

MÔ HÌNH HEDONIC VÀ PHẦN MỀM CHO BÀI TOÁN XÁC ĐỊNH GIÁ ĐẤT, CÁC YẾU CỐ ẢNH HƯỞNG ĐẾN GIÁ ĐẤT

Trần Đức Quỳnh^{1*}, Bùi Nguyên Hạnh²

¹*Khoa Công nghệ Thông tin, Học viện Nông nghiệp Việt nam*

²*Khoa Quản lý đất đai, Học viện Nông nghiệp Việt Nam*

Email*: tdquynh@vnua.edu.vn

Ngày gửi bài: 22.07.2015

Ngày chấp nhận: 03.09.2015

TÓM TẮT

Ước lượng giá đất và xác định các yếu tố ảnh hưởng đến giá đất là một vấn đề hết sức quan trọng trong quy hoạch và sử dụng đất, trong định giá tài sản cầm cố cho vay của ngân hàng,... Hiện nay chúng ta ước lượng giá đất dựa trên các phương pháp truyền thống. Các phương pháp này chủ yếu nhờ sự phân tích và can thiệp của nhân viên định giá nên rất khó tránh khỏi hiện tượng sai lầm do chủ quan hoặc không minh bạch. Bài báo này trình bày phương pháp định giá và lựa chọn các đặc trưng ảnh hưởng đến giá đất dựa trên mô hình Hedonic. Đây là một mô hình được phát biểu tổng quát bởi Lancaster từ năm 1966 và đã được ứng dụng rộng rãi trong nhiều lĩnh vực kinh tế, bao gồm cả định giá đất. Trên mỗi địa bàn, người ta thường lựa chọn các mô hình cụ thể khác nhau và thử nghiệm chúng. Trong nghiên cứu này, chúng tôi xét hai dạng mô hình là mô hình tuyến tính và mô hình bậc hai kèm theo kĩ thuật phạt để có thể lựa chọn các đặc trưng ảnh hưởng đến giá đất. Việc tìm các tham số cho mô hình này sao cho sai số là nhỏ nhất được thực hiện nhờ việc giải một bài toán tối ưu lồi. Chúng tôi cũng đã phát triển một chương trình máy tính thử nghiệm cho phép tính toán tối ưu các hệ số của mô hình và ước lượng sai số. Kết quả số với số liệu được điều tra trên địa bàn quận Long Biên cho thấy phương pháp đề xuất là chấp nhận được và rất có tiềm năng phát triển.

Từ khóa: Định giá đất, lựa chọn đặc trưng, mô hình hedonic, tối ưu lồi.

Hedonic Model and Software for Land Pricing and Selecting Key Feature of Land Price

ABSTRACT

Land pricing and determining the key factors of the land price are useful for land use planning and for the banking. At present, land price estimation is relied on traditional methods. These methods largely depend on analysis and intervention by pricing staff, thus, may not be free of subjective mistakes and non-transparency. In this study, we proposed a method for land pricing based on hedonic model. This model was introduced by Lancaster in 1966 and has many applications in economics including land pricing. Usually, for each region, various particular mathematical models are chosen and tested. In this paper, we investigated a linear model and a second order model in combination with penalty technique to select the key factors affecting land price. The optimal parameters were obtained by solving convex optimization problems. A demo software to calculate the optimal parameters and to evaluate the error was developed. Numerical results with data collected from Long Bien district showed that the proposed method is potentially promising.

Keywords: Convex optimization, feature selection, Hedonic model, land pricing.

1. ĐẶT VẤN ĐỀ

Định giá đất được hiểu là sự ước tính về giá trị của quyền sử dụng đất bằng hình thái tiền tệ đối với một thửa đất (khu đất) cụ thể cho một mục đích sử dụng đã được xác định, tại một thời điểm xác định.

Ở Việt Nam, Luật đất đai 1993 quy định: “Nhà nước xác định giá các loại đất để tính thuế chuyển quyền sử dụng đất, thu tiền khi giao đất hoặc cho thuê đất, tính giá trị tài sản khi giao đất, bồi thường thiệt hại khi thu hồi”... Từ đó giá đất và việc xác định giá đất trở thành một trong những nội dung quan trọng của quản lý nhà nước về đất đai.

Nghị định số 44/2014/NĐ - CP ngày 15 tháng 05 năm 2014 quy định về các phương pháp định giá đất tại Việt Nam bao gồm 5 phương pháp:

1. *Phương pháp so sánh trực tiếp*: định giá đất thông qua việc phân tích mức giá của các thửa đất trống tương tự về mục đích sử dụng đất, vị trí, khả năng sinh lợi, điều kiện kết cấu hạ tầng, diện tích, hình thể, tính pháp lý về quyền sử dụng đất (sau đây gọi là thửa đất so sánh) đã chuyển nhượng trên thị trường, trúng đấu giá quyền sử dụng đất để so sánh, xác định giá của thửa đất cần định giá.

Phương pháp so sánh trực tiếp được áp dụng để định giá đất khi trên thị trường có các thửa đất so sánh đã chuyển nhượng trên thị trường, trúng đấu giá quyền sử dụng đất;

Nhược điểm của phương pháp này xuất phát từ cơ chế hai giá trong thị trường đất đai Việt Nam, giá khai trong hợp đồng mua bán chuyển nhượng thường không sát với giá thực tế trên thị trường, chính vì vậy, việc dùng giá này để so sánh và xác định giá của thửa đất chờ định giá sẽ khó có kết quả chính xác. Bên cạnh đó, khó có hai thửa đất có đặc điểm hoàn toàn giống nhau để so sánh, vì vậy người định giá dùng kinh nghiệm của mình để hiệu chỉnh các sai khác, do vậy khó tránh khỏi sự chủ quan áp đặt, dẫn đến việc cùng một mảnh đất nhưng mỗi người định giá sẽ đưa ra một mức giá khác nhau.

