



TERM PAPER

On

**SEEMINGLY UNRELATED
REGRESSION**

By

Anomita Ghosh

Nitin Kumar Sinha

Sudipta Ghosh

Udayan Rathore

ECONOMETRICS II

Instructor

Mr. Subrata Sarkar

INTRODUCTION

The SUR system proposed by Arnold Zellner, comprises several individual relationships that are linked by the fact that their disturbances or the error terms are correlated. Such models have found wide applications. For example, demand functions can be estimated for different households for a given commodity. The correlation among the equation disturbances can come from many sources like correlated shocks to household income. Also, one can model the demand of a household for different commodities, but adding up constraints leads to restrictions on the parameters of the different equations. Equations explaining some phenomenon in different cities, states, countries, firms or industries provide a natural application as these various entities are likely to be subjected to spillovers from economywide or worldwide shocks. One can also consider the regression of fertility across all states of India, s , period t on literacy, income and women labor force participation in which the error terms of different states in the same period may be correlated because the same policy may hit different states. Then the seemingly unrelated regressions actually become correlated.

MOTIVATION

There are two main motivations for using 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.(Roger Moon,et al (2006))

The popularity of SUR is related to its applicability to a large class of modeling and testing problems and also the relative ease of estimation. The increased availability of data representing a sample of cross sectional units observed over several time periods provides researchers with a rich source of information.

THE BASIC LINEAR SUR MODEL

_Suppose for the ith individual we have the following relationship:

$$Y_{it} = B_{i1} X_{it1} + B_{i2} X_{it2} + \dots + B_{iki} X_{itki} + \epsilon_{it} \quad i=1(1)N, \\ t=1(1)T$$

- The B's are allowed to vary across the individuals but they are constant over time.
- Each individual can have different no. of explanatory variables .
(individual I has k_i variables)
- However each individual has same no. of observations.

For the ith individual, the above relationship written in matrix notation is as follows:

$$Y_i = X_i B_i + \epsilon_i \text{ for all } i.$$

Where the dimension of Y_i is $T \times 1$,

the dimension of X_i is $T \times k_i$,

the dimension of B_i is $k_i \times 1$,

the dimension of ϵ_i is $T \times 1$.

Now stacking all the N individuals, we have

- Matrix representation of the model

$$y_1 = X_1 \beta_1 + \epsilon_1$$

$$y_2 = X_2 \beta_2 + \epsilon_2$$

⋮

$$y_n = X_n \beta_n + \epsilon_n$$

$$\rightarrow \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{nT \times 1} = \begin{pmatrix} X_1 & 0_{T \times k_2} & \dots & 0_{T \times k_n} \\ 0_{T \times k_1} & X_2 & \dots & 0_{T \times k_n} \\ \vdots & \vdots & \dots & \vdots \\ 0_{T \times k_1} & 0_{T \times k_2} & \dots & X_n \end{pmatrix}_{nT \times k_*} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}_{k_* \times 1} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}_{nT \times 1},$$

where $k_* = \sum_i k_i$.

$$\rightarrow y_* = X_* \beta_* + \epsilon_*$$

- Each individual regression appears to be an usual regression problem, OLS or GLS.
- Each individual has his own β_i 's but the equations may be related through the error term which captures the effect of other factors.

Before going to estimation, we will consider the assumptions to be made on the explanatory variables (X 's), the error terms (ϵ 's) and the relationship between X and ϵ . Here we assume that X and ϵ are uncorrelated, similar to CLRM.

ASSUMPTIONS ON THE EXPLANATORY VARIABLES

- Rank of $X_i = k_i < T$.
- X_i is of full column rank. X 's are fixed, nonstochastic and without any measurement error.
- $r(X_i' X_i) = k_i$. $X_i' X_i$ has full rank k_i and so is invertible.
- $X' X$ has full rank K and hence is also invertible.

ASSUMPTIONS REGARDING THE ERRORS

Within individuals:

$$E(\epsilon_{it}) = 0 \text{ for all } i, t.$$

$$E(\epsilon_{it} \epsilon_{is}) = \sigma_{ij} \text{ if } t=s \\ = 0 \text{ if } t \neq s$$

- There is homoscedasticity within individuals (σ not indexed by t) but heteroscedasticity across individuals (σ indexed by i).
- No direct autocorrelation.

Across individuals

$$E(\epsilon_{it} \epsilon_{js}) = \sigma_{ij} \quad \text{if } t=s, i \neq j$$

$$= 0 \quad \text{if } t \neq s, i \neq j$$

- There exists contemporaneous correlation.
- But there is no cross autocorrelation.

$$E(\epsilon_i \epsilon_j') = E \left(\begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{iT} \end{pmatrix} \begin{pmatrix} \epsilon_{j1} & \epsilon_{j2} & \dots & \epsilon_{jT} \end{pmatrix} \right) = E \begin{pmatrix} \epsilon_{i1} \epsilon_{j1} & \epsilon_{i1} \epsilon_{j2} & \dots & \epsilon_{i1} \epsilon_{jT} \\ \epsilon_{i2} \epsilon_{j1} & \epsilon_{i2} \epsilon_{j2} & \dots & \epsilon_{i2} \epsilon_{jT} \\ \vdots & \vdots & & \vdots \\ \epsilon_{iT} \epsilon_{j1} & \epsilon_{iT} \epsilon_{j2} & \dots & \epsilon_{iT} \epsilon_{jT} \end{pmatrix}$$

$$= \begin{bmatrix} \sigma_{ij} & 0 & \dots & 0 \\ 0 & \sigma_{ij} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sigma_{ij} \end{bmatrix} = \sigma_{ij} I_T$$

The technique of estimation that we apply in SUR depends on the structure of the variance covariance matrix.

