eivreg — Errors-in-variables regression

Syntax

eivreg depvar [ indepvars ] [ if ] [ in ] [ weight ] [ , options ]

options

Model

_reliab_(indepvar # [ indepvar # [ . . . ] ])

set confidence level; default is level(95)

display_options

control column formats, row spacing, line width, display of omitted
variables and base and empty cells, and factor-variable labeling

_coeflegend

display legend instead of statistics

_indepvars_ may contain factor variables; see [U] 11.4.3 Factor variables.

bootstrap, by, jackknife, rolling, and statsby are allowed; see [U] 11.1.10 Prefix commands.

Weights are not allowed with the bootstrap prefix; see [R] bootstrap.

aweight s and fweight s are allowed; see [U] 11.1.6 weight.
coeflegend does not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.

Menu

Statistics > Linear models and related > Errors-in-variables regression

Description

eivreg fits errors-in-variables regression models.

Options

_reliab_(indepvar # [ indepvar # [ . . . ] ])

specifies the measurement reliability for each independent
variable measured with error. Reliabilities are specified as pairs consisting of an independent
variable name (a name that appears in _indepvars_) and the corresponding reliability _r_, 0 < _r_ ≤ 1.
Independent variables for which no reliability is specified are assumed to have reliability 1. If the
option is not specified, all variables are assumed to have reliability 1, and the result is thus the
same as that produced by _regress_ (the ordinary least-squares results).
Reporting
level(#); see [R] estimation options.

display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nolabel, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] estimation options.

The following option is available with eivreg but is not shown in the dialog box: coeflegend; see [R] estimation options.

Remarks and examples


Errors-in-variables regression models are useful when one or more of the independent variables are measured with additive noise. Standard regression (as performed by regress) would underestimate the effect of the variable, and the other coefficients in the model can be biased to the extent that they are correlated with the poorly measured variable. You can adjust for the biases if you know the reliability:

\[ r = 1 - \frac{\text{noise variance}}{\text{total variance}} \]

That is, given the model \( y = X\beta + u \), for some variable \( x_i \) in \( X \), the \( x_i \) is observed with error, \( x_i = x_i^* + e \), and the noise variance is the variance of \( e \). The total variance is the variance of \( x_i \).

Example 1

Say that in our automobile data, the weight of cars was measured with error, and the reliability of our measured weight is 0.85. The result of this would be to underestimate the effect of weight in a regression of, say, price on weight and foreign, and it would also bias the estimate of the coefficient on foreign (because being of foreign manufacture is correlated with the weight of cars). We would ignore all of this if we fit the model with regress:

```
. use http://www.stata-press.com/data/r13/auto
(1978 Automobile Data)
. regress price weight foreign
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 74</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>316859273</td>
<td>2</td>
<td>158429637</td>
<td>F( 2, 71) = 35.35</td>
</tr>
<tr>
<td>Residual</td>
<td>318206123</td>
<td>71</td>
<td>4481776.38</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>635065396</td>
<td>73</td>
<td>8699525.97</td>
<td>R-squared = 0.4989</td>
</tr>
</tbody>
</table>

| price | Coef.   | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|-------|---------|-----------|-------|------|----------------------|
| weight | 3.320737 | .3958784  | 8.39  | 0.000| 2.531378, 4.110096  |
| foreign| 3637.001 | 668.583   | 5.44  | 0.000| 2303.885, 4970.118  |
| _cons  | -4942.844| 1345.591  | -3.67 | 0.000| -7625.876, -2259.812|
With `eivreg`, we can account for our measurement error:

```
.eivreg price weight foreign, r(weight .85)
```

<table>
<thead>
<tr>
<th>assumed variable</th>
<th>reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td>0.8500</td>
</tr>
<tr>
<td>foreign</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

**Errors-in-variables regression**

- Number of obs = 74
- F( 2, 71) = 50.37
- Prob > F = 0.0000
- R-squared = 0.6483
- Root MSE = 1773.54

|              | Coef. | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|--------------|-------|-----------|-------|-------|---------------------|
| price        |       |           |       |       |                     |
| weight       | 4.31985 | .431431  | 10.01 | 0.000 | 3.459601 - 5.180099 |
| foreign      | 4637.32 | 624.5362 | 7.43  | 0.000 | 3392.03 - 5882.609  |
| _cons        | -8257.017 | 1452.086 | -5.69 | 0.000 | -11152.39 - 5361.639 |

The effect of weight is increased—as we knew it would be—and here the effect of foreign manufacture is also increased. A priori, we knew only that the estimate of `foreign` might be biased; we did not know the direction.

**Technical note**

Swept under the rug in our example is how we would determine the reliability, $r$. We can easily see that a variable is measured with error, but we may not know the reliability because the ingredients for calculating $r$ depend on the unobserved noise.

For our example, we made up a value for $r$, and in fact we do not believe that weight is measured with error at all, so the reported `eivreg` results have no validity. The `regress` results were the statistically correct results here.

But let’s say that we do suspect that weight is measured with error and that we do not know $r$. We could then experiment with various values of $r$ to describe the sensitivity of our estimates to possible error levels. We may not know $r$, but $r$ does have a simple interpretation, and we could probably produce a sensible range for $r$ by thinking about how the data were collected.

If the reliability, $r$, is less than the $R^2$ from a regression of the poorly measured variable on all the other variables, including the dependent variable, the information might as well not have been collected; no adjustment to the final results is possible. For our automobile data, running a regression of `weight` on `foreign` and `price` would result in an $R^2$ of 0.6743. Thus the reliability must be at least 0.6743 here. If we specify a reliability that is too small, `eivreg` will inform us and refuse to fit the model:

```
.eivreg price weight foreign, r(weight .6742)
reliability r() too small
r(399);
```

Returning to our problem of how to estimate $r$, too small or not, if the measurements are summaries of scaled items, the reliability may be estimated using the `alpha` command; see [MV] `alpha`. If the score is computed from factor analysis and the data are scored using `predict`’s default options (see [MV] `factor postestimation`), the square of the standard deviation of the score is an estimate of the reliability.
Consider a model with more than one variable measured with error. For instance, say that our model is that price is a function of weight, foreign, and mpg and that both weight and mpg are measured with error.

```
eivreg price weight foreign mpg, r(weight .85 mpg .9)
```

<table>
<thead>
<tr>
<th>variable</th>
<th>reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td>0.8500</td>
</tr>
<tr>
<td>mpg</td>
<td>0.9000</td>
</tr>
<tr>
<td>*</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Errors-in-variables regression

|                    | Coef.       | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|--------------------|-------------|-----------|-------|------|----------------------|
| price              |             |           |       |      |                      |
| weight             | 12.88302    | .6820532  | 18.89 | 0.000| 11.52271              |
| foreign            | 8268.951    | 352.8719  | 23.43 | 0.000| 7565.17               |
| mpg                | 999.2043    | 73.60037  | 13.58 | 0.000| 852.413               |
| _cons              | -56473.19   | 3710.015  | -15.22| 0.000| -63872.58            |

Root MSE = 496.41

Stored results

eivreg stores the following in e():

**Scalars**
- e(N): number of observations
- e(df_m): model degrees of freedom
- e(df_r): residual degrees of freedom
- e(r2): R-squared
- e(F): F statistic
- e(rmse): root mean squared error
- e(rank): rank of e(V)

**Macros**
- e(cmd): eivreg
- e(cmdline): command as typed
- e(depvar): name of dependent variable
- e(rellist): indepvars and associated reliabilities
- e(utype): weight type
- e(wexp): weight expression
- e(properties): b V
- e(predict): program used to implement predict
- e(asbalanced): factor variables fvset as asbalanced
- e(asobserved): factor variables fvset as asobserved

**Matrices**
- e(b): coefficient vector
- e(Cns): constraints matrix
- e(V): variance–covariance matrix of the estimators

**Functions**
- e(sample): marks estimation sample
Methods and formulas

Let the model to be fit be

\[ y = X^* \beta + e \]
\[ X = X^* + U \]

where \( X^* \) are the true values and \( X \) are the observed values. Let \( W \) be the user-specified weights. If no weights are specified, \( W = I \). If weights are specified, let \( v \) be the specified weights. If \texttt{fweight} frequency weights are specified, then \( W = \text{diag}(v) \). If \texttt{aweight} analytic weights are specified, then \( W = \text{diag}\{v/(1'v)(1'1)\} \), meaning that the weights are normalized to sum to the number of observations.

The estimates \( \hat{b} \) of \( \beta \) are obtained as \( A^{-1}X'Wy \), where \( A = X'WX - S \). \( S \) is a diagonal matrix with elements \( N(1 - r_i)s_i^2 \). \( N \) is the number of observations, \( r_i \) is the user-specified reliability coefficient for the \( i \)th explanatory variable or 1 if not specified, and \( s_i^2 \) is the (appropriately weighted) variance of the variable.

The variance–covariance matrix of the estimators is obtained as \( s^2 A^{-1}X'WXA^{-1} \), where the root mean squared error \( s^2 = (y'Wy - \hat{b}A\hat{b}')/(N - p) \), where \( p \) is the number of estimated parameters.

References


Also see

[R] eivreg postestimation — Postestimation tools for eivreg
[R] regress — Linear regression
[SEM] example 24 — Reliability
[U] 20 Estimation and postestimation commands