
Research article

A New Approach for Model Selection with Two Qualitative Regressors

Amr R. Kamel^{1,*}, Abdulaziz A. Alqarni²

¹ Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt.

¹ Department of Basic Sciences, Elgazeera High Institute for Computers and Information Systems, Ministry of Higher Education, Cairo, Egypt.

² Department of General Studies, Jeddah College of Technology, Technical and Vocational Training Corporation, Jeddah, Saudi Arabia.

*Correspondence: amr_ragab@pg.cu.edu.eg

Abstract: In social science and statistical literature, ordinal qualitative dependent variable models have received substantial attention in terms of theory and application. However, linear models with ordinal qualitative regressors have been overlooked. In this paper, we propose an approach to regression model selection with two-qualitative regressors, which will be used to build a technique for selecting qualitative variables based on the mean square of the prediction error (MSEP). Several data sets were simulated using two models, and the technique was found to choose the best model in all cases. The findings show that significant improvements in bias and efficiency may be made when compared to other estimates. To diagnose the type of interaction, certain graphs are provided. The proposed approach achieves better accuracy with reduced MSEP, improved stability, and faster convergence compared to other models.

Keywords: Classification Performances; Model Selection; Modeling of Interaction; MSEP; Neighbouring Model; Ordinal Qualitative Variables; Statistical Inference.

Mathematics Subject Classification: 62H30; 76B15; 62J20.

Received: 28 October 2022; **Revised:** 19 November 2022; **Accepted:** 10 December 2022; **Published:** 14 December 2022.

1. Introduction

After proving that children's height does not seem to mirror that of their parents but rather regresses to the average population, Sir Francis Galton first used the term "regression" in 1885 [1]. Currently, the phrase "regression analysis" refers to a wide range of statistical methods for predicting

the value of one or more dependent (or response) variables from a set of independent variables and modelling variable relationships (or predictors).

The Double Classification [2] method is a conventional technique to the regression problem with two qualitative regressors, where a model is initially believed to be appropriate and then particular parameter hypotheses are evaluated to compare the effects for the various levels. This method of problem-solving does not include the concept of model selection. The Johnson et al. [3] structures are one of the structures for the interactions that are taken into consideration.

Let x_1 and x_2 be two qualitative variables that can take n_1 and n_2 different categorical values respectively. These categories will be denoted by $i_1, i_2, \dots, i_{\bar{n}_1}$ and $i'_1, i'_2, \dots, i'_{\bar{n}_2}$. Suppose that observations have been made on a quantitative random variable Y at $m = n_1 \times n_2$ points in the set, $\{(i_j, i'_j): 1 \leq j \leq \bar{n}_1, 1 \leq j' \leq \bar{n}_2\}$. Without loss of generality we will assume $n_1 = \bar{n}_1$ and $n_2 = \bar{n}_2$, furthermore at the point (i_j, i'_j) and n_{i_j, i'_j} observations are made. Traditionally, a model such as the following is assumed:

$$E(Y_{i_j, i'_j, k}) = \mu_{i_j, i'_j} = \mu + \alpha_{i_j} + \beta_{i'_j} + \gamma_{i_j, i'_j} \quad (1)$$

Observations are assumed to be uncorrelated. A model like the previous one is over parameterized (it contains the maximum number of parameters, that is, $n_1 + n_2 + (n_1 \times n_2) + 1$ parameters), which implies a reduction in the precision of the estimates, since the variance of the estimators grows as the number of parameters in the model grows. It is known that the least squares estimates in Equation (1) is,

$$\hat{\mu} = \overline{Y \dots} = 1/n \sum_{j=1}^{n_1} \sum_{j'=1}^{n_2} \sum_{k=1}^{n_{i_j, i'_j}} Y_{i_j, i'_j, k} \quad (2)$$

These estimators have interesting properties as they are least squares [4]. Considering different regression equations is equivalent to considering different partitions on the value space of the regressors. In the very simple case of $n_1 = 3$ and $n_2 = 2$ and the interactions equal to zero (additive model) [5]. The equations that define the possible models are:

1. Model with one parameter.
2. Models with two parameters.
3. Models with three parameters
4. Model with four parameters.

The number of parameters in the model grows as the number of classes grows in the partition that the model induces in the value space $x = x_1 \times x_2$. Counting the number of models with p parameters would be equivalent to counting the partitions where the sum of the class numbers of the partitions in the value spaces of x_1 and x_2 equals $p + 1$.

Proposition 1: Let x_1 and x_2 be two qualitative variables with n_1 and n_2 different values ($n_1 \leq n_2$). So there are:

$$\sum_{a=p+1-\text{Min}(n_2,p)}^{\text{Min}(n_1,p)} \{ [\sum_{j=1}^a \frac{(-1)^{a-j}}{j!(a-j)!} j^{n_1}] [\sum_{j=1}^{p+1-a} \frac{(-1)^{p+1-a-j}}{j!(p+1-a-j)!} j^{n_2}] \} \quad (3)$$

Different additive models with p parameters.