2. *Phương pháp chiết trừ*: định giá đất đối với thửa đất có tài sản gắn liền với đất bằng

cách loại trừ phần giá trị tài sản gắn liền với đất ra khỏi tổng giá trị bất động sản (bao gồm giá trị đất và giá trị tài sản gắn liền với đất).

Phương pháp chiết trừ được áp dụng để định giá đối với thửa đất có tài sản gắn liền với đất trong trường hợp có đủ số liệu về giá các bất động sản (gồm đất và tài sản gắn liền với đất) tương tự với thửa đất cần định giá đã chuyển nhượng trên thị trường, trúng đấu giá quyền sử dụng đất;

Về cơ bản, cơ chế của phương pháp này cũng là dùng nguyên tắc so sánh, nên những khó khăn vướng mắc của phương pháp so sánh trực tiếp cũng là những khó khăn của phương pháp này.

3. *Phương pháp thu nhập*: định giá đất tính bằng thương số giữa mức thu nhập ròng thu được bình quân một năm trên một đơn vị diện tích đất so với lãi suất tiền gửi tiết kiệm bình quân một năm tính đến thời điểm định giá đất của loại tiền gửi VNĐ kỳ hạn 12 tháng tại ngân hàng thương mại nhà nước có mức lãi suất tiền gửi tiết kiệm cao nhất trên địa bàn cấp tỉnh.

Phương pháp thu nhập được áp dụng để định giá đối với thửa đất xác định được các khoản thu nhập, chi phí từ việc sử dụng đất;

Phương pháp này tính giá đất dựa vào thu nhập mang lại từ mảnh đất đó, nhưng trong thực tế cùng một mảnh đất nhưng các chủ sử dụng khác nhau sử dụng thì tùy vào năng lực của từng người chủ đầu tư mà kết quả kinh doanh (thu nhập) là khác nhau, như vậy dẫn đến việc cùng một mảnh đất nhưng khi xác định giá theo phương pháp thu nhập đối với các người sử dụng khác nhau lại cho các mức giá khác nhau, thậm chí nếu như người chủ đất làm ăn thua lỗ (không có thu nhập, hoặc có thể âm) thì giá đất là bằng 0 hoặc giá đất âm?

4. *Phương pháp thặng dư*: định giá đất đối với thửa đất có tiềm năng phát triển theo quy hoạch sử dụng đất, quy hoạch xây dựng hoặc được phép chuyển mục đích sử dụng đất để sử dụng cao nhất và tốt nhất bằng cách loại trừ tổng chi phí ước tính ra khỏi tổng doanh thu phát triển giả định của bất động sản.

Phương pháp thẳng dư được áp dụng để định giá đối với thửa đất có tiềm năng phát triển do thay đổi quy hoạch hoặc chuyển mục đích sử dụng đất khi xác định được tổng doanh thu phát triển giả định và tổng chi phí ước tính;

Độ chính xác của phương pháp này tùy thuộc vào việc ước tính giá trị của tài sản trong tương lai. Xác định được giá đất hiện tại xác định vốn đã là một vấn đề rất khó khăn, vì vậy việc ước tính tổng giá trị tài sản không phải ở thời điểm hiện tại mà là ở tương lai xa với rất nhiều biến động, quy hoạch mà tại thời điểm định giá không thể lường được lại càng khó khăn hơn nhiều, do vậy mức giá đưa ra khả năng có độ chính xác không cao là rất lớn.

5. *Phương pháp hệ số điều chỉnh giá đất* là phương pháp định giá đất bằng cách sử dụng hệ số điều chỉnh giá đất nhân (x) với giá đất trong bảng giá đất do Ủy ban nhân dân tỉnh, thành phố trực thuộc Trung ương (sau đây gọi là Ủy ban nhân dân cấp tỉnh) ban hành.

Phương pháp hệ số điều chỉnh giá đất được áp dụng để định giá đất đối với trường hợp:

- Thuê đất thu tiền hàng năm mà phải xác định lại đơn giá thuê đất để điều chỉnh cho chu kỳ tiếp theo; giá khởi điểm để đấu giá quyền sử dụng đất khi Nhà nước cho thuê đất thu tiền thuê đất hàng năm;

- Trường hợp tính tiền bồi thường khi Nhà nước thu hồi đất đối với dự án có các thửa đất liền kề nhau, có cùng mục đích sử dụng, khả năng sinh lợi và thu nhập từ việc sử dụng đất tương tự nhau.

- Phương pháp hệ số điều chỉnh giá đất cũng được áp dụng để định giá đất cho các thửa đất hoặc khu đất của dự án có giá trị (tính theo giá đất trong bảng giá đất) dưới 30 tỷ đồng đối với các thành phố trực thuộc Trung ương; dưới 10 tỷ đồng đối với các tỉnh miền núi, vùng cao; dưới 20 tỷ đồng đối với các tỉnh còn lại đối với các trường hợp sau:

- Tính tiền sử dụng đất khi Nhà nước công nhận quyền sử dụng đất của hộ gia đình, cá nhân đối với phần diện tích đất ở vượt hạn mức; cho phép chuyển mục đích sử dụng đất từ đất nông nghiệp, đất phi nông nghiệp không phải là đất ở

sang đất ở đối với phần diện tích vượt hạn mức giao đất ở cho hộ gia đình, cá nhân.

