Exercises for Chapter 6 of Vinod's "HANDS-ON INTERMEDIATE ECONOMETRICS USING R"

H. D. Vinod Professor of Economics,
Fordham University, Bronx, New York 10458

Abstract

These are exercises to accompany the above-mentioned book having the URL: (http://www.worldscibooks.com/economics/6895.html). At this time all of the following exercises are suggested by H. D. Vinod (HDV) himself. Vinod invites readers to suggest new and challenging exercises (along with full and detailed answers) dealing with the discussion in Chapter 6 and related discussion from econometric literature by sending e-mail to vinod@fordham.edu. If the answers involve R programs, they must work. Readers may also suggest improvements to the answers and/or hints to existing exercises. If we include exercises and improvements suggested by readers, we promise to give credit to such readers by name. Furthermore, we will attach the initials of readers to individual exercises to identify the reader. Some R outputs are suppressed for brevity.

6 Exercises Mostly Based on Chapter 6 (Simultaneous Equations Models) of the text

6.1 Exercise (Haavelmo Model)

Use the following initial definitions of data for consumption and income and re-estimate snippets 6.1.1 and 6.1.2. Compare the results to the ones in the text.

```
rm(list=ls()) #rm means remove them all to clean up old stuff
print(paste("Following executed on", date()))
#date-time stamp
library(strucchange)
data("USIncExp")
attach(as.data.frame(USIncExp))
C=expenditure
Y=income
summary(cbind(C,Y))
```

6.2 Exercise (Klein I model)

Estimate the Klein I model by various simultaneous equation estimation methods including OLS, SUR and 2SLS by using Henningsen and Hamann's R package [2] called 'systemfit.' Write the system in terms of 3 equations only after substituting out the identities. Use the package 'sem' [1] to estimate the models by LIML and FIML methods and compare them with the iterated methods suggested in Appendix C of the Vignette accompanying the package 'systemfit' for the Klein I data. New version of 'sem' omits LIML and FIML.

```
library(systemfit)
data("KleinI")
eqConsump = consump ~ corpProf + corpProfLag + wages
eqInvest = invest ~ corpProf + corpProfLag + capitalLag
eqPrivWage = privWage ~ gnp + gnpLag + trend
inst = ~govExp + taxes + govWage + trend + capitalLag + corpProfLag +
 gnpLag
system = list(Consumption = eqConsump, Investment = eqInvest,
 PrivateWages = eqPrivWage)
#OLS estimation:
kleinOls = systemfit(system, data = KleinI)
round(coef(summary(kleinOls)), digits = 3)

#2SLS estimation:
klein2sls = systemfit(system, method = "2SLS", inst = inst,
 data = KleinI, methodResidCov = "noDfCor")
rround(coef(summary(klein2sls)), digits = 3)
```

```
#3SLS estimation:
klein3sls = systemfit(system, method = "3SLS", inst = inst,
 data = KleinI, methodResidCov = "noDfCor")
rround(coef(summary(klein3sls)), digits = 3)

#Iterated 3SLS estimation:
kleinI3sls = systemfit(system, method = "3SLS", inst = inst,

 data = KleinI, methodResidCov = "noDfCor", maxit = 500)
rround(coef(summary(kleinI3sls)), digits = 3)
```

Use appendix C of the Vignette to get the necessary code for LIML and FIML comparisoins.

6.3 Exercise (Food market demand supply model)

Download the vignette accompanying the R package called 'systemfit' and use the data called 'Kmenta.' Estimate the demand equation and supply equation with one exogenous variable in each equation. Estimate the model with 'seemingly unrelated regressions' (SUR), two and three stage least squares method. Indicate the instrumental variables used. Note which variable is 'absent' in which equation. Use the following notation used in the vignette which comes with the 'systemfit' package for variable names and coefficients, when appropriate.

consump = β_1 + β_2 price + β_3 income

consump = β_4 + β_5 price + β_6 farmPrice +β_7 trend

```
library("systemfit")
vignette("systemfit") #a beautiful pdf file with useful info
data("Kmenta")
attach(Kmenta)
eqDemand = consump ~ price + income #demand eq exo income
eqSupply = consump ~ price + farmPrice + trend #supp eq
eqSystem = list(demand = eqDemand, supply = eqSupply)
#above defines the equation system with a list
fitols = systemfit(eqSystem)
#when no method is specified it gives OLS
print(fitols)
```

```
b.ols=coef(fitols)
fitsur = systemfit(eqSystem, method = "SUR")
summary(fitsur, residCov = FALSE, equations = FALSE)
#these choices give compact output shown below
b.sur=coef(fitsur)
#above is SUR method
fit3sls = systemfit(eqSystem, method = "3SLS",
  inst = ~income +  farmPrice + trend)
b.3sls=coef(fit3sls)
fit3sls2 = systemfit(eqSystem, method = "3SLS",
  inst = list(~farmPrice +
  trend, ~income + farmPrice + trend))
b.3sls2=coef(fit3sls2)
options(digits=4)  #allows compact printing
rbind(b.ols, b.sur, b.3sls, b.3sls2)
```

