Description

_xtintreg_ fits a random-effects regression model in which the dependent variable may be measured as point data, interval data, left-censored data, or right-censored data. The dependent variable must be specified using two *depvars* that indicate how the dependent variable was measured. The user can request that a likelihood-ratio test comparing the panel interval regression model with the pooled model be conducted at estimation time.

Quick start

Regression on *x* of an interval-measured dependent variable with lower endpoint *y_lower* and upper endpoint *y_upper* using _xtset_ data

```
xhintreg y_lower y_upper x
```

Add indicators for levels of categorical variable *a* as covariates

```
xhintreg y_lower y_upper x i.a
```

Perform likelihood-ratio test against pooled model

```
xhintreg y_lower y_upper x i.a, intreg
```

Menu

Statistics > Longitudinal/panel data > Censored outcomes > Interval regression (RE)
xtintreg — Random-effects interval-data regression models

Syntax

```
x	intreg  depvar_lower  depvar_upper  [  indepvars  ]  [  if  ]  [  in  ]  [  weight  ]  [ ,  options  ]
```

The values in `depvar_lower` and `depvar_upper` should have the following form:

<table>
<thead>
<tr>
<th>Type of data</th>
<th><code>depvar_lower</code></th>
<th><code>depvar_upper</code></th>
</tr>
</thead>
<tbody>
<tr>
<td>point data</td>
<td><code>a</code></td>
<td><code>a</code></td>
</tr>
<tr>
<td>interval data</td>
<td><code>[a,b]</code></td>
<td><code>a</code></td>
</tr>
<tr>
<td>left-censored data</td>
<td><code>(-∞, b)</code></td>
<td><code>.</code></td>
</tr>
<tr>
<td>right-censored data</td>
<td><code>[a, +∞)</code></td>
<td><code>a</code></td>
</tr>
<tr>
<td>missing</td>
<td><code>.</code></td>
<td><code>.</code></td>
</tr>
</tbody>
</table>

**options**

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td><code>noconstant</code></td>
</tr>
<tr>
<td><code>offset(varname)</code></td>
</tr>
<tr>
<td><code>constraints(constraints)</code></td>
</tr>
<tr>
<td>SE</td>
</tr>
<tr>
<td><code>vce(vctype)</code></td>
</tr>
<tr>
<td>Reporting</td>
</tr>
<tr>
<td><code>level(#)</code></td>
</tr>
<tr>
<td><code>lrmodel</code></td>
</tr>
<tr>
<td><code>intreg</code></td>
</tr>
<tr>
<td><code>nocnsreport</code></td>
</tr>
<tr>
<td><code>display_options</code></td>
</tr>
<tr>
<td>Integration</td>
</tr>
<tr>
<td><code>intmethod(intmethod)</code></td>
</tr>
<tr>
<td><code>intpoints(#)</code></td>
</tr>
<tr>
<td>Maximization</td>
</tr>
<tr>
<td><code>maximize_options</code></td>
</tr>
<tr>
<td><code>collinear</code></td>
</tr>
<tr>
<td><code>coeflegend</code></td>
</tr>
</tbody>
</table>

A panel variable must be specified; use `xtset`; see [XT] xtset.

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

`depvar_lower`, `depvar_upper`, and `indepvars` may contain time-series operators; see [U] 11.4.4 Time-series varlists.

by and `stataby` are allowed; see [U] 11.1.10 Prefix commands.

iweights are allowed; see [U] 11.1.6 weight. Weights must be constant within panel.

`collinear` and `coeflegend` do not appear in the dialog box.

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

- **Model**
  
  - `noconstant, offset(varname), constraints(constraints);` see [R] Estimation options.
  
- **SE**
  
  - `vce(vcetype)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`oim`) and that use bootstrap or jackknife methods (`bootstrap, jackknife`); see [XT] vce options.
  
- **Reporting**
  
  - `level(#)`, `lrmodel`; see [R] Estimation options.
  
  - `intreg` specifies that a likelihood-ratio test comparing the random-effects model with the pooled (`intreg`) model be included in the output.
  
