Queen's Economics Department Working Paper No. 17

PRIOR ADJUSTMENT: AN EXTENTION OF THE FRISCH-WAUGH THEOREM TO THE METHOD OF THE "TWO-STAGE LEAST SQUARES"

J.C.R. Rowley
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

3-1970

PRIOR ADJUSTMENT: AN EXTENSION OF THE FRISCH-WAUGH THEOREM TO THE METHOD OF "TWO-STAGE LEAST-SQUARES"

J. C. R. Rowley

Discussion Paper No. 17

March 1970

Department of Economics,
Queen's University,
Kingston, Ontario,
Canada.

The problems of seasonal adjustment and other forms of prior adjustment have seldom been integrated into a general framework of estimation. A well-known result, due to Frisch and Waugh, has been used to demonstrate how linear seasonal influences might be treated in the context of the general linear statistical model. One recent generalization of the Frisch-Waugh theorem indicates that all prior adjustment should take account of the estimating technique that would be used if all factors were treated at the same time. We establish a proposition that one form of prior adjustment is consistent with two different estimating techniques that are in common use.

Some Earlier Results

The parameters of a linear model are partitioned into two groups which form the elements of two column vectors β_1 and β_2. Columns of the signal matrix are rearranged to agree with this partition. Let X represent the adjusted signal matrix and let X_1 and X_2 be submatrices associated with β_1 and β_2. The model may be written in the following form:

$$(1) \qquad y = X\beta + u = X_1\beta_1 + X_2\beta_2 + u.$$

If K_1 and K_2 are the number of variables recorded in X_1 and X_2 and if K is their sum, these values are assumed to be less than or equal to the size of the sample T. A linear transformation, represented by the matrix H, is applied to this adjusted model. H is chosen so that HX_1 is a null matrix and HX_2 has rank K_2.

$$Hy = HX_1\beta_1 + HX_2\beta_2 + Hu$$

(2) $\quad Hy = HX_2\beta_2 + Hu.$

Both (1) and (2) involve the vector β_2 and the principle of least-squares may be applied to either equation in order to estimate this vector. Let $\hat{\beta}_2$ and $\tilde{\beta}_2$ represent the resulting estimators. 0 is a null matrix of order K_2 by K, I_2 is the identity matrix of order K_2 by K_2 and X is the partitioned matrix $(X_1 \quad X_2)$ which is assumed to have rank K.

(3)
$$\tilde{\beta}_2 \equiv \{(HX_2)'(HX_2)\}^{-1} (HX_2)'(Hy)$$
$$= (X_2'H'HX_2)^{-1} X_2' H' Hy$$

(4)
$$\hat{\beta}_2 \equiv (0 \quad I_2)(X'X)^{-1}X'y$$
$$= (0 \quad I_2)\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1} \begin{pmatrix} X_1'y \\ X_2'y \end{pmatrix}$$

Define

(5)
$$P_1 \equiv I - X_1(X_1'X_1)^{-1} X_1'$$

Notice that P_1X_1 is a null matrix and P_1 is a symmetric idempotent matrix. If P_1X_2 has rank K_2, two desirable results follow. First, P_1 satisfies the two constraints placed on the transformation H. Second, the theorem for the partitioned inverse may be used. I_1 is the identity matrix of order K_1 by K_1.

$$(X'X)^{-1} = \begin{pmatrix} (X_1'X_1)^{-1}\{I_1 + X_1'X_2(X_2'P_1X_2)^{-1}X_2'X_1(X_1'X_1)^{-1}\} & : & -(X_1'X_1)^{-1}X_1'X_2(X_2'P_1X_2)^{-1} \\ -(X_2'P_1X_2)^{-1} X_2'X_1(X_1'X_1)^{-1} & : & (X_2' P_1X_2)^{-1} \end{pmatrix}$$

$$\hat{\beta}_2 = -(X_2'P_1X_2)^{-1}X_2'X_1(X_1'X_1)^{-1}X_1'y + (X_2'P_1X_2)^{-1}X_2'y$$
$$= (X_2'P_1X_2)^{-1}X_2'\{-X_1(X_1'X_1)^{-1}X_1 + I\} y$$
$$= (X_2'P_1X_2)^{-1}X_2'P_1y$$

$$(6) \qquad \hat{\beta}_2 = (X_2'P_1'P_1X_2)^{-1} X_2'P_1'P_1y \qquad \text{since } P_1 \text{ is symmetric and idempotent.}$$

$$P_1y = y - X_1(X_1'X_1)^{-1}X_1'y$$

$$P_1X_2 = X_2 - X_1(X_1'X_1)^{-1}X_1'X_2$$

P_1y and P_1X_2 are the vectors of residuals obtained when y and X_2 are regressed on X_1.

