Queen's Economics Department Working Paper No. 66

MOVING-AVERAGE TRANSFORMATIONS IN CLASSICAL LINEAR MODELS

J.C.R. Rowley
Queen's University

D.A. Wilton
Queen's University

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

11-1971

MOVING-AVERAGE TRANSFORMATIONS IN CLASSICAL LINEAR MODELS

J. C. R. Rowley and D. A. Wilton
Queen's University

Discussion Paper No. 66

November 1971

MOVING-AVERAGE TRANSFORMATIONS IN THE CLASSICAL LINEAR MODEL

Economists are frequently compelled to use data which take inappropriate forms. One particular deficiency is discussed below; namely, the prior adjustment of economic time-series by moving-average transformations. This discussion is restricted to regression analyses and data which satisfy the conditions for the classical linear model.

Let y denote a vector which contains n observations for the dependent variable in a particular study and let X represent a matrix, of order n by k, which contains observations for k explanatory variables. Both y and X are assumed to be unavailable. Instead, they have been subjected to a prior moving-average transformation with known non-stochastic weights. This transformation can be represented by a non-stochastic matrix G, say, which has order $(n-q)$ by n where q is the length of the moving-average. If $\{d_{ij}\}$ is a collection of known weights, G will have the following form.

$$(1) \qquad G = \begin{bmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{iq} & \searrow & & 0 \\ & d_{21} & d_{22} & \cdots & d_{2,q-1} & d_{2q} & \searrow & \\ & & d_{31} & \cdots & & & & \\ 0 & & & \searrow & d_{n-q,1} & & \cdots & d_{n-q,q} \end{bmatrix}$$

Then the elements of Gy and GX represent the data in the form in which they are available for use. G and GX are assumed to have ranks $(n-q)$ and k respectively.

A "micro-equation" (2) and a "macro-equation" (3) must be distinguished. The former is the specification which would be used if data were available in an appropriate form whereas the latter is the specification which is actually used.[1]

$$(2) \qquad y = X\beta + u \quad ,$$

where β is a vector of k parameters and u is a vector of random errors.

$$(3) \qquad Gy = GX\beta + Gu.$$

If X is a non-stochastic matrix, u has a zero mean vector and u has a scalar matrix for its dispersion matrix, then (2) satisfies the conditions for the classical linear model and (3) satisfies the conditions for Aitken's generalization[2] of this model. These assumptions are maintained throughout the discussion which follows. They are supplemented by the requirement that the distribution of errors be gaussian.

The presence of zeros in the transformation G is inessential[3] and this matrix can represent any matrix of appropriate order and rank. We shall require that the transpose of G is not equal to the Moore-Penrose inverse[4] of G. In this excluded case, the macro-equation would satisfy the conditions for the classical linear model. One popular collection of transformations is affected by this exclusion; namely, the "groupings"[5] for which G is quasidiagonal with non-zero blocks provided by row vectors having unit euclidean norms. Exclusion permits attention to be focused on the problems of autocorrelation and heteroscedasticity which are introduced by the prior transformation of data.

Let $E(\theta)$ and $D(\theta)$ represent the mean vector and dispersion matrix of an arbitrary stochastic vector θ. Let $\hat{\beta}$ represent the least-squares estimator of β from the macro-equation.

$$(4) \qquad \hat{\beta} = (X'HX)^{-1} X'Hy \qquad \text{if } H \equiv G'G.$$

$$(5) \quad E(\hat{\beta}) = \beta \qquad \text{if } E(u) = 0.$$

$$(6) \quad D(\hat{\beta}) = \sigma^2(X'HX)^{-1}X'HHX(X'HX)^{-1}$$

if $D(u) = \sigma^2 I$ for some finite scalar σ^2.

The Gauss-Markov theorem[6] indicates that the diagonal elements of $D(\hat{\beta})$ cannot be less than the corresponding ones of the dispersion matrix for the least-squares estimators which are based on the micro-equation (2). Orcutt has discussed this potential loss of efficiency and it will not be reviewed here as the more efficient estimator cannot be calculated if the micro-data is unavailable. Both estimators are unbiased. Our primary concern is with the influence of non-spherical errors, which are directly attributable to the prior transformation, upon conventional test statistics that are calculated on the false presumtion that the errors are spherical.