Proof: It is known that if x is a qualitative variable with m different categories, the number of models with p parameters (number of partitions with p classes available in literature and cited by Kim, Byungchan [6]. Therefore, the quantities;

$$H_1 = \sum_{j=1}^a \frac{(-1)^{a-j}}{j!(a-j)!} j^{n_1} \quad (4)$$

$$H_2 = \sum_{j=1}^{p+1-a} \frac{(-1)^{p+1-a-j}}{j!(p+1-a-j)!} j^{n_2} \quad (5)$$

Represent the number of partitions with a classes in the value space of x_1 and with $p + 1 - a$ classes in the value space of x_2 ; being its sum equal to $p + 1$. If M is an additive model defined by the equation that induces partitions such as those considered, it will have p parameters. For a fixed, each partition with a classes in the value space of x_1 is combined with each of the H_2 partitions with $p + 1 - a$ classes in the value space of x_2 , that is, for a fixed the number of models with p parameters it will be $H = H_1 \cdot H_2$. But the number of classes of the partitions in the value space of x_1 can be at least equal to $p + 1 - \text{Min}(n_2, p)$ and at most $\text{Min}(n_1, p)$. now adding for all possible values of the formula given in Equation (3) is obtained. The number of additive models will serve as a reference to assess how large the number of possible models is.

The rest of the paper is organized as follows: After this introduction, Section 2 presents a brief overview of qualitative analysis. While Section 3 provides modeling of interaction, validation and variable selection procedures, and evaluation of classification performances. In Section 4, the Regressor Selection is presented. Section 5 discusses the proposed approach to model selection, definition of neighbouring model, and algorithm of proposed approach with two qualitative regressors. Section 6 presents the Graphical analysis through Examples and Simulation results. Finally, Section 7 offers the concluding remarks.

2. Qualitative Analysis

Cowie and McKeown [7] established the Qualitative Analysis (QA) technique for identifying local patterns across numerous assessors in time-continuous assessments. Feel traces were employed in their research [8]. A graphical user interface (GUI) with axes corresponding to distinct emotional qualities is used in the perceptual evaluation. The axes extreme values correlate to those emotional qualities extreme values. An evaluator observes the stimuli, interprets the emotional content, and changes the mouse pointer to match the emotional attribute's perceived degree. The interface records the cursor's location in real time, providing emotional traces.

The QA converts continuous assessments into ordinal matrices that capture relative trends between evaluators. The individual matrices (IMs) are formed in the first stage to record the relative trends in an evaluator's traces. Figure 1 (a) depicts the procedure, in which the traces are first segmented into N bins of equal length (3s in this study). The average value of the trace within each

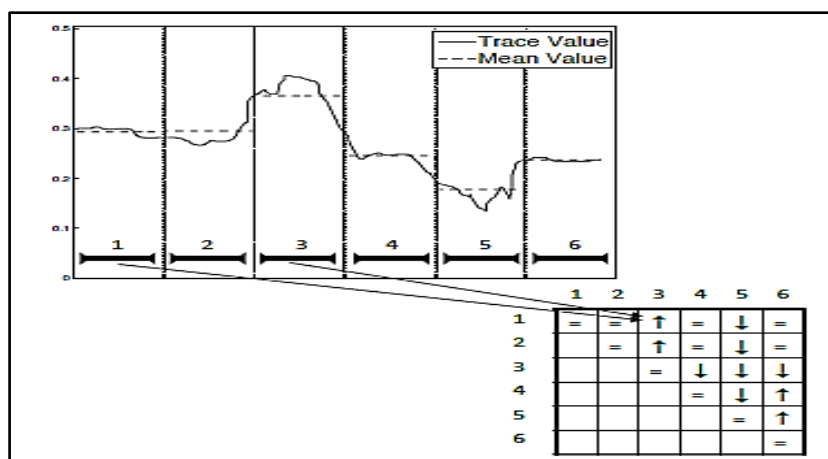
bin is then estimated, and this value is denoted as b_i with $i \in \{1 \dots N\}$. The IMs are created with relative comparisons between the values of the bins. If $i < j$, we define a fall when $b_i - b_j$ is greater than a threshold Equation (6), and a rise when $b_j - b_i$ is greater than a threshold Equation (7). Otherwise, we consider that the bins are similar Equation (8). $t_{threshold}$ is a parameter of the QA method.

