## Linear Systems of Equations

$$\underset{\sim}{y}_1 = X_1 \underset{\sim}{\beta}_1 + \underset{\sim}{u}_1$$

$\underset{\sim}{y}_1 \quad T \times 1$

$X_1 \quad T \times k_1$

$\underset{\sim}{\beta}_1 \quad k \times 1$

$\underset{\sim}{u}_1 \quad T \times 1$

$$\underset{\sim}{y}_2 = X_2 \underset{\sim}{\beta}_2 + u_2 \qquad \underset{\sim}{y}_2 \; T \times 1 \quad X_2 \; T \times k_2$$

$$\vdots$$

$$\underset{\sim}{y}_m = X_m \underset{\sim}{\beta}_m + u_m \qquad y_m \; T \times 1 \quad X_m \; T \times k_m$$

Hence, there are M Equations and each equation has T observations

The equations may contain different variables, or same variables for different entities.

The latter was the case for the famous Grunfeld (1958) example.

$$I_{it} = \beta_{1i} + \beta_{2i} F_{it} + \beta_{3i} C_{it} + u_{it}$$

$I_{it}$ was investment exp for firm i at time t

$F_{it}$ anticipated profit for i at t
$C_{it}$ capital stock for i and t.

There were a number ~~fo~~ of firms

The idea was that each firm had a seperate investment function, but respond similarly to external events that occur in time.

That means their errors (omitted events) are contemporaneously correlated.

e.g., macro economic events (shocks) are likely to affect all firms in the periods that they occur.

## Model

The $i^{th}$ firm

$$\underset{\sim}{y}_i = X_i \beta_i + \underset{\sim}{u}_i \qquad i = 1, 2, \ldots, M$$

$$\underset{\sim}{u}_{MT \times 1} = \begin{pmatrix} \underset{\sim}{u}_1 \\ \underset{\sim}{u}_2 \\ \vdots \\ \underset{\sim}{u}_m \end{pmatrix}$$

Let $E[\underset{\sim}{u} \mid X_1, X_2, \ldots, X_m] = 0$
(strict exog.)

Homosked. within firm
$$E[\underset{\sim}{u}_m \underset{\sim}{u}_m^T] = \sigma_{mm} I_T$$

Each Eq. has $k_i$ regressors
so, let $K = \sum_{i=1}^{m} k_i$ Be total # in system.

$T \times k_i \quad \forall i$

Also, $E[u_{it} u_{js} \mid X_1, X_2, \ldots, X_m] =$

$$\begin{matrix} \sigma_{ij} & t = s \\ 0 & t \neq s \end{matrix}$$

So, $E[\underset{\sim}{u}_i \underset{\sim}{u}_j^T] = \sigma_{ij} I_T$

$$E[\underset{\sim}{u}\,\underset{\sim}{u}^T] = E\begin{bmatrix} \underset{\sim}{u}_1 u_1^T & \underset{\sim}{u}_1 u_2^T & \cdots & \underset{\sim}{u}_1 u_m^T \\ & \ddots & & \\ & & \ddots & \\ & & & \underset{\sim}{u}_m \underset{\sim}{u}_m^T \end{bmatrix}$$

$$\equiv \Omega = \begin{bmatrix} \sigma_{11} I_T & \sigma_{12} I_T & \cdots & \sigma_{1m} I_T \\ \sigma_{21} I_T & \sigma_{22} I_T & & \\ & & \ddots & \\ & & & \sigma_{mm} I_T \end{bmatrix}$$