- Tính tiền thuê đất đối với đất nông nghiệp vượt hạn mức giao đất, vượt hạn mức nhận chuyển quyền sử dụng đất nông nghiệp của hộ gia đình, cá nhân;

- Tính tiền sử dụng đất khi Nhà nước giao đất có thu tiền sử dụng đất không thông qua hình thức đấu giá quyền sử dụng đất; công nhận quyền sử dụng đất, cho phép chuyển mục đích sử dụng đất đối với tổ chức mà phải nộp tiền sử dụng đất;

- Tính tiền thuê đất đối với trường hợp Nhà nước cho thuê đất không thông qua hình thức đấu giá quyền sử dụng đất;

- Tính giá trị quyền sử dụng đất khi cổ phần hóa doanh nghiệp nhà nước mà doanh nghiệp cổ phần sử dụng đất thuộc trường hợp Nhà nước giao đất có thu tiền sử dụng đất, cho thuê đất trả tiền thuê đất một lần cho cả thời gian thuê; tính tiền thuê đất đối với trường hợp doanh nghiệp nhà nước cổ phần hóa được Nhà nước cho thuê đất trả tiền thuê đất hàng năm;

- Tổ chức kinh tế, tổ chức sự nghiệp công lập tự chủ tài chính, hộ gia đình, cá nhân, người Việt Nam định cư ở nước ngoài, doanh nghiệp có vốn đầu tư nước ngoài đang được Nhà nước cho thuê đất trả tiền thuê đất hàng năm được chuyển sang thuê đất trả tiền thuê đất một lần cho cả thời gian thuê và phải xác định lại giá đất cụ thể để tính tiền thuê đất tại thời điểm có quyết định cho phép chuyển sang thuê đất theo hình thức trả tiền thuê đất một lần cho cả thời gian thuê

- Người mua tài sản được Nhà nước tiếp tục cho thuê đất trong thời hạn sử dụng đất còn lại theo giá đất cụ thể, sử dụng đất đúng mục đích đã được xác định trong dự án.

Phương pháp này về cơ bản cũng là phương pháp so sánh trực tiếp, nên những khó khăn của phương pháp so sánh trực tiếp cũng chính là khó khăn của phương pháp này.

Ngoài các phương pháp truyền thống như các phương pháp đang được sử dụng ở Việt nam, nhiều nước phát triển trên thế giới hiện nay như Anh, Pháp, Mỹ đang nghiên cứu và áp dụng khá rộng rãi các phương pháp có sử dụng đến các mô

hình toán học để xác định giá trị đất đai. Từ năm 1963, Bailey và các cộng sự của ông đã đưa ra phương pháp xác định giá trị của một bất động sản dựa vào một hàm hồi quy tuyến tính (Bailey et al., 1963). Đến năm 1964, Alonso một nhà nghiên cứu nổi tiếng trong lĩnh vực bất động sản cho rằng giá trị của một bất động sản phụ thuộc vào yếu tố vị trí của nó (Alonso, 1964). Ông được người ta nhắc đến với câu nói: *vị trí, vị trí và vị trí* để xác định giá trị của một bất động sản. Năm 1966, trong lý thuyết về khách hàng, Lancaster đưa ra một mô hình mà theo đó thì giá trị của một tài sản là phụ thuộc vào một số đặc tính của tài sản đó (Lancaster, 1966). Đến năm 1974, Rosen tận dụng các ý tưởng của Bailey và Lancaster để đưa ra một mô hình toán học gọi là mô hình hodenic để xác định giá trị của một loại sản phẩm và phân tích giá trị cân bằng trên thị trường của sản phẩm đó (Rosen, 1974). Trong suốt các quãng thời gian sau đó có nhiều nghiên cứu ứng dụng mô hình hodenic vào việc xác định giá bất động sản được thực hiện bởi các nhà nghiên cứu ở các quốc gia khác nhau. Trong các công trình đó chúng ta có thể kể đến công trình của Englund (1998) nghiên cứu về các phương pháp định giá bất động sản trên thị trường Thụy Điển. Công trình của Zhou and Sornette (2008) nghiên cứu, phân tích thị trường bất động sản và bong bóng bất động sản ở Lasvegas, Mỹ, công trình của Gouriéroux and Laferrère (2009) phân tích về mô hình hodenic để định giá bất động sản và các kết quả có được từ thị trường bất động sản của Pháp.

Lợi thế của mô hình hodenic là tính linh hoạt, có thể sử dụng ngay được các dữ liệu giao dịch đang diễn ra và dữ liệu về các đặc tính của một bất động sản có thể được khai thác ở rất nhiều nguồn khác nhau, kết quả của phương pháp này được đánh giá là đáng tin cậy cho việc xác định giá trị của một bất động sản. Ngoài những ưu điểm trên phương pháp sử dụng mô hình hodenic còn có thể giúp cho chúng ta dự đoán được bong bóng cục bộ của thị trường bất động sản khi giá trị giao dịch trên thị trường của khu vực đó cao hơn nhiều so với giá trị tính toán cho bởi mô hình. Ngoài ra thì khi xây dựng mô hình và đánh giá sai số chúng ta có thể lựa

chọn được các đặc trưng ảnh hưởng nhiều đến giá đất trên quan điểm sử dụng ít đặc tính nhất và sai số là nhỏ nhất.

