

# Fuzzy Ridge Regression with non symmetric membership functions and quadratic models

S. Donoso, N. Marín, and M.A. Vila

IDBIS Research Group - Dept. of Computer Science and A. I.,  
E.T.S.I.I. - University of Granada, 18071, Granada, Spain  
sdonosos@vtr.net, {nicm|vila}@decsai.ugr.es  
<http://frontdb.ugr.es>

**Abstract.** Fuzzy regression models has been traditionally considered as a problem of linear programming. The use of quadratic programming allows to overcome the limitations of linear programming as well as to obtain highly adaptable regression approaches. However, we verify the existence of multicollinearity in fuzzy regression and we propose a model based on Ridge regression in order to address this problem.

## 1 Introduction

Regression analysis tries to model the relationship among one dependent variable and one or more independent variables. During the regression analysis, an estimate is computed from the available data though, in general, it is very difficult to obtain an *exact* relation.

Probabilistic regression assumes the existence of a crisp aleatory term in order to compute the relation. In contrast, fuzzy regression (first proposed by Tanaka et al. [15]) considers the use of fuzzy numbers.

The use of fuzzy numbers improves the modeling of problems where the output variable (numerical and continuous) is affected by imprecision. Even in absence of imprecision, if the amount of available data is small, we have to be cautious in the use of probabilistic regression. Fuzzy regression is also a practical alternative if our problem does not fulfill the suppositions of probabilistic regression (as, for example, that the coefficient of the regression relation must be constant).

Fuzzy regression analysis (with crisp input variables and fuzzy output variable) can be categorized in two alternative groups:

- Proposals based on the use of *possibility* concepts [10–13, 16, 17].
- Proposals based on the minimization of central values, mainly through the use of the least squares method[7, 9].

Possibilistic regression is frequently carried out by means of the use of linear programming. Nevertheless, implemented in such a way, this method does not consider the optimization of the central tendency and usually derives a high number of crisp estimates.

In this work we introduce a proposal where both approaches of fuzzy regression analysis are integrated. We also show that the use of quadratic programming can improve the management of multicollinearity among input variables. To address this problem, we propose a new version of Fuzzy Ridge Regression.

The paper is organized as follows: next section presents new regression models based on the use of quadratic programming, section 3 describes a new version of Fuzzy Ridge Regression based on the methods of section 2, section 4 is devoted to presents and example, and, finally, section 5 concludes the paper.

## 2 Fuzzy Linear Regression

Let  $X$  be a data matrix of  $m$  variables  $X_1, \dots, X_m$ , with  $n$  observations each one (all of them real numbers), and  $Y_i$  ( $i = 1, \dots, n$ ) be a fuzzy set characterized by a LR membership function  $\mu_{Y_i}(x)$ , with center  $y_i$ , left spread  $p_i$ , and right spread  $q_i$  ( $Y_i = (y_i, p_i, q_i)$ ).

The problem of fuzzy regression is to find fuzzy coefficients  $A_j = (a_j, c_{Lj}, c_{Rj})$  such that the following model holds:

$$Y_i = \sum_{j=1}^m A_j X_{ij} \quad (1)$$

The model formulated by Tanaka et al. [15] considers that the (fuzzy) coefficients which have to be estimated are affected by imprecision. This model intends to minimize the imprecision by the following optimization criterion [14]:

$$\text{Min} \sum_{i=1}^n \sum_{j=1}^m (c_{Lj} + c_{Rj}) |X_{ij}| \quad (2)$$

subject to usual condition that, at a given level of possibility ( $h$ ), the  $h$ -cut of the estimated value  $\tilde{Y}_i$  contains the  $h$ -cut of the empiric value  $Y_i$ . This restriction can be expressed by means of the following formulation[1]:

$$\sum_j^m a_j X_{ij} + (1-h) \sum_j^m c_{Rj} |X_{ij}| \geq y_i + (1-h)q_i \quad \text{for } i = 1, \dots, n \quad (3)$$

$$\sum_j^m a_j X_{ij} - (1-h) \sum_j^m c_{Lj} |X_{ij}| \leq y_i - (1-h)p_i \quad \text{for } i = 1, \dots, n \quad (4)$$

$$c_{Rj}, c_{Lj} \geq 0 \quad \text{for } j = 1, \dots, m \quad (5)$$

where  $h$  is a degree of possibility for the estimate, such that

$$\mu(Y_i) \geq h \quad \text{for } i = 1, \dots, n \quad (6)$$

The aforementioned formulation arises from the application of Zadeh's Extension Principle[18] and has been proved by Tanaka[15].

