ can be approximated locally by $c_j(u) \approx c_j + b_j(u - u_0)$, with u located in the neighborhood of a given point u_0. Write $\boldsymbol{\beta} = (\mathbf{c}^T, \mathbf{b}^T)^T$, and let $\boldsymbol{\beta}_0(.) = (\mathbf{c}^T(.), \mathbf{c}'(.)^T)^T$ be the vector of the true parameter functions. The local quantile regression estimator $\widehat{\boldsymbol{\beta}}$=$(\widehat{\mathbf{c}}^T, \widehat{\mathbf{b}}^T)^T$ is obtained by minimizing

$$\frac{1}{n}\sum_{i=1}^{n} \rho_\tau(T_i - \mathbf{Z}_i^T\boldsymbol{\beta})K(\frac{U_i - u_0}{h}), \tag{5.2}$$

where $\rho_\tau(y) = y[\tau - I(y < 0)]$ is the loss function for quantile regression, $\mathbf{Z}_i = (\mathbf{X}_i^T, \mathbf{X}_i^T(U_i - u_0))^T$, $K(\cdot)$ is a bounded kernel function, and $h = h_n > 0$ is a bandwidth parameter. The first-order condition of the minimisation yields the sample estimating equation

$$\ell_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \phi_\tau(T_i - \mathbf{Z}_i^T\boldsymbol{\beta})\mathbf{Z}_i^* K(\frac{U_i - u_0}{h}) = 0, \tag{5.3}$$

where $\phi_\tau(y) = \tau - I(y < 0)$, the derivative function of $\rho_\tau(y)$, $\mathbf{Z}_i^* = (\mathbf{X}_i^T, \mathbf{X}_i^T(U_i - u_0)/h)^T = \mathbf{H}^{-1}\mathbf{Z}_i$, $\mathbf{H} = \mathbf{I}_p \otimes diag(1, h)$, and $\mathbf{I}_p$ is a p-dimensional identity matrix.

Our estimation is based on the Epanechnikov kernel $K(u) = 3(1 - u^2)(|u| \leq 1)/4$ with bandwidth $h = 0.3$. We consider $\tau = 0.25, 0.50, 0.75$. At each specified level of τ, we first estimate a full model containing all of the twelve housing attributes as in (5.1). We define a significant housing attribute as one whose coefficient's interval estimate at the 95%

Table 3: Significant housing attributes at the three quantiles

	AGE	DIR	$SEAVF$	$SEAVS$	$SEAVM$	$GARDV$
$\tau = 0.25$	$\surd$	$\surd$	$\surd$	$\surd$	–	–
$\tau = 0.50$	–	$\surd$	$\surd$	$\surd$	–	–
$\tau = 0.75$	–	$\surd$	$\surd$	$\surd$	–	$\surd$
	$MONV$	$STRN$	$STRI$	$INDUSTV$	$CEMV$	$LGAREA$
$\tau = 0.25$	–	–	$\surd$	–	–	$\surd$
$\tau = 0.50$	–	–	–	$\surd$	$\surd$	$\surd$
$\tau = 0.75$	–	–	$\surd$	$\surd$	$\surd$	$\surd$

confidence level based on the full model does not contain 0 for at least 80% of the values of U. Table 3 shows the significant housing attributes at $\tau = 0.25, 0.50$ and 0.75.

One interesting finding is that not all housing attributes are significant at all three levels of τ, corroborating our postulation of the heterogeneity of marginal effects of housing attributes across the distribution of housing prices. For example, AGE has a significant impact on housing prices only at $\tau = 0.25$, whereas $GARDV$ is a significant attribute only at $\tau = 0.75$, which implies that the age of a building generally only affects prices of low-market dwellings and the view of a garden generally concerns only buyers of up-market dwellings. On the other hand, $INDUSTV$ and $CEMV$ are significant at τ=0.50 and τ=0.75, but insignificant at τ=0.25, indicating that views of the industrial plants and cemetery affect medium- and higher-priced dwellings, but are unlikely to affect lower-priced dwellings. Attributes including DIR, $SEAVF$, $SEAVS$ and $LGAREA$ significantly impact dwelling prices at all three levels of τ, whereas the opposite is observed for $SEAVM$, $MONV$ and $STRN$. Interestingly, $STRI$ is a significant attribute except for $\tau = 0.50$, indicating that internal street view is an important factor for both lower- and higher-priced dwellings but not for medium-priced dwellings.