VARIANCE COVARIANCE MATRIX OF ϵ

$$\begin{aligned}
E(\varepsilon_* \varepsilon_*') &= E \begin{pmatrix} \varepsilon_1 \varepsilon_1' & \varepsilon_1 \varepsilon_2' & \dots & \varepsilon_1 \varepsilon_n' \\ \varepsilon_2 \varepsilon_1' & \varepsilon_2 \varepsilon_2' & \dots & \varepsilon_2 \varepsilon_n' \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_n \varepsilon_1' & \varepsilon_n \varepsilon_2' & \dots & \varepsilon_n \varepsilon_n' \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{11} I_T & \sigma_{12} I_T & \dots & \sigma_{1n} I_T \\ \sigma_{21} I_T & \sigma_{22} I_T & \dots & \sigma_{2n} I_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} I_T & \sigma_{n2} I_T & \dots & \sigma_{nn} I_T \end{pmatrix} = \Sigma \otimes I_T \equiv \Omega
\end{aligned}$$

There are several cases of estimation.

CASE1- Simple case

$\sigma_{ii} = \sigma^2$ for all i .

$\sigma_{ij} = 0$ for all $i \neq j$.

Hence We Have $\Sigma \otimes I_T = \sigma^2 I_{NT}$

- Therefore OLS is efficient.

On running the aforesaid model in SAS the OLS results thus obtained were:

Also $\hat{\beta}_{OLS}$ on the full model \leftrightarrow $\hat{\beta}_{OLS}$ on each individual.

$$\begin{aligned}
\hat{\beta}_* &= \begin{pmatrix} \hat{\beta}_1^* \\ \hat{\beta}_2^* \\ \vdots \\ \hat{\beta}_n^* \end{pmatrix} = (X_*' X_*)^{-1} X_*' y_* \\
&= \begin{pmatrix} (X_1' X_1)^{-1} & 0 & \dots & 0 \\ 0 & (X_2' X_2)^{-1} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & (X_n' X_n)^{-1} \end{pmatrix} \begin{pmatrix} X_1' y_1 \\ X_2' y_2 \\ \vdots \\ X_n' y_n \end{pmatrix} \\
&= \begin{pmatrix} (X_1' X_1)^{-1} X_1' y_1 \\ (X_2' X_2)^{-1} X_2' y_2 \\ \vdots \\ (X_n' X_n)^{-1} X_n' y_n \end{pmatrix} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{pmatrix}
\end{aligned}$$

- Pooling does not help. Each individual coefficient can be estimated separately.

The conclusion remains the same even if σ_{ii} 's are different for different i's, provided $\sigma_{ij} = 0$ for all $i, j \quad i \neq j$.

- When the variance -covariance matrix is of the form $\Sigma \otimes I_T \neq \sigma^2 I_{NT}$ then GLS should have been applied. However for each individual, homoskedasticity holds hence OLS can be applied to each individual.
- The individuals are themselves not connected. Hence OLS is efficient in this case also.

Here also $\hat{\beta}_{OLS}$ on the full model \leftrightarrow $\hat{\beta}_{OLS}$ on each individual.

Example:

Suppose we consider the following regression:

$$I_{ITCt} = B_0 + B_1 \text{profit}_{ITCt} + B_2 \text{sales}_{ITCt} + B_3 \text{corptax}_{ITCt} + \epsilon_{ITCt}$$

$$I_{TATAt} = \gamma_0 + \gamma_1 \text{profit}_{TATAt} + \gamma_2 B_2 \text{sales}_{TATAt} + \gamma_3 \text{corptax}_{TATAt} + \epsilon_{TATAt}$$

Here we consider cross sectional unit i as a group and time series unit t as firm in that group. The two groups considered are ITC and TATA. They dominate the market in completely different segments and thus there is a natural justification for assuming $\sigma_{ij} = 0$.

CASE 2- General case (no restriction on Σ)

If $\Sigma \neq \sigma^2 I_T$ then we apply GLS.

Let $\Omega^{-1} = (\Sigma \otimes I_T)^{-1} = \Sigma^{-1} \otimes I_T$. Then, the (infeasible) GLS estimator $\tilde{\beta}_* = (X_*' \Omega^{-1} X_*)^{-1} X_*' \Omega^{-1} y_*$ is unbiased, efficient, consistent and asymptotically efficient.

- All the coefficients of all the individuals are estimated together.
- It does not boil down to individual GLS estimation.
- The GLS estimator defined above is not feasible in the sense that it depends on the unknown variance- covariance matrix Σ .