### 2.1 Use of Quadratic Programming

Our first approximation to the use of quadratic programming in fuzzy regression analysis is based on the interval model proposed by Tanaka and Lee[14].

If we want to minimize the extensions, taking into account that we use non symmetrical triangular membership functions, and we want to consider the minimization of the deviation with respect to the central tendency, we have the objective function

$$J = k_1 \sum_{i=1}^n (y_i - a' X_i)^2 + k_2 (c_L X' X c'_L + c_R X' X c'_R) \quad (7)$$

where  $k_1$  and  $k_2$  are weights  $\in [0, 1]$  that perform a very important role: they allow to give more importance to the central tendency ( $k_1 > k_2$ ) or to the reduction of the estimate's uncertainty ( $k_1 < k_2$ ) in the process.

The model with this objective function (7) and restrictions (3)-(5) will be called Extended Tanaka Model (ETM) in this paper and, with the parameters,  $ETM(k_1, k_2)$ .

Let us now focus not in the minimization of the uncertainty of the estimated results but on the quadratic deviation with respect to the empiric data. That is, we will contrast the estimated spreads with respect to the spreads of the output data ( $p_i$  and  $q_i$ ).

According to this new criterion, the objective function represents the quadratic error for both the central tendency and each one of the spreads:

$$J = k_1 \sum_{i=1}^n (y_i - a' X_i)^2 + k_2 \left( \sum_{i=1}^n (y_i - p_i - (a' - c'_L) X_i)^2 + \sum_{i=1}^n (y_i + q_i - (a' + c'_R) X_i)^2 \right) \quad (8)$$

The model with objective function (8) and restrictions (3)-(5) will be called Quadratic Possibilistic Model (QPM) in this paper and, with the parameters,  $QPM(k_1, k_2)$ . It can be proven that this last model does not depend on the data unit.

One of the main criticisms to possibilistic regression analysis is that as the number of available data increases the length of estimated spreads also increases.

In this context, we propose a third new model, called Quadratic Non-Possibilistic (QNP), which considers the objective function (8) and which only incorporates the restriction (5).

**Example 1** *We experiment with data taken from Tanaka's paper[14], where X goes from 1 to 8.*

First, we have applied the model of Kim[8] and Chang[2] with X varying from 1 to 22. The results of this analysis are depicted in Fig. 1. As can be observed, when X=15, the three curves converge ( $a_i = a_i - c_{Li} = a_i + c_{Ri}$ ). With values higher than 15, the relationship among extreme points in the estimated

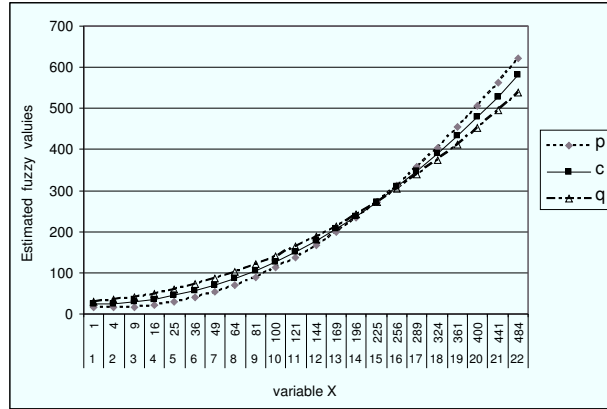


Fig. 1. Predictions with methods of Kim [8] and Chang [2]

membership functions reverses, so that the left extreme is higher than the right extreme (which has no sense).

The same experimentation with Model QNP is depicted in Figure 2. Model QNP forces the estimate's structure to be the same for both the central tendency and the fuzzy extremes. This fact, which can be seen as a restriction in the behavior of the spreads, guarantees that the inconsistencies of the previous example do not appear.