Figures 5-7 provide plots of the estimated coefficient functions, and the corresponding 95% confidence intervals of the coefficients for the three quantiles. Some discussions of the results are in order. The attribute AGE affects only the lower-priced dwellings, and its effect is negative and tends to decrease as floor level increases (Figure 5(b)). At all three levels of τ, DIR is found to be a more important attribute for lower and higher floor condos than for middle floor condos (Figures 5(c), 6(b) and 7(b)). Generally speaking, the attribute $SEAVF$ has a positive, and relatively stable, effect on prices; as is expected, its effect is usually stronger than that of $SEAVS$ (Figures 5(d), 6(c), 7(c), 5(e), 6(d) and

7(d)). Exceptions occur at very high floor levels, where the impact of $SEAVS$ is found to increase substantially and can sometimes exceed that of $SEAVF$. A view of the garden, represented by $GARV$, affects only higher-priced dwellings, and its effect on prices is positive except for very high floor condos (Figure 7(e)). The attribute $STRI$ representing internal street view affects only lower-priced and higher-priced condos and its impact on prices is largely negative (Figures 5(f) and 7(f)). On the other hand, views of industrial plants and cemeteries, represented by $INDUSTV$ and $CEMV$ respectively, have negative effects only on the medium- and higher-priced condos (Figures 6(e), 6(f), 7(g) and 7(h)). The attribute $LGAREA$ affects prices positively for condos at all three price tiers and its effect is most substantial for high floor condos (Figures 5(g), 6(g) and 6(i)).

