"Systems of Equations"

(A) The "Seemingly Unrelated Regression Equations" Model

Often, we have an economic model comprising several equations. The presence of the other equations provides additional information that can be used to help with estimation of any particular equation.

* This may enable us to improve asymptotic efficiency of estimators
* May imply that regressors are random & so OLS is inconsistent
* Motivates construction of different estimator(s).

First, consider case where there are several equations, but regressors not random.

Example 1:

Demand for food in different Provinces –

$$QF_{it} = \beta_{1i} + \beta_{2i}(PF_{it} / CPI_{it}) + \beta_{3i} RDI_{it} + \varepsilon_{it}$$

$$(i = 1, 2, ---, P)$$

A set of P regression equations. Will there be some correlation between ε_{it} & ε_{jt}?

Example 2: (CAPM)

$$(r_{it} - r_{ft}) = \beta_{1i} + \beta_{2i}(r_{mt} - r_{ft}) + \varepsilon_{it}$$

$$(i = 1, 2, ----, N)$$

A set of N equations. Why would ε_{it} and ε_{jt} be independent of each other?

Example 3:

Production fctns. : cross-section data for several years.

$$\log Q_{it} = \beta_{1t} + \beta_{2t} \log L_{it} + \beta_{3t} \log K_{it} + \varepsilon_{it}$$

$$(i = 1, ---, n \quad ; \quad t = 1, 2, 3)$$

In general:

$$\begin{aligned} y_1 &= X_1 \beta_1 + \varepsilon_1 \\ y_2 &= X_2 \beta_2 + \varepsilon_2 \\ &\vdots \\ y_M &= X_M \beta_M + \varepsilon_M \end{aligned}$$

(β_i is $k_i \times 1$)
$T > k_i$
$K = \sum_i k_i$

(M equations; T observations)

N.B.: * Can't just "pool" the data – different coefficient vectors for each equation.

* May want to test restrictions across equations (maybe impose them).

So the SURE Model is:

$$y_i = X_i \beta_i + \varepsilon_i \quad ; \quad i = 1, \ldots, M$$

$$\varepsilon = [\varepsilon_1', \varepsilon_2', \ldots, \varepsilon_M']'$$

$$E(\varepsilon) = 0 \quad ; \quad E(\varepsilon_{it}\varepsilon_{js}) = \sigma_{ij} \; ; \; t = s$$
$$= 0 \; ; \; t \neq s$$

So:

$$E(\varepsilon_i \varepsilon_j') = \sigma_{ij} I_T$$

& $$E(\varepsilon \varepsilon') = \Omega = \begin{bmatrix} \sigma_{11} I_T & \sigma_{12} I_T & \cdots & \sigma_{1M} I_T \\ \sigma_{21} I_T & \sigma_{22} I_T & & \vdots \\ \vdots & & \ddots & \\ \sigma_{M1} I_T & \cdots & & \sigma_{MM} I_T \end{bmatrix}$$

We can re-write the model as:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & & O \\ & X_2 & \\ O & & \ddots & \\ & & & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_M \end{bmatrix}$$

or, $$y = X\beta + \varepsilon$$

& $E(\varepsilon) = 0$; $\text{cov.}(\varepsilon) = E(\varepsilon\varepsilon') = \Omega$.

Clearly, OLS will be unbiased & consistent, but inefficient. This looks like a GLS problem. (Zellner, 1962).

Note that we can write:

$$\text{cov.}(\underset{\sim}{\varepsilon}) = \Omega = \begin{bmatrix} \sigma_{11} I_T & \sigma_{12} I_T & \cdots & \sigma_{1m} I_T \\ \sigma_{21} I_T & \ddots & & \vdots \\ \vdots & & \ddots & \\ \sigma_{m1} I_T & \cdots & & \sigma_{mm} I_T \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{m1} & \cdots & & \sigma_{mm} \end{bmatrix} \otimes I_T$$

Kronecker Product

$$= \Sigma \otimes I_T; \text{ say.}$$

Applying GLS:

$$\hat{\beta} = [X'\Omega^{-1}X]^{-1} X'\Omega^{-1}y$$

$$= [X'(\Sigma \otimes I_T)^{-1}X]^{-1} X'(\Sigma \otimes I_T)^{-1}y$$

$$= [X'(\Sigma^{-1} \otimes I_T)X]^{-1} X'[\Sigma^{-1} \otimes I_T]y$$

(Note advantage of this expression – dimensions)

As usual, to obtain $\hat{\beta}$ we need to know Ω (or Σ).