<u>GLS</u>

$$\begin{pmatrix} \underset{\sim}{y}_1 \\ \underset{\sim}{y}_2 \\ \vdots \\ \underset{\sim}{y}_m \end{pmatrix} = \begin{bmatrix} X_1 & 0 & 0 & \cdots & 0 \\ 0 & X_2 & 0 & \cdots & 0 \\ \vdots & & & & \\ \vdots & & & & \\ 0 & 0 & \cdot & 0 & X_m \end{bmatrix} \begin{matrix} \underset{\sim}{\beta}_1 \\ \underset{\sim}{\beta}_2 \\ \vdots \\ \underset{\sim}{\beta}_m \end{matrix} + \begin{matrix} \underset{\sim}{u}_1 \\ \underset{\sim}{u}_2 \\ \vdots \\ u_m \end{matrix}$$

$$\underset{mT\times 1}{\underset{\sim}{y}} = \underset{mT\times K}{X}\,\underset{K\times 1}{\underset{\sim}{\beta}} + \underset{mT\times 1}{\underset{\sim}{u}}$$

Set

$$\underset{m \times m}{\Sigma} \equiv \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \cdot & \sigma_{2m} \\ \cdot & & & \\ \vdots & & & \\ \sigma_{m1} & \cdot & \cdot\ \cdot & \sigma_{mm} \end{bmatrix}$$

$$\Omega = \Sigma \otimes I_T$$

$\otimes$ is a Kronecker Product

$A \otimes B$ with

$$A = \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix} \qquad B = \begin{matrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{matrix}$$

$$= \begin{matrix} a_{11} B & a_{12} B \\ a_{21} B & a_{22} B \end{matrix}$$

$$= \begin{bmatrix} a_{11} b_{11} & a_{11} b_{12} & b_{11} b_{13} & a_{21} b_{11} & \cdots a_{21} b_{13} \\ a_{11} b_{21} & a_{11} b_{22} & a_{1} b_{23} & a_{21} b_{21} & a_{21} b_{23} \\ a_{21} b_{11} & a_{21} b_{12} & a_{21} b_{13} & - & \\ a_{21} b_{21} & a_{21} b_{22} & a_{21} b_{23} & & - \end{bmatrix}$$

## Algebra

- NOTE: For square & invertible matrices $A$ & $B$.

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

otherwise,

$$(A \otimes B)^{T} = A^{T} \otimes B^{T}$$

$$(A \otimes B)(C \otimes D) = AC \otimes BD$$

$$GLS \Rightarrow \quad \hat{\beta}_{GLS} = (X^{T} \Omega^{-1} X)^{-1} X^{T} \Omega^{-1} y$$

$$= (X^{T} (\Sigma \otimes I)^{-1} X)^{-1} X^{T} (\Sigma \otimes I)^{-1} y$$

$$= (X^{T} (\Sigma^{-1} \otimes I) X)^{-1} X^{T} (\Sigma^{-1} \otimes I) \underset{\sim}{y}$$

This estimator is more efficient than OLS eq. by eq. except

(1) if the equations really are uncorrelated, i.e.,

$$\sigma_{ij} = 0 \quad i \neq j$$

~~Prbm~~

(2) If the equations have identical explanatory variables

$$X_i = X_j$$

In general, efficiency improves the higher the contemporaneous correlation and the more dissimilar the regressors between equations.

The equivalence of OLS and SUR when regressors are the same in each equation can be mitigated if cross-equation restrictions are imposed (e.g., cross-price elasticities are same in any ~~two demand~~ pair of demand eq.s when the goods are complements or substitutes)

# Feasible GLS

$\Sigma$ is unknown and is estimated as follows.

$$\hat{\sigma}_{ij} = \hat{\underset{\sim}{u}}_i^T \hat{\underset{\sim}{u}}_j / \tau_{ij}$$

where $\hat{\underset{\sim}{u}}_i$ are the LS residuals from estimating eq. $i$

Any of these consist.

$$\tau_{ij} = \begin{cases} T \\ (T-k_i)^{1/2}(T-k_j)^{1/2} \\ T - \max(k_i, k_j) \end{cases}$$

FGLS $\hat{\beta}_{FGLS} \overset{a}{\sim} N\left(\underset{\sim}{\beta}, \left(X(\Sigma^{-1} \otimes I)X\right)^{-1}\right)$

And $\therefore$ Wald tests can be performed in the usual manner. Either within an equation or across equations or both.

The Model can be modified to allow for autocorrelation within equations or heteroskedasticity within eq.

Specification Tests

If $\sigma_{ij} = 0$ then eq. by eq. OLS is efficient and easy.

$$LM = n \sum_{i=2}^{M} \sum_{j=1}^{M-1} r_{ij}^2 \sim \chi^2_{\frac{M(M-1)}{2}}$$

if Ho True

$r_{ij}$ are the ~~s~~ sample correlations among the LS residuals

$H_0: \sigma_{ij} = 0 \quad \forall i \neq j$

$H_A$: not $H_0$:

In the Grunfeld example, you might want to test whether the Model could be __pooled__.

$$\underset{MT\times 1}{\underset{\sim}{y}} = \underset{MT\times 3}{\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix}} \underset{\sim}{\beta} + \underset{MT\times 1}{\underset{\sim}{u}}$$

That is each firm has same intercept, profit slope, & Cap. Slope.

let $M=10$

$R_i = 3 \quad i = 1, \dots, M.$

Wald Test is.

$$\frac{(SSR_R - SSR_u)/27}{SSR_u / MT-K} \overset{a}{\sim} F_{27, MT-K} \text{ if.}$$

pooled model is correct.

correlated with other regressors, then you should use the fixed effects estimator to obtain consistent parameter estimates of your slopes.

In the Grunfeld data, a p-value less than 5% indicates that the Breusch-Pagan test rejects the hypothesis that the effects are not random (in other words, the effects **are** random). For the Hausman test, the p-value is greater than 5%. The random effects do not appear to be correlated with the regressors and random effects can be used.

### 15.2.4 SUR

The acronym SUR stands for **seemingly unrelated regression equations**. SUR is another way of estimating panel data models that are long (large T) but not wide (small N). More generally though, it is used to estimate systems of equations that do not necessarily have any parameters in common and are hence unrelated. In the SUR framework, each firm in your sample is parametrically different; each firm has its own regression function, i.e., different intercept and slopes. Firms are not totally unrelated, however. In this model the firms are linked by what is **not** included in the regression rather than by what is. The firms are thus related by unobserved factors and SUR requires us to specify how these omitted factors are linked in the system's error structure.

In the basic SUR model, the errors are assumed to be homoscedastic and linearly independent within each equation, or in our case, each firm. The error of each equation may have its own variance. Most importantly, each equation (firm) is correlated with the others in the same time period. The latter assumption is called **contemporaneous correlation**, and it is this property that sets SUR apart from other models.

In the context of the two firm Grunfeld model in (15.4) this would mean that $Var[e_{GE,t}] = \sigma^2_{GE}; Var[e_{W,t}] = \sigma^2_W$; $Cov(e_{GE,t}, e_{W,t}) = \sigma_{GE,W}$ for all time periods; and $Cov(e_{i,t}, e_{i,s}) = 0$ for $t \neq s$ for each firm, $i = GE, W$. So in the SUR model you essentially have to estimate a variance for each individual and a covariance between each pair of individuals. These are then used to construct a generalized least squares estimator of the equations parameters.

Even though SUR requires a $T$ and an $N$ dimemsion, it is not specifically a panel technique. This is because the equations in an SUR system may be modeling different behaviors for a single individual rather than the same behavior for several individuals. As mentioned before, it is best used when panels are long and narrow since this gives you more observations to estimate the equations variances and the cross equation covariances. More time observations reduces the sampling variation associated with these estimates, which in turn improves the performance of the feasible generalized least squares estimator. If your panel dataset has a very large number of individuals and only a few years, then FGLS may not perform very well in a statistical sense. In the two firm Grunfeld example, N=2 and T=20 so we needn't worry about this warning too much, although the asymptotic inferences are based on $T$ (and not $N$) being infinite.

