Lecture 14
SUR

Panel Data Sets

• A panel data set, or longitudinal data set, is one where there are repeated observations on the same units. Now, we have $\{y_{it}, \mathbf{x}_{it}\}$, where $i = 1, 2, ..., N$ and $t = 1, 2, ..., T_i$, usually, $N > T_i$.

• The units –the *i's*– may be individuals, households, firms, countries, or any set of entities that remain stable through time.

• Repeated observations create a potentially very large panel data sets. With N units and T time periods $\Rightarrow$ Number of observations: NT.
– Advantage: Large sample! Great for estimation.
– Disadvantage: Dependence! Observations are likely not independent

• Modeling the potential dependence creates different models.

Panel Data Sets

• The National Longitudinal Survey (NLS) of Youth is an example. The same respondents were interviewed every year from 1979 to 1994. Since 1994 they have been interviewed every two years.

• The *CRSP* database has daily and monthly stock and index returns from 1962 on for over 5,000 stocks (N=5,000 and T(monthly)=600).

Panel Data Sets

• Panel data sets are often very large. If there are N units and T time periods, the potential number of observations is NT (for the CRSP dataset, we have over 3 million observations). Potentially, great for estimation!

$$\begin{array}{cc} & \text{Cross section} \\ \begin{array}{c}\text{Time}\\ \text{series}\end{array} & \begin{bmatrix} y_{11} & y_{21} & \cdots & y_{i1} & \cdots & y_{N1} \\ y_{12} & y_{22} & \cdots & y_{i2} & \cdots & y_{N2} \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ y_{1t} & y_{2t} & \cdots & y_{it} & \cdots & y_{Nt} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ y_{1T} & y_{2T} & \cdots & y_{iT} & \cdots & y_{NT} \end{bmatrix} = [\mathbf{y}_1\ \mathbf{y}_2 \ldots \mathbf{y}_i \ldots \mathbf{y}_N] \end{array}$$

Panel Data Sets

- Notation:

$$y_1 = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1t} \\ \vdots \\ y_{1T} \end{bmatrix};\ldots; y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{it} \\ \vdots \\ y_{iT} \end{bmatrix} \quad X_1 = \begin{bmatrix} x_{11} & x_{21} & \ldots & x_{k1} \\ x_{12} & x_{22} & \ldots & x_{k2} \\ \vdots & \vdots & \ldots & \vdots \\ x_{1t} & x_{2t} & \ldots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1T_1} & x_{2T_1} & \ldots & x_{kT_1} \end{bmatrix}; \ldots; X_i = \begin{bmatrix} w_{11} & w_{21} & \ldots & w_{k1} \\ w_{12} & w_{22} & \ldots & x_{k2} \\ \vdots & \vdots & \ldots & \vdots \\ w_{1t} & w_{2t} & \ldots & w_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ w_{1T_j} & w_{2T_j} & \ldots & w_{kT_j} \end{bmatrix}$$

- A standard panel data set model stacks the $\mathbf{y}_i$'s and the $\mathbf{x}_i$'s:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{c} + \boldsymbol{\varepsilon}$$

$\mathbf{X}$ is a $\Sigma_i T_i$xk matrix

$\boldsymbol{\beta}$ is a kx1 matrix

$\mathbf{c}$ is $\Sigma_i T_i$x1 matrix, associated with unobservable variables.

$\mathbf{y}$ and $\boldsymbol{\varepsilon}$ are $\Sigma_i T_i$x1 matrices

Panel Data Sets

- Longitudinal data (Large *N*)
  - National longitudinal survey of youth (NLS)
  - British household panel survey (BHPS)
  - Panel Study of Income Dynamics (PSID)
- Time series cross section (TSCS) data (Large *T*)
  - Grunfeld's investment data
  - Penn world tables
- Financial data
  - COMPUSTAT provides financial data by firm (*N* =99,000) and by quarter (*T* = 1962:I, 1962:II, ...,)
  - Exchange rate data, essentially infinite *T*, *N*=160+
  - *Datastream* provides economic and financial data for countries. It also covers bonds and stock markets around the world.
  - *OptionMetrics* is a database of historical prices, implied volatility for listed stocks and option markets.