PROPOSITION ONE (The Frisch-Waugh Theorem)

Least-squares estimators for a subset of parameters may be derived from a transformed model. This transformation has two important characteristics. The part of the signal matrix that is not associated with the chosen subset is annihilated. Observations for the dependent variable and the remaining independent variables are replaced by residuals obtained from a prior regression of these variables upon the excluded variables.

Frisch and Waugh (1933) were concerned with the prior adjustment of time-series to eliminate linear trends. Their result was extended by Tintner (1957) to include the elimination of polynomials with known degree and unknown coefficients. Kloek (1961) indicated the optimal transformation for the prior elimination of a constant. In this simple case, P_1 is the difference between an identity matrix and a matrix having the value $^1/_T$ for all its elements. Premultiplication of any matrix by this choice of P_1 replaces each element in the second matrix by its deviation from the corresponding column's mean value. Lovell (1963, 1966) illustrated the use of the Frisch-Waugh theorem in the adjustment of time-series for seasonal factors that affect the dependent variable linearly. He indicated the following generalization.

PROPOSITION TWO.

The application of Aitken's generalization of the least-squares principle to a linear model, in which the dispersion matrix of the errors is known except for a scalar factor, will yield identical estimators of a subset of parameters to those estimators derived by the application of the ordinary least-squares principle to a transformed model. Two adjustments to the original model are required in order to obtain this transformed model. First, the model is modified so that the resulting errors are spherical. Then the Frisch-Waugh procedure can be followed.

Let $\sigma^2 V$ represent the dispersion matrix of the original errors in the specification represented by (1), where V is a known symmetric positive-definite matrix. Then a non-singular matrix N can be found such that $(N'N)$ is the inverse of V and, consequently, (NVN') is a unit matrix. The dispersion matrix of Nu is $\sigma^2 I$.

$$Ny = NX_1\beta_1 + NX_2\beta_2 + Nu$$

or,

(7) $$y_* = X_{1*}\beta_1 + X_{2*}\beta_2 + u_* ,$$ with an obvious change in notation. This equation has a similar form to (1) and the new optimal linear transformation, applied to (7) so that the component involving β_2 is annihilated, is P_{1*} which is defined below.

$$P_{1*} \equiv I - X_{1*}(X_{1*}'X_{1*})^{-1} X_{1*}'$$

$$= I - NX_1(X_1'V^{-1}X_1)^{-1} X_1'N'$$

The complete transformation is represented by $P_{1*}N$, or $N\{I - X_1(X_1'V^{-1}X_1)^{-1}X_1'V^{-1}\}$, and the common estimator is

$$[X_2'\{V^{-1} - V^{-1} X_1(X_1'V^{-1}X_1)^{-1}X_1'V^{-1}\}X_2]^{-1} X_2'\{V^{-1} - V^{-1}X_1(X_1'V^{-1}X_1)^{-1} X_1'V^{-1}\}y.$$

The Method of "Two-Stage Least-Squares"

Consider a "mixed" linear model in which some explanatory variables are correlated with contemporaneous errors whereas other explanatory variables are either non-stochastic or uncorrelated with contemporaneous errors. This model can be represented by the following equation after appropriate rearrangement of the columns of the signal matrix. Let Y and Z_1 represent matrices containing observations for the two sets of explanatory variables and let α and δ represent the vectors arising from the concomitant partition of parameters.

$$(8) \qquad y = Y\alpha + Z_1\delta + \xi$$

$$y = (Y \quad Z_1)\begin{pmatrix}\alpha \\ \delta\end{pmatrix} + \xi$$

Assume that the variables recorded in Z_1 are a subset of a number of variables that are associated linearly with either the dependent variable or the explanatory variables recorded in Y. Without loss of generality, we can write the following identity.

$$(9) \qquad Z_1 \equiv Z\begin{pmatrix}I \\ 0\end{pmatrix}$$

Let observations for a particular subset of the variables recorded in Z_1 be collected in the matrix Z_*. Again, without loss of generality, we can write the following identity.

$$(10) \qquad Z_* \equiv Z_1\begin{pmatrix}I \\ 0\end{pmatrix} \qquad \text{and} \quad Z_* = Z\begin{pmatrix}I \\ 0\end{pmatrix}$$

The three identity matrices of (9) and (10) are unsubscripted but will have different orders.