Two Special Cases –

1. Errors are unrelated across equations.

In this case, $\sigma_{ij} = 0$; for all $i \neq j$.

& $\Sigma = \begin{pmatrix} \sigma_{11} & & 0 \\ & \sigma_{22} & \\ 0 & & \sigma_{mm} \end{pmatrix}$; $\Omega = \begin{pmatrix} \sigma_{11} I & & 0 \\ & \ddots & \\ 0 & & \sigma_{mm} I \end{pmatrix}$,

Just estimate equations separately by OLS & get efficient estimates.

2: $X_1 = X_2 = \cdots = X_m = Z$ (say)

("Multivariate regression model".)

So, $X = \begin{bmatrix} Z & & 0 \\ & Z & \\ 0 & & Z \end{bmatrix} = (I_m \otimes Z)$

$$\hat{\beta} = [X'(\Sigma^{-1} \otimes I_T)X]^{-1} [X'(\Sigma^{-1} \otimes I_T)y]$$

$$= [(I_m \otimes Z)'(\Sigma^{-1} \otimes I)(I_m \otimes Z)]^{-1} \cdot [(I_m \otimes Z)'(\Sigma^{-1} \otimes I_T)y]$$

$$= (\Sigma^{-1} \otimes Z'Z)^{-1}(\Sigma^{-1} \otimes Z)'y$$

$$= (\Sigma \otimes (Z'Z)^{-1})(\Sigma^{-1} \otimes Z')y$$

$$= (I_m \otimes (Z'Z)^{-1}Z'y)$$

But this is just:

$$\hat{\beta} = \begin{bmatrix} b_1 & & 0 \\ & b_2 & \\ 0 & & \ddots \; b_m \end{bmatrix} \; ; \quad \text{OLS for each equation.}$$

What if Ω (or Σ) unknown?

Estimate Σ consistently – lots of ways of doing this. One possibility is:

* Estimate each equation by OLS & get

$\underset{\sim}{e}_1, \underset{\sim}{e}_2, ---, \underset{\sim}{e}_m$.

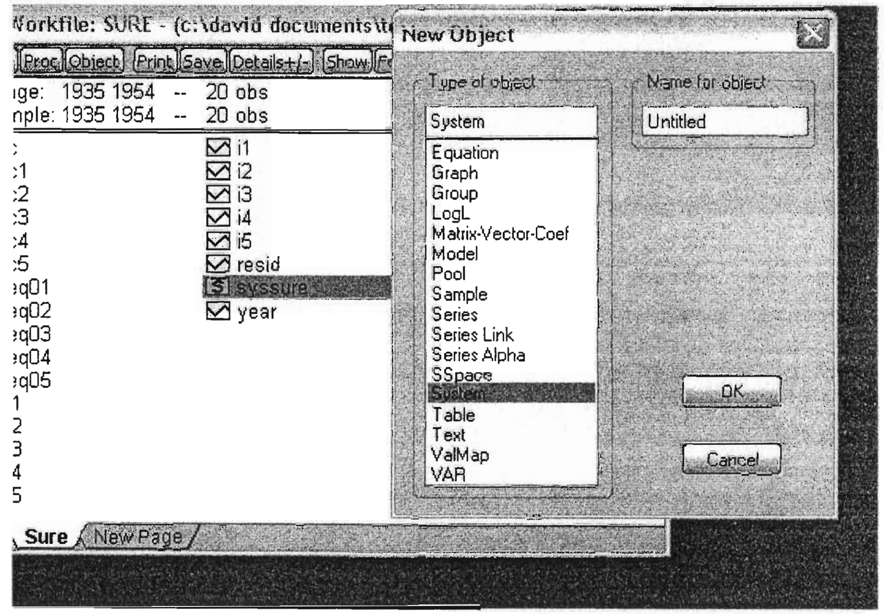
* Obtain $\hat{\sigma}_{ii} = \underset{\sim}{e}_i' \underset{\sim}{e}_i / T \qquad (i = 1, ---, m)$