Balanced and Unbalanced Panels

• Notation:

$y_{i,t}$, $i = 1,\ldots,N$; $t = 1,\ldots,T_i$

• Mathematical and notational convenience:

- Balanced: NT

(that is, every unit is surveyed in every time period.)

- Unbalanced: $\sum_{i=1}^{N} T_i$

Q: Is the fixed T_i assumption ever necessary? SUR models.

• The NLS of Youth is *unbalanced* because some individuals have not been interviewed in some years. Some could not be located, some refused, and a few have died. CRSP is also *unbalanced,* some firms are listed from 1962, others started to be listed later.

Panel Data Model: CLM Revisited

• The DGP of the CLM is slightly modified:

(**A1**) $y_{it} = \mathbf{x}_{it}'\beta_i + \varepsilon_{it}$

$i = 1, 2, \ldots, N$ - we have N *individual, groups* or *firms*.

$t = 1, 2, \ldots, T_i$ - usually, $N > T_i$.

That is, the classical linear relation applies to each of N equations and T observations. If we assume **(A2)** to **(A4)**, the y_i's are independent. No gain from a system estimation ⇒ N OLS estimations are all we need!

Example: The CAPM:

$$r_{it} - r_{ft} = \alpha_i + \beta_i(r_{mt} - r_{ft}) + \varepsilon_{it}$$

In the CAPM, $\mathbf{x}_{it} = \mathbf{x}_t$ ⇒ Explanatory variables are common across i.

Note: In economics, N is traditionally small – 50 states, few developed countries. But, no for the CAPM: N is in the thousands!

Panel Data Model: CLM Revisited - Notation

• Rewrite (**A1**) DGP using matrix notation.

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i \qquad i = 1, 2,..., N$$

- Dimensions:

$\mathbf{X}_i$ is a T_ixk matrix

$\boldsymbol{\beta}_i$ is a kx1 matrix

$\mathbf{y}_i$ and $\boldsymbol{\varepsilon}_i$ are T_ix1 matrices

• Now, stacking all the equations:

$$\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- Dimensions:

$\mathbf{X}$ is a $\Sigma_i T_i$xNk matrix (if T_i=T for all i => $\mathbf{X}$ is NTxNk)

$\boldsymbol{\beta}$ is a Nkx1 matrix

$\mathbf{y}$ and $\boldsymbol{\varepsilon}$ are $\Sigma_i T_i$x1 matrices

Panel Data Model: CLM Revisited – (A3')

• DGP: (**A1**) $\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

where $\mathbf{X}$ is a $\Sigma_i T_i$xNk matrix, $\boldsymbol{\beta}$ is a Nkx1 matrix, and $\mathbf{y}$ and $\boldsymbol{\varepsilon}$ are $\Sigma_i T_i$x1 matrices

• General formulation for covariance matrix: (**A3'**) $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}] = \mathbf{V}$

Note: $\mathbf{V}$ is an $\Sigma_i T_i$x$\Sigma_i T_i$ matrix (if T_i=T for all i, then it is an NT x NT matrix): Huge!

• We can have different elements in (**A3'**):

(1) Standard groupwise heteroscedasticity (diagonal elements)

(2) Autocorrelated errors (off-diagonal i elements)

(3) Contemporaneously cross-correlated errors (off-diagonal ij elements)

(4) Time-varying cross-correlated errors (off-diagonal ij elements)

Seemingly Unrelated Regressions (SUR)

• In the SUR model we assume a specific form for **V**:

(**A1**) $\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

(**A2**) $E[\boldsymbol{\varepsilon}_i|\mathbf{X}] = \mathbf{0},$

(**A3'**) $\text{Var}[\varepsilon_i|\mathbf{X}] = \sigma_i^2\mathbf{I}_T = \sigma_{ii}\mathbf{I}_T$ – groupwise heteroscedasticity.

