THE ECONOMIC AND SOCIAL RESEARCH INSTITUTE

Memorandum Series No. 171 .

A Pitfall in Using TSLS with Linear Restrictions in TROLL

by

Denis Conniffe
John FitzGerald
Gerard Hughes

February 1985

IR£2

A Pitfall in Using TSLS with Linear Restrictions in TROLL*

Suppose we wish to estimate the following equation by TSLS,

$$Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e \qquad 1$$

because x_3 is correlated with the error term, and to impose the restriction that $\beta_1 + \beta_2 + \beta_3 = 1$. Assume that in addition to the predetermined variables in equation 1 we wish to use another variable, z, to instrument the endogenous variable x_3. The format commands to do so using the TSLS option in the second edition of TROLL (1980, p. 11-44) are:

CRINST x_1 x_2 z;

DOINST β_3;

The CRINST command specifies the predetermined variables to be used in the first stage regression and the DOINST command specifies that estimated values from the first stage regression are to be inserted for x_3 for the second stage regression.

If the restriction is imposed on the coefficient of one of the predetermined variables in equation 1, for example x_1 (i.e. $\beta_1 = 1 - \beta_2 - \beta_3$) the equation to be estimated at the second stage can be written as:

$$(y - x_1) = \alpha + \beta_2\ (x_2 - x_1) + \beta_3\ (\hat{x}_3 - x_1) + e \qquad 2$$

This equation has one endogenous variable, $(\hat{x}_3 - x_1)$ on the right hand side and the DOINST command correctly identifies it. However, if the restriction is imposed on the coefficient of the endogenous variable, x_3 (i.e., $\beta_3 = 1 - \beta_1 - \beta_2$), the equation to be estimated at the second stage is:

$$(y - x_3) = \alpha + \beta_1\ (x_1 - \hat{x}_3) + \beta_2\ (x_2 - \hat{x}_3) + e \qquad 3$$

This equation has two endogenous variables, $(x_1 - \hat{x}_3)$ and $(x_2 - \hat{x}_3)$, on the right hand side due to the imposition of the restriction on the endogenous variable x_3 and the DOINST command identifies neither of them. The computer program will, therefore, ignore the results of the first stage regression and

*We would like to thank Patrick Honohan for his comments on the TSLS results in an earlier draft of Hughes (1985) which led us to uncover the pitall in TROLL described in this memorandum.

run an OLS regression on the equation

$$(y - x_3) = \alpha + \beta_1 (x_1 - x_3) + \beta_2 (x_2 - x_3) + e \qquad 4$$

The coefficient estimates for this equation will differ from those for equation 2 because the first stage results are not taken into account in equation 4.

The easiest way around the difficulty posed by additional endogenous variables being created when restrictions are imposed on variables which are to be instrumented in TSLS is to include in the DOINST command the coefficients of all predetermined variables appearing in the second stage regression. This will ensure that the coefficient estimates are invariant to the way in which the restrictions are imposed because it forces the program to regress all of the variables on the right hand side of the second stage regression equation on the same set of instruments. Variables which are unaffected by the restrictions will be unchanged because the regression of a variable on a set of regressors which includes itself simply reproduces that variable whereas variables which are affected by the restriction will be replaced by a transformation which takes account of the first stage estimates of endogenous variables on which restrictions are imposed.

Following this procedure the appropriate format commands when the restriction in our model is placed on x_3 above are

CRINST x_1 x_2 z;

DOINST β_1 β_2;

The results of the first stage regression will be

$$x_1 = f (x_1, x_2, z) = x_1 \qquad 5$$

$$x_2 = g (x_1, x_2, z) = x_2 \qquad 6$$

$$x_3 = h (x_1, x_2, z) = \hat{x}_3 \qquad 7$$

where the regressions are linear but general notation is used for brevity. The second stage regression equation is therefore:

$$[y - h (x_1, x_2, z)] = \alpha + \beta_1 [f(x_1, x_2, z) - h (x_1, x_2, z)] + \beta_2 [g (x_1, x_2, z) - h(x_1, x_2, z)] + e \qquad 8$$

which reduces to equation 3 when the results for equations 5-7 are substituted.

