International Journal of Research Studies in Computer Science and Engineering (IJRSCSE)
Volume 10, Issue 2, 2024, PP 50-57
ISSN 2349-4840 (Print) & ISSN 2349-4859 (Online)
DOI: http://dx.doi.org/10.20431/2349-4859.1002005
www.arcjournals.org

Using Seemingly Unrelated Regression with Data Sets that Contain Multiple Measurements in Individual Sampling Units

West P.W[1*], Ratkowsky D. A[2]

[1]*Faculty of Science and Engineering, Southern Cross University, Lismore, NSW, Australia*

[2]*Tasmanian Institute of Agriculture, University of Tasmania, Hobart, Tasmania, Australia*

***Corresponding Author:** ***West P.W**, Faculty of Science and Engineering, Southern Cross University, Lismore, NSW, Australia. **Email:** pwest@nor.com.au*

Abstract: *A data set may contain multiple measurements from each sampling unit within it. If ordinary least-squares (OLS) regression is then applied to fit a model to the data, the covariance matrix of the fitted parameters is usually underestimated. Methods including adjusted ordinary least-squares (AOLS) regression and mixed-modelling have been used to overcome this problem. However, seemingly unrelated regression (SUR) does not seem to have been used previously in this context. Using two examples of forestry interest, simulation studies were undertaken to explore the efficacy of SUR in this respect. AOLS and SUR both involve a two-stage modelling process requiring there be at least one more observation in each sampling unit than there are parameters in the first-stage model. All methods tested here produced unbiased estimators of the model parameters. AOLS, SUR and mixed models gave unbiased estimators of the parameter covariance matrix. The latter two were appreciably, and equally, more efficient estimators than AOLS. It was concluded that the two-stage model approach with SUR holds considerable potential for use with such data sets.*

1. INTRODUCTION

It is common to find a data set in which several measurements have been made on each of the sampling units within it. Such data sets are often said to contain 'multiple' or 'repeated' measurements, or the data are said to be 'clustered' within the sampling units. If the multiple measurements are made at different times, the data are often termed 'longitudinal'. If regression analysis is to be applied with such data sets, allowance must be made for the covariance amongst the residuals in each sampling unit that will arise from the fitted regression. If this is not done, bias may occur in the estimated covariance matrix of the parameter estimates from the fitted model.

Various approaches have been proposed to deal with this problem (e.g. West et al. 1984; West, 1995; Pinheiro & Bates, 2000; Litière et al., 2007; Galbraith et al., 2010; Fitzmaurice et al. 2011; McNeish et al., 2017). An older approach has been to use generalized least-squares (GLS) regression. This involves incorporation, into least-squares regression theory, a weighting matrix designed to reflect, and allow for, the variance-covariance matrix of the error term of the model. Various works (e.g. Ferguson & Leech, 1978; Davis & West, 1981; West et al., 1986; Kauermann & Carroll, 2001; Fitzmaurice at al., 2011; McNeish et al., 2017; West & Ratkowsky, 2022) describe this approach and ways that have been developed to estimate the matrix. A more recent approach, which has become widely adopted, is the use of 'mixed' models. Such models include 'fixed' effects; these relate the response variable of the model to predictor variable(s), irrespective of the sampling unit to which each observation belongs. They also include 'random' effects, which are particular to the sampling units and determine how the relationship between response and predictors is affected by the presence of the multiple measurements within the sampling units. However, study of various of the works cited above will confirm that the approach to be used with such data sets should depend on the circumstances of the data themselves; there is no certainty that mixed models will necessarily yield better estimates of the parameters, or their covariance matrix, than other methods.

McNeish et al. (2017) criticized strongly the frequent insistence over recent times that mixed models should be used to deal with such data sets. Further to their work, the present authors undertook

simulation studies to examine the efficacy of several regression methods with such data sets, using two examples which employed multiple linear regressions (West & Ratkowsky, 2022). To estimate the model parameters and their covariance matrix, they compared four methods, namely ordinary least-squares (OLS) regression, generalised least-squares (GLS) regression using the 'sandwich' estimator (Kauermann & Carroll, 2001; Fitzmaurice et al., 2011; McNeish et al., 2017), 'adjusted ordinary least-squares' (AOLS) regression (Davis & West, 1981; West et al., 1986) and the use of mixed models.