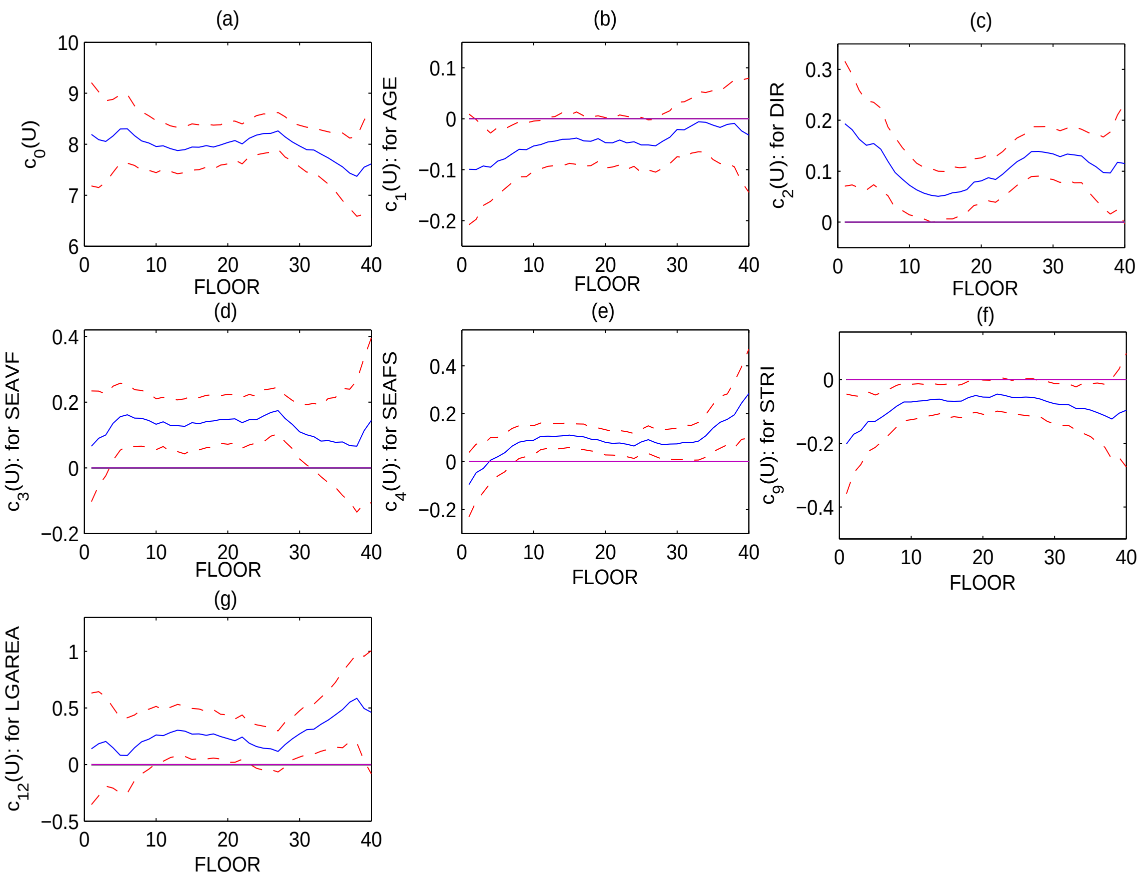

Figure 5: Plots of VC-QR estimated coefficient functions (solid curve) with 95% confidence interval (dashed curves) for $\tau = 0.25$.