$E[\varepsilon_{it}\,\varepsilon_{jt}\,|\mathbf{X}] = \sigma_{ij}$ – contemporaneous correlation

$E[\varepsilon_{it}\,\varepsilon_{is}\,|\mathbf{X}] = 0\ (t{\neq}s)$ – no autocorrelation

$E[\varepsilon_{it}\,\varepsilon_{js}\,|\mathbf{X}] = 0\ (t{\neq}s)$ – no time-varying cross-correlation

(**A4**) Rank(**X**) = full rank Nk

• In (**A1**)-(**A4**), we have a GR model.

• In (**A1**), individual *i* seems independent of individual *j*. But, they are not. They are related through the covariance matrix in (**A3'**).

SUR: Formulation

• Q: What kind of theoretical structure produces a SUR DGP?

A: We need a model where there is a specific, heteroscedastic *i* factor and a common factor to all individuals. This common factor causes contemporaneous correlation only. It causes no correlations over time.

In finance, the variation in excess returns is affected both by firm specific factors and by the economy as a whole.

• The SUR model is a GR model. A rich model with (assume T_i=T):

(1) Different coefficient vectors for each *i* $\Rightarrow Nk$ parameters

(2) Different variances for each *i* $\Rightarrow N$ parameters

(3) Correlation across *i* at each *t* $\Rightarrow N(N\text{-}1)/2$ parameters

<u>Note</u>: We have NT observations to estimate $(Nk + N + N(N\text{-}1)/2)$ parameters. We need T to be reasonably big.

SUR: Formulation

• In (**A1**), individual *i* seems independent of individual *j*. But, they are not. They are related through the covariance matrix in (**A3'**).

• Rewrite the contemporaneous correlation structure in (**A3'**):

$$E[\varepsilon_{it}\,\varepsilon_{jt}\,|\mathbf{X}] = \sigma_{ij} \qquad \text{–contemporaneous correlation}$$

$$E[\varepsilon_{it}\,\varepsilon_{js}\,|\mathbf{X}] = 0 \qquad \text{when } t \neq s$$

$$\mathrm{E}(\varepsilon_i \varepsilon_j{}') = E\begin{pmatrix} \varepsilon_{1i}\varepsilon_{1j} & \varepsilon_{1i}\varepsilon_{2j} & \cdots & \varepsilon_{1i}\varepsilon_{Tj} \\ \varepsilon_{2i}\varepsilon_{1j} & \varepsilon_{2i}\varepsilon_{2j} & \cdots & \varepsilon_{2i}\varepsilon_{Tj} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{Ti}\varepsilon_{1j} & \varepsilon_{Ti}\varepsilon_{2j} & \cdots & \varepsilon_{Ti}\varepsilon_{Tj} \end{pmatrix} = \begin{pmatrix} \sigma_{ij} & 0 & \cdots & 0 \\ 0 & \sigma_{ij} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{ij} \end{pmatrix}$$

$$= \sigma_{ij} I_T$$

• This covariance matrix for the model is an NTxNT matrix (if $T_i=T$). To get the final SUR formulation, stack the equations ⇒ GR model.

Example: The SUR Model (2x2 case)

Consider a two equation system.

$$\mathbf{y_1} = \mathrm{X_1}\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1$$
$$\mathbf{y_2} = \mathrm{X_2}\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2$$

Now, stack these two equations: $\begin{bmatrix} \mathrm{y}_1 \\ \mathrm{y}_2 \end{bmatrix} = \begin{bmatrix} \mathrm{X}_1 & 0 \\ 0 & \mathrm{X}_2 \end{bmatrix}\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$

The disturbance covariance matrix is

$$\mathrm{Var}\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} = \mathrm{E}\begin{bmatrix} \varepsilon_1\varepsilon_1{}' & \varepsilon_1\varepsilon_2{}' \\ \varepsilon_2\varepsilon_1{}' & \varepsilon_2\varepsilon_2{}' \end{bmatrix} = \begin{pmatrix} \sigma_{11}\mathbf{I} & \sigma_{12}\mathbf{I} \\ \sigma_{12}\mathbf{I} & \sigma_{22}\mathbf{I} \end{pmatrix}$$

$$\sigma^2\boldsymbol{\Omega} = \mathbf{V}.$$

(The σ^2 is there just to use the notation we are used to for the GR model.

