Cowles Foundation Paper

Econometrica, Vol. 30, No. 2 (April, 1962)

PARTIAL TRACE CORRELATIONS

BY JOHN W. HOOPER[1]

1. INTRODUCTION

THIS PAPER is a sequel to [**5**] in which a generalized correlation coefficient (the trace correlation) for multi-equation models was presented. This paper is concerned with the development of partial trace correlations. The relation between partial trace correlations and the trace correlation is analogous to the relation between partial correlation coefficients and the multiple correlation coefficient in a single-equation model.[2] In the latter the multiple correlation coefficient measures the extent to which the regression relationship on all of the independent variables accounts for the observed variation in the dependent variable. In addition there are partial correlation coefficients which measure the extent to which the regression on a particular independent variable explains the observed variation in the dependent variable, holding the influence of the other independent variables constant. Similarly in multi-equation models there is the trace correlation[3] which is a measure of the extent to which the regression relationship explains the variation in the set of jointly dependent variables. The purpose of this paper is to develop the analogue of partial correlation coefficients for multi-equation models, i.e., to develop a measure of the degree to which the regression relationship on a subset of the independent variables accounts for the variation in the jointly dependent variables, while holding the influence of the other independent variables constant.

As in the case of the trace correlation the basic concepts are developed through the use of canonical correlation theory. This is done in Section 2, and in Section 3 partial trace correlations are defined. In Section 4 the asymptotic sampling variances are given.

2. CANONICAL CORRELATION THEORY APPLIED TO CONDITIONAL SETS OF VARIABLES

Let there be two sets of variables $y_1, \ldots, y_M$ and $x_1, \ldots, x_\Lambda$, where there are T observations on each of these variables available, so that the aggregate set of data can be written in the form of a $T \times M$ matrix Y and a $T \times \Lambda$

[1] I am indebted to Professors K. J. Arrow, H. Chernoff, and W. Fisher for their constructive comments on an earlier draft of this paper. Any errors which remain are the sole responsibility of the author. This paper is based on Ch. IV in [**4**].

[2] A clear account of this relationship is given in [**1**, pp. 27–34].

[3] The reader is referred to [**5**] for the development of the trace correlation.

matrix X.[4] Partition the X matrix into

$$X = [X_1 \; X_2] \tag{2.1}$$

where X_1 has Λ_1 columns of T elements each and X_2 has Λ_2 columns ($\Lambda_1 + \Lambda_2 = \Lambda$). In order to define the canonical partial correlations we first consider the linear regression of Y on X_2 and X_1 on X_2 which can be represented as

$$\begin{aligned} Y &= X_2\beta_1 + \bar{V}_1 , \\ X_1 &= X_2\beta_2 + \bar{V}_2 , \end{aligned} \tag{2.2}$$

where β_1 and β_2 are unknown parameter matrices of order $\Lambda_2 \times M$ and $\Lambda_2 \times \Lambda_1$, respectively, and where $\bar{V}_1 = [v_{t\mu}]$ and $\bar{V}_2 = [v_{t\lambda_1}]$ are $T \times M$ and $T \times \Lambda_1$ matrices of parent disturbances.[5] We then find that the estimated relationships in (2.2) are

$$\begin{aligned} Y &= X_2\hat{\beta}_1 + V_1 , \\ X_1 &= X_2\hat{\beta}_2 + V_2 , \end{aligned} \tag{2.3}$$

where $\hat{\beta}_1 = (X_2'X_2)^{-1}X_2'Y$ and $\hat{\beta}_2 = (X_2'X_2)^{-1}X_2'X_1$ are the least squares estimators and V_1 and V_2 are matrices of estimated residuals. The sample canonical correlations may be regarded as the zero-order canonical correlations between the two conditioned sets of variables $V_1 = Y - X_2\hat{\beta}_1$ and $V_2 = X_1 - X_2\hat{\beta}_2$. The squared canonical correlations, all of which lie between zero and unity,[6] are the latent roots of the matrix $C_1^{-1}W'C_2^{-1}W$ where

$$\begin{aligned} C_1 &= [Y'Y - Y'X_2(X_2'X_2)^{-1}X_2'Y] , \\ C_2 &= [X_1'X_1 - X_1'X_2(X_2'X_2)^{-1}X_2'X_1] , \\ W &= [X_1'Y - X_1'X_2(X_2'X_2)^{-1}X_2'Y] . \end{aligned} \tag{2.4}$$

To apply these results to simultaneous equation systems we consider the following reduced-form system of equations,

$$Y = X\Pi + \bar{V} \tag{2.5}$$

where Π is the matrix of parent reduced-form coefficients and $\bar{V}$ the $T \times M$

[4] We also assume that Y is of rank M and X of rank Λ and that all variables are measured as deviations from their means.