$$b_i - b_j > t_{threshold} \quad (6)$$

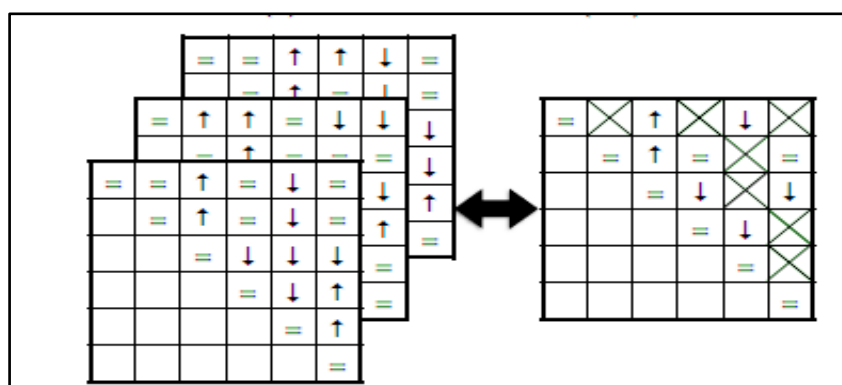
$$b_j - b_i > t_{threshold} \quad (7)$$

$$|b_j - b_i| < t_{threshold} \quad (8)$$

The IMs are then combined to generate a consensus matrix (CM) in the following stage. The technique is depicted in Figure 1 (b), which tries to capture evaluator agreement. If "X%" of the evaluators agree on a trend (entries from various IMs), the CM entry relating to the trend is established (rise, fall or similar). Another aspect of the QA process is the variable X, which is referred to as the tolerance agreement. Segments lacking consensus (X in Figure 1 (b)) are entries that fail to establish an agreement. The QA approach was used by Parthasarathy et al. [9] to create preference learning algorithms to rate emotions.



(a) Individual matrices.



(b) Consensus matrix.

Figure 1: Individual matrices generation from time-continuous traces, as well as consensus matrix generation.

3. Modeling of Interaction

When working with two qualitative regressors and wanting to perform a complete analysis, it is necessary to determine the presence or not of the interaction term and estimate the variance. The usual model with interactions is over-parameterized and as already mentioned this is a big drawback. The consideration of certain forms of interactions proposed by many scientists including Flache et al. [10] produces a decrease in the number of parameters. This consideration is tremendously important when there is only one observation for combinations of treatments, because then the classical theory of linear models cannot be used, since the estimation of the variance of the error has to be obtained from the sum of squares of the interactions.

The different models that are proposed to be considered others are, Additive model; Concurrent model; Regression model per column; Regression model per row; Fifth model; Mandel's model; Seventh, eighth, ninth and tenth models. Subject to restrictions;

$$\begin{aligned}\sum_{j=1}^{n_1} \alpha_{ij} &= \sum_{j'=1}^{n_2} \beta_{i'j'} = \sum_{j=1}^{n_1} \theta_{ij} = \sum_{j'=1}^{n_2} \vartheta_{i'j'} = 0, \\ \sum_{j=1}^{n_1} \theta_{ij}^2 &= \sum_{j'=1}^{n_2} \vartheta_{i'j'}^2 = \sum_{l=1} U_{li_j}^2 = \sum_{l=1} V_{li'j'}^2 = 1, \\ \sum_{j'=1}^{n_2} \gamma_{ij,i'j'}^2 &= \sum_{j'=1}^{n_2} \gamma_{i'j',i'j'} = 0.\end{aligned}$$

The number of parameters in any model listed above will be equal to the number of parameters of the additive model contained in it, plus the number of parameters provided by the interactions.

3.1 Validation and variable selection procedures

Cross-validation techniques are required for classifiers, just as they are for regression models, in order to evaluate their prediction classification skills on unknown objects. Because the modelled response is qualitative rather than quantitative, the prediction ability estimate of classification models is obviously conducted on different parameters than regression approaches. In any instance, a variety of metrics can be employed, such as the percentage of properly identified items relative to the total number of accessible objects or the proportion of correctly classified objects in a particular category of interest. Even if these parameters can be calculated using the same procedures used to validate regression models (single evaluation set, leave-one-out, leave-more-out, repeated training/test splitting, bootstrap), when classification models are validated, the percentage of objects retained in each cross validation group must be taken into account.

Consider a data set containing two classes (A and B), as well as a cross-validation technique in which groups of objects are removed from the training set, one at a time, and the classification model is tested. The validation result will be unsuccessful if the entire class A is removed from the data set during validation (all objects belonging to A are used to test the model); in fact, the model will be built without the removed class's objects (the model will not consider class A) and thus will not

recognise objects belonging to that class. A proper validation approach, on the other hand, should at the very least keep objects from all of the considered classes in each training group.

However, the number of items used to construct a classification model is typically a significant consideration, since a small number of objects cannot adequately reflect all of the elements that influence class variability. If the ratio between the number of objects and the number of variables is high, however, various classification approaches, such as discriminant analysis, can be applied. If the number of objects cannot be increased, variable selection methods can be used to decrease the number of descriptors. In reality, to improve classification performance and pick the most discriminating descriptors, classification approaches can be used with variable selection tools. Even while more advanced techniques, such as genetic algorithms, can (and have) been used, the bulk of classification selection procedures are based on stepwise discriminant analysis or similar schemes. In most stepwise analyses, error percentages are employed as an informal stopping criteria; if a subset of s variables out of p produces a smaller error than the entire set of variables, the s variables can be regarded sufficient for class separation. The classification performance of multiple subgroups of decreasing sizes may then be compared. The Wilks' lambda [11] is a typical approach for picking the optimal subset of variables for splitting groups. It is defined as:

$$\Lambda = \frac{|W|}{|W+B|}, \quad (9)$$

where W and B are the within and between sum of squares, respectively. Wilks' lambda is a number that varies from 0 to 1, with values near to 0 indicating that the group means diverge. As a result, the Wilks' lambda values of the variables with the lowest Wilks' lambda values can be kept in the classification model.

