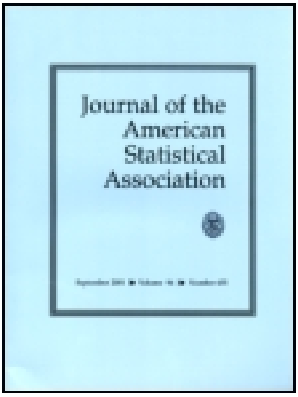

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias

Arnold Zellner [a]
[a] University of Wisconsin , USA
Published online: 10 Apr 2012.

To cite this article: Arnold Zellner (1962) An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias, Journal of the American Statistical Association, 57:298, 348-368

To link to this article: http://dx.doi.org/10.1080/01621459.1962.10480664

AN EFFICIENT METHOD OF ESTIMATING SEEMINGLY UNRELATED REGRESSIONS AND TESTS FOR AGGREGATION BIAS*

ARNOLD ZELLNER
University of Wisconsin

In this paper a method of estimating the parameters of a set of regression equations is reported which involves application of Aitken's generalized least-squares [1] to the whole system of equations. Under conditions generally encountered in practice, it is found that the regression coefficient estimators so obtained are at least asymptotically more efficient than those obtained by an equation-by-equation application of least squares. This gain in efficiency can be quite large if "independent" variables in different equations are not highly correlated and if disturbance terms in different equations are highly correlated. Further, tests of the hypothesis that all regression equation coefficient vectors are equal, based on "micro" and "macro" data, are described. If this hypothesis is accepted, there will be no aggregation bias. Finally, the estimation procedure and the "micro-test" for aggregation bias are applied in the analysis of annual investment data, 1935–1954, for two firms.

1. INTRODUCTION

GIVEN a set of regression equations, we consider the problem of estimating regression coefficients efficiently. It is only under special conditions, stated explicitly below, that classical least-squares applied equation-by-equation yields efficient coefficient estimators. For conditions generally encountered, we propose an estimation procedure which yields coefficient estimators at least asymptotically more efficient than single-equation least-squares estimators. In this procedure regression coefficients in all equations are estimated simultaneously by applying Aitken's generalized least-squares [1] to the whole system of equations. To construct such Aitken estimators, we employ estimates

* This paper was written when the author was visiting Fulbright Professor at the Netherlands School of Economics and Associate Professor on leave, University of Wisconsin. It is with much pleasure that he acknowledges the benefit derived from discussions (and cigars) with Professor H. Theil and Mr. P. J. M. van den Bogaard. Thanks are also due to Miss E. van der Hoeven for assistance with the calculations and expert typing.

of the disturbance terms' variances and covariances based on the residuals derived from an equation-by-equation application of least-squares.[1] While we apply this estimation procedure in the analysis of temporal cross-section data, annual micro-investment data, 1935–1954, we recognize that the procedure is more generally applicable. For example, it can be applied in the analysis of data provided by a single cross-section budget study when regressions for several commodities are to be estimated. Another application would be in time-series regression analyses of the demands for a variety of consumption (or investment) goods. A fourth application is to regression equations in which each equation refers to a particular classification category and the observations refer to different points in space, as in Barten and Koerts' analysis of voters' transitions from party to party within various voting districts [2].

Further, within the estimation framework a test of the equality of regression coefficient vectors, and thus of the absence of one important type of aggregation bias, is described and applied in the analysis of micro-investment relations. Like the estimation procedure, this testing procedure is more generally applicable. Finally, a procedure for testing for aggregation bias which utilizes just macro-data is developed.

The plan of the paper is as follows. In Section 2 we describe the system and the proposed estimation procedure. Section 3 is devoted to establishing the properties of estimators constructed in Section 2 and to providing an explicit statement of the gain in efficiency over single-equation least-squares estimation. We then turn to some aspects of the aggregation problem in Section 4, in particular to consideration of two tests for aggregation bias, one employing micro-data, the other macro-data. Then the estimation and one testing procedure are applied in Section 5. Lastly, we present some concluding remarks in Section 6.

2. EFFICIENT ESTIMATION OF SEEMINGLY UNRELATED REGRESSION EQUATIONS

Let

$$y_\mu = X_\mu \beta_\mu + u_\mu \tag{2.1}$$

be the μ'th equation of an M equation regression system with y_μ a $T\times 1$ vector of observations on the μ'th "dependent" variable, X_μ a $T\times l_\mu$ matrix with rank l_μ, of observations on l_μ "independent" nonstochastic variables, β_μ a $l_\mu\times 1$ vector of regression coefficients and u_μ, a $T\times 1$ vector of random error terms, each with mean zero. The system of which (2.1) is an equation may be written as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_M \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_M \end{bmatrix} \tag{2.2}$$

$$y = X\beta + u \tag{2.3}$$

[1] This procedure, modified in certain respects, has been applied to estimate the parameters of "simultaneous equation" econometric models in reference [13].

where $y \equiv [y_1' y_2' \cdots y_M']'$, $\beta \equiv [\beta_1' \beta_2' \cdots \beta_M']'$, $u \equiv [u_1' u_2' \cdots u_M']'$ and X represents the block-diagonal matrix on the r.h.s. of (2.2). The $MT \times 1$ disturbance vector in (2.2) and (2.3) is assumed to have the following variance-covariance matrix:

$$\Sigma = V(u) = \begin{bmatrix} \sigma_{11}I & \sigma_{12}I & \cdots & \sigma_{1M}I \\ \sigma_{21}I & \sigma_{22}I & \cdots & \sigma_{2M}I \\ \vdots & \vdots & & \vdots \\ \sigma_{M1}I & \sigma_{M2}I & \cdots & \sigma_{MM}I \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2M} \\ \vdots & \vdots & & \vdots \\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MM} \end{bmatrix} \otimes I \qquad (2.4)$$
$$= \Sigma_c \otimes I,$$

where I is a unit matrix of order $T \times T$ and $\sigma_{\mu\mu'} = E(u_{\mu t} u_{\mu' t})$ for $t = 1, 2, \cdots, T$ and $\mu, \mu' = 1, 2, \cdots, M$.

In temporal cross-section regressions, t represents time and (2.3) implies constant variances and covariances from period to period as well as the absence of any auto or serial correlation of the disturbance terms. The $\sigma_{\mu\mu'}$ with $\mu = \mu'$ are then the variances and with $\mu \neq \mu'$ the covariances of the micro-units' disturbance terms (or dependent variables) for any time period. In a single cross-section budget study where t represents the t'th household and each equation "explains" expenditure on a particular commodity, $\sigma_{\mu\mu'}$ is the covariance between the disturbance term in the equation for commodity μ (or expenditure on commodity μ) and that in the equation for commodity μ' (or expenditure on μ') while $\sigma_{\mu\mu}$ is the variance of the disturbance term in the equation for expenditure on commodity μ (or alternatively, the variance of expenditure on commodity μ). The form of (2.4) implies that the $\sigma_{\mu\mu'}$ are the same for all households and that there is no correlation between different households' disturbances (or expenditures). Lastly, in application to geographic problems, t stands for the t'th geographic region and the form of (2.3) is such that there are correlations between disturbances or dependent variables relating to a particular region but not to different regions. Also disturbance variances and covariances are assumed to be constant from region to region.

