A NOTE ON STANDARD PARTIAL REGRESSION COEFFICIENTS

BU-13[illegible]-M

Walter T. Federer

October, [illegible]

ABSTRACT

The range of values of the correlation coefficient between two independent variables, given the two correlation-coefficients of the dependent with each of the two independent variables, is investigated both empirically and analytically. Likewise, the range of values of the standard partial regression coefficient is investigated both empirically and analytically for the case of equal correlation coefficients of the dependent and with each of the two independent variables. Some additional results were obtained for the case of unequal correlation coefficients and on the relationships among the various statistics in the multiple correlation coefficient. The methods used could be employed to investigate the problems considered in the paper when there are three or more independent variables.

A NOTE ON STANDARD PARTIAL REGRESSION COEFFICIENTS*

BU-139-M Walter T. Federer October, 1961

Although a large number of the books in statistics discuss the computational aspects of standard partial regression coefficients, very few authors (e.g., Fisher, Statistical Methods for Research Workers, 1936; Wilks, Mathematical Statistics, 1944; Snedecor, Statistical Methods, 1946; Steel and Torrie, Principles and Procedures of Statistics, 1960) have anything to say about the interpretation of partial or standard partial regression coefficients. Apparently no one has bothered to describe the range of standard partial regression coefficients in multiple regression.

The purpose of this note is to investigate some aspects of the range of values for standard partial regression coefficients for two independent variates and one dependent variate in multiple regression. En route to this investigation some interrelationships among the total or zero order correlations are obtained.

Snedecor (loc. cited) does comment on the standard partial regression coefficient for one independent variate. In particular he states that the standard partial regression coefficient, b'_{yx}, of Y on X is:

$$b'_{yx} = b_{yx}\frac{s_x}{s_y} = \frac{\Sigma xy}{\Sigma x^2}\sqrt{\frac{\Sigma x^2}{\Sigma y^2}} = r_{yx} \quad .$$

Hence, the range of b'_{yx} must be between minus one and plus one.

For the case of two independent variates, X_1 and X_2, and one dependent variate, Y, the standard partial regression coefficients may be expressed in terms of the total correlation coefficients as:

$$b'_{y1\cdot 2} = \frac{r_{y1}-r_{12}r_{y2}}{1-r_{12}^2}$$

and

$$b'_{y2\cdot 1} = \frac{r_{y2}-r_{12}r_{y2}}{1-r_{12}^2} \quad .$$

Also,

$$R^2 = \frac{r_{y1}^2+r_{y2}^2-2r_{12}r_{y1}r_{y2}}{1-r_{12}^2} \quad .$$

---

*Biometrics Unit, Plant Breeding Department, Cornell University, Ithaca, N. Y.

In passing, we note some facts about the correlation coefficient r_{12}. The multiple correlation coefficient squared takes on its minimum value when

$$r_{12} = \frac{r_{y1}^2+r_{y2}^2}{2r_{y1}r_{y2}} \pm \sqrt{\left(\frac{r_{y1}^2+r_{y2}^2}{2r_{y1}r_{y2}}\right)^2-1}$$

$$= \frac{r_{y1}}{r_{y2}} \text{ and } \frac{r_{y2}}{r_{y1}} \quad .$$

One of the above roots will be inadmissible whenever $r_{y1} \neq r_{y2}$, because it is outside the range minus one to plus one. Hence, the value of r_{12} lying between minus and plus one is the one yielding a minimum value of R^2. If $r_{y1}=r_{y2}$, then $r_{12}=1$ is the value of r_{12} yielding a minimum value of R^2. (The above result was obtained by differentiating R^2 with respect to r_{12}, setting the resulting equation equal to zero, and solving for r_{12}. Since the equation is quadratic in r_{12}, two roots are obtained when $r_{12}= \pm 1$, the value of R^2 tends to a minimum in the limit as r_{12}^2 tends to unity.)