3.2 Evaluation of classification performances

As previously stated, numerous factors may be used to estimate the quality of classification models for both fitting and validation [12]. These factors, of course, are connected to the occurrence of mistakes in the results (Objects assigned to the incorrect classes), even if errors might be weighted differently depending on the classification goals. The confusion matrix, which is a square matrix of dimensions $G \times G$, where G is the number of classes, may be used to calculate all classification indices. Table 1 shows a general depiction of a confusion matrix, with each row n_{gk} denoting the number of objects belonging to class g that are allocated to class k . As a result, the n_{gg} diagonal elements indicate successfully categorised items, whereas the off-diagonal elements represent incorrectly classified things. Because n_{gk} differs from n_{kg} , the confusion matrix is frequently asymmetric, i.e. the number of objects belonging to class g and allocated to class k is not always equal to the number of objects belonging to k and assigned to g .

Table1: General representation of a confusion matrix

		Assigned class				
		1	2	3	...	G
True class	1	n_{11}	n_{12}	n_{13}	...	n_{1G}
	2	n_{21}	n_{22}	n_{23}	...	n_{2G}
	3	n_{31}	n_{32}	n_{33}	...	n_{3G}

	G	n_{G1}	n_{G2}	n_{G3}	...	n_{GG}

We can get an indication of how a classification model is functioning by looking at the confusion matrix (based on fitting or verified results); of course, some more useful indices may be constructed to synthesize this information. The non-error rate (NER) can be defined as follows:

$$\text{NER} = \frac{\sum_{g=1}^G n_{gg}}{n}, \quad (10)$$

where n is the total number of objects. The non-error rate (also known as accuracy or classification rate) is the most basic indicator of a classification model's quality, and it measures the proportion of objects that are properly assigned. The error rate (ER) is the NER complementary index; it is the percentage of incorrectly assigned objects and is defined as:

$$\text{ER} = \frac{n - \sum_{g=1}^G n_{gg}}{n} = 1 - \text{NER}. \quad (11)$$

NER and ER can merely explain a model's performance, but a classification tool's output should be regarded adequate from a statistical standpoint when the classification ability is much more than that produced by random class assignment. As a result, model efficiency may be assessed by contrasting ER with the no-model error rate (NOMER), which is the error rate achieved by allocating all objects to the biggest class and can be computed as follows:

$$\text{NOMER} = \frac{n - n_M}{n}, \quad (12)$$

where n_M is the number of objects belonging to the largest class. On the other hand, the error rate may also be compared to the error produced by assigning people to one of the stated classes at random:

$$\text{Random ER} = \frac{\sum_{g=1}^G \left(\frac{n - n_g}{n}\right) \cdot n_g}{n}, \quad (13)$$

where n_g is the number of objects belonging to the g^{th} class:

$$n_g = \sum_{k=1}^G n_{gk}$$

4. Regressor Selection

Determine which regressors will be used to explain the models output. A regressor is a function of the measured data, such as the system's previous and last inputs and/or outputs. The regressor might be chosen entirely based on measurement data or in conjunction with information gleaned from other sources, such as physical laws. When relevant regressors are located, the tasks of selecting a model type and estimating model parameters become significantly easier. On several of the duties listed above, we'll go through them in further depth. Assume we have a specific measured output signal $y(t)$ and the corresponding input signal $u(t)$. Let Z^N denote the measured input/output data for time $t = 1, \dots, N$.

It has turned out to be useful to describe the relation between Z^N and the output $y(t)$ using a concatenation of two mappings [13]. The first maps the data Z^N of growing dimension into a regression vector $\varphi(t) = \varphi(Z^N)$ of fixed dimension. The elements of $\varphi(t)$ will be denoted regressors. The second mapping, parameterized with θ , maps the regression vector $\varphi(t)$ onto the output $y(t)$;

$$y(t) = g(\varphi(t), \theta). \quad (14)$$

Useful choices of the map $\varphi(Z^N)$ include;

$$\varphi(t) = \begin{bmatrix} y(t-T) \\ y(t-2T) \\ \dots \\ y(t-k_y T) \\ u(t) \\ u(t-T) \\ \dots \\ u(t-k_u T) \end{bmatrix}. \quad (15)$$

Where k_y and k_u are parameters to decide and T is the sampling period. Also, non-linear mappings from the measured data to the regressors, such as polynomials of the regressors in Equation (15) or other sophisticated input and output signal functions, might also be effective. The mapping proposal is frequently aided by system understanding and creativity. The majority of this thesis's talks will focus on possible regressors like Equation (15). Regressor selection is the process of deciding which recommended regressors to include in the system model.

5. Proposed Approach to Model Selection

It is considered that observations $Y_{i_j, i'_j, k}$ are made on a random variable Y , which satisfy the regression equation:

$$Y_{i_j, i'_j, k} = f(i_j, i'_j) + \varepsilon_{i_j, i'_j, k} \quad (16)$$

They are assumed to be $\varepsilon_{i_j, i'_j, k}$ random (unobservable) errors with zero expectation and variance σ^2 and that they are unrelated.