EXAMPLE:

Suppose we consider the following regression model.

$$\text{Hrearn} = B_0 + B_1 \text{educ} + B_2 \text{exper} + B_3 \text{exper}^2 + B_4 \text{tenure} + B_5 \text{tenure}^2 + B_6 \text{union} + B_7 \text{south} + B_8 \text{married} + B_9 \text{white} + B_{10} \text{male} + \epsilon_{it}$$

$$\text{Hrbenefits} = \gamma_0 + \gamma_1 \text{educ} + \gamma_2 \text{exper} + \gamma_3 \text{exper}^2 + \gamma_4 \text{tenure} + \gamma_5 \text{tenure}^2 + \gamma_6 \text{union} + \gamma_7 \text{south} + \gamma_8 \text{married} + \gamma_9 \text{white} + \gamma_{10} \text{male} + u_{it}$$

Where Hrearn is hourly earnings of the workers.

Hrbenefits is hourly benefits of the workers.

- The errors across the two equations appear to be positively correlated because the same unobservables such as ability that lead to higher earnings also leads to higher benefits.

CONSEQUENCES OF APPLYING OLS IN THE GENERAL MODEL

- $\hat{\beta}_{\text{OLS}}$ is still unbiased.
- But $\hat{\beta}_{\text{OLS}}$ is not efficient. $V(\hat{\beta}_{\text{OLS}}) - V(\hat{\beta}_{\text{GLS}})$ is a nonnegative definite matrix.

There are two important cases where OLS and GLS will give identical estimates.

- $\sigma_{ij}=0$ no matter what σ_{ii} is.
- $X_i = X$ or $X_i = A^i X$ with $\det A^i \neq 0$ (i.e. same no. of variables for each i, same explanatory variables for each I, same values of the variables for each i).

Examples

- Suppose we consider the following regression model.

$$\ln \text{EXS}_{it} = \alpha_0 + \alpha_1 \text{trend}_{it} + \epsilon_{it}$$

Where $\ln \text{EXS}_{it}$ is the expenditure on services of country i at time t. Trend variable is the same for everybody so we can estimate the above equation by OLS.

- Suppose we consider the CAPM equation.

The rate of return

$$R_{it} - r_{ft} = \alpha_i + B_i (r_{mt} - r_{ft}) + u_{it}$$

Where R_{it} is the rate of return on security i.

r_{ft} is the riskfree rate of return on the govt. bond.

r_{mt} is the market rate of return.

B_i is a measure of the volatility of security.

Here the regressor is the same for all the securities. Hence we can estimate the above equation by OLS.

- $\hat{\beta}_{OLS}$ is consistent (plim $\hat{\beta}_{OLS} = \beta_i$) By using this property of $\hat{\beta}_{OLS}$

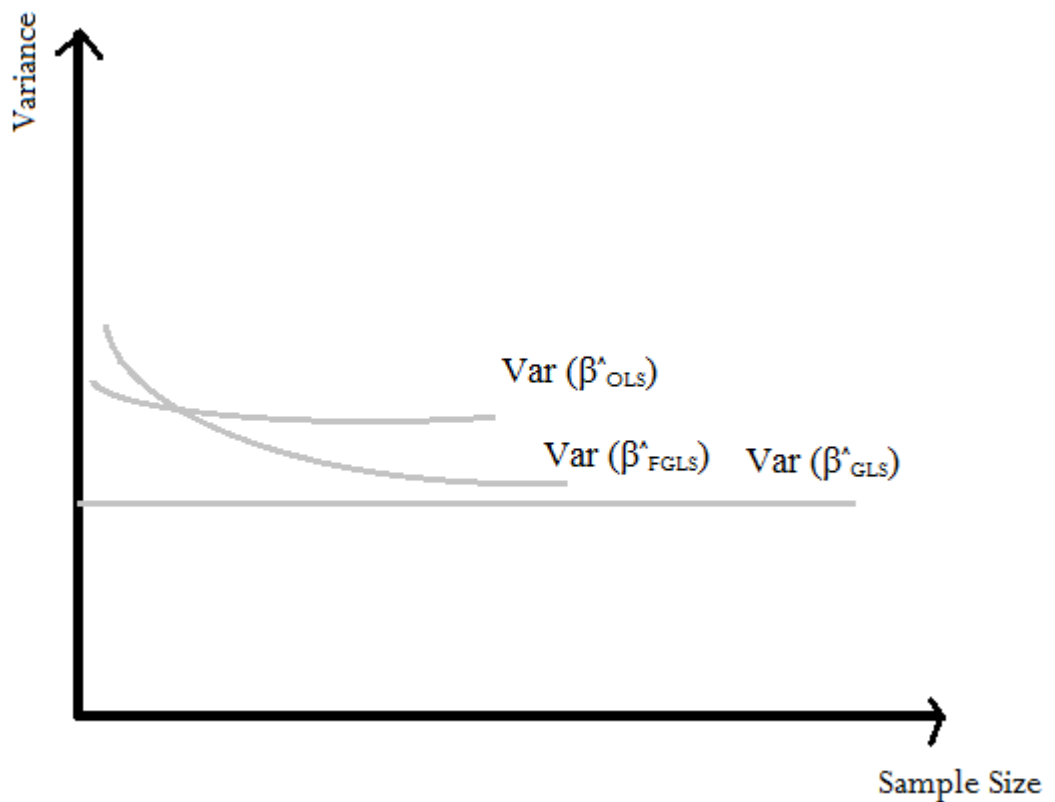
and the fact that $\hat{\beta}_{OLS} = \hat{\beta}_{i, OLS}$, we can get consistent estimates of σ_{ij} ($\hat{\sigma}_{ij} = (\hat{\epsilon}_i' \hat{\epsilon}_j) / T$) and hence $\hat{\Sigma} = [\hat{\sigma}_{ij}]$ and thereby can use feasible GLS (FGLS) which is a GLS estimator obtained by replacing Σ by its consistent estimate.