Define the following two symmetric, non-stochastic, idempotent matrices.

(11) $P \equiv I - Z(Z'Z)^{-1} Z'$ provided $(Z'Z)$ is non-singular

(12) $M \equiv I - P = Z(Z'Z)^{-1} Z'$

$$MZ_1 = Z(Z'Z)^{-1}Z'Z \begin{pmatrix} I \\ 0 \end{pmatrix} = Z_1$$

$MP = 0$ since P is idempotent.

$$y = (M + P)Y\alpha + Z_1\delta + \xi$$

$$= MY\alpha + Z_1\delta + PY\alpha + \xi$$

Define (13) $e = PY\alpha + \xi$

(14) $y = MY\alpha + Z_1\delta + e$

or,

$$y = (MY \quad Z_1) \begin{pmatrix} \alpha \\ \delta \end{pmatrix} + e$$

Suppose the principle of least-squares is applied to this revised specification (14) instead of the original form (8). Let the resulting estimators be denoted $\tilde{\alpha}$ and $\tilde{\delta}$.

(15) $$\begin{pmatrix} \tilde{\alpha} \\ \tilde{\delta} \end{pmatrix} = \begin{pmatrix} Y'M'MY & Y'M'Z_1 \\ Z_1'MY & Z_1'Z_1 \end{pmatrix}^{-1} \begin{pmatrix} Y'M'y \\ Z_1'y \end{pmatrix}$$

$$= \begin{pmatrix} Y'MY & Y'Z_1 \\ Z_1'Y & Z_1'Z_1 \end{pmatrix}^{-1} \begin{pmatrix} Y'My \\ Z_1' y \end{pmatrix}$$

or,

$$(16) \quad \begin{pmatrix} \tilde{\alpha} \\ \tilde{\delta} \end{pmatrix} = \{(Y \quad Z_1)'Z(Z'Z)^{-1}Z'(Y \quad Z_1)\}^{-1} (Y \quad Z_1)'Z(Z'Z)^{-1}Z'y$$

These estimators have become known as "two-squares least-squares estimators" (Theil, 1958) or "generalized classical least-squares estimators" (Basmann, 1957). The first nomenclature stems from the use of the following orthogonal partition of Y.

$$\begin{aligned} Y &= \{Z(Z'Z)^{-1}Z'\}Y + \{I - Z(Z'Z)^{-1}Z'\}Y \\ &= MY + PY \end{aligned}$$

(MY) is the estimated signal component when Y is regressed on Z and (PY) is the matrix of residuals from the same regression.

These estimators will exist if the inverse in (16) exists. A necessary condition for this existence is that the number of variables recorded in $(Y : Z_1)$ must be less than, or equal to, the number of variables recorded in Z. That is, the number of additional variables introduced in Z must either exceed, or equal, the number of variables recorded in Y. This condition cannot be fulfilled by an arbitrary increase in these additional variables since choice is restricted by the constraints that $(Z'Z)$ is non-singular and $\{(Y : Z_1)'Z\}$ has full rank.

Prior Adjustment

Since the two-stage least-squares estimators are obtained from the application of the principle of least-squares to the specification (14), we can apply the Frisch-Waugh result to that equation. If Z_* is the matrix annihilated by the optimal transformation in this context, (5) implies that this optimal transformation may be characterized by the matrix

$\{I - Z_*(Z_*'Z_*)^{-1} Z_*'\}$. Denote this as Q.