SUR Estimation: OLS and GLS

• Since OLS is consistent, each equation can be fit by OLS. HAC estimator can be used for inferences.

• GLS Estimation: $\hat{\beta}_{FGLS} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y$

Q: Why do GLS? Efficiency improvement.

$$E\left((\hat{\beta}_{GLS}-\beta)(\hat{\beta}_{GLS}-\beta)'\right) = E\left((X'V^{-1}X)^{-1}X'V^{-1}ee'V^{-1}X(X'V^{-1}X)^{-1}\right)$$
$$= (X'V^{-1}X)^{-1}X'V^{-1}VV^{-1}X(X'V^{-1}X)^{-1}$$
$$= (X'V^{-1}X)^{-1}$$

• Gains to GLS:

- Efficiency gains increase as the cross equation correlation increases.
- But, no gains if identical regressors –for example, in the CAPM.

$\Rightarrow$ GLS is the same as OLS.

SUR: GLS Estimation

• Derivation of the GLS estimator for the 2x2 case:

$$\hat{\beta}_{GLS} = (X'V^{-1}X)^{-1}X'V^{-1}y = \left(\begin{pmatrix} X_1' & 0 \\ 0 & X_2' \end{pmatrix} \begin{pmatrix} \sigma_1^2 I_T & \sigma_{12} I_T \\ \sigma_{21} I_T & \sigma_2^2 I_T \end{pmatrix}^{-1} \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \right)^{-1} \begin{pmatrix} X_1' & 0 \\ 0 & X_2' \end{pmatrix} \begin{pmatrix} \sigma_1^2 I_T & \sigma_{12} I_T \\ \sigma_{21} I_T & \sigma_2^2 I_T \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$= \left(\begin{pmatrix} X_1' & 0 \\ 0 & X_2' \end{pmatrix} \begin{pmatrix} \frac{\sigma_{22}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} I_T & \frac{-\sigma_{12}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} I_T \\ \frac{-\sigma_{21}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} I_T & \frac{\sigma_{11}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} I_T \end{pmatrix} \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \right)^{-1} \begin{pmatrix} X_1' & 0 \\ 0 & X_2' \end{pmatrix} \begin{pmatrix} \frac{\sigma_{22}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} I_T & \frac{-\sigma_{12}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} I_T \\ \frac{-\sigma_{21}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} I_T & \frac{\sigma_{11}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} I_T \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\sigma_{22}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} X_1'X_1 & \frac{-\sigma_{12}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} X_1'X_2 \\ \frac{-\sigma_{21}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} X_2'X_1 & \frac{\sigma_{11}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} X_2'X_2 \end{pmatrix}^{-1} \begin{pmatrix} \frac{\sigma_{22}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} X_1'y_1 - \frac{\sigma_{12}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} X_1'y_2 \\ \frac{-\sigma_{21}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} X_2'y_1 + \frac{\sigma_{11}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} X_2'y_2 \end{pmatrix}$$

Notation: Kronecker Products

• A Kronecker product is a matrix product, denoted A ⊗ B, in which in the result, each element of A multiplies the entire matrix B. That is, A ⊗ B creates a matrix of matrices.

$$A \otimes B = E \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1K}B \\ a_{21}B & a_{22}B & \cdots & a_{2K}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{T1}B & a_{T2}B & \cdots & a_{TK}B \end{pmatrix}$$

<u>Note</u>: There is no requirement for conformability in this operation. The Kronecker product can be computed for any pair of matrices.

• In the SUR case $\mathbf{V} = \mathbf{\Sigma} \otimes \mathbf{I}_T$

Notation: Kronecker Products

• For the Kronecker product,

$(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$ (This is the important result for GLS.)