$$\hat{\beta}_{FGLS} = [X_*' (\hat{\Sigma}^{-1} \otimes I_T) X_*]^{-1} X_*' (\hat{\Sigma}^{-1} \otimes I_T) y_*;$$

- FGLS estimators cannot be said to be unbiased. But they are consistent and asymptotically equivalent to its infeasible counterpart (i.e.FGLS has same asymptotic variance covariance matrix as GLS.) The results of running FGLS on our model gives the following results:

Now we wanted to compare the relative efficiency of $\hat{\beta}_{OLS}$,

$$\hat{\beta}_{GLS} \text{ and } \hat{\beta}_{FGLS}.$$



ITERATED FGLS:

An alternative FGLS estimator is the iterated FGLS estimator. The IFGLS estimator used most often is Zellner's iterated SUR (ISUR) estimator. The steps involved in using the ISUR estimator are as follows.

1. Estimate the parameters of the big equation using Zellner's SUR estimator described above.
2. Use the parameter estimates from this regression to compute the residuals for each of the N-equations.

3. Use the residuals to obtain new estimates of the variances and covariances of the disturbances for the N-equations, and therefore a new estimate of β and Σ .
4. Use the new estimate of Σ to repeat step 1 and obtain new parameter estimates.
5. Repeat steps 2, 3, and 4. (Each time you obtain new parameter estimates this completes an iteration).
6. Continue to iterate until convergence is achieved. Convergence is achieved when the change in the parameter estimates is very small. Very small is defined by a predetermined criterion. This last set of parameter estimates are the ISUR estimates.

Properties of the ISUR Estimator :

The ISUR estimator has the same asymptotic properties as the SUR estimator. However, there is an ongoing debate about whether the ISUR or SUR estimator yields better estimates in small samples. Most econometricians seem to prefer the ISUR estimator. One reason for this is given below.

SUR WITH FIRST ORDER AUTOREGRESSIVE PROCESS:

Till now we have assumed that error terms are random, but it is not the case always. Errors may show a pattern. When the error term is related to the previous error term, it can be written in an algebraic equation.

$\epsilon_t = \rho \epsilon_{(t-1)} + v_t$, where ρ is the autocorrelation coefficient between the

two disturbance terms, and v_t is the disturbance term for the autocorrelation. This is known as an Autoregressive Process. The v_t is needed within the equation because although the error term is less random, it still has a slight random effect.

We will assume error ϵ_{it} is generated by a SIMPLE (ϵ_{it} depends only on its own past error), STATIONARY ($|\rho_i| < 1$) and FIRST ORDER (ϵ_{it} depends only on the $\epsilon_{i(t-1)}$) .i.e $\epsilon_{it} = \rho_i \epsilon_{i(t-1)} + v_{it}$

The assumptions made on error terms earlier will apply to v_{it} .

$$E(v_{it} v_{is}) = \sigma_{ii} ; t=s$$

$$= 0 ; t \neq s$$

$$E(v_{it} v_{js}) = \sigma_{ij} ; t=s$$

$$= 0 ; t \neq s$$

. Now knowing the error structure of v_{it} and the relation b/w v_{it} and ϵ_{it} we want to derive the error structure of ϵ_{it}

$$\epsilon_{it} = \rho_i \epsilon_{i(t-1)} + v_{it} \dots\dots\dots (A)$$

$$\epsilon_{it} = v_{it} + \rho_i \epsilon_{i(t-1)}$$

$$= v_{it} + \rho_i [v_{i(t-1)} + \rho_i \epsilon_{i(t-2)}]$$

Continuing in this way we will obtain

$$\epsilon_{it} = v_{it} + \rho_i v_{i(t-1)} + \rho_i^2 v_{i(t-2)} + \dots\dots\dots$$

$$+ \rho_i^s v_{i(t-s)} + \rho_i^{s+1} v_{i(t-s-1)} + \dots\dots\dots$$

Now calculating $E(\epsilon_{it} \epsilon_{j(t-s)})$, we will get

$$E(\epsilon_{it} \epsilon_{j(t-s)}) = \rho_i^s \sigma_{ij} (1/1-\rho_i\rho_j)$$

For the whole model we will get

$$E(\epsilon_i \epsilon_j') =$$

$$\begin{bmatrix} \epsilon_{i1}\epsilon_{j1} & \epsilon_{i1}\epsilon_{j2} & \dots & \epsilon_{i1}\epsilon_{jT} \\ \epsilon_{i2}\epsilon_{j1} & \epsilon_{i2}\epsilon_{j2} & \dots & \epsilon_{i2}\epsilon_{jT} \\ \vdots & \vdots & \vdots & \vdots \\ \epsilon_{iT}\epsilon_{j1} & \epsilon_{iT}\epsilon_{j2} & \dots & \epsilon_{iT}\epsilon_{jT} \end{bmatrix}$$

Now substituting for each entries in above matrix by using

$$E(\epsilon_{it} \epsilon_{j(t-s)}) = \rho_i^s \sigma_{ij} (1/1-\rho_i\rho_j)$$

We will obtain our var- cov matrix U_{ij} (say)

$$\text{Therefore } U_{ij} = \sigma_{ij} (1/1-\rho_i\rho_j)^* B,$$

4. Run regression on equation A and estimate V_{it} , V_{jt} .
5. Using the estimates of V_{it} , V_{jt} calculate $\hat{\sigma}_{ij}$
6. Then we will obtain the first value of $\hat{\beta}_{fgls}$. Continue this process. This is called Iterated FGLS which will asymptotically tends to β_{gls} .