When estimating an SUR model, the data have to be arranged in a slightly different way than in the preceding panel examples. Basically, they need to be arranged as a time series (not a panel) with different firms variables listed separately. Hill et al. [2007] have done this for us in the *grunfeld2.gdt* data set. The **gretl** script to estimate the two firm SUR model using this data is

```
open "c:\Program Files\gretl\data\poe\grunfeld2.gdt"

system name="Grunfeld"
equation inv_ge const v_ge k_ge
equation inv_we const v_we k_we
end system

estimate "Grunfeld" method=sur
```

Since SUR is a method of estimating a system of equations (just like you did in chapter 11), the same syntax is used here. It consists of a block of code that starts with the `system name="Grunfeld"` line. One advantage naming your system is that results are attached to it and you can perform subsequent computations based on them. For instance, with a saved set of equations you can impose restrictions on a single equation in the model or impose restrictions across equations. This is accomplished using the `restrict` statement.

Following the system name, each equation is put on a separate line. Notice that each equation is identified using `equation` which is followed by the dependent variable and then the independent variables which include a constant. Close the system block using the `end system` command. The system is then estimated using the line `estimate "Grunfeld" method=sur`. Executing this script yields Figure 15.3 below.

The test to determine whether there is sufficient contemporaneous correlation is simple to do from the standard output. Recall from *POE* that the test is based on the squared correlation

$$r^2_{GE,W} = \frac{\hat{\sigma^2}_{GE,W}}{\hat{\sigma^2}_{GE}\hat{\sigma^2}_{W}} \tag{15.8}$$

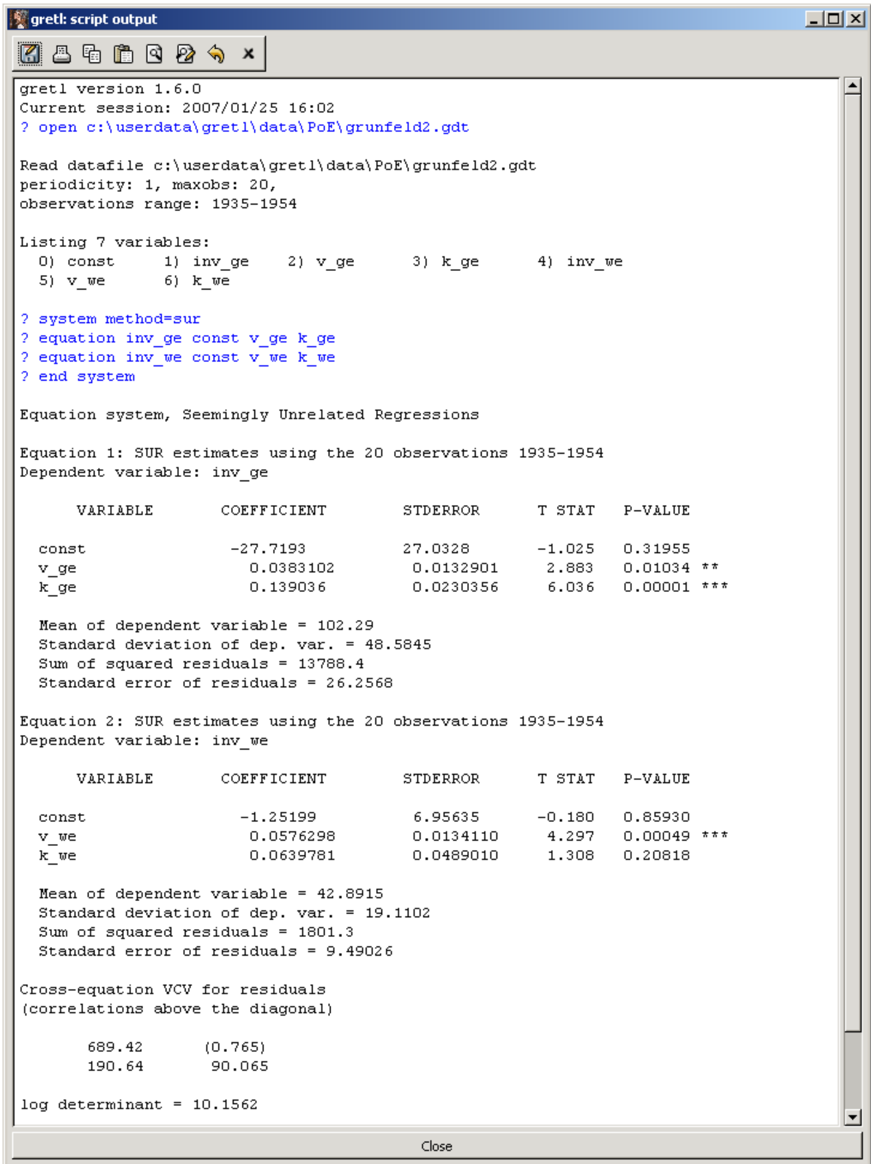
A little caution is required here. The squared correlations are supposed to be computed based on the residuals from the *least squares estimator*, not SUR. The "Cross-equation VCV for residuals" in the output in Figure 15.3 is computed based on SUR residuals. So, you'll need to change the estimation method to `ols` and rerun the script to get the right inputs for this statistic. The new script is:

```
open "c:\Program Files\gretl\data\poe\grunfeld2.gdt"

system name="Grunfeld"
equation inv_ge const v_ge k_ge
equation inv_we const v_we k_we
end system

estimate "Grunfeld" method=ols
```