[5] It is further assumed that the random terms of $\bar{V}_1$ and $\bar{V}_2$ have the following usual properties:

(i) $$Ev_{t\mu} = Ev_{t\lambda_1} = 0$$

(ii) $$E(v_{t\mu}v_{t'\mu'}) = \begin{cases}\sigma_{\mu\mu'} , \\ 0 ,\end{cases} \qquad E(v_{t\lambda_1}v_{t'\lambda_1'}) = \begin{cases}\sigma_{\lambda_1\lambda_1'} , & t = t' , \\ 0 , & t \neq t' ,\end{cases}$$

for t, $t' = 1, \ldots, T$, $\mu, \mu' = 1, \ldots, M$, and $\lambda_1', \lambda_1 = 1, \ldots, \Lambda_1$.

[6] Cf. [9] for a statement of these results.

matrix of parent reduced-form disturbances. We partition the Π matrix to conform to the partitioning of the X matrix and obtain

$$Y = X_1\Pi_1 + X_2\Pi_2 + \bar{V}\,, \tag{2.6}$$

where Π_1 and Π_2 are $\Lambda_1 \times M$ and $\Lambda_2 \times M$ matrices of reduced-form coefficients, respectively. The estimates of these coefficients, given by P_1 and P_2, as obtained by the method of least squares, are

$$\begin{aligned} P_1 &= C_2^{-1}W\,, \\ P_2 &= (X_2'X_2)^{-1}[X_2'Y-(X_2'X_1)C_2^{-1}W]\,. \end{aligned} \tag{2.7}$$

The estimated reduced-form equations can then be written as

$$Y = X_1P_1 + X_2P_2 + V\,, \tag{2.8}$$

where V is a $T \times M$ matrix of calculated residuals. It then follows after multiplication and some simplification that the estimated moment matrix of the jointly dependent variables can be written as

$$Y'Y = W'C_2^{-1}W + Y'X_2(X_2'X_2)^{-1}X_2'Y + V'V \tag{2.9a}$$

or[7]

$$C_1 = W'C_2^{-1}W + V'V\,. \tag{2.9b}$$

The left-hand side of (2.9b), C_1, is the estimated conditional (conditioned on X_2) moment matrix of the jointly dependent variables which means that it is the variation left unexplained by the regression of Y on X_2. We notice from the right-hand side of (2.9b) that this unexplained variation in the y's given X_2 can be divided into two parts, viz., $W'C_2^{-1}W$, which can be interpreted as that part of the variation of the variables in V_1 which would be explained by the regression of V_1 on V_2, and $V'V$, which is the observed moment matrix of unexplained residuals. However, C_2 is also a conditional moment matrix since it represents that part of the variation in the x_1's which is not accounted for by the regression of X_1 on X_2. Thus we see that $W'C_2^{-1}W$ represents in matrix form that part of the variation in the jointly dependent variables which is explained by the regression relationship on the exogenous variables contained in the set X_1, after the influence of the exogenous variables contained in the set X_2 has been eliminated from both Y and X_1.

It now easily follows from (2.9b) that

$$C_1^{-1}W'C_2^{-1}W = I - F\,, \tag{2.10}$$

[7] See (2.4) for the definition of C_1.