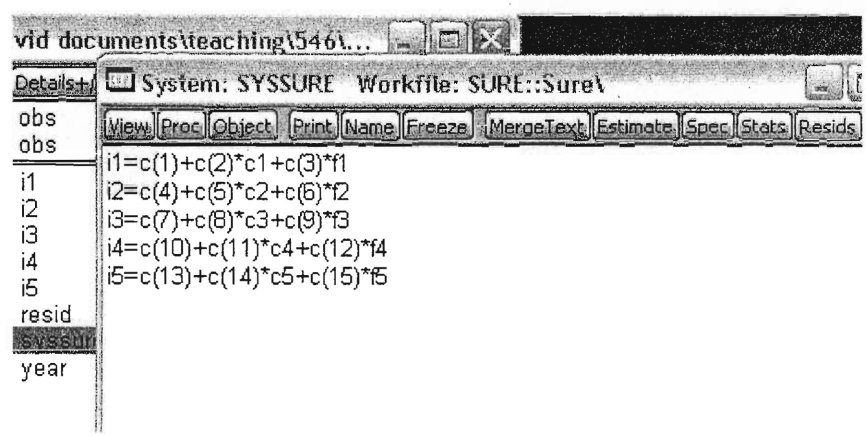
Obtain $\hat{\sigma}_{ij} = \underset{\sim}{e}_i' \underset{\sim}{e}_j / T \qquad (i \neq j)$

Replace Σ by

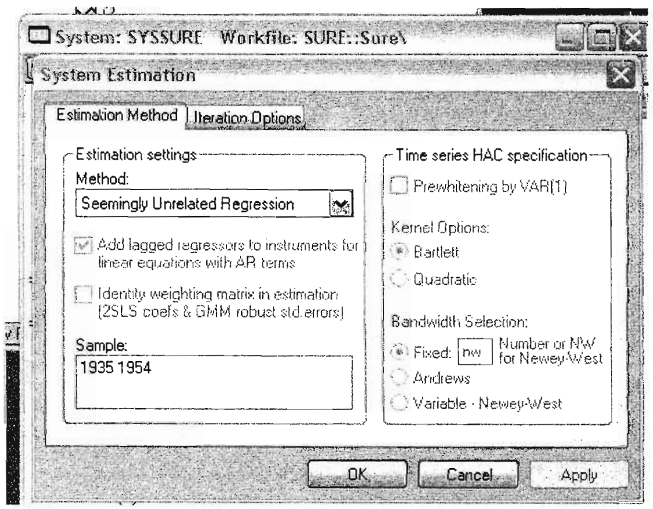
$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1m} \\ \vdots & & \ddots & \vdots \\ \hat{\sigma}_{m1} & \cdots & & \hat{\sigma}_{mm} \end{pmatrix}$$

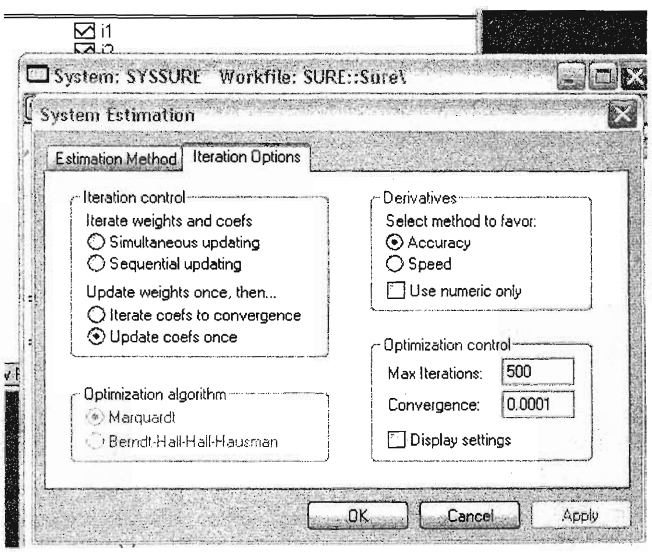
& then apply GLS.