Figure 15.3: The results from the two firm model estimated as seemingly unrelated regression equations





```
gretl version 1.6.0
Current session: 2007/01/25 16:02
? open c:\userdata\gretl\data\PoE\grunfeld2.gdt

Read datafile c:\userdata\gretl\data\PoE\grunfeld2.gdt
periodicity: 1, maxobs: 20,
observations range: 1935-1954

Listing 7 variables:
  0) const      1) inv_ge    2) v_ge      3) k_ge      4) inv_we
  5) v_we       6) k_we

? system method=sur
? equation inv_ge const v_ge k_ge
? equation inv_we const v_we k_we
? end system

Equation system, Seemingly Unrelated Regressions

Equation 1: SUR estimates using the 20 observations 1935-1954
Dependent variable: inv_ge

      VARIABLE       COEFFICIENT        STDERROR      T STAT    P-VALUE

  const               -27.7193          27.0328       -1.025    0.31955
  v_ge                  0.0383102        0.0132901     2.883    0.01034 **
  k_ge                  0.139036         0.0230356     6.036    0.00001 ***

  Mean of dependent variable = 102.29
  Standard deviation of dep. var. = 48.5845
  Sum of squared residuals = 13788.4
  Standard error of residuals = 26.2568

Equation 2: SUR estimates using the 20 observations 1935-1954
Dependent variable: inv_we

      VARIABLE       COEFFICIENT        STDERROR      T STAT    P-VALUE

  const                -1.25199          6.95635      -0.180    0.85930
  v_we                  0.0576298        0.0134110     4.297    0.00049 ***
  k_we                  0.0639781        0.0489010     1.308    0.20818

  Mean of dependent variable = 42.8915
  Standard deviation of dep. var. = 19.1102
  Sum of squared residuals = 1801.3
  Standard error of residuals = 9.49026

Cross-equation VCV for residuals
(correlations above the diagonal)

       689.42       (0.765)
       190.64        90.065

log determinant = 10.1562
```

and the resulting cross-equation variance covariance for the residuals is

```
Cross-equation VCV for residuals
(correlations above the diagonal)

       777.45      (0.729)
       207.59       104.31
```

