Seemingly Unrelated Regressions

Hyungsik Roger Moon
Department of Economics
University of Southern California
moonr@usc.edu

Benoit Perron
Département de sciences économiques, CIREQ, and CIRANO
Université de Montréal
benoit.perron@umontreal.ca

July 2006

Abstract

This article considers the seemingly unrelated regression (SUR) model first analyzed by Zellner (1962). We describe estimators used in the basic model as well as recent extensions.

Seemingly unrelated regressions

A seemingly unrelated regression (SUR) system comprises several individual relationships that are linked by the fact that their disturbances are correlated. Such models have found many applications. For example, demand functions can be estimated for different households (or household types) for a given commodity. The correlation among the equation disturbances could come from several sources such as correlated shocks to household income. Alternatively, one could model the demand of a household for different commodities, but adding-up constraints leads to restrictions on the parameters of different equations in this case. On the other hand, equations explaining some phenomenon in different cities, states, countries, firms or industries provide a natural application as these various entities are likely to be subject to spillovers from economy-wide or worldwide shocks.

There are two main motivations for use of SUR. The first one is to gain efficiency in estimation by combining information on different equations. The second motivation is to impose and/or test restrictions that involve parameters in different equations. Zellner (1962) provided the seminal work in this area, and a thorough treatment is available in the book by Srivastava and Giles (1987). A recent survey can be found in Fiebig (2001) . This chapter selectively overviews the SUR model, some of the estimators used in such systems and their properties, and several extensions of the basic SUR model. We adopt a Classical perspective, although much Bayesian analysis has been done with this model (including Zellner's contributions).

Basic linear SUR model

Suppose that y_{it} is a dependent variable, $x_{it} = (1, x_{it,1}, x_{it,2}, ..., x_{it,K_i-1})'$ is a K_i-vector of explanatory variables for observational unit i, and u_{it} is an unobservable error term, where the double index it denotes the t^{th} observation of the i^{th} equation in the system. Often t denotes time and we will refer to this as the time dimension, but in some applications, t could have other interpretations, for example as a location in space. A classical linear SUR model is a system of linear regression equations,

$$\begin{aligned} y_{1t} &= \beta_1' x_{1t} + u_{1t} \\ &\vdots \\ y_{Nt} &= \beta_N' x_{Nt} + u_{Nt} \end{aligned}$$

where $i = 1, \cdots, N$, and $t = 1, ..., T$. Denote $L = K_1 + \cdots + K_N$. Further simplification in notation can be accomplished by stacking the observations either in the t dimension or for each i. For example, if we stack for each observation t, let $Y_t = [y_{1t}, ..., y_{Nt}]'$, $\tilde{X}_t = diag\left(x_{1t}, x_{2t}, ..., x_{NT}\right)$, a block-diagonal matrix with $x_{1t}, ..., x_{NT}$ on its diagonal, $U_t = [u_{1t}, ...u_{Nt}]'$, and $\beta = \left[\beta_1', ..., \beta_N'\right]'$. Then,

$$Y_t = \tilde{X}_t' \beta + U_t. \tag{1}$$

Another way to present the SUR model is to write it in a form of a multivariate regression with parameter restrictions. For this, define $X_t = [x'_{1t}, x'_{2t}, ..., x'_{NT}]'$ and $A(\beta) = diag(\beta_1, ..., \beta_N)$ to be a $(L \times N)$ block diagonal coefficient matrix. Then, the SUR model in (1) can be rewritten as

$$Y_t = A(\beta)' X_t + U_t, \tag{2}$$

and the coefficient $A(\beta)$ satisfies

$$vec(A(\beta)) = G\beta, \tag{3}$$

for some $(NL \times L)$ full rank matrix G. In the special case where $K_1 = \cdots = K_N = K$, we have $G = diag(i_1, \ldots, i_N) \otimes I_K$, where i_j denotes the j'th column of the $N \times N$ identity matrix I_N.

