Econometrics Working Paper EWP0710
ISSN 1485-6441

University of Victoria
Department of Economics

Bayesian Fuzzy Regression Analysis and Model Selection: Theory and Evidence

Hui Feng
Department of Economics, Business & Mathematics
King's University College
University of Western Ontario
&
David E. Giles
Department of Economics
University of Victoria

December 2007

Abstract

In this study we suggest a Bayesian approach to fuzzy clustering analysis – the Bayesian fuzzy regression. Bayesian Posterior Odds analysis is employed to select the correct number of clusters for the fuzzy regression analysis. In this study, we use a natural conjugate prior for the parameters, and we find that the Bayesian Posterior Odds provide a very powerful tool for choosing the number of clusters. The results from a Monte Carlo experiment and two real data applications of Bayesian fuzzy regression are very encouraging.

Keywords: Bayesian posterior odds, model selection, fuzzy regression, fuzzy clustering

JEL Classifications: C1, C6, C8, C90

Contact Author: Hui Feng; Department of Economics, Business and Mathematics, King's University College at University of Western Ontario, London, Ontario, Canada N5A 2M3
E-mail: hfeng8@uwo.ca; FAX: (519) 433 0353; Phone: (519) 433 3491 *ext*. 4435

1. Introduction

Recent developments in econometric modelling have emphasized a variety of non-linear specifications. Parametric examples of these include various regime-switching models (the threshold and Markov switching autoregressive models): threshold autoregressive (TAR) models, self-exciting threshold autoregressive (SETAR) models, smoothing threshold autoregressive (STAR) models, and various others. In addition, non-parametric and semi-parametric models are widely used, though the well-known "curse of dimensionality" can place some limitations on their use with multivariate data. The use of fuzzy clustering analysis in the context of econometric modelling is a rather new approach within the class of nonlinear econometric models. Fuzzy clustering analysis is a very flexible and powerful technique that is used widely in the pattern recognition literature, for example, and has recently been applied in the area of modelling and forecasting economic variables (see Giles and Draeseke, 2003; Giles and Mosk, 2003; Feng, 2007; Giles and Stroomer, 2006; and Feng, 2007).

In fuzzy regression modelling, we use the explanatory variables to partition the sample into a pre-assigned number of clusters. The membership functions are calculated using some algorithm for each of the observations, and these provide weights between zero and one which indicate the degree to which each observation belongs to each cluster. The fuzzy clustering regression is then obtained by taking the weighted average of the regression results from each of the clusters, using the membership functions as weights. To date, researchers have treated the number of fuzzy clusters as being pre-determined, and no formal procedures have been used to determine the "optimal" number of clusters. In practice, the number of clusters is set to be between one and four, with one simply being the usual case where the full sample is used (standard regression).

We observe that the choice of a particular value for the number of clusters implies the choice of a particular model specification. Altering the number of clusters changes the number of "sub-models" that are fitted to the data and subsequently combined into a final result. We approach this model-selection problem from a Bayesian perspective, and propose the use of the Bayesian Posterior Odds to determine the number of fuzzy clusters used in the fuzzy regression analysis. In fuzzy regression analysis, the models fitted to each cluster can be estimated by any appropriate technique. Ordinary Least Squares is a typical choice. Nonlinearity is modeled successfully because the fuzzy combination of the regression results from each cluster involves weights that vary at each data-point. Here, however, in order to be consistent in our overall approach we use Bayesian estimation for each cluster's sub-model. More specifically, we adapt the standard Bayesian estimator based on the natural-conjugate prior p.d.f. to our clustering context. As a result, we refer to this overall modelling methodology as Bayesian Fuzzy Regression analysis.

This paper is organized as follows. The next section introduces fuzzy clustering analysis. Section 3 discusses some basic concepts associated with Bayesian inference, and sets up the Bayesian Posterior Odds analysis in a general model selection framework. Section 4 derives the Bayesian Posterior Odds under the natural conjugate prior for choosing the number of fuzzy clusters. The design and results of an extensive Monte Carlo experiment are presented in section 5, and two simple applications using real data are discussed in section 6. The last section offers our conclusions and some further research suggestions.

1. Fuzzy Clustering Analysis

A classical set can be viewed as a "crisp" set, in the sense that it has a clearly defined boundary. For example, the set *E* could be defined as any integer that is greater than 10. The membership of a "crisp" set requires the individual element either be a member or not - the "degree of membership" is either unity or zero.:

$$Mu(x) = \{\begin{matrix} 1 \\ 0 \end{matrix} \quad \text{if} \quad \begin{matrix} x \in E \\ x \notin E \end{matrix} \quad . \tag{1}$$

On the other hand, a fuzzy set is just as the name implies: "without a crisp boundary". The difference between a fuzzy set and a classical set lies in the nature of the membership for each element. Zadeh (1965) defined the meaning of the membership for fuzzy sets to be a continuous number between zero and one. This means that any element can be associated with one or more clusters. Further, all of these membership values added together should equal unity. Generally, this association involves different degrees of membership with each of the fuzzy sets. Just as this makes the boundaries of the sets fuzzy, it makes the location of the centroid of the set fuzzy as well.

The "Fuzzy c-Means" (FCM) algorithm, which was developed and improved by Dunn (1974, 1977) and Bezdek (1973, 1981), is frequently used in pattern recognition. The objective is to partition the data into fuzzy sets or clusters, to locate these clusters, and to quantify the degree of membership of every data-point with every cluster. It is based on minimization of the following functional:

$$J(U,v)=\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^{m}(d_{ik})^{2} \quad , \tag{2}$$

where u_{ik} is the "degree of membership" of data-point "k" in cluster "i", and d_{ik} is the distance between data x_k and the *i-th* cluster center v_i. "c" is the number of clusters presumed, and m is any real number greater than 1, which measures the degree of the fuzziness. In the limit, as m approaches unity, the membership becomes crisp. A common choice is $m = 2$. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ik} and the cluster centers v_i by:

$$u_{ik} = 1/\{\sum_{j=1}^{n}[(d_{ik})^{2}/(d_{jk})^{2}]^{1/(m-1)}\} \tag{3}$$

and,

$$v_i = [\sum_{k=1}^{n}(u_{ik})^{m}x_k]/[\sum_{k=1}^{n}(u_{ik})^{m}] \;;\; i = 1, 2, \ldots, \text{c.} \tag{4}$$

Over the years, there have been numerous developments in fuzzy set studies following the research by Zadeh (1965). Many applications can be found in the areas of computer science, systems analysis, electrical and electronic engineering, pattern recognition and psychology. In recent years, various applications of fuzzy sets in the area of econometrics have been led by Giles and his co-workers. See Giles (2005), Draeseke and Giles (2002), Giles and Draeseke (2003), Giles and Mosk (2003), Chen and Giles (2004), Giles (2005), Giles and Feng (2005), and Giles and Stroomer (2004, 2006). In these applications, "fuzzy regression", analysis is introduced. After clustering the data on the basis of the explanatory variables, separate regressions are fitted over each of the clusters. The fuzzy regression is obtained by combining the results from each cluster using the membership functions as the weights. As these weights vary continuously over the sample, even if linear models are fitted to each cluster, the resulting model can capture complex non-linearities in the data. See Giles and Draeseke (2003) for full details. We have written code for commands in the SHAZAM (2004) econometric package for the analysis in this paper.

For the fuzzy clustering analysis, the number of cluster "*c*" and the fuzziness parameter "*m*" have to be set before the actual regression analysis. Usually, for researchers, the common choice of "*m*" is 2 and "*c*" is from 1 up to 4, and seldom does the value of "*c*" exceed 6. However, these choices are informal, and in particular, no formal procedures have been introduced for the choice of these parameters. Different values of "*c*" and "*m*" imply that different numbers of clusters and different degrees of fuzziness will be used for the regression analysis, which means that different fuzzy models are going to be employed. This suggests a model-selection problem. What is the best choice for "*c*" and "*m*" for a given sample of data when applying fuzzy regression analysis? In order to simplify the analysis, in this paper we fix the value of "*m*" to be 2, and view the selection of "*c*" as the model-selection problem. Our analysis could be extended to the case where both "*m*" and "*c*" are to be selected, though the fact that the former parameter can take a continuum of values would complicate matters. Here, we adopt a Bayesian model-selection approach, and use Bayesian Posterior Odds analysis to select the value of "*c*".

2. Bayesian Posterior Odds Analysis—The Choice of "*c*"

3.1 Bayesian Model Selection

In order to discuss model selection *via* Bayesian Posterior Odds analysis, we first need to introduce some basic notation and concepts. Let M denote the model space, and M_i denote the i^{th} candidate model. The number of models is m^*, a finite or countably infinite value. So, $M = \{M_i\}$ $i=1, 2, 3\ldots m^*$. There are two sets of prior information that we need to consider in our Bayesian analysis here: the prior information about the models and the prior information about the parameters of those models.

First, assigning a prior mass function over *M*, let $p(M_i)$ denote the prior probability that the *i-th* model is the 'true' model, where

$$0 \leq p(M_i) \leq 1$$

and

$$\sum_{i=1}^{m^*} p(M_i) = 1 \ .$$

Suppose that the i^{th} model has a vector of parameters, θ_i, belonging to a parameter space, Ω_i, i = 1, 2, …., m^*. Let $p(\theta_i \mid M_i)$ denote the prior p.d.f. for the parameters of the i^{th} model. For each model in *M* the joint data density is given by $p(y \mid \theta_i, M_i)$, and viewed as a function of the parameters this is the usual likelihood function, $l(\theta_i \mid y, M_i)$. The conditional (on the model) data density is obtained as

$$p(y \mid M_i) = \int_{\Omega_i} p(y \mid \theta_i, M_i) p(\theta_i \mid M_i) d\theta_i , \qquad (5)$$

and the marginal data density is

$$p(y) = \sum_{j=1}^{m^*} p(M_j) p(y \mid M_j) \ . \qquad (6)$$

Applying Bayes' Theorem, the posterior probability of model M_i is

$$p(M_i \mid y) = \frac{p(M_i) p(y \mid M_i)}{p(y)} \propto p(M_i) p(y \mid M_i) \qquad . \qquad (7)$$

Note that the posterior probability of M_i will be calculated incorrectly if the calculation of $p(y)$ is incorrect, and this will be the case in the (likely) event that *M* is not fully specified. So, we will use the Bayesian Posterior Odds (BPO) to select the best model among a handful of candidate models because, even if *M* is incompletely specified, the Bayesian Posterior Odds will always be correct, for the following reason.