Also, we may note that since r_{12} is a function of R^2, r_{y1}, and r_{y2}, the range of values possible for r_{12} may be shorter than the entire range from minus to plus unity. The range of values for r_{12} is found from the multiple regression equation with R^2 set equal to unity, thus $-(1-r_{12}^2)R^2+r_{y1}^2+r_{y2}^2-2r_{12}r_{y1}r_{y2}=0$. The two roots of r_{12} from the above quadratic equation in r_{12} are:

$$r_{12} = r_{y1}r_{y2} \pm \sqrt{(R^2-r_{y1}^2)(R^2-r_{y2}^2)}/R^2 \quad .$$

When R^2 is set equal to one, the smaller root is the lowest value possible for r_{12} and the larger root is the highest value possible for r_{12}.

The minimum value of R^2 may be obtained from the two roots of R^2 from the term under the radical above equated to zero, or $R^4-R^2(r_{y1}^2+r_{y2}^2)+r_{y1}^2r_{y2}^2=0$. The two roots are

$$2R^2 = r_{y1}^2+r_{y2}^2 \sqrt{(r_{y1}^2+r_{y2}^2)^2-4r_{y1}r_{y2}^2}$$

or

$$R^2 = r_{y1}^2 \text{ or } r_{y2}^2 \quad ,$$

whichever is the larger.

Two particular examples using the above results were computed for varying values of r_{12} (Figure 1). For the first example, r_{y1} was set equal to 0.8 and r_{y2} was set equal to 0.2. Then the range of values for r_{12} may be computed from:

$$r_{12} = r_{y1}r_{y2} \pm \sqrt{(1-r_{y1}^2)(1-r_{y2}^2)}$$

$$= .16 \pm \sqrt{.36(.96)} = -.4279 \text{ or } +.7478 \;.$$

Then,

$$-.4279 \leq r_{12} \leq .7478 \;.$$

The value of r_{12} yielding a minimum value of R^2 is $r_{12}=.2/.8=.25$. This minimum value is

$$.64 = r_{y1}^2 = \left\{.64+.04-2(.25)(.2)(.8)\right\} / (1-.25^2) \;.$$

As a second example, take $r_{y1}=.2=r_{y2}$. Here we find that the range of values for r_{12} is obtained from:

$$r_{12} = .2(.2) \pm \sqrt{.96^2} = .04 \pm .96 \;.$$

Therefore,

$$-.92 \leq r_{12} \leq 1 \;.$$

The value of r_{12} yielding a minimum value of R^2 is

$$r_{12} = r_{y1}/r_{y2} = 1 \;.$$

The minimum value of R^2 is obtained in the limit as r_{12} tends to unity and is $r_{y1}^2=r_{y2}^2= .04$.

The range of values for $b'_{y1\cdot 2}$ and $b'_{y2\cdot 1}$ for the first example were investigated empirically (see Figure 2) and found to be

$$+.787 \leq b'_{y1\cdot 2} \leq 1.477 \quad (\text{for } r_{12}=.127 \text{ and } r_{12}=.7478)$$

and

$$-.90 \leq b'_{y1\cdot 2} \leq .65 \quad (\text{for } r_{12}= -.4279 \text{ and } r_{12}=.7478)$$

For the second example, the empirical investigation indicated that the range of values for the standard partial regression coefficients was

$$0.1 \leq b'_{y1\cdot2} = b'_{y2\cdot1} \leq 2.5$$

In both cases the range is bounded within rather narrow limits but not between plus and minus unity as was the case for one independent variate.

Something more than the empirical investigation can be done. First consider the case for $r_{y1}=r_{y2}$. Then, the standard partial regression coefficients may be written as

$$b'_{y1\cdot2} = \frac{(r_{y1}=r_{y2})}{1+r_{12}} = b'_{y\cdot21} \quad ,$$

the range of possible values for r_{12} is

$$2r^2_{y1}-1 \leq r_{12} \leq 1$$

The lower limit of r_{12} tends to minus unity as r_{y1} tends to zero. Since $b'_{y1\cdot2} = r_{y1}/(1+r_{12})$ is either an increasing (r_{y1} negative) or a decreasing (r_{y1} positive) function as r_{12} varies from its lower limit to its upper limit, the range of values for $b'_{y1\cdot2}=b'_{y2\cdot1}$ may be found by inserting the values of the two end-points for r_{12}. Doing this, we find the range on the standard partial regression coefficient to be:

$$\frac{r_{y1}}{2} \leq b'_{y1\cdot2} = b'_{y2\cdot1} \leq 1/2r_{y1}$$

for r_{y1} positive. (The inequalities are reversed for r_{y1} negative.) As r_{y1} tends to zero, the right hand side tends to infinity. However, for r_{y1} greater than one-tenth, say, the upper bound is less than five.