If $\mathbf{A}$ is $M\times M$ and $\mathbf{B}$ is $n \times n$, then

$|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^n \text{ x } |\mathbf{B}|^M$,

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

$\text{trace}(\mathbf{A}\otimes\mathbf{B}) = \text{tr}(\mathbf{A}) \text{ x } \text{tr}(\mathbf{B})$.

For $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ such that the products are defined is

$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$.

• Then, in the SUR case, the GLS estimator becomes

$$\hat{\beta}_{GLS} = (X'V^{-1}X)^{-1}X'V^{-1}y = (X'[\Sigma\otimes I]^{-1}X)^{-1}X'[\Sigma\otimes I]^{-1}y$$
$$= (X'[\Sigma^{-1}\otimes I]X)^{-1}X'[\Sigma^{-1}\otimes I]y$$

SUR – Special Case: Identical Regressors

• Back to the 2x2 case. Now, suppose the equations involve the same **X** matrices. A typical example, the CAPM.

For the two equation model, if $X_1 = X_2$, then

$$[X'V^{-1}X]^{-1} = \begin{bmatrix} \sigma^{11}X'X & \sigma^{12}X'X \\ \sigma^{12}X'X & \sigma^{22}X'X \end{bmatrix}^{-1} = (\Sigma^{-1}\otimes X'X)^{-1}$$

Using our result for the inverse of a Kronecker product, this is

$$\Sigma\otimes(X'X)^{-1} = \begin{bmatrix} \sigma_{11}(X'X)^{-1} & \sigma_{12}(X'X)^{-1} \\ \sigma_{12}(X'X)^{-1} & \sigma_{22}(X'X)^{-1} \end{bmatrix}.$$

$$X'V^{-1}y = \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix}' \begin{pmatrix} \sigma^{11}I & \sigma^{12}I \\ \sigma^{12}I & \sigma^{22}I \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{bmatrix} \sigma^{11}X'y_1 + \sigma^{12}X'y_2 \\ \sigma^{12}X'y_1 + \sigma^{22}X'y_2 \end{bmatrix}$$

We have a useful result from least squares algebra: $X'y = X'Xb$. By using this, we get a simpler result,

$$X'V^{-1}y = \begin{bmatrix} \sigma^{11}X'Xb_1 + \sigma^{12}X'Xb_2 \\ \sigma^{12}X'Xb_1 + \sigma^{22}X'Xb_2 \end{bmatrix}.$$

SUR – Special Case: Identical Regressors

Now, multiply out $[X'V^{-1}X]^{-1}\,X'V^{-1}y$. All of the $X'X$ terms will cancel out, as matrices multiply their inverses and we can collect like terms.. What remains is

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} (\sigma_{11}\sigma^{11} + \sigma_{12}\sigma^{12})b_1 + (\sigma_{11}\sigma^{12} + \sigma_{12}\sigma^{11})b_2 \\ (\sigma_{12}\sigma^{11} + \sigma_{22}\sigma^{12})b_1 + (\sigma_{12}\sigma^{12} + \sigma_{22}\sigma^{22})b_2 \end{pmatrix}.$$

Now (the rabbit in the hat).

$\sigma^{11} = \sigma_{22}/(\sigma_{11}\sigma_{22} - \sigma_{12}^2)$,
$\sigma^{22} = \sigma_{11}/(\sigma_{11}\sigma_{22} - \sigma_{12}^2)$, and
$\sigma^{12} = -\sigma_{12}/(\sigma_{11}\sigma_{22} - \sigma_{12}^2)$.

Multiplying things out once again, we get

$\sigma_{11}\sigma^{11}+\sigma_{12}\sigma^{12} = 1$ (!),
$\sigma_{11}\sigma^{12}+\sigma_{12}\sigma^{11} = 0$ (!!),
$\sigma_{12}\sigma^{11}+\sigma_{22}\sigma^{12} = 0$ (!!!) and
$\sigma_{12}\sigma^{12}+\sigma_{22}\sigma^{2} = 1$ (☺!!!!)

So, GLS equals OLS in this model.