In a formal sense we now regard (2.2) or (2.3) as a single-equation regression model and apply Aitken's generalized least-squares [1]. That is, we pre-multipy both sides of (2.3) by a matrix H which is such that $E(Huu'H') = H\Sigma H' = I$. In terms of transformed variables, the original variables pre-multiplied by H, the system now satisfies the usual assumptions of the least-squares model. Thus application of least-squares[2] will yield, as is well-known, a best linear unbiased estimator, which is

$$b^* = (X'H'HX)^{-1}X'H'Hy = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y. \qquad (2.5)$$

In constructing this estimator, we need the inverse of Σ which is given by:

[2] The quadratic form to be minimized in the Aitken approach is not the sum of squares of the original disturbance terms but, as is well-known, that of the transformed disturbances, namely $u'H'Hu$, or $u'\Sigma^{-1}u$. As will be pointed out below, there are good common-sense reasons for applying least squares to the transformed variables, reasons which make clear why it is that the Aitken estimator is more efficient than the classical least-squares estimator based on the original variables.

$$\Sigma^{-1} = V^{-1}(u) = \begin{bmatrix} \sigma^{11}I & \cdots & \sigma^{1M}I \\ \vdots & & \vdots \\ \sigma^{M1}I & \cdots & \sigma^{MM}I \end{bmatrix} = \Sigma_c^{-1} \otimes I. \tag{2.6}$$

Then the Aitken estimator of the coefficient vector, given in (2.5), is

$$b^* = \begin{bmatrix} b_1^* \\ b_2^* \\ \vdots \\ b_M^* \end{bmatrix} = \begin{bmatrix} \sigma^{11}X_1'X_1 & \sigma^{12}X_1'X_2 & \cdots & \sigma^{1M}X_1'X_M \\ \sigma^{21}X_2'X_1 & \sigma^{22}X_2'X_2 & \cdots & \sigma^{2M}X_2'X_M \\ \vdots & \vdots & & \vdots \\ \sigma^{M1}X_M'X_1 & \sigma^{M2}X_M'X_2 & \cdots & \sigma^{MM}X_M'X_M \end{bmatrix}^{-1} \times \begin{bmatrix} \sum_{\mu=1}^{M} \sigma^{1\mu}X_1'y_\mu \\ \vdots \\ \sum_{\mu=1}^{M} \sigma^{M\mu}X_M'y_\mu \end{bmatrix} \tag{2.7}$$

and the variance-covariance matrix of the estimator b^* is easily shown to be $(X'\Sigma^{-1}X)^{-1}$ or

$$V(b^*) = \begin{bmatrix} \sigma^{11}X_1'X_1 & \sigma^{12}X_1'X_2 & \cdots & \sigma^{1M}X_1'X_M \\ \sigma^{21}X_2'X_1 & \sigma^{22}X_2'X_2 & \cdots & \sigma^{2M}X_2'X_M \\ \vdots & \vdots & & \vdots \\ \sigma^{M1}X_M'X_1 & \sigma^{M2}X_M'X_2 & \cdots & \sigma^{MM}X_M'X_M \end{bmatrix}. \tag{2.8}$$

The estimator in (2.7) possesses all of the usual optimal properties of Aitken estimators; that is, it is a best linear unbiased estimator.[3] Further, with an added normality assumption, it is also a maximum-likelihood estimator. It is to be noted that (2.6) is identical with estimators provided by single-equation least-squares if the disturbance terms have a diagonal variance-covariance matrix, i.e., if $\sigma_{\mu\mu'}=\sigma_{\mu'\mu}=0$ for $\mu'\neq\mu$. Also, if $X_1=X_2=\cdots=X_M$, (2.6) "collapses" to yield single-equation least-squares estimators even if disturbance terms in different equations are correlated ($\sigma_{\mu'\mu}\neq 0$), and these are, as is well-known, the same as maximum-likelihood estimators. However, when the X_μ are not all the same and when the disturbance terms in different equations are correlated, the estimator in (2.6) will differ from the single-equation least-squares estimators.

If Σ is unknown, as it usually is, it is impossible to use (2.6) and (2.7) in practice. What we propose to do is to employ an estimate of $\{\sigma^{\mu\mu'}\}$ in constructing the Aitken estimator. This estimate is,[4]

[3] The single-equation least-squares estimator of the coefficient vector in (2.2) is a member of the linear class of estimators to which the Aitken estimator belongs.

[4] Here, for simplicity, we assume that there are κ "independent" variables in each regression. Then the variance estimators in (2.9) will be unbiased. However, the covariance estimators will only be asymptotically unbiased. When the independent variables in different equations are highly correlated, as is the case in many applications, the small-sample bias in the covariance estimators in (2.9) will be small. In reference [14, p. 14] an unbiased covariance estimator is presented.

$$(T-l)\Sigma_e = (T-l)\{s_{\mu\mu'}\} = \{\hat{u}_\mu' \hat{u}_{\mu'}\} = \{(y_\mu - X_\mu\hat{\beta}_\mu)'(y_{\mu'} - X_{\mu'}\hat{\beta}_{\mu'})\} \quad (2.9)$$

where $\hat{\beta}_\mu$ is the usual single-equation least-squares estimator, $(X_{\mu'}X_\mu)^{-1}X_{\mu'}y_\mu$. Thus (2.9) is an estimate of the disturbance variance-covariance matrix formed from the single-equation least-squares residuals. Given that we have the estimate $\{s_{\mu\mu'}\}$, we can obtain by inversion the matrix $\{s^{\mu\mu'}\}$ the elements of which are employed to form the estimator:

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{bmatrix} = \begin{bmatrix} s^{11}X_1'X_1 & s^{12}X_1'X_2 & \cdots & s^{1M}X_1'X_M \\ s^{21}X_2'X_1 & s^{22}X_2'X_2 & \cdots & s^{2M}X_2'X_M \\ \vdots & \vdots & & \vdots \\ s^{M1}X_M'X_1 & s^{M2}X_M'X_2 & \cdots & s^{MM}X_M'X_M \end{bmatrix}^{-1} \begin{bmatrix} \sum_{\mu=1}^{M} s^{1\mu}X_1'y_\mu \\ \vdots \\ \sum_{\mu=1}^{M} s^{M\mu}X_M'y_\mu \end{bmatrix}. \quad (2.10)$$

It will be shown that $b=b^*+0(T^{-1})$, that $T^{1/2}(b-\beta)$ and $T^{1/2}(b^*-\beta)$ have the same asymptotic normal distribution, and that the moment matrix of b is:

$$V(b) = \begin{bmatrix} s^{11}X_1'X_1 & s^{12}X_1'X_2 & \cdots & s^{1M}X_1'X_M \\ s^{21}X_2'X_1 & s^{22}X_2'X_2 & \cdots & s^{2M}X_2'X_M \\ \vdots & \vdots & & \vdots \\ s^{M1}X_M'X_1 & s^{M2}X_M'X_2 & \cdots & s^{MM}X_M'X_M \end{bmatrix}^{-1} + o(T^{-1}), \quad (2.11)$$

where $0(T^{-1})$ denotes a quantity which is of the order T^{-1} in probability and $o(T^{-1})$ denotes terms of higher order of smallness than T^{-1}.

3. PROPERTIES OF THE TWO-STAGE AITKEN ESTIMATOR

3.1. Moment Matrix and Asymptotic Distribution

We now turn to providing proofs of the statements made at the end of Section 2 regarding the properties of the estimator in (2.9).

Let $\Sigma_e=(\Sigma_c+\Delta_1)\otimes I$ be the estimated disturbance covariance matrix where $\Sigma_c\otimes I$ is given by (2.3) and Δ_1 is a matrix whose elements are the sampling errors of the single-equation least-squares estimators of the elements of Σ_c and these sampling errors are known to be $0(T^{-1/2})$ in probability ($\equiv$i.p.). Thus,

$$\begin{aligned} \Sigma_e^{-1} &= (\Sigma_c + \Delta_1)^{-1} \otimes I = [\Sigma_c^{-1} - \Sigma_c^{-1}\Delta_1\Sigma_c^{-1} + \cdots] \otimes I \\ &= \Sigma^{-1} - \Delta_2 \cdots \end{aligned} \quad (3.1)$$

and

$$\Delta_2 = \left\{ \sum_{i=1}^{M} \sum_{j=1}^{M} \sigma^{\mu i} \delta_{ij}^{(1)} \sigma^{j\mu'} \right\} \otimes I = \{\delta_{\mu\mu'}^{(2)}\} \otimes I$$

where $\delta_{\mu\mu'}^{(2)}$ is $0(T^{-1/2})$ i.p. and terms of higher order of smallness have been neglected. Now the two-stage Aitken estimator is

$$b = (X'\Sigma_e^{-1}X)^{-1}X'\Sigma_e^{-1}y$$

or

$$b - \beta = (X'\Sigma_e^{-1}X)^{-1}X'\Sigma_e^{-1}u$$

where $y \equiv (y_1', y_2', \cdots, y_M')'$, $u \equiv (u_1', u_2', \cdots, u_M')'$, $b \equiv (b_1', b_2', \cdots, b_M')'$, $\beta \equiv (\beta_1', \beta_2', \cdots, \beta_M')'$ and X denotes the block-diagonal matrix on the r.h.s. of (2.2). Then utilizing (3.1), we have