The function $f(i_j, i'_j)$ belongs to a set M defined by:

$$M = \{g(x_1, x_2, \beta) : \beta \in \beta\} \quad (17)$$

With;

$$g(x_1, x_2, \beta) = \sum_r 1_{C_r^g}(x_1, x_2) \mu_{i_j, i'_j} \quad (18)$$

$$\beta^t = (\mu, \alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_{p_1}}, \beta_{i_{r_1}}, \beta_{i'_{r_2}}, \dots, \beta_{i'_{r_2}} : \gamma_1, \gamma_2, \dots, \gamma_{p_3}).$$

Where $\forall \gamma_l = \gamma_{li_j, i'_j}$ for some (i_j, i'_j) and is any of the types of interactions described and is given by one of the expressions from the different models given above.

$$N = \bigcup_r C_r^g$$

So the selection of the model is given by the selection of a function g in M to approximate f . The set of observations will be used to calculate an estimator $\hat{\beta}$ of the vector of parameters β and select a function $g(x_1, x_2, \hat{\beta})$.

In the selection of the function $g(x_1, x_2, \hat{\beta})$, the minimization criterion of mean square of the prediction error (MSEP) or mean squared error (MSE) will be used. The vector of parameters β will be estimated by the Least Squares (LS) Method.

In the first stage of the approach, the models with a number of parameters less than or equal to a number p_0 set by the user appropriately and within the permissible limits will be analyzed. Subsequently, a way must be found to reduce the number of models to be compared and the idea would be to make a reduction in such a way that between one step of the approach and another the analyzed models do not change abruptly in terms of the number of parameters. In achieving this last objective, the concept of the neighbour model that is given below plays a fundamental role.

5.1 Definition of neighbouring model

Let $\delta \in \{\alpha; \beta; \gamma(x_1, x_2)\}$, let M_0 be a model, and let π_0 be the partition determined by the function $g_0(x_1, x_2, \beta)$ that defines the model M_0 . sean C_{ok} the classes of this partition [14]. It is said that it is a neighbour model of M_V^δ . M_0 According to δ , if it is true:

1. There is one and only one δ and one and only one class C_{ok} such that:

$$C_{ok} = (C_{ok'})^\delta \cup (C_{vk''})^\delta,$$

where the superscript δ has been used to indicate that the class is affected only according to that parameter.

2. For all $r \neq k$ there exists r' , such that $k' \neq r' \neq k''$, for which it is true:

$$Cor = (Cvr)\delta$$

3. Furthermore, the form of the interaction of the model is the same as that of the M_V^δ . M_O Model.

From the way the neighbour model has been defined, it is intuitively clear that it has one more parameter than M_O . In a problem with two qualitative variables, there are three parameter names, which have been represented as follows:

- $\alpha \rightarrow$ levels of the first variable.
- $\beta \rightarrow$ levels of the second variable.
- $\gamma \rightarrow$ Interactions.

To build a selection approach, based on neighbouring models, it is necessary to decide in what order the nominations are taken to increase the number of parameters in the model. Since there is no preference between one factor and another, this order is irrelevant. As you have to decide on an order, it is proposed: $\alpha \rightarrow \beta \rightarrow \gamma$, for more details see [15].

5.2 Algorithm of proposed approach with two qualitative regressors

Let $\gamma_1(x_1, x_2)$ be the form of interaction in the model $M_l(\cdot)$, $l = 1, 2, \dots, 10$. Let $p(l)$ be the maximum number of allowable parameters for the model $M_1(\cdot)$. Now we let;

$$M_l^0 = \{(M_l(p) : p \leq p_0; 1 \leq p_0 \leq p(l)\}, \quad (19)$$

Let;

$$\hat{r}(M_l(p)) = \hat{r}(l, p(l)).$$

An estimate of the MSEP for the $M_l(p)$ model;

- Calculate the number of possible additive models for each $p = 1, 2, \dots, p(l)$. Being l a fixed number.
- According to what is determined in the previous point, the computing facilities and the time and effort that one is willing to use, select a value p_0 and thus the set that M_l^0 we will call the set of basic models will be determined.

In what follows, the following notation will be used, for example $(M_l^0)_v^\alpha$ is a neighbour model of according to the partitions M_l^0 corresponding to the parameter α . That is, a parameter has been increased with respect to the number it contained M_l^0 but the increase is made in the denomination α . The class formed by the neighbours of according to the parameter α , will be denoted by $M_l^{v,\alpha}$.

