EFFECTS OF INTERCORRELATION
UPON MULTIPLE CORRELATION
AND REGRESSION MEASURES

by

Karl A. Fox and James F. Cooney, Jr.

United States Department of Agriculture

Agricultural Marketing Service

Washington 25, D. C.

April 1954

PREFACE

From certain viewpoints intercorrelation (that is, correlation between independent variables) is not a major problem in statistical analysis. Routine instructions for solving multiple-regression problems include formulas for net regression coefficients and standard errors which automatically take account of the effects of intercorrelation. Nevertheless, research workers are frequently surprised when two analyses showing nearly the same direct correlations between the dependent and each independent variable yield widely differing net regression and multiple correlation coefficients.

Many of the three-variable calculations discussed in this paper were developed by the senior author in May 1947 to explain such happenings in a precise way. Late in 1952 the junior author rechecked and extended the three-variable calculations. He also developed representative calculations for the four-variable case, setting up the intercorrelation formulas in a matrix notation which permits generalization to any number of variables.

The authors are indebted to Frederick V. Waugh for helpful suggestions on the four-variable and general cases. The four-variable computations were carried out by Jacqueline Spiro.

CONTENTS

EFFECTS OF INTERCORRELATION UPON MULTIPLE CORRELATION AND REGRESSION MEASURES

by

Karl A. Fox, Chief, Statistical and Historical Research Branch,
and James F. Cooney, Jr., Mathematical Statistician
Agricultural Marketing Service

Trained statisticians have known the effects of intercorrelation in a general way for some 60 years. They have given particular attention to the extreme case ("multicollinearity") in which two or more independent variables are so highly correlated that their separate effects cannot be distinguished. At the other extreme, where there is no intercorrelation, the effects of the different independent variables are strictly additive.

Many users of the graphic method of regression analysis know that intercorrelation between independent variables tends to delay the convergence of successive graphic approximations toward the mathematical solution.

Trained statisticians are also aware that increasing levels of intercorrelation are reflected in increasing standard errors of net regression coefficients--that is, high intercorrelation tends to mean lowered reliability for the individual regression constants. But apparently it is safe to say that most students of elementary statistics and most persons who make regression analyses as an adjunct to their applied work have only a vague idea of the effects of intercorrelation and frequently get results from multiple-correlation analyses that they are unable to explain. More concrete information on the effects of intercorrelation through its whole range of variation and not merely at the points 0 and 1 is therefore expected to be useful. This information for the three-variable case is shown by means of charts and tables for a number of pairs of values of the simple (or gross) correlation coefficients between the dependent variable and each of the two independent variables. The four-variable case is treated for a more limited range of numerical values but in a notation which permits generalization of the results to any number of variables.

SUMMARY

With given values for r_{12} and r_{13} and a specified size of sample, all of the correlation measures in a three-variable problem can be expressed simply as functions of r_{23}. In certain cases as r_{23} increases the coefficient of multiple determination declines continuously. In other cases it trends down to a minimum value and then increases. For given values of r_{12} and r_{13}, r_{23} can take only a limited range of values. In some cases as the degree of intercorrelation increases beyond a certain level, the "weaker" of the two partial regression coefficients changes sign. Within a considerable range of values of r_{23} in the region of this sign change, the value of the corresponding partial regression coefficient, for samples of about 20 observations, does not differ significantly from zero.

The four-variable case is more complicated because six simple correlation coefficients are involved. The same approach can be used, however, by specifying five of these and allowing the sixth one (r_{34}) to vary over its entire range of possible values. When each of the three simple intercorrelation coefficients is very high the values of the partial regression coefficients are very unstable and, in samples of about 20 observations, are smaller than their standard error in all but a small portion of the range of permissible values of r_{34}. As r_{23} and r_{24} take on smaller values, the partial regression coefficients acquire greater stability and exceed their standard error over a large part of the permissible range of r_{34}.

EFFECTS OF INTERCORRELATION IN THE THREE-VARIABLE CASE

The Approach

The general problem of three-variable regression analysis is to estimate the values of a dependent variable, X_1, based on given values of two independent variables, X_2 and X_3. Assume that we have already calculated the direct correlation coefficients (r_{12} and r_{13}) between X_1 and X_2 on the one hand and X_1 and X_3 on the other for a number of different problems. Suppose that several of the analyses, based on entirely different sets of data, have yielded the same values of r_{12} and r_{13}. Nevertheless, in each case, we may obtain different values for the multiple and partial correlation coefficients and for the net regression, or beta, coefficients. The coefficient of multiple determination, $R^2_{1.23}$, may vary from almost 1.0 down to 0.5, or lower.

We then ask, Why do these differences occur? In the three-variable case, these variations can be wholly explained by variations in the value of the intercorrelation coefficient, r_{23}, between the independent variables X_2 and X_3.

Basic Formulas

Several standard methods of calculating multiple correlation and regression coefficients start out from the determinant of simple correlation coefficients, which for the three-variable case is as follows:

$$\triangle_3 = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix} = 1 + 2\, r_{12}\, r_{13}\, r_{23} - r_{12}^2 - r_{13}^2 - r_{23}^2 \,. \qquad (1)$$

All the basic correlation and regression measures in the three-variable case can be derived from the values of the three simple correlation coefficients, with the following exceptions which are trivial in the present context: First, if we wish to talk about net regression coefficients in terms of original units (pounds, dollars, and so on) rather than normalized or standard deviation units we must multiply the beta coefficients by the ratio of the standard deviation of the dependent variable to that of the particular independent variable concerned. Second, the standard error of the beta coefficient, or of the corresponding net regression coefficient, is affected by the number of observations in the sample.

It may be seen from these formulas that once we have fixed the values of r_{12} and r_{13}, the various measures can be expressed simply as functions of r_{23}. In the charts and tables that follow, the value of each correlation measure was calculated for series of values of the intercorrelation coefficient r_{23} covering all or nearly all of the range of possible values of that coefficient, given the stated values of r_{12} and r_{13}. 1/

Discussion of Charts and Tables

In figure 1 the values of r_{12} and r_{13} were chosen in such a way that the coefficient of multiple determination is 0.98 when the intercorrelation coefficient is zero. As r_{23} increases, the value of $R_{1.23}^2$ declines continuously, approaching a lower limit of 0.49 as r_{23} approaches 1. At this point the variable X_3 adds nothing to the explanation of X_1 that is not already given by the single independent variable X_2. The partial correlation coefficient $r_{12.3}$ decreases continuously as r_{23} increases, approaching zero as r_{23} approaches 1. The beta coefficient also decreases through this entire range and its standard error increases. By the time r_{23} exceeds 0.7, the beta coefficient (based on an assumed 20 observations) is no longer significantly different from zero at the commonly used 5-percent probability level.

Figure 2 illustrates the fact that the values of r_{12} and r_{13} set certain limits upon the range of values which r_{23} may take. Obviously, if X_2 and X_3 are both closely correlated with X_1 they have some degree of correlation with each other. The exact nature of the limits set upon r_{23} by the values of r_{12} and r_{13} is shown in appendix note 1. In this particular case, r_{23} cannot be lower than 0.62.

Figure 3 shows a result that may be surprising to many applied workers. As intercorrelation increases beyond a certain level, the "weaker" of the two partial regression coefficients changes sign from positive to negative. This

---

1/ In the charts and tables that follow, r_{12} and r_{13} are always taken as positive, and the corresponding values of r_{23} and other measures are predominantly positive. If the same absolute values of r_{12} and r_{13} are taken with negative signs, the corresponding values of $R_{1.23}^2$, r_{23}, $S_{\beta_{12.3}}$, and $S_{\beta_{13.2}}$ are the same as before; absolute values of $\beta_{12.3}$, $\beta_{13.2}$, $r_{12.3}$ and $r_{13.2}$ are the same as before but with the opposite sign. If we take r_{12} positive and r_{13} negative, r_{23} will be predominantly negative. Values of $R_{1.23}^2$, $S_{\beta_{12.3}}$, $S_{\beta_{13.2}}$, $\beta_{12.3}$ and $r_{12.3}^2$ will be the same as in the first case (r_{12} and r_{13} both positive); and the absolute values of $\beta_{13.2}$ and $r_{13.2}$ will be the same as in the first case but with opposite sign. Finally, if we take r_{12} negative and r_{13} positive, the values of r_{23} will be predominantly negative; those of $R_{1.23}^2$, $S_{\beta_{12.3}}$, $S_{\beta_{13.2}}$, $\beta_{13.2}$ and $r_{13.2}$ will be the same as in the first case; and the absolute values of $\beta_{12.3}$ and $r_{12.3}$ will be the same as in the first case but with opposite sign. As the absolute values of all the measures are unchanged by these interchanges of signs, the figures tabulated below each chart can be used for all four cases with appropriate changes in signs.

Equations (2) through (7) define the various correlation and regression measures in terms of the three simple correlation coefficients:

$$R^2_{1.23} = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}\, r_{13}\, r_{23}}{1 - r_{23}^2} \tag{2}$$

$$\beta_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{1 - r_{23}^2} \tag{3}$$

$$b_{12.3} = \beta_{12.3} \cdot \frac{S_1}{S_2}\,, \tag{3.1}$$

where S_1 and S_2 are the standard deviations of X_1 and X_2 respectively.

$$\beta_{13.2} = \frac{r_{13} - r_{12}\, r_{23}}{1 - r_{23}^2} \tag{4}$$

$$b_{13.2} = \beta_{13.2} \cdot \frac{S_1}{S_3}\,, \tag{4.1}$$

where S_3 is the standard deviation of X_3.

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1-r_{23}^2)\,(1-r_{13}^2)}} \tag{5}$$

$$r_{13.2} = \frac{r_{13} - r_{12}\, r_{23}}{\sqrt{(1-r_{23}^2)\,(1-r_{12}^2)}} \tag{6}$$

$$S_{\beta_{12.3}} = S_{\beta_{13.2}} = \frac{\sqrt{1-R^2_{1.23}}}{(1-r^2_{23})\sqrt{N-3}} = \frac{\sqrt{1+2r_{12}r_{13}r_{23}-r_{12}^2-r_{13}^2-r_{23}^2}}{(1-r^2_{23})\sqrt{N-3}} \tag{7}$$

$$S_{b_{12.3}} = S_{\beta_{12.3}} \cdot \frac{S_1}{S_2}\,, \text{ and} \tag{7.1}$$

$$S_{b_{13.2}} = S_{\beta_{13.2}} \cdot \frac{S_1}{S_3}\,. \tag{7.2}$$

change in sign occurs at a value of r_{23} somewhat above the value of the lower of the two simple correlation coefficients, r_{12} and r_{13}. Within a considerable range of values of r_{23} in the region of this sign-change, the value of the corresponding beta coefficient would not differ significantly from zero. Other features illustrated in figure 3 are (1) that the "stronger" of the two regression coefficients increases for a time as the intercorrelation increases, and (2) that the coefficient of multiple determination trends down to a minimum at some value of r_{23} greater than the lower of the two direct coefficients and then increases again.