All these methods produced unbiased estimators of the parameters. OLS and the sandwich estimator both underestimated the parameter standard errors. For the mixed models, it was found difficult to choose the most appropriate form of the model to apply. However, eventually a form was identified that generally provided an unbiased estimator of the parameter standard errors, although it sometimes gave problems when converging to a solution. AOLS gave an unbiased estimator of the standard errors, but occasionally yielded an estimate of the parameter covariance matrix that was not positive-definite. However, when each method worked satisfactorily, the mixed modelling approach provided an appreciably more precise estimator of the parameters than did AOLS, with parameter standard errors often being about 35% less than those with AOLS (West & Ratkowsky, 2022, Table 2).

The present work undertakes simulations with the same two examples as used by West & Ratkowsky (2022). It examines the efficacy of using a GLS approach to fit the models that has not been used before in this context, namely, seemingly unrelated regression (SUR or SURE) (Zellner, 1962; Srivastava & Giles, 1987). The application of SUR is compared with the use of both mixed models and AOLS. Both SUR and AOLS involve a two-stage modelling process. The first stage involves fitting a model separately in each and every sampling unit, using the repeated measurements in each. The estimates of each parameter from all those first-stage models are then related to predictor variables that are properties of the sampling units themselves to develop second-stage models. The first- and second-stage models are then pooled and fitted as a single model to all the data from all the sampling units. This process means that both the SUR and AOLS systems are constrained in that each and every one of the sampling units in their data must have at least one more observation than there are parameters in the first-stage model so that it may be fitted in each sampling unit. The mixed model system is not constrained in that way and is relevant whenever there is simply more than one observation in some or all of the sampling units.

2. METHODS

2.1. Examples

The first of the two examples used here concerned stem wood volume production in plantation forests of radiata pine (*Pinus radiata* D. Don) in South Australia (Ferguson & Leech, 1978). The data consisted of several measurements of the total wood volume in tree stems (V, $m^3\ ha^{-1}$) in each of a set of forest plots (the sampling units), measured at various ages (A, tens of years) when the forest was 10–55 years old. The number and ages of measurements varied between plots. The first-stage model fitted in each sampling unit was

$$\ln(V) = p_1 + p_2/A \qquad (1)$$

where $\ln(\cdot)$ denotes natural logarithms and p_1 and p_2 were parameters. The parameter estimates from these models were then related to a measure of the productive capacity for growth of the site on which the trees of each plot were growing. This predictor (P, $m^3\ ha^{-1}$) was particular to the plot concerned and was determined as the stem wood volume in the plot at 10 years of age; the use of such 'phytocentric' measures of site productive capacity is common in forestry science (West, 2015, Sect. 8.7). The second-stage models were then

$$p_1 = p_{11} + p_{12} \ln(P) \qquad (1a)$$

and

$$p_2 = p_{21} \ln(P) \qquad (1b)$$

where p_{11}, p_{12} and p_{21} were parameters. Substituting Equations (1a, b) into Equation (1) yielded the full model relating wood volume to age and site productive capacity as

$$\ln(V) = p_{11} + p_{12} \ln(P) + p_{21} \ln(P)/A \qquad (2)$$

After some consideration, this was used as the basis for the mixed model

$$\ln(V) = (p_{11} + v_{11i}) + p_{12} \ln(P) + (p_{21} + v_{21i}) \ln(P)/A \qquad (3)$$

where v_{11i} and v_{21i} were random effects for the ith sampling unit.

The second example (Grassia & De Boer, 1980) concerned the change between six months and eight years of age in the number of molar teeth (M) found at various ages (A, tens of years) in the mouths of each of a number of Australian wallabies (the sampling units) of different sexes (S, 1=male, −1=female). The first-stage model was

$$M = q_1 + q_2 \ln(A) \qquad (4)$$

where q_1 and q_2 were parameters. The second-stage models were

$$q_1 = q_{11} + q_{12} S \qquad (4a)$$

and

$$q_2 = q_{21} \qquad (4b)$$

where q_{11}, q_{12} and q_{21} were parameters. This yielded the full model relating tooth number to wallaby age and sex as

$$M = q_{11} + q_{12} S + q_{21} \ln(A) \qquad (5)$$

The mixed model then tested was

$$M = (q_{11} + w_{11i}) + q_{12} S + q_{21} \ln(A) \qquad (6)$$

where w_{11i} was a random effect for the ith wallaby.