9





10

System: SYSSURE
Estimation Method: Seemingly Unrelated Regression
Date: 03/09/07 Time: 08:51
Sample: 1935 1954
Included observations: 20
Total system (balanced) observations 100
Linear estimation after one-step weighting matrix

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-162.3641	89.45923	-1.814951	0.0731
C(2)	0.382746	0.032768	11.68047	0.0000
C(3)	0.120493	0.021629	5.570868	0.0000
C(4)	0.504304	11.51283	0.043804	0.9652
C(5)	0.308545	0.025864	11.92971	0.0000
C(6)	0.069546	0.016898	4.115732	0.0001
C(7)	-22.43891	25.51859	-0.879316	0.3817
C(8)	0.037291	0.012263	3.040936	0.0031
C(9)	0.130783	0.022050	5.931272	0.0000
C(10)	1.088877	6.258804	0.173975	0.8623
C(11)	0.057009	0.011362	5.017416	0.0000
C(12)	0.041506	0.041202	1.007400	0.3166
C(13)	85.42325	111.8774	0.763543	0.4473
C(14)	0.101478	0.054784	1.852344	0.0674
C(15)	0.399991	0.127795	3.129956	0.0024

Determinant residual covariance 6.18E+13

Equation: I1=C(1)+C(2)*C1+C(3)*F1
Observations: 20

R-squared	0.920742	Mean dependent var	608.0200
Adjusted R-squared	0.911417	S.D. dependent var	309.5746
S.E. of regression	92.13828	Sum squared resid	144320.9
Durbin-Watson stat	0.936490		

Equation: I2=C(4)+C(5)*C2+C(6)*F2
Observations: 20

R-squared	0.911862	Mean dependent var	86.12350
Adjusted R-squared	0.901493	S.D. dependent var	42.72556
S.E. of regression	13.40980	Sum squared resid	3056.985
Durbin-Watson stat	1.917509		

Equation: I3=C(7)+C(8)*C3+C(9)*F3
Observations: 20

R-squared	0.687636	Mean dependent var	102.2900
Adjusted R-squared	0.650887	S.D. dependent var	48.58450
S.E. of regression	28.70654	Sum squared resid	14009.12
Durbin-Watson stat	0.962757		

Equation: I4=C(10)+C(11)*C4+C(12)*F4
Observations: 20

R-squared	0.726429	Mean dependent var	42.89150
Adjusted R-squared	0.694244	S.D. dependent var	19.11019
S.E. of regression	10.56701	Sum squared resid	1898.249
Durbin-Watson stat	1.259005		

Equation: I5=C(13)+C(14)*C5+C(15)*F5
Observations: 20

R-squared	0.421959	Mean dependent var	405.4600
Adjusted R-squared	0.353954	S.D. dependent var	129.3519
S.E. of regression	103.9692	Sum squared resid	183763.0
Durbin-Watson stat	1.017982		

Maximum Likelihood Estimation:

In practice, iterating the GLS estimator between (conditional) estimates of β and Σ yields the MLE if the errors are Normal. (This what EViews does.)

We can also consider MLE directly.

The model is:

$$\underset{(TM\times 1)}{\underset{\sim}{y}} = \underset{(TM\times K)(K\times 1)}{X\underset{\sim}{\beta}} + \underset{(TM\times 1)}{\underset{\sim}{\varepsilon}} \quad ; K = \sum_i K_i$$

$$E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0} \quad ; \quad V(\underset{\sim}{\varepsilon}) = \Sigma \otimes I_T$$

& $\underset{\sim}{\varepsilon} \sim$ Normal : $\underset{\sim}{\varepsilon} \sim N[\underset{\sim}{0}, \Sigma \otimes I_T]$.

So, $L = (2\pi)^{-TM/2} |\Sigma \otimes I_T|^{-1/2}$

$$\cdot \exp\{-\tfrac{1}{2}(y - X\beta)'(\Sigma \otimes I_T)^{-1}(y - X\beta)\}$$

* So,

$$\log L = -\frac{TM}{2}\log(2\pi) - \frac{T}{2}\log|\Sigma|$$
$$- \tfrac{1}{2}(y - X\beta)'(\Sigma^{-1} \otimes I_T)(y - X\beta)$$

* As usual, we have a highly non-linear estimation problem, even though the model itself is linear in the parameters.
* Tests of restrictions on the coefficients of the model (within, or across, equations) are readily performed using the Wald test.
* With some effort, we can show that

$$\tilde{\log L} = \text{const.} - T/2 \log|\tilde{\Sigma}|$$

($|\tilde{\Sigma}|$ is reported in EViews output)

This is helpful, as it makes it easy to construct a LRT of the hypothesis that Σ is diagonal.