The income variable is absent from the supply equation and farmPrice and trend variables are absent from the demand equation. These absences permit identification of the system of equations. We have set up the estimated coefficients of all models into a comparable table by using the 'rbind' command to bind the rows together, while retaining the suitable column headings.

```
> summary(fitsur, residCov = FALSE, equations = FALSE)
systemfit results
method: SUR

        N DF SSR detRCov OLS-R2 McElroy-R2
system 40 33 170   0.879  0.683      0.789

        N DF   SSR  MSE RMSE    R2 Adj R2
demand 20 17  65.7 3.86 1.97 0.755  0.726
supply 20 16 104.1 6.50 2.55 0.612  0.539


Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
demand_(Intercept)  99.3329     7.5145   13.22  2.3e-10 ***
demand_price        -0.2755     0.0885   -3.11  0.00633 **
```

```
demand_income        0.2986     0.0419    7.12  1.7e-06 ***
supply_(Intercept)  61.9662    11.0808    5.59  4.0e-05 ***
supply_price         0.1469     0.0944    1.56  0.13941
supply_farmPrice     0.2140     0.0399    5.37  6.3e-05 ***
supply_trend         0.3393     0.0679    5.00  0.00013 ***
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1

>rbind(b.ols, b.sur, b.3sls, b.3sls2)
        demand_(Intercept) demand_price demand_income supply_(Intercept)
b.ols                99.90      -0.3163        0.3346              58.28
b.sur                99.33      -0.2755        0.2986              61.97
b.3sls               94.63      -0.2436        0.3140              52.20
b.3sls2             243.68      -1.5685        0.1446              49.60
        supply_price supply_farmPrice supply_trend
b.ols         0.1604           0.2481       0.2483
b.sur         0.1469           0.2140       0.3393
b.3sls        0.2286           0.2282       0.3611
b.3sls2       0.2394           0.2555       0.2529
```

6.4 Exercise (Notation System)

Describe the notation system used in the discussion of simultaneous equation models with the help of a concrete example of the food market demand supply model used in an earlier exercise. Write out all matrices and vectors with actual data numbers for all equations y_j from (6.1.4) of the text. Explicitly show all $Y_j^*, X_j^*, Y_j, X_j, Z_j$ matrices and vectors as the case may be for all j. Since most vectors and matrices are large, always include initial 3 rows and ending 3 rows with dot dot dot in between to represent intermediate rows. Be sure to include all important subcomponents to indicate that you understand the composition of the vectors and matrices such as those in eq. (6.1.10) of the text. State the values of M_j, K_j, M_j^*, K_j^* also. State which equation is over, just or under identified based on order condition (Hint: based on the criterion in (6.4.11) of the text).

ANSWER: (Hint: Sec. 6.1.1 has items (i) to (xiii)) Previous question provides us with the equations of the model:

From the description of the data (Kmenta) we know that there are 20 observations and that the exogenous variables are income, farmPrice, and trend; the endogenous variables are price and consump were generated by simulation.

	consump	price	income	farmPrice	trend
1	98.485	100.323	87.4	98.0	1
2	99.187	104.264	97.6	99.1	2
3	102.163	103.435	96.7	99.1	3
4	101.504	104.506	98.2	98.1	4
5	104.240	98.001	99.8	110.8	5
6	103.243	99.456	100.5	108.2	6
7	103.993	101.066	103.2	105.6	7
8	99.900	104.763	107.8	109.8	8
9	100.350	96.446	96.6	108.7	9
10	102.820	91.228	88.9	100.6	10
11	95.435	93.085	75.1	81.0	11
12	92.424	98.801	76.9	68.6	12
13	94.535	102.908	84.6	70.9	13
14	98.757	98.756	90.6	81.4	14
15	105.797	95.119	103.1	102.3	15
16	100.225	98.451	105.1	105.0	16
17	103.522	86.498	96.4	110.5	17
18	99.929	104.016	104.4	92.5	18
19	105.223	105.769	110.7	89.3	19
20	106.232	113.490	127.1	93.0	20

There are two endogenous variables consump and price implying that M=2 and hence there are 2 equations in the system of equations. Let us discuss the notation for M=2 equations.

