International Journal of Economic Research

ISSN : 0972-9380

available at http: www.serialsjournal.com

Volume 14 • Number 8 • 2017

Clustering on The Basis of Regression Equations: The General Case

Jagabandhu Saha[1]

[1]*Department of Economics and Politics, Visva-Bharati University, India. E-mail: saha_jagabandhu@yahoo.com*

Abstract: While Chow test is applicable for testing the equality between sets of coefficients in two linear regressions, this paper attempts to construct a test procedure not only to compare the equality between the sets of coefficients in two linear regressions, but also, in case they are not equal, to provide detailed informations about the inequality of the sets. Not only that, in this paper attempt is also made to accomplish all these for not just two linear regressions but for the two linear regressions of all possible pairs of linear regressions out of any number of given linear regressions, and then the results of all these comparisons are used in order to form clusters among the regressions on the basis of some principle stated therein. The procedure is then illustrated through comparison of Engel curves on food for different Socioeconomic groups of Rural India, using NSSO data.

1. INTRODUCTION

It is a common practice to test the equality between sets of coefficients in two linear regressions by Chow Test (Chow 1960) [1]. There is no problem if in the Chow Test the null hypothesis of equality between the sets of coefficients is not rejected (as in the examples in his paper [1]). But if rejected, then, naturally, one is probed to the questions:

(a) at which component/s the sets differ, and

(b) for each of these components, between the two coefficients of the two regressions concerned, which one is larger/smaller.

Chow test, however, does not provide any answer to these questions at all. These problems can be resolved with some modifications of the model (Saha and Pal 2014) [2]. Saha and Pal introduced the concept of "component wise complete comparison" (CCC)[2] in order to overcome this problem. The test procedure for CCC between every two successive regressions out of any number of given successive regressions was developed. Now, this paper attempts to construct a test procedure for CCC between not only every two successive regressions out of any number of given successive regressions, but between the two linear regressions of *all possible pairs of regressions out of any number of given regressions*, and then uses all these comparisons in order to form clusters among the regressions on the basis of this simple principle: *all those regressions which satisfy the condition that the vectors of coefficients of any two of these regressions do not differ from each other significantly will form a cluster.* The rest of the paper can be outlined as follows. In Section 2, we put the problem in the formal terms, problem of finding test procedure for CCC between the two regressions of all possible pairs of regressions out of several given regressions. Section 3 is devoted for the methodology for solving this problem and then for discussion about process of clustering. In Section 4 we consider a numerical example in order to illustrate the methodology and the process of clustering while in Section 5 we present our conclusions.

2. THE MODEL

We consider the problem of finding test procedure for CCC between the two regressions of all possible pairs of regressions out of m given regressions as follows:

$$y^{(1)} = a_1^{(1)} + a_2^{(1)}x_2^{(1)} + a_3^{(1)}x_3^{(1)} + ... + a_k^{(1)}\, x_k^{(1)} + u^{(1)},$$

$$y^{(2)} = a_1^{(2)} + a_2^{(2)}x_2^{(2)} + a_3^{(2)}x_3^{(2)} + ... + a_k^{(2)}\, x_k^{(2)} + u^{(2)},$$

....

$$y^{(m)} = a_1^{(m)} + a_2^{(m)}\, x_2^{(m)} + a_3^{(m)}x_3^{(m)} + ... + a_k^{(m)}x_k^{(m)} + u^{(m)}, \qquad ...(1)$$

where, n_1, n_2, ..., n_m are the nos. of observations for these regressions. We, assume as in Chow test that the components of each of $u^{(1)}$, $u^{(2)}$, ..., $u^{(m)}$ are iid $N(0, \sigma^2)$. Also the vectors $u^{(1)}$, $u^{(2)}$, ..., $u^{(m)}$ are mutually independent.

Now, we proceed as follows. Firstly, for CCC between first and second regressions in (1), first and third regressions, ..., first and *m*-th regressions, one requires to decide whether the differentials:

$$c_j^{12} = a_j^2 - a_j^1 < 0 \text{ or } = 0 \text{ or } > 0, \text{ for all } j = 1, 2, ..., k,$$

$$c_j^{13} = a_j^3 - a_j^1 < 0 \text{ or } = 0 \text{ or } > 0, \text{ for all } j = 1, 2, ..., k,$$

....

$$c_j^{1m} = a_j^m - a_j^1 < 0 \text{ or } = 0 \text{ or } > 0, \text{ for all } j = 1, 2, ..., k.$$

Similarly, for second and third regressions, second and fourth regressions, ..., second and *m*-th regressions, one requires to decide whether the differentials:

$$c_j^{23} = a_j^3 - a_j^2 < 0 \text{ or } = 0 \text{ or } > 0, \text{ for all } j = 1, 2, ..., k,$$

$$c_j^{24} = a_j^4 - a_j^2 < 0 \text{ or } = 0 \text{ or } > 0, \text{ for all } j = 1, 2, ..., k,$$

....

$$c_j^{2m} = a_j^m - a_j^2 < 0 \text{ or } = 0 \text{ or } > 0, \text{ for all } j = 1, 2, ..., k.$$

....