Let $[p(M_i)/p(M_j)]$ be the prior odds between model *i* and *j*. Then

$$BPO_{ij} = \frac{p(M_i \mid y)}{p(M_j \mid y)} = \{\frac{p(M_i)p(y \mid M_i)/p(y)}{p(M_j)p(y \mid M_j)/p(y)}\} = [\frac{p(M_i)}{p(M_j)}][\frac{p(y \mid M_i)}{p(y \mid M_j)}]. \qquad (8)$$

That is:

Bayesian Posterior Odds = [Prior Odds] × ["Bayes Factor"].

These posterior odds are independent of *p*(*y*), and hence of the specification of *M*. Of course, if *M* is complete, we could use the BPO to calculate individual posterior probabilities which reflect our beliefs about each model being a true model, given the sample information. In the context of the model selection problem, after obtaining the Bayesian Posterior Odds, generally we will still want to come to a conclusion of "rejecting" or "accepting" one model compared with some other competing model. This is a *two-action* problem (Zellner, 1971, p.291).

In general terms, let us consider two competing hypotheses, H_0 and H_1. There are two possible states of the world – either H_0 is true or H_1 is true. Let $\hat{H}_i$ denote the action of choosing the i^{th} hypothesis (*i* = 1, 2). If we accept the true hypothesis or we reject the false hypothesis, we will incur zero loss. However, our loss would be the positive amount $L(H_1, \hat{H}_0)$ if we accept the false null hypothesis, and our loss would be $L(H_0, \hat{H}_1)$ if we reject the true hypothesis. Which model will be accepted depends on which model minimizes posterior expected loss:

if $E(L \mid \hat{H}_0) < E(L \mid \hat{H}_1)$, Accept H_0.

If $E(L \mid \hat{H}_1) < E(L \mid \hat{H}_0)$, Accept H_1.

It is well-known that under any symmetric loss function[1], H_0 will be accepted only if

$$p(H_0 \mid y) > p(H_1 \mid y), \text{ or } \frac{p(H_0 \mid y)}{p(H_1 \mid y)} > 1.$$

[1] We could assume that the loss is asymmetric. This would simply alter the "threshold" value for the BPO.

Again, it should be emphasized that this determination of the rankings of the models through the BPO's does not require the calculation of the individual posterior probabilities for the individual models, and so it is not affected if the model space is under-specified.

3.2 Bayesian Estimation

In the above discussion, each model was parameterized by a vector of parameters, about which prior information was needed in order to construct the BPO. It is natural, therefore, to approach the estimation of these parameters from a Bayesian perspective. Such estimation is necessary in order to complete our fuzzy regression analysis.

The Bayes and Minimum Expected Loss (MEL) estimators coincide if the posterior expected loss is finite, so the discussion here focuses on MEL estimation, strictly speaking. Let us suppose that model M_i has been selected. Applying Bayes' Theorem, we first obtain the posterior p.d.f. for θ_i as:

$$p(\theta_i \mid y, M_i) \propto p(\theta_i \mid M_i) p(y \mid \theta_i, M_i) . \tag{9}$$

Then the MEL estimator of θ_i is the $\hat{\theta}_i$ that minimizes the posterior expected loss,

$$\int_{\Omega_i} L(\theta_i, \hat{\theta}_i) p(\theta_i \mid y, M_i) . \tag{10}$$

It is well known that if the loss function is quadratic, then $\hat{\theta}_i$ is the mean of $p(\theta_i \mid y, M_i)$; for an absolute error loss this MEL estimator is the median of $p(\theta_i \mid y, M_i)$; and $\hat{\theta}_i$ is the mode of the posterior density in the case of a "zero-one" loss function (Zellner, 1971, 24-25). In what follows in the next section our choice of prior p.d.f. for the parameters results in a posterior density whose characteristics ensure that the same MEL estimator arises under each of these three particular loss functions.

3. Bayesian Fuzzy Analysis

As was indicated in section 1, different choices for the number of fuzzy clusters, “*c*”, generate different fuzzy regression models of the type. For example, letting the value of “*c*” run from 1 to 4 implies that there are four possible fuzzy models. The first model is the one-cluster fuzzy model (*i.e.*, all of the data are used, and we have a conventional regression situation); the second model is based on two fuzzy clusters over the sample; and the third and fourth models assume there are respectively three and four fuzzy clusters over the sample range. The regressions that we fit over each cluster can be of any kind, depending on the nature of the data. In our case they will be multiple linear regressions.

Let:

$$M_1: \quad y_{11} = X_{11}\beta_{11} + u_{11} \qquad ; \ u_{11} \sim \mathrm{N}(0, \ \sigma_1 \mathrm{I}_{n11})$$

$$M_2: \quad y_{21} = X_{21}\beta_{21} + u_{21} \qquad ; \ u_{21} \sim \mathrm{N}(0, \ \sigma_{21} \mathrm{I}_{n21})$$

$$y_{22} = X_{22}\beta_{22} + u_{22} \qquad ; \ u_{22} \sim \mathrm{N}(0, \ \sigma_{22} \mathrm{I}_{n22})$$

$$M_3: \quad y_{31} = X_{31}\beta_{31} + u_{31} \qquad ; \ u_{31} \sim \mathrm{N}(0, \ \sigma_{31} \mathrm{I}_{n31})$$

$$y_{32} = X_{32}\beta_{32} + u_{32} \qquad ; \ u_{32} \sim \mathrm{N}(0, \ \sigma_{32} \mathrm{I}_{n32})$$

$$y_{33} = X_{33}\beta_{33} + u_{33} \qquad ; \ u_{33} \sim \mathrm{N}(0, \ \sigma_{33} \mathrm{I}_{n33})$$

$$M_4: \quad y_{41} = X_{41}\beta_{41} + u_{41} \qquad ; \ u_{41} \sim \mathrm{N}(0, \ \sigma_{41} \mathrm{I}_{n41})$$

$$y_{42} = X_{42}\beta_{42} + u_{42} \qquad ; \ u_{42} \sim \mathrm{N}(0, \ \sigma_{42} \mathrm{I}_{n42})$$

$$y_{43} = X_{43}\beta_{43} + u_{43} \qquad ; \ u_{43} \sim \mathrm{N}(0, \ \sigma_{43} \mathrm{I}_{n43})$$

$$y_{44} = X_{44}\beta_{44} + u_{44} \qquad ; \ u_{44} \sim \mathrm{N}(0, \ \sigma_{44} \mathrm{I}_{n44})$$

X_{ij}’s are each ($n_{ij} \times k$) matrices, each with rank k. The β_{ij} are each ($k \times 1$) coefficient vectors, and the u_{ij} are ($n_{ij} \times 1$) vectors of random error terms. β_{ij} and σ_{ij} are the coefficient vector and error standard deviation for the j^{th} cluster of model i, for $i, j = 1, 2, 3, 4$.

The prior probabilities associated with each model are denoted $p(M_1)$, $p(M_2)$, $p(M_3)$ and $p(M_4)$. The prior p.d.f.’s for the parameters β_{ij} and σ_{ij} (i, j = 1, 2, 3, 4) are taken to be the natural conjugate prior density in this case. Obviously, other possibilities could be considered, but this choice will suffice to illustrate the methodology and it provides an interesting (and mathematically tractable) benchmark. The following methodology would be unaltered if alternative prior p.d.f.'s were used, though the specific results would change, of course. In our case, we have:

$$p(\beta_{ij},\sigma_{ij}) = p(\beta_{ij} \mid \sigma_{ij})\, p(\sigma_{ij})$$

$$p(\beta_{ij} \mid \sigma_{ij}) = \frac{\left|C_{ij}\right|^{1/2}}{(2\pi)^{k_{ij}/2}\,\sigma_i^{\,k_{ij}}} \exp[-\frac{1}{2\sigma_{ij}{}^{2}}(\beta_{ij} - \bar{\beta}_{ij})' C_{ij} (\beta_{ij} - \bar{\beta}_{ij})] \qquad (11)$$

$$p(\sigma_{ij}) = \frac{K_{ij}}{\sigma_{ij}^{q_{ij}+1}} \exp(-\frac{q_{ij}\bar{s}_{ij}^{2}}{2\sigma_{ij}^{2}}) \ , \quad i, j = 1, 2, 3, 4$$

where the normalizing constant is $K_{ij} = 2(q_{ij}\bar{s}_{ij}^{2}/2)^{q_{ij}/2}/\Gamma(q_{ij}/2)$. That is, the conditional prior p.d.f. for β_{ij} given σ_{ij} is multivariate normal with prior mean vector $\bar{\beta}_{ij}$ and covariance matrix $\sigma_{ij}^{2}C_{ij}^{-1}$. The marginal prior information for σ_{ij} is represented by an inverted gamma density, with parameters q_{ij} and $\bar{s}_{ij}^{2}$ to be assigned values by the investigator. For this marginal prior to be proper, we need $0 < q_{ij}$, $\bar{s}_{ij}^{2} < \infty$; $i, j = 1, 2, 3, 4$.

We can now proceed with the Bayesian Posterior Odds calculations. Equation (8) gives us the formula for the BPO. However, in order to get the BPO we need to derive the conditional data densities for each of the models. Given the assumption of normal errors in all of the sub-models, the likelihood function for Model 1, for example, is:

$$p(y_{11} \mid \beta_{11}, \sigma_{11}, M_1) = \frac{1}{(2\pi)^{n11/2}} \frac{1}{\sigma_{11}{}^{n11}} \exp\{\frac{1}{2\sigma_{11}^{2}}[v_{11}s_{11}^{2} + (\beta_{11} - \hat{\beta}_{11})' X_{11}{}' X_{11}(\beta_{11} - \hat{\beta}_{11})]\} \qquad (12)$$

where $\hat{\beta}_{11} = (X_{11}{}' X_{11})^{-1} X_{11}{}' y_{11}$, $v_{11} = n_{11} - k$ and $v_{11}s_{11}^{2} = (y_{11} - X_{11}\hat{\beta}_{11})'(y_{11} - X_{11}\hat{\beta}_{11})$.

The likelihoods associated with the various clusters for the multi-cluster models follow in an obvious manner.