Now consider the case for $r_{y1} > r_{y2}$ and both of the same sign. Then, the value of r_{12} yielding a minimum value of R^2 is r_{y2}/r_{y1}. The values of $b'_{y1\cdot2}$ and $b'_{y2\cdot1}$ at $r_{12}=r_{y2}/r_{y1}$ are:

$$b'_{y1\cdot2} = \frac{r_{y1}-r_{y2}(r_{y2}/r_{y1})}{1-(r_{y2}/r_{y1})^2} = r_{y1}$$

and

$$b'_{y2\cdot1} = \frac{r_{y2} - r_{y1}(r_{y2}/r_{y1})}{1-(r_{y2}/r_{y1})^2} = 0$$

For $r_{y1}r_{y2}$ = a positive number the larger root of r_{12} is

$$r_{12} = r_{y1}r_{y2} + \sqrt{(1-r_{y1}^2)(1-r_{y2}^2)} = a_1$$

and the smaller root is

$$r_{12} = r_{y1}r_{y2} - \sqrt{(1-r_{y1}^2)(1-r_{y2}^2)} = a_2$$

at these end points

$$b'_{y1\cdot2} = \frac{r_{y1}-r_{y2}a_1}{1-a_1^2}$$

and

$$b'_{y1\cdot2} = \frac{r_{y1}-r_{y2}a_2}{1-a_2^2}$$

Likewise, the values of $b'_{y2\cdot1}$ at the end values for r_{12} are:

$$b'_{y2\cdot1} = \frac{r_{y2}-r_{y1}a_1}{1-a_1^2}$$

and

$$b'_{y2\cdot1} = \frac{r_{y2}-r_{y1}a_2}{1-a_2^2}$$

As indicated in Figure 2, the minimum value of $b'_{y1\cdot2}$ in the first example is not obtained at any of the three points (two end points for r_{12} and value of r_{12} yielding a minimum R^2) listed above. However, the range of $b'_{y2\cdot1}$ is obtained by observing what happens at the two end points for r_{12}.

Given that $r_{y1} > r_{y2}$ and r_{y1} is positive, the value of r_{12} yielding a minimum value of $b'_{y1\cdot2}$ for r_{y1} and r_{y2} fixed is obtained as follows:

$$\frac{\partial b'_{y1\cdot2}}{\partial r_{12}} = \frac{-r_{y2}(1-r_{12}^2)+(r_{y1}-r_{y2}r_{12})(2r_{12})}{(1-r_{12}^2)^2} = 0$$

Solving for the r_{12}, we find:

$$r_{12} = \frac{r_{y1} \pm \sqrt{r_{y1}^2 - r_{y2}^2}}{r_{y2}}$$

(For the first numerical example above

$$r_{12} = \frac{.8 \pm \sqrt{.60}}{.2} = \frac{.8 \pm .7746}{.2} = 7.873 \quad \text{or} \quad .127 \quad .$$

Therefore, $r_{12}=.127$.)

From the above we note that for $r_{y1} > r_{y2}$ and both positive the lowest value for $b'_{y1\cdot 2}$ is obtained when

$$r_{12} = \frac{r_{y1} + \sqrt{r_{y1}^2 - r_{y2}^2}}{r_{y2}}$$

and the largest value of $b'_{y1\cdot 2}$ is obtained when r_{12} equals one of its two end point values, i.e.,

$$r_{12} = r_{y1}r_{y2} \pm \sqrt{(1-r_{y1}^2)(1-r_{y2}^2)} \quad .$$

We cannot obtain the value of r_{12} minimizing $b'_{y2\cdot 1}$ in the above manner because the number under the radical would be negative. Hence, we look at the two end values of r_{12} to obtain the range on $b'_{y2\cdot 1}$.