$$(17) \qquad Q \equiv I - Z_*(Z_*'Z_*)^{-1}Z_*'$$

$$(18) \qquad Qy = QMY\alpha + QZ_1\delta + Qe$$

$$QM = \{I - Z_*(Z_*'Z_*)^{-1}Z_*'\}Z(Z'Z)^{-1}Z'$$

$$= Z(Z'Z)^{-1}Z' - Z_*(Z_*'Z_*)^{-1} Z_*' Z(Z'Z)^{-1}Z'$$

$$= Z(Z'Z)^{-1} Z' - Z_*(Z_*Z_*)^{-1}Z_* \qquad \text{using (10)}$$

or,

$$(19) \qquad QM = \{I - Z_*(Z_*'Z_*)^{-1}Z_*'\} - \{I - Z(Z'Z)^{-1}Z'\}$$

$$QMY = \{I - Z_*(Z_*'Z_*)^{-1}Z_*'\}Y - \{I - Z(Z'Z)^{-1}Z'\}Y$$

The necessary condition for the existence of the two-stage least-squares estimators implies that (QM) is not a null matrix except in the trivial case when Y is empty.

$$MQ = Z(Z'Z)^{-1}Z'\{I - Z_*(Z_*Z_*)^{-1} Z_*'\}$$

$$= Z(Z'Z)^{-1}Z' - Z(Z'Z)^{-1}Z'Z_*(Z_*'Z_*)^{-1}Z_*'$$

$$= Z(Z'Z)^{-1}Z' - Z_*(Z_*'Z_*)^{-1}Z_*' \qquad \text{using (10)}$$

$$= QM$$

Substitution for (QM) in (18) yields the following equation

$$(20) \qquad Qy = M(QY)\alpha + QZ_1\delta + Qe$$

The introduction of some additional terminology facilitates a simple statement of two results established above. We shall use the terms "endogenous", "predetermined" and "included predetermined" to describe the variables recorded in the matrices $(y : Y)$, Z and Z_1 respectively.

PROPOSITION THREE.

The "two-stage least-squares estimators" for the parameters associated with endogenous variables and some included predetermined variables in a linear model may be obtained from application of the principle of least-squares to an adjusted model. In order to obtain the latter model, different transformations are applied to two distinct sets of variables and a third set is annihilated. The dependent variable and included predetermined variables are replaced by the residuals obtained when each is regressed on the predetermined variables that are to be annihilated. The explanatory endogenous variables are replaced by the differences between two sets of residuals obtained from two prior regressions in which these variables are regressed on all predetermined variables and on the predetermined variables that are to be annihilated.

PROPOSITION FOUR.

The "two-stage least-squares estimators" for the parameters associated with endogenous variables and some included predetermined variables in a linear model may be obtained by application of the method of "two-stage least-squares" to a transformed model. In order to obtain the latter model, observations for each variable are replaced by the residuals obtained from a prior regression of that variable upon the particular set of included predetermined variables that are to be annihilated.

The fourth proposition is especially important for the econometrician who seeks to remove linear seasonal factors in each variable before further investigation. If these factors are eliminated by a prior application of the least-squares principle, the Frisch-Waugh theorem can be extended to the method of two-stage least-squares. The strict order required by Lovell for his extension of the theorem to Aitken's generalized estimators indicates that the choice of estimating procedure affects the choice of prior adjustment and this is illustrated by the third proposition. However, the commutivity of the two transformations for the two-stage least-squares estimates eliminates the order in which they need to be applied. This result does not hold for other members of the general k-class of estimators.

REFERENCES

R. FRISCH and F. V. WAUGH:

"Partial time regressions as compared with individual trends",
Econometrica,
Volume 1, 1933.
Pages 387-401.

G. TINTNER:

"Econometrics".
John Wiley, 1957.
Pages 301-308.

T. KLOEK:

"Note on convenient matrix notation in multivariate statistical analysis and in the theory of linear aggregation".

International Economic Review,
Volume 2, Number 3, September 1961.
Pages 351-360.

M. C. LOVELL:

(1) "Seasonal adjustment of economic time series and multiple regression analysis".

Journal of the American Statistical Association,
Volume 58, December 1963.
Pages 993-1010.

(2) "Alternative axiomatizations of seasonal adjustment".

Journal of the American Statistical Association,
Volume 61, September 1966.
Pages 800-802.

R. L. BASMANN:

"A generalized classical method of linear estimation of coefficients in a structural equation".

Econometrica,
Volume 25, January 1957.
Pages 77-83.

H. THEIL:

"Economic Forecasts and Policy". 1st Revised Edition.
North Holland, 1958.