2. VẬT LIỆU VÀ PHƯƠNG PHÁP

2.1. Phương pháp nghiên cứu

Để hoàn thành nghiên cứu này chúng tôi sử dụng các phương pháp sau đây

Phương pháp phân tích: phương pháp này sử dụng trong nghiên cứu các yếu tố ảnh hưởng đến giá trị của một khu đất và lựa chọn dạng hàm hồi quy phù hợp.

Phương pháp thống kê số liệu: sử dụng trong thu thập số liệu phục vụ việc thử nghiệm mô hình, số liệu được lấy bao gồm các số liệu sơ cấp được điều tra trên địa bàn quận Long Biên. Ngoài các số liệu sơ cấp chúng tôi còn sử dụng điều tra số liệu thứ cấp, các số liệu thứ cấp này là thông tin về các thửa đất được giao dịch trên địa bàn quận Long biên trong năm 2013 và 2014.

Phương pháp chuyên gia: được sử dụng để đưa ra các yếu tố có ảnh hưởng đến giá đất và thiết kế bảng hỏi phục vụ cho việc điều tra số liệu sơ cấp. Chúng tôi sử dụng ý kiến chuyên gia trong các lĩnh vực quản lý đất đai và thị trường bất động sản.

Phương pháp mô hình: Mô hình được sử dụng trong nghiên cứu là các mô hình toán học. Các mô hình này bao gồm mô hình Hedonic và mô hình hồi quy tuyến tính, mô hình hồi quy bậc hai và mô hình lựa chọn đặc trưng; đây là một mô hình kết hợp giữa mô hình hedonic, mô hình hồi quy, mô hình tối ưu.

2.2. Vật liệu nghiên cứu

Để thử nghiệm các phương pháp và mô hình đề xuất, chúng tôi sử dụng bộ số liệu được điều tra trên địa bàn quận Long Biên. Bộ số liệu này bao gồm thông tin của 82 thửa đất (bao gồm cả công trình trên đất nếu có), các thông tin này bao gồm: giá giao dịch, diện tích của khu đất, vị trí của khu đất chia theo quy định của Bộ tài nguyên và môi trường, độ rộng đường vào nhà, độ thuận tiện của lối vào nhà (0: đi bộ, 1: đi xe máy, 2: đi được ô tô), khoảng cách đến trung

tâm thành phố, khoảng cách đến trường học gần nhất, đánh giá chất lượng trường học, đánh giá chất lượng dịch vụ y tế, đánh giá tình trạng số đỏ (có hoặc không có), khoảng cách đến chợ gần nhất, khoảng cách đến trung tâm quận, độ rộng mặt tiền của thửa đất, tổng diện tích sàn của nhà, đặc điểm nhà (kiên cố, bán kiên cố, tạm).

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Phương pháp mô hình Hedonic cho định giá đất

Một cách tổng quát, trong mô hình hoderic giá của một thửa đất (có thể bao gồm cả các công trình trên thửa đất) phụ thuộc vào các đặc trưng của miếng đất như vị trí so với trung tâm, gần đường, gần các khu tiện ích và giá trị của các công trình trên miếng đất đó như diện tích nhà, số phòng ngủ, số tầng,... Mục tiêu của mô hình là xác định được một biểu diễn về giá của miếng đất dựa vào những đặc trưng nói trên. Các mô hình để xác định giá có thể là các mô hình đơn giản như mô hình tuyến tính hay các mô hình phức tạp hơn như mô hình mũ, mô hình loga... Việc lựa chọn các mô hình này sẽ được đánh giá và tùy theo từng bộ số liệu cho phù hợp. Trong nghiên cứu này của chúng tôi, hai dạng hàm giá được xây dựng và so sánh.

a. Mô hình tuyến tính

Mô hình tuyến tính là một mô hình đơn giản và được sử dụng nhiều trong bài toán xác định giá đất. Trong các nghiên cứu về giá đất có sử dụng đến mô hình tuyến tính chúng ta có thể kể đến các nghiên cứu của Gouriéroux, Laferrère (2009); Bartik (1987); Cebula (2009).

Giả sử giá của mảnh đất là Y và các đặc tính ảnh hưởng đến giá của mảnh đất như diện tích của thửa đất, độ rộng mặt tiền, độ rộng đường vào nhà, tình trạng pháp lý của khu đất, tiện ích của khu dân cư (điều kiện vệ sinh, điều kiện trường học, y tế), khoảng cách đến trung tâm phường, quận, thành phố... được lượng hóa và kí hiệu là $X_1, X_2, \dots, X_n$. Ta cần xây dựng hàm giá của mảnh đất là một hàm tuyến tính theo các biến trên, nghĩa là ta cần tìm một công thức:

$$Y = a_0 + \sum_{k=1}^n a_k \cdot X_k$$

Qua điều tra số liệu ta thu thập được m bộ số liệu và giả sử $Y^i, X_1^i, X_2^i, \dots, X_n^i, i=1, 2, \dots, m$ là các số liệu thu thập được của bộ số liệu thứ i . Ta cần đi tìm các hệ số $a_i, i=0, 1, 2, \dots, n$ sao cho sai số là nhỏ nhất. Điều này dẫn đến việc giải một bài toán tối ưu như sau:

$$\min \left\{ \sum_{i=1}^m \left(a_0 + \sum_{k=1}^n a_k X_k^i - Y^i \right)^2 \right\}$$