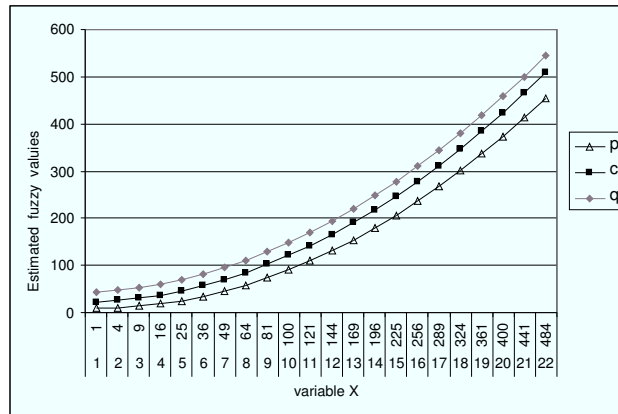


Fig. 2. Predictions with method QNP

This predictive capability of the proposed model overcomes the limitation analyzed by Kim et. al [8], where the capability of prediction is restricted only to probabilistic models.

Outliers pay a determining role in the estimation of the extensions in possibilistic regression. That is the reason why we propose the use of an alternative model, the aforementioned QNP, where restrictions 3-5 are reduced to only restriction 5. In this new model, the estimations of extensions represent the whole set of data extensions, and not only outliers, as in the possibilistic case.

### 3 Fuzzy Ridge Regression

The approach based on quadratic programming analyzed in previous sections has the additional advantage of allowing the management of multicollinearity. With this approach, we can set regression methods which deal with the problem of multicollinearity among input variables, as for example, fuzzy ridge regression.

In the seminal paper of fuzzy regression, Tanaka et al. [15] stated about their concrete example “the fact that  $A_4$  and  $A_5$  are negative depends on the strong correlations between variables  $X_4$  and  $X_5$ ”. Actually, the correlation between  $X_4$  and  $X_5$  is 0.95, much higher than any other value of correlation in  $Y$  and  $X_i$ , which indicates a very high multicollinearity. It can be assumed that the same distortion effect that affects probabilistic regression can be found in fuzzy regression.

The most popular probabilistic regression techniques that are usually used to deal with multicollinearity are *Principal Component Regression* and the *Ridge Regression*. Recently, papers about Fuzzy Ridge Regression has appeared in the literature which use an approach closely related to the support vector machine proposed by Vapnik[5, 6].

In the area of probabilistic regression, Ridge regression can be seen as a correction of the matrix  $X'X$ . This matrix, in presence of multicollinearity, has values close to zero. It can be proven that the expected value for estimations  $\tilde{a}'\tilde{a}$  is

$$E(\tilde{a}'\tilde{a}) = a'a + \sigma^2 \sum_i \left( \frac{ct}{\lambda_i} \right) \quad (9)$$

where  $\lambda_i$  are the eigenvalues values of  $X'X$  and  $ct$  is a constant. If these values are close to 0, the expected value for  $a'a$  increases a lot, producing coefficients with high absolute value and with the opposite sign, as the comment of Tanaka et al. suggests.

The introduction of a small positive value in the diagonal of  $X'X$  moves the least value of  $\lambda_i$  far from zero, and, thus, the expected value for  $\tilde{a}'\tilde{a}$  decreases.

The Ridge Regression can be seen as the addition of a new factor to the objective function. This factor depends on a parameter  $\lambda$ , called Ridge parameter. Ridge regression minimizes the conventional criterion of least squares in the following way [4]:

$$a^{ridge} = \min_a \left[ \sum_i (y_i - \sum_j X_{ij} a_j)^2 + \lambda \sum_j a_j^2 \right] \quad (10)$$

The Ridge solutions are not equivariant under changes in the scale of the inputs.

The model of Fuzzy Ridge Regression (FRR), introduced firstly in our work [3], is formalized with the following objective function:

$$\begin{aligned} & k_1 \sum_{i=1}^n (y_i - a' X_i)^2 + k_2 (\sum_{i=1}^n (y_i - p_i - (a' - c'_L) X_i)^2 + \\ & + \sum_{i=1}^n (y_i + q_i - (a' + c'_R) X_i)^2) + \lambda (a' a) \end{aligned} \quad (11)$$

where the *penalty* only acts on the vector of central values. This model will be called FRR (fuzzy ridge regression).