From (5), for the first model, which has one cluster, the conditional data density is:

$$p(y \mid M_1) = \iint p(y_{11} \mid \beta_{11}, \sigma_{11}, M_1)\, p(\beta_{11} \mid \sigma_{11})\, p(\sigma_{11})\, d\beta_{11} d\sigma_{11} \qquad .$$

Under the (reasonable) assumption that clusters are independent, the conditional data density for Model 2, with two clusters, is:

$$p(y|M_2)=\iint p(y_{21}|\beta_{21},\sigma_{21},M_2)p(\beta_{21}|\sigma_{21})p(\sigma_{21})\,d\beta_{21}d\sigma_{21}\iint p(y_{22}|\beta_{22},\sigma_{22},M_2)p(\beta_{22}|\sigma_{22})p(\sigma_{22})\,d\beta_{22}d\sigma_{22} \tag{13}$$

For Model 3 with three clusters this density is:

$$\begin{aligned}p(y|M_3)=&\iint p(y_{31}|\beta_{31},\sigma_{31},M_3)p(\beta_{31}|\sigma_{31})p(\sigma_{31})\,d\beta_{31}d\sigma_{31}\iint p(y_{32}|\beta_{32},\sigma_{32},M_3)p(\beta_{32}|\sigma_{32})p(\sigma_{32})\,d\beta_{32}d\sigma_{32}\\&\times\iint p(y_{33}|\beta_{33},\sigma_{33},M_3)p(\beta_{33}|\sigma_{33})p(\sigma_{33})\,d\beta_{33}d\sigma_{33}\end{aligned}$$

For Model 4 with four clusters it is:

$$\begin{aligned}p(y|M_4)=&\iint p_{41}(y|\beta_{41},\sigma_{41},M_4)p(\beta_{41}|\sigma_{41})p(\sigma_{41})\,d\beta_{41}d\sigma_{41}\iint p(y_{42}|\beta_{42},\sigma_{42},M_4)p(\beta_{42}|\sigma_{42})p(\sigma_{42})\,d\beta_{42}d\sigma_{42}\\&\times\iint p(y_{43}|\beta_{43},\sigma_{43},M_4)p(\beta_{43}|\sigma_{43})p(\sigma_{43})\,d\beta_{43}d\sigma_{43}\iint p(y_{44}|\beta_{44},\sigma_{44},M_4)p(\beta_{44}|\sigma_{44})p(\sigma_{44})\,d\beta_{44}d\sigma_{44}\end{aligned} \tag{14}$$

where, as before, the β_{ij} and σ_{ij} are the parameters for the j^{th} cluster of model *i*, for *i*, *j* = 1, 2, 3, 4. Using the results of Zellner (1971, p.308):

$$\begin{aligned}p(y|M_1)&=\iint p(y_{11}|\beta_{11},\sigma_{11},M_1)\,p(\beta_{11}|\sigma_{11})\,p(\sigma_{11})\,d\beta_{11}d\sigma_{11}\\&=\iint\frac{1}{(2\pi)^{n/2}}\frac{1}{{\sigma_{11}}^{n}}\exp\{\frac{1}{2\sigma_{11}^2}[v_{11}s_{11}^2+(\beta_{11}-\hat{\beta}_{11})'X_{11}{}'X_{11}(\beta_{11}-\hat{\beta}_{11})]\}\\&\quad\times\frac{|C_{11}|^{1/2}}{(2\pi)^{k_{11}/2}{\sigma_{11}}^{k_{11}}}\exp[-\frac{1}{2{\sigma_{11}}^2}(\beta_{11}-\bar{\beta}_{11})'C_{11}(\beta_{11}-\bar{\beta}_{11})]\\&\quad\times\frac{K_{11}}{\sigma_{11}^{q_{11}+1}}\exp(-\frac{q_{11}\bar{s}_{11}^2}{2\sigma_{11}^2})\,d\beta_{11}d\sigma_{11}\\&=\frac{1}{2}(2\pi)^{-n/2}K_{11}\left\|\frac{|C_{11}|}{|A_{11}|}\right\|^{1/2}2^{\frac{n_{11}+q_{11}}{2}}\Gamma(\frac{n_{11}+q_{11}}{2})(q_{11}\bar{s}_{11}^2+v_{11}s_{11}^2+Q_{11a}+Q_{11b})^{-\frac{n_{11}+q_{11}}{2}}\end{aligned} \tag{15}$$

where $A_{11}=C_{11}+X_{11}{}'X_{11}$

$$\begin{aligned}\tilde{\beta}_{11}&=A_{11}^{-1}(C_{11}\bar{\beta}_{11}+X_{11}{}'X_{11}\hat{\beta}_{11})\\Q_{11a}&=(\bar{\beta}_{11}-\tilde{\beta}_{11})'C_{11}(\bar{\beta}_{11}-\tilde{\beta}_{11})\\Q_{11b}&=(\hat{\beta}_{11}-\tilde{\beta}_{11})X_{11}{}'X_{11}(\hat{\beta}_{11}-\tilde{\beta}_{11})\end{aligned} \tag{16}$$

$$K_{11} = \frac{2(q_{11}\bar{s}_{11}^{2} / 2)^{q_{11}/2}}{\Gamma(q_{11}/2)}$$

$$v_{11} = n_{11} - k \ .$$

Similar operations are used to determine the conditional data densities for the other three models. Given the independence of the clusters, the results for the other models with more than one cluster can be written as the product of the integrals for each of the clusters. So, using notation that is an obvious generalization of that in (16):

$$\begin{aligned} p(y \mid M_2) &= \frac{1}{2} K_{21} \left| \frac{|C_{21}|}{|A_{21}|} \right|^{1/2} 2^{\frac{n_{21}+q_{21}}{2}} \Gamma(\frac{n_{21}+q_{21}}{2})(q_{21}\bar{s}_{21}^{2} + v_{21}s_{21}^{2} + Q_{21a} + Q_{21b})^{-\frac{n_{21}+q_{21}}{2}} \\ &\quad \times \frac{1}{2} K_{22} \left| \frac{|C_{22}|}{|A_{22}|} \right|^{1/2} 2^{\frac{n_{22}+q_{22}}{2}} \Gamma(\frac{n_{22}+q_{22}}{2})(q_{22}\bar{s}_{22}^{2} + v_{22}s_{22}^{2} + Q_{22a} + Q_{22b})^{-\frac{n_{22}+q_{22}}{2}} \\ &= \frac{1}{4}(2\pi)^{-n/2} K_{21}K_{22} \left[\frac{|C_{21}||C_{22}|}{|A_{21}||A_{22}|} \right]^{1/2} \\ &\quad \times \frac{(q_{21}\bar{s}_{21}^{2} + v_{21}s_{21}^{2} + Q_{21a} + Q_{21b})^{-\frac{n_{21}+q_{21}}{2}} (q_{22}\bar{s}_{22}^{2} + v_{22}s_{22}^{2} + Q_{22a} + Q_{22b})^{-\frac{n_{22}+q_{22}}{2}}}{2^{\frac{n_{21}+q_{21}}{2}} \Gamma(\frac{n_{21}+q_{21}}{2}) 2^{\frac{n_{22}+q_{22}}{2}} \Gamma(\frac{n_{22}+q_{22}}{2})} \end{aligned} \qquad (17)$$

Then, using these results in (8), the Bayesian Posterior Odds between Model 1 and Model 2 are:

$$BPO_{12} = [\frac{p(M_1)}{p(M_2)}] \times [\frac{p(y\,|\,M_1)}{p(y\,|\,M_2)}]$$

$$= 2 \times [\frac{p(M_1)}{p(M_2)}] \times \frac{K_{11}}{K_{21}K_{22}} \left[\frac{|C_{11}\|A_{21}\|A_{22}|}{|A_{11}\|C_{21}\|C_{22}|}\right]^{1/2} \frac{2^{\frac{n_{11}+q_{11}}{2}}\Gamma(\frac{n_{11}+q_{11}}{2})}{2^{\frac{n_{21}+q_{21}}{2}}\Gamma(\frac{n_{21}+q_{21}}{2})2^{\frac{n_{22}+q_{22}}{2}}\Gamma(\frac{n_{22}+q_{22}}{2})}$$

$$\times \frac{(q_{21}\bar{s}_{21}^2 + v_{21}s_{21}^2 + Q_{21a} + Q_{21b})^{-\frac{n_{21}+q_{21}}{2}} (q_{22}\bar{s}_{22}^2 + v_{22}s_{22}^2 + Q_{22a} + Q_{22b})^{-\frac{n_{22}+q_{22}}{2}}}{(q_{11}\bar{s}_{11}^2 + v_{11}s_{11}^2 + Q_{11a} + Q_{11b})^{-\frac{n_{11}+q_{11}}{2}}}$$

$$= 2 \times [\frac{p(M_1)}{p(M_2)}] \times \left[\frac{|C_{11}\|A_{21}\|A_{22}|}{|A_{11}\|C_{21}\|C_{22}|}\right]^{1/2} \times (\frac{{\delta_{11}}^{n_{21}}}{{\delta_{21}}^{n_{21}}{\delta_{22}}^{n_{22}}})^{-1/2}$$

$$\times \frac{{\bar{s}_{11}}^2/\delta_{11}}{({\bar{s}_{21}}^2/\delta_{21})({\bar{s}_{22}}^2/\delta_{22})} \frac{f_{q_{11},n_{11}}({\bar{s}_{11}}^2/\delta_{11})}{f_{q_{21},n_{21}}({\bar{s}_{21}}^2/\delta_{21})\, f_{q_{22},n_{22}}({\bar{s}_{22}}^2/\delta_{22})}$$

where

$$K_{ij} = \frac{2(q_{ij}\bar{s}_{ij}^2/2)^{q_{ij}/2}}{\Gamma(q_{ij}/2)}$$

$$\delta_{ij} = (v{s_{ij}}^2 + Q_{ija} + Q_{ijb})/n_{ij}\,, \qquad (18)$$

and $f_{q_{ij},n_{ij}}({\bar{s}_1}^2/\delta_{ij})$ denotes the ordinate of the *F* p.d.f. with q_{ij} and n_{ij} degrees of freedom. The other notation in these BPO formulae is again an obvious generalization of that used in (16).