An Example

Recently one of us wished to estimate the following wage equation for the period 1954 (2) to 1980 (4)

$$d\ \ln(AHEADJ) = A_o + A_1 d\ln(1 + s) + A_2 d\ln(1 - t) + A_3 \ln(Q/\hat{Q}) + A_4 d\ln P_q + A_5 d\ln P_{q-1} + A_6 d\ln P_c + A_7 SI + A_8 S2 + A_9 S3 \quad (9)$$

where (AHEADJ) is average hourly earnings in Transportable Goods Industries adjusted for overtime earnings;

(1 + s) is the employer payroll tax rate plus one;

(1 - T) is the proportion of earnings retained by employees after income tax and PRSI deductions;

$(Q/\hat{Q})$ is the ratio of the index of industrial production to its trend value;

P_q is the wholesale price index for the output of industry;

P_{q-1} is the wholesale price index for the output of industry lagged one quarter;

P_c is the consumer price index;

SI, S2 S3 are seasonal dummy variables;

The labour market model from which this equation is derived suggested the following restrictions should be imposed:

$$A_1 + A_4 + A_5 = 0,\ A_4 + A_5 + A_6 = 1,\ \text{and}\ A_2 + A_6 = 0$$

Since the two current price variables, P_q and P_c, were thought to be correlated with the error term it was decided to estimate the equation by TSLS using the predetermined variables in the wage equation plus import prices (MPI), labour productivity (LABPRD) and a dummy variable for the third quarter of 1975 (D753) to instrument the two price variables. The regression commands which were used were

```
CRINST DLOGMPI DLOGMPI (-1) DLLABRD DLOGS DLOGT3 LOGTQ DLPPI1 S1 S2 S3
DLAHEADJ (-1) DLAHEADJ (-4) D753 DLOGCPI (-4);
DOINST A4 A6;
```

Imposing the restrictions $A_1 = -(A_4 + A_5)$. $A_6 = (1 - A_4 - A_5)$, $A_2 = -(1 - A_4 - A_5)$ gave the estimates shown for the first equation in the table. Imposing the restrictions as $A_2 = -(1 + A_1)$, $A_5 = -(A_1 + A_4)$, $A_6 = (1 + A_1)$ or $A_2 = -(1 + A_1)$ $A_4 = -(A_1 + A_5)$ $A_6 = (1 + A_1)$ gave the estimates shown for the second and third equations. The result of the inappropriate DOINST command was that equation 1 was estimated treating only p_q as endogenous, equation 2 was estimated treating only p_c as endogenous while equation 3 treated neither as endogenous and was estimated by OLS. When the DOINST command specifies that all predetermined variables on the right hand side of equation 9 are to be instrumented in the second stage regression it is written as:

DOINST A1 A2 A3 A5 A7 A8 A9;

Some of the coefficients specified in this command will be redundant due to the imposition of the restrictions but it is easier to include them all rather than work out which may not be needed. The coefficient estimates which emerged from the second stage regressions when equation 9 was estimated by TSLS using the modified DOINST command and the three permutations of the restrictions specified above were identical, as was to be expected. The correct TSLS estimate of equation 9 is shown as equation 4 in the table.

TSLS Results for Wage Equation with Restrictions and DOINST Commands Specified in the Text, 1954 (2) - 1980 (4). Dependent Variables is d ln AHEADJ.

	Equation			
Variable	(1)	(2)	(3)	(4)
Constant	0.015 (4.40)	0.016 (4.68)	0.015 (4.64)	0.016 (4.43)
d ln (1+s)	−0.471 (2.86)	−0.572 (4.81)	−0.478 (4.73)	−0.488 (2.90)
d ln (1−T)	−0.529 (3.21)	−0.428 (3.60)	−0.522 (5.17)	−0.512 (3.04)
ln (Q/Q)	0.232 (1.56)	0.162 (1.05)	0.217 (1.48)	0.173 (1.13)
d ln P_q	0.211 (1.08)	0.555 (3.10)	0.277 (2.76)	0.468 (2.16)
d ln P_{q-1}	0.260 (2.65)	0.017 (0.13)	0.201 (2.17)	0.020 (0.16)
d ln P_c	0.529 (3.20)	0.428 (3.60)	0.522 (5.17)	0.512 (3.04)
S1	−0.008 (1.58)	−0.012 (2.19)	−0.009 (1.78)	−0.011 (2.12)
S2	−0.003 (0.75)	−0.003 (0.61)	−0.003 (0.73)	−0.003 (0.65)
S3	−0.011 (2.43)	−0.011 (2.32)	−0.011 (2.43)	−0.011 (2.22)
R^2	0.491	0.452	0.494	0.436
F	16.09	13.73	16.27	14.66
SER	0.017	0.017	0.017	0.017
DW	1.98	1.88	1.96	1.94
Cond. (X)	4.45	4.20	3.95	4.62

REFERENCES

HUGHES, Gerard, 1985, Payroll Tax Incidence, the Direct Tax Burden, and the Rate of Return on State Pension Contributions in Ireland. Dublin: Economic and Social Research Institute, Paper No. 120.

TROLL, 1980. TROLL USER's Guide, Second Edition. MIT Information Processing Services.