Subscript j=1 for the first equation:

$$consump = \beta_1 + \beta_2 * price + \beta_3 * income + e_1$$

This equation has $M_1 = 1$ since there is one endogenous variable on the RHS of this equation. It has $K_1 = 2$ or two exogenous variables on the RHS. Let ι denote a column of ones used to generate the intercept. Then the two exogenous variables are ι and income.

Based on (6.1.4) (page 263) of the text, we want to write this as:

$$y_1 = Y_1\gamma_1 + X_1\beta_1 + u_1 = Z_1\delta_1 + u_1$$

There is some unavoidable ambiguity, since the notation β_1 in (6.1.4) of the text is distinct from the same notation used for the demand equation intercept.

Lower case y_1 for consump and upper case Y_1 for price are 20×1 vectors. γ_1 is a scalar containing β_2.

Upper case X_1 for included exogenous variables is the 20×2 matrix containing two columns: ι and income. β_1 is a 2×1 column vector $(\beta_3, \beta_1)'$ in the notation of the first (demand) equation. Finally $u_1 = e_1$.

Now we turn to starred or omitted exogenous variables from the RHS. $K_1^* = 2$ for priceFarm and trend. Hence X_1^* is 20×2 matrix.

Using the data we have:

$$\begin{bmatrix} 98.485 \\ 99.187 \\ 102.163 \\ \vdots \\ 99.929 \\ 105.223 \\ 106.232 \end{bmatrix} = \begin{bmatrix} 100.323 \\ 104.264 \\ 103.435 \\ \vdots \\ 104.016 \\ 105.769 \\ 113.490 \end{bmatrix} \begin{bmatrix} \beta_2 \end{bmatrix} + \begin{bmatrix} 87.4 & 1 \\ 97.6 & 1 \\ 96.7 & 1 \\ \vdots & \\ 104.4 & 1 \\ 110.7 & 1 \\ 127.1 & 1 \end{bmatrix} \begin{bmatrix} \beta_3 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_{1.1} \\ e_{1.2} \\ e_{1.3} \\ \vdots \\ e_{1.18} \\ e_{1.19} \\ e_{1.20} \end{bmatrix}$$

Subscript j=2 for the second equation:

$$consump = \beta_4 + \beta_5 * price + \beta_6 * farmPrice + \beta_7 * trend + e_2$$

This equation has $M_2 = 1$ since there is one endogenous variable (price) on the RHS of this equation. It has $K_2 = 3$ or three exogenous variables on the RHS: ι, farmPrice and trend.

Based on (6.1.4) of the text, (page 263) want to write this as:

$$y_2 = Y_2\gamma_2 + X_2\beta_2 + u_2 = Z_2\delta_2 + u_2$$

Lower case y_2 = consump is 20×1, upper case Y_2 for price is also 20×1. γ_2 is a scalar containing β_5

Upper case X_2 for included exogenous variables is the 20×3 matrix containing three columns: ι, farmPrice and trend. β_2 is a 3×1 column vector $(\beta_6, \beta_7, \beta_4)'$ in the notation of the second (supply) function. Finally $u_2 = e_2$.

Now we turn to starred or omitted exogenous variables from the RHS. $K_2^* = 1$ for income. Hence X_2^* is 20×1 vector.

In terms of data the second (supply) equation looks like:

$$\begin{bmatrix} 98.485 \\ 99.187 \\ 102.163 \\ \vdots \\ 99.929 \\ 105.223 \\ 106.232 \end{bmatrix} = \begin{bmatrix} 100.323 \\ 104.264 \\ 103.435 \\ \vdots \\ 104.016 \\ 105.769 \\ 113.490 \end{bmatrix} \begin{bmatrix} \beta_5 \end{bmatrix} + \begin{bmatrix} 98.0 & 1 & 1 \\ 99.1 & 2 & 1 \\ 99.1 & 3 & 1 \\ \vdots & & \\ 92.5 & 18 & 1 \\ 89.3 & 19 & 1 \\ 93.0 & 20 & 1 \end{bmatrix} \begin{bmatrix} \beta_6 \\ \beta_7 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} e_{2.1} \\ e_{2.2} \\ e_{2.3} \\ \vdots \\ e_{2.18} \\ e_{2.19} \\ e_{2.20} \end{bmatrix}$$

The Z_1 is 20×3 matrix with $M_1 + K_1 = 1 + 2 = 3$ columns $Z_1 = [Y_1 | X_1]$. Similarly, Z_2 is the 20×4 matrix with $M_2 + K_2 = 1 + 3 = 4$ columns. or $Z_2 = [Y_2 | X_2]$. We omit writing them explicitly with numbers for brevity.

