



## Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting

Jae Joon Ahn<sup>a</sup>, Hyun Woo Byun<sup>a</sup>, Kyong Joo Oh<sup>a,\*</sup>, Tae Yoon Kim<sup>b,1</sup>

<sup>a</sup> Department of Information and Industrial Engineering, Yonsei University, 134, Shinchon-Dong, Seodaemun-Gu, Seoul 120-749, South Korea

<sup>b</sup> Department of Statistics, Keimyung University, Daegu 704-701, South Korea

### ARTICLE INFO

#### Keywords:

Ridge regression  
Genetic algorithm  
Real estate market

### ABSTRACT

This study considers real estate appraisal forecasting problem. While there is a great deal of literature about use of artificial intelligence and multiple linear regression for the problem, there has been always controversy about which one performs better. Noting that this controversy is due to difficulty finding proper predictor variables in real estate appraisal, we propose a modified version of ridge regression, i.e., ridge regression coupled with genetic algorithm (GA-Ridge). In order to examine the performance of the proposed method, experimental study is done for Korean real estate market, which verifies that GA-Ridge is effective in forecasting real estate appraisal. This study addresses two critical issues regarding the use of ridge regression, i.e., when to use it and how to improve it.

© 2012 Elsevier Ltd. All rights reserved.

### 1. Introduction

In recent years, interest in performance of real estate markets and real estate investment trusts (REITs) has grown up so fast and tremendously as they are usually required for asset valuation, property tax, insurance estimations, sales transactions, and estate planning. Conventionally, sales comparison approach has been widely accepted to forecast residential real estate. The sales comparison grid method, however, is often questioned for relying too much on subjective judgments for obtaining reliable and verifiable data (Wiltshaw, 1995). As a consequence, multiple linear regression (MLR) based on related predictors has been considered as a rigorous alternative enhancing predictability of real estate and property value, which immediately faces criticism such as nonlinearity within the data, multicollinearity issues in the predictor variables and the inclusion of outlier in the sample. As is often the case with other financial forecasting problems, this criticism has prompted researchers to resort to artificial neural network (ANN) as another logical alternative (Ahn, Lee, Oh, & Kim, 2009; Chen & Du, 2009; Dong & Zhou, 2008; Lee, Booth, & Alam, 2005; Lu, 2010; Oh & Han, 2000; Versace, Bhatt, Hinds, & Shiffer, 2004). The follow-up studies observe, however, that either ANN or MLR fails to report a dominating performance than the other, i.e., ANN excels MLR in some cases while MLR excels ANN in other cases (Dehghan, Sattari, Chehreh, & Aliabadi, 2010; Hua, 1996; Nguyen

& Cripps, 2001; Worzala, Lenk, & Silva, 1995). In this study, it will be shown that this confusing episode appears due to difficulty finding proper predictor variables and could be resolved quite successfully by a modified version of ridge regression, i.e., ridge regression coupled with genetic algorithm (GA-Ridge).

Theoretically as well as practically, there has been widespread strong objection to arbitrary use of ridge regression. The main criticisms are twofold. Firstly, though it is well known that ridge regression is effective for the case where the unknown parameters (or the linear coefficients) are known a priori to have small modulus values, it is hard to obtain or implement such prior information. Secondly, blind use of ridge regression is likely to change any non-significant predictor variable into significant one easily. Our study addresses these two critical issues and proposes GA-Ridge as a measure that takes care of them nicely.

The rest of the study is divided as follows. Section 2 discusses background of this article involving difficulty finding proper predictor variables in real estate forecasting. Section 3 is devoted to detailed description of the proposed GA-Ridge and discusses its effectiveness for handling the two critical issues of ridge regression. In Section 4, GA-Ridge is experimented in the Korean real estate market to demonstrate its effectiveness. Lastly, the concluding remarks are given in Section 5.

### 2. Background

#### 2.1. Predictor variable for real estate forecasting

Forecasting of asset pricing is a major issue in real estate practice (Bourassa, Cantoni, & Hoesli, 2010; Chica-Olmo, 2007;

\* Corresponding author. Tel.: +82 2 2123 5720; fax: +82 2 364 7807.

E-mail addresses: [redsamzang@yonsei.ac.kr](mailto:redsamzang@yonsei.ac.kr) (J.J. Ahn), [next1219@nate.com](mailto:next1219@nate.com) (H.W. Byun), [johanoh@yonsei.ac.kr](mailto:johanoh@yonsei.ac.kr) (K.J. Oh), [tykim@kmu.ac.kr](mailto:tykim@kmu.ac.kr) (T.Y. Kim).

<sup>1</sup> Tel.: +82 53 580 5533.

McCluskey & Anand, 1999; O’Roarty et al., 1997; Peterson & Flanagan, 2009; Tay & Ho, 1994; Wilson, Paris, Ware, & Jenkins, 2002). Property development relies on prediction of expected costs and returns (Allen, Madura, & Springer, 2000; Evans, James, & Collins, 1993; Juan, Shin, & Perng, 2006; McCluskey & Anand, 1999; Pivo & Fisher, 2010). Property and facilities managers need forecasts of supply and demand as well as of cost and return. Funds and investment managers rely on forecasts of the present and future values of real estate in terms of economic activities. In these real estate forecasting problems, there has been a hot controversy over superiority of ANN over MLR as proper tool since use of ANN for residential valuation was first suggested by Jensen (1990). Rossini (1997) seeks to assess the application of ANN and MLR to residential valuation and supports the use of MLR while Do and Grudnitski (1992) suggests ANN as a superior technique. Worzala et al. (1995) notices that while ANN slightly outperforms MLR in some cases, the difference between the two is insignificant. Hua (1996) and Brooks and Tsolacos (2003) support ANN over MLR with some cautionary note on predictor variables and McGreal, Berry, McParland, and Turner (2004) expresses skepticism about ANN. Noting that ANN is designed mainly for the purpose of modeling any functional relationship (or ANN is designed mainly to correct modeling bias or assumption), ANN is expected to excel MLR when there is significant modeling bias from linear model, while MLR is expected to excel ANN otherwise. Thus no clear-cut superiority between ANN and MLR implicitly suggests that other source of trouble might exist than incorrect modeling in real estate appraisal forecasting.