$$\begin{aligned} b - \beta &= [X'(\Sigma^{-1} - \Delta_2)X]^{-1}X'(\Sigma^{-1} - \Delta_2)u \\ &= \{[X'\Sigma^{-1}X][I - (X'\Sigma^{-1}X)^{-1}X'\Delta_2 X]\}^{-1}X'(\Sigma^{-1} - \Delta_2)u \qquad (3.2) \\ &= [(X'\Sigma^{-1}X)^{-1} + (X'\Sigma^{-1}X)^{-1}(X'\Delta_2 X)(X'\Sigma^{-1}X)^{-1} + \cdots]X'(\Sigma^{-1} - \Delta_2)u \end{aligned}$$

where terms of higher order of smallness than $0(T^{-1\frac{1}{2}})$ have been deleted in the square brackets. Rearranging terms, we find

$$b - \beta = b^* - \beta + \Delta_3, \qquad (3.3)$$

where b^* is the "pure" Aitken estimator and

$$\Delta_3 = -(X'\Sigma^{-1}X)^{-1}X'\Delta_2 u + (X'\Sigma^{-1}X)^{-1}X'\Delta_2 X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}u, \qquad (3.4)$$

terms of higher order of smallness being again neglected. Considering the second term on the r.h.s. of (3.3) we observe that $X'\Sigma^{-1}u$ is $0(T^{1/2})$ i.p., which means that the term as a whole is $0(T^{-1})$ i.p. The same applies to the first term, for the order of $X'\Delta_2 u$ is that of Δ_2 multiplied by that of $X'u$, that is, $0(1)$.

If we then take the expectation of both sides of (3.2), we find that the bias of b is at most of $0(T^{-1})$. Furthermore, since $b^* - \beta$ is $0(T^{-1/2})$ i.p. and Δ_3 is $0(T^{-1})$ i.p., the asymptotic covariance matrix of $b - \beta$ is the same as that of $b^* - \beta$. Finally, since it is known that under general conditions the asymptotic distribution of $T^{1/2}(b^* - \beta)$ is normal, the asymptotic distribution of $T^{1/2}(b - \beta)$ is the same as that of $T^{1/2}(b^* - \beta)$, because the difference of these two quantities, $T^{1/2}\Delta_3$, has zero probability limit; see the convergence theorem in reference [5, p. 254].

3.2. The Gain in Efficiency

Since the Aitken estimator of $\{\beta_2', \beta_2', \cdots, \beta_M'\}'$ in (2.2) differs from that derived by application of least-squares equation-by-equation, it must be the case that the Aitken estimator is more efficient. Essentially, this gain in efficiency occurs because in estimating the coefficients of a single equation, the Aitken procedure takes account of zero restrictions on coefficients occurring in other equations. These zero restrictions can be seen clearly if the system in (2.2) is rewritten as:

$$(y_1 y_2 \cdots y_M) = (X_1 X_2 \cdots X_M)\begin{bmatrix} \beta_1 & 0 & \cdots & 0 \\ 0 & \beta_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \beta_M \end{bmatrix} + (u_1 u_2 \cdots u_M). \qquad (3.5)$$

It is instructive to consider a system[5] with a disturbance covariance matrix such that $\sigma_{\mu\mu}=\sigma^2$ and $\sigma_{\mu\mu'}=\sigma^2\rho$ for $\mu\neq\mu'$, or $\Sigma_c=\sigma^2\,[(1-\rho)I+\rho ee']$ where I is a unit matrix of size $M\times M$ and $e'=[1, 1, \cdots, 1]$, a $1\times M$ vector. Then $\Sigma_c^{-1}=\alpha I-\gamma ee'$ with $\alpha^{-1}=\sigma^2(1-\rho)$ and $\gamma=\alpha\rho/[1+(M-1)\rho]$. Then for the covariance matrix of the estimator b^*, we have

$$V(b^*) = [X'(\Sigma_c^{-1}\otimes I)X]^{-1}$$

$$= \begin{bmatrix} (\alpha-\gamma)X_1'X_1 & -\gamma X_1'X_2 & \cdots & -\gamma X_1'X_M \\ -\gamma X_2'X_1 & (\alpha-\gamma)X_2'X_2 & \cdots & -\gamma X_2'X_M \\ \vdots & \vdots & & \vdots \\ -\gamma X_M'X_1 & -\gamma X_M'X_2 & \cdots & (\alpha-\gamma)X_M'X_M \end{bmatrix}^{-1}. \tag{3.6}$$

For the two dimensional case ($M=2$) the covariance matrix of the first equation's coefficient vector estimator is:

$$V(b_1^*) = \left[(\alpha-\gamma)X_1'X_1 - \frac{\gamma^2}{\alpha-\gamma}X_1'X_2(X_2'X_2)^{-1}X_2'X_1\right]^{-1} \tag{3.7}$$

and it can be shown that [cf. 14]

$$|V(b_1^*)| = \frac{(1-\rho^2)^{l_1}}{\prod_{\mu=1}^{l_1}(1-\rho^2 r_\mu^2)}\,|\sigma^2(X_1'X_1)^{-1}| \tag{3.8}$$

where l_1 is the number of independent variables in the first equation ($l_1\leq l_2$) and r_μ is the μ'th canonical correlation coefficient associated with the sets of variables in X_1 and X_2. Since $0\leq r_\mu^2\leq 1$, it is clear that the generalized variance of b_1^* will be smaller than or equal to $|\sigma^2(X_1'X_1)^{-1}|$, the generalized variance of the "single-equation" least squares estimator of the first equation's coefficient vector. If $r_\mu=0$ for all μ, as would be the case if $X_1'X_2=0$, the expression in (3.6) reduces to $(1-\rho^2)^{l_1}|\sigma^2(X_1'X_1)^{-1}|$ which represents the minimal generalized variance for given ρ and σ^2.

Further from (3.6), with $X_\mu'X_{\mu'}=0$ for $\mu\neq\mu'$, we obtain

$$V(b_1^*) = \left[\frac{1-\rho}{1-\dfrac{\rho}{1+\rho(M-1)}}\right]\sigma^2(X_1'X_1)^{-1} \tag{3.9}$$

and thus as the number of equations, M, approaches infinity with $X_\mu'X_{\mu'}=0$ for $\mu, \mu'=1, 2, \cdots, M$ and $\mu\neq\mu'$, then $V(b_1^*)$ approaches $(1-\rho)\sigma^2(X_1'X_1)^{-1}$.

4. TESTING FOR AGGREGATION BIAS

4.1. Testing with Micro-Data

It is, of course, possible to develop tests of a variety of hypotheses about the

[5] This case was suggested to the author by one of the *Journal's* editors (cf. [14] for additional results). Since we have illustrative purposes in mind, we neglect the fact that elements of the disturbance covariance matrix must be estimated.

coefficient vector in (2.2). One particularly important hypothesis in the case that $X_1, X_2, \cdots, X_M$ are all of the same size and represent matrices of observations on particular variables relating to different micro-units is the following one:

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_M. \tag{4.1}$$

The hypothesis in (4.1) states that micro-units are homogeneous insofar as their regression coefficient vectors are concerned. Further if (4.1) is valid, there will be no aggregation bias involved in simple linear aggregation [7, 11]. That is, with simple linear aggregation, we form[6]

$$\bar{y} = \sum_{\mu} y_\mu, \qquad \overline{X} = \sum_{\mu} X_\mu$$

and estimate $\bar{\beta}$ in:

$$\bar{y} = \overline{X}\bar{\beta} + \bar{u} \tag{4.2}$$

where

$$\bar{u} = \sum_{\mu} u_\mu.$$

The expectation of the least-squares estimator of $\bar{\beta}$ is given by:

$$E\bar{b} = \sum_{\mu} B_\mu \beta_\mu \tag{4.3}$$

where $B_\mu = (\overline{X}'\overline{X})^{-1}\overline{X}'X_\mu$. Clearly

$$\sum_{\mu} B_\mu = I \quad \text{since} \quad \sum_{\mu} X_\mu \equiv \overline{X}.$$

Thus if hypothesis (4.1) is true, the expectation of the macro-estimator $\bar{b}$ will be equal to the micro-parameter vector.

