Statistical and Econometric Methods for Transportation Data Analysis

Chapter 5 – Simultaneous-Equation Models

Example 5.1
Three-Stage Least Squares (3SLS)

Speed data are collected data from a 6-lane freeway with 3-lanes in each direction separated by a large median (each direction is considered separately). At the point where the data were gathered, highly variable seasonal weather conditions were present. As a consequence, seasonal factors are expected to play a role. The data were collected over a period of a year, and the mean speeds, by lane, were the mean of the spot speeds gathered over one-hour periods (2,575 observations). The equation system is written as,

$$s_R = \beta_R Z_R + \lambda_R s_C + \varepsilon_R$$
$$s_C = \beta_C Z_C + \lambda_C s_L + \tau_C s_R + \varepsilon_C$$
$$s_L = \beta_L Z_L + \lambda_L s_C + \varepsilon_L$$

where *s*'s are the mean speeds (over a one-hour period in kilometers/hr) for the right-most lane (subscript *R*) relative to the direction of travel (the slow lane), the center lane (subscript *C*) and the left lane (subscript *L*), *Z*'s are vectors of exogenous variables influencing the mean speeds in the corresponding lanes, β's are vectors of estimable parameters, λ's and τ's are estimable scalars, and ε's are disturbance terms. Estimate this model (or some statistically defensible alternative) using 3SLS.

Then, estimate the following equation system using seemingly unrelated regression estimation SURE as in assignment #1:

$$s_R = \beta_R Z_R + \varepsilon_R$$
$$s_C = \beta_C Z_C + \varepsilon_C$$
$$s_L = \beta_L Z_L + \varepsilon_L$$

In your write-up include:

1. The results of your best model specifications.

2. A discussion of the logical process that led you to the selection of your final specification (the theory behind the inclusion of your selected variables). Include *t*-statistics and justify the signs of your variables.

3. A brief comparison of 3SLS and SURE results.

Data in file "Ex5-1.txt":