2.2. Simulation Data Sets

For each of the *P. radiata* and wallaby examples, 1,000 simulated data sets were constructed for use in comparing results obtained with the various regression estimators being considered here. In each simulated data set for the *P. radiata* example, there were 20 plots, each with a productive capacity within the range 57–253 m^3 ha^{-1} and each with 3–9 observations of volume at ages within the range 10–55 yr. In the wallaby case, each simulated data set had 43 wallabies, 17 male and 26 female, with 3−24 observations from each at randomly chosen ages between 0.5–8 yr, but with ages at least 73 days apart.

The simulated data sets were based on chosen parameter values for Models (2) and (5) and assumed matrices representing the covariances between the residuals of the models. Sets of multivariate random normal deviates were determined to construct each simulated data set. Full details describing these processes are given by West & Ratkowsky (2022).

2.3. Simulations

For each simulated data set, the full models with fixed effects only (2 and 5) were fitted using OLS and AOLS regression. The mathematical details of AOLS are shown in the Appendix of West & Ratkowsky (2022). The first-stage models (1 and 4) were fitted independently to the data of each sampling unit (plot or wallaby) using OLS regression. Their parameter estimates were then treated as sets of response vectors (two sets for each example) and the second-stage models (1a, b and 4a, b) were then fitted jointly using SUR. The mathematical details of the methods used with SUR are given in the Appendix. The models with random effects (3 and 6) were fitted using the maximum likelihood methods of the NLMIXED procedure of the SAS® statistical package[1]. The OLS, AOLS and SUR models were fitted using computer programmes written by one of us (PWW); note that the MODEL procedure of the SAS package may be used also to fit SUR models.

These processes yielded 1,000 estimates of the model parameters and their covariance matrices for each of the *P. radiata* and wallaby examples and for each of the four estimation methods, OLS, AOLS, mixed models and SUR.

3. RESULTS

Tables 1 and 2 show results copied from Tables 1 and 2, respectively, of West & Ratkowsky (2022). Those results had been obtained from the 1,000 simulations in both examples for OLS, AOLS and mixed models. The results obtained here for SUR have been added to both tables.

[1]Documentation for the SAS statistical package is available at https://support.sas.com/en/documentation.html (accessed October 2024).

The results of Table 1 suggest that all methods yielded estimates of the parameters with little bias. For the present SUR results, this was confirmed as follows using the following formal test as described at Equation (27) of West et al. (1986). Suppose one has a $B\times1$ vector, β^*, of parameter estimates from a model determined using a data set containing N observations and with a corresponding estimate of their $B\times B$ covariance matrix, $\hat{V}$. Suppose it is desired to determine if those parameter estimates differ significantly from some chosen $B\times1$ vector, β, of parameter values. Then, the test statistic $\tilde{F}$, determined as

Table 1. *For 1,000 simulations of both examples considered here, with each of four estimation methods (OLS, AOLS, mixed models and SUR as discussed in the text), results are shown for the population values of the model parameters assumed when constructing the simulation data sets and the average of the 1,000 estimates of the parameters obtained with each method. Results for OLS, AOLS and the mixed models are the same as shown in Table 1 of West & Ratkowsky (2022)*

Param-eter	Population value	OLS	AOLS	Mixed models	SUR
P. radiata example					
p_{11}	3.957	3.930	3.930	3.962	3.965
p_{12}	0.697	0.704	0.704	0.698	0.697
p_{21}	-0.511	-0.518	-0.518	-0.519	-0.518
Wallaby example					
q_{11}	9.470	9.468	9.468	9.479	9.479
q_{12}	-0.711	-0.709	-0.709	-0.712	-0.712
q_{21}	1.480	1.474	1.474	1.481	1.481

Table 2. *Results for simulations as in Table 1 for the standard errors of the 1,000 parameter estimates obtained for each estimation method (shown as Population value) and the average of the 1,000 estimates of the standard error obtained with each method (shown as Simulation average)*

