Journal of Econometrics 6 (1977) 147–164. © North-Holland Publishing Company

AN APPROXIMATION TO THE FINITE SAMPLE DISTRIBUTION OF ZELLNER'S SEEMINGLY UNRELATED REGRESSION ESTIMATOR

Peter C.B. PHILLIPS*

University of Birmingham, Birmingham, England

Received September 1975, final version received October 1976

A multiple equation model of the seemingly unrelated regressions type is considered. We derive an Edgeworth expansion up to $0(T^{-1})$, where T is the sample size, of the finite sampls distribution function of the seemingly unrelated regression estimator of the parameters in thie model. We examine the two-equation case where our results can be related to exact theory in the special case of orthogonal exogenous variables and we take as a particular numerical example Zellner's original application to micro-investment functions.

1. Introduction

Since the publication of Zellner's original paper (1962) on the estimation of seemingly unrelated regression equations, a number of papers have appeared that deal with various aspects of the finite sample distribution of the seemingly unrelated regression estimator (SURE) in this model. In an important paper Zellner (1963) has himself derived the finite sample distribution of the coefficient estimator in the special two-equation case where the exogenous variables in different equations are orthogonal and the disturbances are normally distributed; Zellner also compared the exact second-moment matrix of the estimator in this case with that of the single-equation least-squares (SELS) estimator [see also Zellner (1972)]. In a more general context, Kakwani (1967) has given conditions under which the SURE is unbiased.[1] Experimental evidence on the small sample

*I wish to acknowledge, with thanks, the comments of Yusaku Kataoka on earlier versions of this paper, which have saved me from a number of errors (including an important mistake in an earlier section on alternative covariance matrix estimates) and improved the presentation of the paper. I am grateful also to Professor Arnold Zellner for further suggestions. The research was supported by the Social Science Research Council under Grant Number HR 3432/1.

[1]Professor Arnold Zellner has kindly brought my attention to two more recent studies reporting exact results by Metha and Swamy (1975) and Kataoka (1974). Metha and Swamy derive the exact second moments of the SURE in a two-equation model without assuming pairwise orthogonal exogenous variables while Kataoka, who assumes pairwise orthogonal exogenous variable, derives the exact second moments of the SURE (using a covariance estimator from a restricted regression) in a two-equation model, as well as the exact distribution and second moments of the SURE in a system of several equations.

behaviour of the SURE is also available and Kmenta and Gilbert (1968) compare the sampling distribution of the SURE in a number of specific models with that of the maximum-likelihood estimator (under normality assumptions) and the SELS estimator. The results of Kmenta and Gilbert suggest that the asymptotic properties of the SURE [Zellner (1962) and Zellner–Huang (1962)] carry over well in small samples, although their experiments do not lend support to all of Zellner's exact results for models with orthogonal exogenous variables. In particular, when the exogenous variables are highly correlated Kmenta and Gilbert do not observe an efficiency gain in the SURE relative to the SELS estimator as the sample size increases.[2]

In the present paper we revisit the Zellner model and derive an asymptotic series expansion of the Edgeworth type as an approximation to the finite sample distribution of the SURE. This approximation helps to provide some further evidence of the finite sample behaviour of the SURE.

2. An approximation in the general case

We will work with the model

$$y_t = Ax_t + u_t, \qquad t = 1, \ldots, T, \tag{1}$$

where y_t is an $n \times 1$ vector of endogenous variables, x_t is an $m \times 1$ vector of non-random exogenous variables, and the $u_t (t = 1, \ldots, T)$ are mutually independent normally distributed random vectors with zero mean and non-singular covariance matrix $\Sigma = [(\sigma_{ij})]$. We write (1) as $Y' = AX' + U'$, where for example $Y' = [y_1, \ldots, y_T]$, and assume that X has full rank and $T > m+n$. A is a matrix of unknown coefficients which we assume can be parameterised in the form[3]

$$\text{vec}(A) = S\alpha - s, \tag{2}$$

where vec(A) is the vector formed by taking the direct sum of the rows of A, S is an $nm \times q$ matrix whose elements are known constants and whose rank is q, and s is a vector of known constants. In (2) α is taken as the vector of basic parameters and the model then includes the seemingly unrelated regression model as a special case as well as Malinvaud's general linear model (1970, pp. 289–296) which allows for the same parameters to occur in more than one equation.

The Aitken estimator of α which minimises the quadratic form

[2]Note also that the role of the exogenous variables in determining the efficiency gain of the Aitken estimator relative to SELS is considered in Zellner–Huang (1962).

[3]Cf. Sargan (1976, app. C, p. 1).

$$\{\operatorname{vec}(Y')-(I\otimes X)(S\alpha-s)\}'(\Sigma^{-1}\otimes I)\{\operatorname{vec}(Y')-(I\otimes X)(S\alpha-s)\}$$

is given by

$$\hat{\alpha} = \{S'(\Sigma^{-1}\otimes X'X)S\}^{-1}\{S'(\Sigma^{-1}\otimes X')\operatorname{vec}(Y')+S'(\Sigma^{-1}\otimes X'X)s\}. \tag{3}$$

We let $\hat{A}$ denote the corresponding estimator of A and then

$$\begin{aligned}\operatorname{vec}(\hat{A})-\operatorname{vec}(A) &= S(\hat{\alpha}-\alpha)\\ &= S\{S'(\Sigma^{-1}\otimes X'X)S\}^{-1}\{S'(\Sigma^{-1}\otimes X')\operatorname{vec}(Y')\\ &\quad +S'(\Sigma^{-1}\otimes X'X)(s-S\alpha)\}\\ &= S\{S'(\Sigma^{-1}\otimes X'X)S\}^{-1}\{S'(\Sigma^{-1}\otimes X')\operatorname{vec}(U')\}.\end{aligned}$$