Variable Number	Description
X1	Mean speed in the right lane in kilometers per hour (gathered over a one-hour period)
X2	Mean speed in the center lane in kilometers per hour (gathered over a one-hour period)
X3	Mean speed in the left lane in kilometers per hour (gathered over a one-hour period)
X4	Traffic flow in right lane (vehicles per hour)
X5	Traffic flow in center lane (vehicles per hour)
X6	Traffic flow in left lane (vehicles per hour)
X7	Proportion of passenger cars (including pick-up trucks and minivans) in the right lane
X8	Proportion of passenger cars (including pick-up trucks and minivans) in the center lane
X9	Proportion of passenger cars (including pick-up trucks and minivans) in the left lane
X10	Month that speed data was collected (1=January, 2=February, etc.)
X11	Hour in which data was collected (the beginning hour of the one-hour data collection period)

```
--> read;nvar=11;nobs=2575;file=D:Ex5-1.txt$
--> dstat;rhs=x1,x2,x3$
                          Descriptive Statistics
            All results based on nonmissing observations.
Variable        Mean         Std.Dev.        Minimum         Maximum      Cases
------------------------------------------------------------------------------
X1        .116255064D+03  .105706468D+02  .357700000D+02  .127960000D+03   2575
X2        .124934377D+03  .970367813D+01  .545100000D+02  .132680000D+03   2575
X3        .128996381D+03  .117778716D+02  .396400000D+02  .141750000D+03   2575

--> create;trucksR=(1-x7)*x4$
--> create;trucksC=(1-x8)*x5$
--> create;trucksL=(1-x9)*x6$
--> create;if(x10>10|x10<3)winter=1$
--> create;if(x10>2&x10<6)spring=1$
--> create;if(x11>600&x11<1100)ampeak=1$
--> create;if(x11>1600&x11<2000)pmpeak=1$
--> create;if(x7<0.81)tr20R=1$
--> create;if(x8<0.81)tr20C=1$
--> dstats;rhs=tr20R,tr20C$
                          Descriptive Statistics
            All results based on nonmissing observations.
Variable        Mean         Std.Dev.        Minimum         Maximum      Cases
------------------------------------------------------------------------------
TR20R     .914951456D+00  .279008101D+00  .000000000D+00  .100000000D+01   2575
TR20C     .104854369D+00  .306425186D+00  .000000000D+00  .100000000D+01   2575

--> 3sls;lhs=x1,x2,x3
    ;Eq1=one,x2,tr20R,winter,spring,ampeak,trucksR
    ;Eq2=one,x1,x3,pmpeak,trucksC
    ;Eq3=one,x2,x6
    ;Inst=x4,x5,x6,winter,spring,ampeak,pmpeak,tr20R,tr20C,x7,x8,x9,
    x11,trucksR,trucksC,trucksL
    ;maxit=1$
Iteration    0, 3SLS          =    1.000000
Iteration    1, 3SLS          =    2.516680

+------------------------------------------------------------------------+
| Estimates for equation: X1                                             |
| InstVar/GLS least squares regression    Weighting variable = none      |
| Dep. var. = X1       Mean=   .4514759770E-01, S.D.=   10.57064675      |
| Model size: Observations =    2575, Parameters =   7, Deg.Fr.=   2568  |
| Residuals:  Sum of squares= .5446923658D+05, Std.Dev.=        4.60551  |
| Fit:        R-squared=  .810101, Adjusted R-squared =          .80966  |
|             (Note:  Not using OLS.  R-squared is not bounded in [0,1]  |
| Model test: F[  6,   2568] = 1825.83,    Prob value =          .00000  |
| Diagnostic: Log-L =  -7582.9419, Restricted(b=0) Log-L =   -9725.3252  |
|             LogAmemiyaPrCrt.=    3.057, Akaike Info. Crt.=      5.895  |
| Durbin-Watson Stat.=     1.2629       Autocorrelation   =      .3685  |
+------------------------------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	-23.45421780	1.8507657	-12.673	.0000	
X2	1.112370386	.14502620E-01	76.701	.0000	124.93438
TR20R	-.1146796292	.29993767	-.382	.7022	.91495146
WINTER	-.5198830160	.18498440	-2.810	.0049	.42679612
SPRING	-.8525645458	.27376237	-3.114	.0018	.97864078E-01
AMPEAK	-.6382665028	.20379124	-3.132	.0017	.18252427
TRUCKSR	.1620186604E-01	.14132440E-02	11.464	.0000	77.475219

Estimates for equation: X2
InstVar/GLS least squares regression Weighting variable = none
Dep. var. = X2 Mean= .4851820454E-01, S.D.= 9.703678128
Model size: Observations = 2575, Parameters = 5, Deg.Fr.= 2570
Residuals: Sum of squares= .5849447751D+05, Std.Dev.= 4.77080
Fit: R-squared= .758188, Adjusted R-squared = .75781
(Note: Not using OLS. R-squared is not bounded in [0,1]
Model test: F[4, 2570] = 2014.52, Prob value = .00000
Diagnostic: Log-L = -7674.7359, Restricted(b=0) Log-L = -9504.9670
LogAmemiyaPrCrt.= 3.127, Akaike Info. Crt.= 5.965
Durbin-Watson Stat.= 1.5964 Autocorrelation = .2018

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	19.37310896	1.1445991	16.926	.0000	
X1	-.1167705732E-01	.10851777E-01	-1.076	.2819	116.25506
X3	.8288650092	.13084766E-01	63.346	.0000	128.99638
PMPEAK	-.1560351464E-01	.40878633E-01	-.382	.7027	.14252427
TRUCKSC	-.9591875570E-04	.13701144E-02	-.070	.9442	13.779988

Estimates for equation: X3
InstVar/GLS least squares regression Weighting variable = none
Dep. var. = X3 Mean= .5009568178E-01, S.D.= 11.77787156
Model size: Observations = 2575, Parameters = 3, Deg.Fr.= 2572
Residuals: Sum of squares= .8618792363D+05, Std.Dev.= 5.78879
Fit: R-squared= .758337, Adjusted R-squared = .75815
(Note: Not using OLS. R-squared is not bounded in [0,1]
Model test: F[2, 2572] = 4035.46, Prob value = .00000
Diagnostic: Log-L = -8173.7680, Restricted(b=0) Log-L = -10003.7896
LogAmemiyaPrCrt.= 3.513, Akaike Info. Crt.= 6.351
Durbin-Watson Stat.= 1.5985 Autocorrelation = .2008

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	-23.69019051	1.6353961	-14.486	.0000	
X2	1.221993379	.13025929E-01	93.812	.0000	124.93438
X6	.2024043224E-03	.14880520E-03	1.360	.1738	96.831845