  - `nocnsreport`; see [R] Estimation options.
  
- **Integration**
  
  - `intmethod(intmethod), intpoints(#);` see [R] Estimation options.
  
- **Maximization**
  
  - `maximize_options: difficult, technique(algorithm_spec), iterate(#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), nonrtolerance, and from(init_specs);` see [R] Maximize. These options are seldom used.

The following options are available with xtintreg but are not shown in the dialog box:

- `collinear, coeflegend;` see [R] Estimation options.

**Remarks and examples**

xtintreg may be used to fit a random-effects interval regression model. Consider the linear regression model with panel-level random effects

\[ y_{it} = \mathbf{x}_{it} \beta + \nu_i + \epsilon_{it} \]

for \( i = 1, \ldots, n \) panels, where \( t = 1, \ldots, n_i \). The random effects, \( \nu_i \), are i.i.d., \( N(0, \sigma^2_{\nu}) \), and \( \epsilon_{it} \) are i.i.d., \( N(0, \sigma^2_{\epsilon}) \) independently of \( \nu_i \). The observed data consist of the couples, \((y_{1it}, y_{2it})\), such that all that is known is that \( y_{1it} \leq y_{it} \leq y_{2it} \), where \( y_{1it} \) is possibly \(-\infty\) and \( y_{2it} \) is possibly \(+\infty\).

**Example 1: Random-effects interval regression**

We begin with the nlswork dataset described in [XT] xt and create two fictional dependent variables, where the wages are instead reported sometimes as ranges. The wages have been adjusted to 1988 dollars and have further been recoded such that some of the observations are known exactly, some are left-censored, some are right-censored, and some are known only in an interval.
We wish to fit a random-effects interval regression model of adjusted (log) wages. We specify the \texttt{intreg} option to test our random-effects model against our pooled estimator.

\begin{verbatim}
. use https://www.stata-press.com/data/r16/nlswork5
(National Longitudinal Survey. Young Women 14–26 years of age in 1968)
. xtintreg ln_wage1 ln_wage2 i.union age grade not_smsa south##c.year, intreg
\end{verbatim}

Random-effects interval regression

<table>
<thead>
<tr>
<th>Number of obs</th>
<th>19,224</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncensored</td>
<td>4,810</td>
</tr>
<tr>
<td>Left-censored</td>
<td>4,781</td>
</tr>
<tr>
<td>Right-censored</td>
<td>4,848</td>
</tr>
<tr>
<td>Interval-cens.</td>
<td>4,785</td>
</tr>
</tbody>
</table>

Group variable: idcode

| Number of groups | 4,148 |

Random effects \( u_i \) ~ Gaussian

<table>
<thead>
<tr>
<th>Obs per group:</th>
</tr>
</thead>
<tbody>
<tr>
<td>min = 1</td>
</tr>
<tr>
<td>avg = 4.6</td>
</tr>
<tr>
<td>max = 12</td>
</tr>
</tbody>
</table>

Integration method: mvaghermite

| Integration pts. | 12 |

Log likelihood = -23260.672

| Wald chi2(7) | 2461.69 |
| Prob > chi2   | 0.0000 |

| Coef.      | Std. Err. | z      | P>|z|   | [95% Conf. Interval] |
|-------------|-----------|--------|-------|----------------------|
| 1.union     | .1229681  | .0092943| 13.23 | 0.000                | .1047516 .1411846 |
| age         | .0096333  | .0019   | 5.07  | 0.000                | .0059094 .0133572 |
| grade       | .0756045  | .0023828| 31.73 | 0.000                | .0709343 .0802747 |
| not_smsa    | -.1481304 | .011433 | -12.96| 0.000                | -.1705387 -.1257221|
| 1.south     | -.3586443 | .0977512| -3.67 | 0.000                | -.5502331 -.1670555|
| year        | .0029219  | .0020353| 1.44  | 0.151                | -.0010671 .0069109 |
| south#c.year| .0032699  | .0012076| 2.71  | 0.007                | .000903  .0056368 |
| _cons       | .2747391  | .1141328| 2.41  | 0.016                | .0510429 .4984352 |