The characteristics of figure 3 are repeated in the data shown in tables 4 and 5 and in figure 4. Each shows minimum values 2/ for the coefficient of multiple determination, $R^2_{1.23}$, the stronger partial correlation coefficient, $r_{12.3}$, and the stronger regression coefficient, $\beta_{12.3}$, and each shows a sign change for the weaker coefficients, $\beta_{13.2}$ and $r_{13.2}$. As the spread between r_{12} and r_{13} increases, so also does the range of permissible values of r_{23}. When r_{13} falls to 0.3, r_{23} can take on small negative values as well as positive values. If r_{13} equals 0.1, then r_{23} can take values slightly lower than -0.3.

The summary tables contain values of the "t-ratios", that is, ratios of the respective net regression coefficients to their standard errors. As in each of the last four cases $\beta_{12.3}$ has a minimum and $S_{\beta_{12.3}}$ has a maximum, the corresponding t-ratio has a minimum beyond which it rises again with further increases in r_{23}.

2/ That is, minima in the mathematical sense of points at which the slope with respect to r_{23} is zero and becomes positive as r_{23} increases and negative as r_{23} decreases. Further information on these minimum values is given in appendix note 2.

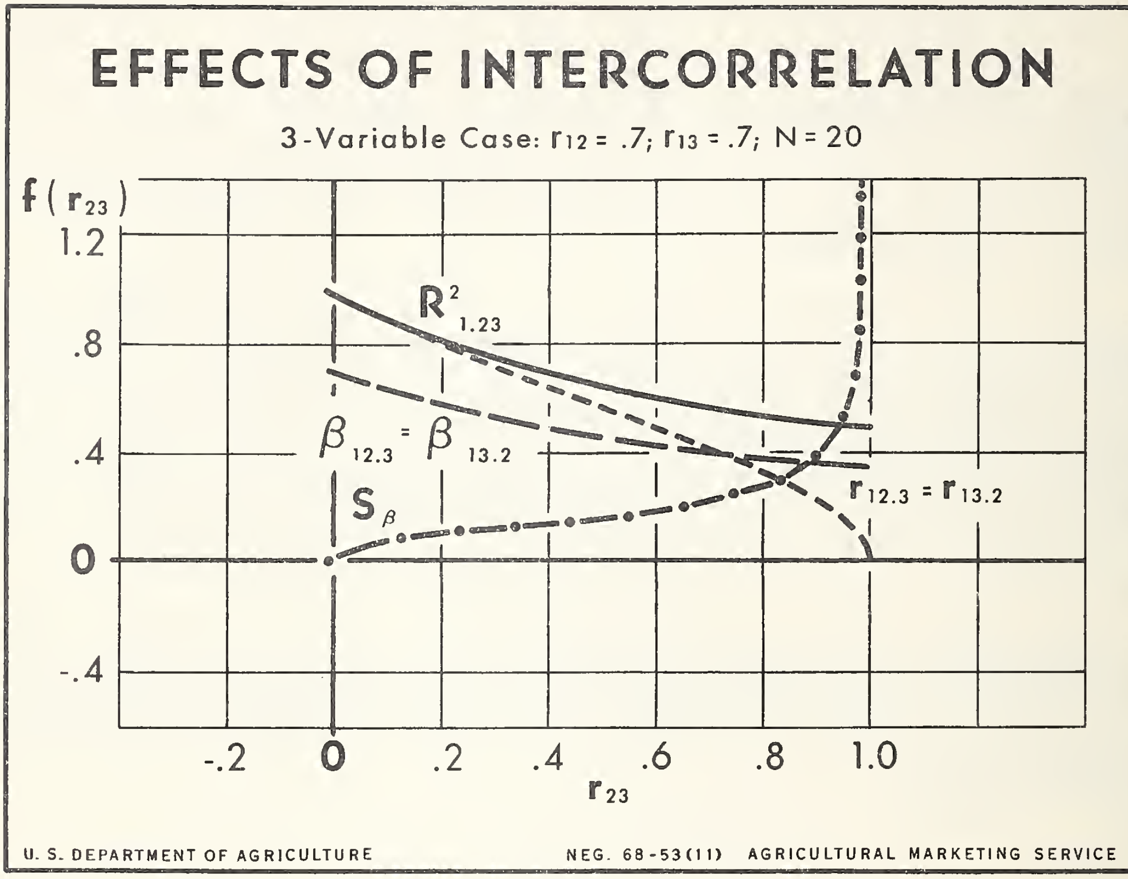

Figure 1

Table 1.- Data for case in which r_{12} = 0.7, r_{13} = 0.7, and N = 20

r_{23}	$R^2_{1.23}$	$\beta_{12.3}$	$\beta_{13.2}$	$r_{12.3}$	$r_{13.2}$	S_β 1/	t-ratio for - $\beta_{12.3}$	t-ratio for - $\beta_{13.2}$
-0.020 2/	1.0000	0.7143	0.7143	1.0000	1.0000	0	---	---
0	.9800	.7000	.7000	.9803	.9803	0.0346	20.2312	20.2312
.100	.8909	.6364	.6364	.8866	.8866	.0806	7.8958	7.8958
.200	.8167	.5833	.5833	.8003	.8003	.1058	5.5132	5.5132
.300	.7538	.5385	.5385	.7193	.7193	.1261	4.2704	4.2704
.400	.7000	.5000	.5000	.6417	.6417	.1449	3.4507	3.4507
.500	.6533	.4667	.4667	.5659	.5659	.1649	2.8302	2.8302
.600	.6125	.4375	.4375	.4901	.4901	.1887	2.3185	2.3185
.700	.5765	.4118	.4118	.4118	.4118	.2209	1.8642	1.8642
.800	.5444	.3889	.3889	.3267	.3267	.2728	1.4256	1.4256
.900	.5158	.3684	.3684	.2249	.2249	.3872	.9514	.9514
.950	.5026	.3590	.3590	.1570	.1570	.5478	.6553	.6553
.980	.4949	.3535	.3535	.0985	.0985	.8655	.4084	.4084
.990	.4925	.3518	.3518	.0695	.0695	1.2278	.2865	.2865
.999	.4902	.3502	.3502	.0219	.0219	3.1623	.1107	.1107
1.000 3/	---	---	---	---	---	---	0	0

1/ Identical for each β.

2/ Lowest possible value of r_{23}.

3/ Highest possible value of r_{23}.

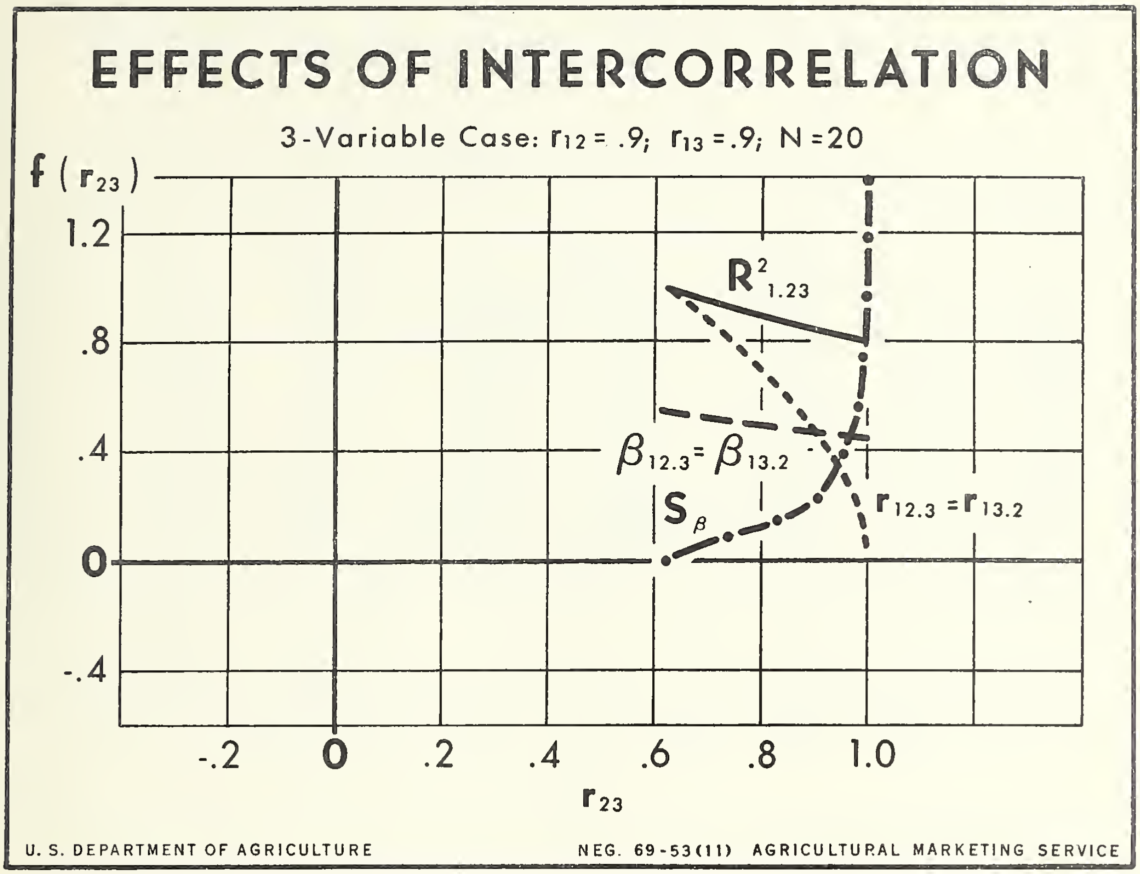

Figure 2

Table 2.- Data for case in which r_{12} = 0.9, r_{13} = 0.9, and N = 20

r_{23}	$R^2_{1.23}$	$\beta_{12.3}$	$\beta_{13.2}$	$r_{12.3}$	$r_{13.2}$	S_β 1/	t-ratio for - $\beta_{12.3}$	t-ratio for - $\beta_{13.2}$
0.620 2/	1.0000	0.5556	0.5556	1.0000	1.0000	0	---	---
.700	.9529	.5294	.5294	.8710	.8710	0.0735	7.2073	7.2073
.800	.9000	.5000	.5000	.6883	.6883	.1277	3.9154	3.9154
.900	.8526	.4737	.4737	.4737	.4737	.2135	2.2183	2.2183
.950	.8308	.4615	.4615	.3306	.3306	.3195	1.4444	1.4444
.980	.8182	.4545	.4545	.2076	.2076	.5193	.8752	.8752
.990	.8141	.4523	.4523	.1463	.1463	.7431	.6087	.6087
.999	.8000	.4500	.4500	.0462	.0462	2.0000	.2250	.2250
1.000 3/	---	---	---	---	---	---	---	---

1/ Identical for each β.

2/ Lowest possible value of r_{23}.

3/ Highest possible value of r_{23}.