Example- In the regression of Hrearn and Hrbenefits on educ, exper, $exper^2$, tenure, $tenure^2$, union, south, married, white, male we have omitted the variable ability which will be captured by the error term. However ability cannot be developed all at once, it sometimes depends on how experienced one is and experience is accumulated over the years. (Arrow's learning by doing phenomenon). Thus the error term may not be completely random, it may show a pattern.

SUR WITH SINGULAR VARIANCE MATRICES

Generally we assume that the variance matrix of u_i i.e Ω is not singular. However in consumer and producer theory applications this assumption is not always true in the original structural equations because of additivity constraints

Suppose that for a given year, each firm in a particular industry uses 3 inputs, capital (K), labor (L) and materials (M). Because of regional variation and differential tax concessions, firms across U.K. may face different prices for these inputs. Let p_{iK} denote the price of capital to firm i , p_{iL} denote the price of labor to firm i , p_{iM} denote the price of materials

to firm i . For each firm i , let s_{iK} be the cost share of capital, s_{iL} be the cost share of labor, s_{iM} be the cost share of materials. By definition, $s_{iK} + s_{iL} + s_{iM} = 1$.

Suppose we consider the following set of cost share equations.

$$s_{iK} = B_{10} + B_{11} \log p_{iK} + B_{12} \log p_{iL} + B_{13} \log p_{iM} + u_{iK} \dots 1)$$

$$s_{iL} = B_{20} + B_{12} \log p_{iK} + B_{22} \log p_{iL} + B_{23} \log p_{iM} + u_{iL} \dots 2)$$

$$s_{iM} = B_{30} + B_{13} \log p_{iK} + B_{23} \log p_{iL} + B_{33} \log p_{iM} + u_{iM} \dots 3)$$

where the symmetry restrictions from production theory have been imposed. The random errors u_{ig} , $g=K,L,M$ affect production but are unobservable to the modeler. For a SUR analysis, we would assume that $E(u_i/p_i) = 0$, where $u_i = (u_{iK}, u_{iL}, u_{iM})'$

and $p_i = (p_{iK}, p_{iL}, p_{iM})'$.

Because the cost shares must sum to 1 for each i , we must have $B_{10} + B_{20} + B_{30} = 1$

$$B_{11} + B_{12} + B_{13} = 0$$

$$B_{12} + B_{22} + B_{23} = 0$$

$$B_{13} + B_{23} + B_{33} = 0.$$

$$u_{iK} + u_{iL} + u_{iM} = 0.$$

This last restriction implies that $\Omega = \text{var}(u_i)$ has rank 2. We can drop one of the equations say the equation for materials and analyse the equations

for labor and capital. We can express the restrictions on the B's in the first two equations as

$$B_{13} = -B_{11} - B_{12} \dots (4)$$

$$B_{23} = -B_{12} - B_{22} \dots (5)$$

Plugging equations (4) and (5) in (1) and (2) we get,

$$s_{iK} = B_{10} + B_{11} \log(p_{iK}/p_{iM}) + B_{12} \log(p_{iL}/p_{iM}) + u_{iK}$$

$$s_{iL} = B_{20} + B_{12} \log(p_{iK}/p_{iM}) + B_{22} \log(p_{iL}/p_{iM}) + u_{iL}$$

We now have a two equation system with variance matrix of full rank.

We redefine $u_i = (u_{iK}, u_{iL})'$ $y_i = (s_{iK}, s_{iL})'$.

If we take $B = (B_{10}, B_{11}, B_{12}, B_{20}, B_{22})$ then $X_i =$

$$\begin{bmatrix} 1 & \log(p_{iK}/p_{iM}) & \log(p_{iL}/p_{iM}) & 0 & 0 \\ 0 & 0 & \log(p_{iK}/p_{iM}) & 1 & \log(p_{iL}/p_{iM}) \end{bmatrix}$$

Suppose we assume $E(u_{ig}/p_{iK}, p_{iL}, p_{iM}) = 0$ $g=K,L$.

Under the above conditions, OLS and FGLS estimators are consistent.

However System OLS is not OLS equation by equation since B_{12} shows up in both the equations.

If $v(u_i/p_i)$ is constant then FGLS is asymptotically efficient else we have to use the robust variance-covariance matrix estimator for FGLS.

Our Model:

Here to illustrate the workings of a SUR estimation, we use a model in which we regress investment(I) made by three firms on their market capital (mktcap) (which is the product of number of outstanding shares and their prices) and net fixed assets(nfa)(which includes its net stock of physical capital like buildings and land, plants and machinery and transport communication equipments etc). The firms considered here are Ashok Leyland, Mahindra and Mahindra and Tata Motors. Since all these three firms belong to the automotive sector, any change in market conditions or policy regulations will affect all the companies in this sector. So we can expect the errors of all the three above equations to be contemporaneously correlated. Thus the model suits the framework of SUR estimation.