The two-stage estimator of α is now obtained by replacing Σ is (3) with an estimate of Σ based on the residuals of a preliminary least-squares regression on (1). We will use the estimate[4]

$$\begin{aligned}\Sigma^* &= \frac{1}{T-m}\,Y'\{I-X(X'X)^{-1}X'\}\,Y\\ &= \frac{1}{T-m}\,U'\{I-X(X'X)^{-1}X'\}\,U,\end{aligned}$$

and then the corresponding estimates of A and α, which we denote by A^* and α^*, respectively, satisfy

$$\operatorname{vec}(A^*-A) = S\{S'(\Sigma^{*-1}\otimes X'X)S\}^{-1}\{S'(\Sigma^{*-1}\otimes X')\operatorname{vec}(U')\},$$

and

$$\alpha^*-\alpha = \{S'(\Sigma^{*-1}\otimes X'X)S\}^{-1}\{S'(\Sigma^{*-1}\otimes X')\operatorname{vec}(U')\}. \tag{4}$$

If $M = X'X/T$ converges as $T\to\infty$ to a finite non-singular matrix $\bar{M}$ then the limiting distribution of $T^{\frac{1}{2}}(\alpha^*-\alpha)$ is normal with zero mean and covariance matrix $\{S'(\Sigma^{-1}\otimes\bar{M})S\}^{-1}$. Setting $\Sigma^* = \Sigma+\Delta\Sigma$ we can write (4) as

[4]Cf. Zellner (1963). Renormalising Σ^* by $1/T$ rather than $1/(T-m)$ affects terms of $0_p(T^{-2})$ in the expansion of the estimates of α and A in powers of $1/T^{\frac{1}{2}}$. This does not then affect the first two terms of the Edgeworth expansion [that is, terms up to $0(T^{-1})$].

$$\alpha^* - \alpha = [S'\{(\Sigma + \Delta\Sigma)^{-1} \otimes M\}S]^{-1} \left[S'\{(\Sigma + \Delta\Sigma)^{-1} \otimes I\} \operatorname{vec} \frac{U'X}{T} \right],$$

$$= e_T(p, w) \tag{5}$$

where $p = \operatorname{vec}(U'X/T)$, and w is a vector of the distinct elements of

$$\Delta\Sigma = (T-m)^{-1} U'\{I - X(X'X)^{-1}X'\}U - \Sigma.$$

Thus, the error in the estimate α^* can be written as a function of p [whose distribution is normal with zero mean and covariance matrix $(\Sigma \otimes M)/T$)] and w (whose elements are statistically independent of p). Moreover, the elements of the error function e_T satisfy the derivative conditions in Sargan (1975), and $T^{\frac{1}{2}}w$ has bounded moments of all orders as $T \to \infty$ so that by Sargan's (1975) approximation theorem the distribution of $T^{\frac{1}{2}}(\alpha^* - \alpha)$ admits a valid Edgeworth expansion.[5] In what follows we will derive this expansion up to $0(T^{-1})$ and our method of approach, which is similar to that in Phillips (1975) and Sargan (1976), involves the expansion of the characteristic function of a linear combination of the components of $T^{\frac{1}{2}}(\alpha^* - \alpha)$ $(A^* - A)$ in powers of $1/T^{\frac{1}{2}}$.

Our first step is to obtain a more convenient representation of (5) by expanding $(\Sigma + \Delta\Sigma)^{-1}$. We have

$$\Sigma^{*-1} = (\Sigma + \Delta\Sigma)^{-1}$$

$$= \Sigma^{-1} - \Sigma^{-1}(\Delta\Sigma)\Sigma^{-1} + \Sigma^{-1}(\Delta\Sigma)\,\Sigma^{-1}(\Delta\Sigma)\Sigma^{-1} + 0_p(T^{-\frac{3}{2}})$$

$$= \Sigma^{-1} + (\Delta\Sigma^{-1}) + 0_p(T^{-\frac{3}{2}}), \quad \text{say.}$$

Then, setting $F^* = S'(\Sigma^{*-1} \otimes M)S$ and $F = S'(\Sigma^{-1} \otimes M)S$ we have

[5]The analysis in Sargan (1975) pertains to a marginal distribution so that we work later in the paper with the linear combination $h'(\alpha^* - \alpha)$ of the components of $\alpha^* - \alpha$. Denoting this linear combination by η_T and taking first derivatives at the origin we have

$$\eta_p = \partial\eta_T(0)/\partial_p = (\Sigma^{-1} \otimes I)\, S[S'(\Sigma^{-1} \otimes M)S]^{-1}h,$$

so that

$$\eta_p'\,\eta_p = h'[S'(\Sigma^{-1} \otimes M)S]^{-1}\,[S'(\Sigma^{-2} \otimes I)S]\,[S'(\Sigma^{-1} \otimes M)S]^{-1}h,$$

which is bounded above zero as $T \to \infty$; and $\eta_w = \partial\eta_T(0)/\partial w = 0$, being linear in the elements of p. With Σ non-singular, it is clear that $\eta_T(\,.\,)$ has continuous derivatives up to the fourth order (at least) in a fixed neighbourhood of the origin. Moreover, for large enough T, these derivatives are bounded uniformly in T (as $T \to \infty$) since S and Σ are independent of T and M has a finite limit as $T \to \infty$ by assumption. Finally, we note that $U'\{I - X(X'X)^{-1}\,X'\}\,U$ is Wishart $(\Sigma, T-m)$ so that all cumulants of w exist. But the components of w are standardised statistics and the cumulants of Tw are of $0(T)$ as $T \to \infty$. From this it is clear that the cumulants of $T^{\frac{1}{2}}w$ are bounded as $T \to \infty$. This verifies Assumptions 1–4 in Sargan (1975, p. 327).