One possible source of trouble is difficulty finding significant and reliable predictors as discussed by several authors. Rossini (1997) noticed that quantitative predictor variables such as past sale price, land area, rooms and year of construction tend to suffer from lack of qualitative measures, while qualitative predictor variables such as building style and environments are frequently rather simplistic and fail to capture sufficient information. Similar observations are made by Brooks and Tsolacos (2003) and McGreal et al. (2004). In particular, Brooks and Tsolacos (2003) noticed that significant predictors depend on the used methodology. These discussions altogether suggest clearly that finding a proper set of predictor variables is hard in real estate appraisal and it would be highly desirable to take care of this predictor selection problem technically.

## 2.2. Ridge regression

Ridge regression is known as a very useful tool for alleviating multicollinearity problem (Walker & Birch, 1988). Its formal formulation is given as one of least squares subject to a specific type of restrictions on the parameters. The standard approach to solve an overdetermined system of linear equations:

$$\mathbf{Y} = \mathbf{X}\beta$$

is known as linear least squares and seeks to minimize the residual:

$$\|\mathbf{Y} - \mathbf{X}\beta\|^2,$$

where  $\mathbf{Y}$  is  $n \times 1$  vector,  $\mathbf{X}$  is  $n \times p$  matrix ( $n \geq p$ ),  $\beta$  is  $p \times 1$  vector and  $\|\cdot\|$  is Euclidean norm. However, the matrix  $\mathbf{X}$  may be ill conditioned or singular yielding a non-unique solution. In order to give preference to a particular solution with desirable properties, the regularization term is included in this minimization:

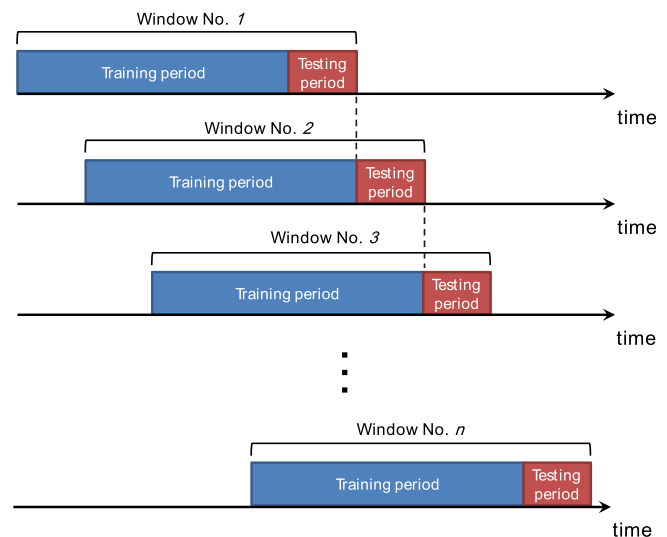
$$\|\mathbf{Y} - \mathbf{X}\beta\|^2 + k^2\|\beta\|^2.$$

This regularization improves the conditioning of the problem, thus enabling a numerical solution. An explicit solution, denoted by  $\hat{\beta}$ , is given by

$$\hat{\beta} = \hat{\beta}(k) = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}, \quad (1)$$

**Table 1**  
Training and testing period for moving window scheme.

Window number	Training period	Testing period
1	1996.07–2004.12	2005.01–2005.06
2	1997.01–2005.06	2005.07–2005.12
3	1997.07–2005.12	2006.01–2006.06
4	1998.01–2006.06	2006.07–2006.12
5	1998.07–2006.12	2007.01–2007.06
6	1999.01–2007.06	2007.07–2007.12
7	1999.07–2007.12	2008.01–2008.06
8	2000.01–2008.06	2008.07–2008.12
9	2000.07–2008.12	2009.01–2009.06
10	2000.01–2009.06	2009.07–2009.12



**Fig. 1.** Moving window scheme.

where  $k$  is a positive number. In applications, the interesting values of  $k$  usually lie in the range of  $(0, 1)$ . This procedure is called ridge regression.

It is well known that ridge regression can be regarded as an estimation of  $\beta$  from the data subject to prior knowledge that smaller values in modulus of the  $\beta$ s are more likely than larger values, and that larger and larger values of the  $\beta$ s are more and more unlikely. Thus ridge regression is quite useful when smaller values in modulus of the  $\beta$ s are expected more than larger values. In this context, one major drawback of ridge regression is the “unchecked arbitrariness” when it is implemented in practice. Indeed the characteristic effect of the ridge regression procedure is to change any non-significant estimated  $\beta$  to the significant estimated  $\beta$  and hence it is questionable that much real improvement can be really achieved by such a procedure. Refer to Draper and Smith (1981).

## 2.3. Genetic algorithm

GA is a stochastic search technique that can explore large and complicated spaces on the ideas from natural genetics and evolutionary principle (Goldberg, 1989; Holland, 1975; Oh, Kim, & Min, 2005). It has been demonstrated to be effective and robust in searching very large spaces in a wide range of applications (Koza, 1993). GA is particularly suitable for multi-parameter optimization problems with an objective function subject to numerous hard and soft constraints. GA performs the search process in four

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات