KM34(1990)
pag 65 - 85

REGRESSION MODELS FOR CHANNELING

John P. van de Geer*

ABSTRACT

Given a variable y_1 and another set of variables Y_2. It is assumed that variation in y_1 depends primarily on variation in variables X_1, and that Y_2 depends primarily on X_2.
The problem is: is it possible to channel the effect of X_2 on y_1 through Y_2? This would imply that prediction of y_1 on the basis of (X_1, Y_2) does not become better if y_1 is predicted from (X_1, X_2, Y_2).

Such channeling solutions have been used mainly in econometrics, but could be applicable in many other fields of research.

The paper discusses: under what conditions is a perfect solution for channeling feasible (called the "just identifiable" case). This is followed by a discussion of two methods which can be applied if a perfect solution is not feasible (the "overidentified" case). They are: two-stages least squares (2SLS) and least generalized residual variance (LGRV). The two methods are explained in terms of algebraic and geometrical properties of vectors (not in terms of their mathematical-statistical properties).

The main focus of the paper is on the situation where Y_2 contains only one variable y_2, and where not only the effect of X_2 on y_1 is channeled through y_2, but also the effect of X_1 on y_2 is channeled through y_1 (non-recursive path diagram). A numerical example is added.

KEY WORDS

identifiability
least generalized residual variance
multiple regression
non-recursive path diagram
two stages least squares

---

* Prof.dr. J.P. van de Geer
Rijksuniversiteit Leiden
Vakgroep Datatheorie, FSW
Wassenaarseweg 52
2333 AK Leiden
Tel. 071-273829

1 STATEMENT OF THE PROBLEM

1.1 Path diagrams

In a path diagram, variables usually are pictured as little circles. Arrows between those circles are used to indicate that some variable has an effect on some other variables. See Figure 1A for a simple example. The figure just says that y_1 and y_2 depend upon the variables in X; the figure represents a multiple regression solution of y_1 and y_2 on X. In the figure, X has been partitioned into two subsets $X = (X_1, X_2)$. The reason for this partitioning needs not concern us for the moment; it will be explained in the sequel.

Obviously, the purpose of a path diagram is to explain the intercorrelations between variables in terms of a simple "model". A simple model requires that one should not draw more arrows than necessary. In this paper we study the possibility of simplifying a model by means of "channeling".

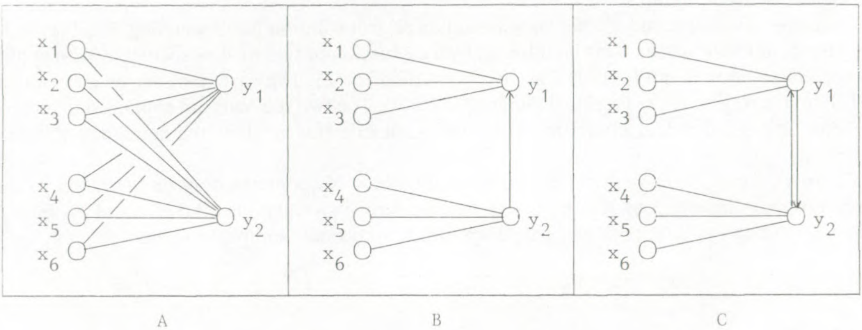

Figure 1 Figure A shows path diagrams for regression of y_1 and of y_2 on $X = (x_1, .., x_6)$. The lines in the diagram are arrows pointing towards y_1 or to y_2. Figure B gives the diagram in which the effect of $X_2 = (x_4\ x_5\ x_6)$ on y_1 is channeled through y_2. Figure C contains the diagram of Figure B but shows in addition the diagram where the effect of $X_1 = (x_1\ x_2\ x_3)$ on y_2 is channeled through y_1, which results into the double arrow between y_1 and y_2 (non-recursive path diagram).

1.2 Channeling

Would it be possible to channel the effect of X_2 on y_1 through y_2, so that the latter variable obtains the role of transmittor, as shown in Figure 1B? We call this a "channeling" solution.

Obviously, the diagram in Figure 1B is a simplification compared to Figure 1A. But the advantage of channeling should not be that it merely simplifies the path diagram. There should be substantive arguments as well. We shall formulate them in terms of abstract conditions - illustrated, between brackets, by a reference to the empirical example described in section 6.

(i) X_1 should be an obvious set of predictors of y_1 (a child's occupational aspiration can be predicted from the child's background).

(ii) Prediction of y_1 will improve if y_2 is added to the set of predictors (X_1, y_2). (A child's aspiration can be better predicted if we know not only the child's background, but also the aspiration of its best friend).

(iii) Prediction of y_1 will also improve if X_2 is added to X_1 in the set of predictors (the child's aspiration can be predicted better if we know not only its own background, but also that of its best friend).

(iv) However - and this is the crucial point for channeling - prediction of y_1 on the basis of (X_1, y_2) will not improve if X_2 is added to this set of predictors which then would become (X_1, X_2, y_2). (Prediction of a child's aspiration based on the child's own background and its friend's aspiration does not become better if we also know its friend's background).

However, there is another aspect. For Figure 1B we may write a multiple regression equation

$$y_1 = X_1 b_{11} + y_2 c_{21} + v_1 \tag{1}$$

in which b_{11} is a vector of weights, c_{21} is a single weight, and v_1 a random vector. Multiple regression requires that v_1 is uncorrelated with X_1 and y_1 . But this multiple regression solution ignores X_2 , and therefore does not forbid that v_1 is correlated with X_2 .

For a channeling solution we may require instead a regression equation in which v_1 must be uncorrelated with X_1 and X_2 (not necessarily with y_2).

The argument is that the set X forms the ultimate set of predictors, and that the term $y_2 c_{21}$ in eq. (1) above is nothing but a substitute for a term $X_2 b_{21}$ in an equation for multiple regression of y_1 on $X = (X_1, X_2)$.

This difference between multiple regression and channeling will be further discussed in section 3.

1.3 Non-recursive path-diagrams

Of course, given Figure 1B, one might ask: is it also possible to channel the effect of X_1 on y_2 through y_1? We then obtain the diagram of Figure 1C.

This diagram is non-recursive. In general, a recursive diagram requires that, if one starts at some variable and follows a path indicated by arrows, it is not possible to come back at the initial variable. The diagram in Figure 1C is non-recursive: starting at y_1 and following arrows, one may come back to y_1 again. The double arrow between y_1 and y_2 is only the simplest example of a non-recursive path. (E.g., a triangular non-recursive path would imply that there is some variable y_1 with an arrow directed to y_2 , that there is an arrow from y_2 to y_3, and that there is again an arrow from y_3 to y_1).

Clearly, a non-recursive diagram, such as in Figure 1C, can be acceptable only if y_1 and y_2 are synchronous variables: things which happen later in time cannot have an effect on things which happen earlier. But it also may be true that variables y_1 and y_2 are aggregated over time (which means: at some moments y_1 influences y_2, and at other moments y_2 influences y_1 - but aggregated over time, one finds that y_1 and y_2 influence each other).

1.4 Objectives of this paper

The objectives of this paper are:

(i) to introduce the channeling solution (compared to the multiple regression solution) in terms of data theory (instead of using a mathematical-statistical approach) - section 3;

(ii) to show under what conditions a unique channeling solution can be identified - section 4;

(iii) to discuss two compromise solutions for the situation where a channeling solution cannot be uniquely identified because there are too many restrictions (overidentification) - section 5;

(iv) to display a numerical example with a non-recursive path diagram - section 6.

2 NOTATION

In this paper we shall assume that all variables are standardized in such a way that they have zero mean and that their sum of squares is equal to unity. The latter assumption is just for notational convenience (classical standardization requires that the sum of squares is equal to n = number of elements in a vector).

In addition, we use a special notation which is summarized in the Appendix. A more complete treatment of this notation can be found in Van de Geer (1986).

3 CHANNELING AND MULTIPLE REGRESSION

3.1 Multiple regression

An equation for multiple regression of y_1 on X_1 and y_2 can be written as

$$y_1 = X_1\underline{b}_{11} + y_2\underline{c}_{21} + v_1 \tag{2}$$

in the same format as equation (1) in section 1.3. But we may also write, in the notation adopted for this paper:

$$y_1 = y_1\underline{P}(X_1, y_2) + y_1\underline{A}(X_1, y_2) = y_1\underline{P}(X_1, y_2) + v_1 \tag{3}$$

where

$$y_1\underline{P}(X_1, y_2) = y_1\underline{P}X_1 + y_1\underline{P}(y_2\underline{A}X_1) = X_1b_{11} + y_2c_{21}\underline{A}X_1. \tag{4}$$

Development of equation (4) reveals that

$$\underline{c}_{21} = c_{21}$$
$$\underline{b}_{11} = b_{11} - b_{12}c_{21}$$

where b_{11} gives the weights for regression of y_2, on X_1, and b_{12} the weights for regression of y_2 on X_2.

Equation (3) shows that

$$v_1 = y_1\underline{A}(X_1, y_2) = (y_1\underline{A}X_1)\underline{A}(y_2\underline{A}X_1)$$

and that v_1 therefore is uncorrelated with X_1 and with y_2. Note that v_1 is not uncorrelated with X_2 - in fact, X_2 plays no role in the equations above.

3.2 Channeling

Here the regression equation can be written as

$$y_1 = y_1\underline{P}X_1 + y_2c_{21}\underline{A}X_1 + u_1 = X_1b_{11} + y_2c_{21}\underline{A}X_1 + u_1 \tag{5}$$

with the same solution for regression weights b_{11}, as in eq. (4), but with a different solution for c_{21}, and therefore also a solution for u_1 which differs from that for v_1. The point is that c_{21} now must be chosen in such a way that u_1 becomes orthogonal to X_1 and X_2 (but not necessarily orthogonal to y_2), whereas v_1 was orthogonal to X_1 and y_2 (but not necessarily orthogonal to X_2).

This can be achieved by taking c_{21} in eq. (5) such that

$$(y_1\underline{A}X_1)\underline{P}(X_2\underline{A}X_1) = (y_2c_{21}\underline{A}X_1)\underline{P}(X_2\underline{A}X_1). \tag{6}$$

In words: $y_1\underline{A}X_1$ and $y_2c_{21}\underline{A}X_1$ must have identical projections on $X_2\underline{A}X_1$. Such a solution is not always feasible (this will be further discussed in section 4). If the solution is feasible, it must be true that the difference vector $(y_1\text{-}y_2c_{21})\underline{A}X_1$ has zero projection on $X_2\underline{A}X_1$. It thus follows that

$$u_1 = ((y_1\text{-}y_2c_{21})\underline{A}X_1)\underline{A}(X_2\underline{A}X_1) \tag{7}$$

which confirms that u_1 is orthogonal to X. To spell it out: it is true by definition that $(y_1\text{-}y_2c_{21})\underline{A}X_1$ is orthogonal to X_1, and eq. (7) adds the further requirement that u_1 is the component of $(y_1\text{-}y_2c_{21})\underline{A}X_1$ which is orthogonal to $X_2\underline{A}X_1$, and therefore orthogonal to X_2.

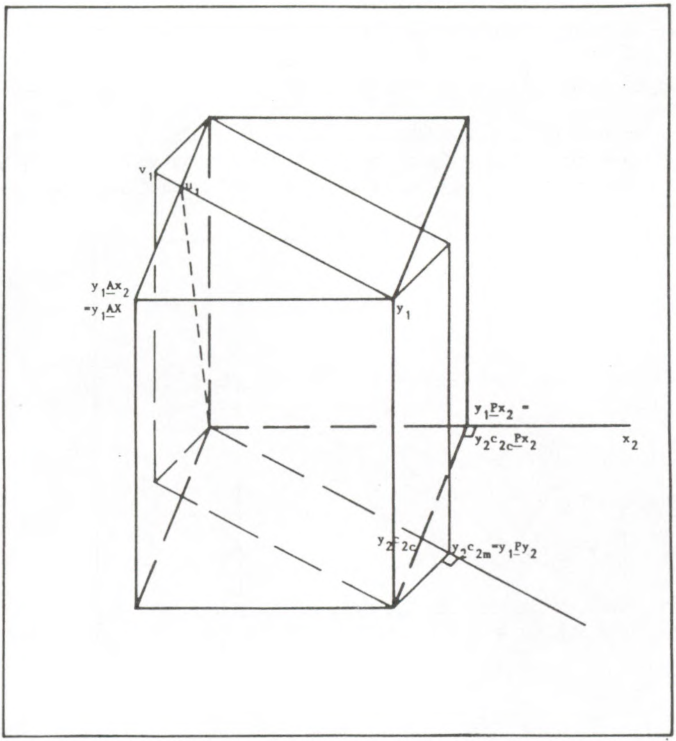

Figure 2 Comparison of multiple regression and a perfect channeling solution. All vectors in the graph are projections on the plane orthogonal to X_1. Further explanation of the figure in section 3.3.

3.3 Graph of the two solutions

The multiple regression solution and the channeling solution are graphed in Figure 2. In this figure it is assumed that $X_2\underline{A}X_1$ is the one-dimensional vector $x_2\underline{A}X_1$. The graph then shows a three dimensional vector-constellation, spanned by the vectors $x_2\underline{A}X_1$, $y_1\underline{A}X_1$, and $y_2\underline{A}X_1$. In the figure, labels for vectors have been simplified by omitting the specification "$\underline{A}X_1$". E.g., y_1 in the figure stands for the vector $y_1\underline{A}X_1$, label x_2 for the vector $x_2\underline{A}X_1$. Also, the figure distinguishes between the weights c_{2m} and c_{2c} for the multiple regression solution, and the channeling solution, respectively. The label y_2c_{21} therefore stands for the vector $y_2c_{21}\underline{A}X_1$, etc.

Figure 2 shows two interlocking rectangular blocks. These blocks have sides which are rectangles. One block, drawn with light lines, gives the multiple regression solution. This block is defined by the vectors $y_1\underline{A}X_1$, v_1, and $y_2c_{2m}\underline{A}X_1$. It is typical of this block that $y_2c_{2m}\underline{A}X_1$ is the projection of y_1 on $y_2\underline{A}X_1$.

The other block, drawn with heavy lines, gives the channeling solution. This block is defined by the vectors $y_1\underline{A}X_1$, its projection $y_1\underline{P}(x_2\underline{A}X_1)$, and its anti-projection $y_1\underline{A}(x_2\underline{A}X_1) = y_1\underline{A}X$.

In other words: we have a rectangular decomposition

$$y_1\underline{A}X_1 = y_2c_{2m}\underline{A}X_1 + v_1 \tag{8}$$

which corresponds to the multiple regression solution. And we also have the decomposition

$$y_1\underline{A}X_1 = y_2c_{2c}\underline{A}X_1 + u_1 \tag{9}$$

where this decomposition describes a parallelogram. This is further illustrated in Figure 3. This figure shows the plane spanned by $y_1\underline{A}X_1$, u_1, and v_1. The figure shows that $y_1\underline{A}X_1$, v_1, and $y_2c_{12m}\underline{A}X_1$ form a rectangle, whereas $y_1\underline{A}X_1$, u_1, and $y_2c_{2c}\underline{A}X_1$ form a parallelogram.

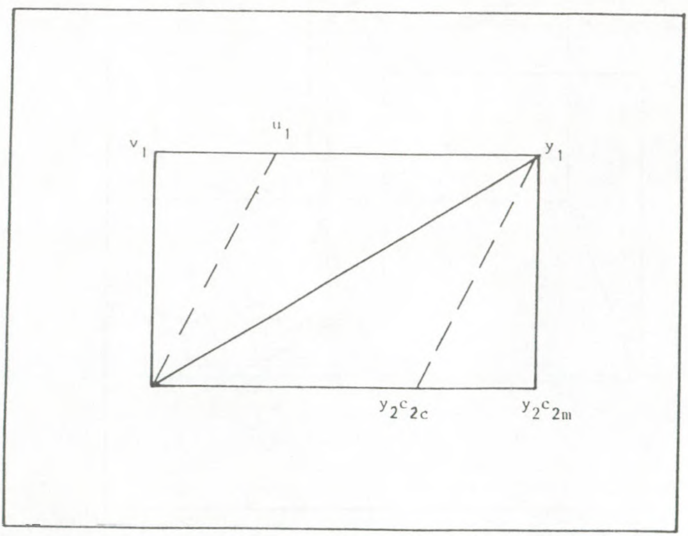

Figure 3 In Fig. 2 the vectors u_1, v_1, and $y_1\underline{A}X_1$ are located in a plane. Fig. 3 shows this plane.

Figure 3 therefore also shows that in the multiple regression solution (with the rectangle) SSv_1 is unconditionally minimized, whereas in the channeling solution (with the parallelogram) SSu_1 will be larger than SSv_1 (and is minimized conditionally: u_1 must be orthogonal to $x_2\underline{A}X_1$). In fact, in Figure 2 it is shown that u_1 is a vector in the side of the heavier drawn block, where this side is orthogonal to $y_1P(x_2\underline{A}X_1) = y_2c_{2c}\underline{P}(x_2\underline{A}X_1)$.

3.4 Feasibility of the channeling solution

Figures 2 and 3 picture a perfect channeling solution, in which $y_1P(x_2\underline{A}X_1) = y_2c_{2c}\underline{P}(x_2\underline{A}X_1)$. Such a solution is not always feasible. The next problem, therefore, will be to find out under what conditions such a solution is feasible if $X_2\underline{A}X_1$ and Y_2AX_1 are higher-dimensional (instead of uni-dimensional).

4 IDENTIFIABILITY OF THE CHANNELING SOLUTION

4.1 Basic requirement

The basic requirement for a channeling solution is:

$$y_1\underline{A}X_1 = Y_2c_{21}\underline{A}X_1 + u_1 \tag{10}$$

where c_{21} is a vector of weights, and where u_1 must be orthogonal to $X_2\underline{A}X_1$.

In situations where a unique solution for c_{21} exists, eq. (10) is said to be "just identifiable". If there are no solutions for c_{21} such that u_1 obeys the basis requirement, eq. (10) is said to be "overidentified". If there is more than one possible solution, the equation is called "underidentified".

4.2 Simplification

Under what conditions can a solution for c_{21} be identified? To answer this question, we need a number of preliminary simplifications.

(1) The first one is that, without loss of generality, $Y_2\underline{A}X_1$ may be replaced by an orthogonal and unit-normalized basis Π. This means that the space spanned by Y_2AX_1 is identical to the space spanned by Π, where $\Pi'\Pi = I$.

(ii) Secondly, $X_2\underline{A}X_1$ can also be replaced by an orthogonal and unit-normalized basis Ψ (Ψ spans the same space as $X_2\underline{A}X_1$, with $\Psi'\Psi = I$).

(iii) Thirdly, we may require in addition that Π and Ψ correspond to the canonical solution between $Y_2\underline{A}X_1$ and $X_2\underline{A}X_1$. This is further specified in Table 1. Here, the matrix Π (with m columns) is partitioned as $\Pi = (\Pi_1, \Pi_2)$ with m* and m** columns, respectively (m* + m** = m). Similarly for Ψ (with k columns), partitioned as $\Psi = (\Psi_1, \Psi_2)$, with k* and k** columns, respectively, and where k* + k** = k, but also k* = m*.

The canonical solution implies that $\Pi'_1\Psi_1 = \Omega$ is a square and diagonal matrix, with on its diagonal the non-zero canonical correlations between $Y_2\underline{A}X_1$ and $X_2\underline{A}X_1$. The matrices $\Pi'_1\Psi_2$ and $\Pi'_2\Psi_1$ (if they exist) are zero matrices.

Table 1 Correlations between auxiliary variables: most general case

η	Π_1	Π_2	Ψ_1	Ψ_2
1	a'_1	a'_2	b'_1	b'_2
a_1	I	0	Ω	0
a_2	0	I	0	0
b_1	Ω	0	I	0
b_2	0	0	0	I

Table 1 also contains the correlations between $y_1\underline{A}X_1$ and (Π_1, Π_2), indicated as the vectors a_1 and a_2. Similarly for the correlations between $y_1\underline{A}X_1$ and (Ψ_1, Ψ_2), with notation b_1 and b_2. The symbol η is used for the unit-normalized version of $y_1\underline{A}X_1$ (so that $\eta'\eta = 1$).

In the sequel we shall assume that the correlations in a_1, a_2, and b_1, b_2 are not all equal to zero.

4.3 Just-identifiable solution

The just-identifiable solution is defined as a unique solution for d, such that

$$\Psi'(\eta, \Pi)d = 0. \tag{11}$$

This solution can be identified if Π_2 and Ψ_2 do not exist, as in Table 2.
The solution then becomes:

$$d' = (1, -b'_1\Omega^{-1}). \tag{12}$$

Table 2 Just-identifiable case

η	Π_1	Ψ_1
1	a'_1	b'_1
a_1	I	Ω
b_1	Ω	I

In fact, eq. (11), in this special case, defines $k^* = m^*$ homogeneous equations with $(k^* + 1) = (m^* + 1)$ unknowns, so that a solution for d can be identified uniquely (up to an arbitrary proportionality coefficient: this explains why the first element of d' can be set equal to unity).

The solution implies:

$$u_1 = (\eta, \Pi_1)d$$

where u_1 is uncorrelated with X_1 (because η and Π_1 are uncorrelated with X_1), and where u_1 is also uncorrelated with X_2 (because eq. (11) requires that u_1 is uncorrelated with Ψ_1, and therefore uncorrelated with $X_2\underline{A}X_1$, and therefore uncorrelated with X_2).

4.4 Overidentification

This situation arises if Π_2 does not exist, as shown in Table 3. Now eq. (11) gives $m > k^*$ homogeneous equations with k^* unknowns. There is no solution for d, because there are too many restrictions in eq. (11).

Nevertheless, one might look for a compromise solution of d. Such solutions will be discussed in section 5.

Table 3 Over-identified case

η	Π_1	Ψ_1	Ψ_2
1	a'_1	b'_1	b'_2
a_1	I	Ω	0
b_1	Ω	I	0
b_2	0	0	I

Table 4 Under-identified case

η	Π_1	Π_2	Ψ_1
1	a'_1	a'_2	b'_1
a_1	I	0	Ω
a_2	0	I	0
b_1	Ω	0	I

4.5 Under-identification

This situation arises if Ψ_2 does not exist, as shown in Table 4. Eq. (11) now specifies m^* homogeneous equations with more than $(m^* + 1) = (k^* + 1)$ unknowns. As a result there will be more than one solution for d. These solutions can be summarized as

$$d' = (1, -b'_1\Omega\text{-}1, g'_3) \tag{13}$$

where g'_3 is any arbitrary row of weights.

Of course, we may compromise. One possibility is to take $g_3 = 0$, which comes to the same thing as ignoring Π_2, so that we are back to the situation in Table 2.

Another compromise would be that we take g_3 in such a way that SSu_1 is minimized. This will be true if we take $g_3 = a_2$. This solution is a compromise between a perfect channeling solution with respect to Π_1, augmented by a multiple regression solution with respect to Π_2.

4.6 Both Π_2 and Ψ_2 exist

This is the situation of Table 1. One possible compromise would be to ignore Π_2 and Ψ_2, so that Table 1 becomes reduced to Table 2, and we have a pseudo-just identifiable solution. Or, we may ignore Π_2, so that Table 1 is reduced to the overidentified case, and possible compromises are shown in section 5. Or, we may ignore Ψ_2, so that Table 1 is reduced to the under-identified case, and the compromise suggested in section 4.5 would apply.

4.7 Conclusion about identifiability

There are two problems to be considered.

(i) The first one is that the rank of Π or Ψ might be smaller than the number of variables in Y_2 or X_2. This means that there is linear dependence among the variables in $Y_2\underline{A}X_1$ or in $X_2\underline{A}X_1$. If $Y_2\underline{A}X_1$ has deficient rank, it follows that a weighted sum $Y_2c_{21}\underline{A}X_1$ is not uniquely defined in terms of the weights c_{21}. In fact, those weights can be chosen in different ways without effect on the values in $Y_2c_{21}\underline{A}X_1$. We do not consider this as an instance of "unidentifiability". Linear dependence among predictors always can be handled (e.g., by applying "truncation" methods, such as "principal component truncation").
The same problem arises in fact if there is linear dependence among X_1, so that there is no unique solution for b_{11}, where $y_1\underline{P}X_1 = X_1b_{11}$. Although then one can take different values for the weights b_{11}, the vector $y_1\underline{P}X_1$ remains uniquely defined.

(ii) Even if $Y_2\underline{A}X_1$ and $X_2\underline{A}X_1$ have the same rank $k = m$ as Y_2 and X_2 themselves, it does not follow that there are no zero canonical correlations between $Y_2\underline{A}X_1$ and $X_2\underline{A}X_1$. Knowledge of the number of variables in Y_2 and X_2 therefore is not sufficient to decide whether the situation is just-identifiable, overidentified, or under-identified. Goldberger (1964) is not complete in this respect: his exposition ignores the possible difference between k and k*, and between m and m*.

5 OVERIDENTIFICATION

5.1 Introduction

The overidentified case receives a lot of attention in econometric literature. Compromises for a channeling solution then are introduced mainly on the basis of mathematical-statistical arguments (unbiasedness, efficiency, sufficiency). Such a treatment makes it somewhat difficult what these compromises mean in terms of data theory, by which we mean an understanding in terms of the algebra and geometry of the vectors involved. The latter also makes it easier to see how their compromises are related to other multivariate techniques, more common in the social sciences.

We shall discuss two compromises: (1) two-stages least squares (2SLS), and (2) least generalized residual variance (LGRV). Our discussion is based mainly on Goldberger's textbook (1964).

5.2 Two-stages least squares (2SLS)

In section 3.3 it was explained that the channeling solution takes a weight c_2 for y_2, in such a way that $y_1\underline{P}(X_2\underline{A}X_1)$ and $y_2c_2\underline{P}(X_2\underline{A}X_1)$ become identical (eq. 6 and Figure 2). If Y_2 contains more vectors, the same solution remains valid, on the understanding that c_2 now is a vector of weights. This produces the just-identifiable solution, if it exists.

However, in the situation of Table 3 no such solution exists. The reason is simply that $y_1\underline{P}(X_2\underline{A}X_1) = y_1\underline{P}\Psi_1 + y_1\underline{P}\Psi_2$ - a sum of two non-zero vectors.
But $Y_2c_2\underline{P}(X_2\underline{A}X_1) = Y_2c_2\underline{P}\Psi_1 + y_1\underline{P}\Psi_2$, where the latter vector is a zero vector for any choice of c_2. It follows directly that $Y_1\underline{P}\Psi$ and $Y_2c_2\underline{P}\Psi$ never can be identical.

2SLS takes the compromise solution for c_2 which results in equality of the two projections $y_1\underline{P}\Psi_1$ and $Y_2c_2\underline{P}\Psi_1$. In other words, this solution just ignores Ψ_2. This means that Table 3 is reduced to Table 2 where a just-identifiable solution is possible. This solution has weights

$$d' = (1, -b'_1\Omega^{-1}) \tag{14}$$

which is the same expression as in eq. (12).

The geometry of this solution is illustrated in Figure 2, for the special case that Y_2 is one-dimensional (so that $\Pi_1 = y_2$), and replacing the vector $x_2\underline{A}X_1$ in Figure 2 by the single vector ψ_1. More generally, the compromise says that $\Pi = \Pi_1$ has dimensionality $m = m^*$, and that Ψ_1 also has dimensionality $k^* = m^*$. It then follows that the projection of Π on Ψ_1 also has dimensionality m^*. But (η,Π) has dimensionality (m^*+1), whereas the projection of (η,Π) on Ψ_1 still has dimensionality $k^* = m^*$. It follows that there must be a direction in the space spanned by (η,Π) with zero projection on Ψ_1. This direction is specified in eq. (14); it is the direction of the vector

$$(\eta,\Pi)d = \eta - \Pi\Omega^{-1}b_1. \tag{15}$$

The vector in eq. (15) has the same direction as u_1. In fact, u_1 is defined by the equation

$$y_1\underline{A}X_1 = Y_2c_{21}\underline{A}X_1 + u_1 \tag{16}$$

whereas eq. (15) can be re-written as

$$\eta = \Pi\Omega^{-1}b_1 + (\eta,\Pi)d. \tag{17}$$

The two equations (16) and (17) differ in a proportionality constant, because η is the unit-normalized version of $y_1\underline{A}X_1$, or $\eta = y_1\underline{A}X_1/SS(y_1\underline{A}X_1)^{1/2}$. It follows that

$$(\eta,\Pi)d = u_1/SS(y_1\underline{A}X_1)^{1/2}. \tag{18}$$

One should note that there is not necessarily a unique 2SLS solution if $X_2\underline{A}X_1$ has larger rank than $Y_2\underline{A}X_1$. The fact that $k>m$ does not guarantee that there are no zero canonical correlations. In other words: the fact that $k>m$ does not guarantee that Π_2 does not exist.

The name "two-stages least squares" can be explained by taking the projection of y_1 on X_1 as the first stage, followed by taking the projection of $y_1\underline{A}X_1$ on $X_2\underline{A}X_1$ as the second step. Obviously, the latter projection is identical to the projection of y_1 on $X_2\underline{A}X_1$.

5.3 Least generalized residual variance (LGRV)

Define $Y = (y_1, Y_2)$. In the just-identifiable case there is a unique solution for weights w, such that $Yw\underline{A}X_1$ has zero projection on X_2AX_1. This solution implies that the angle between $Yw\underline{A}X_1 = u_1$ and the space spanned by $X_2\underline{A}X_1$ is an angle of 90° degrees.

In the overidentified case no such solution for w can be found. But it would be a compromise to take the solution for w in such a way that the angle between $Yw\underline{A}X_1$ and its projection on $X_2\underline{A}X_1$ is as much as possible close to an angle of 90°. This is equivalent with saying that w gives weights for $Y\underline{A}X_1$ which correspond to the *smallest canonical correlation* between $Y\underline{A}X_1$ and $X_2\underline{A}X_1$.

Now it will be true for any arbitrary choice of w that $Yw\underline{A}X$ and $Yw\underline{P}(X_2\underline{A}X_1)$ are orthogonal vectors, with sum vector

$$Yw\underline{A}X_1 = Yw\underline{A}X + Yw\underline{P}(X_2AX_1). \tag{19}$$

The proof of the theorem used in eq. (19) can be derived from the Appendix of this paper. It follows that the compromise based on the smallest canonical correlation between $Y\underline{A}X_1$ and $X_2\underline{A}X_1$

is equivalent to the solution for weights w which give the *largest* canonical correlation between $Yw\underline{A}X_1$ and $Yw\underline{A}X$.

Figure 4 gives a geometrical illustration. The figure shows a rectangular block, in which the diagonal vector Yw is decomposed into three orthogonal components: $Yw\underline{P}X_1$, $Yw\underline{P}(X_1\underline{A}X_1)$, and $Yw\underline{A}X$. In the just-identifiable case, this block collapses into a single plane rectangle, with sides $YwAX = Yw\underline{A}X_1$ and $Yw\underline{P}X = Yw\underline{P}X_1$, with diagonal $Yw = Yw\underline{A}(X_2AX_1)$, and with $Yw\underline{P}(X_2\underline{A}X_1) = 0$.

The rationale of the LGRV compromise must be clear. Yw has a component $Yw\underline{A}X_1$ (uncorrelated with X_1), and this component should be as much as possible similar to the component $Yw\underline{A}X$ (the component of Yw not predictable from X as a whole).

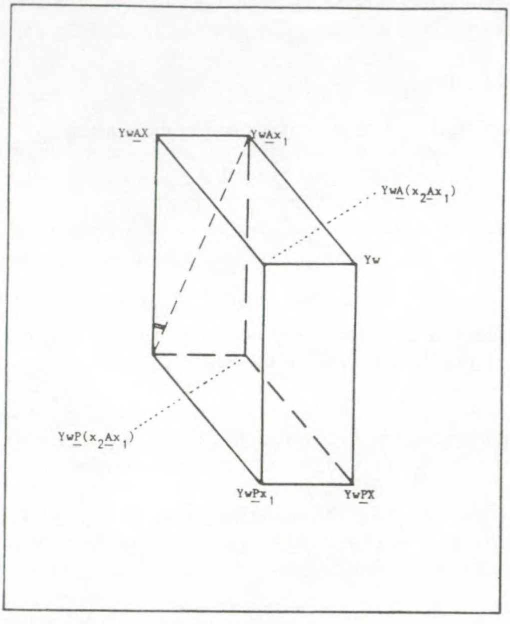

Figure 4 Decomposition of a vector Yw into orthogonal components:
$Yw = Yw\underline{P}X + Yw\underline{A}X$, $Yw = Yw\underline{P}x_1 + Yw\underline{A}x_1$, and
$Yw = Yw\underline{P}(x_2\underline{A}x_1) + Yw\underline{A}(x_2\underline{A}x_1)$.