$$F^* = S'(\Sigma^{-1}\otimes M)S+S'((\Delta\Sigma^{-1})\otimes M)S+0_p(T^{-\frac{3}{2}})$$

$$= \{I+S'((\Delta\Sigma^{-1})\otimes M)SF^{-1}\}F+0_p(T^{-\frac{3}{2}}),$$

so that

$$F^{*-1} = F^{-1}\{I+S'((\Delta\Sigma^{-1})\otimes M)SF^{-1}\}^{-1}+0_p(T^{-\frac{3}{2}})$$

$$= F^{-1}-F^{-1}S'((\Delta\Sigma^{-1})\otimes M)SF^{-1}$$

$$+F^{-1}S'((\Delta\Sigma^{-1})\otimes M)SF^{-1}S'((\Delta\Sigma^{-1})\otimes M)SF^{-1}+0_p(T^{-\frac{3}{2}}).$$

Hence

$$\alpha^*-\alpha = \{F^{-1}-F^{-1}S'((\Delta\Sigma^{-1})\otimes M)SF^{-1}$$

$$+F^{-1}S'((\Delta\Sigma^{-1})\otimes M)SF^{-1}S'((\Delta\Sigma^{-1})\otimes M)SF^{-1}\}$$

$$\times S'\left\{(\Sigma^{-1}\otimes I)\operatorname{vec}\left(\frac{U'X}{T}\right)+((\Delta\Sigma^{-1})\otimes I)\operatorname{vec}\left(\frac{U'X}{T}\right)\right\}$$

$$+0_p(T^{-2}),$$

and introducing the notation

$$G = SF^{-1}S' \quad\text{and}\quad \bar{p} = \operatorname{vec}\left(\frac{U'X}{T^{\frac{1}{2}}}\right),$$

we obtain

$$T^{\frac{1}{2}}(\alpha^*-\alpha) = F^{-1}S'(\Sigma^{-1}\otimes I)\bar{p}+F^{-1}S'(\Delta\Sigma^{-1})\otimes I)\bar{p}$$

$$-F^{-1}S'((\Delta\Sigma^{-1})\otimes M)G(\Sigma^{-1}\otimes I)\bar{p}$$

$$-F^{-1}S'((\Delta\Sigma^{-1})\otimes M)G((\Delta\Sigma^{-1})\otimes I)\bar{p}$$

$$+F^{-1}S'((\Delta\Sigma^{-1})\otimes M)G((\Delta\Sigma^{-1})\otimes M)G(\Sigma^{-1}\otimes I)\bar{p}$$

$$+0_p(T^{-\frac{3}{2}})$$

$$= B\bar{p}+0_p(T^{-\frac{3}{2}}), \quad \text{say.} \tag{6}$$

We now let $a_h = T^{\frac{1}{2}}h'(\alpha^*-\alpha)$, where h is a constant q-vector and we denote by $f_1(\bar{p})$ and $f_2(w)$ the probability density functions of $\bar{p}$ and w, respectively. Then the characteristic function of a_h is given by

$$\psi(s) = \int \exp\,[ia_h s] f_1(\bar{p}) f_2(w) \mathrm{d}\bar{p} \mathrm{d}w,$$

where the integration is over the entire $(\bar{p}, w)$ space. But, from (6), $a_h = h'B\bar{p} + 0_p(T^{-\frac{1}{2}})$, and

$$\begin{aligned} B = {} & F^{-1}S'(\Sigma^{-1}\otimes I)+F^{-1}S'((\Delta\Sigma^{-1})\otimes I) \\ & -F^{-1}S'((\Delta\Sigma^{-1})\otimes M)G(\Sigma^{-1}\otimes I) \\ & -F^{-1}S'((\Delta\Sigma^{-1})\otimes M)G((\Delta\Sigma^{-1})\otimes I) \\ & +F^{-1}S'((\Delta\Sigma^{-1})\otimes M)G((\Delta\Sigma^{-1})\otimes M)G(\Sigma^{-1}\otimes I) \end{aligned} \tag{7}$$

is a function only of w so that

$$\psi(s) = E\left\{\int \exp\,[\mathrm{i}(h'B\bar{p})s] f_1(\bar{p})\mathrm{d}\bar{p}\right\} + 0(T^{-\frac{1}{2}}), \tag{8}$$

where the integral within the expectation is over the $\bar{p}$-space. Now, from the normality of $\bar{p}$, we have

$$\int \exp\,[\mathrm{i}(h'B\bar{p})s] f_1(\bar{p})\mathrm{d}\bar{p} = \exp\,[-(s^2/2)h'B(\Sigma\otimes M)B'h]. \tag{9}$$

Using (7), the exponent on the right-hand side of (9) can be expanded as

$$\begin{aligned} h'B(\Sigma\otimes M)B'h = {} & h'F^{-1}S'(\Sigma^{-1}\otimes M)SF^{-1}h \\ & +2h'F^{-1}S'((\Delta\Sigma^{-1})\otimes M)SF^{-1}h \\ & -2h'F^{-1}S'((\Delta\Sigma^{-1})\otimes M)G(\Sigma^{-1}\otimes M)SF^{-1}h \\ & -2h'F^{-1}S'((\Delta\Sigma^{-1})\otimes M)G((\Delta\Sigma^{-1})\otimes M)SF^{-1}h \\ & +2h'F^{-1}S'((\Delta\Sigma^{-1})\otimes M)G((\Delta\Sigma^{-1})\otimes M) \\ & \times G(\Sigma^{-1}\otimes M)SF^{-1}h \\ & +h'F^{-1}S'(((\Delta\Sigma^{-1})\Sigma(\Delta\Sigma^{-1}))\otimes M)SF^{-1}h \\ & -2h'F^{-1}S'((\Delta\Sigma^{-1})\otimes M)G((\Delta\Sigma^{-1})\otimes M)SF^{-1}h \\ & +h'F^{-1}S'((\Delta\Sigma^{-1})\otimes M)G(\Sigma^{-1}\otimes M) \\ & \times G((\Delta\Sigma^{-1})\otimes M)SF^{-1}h+0_p(T^{-\frac{3}{2}}). \end{aligned}$$