Lastly, for $(m - 1)$-th and *m*-th regressions, one requires to decide whether the differentials:

$$c_j^{m-1,m} = a_j^m - a_j^{m-1} < 0 \text{ or } = 0 \text{ or } > 0, \text{ for all } j = 1, 2, ..., k.$$

A moment's reflection shows that the desired CCC, *i.e.*, CCC for all pairs of regressions out of the m regressions in (1), will be over when all the decisions enlisted just above, in $(m-1)$ groups, so to say, are completed.

3. THE METHODOLOGY

As indicated at the end of the last Section, Methodology consists of $(m-1)$ steps as follows:

1. combine the m regressions in (1) into a single regression equation model; modify the combined model in such way that the differentials c_j^{12}, c_j^{13}, ..., c_j^{1m}, for all $j = 1, 2, ..., k$, appear as regression coefficients in the modified model, and then, run regression with the modified model and perform tests on the regression coefficients of this model (particularly, on the differentials concerned) and decide for each of these differentials whether it is : < 0 or = 0 or > 0. This will complete CCC for $(m-1)$ pairs of regressions: first and second, first and third, ..., first and m-th.
2. do exactly similar as in the step 1. but with the last $(m-1)$ regressions in (1). This will cover CCC for $(m-2)$ pairs of regressions: second and third, second and fourth, ..., second and m-th.

   ....

   $m-1$). Lastly, do exactly similar as in the step 1. But with only the last two regressions in (1). This will cover CCC for one pair (the last pair) of regressions: $(m-1)$-th and m-th.

Thus CCC for all possible pairs of regressions is complete.

Let us work out these steps.

1. We combine the regressions in (1) as follows:

$$\begin{pmatrix} y_1^{(1)} \\ \vdots \\ y_{n_1}^{(1)} \\ y_1^{(2)} \\ \vdots \\ y_{n_2}^{(2)} \\ \vdots \\ y_1^{(m)} \\ \vdots \\ y_{nm}^{(m)} \end{pmatrix} \begin{pmatrix} 1 & x_{21}^{(1)} & \cdots & x_{k_1}^{(1)} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{2,n_1}^{(1)} & \cdots & x_{k,n_1}^{(1)} & 0 & 0 & \cdots & 0 & \cdots & 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & x_{21}^{(2)} & \cdots & x_{k_1}^{(2)} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & x_{2,n_2}^{(2)} & \cdots & x_{k,n_2}^{(2)} & \cdots & 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & x_{21}^{(m)} & \cdots & x_{k_1}^{(m)} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & x_{2,nm}^{(m)} & \cdots & x_{k,nm}^{(m)} \end{pmatrix} \begin{pmatrix} a_1^{(1)} \\ \vdots \\ a_k^{(1)} \\ a_1^{(2)} \\ \vdots \\ a_k^{(2)} \\ \vdots \\ a_k^{(m)} \\ \vdots \\ a_k^{(m)} \end{pmatrix} + \begin{pmatrix} u_1^{(1)} \\ \vdots \\ u_{n_1}^{(1)} \\ u_1^{(2)} \\ \vdots \\ u_{n_2}^{(2)} \\ \vdots \\ u_1^{(m)} \\ \vdots \\ u_{nm}^{(m)} \end{pmatrix} \quad ...(2)$$

where, in (2) and subsequently, $n_i = n_i$, for all $i = 1, 2, ..., m$.

