
INST (INST=(*list of instruments*),
 ROBUSTSE, SILENT, TERSE, UNNORM, WEIGHT=*variable name*)
dependent variable name independent variables ;

Function:

INST obtains single equation instrumental variable estimates. By choosing an appropriate list of instrumental variables, INST will obtain conventional two stage least squares estimates. Options allow you to obtain weighted estimates to correct for heteroskedasticity, or to obtain standard errors which are robust in the presence of heteroskedasticity of the disturbances.

Usage:

In the basic INST statement, list the dependent variable first and then the independent variables which are in the equation. Include an option INST= (list of variables to be included as instruments. The list of instruments must include any exogenous variables in the equation, in particular the constant, C, as well as any additional instruments you may wish to specify. There must be at least as many instrumental variables as there are independent variables in the equation to meet the rank condition for identification. Any observations with missing values will be dropped from the sample.

Two stage least squares is INST with all the exogenous variables in the complete model listed as instruments and no other variables. Valid estimation can be based on fewer instruments when a complete model involves a large number of exogenous variables, or estimates can be made even when the rest of a simultaneous model is not fully specified. In these cases, the estimator is instrumental variables, but not really two stage least squares.

If there are exactly as many instruments as independent variables specified, the resulting estimator is classic instrumental variables, that is,

$$d = (Z'X)^{-1}Z'y$$

where Z is the matrix of instruments, X the matrix of independent variables, and y the dependent variable. (For the more general case, see the formulas below).

Instrumental variable estimation may also be obtained with the AR1 procedure and the LSQ procedure (for nonlinear two and three stage least squares). In these cases, include the list of instruments in the INST option. See those commands for further information.

The list of independent variables on the INST command may include PDL variables, however, you are responsible for seeing that there are enough appropriate instruments for these variables after scrambling. If the PDL variable is exogenous, the safest list would include all the lags of the variable over which the PDL is defined. These variables may be highly collinear, but will cause no problems due to TSP's use of the generalized inverse when computing regressions - a subset of the variables which includes all the information in them will be used. If the PDL variable is endogenous, you must include enough instruments for all the resulting scrambled variables to meet the order condition for identification.

Options:

INST=(*list of instrumental variables*)

INST

NORM/UNNORM tells whether the weights are to be normalized so that they sum to the number of observations. This has no effect on the coefficient estimates and most of the statistics, but it makes the magnitude of the unweighted and weighted data the same on average, which may help in interpreting the results. The **NORM** option has no effect if the **WEIGHT** option has not been specified.

ROBUSTSE/NOROBUST causes the variance of the coefficient estimates, the standard errors, and associated t-statistics to be computed using the formulas suggested by White, among others. These estimates of the variance are consistent even when the disturbances are not homoskedastic (although they must be independent), and when their variances are correlated with the independent variables in the model. See the references for the exact formulas.

SILENT/NOSILENT suppresses all printed output.

TERSE/NOTERSE suppresses printing of everything but the $E'P_2E$ objective function and the table of coefficients.

WEIGHT= the name of a series which will be used to weight the observations. The data and the instruments are multiplied by the square roots of the weighting series before the regression is computed, so that the weighting series should be proportional to the inverses of the variances of the disturbances. If the weight is zero for a particular observation, that observation is not included in the computations nor is it counted in determining degrees of freedom.

Examples:

This example estimates the consumption function for the illustrative model, using the constant, trend, government expenditures (G), and the log of the money supply (LM) as instruments:

```
INST CONS,C,GNP INVR C,G,TIME,LM ;
```

Using population as weights, the following example regresses the fraction of young people living alone on various other demographic characteristics across states. Population is proportional to the inverse of the variance of per capita figures.

```
INST (WEIGHT=POP INST=(C, URBAN, CATHOLIC, SERVEMP, SOUTH)  
      YOUNG, C,RSALE, URBAN, CATHOLIC ;
```

Other examples of the **INST/2SLS** command:

```
INST (ROBUSTSE INST=C LOGR LOGR(-1) LOGR(-2) LOGR(-3)) LOGP C LOGP(-1) LOGR ;  
2SLS(INST=(C,LM(-1)-LM(-3))) TBILL C RATE(4,12,FAR) ;
```

Note that the constant (C) must always be named explicitly as an instrument if it is needed.