But

$$G(\Sigma^{-1}\otimes M)G = [S\{S'(\Sigma^{-1}\otimes M)S\}^{-1}S'](\Sigma^{-1}\otimes M)$$
$$\times [S\{S'(\Sigma^{-1}\otimes M)S\{^{-1}S']$$
$$= S\{S'(\Sigma^{-1}\otimes M)S\}^{-1}S'$$
$$= G, \tag{10}$$

and

$$F^{-1}S'(\Sigma^{-1}\otimes M)SF^{-1} = F^{-1},$$

so that after cancellation

$$h'B(\Sigma\otimes M)B'h = h'F^{-1}h+h'F^{-1}S'(((\Delta\Sigma^{-1})\,\Sigma(\Delta\Sigma^{-1}))\otimes M)SF^{-1}h$$
$$-h'F^{-1}S'((\Delta\Sigma^{-1})\otimes M)\,G((\Delta\Sigma^{-1})\otimes M)SF^{-1}h$$
$$+0_p(T^{-\frac{3}{2}}).$$

Hence, returning to (9) we obtain

$$\exp[-(s^2/2)h'B(\Sigma\otimes M)Bh] = \exp[-(s^2/2)h'F^{-1}h]$$
$$\times[1-(s^2/2)\{h'F^{-1}S'$$
$$\times(((\Delta\Sigma^{-1})\Sigma(\Delta\Sigma^{-1}))\otimes M)SF^{-1}h$$
$$-h'F^{-1}S'((\Delta\Sigma^{-1})\otimes M)$$
$$\times G((\Delta\Sigma^{-1})\otimes M)SF^{-1}h\}]+0_p(T^{-\frac{3}{2}}). \tag{11}$$

Since

$$\Delta\Sigma^{-1} = -\Sigma^{-1}(\Delta\Sigma)\Sigma^{-1}+\Sigma^{-1}(\Delta\Sigma)\Sigma^{-1}(\Delta\Sigma)\Sigma^{-1}$$
$$= -\Sigma_a+\Sigma_b, \quad \text{say,}$$

we can write the right-hand side of (11) in the form

$$\exp[-(s^2/2)h'F^{-1}h][1-(s^2/2)\{h'F^{-1}S'(\Sigma_b\otimes M)SF^{-1}h$$
$$-h'F^{-1}S'(\Sigma_a\otimes M)G(\Sigma_a\otimes M)SF^{-1}h\}]+0_p(T^{-\frac{3}{2}}).$$

Taking expectations we now obtain the following expansion of the characteristic function of a_h:

$$\psi(s) = \exp[-(s^2/2)h'F^{-1}h][1-(s^2/2)\{h'F^{-1}S'(E(\Sigma_b)\otimes M)SF^{-1}h$$
$$-h'F^{-1}S'E\{(\Sigma_a\otimes M)G(\Sigma_a\otimes M)\}SF^{-1}h\}]+0(T^{-\frac{3}{2}}). \qquad (12)$$

To evaluate the expectations in (12) we first write $\Sigma^{-1} = KK'$ for some non-singular matrix K, and then

$$\Sigma_b = K(K'\Delta\Sigma K)(K'\Delta\Sigma K)K'$$
$$= \left(\frac{1}{T-m}\right)^2 K(\Delta L)^2K', \qquad (13)$$

where

$$\Delta L = K'U'\{1-X(X'X)^{-1}X'\}UK-(T-m)I.$$

But, $U'\{I-X(X'X)^{-1}X'\}U$ is Wishart $(\Sigma, T-m)$ so that $K'U'\{I-X(X'X)^{-1}X'\}UK$ is Wishart $(I, T-m)$, and we have [Anderson (1958, p. 161)]:

$$E[(\Delta L^2)_{ij}] = E\left\{\sum_{k=1}^{n}(\Delta L)_{ik}(\Delta L)_{kj}\right\}$$
$$= \sum_{k=1}^{n}(T-m)\{\delta_{ik}\delta_{kj}+\delta_{ij}\delta_{kk}\},$$

where δ denotes the Kronecker delta. Thus

$$E(\Delta L^2) = (T-m)(1+n)I,$$

and from (13) we find

$$E(\Sigma_b) = \frac{1+n}{T-m}\Sigma^{-1}. \qquad (14)$$

We now turn to the expectation in the third term in the square brackets in (12). We will consider the expression

$$E\{S'(\Sigma_a\otimes M)G(\Sigma_a\otimes M)S\}$$
$$= E\{S'(\Sigma_a\otimes M)S\{S'(\Sigma^{-1}\otimes M)S\}^{-1}S'(\Sigma_a\otimes M)S\}. \qquad (15)$$

We partition the columns of S' into n blocks of m as

$$S' = [S_1 \,\vdots\, S_2 \,\vdots\, \cdots \,\vdots\, S_n],$$

and then

$$S'(\Sigma_a \otimes M)S = \sum_{j=1}^{n} \sum_{k=1}^{n} \sigma_{jk}^a S_j M S_k',$$

where σ_{jk}^a is the (j, k)th element of Σ_a. Hence, (15) becomes

$$E\left[\left(\sum_{j,l} \sigma_{jl}^a S_j M S_l'\right)\{S'(\Sigma^{-1}\otimes M)S\}^{-1}(\Sigma_{rp}\sigma_{rp}^a S_r M S_p')\right]$$

$$= \sum_{j,l} \sum_{r,p} E(\sigma_{jl}^a \sigma_{rp}^a)(S_j M S_l')\{S'(\Sigma^{-1}\otimes M)S\}^{-1}(S_r M S_p'). \tag{16}$$