| /sigma_u    | .3044775  | .0052644| 57.84 | 0.000                | .2941594 .3147956 |
| /sigma_e    | .3516248  | .00307  | 114.54| 0.000                | .3456078 .3576418 |

| rho         | .4285095  | .0101261| .4087613 | .4484385 |

LR test of sigma_u=0: chibar2(01) = 2683.77  Prob >= chibar2 = 0.000

The results from an interval regression can be interpreted as we would those from a linear regression. Because the dependent variable is log transformed, the coefficients can be interpreted in terms of a percentage change. We see, for example, that on average, union members make 12.3% more than nonunion members.
The output also includes the overall and panel-level variance components (labeled \( \sigma_e \) and \( \sigma_u \), respectively) together with \( \rho \) (labeled \( \text{rho} \)),

\[
\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}
\]

which is the proportion of the total variance contributed by the panel-level variance component.

When \( \text{rho} \) is zero, the panel-level variance component is unimportant, and the panel estimator is not different from the pooled estimator. A likelihood-ratio test of this is included at the bottom of the output. This test formally compares the pooled estimator (interval regression) with the panel estimator. In this case, we reject the null hypothesis that there are no panel-level effects.

Technical note

The random-effects model is calculated using quadrature, which is an approximation whose accuracy depends partially on the number of integration points used. We can use the quadchk command to see if changing the number of integration points affects the results. If the results change, the quadrature approximation is not accurate given the number of integration points. Try increasing the number of integration points using the \text{intpoints}() option and run quadchk again. Do not attempt to interpret the results of estimates when the coefficients reported by quadchk differ substantially. See [XT] quadchk for details and [XT] xtprobit for an example.

Because the xtintreg likelihood function is calculated by Gauss–Hermite quadrature, on large problems the computations can be slow. Computation time is roughly proportional to the number of points used for the quadrature.

Stored results

xtintreg stores the following in \( e() \):

Scalars

- \( e(N) \): number of observations
- \( e(N_g) \): number of groups
- \( e(N_unc) \): number of uncensored observations
- \( e(N_lc) \): number of left-censored observations
- \( e(N_rc) \): number of right-censored observations
- \( e(N_int) \): number of interval observations
- \( e(k) \): number of parameters
- \( e(k_aux) \): number of auxiliary parameters
- \( e(k_eq) \): number of equations in \( e(b) \)
- \( e(k_eq_model) \): number of equations in overall model test
- \( e(df_m) \): model degrees of freedom
- \( e(ll) \): log likelihood
- \( e(ll_0) \): log likelihood, constant-only model
- \( e(ll_c) \): log likelihood, comparison model
- \( e(chi2) \): \( \chi^2 \)
- \( e(chi2_c) \): \( \chi^2 \) for comparison test
- \( e(rho) \): \( \rho \)
- \( e(sigma_u) \): panel-level standard deviation
- \( e(sigma_e) \): standard deviation of \( \epsilon_{it} \)
- \( e(n_quad) \): number of quadrature points
- \( e(g_min) \): smallest group size
### Methods and formulas

Assuming a normal distribution, \( N(0, \sigma^2_\nu) \), for the random effects \( \nu_i \), we have the joint (unconditional of \( \nu_i \)) density of the observed data for the \( i \)th panel

\[
f\left\{ (y_{1i1}, y_{2i1}), \ldots, (y_{1in_i}, y_{2in_i}) | \mathbf{x}_{1i1}, \ldots, \mathbf{x}_{in_i} \right\} = \int_{-\infty}^{\infty} \frac{e^{-\nu_i^2/2\sigma^2_\nu}}{\sqrt{2\pi}\sigma_\nu} \left\{ \prod_{t=1}^{n_i} F(y_{1it}, y_{2it}, \mathbf{x}_{it}\beta + \nu_i) \right\} d\nu_i
\]
where