Similarly, the Bayesian Posterior Odds between Model 1 and Model 3 are:

$$
\begin{aligned}
BPO_{13} &= [\frac{p(M_1)}{p(M_3)}]\times[\frac{p(y\,|\,M_1)}{p(y\,|\,M_3)}] \\
&= [\frac{p(M_1)}{p(M_3)}]\times 2^{3-1}\,\frac{K_{11}}{K_{31}K_{32}K_{33}}\left[\frac{|C_{11}\|A_{31}\|A_{32}\|A_{33}|}{|A_{11}\|C_{31}\|C_{32}\|C_{33}|}\right]^{1/2} \\
&\quad\times\frac{2^{\frac{n_{11}+q_{11}}{2}}\Gamma(\frac{n_{11}+q_{11}}{2})}{2^{\frac{n_{31}+q_{31}}{2}}\Gamma(\frac{n_{31}+q_{31}}{2})2^{\frac{n_{32}+q_{32}}{2}}\Gamma(\frac{n_{32}+q_{32}}{2})2^{\frac{n_{33}+q_{33}}{2}}\Gamma(\frac{n_{33}+q_{33}}{2})} \\
&\quad\times(q_{31}\bar{s}_{31}^2+v_{31}s_{31}^2+Q_{31a}+Q_{31b})^{-\frac{n_{31}+q_{31}}{2}} \\
&\quad\times\frac{(q_{22}\bar{s}_{32}^2+v_{22}s_{32}^2+Q_{32a}+Q_{32b})^{-\frac{n_{32}+q_{32}}{2}}(q_{33}\bar{s}_{33}^2+v_{33}s_{33}^2+Q_{33a}+Q_{33b})^{-\frac{n_{33}+q_{33}}{2}}}{(q_{11}\bar{s}_{11}^2+v_{11}s_{11}^2+Q_{11a}+Q_{11b})^{-\frac{n_{11}+q_{11}}{2}}} \\
&= 2\times[\frac{p(M_1)}{p(M_3)}]\times\left[\frac{|C_{11}\|A_{31}\|A_{32}\|A_{33}|}{|A_{11}\|C_{31}\|C_{32}\|C_{33}|}\right]^{1/2}\times(\frac{\delta_{11}}{\delta_{31}\delta_{32}\delta_{33}})^{-n/2}\times\frac{\bar{s}_{11}{}^2/\delta_{11}}{(\bar{s}_{31}{}^2/\delta_{31})(\bar{s}_{32}{}^2/\delta_{32})(\bar{s}_{33}{}^2/\delta_{33})} \\
&\quad\times\frac{f_{q_{11},n_{11}}(\frac{\bar{s}_{11}{}^2}{\delta_{11}})}{f_{q_{31},n_{31}}(\frac{\bar{s}_{31}{}^2}{\delta_{31}})\,f_{q_{32},n_{32}}(\frac{\bar{s}_{32}{}^2}{\delta_{32}})\,f_{q_{33},n_{33}}(\frac{\bar{s}_{33}{}^2}{\delta_{33}})}
\end{aligned}
\qquad (19)
$$

The Bayesian Posterior Odds between Model 2 and Model 3 are:

$$\begin{aligned} BPO_{23} &= [\frac{p(M_2)}{p(M_3)}]\times[\frac{p(y\,|\,M_2)}{p(y\,|\,M_3)}] \\ &= 2^{3-2}\times[\frac{p(M_2)}{p(M_3)}]\times\frac{K_{21}K_{22}}{K_{31}K_{32}K_{33}}\left[\frac{|C_{21}||C_{22}||A_{31}||A_{32}||A_{33}|}{|A_{21}||A_{22}||C_{31}||C_{32}||C_{33}|}\right]^{1/2} \\ &\quad\times\frac{2^{\frac{n_{21}+q_{21}}{2}}\Gamma(\frac{n_{21}+q_{21}}{2})2^{\frac{n_{22}+q_{22}}{2}}\Gamma(\frac{n_{22}+q_{22}}{2})}{2^{\frac{n_{31}+q_{31}}{2}}\Gamma(\frac{n_{31}+q_{31}}{2})2^{\frac{n_{32}+q_{32}}{2}}\Gamma(\frac{n_{32}+q_{32}}{2})2^{\frac{n_{33}+q_{33}}{2}}\Gamma(\frac{n_{33}+q_{33}}{2})} \\ &\quad\times(q_{31}\bar{s}_{31}^2+v_{31}s_{31}^2+Q_{31a}+Q_{31b})^{-\frac{n_{31}+q_{31}}{2}} \\ &\quad\times\frac{(q_{22}\bar{s}_{32}^2+v_{22}s_{32}^2+Q_{32a}+Q_{32b})^{-\frac{n_{32}+q_{32}}{2}}(q_{33}\bar{s}_{33}^2+v_{33}s_{33}^2+Q_{33a}+Q_{33b})^{-\frac{n_{33}+q_{33}}{2}}}{(q_{21}\bar{s}_{21}^2+v_{21}s_{21}^2+Q_{21a}+Q_{21b})^{-\frac{n_{21}+q_{21}}{2}}(q_{22}\bar{s}_{22}^2+v_{22}s_{22}^2+Q_{22a}+Q_{22b})^{-\frac{n_{22}+q_{22}}{2}}} \\ &= [\frac{p(M_2)}{p(M_3)}]\times\left[\frac{|C_{21}||C_{22}||A_{31}||A_{32}||A_{33}|}{|A_{21}||A_{22}||C_{31}||C_{32}||C_{33}|}\right]^{1/2}\times(\frac{{\delta_{21}}^{n_{21}}{\delta_{22}}^{n_{22}}}{{\delta_{31}}^{n_{31}}{\delta_{32}}^{n_{32}}{\delta_{33}}^{n_{33}}})^{-1/2} \\ &\quad\times\frac{({\bar{s}_{21}}^2/\delta_{21})({\bar{s}_{22}}^2/\delta_{22})}{({\bar{s}_{31}}^2/\delta_{31})({\bar{s}_{32}}^2/\delta_{32})({\bar{s}_{33}}^2/\delta_{33})}\frac{f_{q_{21},n_{21}}(\frac{{\bar{s}_{21}}^2}{\delta_{21}})f_{q_{22},n_{22}}(\frac{{\bar{s}_{22}}^2}{\delta_{22}})}{f_{q_{31},n_{31}}(\frac{{\bar{s}_{31}}^2}{\delta_{31}})f_{q_{32},n_{32}}(\frac{{\bar{s}_{32}}^2}{\delta_{32}})f_{q_{33},n_{33}}(\frac{{\bar{s}_{33}}^2}{\delta_{33}})} \end{aligned}$$

(20)

Further details relating to the 4-cluster model are available from the authors on request, to conserve space.

After we calculate the Bayesian Posterior Odds, under a symmetric loss function, if the Bayesian Posterior Odds are greater than 1, say $\text{BPO}_{12} > 1$, then Model 1 is preferred to Model 2, *etc*. The prior information becomes diffuse in this Natural Conjugate Prior case as $|C_{ij}| \to 0$ and $q_{11} = q_{21} = q_{22} \to 0$. Under these condition, the Bayesian estimator collapses to the OLS estimator. In this case it is readily shown that under some mild conditions, and with prior odds for the model of unity, choosing Model 1 if BPO_{12} exceeds unity (as would be the case under a symmetric loss function) is equivalent to choosing the model with the higher R^2. This last result, and a comprehensive discussion of the roles of the various terms in BPO_{12} can be found in Zellner (1971, pp.309-312). It is also well known that the BPO become indeterminate in the exactly diffuse prior case, a point that has been taken into account in the next two sections.

In summary, our Bayesian Fuzzy Regression analysis involves clustering the data into *c* fuzzy clusters, where the optimal value of *c* is determined by BPO analysis. Then, Bayesian regression models are fitted over each of the *c* clusters, and the results are combined using the membership functions as weights.

4. Monte Carlo Experiment

5.1 Experimental Design

There are four parts to the Monte Carlo experiment that we have conducted. The objective of the experiment is to assess the performance of the above BPO analysis in selecting the appropriate number of clusters to use in fuzzy regression analysis. The first set of experiments involves a one-cluster data generating process. Ideally, in this case the BPO should favor a fuzzy model with one cluster all of the time. The other three sets of experiments involve two-cluster, three-cluster and four-cluster data generating processes. The programming for the Monte Carlo experiment has been done with SHAZAM (2001) code.

The sample sizes that have been considered are n = 24, 60, 120, 240, 480 and 1200 for every part of the experiment. The number of Monte Carlo repetitions is set to 1000, and this number was found to be more than sufficient to ensure the stability of the results. We made the prior information for the parameters minimal by setting q_{ij} = 0.01 and C_{ij} = 0.001 (for all *i*, *j*). The values for the other parameters of the prior are $\overline{\beta}_{ij}$ = 0.2 and $\overline{s}_{ij}^{2}$ = 2 (for all *i*, *j*), although the results are fully robust to these last choices. We assign equal prior probabilities for each of the competing models, so that we do not favor any particular model before the analysis. In order to make the results more general, in each part of the experiment we let the variance of the data generating process vary between 1 and 10, which helps us to increase the data generating process's degree of fuzziness and this in turn provides a significant challenge for the BPO analysis.

The data generating process that has been used comprises a number of separate line segments, one for each cluster in the underlying process:

$$y_{ij} = \alpha_j + \beta_j x_{ij} + \varepsilon_{ij} \;;\; i = 1, 2, 3, \ldots, (n/c) \;;\; j = 1, 2, \ldots, c.$$

In other words, the data may be generated using one, two, three or four clusters. The regressor series is $\{x_i\}$ =$\{(i / 100)$: $i = 1, 2, \ldots, n\}$, and $\{x_{ij}\} = \{x_i : i = (n / c)+ (j -1), \ldots.., (jn / c)\}$; $j = 1, 2, \ldots, c$. The intercept parameters take the values {1}, {1, 5}, {1, 5, 6} and {1, 5, 6, 9} for the cases of one to four clusters respectively. Two cases are considered as far as the slope coefficients in the data-generating process are concerned. The first (the "big slope" case) has β values of {1}, {1, -5}, {1, -5, 3 } and {1, -5, 3, -8} for the cases of one to four clusters respectively. The second (the "small slope" case) has β values of {0.1}, {0.1, -0.5}, {0.1, -0.5, 0.3 }and {0.1, -0.5, 0.3, -0.8} for the cases of one to four clusters respectively. The error terms, "ε_{ij}", are generated as being independent and normally distributed, with mean zero. Different error variances with the values 1, 2, 3, …., 10 are considered.

Figures 1 and 2 show the four data generating processes with a sample size of 240. In each case here the variance for the data generating process is 3. These graphs provide an indication of the degree of fuzziness of the data that are used in the Monte Carle experiment, and they show how difficult it would be for us in real life to determine the correct number of clusters just by simply looking at a data plot alone. These plots are only a small part of all the Monte Carlo experiments we undertook in this research.

5.2 Experiment Results

5.2.1 Results with "Big" Slopes

Our results reported in the various tables show the probability that the BPO analysis selects the true model over the 1000 repetitions in any part of the experiment. Tables 1 to 3 provide the results for the three sets of experiments when the data generating process runs from two clusters to four clusters with the "big" slope parameters defined in section 5.1.[2] In all of the following tables, P_{ij} denotes the probability that the BPO favour model i over model j based on the 1000 repetitions.