But

$$\Sigma_a = \Sigma^{-1}(\Delta\Sigma)\Sigma^{-1} = K(K'\Delta\Sigma K)K' = \frac{1}{T-m} K(\Delta L)K'.$$

Then using k_i' to denote the ith row of K, we have

$$E(\sigma_{jl}^a \sigma_{rp}^a) = \left(\frac{1}{T-m}\right)^2 E\{k_j'(\Delta L)K_l k_r'(\Delta L)k_p\}$$

$$= \left(\frac{1}{T-m}\right)^2 \sum_{b,c=1}^{n} \sum_{d,e=1}^{n} E((\Delta L)_{bc}(\Delta L)_{de})k_{jb}k_{lc}k_{rd}k_{pe}$$

$$= \left(\frac{1}{T-m}\right)^2 \sum_{b,c} \sum_{d,e} (T-m)(\delta_{bd}\delta_{ce}+\delta_{be}\delta_{cd})k_{jb}k_{lc}k_{rd}k_{pe}$$

$$= \left(\frac{1}{T-m}\right)\{(k_j'k_r)(k_l'k_p)+(k_j'k_p)(k_l'k_r)\}$$

$$= \left(\frac{1}{T-m}\right)\{\sigma^{jr}\sigma^{lp}+\sigma^{jp}\sigma^{lr}\},$$

where, for instance, σ^{jr} denotes the (j, r)th element of Σ^{-1}. Thus, (16) is

$$\left(\frac{1}{T-m}\right) \sum_{j,l} \sum_{rp} \{\sigma^{jr}\sigma^{lp}+\sigma^{jp}\sigma^{lr}\}(S_j M S_l')\{S'(\Sigma^{-1}\otimes M)S\}^{-1}(S_r M S_p'), \tag{17}$$

which we denote by

$$\left(\frac{1}{T-m}\right) D.$$

From (12), (14), (15) and (17) we now have

$$\psi(s) = \exp[-(s^2/2)h'F^{-1}h]$$
$$\times\left[1-\frac{s^2}{2}\left\{h'F^{-1}S'\left(\left(\frac{1+n}{T-m}\right)\Sigma^{-1}\otimes M\right)SF^{-1}h\right.\right.$$
$$\left.\left.-\left(\frac{1}{T-m}\right)h'F^{-1}DF^{-1}h\right\}\right]+0(T^{-\frac{3}{2}})$$
$$= \exp[-(s^2/2)h'F^{-1}h]$$
$$\left[1-\frac{s^2}{2}\left\{\left(\frac{1+n}{T-m}\right)h'F^{-1}h-\left(\frac{1}{T-m}\right)h'\ \Phi h\right\}\right]+0(T^{-\frac{3}{2}}), \tag{18}$$

where

$$\Phi = F^{-1}DF^{-1}. \tag{19}$$

Inverting (18) and using the fact that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}(is)^r\exp[-(s^2w^2/2)]\exp[-isx]\mathrm{d}s = \left(\frac{1}{w}\right)^{r+1}H_r\left(\frac{x}{w}\right)i\left(\frac{x}{w}\right)$$
$$= (-1)^r\left(\frac{1}{w}\right)^{r+1}i^{(r)}\left(\frac{x}{w}\right),$$

where $H_r(.)$ is the rth Hermite polynomial and $i(.)$ is the standard normal density, we obtain

$$P(T^{\frac{1}{2}}h'(\alpha^*-\alpha)\leqq x) = I\left(\frac{x}{w}\right)+\tfrac{1}{2}\left\{\left(\frac{1+n}{T-m}\right)w^2-\left(\frac{1}{T-m}\right)h'\Phi h\right\}$$
$$\times\left(\frac{1}{w^2}\right)i'\left(\frac{x}{w}\right)+0(T^{-\frac{3}{2}}), \tag{20}$$

where $I(.)$ denotes the standard normal distribution function, and

$$w^2 = h'F^{-1}h = h'\{S'(\Sigma^{-1}\otimes M)S\}^{-1}h.$$

The limit of w^2 as $T\rightarrow\infty$ is then the asymptotic variance of $T^{\frac{1}{2}}h'(\alpha^*-\alpha)$

The right-hand side of (20) reduces to [up to $0(T^{-1})$]

$$I\left(\frac{x}{w}\right)-i\left(\frac{x}{w}\right)\frac{x}{w}\left(\frac{1}{2(T-m)}\right)\left\{(1+n)-\frac{h'\Phi h}{w^2}\right\}. \tag{21}$$

An alternative representation of the distribution function of $T^{\frac{1}{2}}h'$ $(\alpha^*-\alpha)$ which is correct up to the same order of smallness in $1/T^{\frac{1}{2}}$ is [cf. Sargan (1975)]

$$I\left(\frac{x}{w}+g\frac{x}{w}\right), \tag{22}$$

where

$$g=-\left(\frac{1}{2(T-m)}\right)\left\{(1+n)-\frac{h'\Phi h}{w^2}\right\}.$$

Setting $x=0$ in (21) we obtain $0.5+0(T^{-\frac{3}{2}})$ so that the distribution of $T^{\frac{1}{2}}h'(\alpha^*-\alpha)$ is median-unbiased up to $0(T^{-1})$. Moreover, the approximation to the probability density of $T^{\frac{1}{2}}h'(\alpha^*-\alpha)$ corresponding to (21) is

$$i\left(\frac{x}{w}\right)\frac{1}{w}\left\{(1+g)-g\left(\frac{x}{w}\right)^2\right\},$$

up to the same order in $1/T^{\frac{1}{2}}$. Hence, to $0(T^{-1})$ the distribution is symmetric. These results square with those in Kakwani (1967). We note also that there is no term of $0(T^{-\frac{1}{2}})$ in (21) which suggests that even for moderate sample sizes the distribution of the SURE may be quite close to that of the Aitken estimator.