\[ F(y_{1it}, y_{2it}, \Delta_{it}) = \begin{cases} 
\left(\frac{1}{\sqrt{2\pi\sigma_c}}\right)^{-1} e^{-((y_{1it} - \Delta_{it})^2)/(2\sigma_c^2)} & \text{if } (y_{1it}, y_{2it}) \in C \\
\Phi\left(\frac{y_{2it} - \Delta_{it}}{\sigma_c}\right) & \text{if } (y_{1it}, y_{2it}) \in L \\
1 - \Phi\left(\frac{y_{1it} - \Delta_{it}}{\sigma_c}\right) & \text{if } (y_{1it}, y_{2it}) \in R \\
\Phi\left(\frac{y_{2it} - \Delta_{it}}{\sigma_c}\right) - \Phi\left(\frac{y_{1it} - \Delta_{it}}{\sigma_c}\right) & \text{if } (y_{1it}, y_{2it}) \in I 
\end{cases} \]

where \( C \) is the set of noncensored observations \((y_{1it} = y_{2it} \text{ and both nonmissing})\), \( L \) is the set of left-censored observations \((y_{1it} \text{ missing and } y_{2it} \text{ nonmissing})\), \( R \) is the set of right-censored observations \((y_{1it} \text{ nonmissing and } y_{2it} \text{ missing})\), \( I \) is the set of interval observations \((y_{1it} < y_{2it} \text{ and both nonmissing})\), and \( \Phi() \) is the cumulative normal distribution.

The panel-level likelihood \( l_i \) is given by

\[
l_i = \int_{-\infty}^{\infty} e^{-\nu_i^2/2\sigma^2} \left\{ \prod_{t=1}^{n_i} F(y_{1it}, y_{2it}, x_{it}\beta + \nu_i) \right\} d\nu_i
\]

\[
eq \int_{-\infty}^{\infty} g(y_{1it}, y_{2it}, x_{it}, \nu_i) d\nu_i
\]

This integral can be approximated with \( M \)-point Gauss–Hermite quadrature

\[
\int_{-\infty}^{\infty} e^{-x^2} h(x) dx \approx \sum_{m=1}^{M} w^*_m h(a^*_m)
\]

This is equivalent to

\[
\int_{-\infty}^{\infty} f(x) dx \approx \sum_{m=1}^{M} w^*_m \exp\left\{ (a^*_m)^2 \right\} f(a^*_m)
\]

where the \( w^*_m \) denote the quadrature weights and the \( a^*_m \) denote the quadrature abscissas. The log likelihood, \( L \), is the sum of the logs of the panel-level likelihoods \( l_i \).

The default approximation of the log likelihood is by adaptive Gauss–Hermite quadrature, which approximates the panel-level likelihood with

\[
l_i \approx \sqrt{2\hat{\sigma}_i} \sum_{m=1}^{M} w^*_m \exp\left\{ (a^*_m)^2 \right\} g(y_{1it}, y_{2it}, x_{it}, \sqrt{2\hat{\sigma}_i} a^*_m + \hat{\mu}_i)
\]

where \( \hat{\sigma}_i \) and \( \hat{\mu}_i \) are the adaptive parameters for panel \( i \). Therefore, using the definition of \( g(y_{1it}, y_{2it}, x_{it}, \nu_i) \), the total log likelihood is approximated by
\[ L \approx \sum_{i=1}^{n} w_i \log \left[ \sqrt{2\hat{\sigma}_i} \sum_{m=1}^{M} w_m^* \exp \left\{ (a_m^*)^2 \right\} \exp \left\{ -\left( \sqrt{2\hat{\sigma}_i} a_m^* + \hat{\mu}_i \right)^2 / 2\sigma^2 \right\} \right] \\
\times \sqrt{2\pi} \nu ^{1/2} \prod_{t=1}^{n_i} F\left( y_{1it}, y_{2it}, x_{it} \beta + \sqrt{2\hat{\sigma}_i} a_m^* + \hat{\mu}_i \right) \]  

(1)

where \( w_i \) is the user-specified weight for panel \( i \); if no weights are specified, \( w_i = 1 \).