Assumption:

In the classical linear SUR model, we assume that for each $i = 1, ..., N$, $x_i = [x_{i1}, ..., x_{iT}]'$ is of full rank K_i, and that conditional on all the regressors $X' = [X_1, ..., X_T]$, the errors U_t are *iid* over time with mean zero and homoskedastic variance $\Sigma = E(u_t u'_t \mid X)$. Furthermore, we assume that Σ is positive definite and denote by σ_{ij} the $(i,j)^{th}$ element of Σ, that is, $\sigma_{ij} = E(u_{it}u_{jt}|X)$.

Under this assumption, the covariance matrix of the entire vector of disturbances $U' = [U_1, ..., U_T]$ is given by $E\left[vec(U)(vec(U))'\right] = \Sigma \otimes I_T$.

Estimation of β:

In this section we summarize four estimators of β that have been widely used in applications of the classical linear SUR. Other estimators (such as Bayes, empirical Bayes, or shrinkage estimators) have also been proposed. Interested readers should refer to Srivastava and Giles (1987) and Fiebig (2001).

1. Ordinary least squares (OLS) estimator:

The first estimator of β is the ordinary least squares (OLS) estimator of Y_t on regressor X_t,

$$\hat{\beta}_{OLS} = \left(\sum_{t=1}^{T} \tilde{X}_t \tilde{X}'_t\right)^{-1} \sum_{t=1}^{T} \tilde{X}_t Y_t.$$

This is just the vector that stacks the equation-by-equation OLS estimators, $\hat{\beta}_{OLS} = \left(\hat{\beta}'_{1,OLS}, ..., \hat{\beta}'_{N,OLS}\right)'$, where $\hat{\beta}_{i,OLS} = \left(\sum_{t=1}^{T} x_{it}x'_{it}\right)^{-1} \sum_{t=1}^{T} x_{it}y_{it}$.

2. Generalized least squares (GLS) and feasible GLS (FGLS) estimator:

When the system covariance matrix Σ is known, the GLS estimator of β is

$$\hat{\beta}_{GLS} = \left(\sum_{t=1}^{T} \tilde{X}_t \Sigma^{-1} \tilde{X}_t'\right)^{-1} \sum_{t=1}^{T} \tilde{X}_t \Sigma^{-1} Y_t.$$

When the covariance matrix Σ is unknown, a feasible GLS (FGLS) estimator is defined by replacing the unknown Σ with a consistent estimate. A widely used estimator of Σ is

$$\hat{\Sigma} = (\hat{\sigma}_{ij}),$$

where $\hat{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T} \hat{e}_{it}\hat{e}_{jt}$ and $\hat{e}_{kt}$ is the OLS residuals of the k^{th} equation, that is, $\hat{e}_{kt} = y_{kt} - \hat{\beta}_{k,OLS}' x_{kt}$, $k = i, j$. Then

$$\hat{\beta}_{FGLS} = \left(\sum_{t=1}^{T} \tilde{X}_t \hat{\Sigma}^{-1} \tilde{X}_t'\right)^{-1} \sum_{t=1}^{T} \tilde{X}_t \hat{\Sigma}^{-1} Y_t.$$

The FGLS estimator is a two-step estimator where OLS is used in the first step to obtain residuals $\hat{e}_{kt}$ and an estimator of Σ. The second step compute $\hat{\beta}_{FGLS}$ based on the estimated Σ in the first step. This estimator is sometimes referred to as the restricted estimator as opposed to the unrestricted estimator proposed by Zellner that uses the residuals from an OLS regression of (2) without imposing the coefficient restrictions (3), i.e. from regressing each regressand on all distinct regressors in the system.