3. Some comparisons with SELS

The approximation given by the first two terms of (21) or by (22) can be used to compare the finite sample properties of the SURE with those of the SELS estimator. We denote the latter estimator by $\tilde{\alpha}$, and then

$$P(T^{\frac{1}{2}}h'(\tilde{\alpha}-\alpha)\leqq x)=I\left(\frac{x}{w_1}\right),$$

where

$$w_1^2=h'\{S'(I\otimes M)S\}^{-1}\{S'(\Sigma\otimes M)S\}\{S'(I\otimes M)S\}^{-1}h.$$

Then the difference between the estimators in terms of concentration in an interval symmetric about the true value is

$$P(|T^{\frac{1}{2}}h'(\alpha^*-\alpha)| \leqq x)-P(|T^{\frac{1}{2}}h'(\tilde{\alpha}-\alpha)| \leqq x)$$

$$= \left\{I\left(\frac{x}{w}(1+g)\right) - I\left(-\frac{x}{w}(1+g)\right)\right\}$$

$$- \left\{I\left(\frac{x}{w_1}\right) - I\left(-\frac{x}{w_1}\right)\right\} + 0(T^{-\frac{3}{2}})$$

$$= 2\left\{I\left(\frac{x}{w}(1+g)\right) - I\left(\frac{x}{w_1}\right)\right\} + 0(T^{-\frac{3}{2}})$$

$$= 2i(X)\left\{\frac{x}{w}(1+g) - \frac{x}{w_1}\right\} + 0(T^{-\frac{3}{2}}),$$

where X lies between $x(1+g)/w$ and x/w_1. Thus, to the stated order of approximation the SURE is the more concentrated about the true value if

$$\frac{1}{w}(1+g) > \frac{1}{w_1}\,.$$

That is if

$$T-m > \frac{\{(1+n)-h'\Phi h/w^2\}w_1}{2(w_1-w)}\,. \tag{23}$$

In view of the complexity of the matrix Φ it is difficult to draw useful general inferences from the above. One case where Φ has a very simple form and where it is possible to compare the implications of the above with exact results is Zellner's two-equation case with orthogonal exogenous variables. For, in this case, if there are m_1 exogenous variables in the first equation and m_2 in the second, we have

$$S = \begin{bmatrix} I_{m_1} & 0 \\ 0 & 0 \\ 0 & I_{m_2} \end{bmatrix},$$

so that

$$\{S'(\Sigma^{-1}\otimes M)S\}^{-1} = \begin{bmatrix} \frac{1}{\sigma^{11}}\left(\frac{X_1'X_1}{T}\right)^{-1} & 0 \\ 0 & \frac{1}{\sigma^{22}}\left(\frac{X_2'X_2}{T}\right)^{-1} \end{bmatrix},$$

and

$$\Phi = \begin{bmatrix} \frac{2}{\sigma^{11}}\left(\frac{X_1'X_1}{T}\right)^{-1} & 0 \\ 0 & \frac{2}{\sigma^{22}}\left(\frac{X_2'X_2}{T}\right)^{-1} \end{bmatrix},$$

where X_i is the matrix of T observations on the m_i exogenous variables in equation i ($i = 1, 2$). We consider the vector of coefficients in the first equation so that setting

$$h' = (h_1', 0),$$

where h_1 has m_1 components, we have

$$w^2 = \frac{1}{\sigma^{11}} h_1' \left(\frac{X_1'X_1}{T}\right)^{-1} h_1,$$

and

$$h'\Phi h = \frac{2}{\sigma^{11}} h_1' \left(\frac{X_1'X_1}{T}\right)^{-1} h_1.$$

Thus, (22) becomes for this case

$$I\left(\frac{x}{w}\left(1 - \frac{1}{2(T-m)}\right)\right),$$

and the condition (23) for the superior concentration of the SURE is [up to an error of $0(T^{-\frac{3}{2}})$]

$$T-m > \frac{w_1}{2(w_1 - w)}. \tag{24}$$

But

$$w_1^2 = \sigma_{11} h_1' \left(\frac{X_1'X_1}{T}\right)^{-1} h_1,$$

so that (24) is

$$T-m > \frac{(\sigma_{11})^{\frac{1}{2}}}{2\{(\sigma_{11})^{\frac{1}{2}} - (1/\sigma^{11})^{\frac{1}{2}}\}}.$$

Setting $\rho = \sigma_{12}/(\sigma_{11}\sigma_{22})^{\frac{1}{2}}$ we get

$$T-m > \frac{1}{2\{1-(1-\rho^2)^{\frac{1}{2}}\}} = \frac{1+(1-\rho^2)^{\frac{1}{2}}}{2\rho^2}. \tag{25}$$

On the other hand, from exact results [Zellner (1963, 1972)] we know that the SURE has smaller variance than the SELS estimator when

$$T-m > (1+\rho^2)/\rho^2. \tag{26}$$

As is clear from table 1 below, (25) implies very similar values of $T-m$ for an efficiency gain from the SURE relative to SELS. We note also in the table that (25) performs a little better (although there is not much between them) than

$$T-m > (1-\rho^2)/\rho^2, \tag{27}$$

which is the condition derived from the Nagar approximation,[6]

$$w^2(1-2g) = w^2\left(1+\frac{1}{T-m}\right),$$

to the variance of the SURE in this case.