SUR: Estimation by FGLS

- In general, **V** is unknown. We need to estimate it. We need to estimate N variances (the σ_{ii}'s) and $N(N\text{-}1)/2$) covariances (the σ_{ij}'s)
- We can use the usual FGLS two-step method or we can use ML.

(1) Two-step FGLS is essentially the same as the group-wise heteroscedastic model, starting with OLS to get the e's.

In the 2x2 example:

(1) $\hat{\sigma}_1^2 = \frac{\sum_{t=1}^{T} e_{t1}^2}{T-K} = \frac{e_1'e_1}{T-K}; \quad \hat{\sigma}_2^2 = \frac{\sum_{t=1}^{T} e_{t2}^2}{T-K} = \frac{e_2'e_2}{T-K}; \quad \hat{\sigma}_{12} = \frac{\sum_{t=1}^{T} e_{t1}e_{t2}}{T-K} = \frac{e_1'e_2}{T-K}$

(2) $\hat{\beta}_{FGLS} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y$

(2) Maximum likelihood estimation for normally distributed errors: Just iterate FGLS.

SUR: Inference About the Coefficient Vectors

- Usually based on Wald statistics. F is OK, but JF = Wald is often simpler, and is more common.

- If the estimator is MLE, the LR statistic is given by:
  LR = T * $\{\log|\mathbf{S}_{\text{restricted}}| - \log|\mathbf{S}_{\text{unrestricted}}|\}$

SUR: Pooling (Aggregation)

Q: When can we aggregate the data? Aggregation is great for estimation. Instead of having *T*, we have *NT* observations!

• A special case in which all of the $\boldsymbol{\beta}_i$'s are the same. That is,

$$\mathbf{y}_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{it}$$

• Pooling is a restricted version of the SUR model: H_0: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = ... = \boldsymbol{\beta}_N$.

• This null hypothesis can be tested: LR test, *F*-test. *F*-test:

$$F = \frac{(RSS_{Pool} - RSS_U)/(N-1)}{RSS_U/(NT-N-K)} \sim F_{N-1,NT-N-K}$$

This the original question in Zellner/Grunfeld papers: the effect of aggregation The idea was to test this proposition –i.e., all coefficient vectors are the same-, so the regression could be pooled.

SUR: Aggregation - Inference

• Testing a hypothesis about $\boldsymbol{\beta}$. The usual results for GLS. Using an estimate of

$$[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}.$$

Use the one we computed to obtain the FGLS or ML estimates.

• Tests are "asymptotic-t" or Wald tests.

• It is easy to test hypotheses about $\boldsymbol{\Sigma}$. Use a likelihood ratio test.

<u>Note</u>: Zellner (1962) was the developer of this model and estimation technique: "*An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias,*" JASA, 1962, pp. 500-509.

Arnold Zellner (1927-2010, USA)

Application: Volume and Returns

• Chuang & Susmel (2010, JBF). A bivariate SUR model is estimated to investigate the causal relation between portfolio volume and market returns across the low and high institutional ownership portfolios within each size and volume quartile over the period from January 1996 to May 2007, in Taiwan:

$$V_{ij,t} = \alpha_{ij} + \beta_{ij1} DAVR_{m,t} + \beta_{ij2} DMAD_{ij,t} + \sum_{k=1}^{K} \gamma_{ijk} R_{m,t-k} + \varepsilon_{ij,t},$$

$j = l$ and h (Low and High ownership); $i = 1, \ldots, 4$ (Portfolio Size)

$V_{ij,t}$: Value-weighted detrended trading volume of portfolio ij,
$R_{m,t}$: Return on a value-weighted Taiwanese market index,
$DAVR_{m,t}$: Detrended absolute value of $R_{m,t}$
$DMAD_{ij,t}$: Detrended value-weighted average of the beta-adjusted differences between the returns of stocks in portfolio ij and R_m.
P_{ij} : Value-weighted portfolio of size i and institutional ownership j.