Perform the following algorithm steps:

1. Determine;

$$M_l^0 = \frac{Avg \ Min \ \hat{r}(l, p)}{M \in M_l^0}.$$

Let;

$$M_l^0 = M_l^0(p').$$

where p' represents the number of parameters in M_l^0

2. Let;

$$M_l^{v,\alpha} = \{(M_l^0)_v^\alpha\}.$$

Determine;

$$M_l^{1,\alpha} = \frac{\text{Avg Min } \hat{r}(l, p' + 1)}{M \in M_l^{v,\alpha}}.$$

Then we have,

$$M_l^{1,\alpha} = M_l^{1,\alpha}(p' + 1).$$

3. Let;

$$M_l^{v,\beta} = \{(M_l^{1,\alpha})_v^\beta\}.$$

Determine;

$$M_l^{2,\beta} = \frac{\text{Avg Min } \hat{r}(l, p' + 2)}{M \in M_l^{v,\beta}}.$$

4. Let;

$$M_l^{v,\gamma(x_1,x_2)} = \{(M_l^{2,\beta})_v^{\gamma(x_1,x_2)}\}.$$

Determine;

$$M_l^{3,\gamma(x_1,x_2)} = \frac{\text{Avg Min } \hat{r}(l, p' + 3)}{M \in M_l^{v,\gamma(x_1,x_2)}}.$$

5. Let;

$$M_l^{v,\alpha} = \{(M_l^{3,\gamma(x_1,x_2)})_v^\alpha\}.$$

Determine;

$$M_l^{4,\alpha} = \frac{\text{Avg Min } \hat{r}(l, p' + 4)}{M \in M_l^{v,\alpha}}.$$

6. Let;

$$M_l^{v,\beta} = \{(M_l^{4,\alpha})_v^\beta\}.$$

Determine;

$$M_l^{5,\alpha} = \frac{\text{Avg Min } \hat{r}(l, p' + 5)}{M \in M_l^{v,\alpha}}.$$

7. Let;

$$M_l^{v,\gamma(x_1,x_2)} = \{(M_l^{5,\alpha})_v^{\gamma(x_1,x_2)}\}.$$

Determine;

$$M_l^{6,\gamma(x_1,x_2)} = \frac{\text{Avg Min } \hat{r}(l, p' + 6)}{M \in M_l^{v,\gamma(x_1,x_2)}}.$$

8. Repeat steps 5, 6 and 7. In this way, for $q \in \{z: 0 \leq z\}$, steps $2 + 3q$, $3 + 3q$ and $4 + 3q$, would be determined by:

⋮
⋮

$2 + 3q$.-

$$M_l^{v,\alpha} = \{(M_l^{3q,\gamma(x_1,x_2)})_v^\alpha\}.$$

Determine;

$$M_l^{1+3q,v} = \frac{\text{Avg Min } \hat{r}(l, p' + 1 + 3q)}{M \in M_l^{v,\alpha}}.$$

Let; $3 + 3q$.-

That is, the median of the l - m -th residue is a function of α_l . What $\sum_{l=1}^3 \alpha_l = 0$. It has to;

$$\begin{aligned} \text{Med } r_{lm} &= \alpha_2 k_m - \alpha_3 k_m, \\ &= -\text{Med } r_{2m} - \text{Med } r_{3m}. \end{aligned}$$

A graph of residuals showing these against the $\hat{\alpha}_l$ against its indices, it can provide invaluable information to guide subsequent explorations.

Some graphical illustrations are given in Figure 2 (graph 1-4) a graph like number 2 does not suggest that relationship and, therefore, no changes in that sense. While the graphs numbers 1, 3 and 4, if they suggest changes, and then the model should be explored by making said change in the interaction modelling.

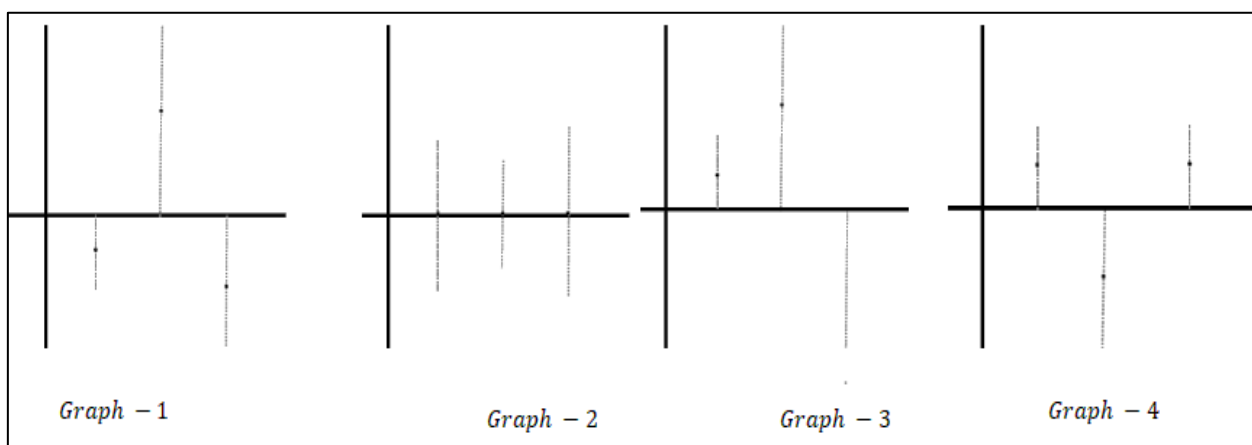


Figure 2: The Correct Model in Case 1.