Then you compute

$$r^2_{GE,W} = \frac{207.59^2}{(777.45)(104.31)} = 0.729 \tag{15.9}$$

Notice that **gretl** produces this number for you in the upper diagonal of the matrix and places it in parentheses. Using the given computation the test statistic is

$$LM = Tr^2_{GE,W} \xrightarrow{d} \chi^2_{(1)} \tag{15.10}$$

provided the null hypothesis of no correlation is true. The arithmetic is $(20 * 0.729) = 14.58$

The `restrict` command can be used to impose the cross-equation restrictions on a system of equations that has been previously defined and named. The set of restrictions is started with the keyword `restrict` and terminated with `end restrict`. Some additional details and examples of how to use the `restrict` command are given in section 6.1. Each restriction in the set is expressed as an equation. Put the linear combination of parameters to be tested on the left-hand-side of the equality and a numeric value on the right. Parameters are referenced using `b[i,j]` where $i$ refers to the equation number in the system, and $j$ the parameter number. So, to equate the intercepts in equations one and two use the statement

$$b[1,1] - b[2,1] = 0 \tag{15.11}$$

The full syntax for testing the full set of cross-equation restrictions

$$\beta_{1,GE} = \beta_{1,W}, \quad \beta_{2,GE} = \beta_{2,W}, \quad \beta_{3,GE} = \beta_{3,W} \tag{15.12}$$

on equation 15.4 is shown in Table 15.1: **Gretl** estimates the two equation SUR subject to the restrictions. Then it computes an F-statistic of the null hypothesis that the restrictions are true versus the alternative that at least one of them is not true. It returns the computed F-statistic and its p-value. A p-value less than the desired level of significance leads to a rejection of the hypothesis.

The **gretl** output from this test procedure is

```
F test for the specified restrictions:
  F(3,34) = 2.92224 with p-value 0.0478934
```

which matches the results in the text. At the 5% level of significance, the equality of the two equations is rejected.

Table 15.1: Script for imposing cross-equation restrictions in an SUR model

```
system name="Grunfeld"
equation inv_ge const v_ge k_ge
equation inv_we const v_we k_we
end system

restrict "Grunfeld"
b[1,1]-b[2,1]=0
b[1,2]-b[2,2]=0
b[1,3]-b[2,3]=0
end restrict

estimate "Grunfeld" method=sur --geomean
```

## 15.3 NLS Example

Hill et al. [2007] provides a subset of National Longitudinal Survey which is conducted by the US Department of Labor. The database includes observations on women, who in 1968, were between the ages of 14 and 24. It then follows them through time, recording various aspects of their lives annually until 1973 and bi-annually afterwards. Our sample consists of 716 women observed in 5 years (1982, 1983, 1985, 1987 and 1988). The panel is balanced and there are 3580 total observations.

Two models are considered in equations (15.13) and (15.14) below.

$$\begin{aligned} ln(WAGE)_{it} = \beta_{1i} + \beta_2 exper_{it} + \beta_3 exper^2_{it} + \beta_4 tenure_{it} \\ + \beta_5 tenure^2_{it} + \beta_6 south_{it} + \beta_7 union_{it} + e_{it} \end{aligned} \tag{15.13}$$

$$\begin{aligned} ln(WAGE)_{it} = \beta_{1i} + \beta_2 exper_{it} + \beta_3 exper^2_{it} + \beta_4 tenure_{it} \\ + \beta_5 tenure^2_{it} + \beta_6 south_{it} + \beta_7 union_{it} + \\ + \beta_8 black_{it} + \beta_9 educ_{it} + e_{it} \end{aligned} \tag{15.14}$$

The first model (15.13) is estimated using fixed effects. Race (`black`) and education (`educ`) are added to form the model in (15.14). Since these variables do not change for individuals in the sample, their influences cannot be estimated using fixed effects. So, this equation is estimated using random effects using the script below:

```
open "c:\Program Files\gretl\data\poe\nels_panel.gdt"
panel lwage const exper exper2 tenure tenure2 south union
```