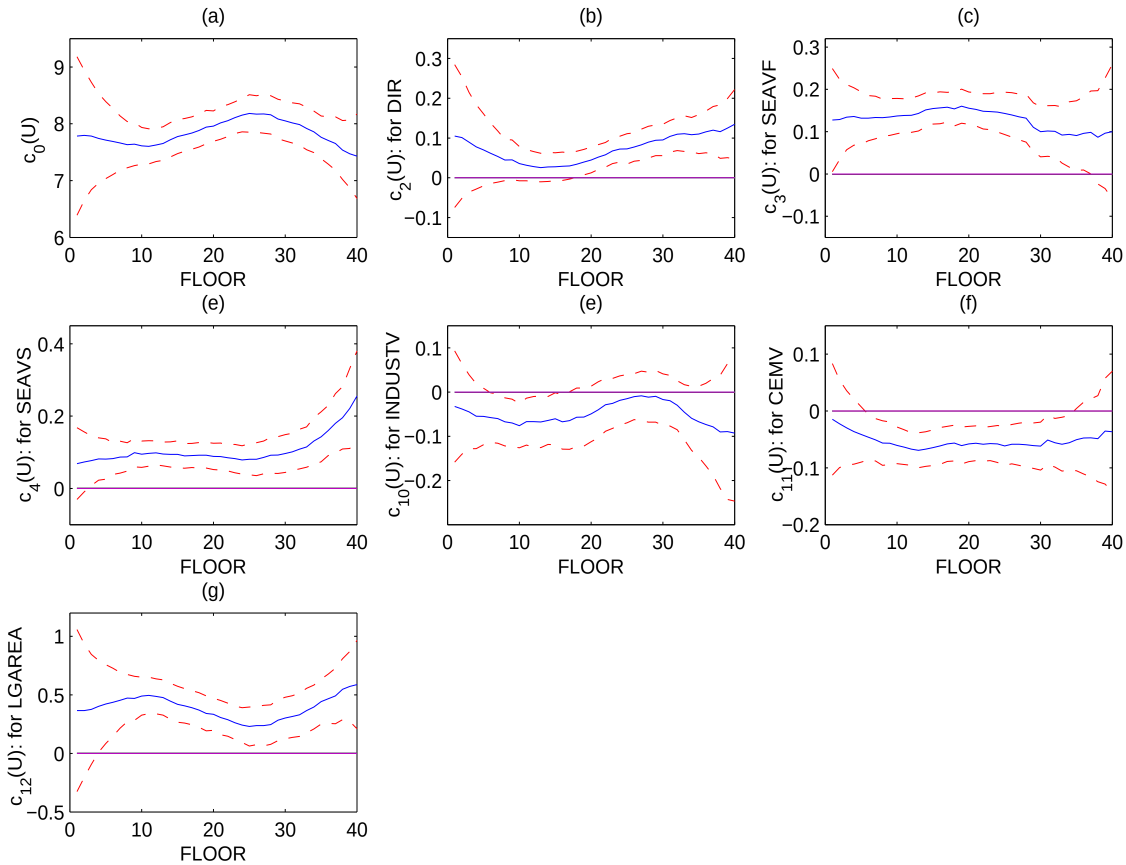

Figure 6: Plots of VC-QR estimated coefficient functions (solid curve) with 95% confidence interval (dashed curves) for $\tau = 0.50$.

6 Conclusions

Recent years have seen an increase in the application of nonparametric and semi-parametric techniques in economics. Unfortunately, the literature on applied econometrics has not kept pace with the theoretical advances, the varying coefficient model being one such important development. This paper has taken steps in this direction by demonstrating the advantages of VC modeling in hedonic house price modeling using data from the Hong Kong property market. We have also utilised a VC-QR approach to examine the heterogeneity of the marginal effects of housing attributes across the distribution of housing prices. We have found that price behaviour can often be vastly heterogenous across dwellings at different ends of the market.

As discussed, the main advantages of the VC model lie in its ability to avoid the 'curse of dimensionality' and its easy interpretability similar to an ordinary linear regression. There have been some important theoretical advances in the VC modeling approach in the statistics literature that are largely ignored by econometricians. The paper is an attempt to put the case for the inclusion of the VC model in the econometrician's repertoire. It is our hope that this paper will help increase the awareness of applied economists and real estate analysts to this useful modeling approach. Certainly, further exploration of this approach to modeling economic data seems justified.