Đây là một bài toán tối ưu lồi, khả vi và không khó khăn để giải bài toán này bằng nhiều công cụ khác nhau. Cách phổ biến nhất là đi tìm các hệ số của hàm hồi quy tuyến tính theo phương pháp bình phương tối thiểu. Tuy nhiên thực chất là chúng ta muốn đánh giá tổng sai số, trong khi biểu thức trên là tổng bình phương sai số nên chúng tôi thay vì giải quyết bài toán trên chúng tôi sẽ giải quyết bài toán tối ưu như sau:

$$\min \left\{ \sum_{i=1}^m \left| a_0 + \sum_{k=1}^n a_k X_k^i - Y^i \right| \right\}$$

Bài toán này vẫn là một bài toán tối ưu lồi nhưng việc giải nó thì phức tạp hơn vì tính chất không khả vi của hàm mục tiêu.

b. Mô hình bậc hai

Thực tế thì mô hình hồi quy tuyến tính là đơn giản về phương pháp giải nhưng lại khó cho ra một sai số đủ tốt vì hàm giá có thể là một hàm số phi tuyến. Trong mục này chúng tôi đề xuất xấp xỉ hàm giá bằng một hàm bậc hai của các biến. Với các giả thiết và điều kiện nêu như trong phần hàm tuyến tính thì ta phải đi tìm các hệ số cho hàm số sau

$$Y = a_0 + \sum_{k=1}^n a_k \cdot X_k + \sum_{k=1}^n \sum_{l=k}^n a_{kl} \cdot X_k X_l$$

Các hệ số a_k, a_{kl} thu được từ việc giải bài toán tối ưu

$$\min \left\{ \sum_{i=1}^m \left| a_0 + \sum_{k=1}^n a_k X_k^i + \sum_{k=1}^n \sum_{l=k}^n a_{kl} X_k^i X_l^i - Y^i \right| \right\}$$

Mặc dù mô hình bậc 2 tổng quát hơn mô hình tuyến tính nhưng lại có nhược điểm chứa nhiều tham số nên khi dung lượng mẫu không đủ lớn, dễ xảy ra hiện tượng sai số không ổn định.

Trong mục này, chúng tôi đã trình bày về phương pháp mô hình Hedonic và hai dạng mô hình được nghiên cứu trong bài báo này. Lợi thế của mô hình hodenic là tính linh hoạt, có thể sử dụng ngay được các dữ liệu giao dịch đang diễn ra và dữ liệu về các đặc tính của một bất động sản có thể được khai thác ở rất nhiều nguồn khác nhau, kết quả của phương pháp này được đánh giá là đáng tin cậy cho việc xác định giá trị của một bất động sản. Ngoài những ưu điểm trên phương pháp sử dụng mô hình hodenic còn có thể giúp cho chúng ta dự đoán được bong bóng cục bộ của thị trường bất động sản khi giá trị giao dịch trên thị trường của khu vực đó cao hơn nhiều so với giá trị tính toán cho bởi mô hình.

3.2. Lựa chọn đặc trưng ảnh hưởng đến giá đất dựa trên mô hình Hedonic

Trong mục này chúng ta xét tổng quát giá đất là một hàm số của các biến đầu vào các đặc trưng của mảnh đất $Y = f(X_1, X_2, \dots, X_n)$. Với mong muốn sử dụng một số ít các đặc tính mà vẫn có thể xây dựng được hàm giá có sai số chấp nhận được. Hơn nữa trong thực tế khi xây dựng hàm giá nếu chúng ta sử dụng quá nhiều thuộc tính thì các hàm đó có sai số thường không ổn định. Cụ thể sai số trên tập dữ liệu huấn luyện có thể rất nhỏ nhưng sai số trên tập kiểm tra lại rất lớn. Để hạn chế khả năng trên, cần thiết phải xây dựng một quy tắc để chọn ra những đặc tính có ảnh hưởng nhiều nhất đến mô hình. Về mặt lý thuyết, có nhiều cách lựa chọn các đặc tính ảnh hưởng đến giá, chẳng hạn chúng ta có thể tính hiệp phương sai của các đặc tính và giá rồi chọn ra những đặc tính có hiệp phương sai lớn nhất. Cách làm như trên có thể dẫn đến những đặc tính có mối quan hệ tuyến tính với nhau đều có hiệp phương sai với giá rất lớn và đều được lựa chọn, trong khi đó thì chúng ta chỉ cần lựa chọn một đặc tính là đủ. Để hạn chế

nhược điểm trên, người ta đề xuất các phương pháp tính toán từng bước một để thêm vào các đặc trưng dựa trên việc kết hợp tính toán hiệp phương sai và ước lượng giá theo từng biến.

Cách tiếp cận của chúng tôi dựa trên kỹ thuật tối ưu với quan điểm tất cả được phản ánh trong sai số của mô hình, những đặc tính không cần thiết và không được lựa chọn vào mô hình thì có hệ số liên quan đến nó trong mô hình bằng 0. Ta kí hiệu tất cả các tham số trong hàm giá thành một véc tơ $\beta = (\beta_1, \beta_2, \dots, \beta_l)$ và gọi $|\beta|_0$ là số những phần tử khác không của β (chuẩn không). Do đó, cần giải quyết bài toán tối ưu sau đây

$$\min \left\{ \sum_{i=1}^m |f(X_1^i, X_2^i, \dots, X_n^i) - Y^i| + \lambda \cdot |\beta|_0 \right\}$$

Hàm mục tiêu của bài toán là tổng của hai phần, phần đầu chính là sai số của mô hình, phần sau là phần tính số đặc tính được đưa vào mô hình. Mong muốn của chúng ta là số đặc tính ít và sai số là nhỏ nhất. Nếu trong mô hình càng nhiều đặc tính thì giá trị hàm mục tiêu sẽ bị phạt một lượng càng lớn. Hệ số λ là một hệ số dương và được gọi là hệ số phạt, khi giá trị này càng lớn sẽ càng ít đặc tính được đưa vào trong mô hình tính toán.