Since $2 = K_1^* \geq M_1 = 1$ demand equation is overidentified according to equation (6.4.11) of the text. Also, since $1 = K_2^* \geq M_2 = 1$, the second (supply) equation is exactly or just identified.

6.5 Exercise (Structure and Reduced Form)

Use an extended Haavelmo model after including the government expenditures G as additional endogenous variable explained by an exogenous time trend and write out the structural equations as in (6.1.17) and its reduced form (6.1.20) of the text. Show the revised matrices as is done in equations (6.1.18) and (6.1.19) and the Π matrix in (6.1.21) with symbols, not numbers.

6.6 Exercise (Structure, Reduced Form Notation)

Use the following R code to define 3 endogenous and 3 exogenous variables where we denote the (dot dot dot) by -99. Assume that T=50 is the sample size.

```
en1=c(3,4,-99,9)
en2=c(4,6,-99,6)
```

```
en3=c(1,2,-99,19)
ex1=c(7,8,-99,20)
ex2=c(9,12,-99,21)
ex3=c(13,14,-99,22)
```

Structural equations are given as:

en1= a + b en3 + q ex3

en3= c + d ex1 + w en1

en2= e + f ex1 + g ex2 + h ex3 + k en3

Comment on the endogeneity problem if the term 'w en1' were absent from the second equation.

Write out all matrices and vectors with actual data numbers for all equations y_j from (6.1.4) of the text. Explicitly show all $Y_j^*, X_j^*, Y_j, X_j, Z_j$ matrices and vectors as the case may be for all j. Be sure to include all important subcomponents to indicate that you understand the composition of the vectors and matrices such as those in eq. (6.1.10) of the text. State the values of M_j, K_j, M_j^*, K_j^* also. State which equation is over, just or under identified based on order condition (Hint: based on the criterion in (6.4.11) of the text). Write the structure as (6.1.17) of the text and the reduced form as (6.1.21) of the text.

ANSWER: If the term 'w en1' were absent, there is no endogeneity problem if we simply substitute the RHS of second equation on the right sides of the first and third equation. After such substitution all equations have only exogenous variables on the RHS.

However, since the term 'w en1' is present, follow the methods in the answer to an earlier question on notation system.

6.7 Exercise (Structure and Reduced Form)

Use the food market demand supply model data to explicitly write in terms of data the structural equations in (6.1.17) and its reduced form (6.1.20) of the text. Explicitly show the matrices as is done in equations (6.1.18) and (6.1.19) and the Π matrix in (6.1.21) with data numbers, not symbols.

6.8 Exercise (Assumptions)

Describe the assumptions and simultaneous equations bias with the help of Kmenta's model or another concrete example. (Hint: assumptions are in Sec.

6.1.5 and bias is in Sec. 6.1.2)

6.9 Exercise (GIV estimator)

Is the GLS estimator always superior to OLS? Discuss the difference between the IV and GIV estimators. Write the formula for the covariance matrix of the latter. (Hint: see pages 272-275.)

6.10 Exercise (k-class and LIML)

Describe the various special cases of the k-class estimator. Discuss the eigenvalues and eigenvectors of a matrix involved in LIML estimation. What choice of k in k-class estimator in (eq. 6.3.3) yields the LIML estimator? How is the k for the LIML related to the k of the 2SLS estimator?

ANSWER: First, note that LIML is a special case of the k-class estimator with k equal to the smallest eigenvalue of (6.3.4). The 2SLS is also a special case of k-class when k=1. Now we further explain the LIML theory. Consider jth structural equation, (6.1.4): $y_j = Y_j\gamma_j + X_j\beta_j + u_j$. Using the notation on page 281 define an artificial variable $Y_{0j} = (y_j, Y_j)$ having an artificial parameter vector $\gamma_{0j} = (-1, \gamma)'$. Now rewrite the equation in terms of the artificial variable as: $Y_{0j}\gamma_{0j} + X_j\beta_j + u_j$. Anderson-Rubin result from 1950's that maximization of the likelihood for the $j-th$ equation is equivalent to minimization of a 'variance ratio' of two error sum of squares (ESS). The numerator of the ratio has the ESS when the artificial variable is regressed on X_j. The denominator has a similar ESS when the artificial variable is regressed on the entire matrix of all exogenous variables from all equations, $X = [X_j\, X_j^*]$. It is well known that when we add regressors, the ESS cannot decrease. Hence the ESS in the numerator is $\geq$ ESS in the denominator. These ESS in turn depend on projection (hat) matrices $H_{Xj} = X_j(X_j'X_j)^{-1}X_j'$ and $H_X = X(X'X)^{-1}X'$. Recall that any residual is identity minus the hat matrix times y. Thus, it can be shown that the variance ratio to be minimized is defined as:

$$\rho = \frac{\gamma_{0j}'W_{0j}\gamma_{0j}}{\gamma_{0j}'W_{1j}\gamma_{0j}} \qquad (1)$$

where the matrices W_{0j} and W_{1j} are as defined on page 281 by using the projection matrices. The ρ is a ratio of quadratic forms in the parameter vector γ_{0j}. Minimizing ρ with respect to the artificial parameter γ_{0j} leads to the first order condition: $(W_{0j} - \lambda W_{1j})\gamma_{0j} = 0$. A nontrivial solution cannot exist unless the matrix $(W_{0j} - \lambda W_{1j})$ is singular. Hence the determinant $|W_{0j} - \lambda W_{1j}| = 0$ must hold. The solution is a minimum provided λ is the smallest eigenvalue of the $(W_{1j})^{-1}W_{0j}$ matrix.

Since the ESS in the numerator is $\geq$ ESS in the denominator, the minimum eigenvalue $\lambda_{min} \geq 1$. Hence the k for 2SLS is no greater than the k for LIML.

6.11 Exercise (LIML for food market model)

Describe the food market demand model and its LIML estimation Explicitly show the matrices involved for eigenvalues and eigenvectors of in LIML estimation. Explicitly show the numbers for W_{0j}, W_{1j} matrices with sufficient detail showing that you understand their composition.

6.12 Exercise (Identification)

Describe the 2×3 submatrices of the Π matrix involved in identification of a system of simultaneous equations. Discuss the role of the rank of Π_{21} and conditions for overidentification.

6.13 Exercise (Identification for food market model)

Use the numerical data and fill with numbers each of the matrices in eq. (6.4.5) involved in identification of the food market model.

6.14 Exercise (Wold Recursive Model)

Describe the Cobweb model for agricultural supply as an example of the Wold recursive system. Describe the covariance matrix of the stacked system by using Kronecker product of matrices. (Hint: see eq. 6.5.2)

6.15 Exercise (3SLS as feasible GLS)

Show that 3SLS estimator is a feasible GLS estimator applied to the entire system of simultaneous equations. Discuss the efficiency properties of 3SLS.

6.16 Exercise (3SLS for food market model)

Compute the 3SLS estimator for the food market model explicitly showing what system is being estimated. Also apply the Hausman specification test for the validity of 3SLS.

Hint: The answer is provided in the context of answers to a later question.

6.17 Exercise (Iterated 3SLS for food market model)

Compute the iterated 3SLS estimator for the food market model explicitly showing what system is being estimated and how it is iterated.

Hint: The answer is provided in the context of answers to a later question.

6.18 Exercise (Testing Restrictions for the food market model)

Recall the specification from the vignette used in the package 'systemfit' available by using the command "vignette("systemfit")".

consump = β_1 + β_2 price + β_3 income

consump = β_4 + β_5 price + β_6 farmPrice +β_7 trend

Now test the cross-equation linear restriction that the price coefficient in the demand equation is the negative of the coefficient of 'farmPrice' in the supply equation. That is: $\beta_2 + \beta_6 = 0$. Would the F test from Chapter 3 apply here? Report the results of both that F test and the correct likelihood ratio test. Also apply the Hausman specification test for the validity of 3SLS.

Hint: Follow the method discussed in the vignette. It involves first defining the restriction and then fitting the 2SLS model by two methods, 2SLS and restricted 2SLS. Then apply the F test from eq. (3.2.17) from the text.

$$F_{(r,df)} = [(\text{ReRSS} - \text{UnRSS})/r]/[\text{UnRSS})/(df_{unr})], \tag{2}$$

where (df_{unr}) denotes the degrees of freedom (df) of the unrestricted model. Recall that in single equation models we have ReRSS > UnRSS, that is, restrictions always increase the residual sum of squares (RSS) or worsen the fit. The testing issue is whether that worsening is statistically significant or the test statistic exceeds the critical value from F table with the indicated df. However this strategy from Chapter 3 does not work for simultaneous equations and we need to consider log of the determinant of the covariance matrix of residuals (not just sum of squares) of the two models.