Application: Volume and Returns

• Tests statistics:
- W-$K(\gamma)$ ~ χ^2 with K degrees of freedom under H_0: $\gamma_{ijk} = 0$, for all k.
- W-$1(\gamma)$ ~ χ^2_1 under H_0: $\sum_k \gamma_{ijk} = 0$.
- W-$1(\gamma_{il}=\gamma_{ih})$ ~ χ^2_1 H_0: $\sum_k \gamma_{ilk} = \sum_k \gamma_{ihk}$.
- $Q(12)$: Ljung-Box Q-statistic with up to 12 lags for the residuals in each regression.

Application: Volume and Returns

Panel A: Size-institutional ownership portfolios

P_{iI} vs. P_{iA}	P_{1I} vs. P_{1A}		P_{2I} vs. P_{2A}		P_{3I} vs. P_{3A}		P_{4I} vs. P_{4A}	
Dependent variable	V_{1A}	V_{1I}	V_{2A}	V_{2I}	V_{3I}	V_{3A}	V_{4I}	V_{4A}
Lag Length	2	2	2	2	3	3	2	2
$W\text{-}K(\gamma)$	34.7538***	27.7211***	18.5454***	16.2424***	48.1831***	27.6394***	20.2468***	8.4529**
$\sum_s \gamma_p$	0.0490	0.0313	0.0447	0.0230	0.0556	0.0400	0.0249	0.0149
$W\text{-}1(\gamma)$	34.7383***	27.6295***	18.3037***	16.0131***	44.1570***	27.0989***	19.6947***	7.9265***
Do investors trade overconfidently?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$W\text{-}1(\gamma_I = \gamma_A)$	15.6096***		7.2354***		5.4767**		7.1466***	
Do individual investors trade more overconfidently than institutional investors?	Yes		Yes		Yes		Yes	
$\bar{R}^2$	0.2462	0.1545	0.1811	0.2171	0.2416	0.2091	0.2523	0.2186
$Q(12)$	5.5352	8.4416	4.1323	6.0625	4.6436	8.5629	14.6023	12.7657

Panel B: Volume-institutional ownership portfolios

P_{iI} vs. P_{iA}	P_{1I} vs. P_{1A}		P_{2I} vs. P_{2A}		P_{3I} vs. P_{3A}		P_{4I} vs. P_{4A}	
Dependent variable	V_{1A}	V_{1I}	V_{2A}	V_{2I}	V_{3I}	V_{3A}	V_{4I}	V_{4A}
Lag Length	2	2	2	2	3	3	3	3
$W\text{-}K(\gamma)$	15.9248***	14.9392***	19.4521***	11.0707***	34.3996***	12.8443***	26.5099***	7.7262*
$\sum_s \gamma_p$	0.0420	0.0230	0.0302	0.0142	0.0431	0.0230	0.0363	0.0188
$W\text{-}1(\gamma)$	15.4173***	14.6576***	18.2309***	11.0588***	31.7993***	11.4731***	25.3486***	7.4171***
Do investors trade overconfidently?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$W\text{-}1(\gamma_I = \gamma_A)$	4.6659**		7.6575***		11.3149***		8.4940***	
Do individual investors trade more overconfidently than institutional investors?	Yes		Yes		Yes		Yes	
$\bar{R}^2$	0.0423	0.2104	0.2779	0.2456	0.2071	0.2181	0.1414	0.1528
$Q(12)$	6.4983	8.1150	11.2867	11.8460	12.7558	10.8780	9.8381	47.357

Ooops!: OLS instead of SUR

• OLS is consistent and unbiased. But, it is inefficient.

• Q: What happens if we use OLS (**b** and Var_{OLS}[**b**])?

We know Var_{OLS}[**b**] is incorrect (we should have used the sandwich estimator). We can calculate the relative efficiency of OLS relative to SUR (GLS).