Việc giải quyết bài toán trên sẽ yêu cầu những thuật toán tối ưu phức tạp vì hàm mục tiêu là không liên tục, thông thường có thể giải quyết nó bằng các thuật toán giải địa phương hoặc xấp xỉ bài toán bằng một bài toán dễ giải hơn mà không làm thay đổi tính chất. Một trong những cách là thay thế chuẩn không bằng chuẩn một, $|\beta|_1 = \sum_k |\beta_k|$ và ta có một bài toán tối ưu lồi, không khả vi

$$\min \left\{ \sum_{i=1}^m |f(X_1^i, X_2^i, \dots, X_n^i) - Y^i| + \lambda \cdot |\beta|_1 \right\}$$

Có nhiều công trình nghiên cứu về xử lý dữ liệu đã cho thấy kết quả tính toán với chuẩn một là rất gần kết quả với chuẩn không. Ngoài ra, về mặt lý thuật người ta chứng minh được bài toán với chuẩn không là tương đương với

$$\min \left\{ \sum_{i=1}^m \left| a_0 + \sum_{k=1}^n a_k X_k^i - Y^i \right| + \lambda \left(|a_0| + \sum_{k=1}^n |a_k| \right) \right\} \quad (1)$$

$$\min \left\{ \sum_{i=1}^m \left| a_0 + \sum_{k=1}^n a_k X_k^i + \sum_{k=1}^n \sum_{l=k}^n a_{kl} X_k^i X_l^i - Y^i \right| + \lambda \left(|a_0| + \sum_{k=1}^n |a_k| + \sum_{k=1}^n \sum_{l=k}^n |a_{kl}| \right) \right\} \quad (2)$$

một bài toán với chuẩn $|\beta|_p = \left(\sum_k |\beta_k|^p \right)^{\frac{1}{p}}$ ($p > 0$ đủ nhỏ).

Áp dụng cho mô hình tuyến tính và mô hình bậc 2 đã nêu trong mục 2.1, lần lượt phải giải các bài toán tối ưu (1), (2):

3.3. Kết quả tính toán và thảo luận

Để thử nghiệm các kết quả tính toán chúng tôi dùng 82 bộ số liệu điều tra trên địa bàn quận Long Biên. Số liệu được thu thập từ các phường Bồ Đề, Ngọc Lâm, Cự Khối, Thượng Thanh, Long Biên, Việt Hưng, Thạch Bàn. Các thông tin sau khi được điều tra và lượng hóa được đưa vào các biến với kí hiệu như sau

- X1: Diện tích khu đất
- X2: Độ rộng đường vào nhà
- X3: Độ thuận tiện của lối vào nhà (0: đi bộ, 1: đi xe máy, 2: đi được ô tô)
- X4: khoảng cách đến trung tâm thành phố
- X5: Khoảng cách đến trường học gần nhất
- X6: Chất lượng vệ sinh của khu phố
- X7: Đánh giá chất lượng trường học
- X8: Đánh giá chất lượng dịch vụ y tế
- X9: Đánh giá trình trạng số đỗ (có hoặc không có)
- X10: Khoảng cách đến chợ gần nhất
- X11: Khoảng cách đến trung tâm quận
- X12: Độ rộng mặt tiền của thửa đất
- X13: Tổng diện tích sàn của nhà
- X14: Đặc điểm nhà (kiên cố, bán kiên cố, tạm)
- Y: Giá bất động sản chuyển nhượng

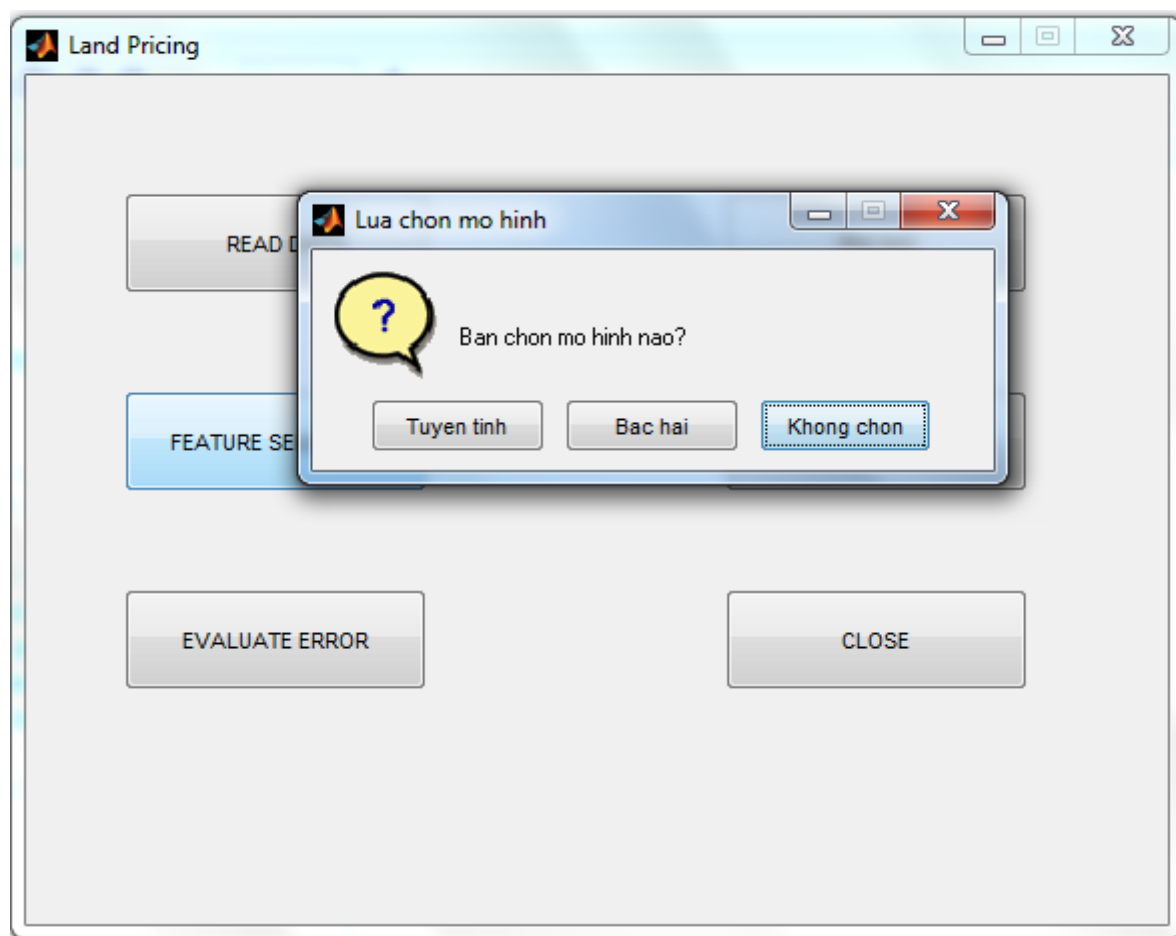
Chúng tôi xét hai dạng mô hình là mô hình tuyến tính và mô hình bậc 2 với hàm toán học biểu diễn giá lần lượt cho bởi công thức sau:

$$Y = a_0 + \sum_{k=1}^{14} a_k \cdot X_k \text{ và}$$

$$Y = a_0 + X_1 \cdot (a_1 + \sum_{k=2}^{12} a_k \cdot X_k) + X_{13} (a_{13} + a_{14} X_{14})$$

Một chương trình máy tính để tính toán các tham số tối ưu và để lựa chọn các đặc trưng ảnh hưởng đến giá đất được xây dựng trên Matlab. Việc giải bài toán tối ưu lỗi (1), (2) được thực hiện bằng cách gọi phần mềm CVX (Stephen Boyd, <http://cvxr.com/cvx/>), đây là một phần mềm chuyên dụng và hiện nay được sử dụng ở hơn 190 trường đại học tại Mỹ. Hình ảnh giao diện chương trình được thể hiện trong hình 1.

Chương trình có các chức năng đọc dữ liệu vào từ một tệp excel. Trích chọn ra các đặc tính ảnh hưởng nhiều nhất đến giá đất và đánh giá sai số. Để lựa chọn được các đặc trưng khác nhau, chúng tôi thay đổi tham số phạt trong các mô hình với các giá trị khác là $\lambda = 0$, $\lambda = 1,0$; $\lambda = 5,0$; $\lambda = 10$; $\lambda = 100$. Kết quả về số đặc tính cần được sử dụng để xây dựng mô hình được tổng hợp ở bảng 1. Sai số trung bình của hai mô hình trên toàn bộ số liệu được tổng hợp trong bảng 2. Mặt khác, để đánh giá sai số của các mô hình phục vụ cho việc ước lượng giá đất chúng tôi lựa chọn ngẫu nhiên một số số liệu trong bộ dữ liệu để xây dựng mô hình (dữ liệu huấn luyện), sau đó đánh giá sai số trung bình với những số liệu còn lại (dữ liệu kiểm tra), quá trình này được lặp đi lặp lại 100 lần và sau đó tính trung bình sai số. Số lượng dữ liệu để xây dựng mô hình có thể lựa chọn ở 3 mức là 70%, 80% hoặc 90% số liệu có trong bộ dữ liệu ban đầu. Kết quả của mô hình tuyến tính được thể hiện trong bảng 3 trong khi kết quả của mô hình bậc hai được thể hiện trong bảng 4.



Hình 1. Hình ảnh giao diện của chương trình

Bảng 1. Kết quả về số đặc tính được sử dụng để xây dựng mô hình

Loại mô hình	$\lambda = 0$	$\lambda = 1,0$	$\lambda = 5,0$	$\lambda = 10$	$\lambda = 100$
Tuyến tính	14	13	10	8	3
Bậc hai	14	14	14	13	10

Bảng 2. Kết quả về sai số trên bộ số liệu đối với hai mô hình

Loại mô hình	$\lambda = 0$	$\lambda = 1,0$	$\lambda = 5,0$	$\lambda = 10$	$\lambda = 100$
Tuyến tính	17,60	17,97	24,64	29,79	41,86
Bậc hai	17,78	17,75	17,55	19,06	19,96

Bảng 3. Kết quả về sai số trên tập dữ liệu kiểm tra với mô hình tuyến tính

Số dữ liệu huấn luyện	$\lambda = 0$	$\lambda = 1,0$	$\lambda = 5,0$	$\lambda = 10$	$\lambda = 100$
70%	28,23	26,21	33,44	45,53	48,26
80%	26,67	25,09	31,49	41,50	49,61
90%	25,73	23,31	29,62	38,47	45,49

Bảng 4 . Kết quả về sai số trên tập dữ liệu kiểm tra với mô hình bậc hai

Số dữ liệu huấn luyện	$\lambda = 0$	$\lambda = 1,0$	$\lambda = 5,0$	$\lambda = 10$	$\lambda = 100$
70%	32,36	32,05	33,35	31,40	30,02
80%	29,50	28,89	29,37	29,70	29,19
90%	26,39	28,62	28,18	28,67	29,18

Từ các bảng trên ta có thể nhận xét về kết quả với bộ số liệu hiện có như sau:

- Trên toàn bộ số liệu, mô hình bậc hai cho sai số nhỏ hơn mô hình tuyến tính nhưng khi chia dữ liệu thành hai tập là tập dữ liệu huấn luyện và dữ liệu test thì sai số trên tập dữ liệu test của mô hình tuyến tính tốt hơn.