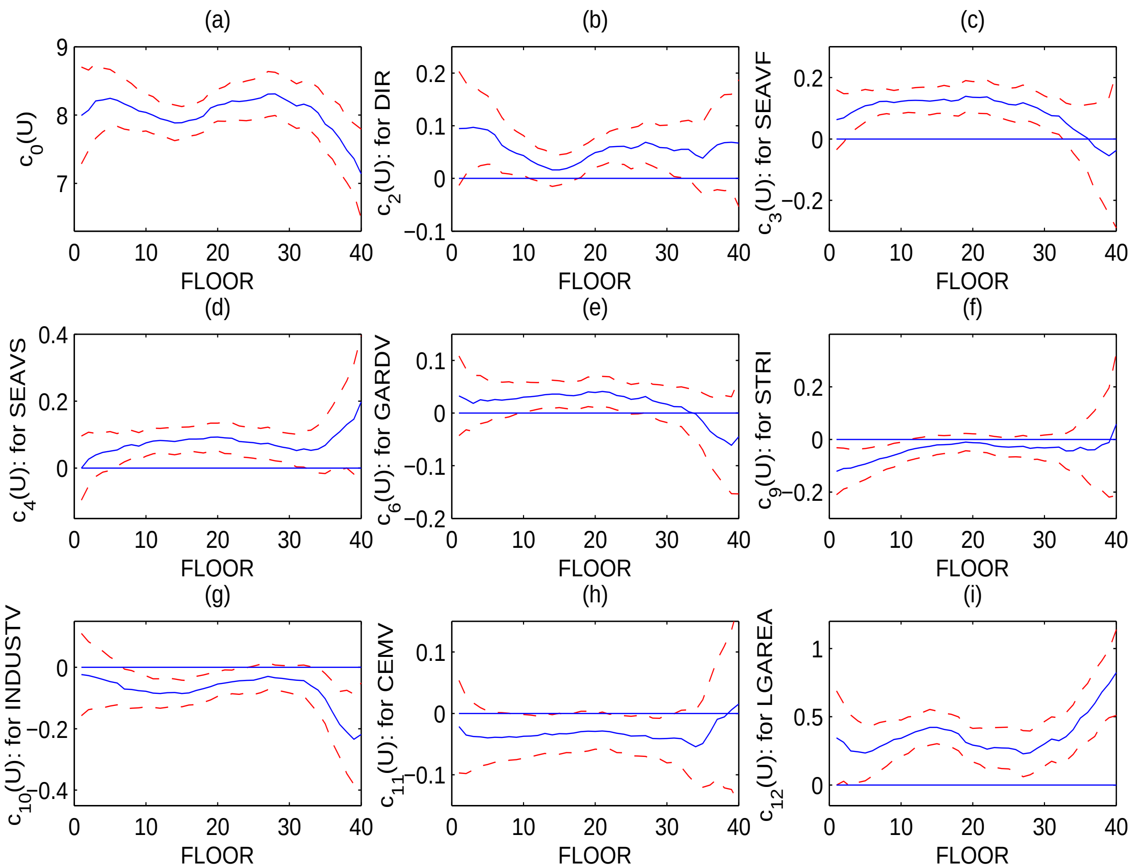

Figure 7: Plots of VC-QR estimated coefficient functions (solid curve) with 95% confidence interval (dashed curves) for $\tau = 0.75$.

References

Anglin, P. and Gencay, R. (1996). Semiparametric estimation of a hedonic price function. *Journal of Applied Econometrics*, **11**, 633-648.

Bucchianeri, G.W. (2008). Has Sars infected the housing market? Evidence from Hong Kong. *Journal of Urban Economics*, **63**, 74–95.

Bao, H.X.H. and Wan, A.T.K. (2004). On the use of spline smoothing in estimating hedonic housing price models: empirical evidence using Hong Kong data. *Real Estate Economics*, **32**, 487–597.

Bao, H.X.H. and Wan, A.T.K. (2007). Improved estimators of hedonic housing price models. *Journal of Real Estate Research*, **29**, 267–302.

Cai, Z. (2007). Trending time varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, **136**, 163–188.

Cai, Z. and Xu, X. (2009). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, **103**, 1595–1608.

Camlong-Viot, C., Rodriguez-Poo, J. and Vieu, P. (2006). Nonparametric and semiparametric estimation of additive models with both discrete and continuous variables under dependence. In *Current Advances and Trends in Nonparametric Statistics*. Physica-Verlag, Heidelberg, 155–178.

Diebold, F.X. and Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, **13**, 134-144.

Fan, J. (1993). Local linear regression smothers and their minimax. *Annals of Statistics*, **21**, 196–216.

Fan, J. and Gijbels, I., (1996). *Local Polynomial Modeling and Its Applications*, Chapman and Hall, London.

Fan, J., Härdle, W. and Mammen, E. (1998). Direct estimation of low-dimensional components in additive models. *Annals of Statistics*, **26**, 943–971.

Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*, **99**, 710–723.

Fan, J., Lin, H. and Zhou, Y. (2006). Local partial-likelihood estimation for lifetime data. *Annals of Statistics*, **34**, 290–325.

Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics*, **27**, 715–731.

Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface*, **1**, 179–195.

Goodman, A.C., (1978). Hedonic price, price indices and housing markets. *Journal of Urban Economics*, **5**, 471–484.

Halvorsen, R., and Pollakowski, H., (1981). Choice of functional form for hedonic price equations. *Journal of Urban Economics*, **10**, 37–47.

Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models*, Chapman and Hall, New York.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society (Series B)*, **55**, 757-796.

Henderson, D.J. and Parmeter, C.F. (2015). *Applied Nonparametric Econometrics*, Cambridge University Press, New York.

Honda, T., (2004). Quantile regression in varying coefficient models. *Journal of Statistical Planning and Inference*, **121**, 113–125.

Kain, J.F. and Quigley, J.M. (1970). Measuring the value of housing quality. *Journal of the American Statistical Association*, **65**, 532–548.

Kim, H.G., Hung, K.C. and Park, S.Y. (2015). Determinants of housing prices in Hong Kong: A Box-Cox quantile regression approach. *Journal of Real Estate Finance and Economics*, **50**, 270–287.