Case 2: Consider the situation of case 1 but now the true model is the one given in b of column 3 in Equation (17). The result is similar but the dependency arises with respect to β_j ;

$$\text{Med } r_{i,j} = k_i \beta_j.$$

The residuals should be plotted against or against their indices $\hat{\beta}_j$. These graphs can be useful in almost all situations except the one where the model given in column 5 in Equation (17) is involved.

6.2 Simulation Results

All the models referred to above except 1 and 10; they are non-linear; therefore, it seems reasonable to reject the hypothesis of constant variance. This has been considered in a programming language made in Turbo Pascal “version 7.0”. Tables 2, 3 and 4 summarizes the twenty data sets were simulated with two models (ten with each) and it was obtained that the approach selected the optimal model in all cases; this is also shown in Figure 3. Moreover, if the data contains outlier values, conducting a meaningful statistical inference would be difficult, see e.g. [18, 19, 20, 21] for handling and solving this problem in several models.

Table 2: Model $\rightarrow E(Y_{i,j,i',j'}) = \mu + \alpha_{ij} + \beta_{i'j'} + \lambda\alpha_{ij}\beta_{i'j'}$

MSEP	Selected the optimal model	MSEP of the optimal model
4.535	x	-
4.291	-	6.292
8.114	x	-
2.343	-	3.715
6.091	x	-
4.310	-	9.184
6.653	-	7.083
3.073	x	-
10.344	x	-
5.416	-	2.952

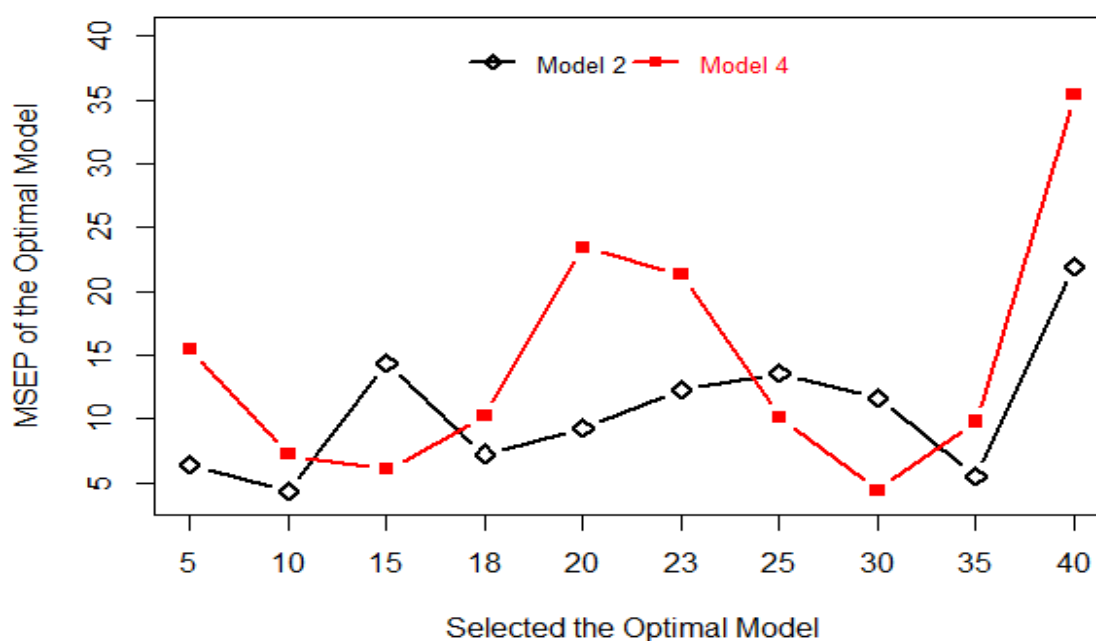


Figure 3: Approach to select the optimal model

Table 3: Model $\rightarrow E(Y_{i,j,i',j'}) = \mu + \alpha_{ij} + \beta_{i'j'} + c_i\beta_{i'j'}$

MSEP	Selected the optimal model	MSEP of the optimal model
13.605	-	10.932
12.874	x	-
24.343	-	5.824
7.028	x	-
18.272	x	-
12.931	-	20.291
19.958	x	-
9.220	x	-
31.033	-	23.681
16.248	x	-

Table 4: Model $\rightarrow E(Y_{i,j,i',j'}) = \mu + \alpha_{ij} + \beta_{i'j'} + c_i\beta_{i'j'}$

MSEP	Selected the optimal model	MSEP of the optimal model
35.374	x	-
33.473	-	31.723
43.872	-	18.801
18.272	x	-
47.506	-	53.271
33.620	x	-
51.891	-	49.061
63.292	x	-
70.872	-	10.932
80.687	x	-

7. Concluding Remarks

In this paper, we propose a new two-qualitative regressors approach to regression model selection, within a predefined class of models, which will be utilized to develop a method for selecting qualitative variables the approach is based on the minimization of an estimate of the MSEP. Some interaction structures are considered, since two models were used to simulate several data sets, and the technique was found to identify the optimal model in all cases. When compared to other estimates, the results suggest that significant improvements in bias and efficiency can be made. Some graphs are proposed to diagnose the form of the interaction. The proposed approach achieves better accuracy with reduced MSEP, improved stability, and faster convergence compared to other models.

