Two-stage Least Squares revisited

Two-stage least-squares (TSLS) is particularly useful for estimating econometric models of the form

$$\mathbf{y} = \mathbf{B}\mathbf{y} + \mathbf{\Gamma}\mathbf{x} + \mathbf{u},$$

where $\mathbf{y} = (y_1, y_2, ..., y_p)$ is a set of endogenous or jointly dependent variables, $\mathbf{x} = (x_1, x_2, ..., x_q)$ is a set of exogenous or predetermined variables uncorrelated with the error terms $\mathbf{u} = (u_1, u_2, ..., u_p)$, and $\mathbf{B}$ and $\mathbf{\Gamma}$ are parameter matrices.

A typical feature of the above model is that not all y-variables and not all x-variables are included in each equation.

A necessary condition for identification of each equation is that, for every y-variable on the right side of the equation, there must be at least one x-variable excluded from that equation. There is also a sufficient condition for identification, the so-called rank condition, but this is often difficult to apply in practice. For further information on identification of interdependent systems, see, e.g., Goldberger (1964, pp. 313-318).

Klein's (1950) Model I is a classical econometric model that has been used extensively as a benchmark problem for studying econometric methods. It is an eight-equation system based on annual data for the United States in the period between the two World Wars. It is dynamic in the sense that elements of time play important roles in the model.

The three behavioral equations of Klein's Model are

$$\begin{aligned} C_t &= a_1 P_t + a_2 P_{t-1} + a_3 W_t + u_1 \\ I_t &= b_1 P_t + b_2 P_{t-1} + b_3 K_{t-1} + u_2 \\ W_t^* &= c_1 E_t + c_2 E_{t-1} + c_3 A_t + u_3 \end{aligned}$$

In addition to these stochastic equations, the model includes five identities (definitional equations):

$$\begin{aligned} P_t &= Y_t - W_t \\ Y_t &= C_t + I_t + G_t - T_t \\ K_t &= K_{t-1} + I_t \\ W_t &= W_t^* + W_t^{**} \\ E_t &= Y_t + T_t - W_t^{**} \end{aligned}$$

The endogenous variables are:

Ct Aggregate Consumption (y_1)

It Net Investment (y_2)

Wt* Private Wage Bill (y_3)

Pt Total Profits (y_4)

Yt Total Income (y_5)

Kt End-of-year Capital Stock (y_6)

Wt Total Wage Bill (y_7)

Et Total Production of Private Industry (y_8)

The predetermined variables are the exogenous variables:

Wt** Government Wage Bill (x_1)

Tt Taxes (x_2)

Gt Government Non-Wage Expenditure (x_3)

At Time in Years from 1931 (x_4)

and the lagged endogenous variables $P_{t-1}\,(x_5)$, $K_{t-1}\,(x_6)$ and $E_{t-1}\,(x_7)$. All variables except A_t are in billions of 1934 dollars. Annual time series data for 1921–1941 has been computed from Theil's (1971) Table 9.1. These data are given in file **KLEIN.RAW**. Files for this example can be found in the **PRELIS Examples** folder.

In the consumption function there are 2 *y*-variables included on the right side, namely P_t and W_t and 6 *x*-variables excluded namely W_t^{**}, T_t, G_t, A_t, K_{k-1} and E_{t-1} so that the order condition is fulfilled. Similarly, it can be verified that the order condition is met also for the other two equations.

To estimate the consumption function, we use C_t as the *y*-variable, P_t, P_{t-1} , and W_t as *x*-variables, and all the predetermined variables as *z*-variables. A PRELIS command file to do this is (see **KLEIN1.PRL**):

```
Estimating Klein's Consumption Function
Data NI=15
Labels
C P_1 W* I K_1 E_1 W** T A P K E W Y G
Rawdata=KLEIN.RAW
Continuous All
RG C on P P_1 W with W** T G A P_1 K_1 E_1
Output Matrix=CM
```

The estimated consumption equation is given in the output as:[1]

```
Estimated Equations

        C = 16.555 + 0.0173*P + 0.216*P_1 + 0.810*W + Error, R² = 0.977
 Standerr  (1.468)  (0.131)    (0.119)     (0.0447)
 t-values   11.277   0.132      1.814       18.111
 P-values   0.000    0.897      0.086       0.000

 Error Variance = 1.290

 Instrumental Variables: W** T G A P_1 K_1 E_1
```

Note that one cannot estimate this equation by regressing C on P_t, P_{t-1} , and W_t as u_1 is not uncorrelated with P_t and W_t. Instead of

```
RG C on P P_1 W with W** T G A P_1 K_1 E_1
```

one can write

```
Equation C = P P_1 W with W** T G A P_1 K_1 E_1
```

and the word Equation can be abbreviated as Eq. Upper case or lower case can be used.

The intercept term in the equation will always be estimated by PRELIS. To estimate the equation without the intercept term use the SIMPLIS or LISREL command language and read a covariance matrix. To estimate the equation in standardized form, read a correlation matrix or specify MA = KM instead in LISREL.

The following PRELIS command file will add the residual to the variables and compute the covariance matrix of the extended set of variables (see file **KLEIN2.PRL**):

```
Estimating Klein's Consumption Function
Data NI=15
Labels
C P_1 W* I K_1 E_1 W** T A P K E W Y G
Rawdata=KLEIN.RAW
Continuous All
Equation C = P P_1 W with W** T G A P_1 K_1 E_1 Res=U
Output Matrix=CM
```

One can estimate all three behavioral equations in Klein's Model in one run. The following command file (see file **KLEIN3.PRL**) will do that and add the three residuals to the raw data in a file **KLEINEXT.RAW**.

```
Estimating Klein's Model I with TSLS
Data NI=15
Labels
C P_1 W* I K_1 E_1 W** T A P K E W Y G
Rawdata=KLEIN.RAW
Continuous All
EQ C = P P_1 W with W** T G A P_1 K_1 E_1 RES=U_1
EQ I = P P_1 K_1 with W** T G A P_1 K_1 E_1 RES=U_2
EQ W* = E E_1 A with W** T G A P_1 K_1 E_1 RES=U_3
```

[1] These results agree with those given in Goldberger (1964, p. 336) and Theil (1971, p. 458), except that these authors do not give the estimate of the intercept and they divide by N in (3.5) instead of $N-1-p$

```
Output Matrix=CM Raw=KLEINEXT.RAW
```

This gives the following result.

```
    C = 16.555 + 0.0173*P + 0.216*P_1 + 0.810*W + Error, R² = 0.977
 Standerr           (0.131)    (0.119)     (0.0447)
 t-values            0.132      1.814       18.111
 P-values            0.897      0.086       0.000

 Error Variance = 1.290

 Instrumental Variables: W** T G A P_1 K_1 E_1

        I = 20.278 + 0.150*P + 0.616*P_1 - 0.158*K_1 + Error, R² = 0.885
 Standerr           (0.193)   (0.181)     (0.0402)
 t-values            0.780     3.404      -3.930
 P-values            0.445     0.003       0.001

 Error Variance = 1.709

 Instrumental Variables: W** T G A P_1 K_1 E_1

       W* = 1.500 + 0.439*E + 0.147*E_1 + 0.130*A + Error, R² = 0.987
 Standerr  (0.543) (0.0396)  (0.0432)    (0.0324)
 t-values   2.765   11.082    3.398       4.026
 P-values   0.013   0.000     0.003       0.001

 Error Variance = 0.589

 Instrumental Variables: W** T G A P_1 K_1 E_1
```