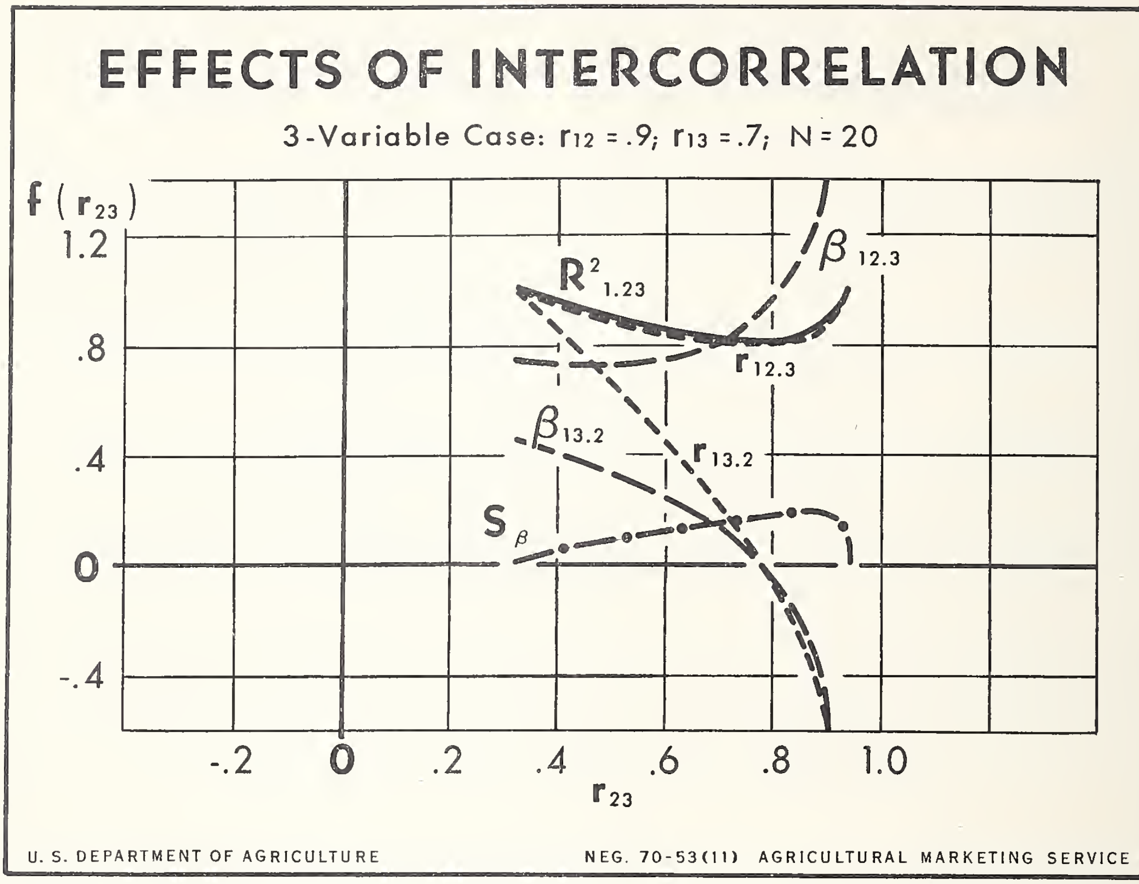

Figure 3

Table 3.- Data for case in which r_{12} = 0.9, r_{13} = 0.7 and N = 20

r_{23}	$R^2_{1.23}$	$\beta_{12.3}$	$\beta_{13.2}$	$r_{12.3}$	$r_{13.2}$	S_β 1/	t-ratio for - $\beta_{12.3}$	t-ratio for - $\beta_{13.2}$
0.3187 2/	1.0000	0.7535	0.4599	1.0000	1.0000	0	---	---
.4000	.9488	.7381	.4048	.9473	.8511	0.0599	12.3222	6.7579
.4500	.9191	.7335	.3695	.9172	.7578	.0772	9.5013	4.7863
.5000	.8933	.7333	.3333	.8892	.6623	.0917	7.9967	3.6347
.5500	.8703	.7384	.2939	.8635	.5632	.1046	7.0593	2.8098
.6000	.8500	.7500	.2500	.8402	.4588	.1175	6.3830	2.1277
.6500	.8329	.7706	.1991	.8200	.3472	.1305	5.9050	1.5257
.7000	.8196	.8039	.1373	.8039	.2249	.1442	5.5749	.9521
.7500	.8114	.8571	.0571	.7938	.0867	.1592	5.3838	.3587
.8000	.8111	.9444	-.0556	.7935	-.0765	.1758	5.3720	-.3163
.8500	.8252	1.0991	-.2342	.8107	-.2831	.1925	5.7096	-1.2166
.9000	.8737	1.4211	-.5789	.8673	-.5789	.1977	7.1882	-2.9282
.9413 3/	1.0000	2.1149	-1.2912	1.0000	-1.0000	0	---	---

1/ Identical for each β .

2/ Lowest possible value of r_{23}.

3/ Highest possible value of r_{23}.

Table 4.- Data for case in which $r_{12} = 0.9$, $r_{13} = 0.5$ and N = 20

r_{23}	$R^2_{1.23}$	$\beta_{12.3}$	$\beta_{13.2}$	$r_{12.3}$	$r_{13.2}$	S_β 1/	t-ratio for - $\beta_{12.3}$	t-ratio for - $\beta_{13.2}$
0.0725 2/	1.0000	0.8684	0.4371	1.0000	1.0000	0	---	---
.1000	.9798	.8586	.4141	.9864	.9454	0.0346	24.8150	11.9682
.2000	.9167	.8333	.3333	.9428	.7492	.0714	11.6709	4.6681
.3000	.8681	.8242	.2527	.9079	.5532	.0922	8.9393	2.7408
.4000	.8333	.8333	.1667	.8819	.3504	.1082	7.7015	1.5407
.5000	.8133	.8667	.0667	.8667	.1325	.1208	7.1747	.5522
.6000	.8125	.9375	-.0625	.8661	-.1147	.1311	7.1510	-.4767
.7000	.8431	1.0784	-.2549	.8892	-.4176	.1345	8.0178	-1.8952
.8000	.9444	1.3889	-.6111	.9623	-.8413	.0954	14.5587	-6.4057
.8275 3/	1.0000	1.5428	-.7766	1.0000	-1.0000	0	---	---

1/ Identical for each β.

2/ Lowest possible value of r_{23}.

3/ Highest possible value of r_{23}.

Table 5.- Data for case in which $r_{12} = 0.9$, $r_{13} = 0.3$ and N = 20

r_{23}	$R^2_{1.23}$	$\beta_{12.3}$	$\beta_{13.2}$	$r_{12.3}$	$r_{13.2}$	S_β 1/	t-ratio for - $\beta_{12.3}$	t-ratio for - $\beta_{13.2}$
-0.1458 2/	1.0000	0.9642	0.4406	1.0000	1.0000	0	---	---
-.1000	.9636	.9394	.3939	.9798	.8992	0.0465	20.2022	8.4710
0	.9000	.9000	.3000	.9435	.6882	.0767	11.7340	3.9113
.1000	.8545	.8788	.2121	.9166	.4842	.0930	9.4495	2.2806
.2000	.8250	.8750	.1250	.8987	.2810	.1035	8.4541	1.2077
.3000	.8110	.8901	.0330	.8901	.0722	.1105	8.0552	.2986
.4000	.8143	.9286	-.0714	.8921	-.1502	.1141	8.1385	-.6258
.5000	.8400	1.0000	-.2000	.9079	-.3974	.1120	8.9286	-1.7857
.6000	.9000	1.1250	-.3750	.9439	-.6883	.0959	11.7310	-3.9103
.6858 3/	1.0000	1.3107	-.5988	1.0000	-1.0000	0	---	---

1/ Identical for each β.

2/ Lowest possible value of r_{23}.

3/ Highest possible value of r_{23}.

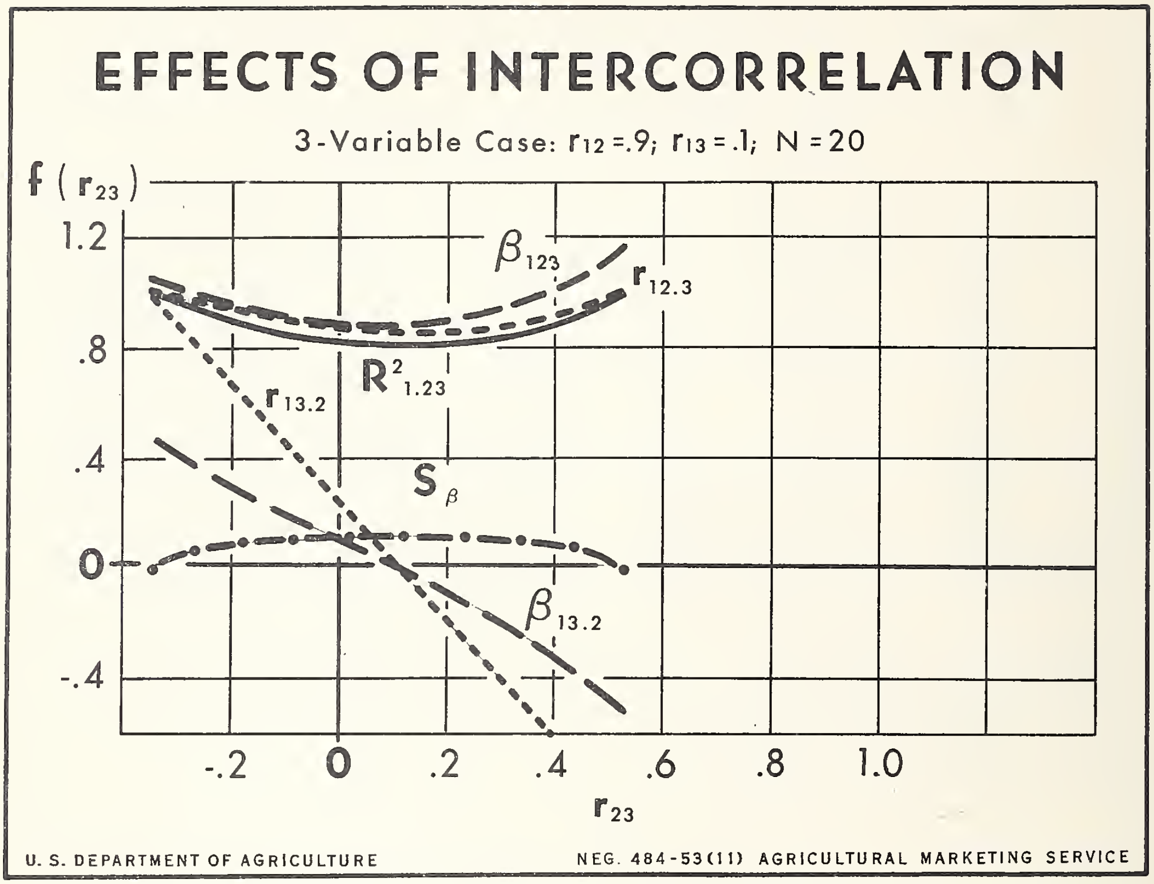

Figure 4

Table 6.- Data for case in which $r_{12} = 0.9$, $r_{13} = 0.1$, and N = 20

r_{23}	$R^2_{1.23}$	$\beta_{12.3}$	$\beta_{13.2}$	$r_{12.3}$	$r_{13.2}$	S_β 1/	t-ratio for - $\beta_{12.3}$	t-ratio for - $\beta_{13.2}$
-0.3437 2/	1.0000	1.0595	0.4641	1.0000	1.0000	0	---	---
-.3000	.9604	1.0220	.4066	.9798	.8899	0.0510	20.0392	7.9725
-.2000	.8917	.9583	.2917	.9437	.6556	.0812	11.8017	3.5924
-.1000	.8465	.9192	.1919	.9192	.4381	.1010	9.1010	1.9000
0	.8200	.9000	.1000	.9045	.2294	.1030	8.7379	.9709
.1000	.8101	.8990	.0101	.8990	.0231	.1122	8.0125	.0900
.2000	.8167	.9166	-.0833	.9027	-.1873	.1058	8.6635	-.7873
.3000	.8418	.9560	-.1868	.9166	-.4089	.1010	9.4653	-1.8495
.4000	.8930	1.0000	-.3023	.9319	-.6432	.0854	11.7096	-3.5398
.5000	.9733	1.1333	-.4667	.9864	-.9272	.0458	24.7445	10.1900
.5237 3/	1.0000	1.1680	-.5116	1.0000	-1.0000	0	---	---

1/ Identical for each β.
2/ Lowest possible value of r_{23}.
3/ Highest possible value of r_{23}.