To modify the model (2) in order to get the differentials $c_j^{12}, c_j^{13}, ..., c_j^{1m}$, for all $j = 1, 2, ..., k$, as regression coefficients, we rewrite it as:

$$\begin{pmatrix} y_1^{(1)} \\ \vdots \\ y_{n_1}^{(1)} \\ y_1^{(2)} \\ \vdots \\ y_{n_2}^{(2)} \\ \vdots \\ y_1^{(m)} \\ \vdots \\ y_{nm}^{(m)} \end{pmatrix} \begin{pmatrix} 1 & x_{21}^{(1)} & \cdots & x_{k_1}^{(1)} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{2,n_1}^{(1)} & \cdots & x_{k,n_1}^{(1)} & 0 & 0 & \cdots & 0 & \cdots & 0 & 1 & \cdots & 0 \\ 1 & x_{21}^{(2)} & \cdots & x_{k_1}^{(2)} & 1 & x_{21}^{(2)} & \cdots & x_{k_1}^{(2)} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{2,n_2}^{(2)} & \cdots & x_{k,n_2}^{(2)} & 1 & x_{2,n_2}^{(2)} & \cdots & x_{k,n_2}^{(2)} & \cdots & 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \cdots & \vdots \\ 1 & x_{21}^{(m)} & \cdots & x_{k_1}^{(m)} & 0 & 0 & \cdots & 0 & \cdots & 1 & x_{21}^{(m)} & \cdots & x_{k_1}^{(m)} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{2,nm}^{(m)} & \cdots & x_{k,nm}^{(m)} & 0 & 0 & \cdots & 0 & \cdots & 1 & x_{2,nm}^{(m)} & \cdots & x_{k,nm}^{(m)} \end{pmatrix} \begin{pmatrix} 2_1^{(1)} \\ \vdots \\ 2_k^{(1)} \\ c_1^{12} \\ \vdots \\ c_k^{12} \\ \vdots \\ c_1^{1m} \\ \vdots \\ c_k^{1m} \end{pmatrix} + \begin{pmatrix} u_1^{(1)} \\ \vdots \\ u_{n_1}^{(1)} \\ u_1^{(2)} \\ \vdots \\ u_{n_2}^{(2)} \\ \vdots \\ u_1^{(m)} \\ \vdots \\ u_{nm}^{(m)} \end{pmatrix} \quad ...(3)$$

Now, we run regression with (3) and carry out tests as prescribed above.

2. We combine the last (m-1) regressions in (1) as follows:

$$\begin{pmatrix} y_1^{(2)} \\ \vdots \\ y_{n_2}^{(2)} \\ y_1^{(3)} \\ \vdots \\ y_{n_3}^{(3)} \\ \vdots \\ y_1^{(m)} \\ \vdots \\ y_{nm}^{(m)} \end{pmatrix} \begin{pmatrix} 1 & x_{21}^{(2)} & \cdots & x_{k_1}^{(2)} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{2,n_2}^{(2)} & \cdots & x_{k,n_2}^{(2)} & 0 & 0 & \cdots & 0 & \cdots & 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 & 1 & x_{21}^{(3)} & \cdots & x_{k_1}^{(3)} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & \cdots & 0 & 1 & x_{2,n_3}^{(3)} & \cdots & x_{k,n_3}^{(3)} & \cdots & 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \cdots & \vdots \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & x_{21}^{(m)} & \cdots & x_{k_1}^{(m)} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & x_{2,nm}^{(m)} & \cdots & x_{k,nm}^{(m)} \end{pmatrix} \begin{pmatrix} a_1^{(2)} \\ \vdots \\ a_k^{(2)} \\ a_1^{(3)} \\ \vdots \\ a_k^{(3)} \\ \vdots \\ a_1^{(m)} \\ \vdots \\ a_k^{(m)} \end{pmatrix} + \begin{pmatrix} u_1^{(2)} \\ \vdots \\ u_{n_2}^{(2)} \\ u_1^{(3)} \\ \vdots \\ u_{n_3}^{(3)} \\ \vdots \\ u_1^{(m)} \\ \vdots \\ u_{nm}^{(m)} \end{pmatrix} \quad ...(4)$$

To modify the model (4) in order to get the differentials $c_j^{23}, c_j^{24}, ..., c_j^{2m}$, for all $j = 1, 2, ..., k$, as regression coefficients, we rewrite it as:

$$
\begin{pmatrix} y_1^{(2)} \\ \vdots \\ y_{n_2}^{(2)} \\ y_1^{(3)} \\ \vdots \\ y_{n_3}^{(3)} \\ \vdots \\ y_1^{(m)} \\ \vdots \\ y_{nm}^{(m)} \end{pmatrix}
\begin{pmatrix}
1 & x_{21}^{(2)} & \cdots & x_{k_1}^{(2)} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\
1 & x_{2,n_2}^{(2)} & \cdots & x_{k,n_2}^{(2)} & 0 & 0 & \cdots & 0 & \cdots & 0 & 1 & \cdots & 0 \\
1 & x_{21}^{(3)} & \cdots & x_{k_1}^{(3)} & 1 & x_{21}^{(3)} & \cdots & x_{k_1}^{(3)} & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\
1 & x_{2,n_3}^{(3)} & \cdots & x_{k,n_3}^{(3)} & 1 & x_{2,n_3}^{(3)} & \cdots & x_{k,n_3}^{(3)} & \cdots & 0 & 0 & \cdots & \vdots \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \cdots & \vdots \\
1 & x_{21}^{(m)} & \cdots & x_{k_1}^{(m)} & 0 & 0 & \cdots & 0 & \cdots & 1 & x_{21}^{(m)} & \cdots & x_{k_1}^{(m)} \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\
1 & x_{2,nm}^{(m)} & \cdots & x_{k,nm}^{(m)} & 0 & 0 & \cdots & 0 & \cdots & 1 & x_{2,nm}^{(m)} & \cdots & x_{k,nm}^{(m)}
\end{pmatrix}
\begin{pmatrix} a_1^{(2)} \\ \vdots \\ a_k^{(2)} \\ c_1^{23} \\ \vdots \\ c_k^{23} \\ \vdots \\ c_1^{2m} \\ \vdots \\ c_k^{2m} \end{pmatrix}
+
\begin{pmatrix} u_1^{(2)} \\ \vdots \\ u_{n_2}^{(2)} \\ u_1^{(3)} \\ \vdots \\ u_{n_3}^{(3)} \\ \vdots \\ u_1^{(m)} \\ \vdots \\ u_{nm}^{(m)} \end{pmatrix}
\quad \ldots(5)
$$

Now, we run regression with (5) and carry out tests as prescribed above.

....

m – 1). We combine the last two regressions in (1) as follows:

$$
\begin{pmatrix} y_1^{(m-1)} \\ \vdots \\ y_{n,m-1}^{(m-1)} \\ y_1^{(m)} \\ \vdots \\ y_{nm}^{(m)} \end{pmatrix}
\begin{pmatrix}
1 & x_{21}^{(m-1)} & \cdots & x_{k_1}^{(m-1)} & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\
1 & x_{2,(n,m-1)}^{(m-1)} & \cdots & x_{k,(n,m-1)}^{(m-1)} & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 1 & x_{21}^{(m)} & \cdots & x_{k_1}^{(m)} \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\
0 & 0 & \cdots & 0 & 1 & x_{2,nm}^{(m)} & \cdots & x_{k,nm}^{(m)}
\end{pmatrix}
\begin{pmatrix} a_1^{(m-1)} \\ \vdots \\ a_k^{(m-1)} \\ a_1^{(m)} \\ \vdots \\ a_k^{(m)} \end{pmatrix}
+
\begin{pmatrix} u_1^{(m-1)} \\ \vdots \\ u_{n,m-1}^{(m-1)} \\ u_1^{(m)} \\ \vdots \\ u_{nm}^{(m)} \end{pmatrix}
\quad \ldots(6)
$$

In order to get the differentials $c_j^{m-1,m}$, for all j = 1, 2, ..., k, we rewrite (6) as:

$$
\begin{pmatrix} y_1^{(m-1)} \\ \vdots \\ y_{n,m-1}^{(m-1)} \\ y_1^{(m)} \\ \vdots \\ y_{nm}^{(m)} \end{pmatrix}
\begin{pmatrix}
1 & x_{21}^{(m-1)} & \cdots & x_{k_1}^{(m-1)} & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\
1 & x_{2,(n,m-1)}^{(m-1)} & \cdots & x_{k,(n,m-1)}^{(m-1)} & 0 & 0 & \cdots & 0 \\
1 & x_{21}^{(m)} & \cdots & x_{k_1}^{(m)} & 1 & x_{21}^{(m)} & \cdots & x_{k_1}^{(m)} \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\
1 & x_{2,nm}^{(m)} & \cdots & x_{k,nm}^{(m)} & 1 & x_{2,nm}^{(m)} & \cdots & x_{k,nm}^{(m)}
\end{pmatrix}
\begin{pmatrix} a_1^{(m-1)} \\ \vdots \\ a_k^{(m-1)} \\ c_1^{m-1,m} \\ \vdots \\ a_k^{m-1,m} \end{pmatrix}
+
\begin{pmatrix} u_1^{(m-1)} \\ \vdots \\ u_{n,m-1}^{(m-1)} \\ u_1^{(m)} \\ \vdots \\ u_{nm}^{(m)} \end{pmatrix}
\quad \ldots(7)
$$

Now, we run regression with (7) and carry out tests as prescribed above.

