Artha Vijnana
Sept 1988 V 30 N 3, 277-286

Relationship Between Sub-Period and Total Period Regression

Pradeep Apte

The problem with which we deal here may be stated as follows. Suppose we have two sample-data sets, each pertaining to successive periods or groups. Let these be denoted as $[X_1, Y_1]$ with n_1 observations and $[X_2, Y_2]$ with n_2 observations. We first compute regression slope-coefficient using these two data sets separately and denote these as β_1, and β_2 for the first and second samples respectively.

Now combining these two sub-period data sets we have yet another sample i.e. $[(X_1, X_2); (Y_2, Y_2)]$. With $n_1 + n_2$ number of observations and calculate slope coefficient again, denoted by β. What relationship does exist between β_1, β_2 and β?

Relevance of such a problem can be easily imagined. Research workers often handle the situations where the given data is divided into sub-groups and regression parameters for the sub-groups taken separately and these subgroups taken or pooled together are calculated. For example; growth rates, incremental capital output ration (ICOR) consumption function etc. are often estimated, using regression, for, sub-periods as well as for the total period. Generally, (perhaps due to experience with arithmatic averages), one expects that growth rate or ICOR, for the total period will be some-where between the subperiod parameter estimates. [1] But this does not always happen and a seemingly paradoxical situation emerges, where the parameter of the 'total' period lies entirely above or below the subgroup parameters. Such a result needs some straightforward explanation which standard text-books on statistics do not often provide. [2]

A similar question has been dealt with by Placket (1950). [3] But the question was posed, somewhat differently, as follows. If we have a sample which is initially given and later