EFFECTS OF INTERCORRELATION IN THE FOUR-VARIABLE CASE

Differences from the Three-Variable Case

The problem of intercorrelation in the four-variable case is much more complicated because there are now three intercorrelation coefficients instead of one. We have three independent variables, X_2, X_3, and X_4, and we may have intercorrelation between X_2 and X_3, X_2 and X_4, and X_3 and X_4.

Following the approach used in the three-variable case, let us suppose that we have a large number of four-variable regression analyses on different sets of data. We select a number of these analyses in which the values of r_{12}, r_{13}, and r_{14} (the direct or simple correlation coefficients between the dependent and each independent variable) are about the same. Nevertheless, we find that the partial correlation and net regression coefficients are different in each case. These differences are due to the varying degrees of intercorrelation, represented by combinations of values of r_{23}, r_{24}, and r_{34}.

A systematic exploration of the effects of intercorrelation in the four-variable case would involve a great deal of labor. One possibility would be to fix the values of r_{23} and r_{24} and to trace the effects of variations in r_{34} upon the different regression measures. Except for a change in notation, such a demonstration would apply equally well to changes in either of the other intercorrelation coefficients, r_{23} or r_{24}. Before doing this, however, we shall illustrate the complications of the four-variable case in terms of the basic formulas for correlation and regression coefficients.

Basic Formulas

It will be convenient at this point to introduce a determinant notation, which avoids excessive rewriting of the simple correlation coefficients. This notation can be extended to five or more variables, and also to the three-variable case previously considered.

In the three-variable case, the three-rowed determinant of correlation coefficients

$$\triangle_3 = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix} \qquad (1)$$

can be made to yield nine different two-rowed determinants by deleting one column and one row of $\triangle_3$. Suppose we call the two-row determinant obtained by deleting the first column and the first row $\triangle_{11}$ $(= 1 - r_{23}^2)$; that obtained by deleting the first column and the second row, $\triangle_{12}$ $(= r_{12} - r_{13}\, r_{23})$; and so on. The complete set of two-row determinants, which we call the $\triangle_{ij}$'s, is as follows:

$$\overset{3}{\triangle}_{11} = 1 - r_{23}^2 \tag{1.1}$$

$$\overset{3}{\triangle}_{22} = 1 - r_{13}^2 \tag{1.2}$$

$$\overset{3}{\triangle}_{33} = 1 - r_{12}^2 \tag{1.3}$$

$$\overset{3}{\triangle}_{12} = r_{12} - r_{13}\, r_{23} = \overset{3}{\triangle}_{21} \tag{1.4}$$

$$\overset{3}{\triangle}_{13} = -\,(r_{13} - r_{12}\, r_{23}) = \overset{3}{\triangle}_{31} \tag{1.5}$$

$$\overset{3}{\triangle}_{23} = (r_{23} - r_{12}\, r_{13}) = \overset{3}{\triangle}_{32} \tag{1.6}$$

All of the formulas for correlation and regression measures given in the preceding section for the three-variable case can be stated in terms of $\overset{3}{\triangle}$ and the $\overset{3}{\triangle}_{ij}$'s, as follows:

$$R^2_{1.23} = 1 - \frac{\overset{3}{\triangle}}{\overset{3}{\triangle}_{11}} \tag{2$'$}$$

$$\beta_{12.3} = \frac{\overset{3}{\triangle}_{12}}{\overset{3}{\triangle}_{11}} \tag{3$'$}$$

$$\beta_{13.2} = -\,\frac{\overset{3}{\triangle}_{13}}{\overset{3}{\triangle}_{11}} \tag{4$'$}$$

$$r_{12.3} = \frac{\overset{3}{\triangle}_{12}}{\sqrt{\overset{3}{\triangle}_{11} \cdot \overset{3}{\triangle}_{22}}} \tag{5$'$}$$

$$r_{13.2} = \frac{\overset{3}{\triangle}_{13}}{\sqrt{\overset{3}{\triangle}_{11} \cdot \overset{3}{\triangle}_{33}}} \tag{6$'$}$$

$$s_{\beta_{12.3}} = s_{\beta_{13.2}} = \frac{\sqrt{\overset{3}{\triangle}}}{\overset{3}{\triangle}_{11}\,\sqrt{N-3}} \tag{7$'$}$$

Once we have fixed the values of r_{12} and r_{13}, $\overset{3}{\triangle}$ and all but two of the $\overset{3}{\triangle}_{ij}$'s are functions only of r_{23}; these two, $\overset{3}{\triangle}_{22}$ and $\overset{3}{\triangle}_{33}$, are constants. Each of the six formulas in this paragraph involves $\overset{3}{\triangle}_{11}$, which changes with r_{23}, and another determinant which also changes with r_{23}. This means that we cannot vary r_{23} arbitrarily without consistently varying $\triangle_3$ and the $\triangle_{3\,ij}$'s.

In the four-variable case, the determinant of correlation coefficients is

$$\triangle_4 = \begin{vmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{12} & 1 & r_{23} & r_{24} \\ r_{13} & r_{23} & 1 & r_{34} \\ r_{14} & r_{24} & r_{34} & 1 \end{vmatrix} \tag{8}$$

The determinant of the three intercorrelation coefficients is

$$\triangle_{4\,11} = \begin{vmatrix} 1 & r_{23} & r_{24} \\ r_{23} & 1 & r_{34} \\ r_{24} & r_{34} & 1 \end{vmatrix} = 1 + 2\, r_{23}\, r_{24}\, r_{34} - r_{23}^{2} - r_{24}^{2} - r_{34}^{2} . \tag{9}$$

Each of the 15 other possible $\triangle_{4\,ij}$'s is now also a three-rowed determinant. The formulas in the preceding paragraph still apply, with an appropriate change in notation: For example,

$$\beta_{12.34} = \frac{\triangle_{4\,12}}{\triangle_{4\,11}} = \frac{\begin{vmatrix} r_{12} & r_{23} & r_{24} \\ r_{13} & 1 & r_{34} \\ r_{14} & r_{34} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{23} & r_{24} \\ r_{23} & 1 & r_{34} \\ r_{24} & r_{34} & 1 \end{vmatrix}} \tag{10}$$

All of the $\triangle_{4\,ij}$'s for which $i \neq j$ involve all three intercorrelation coefficients. $\triangle_{4\,22}$, $\triangle_{4\,33}$, and $\triangle_{4\,44}$ each contain only one of these coefficients. But this last point is not very helpful, as each formula (2) through (7) includes either $\triangle_4$ itself or a $\triangle_{4\,ij}$ for which $i \neq j$.

We noted in the three-variable case that the values assumed for r_{12} and r_{13} impose certain limits upon the values which might be assumed by r_{13}. Similarly, the values assumed for r_{12} and r_{14} impose restrictions on r_{24}, and those assumed for r_{13} and r_{14} impose restrictions on r_{34}. For example, if $r_{12} = r_{13} = r_{14} = 0.7$, each of the three intercorrelation coefficients may range from - 0.02 to 1.0. However, the values of r_{23} and r_{24} also set limits to the permissible values of r_{34}. A consistent set of limits for the six simple r's can be derived from the fact that $\triangle\!\!\!\!4_{22}$, $\triangle\!\!\!\!4_{33}$, $\triangle\!\!\!\!4_{44}$, and $\triangle\!\!\!\!4_{11}$ must all lie between 0 and 1

Discussion of Charts and Tables

Figures 5 through 10 and tables 7 through 13 provide some insights into the effects of intercorrelation in the four-variable case. The first five cases assume that all three of the direct correlation coefficients r_{12}, r_{13}, and r_{14} are equal to 0.7. Two of the intercorrelation coefficients, r_{23} and r_{24}, are then set equal to 0.9, 0.7, 0.5, 0.3 and 0.1, respectively. In each case, the third intercorrelation coefficient, r_{34}, is allowed to vary over its entire range of possible values given the values of the other five coefficients and the basic requirements $R^2_{1.234} \leq 1$ and $|r_{34}| \leq 1$.

The fact that we have set all three of the direct coefficients equal to one another, and two of the intercorrelation coefficients equal to each other, produces several symmetries in the results. One is that $\beta_{13.24}$ and $\beta_{14.23}$ are equal in each case. Another is that at the point where r_{34} is equal to r_{23} and r_{24}, all three beta coefficients are equal.

Figure 5 reflects a very high degree of intercorrelation. One symptom of this is the fact that $\triangle\!\!\!\!4_{11}$, the determinant of intercorrelation coefficients, takes on very small values-- 0.036 or less--over its entire range. In figure 10, in contrast, the value of $\triangle\!\!\!\!4_{11}$ reaches a peak of 1.0 when all three intercorrelation coefficients are zero, and exceeds 0.5 over a considerable range of values of r_{34}. In fact, figure 5 approaches the extreme of multicollinearity to which Frisch gave so much attention in the early thirties. The values of the beta coefficients are very unstable and are smaller than their standard errors in all but a small portion of the range of permissible values of r_{34}. And the range of permissible values of r_{34} is limited.

Figure 6, in which r_{23} and r_{24} equal 0.7, shows a greater stability of the beta coefficients with respect to given changes in r_{34} than does figure 5. The standard error of the beta coefficients is also more stable than in the preceding chart. The beta coefficients exceed their standard errors over a considerably wider range of values, although they do not reach twice the level of their standard errors anywhere in the permissible range. In both of these figures the behavior of $\beta_{12.34}$ corresponds to that of the weaker coefficient in some of the three-variable charts. When the level of r_{34} drops significantly below the levels of r_{23} and r_{24}, $\beta_{12.34}$ changes sign from positive to negative. Visually, it appears that $\beta_{13.24}$ ($= \beta_{14.23}$) is a reflection of $\beta_{12.34}$ about the particular level at which all three intercorrelation coefficients, and hence all three beta coefficients, are equal. The value of the betas at this point of equality increases from one case to the next as the paired intercorrelation coefficients (r_{23} and r_{24}) decrease.

Figure 7 shows still greater stability in the values of the beta coefficients and their standard errors. The beta coefficients exceed their standard errors over a large part of the permissible range and for certain limited values of r_{34}, near zero, they exceed two standard errors.

The data given in table 10 show increasing stability of the beta coefficients and their standard errors within the range of permissible values. However, that range itself is somewhat reduced. Apparently, as r_{23} and r_{24} are lowered, r_{34} must remain significantly above zero if the other constraints on the various correlation measures are to be met. If all three intercorrelation coefficients were zero the coefficient of multiple determination, $R_{1.234}^2$, should be equal to the sum of squares of the three direct correlation coefficients--in this case, $3 \times (0.7)^2 = 1.47$. As this is an impossible value of $R_{1.234}^2$, the value of r_{34} which leads to it is not permissible. While the ratios of the beta coefficients to their standard errors are greater than 1 over most of the range of permissible values, at no point does any one of these ratios exceed 2.0.