DATA:

The data for the above said regression analysis is obtained from (prowess 4.12). the analysis involves the usage of data on number of outstanding shares and their closing NSE prices from 1995 to 2012 for all the three companies said above. Along with it data on investment, net

fixed assets and current assets for the same period were collected from moneycontrol.com.

The data set is given below:

ASHOK LEYLAND:

Year	Investment(Rs million)	MktCap(Rs million)	netfixedassets(Rs million)
1996	781.4	18475.7	5785.5
1997	583.1	14806.7	6918.8
1998	484.8	10150.6	8072.7
1999	624.7	4816.64	8048.4
2000	1203.8	5060.45	8935.3
2001	1179.7	8212.08	9150.5
2002	1173.3	5637.25	8915.4
2003	1575.6	9555.98	9561
2004	1465.9	11482.6	9025.4
2005	2291.9	30101	8748.3
2006	3681.8	25034.6	8938.5
2007	2211	49229.9	9432.7
2008	6099	50836.6	13070.3
2009	2635.6	47094	15255.4
2010	3261.5	24145.6	33991.3
2011	12300	74232.2	42495.7
2012	15844.9	75629.7	46337.9

MAHINDRA & MAHINDRA:

year	investment (Rs million)	market capital (Rs million)	net fixed assets (Rs million)
1996	4144.4	16309.5	3277.6
1997	6093.4	26110.1	3893.4
1998	6560.3	31963.2	7130.7
1999	8104.7	28184	9149.6
2000	8229.8	23361.7	10160.5
2001	7099.9	34515.4	10691.8
2002	7472.6	13247.1	12097
2003	8564.8	13230.8	11884.9
2004	10480.9	11519.7	14137.8
2005	11515.3	53897.6	13532
2006	16828.7	57644.7	13641.4
2007	22009.3	151057	13752.6
2008	40446.8	191488	15905.7
2009	55853.9	171294	18144.5
2010	61315.9	106970	25676
2011	89413.3	313136	27385.2
2012	103105	429973	30686.1

TATA MOTORS:

year	investment (Rs million)	market capital (Rs million)	net fixed assets (Rs million)
1996	6609	58662.7	11743.8
1997	8265.7	108860	15801.1
1998	8371.1	89201.2	20588
1999	10293.8	74421.8	25106.6
2000	12036.6	43122.7	26952.1
2001	14050.3	34873.2	35922.6
2002	11899.2	16707.4	35474.7
2003	12718	40302.5	33377.8
2004	30814.4	49786.7	31759.2
2005	29120.6	171485	29392.7
2006	20151.5	149657	31567.1
2007	24770	356744	35689.7
2008	49102.7	280629	38732.6
2009	129681	240053	53801.2
2010	223369	81104.8	76391.1
2011	226242	383685	111988
2012	204936	671952	134108

Emperical results:

CASE I :

The simple case: $\sigma_{ii} = \sigma^2$ for all i ,

$\sigma_{ij} = 0$ for all i, j such that $i \neq j$.

This implies that $\Sigma \otimes I_T = \sigma^2 \mathbf{I}_{NT}$

Here we can run simple OLS because the structure is homoscedastic and OLS estimation becomes efficient.

The commands for the above said regression are:

```
proc syslin data=sasuser.firms;
aleyand: model i1 =mtcap1 nfa1 asset1;
mahindra: model i2 =mtcap2 nfa2 asset2;
tata: model i3 =mtcap3 nfa3 asset3;
run;
```

RESULTS OBTAINED:

The SYSLIN Procedure
Ordinary Least Squares Estimation

Model	ALEYAND
Dependent Variable	i1
Label	i1

Parameter Estimates

Variable	Parameter DF	Standard Estimate	Error	t Value	Pr > t	Variable Label
Intercept	1	-1650.70	617.2134	-2.67	0.0181	Intercept
mtcap1	1	0.083894	0.024884	3.37	0.0046	mtcap1
nfa1	1	0.183989	0.045824	4.02	0.0013	nfa1

Model MAHINDRA
 Dependent Variable i2
 Label i2

Parameter Estimates

Variable	Parameter DF	Standard Estimate	Error	t Value	Pr > t	Variable Label
Intercept	1	-4540.26	2466.456	-1.84	0.0869	Intercept
mtcap2	1	0.122712	0.026082	4.70	0.0003	mtcap2
nfa1	1	1.341665	0.243745	5.50	<.0001	nfa1

Model TATA
 Dependent Variable i3
 Label i3

Parameter Estimates

Variable	Parameter DF	Standard Estimate	Error	t Value	Pr > t	Variable Label
Intercept	1	-38640.5	13368.10	-2.89	0.0119	Intercept
mtcap3	1	-0.11165	0.075409	-1.48	0.1609	mtcap3
nfa3	1	2.669274	0.391279	6.82	<.0001	nfa3

CASE 2:

Here σ_{ii} 's are different for different firms but σ_{ij} 's are equal to zero, i.e

$$\sigma_{ii} = \sigma_{ii} \text{ for all } i=1,2,\dots,N$$

but $\sigma_{ij} = 0$ for all $i=j, i,j= 1,2,\dots,N$

Here $\sum \otimes I_T = \sigma^2 I_T$

The results on running SUR estimation i.e FGLS with the restriction that the variance-covariance matrix is a diagonal matrix gives the following result:

The commands for the above said regression are:

```
proc syslin data=sasuser.firms sdiag sur;
aleyand: model i1 =mtcap1 nfa1 asset1;
mahindra: model i2 =mtcap2 nfa2 asset2;
tata: model i3 =mtcap3 nfa3 asset3;
run;
```

RESULTS OBTAINED:

Seemingly Unrelated Regression Estimation

Cross Model Covariance

	ALEYLAND	MAHINDRA	TATA
ALEYLAND	2457960	0	0
MAHINDRA	0	40461426	0
TATA	0	0	1.0611E9

The SYSLIN Procedure
Seemingly Unrelated Regression Estimation

Model ALEYAND
Dependent Variable i1
Label i1

Parameter Estimates

Variable	Parameter DF	Standard Estimate	Error	t Value	Pr > t	Variable Label
Intercept	1	-1650.70	617.2134	-2.67	0.0181	Intercept
mtcap1	1	0.083894	0.024884	3.37	0.0046	mtcap1
nfa1	1	0.183989	0.045824	4.02	0.0013	nfa1

Model MAHINDRA
Dependent Variable i2
Label i2

Parameter Estimates

Variable	Parameter DF	Standard Estimate	Error	t Value	Pr > t	Variable Label
Intercept	1	-4540.26	2466.456	-1.84	0.0869	Intercept
mtcap2	1	0.122712	0.026082	4.70	0.0003	mtcap2
nfa1	1	1.341665	0.243745	5.50	<.0001	nfa1

Model TATA
 Dependent Variable i3
 Label i3

Parameter Estimates

Variable	Parameter DF	Standard Estimate	Error	t Value	Pr > t	Variable Label
Intercept	1	-38640.5	13368.10	-2.89	0.0119	Intercept
mtcap3	1	-0.11165	0.075409	-1.48	0.1609	mtcap3
nfa3	1	2.669274	0.391279	6.82	<.0001	nfa3

Thus we see that the parameter estimates obtained in the earlier case is equal to those obtained in this case. So our result mentioned earlier holds.

CASE 3:

This the general case where there are no restrictions on the Variance-covariance matrix i.e Σ . For this model, the assumption of unrestricted variance covariance matrix hold good as all the firms belong to the automotive sector and any policy change or change in market conditions in this sector affects the firms and makes their errors correlated.

Here we run FGLS without any restriction on the Σ matrix.

The commands for the above said regression are:

```
proc syslin data=sasuser.firms sur;  
aleyand: model i1 =mtcap1 nfa1 asset1;  
mahindra: model i2 =mtcap2 nfa2 asset2;  
tata: model i3 =mtcap3 nfa3 asset3;  
run;
```

RESULTS OBTAINED:

The SYSLIN Procedure
Seemingly Unrelated Regression Estimation

Cross Model Covariance

	ALEYLAND	MAHINDRA	TATA
ALEYLAND	2457960	-5454215	-4.092E7
MAHINDRA	-5454215	40461426	1.4209E8
TATA	-4.092E7	1.4209E8	1.0611E9

Model ALEYAND
Dependent Variable i1
Label i1

Parameter Estimates

Variable	Parameter DF	Standard Estimate	Error	t Value	Pr > t	Variable Label
Intercept	1	-1993.97	599.4260	-3.33	0.0050	Intercept
mtcap1	1	0.080975	0.019980	4.05	0.0012	mtcap1
nfa1	1	0.212450	0.039274	5.41	<.0001	nfa1

Model MAHINDRA
 Dependent Variable i2
 Label i2

Parameter Estimates

Variable	Parameter DF	Standard Estimate	Error	t Value	Pr > t	Variable Label
Intercept	1	-2828.49	2395.170	-1.18	0.2573	Intercept
mtcap2	1	0.144008	0.022897	6.29	<.0001	mtcap2
nfa1	1	1.085418	0.215314	5.04	0.0002	nfa1

Model TATA
 Dependent Variable i3
 Label i3

Parameter Estimates

Variable	Parameter DF	Standard Estimate	Error	t Value	Pr > t	Variable Label
Intercept	1	-32791.8	12980.03	-2.53	0.0242	Intercept
mtcap3	1	-0.03413	0.058701	-0.58	0.5702	mtcap3
nfa3	1	2.241070	0.329630	6.80	<.0001	nfa3

With the above assumptions still holding running iterated FGLS will give similar results if the sample size is large enough. Here our sample size is not large so results are varying from the results obtained from FGLS estimation.

The commands for the above said regression are:

```
proc syslin data=sasuser.firms itsur;
aleyand: model i1 =mtcap1 nfa1 asset1;
mahindra: model i2 =mtcap2 nfa2 asset2;
tata: model i3 =mtcap3 nfa3 asset3;
run;
```

RESULTS OBTAINED:

The SYSLIN Procedure
Iterative Seemingly Unrelated Regression Estimation

Parameter Estimates

Variable	Parameter			t Value	Pr > t	Variable
	DF	Estimate	Standard Error			Label
Intercept	0	-2970.09	353.6361	-8.40	<.0001	Intercept
mtcap1	0	0.066046	0.016982	3.89	0.0016	mtcap1
nfa1	0	0.305876	0.032895	9.30	<.0001	nfa1