An extension of the model, including the extensions, is

$$\begin{aligned} & k_1 \sum_{i=1}^n (y_i - a' X_i)^2 + k_2 (\sum_{i=1}^n (y_i - p_i - (a' - c'_L) X_i)^2 + \\ & + \sum_{i=1}^n (y_i + q_i - (a' + c'_R) X_i)^2) + \lambda (\sum_{j=1}^m (k_3 a_j^2 + k_4 ((a_j - c_{Lj})^2 + (a_j + c_{Rj})^2))) \end{aligned} \quad (12)$$

where  $k_3 y k_4$  are constants to weight the terms, and the spreads are  $a_j - c_{Lj}$  and  $a_j + c_{Rj}$ . This model will called  $EFRR_\lambda(k_3, k_4)$  (extension of fuzzy ridge regression with parameters  $\lambda$ ,  $k_3$  y  $k_4$ ).

There exist many proposals to choose  $\lambda$ . Many of them suggest varying the parameter in a certain interval, checking the behavior of the coefficients, and choosing  $\lambda$  when the estimates remain stable.

A more general approach can be proposed, where the  $\lambda$  Ridge parameter depends on each variable ( $\lambda_j$  with  $j = 1, \dots, m$ ). In this case, the objective function is as follows

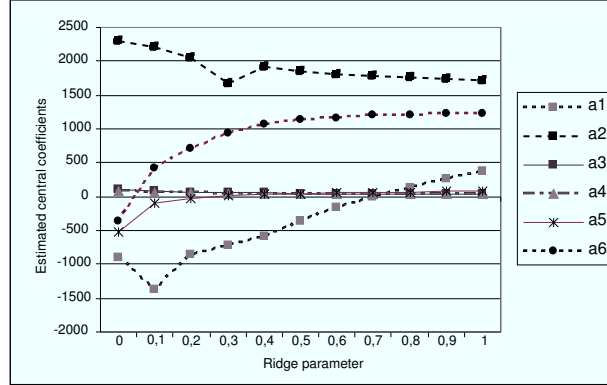
$$\begin{aligned} & k_1 \sum_{i=1}^n (y_i - a' X_i)^2 + k_2 (\sum_{i=1}^n (y_i - p_i - (a' - c'_L) X_i)^2 + \\ & + \sum_{i=1}^n (y_i + q_i - (a' + c'_R) X_i)^2) + (\sum_{j=1}^m \lambda_j (k_3 a_j^2 + k_4 ((a_j - c_{Lj})^2 + (a_j + c_{Rj})^2))) \end{aligned} \quad (13)$$

called GFRR (generalized fuzzy regression model) with the parameters  $\lambda_j$ ,  $k_3$  and  $k_4$ .

### 3.1 Examples

Let use introduce an example of use of the previously described methods.

**Example 2** *The example, similar to the one used for Tanaka [15] to illustrate the problem of multicollinearity, will be used here to experiment with the previously defined Fuzzy Ridge Regression model. We will use the method QPM, with  $k_1 = 1$  and  $k_2 = 1$  for our calculus.*



**Fig. 3.** Central coefficients,  $a_i$ , as  $\lambda_i$  increases

Figure 3 depicts the trajectory of the coefficients' centers, when the  $\lambda_i$  parameters are function of the diagonal of the matrix  $X'X$  (from 0 to 1 with increments of 0.1).

According to the example of Tanaka et al. ( $Y$  is the price of a house), all the coefficients must be positive (maybe with the exception of the number of Japanese rooms) because as the value of the variable increases the value of the house must also increase. The regression analysis, either of least squares or our fuzzy regression, initially produces some negative coefficients. However, three coefficients, which initially have negative values, reach positive values.

If we suppose that the  $\lambda$  coefficients are constant, varying from 0 to 55, we have the trajectory for the coefficients' centers depicted in figure 4. As can be observed, one of the coefficient remains negative while the other two become positive.