Let $\hat{S}$ represent the residual sum of squares for the least-squares fit obtained from the macro-equation.

$$(7) \qquad \hat{S} \equiv (Gy - GX\hat{\beta})'(Gy - GX\hat{\beta})$$

$$= y'\{H - HX(X'HX)^{-1} X'H\}y$$

$$(8) \qquad \hat{S} = u'\{H - HX(X'HX)^{-1} X'H\}u$$

The gram matrix H is idempotent if and only if the transpose of G is equal to its Moore-Penrose inverse.[7] If H is idempotent, then $\hat{S}$ is proportional to a chi-square variate with $(n - q - k)$ degrees of freedom.[8]

If the explanatory variables are partitioned into two categories, observations for these categories might be assembled in the matrices X_1 and

X_2 respectively when available. The parameters can be partitioned similarly.

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

$$(2) \qquad Gy = GX_1\beta_1 + GX_2\beta_2 + Gu.$$

Consider the linear hypothesis $\mathcal{H}$ that the k_2 elements of β_2 are zero.[9] The least-squares estimator which is constrained by this hypothesis is given by

$$(9) \qquad \tilde{\beta}_1 = (X_1'HX_1)^{-1}X_1'Hy \qquad \text{and} \quad \tilde{\beta}_2 = 0 ,$$

and the constrained sum of squared residuals for this fit is

$$(10) \qquad \tilde{S} \equiv (Gy - GX_1\tilde{\beta}_1)'(Gy - GX_1\tilde{\beta}_1).$$

$$= y'\{H - HX_1(X_1'HX_1)^{-1}X_1'H\}y$$

$$(11) \qquad \tilde{S} = u'\{H - HX_1(X_1'HX_1)^{-1}X_1'H\}u \qquad \text{under } \mathcal{H}.$$

If H is idempotent, then $\tilde{S}$ is proportional to a chi-squared variate with $(n - q - k_2)$ degrees of freedom under the null hypothesis.

$$(12) \qquad \frac{\tilde{S} - \hat{S}}{\hat{S}} = \frac{y'\{HX(X'HX)^{-1}X'H - HX_1(X_1'HX_1)^{-1}X_1'H\}y}{y'\{H - HX(X'HX)^{-1}X'H\}y}$$

$$= \frac{y'\ Ay}{y'\ By} \qquad \text{with implicit definitions}$$

of A and B derived from (12). The numerator and denominator of this ratio are independent if and only if the product of A and B is a null matrix by Craig's theorem.[10]

The ratio is an appropriate basis for a significance test of the null hypothesis if it is distributed as Fisher's F when adjusted for degrees of freedom. Necessary and sufficient conditions for appropriateness are, therefore, given by the symmetry and idempotency of A and B and the mutual orthogonality of these two matrices. These three conditions hold if the gram matrix H is idempotent; that is, if the transpose of G is equal to its Moore-Penrose inverse which is denoted G^-. We have excluded this case and all three conditions will usually be violated by an arbitrary choice of G. Three equivalent approaches might be taken to overcome this problem.

APPROACH ONE.

Aitken suggested a generalized least-squares estimator[11] which is weighted by the inverse of the dispersion matrix for the macro-errors Gu. The test statistic is based on the residual sums of squares for this generalized least-squares fit in unconstrained and constrained cases. These sums are weighted by the same matrix as the Aitken estimators, which can be chosen as the inverse of GG' since the scale of this matrix leaves both estimators and residual sums of squares unaffected. Let $\hat{\beta}_g$ and $\tilde{\beta}_{1g}$ represent the Aitken estimators and let $\hat{S}_g$ and $\tilde{S}_g$ represent the two weighted sums of squared residuals.

$$(13) \qquad \hat{\beta}_g = (X'H_gX)^{-1}X'H_gy \qquad \text{where } H_g \equiv G'(GG')^{-1}G$$

$$(14) \qquad \tilde{\beta}_{1g} = (X_1'H_gX_1)^{-1}X_1'H_gy$$

$$(15) \qquad \hat{S}_g \equiv (Gy - GX\hat{\beta}_g)'(GG')^{-1}(Gy - GX\hat{\beta}_g)$$

$$= y'\{H_g - H_gX(X'H_gX)^{-1}X'H_g\}y$$

$$(16) \qquad \hat{S}_g = u'\{H_g - H_gX(X'H_gX)^{-1}X'H_g\}u$$

$$(17) \qquad \tilde{S}_g = y'\{H_g - H_gX_1(X_1'H_gX_1)^{-1}X_1'H_g\}u$$

$$(18) \qquad \tilde{S}_g = u'\{H_g - H_gX_1(X_1'H_gX_1)^{-1}X_1'H_g\}u \qquad \text{under } \mathcal{X}.$$

$$(19) \qquad \frac{\tilde{S}_g - \hat{S}_g}{\hat{S}_g} = \frac{u' A_g u}{u' B_g u} \qquad \text{under } * ,$$

where A_g and B_g are defined implicitly by reference to (16) and (17). This new ratio is distributed as Fisher's F, when adjusted for degrees of freedom, under the null hypothesis. It can be readily confirmed that A_g and B_g are symmetric and idempotent and the product A_gB_g is a null matrix.