First, by way of illustration, consider the section of Table 1 corresponding to n = 240. Across the columns, the variance for the DGP increases from 1 to 10. When the variance equals 1, we have:

$$M_1 \succ M_2 \text{, with 0\% probability } \Leftrightarrow M_1 \prec M_2 \text{ with 100\% probability.}$$

[2] We found that when the data generating process involves only one cluster, the BPO identify the correct model 100% of the time, regardless of the level of the variance in the data generating process.

$M_1 \succ M_3$, with 0% probability $\Leftrightarrow M_1 \prec M_3$ with 100% probability.

$M_1 \succ M_4$, with 0% probability $\Leftrightarrow M_1 \prec M_4$ with 100% probability.

$M_2 \succ M_3$, with 100% probability.

$M_2 \succ M_4$, with 100% probability.

$M_3 \succ M_4$, with 100% probability.

As a result, we can rank the four models in the order: $M_2 \succ M_3 \succ M_4 \succ M_1$, where $A \succ B$ indicates "*A* is preferred to *B*"

In the illustrative case being discussed here, the BPO analysis correctly selects the model with two fuzzy clusters with 100% probability. We find a similar ranking among the models when the variance is 2. As the variance increases to 3, the two-cluster fuzzy model still is chosen correctly 70% of the time among the four models, though the ranking of the three-cluster and four-cluster fuzzy models is now reversed, and the three-cluster fuzzy model is chosen over the one-cluster fuzzy model with 30% probability. Nevertheless, the result that the two-cluster fuzzy model is the best among the four models still holds. As the variance increases to 4, the probability that the two-cluster fuzzy model is chosen correctly over the one-cluster fuzzy model decreases to 80%, and overall the two-cluster fuzzy model is still the best. As the degree of fuzziness in the data generating process increases further (with the variance greater than or equal to 5), the one-cluster fuzzy model rules the rest of the three models including the true model with two fuzzy clusters. Perhaps not surprisingly, when the data are extremely diffuse, there is a tendency for the model selection procedure to eventually infer that there is just a single cluster of data.

Second, as we move to the next section in Table 1 (where the sample size increases to 480 and the data generating process is still two clusters), we find that the BPO's ability to detect the true model increases dramatically. Even with a variance of 10, the fuzzy two-cluster model is correctly chosen 100% of the time. In this case we see that the true model dominates the other three models even in the fuzziest case considered, with 100% probability. Similar results emerge in Tables 2 and 3.

Across Tables 1 to 3 we see that the ability of the BPO analysis to select the true model increases as the number of observations increases in the sample. The results all across the three tables for the case where the n = 24 are always poor. However, for n > 60 we begin to see some improvement and when the sample size is above 240 and around 480, we find the BPO pick the true model nearly every time when the degree of the fuzziness is moderate. The value of the variance we use for the data generating process also affects the results and as the data become fuzzier, the probability of the BPO picking the true model decreases.

Another important result is that the BPO tend to favour models based on few clusters. In Table 1 we see that even when the variance increases to 10, the three-cluster and four-cluster fuzzy models are never ranked above the true two-cluster fuzzy model. In Table 2, the four-cluster fuzzy model is never ranked above the true three-cluster fuzzy model.

In practice, sometimes when we look at a plot of the data, we may get a sense that there should be more than one cluster and as a result, the one-cluster model can be eliminated at the every beginning. Over all, the one-cluster model suggests that there is no nonlinearity for the data. The next part of the Monte Carlo experiment relates to the case where the data have been generated with the number of clusters being greater than or equal to 2. In this case, we decrease the model space to three models – those based on two fuzzy clusters, three fuzzy clusters and four fuzzy clusters. The results are similar to those discussed already for the case where we have four candidate models.

Table 4, 5 and 6 provide the results for this part of the experiment as we keep the other test standards the same as before. We see the BPO's performance of selecting the true model has been enhanced. All of the other features noted for the first four tables also hold in this case, and the sample size only needs to exceed 120 for the true model to be selected with high probability. In Table 4, we see that in all cases where n > 60, the BPO select model 2 (the correct model) with 100% probability. In Table 6, we see that as the sample size is increased to 240, we can select the correct model nearly all of the time. The only result that changes a little in this case is the ranking of the other two models.

5.2.2 Results with "Small" Slopes

In the following part of the Monte Carlo experiment, we have used the "small slope" coefficient vector for the data-generating process, as defined in section 5.1. From the data generating process, we can see that the degree of the fuzziness added to the sample comes from two sources:

one is the variance of the error terms that we assign for the data-generating process, and the other is through the slope parameters. As we see from Figure 2 (a) to (d), with the "small slopes" the data become more and more "cloudy" and it is very difficult to determine the correct number of clusters visually. This part of the Monte Carlo experiment provides an even more stringent test of the BPO analysis. The variance value for the data generation process again runs from 1 to 10, and the sample size runs from 24 to 1200, and there are 1000 repetitions.

Tables 7 to 9 provide the probabilities that the BPO favor the first model over the second model over the 1000 repetitions. The four fuzzy models we are considering are: one-cluster, two-cluster, three-cluster and four-cluster fuzzy models[3]. The data generating process for those three tables as before runs from two cluster to four clusters. For the three tables here, the results in the case are similar to those in the "big slope" case — the performance of the BPO increases as the sample size increases and decreases as the value of the variance increases. Tables 10 to 12 give the results for the BPO analysis when we reduce the model space by dropping the model with one cluster. Again, the results are similar to those in section 5.2.1.

The main difference between the results for the "big slope" and "small slope" parts of the Monte Carlo experiment is the understandable reduction in the ability of the BPO analysis to choose the true model in the latter case. Overall, as the number of observations increases, the BPO's ability to pick the true model becomes stronger. It is a "consistent" selection procedure. Some experimentation with the specification of the prior information about the models' parameters indicated that our results are quite robust in this respect.

5. Two Real Data Applications

In this section, we apply the BPO analysis to two real data sets.

6.1 Motorcycle Data

This well-known data set relates to the head acceleration of a PMTO (*post mortem* human test object), measured after a simulated motorcycle impact (Schmidt, *et al*., 1981). Figure 3 provides a

[3] Again, regardless of the level of the variation in the data generating process with one cluster, the model picked by the BPO is *always* the correct one.

plot of the data. Due to the nature of the experiment, there are many observations at the same time points From a first look at the data, we might guess that the number of clusters for this set of data might be smaller than four. To be consistent with our previous experiment, the model space in this BPO analysis is set to include four models: models with one to four fuzzy clusters. We assign each model a prior probability of 0.25.

First, we cluster the sample into two clusters, three clusters and four clusters respectively according to the *X* variable, which in this example is the time after impact. The values of the membership functions are generated accordingly for each model. Employing a relatively uninformative natural conjugate prior for each model's parameters, we get the Bayesian fuzzy regression in the weighted average fashion using the membership functions as the weights. The fuzzy model with three clusters is chosen by the BPO. In Figure 3 we see that the three-cluster and four-cluster fuzzy models clearly outperform the fuzzy models with one cluster and two clusters in terms of the fit to the original data. When the fit in the left tail of the data is considered, there is an informal preference for the three-cluster model.

The posterior probabilities for the models in Table 13 confirm this, being 0.98 and 0.02 for the three-cluster and four-cluster models respectively, and essentially zero for the other models. This conclusion is very robust to the choice of prior masses on the model space. Figure 4 shows the fitted regressions from the three-cluster fuzzy model and standard nonparametric kernel regression. The fuzzy three-cluster model out-performs the nonparametric model not only in terms of fit (RMSE = 26.75 and 37.99 respectively), but also with respect to the shape of the fitted regressions in the middle range of the data.

6.2 Journal Subscriptions Data

Our second application relates to prices and the number of library subscriptions for 180 economics journals for the year 2000 (Bergstrom, 2001). Our objective is to fit a regression to explain the number of subscriptions to a journal as a function of its price. Among the four possible fuzzy models, the BPO analysis selects the model based on four fuzzy clusters. In Figure 5, we show the fits of all the candidate models. Clearly, the model with four fuzzy clusters best describes the data. In Figure 6, we plot both the preferred fuzzy regression and a nonparametric kernel regression. The fuzzy four-cluster model outperforms the nonparametric model in the left tail and also the middle range of the data. The RMSE from the fuzzy four-cluster model is 174.33 and that from the nonparametric approach is 178.91.

6. Conclusions

In this study, we set up a standard Bayesian approach to the model selection problem of choosing the number of fuzzy clusters to be used in the context of the recently developed fuzzy regression analysis. Using the Bayesian Posterior Odds to select the number of clusters to be used improves the fuzzy regression analysis in an important way. The use of a Bayesian approach to both model selection and regression estimation illustrates one of merits of Bayesian inference – namely its unity, and the flexibility with which prior information about the model space and the parameter space can be incorporated into the analysis.

Our Monte Carlo experiments show how powerful this approach can be, especially with moderate to large samples. Even when the sample data visually appear to be generated from one “regime”, the BPO analysis is very successful in determining the true number of underlying clusters. Two brief applications with real data are also extremely encouraging. Compared with standard non-parametric kernel regression, the Bayesian fuzzy regression captures the nonlinearity of the data extremely well. The sample sizes in these two applications does not exceed 180, showing that in practice Bayesian fuzzy regression can give a highly satisfactory performance even when the sample size is modest.

There are still some limitations to the application of this Bayesian fuzzy regression approach as is evident in the above results. Although the theory of this approach has been set up for the general multivariate case, the applications considered in this study focus only on the univariate case. None the less, the evidence provided in this research lends credibility to Bayesian Fuzzy Regression analysis, and especially to the use of Bayesian Posterior Odds to select the number of fuzzy clusters that are to be used. Other recent applications of fuzzy regression have shown that the methodology performs well when a frequentist approach to inference is taken in the multivariate case, and there is every reason to suppose that this will be held within a Bayesian framework.

References:

Bezbek, J. C. (1973), Fuzzy Mathematics in Pattern Classification, Ph.D. Thesis, Applied Mathematics Center, Cornell University, Ithaca, NY.

Bezbek, J. C. (1981), *Pattern Recognition With Fuzzy Objective Function Algorithms*, Plenum Press, New York.

Chen, J. & D. E. A. Giles (2004), "Gender Convergence in Crime: Evidence From Canadian Adult Offense Charge Data", *Journal of Criminal Justice,* 32, 593-606.

Dunn, J. C. (1974), "Well Separated Clusters and Optimal Fuzzy Partitions", *Journal of Cybernetics*, 4, 95-104.