Simple 2x2 setting:

$$Y_{1t} = \beta_{11} + \beta_{12} X_{1t} + \varepsilon_{1t}$$
$$Y_{2t} = \beta_{21} + \beta_{22} X_{2t} + \varepsilon_{2t} \quad \textit{for } t = 1,2,\ldots,T$$

Ooops!: OLS instead of SUR

• We can show that

$$\operatorname{var}\left(\hat{\beta}_{12,OLS}\right)=\frac{\sigma_{11}}{m_{x_1x_1}} \qquad \operatorname{var}\left(\hat{\beta}_{22,OLS}\right)=\frac{\sigma_{22}}{m_{x_2x_2}}$$

$$where \quad m_{x_1x_1} = \sum_{t=1}^{T}\left(X_{it}-\overline{X}_i\right)\left(X_{jt}-\overline{X}_j\right) \quad for\ i,\ j = 1,2$$

$$\operatorname{var}\begin{pmatrix}\hat{\beta}_{12,GLS}\\ \hat{\beta}_{22,GLS}\end{pmatrix}=\left(\sigma_{11}\sigma_{22}-\sigma_{12}^2\right)\begin{bmatrix}\sigma_{22}m_{x_1x_1} & -\sigma_{12}m_{x_1x_2}\\ -\sigma_{12}m_{x_1x_2} & \sigma_{11}m_{x_2x_2}\end{bmatrix}^{-1}$$

$$\operatorname{var}\left(\hat{\beta}_{12,GLS}\right)=\frac{\left(\sigma_{11}\sigma_{22}-\sigma_{12}^2\right)\sigma_{11}m_{x_2x_2}}{\sigma_{11}\sigma_{22}m_{x_1x_1}m_{x_2x_2}-\sigma_{12}^2m_{x_1x_2}^2}$$

$$\operatorname{var}\left(\hat{\beta}_{22,GLS}\right)=\frac{\left(\sigma_{11}\sigma_{22}-\sigma_{12}^2\right)\sigma_{22}m_{x_1x_1}}{\sigma_{11}\sigma_{22}m_{x_1x_1}m_{x_2x_2}-\sigma_{12}^2m_{x_1x_2}^2}$$

Ooops!: OLS instead of SUR

• Using $\rho = \sigma_{12}/\left(\sigma_{11}\sigma_{22}\right)^{1/2}$ and $r = m_{x_1x_2}/\left(m_{x_1x_1}m_{x_2x_2}\right)^{1/2}$ show that

$$\frac{\operatorname{var}\left(\hat{\beta}_{12,GLS}\right)}{\operatorname{var}\left(\hat{\beta}_{12,OLS}\right)}=\frac{1-\rho^2}{1-\rho^2r^2}$$

• We can differentiate with respect to $\theta = \varrho^2$ and show it is a non-increasing function of θ.

• We can differentiate with respect to $\lambda = \mathrm{r}^2$ and show it is a non-decreasing function of λ.

Ooops!: OLS instead of SUR

- Efficiency table

		ρ										
		0.00	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**	**1.00**
	0.00	1.00	0.99	0.96	0.91	0.84	0.75	0.64	0.51	0.36	0.19	0.00
	0.10	1.00	0.99	0.96	0.91	0.84	0.75	0.64	0.51	0.36	0.19	0.00
	0.20	1.00	0.99	0.96	0.91	0.85	0.76	0.65	0.52	0.37	0.20	0.00
	0.30	1.00	0.99	0.96	0.92	0.85	0.77	0.66	0.53	0.38	0.20	0.00
	0.40	1.00	0.99	0.97	0.92	0.86	0.78	0.68	0.55	0.40	0.22	0.00
r	**0.50**	1.00	0.99	0.97	0.93	0.88	0.80	0.70	0.58	0.43	0.24	0.00
	0.60	1.00	0.99	0.97	0.94	0.89	0.82	0.74	0.62	0.47	0.27	0.00
	0.70	1.00	0.99	0.98	0.95	0.91	0.85	0.78	0.67	0.52	0.32	0.00
	0.80	1.00	1.00	0.99	0.97	0.94	0.89	0.83	0.74	0.61	0.39	0.00
	0.90	1.00	1.00	0.99	0.98	0.97	0.94	0.90	0.85	0.75	0.55	0.00
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00