Kim, M.O. (2007). Quantile regression with varying-coefficients. *Annals of Statistics*, **35**, 92–108.

Knight, J.R., Dombrow, J. and Sirmans, C.F., (1995). A varying-coefficient approach to constructing house price indexes. *Real Estate Economics*, **23**, 187–205.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, **74**, 132–157.

Li, Q. and Racine, J.S. (2007). *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, New Jersey.

Mak, S., Choy, L. and Ho, W. (2010). Quantile regression estimates of Hong Kong real estate prices. *Urban Studies*, **47**, 2461-2472.

Magnus, J.R., Wan, A.T.K. and Zhang, X. (2011) Weighted average least squares estimation with nonspherical disturbances and an application to the Hong Kong housing market. *Computational Statistics and Data Analysis*, **55**, 1331-1341.

Martins-Filho, C. and Bin, O. (2005). Estimation of hedonic price functions via additive nonparametric regression. *Empirical Economics*, **30**, 99-114.

Mason, C. and Quigley, J.M. (1996). Nonparametric hedonic housing prices. *Housing Studies*, **11**, 387-406.

McMillen, D.P. and Thorsnes (2000). The reaction of housing prices to information on superfund sites: a semiparametric analysis of the Tacoma, Washington market. *Advances in Econometrics*, **14**, 201–228.

Messe, R. and Wallace, N. (1991). Nonparametric estimation of dynamic hedonic price models and the construction of residential housing price indices. *AREUEA Journal*, 19, 308-332.

Pace, R.K., (1993). Nonparametric methods with applications to hedonic models. *Journal of Real Estate Finance and Economics*, **7**, 185–204.

Pace, R.K., (1995). Parametric, semiparametric, and nonparametric estimation of characteristic values within mass assessment and hedonic pricing models. *Journal of Real Estate Finance and Economics*, **11**: 195–217.

Pagan, A.R. and Ullah, A. (1999). *Nonparametric Econometrics*, Cambridge University Press, New York.

Park, B.U., Mammen, E., Lee, Y.K. and Lee, E.R. (2015). Varying coefficient regression models: a review and new development, *International Statistical Review*, **83**, 36–64.

Parmeter, C.F., Henderson, D.L. and Kumbhakar, S.C. (2007). Nonparametric estimation of a hedonic price function. *Journal of Applied Econometrics*, **22**, 695-699.

Pavlov, A.D. (2000). Space-varying regression coefficients: a semiparametric approach applied to real estate market. *Real Estate Economics*, **28**, 249-283.

Racine, J.S. and Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, **119**, 99130.

Rosen, S., (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, **82**, 34–55.

Ruppert, D. and Wand, M.P. (1994). Multivariate weighed least squares regression. *Annals of Statistics*, **22**, 1346–1370.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society (Series B)*, **50**, 413–436.

Stone, C.J. (1986). The dimensionality reduction principle for generalized additive models. *Annals of Statistics*, **14**, 590–606.

Sunding, D.L. and Swoboda, A.M. (2010). Hedonic analysis with locally weighted regression: an application to the shadow cost of housing regulation in Southern California. *Regional Science and Urban Economics* **40**, 550-573.

Witte, A.D., Sumba, H.J. and Erekson, H. (1979). An estimate of a structural hedonic price model of the housing market: an application of Rosen's theory of implicit markets. *Econometrica*, **47**, 1151-1173.

Xie, S., Zhou, Y. and Wan, A.T.K. (2014). A varying-coefficient expectile model for estimating Value at Risk. *Journal of Business and Economic Statistics*, **32**, 576–592.

Yatchew, A. (2003). *Semiparametric Regression for the Applied Econometrician*, Cambridge University Press, New York.

Zhang, W., Lee, S.Y. and Song, X. (2002). Local polynomial fitting in semivarying coefficient model. *Journal of Multivariate Analysis*, **82**, 166–188.

Zietz, J., Zietz, E.N. and Sirmans, C.F. (2008). Determinants of housing prices: a quantile regression approach. *Journal of Real Estate Finance and Economics*, **37**, 317–333.