Param-eter	OLS		AOLS		Mixed model		SUR	
	Popul-ation value	Simulation average	Popul-ation value	Simulation average	Popul-ation value	Simulation average	Popul-ation value	Simulation average
P. radiata example								
p_{11}	0.358	0.167	0.358	0.348	0.194	0.173	0.200	0.185
p_{12}	0.071	0.034	0.071	0.070	0.039	0.035	0.040	0.037
p_{21}	0.028	0.013	0.028	0.028	0.021	0.021	0.021	0.021
Wallaby example								
q_{11}	0.318	0.118	0.318	0.301	0.244	0.202	0.239	0.232
q_{12}	0.216	0.061	0.216	0.222	0.201	0.201	0.126	0.127
q_{21}	0.111	0.085	0.111	0.105	0.030	0.012	0.028	0.027

$$\tilde{F} = (\beta - \beta^*)' \hat{V}^{-1} (\beta - \beta^*)/B, \qquad (7)$$

is distributed as the Fisher F-distribution, $F_{B,N-B}$.

This test was applied to the present SUR results by testing the departure of the vector of means of the N=1,000 simulation estimates (shown in the 'SUR' column of Table 1) of the parameters (B=3 for both examples) from the vector of their population values (shown in the 'Population value' column) and using the observed covariance matrix of the 1,000 estimates. There was no significant difference between the two parameter vectors ($p<0.001$), suggesting that SUR provided an unbiased estimator of the parameters. It had been concluded in the previous work that this was true also for the OLS, AOLS and mixed model results.

Table 2 compares the average of the 1,000 simulation estimates of the standard errors of each of the parameters (Simulation average columns) of the two examples against the observed standard errors of the 1,000 estimates (Population value columns). As expected, OLS grossly underestimated the standard errors for it took no account of the correlation between the residuals from the fitted model in each of the sampling units. The AOLS, mixed models and SUR methods all closely estimated their population standard errors, with the exception of q_{21} in the wallaby example where the mixed model grossly under-estimated the population value. It was found by West & Ratkowsky (2022) that replacing Model (6) with an alternative,

$$M = q_{11} + (q_{12} + w_{12i})\, S + (q_{21} + w_{21i})\, \ln(A), \qquad (8)$$

where w_{12i} and w_{21i} were random effects for the ith wallaby, would overcome this problem. However, this version of the model failed to converge satisfactorily in about 18% of the simulation data sets.

The results of Table 2 suggest that SUR closely estimated the population standard errors. To undertake a formal test of those results, the parameter estimates and the corresponding estimates of their covariance matrix were tested, using Equation (7), for each individual simulation result, using B=3 for both examples and the number of observations in the SUR fit to the second-stage models, that is N=20 for the *P. radiata* example and N=43 for the wallaby example. Table 3 shows the proportion of times that a statistically significant deviation was found between the observed parameter estimates and their population values. Those proportions correspond closely to what would be expected from F tests at p=0.1, 0.05 or 0.01. That is to say, it appears that SUR provided an unbiased estimator of the covariance matrix of the parameter estimates, consistent with the standard error comparisons in Table 2. In the previous work, this appeared to be true also for AOLS and the mixed models, or at least for the satisfactorily converged cases with Model (8) in the wallaby example.

Also of note, for both examples, is that the standard error population values shown in Table 2 are much lower for the mixed models and SUR than they are for AOLS; they average about 40% lower over all the parameters in both examples. That is to say, both mixed models and SUR provided appreciably more efficient estimators than AOLS.

Table 3. *For both examples, the proportion of times that, from 1,000 simulations using the SUR method, the estimates of the parameters of the model differed significantly from their population values, at three arbitrarily chosen probability values*

Probability	Example	
	P. radiata	**Wallaby**
0.1	0.111	0.088
0.05	0.047	0.045
0.01	0.008	0.009

4. Discussion And Conclusions

SUR was devised initially (Zellner, 1962) to deal with data sets where several different response variables had been measured on a set of sampling units. It would be possible to simply fit a model for each of those variables separately with OLS. However, SUR allows that there may be some level of intercorrelation amongst the residuals of each of those models; the additional information then used by SUR to fit the models jointly should render it a more efficient estimator than OLS, at least asymptotically (Srivastava & Giles, 1987, p. 11). The present authors reviewed some 14 examples from their principal area of interest, forest science, where a comparison had been made between results when several models were fitted either separately with OLS or jointly with SUR (West & Ratkowsky, 2023). None of these involved a two-stage model, but all were small sample examples. It was concluded that SUR might sometimes be a little more efficient than OLS, but there was generally little advantage gained in the goodness-of-fit to the data and sometimes there was a slight disadvantage.

