Title

intro — Introduction to longitudinal/panel data manual

Description

This entry describes this manual and what has changed since Stata 9.

Remarks

This manual documents the `xt` commands and is referred to as [XT] in cross-references.

Following this entry, [XT] **xt** provides an overview of the `xt` commands. The other parts of this manual are arranged alphabetically. If you are new to Stata's `xt` commands, we recommend that you read the following sections first:

[XT] **xt**	Introduction to xt commands
[XT] **xtset**	Declare a dataset to be panel data
[XT] **xtreg**	Fixed-, between-, and random-effects, and population-averaged linear models

Stata is continually being updated, and Stata users are always writing new commands. To find out about the latest cross-sectional time-series features, type `search panel data` after installing the latest official updates; see [R] **update**.

What's new

This section is intended for previous Stata users. If you are new to Stata, you may as well skip it.

1. New command `xtset` declares a dataset to be panel data and designates the variable that identifies the panels. In previous versions of Stata, you specified options `i(`*groupvar*`)` and sometimes `t(`*timevar*`)` to identify the panels. You specified the `i()` and `t()` options on the `xt` command you wanted to use. Now you "`xtset` *groupvar*" or "`xtset` *groupvar timevar*" first. The values you set will be remembered from one session to the next if you save your dataset.

   `xtset` also provides a new feature. `xtset` allows option `delta()` to specify the frequency of the time-series data, something you will need to do if you are using Stata's new date/time variables.

   Finally, you can still specify old options `i()` and `t()`, but they are no longer documented. Similarly, old commands `iis` and `tis` continue to work but are no longer documented. See [XT] **xtset**.

2. New estimation commands `xtmelogit` and `xtmepoisson` fit nested, hierarchical, and mixed models with binary and count responses; i.e., you can fit logistic and Poisson models with complex, nested error components. Syntax is the same as for Stata's linear mixed-model estimator, `xtmixed`. To fit a model of graduation with a fixed coefficient on `x1` and random coefficient on `x2` at the school level, and with random intercepts at both the school and class-within-school level, you type

   ```
   . xtmelogit graduate x1 x2 || school: x2 || class:
   ```

   `predict` after `xtmelogit` and `xtmepoisson` will calculate predicted random effects. See [XT] **xtmelogit**, [XT] **xtmelogit postestimation**, [XT] **xtmepoisson**, and [XT] **xtmepoisson postestimation**.

3. New estimation commands are available for fitting dynamic panel-data models:

    a. Existing estimation command `xtabond` fits dynamic panel-data models by using the Arellano–Bond estimator but now reports results in levels rather than differences. Also, `xtabond` will now compute the Windmeijer biased-corrected two-step robust VCE. See [XT] **xtabond**.

    b. New estimation command `xtdpdsys` fits dynamic panel-data models by using the Arellano–Bover/Blundell–Bond system estimator. `xtdpdsys` is an extension of `xtabond` and produces estimates with smaller bias when the AR process is too persistent. `xtpdsys` is also more efficient than `xtabond`. Whereas `xtabond` uses moment conditions based on the differenced errors in producing results, `xtpdsys` uses moment conditions based on differences and levels. See [XT] **xtdpdsys**.

    c. New estimation command `xtdpd` fits dynamic panel-data models extending the Arellano–Bond or the Arellano–Bover/Blundell–Bond system estimator and allows a richer syntax for specifying models and so will fit a broader class of models then either `xtabond` or `xtdpdsys`. `xtdpd` can be used to fit models with serially correlated idiosyncratic errors, whereas `xtdpdsys` and `xtabond` assume no serial correlation. `xtdpd` can be used with models where the structure of the predetermined variables is more complicated than that assumed by `xtdpdsys` or `xtabond`. See [XT] **xtdpd**.

    d. New postestimation command `estat abond` tests for serial correlation in the first-differenced errors. See [XT] **xtabond postestimation**, [XT] **xtdpdsys postestimation**, and [XT] **xtdpd postestimation**.

    e. New postestimation command `estat sargan` performs the Sargan test of overidentifying restrictions. See [XT] **xtabond postestimation**, [XT] **xtdpdsys postestimation**, and [XT] **xtdpd postestimation**.

4. Existing estimation command `xtreg, fe` now accepts `aweights`, `fweights`, and `pweights`. Also, new option `dfadj` specifies that the cluster–robust VCE be adjusted for the within transform. This was previously the default behavior. See [XT] **xtreg**.

5. Existing estimation commands `xtreg, fe` and `xtreg, re` used to be willing to produce cluster–robust VCEs when the panels were not nested within the clusters. Sometimes this VCE is consistent and other times it is not. You must now specify the new `nonest` option to obtain a cluster–robust VCE when the panels are not nested within the clusters.

6. The numerical method used to evaluate distributions, known as quadrature, has been improved. This method is used by the `xt` random-effects estimation commands `xtlogit`, `xtprobit`, `xtcloglog`, `xtintreg`, `xttobit`, and `xtpoisson, re normal`.

    a. For the estimation commands, the default method is now `intmethod(mvaghermite)`. The old default was `intmethod(aghermite)`.

    b. Option `intpoints(#)` for the commands now allows up to 195 quadrature points. The default is 12, and the old upper limit was 30. (Models with large random effects often require more quadrature points.)

    c. The estimation commands may now be used with constraints regardless of the quadrature method chosen.

    d. Command `quadchk`, for use after estimation to verify that the quadrature approximation was sufficiently accurate, now produces a more informative comparison table. Before, four fewer and four more quadrature points were used, and that was reasonable if the number of quadrature points was, say, $n_q = 12$. Now you may specify significantly larger n_q and the ± 4 is not useful. Now `quadchk` uses $n_q - \texttt{int}(n_q/3)$ and $n_q + \texttt{int}(n_q/3)$.

e. `quadchk` has new option `nofrom` that forces refitted models to start from scratch rather than starting from the previous estimation results. This is important if you use the old `intmethod(aghermite)`, which is sensitive to starting values, but not important if you are using the new default `intmethod(mvaghermite)`.

See [XT] **quadchk**.

7. All `xt` estimation commands now accept option `vce(`*vcetype*`)`. As mentioned in the [U] **1.3.3 What's new in statistics (general)**, `vce(robust)` and `vce(cluster` *varname*`)` are the right ways to specify the old `robust` and `cluster()` options, and option `vce()` allows other VCE calculations as well.
8. Existing estimation command `xtcloglog` has new option `eform` that requests exponentiated coefficients be reported; see [XT] **xtcloglog**.
9. Existing estimation command `xthtaylor` now allows users to specify only endogenous time-invariant variables, only endogenous time-varying variables, or both. Previously, both were required. See [XT] **xthtaylor**.
10. Most `xt` estimation commands have new option `collinear`, which specifies that collinear variables are not to be removed. Typically, you do not want to specify this option. It is for use when you specify constraints on the coefficients such that, even though the variables are collinear, the model is fully identified. See [XT] **estimation options**.
11. Existing command `xtdes` has been renamed to `xtdescribe`. `xtdes` continues to work as a synonym for `xtdescribe`. See [XT] **xtdescribe**.
12. The [XT] manual has an expanded glossary.

For a complete list of all new features in Stata 10, see [U] **1.3 What's new**.

Also See

[U] **1.3 What's new**

[R] **intro** — Introduction to base reference manual