tsfill — Fill in gaps in time variable

Syntax

    tsfill [, full]

You must `tsset` your data before using `tsfill`; see [TS] tsset.

Menu

Statistics > Time series > Setup and utilities > Fill in gaps in time variable

Description

`tsfill` is used after `tsset` to fill in gaps in time-series data and gaps in panel data with new observations, which contain missing values. For instance, perhaps observations for `timevar = 1, 3, 5, 6, . . . , 22` exist. `tsfill` would create observations for `timevar = 2` and `timevar = 4` containing all missing values. There is seldom reason to do this because Stata’s time-series operators consider `timevar`, not the observation number. Referring to `L.gnp` to obtain lagged `gnp` values would correctly produce a missing value for `timevar = 3`, even if the data were not filled in. Referring to `L2.gnp` would correctly return the value of `gnp` in the first observation for `timevar = 3`, even if the data were not filled in.

Option

`full` is for use with panel data only. With panel data, `tsfill` by default fills in observations for each panel according to the minimum and maximum values of `timevar` for the panel. Thus if the first panel spanned the times 5–20 and the second panel the times 1–15, after `tsfill` they would still span the same periods; observations would be created to fill in any missing times from 5–20 in the first panel and from 1–15 in the second.

If `full` is specified, observations are created so that both panels span the time 1–20, the overall minimum and maximum of `timevar` across panels.

Remarks and examples

Remarks are presented under the following headings:

- Using `tsfill` with time-series data
- Using `tsfill` with panel data
- Video example
Using tsfill with time-series data

You have monthly data, with gaps:

. use http://www.stata-press.com/data/r13/tsfillxmpl
. tsset
time variable: mdate, 1995m7 to 1996m3, but with gaps
delta: 1 month
. list mdate income

<table>
<thead>
<tr>
<th>mdate</th>
<th>income</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995m7</td>
<td>1153</td>
</tr>
<tr>
<td>1995m8</td>
<td>1181</td>
</tr>
<tr>
<td>1995m11</td>
<td>1236</td>
</tr>
<tr>
<td>1995m12</td>
<td>1297</td>
</tr>
<tr>
<td>1996m1</td>
<td>1265</td>
</tr>
<tr>
<td>1996m3</td>
<td>1282</td>
</tr>
</tbody>
</table>

You can fill in the gaps by interpolation easily with tsfill and ipolate. tsfill creates the missing observations:

. tsfill
. list mdate income

<table>
<thead>
<tr>
<th>mdate</th>
<th>income</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995m7</td>
<td>1153</td>
</tr>
<tr>
<td>1995m8</td>
<td>1181</td>
</tr>
<tr>
<td>1995m9</td>
<td>.</td>
</tr>
<tr>
<td>1995m10</td>
<td>.</td>
</tr>
<tr>
<td>1995m11</td>
<td>1236</td>
</tr>
<tr>
<td>1996m1</td>
<td>1265</td>
</tr>
<tr>
<td>1996m2</td>
<td>.</td>
</tr>
<tr>
<td>1996m3</td>
<td>1282</td>
</tr>
</tbody>
</table>

We can now use ipolate (see [D] ipolate) to fill them in:

. ipolate income mdate, gen(ipinc)
. list mdate income ipinc

<table>
<thead>
<tr>
<th>mdate</th>
<th>income</th>
<th>ipinc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995m7</td>
<td>1153</td>
<td>1153</td>
</tr>
<tr>
<td>1995m8</td>
<td>1181</td>
<td>1181</td>
</tr>
<tr>
<td>1995m9</td>
<td>.</td>
<td>1199.3333</td>
</tr>
<tr>
<td>1995m10</td>
<td>.</td>
<td>1217.6667</td>
</tr>
<tr>
<td>1995m11</td>
<td>1236</td>
<td>1236</td>
</tr>
<tr>
<td>1995m12</td>
<td>1297</td>
<td>1297</td>
</tr>
<tr>
<td>1996m1</td>
<td>1265</td>
<td>1265</td>
</tr>
<tr>
<td>1996m2</td>
<td>.</td>
<td>1273.5</td>
</tr>
<tr>
<td>1996m3</td>
<td>1282</td>
<td>1282</td>
</tr>
</tbody>
</table>
Using tsfill with panel data

You have the following panel dataset:

```
. use http://www.stata-press.com/data/r13/tsfillxmpl2, clear
. tsset
    panel variable:  edlevel (unbalanced)
    time variable:  year, 1988 to 1992, but with a gap
    delta:  1 unit
. list edlevel year income
    edlevel year income
    1.  1 1988  14500
    2.  1 1989  14750
    3.  1 1990  14950
    4.  1 1991  15100
    5.  2 1989  22100
    6.  2 1990  22200
    7.  2 1992  22800
```

Just as with nonpanel time-series datasets, you can use `tsfill` to fill in the gaps:

```
. tsfill
. list edlevel year income
    edlevel year income
    1.  1 1988  14500
    2.  1 1989  14750
    3.  1 1990  14950
    4.  1 1991  15100
    5.  2 1989  22100
    6.  2 1990  22200
    7.  2 1991  .   ← new
    8.  2 1992  22800
```

You could instead use `tsfill` to produce fully balanced panels with the `full` option:

```
. tsfill, full
. list edlevel year income, sep(0)
    edlevel year income
    1.  1 1988  14500
    2.  1 1989  14750
    3.  1 1990  14950
    4.  1 1991  15100
    5.  1 1992  .   ← new
    6.  2 1988  .   ← new
    7.  2 1989  22100
    8.  2 1990  22200
    9.  2 1991  .   ← new
   10.  2 1992  22800
```
Video example

Time series, part 1: Formatting dates, tsset, tsreport, and tsfill

Also see

[TS] tsset — Declare data to be time-series data
[TS] tsappend — Add observations to a time-series dataset