Dunn, J. C. (1977), "Indices of Partition Fuzziness and the Detection of Clusters in Large Data Sets", in M. Gupta and G. Seridis (eds.), *Fuzzy Automata and Decision Processes*, Elsevier, New York.

Draeseke, R. and D. E. A. Giles, 2002, "Modelling the New Zealand Underground Economy Using Fuzzy Logic Techniques", *Mathematics and Computers in Simulation*, 59, 115-123.

Feng, H. (2007), "Forecast Comparison Between Two Nonlinear Models in Applied Financial Econometrics", under review *Applied Economics*.

Feng, H. (2007), "Forecasting Comparison between Two Nonlinear Models: Fuzzy Regression *vs*. SETAR", under review *Journal of Economic Modelling*.

Giles, D. E. A. (2005), "Output Convergence and International Trade: Time-Series and Fuzzy Clustering Evidence for New Zealand and Her Trading Partners, 1950-1992", *Journal of International Trade and Economic Development*, 14, 93-114.

Giles, D. E. A. and R. Draeseke (2003), "Econometric Modelling Using Fuzzy Pattern Recognition via the Fuzzy c-Means Algorithm", in D. E. A. Giles (ed.), *Computer Aided Econometrics*, Marcel Dekker, New York, 407-450.

Giles, D. E. A. and H. Feng, (2005), "Output and Well-Being in Industrialized Nations in the Second Half of the 20th Century: Testing for Convergence Using Fuzzy Clustering Analysis", *Structural Change & Economic Dynamics*, 2005, 16, 285-308.

Giles, D. E. A. and C. A. Mosk (2003), "A Long-Run Environmental Kuznets Curve for Enteric CH_4 Emissions in New Zealand: A Fuzzy Regression Analysis", Econometrics Working Paper EWP0307, Department of Economics, University of Victoria.

Giles, D. E. A. and C. Stroomer (2004), “Identifying the Cycle of a Macroeconomic Time-Series Using Fuzzy Filtering”, Econometrics Working Paper EWP0406, Department of Economics, University of Victoria.

Giles, D. E. A. and C. Stroomer (2006), “Does Trade Openness Affect the Speed of Output Convergence? Some Empirical Evidence”, *Empirical Economics*, 31, 883-903.

MacQueen, J. (1967), “Some Methods for Classification and Analysis of Multivariate Observations”, in J. M. Le Cam and J. Neyman (eds.), *Proceedings of the 5th Berkeley Symposium in Mathematical Statistics and Probability*, University of California Press, Berkeley CA, 281-297.

Ruspini, E. (1970), “Numerical Methods for Fuzzy Clustering”, *Information Science*, 2, 319-350.

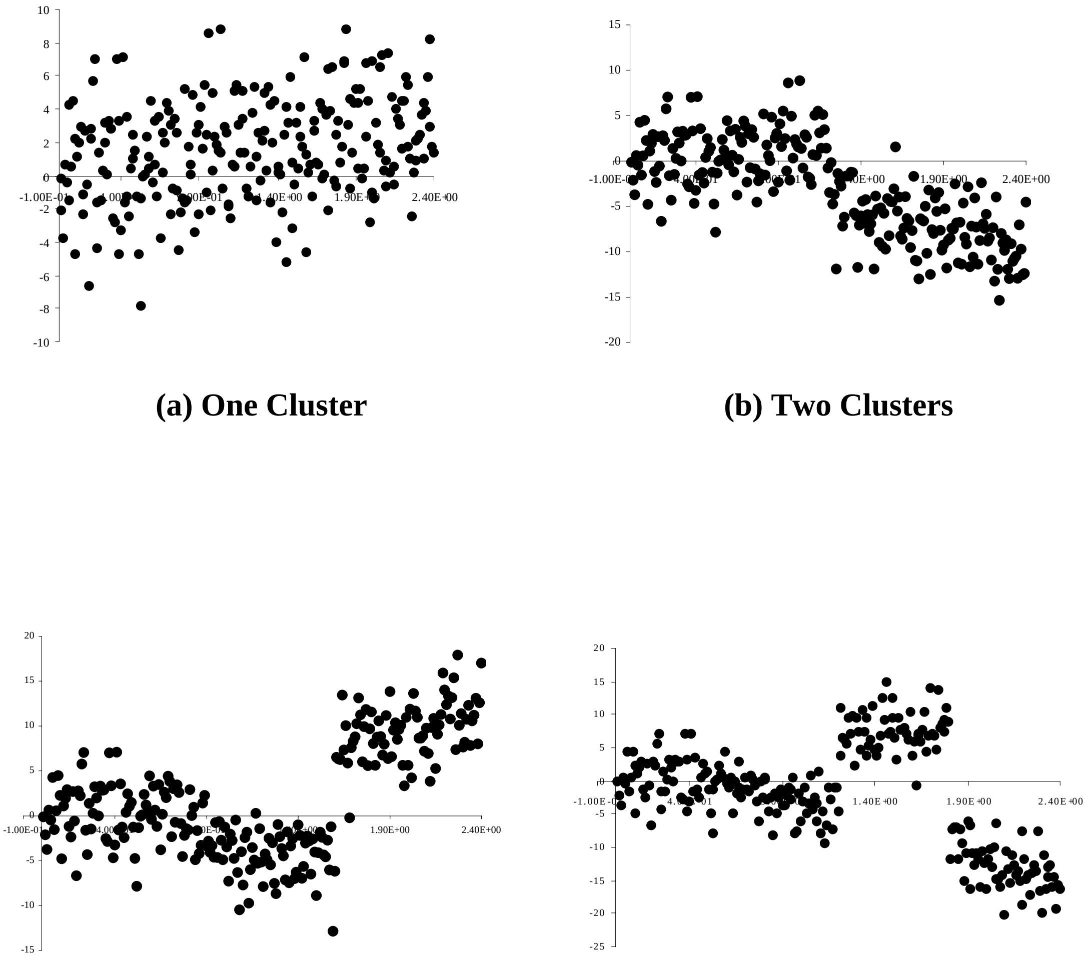
SHAZAM (2001), SHAZAM Econometrics Package, User's Guide, Version 9, Northwest Econometrics, Vancouver, B.C.

Shepherd, D. and F. K. C. Shi (1998), “Economic Modelling With Fuzzy Logic”, paper presented at the CEFES ’98 Conference, Cambridge, U.K..

Zadeh, L. A. (1965), “Fuzzy Sets”, *Information and Control*, 8, 338-353.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

Figure 1. Data Generating Process – “Big” Slope Case: n = 240



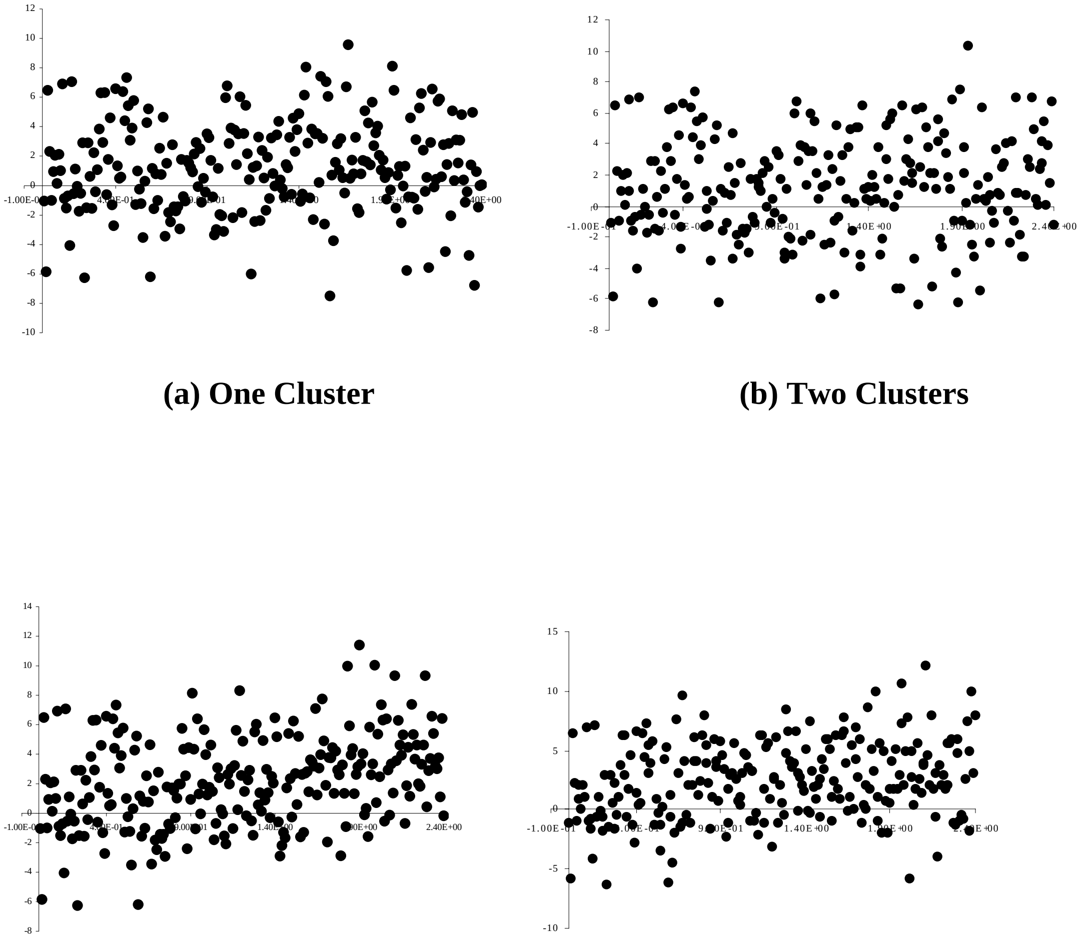
(a) One Cluster

(b) Two Clusters

(c) Three Clusters

(d) Four Clusters

Figure 2. Data Generating Process – “Small” Slope Case: n = 240



(a) One Cluster

(b) Two Clusters

(c) Three Clusters

(d) Four Clusters

Table 1. The Bayesian Posterior Odds Monte Carlo Experiment for Two Clusters

Data Generation Process: *c* = 1, 2, 3, 4

		σ = 1	σ = 2	σ = 3	σ = 4	σ = 5	σ = 6	σ = 7	σ = 8	σ = 9	σ = 10
n=24	P_{12}	0.2	0.8	0.9	0.9	0.9	0.9	1	0.9	1	1
	P_{13}	1	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=60	P_{12}	0.2	0.9	1	1	1	1	1	1	1	1
	P_{13}	1	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=120	P_{12}	0.1	0.8	1	1	1	1	1	1	1	1
	P_{13}	0.7	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=240	P_{12}	0	0	0	0.2	0.6	0.8	0.9	0.9	1	1
	P_{13}	0	0	0.7	1	1	1	1	1	1	1
	P_{14}	0	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=480	P_{12}	0	0	0	0	0	0	0	0	0	0
	P_{13}	0	0	0	0	0	0	0	0.2	0.4	0.7
	P_{14}	0	0	0	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	0	1	1	1	1	1	1	1	1	1
n=1200	P_{12}	0	0	0	0	0	0	0	0	0	0
	P_{13}	0	0	0	0	0	0	0	0	0	0
	P_{14}	0	0	0	0	0	0	0	0	0	0
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	0	0	0	0	1	1	1	1	1	1