In testing (4.1), it is necessary to use a test statistic which takes account of the fact that the disturbances in the micro-regressions are correlated. Fortunately such a test has been described in the literature [10, p. 82]. The test statistic, employed for testing such restrictions on regression systems, is given by:

$$F_{q,n-m} = \frac{n-m}{q} \times \frac{y'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}C'[C(X'\Sigma^{-1}X)^{-1}C']^{-1}C(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y}{y'\Sigma^{-1}y - y'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y} \tag{4.4}$$

where n is the number of observations on y, m the number of independent variables, q the number of restrictions on the system, and C the matrix of the restrictions, $C\beta = 0$. In terms of the system in (2.2), $n = MT$, $m = Ml$ and $q = (M-1)l$. The restrictions given by the hypothesis (4.1) can be expressed as follows:

[6] This convenient matrix representation of this aspect of the aggregation problem is presented by Kloek [7].

$$C\beta = \begin{bmatrix} I & -I & 0 \cdots 0 & 0 & 0 \\ 0 & I & -I \cdots 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 \cdots I & -I & 0 \\ 0 & 0 & 0 \cdots 0 & I & -I \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_M \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}. \tag{4.5}$$

The unit and zero matrices in (4.5) are of order $l \times l$. Thus there are $q=(M-1)l$ restrictions, as stated above. Roy's [10, p. 82] very elegant derivation of this test does not involve the likelihood-ratio approach. However, as shown in a straight-forward manner in Appendix A, the likelihood-ratio approach leads to the same test statistic. If the disturbance covariance matrix were known (4.4) would give an exact test of the hypothesis in (4.1). When an estimate of this matrix is employed in constructing the test statistic, we show in Appendix B that the resulting statistic, say $\tilde{F}$, is equal to the statistic in (4.4) plus an error which is $0(n^{-1/2})$ in probability. Then by a theorem in [5, p. 254], $\tilde{F}$ will have the same asymptotic distribution as $F_{q,n-m}$. But, as shown in Appendix A, $-2 \log \lambda = qF_{q,n-m} + 0(n^{-1})$ where λ is the likelihood ratio for testing the hypothesis in (4.1). It is known [8, p. 259 and 12, p. 151] that $-2 \log \lambda$, and thus $qF_{q,n-m}$ (and $q\tilde{F}$), is asymptotically distributed as $\chi_q^2 = \chi_{(M-1)l}^2$, where $l(M-1)$ is the number of restrictions involved in (4.1).

For small samples there is some question about how to proceed. We can compute $q\tilde{F}$ and use $q\tilde{F}$'s asymptotic distribution, χ_q^2, assuming that the asymptotic results apply. Another alternative, which may be better, would be to assume that $\tilde{F}$'s distribution is closely approximated by that of $F_{q,n-m}$.[7]

4.2. Testing for Aggregation Bias with Macro-Data

When just macro-data are available, it is customary to estimate a macro-relation, for example that in (4.2), and then to proceed as if no aggregation bias were present. Obviously it would be desirable to have a test of the hypothesis of no aggregation bias, particularly one which employs just macro-data. A test of this sort is developed below. Initially, we restrict ourselves to consideration of a simple system involving two micro-regressions, each with one independent variable:

$$\begin{cases} y_1(t) = \beta_{11}x_{11}(t) + \beta_{10} + u_1(t) \\ y_2(t) = \beta_{21}x_{21}(t) + \beta_{20} + u_2(t). \end{cases} \tag{4.6}$$

The corresponding macro-relation, obtained by adding these two micro-equations, is:

$$\bar{y}(t) = \left[\frac{\beta_{11}x_{11}(t) + \beta_{21}x_{21}(t)}{x_{11}(t) + x_{21}(t)}\right]\bar{x}(t) + \beta_0 + \bar{u}(t) \tag{4.7}$$

where, as before, a bar over a variable denotes a sum of micro-variables. Now it is seen that the coefficient of $\bar{x}(t)$ in (4.9) is a weighted average of β_{11} and

[7] A similar problem arises in connection with the small-sample properties of identifiability test statistics in reference [3].

β_{21} with weights $w_1(t)=x_{11}(t)/[x_{11}(t)+x_{21}(t)]$ and $1-w_1(t)$. Introducing these weights explicitly, (4.7) becomes:

$$\bar{y}(t) = \beta_{21}\bar{x}(t) + (\beta_{11} - \beta_{21})w_1(t)\bar{x}(t) + \beta_0 + \bar{u}(t). \tag{4.8}$$

If data are available giving $w_1(t)$, $t=1, 2, \cdots, T$, it is possible to form the variable $w_1(t)\bar{x}(t)$, run the regression in (4.8), and test the hypothesis that the coefficient of $w_1(t)\bar{x}(t)$ is equal to zero. This is a test of the hypothesis $\beta_{11}=\beta_{21}$ and if accepted as true means that no aggregation bias is present. In a practical case, $w_1(t)$ might be a firm's proportion of industry sales in year t. Data on market shares might be available while micro-data on certain "dependent" variables might not be available.

If data on $w_1(t)$ are not available, it may be that an investigator is willing to stipulate on *a priori* theoretical grounds or on some other basis that $w_1(t)$ is a function of a variable for which data are available. To be specific suppose $w_1(t)=\alpha_0+\alpha_1 Z(t)$.[8] On substituting in (4.8) we obtain:

$$\bar{y}(t) = [\beta_{21} + \alpha_0(\beta_{11} - \beta_{21})]\bar{x}(t) + (\beta_{11} - \beta_{21})\alpha_1 Z(t)\bar{x}(t) + \beta_0 + \bar{u}(t). \tag{4.9}$$

Now, regressing[9] $\bar{y}(t)$ on $\bar{x}(t)$ and $Z(t)\bar{x}(t)$ and testing the hypothesis that the coefficient of the second variable is zero constitutes a test of the hypothesis of equality of micro-parameters and thus of no aggregation bias.

This testing procedure can easily be extended to cover more complicated specifications. For example, if there are M micro-regressions in (4.6) rather than two, (4.7) becomes:

$$\bar{y}(t) = \beta_{M1}\bar{x}(t) + \sum_{i=1}^{M-1} (\beta_{i1} - \beta_{M1})w_i(t)\bar{x}(t) + \beta_0 + \bar{u}(t) \tag{4.10}$$

where $w_i(t)=x_{i1}(t)/\bar{x}(t)$. If we now have $w_i(t)=\alpha_{0i}+\alpha_i Z(t)$, this last expression becomes

$$\begin{aligned}\bar{y}(t) = &\left[\beta_{M1} + \sum_{i=1}^{M-1} (\beta_{i1} - \beta_{M1})\alpha_{0i}\right]\bar{x}(t) \\ &+ \left[\sum_{i=1}^{M-1} (\beta_{i1} - \beta_{M1})\alpha_i\right] Z(t)\bar{x}(t) + \beta_0 + \bar{u}(t).\end{aligned} \tag{4.11}$$

Again a simple regression of $\bar{y}(t)$ on $\bar{x}(t)$ and $Z(t)\bar{x}(t)$ is all that is needed to test the hypothesis of micro-parameter equality. Further, the procedure can be extended to apply to systems with more than one independent variable in each regression. In all cases, however, the application of the test is conditional upon there being meaningful relations between the weights, the w_i, and some variable or variables for which data are available.

5. APPLICATION OF METHODS TO INVESTMENT DEMAND

To illustrate the methods described above, we utilize the investment equation developed by Grunfeld [6] and described in Boot and de Witt [4]. Grunfeld's

---

[8] If this relation is stochastic, say, $w_1(t)=\alpha_0+\alpha_1 Z(t)+v(t)$, where $v(t)$ is a stochastic disturbance term, the approach shown below leads to a regression model in which one (or some) of the independent variables have "measurement error." Problems of estimation and testing associated with such models are not considered in this paper.

[9] It is assumed that $Z(t)$ is an exogenous variable. Or alternatively, $Z(t)$ can be a polynomial in exogenous variables.

investment function involves a firm's current gross investment, $I(t)$, being dependent on the firm's beginning-of-year capital stock, $C(t-1)$ and the value of its outstanding shares at the beginning of the year, $F(t-1)$. That is, the micro-investment function is:

$$I(t) = \alpha_0 + \alpha_1 C(t-1) + \alpha_2 F(t-1) + u(t), \qquad t = 1, 2, \cdots, T. \quad (5.1)$$

Herein we present estimates of (5.1) for two firms, General Electric and Westinghouse, by the method described above and by single-equation least-squares. The annual data, 1935–1954, are taken from reference [4].