In any case, the availability of more reliable coefficients permits a better knowledge of the function we are looking for, and, consequently, better conditions for the use with predictive aims.

In order to end this section, let us compare our model with the model of Hong and Hwang[5,6]. These authors do a *dual estimation* of the coefficients, with the relation:

$$\beta_{dual} = Y'(XX' + I\lambda)^{-1}X \tag{14}$$

where  $I$  is the identity matrix of range  $n$  and  $\lambda$  is a constant, the Ridge coefficient, whose values increase in value from 0.

If we take the same data, and make  $\lambda$  increase from 0 to 1.5 (with increments of 0.1) we obtain the results depicted in figure 5.

These results must be contrasted with those of figure 4, where the ridge parameter is also constant. As can be observed, central coefficients have a similar behavior in both graphics: there is a positive coefficient which converges to (ap-

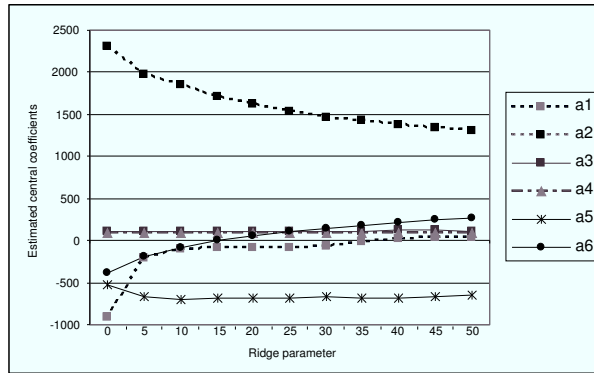


Fig. 4. Ridge central coefficients,  $a_i$ , as  $\lambda$  increases

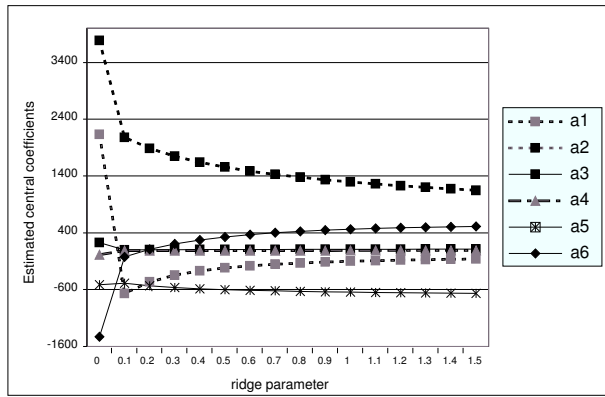


Fig. 5. Dual Ridge Central Coefficients of Hong and Hwang,  $Beta_{dual}$  as  $\lambda$  increases



proximately) 1300 and a negative coefficient which converges to (approximately) -600. The other coefficients are close to zero.

However, the main difference is in the central coefficient  $a_1$ . With the method of Hong and Hwang, this coefficient has a high value when  $\lambda = 0$  and is -600 when  $\lambda = 0.1$ . This fact does not occurs with our method.

Let us now present a second example with a higher amount of variables. We have ten demographic groups as input data and the output is the saving (positive or negative) of the whole population. For the sake of space we omit the table with numerical data.

Results with model EFRR(1,1), normalizing data according to the maximum value, have been computed with lambda varying from 0 to 1 and are shown in figure 6.

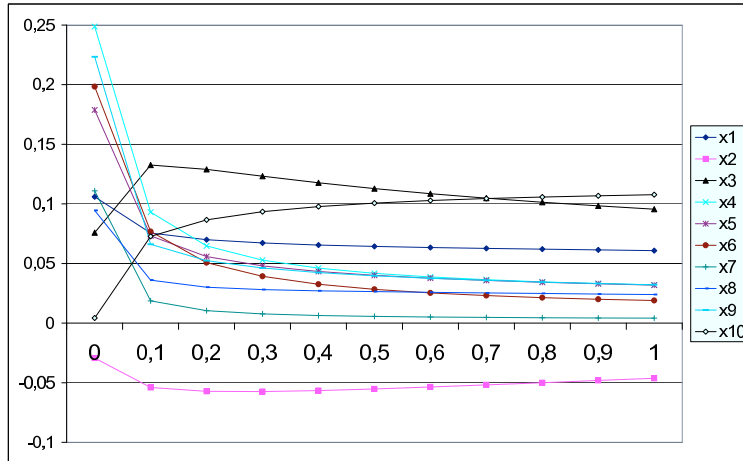


Fig. 6. Ridge tracing, example 2

As we can see, the step from 0 to 0.1 produces the best coefficients adjustment, which, on the other hand, have a quite stable behavior. Notice that coefficients of variables  $x_3$  and  $x_{10}$  increase their value from aprox. 0 and 0.1 to the highest values among the coefficients.