The correct strategy as explained in Henningsen and Hamann's vignette and in advanced Econometric texts is to use the following likelihood ratio test statistic:

$$LR = T(log|\hat{\Sigma}_{Restr}| - log|\hat{\Sigma}_{UnRestr}|), \quad (3)$$

where T is the sample size, and $\hat{\Sigma}$ refer to covariance matrices of residuals of all equations with the subscript 'Restr' for the restricted model and 'UnRestr' for the unrestricted model. It is known that the likelihood ratio is asymptotically distributed as a Chi-square random variable. The package 'systemfit' does this calculation conveniently by the function 'lrtest' for likelihood ratio test as shown below and reports the p-value of the Chi-square test, so one need not look up any Tables for the Chi-square.

```
rm(list=ls()) #rm means remove them all to clean up old stuff
library(systemfit)
data("Kmenta")
attach(Kmenta)
eqDemand = consump ~ price + income #demand eq exo income
eqSupply = consump ~ price + farmPrice + trend #supp eq
eqSystem = list(dd = eqDemand, sup = eqSupply)
#above defines the equation system with a list
fitols = systemfit(eqSystem)
#when no method is specified it gives OLS
print(fitols)
b.ols=coef(fitols)
fit3sls = systemfit(eqSystem, method = "3SLS",
  inst = ~income +  farmPrice + trend)
b.3sls=coef(fit3sls);summary(fit3sls)
fitI3sls = systemfit(eqSystem, method = "3SLS",
inst = ~income + farmPrice + trend, data = Kmenta,
maxit = 250)
b.Iter3sls=coef(fitI3sls);summary(fitI3sls)
fit2sls = systemfit(eqSystem, method = "2SLS", inst = ~income +
 farmPrice + trend, data = Kmenta)
b.2sls=coef(fit2sls);summary(fit2sls)
restrict = "dd_price + sup_farmPrice = 0"
fit2slsRmat = systemfit(eqSystem, method = "2SLS",
inst = ~income +  farmPrice + trend, data = Kmenta,
```

```
 restrict.matrix = restrict)
b.restr.2sls=coef(fit2slsRmat);summary(fit2slsRmat)
options(digits=4)  #allows compact printing
rbind(b.ols, b.3sls, b.Iter3sls, b.2sls,b.restr.2sls)
```

Note that the command 'eqSystem = list(dd = eqDemand, sup = eqSupply)' specifies the identifier 'dd' for the demand equation and 'sup' for the supply equation. These are used by the software to identify the coefficients of the same regressor (say price) appearing in many equations. It is good to choose short equation identifies, since they take up space in printing of results. The outputs of 'summary' functions are suppressed for brevity. The coefficients of all models are collected as objects with self descriptive names starting with 'b.' for coefficients. These outputs show that iterated 3SLS is close to 3SLS and that the restricted 2SLS model does satisfy the restriction $\beta_2 + \beta_6 = 0$. The coefficients 'dd_price' is 0.2540 and that of 'sup_farmPrice' equals -0.2540.

```
             dd_(Intercept) dd_price dd_income sup_(Intercept) sup_price
b.ols                 99.90  -0.3163    0.3346           58.28    0.1604
b.3sls                94.63  -0.2436    0.3140           52.20    0.2286
b.Iter3sls            94.63  -0.2436    0.3140           52.66    0.2266
b.2sls                94.63  -0.2436    0.3140           49.53    0.2401
b.restr.2sls          95.39  -0.2540    0.3170           49.77    0.2394
             sup_farmPrice sup_trend
b.ols               0.2481    0.2483
b.3sls              0.2282    0.3611
b.Iter3sls          0.2234    0.3800
b.2sls              0.2556    0.2529
b.restr.2sls        0.2540    0.2519
```

The code above shows how the package 'systemfit' expects us to state the restrictions. The variable name 'price' is prefixed with 'dd' and an underscore by the software itself. It identifies the coefficient of 'price' in the demand equation. Similarly, for the coefficient of 'farmPrice.' The restriction is conveyed to the software by the intuitive command: '(restrict = "dd_price + sup_farmPrice = 0")' above. It is interesting that the restricted RSS is not larger than unrestricted RSS here for cross equation restrictions. In fact the single equation F statistic of Chapter 3 is meaningless (negative) here. The correct approach involves a likelihood ratio test explained in the software vignette.

```
UnRSS=sum(resid(fit2sls)^2)# 162.4
ReRSS=sum(resid(fit2slsRmat)^2)# 161.7
df=summary(fit2sls)$df[2]  #7, 33
#summary(fit2slsRmat)$df  # 6, 34
r=1
Fstat=((ReRSS-UnRSS)/r ) /(UnRSS/df );Fstat
lrtest(fit2sls, fit2slsRmat)
```