Table 2. The Bayesian Posterior Odds Monte Carlo Experiment for Three Clusters Data Generation Process: c = 1, 2, 3, 4

		$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 6$	$\sigma = 7$	$\sigma = 8$	$\sigma = 9$	$\sigma = 10$
n=24	P_{12}	0.9	0.9	0.9	1	1	1	1	1	1	1
	P_{13}	1	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	0.9	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=60	P_{12}	0	0.5	0.9	0.9	1	1	1	1	1	1
	P_{13}	0	0.9	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	0.2	0.9	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=120	P_{12}	0	0	0	0.2	0.5	0.7	0.8	0.9	1	1
	P_{13}	0	0	0	0.5	0.9	1	1	1	1	1
	P_{14}	0	1	1	1	1	1	1	1	1	1
	P_{23}	0	0	0.4	0.8	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=240	P_{12}	0	0	0	0	0	0	0	0	0	0.1
	P_{13}	0	0	0	0	0	0	0	0	0.1	0.4
	P_{14}	0	0	0	0.9	1	1	1	1	1	1
	P_{23}	0	0	0	0	0	0.1	0.2	0.5	0.7	0.8
	P_{24}	0	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=480	P_{12}	0	0	0	0	0	0	0	0	0	0
	P_{13}	0	0	0	0	0	0	0	0	0	0
	P_{14}	0	0	0	0	0	0	0	0.3	0.9	1
	P_{23}	0	0	0	0	0	0	0	0	0	0
	P_{24}	0	0	0.2	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=1200	P_{12}	0	0	0	0	0	0	0	0	0	0
	P_{13}	0	0	0	0	0	0	0	0	0	0
	P_{14}	0	0	0	0	0	0	0	0	0	0
	P_{23}	0	0	0	0	0	0	0	0	0	0
	P_{24}	0	0	0	0	0	0	0	0	0.5	1
	P_{34}	1	1	1	1	1	1	1	1	1	1

Table 3. The Bayesian Posterior Odds Monte Carlo Experiment for Four Clusters Data Generation Process: c = 1, 2, 3, 4

		σ = 1	σ = 2	σ = 3	σ = 4	σ = 5	σ = 6	σ = 7	σ = 8	σ = 9	σ = 10
n=24	P_{12}	0.4	0.9	0.9	0.9	0.9	1	1	1	1	1
	P_{13}	1	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=60	P_{12}	0	0	0.4	0.7	0.9	0.9	0.9	1	1	1
	P_{13}	0	0.8	1	1	1	1	1	1	1	1
	P_{14}	0.6	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=120	P_{12}	0	0	0	0	0	0.2	0.4	0.5	0.7	0.8
	P_{13}	0	0	0	0.2	0.7	0.9	1	1	1	1
	P_{14}	0	0	0.9	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	0.1	1	1	1	1	1	1	1	1	1
	P_{34}	0	1	1	1	1	1	1	1	1	1
n=240	P_{12}	0	0	0	0	0	0	0	0	0	0
	P_{13}	0	0	0	0	0	0	0	0	0	0.1
	P_{14}	0	0	0	0	0	0.4	0.9	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	0	0	0.5	1	1	1	1	1	1	1
	P_{34}	0	0	0.1	1	1	1	1	1	1	1
n=480	P_{12}	0	0	0	0	0	0	0	0	0	0
	P_{13}	0	0	0	0	0	0	0	0	0	0
	P_{14}	0	0	0	0	0	0	0	0	0	0
	P_{23}	1	1	1	1	0.9	0.9	0.9	0.9	0.9	0.9
	P_{24}	0	0	0	0	0	0	0.5	1	1	1
	P_{34}	0	0	0	0	0	0	0.1	0.9	1	1
n=1200	P_{12}	0	0	0	0	0	0	0	0	0	0
	P_{13}	0	0	0	0	0	0	0	0	0	0
	P_{14}	0	0	0	0	0	0	0	0	0	0
	P_{23}	1	1	1	1	1	1	1	1	1	0.9
	P_{24}	0	0	0	0	0	0	0	0	0	0
	P_{34}	0	0	0	0	0	0	0	0	0	0

Table 4. The Bayesian Posterior Odds Monte Carlo Experiment for Two Clusters Data Generation Process: c = 2, 3, 4

		$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 6$	$\sigma = 7$	$\sigma = 8$	$\sigma = 9$	$\sigma = 10$
n=24	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=60	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=120	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=240	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=480	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	0	1	1	1	1	1	1	1	1	1
n=1200	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	0	0	0	0	1	1	1	1	1	1

Table 5. The Bayesian Posterior Odds Monte Carlo Experiment for Three Clusters Data Generation Process: c = 2, 3, 4

		$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 6$	$\sigma = 7$	$\sigma = 8$	$\sigma = 9$	$\sigma = 10$
n=24	P_{23}	0.9	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=60	P_{23}	0.2	0.9	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=120	P_{23}	0	0	0.4	0.8	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=240	P_{23}	0	0	0	0	0	0.1	0.2	0.5	0.7	0.8
	P_{24}	0	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=480	P_{23}	0	0	0	0	0	0	0	0	0	0
	P_{24}	0	0	0.2	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=1200	P_{23}	0	0	0	0	0	0	0	0	0	0
	P_{24}	0	0	0	0	0	0	0	0	0.5	1
	P_{34}	1	1	1	1	1	1	1	1	1	1

Table 6. The Bayesian Posterior Odds Monte Carlo Experiment for Four Clusters Data Generation Process: c = 2, 3, 4

		$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 6$	$\sigma = 7$	$\sigma = 8$	$\sigma = 9$	$\sigma = 10$
n=24	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=60	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=120	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	0.1	1	1	1	1	1	1	1	1	1
	P_{34}	0	1	1	1	1	1	1	1	1	1
n=240	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	0	0	0.5	1	1	1	1	1	1	1
	P_{34}	0	0	0.1	1	1	1	1	1	1	1
n=480	P_{23}	1	1	1	1	0.9	0.9	0.9	0.9	0.9	0.9
	P_{24}	0	0	0	0	0	0	0.5	1	1	1
	P_{34}	0	0	0	0	0	0	0.1	0.9	1	1
n=1200	P_{23}	1	1	1	1	1	1	1	1	1	0.9
	P_{24}	0	0	0	0	0	0	0	0	0	0
	P_{34}	0	0	0	0	0	0	0	0	0	0

Table 7. The Bayesian Posterior Odds Monte Carlo Experiment for Two Clusters Data Generation Process With Small Slopes: c = 1, 2, 3, 4

		σ = 1	σ = 2	σ = 3	σ = 4	σ = 5	σ = 6	σ = 7	σ = 8	σ = 9	σ = 10
n=24	P_{12}	0.1	0.6	0.9	0.9	0.9	0.9	0.9	0.9	1	1
	P_{13}	0.9	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=60	P_{12}	0	0.4	0.8	0.9	1	1	1	1	1	1
	P_{13}	0.7	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=120	P_{12}	0	0.2	0.7	0.9	1	1	1	1	1	1
	P_{13}	0.1	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=240	P_{12}	0	0	0.5	0.8	0.9	1	1	1	1	1
	P_{13}	0	0.9	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=480	P_{12}	0	0	0.4	0.8	0.9	1	1	1	1	1
	P_{13}	0	0.8	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=1200	P_{12}	0	0	0.1	0.6	0.9	1	1	1	1	1
	P_{13}	0	0.1	0.9	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1

Table 8. **The Bayesian Posterior Odds Monte Carlo Experiment for Three Clusters Data Generation Process With Small Slopes: c = 1, 2, 3, 4**

		σ = 1	σ = 2	σ = 3	σ = 4	σ = 5	σ = 6	σ = 7	σ = 8	σ = 9	σ = 10
n=24	P_{12}	1	1	1	1	1	1	1	1	1	1
	P_{13}	1	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=60	P_{12}	1	1	1	1	1	1	1	1	1	1
	P_{13}	1	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=120	P_{12}	1	1	1	1	1	1	1	1	1	1
	P_{13}	1	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	0.9	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=240	P_{12}	0.1	0.9	1	1	1	1	1	1	1	1
	P_{13}	0	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	0.1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=480	P_{13}	0	0	0.5	0.9	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	0	0	0.7	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=1200	P_{12}	0	0	0	0	0	0	0	0	0	0.1
	P_{13}	0	0	0	0	0	0	0	0	0.1	0.3
	P_{14}	0	0	1	1	1	1	1	1	1	1
	P_{23}	0	0	0	0	0	0	0.1	0.4	0.6	0.8
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1

Table 9. The Bayesian Posterior Odds Monte Carlo Experiment for Four Clusters Data Generation Process With Small Slopes: c = 1, 2, 3, 4

		$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 6$	$\sigma = 7$	$\sigma = 8$	$\sigma = 9$	$\sigma = 10$
n=24	P_{12}	1	1	1	1	1	1	1	1	1	1
	P_{13}	1	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=60	P_{12}	0.9	1	1	1	1	1	1	1	1	1
	P_{13}	1	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=120	P_{12}	0.5	1	1	1	1	1	1	1	1	1
	P_{13}	1	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=240	P_{12}	0	0.2	0.8	1	1	1	1	1	1	1
	P_{13}	0	1	1	1	1	1	1	1	1	1
	P_{14}	1	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=480	P_{12}	0	0	0	0	0.3	0.6	0.8	0.9	0.9	1
	P_{13}	0	0	0.1	0.8	1	1	1	1	1	1
	P_{14}	0	1	1	1	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=1200	P_{12}	0	0	0	0	0	0	0	0	0	0
	P_{13}	0	0	0	0	0	0	0	0	0	0
	P_{14}	0	0	0	0.7	1	1	1	1	1	1
	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	0	1	1	1	1	1	1	1	1	1
	P_{34}	0	0.8	1	1	1	1	1	1	1	1