5.4 Comparison between 2SLS and LGRV

First of all, the two compromise solutions become identical to the just-identifiable solution if there is such a solution. In the overidentified case, however, the two solutions comprise in different ways. 2SLS acts as if $X_2\underline{A}X_1$ may be replaced by $(Y_2\underline{A}X_1)\underline{P}(X_2\underline{A}X_1) = Y_2\underline{P}(X_2\underline{A}X_1)$. The result is that u_1 will be orthogonal to X_1 (this is the first "stage"), and, as the second "stage", will be orthogonal to Ψ_1. But u_1 will not necessarily be orthogonal to Ψ_2. In fact, Ψ_2 is "ignored". Or perhaps better: the correlations b_2 between y_1 and ψ_2 are ignored.

LGRV does not ignore Ψ_2. This solution makes u_1 as much as possible orthogonal to both Ψ_1 and Ψ_2, and the solution therefore does depend on the values in b_2 (which are ignored in 2SLS).

The two compromise solutions become identical if the correlations b_2 are zero's, which would mean that y_1 is uncorrelated with directions in X_2 which are independent of X_1 as well as of Y_2.

5.5 Numerical illustration

The following small example is based on the correlation matrix in Table 5. It is assumed that η and Π are uni-dimensional, whereas Ψ has the two dimensions ψ_1 and ψ_2. The table also contains correlations with an additional vector z, defined as the unit-normalized vector $\eta\underline{A}\Pi$. These correlations between z and other vectors can be easily derived from the other correlations given in the table.

Table 5 Numerical example of section 5.5

η	Π_1	ψ_1	ψ_2	z
1	.4	.5	.6	.9165
.4	1	.7	0	0
.5	.7	1	0	.24
.6	0	0	1	.6547
.9165	0	.24	.6547	1

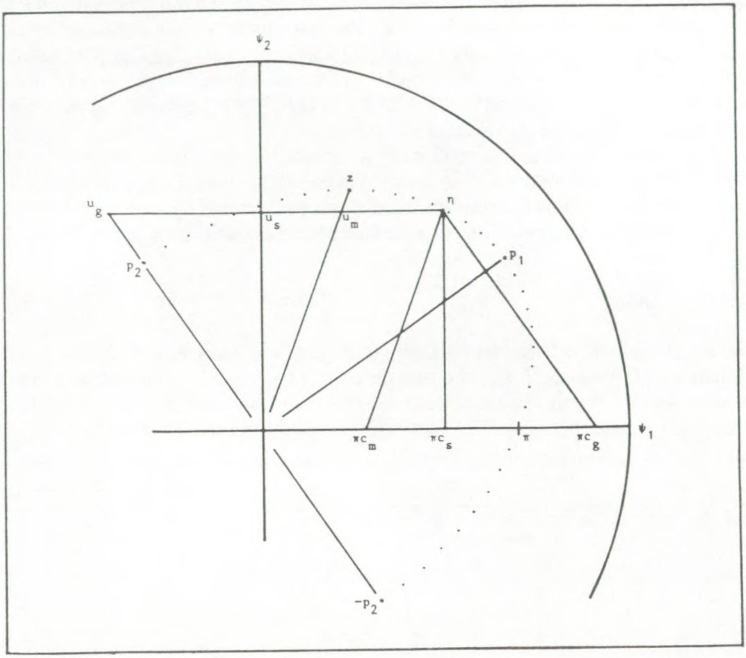

Figure 5 The figure is based on the example of Table 5. It shows projections on the plane spanned by ψ_1 and ψ_2 (located on the unit circle, which is only partly drawn). The 2SLS solution appears as the rectangular decomposition $\eta = \pi c_s + u_s$ (where us must be orthogonal to ψ_1). The multiple regression solution appears as the parallelogram $\eta = \pi c_m + u_m$ (the projection of the corresponding rectangle in Figure 6). The LGRV solution is given by the parallelogram $\eta = \pi c_g + u_g$, where u_g has the same direction as the shortest principal axis p_2 of the dotted ellipse (this ellipse is the projection of the unit circle in Figure 6) - so that u_g is orthogonal to p_1. Only one half of the ellipse has been drawn, cut off by its shortest principal axes p_2 and $-p_2$.

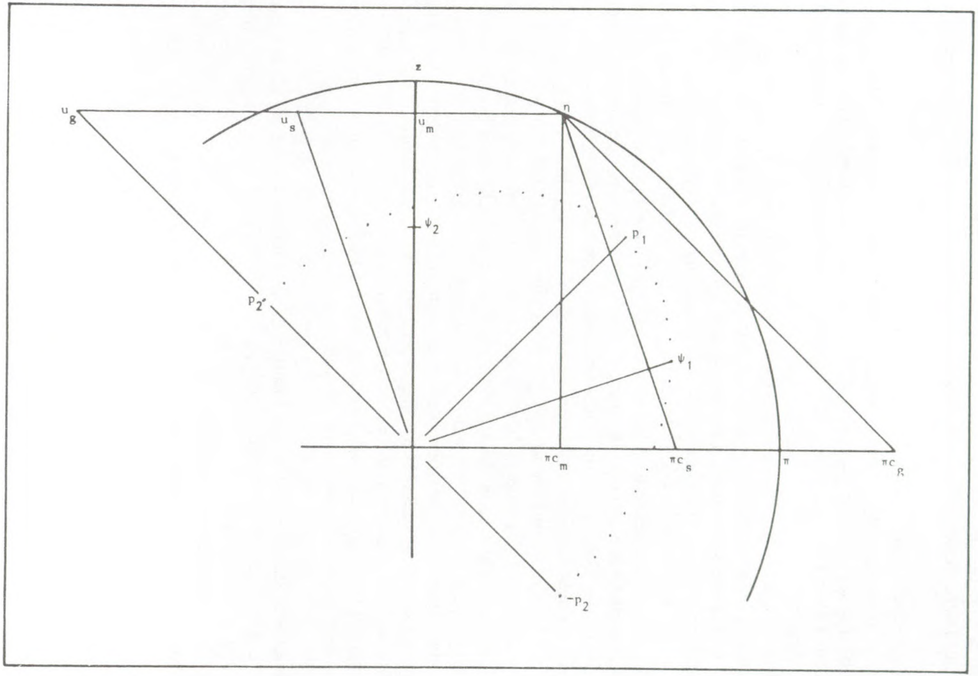

Figure 6 The figure is based on the example of Table 5. It shows projections of vectors on the plane spanned by π, η, and z (these three vectors are loctaed on the unit circle). The multiple regression solution appears as the rectangle $\eta = \pi c_m + u_m$ (projected on Figure 5 as a parallelogram). The 2SLS solution is shown in the paralleogram $\eta = \pi c_s + u_s$, where u_s is orthogonal to (the projection) of ψ_1. The LGRV solution is given as the parallelogram $\eta = \pi c_g + u_g$, where u_g has the same direction as the shortest principal axis p_2 of the ellipse, and therefore is orthogonal to the longest axis p_1.
Note that u_m, u_s, and u_g are vectors located in the plane of drawing (because η and π are located in this plane).