For convenience we relabel the variables as follows:

Firm	$I(t)$	$C(t-1)$	$F(t-1)$	1
General Electric	$y_1(t)$	$x_{11}(t)$	$x_{12}(t)$	$x_{13}(t)$
Westinghouse	$y_2(t)$	$x_{21}(t)$	$x_{22}(t)$	$x_{23}(t)$

The equation system to be estimated is then:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix},$$

where $X_1(t) = [x_{11}(t) x_{12}(t) x_{13}(t)]$, $x_2(t) = [x_{21}(t) x_{22}(t) x_{23}(t)]$, $\beta_1' = [\beta_{11}\beta_{12}\beta_{10}]$ and $\beta_2' = [\beta_{21}\beta_{22}\beta_{20}]$. Let $Z_1 = [y_1 X_1]$ and $Z_2 = [y_2 X_2]$ and $Z = [Z_1 Z_2]$. In our case we have for the submatrices[10] of $Z'Z$:

$$\begin{bmatrix} y_1'y_1 & \vdots & y_1'X_1 \\ \cdots & \vdots & \cdots \\ X_1'y_1 & \vdots & X_1'X_1 \end{bmatrix} = \begin{bmatrix} 254113.50 & \vdots & 1005863.46 & 4093308.29 & 2045.8 \\ \cdots & \vdots & \cdots & \cdots & \cdots \\ & \vdots & 4395946.84 & 15769824.07 & 8003.2 \\ & \vdots & & 78628914.21 & 38826.5 \\ & \vdots & & & 20 \end{bmatrix}$$

$$\begin{bmatrix} y_1'y_2 & \vdots & y_1'X_2 \\ \cdots & \vdots & \cdots \\ X_1'y_2 & \vdots & X_1'X_2 \end{bmatrix} = \begin{bmatrix} 103869.607 & \vdots & 221467.99 & 1531586.94 & 2045.8 \\ \cdots & \vdots & \cdots & \cdots & \cdots \\ 413156.104 & \vdots & 974281.31 & 6153588.29 & 8003.2 \\ 1719503.680 & \vdots & 3369944.27 & 27247303.72 & 38826.5 \\ 857.83 & \vdots & 1712.8 & 13418.2 & 20 \end{bmatrix}$$

$$\begin{bmatrix} y_2'y_2 & \vdots & y_2'X_2 \\ \cdots & \vdots & \cdots \\ X_2'y_2 & \vdots & X_2'X_2 \end{bmatrix} = \begin{bmatrix} 43732.4023 & \vdots & 90592.412 & 643262.570 & 857.83 \\ \cdots & \vdots & \cdots & \cdots & \cdots \\ & \vdots & 220345.72 & 1344261.18 & 1712.8 \\ & \vdots & & 9942109.78 & 13418.2 \\ & \vdots & & & 20 \end{bmatrix}$$

and the remaining one is just the transpose of one already shown above.

[10] Just the upper parts of symmetric matrices are shown. It should be noted that the elements of y_2 are given to two decimal places in the original data whereas all other data are given accurate to one decimal place.

We first compute the single-equation least-squares estimates in the usual way to obtain:

$$\hat{\beta}_1 = \begin{bmatrix} \hat{\beta}_{11} \\ \hat{\beta}_{12} \\ \hat{\beta}_{10} \end{bmatrix} = \begin{bmatrix} 0.151693870 \\ 0.026551189 \\ -9.956306513 \end{bmatrix} \text{ and } \hat{\beta}_2 = \begin{bmatrix} \hat{\beta}_{21} \\ \hat{\beta}_{22} \\ \hat{\beta}_{20} \end{bmatrix} = \begin{bmatrix} 0.092406491 \\ 0.052894127 \\ -0.509390038 \end{bmatrix}.$$

To get the estimated disturbance covariance matrix conveniently, we write

$$[y_1\ y_2] = [X_1\ X_2]\begin{bmatrix} \hat{\beta}_1 & 0 \\ 0 & \hat{\beta}_2 \end{bmatrix} + [\hat{u}_1\ \hat{u}_2]$$

or

$$Y = XB + \hat{U}.$$

Then,

$$\begin{aligned} \hat{U}'\hat{U} &= (Y - XB)'(Y - XB) = Y'Y - B'X'XB \\ &= \begin{bmatrix} y_1'y_1 & y_1'y_2 \\ y_2'y_1 & y_2'y_2 \end{bmatrix} - \begin{bmatrix} \hat{\beta}_1'X_1'X_1\hat{\beta}_1 & \hat{\beta}_1'X_1'X_2\hat{\beta}_2 \\ \hat{\beta}_2'X_2'X_1\hat{\beta}_1 & \hat{\beta}_2'X_2'X_2\hat{\beta}_2 \end{bmatrix} \\ &= \begin{bmatrix} 13216.5899 & 3988.0118 \\ & 1821.2808 \end{bmatrix}, \end{aligned}$$

which is equal to $(T-3)\{s_{\mu\mu'}\}$ where $T=20$, the number of observations on each variable. We now invert this last matrix to obtain $(T-3)^{-1}\{s^{\mu\mu'}\}$:

$$(T-3)^{-1}\{s^{\mu\mu'}\} = \begin{bmatrix} .000223009584 & -.000488319216 \\ & .00161832608 \end{bmatrix}.$$

We can now obtain the estimate of the moment matrix of the two-stage Aitken estimators by forming and inverting the following matrix:

$$\begin{bmatrix} s^{11}X_1'X_1 & s^{12}X_1'X_2 \\ s^{21}X_2'X_1 & s^{22}X_2'X_2 \end{bmatrix}.$$

The inverse of this last matrix is shown below. Elements on the diagonal are estimated coefficient estimator variances while off-diagonal elements are estimated covariances.

$$\left[\begin{array}{ccc:ccc} .0006006 & -.0000360 & -.1704 & .0007562 & -.0000197 & -.0515761 \\ & .0001885 & -.3516 & -.0003914 & .0001447 & -.0635894 \\ & & 789.6028 & .4573140 & -.2731373 & 155.8156 \\ \hdashline & & & .0026914 & -.0005191 & .1177480 \\ & & & & .0001972 & -.0878539 \\ & & & & & 54.2149 \end{array}\right]$$

To obtain two-stage Aitken coefficient estimates, we multiply the last matrix into the following vector:

$$\begin{bmatrix} s^{11}X_1'y_1 + s^{12}X_1'y_2 \\ s^{21}X_2'y_1 + s^{22}X_2'y_2 \end{bmatrix} = (T-3)\begin{bmatrix} 22.565127 \\ 73.180290 \\ 0.037338 \\ \cdots\cdots \\ 38.460988 \\ 293.105260 \\ 0.389245 \end{bmatrix}.$$

The point estimates so obtained along with their estimated variances are shown in Table 1. Also shown are the single-equation least-squares estimates and their estimated variances.

TABLE 1. RESULTS OF TWO-STAGE AITKEN AND SINGLE-EQUATION LEAST-SQUARES ESTIMATION OF MICRO-INVESTMENT FUNCTIONS

Micro-unit	Coefficient of	Two-Stage Aitken Method		Single-Equation Least-Squares	
		Coefficient estimate	Variance of coefficient estimator	Coefficient estimate	Variance of coefficient estimator
General Electric	C-1	.1326	.0006006	.1517	.0006605
	F-1	.0421	.0001885	.0266	.0002423
	1	−32.4807	789.6	−9.9563	984.1
Westinghouse	C-1	.0459	.002691	.0924	.003147
	F-1	.0611	.0001972	.0529	.0002468
	1	−2.0113	54.21	−.5094	64.24

It is seen from the results in Table 1 that application of the estimation procedure described above has resulted in a significant reduction (about 20 per cent) in the estimated coefficient estimator variances as compared with those of single-equation least-squares. The estimated correlation between the disturbances in the two equations is 0.81. Thus from what has been said in Section 3.2, the maximum gain to be expected is approximately $1-(0.81)^2=034$ times the single-equation estimated variances. That the maximum gain was not realized is due to the fact that $X_1'X_2 \neq 0$.

We note also in Table 1 that the point estimates yielded by the two methods differ. This is to be expected since different quadratic forms are minimized in the two approaches and also, obviously, if one method is more efficient than another, the estimates yielded by the two methods cannot always, or even usually, be identical. What it is important to realize is that it makes good sense to use the Aitken quadratic form. In this form we have weighted deviations;

that is, the data in the sample are not all given the same weight but are weighted by elements of the covariance matrix's inverse. In a single-equation case with heteroscedasticity present, this means weighting the square of each deviation by the reciprocal of its variance, an extremely sensible procedure. With the use of direct least-squares, all squared deviations are given the same weight, a rather unsatisfactory weighting of the evidence in the sample. Similar considerations apply to use of Aitken's quadratic form in connection with equation systems.