A circumstance in forest science where SUR has been found useful involves the construction of models that predict the biomasses of each of different parts of individual trees (leaves, stems, roots etc.) in relation to more easily measured tree characteristics, such as stem diameters or heights. Additional models are then added to the system to predict the total tree biomass, models that incorporate the same parameters as used in the models for the separate parts. SUR is used to fit these models jointly and the system ensures that predicted values of the separate parts sum correctly to the predicted total tree biomass. There are numerous examples of this (e.g. Bi et al., 2004; Trautenmuller et al., 2021; Xiong et al., 2023).

The present work suggests that SUR may offer appreciable advantage when applied to data sets that contain multiple measurements of some response variable in individual sampling units. However, the circumstances considered here apply only to models that can be developed through a two-stage process. At the first stage, a relationship between the set of response and predictor variables must be established separately for each and every sampling unit. At the second stage, each parameter estimate from the first-stage models is related to predictor variables that are fixed characteristics of the sampling units themselves. SUR may then be used to fit the second-stage models jointly.

Simulations were carried out here involving a two-stage model system with two small sample examples. The simulations were based on longitudinal data sets with repeated measurements at different ages of a certain characteristic in each sampling unit. When SUR was used to fit the second-stage models jointly, it was found to provide an unbiased estimator of both the parameters of the full model and their variances. This contrasted markedly with results when the full model was fitted using OLS regression, which produced an unbiased estimator of the parameters, but grossly underestimated their variances. This was a consequence of the failure of OLS to account for the correlations between the residuals in each sampling unit that resulted from its fit to the full model. In effect, SUR avoids the consequences of those correlations because its fit to the second-stage models involves only one observation (of each of the first-stage parameters) from each sampling unit.

These results using SUR were compared with simulation results for two other methods. The first, AOLS, is also based on the two-stage model approach, but fits the full model using a GLS approach that involves use of the residuals from the OLS fit (see the Appendix of West & Ratkowsky 2022) and has been shown theoretically to provide an unbiased estimator of the parameters in small samples (West et al., 1986). Indeed, in the present examples, AOLS was found to provide an unbiased estimator of both the parameters and their variances, but it was a markedly less efficient estimator than SUR (Table 2).

The second comparison was with the use of mixed models, a method that also attempts to take account of the correlations amongst residuals in each sampling unit when the full model is fitted. In general, this method was found to provide unbiased estimators of both the parameters and their variances and was equally as efficient as SUR. However, experience with this approach in previous work (West & Ratkowsky, 2022) found that it was both difficult to determine an appropriate parameterisation to use for the mixed models and that the algorithms used to fit them could fail quite often; no such difficulties were encountered in applying the algorithm used here to fit SUR (shown in the Appendix). However, mixed models have an advantage over SUR (and AOLS) in that they are not based on the two-stage model approach. Both SUR and AOLS require that the first-stage model actually be fitted in each sampling unit. This requires that there must be at least one more observation in each sampling unit than there are parameters in the first-stage model. That restriction does not apply in the case of mixed models where some sampling units may have only one or very few observations.

In conclusion, the present results suggest that SUR may offer an opportunity to be used to provide an unbiased and efficient estimator when fitting regression models in data sets that contain multiple measurements in individual sampling units. However, its use is restricted in that the data must be appropriate to allow the two-stage modelling approach. Also, it must be borne in mind that the present work has considered linear models only. Of course, nonlinear models may be used also with SUR, as discussed briefly by Srivastava & Giles (1987, Sect 11.6). Many other issues surround the use of nonlinear regression as various texts discuss (e.g. Ratkowsky, 1983, 1990; Draper & Smith, 1988; Pinheiro & Bates, 1995; Davidian & Giltinan 2003). It is obvious that many other examples need to be considered to develop these conclusions further.