Thus CCC between the two regressions of all possible pairs of regressions out of the m given regressions in (1) are completed.

Having done this one can partition the set of the given regressions into, say, clusters, *every cluster consisting of those regressions which satisfy the condition that the vectors of coefficients of any two of these* *regressions do not differ from each other significantly*. In order to sort out the clusters easily we first present the outcomes of all the above comparisons neatly by introducing a matrix, let it be called *Indicator Matrix* (IM), as:

$$i_{m \times m} = (r_{ij}), \qquad ...(8)$$

where, r_{ij} indicates whether the ith and the jth regressions in (1) differ from each other significantly or not and is defined as follows:

$r_{ij} = 0$, when $i = j$,

0, when $i \neq j$ and the two regressions concerned do not differ from each another significantly,

1, when $i \neq j$ and the two regressions concerned differ from each another significantly,

where, $i, j = 1, 2, ..., m^3$.

It is this IM which will clearly show the clusters: no. of clusters as well as the regressions in each of these clusters. Needless to say that IM is a symmetric matrix with all the diagonal elements as zeros and it will be a null matrix (of order mxm) *iff* all the *m* regressions coincide.

4. ILLUSTRATION

In the context of Engel curve for food, for each of the Socioeconomic groups ST, SC, OBC, OTHERS, ALL, considering regression of "proportion of Monthly Per Capita Expenditure on Food (*p*MPCEF)", on "Monthly Per Capita Expenditure (MPCE)", *i.e.*, regression of pMPCEF on MPCE, totally five regression equations are constructed as follows ($m = 5$). State level data on Rural India on MPCE and MPCEF for the different Socioeconomic groups mentioned are used, the source of data being NSSO Report [3].

Define :

$x^{(i)}$ = MPCE of a State for the i^{th} Socioeconomic group, for all $i = 1, 2, 3, 4, 5$.

$y^{(i)}$ = *p*MPCEF of a State for the ith Socioeconomic group, for all $i = 1, 2, 3., 4, 5$.

Then, the regressions considered are as follows:

$$y^{(i)} = a_1^{(i)} + a_2^{(i)} x_2^{(i)} + u^{(i)}, \qquad ...(9)$$

for all $i = 1, 2, 3, 4, 5$,

and the task is to perform CCC between the two regressions of all possible pairs of regressions out of these five regressions. (The no. of observations for each regression is thirty (no. of States/UTs in India)[4] and, so, we have: $m = 5$, $k = 2$, $n = 30$.)

According to the Methodology described above, we complete the four steps as follows.

1. Firstly, we run the regression (3). The decisions on the basis of the regression results come out as follows:

$$c_1^{12} < 0,\ c_2^{12} > 0,\ c_1^{13} = 0,\ c_2^{13} = 0,\ c_1^{14} = 0,\ c_2^{14} = 0,\ c_1^{15} = 0,\ c_2^{15} = 0.$$

This in turn means that as long as the relationship of MPCE with *p*MPCEF is concerned, the Socioeconomic group ST differs from SC significantly (at both intercept and slope coefficients) but not from the other groups.

2. Secondly, we run the regression similar as in Step 1). but with the *last four regressions in (9)*. The decisions are:

$$c_1^{23} > 0,\ c_2^{23} < 0,\ c_1^{24} > 0,\ c_2^{24} < 0,\ c_1^{25} > 0,\ c_2^{25} < 0.$$

This, in terms of the relationship concerned, means that the Socioeconomic group SC differs from each of OBC, OTHERS and ALL significantly (at both intercept and slope coefficients).

This is quite expected because ST differs from SC significantly but not from OBC, OTHERS and ALL. So, we can say that SC should differ significantly from all other groups.

3. Thirdly, we run the regression similar as in Step 1). but with the *last three regressions in* (9).

The decisions are:

$$c_1^{34} = 0,\ c_2^{34} = 0,\ c_1^{35} = 0,\ c_2^{35} = 0.$$

This in turn means that the Socioeconomic group OBC does not differ from OTHERS and ALL significantly which also is expected from the behaviour of ST itself.