We now turn to an application of the test for micro-parameter equality, described in Section 4.1. In the present application, the numerator of the test statistic is from (4.4):

$$M(T-l)\begin{bmatrix} \sum_{\mu=1}^{2} y_\mu' s^{\mu 1} X_1 & \sum_{\mu=1}^{2} y_\mu' s^{\mu 2} X_2 \end{bmatrix}\begin{bmatrix} X_1'X_1 s^{11} & X_1'X_2 s^{12} \\ X_2'X_1 s^{21} & X_2'X_2 s^{22}\end{bmatrix}^{-1}\begin{bmatrix} I \\ -I\end{bmatrix} \tag{5.2}$$

$$\times\left\{[I \;\; -I]\begin{bmatrix} X_1'X_1 s^{11} & X_1'X_2 s^{12} \\ X_2'X_1 s^{21} & X_2'X_2 s^{22}\end{bmatrix}^{-1}\begin{bmatrix} I \\ -I\end{bmatrix}\right\}^{-1}$$

$$\times [I \;\; -I]\begin{bmatrix} X_1'X_1 s^{11} & X_1'X_2 s^{12} \\ X_2'X_1 s^{21} & X_2'X_2 s^{22}\end{bmatrix}^{-1}\begin{bmatrix} \sum_{\mu=1}^{2} X_1' s^{1\mu} y_\mu \\ \sum_{\mu=1}^{2} X_2' s^{2\mu} y_\mu \end{bmatrix}$$

where $s^{\mu\mu'}$ has been substituted for $\sigma^{\mu\mu'}$; $M=2$, $l=3$, $T-l=17$ and the unit matrices are of size 3×3. Most of the expressions appearing have already been computed in the estimation of the system. However, the second inverse appearing in the expression must be computed. We have

$$A = [I \;\; -I]\begin{bmatrix} X_1'X_1 s^{11} & X_1'X_2 s^{12} \\ X_2'X_1 s^{21} & X_2'X_2 s^{22}\end{bmatrix}^{-1}\begin{bmatrix} I \\ -I\end{bmatrix} = [I \;\; -I]\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22}\end{bmatrix}\begin{bmatrix} I \\ -I\end{bmatrix}$$
$$= [(B_{11}-B_{21})-(B_{12}-B_{22})],$$

where the definition of B_{ij} is obvious and the inverse of A is:[11]

$$(T-3)^{-1}A^{-1} = \begin{bmatrix} 86.805843 & 264.059312 & .125732 \\ & 1573.005191 & .531040 \\ & & .000321 \end{bmatrix}.$$

Then we form

$$\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22}\end{bmatrix}\begin{bmatrix} I \\ -I\end{bmatrix} A^{-1}[I \;\; -I]\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22}\end{bmatrix}$$
$$= \begin{bmatrix} (B_{11}-B_{12})A^{-1}(B_{11}-B_{21}) & (B_{11}-B_{12})A^{-1}(B_{12}-B_{22}) \\ (B_{21}-B_{22})A^{-1}(B_{11}-B_{21}) & (B_{21}-B_{22})A^{-1}(B_{12}-B_{22})\end{bmatrix},$$

a symmetric matrix which we calculated as, $(T-3)^{-1}$ times

[11] The factor $(T-3)^{-1}$ comes in since we have used $(T-3)^{-1}\{s^{\mu\mu'}\}$ rather than $\{s^{\mu\mu'}\}$.

$$\begin{bmatrix} .004370 & .000520 & -3.278245 & .007015 & .000798 & -1.258503 \\ & .002809 & -5.853262 & -.005526 & .002064 & -.957263 \\ & & 13326.484 & 7.392666 & -4.519575 & 2552.105 \\ & & & .039913 & -.007692 & 1.620036 \\ & & & & .002956 & -1.369759 \\ & & & & & 824.893611 \end{bmatrix}$$

The last step in calculation of the numerator is to pre- and postmultiply this last matrix by the vectors shown in (5.2) which are given above. The result is 0.582278 which must be multiplied by $M(T-1)=2(17)=34$ to yield $19.797(T-3)$ which is the value of the numerator of the test statistic.

The denominator of the test statistic in (4.4) is, with $s^{\mu\mu'}$ replacing $\sigma^{\mu\mu'}$:

$$(M-1)l\left\{\left[\sum_{\mu=1}^{2} y_\mu' s^{\mu 1} y_1 + \sum_{\mu=1}^{2} y_\mu' s^{\mu 2} y_2\right] - \left[\sum_{\mu=1}^{2} y_\mu' s^{\mu 1} X_1 \quad \sum_{\mu=1}^{2} y_\mu' s^{\mu 2} X_2\right]\right. \\ \left. \times \begin{bmatrix} X_1' X_1 s^{11} & X_1' X_2 s^{12} \\ X_2' X_1 s^{21} & X_2' X_2 s^{22} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{\mu=1}^{2} X_1' s^{1\mu} y_\mu \\ \sum_{\mu=1}^{2} X_2' s^{2\mu} y_\mu \end{bmatrix}\right\}. \tag{5.3}$$

Everything in this last expression has been computed with the exception of:

$$(T-3)^{-1}\left[\sum_{\mu=1}^{2} y_\mu' s^{\mu 1} y_1 + \sum_{\mu=1}^{2} y_\mu' s^{\mu 2} y_2\right] = 26.000130.$$

Then by direct operations, (5.3) is calculated to be equal to

$$(M-1)l(1.911752)(T-3) = 3(1.911752)(T-3) = 5.735(T-3).$$

We now have:

$$\tilde{F} = \frac{19.797}{5.735} = 3.452.$$

As mentioned above, there are at least two alternative procedures, each with its own approximations, which are candidates for testing the hypothesis in (4.1): (a) we can utilize the fact that $q\tilde{F}$ is, as shown in Appendix B, asymptotically distributed as χ_q^2 with $q=(M-1)1=3$; or (b) we can assume that since $\tilde{F}$ and $F_{q,n-m}$ differ by an amount which is $0(T^{-1/2})$ i.p. the distribution of $\tilde{F}$ will be closely approximated by that of $F_{q,n-m}$. The relevant 95 per cent critical values for (a) and (b) are: $\chi_3^2(.95)=2.605$ and $F_{3,34}(.95)=2.88$. Thus, in this case, both procedures lead to rejection, at the 95 per cent level, of the hypothesis of regression-coefficient vector equality and, therefore, in simple aggregation bias will probably be present.[12]

[12] The macro-test for aggregation is not applied since in this instance no meaningful relations for the weights (see above) are available.

6. CONCLUDING REMARKS

We have presented a method of estimating coefficients in *generally encountered* sets of regression equations which is more efficient than an equation-by-equation application of least-squares. Application of this method to estimate micro-investment functions[13] has led to estimates of coefficient estimator variances about 20 per cent smaller than those of equation-by-equation least-squares. Such a substantial reduction in these variances is indeed a satisfying feature of the application shown above, a feature which will characterize those applications to systems in which the disturbances of different equations are highly correlated and the independent variables of different equations are not highly correlated. Further, while we have applied (and also discussed) the procedure for only the situation involving one regression per micro-unit, it is also possible to extend the method to situations in which these are several regressions per micro-unit.

Lastly, we have described two tests for aggregation bias, one a "micro-test," the other a "macro-test." The micro-test which takes account of the fact that an estimated disturbance covariance matrix is employed in the test statistic involves test of an important hypothesis, namely that, in our application, different micro-units are characterized by the same regression coefficients. Clearly this is important knowledge and the test which provides it should be applied. Finally, the macro-test for aggregation bias, described above, requires knowledge that certain auxiliary relations are true. This is, at present, a weakness of this test. How one proceeds when one is uncertain about the validity of these auxiliary relations is an open question. Then, too, when the auxiliary relations are stochastic, other problems arise, as noted above. These are issues which will receive attention in future work.