3. Gaussian quasi-maximum likelihood estimator (QMLE):

The Gaussian log-likelihood function is

$$L(\beta, \Sigma) = const + \frac{T}{2}\det\Sigma - \frac{1}{2}\sum_{t=1}^{T}\left(Y_t - \tilde{X}_t'\beta\right)' \Sigma^{-1}\left(Y_t - \tilde{X}_t'\beta\right),$$

or equivalently,

$$L(\beta, \Sigma) = const + \frac{T}{2}\det\Sigma - \frac{1}{2}\sum_{t=1}^{T}\left(Y_t - A(\beta)' X_t\right)' \Sigma^{-1}\left(Y_t - A(\beta)' X_t\right),$$

where $A(\beta)$ denotes the coefficient A in (2) with the linear restriction of (3), and the QMLE $\left(\hat{\beta}_{QMLE}, \hat{\Sigma}_{QMLE}\right)$ maximizes $L(\beta, \Sigma)$. When the vector U_t has a normal distribution, this estimator is the maximum likelihood estimator.

4. Minimum distance (MD) estimator:

The idea of the MD estimator is to obtain an estimator of the unrestricted coefficient A in (2), $\hat{A}$, and then, obtain an estimator of β by minimizing the distance between $\hat{A}$ and β in (3). For this, assume that $T > L$ and that the whole regressor matrix X has full rank L. When $\hat{A}$ is the OLS estimator of $A(\beta)$, that is, $\hat{A} = \left(\sum_{t=1}^{T} X_t X_t'\right)^{-1} \sum_{t=1}^{T} X_t Y_t'$, the optimal MD estimator $\hat{\beta}_{MD}$ minimizes the optimal MD objective function

$$Q_{MD}(\beta) = \left(vec\left(\hat{A}\right) - G\beta\right)' \left(\hat{\Sigma}^{-1} \otimes \sum_{t=1}^{T} X_t X_t'\right)\left(vec\left(\hat{A}\right) - G\beta\right).$$

In this case, we have

$$\hat{\beta}_{MD} = \left(G'\left(\hat{\Sigma}^{-1} \otimes \sum_{t=1}^{T} X_t X_t'\right) G\right)^{-1} \left(G'\left(\hat{\Sigma}^{-1} \otimes \sum_{t=1}^{T} X_t X_t'\right) vec\left(\hat{A}\right)\right).$$

Relationship among the estimators:

Some of the above estimators are tightly linked. For example, *if we use the same consistent estimator* $\hat{\Sigma}$, the FGLS and the MD estimators above are identical, that is, $\hat{\beta}_{FGLS} = \hat{\beta}_{MD}$. Also, if we use the QMLE estimator of Σ, $\hat{\Sigma}_{QMLE}$ in place of $\hat{\Sigma}$, $\hat{\beta}_{QMLE}$ is identical to $\hat{\beta}_{FGLS}$ $\left(\text{and to } \hat{\beta}_{MD}\right)$. By the Gauss-Markov theorem, the GLS estimator $\hat{\beta}_{GLS}$ is more efficient than the OLS estimator $\hat{\beta}_{OLS}$ when the system errors are correlated across equations. However, this efficiency gain disappears in some special cases described in Kruskal's theorem (Kruskal, 1968). A well-known special case of this theorem is when the regressors in each equation are the same. For other cases, readers can refer Chapter 14 of Greene (2003) and Davidson and MacKinnon (1993, pp. 294-295). The efficiency gain relative to OLS tends to be larger when the correlation across equations is larger and when the correlation among regressors in different equations is smaller.

Note also that efficient estimators propagate misspecification and inconsistencies across equations. For example, if any equation is misspecified (for example some relevant variable has been omitted), then the entire vector β will be inconsistently estimated by the efficient methods. In this sense, equation-by-equation OLS provides some degree of robustness since it is not affected by misspecification in other equations in the system.