Results are pictured in the two figures 5 and 6. Figure 5 shows the plane spanned by Ψ, with ψ_1 and ψ_2 as two orthogonal radii of the unit-circle. Earlier definitions imply that the projection of Π on this plane must have the same direction as Ψ_1. Figure 6 shows the plane spanned by z and Π as two orthogonal radii of the unit-circle. Here the vector η, located in the plane of drawing of Figure 6, also appears as a radius of the unit-circle.

The unit-circle of Figure 5 projects on Figure 6 as an ellipse, on which the projections of ψ_1 and ψ_2 are located. Conversely, the unit-circle of Figure 6 projects on Figure 5 as an ellipse on which the vectors z, η, and Π are located. The two ellipses have identical size (i.e., their principal axes have the same lengths).

The two figures show three solutions.

(i) Firstly, we have the solutions for regression of η on Π. This solution is shown in Figure 6 as the decomposition of η as a rectangle with sides $\eta\underline{P}\Pi$ and $\eta\underline{A}\Pi$. We write $\eta\underline{A}\Pi = u_m$, and it is easily derived from Table 5 that $u_m = \eta - .4\Pi$. This rectangular decomposition appears in Figure 5 as a parallelogram.

(ii) The 2SLS solution appears in Figure 6 as the decomposition $\eta = \Pi c_s + u_s$. This decomposition has the shape of a parallelogram, with the characteristic feature that u_s must be orthogonal to ψ_1. In Figure 5 the same decomposition appears as a rectangle, with the projection of us orthogonal to ψ_1. The numerical solution is: $u_s = \eta - .714\Pi$.

(iii) The third solution is the LGRV solution, indicated by $u_g = \eta - 1.306\Pi$. The characteristic feature is that u_g has the same direction as the shortest axis of the ellipse, as can be seen in both figures. The rationale of the solution is: since u_g has the same direction as the shortest axis of the projection of the unit-circle, the angle between u_g and its projection on Figure 5 must be larger than the angle between any other radius of the unit-circle in Figure 6 and its projection on Figure 5. The cosine of this angle therefore corresponds to the smallest canonical correlation between (η,Π) and Ψ. In the example, this smallest canonical correlation is equal to .566 (in the figure equal to the length of the shortest principal axis of the ellipse).

The three solutions, u_m, u_s, and u_g, are located on the straight line defined by $\eta - \Pi c$ with different solutions for c. In Figure 6, the solution for c is chosen in such a way that $\eta - \Pi c = u_m$ has minimum length; i.e., u_m is orthogonal to Π. On the other hand, the 2SLS solution is chosen in such a way that the projection of us on Figure 5 has minimum length; i.e., in Figure 5 the projection of u_s is orthogonal to ψ_1. These two solutions, for u_m and u_s, do not depend on the location of the principal axes of the ellipses. They therefore ignore the value of $b_2 = .6$, in the example.

In contrast, the solution for u_g does depend on the location of the shortest axis of the ellipse, and therefore does depend on the value of b_2. In the example, this value is relatively large, which explains why u_g is relatively far away from u_s and u_m. In fact, to the extent that b_2 becomes smaller, u_g will approach u_s. In the limit, when $b_2 = 0$, the vectors u_g and u_s will coincide, because we then have a special just-identifiable case (ψ_2 may be ignored without penalty, because ψ_2 is uncorrelated with η). This limiting case does not imply, however, that u_m becomes the same as $u_s = u_g$.

6 NUMERICAL EXAMPLE

6.1 Introduction

For a real-life numerical example, data have been taken from Duncan, Haller, and Portes (1968). In this study data were collected for 17-year old boys. They refer to three "background" variables $X_1 = (x_1\ x_2\ x_3)$, and one direct variable y_1. The same data were collected for boys which respondents considered to be their "best friend", collected in $X_2 = (x_6\ x_5\ x_4)$ and y_2. Table 6 gives a short description of the variables. Note that the variables (X_1, y_1) are in a sense symmetrical to those in (X_2, y_2). The symmetry is not complete, however: if respondent John says that Peter is his best friend, it does not follow that Peter will say that John is his best friend.

All further analyses are based on the matrix of correlations between the 8 variables described in Table 6. These correlations are given in Table 7.

We will show results of five different analyses.

(A) Multiple regression of y_1 and of y_2 on X.
The diagram for such a solution is given in Figure 1A.

(B) Multiple regression of y_1 on X_1, and of y_2 on X_2.

(C) Multiple regression of y_1 on (X_1, y_2) and of y_2 on (X_2, y_1).
The path-diagram is shown in Figure 1C.

(D) The same diagram as in (C), but now with 2SLS solution.

(E) Same diagram as in (C) or (D) but now with LGRV solution.

Although solutions (C), (D) and (E) have the same path-diagram, their numerical solutions for weights are different. Solution (C) is based on eq. (2) in section 3.1 - (D) is based on the equations in section 5.2 - (E) on those in section 5.3.

Table 6 Variables used in the numerical example of section 6

	respondent	best friend
parental aspiration (degree to which parents encourage their child to have high levels of achievement)	x_1	x_6
family socioeconomic status	x_2	x_5
intelligence of child	x_3	x_4
occupational aspiration child	y_1	y_2

Table 7 Correlations between the variables given in Table 6

x_1	x_2	x_3	x_4	x_5	x_6	y_1	y_2
1.0000	.1839	.0489	.0186	.0782	.1147	.2137	.0839
	1.0000	.2220	.1861	.3355	.1021	.4105	.2598
		1.0000	.2707	.2302	.0931	.3240	.2786
			1.0000	.2950	-.0438	.2930	.3607
				1.0000	.2087	.2995	.5007
					1.0000	.0760	.1988
						1.0000	.4216
							1.0000

Table 8 Weights for the variables in X in the five solutions described in section 6

	MR y_i on X	MR on X_1&X_2	MR y_i on X_i and y_j		2SLS		LGRV	
	β_1	β_{11}	$\beta_{21}c_2$	b_{11}				
x_1	.1423	.1410	.0103	.1307	.0134	.1276	.0136	.1274
x_2	.2780	.3304	.0587	.2718	.0764	.2540	.0773	.2531
x_3	.1888	.2407	.0675	.1762	.0879	.1558	.0890	.1548
		β_{12}						
x_4	.1562	.2287						
x_5	.1059	.2238						
x_6	-.0015	.0393						
	β_2	β_{21}	$\beta_{12}c_1$	b_{22}				
x_1	.0225	.0354						
x_2	.0560	.2018						
x_3	.1115	.2321						
		β_{22}						
x_4	.2163	.2478	.0581	.1897	.0680	.1798	.0684	.1795
x_5	.3671	.4013	.0569	.3446	.0666	.3347	.0669	.3344
x_6	.1130	.1259	.0100	.1159	.0117	.1142	.0118	.1141

6.2 Results

Regression weights for X are given in Table 8: weights which refer to the regression equations for y_1 in the upper half, and those that refer to y_2 in the lower half of the table. The "symmetry" in the variables makes the lower half almost the mirror image of the upper half.