Figure 8 shows a still greater contraction of the range of permissible values of r_{34}. The values of the beta coefficients are quite stable within this limited range, but the standard errors of these coefficients is changing rapidly within it. The t-ratio, β/S_{β} , for $\beta_{12.34}$ exceeds 2.0 toward the lower end of the permissible range of r_{34}; t-ratios for the other beta coefficients do not exceed 1.3 at any value of r_{34}.

It is evident from the above results that to have each of the 3 direct correlation coefficients equal to 0.7 already constitutes a high degree of intercorrelation if one hopes to achieve significant regression coefficients in a four-variable equation involving only 20 or so observations.

In figure 9, the direct correlation coefficients are reduced to 0.5 and two of the intercorrelation coefficients are also set equal to 0.5. This chart may be compared with that of figure 6, in which all 5 of these coefficients were set equal to 0.7. The beta coefficients and their standard errors in figure 9 are considerably more stable and cover a wider range of permissible values than in figure 6. The coefficients exceed their standard errors over most of the permissible range, and the t-ratios for $\beta_{13.24}$ and $\beta_{14.23}$ exceed 2.0 over a sizable range, reaching a maximum of 2.83 when r_{34} reaches its lowest value.

In figure 10 the three direct correlation coefficients are again set equal to 0.5 and two of the intercorrelation coefficients are set at zero. The range of permissible values of r_{34} is about the same as in figure 9 but the degree of stability of the beta coefficients and their standard errors is considerably greater. The coefficient $\beta_{12.34}$ is independent of r_{34}. The t-ratios are greater than 1 over almost the full range of permissible values and exceed 2.0 over considerable portions of this range.

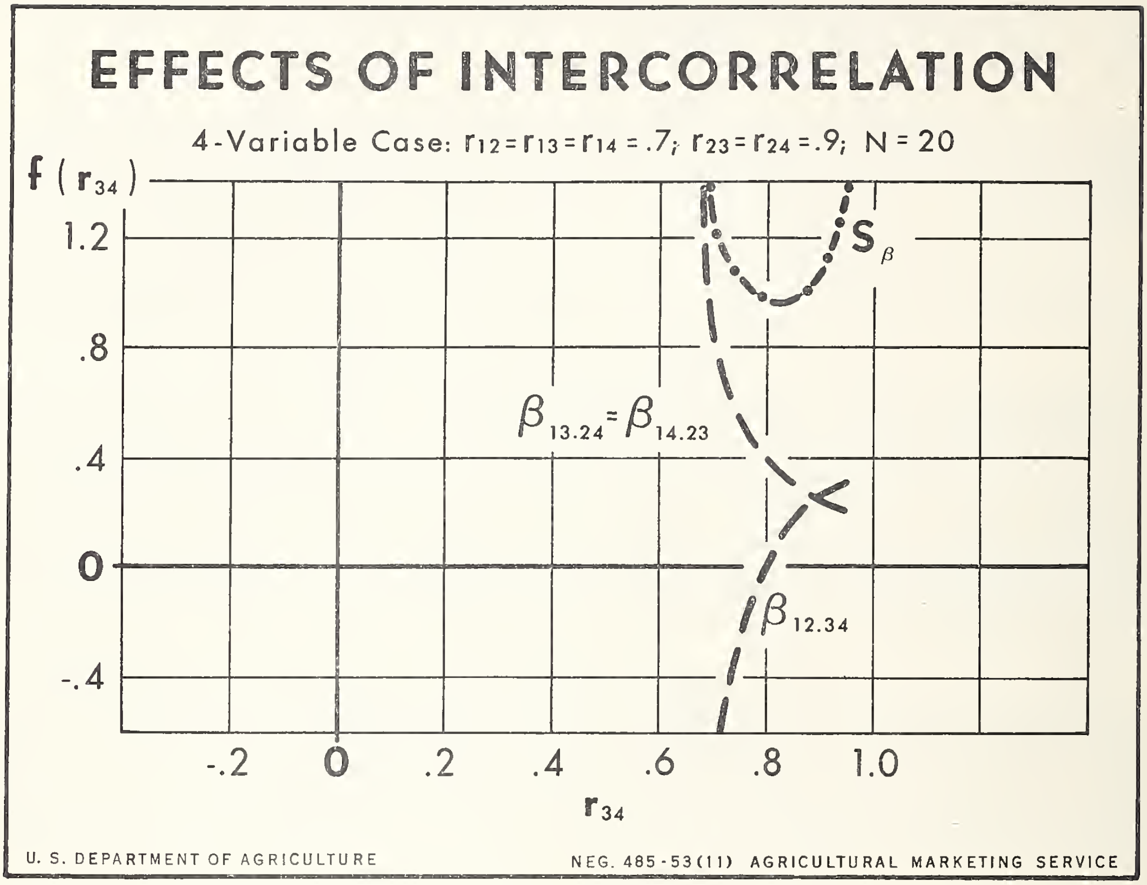

Figure 5

Table 7.- Data for case in which $r_{12} = r_{13} = r_{14} = 0.7$, $r_{23} = r_{24} = 0.9$, and N = 20

r_{34}	$\beta_{12.34}$	$\beta_{13.24}$	$\beta_{14.23}$	S_β 1/	t-ratio for - $\beta_{12.34}$	t-ratio for - $\beta_{13.24}$	t-ratio for - $\beta_{14.23}$
0.6392 2/	-5.8586	3.6436	3.6436	3.0000	-1.9529	1.2145	1.2145
.6500	-3.5000	2.3333	2.3333	2.2105	-1.5834	1.0556	1.0556
.6900	-1.1000	1.0000	1.0000	1.3474	-.8164	.7422	.7422
.6941	-1.0000	.9444	.9444	1.3102	-.7633	.7208	.7208
.7000	-.8750	.8750	.8750	1.2627	-.6930	.6930	.6930
.7500	-.2692	.5385	.5385	1.0429	-.2582	.5163	.5163
.7720	-.1290	.4605	.4605	1.0000	-.1290	.4605	.4605
.8000	0	.3889	.3889	.9721	0	.4000	.4000
.8500	.1522	.3044	.3044	.9825	.1549	.3098	.3098
.8645	.1847	.2862	.2862	1.0000	.1847	.2862	.2862
.9000	.2500	.2500	.2500	1.0827	.2309	.2309	.2309
.9500	.3182	.2121	.2121	1.4044	.2266	.1510	.1510
1.0000 3/	---	---	---	---	---	---	---

1/ Identical for each β.

2/ Lowest possible value of r_{34}.

3/ Highest possible value of r_{34}.

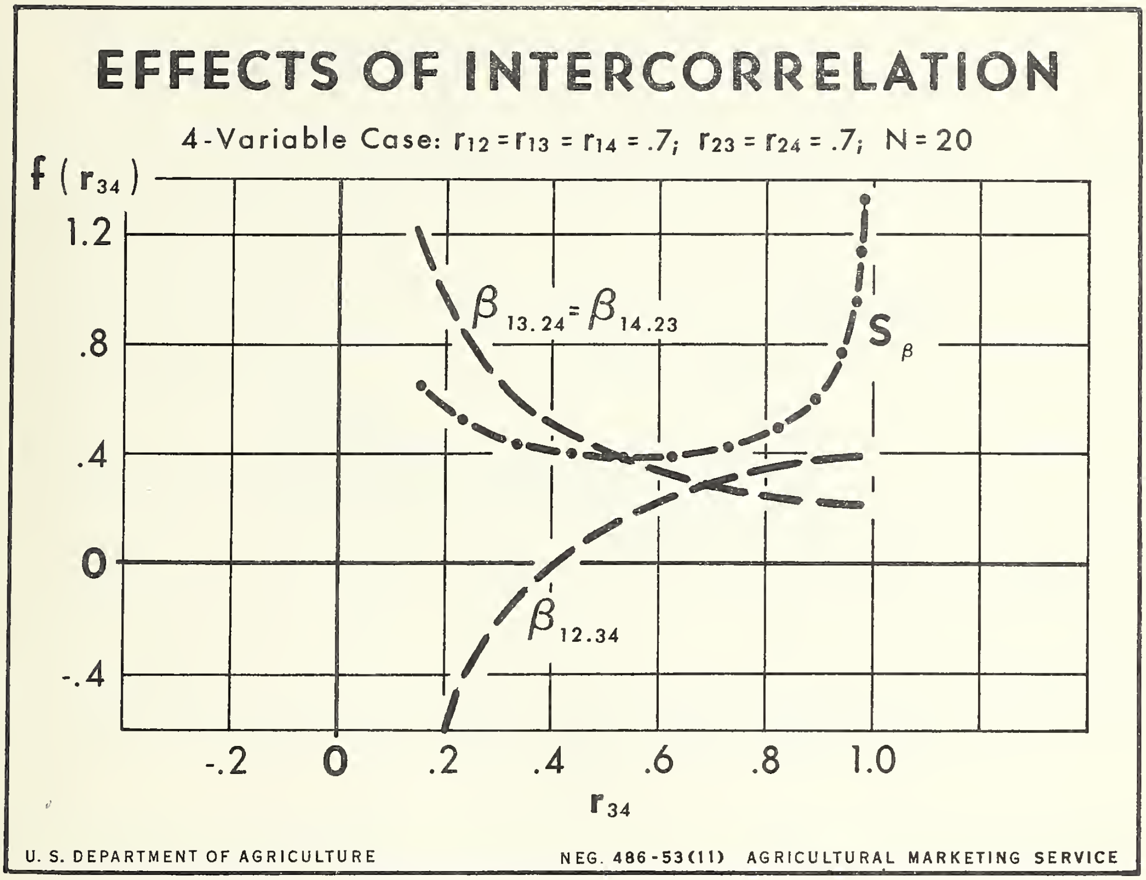

Figure 6

Table 8.- Data for case in which $r_{12} = r_{13} = r_{14} = 0.7$, $r_{23} = r_{24} = 0.7$, and N = 20

r_{34}	$\beta_{12.34}$	$\beta_{13.24}$	$\beta_{14.23}$	S_β 1/	t-ratio for - $\beta_{12.34}$	t-ratio for - $\beta_{13.24}$	t-ratio for - $\beta_{14.23}$
0.1529 2/	-1.0000	1.2143	1.2143	0.6532	-1.5310	1.8590	1.8590
.1900	- .7000	1.0000	1.0000	.5782	-1.2106	1.7294	1.7294
.2000	- .6364	.9546	.9546	.5625	-1.1314	1.6971	1.6971
.3000	- .2188	.6562	.6562	.4622	- .4733	1.4199	1.4199
.4000	0	.5000	.5000	.4167	0	1.2000	1.2000
.5000	.1346	.4038	.4038	.3982	.3381	1.0142	1.0142
.5765	.2071	.3521	.3521	.3973	.5214	.8862	.8862
.6000	.2258	.3387	.3387	.3992	.5657	.8485	.8485
.7000	.2917	.2917	.2917	.4210	.6928	.6928	.6928
.8000	.3415	.2561	.2561	.4772	.7155	.5367	.5367
.9000	.3804	.2283	.2283	.6309	.6030	.3618	.3618
.9592	.3998	.2145	.2145	.9520	.4199	.2253	.2253
.9632	.4010	.2136	.2136	1.0000	.4010	.2136	.2136
.9800	.4060	.2100	.2100	1.3440	.3021	.1562	.1562
1.0000 3/	---	---	---	---	---	---	---

1/ Identical for each β.