Distribution of the estimators:

In the literature on the classical linear SUR, the FGLS estimator $\hat{\beta}_{FGLS}$ is often called the SUR estimator (SURE). The usual asymptotic analysis of the SURE is carried out when the dimension of index t, T, increases to infinity with the dimension of index i, N, kept fixed. For asymptotic theories for large N, T, one can refer to Phillips and Moon (1999**.** Under regularity conditions, the asymptotic distributions as $T \to \infty$ of the aforementioned estimators are:

$$\sqrt{T}\left(\hat{\beta}_{OLS} - \beta\right) \Rightarrow N\left(0, \left[E\left(\tilde{X}_t \tilde{X}_t'\right)\right]^{-1} E\left(\tilde{X}_t \Sigma \tilde{X}_t'\right) \left[E\left(\tilde{X}_t \tilde{X}_t'\right)\right]^{-1}\right)$$

and

$$\begin{aligned} \sqrt{T}\left(\hat{\beta}_{GLS} - \beta\right), \sqrt{T}\left(\hat{\beta}_{FGLS} - \beta\right), \sqrt{T}\left(\hat{\beta}_{MD} - \beta\right) \quad &\Rightarrow \quad N\left(0, \left[E\left(\tilde{X}_t \Sigma^{-1} \tilde{X}_t'\right)\right]^{-1}\right) \\ &\equiv \quad N\left(\left(G'\left(\Sigma^{-1} \otimes E\left(X_t X_t'\right)\right) G\right)^{-1}\right). \end{aligned}$$

It is straightforward to show that the SUR estimator using the information in the system is more efficient (has a smaller variance) than the estimator of the individual equations. Using the above distributional results, it is straightforward to construct statistics to test general nonlinear hypotheses.

Finite sample properties of SURE have been studied extensively either analytically in some restrictive cases (e.g. Zellner, 1963, 1972, Kakwani, 1967), by asymptotic expansions (e.g. Phillips, 1977 and Srivastava and Maekawa, 1995) or by simulation (e.g Kmenta and Gilbert, 1968). Most work has focused on the two-equation case. The above approximations appear to be good descriptions of the finite-sample behavior of the estimators analyzed when the number of observations, T, is large relative to the number of equations, N. In particular, efficient methods provide an efficiency gain in cases where the correlation among disturbances across equations is high and when correlation among regressors across equations is low. Non-normality of disturbances has also been found to deteriorate the quality of the above approximations. Bootstrap methods have also proposed to remedy these documented departures from normality and improve the size of tests.

Extensions

In this section we discuss several extensions of the classical linear SUR model where the assumption on the error terms is no longer satisfied.

Autocorrelation and heteroskedasticity:

As in standard univariate models, non-spherical disturbances can be accommodated by either modelling the residuals or computing robust covariance matrices. In addition to standard dynamic effects, serial correlation can arise in this environment due to the presence of individual effects (see Baltagi, 1980). One could define the equivalent of White (in the case of heteroskedasticity) or HAC (in the case of serial correlation) standard errors to conduct inference with the OLS estimator as in the single-equation framework

For efficiency in estimation, some parametric assumption on the disturbance process is often imposed (see Greene, 2003). For example, in the case of heteroskedasticity, Hodgson, Linton, and Vorkink (2002) propose an adaptive estimator that is efficient under the assumption that the errors follow an elliptical symmetric distribution that includes the normal as a special case. An intermediate approach is to use a restricted (or parametric) covariance matrix to try to capture some efficiency gains in estimation,and then use a nonparametric heteroskedasticity and autocorrelation (HAC) consistent estimator of the covariance matrix to do inference. This two-tier approach (dubbed quasi-FGLS) has been suggested by Creel and Farell (1996) .

Endogenous regressors:

When the regressor X_t in the SUR model is correlated with the error term U_t, one needs instrumental variables (IVs), say, $Z_t = [z'_{1t}, ..., z'_{Nt}]'$ to estimate β. We suppose that the IVs satisfy the usual rank condition. The GMM estimator (or the IV estimator), then, utilizes the moment condition

$$E\left[vec\left(Z_t U_t'\right)\right] = 0.$$

The optimal GMM estimator $\hat{\beta}_{GMM}$ is derived by minimizing the GMM objective function with the optimal choice of weighting matrix given by $\left(\hat{\Sigma}\otimes\left(\sum_{t=1}^{T}Z_tZ_t'\right)\right)^{-1}$,

$$Q_{GMM}(\beta)=\left[\sum_{t=1}^{T}vec\left\{Z_t\left(Y_t-A(\beta)'X_t\right)'\right\}\right]'\left(\hat{\Sigma}\otimes\sum_{t=1}^{T}Z_tZ_t'\right)^{-1}\left[\sum_{t=1}^{T}vec\left\{Z_t\left(Y_t-A(\beta)'X_t\right)'\right\}\right].$$