The following output of the likelihood ratio test by using the function 'lrtest' does not reject the null hypothesis, (p-value of $0.95 > 0.05$) supporting equality of two price coefficients with opposite signs, i.e., the null: $\beta_2 + \beta_6 = 0$.

```
Likelihood ratio test

Model 1: fit2sls
Model 2: fit2slsRmat
  #Df LogLik Df  Chisq Pr(>Chisq)
1   8  -67.6
2   7  -67.6 -1 0.0036       0.95
```

Note that the observed Chi-square statistic 0.0036 is obviously 'small.' The p-value is given under the heading 'Pr(>Chisq),' as is appropriate in the current context.

The Hausman specification test is designed to test the appropriateness of 3SLS itself; and is readily implemented in R as follows:

```
fit2sls = systemfit(eqSystem, method = "2SLS", inst = ~income +
 farmPrice + trend, data = Kmenta)
fit3sls = systemfit(eqSystem, method = "3SLS", inst = ~income +
 farmPrice + trend, data = Kmenta)
hausman.systemfit(fit2sls, fit3sls)
```

The output from the above code is as follows:

```
Hausman specification test for consistency of the 3SLS estimation

data:  Kmenta
Hausman = 2.536, df = 7, p-value = 0.9244
```

Since the p-value exceeds 0.05 we do not reject the null hypothesis that the instrumental variables of each equation are uncorrelated with the disturbance terms of all other equations. That is, we conclude that both 2SLS and 3SLS are consistent, but 3SLS is asymptotically more efficient here. Henningsen and Hamann's 'systemfit' package is indeed a very powerful tool.

6.19 Exercise (Likelihood Function and FIML)

Consider stacked system of M equations. For example, if $M = 3$ we write

$$y = Z\delta + \varepsilon = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} Z_1 & 0 & 0 \\ 0 & Z_2 & 0 \\ 0 & 0 & Z_3 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} + \begin{bmatrix} \in_1 \\ \in_2 \\ \in_3 \end{bmatrix}. \tag{4}$$

Now visualize the general case for any $M \neq 3$ as $y = Z\delta + \varepsilon$. What are the dimensions of various matrices. Also, what is the dimension of the covariance matrix $E\varepsilon\varepsilon'$? Now assume that the errors ε are normally distributed, and write the likelihood function using eq. (10.1.6) on page 421. How would you formulate a maximum likelihood problem for the coefficients δ of all equations together (FIML)?

ANSWER: For the jth structural equation, $y_j = Z_j\delta_j + u_j$, $u_j \sim N(0, \sigma_{jj} I)$, that is, the error covariance matrix is assumed to be $E(u_j u_j') = \sigma^2 I_T$, where I_T denotes the $T \times T$ identity matrix. We write M equations after stacking the $T \times 1$ vector y_1 on top of y_2 on top of y_3, and finally on top of y_M as $y = Z\delta + \varepsilon$. Both y and ε are $MT \times 1$ vectors due to stacking. The dimension of the covariance matrix $E\varepsilon\varepsilon'$ is $MT \times MT$. Note that $E(u_{j'} u_j') = \Omega$, where $j \neq j'$ or covariance matrix across equation errors. Actually, it is given by

$$E\varepsilon\varepsilon' = V_\varepsilon = \Omega \otimes I_T, \tag{5}$$

written in Kronecker product notation $\otimes$. Now turn to the log likelihood function (LL) Writing eq. (10.1.6) in the current notation we have:

$$\text{LL} = -(T/2)\log(2\pi\sigma^2) - 0.5\log|V_\varepsilon| - (2\sigma^2)^{-1}(y - Z\delta)' V_\varepsilon^{-1}(y - Z\delta). \tag{6}$$

The first order condition for maximization of the LL is with respect to δ

is

$$\frac{\partial LL}{\partial \delta} = 0 = Z'V_{\varepsilon}^{-1}(y - Z\delta). \tag{7}$$

Note that we need to replace $V_{\varepsilon} = \Omega \otimes I_T$, by an observable quantity. In other words we need to estimate Ω. We evaluate the first order condition $\frac{\partial LL}{\partial \Omega} = 0$ which leads to the following estimate of Ω

$$\hat{\Omega} = \frac{1}{T}(y_j - Z_j\hat{\delta}_j)'(y_j - Z_j\hat{\delta}_j). \tag{8}$$

It is simply the covariance matrix of residuals for each equation of the system of M equations.