where I is the unit matrix of order M and F is defined as

(2.11) $$F = C_1^{-1} V' V\,.$$

This result implies that

(2.12) $$|(I-F) - r^2 I| = 0$$

and from (2.10) and (2.11) that

(2.13) $$|F - (1-r^2)I| = 0\,.$$

The matrix $C_1^{-1} W' C_2^{-1} W = I - F$ may be regarded as a matrix generalization[8] of the ratio of a product of the conditional variances of a single dependent variable and an independent variable to the squared conditional covariance of these variables in a single equation model. This ratio is equal to $r_{yj\cdot q}^2$ where $r_{yj\cdot q}$ is the partial correlation coefficient between y, the dependent variable, and the jth independent variable, holding the other q independent variables fixed.[9] It follows from (2.12), however, that the latent roots of $I-F$ are r_μ^2 where r_μ^2 is a canonical partial correlation between the y's and the x_1's given the X_2's. Thus we obtain both the matrix generalization, $I-F$, of $r_{yj\cdot q}^2$ and the vector representation, i.e., the vector $[r_\mu^2]$ of M or Λ_1 elements (M if $M \leq \Lambda_1$ and Λ_1 if $M \geq \Lambda_1$) which is the vector of latent roots of $I-F$. By similar reasoning the matrix $F = C_1^{-1} V' V$ may be regarded as the matrix generalization of $1 - r_{yj\cdot q}^2$ and the vector of latent roots of F, i.e., $[1 - r_\mu^2]$, as the vector generalization.

3. PARTIAL TRACE CORRELATIONS

In this section we shall develop a scalar measure of the degree to which the conditioned exogenous variables $[X_1 - X_2(X_2'X_2)^{-1}X_2'X_1]$ account for the variation in the conditioned jointly dependent variables $[Y - X_2(X_2'X_2)^{-1} X_2'Y]$. This scalar measure will be some function f of $I-F$ which takes this matrix into a scalar and may be regarded as a generalization of partial correlation coefficients in single equation models. In order to preserve the essential properties which partial correlation coefficients have, the function f should have the following properties:

(i) $$f[P(I-F)P^{-1}] = f(I-F)$$

(ii) $$f(I-F) + f(F) \equiv 1$$

(iii) $$0 \leq f(I-F) \leq 1\,,$$

where P is a nonsingular transformation matrix. Property (i) ensures that the value of the function f is invariant to the units in which the x and y

[8] See [8, p. 1075] for a similar generalization.

[9] Cf. [1, p. 29].

variables are measured. It also means that we can restrict our attention to functions of those characteristics of $I-F$ which are invariant under similar transformations, i.e., functions of the latent roots[10] of $I-F$. Property (ii) is the natural requirement that the scalar measure of what has been explained of the observed variation in the conditional y's plus what remains unexplained should sum to unity. Property (iii) is that this measure should be between zero and unity.

One such function possessing these properties, and the one we shall adopt, is the trace function. Accordingly we define the *partial trace correlation* to be the positive square root of $\bar{r}^2_{yx_1\cdot x_2}$ where

$$\bar{r}^2_{yx_1\cdot x_2} = (1/M)\ \mathrm{tr}\ (I-F) = (1/M) \sum_{\mu=1}^{M} r_\mu^2 \,. \tag{3.1}$$

We also have that[11]

$$1-\bar{r}^2_{yx_1\cdot x_2} = (1/M)\ \mathrm{tr}\ F = (1/M) \sum_{\mu=1}^{M} (1-r_\mu^2)\ . \tag{3.2}$$

Considering the discussion of $[r_\mu^2]$ as given in Section 2 we may conclude that $\bar{r}^2_{yx_1\cdot x_2}$ can be interpreted as a scalar measure of that part of the variation in the set of conditional jointly dependent variables $(Y|X_2)$, say, that is explained by the systematic part of the regression of $(Y|X_2)$ on a subset of the conditional independent variables $(X_1|X_2)$. Similarly $1-\bar{r}^2_{yx_1\cdot x_2}$ can be interpreted as a measure of that part of the variation in $(Y|X_2)$ which is left unexplained by the regression of $(Y|X_2)$ on $(X_1|X_2)$.

Some special cases of the partial trace correlation are of interest.[12]

When $M = \Lambda_1 = 1$, $\Lambda_2 \geq 1$ the partial trace correlation becomes the single equation partial correlation coefficient. When $\Lambda_2 = 0$ the partial trace correlation becomes the trace correlation and then for $M = \Lambda_1 = 1$ we have the zero-order correlation coefficient (disregarding sign) and for $\Lambda > M = 1$ we have the multiple correlation coefficient. For the case of $M = 1, \Lambda_1 > 1, \Lambda_2 \geq 1$ we have the *multiple partial correlation coefficient.*[13]

One situation in which the partial trace correlation might prove useful is as an aid in the selection of exogenous variables to be used in the model. This is the situation in which either $\Lambda > T$, in which case $(X'X)$ is singular so that neither the reduced-form or the structural parameters can be estimated, or of the case when $T-\Lambda$ is positive but small so that there is but

[10] See [2, pp. 313–314].