Acknowledgment

The authors wish to thank the editor. We also thank anonymous for their encouragement and support. The authors would like to thank the reviewers for their valuable comments and suggestions which improved the paper.

Disclosure Statement

No potential conflict of interest was reported by the authors.

8. References

1. Browne, J. (2016). Inspiration to Perspiration: Francis Galton's Hereditary Genius in Victorian Context. In *Genealogies of genius* (pp. 77-95). Palgrave Macmillan, New York.
2. Ballabio, D., & Todeschini, R. (2009). Multivariate classification for qualitative analysis. *Infrared spectroscopy for food quality analysis and control*, 83, e102.
3. Johnson, D. E., & Graybill, F. A. (1972). An analysis of a two-way model with interaction and no replication. *Journal of the American Statistical Association*, 67(340), 862-868.

4. Snee, R. D., & Rayner, A. A. (1982). Assessing the accuracy of mixture model regression calculations. *Journal of Quality Technology*, 14(2), 67-79.
5. Uyank, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia-Social and Behavioral Sciences*, 106, 234-240.
6. Kim, B. (2012). On the number of partitions of n into k different parts. *Journal of Number Theory*, 132(6), 1306-1313.
7. Cowie, R., & McKeown, G. (2010). Statistical analysis of data from initial labelled database and recommendations for an economical coding scheme. *SEMAINE Report D6b*.
8. Pereira, D. G., Rodrigues, P. C., Mejza, S., & Mexia, J. T. (2012). A comparison between joint regression analysis and the AMMI model: a case study with barley. *Journal of Statistical Computation and Simulation*, 82(2), 193-207.
9. Parthasarathy, S., Cowie, R., & Busso, C. (2016). Using agreement on direction of change to build rank-based emotion classifiers. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(11), 2108-2121.
10. Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., & Lorenz, J. (2017). Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*, 20(4).
11. Belitz, C., & Lang, S. (2008). Simultaneous selection of variables and smoothing parameters in structured additive regression models. *Computational Statistics & Data Analysis*, 53(1), 61-81.
12. Alharbi, A. A., Kamel, A.R., and Atia, S. A. (2022). A New Robust Modeling of Heat and Mass Transfer Process in MHD Based on Adaptive-Network-Based Fuzzy Inference System. *WSEAS Transactions on Heat and Mass Transfer*, 17, 80-96.
13. Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., & Schröder, M. (2000). 'FEELTRACE': An instrument for recording perceived emotion in real time. In *ISCA tutorial and research workshop (ITRW) on speech and emotion*.
14. Kamel, A. R., & Alqarni, A. A. (2020). A new characterization of exponential distribution through minimum chi-squared divergence principle. *Journal of Advanced Research in Applied Mathematics and Statistics*, 5(1&2), 14-26.
15. AL-Tayeb, L. E. (2022) A Strategy for the Selection of Regression Models with Two Qualitative Regressors, *Scientific Journal for Financial and Commercial Studies and Research*, Faculty of Commerce, Damietta University, 3(2)1, 657 - 672.
16. Kamel, A. R., Alqarni, A. A., & Ahmed, M. A. (2022). On the Performance Robustness of Artificial Neural Network Approaches and Gumbel Extreme Value Distribution for Prediction of Wind Speed. *Int. J. Sci. Res. in Mathematical and Statistical Sciences* Vol, 9(4), 5 - 22.

17. Alharbi, Y. S., & Kamel, A. (2022). Fuzzy System Reliability Analysis for Kumaraswamy Distribution: Bayesian and Non-Bayesian Estimation with Simulation and an Application on Cancer Data Set. *WSEAS Transactions on Biology and Biomedicine*, 19, 118-139.
18. Abonazel, M., & Rabie, A. (2019). The impact of using robust estimations in regression models: An application on the Egyptian economy. *Journal of Advanced Research in Applied Mathematics and Statistics*, 4(2), 8-16.
19. Youssef, A. H., Kamel, A. R., & Abonazel, M. R. (2021). Robust SURE estimates of profitability in the Egyptian insurance market. *Statistical journal of the IAOS*, 37(4), 1275-1287. <https://doi.org/10.3233/SJI-200734>.
20. Youssef, A. H., Abonazel, M. R., & Kamel, A. R. (2022). Efficiency Comparisons of Robust and Non-Robust Estimators for Seemingly Unrelated Regressions Model. *WSEAS Transactions on Mathematics*, 21, pp 218-244.
21. Kamel, A. R. (2021). *Handling outliers in seemingly unrelated regression equations model*, MSc thesis, Faculty of graduate studies for statistical research (FGSSR), Cairo University, Egypt.



© 2022 the Author(s), licensee the scientific association for studies and applied research (SASAR). This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>).