First the upper half. The first column gives weights β_1 for regression of y_1 on X. The second column gives weights β_{11} and β_{12} for regression of y_1 on X_1 alone and on X_2 alone, respectively. The fourth column gives weights b_{11} for X_1 in the regression of y_1 on (X_1, y_2). It has been shown in section 3.1 that $b_{11} = \beta_{11} - \beta_{21}c_2$ (where c_2 is the weight given to y_2 in this regression equation - these weights can be found in Table 9). The third column of Table 8 shows values of $\beta_{21}c_2$. Fifth and sixth column have the same interpretation as third and fourth column, but now for the 2SLS solution, whereas the last two columns refer to the LGRV solution.

The lower half of Table 8 has the same interpretation, now with respect to regression equation for y_2.

Table 9 gives the solution for weights c_1 and c_2 (for analyses C, D, and E).

Table 10 shows results for SSu_1 and SSu_2. In all five analyses it will be true that $u_i = u_i\underline{A}X_i$ (i = 1,2), so that $SSu_i = SSu_i\underline{A}X_i$. The table also shows results for $SSu_i\underline{A}X$ (compare section 5.3 and Figure 6), as well as the ratio's $SSu_i\underline{A}X/SSu_i\underline{A}X_i$. Remember that the objective of LGRV is to maximize this ratio which then becomes identical to the squared largest canonical correlation between $Y\underline{A}X$ and $Y\underline{A}X_i$.

Table 9 Weights for y_j in the solutions described in section 6

	MR on X_i and y_j	2SLS	LGRV
c_2	.2909	.3789	.3833
c_1	.2540	.2975	.2990

Table 10 Some criterion results for the numerical example in section 6

	y_i on X	y_i on X_i	yi on (X_i, y_j)	2SLS	LGRV
SSu_1AX_1	.7169	.7552	.6808	.6876	.6883
SSu_1AX	.7169	.7169	.6712	.6797	.6804
SSu_2AX_2	.6682	.6846	.6290	.6306	.6307
SSu_2AX	.6682	.6682	.6252	.6271	.6271
ratio	1	.9493	.9860	.988498	.988504
	1	.9760	.9940	.994418	.994438

6.3 Comments

The most striking result in this example is that 2SLS and LGRV lead to almost identical results. The reason is that $u_i\underline{A}X$ and $u_i\underline{A}X_i$ are already highly correlated even in the multiple regression solution C. In fact, the ratio's in table 10 are all very close to unity (to such an extent that for 2SLS and LGRV we need six decimals to find a difference).

The remaining comments refer to minor technical details.

(i) Firstly, let us go back to the conditions for channeling as described in section 1.2. Those conditions were phrased in terms of squared multiple correlations. In this example, for y_1 predicted from X_1 alone we find a value of .2448. It is improved to .3192 if y_1 is predicted from (X_1, y_2). It also is improved if y_1 is predicted from $X = (X_1, X_2)$: .2831. But prediction from (X, y_2) - with value .3293 - is not really better than prediction from (X_1, y_2).

Similarly for prediction of y_2. From X_2 alone: .3154. From (X_2, y_1): .3711. From X: .3318. From (X, y_2): .3736.

Results above are not very impressive: all differences between squared multiple correlations are rather small. Nevertheless they serve to illustrate what was meant with the conditions described in section 1.2.

Results also show that y_2 is more than just a substitute for X_2 in predicting y_1 : prediction from (X_1, y_2) is better than prediction from X (similarly for prediction of y_2).

(ii) In solution A it is found that $SSu_i\underline{A}X_i = SSu_i\underline{A}X$. The reason is simple enough: in this multiple regression solution it will be true that $u_i = u_i\underline{A}X$, and therefore that $u_i\underline{A}X_i = u_i\underline{A}X$.

(iii) Solution C minimizes $SSu_i\underline{A}X_i$ unconditionally. On the other hand, solution E maximizes the ratio $SSu_i\underline{A}X/SSu_i\underline{A}X_i$.

(iv) Finally, Figure 1C gives the path diagram. It could be specified for the 2SLS solution by writing the 2SLS weights for X_1 (.1276 .2540 .1558) along the arrows from X_1 towards y_1 , and the weights (.1798 .3347 .1142) along the arrows from X_2 to y_2 - these weights can be found in the second 2SLS column of Table 8. In addition, the arrow from y_2 to y_1 obtains numerical value .3789, whereas the arrow from y_1 to y_2 has value .2975 - see Table 9.

APPENDIX : SUMMARY OF NOTATION

1. The notation SSZ is used for the matrix of sums of squares and cross-products of a given matrix Z:

   $SSZ = Z'Z.$
2. Variables are standardized in such a way that SSZ has a diagonal with unit elements.
3. Given two sets of variables X and Y.
   (i) We define $Y\underline{P}X = X(X'X)^{-}X'Y$.
   For $Y\underline{P}X$ read: "Y projected on X". In fact, $Y\underline{P}X$ is just shorthand for the expression above, in which $(X'X)^{-}$ stands for the (generalized) inverse of X'X. One may also say: $Y\underline{P}X$ is the result of application to Y of the operator $X(X'X)^{-}X'$. Geometrically this corresponds to projecting the vectors in Y on the space spanned by the vectors in X.
   (ii) We define: $Y\underline{A}X = Y - Y\underline{P}X = (I-X(X'X)^{-}X')Y$.
   For $Y\underline{A}X$ read: "anti-projections of Y". It is shorthand for application of the operator $(I-X(X'X)^{-}X')$ to Y. Geometrically $Y\underline{A}X$ gives components of Y (in the joint space of Y and X) which are orthogonal to all vectors in the space spanned by X.
   (iii) In the text the following basic theorem is used:
   $SSY = SS(Y\underline{P}X) + SS(Y\underline{A}X)$
   (iv) Also the following two theorems are often used. Let X be partitioned as $X = (X_1, X_2)$. Then:
   $$Y\underline{P}X = Y\underline{P}X_1 + YP(X_2\underline{A}X_1)$$
   $$SS(Y\underline{P}X) = SS(Y\underline{P}X_1) + SS(Y\underline{P}(X_2\underline{A}X_1)).$$

REFERENCES

DUNCAN, O.D., HALLER, A.O., AND PORTES, A. (1968). Peer influences on aspiration: a reinterpretation. *American Journal of Sociology*, 74, 119-137.

GOLDBERGER, A.S. (1964). *Econometric Theory*. New York: Wiley and Sons.

VAN DE GEER, J.P. (1986). *Introduction to linear multivariate data analysis*. Leiden: DSWO Press. (2 vols.)

Ontvangen: 14-10-1988
Geaccepteerd: 25-03-1990