[11] Since $\bar{r}^2_{yx_1\cdot x_2}$ is an arithmetic average of the latent roots of $I-F$, each of which is between zero and unity, it is easily verified that $\bar{r}^2_{yx_1\cdot x_2}$ satisfies the three properties discussed above.

[12] Cf. [9] for a listing of these cases.

[13] See [9]. This is the term used by these authors. There seems to be very little discussion of this potentially useful concept in the literature.

a small number of degrees of freedom for estimating the parameters.[14] In these situations the usual procedure is to perform the estimation using only some subset of the originally selected set of variables. So the problem becomes that of selecting which exogenous variables are to be retained. Although there are several ways to select these variables, all to some extent arbitrary, it would seem that one way which possesses some merit would be to use the partial trace correlations in order to determine those exogenous variables which contribute the most to the explanation of the observed variation in the jointly dependent variables. Then in descending order of importance enough exogenous variables can be selected so as to make the estimation of the parameters possible (when $\Lambda > T$) or more effective (when $T - \Lambda > 0$ but small).

There are also other situations in which one must neglect some of the exogenous variables[15] and in these situations the partial trace correlations and the full trace correlation might be used to determine which variables can be omitted so as to make the unexplained variation ($V'V$) as "small" as possible.[16]

4. THE ASYMPTOTIC SAMPLING VARIANCE

In order to obtain some idea of the significance of a particular partial trace correlation computed from a sample it is desirable to have at least the asymptotic sampling variance. In this section we shall derive this variance to the order of T^{-1} under the assumption that each of the T rows of $[YX_1X_2]$ is an independent random drawing from a $[M+\Lambda]$-dimensional normal parent distribution with zero means.[17] From the definition of the partial trace correlation (3.1) we have

$$M\bar{r}^2_{yx_1 \cdot x_2} = \sum_{\mu=1}^{M} r^2_\mu . \tag{4.1}$$

Taking differentials we obtain

$$M\bar{r}_{yx_1 \cdot x_2} d\bar{r}_{yx_1 \cdot x_2} = \sum_{\mu=1}^{M} r_\mu dr_\mu . \tag{4.2}$$

Squaring and taking expected values we find for large samples that

$$M^2 \bar{\varrho}^2_{yx_1 \cdot x_2} \operatorname{var} \bar{r}_{yx_1 \cdot x_2} = \sum_{\mu,\mu'} \varrho_\mu \varrho_{\mu'} \operatorname{cov} (r_\mu, r_{\mu'}) , \tag{4.3}$$

[14] An excellent discussion of this problem is given in [**7**].

[15] See [**3**, pp. 203–204].

[16] Cf. [**3**, p. 204] for a discussion of the "size" of ($V'V$).

[17] We shall also assume that there are no multiple roots in the population, except possibly zero multiple roots caused by the fact that $\Lambda_1 < M$, which are allowed.

where $\bar{\varrho}^2_{yx_1 \cdot x_2} = (1/M)\mathrm{tr}(I - \Phi)$ is the square of the parent partial trace correlation. Now the problem is solved if we can derive var r_μ and cov $(r_\mu, r_{\mu'})$. These can be derived to the order of T^{-1} but we can eliminate this somewhat tedious derivation by noticing that the two sets of variables $V_1 = Y - X_2\,(X_2'X_2)^{-1}X_2'Y$ and $V_2 = X_1 - X_2(X_2'X_2)^{-1}X_2'X_1$, have the same properties as unconditioned sets of variables. This means that the *large sample* variances and covariances of the canonical correlations between V_1 and V_2 are the same as those between unconditioned sets of variables and so we have the well-known result[18] that

