

tsfill — Fill in gaps in time variable

[Description](#)
[Option](#)[Quick start](#)
[Remarks and examples](#)[Menu](#)
[Also see](#)[Syntax](#)

Description

`tsfill` is used to fill in gaps in time-series data and gaps in panel data with new observations, which contain missing values. `tsfill` is not needed to obtain correct lags, leads, and differences when gaps exist in a series because Stata's time-series operators handle gaps automatically.

Quick start

Add new observations with missing values for missing time periods in a time-series dataset that has been `tsset`

```
tsfill
```

Add new observations with missing values to eliminate gaps in a panel dataset that has been `xtset`

```
tsfill
```

As above, but making the panel strongly balanced

```
tsfill, full
```

Menu

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

Syntax

```
tsfill [ , full ]
```

You must `tsset` or `xtset` your data before using `tsfill`; see [TS] `tsset` and [XT] `xtset`.

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

[stata.com](https://www.stata.com)

Remarks are presented under the following headings:

[Introduction](#)

[Using `tsfill` with time-series data](#)

[Using `tsfill` with panel data](#)

[Video example](#)

Introduction

`tsfill` is used after `tsset` or `xtset` to fill gaps in time-series data and gaps in panel data with new observations. Each new observation contains the appropriate values of the time variable, *timevar*, and, when specified, the panel variable, *panelvar*, and missing values for all other variables in the dataset. 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.

`tsfill` is intended as an intermediate step in a data management process. For example, you may wish to use `tsfill` with time-series data if you plan to interpolate missing values or with panel data if you intend to impute missing values.

You do not need to use `tsfill` to correctly create variables with lags, leads, and differencing, because Stata's time-series operators handle gaps in the series for you; see [U] [11.4.4 Time-series varlists](#). These operators consider *timevar*, not the observation number. For example, suppose we have data on GNP in the years 1989–1991 and 1993–1995. Referring to `L.gnp` to obtain lagged `gnp` values would correctly produce a missing value of lagged `gnp` for *timevar* = 1989 and *timevar* = 1993 even if missing values were not explicitly created using `tsfill`.

Using tsfill with time-series data

You have monthly data, with gaps:

```
. use http://www.stata-press.com/data/r15/tsfillxmpl
. tsset
      time variable:  mdate, 1995m7 to 1996m3, but with gaps
      delta:         1 month
. list mdate income
```

	mdate	income
1.	1995m7	1153
2.	1995m8	1181
3.	1995m11	1236
4.	1995m12	1297
5.	1996m1	1265
6.	1996m3	1282

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

```
. tsfill
. list mdate income
```

	mdate	income	
1.	1995m7	1153	
2.	1995m8	1181	
3.	1995m9	.	← new
4.	1995m10	.	← new
5.	1995m11	1236	
6.	1995m12	1297	
7.	1996m1	1265	
8.	1996m2	.	← new
9.	1996m3	1282	

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

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

	mdate	income	ipinc
1.	1995m7	1153	1153
2.	1995m8	1181	1181
3.	1995m9	.	1199.3333
4.	1995m10	.	1217.6667
5.	1995m11	1236	1236
6.	1995m12	1297	1297
7.	1996m1	1265	1265
8.	1996m2	.	1273.5
9.	1996m3	1282	1282

Using tsfill with panel data

You have the following panel dataset:

```
. use http://www.stata-press.com/data/r15/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 within each panel:

```
. 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	.
8.	2	1992	22800

← new

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	.
6.	2	1988	.
7.	2	1989	22100
8.	2	1990	22200
9.	2	1991	.
10.	2	1992	22800

← new
← new
← new

Video example

[Formatting and managing dates](#)

Also see

[TS] [tsappend](#) — Add observations to a time-series dataset

[TS] [tsset](#) — Declare data to be time-series data

[XT] [xtset](#) — Declare data to be panel data