Then, we have

$$\begin{aligned}\hat{\beta}_{GMM} &= \left\{G'\left(\hat{\Sigma}\otimes\left(\sum_{t=1}^{T}X_tZ_t'\right)\left(\sum_{t=1}^{T}Z_tZ_t'\right)^{-1}\left(\sum_{t=1}^{T}Z_tX_t'\right)\right)^{-1}G\right\}^{-1} \\ &\quad\times G'\left\{\hat{\Sigma}^{-1}\otimes\left(\sum_{t=1}^{T}X_tZ_t'\right)\left(\sum_{t=1}^{T}Z_tZ_t'\right)^{-1}\left(\sum_{t=1}^{T}Z_tX_t'\right)\right\}vec\left(\hat{A}_{2SLS}\right),\end{aligned}$$

where $\hat{A}_{2SLS}=\left\{\left(\sum_{t=1}^{T}X_tZ_t'\right)\left(\sum_{t=1}^{T}Z_tZ_t'\right)^{-1}\left(\sum_{t=1}^{T}Z_tX_t'\right)\right\}^{-1}\left(\sum_{t=1}^{T}X_tZ_t'\right)\left(\sum_{t=1}^{T}Z_tZ_t'\right)^{-1}\left(\sum_{t=1}^{T}Z_tY_t'\right)$ is the two-stage least squares estimator of $A(\beta)$. When X_t is exogenous, so that $X_t=Z_t$, the GMM objective function $Q_{GMM}(\beta)$ and minimum distance objective function $Q_{MD}(\beta)$ are identical, and we conclude that $\hat{\beta}_{GMM}=\hat{\beta}_{MD}$.

Vector autoregressions:

When the index t in the SUR model denotes time and the regressors x_{it} include lagged dependent variables, the classical linear SUR model becomes a vector autoregression model (VAR) with exclusion restrictions. In this case, the regressors X are no longer strictly exogenous, and the assumption in the previous section is violated. A special case is when the order of the lagged dependent variables is one. In this case, for $\{y_{it}\}_t$ to be stationary, it is necessary that the absolute value of the coefficient of y_{it-1} is less than one. If the coefficient of y_{it-1} is one, $\{y_{it}\}_t$ is nonstationary. Nonstationary SUR VAR models have been used in developing tests for unit roots and cointegration in panels with cross-sectional dependence, see for example Chang (2004), Groen and Kleibergen (2003) and Larsson, Lyhagen, and Lothgren (2001).

Seemingly unrelated cointegration regressions:

When the non-constant regressors in X_t are integrated nonstationary variables but the errors in U_t are stationary, we call model (1) (or equivalently (2)) a seemingly unrelated cointegration regression model, see Park and Ogaki (1991), Moon (1999), Mark *et al* (2005), and Moon and Perron (2004). These papers showed that for efficient estimation of β, an estimator of the long-run variance of U_t, not of the spontaneous covariance Σ as in the previous section, should be used in FGLS. In addition some modification of the regression is necessary when the integrated regressors and the stationary errors are correlated. Empirical applications in the main references include tests for purchasing power parity, the relation between national saving and investment, and tests of the forward rate unbiasedness hypothesis.

Nonlinear SUR (NSUR):

An NSUR model assumes that the conditional mean of y_{it} given x_{it} is nonlinear, say $h_i(\beta, x_{it})$, that is, $y_{it} = h_i(\beta, x_{it}) + u_{it}$.Defining $H(\beta, X_t) = (h_1(\beta, x_{1t}), ..., h_N(\beta, x_{Nt}))$, we write the NSUR model in a multivariate nonlinear regression form,

$$Y_t = H(\beta, X_t) + U_t.$$

In this case, we may estimate β using (quasi) MLE assuming that Y_t are Gaussian conditioned on X_t or GMM utilizing the moment condition that $E[g(X_t) U_t'] = 0$ for any measurable transformation g of X_t.