Table 1
Values of $T-m$ for an efficiency gain from the SURE.

ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7
(25)	99.75	24.75	10.85	5.99	3.73	2.50	1.75
(26)	101.00	26.00	12.11	7.25	5.00	3.78	3.04
(27)	99	24.00	10.11	5.25	3.00	1.78	1.04

The two-equation case where $X_1'X_2 \neq 0$ is rather more complicated. In this case we have

$$S'(\Sigma^{-1}\otimes M)S = \frac{1}{T}\begin{bmatrix} \sigma^{11}\, X_1'X_1 & \sigma^{12}\, X_1'X_2 \\ \sigma^{21}\, X_2'X_1 & \sigma^{22}\, X_2'X_2 \end{bmatrix},$$

and writing

$$\{S'(\Sigma^{-1}\otimes M)S\}^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix},$$

we find that the matrix D of (17) can be partitioned as

$$D \quad \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix},$$

[6]This approximation can be developed readily from (21). Cf. Sargan–Mikhail (1971).

where

$$D_{11} = 2(\sigma^{11})^2\left(\frac{X_1'X_1}{T}\right)C_{11}\left(\frac{X_1'X_1}{T}\right)+2\sigma^{11}\sigma^{12}\left(\frac{X_1'X_2}{T}\right)C_{21}\left(\frac{X_1'X_1}{T}\right)$$

$$+2\sigma^{11}\sigma^{12}\left(\frac{X_1'X_1}{T}\right)C_{12}\left(\frac{X_2'X_1}{T}\right)$$

$$+\{(\sigma^{12})^2+\sigma^{11}\sigma^{22}\}\left(\frac{X_1'X_2}{T}\right)C_{22}\left(\frac{X_2'X_1}{T}\right),$$

$$D_{12} = 2\sigma^{11}\sigma^{12}\left(\frac{X_1'X_1}{T}\right)C_{11}\left(\frac{X_1'X_2}{T}\right)$$

$$+\{(\sigma^{12})^2+\sigma^{11}\sigma^{22}\}\left(\frac{X_1'X_2}{T}\right)C_{21}\left(\frac{X_1'X_2}{T}\right)$$

$$+2(\sigma^{12})^2\left(\frac{X_1'X_1}{T}\right)C_{12}\left(\frac{X_2'X_2}{T}\right)$$

$$+2\sigma^{12}\sigma^{22}\left(\frac{X_1'X_2}{T}\right)C_{22}\left(\frac{X_2'X_2}{T}\right),$$

$$D_{21} = D_{12},$$

and

$$D_{22} = \{(\sigma^{12})^2+\sigma^{11}\sigma^{22}\}\left(\frac{X_2'X_1}{T}\right)C_{11}\left(\frac{X_1'X_2}{T}\right)$$

$$+2\sigma^{12}\sigma^{22}\left(\frac{X_2'X_2}{T}\right)C_{21}\left(\frac{X_1'X_2}{T}\right)$$

$$+2\sigma^{12}\sigma^{22}\left(\frac{X_2'X_1}{T}\right)C_{12}\left(\frac{X_2'X_2}{T}\right)$$

$$+2(\sigma^{22})^2\left(\frac{X_2'X_2}{T}\right)C_{22}\left(\frac{X_2'X_2}{T}\right).$$

Then

$$h'\Phi h = h_1'\Phi_{11}h_1,$$

where

$$\Phi_{11} = C_{11}D_{11}C_{11}+C_{12}D_{21}C_{11}+C_{11}D_{12}C_{21}+C_{12}D_{22}C_{21}.$$

A particular example in which the above formulae can be used is Zellner's (1962) original application to the investment function

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

for two firms (General Electric and Westinghouse), where $I(t)$ represent gross investment in year t, $C(t-1)$ the beginning-of-year capital stock and $F(t-1)$ the value of outstanding shares at the beginning of the year. The matrices of sample second moments of the data are given in Zellner (1962), and we assume that the disturbances on the two equations are normally distributed with covariance matrix

$$\begin{bmatrix} 777.4465 & 234.5889 \\ 234.5889 & 107.1342 \end{bmatrix},$$

which is Zellner's estimate from the residuals of a preliminary SELS regression.

With this data and for the appropriate sample size $T = 20$, we have calculated the approximate distribution (21) of the SURE of the coefficients α_1 and α_2 in both equations. In table 2 below we compare this approximation with the distributions of the Aitken estimator ($I(x/w)$) and the SELS estimator ($I(x/w_1)$). Since each of these distributions is symmetric about the origin we consider only a grid of negative values.

In each case we note that the approximate distribution of the SURE is quite close to the distribution of the Aitken estimator and the SELS estimator has greater spread than the SURE. For an interval based on two standard deviations (of the Aitken estimator) on either side of the true parameter value we get in the case of the coefficient α_2 [and up to $0(T^{-1})$ for the estimate α_2^*]:

$$P(|T^{\frac{1}{2}}(\alpha_2^*-\alpha_2)| \geqq 2w) = 0.0524,$$

$$P(|T^{\frac{1}{2}}(\tilde{\alpha}_2-\alpha_2)| \geqq 2w) = 0.0776,$$

for the General Electric equation, and

$$P(|T^{\frac{1}{2}}(\alpha_2^*-\alpha_2)| \geqq 2w) = 0.0526,$$

$$P(|T^{\frac{1}{2}}(\tilde{\alpha}_2-\alpha_2)| \geqq 2w) = 0.0738,$$

for the Westinghouse equation. Thus, in this case the difference between the tail area probabilities of the two estimators appears to be quite large. We note

Table 2
Comparison of finite sample distributions in micro investment equations.