These arguments may be extended to multiple regression with more than two independent variates. From the above, it would appear that standard partial regressions are bounded for $r_{yi} > 0$ $(i=1,2,\cdots)$ and that the upper bounds are fairly low for relatively large total correlation coefficients between the dependent and the independent variables.

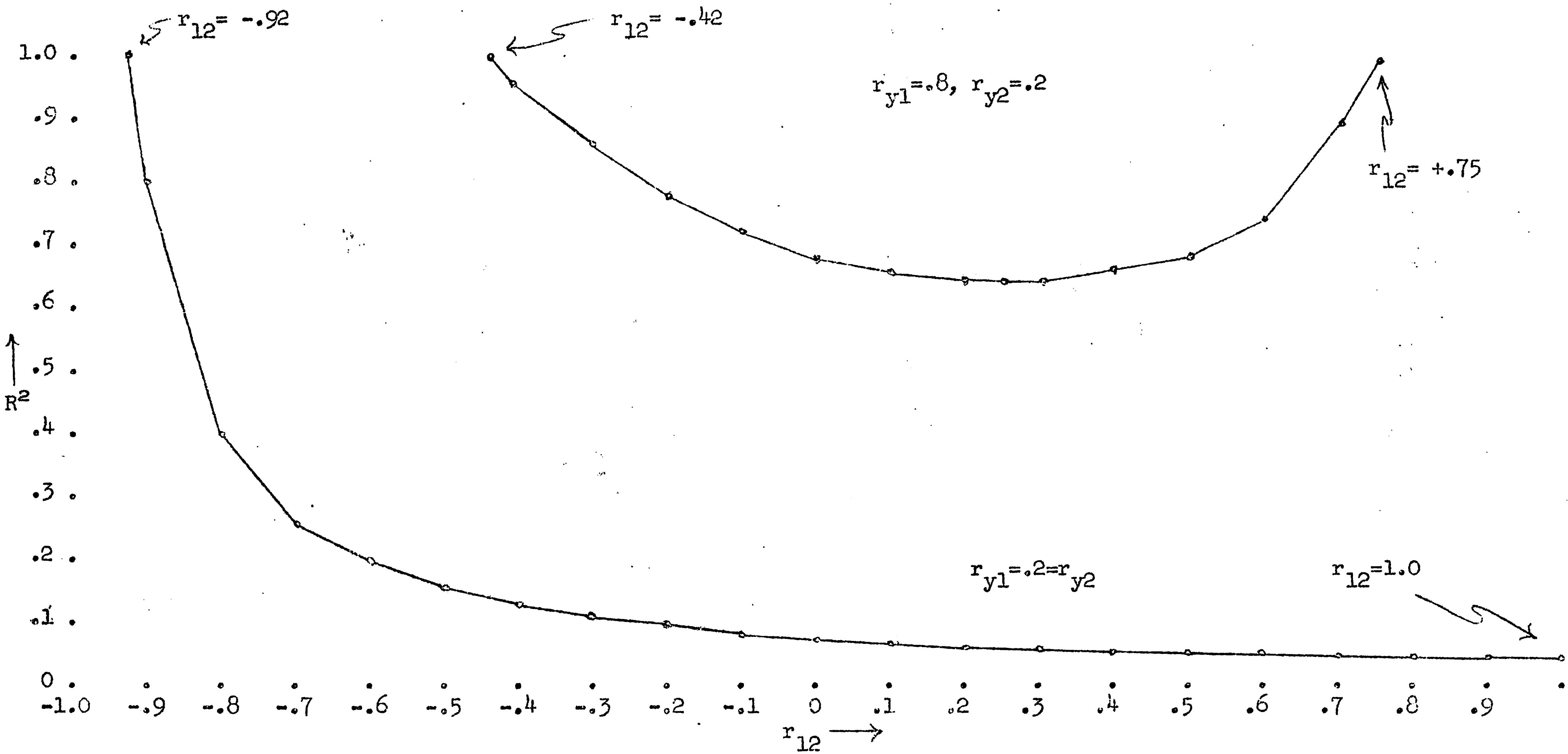

Figure 1. Values of R^2 for varying values of r_{12} for 2 cases, $r_{y1}=.2=r_{y2}$, and $r_{y1}=.8$ and $r_{y2}=.2$.