## 4 Conclusions

In this paper we have tried to validate the use of quadratic programming in order to obtain a good fitness in fuzzy linear regression.

To accomplish this task, we have adapted one existing model (ETM) and we have proposed two new models (QPM and QNP). Method QPM is a good choice when possibilistic restrictions are important in the problem. If we do not want to pay special attention to the possibilistic restrictions, QNP is an appropriate alternative.

We have proposed a special version of Fuzzy Ridge Regression based on our previous study on quadratic methods in order to cope with the multicollinearity problem.

## References

1. Andras Bardossy. Note on fuzzy regression. *Fuzzy Sets and Systems*, 37:65–75, 1990.
2. Yun-Hsi O. Chang. Hybrid regression analysis with reliability and uncertainty measures. *Ph.D. Dissertation, University of Maryland*, 2001.
3. Sergio Donoso. *Anlisis de regresin difusa: nuevos enfoques y aplicaciones*. Tesis doctoral, Universidad de Granada, 2006.
4. T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: data mining, inference, and prediction*. Springer, 2001.
5. D. H. Hong and C. Hwang. Ridge regression procedure for fuzzy models using triangular fuzzy numbers. *Fuzziness and Knowledge-Based Systems*, 12:2:145–159, 2004.
6. D. H. Hong, C. Hwang, and C. Ahn. Ridge estimation for regression models with crisp input and gaussian fuzzy output. *Fuzzy Sets and Systems*, 142:2:307–319, 2004.
7. C. Kao and Chyu C.L. Least squares estimates in fuzzy regression analysis. *European Journal of Operation Research*, pages 426–435, 2003.
8. B. Kim and R. R. Bishu. Evaluation of fuzzy linear regression models by comparison membership function. *Fuzzy Sets and Systems*, 100:343–352, 1998.
9. Kwang Jae Kim, Herbert Moskowitz, and Murat Koksalan. Fuzzy versus statistical lineal regression. *European Journal of Operational Research*, 92:417–434, 1996.
10. Ertunga C. Ozelkan and Lucien Duckstein. Multi-objctive fuzzy regression: a general framework. *Computers and Operations Research*, 27:635–652, 2000.
11. Georg Peters. Fuzzy linear regression with fuzzy intervals. *Fuzzy Sets and Systems*, 63:45–55, 1994.
12. David T. Redden and William H. Woodall. Further examination of fuzzy linear regression. *Fuzzy Sets and Systems*, 79:203–211, 1996.
13. Kazutomi Sugihara, Hiroaki Ishii, and Hideo Tanaka. Interval priorities in ahp by interval regression analysis. *Europeian Journal of Operatin Research*, 158:745–754, 2004.
14. Hideo Tanaka and Haekwan Lee. Interval regression analysis by quadratic programming approach. *IEEE Trans. on Fuzzy Systems*, 6(4), 1998.
15. Hideo Tanaka, S. Uejima, and K. Asai. Linear regression analysis with fuzzy model. *IEEE Trans. on Systems, Man, and Cybernetics*, 12(6):903–907, 1982.
16. Hideo Tanaka and J. Watada. Possibilistic linear systems and their application to the linear regerssion model. *Fuzzy Sets and Systems*, 27(3):275–289, 1998.
17. F-M. Tseng and L. Lin. A quadratic interval logit model for forecasting bankruptcy. *The International Journal of Management Science*, In press.
18. L. A. Zadeh. The concept of a linguistic variable and its application to aproxmate reasoning i, ii, iii. *Information Sciences*, 8-9:199–251, 301–357, 43–80, 1975.