4. Lastly, we run the regression similar as in Step 1). but with the *last two regressions in (9)*. The decisions are:

$$c_1^{45} = 0,\ c_2^{45} = 0.$$

This in turn means that the Socioeconomic group OTHERS does not differ from ALL significantly which also is expected from the behaviour of ST itself.

Thus CCC between the two regressions of all possible pairs of regressions out of these five regressions are completed.

Now, in order to get the clusters in the present context, following the discussions above, we first construct the IM defined by (8) which comes out to be as follows:

$$i_{5\times 5} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \quad \text{...(10)}$$

Hence, representing the Socioeconomic groups concerned, *i.e.*, ST, SC, OBC, OTHERS and ALL as S_1, S_2, S_3, S_4 and S_5 respectively, we get here two clusters as follows:

$$C_1 = \{S_1, S_3, S_4, S_5\} \text{ and}$$

$$C_2 = \{S_2\}.$$

So, it may be concluded that in respect of the relation of MPCE with proportion of MPCE on food, the behaviour of the SC group do not match with that of the other groups while the other groups behave in the same manner.

5. CONCLUSIONS

The above test procedure enables one to perform CCC for the two regressions in each and every pair of regressions possible out of several given regressions. Needless to say, this generalises the Chow test in two directions. This, further, has the following important implications.

Suppose the regressions are now arranged successively in a definite order for investigating existence of structural change. Then, obviously, if this procedure is carried out, CCC between every two successive regressions out of the given regressions, arranged now successively, is automatically done and hence detailed informations regarding all structural changes are obtained, if there is any such at all [5].

Again, once this procedure is performed, *i.e.*, CCC between the two regressions of all possible pairs of regressions out of the several given regressions are done, one can partition the set of the given regressions into, say, clusters, *every cluster consisting of those regressions which satisfy the condition that, as already stated, the vectors of coefficients of any two of these regressions do not differ from each other significantly*. An example has already been illustrated.

It is to be noted that no. of clusters need not always be two; it may be more than two as well as be one when and only when all the regressions coincide, *i.e.*, the Indicator Matrix becomes a null matrix of an appropriate order.

It may be noted that the clustering introduced here is quite different from that which is obtained by using Mahalanobis D^2 Statistic [4] or using *k*-means Method [5]—clustering by means of each of the later approaches is based on a single variable/a vector of variables while that by means of our procedure is based entirely on *relationship of variables*.

It is a hunch that the procedure introduced here can be extended for the purpose of clustering on the basis of not only one relationship of variables but more than one relationship together!

NOTES

1. By complete comparison between any two parameters a and b we mean to decide whether $a < b$ or $a = b$ or $a > b$. By component wise complete comparison (CCC) between two vectors of parameters of the same size $(a_1\, a_2 \ldots a_m)$ and $(b_1\, b_2 \ldots b_m)$ we mean complete comparison between (a_1 and b_1), (a_2 and b_2), ... and (a_m and b_m). By CCC between/of/for two regressions with same no. of parameters we mean CCC between the two vectors of parameters of these regressions. In the paper by Saha and Pal, CCC is done between every two successive regressions out of any number of given successive regressions with same no. of parameters.
2. Of course, IM provides limited knowledge that for every two regressions, whether they coincide or not, nothing more.
3. The no. of States/UTs, as in the Report, is 35. But there 5 for each of which some figure/s is/are missing and hence we have excluded those. Also, it may be noted that the numbers of observations for the regressions need not be the same.
4. It is to be noted that if one requires only CCC between every two successive regressions out of given several successive regressions and nothing more, it is not necessary to carry out the procedure described here; for that

purpose it is sufficient to work out only the procedure laid for that purpose in the paper by Saha and Paul (2014) referred earlier[2].

REFERENCES

Chow, Gregory C. (1960), Tests of Equality between Sets of Coefficients in Two Linear Regressions, *Econometrica*, Vol. 28, No. 3, pp. 591-605.

Saha, J. & Pal, M. (2014), A Modified Chow Test Approach Towards Testing Differences in The Engel Elasticities, *Asian-African Journal of Economics and Econometrics,* Vol. 14, No. 1, 2014; 57-67.

National Sample Survey Organisation (2007), Household Consumer Expenditure Among Socio-Economic Groups: 2004–2005, NSS 61st Round (July 2004-June 2005), *NSSO Report No.* 514(61/1.0/7), Government of India.

Mahalanobis P.C. (1936): On the generalised distance in statistics, *Proceedings of the National Institute of Sciences of India* 2(1): 49–55. Retrieved 2012-05-03.

R in Action, Second Edition, Manning publishing house.