2/ Lowest possible value of r_{34}.

3/ Highest possible value of r_{34}.

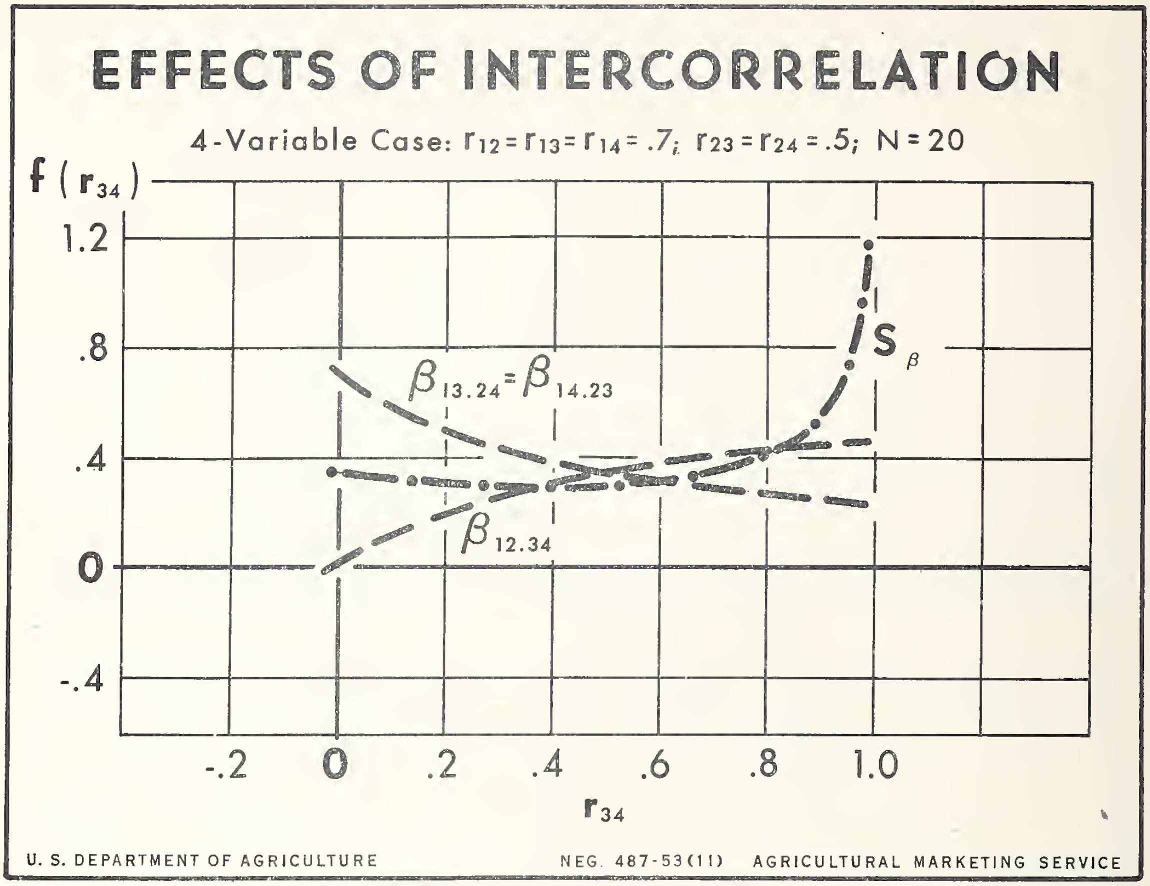

Figure 7

Table 9.- Data for case in which $r_{12} = r_{13} = r_{14} = 0.7$, $r_{23} = r_{24} = 0.5$, and N = 20

r_{34}	$\beta_{12.34}$	$\beta_{13.24}$	$\beta_{14.23}$	S_β 1/	t-ratio for - $\beta_{12.34}$	t-ratio for - $\beta_{13.24}$	t-ratio for - $\beta_{14.23}$
-0.0196 2/	-0.0286	0.7286	0.7286	0.3572	-0.0800	2.0396	2.0396
0	0	.7000	.7000	.3500	0	2.0000	2.0000
.1000	.1167	.5833	.5833	.3224	.3618	1.8091	1.8091
.2000	.2000	.5000	.5000	.3062	.6532	1.6330	1.6330
.2500	.2333	.4667	.4667	.3012	.7746	1.5492	1.5492
.3000	.2625	.4375	.4375	.2981	.8806	1.4676	1.4676
.4000	.3111	.3889	.3889	.2970	1.0474	1.3093	1.3093
.5000	.3500	.3500	.3500	.3031	1.1547	1.1547	1.1547
.6000	.3818	.3182	.3182	.3182	1.2000	1.0000	1.0000
.7000	.4083	.2917	.2917	.3472	1.1762	.8402	.8402
.8000	.4308	.2692	.2692	.4038	1.0667	.6667	.6667
.9000	.4500	.2500	.2500	.5449	.8259	.4588	.4588
.9500	.4586	.2414	.2414	.7538	.6084	.3202	.3202
.9800	.4635	.2365	.2365	1.1763	.3940	.2010	.2010
1.0000 3/	---	---	---	---	---	---	---

1/ Identical for each β.

2/ Lowest possible value of r_{34}.

3/ Highest possible value of r_{34}.

Table 10.- Data for case in which $r_{12} = r_{13} = r_{14} = 0.7$, $r_{23} = r_{24} = 0.3$, and N = 20

r_{34}	$\beta_{12.34}$	$\beta_{13.24}$	$\beta_{14.23}$	S_β 1/	t-ratio for - $\beta_{12.34}$	t-ratio for - $\beta_{13.24}$	t-ratio for - $\beta_{14.23}$
0.1216 2/	0.3877	0.5204	0.5204	0.2749	1.4103	1.8931	1.8931
.2000	.4118	.4804	.4804	.2713	1.5179	1.7707	1.7707
.3000	.4375	.4375	.4375	.2707	1.6162	1.6162	1.6162
.4000	.4590	.4016	.4016	.2747	1.6709	1.4620	1.4620
.5000	.4773	.3712	.3712	.2843	1.6789	1.3057	1.3057
.6000	.4930	.3451	.3451	.3019	1.6330	1.1431	1.1431
.7000	.5066	.3224	.3224	.3324	1.5241	.9699	.9699
.8000	.5185	.3025	.3025	.3895	1.3312	.7766	.7766
.9000	.5291	.2849	.2849	.5287	1.0008	.5389	.5389
1.0000 3/	---	---	---	---	---	---	---

1/ Identical for each β.
2/ Lowest possible value of r_{34}.
3/ Highest possible value of r_{34}.

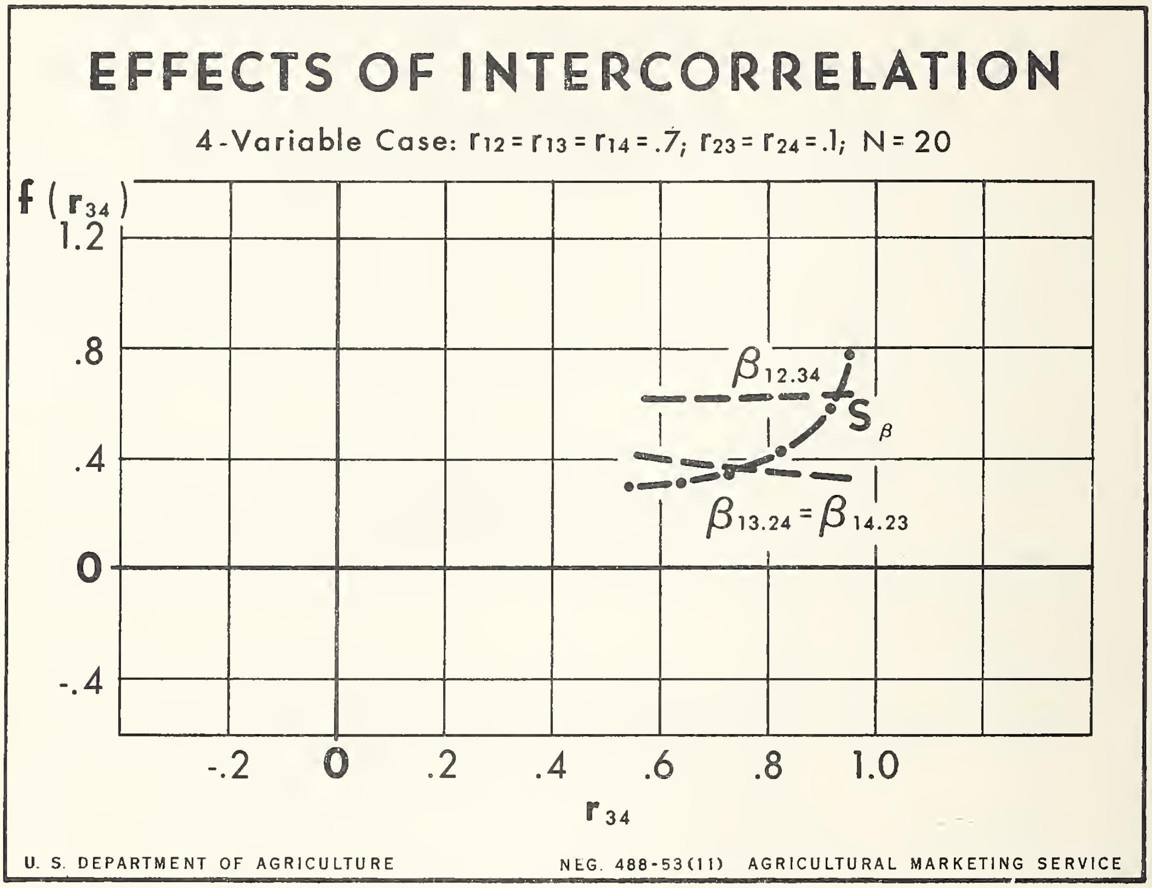

Figure 8

Table 11.- Data for case in which $r_{12} = r_{13} = r_{14} = 0.7$, $r_{23} = r_{24} = 0.1$, and N = 20

r_{34}	$\beta_{12.34}$	$\beta_{13.23}$	$\beta_{14.23}$	S_β 1/	t-ratio for -		
					$\beta_{12.34}$	$\beta_{13.24}$	$\beta_{14.23}$
0.5765 2/	0.6191	0.4048	0.4048	0.3079	2.0105	1.3145	1.3145
.6000	.6202	.3987	.3987	.3133	1.9799	1.2728	1.2728
.7000	.6250	.3750	.3750	.3455	1.0891	1.0854	1.0854
.8000	.6292	.3539	.3539	.4054	1.5522	.8731	.8731
.9000	.6330	.3351	.3351	.5507	1.1494	.6085	.6085
.9500	.6347	.3264	.3264	.7641	.8307	.4272	.4272
1.0000 3/	---	---	---	---	---	---	---

1/ Identical for each β.

2/ Lowest possible value of r_{34}.

3/ Highest possible value of r_{34}.