For example, if $M = 2$ and $T = 3$ for brevity,

$$\Omega = \begin{bmatrix} 9.58 & -3.06 \\ -3.06 & 4.53 \end{bmatrix} \tag{9}$$

$$\Omega \otimes I_3 = \begin{bmatrix} 9.58 & 0.00 & 0.00 & -3.06 & 0.00 & 0.00 \\ 0.00 & 9.58 & 0.00 & 0.00 & -3.06 & 0.00 \\ 0.00 & 0.00 & 9.58 & 0.00 & 0.00 & -3.06 \\ -3.06 & 0.00 & 0.00 & 4.53 & 0.00 & 0.00 \\ 0.00 & -3.06 & 0.00 & 0.00 & 4.53 & 0.00 \\ 0.00 & 0.00 & -3.06 & 0.00 & 0.00 & 4.53 \end{bmatrix} \tag{10}$$

6.20 Exercise (Notation System for reduced form)

Consider a system of two equations with T observations, denoting endogenous variables by y and exogenous variables by x with subscripts:

$y_1 = a + bx_1 + cx_2 + fy_2 + \epsilon_1$

$y_2 = d + ey_1 + \epsilon_2$

Write the dimensions and content of X_2^* and X_2 matrices. Write out the Γ and B matrices of the structure.

ANSWER: X_2^* is $T \times 2$ containing x_1, x_2 and X_2 is a $T \times 1$ matrix. In order to find the other two matrices we write the model carefully as $Y\Gamma + XB = U$ or as

$$
\begin{aligned}
y_1 - f y_2 \quad & -a - b x_1 - c x_2 && = \epsilon_1 && (11)\\
-e y_1 + y_2 \quad & -d && = \epsilon_2 && (12)
\end{aligned}
$$

Now remember we have row-column multiplication. Hence we have:

$$
\Gamma = \begin{bmatrix} 1 & -e \\ -f & 1 \end{bmatrix}, \tag{13}
$$

and

$$
B = \begin{bmatrix} -a & -d \\ -b & 0 \\ -c & 0 \end{bmatrix}. \tag{14}
$$

6.21 Exercise (SUR and Kronecker Products)

Consider a stacked system of M equations with T observations, illustrated for $M = 3$ as:

$$
y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} Z_1\delta_1 \\ Z_2\delta_2 \\ Z_3\delta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}. \tag{15}
$$

Now apply OLS to the entire system. What is the dimension of the variance-covariance matrix of these errors? Write it out in terms of Kronecker product notation. What is the name of the GLS estimator applied to this system?

ANSWER: Define a new Z as stacking of Z_1 on top followed by Z_2 and Z_3 The OLS estimator is $\hat{\delta} = (Z'Z)^{-1}Z'y$. The covariance matrix is $MT \times MT$ and is written as:

$$
E(\epsilon\epsilon') = \begin{bmatrix} E\epsilon_1\epsilon_1' & E\epsilon_1\epsilon_2' & E\epsilon_1\epsilon_3' \\ E\epsilon_2\epsilon_1' & E\epsilon_2\epsilon_2' & E\epsilon_2\epsilon_3' \\ E\epsilon_3\epsilon_1' & E\epsilon_3\epsilon_2' & E\epsilon_3\epsilon_3' \end{bmatrix}. \tag{16}
$$

If the cross equation covariance $Cov(\epsilon_i, \epsilon_j)$ is denoted as σ_{ij}, we can write

$$E(\epsilon\epsilon') = \begin{bmatrix} \sigma_{11}I_T & \sigma_{12}I_T & \sigma_{13}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T & \sigma_{23}I_T \\ \sigma_{31}I_T & \sigma_{32}I_T & \sigma_{33}I_T \end{bmatrix} = \Omega \otimes I_T, \tag{17}$$

which defines the notation Ω. Define its inverse $V = (\Omega^{-1} \otimes I_T)$. Now the GLS estimator is: $\hat{\hat{\delta}}_{GLS} = (Z'VZ)^{-1}Z'Vy$. Zellner called it the seemingly unrelated regressions (SUR) estimator and proved that it is more efficient than OLS for the stacked system. Since V is generally unknown, the 'feasible' version of SUR replaces V by a consistent estimate $\hat{V} = (\hat{\Omega}^{-1} \otimes I_T)$.

References

[1] John Fox, *sem: Structural equation models*, 2009, R package version 0.9-15.

[2] Arne Henningsen and Jeff D. Hamann, *systemfit: A package for estimating systems of simultaneous equations in R*, Journal of Statistical Software **23** (2007), no. 4, 1–40.