REFERENCES

Bi H., Turner J., & Lambert M.J. 2004. Additive biomass equations for native eucalypt forest trees of temperate Australia. Trees., 18, 467−479.

Davidian M., & Giltinan D.M. 2003. Nonlinear models for repeated measurement data: An overview and update. J. Agric. Biol. & Env. Statist, 8, 387−419.

Davis A.W, & West P.W. 1981. Remarks on 'Generalized Least Squares Estimation of Yield Functions' by I.S. Ferguson and J.W. Leech. For. Sci., 27, 233−239.

Draper N.R., & Smith H. 1988. Applied regression analysis. 3rd ed. Wiley, New York.

Ferguson I.S., & Leech J.W. 1978. Generalized least squares estimation of yield functions. For. Sci., 24, 27–42.

Fitzmaurice G.M., Laird N.M., & Ware J.H. 2011. Applied longitudinal analysis. 2nd ed. Wiley Interscience, Hoboken (NJ).

Galbraith S., Daniel J.A., & Vissel B. 2010. A study of clustered data and approaches to its analysis. J. Neuroscience., 30, 10601−10608.

Grassia A., De Boer E.S. 1980. Some methods of growth curve fitting. Mathematical Scientist., 5, 91–103.

Kauermann G., & Carroll R.J. 2001. A note on the efficiency of sandwich covariance estimation J. Amer. Stat. Assoc., 96, 1387−1396.

Litière S., Alonso A., & Molenberghs G. 2007. Type I and type II error under random-effects missspecification in generalized linear mixed models. Biometrics, 63, 1038−1044.

McNeish D., Stapleton L.M, & Silverman R.D. 2017. On the unnecessary ubiquity of hierarchical linear modeling. Psychol. Methods., 22, 114−140.

Pinheiro J.C, & Bates D.M. 1995. Approximations to the log-likelihood function in the nonlinear mixed-effects model. J. Comput. & Graph. Statist., 4, 12–35.

Pinheiro J.C., & Bates D.M. 2000. Mixed-effects models in S and S-Plus. Springer Verlag, New York.

Ratkowsky D.A. 1983. Nonlinear regression modeling. Marcel Dekker, New York.

Ratkowsky D.A. 1990. Handbook of nonlinear regression models. Marcel Dekker, New York.

Srivastava V.K., & Giles D.E.A. 1987. Seemingly unrelated regression equations models. CRC Press, Boca Raton.

Trautenmuller J.W, Netto S.P., Balbinot R., Watzlawick L.F., Dalla Corte A.P., Sanquetta C.R., & Behling A. 2021. Regression estimators for aboveground biomass and its constituent parts of trees in native southern Brazilian forests. Ecol. Indicators., 130, 108025.

West P.W. 1995. Application of regression analysis to inventory data with measurements on successive occasions. For. Ecol. Manage., 71, 227−234.

West P.W. 2015 Tree and forest measurement. 3rd ed. Springer International Publishing, Switzerland.

West P.W., & Ratkowsky D.A. 2022. Simulation studies comparing fixed effect and mixed models in data sets with multiple measurements in individual sampling units. J. Statist. Comp. Simul., 92, 81−100.

West P.W., & Ratkowsky D.A. 2023. A state-space growth model for *Eucalyptus pilularis* in subtropical Australia, fitted with and without seemingly unrelated regression. Australian Forestry., 86, 134−142.

West P.W., Ratkowsky D.A., & Davis A.W. 1984. Problems of hypothesis testing of regressions with multiple measurements from individual sampling units. For. Ecol. Manage., 7, 207−224.

West P.W., Davis A.W., & Ratkowsky D.A. 1986. Approaches to regression analysis with multiple measurements from individual sampling units. J. Statist. Comp. Simul., 26, 149−175.

Xiong N.A, Qiao Y., Ren H.R, Zhang L., Chen R.H., & Wang J. 2023. Comparison of parameter estimation methods based on two additive biomass models with small samples. Forests., 14, 1655.

Zellner A. 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. J. Amer. Stat. Assoc., 57, 348−368.