If Σ is diagonal, we just have a group of M separate linear regressions, & the MLE for each one is just OLS.

When the log-likelihood is evaluated at this restricted MLE, we have

$$\log \hat{L} = \text{const.} - T/2 \log |\hat{\Sigma}|$$

& so

$$LRT = T[\log|\hat{\Sigma}| - \log|\tilde{\Sigma}|]$$
$$= T \log[|\hat{\Sigma}| / |\tilde{\Sigma}|].$$

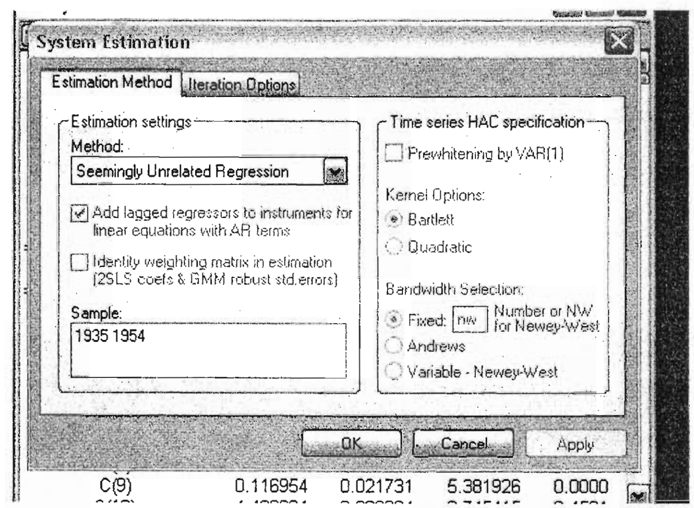
What does $|\hat{\Sigma}|$ look like? In the diagonal case, $\Sigma = \begin{pmatrix} \sigma_{11} & & 0 \\ & \ddots & \\ 0 & & \sigma_{mm} \end{pmatrix}$

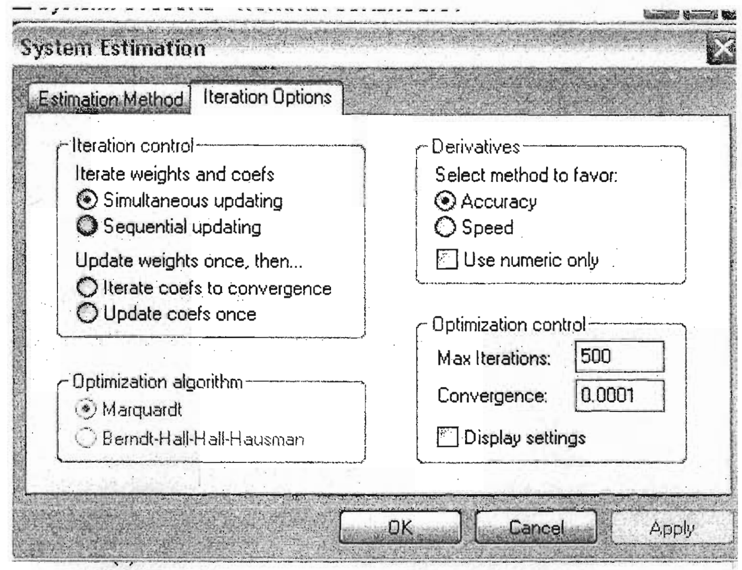
So, $|\hat{\Sigma}| = \prod_{i=1}^{m} \hat{\sigma}_{ii}$ & $\log|\hat{\Sigma}| = \sum_{i=1}^{m} \log \hat{\sigma}_{ii}$

where $\hat{\sigma}_{ii} = e_i' e_i / T$

& e_i is OLS residual vector of eq'n. i.