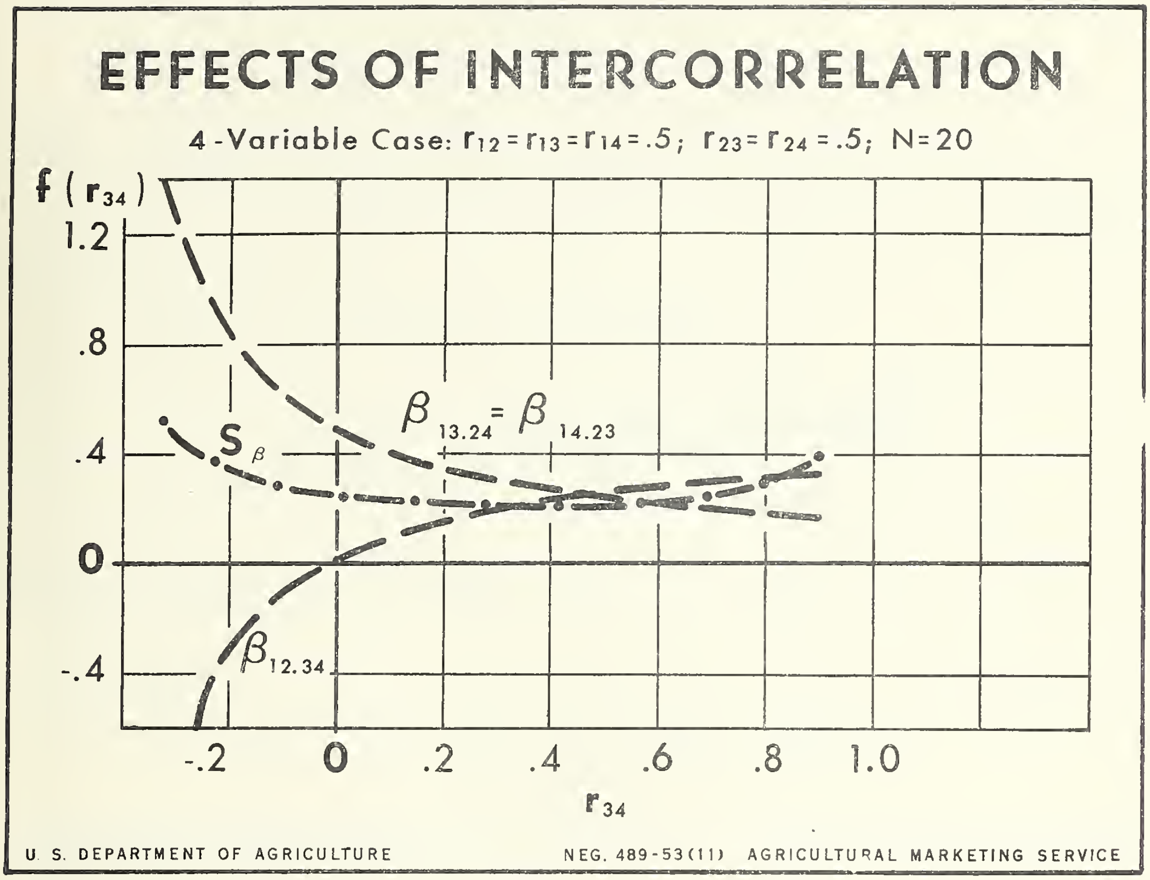

Figure 9

Table 12.- Data for case in which $r_{12} = r_{13} = r_{14} = 0.5$, $r_{23} = r_{24} = 0.5$, and $N = 20$

r_{34}	$\beta_{12.34}$	$\beta_{13.24}$	$\beta_{14.23}$	S_β 1/	t-ratio for - $\beta_{12.34}$	t-ratio for - $\beta_{13.24}$	t-ratio for - $\beta_{14.23}$
-0.3333 2/	-1.0000	1.4999	1.4999	0.5303	-1.8856	2.8285	2.8285
- .3000	- .7500	1.2500	1.2500	.4586	-1.6353	2.7255	2.7255
- .2500	- .5000	1.0000	1.0000	.3873	-1.2910	2.5820	2.5820
- .2000	- .3333	.8333	.8333	.3402	- .9798	2.4495	2.4495
- .1000	- .1250	.6250	.6250	.2827	- .4422	2.2111	2.2111
0	0	.5000	.5000	.2500	0	2.0000	2.0000
.1000	.0833	.4167	.4167	.2303	.3618	1.8091	1.8091
.2000	.1429	.3571	.3571	.2187	.6532	1.6329	1.6329
.3000	.1875	.3125	.3125	.2129	.8806	1.4676	1.4676
.4000	.2222	.2778	.2778	.2122	1.0474	1.3093	1.3093
.5000	.2500	.2500	.2500	.2165	1.1547	1.1547	1.1547
.6000	.2727	.2273	.2273	.2273	1.2000	1.0000	1.0000
.7000	.2917	.2083	.2083	.2480	1.1762	.8401	.8401
.8000	.3077	.1923	.1923	.2885	1.0667	.6667	.6667
.9000	.3214	.1786	.1786	.3892	.8259	.4588	.4588
1.0000 3/	---	---	---	---	---	---	---

1/ Identical for each β.

2/ Lowest possible value of r_{34}.

3/ Highest possible value of r_{34}.

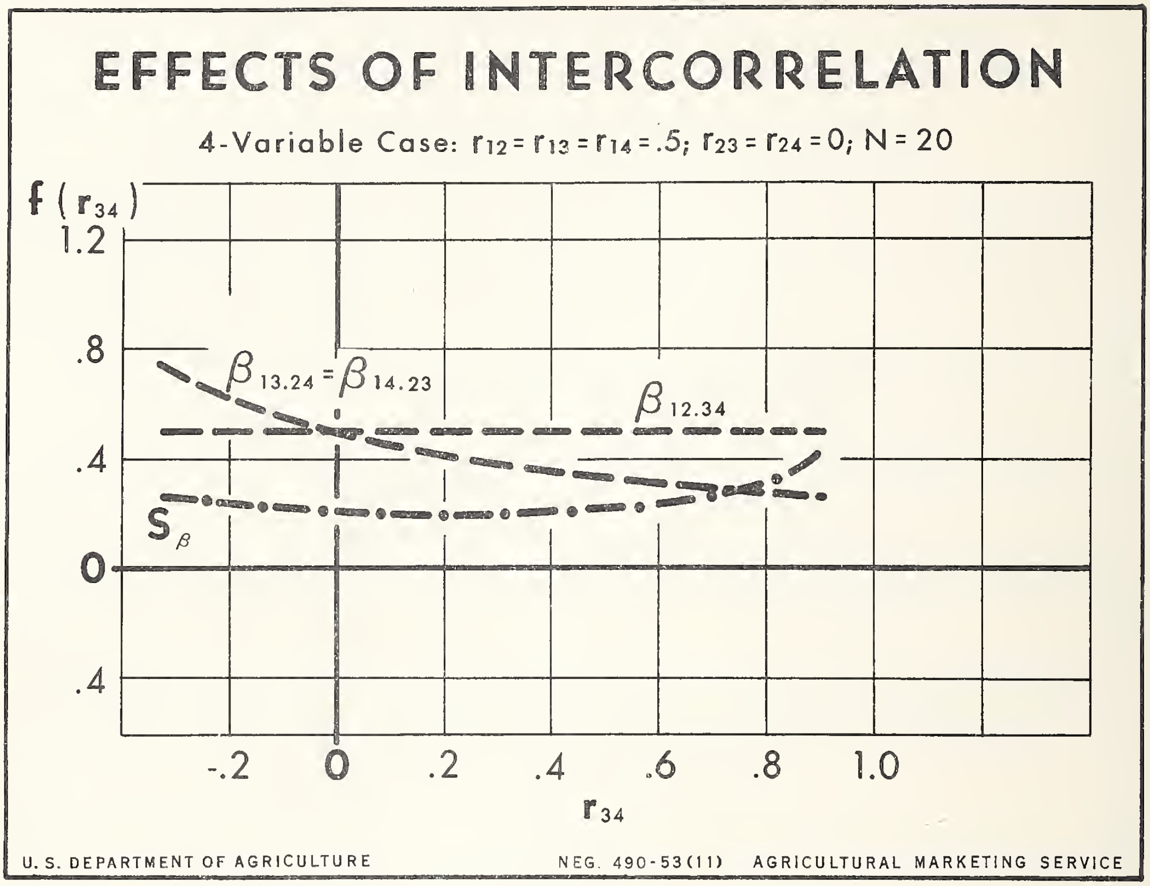

Figure 10

Table 13.- Data for case in which $r_{12} = r_{13} = r_{14} = 0.5$, $r_{23} = r_{24} = 0$, and $N = 20$

r_{34}	$\beta_{12.34}$	$\beta_{13.24}$	$\beta_{14.23}$	S_β 1/	t-ratio for - $\beta_{12.34}$	t-ratio for - $\beta_{13.24}$	t-ratio for - $\beta_{14.23}$
-0.3333 2/	0.5000	0.7500	0.7500	0.2652	1.8856	2.8284	2.8284
- .3000	.5000	.7143	.7143	.2574	1.9429	2.7756	2.7756
- .2500	.5000	.6667	.6667	.2472	2.0226	2.6968	2.6968
- .2000	.5000	.6250	.6250	.2387	2.0949	2.6186	2.6186
- .1000	.5000	.5556	.5556	.2255	2.2172	2.4636	2.4636
0	.5000	.5000	.5000	.2165	2.3094	2.3094	2.3094
.1000	.5000	.4546	.4546	.2109	2.3708	2.1553	2.1553
.2000	.5000	.4167	.4167	.2083	2.4000	2.0000	2.0000
.3000	.5000	.3846	.3846	.2088	2.3950	1.8423	1.8423
.4000	.5000	.3571	.3571	.2125	2.3525	1.6803	1.6803
.5000	.5000	.3333	.3333	.2205	2.2678	1.5118	1.5118
.6000	.5000	.3125	.3125	.2344	2.1333	1.3333	1.3333
.7000	.5000	.2941	.2941	.2582	1.9363	1.1390	1.1390
.8000	.5000	.2778	.2778	.3027	1.6518	.9177	.9177
.9000	.5000	.2632	.2632	.4109	1.2170	.6405	.6405
1.0000 3/	---	---	---	---	---	---	---

1/ Identical for each β.

2/ Lowest possible value of r_{34}.

3/ Highest possible value of r_{34}.

APPENDIX

Note 1 - Limits Imposed on Values of r_{23} by Given Values of r_{12} and r_{23}

By definition, no simple or multiple correlation coefficient can exceed 1 in absolute value.

Let us repeat text equation (2), as follows,

$$R^2_{1.23} = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}\, r_{13}\, r_{23}}{1 - r_{23}^2}, \qquad (2)$$

noting that $R^2_{1.23}$ must lie between 1 and 0.