- Với mô hình tuyến tính, sai số ứng với trường hợp hệ số phạt $\lambda = 1,0$ là tốt nhất và số đặc tính trong trường hợp này là 13, tức là ta không cần dùng hết các đặc tính hiện có nhưng sai số tốt hơn dùng hết toàn bộ số liệu.

- Với mô hình bậc 2, sai số trong trường hợp $\lambda = 10$ là tốt nhất và ta cũng chỉ cần dùng 12 đặc tính (bớt được 2 đặc tính so với bộ số liệu ban đầu).

- Vì sai số trên tập test ứng với trường hợp tập dữ liệu huấn luyện là 90% số liệu, 80% số liệu, 70% số liệu (giảm dần) nên ta có thể nhận xét khi số lượng dữ liệu trong tập huấn luyện càng lớn thì kết quả của mô hình càng chính xác.

4. KẾT LUẬN

Chúng tôi đã nghiên cứu phương pháp ước lượng giá đất dựa trên mô hình hedonic. Hai dạng mô hình toán học được đề xuất là mô hình tuyến tính và mô hình bậc hai. Các kết quả nghiên cứu trên bộ số liệu thu thập ở địa bàn quận Long Biên được trình bày. Mặc dù các sai số khi tính toán trên tập dữ liệu test còn tương đối lớn nhưng với việc xây dựng thành công mô hình và có thể ước lượng với sai số trung bình 23,31% là có thể chấp nhận được. Ngoài ra kết quả của nghiên cứu cũng cho thấy việc sử dụng toàn bộ số liệu cho ra kết quả không tốt bằng việc chỉ sử dụng một số đặc tính của miếng đất. Ưu điểm của phương pháp này còn nằm ở chỗ việc lựa chọn đặc tính ảnh hưởng là hoàn toàn

tự động dựa trên mô hình và sai số tính toán chứ không chỉ phụ thuộc cảm tính của con người. Một chương trình máy tính để phục vụ cho nghiên cứu và phát triển các nghiên cứu về sau cũng đã được chúng tôi phát triển. Chương trình hiện tại có các chức năng đọc dữ liệu, tính toán tham số tối ưu, lựa chọn đặc trưng, đánh giá sai số của mô hình. Trong thời gian tới chúng tôi mong muốn phát triển các mô hình toán học phức tạp hơn với độ chính xác cao hơn. Các chức năng mới của chương trình máy tính như chức năng ước lượng giá cũng sẽ được bổ sung, các nghiên cứu trên bộ số liệu lớn hơn cũng sẽ được chúng tôi thực hiện.

TÀI LIỆU THAM KHẢO

- Alonso W. (1964). Location and Land Use. Harvard University Press, Cambridge, Massachusetts.
- Adair, A.S., J. N. Berry, and W. S. McGreal (1996). Hedonic modeling, housing submarkets and residential valuation, *Journal of Property Research*, 13: 67 - 83
- Bailey, M.J., R.F. Muth, and H.O. Nourse (1963). A regression method for real estate price index construction. *Journal of the American Statistical Association*, 58: 933 - 942.
- Bartik, T. J. (1987). The estimation of demand parameters in hedonic price models, *Journal of Political Economy*, 95(11): 81 - 88.
- Bloomquist, G., and L. Worley (1981). Hedonic prices, demands for urban housing attributes and benefit estimates, *Journal of Urban Economics*, 9: 212 - 221.
- Butler, R. V. (1982). The specification of hedonic indexes for urban housing, *Land Economics*, 58: 94 - 108.
- Cassel, E., and R. Mendelsohn (1985). The choice of functional forms for hedonic price equations: Comment, *Journal of Urban Economics*, 18(2): 135 - 142.
- Cebula R. J. (2009). The Hedonic Pricing Model Applied to the Housing Market of the City of

- Savannah and Its Savannah Historic Landmark District, *The Review of Regional Studies*, 39(1): 9 - 22.
- Englund P. (1998). Improved Price Indexes for Real Estate: Measuring the Course of Swedish Housing Prices, *Journal of urban economics*, 44(171): 96.
- Gouriéroux C., and A. Laferrère (2009). Managing hedonic housing price indexes: The French experience, *Journal of Housing Economics*, 18: 206 - 213
- Ishijima H., and A. Maeda (2013). Real Estate Pricing Models: Theory, Evidence, and Implementation, *Asia - Pacific Finan Markets*, DOI 10.1007/s10690-013-9170-7.
- Lancaster, K. J. (1966). A new approach to consumer theory, *Journal of Political Economy*, 74: 132 - 157.
- Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition, *Journal of Political Economy*, 82(1): 35 - 55.
- Stephen Boyd, [http: //cvxr.com/cvx/](http://cvxr.com/cvx/).
- Zhou W- X., and D. Sornette (2008). Analysis of the real estate market in Las Vegas: Bubble, seasonal patterns, and prediction of the CSW indices, *Physica A*, 387: 243 - 260.