$LRT \xrightarrow{d} \chi^2_v$; $v = \frac{1}{2} M(M-1)$
(if H_0 is true)





System: SYSSURE
Estimation Method: Iterative Seemingly Unrelated Regression
Date: 03/09/07 Time: 08:54
Sample: 1935 1954
Included observations: 20
Total system (balanced) observations 100
Simultaneous weighting matrix & coefficient iteration
Convergence achieved after: 13 weight matrices, 14 total coef iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-173.0379	84.27963	-2.053141	0.0431
C(2)	0.389451	0.031852	12.22679	0.0000
C(3)	0.121953	0.020243	6.024448	0.0000
C(4)	2.378341	11.63135	0.204477	0.8385
C(5)	0.305066	0.026067	11.70320	0.0000
C(6)	0.067451	0.017102	3.943997	0.0002
C(7)	-16.37654	24.96084	-0.656089	0.5135
C(8)	0.037019	0.011770	3.145121	0.0023
C(9)	0.116954	0.021731	5.381926	0.0000
C(10)	4.488934	6.022064	0.745415	0.4581
C(11)	0.053861	0.010294	5.232286	0.0000
C(12)	0.026469	0.037038	0.714661	0.4768
C(13)	138.0101	94.60801	1.458757	0.1483
C(14)	0.088600	0.045278	1.956796	0.0537
C(15)	0.309302	0.117830	2.624987	0.0103

Determinant residual covariance 5.97E+13

Equation: I1=C(1)+C(2)*C1+C(3)*F1
Observations: 20

R-squared	0.919702	Mean dependent var	608.0200
Adjusted R-squared	0.910255	S.D. dependent var	309.5746
S.E. of regression	92.74073	Sum squared resid	146214.3
Durbin-Watson stat	0.936717		

Equation: I2=C(4)+C(5)*C2+C(6)*F2
Observations: 20

R-squared	0.910565	Mean dependent var	86.12350
Adjusted R-squared	0.900043	S.D. dependent var	42.72556
S.E. of regression	13.50807	Sum squared resid	3101.956
Durbin-Watson stat	1.885111		

Equation: I3=C(7)+C(8)*C3+C(9)*F3
Observations: 20

R-squared	0.669021	Mean dependent var	102.2900
Adjusted R-squared	0.630083	S.D. dependent var	48.58450
S.E. of regression	29.54949	Sum squared resid	14843.93
Durbin-Watson stat	0.898029		

Equation: I4=C(10)+C(11)*C4+C(12)*F4
Observations: 20

R-squared	0.701750	Mean dependent var	42.89150
Adjusted R-squared	0.666661	S.D. dependent var	19.11019
S.E. of regression	11.03336	Sum squared resid	2069.496
Durbin-Watson stat	1.124739		

Equation: I5=C(13)+C(14)*C5+C(15)*F5
Observations: 20

R-squared	0.390335	Mean dependent var	405.4600
Adjusted R-squared	0.318610	S.D. dependent var	129.3519
S.E. of regression	106.7753	Sum squared resid	193816.3
Durbin-Watson stat	0.967353		

View Proc Object Print Name Freeze MergeText Estimate Spec Stats Resids

System: SYSSURE
Estimation Method: Iterative Seemingly Unrelated Regression
Date: 03/09/07 Time: 08:56
Sample: 1935 1954
Included observations: 20
Total system (balanced)
Simultaneous weighting
Convergence achieved a

Wald Test

Coefficient restrictions separated by commas

c(2)=c(5)=c(8)=c(11)=c(14)

Examples
C(1)=0, C(3)=2*C(4)

OK Cancel

C(1)				
C(2)				
C(3)				
C(4)				
C(5)				
C(6)				
C(7)				
C(8)	0.037019	0.011770	3.145121	0.0023
C(9)	0.116954	0.021731	5.381926	0.0000
C(10)	4.488934	6.022064	0.745415	0.4581
C(11)	0.053861	0.010294	5.232286	0.0000
C(12)	0.026469	0.037038	0.714661	0.4768
C(13)	138.0101	94.60801	1.458757	0.1483
C(14)	0.088600	0.045278	1.956796	0.0537
C(15)	0.309302	0.117830	2.624987	0.0103

Determinant residual covariance 5.97E+13

17

View | Proc | Object | Print | Name | Freeze | MergeText | Estimate | Spec | Stats