APPROACH TWO.

The Cholesky[12] technique can be used to find a nonsingular matrix N such that N(GG')N' is a scalar matrix. This matrix is used to transform the macro-equation into a form with spherical errors:

$$(20) \qquad NGy = NGX\beta + NGu .$$

$$(21) \qquad D(NGu) = \sigma^2 NGG'N' = \sigma^2 I \text{ , say.}$$

The least-squares theory can be applied to this revised specification.

APPROACH THREE.

Since G has rank (n-q) and order (n-q) by n, its Moore-Penrose inverse G^- is given by[13] $G'(GG')^{-1}$. The macro-equation is transformed by this inverse into another form with non-spherical errors:

$$(22) \qquad G^-Gy = G^-GX\beta + G^-Gu$$

$$(23) \qquad D(G^-Gu) = \sigma^2 G^-G(G^-G)'$$

$$= \sigma^2 G^-G$$

since G^-G is symmetric and idempotent.[14] The least-squares theory can be applied to this further specification even though the errors do not satisfy classical conditions. The unweighted sums of squared residuals are an appropriate basis for the test statistic. Equivalence between the second and third approaches is established by the equality of ($G^{-}G$) and ($G'N'NG$), which follows directly from the definitions of G^{-} and N.

Approach One is seldom used because of the numerical difficulty associated with the inversion of a large matrix (GG'). This difficulty is reduced in severity by the use of the Cholesky technique in Approach Two, which permits the symmetry and other characteristics (e.g. "bandedness") of the matrix to be utilized for efficient calculation, and this approach has enjoyed popular support. Approach Three is novel and might be used increasingly as new algorithms for the Moore-Penrose inverse become available and as the properties of less restrictive generalized inverses are investigated.

FOOTNOTES

1. Rowley and Wilton provide an ample list of studies of wage-determination which implicitly contain this distinction.

2. Rowley, ch. 2.

3. Our primary interest in moving-averages stems from a concomitant investigation of their use in earlier empirical studies of wage-determination.

4. Graybill provides an excellent account of this inverse and other generalized inverses which are less restrictive. Various theorems are cited from Graybill's book in later footnotes and no attempt is made here to attribute these theorems to earlier authors. More advanced expositions are supplied by Pringle and Rayner and contributions to the report edited by Boullion and Odell.

5. Rowley, ch. 7.

6. Rowley, ch. 1.

7. Graybill, th. 6.4.10.

8. Graybill, th. 6.2.7 corollary, and Hogg and Craig, p. 387.

9. This hypothesis is chosen on grounds of simplicity. There is little difficulty in expanding this discussion to a more general linear hypothesis.

10. Hogg and Craig, p. 391.

11. Rowley, ch. 2.

12. Wilkinson, pp. 553-556.

13. Graybill, th. 6.2.16.

14. Graybill, th. 6.2.17.

REFERENCES

Boullion, T. L., and P. L. Odell, *Proceedings of the Symposium on Theory and Application of Generalized Inverses of Matrices*, Lubbock: The Texas Tech Press, 1968.

Graybill, F. A., *Introduction to Matrices with Applications in Statistics*, Belmont: Wadsworth, 1969.

Hogg, R. V., and A. T. Craig, *Introduction to Mathematical Statistics*, Third Edition, New York: Macmillan, 1970.

Orcutt, G. H., *Data Needs for Computer Simulation of Large-Scale Social Systems*, in (ed.) J. E. Beshers, *Computer Methods in the Analysis of Large-Scale Social Systems*, Cambridge: Joint Centre of Urban Studies, MIT-Harvard, 1965, Section VIII.

Pringle, R. M., and A. A. Rayner, *Generalized Inverse Matrices with Applications to Statistics*, London: Griffin, 1971.

Rowley, J. C. R., *Econometric Estimation*, London: Weidenfeld and Nicolson, (to be published in 1971).

Rowley, J. C. R., and D. A. Wilton, *Quarterly Models of Wage Determination: Some New Efficient Estimates*, Discussion Paper No. 51, Institute for Economic Research, Queen's University, Kingston, Ontario, 1971.

Wilkinson, G. N., *The Algebraic Eigenvalue Problem*, Oxford: Clarendon Press, 1965.