$$\text{(4.4)} \qquad \mathrm{var}\, r_\mu = (1/T)(1 - \varrho_\mu^2)^2\,; \qquad \mathrm{cov}\,(r_\mu, r_{\mu'}) = 0 \qquad \text{for } \mu \neq \mu'\,,$$

to the order[19] of T^{-1}. So we obtain

$$\text{(4.5)} \qquad \mathrm{var}\, \bar{r}_{yx_1 \cdot x_2} = \frac{1}{TM^2\, \bar{\varrho}^2_{yx_1 \cdot x_2}} \sum_{\mu=1}^{M} \varrho_\mu^2 (1 - \varrho_\mu^2)^2$$

to our degree of approximation. Furthermore since
$\mathrm{var}\, \bar{r}^2_{yx_1 \cdot x_2} = 4\bar{\varrho}^2_{yx_1 \cdot x_2}\, \mathrm{var}\, \bar{r}_{yx_1 \cdot x_2}$, we also obtain

$$\text{(4.6)} \qquad \mathrm{var}\, \bar{r}^2_{yx_1 \cdot x_2} = \frac{4}{TM^2} \sum_{\mu=1}^{M} \varrho_\mu^2 (1 - \varrho_\mu^2)^2\,.$$

5. FURTHER COMMENTS ON PARTIAL TRACE CORRELATIONS

The concept of the partial trace correlation is but a logical extension of the trace correlation. Thus the same problem of non-uniqueness when definitional equations are present and of using more efficient methods of estimation than unrestricted least squares, as discussed in Section 6 of [**5**] occur. As for the latter problem it can be shown that the partial trace correlation can be extended to take account of a method of estimation such as limited-information or two-stage least squares.[20]

Cowles Foundation for Research in Economics at Yale University

[18] See [**6**, p. 340].

[19] The fact that the large sample variance of canonical partial correlations is the same as that of canonical correlations should cause no surprise since the same result holds as between the finite sample distributions of canonical partial and canonical correlations. More precisely, the distributions of canonical partial correlations in random samples of size $T + 1$ from a $(M + \Lambda_1 + \Lambda_2)$ variate normal population are of the same form as those of canonical correlations in random samples of size $T + 1 - \Lambda_2$. See [**10**] for this result. Thus one might expect that the denominator in (4.4) should be $T - \Lambda_2$. However, as (4.4) is only true to the order of T^{-1}, the denominator causes no error.

[20] See Section 5, Chapter II of [**4**] for this extension. This remark applies only to the estimation of partial trace correlations and not to their sampling distributions.

REFERENCES

[1] ANDERSON, T. W.: *Introduction to Multivariate Statistical Analysis*, New York: John Wiley and Sons, 1958.

[2] BIRKOFF, G., AND S. MACLANE: *A Survey of Modern Algebra*. (Revised ed.), New York: The Macmillan Co., 1953.

[3] CHERNOFF, H., AND H. RUBIN: "Asymptotic Properties of Limited-Information Estimates Under Generalized Conditions," *Studies in Econometric Method*. Edited by T. C. Koopmans and Wm. C. Hood, New York: John Wiley and Sons, 1953.

[4] HOOPER, J. W.: *Correlation Theory as Applied to Simultaneous Equations*, (Unpublished Ph. D. thesis) Stanford University, 1961.

[5] ———: "Simultaneous Equations and Canonical Correlation Theory," *Econometrica*, XXVII (1959), pp. 245–256.

[6] HOTELLING, H.: "Relations Between Two Sets of Variates," *Biometrika*, XXVIII (1936), pp. 321–377.

[7] KLOEK, T., AND L. B. M. MENNES: "Simultaneous Equations Estimation Based on Principal Components of Predetermined Variables," *Econometrica*, XXVIII (1960), pp. 45–61.

[8] LAURENT, A. G.: "A Matrix Definition of the Correlation Between Two Sets of Variables," (Abstract), *Annals of Mathematical Statistics*, XXVIII (1957), pp. 1074–1075.

[9] ROY, S. N., AND J. WHITTLESEY: "Canonical Partial Correlations," (Abstract), *Annals of Mathematical Statistics*, XXIII (1952), p. 642.

[10] ROY, S. N.: "Distribution of Canonical Partial Correlations," (Abstract), *Annals of Mathematical Statistics*, XXIII (1952), p. 647.