Wald Test:
System: SYSSURE

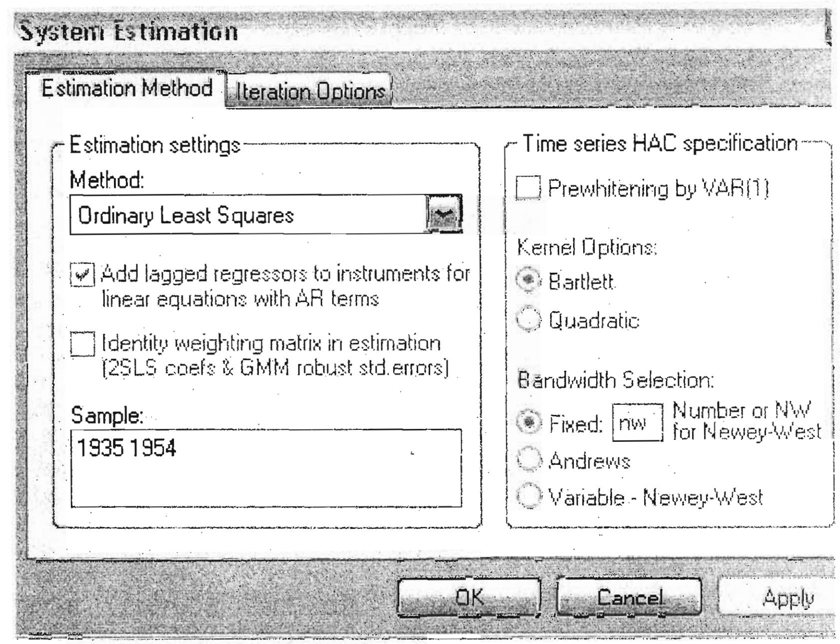
Test Statistic	Value	df	Probability
Chi-square	208.8502	4	0.0000

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(2) - C(14)	0.300851	0.052978
C(5) - C(14)	0.216466	0.052705
C(8) - C(14)	-0.051581	0.044123
C(11) - C(14)	-0.034739	0.042274

Restrictions are linear in coefficients.

18



19

System: SYSOLS Workfile: SURE::Sure\

View Proc Object Print Name Freeze MergeText Estimate Spec Stats Resids

System: SYSOLS
Estimation Method: Least Squares
Date: 03/09/07 Time: 16:01
Sample: 1935 1954
Included observations: 20
Total system (balanced) observations 100

$LRT = 20 \log\left(\frac{9.29}{5.97}\right)$

$= 8.844$

(d.f. = 10)

5% crit. = 18.31

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-149.7825	105.8421	-1.415150	0.1607
C(2)	0.371445	0.037073	10.01933	0.0000
C(3)	0.119281	0.025834	4.617173	0.0000
C(4)	-6.189961	13.50648	-0.458296	0.6479
C(5)	0.315718	0.028813	10.95743	0.0000
C(6)	0.077948	0.019973	3.902602	0.0002
C(7)	-9.956306	31.37425	-0.317340	0.7518
C(8)	0.026551	0.015566	1.705705	0.0917
C(9)	0.151694	0.025704	5.901548	0.0000
C(10)	-0.509390	8.015289	-0.063552	0.9495
C(11)	0.052894	0.015707	3.367658	0.0011
C(12)	0.092406	0.056099	1.647205	0.1032
C(13)	-30.36853	157.0477	-0.193371	0.8471
C(14)	0.156571	0.078886	1.984782	0.0504
C(15)	0.423866	0.155216	2.730808	0.0077

Determinant residual covariance 9.29E+13

Equation: I1=C(1)+C(2)*C1+C(3)*F1
Observations: 20

R-squared	0.921354	Mean dependent var	608.0200
Adjusted R-squared	0.912102	S.D. dependent var	309.5746
S.E. of regression	91.78167	Sum squared resid	143205.9
Durbin-Watson stat	0.937454		

Equation: I2=C(4)+C(5)*C2+C(6)*F2
Observations: 20

R-squared	0.913578	Mean dependent var	86.12350
Adjusted R-squared	0.903411	S.D. dependent var	42.72556