The default method of adaptive Gauss–Hermite quadrature is to calculate the posterior mean and variance and use those parameters for \( \hat{\mu}_i \) and \( \hat{\sigma}_i \) by following the method of Naylor and Smith (1982), further discussed in Skrondal and Rabe-Hesketh (2004). We start with \( \hat{\sigma}_{i,0} = 1 \) and \( \hat{\mu}_{i,0} = 0 \), and the posterior means and variances are updated in the \( k \)th iteration. That is, at the \( k \)th iteration of the optimization for \( l_i \) we use

\[ l_{i,k} \approx \sum_{m=1}^{M} \sqrt{2\hat{\sigma}_{i,k-1}} w_m^* \exp \left\{ (a_m^*)^2 \right\} g\left( y_{1it}, y_{2it}, x_{it}, \sqrt{2\hat{\sigma}_{i,k-1}} a_m^* + \hat{\mu}_{i,k-1} \right) \]

Letting

\[ \tau_{i,m,k-1} = \sqrt{2\hat{\sigma}_{i,k-1}} a_m^* + \hat{\mu}_{i,k-1} \]

\[ \hat{\mu}_{i,k} = \sum_{m=1}^{M} (\tau_{i,m,k-1}) \sqrt{2\hat{\sigma}_{i,k-1}} w_m^* \exp \left\{ (a_m^*)^2 \right\} g\left( y_{1it}, y_{2it}, x_{it}, \tau_{i,m,k-1} \right) l_{i,k} \]

and

\[ \hat{\sigma}_{i,k} = \sum_{m=1}^{M} (\tau_{i,m,k-1})^2 \sqrt{2\hat{\sigma}_{i,k-1}} w_m^* \exp \left\{ (a_m^*)^2 \right\} g\left( y_{1it}, y_{2it}, x_{it}, \tau_{i,m,k-1} \right) l_{i,k} \]

and this is repeated until \( \hat{\mu}_{i,k} \) and \( \hat{\sigma}_{i,k} \) have converged for this iteration of the maximization algorithm. This adaptation is applied on every iteration until the log-likelihood change from the preceding iteration is less than a relative difference of 1e–6; after this, the quadrature parameters are fixed.

The log likelihood can also be calculated by nonadaptive Gauss–Hermite quadrature if the intmethod(ghermite) option is specified. For nonadaptive Gauss–Hermite quadrature, the following formula for the log likelihood is used in place of (1).

\[ L = \sum_{i=1}^{n} w_i \log f \left\{ (y_1i_1, y_2i_1), \ldots, (y_1i_n, y_2i_n) | x_{1i1}, \ldots, x_{ini} \right\} \]

\[ \approx \sum_{i=1}^{n} w_i \log \left\{ \frac{1}{\sqrt{\pi}} \sum_{m=1}^{M} w_m^* \prod_{t=1}^{n_i} F\left( y_{1it}, y_{2it}, x_{it} \beta + \sqrt{2\sigma} a_m^* \right) \right\} \]
Both quadrature formulas require that the integrated function be well approximated by a polynomial of degree equal to the number of quadrature points. Panel size can affect whether

$$\prod_{t=1}^{n_i} F(y_{1it}, y_{2it}, x_{it}\beta + \nu_i)$$

is well approximated by a polynomial. As panel size and $\rho$ increase, the quadrature approximation can become less accurate. For large $\rho$, the random-effects model can also become unidentified. Adaptive quadrature gives better results for correlated data and large panels than nonadaptive quadrature; however, we recommend that you use the `quadchk` command (see [XT] quadchk) to verify the quadrature approximation used in this command, whichever approximation you choose.

References


Also see

[XT] xtintreg postestimation — Postestimation tools for xtintreg
[XT] quadchk — Check sensitivity of quadrature approximation
[XT] xteintreg — Extended random-effects interval regression
[XT] xtreg — Fixed-, between-, and random-effects and population-averaged linear models
[XT] xtset — Declare data to be panel data
[XT] xtobit — Random-effects tobit models
[ME] meintreg — Multilevel mixed-effects interval regression
[R] intreg — Interval regression
[R] tobit — Tobit regression
[ST] stintreg — Parametric models for interval-censored survival-time data
[U] 20 Estimation and postestimation commands