Output:

The output of **INST** begins with an equation title, the name of the dependent variable and the list of instruments. This is followed by statistics on goodness-of-fit: the sum of squared residuals, the standard error of the regression, the R-squared, the Durbin-Watson statistic for autocorrelation of the residuals, and an F-statistic for the hypothesis that all coefficients in the regression except the constant are zero. The objective function $E'P_2E$ is the length of the residual vector projected onto the space of the instruments. This is analogous to the sum of squared residuals in OLSQ -- it can be used to construct a pseudo-F test of nested models. Note that it is zero for exactly identified models (if they have full rank). These statistics are all based on the "structural" residuals, that is residuals computed with the actual values of the right hand side endogenous variables in the model rather than the "fitted" values from a hypothetical first stage regression.

Following this is a table of right hand side variable names, estimated coefficients, standard errors and associated t-statistics. If the variance-covariance matrix has been unsuppressed (see the **SUPRES** command), it is printed after this

table. Finally, if the RESID and PLOTS options are on, a table and plot of the actual and fitted values of the dependent variable and the residuals is printed.

INST also stores most of these results in data storage for later use. The table below lists the results available after an INST command. The fitted values and residuals will only be stored if the RESID option is on (the default).

Name	Type	Length	Variable Description
@LHV	list	1	Name of the dependent variable.
@RNMS	list	#vars	Names of right hand side variables.
@SSR	scalar	1	Sum of squared residuals.
@S	scalar	1	Standard error of the regression.
@S2	scalar	1	Standard error squared.
@YMEAN	scalar	1	Mean of the dependent variable.
@SDEV	scalar	1	Standard deviation of the dependent variable.
@NOB	scalar	1	Number of observations.
@DW	scalar	1	Durbin-Watson statistic.
@RSQ	scalar	1	R-squared.
@ARSQ	scalar	1	Adjusted R-squared.
@FST	scalar	1	F-statistic.
@PHI	scalar	1	$E'P_2E$ - the objective function.
@COEF	vector	#vars	Coefficient estimates.
@SES	vector	#vars	Standard Errors.
@T	vector	#vars	T-statistics.
@VCOV	matrix	#vars* #vars	Variance-covariance of estimated coefficients.
@RES	series	#obs	Residuals = actual - fitted values of the dependent variable.
@FIT	series	#obs	Fitted values of dependent variable.

If the regression includes PDL variables, the following will also be stored:

@SLAG	scalar	1	Sum of the lag coefficients.
@MLAG	scalar	1	Mean lag coefficient (number of time periods).
@LAGF	vector	#lags	Estimated lag coefficients, after "unscrambling".

Method:

Let y be the dependent variable, X be the $(T \times k)$ matrix of independent variables, and Z be the matrix of instrumental variables (the included and excluded exogenous variables for two stage least squares). Then the formulas used to compute the coefficients, their standard errors, and the objective function are the following:

$$P_z = Z(Z'Z)^{-1}Z'$$

$$b = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y \quad \text{or} \quad [X'P_zX]^{-1}X'P_zy$$

$$E = y - Xb \quad ; \quad s^2 = E'E/(T-k) \quad ; \quad @PHI = E'P_zE$$

$$V(b) = s^2[X'Z(Z'Z)^{-1}Z'X]^{-1} \quad \text{or} \quad s^2[X'P_zX]^{-1}$$

The structural residuals E are used to compute all the goodness-of-fit statistics.

INST

References:

Judge et al, **The Theory and Practice of Econometrics**, John Wiley & Sons, New York, 1981, pp. 531-533.

Maddala, G. S., **Econometrics**, McGraw Hill Book Company, New York, 1977, Chapter 11.

Pindyck, Robert S., and Daniel L. Rubinfeld, **Econometric Models and Economic Forecasts**, McGraw Hill Book Company, New York, 1976, Chapter 5.

Theil, Henri, **Principles of Econometrics**, John Wiley & Sons, New York, 1971, Chapter 9.

White, Halbert, "Instrumental Variables Regression with Independent Observations," **Econometrica** 50, March 1982, pp. 483-500.