Model MAHINDRA
Dependent Variable i2
Label i2

Parameter Estimates

Variable	Parameter			t Value	Pr > t	Variable
	DF	Estimate	Standard Error			Label
Intercept	0	3943.258	1251.598	3.15	0.0071	Intercept
mtcap2	0	0.179657	0.019728	9.11	<.0001	mtcap2
nfa2	0	0.392111	0.173201	2.26	0.0400	nfa1

Model TATA
Dependent Variable i3
Label i3

Parameter Estimates

Variable	Parameter			t Value	Pr > t	Variable
	DF	Estimate	Standard Error			Label
Intercept	0	0	.	.	.	Intercept
mtcap3	0	0.075222	0.045488	1.65	0.1190	mtcap3
nfa3	0	1.076851	0.241314	4.46	0.0005	nfa3

The model is not of full rank. Least Squares solutions for the parameters are not unique. Certain statistics will be misleading. A reported degree of freedom of 0 or B means the estimate is biased.

The following parameters have been set to zero. These variables are a linear combination of other variables as shown.

$$\text{Intercept} = +0.0365 * \text{ALEYLAND.Intercept} - 1.226\text{E-}7 * \text{ALEYLAND.mtcap1} - 9.47\text{E-}7 * \text{ALEYLAND.nfa1} - 0.1737 * \text{MAHINDRA.Intercept} + 2.9643\text{E-}7 * \text{MAHINDRA.mtcap2} +$$

INTERPRETATION:

As market capitalization increases, both Ashok Leyland (AL) and Mahindra (M) invest more whereas Tata (T) invests less which is counterintuitive however its coefficient is insignificant at even 10% level. More specifically, as market capitalization increases by Rs. 1 million. Investment by AL increases by Rs. 0.08 million. And for M, it increases by Rs. 0.122 million. The coefficients of AL and M are significant at 5% level. With increase in market capitalization, investment by M increases the maximum.

As net fixed assets increase, investment by all 3 firms increases, all their coefficients are significant at 5% and 1% level. Investment by AL increases by Rs. 0.184 million Rs., or M increases by Rs. 1.342 million. And for T increases by Rs. 2.67 million due to Rs. 1 million. Increase in net fixed assets. Hence investment by T increases the most in response to 1 unit increase in net fixed assets.

ITSUR

As market capitalization increases by 1 million Rs. Investment by AL increases by 0.066 million Rs., for M by 0.18 million Rs and for T by 0.075 million Rs. But t's coefficient is insignificant at 5% level whereas the other 2 coefficients are significant at 5% level. Thus in response to 1 unit increase in market capitalization, M's investment increases the maximum. As net fixed assets increase by 1 million Rs, investment by AL increases

by 0.305 million Rs., for M by 0.392 million Rs. And for T by 1.077 million Rs. All the coefficients are significant at 5% level. Thus in response to 1 unit increase in net fixed assets, investment by T increases the maximum.

Comparison of efficiency between SUR, OLS AND ITSUR

From the standard error of the parameter estimates, we find that ITSUR is the most efficient method of estimation followed by SUR/FGLS and OLS is the most inefficient method of estimation with the standard error of the estimates the highest. This conforms to our a priori expectations.

CONCLUSION:

Thus we see that most of the results of the estimation process of SUR holds with our model. Due to lack of required number of data points the iterated feasible generalized least square result varies from the desired result in estimation but efficiency results still holds good.

The basic linear SUR model can be extended in the following ways:

ENDOGENOUS REGRESSORS- When the regressor X_t in the SUR model is correlated with the error term ϵ_t , one needs instrumental variables say $Z_t = (Z_{1t}', Z_{2t}', \dots, Z_{nt}')'$ to estimate B. The IV's should be correlated with the concerned endogenous regressors and uncorrelated with the error term. The IV's should also satisfy the usual rank condition (No. of IV's should be \geq no. of endogenous regressors). The optimal GMM estimator is derived by minimizing the GMM objective

function with the optimal choice of the weighting matrix given by $[\hat{\Sigma}^{-1} (\sum_{t=1}^T Z_t' Z_t)]^{-1}$.

NONLINEAR SUR- It assumes that the conditional mean of Y_{it} given X_{it} is nonlinear say $h_i(B, X_{it})$. We can estimate B by using quasi MLE assuming that Y_t 's are Gaussian conditioned on X_t or GMM utilizing the moment condition $E[g(X_t) \epsilon_t] = 0$ for any measurable transformation $g(\cdot)$ of X_t .

SUR estimation, today, finds vast usage in various modeling by statisticians and econometricians over the world. It is being combined with TOBIT models and other econometric models for suiting different economic and social scenarios. It is used in growth curve analysis by Aiyi Liu, in the analysis of the effect of devaluation on trade balance by Marc A. Miles and many other economists in their papers. It is used for estimation of the productivity of firms of similar sectors due to world economic scenarios. Another example of SUR estimation is determining the effect of nominal contracting on stock returns. This was done by Kenneth R. French of University of California, Los Angeles, Richard .S. Ruback Massachusetts Institute of Technology and G. William Schwert from University of Rochester in their paper.

REFERENCES

- Seemingly unrelated regressions- Moon, Perron ,July 2006.
- An efficient method of estimating SUR and tests for aggregation bias-Arnold Zellner, Journal of American Statistical Association, Vol. 57, pg 348-368.
- Greene(2008), Econometric Analysis.
- Wooldridge(2001), Econometric analysis of cross section and panel data.