APPENDIX

A. LIKELIHOOD-RATIO TEST FOR MICRO-REGRESSION COEFFICIENT VECTOR EQUALITY

Under the hypothesis (4.1), $\beta_1=\beta_2=\cdots=\beta_M$, the system in (2.2) can be written as

$$\begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_M \end{bmatrix} \beta_1 + \begin{bmatrix} u_1 \\ \cdot \\ \cdot \\ \cdot \\ u_M \end{bmatrix},$$

or

$$y = Z\beta_1 + u.$$

Now transform variables by premultiplication by H where H is such that $E(Huu'H')=\sigma_\omega^2 I$. Let $Hy=\dot{y}$, $HZ=\dot{Z}$ and $Hu=\dot{u}$. Then we have for the likelihood function under the hypothesis, $L(\omega)$,

[13] It is possible to employ the coefficient estimates to obtain a new estimate of the disturbance covariance matrix and then a new set of coefficient estimates, and so on. The small-sample properties of this iterative procedure have not as yet been established.

$$L(\omega) = (2\pi)^{-\frac{1}{2}MT}(\sigma_\omega^2)^{\frac{1}{2}MT} \exp\left[-\tfrac{1}{2}\dot{u}'\dot{u}/\sigma_\omega^2\right]. \tag{A.1}$$

When maximum-likelihood estimators are substituted in (A.1), we obtain

$$L(\hat{\omega}) = (2\pi)^{-\frac{1}{2}MT}(\hat{\sigma}_\omega^2)^{-\frac{1}{2}MT} \exp\left[-\tfrac{1}{2}MT\right]$$

where $\hat{\sigma}_\omega^2 = (1/MT)\hat{\dot{u}}'\hat{\dot{u}} = (1/MT)(\dot{y} - \dot{Z}b_1)'(\dot{y} - \dot{Z}b_1)$ and $b_1 = (\dot{Z}'\dot{Z})^{-1}\dot{Z}'\dot{y}$.

Under the hypothesis Ω involving no restrictions on the coefficients, we have the system in (2.3). Again we transform the variables by premultiplication by H to obtain:

$$\dot{y} = X\beta + \dot{u},$$

where $\dot{y} = Hy$, $\dot{X} = HX$ and $\dot{u} = Hu$. The likelihood function now is:

$$L(\Omega) = (2\pi)^{-\frac{1}{2}MT}(\sigma_\Omega^2)^{-\frac{1}{2}MT} \exp\left[-\tfrac{1}{2}\dot{u}'\dot{u}/\sigma_\Omega^2\right]$$

which upon substitution of maximum-likelihood estimates becomes:

$$L(\hat{\Omega}) = (2\pi)^{-\frac{1}{2}MT}(\hat{\sigma}_\Omega^2)^{-\frac{1}{2}MT} \exp\left[-\tfrac{1}{2}MT\right]$$

where $\hat{\sigma}_n^2 = (1/MT)\dot{u}'\dot{u} = (1/MT)(\dot{y} - \dot{X}b)'(\dot{y} - \dot{X}b)$ and $b = (\dot{X}'\dot{X})^{-1}\dot{X}'\dot{y}$.

The estimated likelihood ratio, λ, is then

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \frac{(\hat{\sigma}_\omega^2)^{-\frac{1}{2}MT}}{(\hat{\sigma}_\Omega^2)^{-\frac{1}{2}MT}}$$

and

$$-2\log\lambda = MT\log\left(\frac{\hat{\sigma}_\omega^2}{\hat{\sigma}_\Omega^2}\right),$$

which is asymptotically distributed as $\chi^2_{(M-1)l}$ [cf. 8, p. 259 and 12, p. 151.].

We must now show that

$$\frac{\hat{\sigma}_\omega^2}{\hat{\sigma}_\Omega^2} = 1 + \frac{q}{n-m} F_{q,n-m} \tag{A.2}$$

where $F_{q,n-m}$ is given by (4.4) and $q = (M-1)l$ and $n - m = M(T-l)$. If (A.2) holds, we can write [cf. 8, p. 262]

$$\begin{aligned} n\log\frac{\hat{\sigma}_\omega^2}{\hat{\sigma}_\Omega^2} &= n\log\left[1 + \frac{q}{n-m}F_{q,n-m}\right] \\ &= \frac{nq}{n-m}F_{q,n-m} - n\left(\frac{q}{n-m}\right)^2 F_{q,n-m}^2 + \cdots \\ &= qF_{q,n-m} + 0(n^{-1}) \end{aligned}$$

and then by the convergence theorem in Cramer [5, p. 254], $n \log \hat{\sigma}_\omega^2/\hat{\sigma}_\Omega^2 = MT \log \hat{\sigma}_\omega^2/\hat{\sigma}_\Omega^2$ and $qF_{q,n-m} = (M-1)lF_{(M-1)l,M(T-l)}$ have the same asymptotic distribution, namely χ_q^2.

To show that (A.2) is valid, we write:

$$\frac{q}{n-m} F_{q,n-m} = \frac{\hat{\sigma}_\omega^2 - \hat{\sigma}_\Omega^2}{\hat{\sigma}_\Omega^2}$$
$$= \frac{(\dot{y} - Zb_1)'(\dot{y} - Zb_1) - (\dot{y} - Xb)'(\dot{y} - Xb)}{(\dot{y} - \dot{X}b)'(\dot{y} - \dot{X}b)} \qquad \text{(A.3)}$$
$$= \frac{b'X'Xb - b_1'Z'Zb_1}{\dot{y}'\dot{y} - b'\dot{X}'\dot{X}b} = \frac{\dot{y}'X(X'X)^{-1}X'\dot{y} - \dot{y}'\dot{Z}(\dot{Z}'Z)^{-1}Z'\dot{y}}{\dot{y}'\dot{y} - \dot{y}'\dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}'\dot{y}}\,.$$

Now the denominator of this last expression is $y'\Sigma^{-1}y - y'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1} \times X'\Sigma^{-1}y$ in terms of the original variables and this is the denominator of $[q/(n-m)]F_{q,n-m}$; see (4.4). The numerator of (A.3) is

$$y'\Sigma^{-1}[X(X'\Sigma^{-1}X)^{-1}X' - Z(Z'\Sigma^{-1}Z)^{-1}Z']\Sigma^{-1}y. \qquad \text{(A.4)}$$

Now

$$Z = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & X_M \end{bmatrix} \begin{bmatrix} I \\ I \\ \cdot \\ \cdot \\ \cdot \\ I \end{bmatrix} = XJ$$

where J is a column of unit matrices. Then (A.4) becomes:

$$y'\Sigma^{-1}X[(X'\Sigma^{-1}X)^{-1} - J(J'X'\Sigma^{-1}XJ)^{-1}J']X'\Sigma^{-1}y$$

or

$$y'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}[X'\Sigma^{-1}X - X'\Sigma^{-1}XJ(J'X'\Sigma^{-1}XJ)^{-1}J'X'\Sigma^{-1}X] \cdot (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y. \qquad \text{(A.5)}$$

For (A.5) to be equal to the numerator of $\{q/(n-m)\}F_{q,n-m}$, we must have

$$X'\Sigma^{-1}X - X'\Sigma^{-1}XJ(J'X'\Sigma^{-1}XJ)^{-1}J'X'\Sigma^{-1}X = C'[C(X'\Sigma^{-1}X)^{-1}C']^{-1}C.$$

That this last equality is true is established by premultiplying both sides by $C(X'\Sigma^{-1}X)^{-1}$ to obtain:

$$C[I - J(J'X'\Sigma^{-1}XJ)^{-1}J'X'\Sigma^{-1}X] = C.$$

Then post-multiplying both sides by J to obtain:

$$C[IJ - J] = CJ.$$

From the definition of C and J, it is seen that both sides of the last equation are just zero matrices. Thus the validity of (A.2) is established.