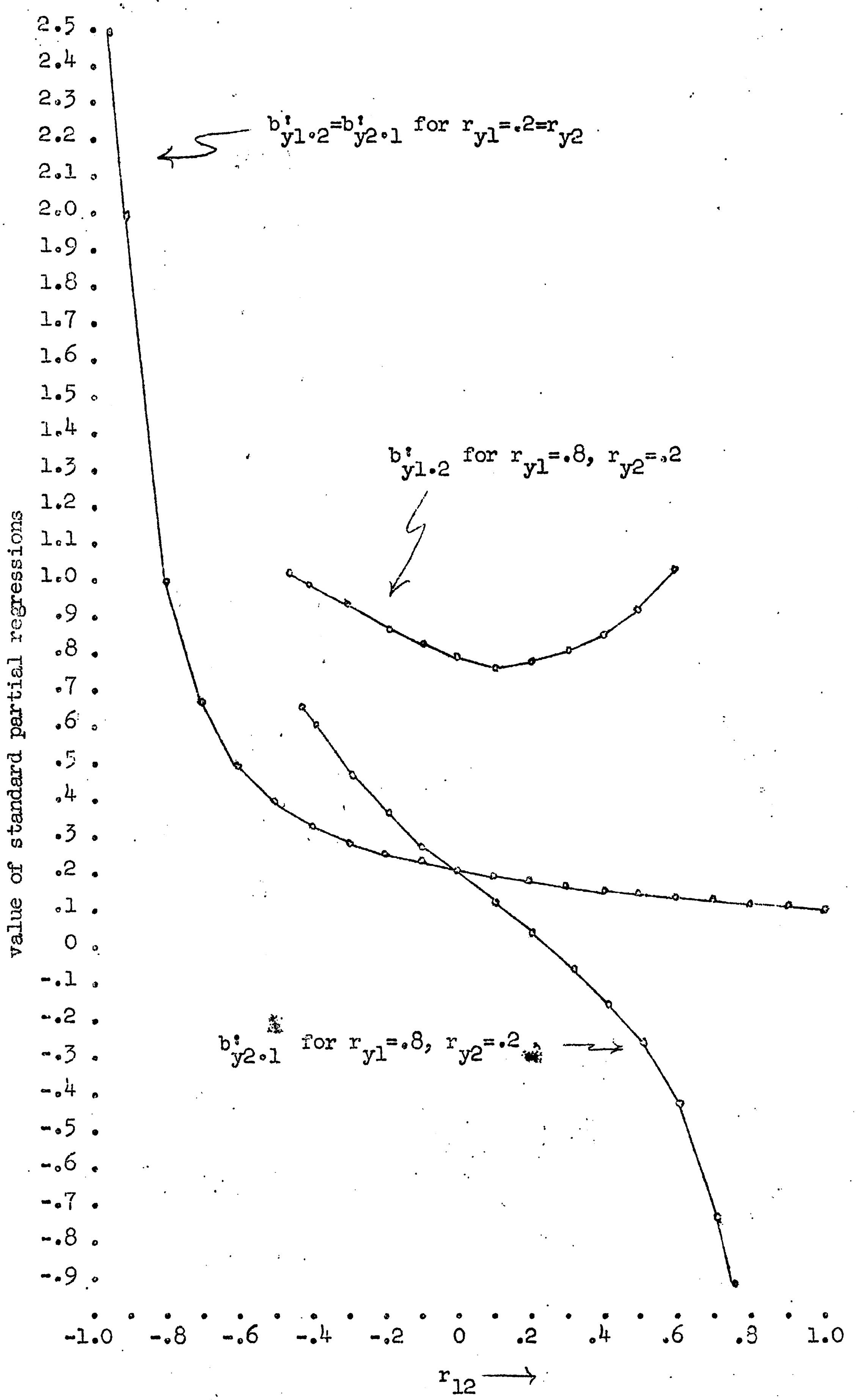

Figure 2. The values of standard partial regression coefficients for varying values of r_{12}.