Table 10. The Bayesian Posterior Odds Monte Carlo Experiment for Two Clusters Data Generation Process With Small Slopes: c = 2, 3, 4

		$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 6$	$\sigma = 7$	$\sigma = 8$	$\sigma = 9$	$\sigma = 10$
n=24	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=60	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=120	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=240	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=480	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=1200	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1

Table 11. The Bayesian Posterior Odds Monte Carlo Experiment for Three Clusters Data Generation Process With Small Slopes: c = 2, 3, 4

		$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 6$	$\sigma = 7$	$\sigma = 8$	$\sigma = 9$	$\sigma = 10$
n=24	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=60	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=120	P_{23}	0.9	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=240	P_{23}	0.1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=480	P_{23}	0	0	0.7	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=1200	P_{23}	0	0	0	0	0	0	0.1	0.4	0.6	0.8
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1

Table 12. The Bayesian Posterior Odds Monte Carlo Experiment for Four Clusters Data Generation Process With Small Slopes: c = 2, 3, 4

		σ = 1	σ = 2	σ = 3	σ = 4	σ = 5	σ = 6	σ = 7	σ = 8	σ = 9	σ = 10
n=24	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=60	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=120	P23	1	1	1	1	1	1	1	1	1	1
	P24	1	1	1	1	1	1	1	1	1	1
	P34	1	1	1	1	1	1	1	1	1	1
n=240	P23	1	1	1	1	1	1	1	1	1	1
	P24	1	1	1	1	1	1	1	1	1	1
	P34	1	1	1	1	1	1	1	1	1	1
n=480	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	1	1	1	1	1	1	1	1	1	1
	P_{34}	1	1	1	1	1	1	1	1	1	1
n=1200	P_{23}	1	1	1	1	1	1	1	1	1	1
	P_{24}	0	1	1	1	1	1	1	1	1	1
	P_{34}	0	0.8	1	1	1	1	1	1	1	1

Table 13. Bayesian Posterior Odds Analysis Report for the Real Data Applications

Bayesian Posterior Probability		
Posterior Probability	Journal Subscriptions	Motorcycle
P1	0.000	0.000
P2	0.000	0.000
P3	0.000	0.998
P4	1.000	0.002

Figure 3. Bayesian Fuzzy Regression for the Motorcycle Data

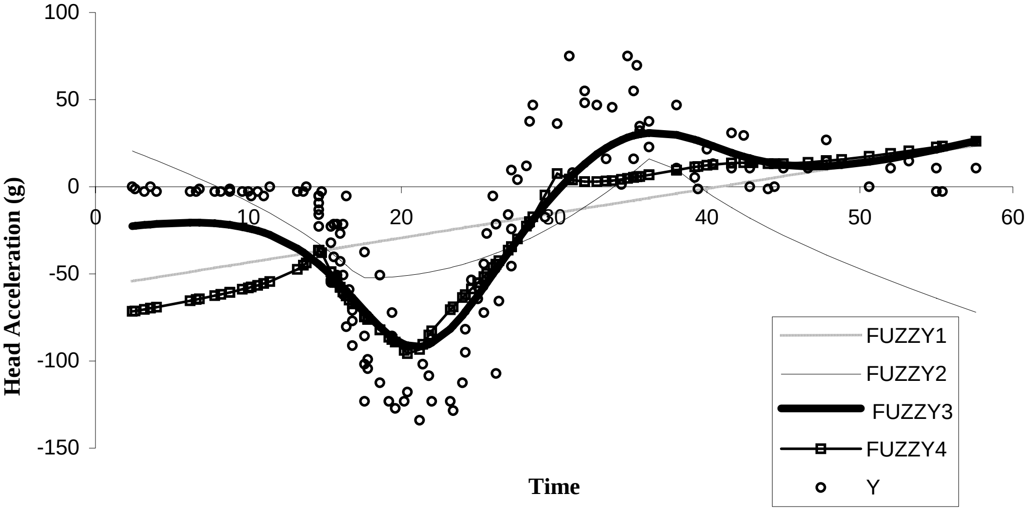


Figure 4. Comparison Between Bayesian Fuzzy Regression and Non-Parametric Regression for the Motorcycle Data

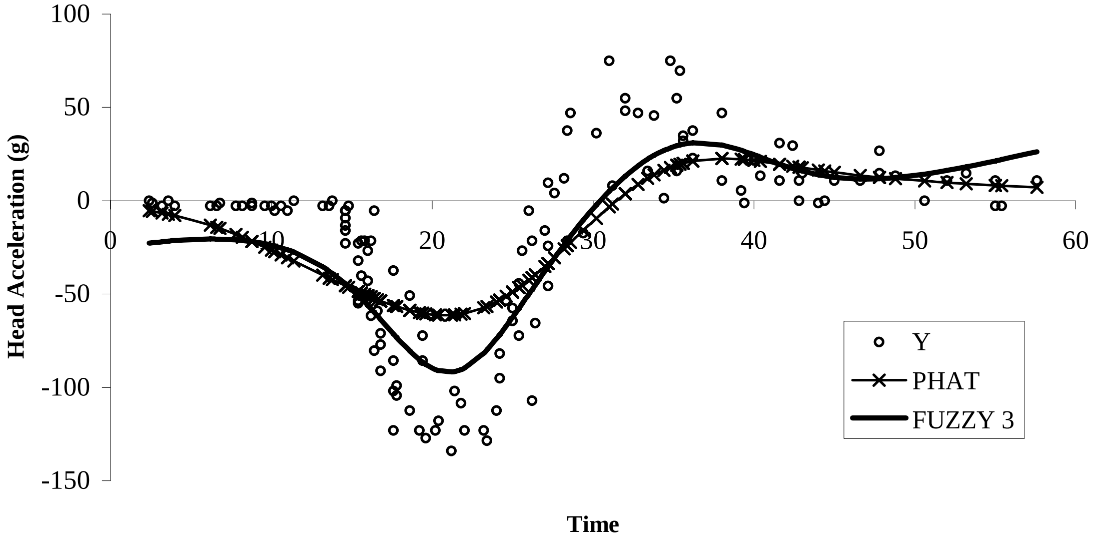


Figure 5. Bayesian Fuzzy Regression for the Journal Subscription Data

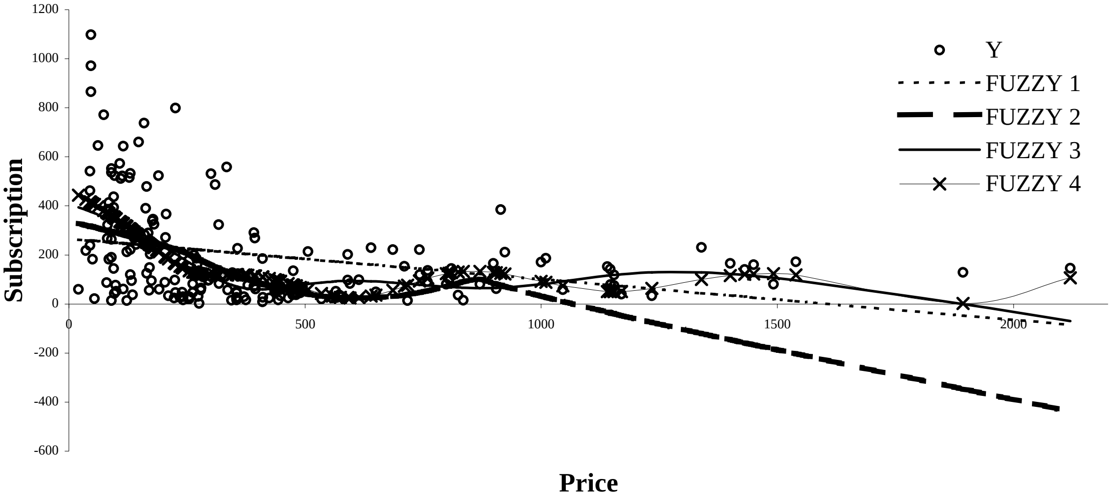
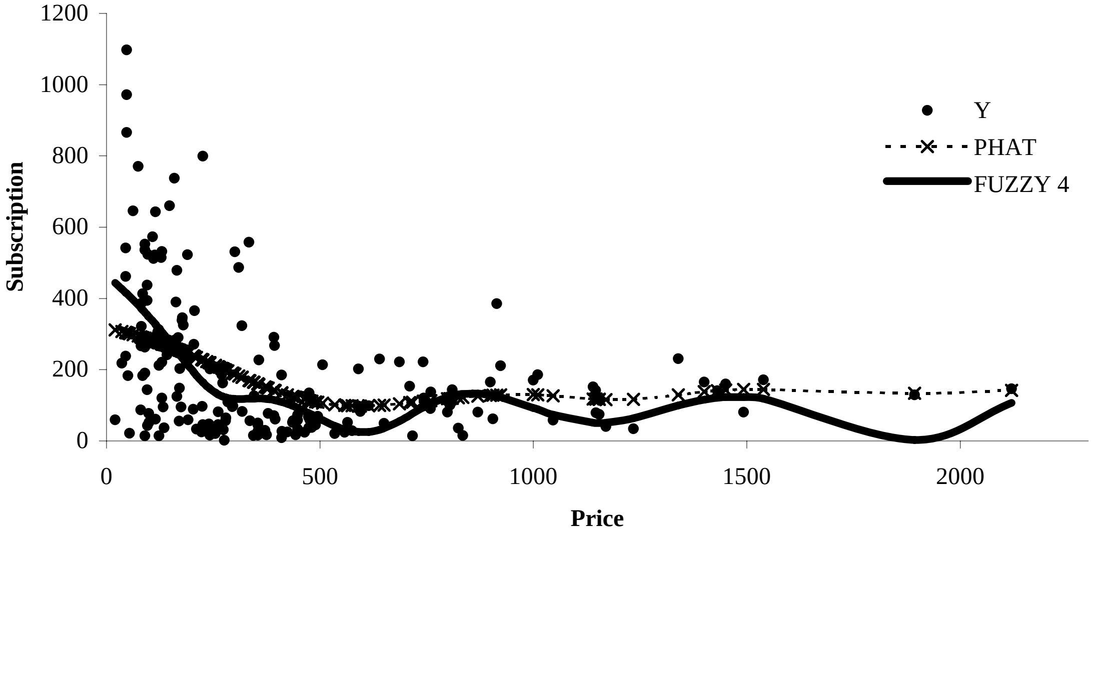


Figure 6. Comparison Between the Bayesian Fuzzy Regression and the Non-Parametric Regression for the Journal Subscription Data


Subscription
Price
Y
PHAT
FUZZY 4