		General Electric				Westinghouse			
		α_1		α_2		α_1		α_2	
x/w	Aitken	SURE	SELS	SURE	SELS	SURE	SELS	SURE	SELS
−3.0	0.0013	0.0015	0.0021	0.0017	0.0040	0.0016	0.0027	0.0017	0.0036
−2.8	0.0025	0.0029	0.0038	0.0032	0.0067	0.0031	0.0047	0.0032	0.0061
−2.6	0.0046	0.0052	0.0066	0.0058	0.0109	0.0055	0.0080	0.0058	0.0100
−2.4	0.0081	0.0090	0.0110	0.0099	0.0171	0.0095	0.0132	0.0099	0.0160
−2.2	0.0139	0.0151	0.0180	0.0164	0.0261	0.0158	0.0209	0.0164	0.0246
−2.0	0.0227	0.0244	0.0283	0.0262	0.0388	0.0254	0.0321	0.0263	0.0369
−1.8	0.0359	0.0382	0.0431	0.0405	0.0561	0.0395	0.0479	0.0406	0.0539
−1.6	0.0547	0.0576	0.0636	0.0605	0.0790	0.0592	0.0694	0.0606	0.0764
−1.4	0.0807	0.0841	0.0910	0.0875	0.1084	0.0860	0.0976	0.0876	0.0105
−1.2	0.1150	0.1187	0.1263	0.1226	0.1448	0.1209	0.1334	0.1227	0.1418
−1.0	0.1586	0.1625	0.1702	0.1665	0.1888	0.1647	0.1774	0.1666	0.1858
−0.8	0.2118	0.2155	0.2229	0.2194	0.2401	0.2177	0.2296	0.2194	0.2374
−0.6	0.2742	0.2774	0.2837	0.2807	0.2982	0.2793	0.2894	0.2808	0.2959
−0.4	0.3445	0.3469	0.3515	0.3493	0.3620	0.3482	0.3556	0.3494	0.3604
−0.2	0.4207	0.4219	0.4244	0.4232	0.4299	0.4227	0.4266	0.4233	0.4291

that most of the difference results from the inefficiency of the SELS estimator. For the tail probability of the SURE is close to that of the Aitken estimator (0.05 here) and, from (40), we find that the disturbance correlation coefficient is $\rho = 0.8128$ so that we would expect the Aitken estimator to have a definite efficiency gain, at least for some of the coefficients.[7]

4. Final comments

Although the results of section 2 are quite general they are limited by the assumptions of non-random exogenous variables and normally distributed disturbances. The former assumption is of some importance since the two-stage estimator is known to be asymptotically less efficient than the Aitken estimator when there are lagged dependent variables amongst the regressors [Maddala (1971)]. Since asymptotic series expansions of the Edgeworth type are known to be valid under more general assumptions than those made in the present paper[8] further research along these lines to include such cases seems desirable.

[7]The precise form of the efficiency gain depends on the elements of the observation matrices X_1 and X_2. Zellner and Huang (1962) show that if we wish to consider the generalised variance of the estimates of the coefficients in a single equation then we can express this gain in terms of the canonical correlation coefficients of X_1 and X_2 as well as ρ.

[8]See Phillips (1975, 1977) and Sargan (1976).

References

Anderson, T.W., 1958, Introduction to multivariate statistical analysis (Wiley, New York).
Kakwani, N.C., 1967, The unbiasedness of Zellner's seemingly unrelated regression equations estimator, Journal of the American Statistical Association 62, 141–142.
Kataoka, Y., 1974, The exact finite sample distributions of joint least squares estimators for seemingly unrelated regression equations, Economic Studies Quarterly 25, 36–44.
Kmenta, J. and R.F. Gilbert, 1968, Small sample properties of alternative estimators of seemingly unrelated regressions, Journal of the American Statistical Association 63, 1180–1200.
Maddala, G.S., 1971, Generalised least squares with an estimated covariance matrix, Econometrica 39, 23–33.
Malinvaud, E., 1970, Statistical methods of econometrics (North-Holland, Amsterdam).
Metha, J.S. and P.A.V.B. Swamy, 1975, Further evidence on the relative efficiencies of Zellner's seemingly unrelated regressions estimator, Special Studies Paper no. 70 (Federal Reserve Board, Washington, DC).
Phillips, P.C.B., 1975, A general theorem in the theory of asymptotic expansions as approximations to the finite sample distributions of econometric estimators, University of Essex Discussion Paper no. 58, Econometrica, forthcoming.
Phillips, P.C.B., 1977, Approximations to some finite sample distributions associated with a first order stochastic difference equation, Econometrica 45, 463–485.
Sargan, J.D., 1975, Gram–Charlier approximations applied to t-ratios of k-class estimators, Econometrica 43, 327–346.
Sargan, J.D., 1976, Econometric estimators and the Edgeworth approximation, Walras lecture at the Econometric Society meeting (San Francisco, CA); Econometrica, forthcoming.
Sargan, J.D. and E.M. Mikhail, 1971, A general approximation to the distribution of instrumental variable estimates, Econometrica 39, 131–169.
Theil, H., 1971, Principles of econometrics (Wiley, New York).
Zellner, A., 1962, An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias, Journal of the American Statistical Association 57, 348–368.
Zellner, A., 1963, Estimators for seemingly unrelated regression equations: Some finite sample results, Journal of the American Statistical Association 58, 977–992.
Zellner, A., 1972, Corrigenda, Journal of the American Statistical Association 67, 255.
Zellner, A. and D.S. Huang, 1962, Further properties of efficient estimators for seemingly unrelated regression equations, International Economic Review 3, 300–313.