APPENDIX

Seemingly Unrelated Regression (SUR)

The principles of SUR as applied here are described at pp. 3–17 of Srivastava and Giles (1987). The model system used is based around M multiple linear regression equations, each of which is based on a set of T observations of M characteristics that have been made on a set of sampling units that are of interest to the observer. The ith (i=1…M) of these equations may be represented as

$$y_i = X_i\beta_i + u_i, \qquad \text{(A1)}$$

where y_i is a T×1 vector of responses of the ith variable being considered, X_i is a T×K_i matrix of predictors for that variable, β_i is a K_i×1 vector of parameters to be estimated and u_i is a T×1 vector of residuals. These models may then be written as

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_M \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_M \end{bmatrix} \qquad \text{(A2)}$$

and expressed in the compact form

$$y = X\beta + u, \qquad \text{(A3)}$$

where y and u are MT×1 vectors, β is a K×1 vector, where K ($=\Sigma_{i=1\ldots M}\ K_i$) is the total number of parameters and X is an MT×K matrix.

SUR then attempts to take account of any intercorrelation that occurs between the residuals of the various models (A1). If this intercorrelation exists, then fitting those models jointly with the SUR techniques may yield a more efficient estimator than if the models were fitted separately. To do this, SUR assumes that the residuals are related between each model through an $MT \times MT$ covariance matrix Ψ, which is the expected value of uu'. In general, values for this matrix are unknown *a priori* and methods to obtain an estimate of it, $\hat{\Psi}$, are given by Srivastava and Giles (1987). One such method there is termed the 'restricted residual' method. The first step in this is to fit each of the M models (A1) separately by ordinary least-squares regression to yield $T \times 1$ vectors of estimates of the residuals, $\hat{u}_1$, $\hat{u}_2$ … $\hat{u}_M$. These vectors are then used to estimate M^2 scalar variance and covariance estimates, $\hat{s}_{ij}$ $(i,j=1 \ldots M)$, as

$$\hat{s}_{ij} = (u_i' \, u_j)/T_{ij}, \qquad \text{(A4)}$$

where

$$T_{ij} = [(T-K_i)(T-K_j)]^{½} \quad \text{(A5)}$$

These scalars may then be arranged to form an $M \times M$ variance-covariance matrix. Occasionally it may be found that this matrix is not positive definite, in which case there are two possible alternatives. The first is to simply replace all the T_{ij} with T, although the estimators $\hat{s}_{ij}$ may not then be unbiased. The second is to obtain unbiased estimators using

$$T_{ij} = \text{tr}(\hat{P}_i \, \hat{P}_j), \qquad \text{(A6)}$$

where

$$\hat{P}_i = I_T - X_i(X_i'X_i)^{-1}X_i', \quad i=1 \ldots M \qquad \text{(A7)}$$

and I_T is a $T \times T$ identity matrix. Neither of these alternatives was found necessary to apply in the present work. Note that in all these computations $\hat{s}_{ij}=\hat{s}_{ji}$. The estimator $\hat{\Psi}$ may then be constructed as the $MT \times MT$ matrix

$$\hat{\Psi} = \begin{bmatrix} \hat{s}_{11} I_T & \cdots & \hat{s}_{1M} I_T \\ \vdots & & \vdots \\ \hat{s}_{1M} I_T & \cdots & \hat{s}_{MM} I_T \end{bmatrix} . \qquad \text{(A8)}$$

The full SUR model (A3) may then be fitted using generalised least-squares regression, where an estimate, $\hat{\beta}$, of the vector of parameters, β, may be determined as

$$\hat{\beta} = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}y \qquad \text{(A9)}$$

and an estimate of their $K \times K$ covariance matrix, $V(\beta)$, as

$$V(\hat{\beta}) = (X'\Psi^{-1}X)^{-1}, \qquad \text{(A10)}$$

as described in Srivastava and Giles (1987) at Equations (2.5) and (2.7). Because this method uses an estimate of the matrix Ψ that is based on observed residuals, it is the most efficient estimator of the model only asymptotically.

Citation: *West P.W, Ratkowsky D. A. Using Seemingly Unrelated Regression with Data Sets that Contain Multiple Measurements in Individual Sampling Units. International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), vol 10, no. 2, 2024, pp. 50-57. DOI: https://doi.org/10.20431/2349-4859.1002005.*