If $R^2_{1.23} = 1$, we have

$$1 - r_{23}^2 = r_{12}^2 + r_{13}^2 - 2r_{12}\, r_{13}\, r_{23}. \qquad (2.1)$$

Rearranging terms, we have

$$r_{23}^2 - 2\, r_{12}\, r_{13}\, r_{23} + (r_{12}^2 + r_{13}^2 - 1) = 0 \qquad (2.2)$$

and, using a standard formula of elementary algebra, we obtain

$$r_{23} = r_{12}\, r_{13} \pm \sqrt{(1-r_{12}^2)\ (1-r_{13}^2)}. \qquad (2.3)$$

If $r_{12} = 0.9$ and $r_{13} = 0.7$, as in figure 3, we have

$$r_{23} = 0.63 \pm \sqrt{(0.19)\ (0.51)} = 0.63 \pm 0.3113;$$

hence, $r_{23} = 0.3187$ or 0.9413. Substituting these values back in equation (2), we obtain $R^2_{1.23} = 1$ in each case.

Only if $r_{12} = r_{13}$ can r_{23} reach the maximum value of 1. But when $r_{23} = 1$, $R^2_{1.23} = r^2_{12}$, which is, in general, less than 1. This is shown in figures 1 and 2.

Thus, if $r_{12} = r_{13} = 0.7$, equation (2.3) gives us

$$r_{23} = 0.49 \pm 0.51 = 1 \text{ or } - 0.02.$$

Substituting - 0.02 for r_{23} in equation (2) we obtain

$$R^2_{1.23} = \frac{0.98 - 0.98\,(-\,0.02)}{1 - (-0.02)^2} = \frac{0.9996}{0.9996} = 1 .$$

But if we substitute $r_{23} = 1$ we obtain

$$R_{1.23} = \frac{0.98 - 0.98(1)}{1 - (1^2)} = \frac{0}{0} ,$$

an indeterminate value.

This indeterminacy can be resolved by applying L'Hopital's rule, 3/ from which we obtain

$$R^2_{1.23} = \frac{-\,0.98}{-\,2(1)} = 0.49 \;(= r_{12}^{\;2}) .$$

<u>Note 2. Minimum Values of Specified Correlation and Regression Measures</u>

Assume that we are given the values of r_{12} and r_{13} and wish to obtain the values of r_{23} at which various coefficients reach their minimum values within the permissible ranges in which (1) $R^2_{1.23}$ is equal to or less than 1 and (2) r_{23} is equal to or less than 1 in absolute value. A necessary condition for a minimum is that the partial derivative with respect to r_{23} of a measure that is a function of r_{23} be zero.

1. Starting with equation (2),

$$R^2_{1.23} = \frac{r_{12}^{\;2} + r_{13}^{\;2} - 2r_{12}\, r_{13}\, r_{23}}{1 - r_{23}^{\;2}}, \qquad (2)$$

we find that $R^2_{1.23}$ reaches a minimum when

3/ See pp. 15-16 of Woods, Frederick S., Advanced Calculus, new ed., 397 pp., illus., New York, 1934, or other standard calculus texts.

we see that the sign is determined by the numerator only.

$$\text{If} \quad r_{23} = \frac{r_{13}}{r_{12}}, \; \beta_{13.2} = 0;$$

$$\text{if} \quad r_{23} < \frac{r_{13}}{r_{12}}, \; \beta_{13.2} > 0; \text{ and}$$

$$\text{if} \quad r_{23} > \frac{r_{13}}{r_{12}}, \; \beta_{13.2} < 0.$$

3. The minima for the partial correlation coefficients, $r_{12.3}$ and $r_{13.2}$, respectively are given at the points where

$$r_{23} = \frac{r_{13}}{r_{12}} \tag{5.1}$$

and

$$r_{23} = \frac{r_{12}}{r_{13}}. \tag{6.1}$$

In tables 3 through 6, $r_{13} < r_{12}$ and $r_{13.2}$ has no minimum in the permissible range of values for r_{23}. However, $r_{12.3}$ has a minimum at the level indicated by equation (5.1).

When $r_{12} = r_{13}$, $r_{12.3}$ and $r_{13.2}$ are equal and reach their lowest point in the permissible range when $r_{23} = 1$.

4. Finally, the minimum value of the standard errors of the beta coefficients can be obtained easily from formula (7),

$$S_{\beta_{12.3}} = S_{\beta_{13.2}} = \frac{\sqrt{1-R^2_{1.23}}}{(1-r_{23}{}^2)\sqrt{N-3}}. \tag{7}$$

When $R^2_{1.23} = 1$, $S_{\beta_{12.3}} = 0$ (provided $r_{23}{}^2$ is less than 1). Thus, in tables 3 through 6, $S_{\beta_{12.3}} = 0$ at two points, one at each end of the permissible range of values of r_{23}. In figures 1 and 2, $S_{\beta_{12.3}} = 0$ at one point, the

$$r_{23} = \frac{(r_{12}^2 + r_{13}^2) \pm (r_{12}^2 - r_{13}^2)}{2\, r_{12}\, r_{13}}, \tag{2.4}$$

that is, when $r_{23} = \frac{r_{12}}{r_{13}}$ or $\frac{r_{13}}{r_{12}}$

When $r_{12} = r_{13}$ (as in figures 1 and 2), $R^2_{1.23}$ reaches its minimum value when $r_{23} = 1$. When $r_{13} \neq r_{12}$, only one of the two values of r_{23} given by equation (2.4) will be less than 1, and hence a permissible value. In figure 3, we have $r_{23} = \frac{0.7}{0.9} = 0.78$; in the data shown in table 4, $r_{23} = 0.56$; and so on. It is clear that these are minimum rather than maximum values.

2. Using the same approach, the minima for the beta coefficients, $\beta_{12.3}$ and $\beta_{13.2}$, respectively, are given at the points where

$$r_{23} = \frac{r_{12} \pm \sqrt{r_{12}^2 - r_{13}^2}}{r_{13}} \tag{3.2}$$

and

$$r_{23} = \frac{r_{13} \pm \sqrt{r_{13}^2 - r_{12}^2}}{r_{12}}. \tag{4.2}$$

If $r_{12} = r_{13}$, the two beta coefficients are identical and reach their low point when $r_{23} = 1$. If $r_{13} > r_{12}$, the low point for $\beta_{12.3}$ is imaginary, while that for $\beta_{13.2}$ occurs at a value somewhat greater than $\frac{r_{13}}{r_{12}}$. The converse is true if $r_{12} > r_{13}$. In the data shown in table 4, for example, $\beta_{13.2}$ has no minimum value in the permissible range; $\beta_{12.3}$ has a minimum value at the point

$$r_{23} = \frac{0.9 \pm \sqrt{0.81 - 0.25}}{0.5} = \frac{0.900 \pm 0.748}{0.5} = 0.304 .$$

The second value of r_{23} is outside the permissible range.

A prominent feature of cases 3 to 6 is the fact that $\beta_{13.2}$ changes sign at a value of r_{23} somewhat greater than that of r_{13}. Referring to equation (4),

$$\beta_{13.2} = \frac{r_{13} - r_{12}\, r_{23}}{1 - r_{23}^2}, \tag{4}$$

lowest permissible value of r_{23}.

Figures 1 and 2 indicate that, if $r_{13} = r_{12}$, $S_{\beta_{12.3}}$ approaches infinity as r_{23} approaches 1. This follows readily from equation (7), since if $R^2_{1.23}$ is less than 1 at the point where $r_{23} = 1$, we have a real number divided by zero.

If we set $\frac{S_{\beta_{12.3}}}{r_{23}} = 0$ we obtain

$$r_{23}^3 - 3r_{12}\, r_{13}\, r_{23}^2 + [2(r_{12}^2 + r_{13}^2) - 1]\, r_{23} - r_{12}\, r_{13} = 0 \,. \qquad (7.3)$$

For given values of r_{12} and r_{13} this can be solved most readily by plotting the values of the function for a range of values of r_{23}. Thus, for table 5, the expression (7.3) becomes

$$r_{23}^3 - 0.81\, r_{23}^2 + 0.8\, r_{23} - 0.27 = 0 \,.$$

This equals zero when r_{23} is approximately 0.424. It is clear from the table that this is a maximum, or upper turning point, rather than a minimum.

Note 3. Relation of Correlation Formulas in Determinant Notations to the P_{ij} Table or Inverse Correlation Matrix

The elements of the inverse of a matrix of simple correlation coefficients may, in the three-variable case, be written as follows:

$$I_{(3)} = \begin{vmatrix} \frac{\Delta^{(3)}_{11}}{\Delta^{(3)}} & \frac{-\Delta^{(3)}_{12}}{\Delta^{(3)}} & \frac{\Delta^{(3)}_{13}}{\Delta^{(3)}} \\ \frac{-\Delta^{(3)}_{12}}{\Delta^{(3)}} & \frac{\Delta^{(3)}_{22}}{\Delta^{(3)}} & \frac{-\Delta^{(3)}_{23}}{\Delta^{(3)}} \\ \frac{\Delta^{(3)}_{13}}{\Delta^{(3)}} & \frac{-\Delta^{(3)}_{23}}{\Delta^{(3)}} & \frac{\Delta^{(3)}_{33}}{\Delta^{(3)}} \end{vmatrix}$$

The array of elements in the inverse is often referred to as " the P_{ij} table" in computation methods such as those developed by Waugh 4/. In this notation, $P_{11} = \frac{\Delta^{(3)}_{11}}{\Delta^{(3)}}$, $P_{12} = \frac{-\Delta^{(3)}_{12}}{\Delta^{(3)}}$, and so on.

4/ Waugh, F.V. A Simplified Method of Determining Multiple Regression Constants, Amer. Statis. Assoc. Jour. 30:694-700. 1935.

To calculate $\beta_{12.3}$ from the inverse or P_{ij} table, we divide $-P_{12}$ by P_{11}. This is equivalent to equation (3');

$$\beta_{12.3} = \frac{-P_{12}}{P_{11}} = \frac{\Delta^{3}_{12}}{\Delta^{3}} \cdot \frac{\Delta^{3}}{\Delta^{3}_{11}} = \frac{\Delta^{3}_{12}}{\Delta^{3}_{11}}. \qquad (3')$$

All the other formulas in the text can be derived in the same fashion from the P_{ij} table. For example.

$$R^2_{1.23} = 1 - \frac{1}{P_{11}} = 1 - \frac{\Delta^{3}}{\Delta^{3}_{11}}, \qquad (2')$$

$$r_{12.3} = \frac{-P_{12}}{\sqrt{P_{11}}\cdot\sqrt{P_{22}}} = \frac{\dfrac{\Delta^{3}_{12}}{\Delta^{3}}}{\dfrac{\sqrt{\Delta^{3}_{11}\cdot\Delta^{3}_{22}}}{\Delta^{3}}} = \frac{\Delta^{3}_{12}}{\sqrt{\Delta^{3}_{11}\cdot\Delta^{3}_{22}}}, \qquad (5')$$

and so on.

The text formulas in determinant notation can be generalized for any number of variables. The same is true of the corresponding formulas in the P_{ij} notation. Thus,

$$\beta_{12.34} = \frac{-P_{12}}{P_{11}} = \frac{\Delta^{4}_{12}}{\Delta^{4}} \cdot \frac{\Delta^{4}}{\Delta^{4}_{11}} = \frac{\Delta^{4}_{12}}{\Delta^{4}_{11}}, \qquad (10)$$

where P_{11} and P_{12} are the first and second elements in the first row of a 4-rowed determinant,

$$I_{(4)} = \left| \frac{\Delta^{4}_{ij}}{\Delta^{4}} \right| , \quad i, j = 1, \ldots. 4.$$