References

[1] Baltagi, B. (1980): On Seemingly Unrelated Regressions with Error Components, *Econometrica*, 48, 1547-1552.

[2] Chang, Y. (2004): Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependency, *Journal of Econometrics*, 120, 263-293.

[3] Creel, M. and M. Farell (1996): SUR Estimation of Multiple Time-Series Models with Heteroskedasticity and Serial Correlation of Unknown Form, *Economic Letters*, 53, 239-245.

[4] Davidson, R. and J. MacKinnon (1993): *Estimation and Inference in Econometrics*, Oxford University Press.

[5] Fiebig, D. G. (2001): Seemingly Unrelated Regression, in Baltagi, B. eds, *A Companion to Theoretical Econometrics*, Backwell Publishers, 0101-121.

[6] Greene, W. (2003): *Econometric Analysis,* 5th ed. New Jersey, Prentice Hall.

[7] Groen, J. and F. Kleibergen (2003): Likelihood-Based Cointegration Analysis in Panels of Vector Error Correction Models. *Journal of Business and Economic Statistics*, *21, 295-318.*

[8] Hodgson, D., O. Linton, and K. Vorkink (2002): Testing the Capital Asset Pricing Model Efficiently under Elliptical Symmetry: A Semiparametric Approach, *Journal of Applied Econometrics*, 17, 619-639.

[9] Larsson, R., J. Lyhagen, and M. Lothgren (2001): Likelihood-based Cointegration Tests in Heterogeneous Panels, *Econometrics Journal*, 4, 109-142.

[10] Kakwani, N.C. (1967): The Unbiasedness of Zellner's Seemingly Unrelated Regression Equations Estimator, *Journal of the American Statistical Association*, 62, 141-142.

[11] Kamenta, J. and R.F. Gilbert (1968): Small Sample Properties of Alternative Estimators for Seemingly Unrelated Regressions, *Journal of the American Statistical Association*, 63, 1180-1200.

[12] Kruskal, W. (1968): When are Gauss-Markov and Least Squares Estimators Identical?, *Annals of Mathematical Statistics*, 39, 70-75.

[13] Mark, N., M. Ogaki, and D. Sul (2005): Dynamic Seemingly Unrelated Cointegrating Regressions, *Review of Economic Studies*, 72, 797-820.

[14] Moon, H.R. (1999): A Note on Fully-Modified Estimation of Seemingly Unrelated Regressions Models with Integrated Regressors, *Economics Letters*, 65, 25–31.

[15] Moon, H.R. and P. Perron (2004): Efficient Estimation of SUR Cointegration Regression Model and Testing for Purchasing Power Parity, *Econometric Reviews,* 23, 293-323.

[16] Park, J. and M. Ogaki (1991): Seemingly Unrelated Canonical Cointegrating Regressions, University of Rochester working paper 280.

[17] Phillips, P.C.B. (1977): Finite sample distribution of Zellner's SURE, *Journal of Econometrics,* 6, 147-164.

[18] Srivastava, V. K. and D. E. A. Giles (1987): *Seemingly Unrelated Regression Equations Models*, New York: Marcel Dekker Inc..

[19] Srivastava, V. K. and K. Maekawa (1995): Efficiency Properties of Feasible Generalized Least Squares Estimators in SURE Models under Non-normal Disturbances, *Journal of Econometrics*, 66, 99-121.

[20] Zellner A. (1962): An Efficient Method of Estimating Seemingly Unrelated Regression Equations and Tests of Aggregation Bias, *Journal of the American Statistical Association*, 57, 500-509.

[21] Zellner A. (1963): Estimators for Seemingly Unrelated Regression Equations: Some Finite Sample Results, *Journal of the American Statistical Association*, 58, 977-992

[22] Zellner A. (1972): Corrigenda, *Journal of the American Statistical Association*, 67, 255