B. DERIVATION OF THE ASYMPTOTIC DISTRIBUTION OF THE TEST STATISTIC EMPLOYED FOR TESTING MICRO-REGRESSION COEFFICIENT VECTOR EQUALITY

In this part we establish that when a consistent estimate Σ_e of Σ is employed in (4.4), the resultant test statistic, say $\tilde{F}$, is equal to $F_{q,n-m}$ plus an error which is of order $n^{-1/2}$ i.p. and thus that they have the same asymptotic distribution. For the denominator of $\tilde{F}$ aside from a multiplicative factor, we have

$$y'\Sigma_e^{-1}y - y'\Sigma_e^{-1}X(X'\Sigma_e^{-1}X)^{-1}X'\Sigma_e^{-1}y \tag{B.1}$$

or, with $\Sigma_e = \Sigma + \Delta_1$,

$$y'(\Sigma + \Delta_1)^{-1}y - y'(\Sigma + \Delta_1)^{-1}X[X'(\Sigma + \Delta_1)^{-1}X]^{-1}X'(\Sigma + \Delta_1)^{-1}y,$$

where Δ_1 is $0(n^{-1/2})$ i.p. We now make the following expansions:

$$(\Sigma + \Delta_1)^{-1} = \Sigma^{-1} - \Sigma^{-1}\Delta_1\Sigma^{-1} + \cdots = \Sigma^{-1} + \Delta_2 \tag{B.2}$$

and

$$\begin{aligned}[X'(\Sigma + \Delta_1)^{-1}X]^{-1} &= [X'(\Sigma^{-1} + \Delta_2)X]^{-1} \\ &= (X'\Sigma^{-1}X)^{-1} - (X'\Sigma^{-1}X)^{-1}X'\Delta_2X(X'\Sigma^{-1}X)^{-1} + \cdots \\ &= (X'\Sigma^{-1}X)^{-1} + \Delta_3,\end{aligned} \tag{B.3}$$

where, i.p., Δ_2 is $0(n^{-1/2})$ and Δ_3 is $0(n^{-1\,1/2})$. Utilizing these results, (B.1) becomes:

$$y'(\Sigma^{-1} + \Delta_2)y - y'(\Sigma^{-1} + \Delta_2)X[(X'\Sigma^{-1}X)^{-1} + \Delta_3]X'(\Sigma^{-1} + \Delta_2)y$$

or

$$y'\Sigma^{-1}y - y'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y + \Delta_4$$

where Δ_4 is $0(n^{1/2})$ i.p. Thus the expression in (B.1) is equal to the quantity appearing in the denominator of (4.4) plus Δ_4, or $\chi^2_{m-n} + \Delta_4$.

For the numerator of $\tilde{F}$, again aside from a multiplicative factor, we have

$$\begin{aligned}&y'\Sigma_e^{-1}X(X'\Sigma_e^{-1}X)^{-1}C'[C(X'\Sigma_e^{-1}X)^{-1}C']^{-1}C'(X'\Sigma_e^{-1}X)^{-1}X'\Sigma_e^{-1}y \\ &\quad = y'(\Sigma^{-1} + \Delta_2)X(X'\Sigma^{-1}X + X'\Delta_2X)^{-1}C'\{C[(X'\Sigma^{-1}X)^{-1} + \Delta_3]C'\}^{-1} \\ &\quad\quad \times C'(X'\Sigma^{-1}X + X'\Delta_2X)^{-1}X'(\Sigma^{-1} + \Delta_2)y\end{aligned} \tag{B.4}$$

where (B.2) and (B.3) have been employed. Now the following expansions are required:

$$\begin{aligned}(X'\Sigma^{-1}X + X'\Delta_2X)^{-1} &= [I + (X'\Sigma^{-1}X)^{-1}X'\Delta_2X]^{-1}(X'\Sigma^{-1}X)^{-1} \\ &= [I - (X'\Sigma^{-1}X)^{-1}X'\Delta_2X + \cdots](X'\Sigma^{-1}X)^{-1} \\ &= (X'\Sigma^{-1}X)^{-1} + \Delta_5,\end{aligned} \tag{B.5}$$

where Δ_5 is $0(n^{-1\,1/2})$ i.p., and

$$\begin{aligned}\{C[(X'\Sigma^{-1}X)^{-1} + \Delta_3]C'\}^{-1} &= \{C(X'\Sigma^{-1}X)^{-1}C' + C\Delta_3C'\}^{-1} \\ &= \{I + [C(X'\Sigma^{-1}X)^{-1}C']^{-1}C\Delta_3C'\}^{-1}[C(X'\Sigma^{-1}X)^{-1}C']^{-1} \\ &= \{I - [C(X'\Sigma^{-1}X)^{-1}C']^{-1}C\Delta_3C' + \cdots\}[C(X'\Sigma^{-1}X)^{-1}C']^{-1} \\ &= [C(X'\Sigma^{-1}X)^{-1}C']^{-1} + \Delta_6,\end{aligned} \tag{B.6}$$

where Δ_6 is $0(n^{1/2})$ i.p. Substituting (B.5) and (B.6) in (B.4), we obtain:

$$y'(\Sigma^{-1} + \Delta_2)X[(X'\Sigma^{-1}X)^{-1} + \Delta_5]C'\{[C(X'\Sigma^{-1}X)^{-1}C']^{-1} + \Delta_6\}C' \\ \times [(X'\Sigma^{-1}X)^{-1} + \Delta_5]X'(\Sigma^{-1} + \Delta_2)y$$

which upon expansion becomes

$$y'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}C'[C(X'\Sigma^{-1}X)^{-1}C']^{-1}C'(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y + \Delta_7,$$

or just the quantity in the numerator of (4.4) plus Δ_7 which is $0(n^{1/2})$ i.p., or $\chi_q^2+\Delta_7$.

Then we obtain from the above results,

$$\begin{aligned}\tilde{F} &= \frac{n-m}{q}\cdot\frac{\chi_q^2 + \Delta_7}{\chi_{n-m}^2 + \Delta_4} = \frac{n-m}{q}\cdot\frac{\chi_q^2}{\chi_{n-m}^2}\left(1 + \frac{\Delta_7}{\chi_q^2}\right)\left(1 + \frac{\Delta_4}{\chi_{n-m}^2}\right)^{-1} \\ &= \frac{n-m}{q}\cdot\frac{\chi_q^2}{\chi_{n-m}^2} + 0(n^{-\frac{1}{2}})\end{aligned}$$

since $(1+\Delta_4/\chi_{n-m}^2)^{-1} = 1-\Delta_4/\chi_{n-m}^2 + \cdots$ and both Δ_7/χ_q^2 and Δ_4/χ_{n-m}^2 are $0(n^{-1/2})$ i.p.

$$\tilde{F} = F_{q,n-m} + 0(n^{-\frac{1}{2}}). \tag{B.7}$$

The result in (B.7) gives us some confidence in employing $\tilde{F}$ for our test statistic; however, it must be recognized that there is still some question about the degrees of freedom associated with $\tilde{F}$ in small samples since in our procedure Σ_e is not an independent estimate of Σ. The small-sample properties of $\tilde{F}$ deserve further investigation. Finally, it is to be noted that since the probability limit of the error in (B.7) is zero, $q\tilde{F}$ will have the same asymptotic distribution as $qF_{q,n-m}$, namely, a χ_q^2 as indicated in part A of the Appendix.

REFERENCES

[1] Aitken, A. C., "On Least-Squares and Linear Combination of Observations," *Proceedings of the Royal Society of Edinburgh*, 55 (1934–35), 42–8.

[2] Barten, A. P., and Koerts, J. "Transition Frequencies in Voting Behaviour." Forthcoming Report of the Econometric Institute.

[3] Basmann, R. L., "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics," *Journal of the American Statistical Association*, 55 (1960), 650–9.

[4] Boot, J. C. G., and de Witt, G. M. "Investment Demand: An Empirical Contribution to the Aggregation Problem," *International Economic Review*, 1 (1960), 3–30.

[5] Cramér, H., *Mathematical Methods of Statistics*. Princeton: Princeton University Press, 1946.

[6] Grunfeld, Y., *The Determinants of Corporate Investment.* Unpublished Ph.D. thesis (University of Chicago, 1958).

[7] Kloek, T., "Note on Convenient Matrix Notations in Multivariate Statistical Analysis and in the Theory of Linear Aggregation." Report 6017 of the Econometric Institute of the Netherlands School of Economics (1960).

[8] Mood, A. M., *Introduction to the Theory of Statistics.* New York: McGraw-Hill Book Company, 1950.

[9] Rao, C. R., *Advanced Statistical Methods in Biometric Research.* New York: John Wiley and Sons, 1952.

[10] Roy, S. N., *Some Aspects of Multivariate Analysis.* New York: John Wiley and Sons, 1957.

[11] Theil, H., *Linear Aggregation of Economic Relations.* Amsterdam: North-Holland Publishing Company, 1954.

[12] Wilks, S. S., *Mathematical Statistics.* Princeton: Princeton University Press, 1943.

[13] Zellner, A., and Theil, H. "Three-Stage Least-Squares: Simultaneous Estimation of Simultaneous Equations." *Econometrica,* 30 (1962), 54-78.

[14] Zellner, A. and Huang, D. S. "Further Properties of Efficient Estimators for Seemingly Unrelated Regression Equations." Systems Formulation and Methodology Workshop Paper 6101, Social Systems Research Institute, University of Wisconsin, 1961, to appear in the *International Economic Review.*