AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

The Stata Journal

Editor
H. Joseph Newton
Department of Statistics
Texas A & M University
College Station, Texas 77843
979-845-3142; FAX 979-845-3144
jnewton@stata-journal.com

Editor
Nicholas J. Cox
Department of Geography
Durham University
South Road
Durham City DH1 3LE UK
n.j.cox@stata-journal.com

Associate Editors

Christopher F. Baum, Boston College
Rino Bellocco, Karolinska Institutet, Sweden and Univ. degli Studi di Milano-Bicocca, Italy
A. Colin Cameron, University of California–Davis
David Clayton, Cambridge Inst. for Medical Research
Mario A. Cleves, Univ. of Arkansas for Medical Sciences
William D. Dupont, Vanderbilt University
Charles Franklin, University of Wisconsin–Madison
Joanne M. Garrett, University of North Carolina
Allan Gregory, Queen's University
James Hardin, University of South Carolina
Ben Jann, ETH Zürich, Switzerland
Stephen Jenkins, University of Essex
Ulrich Kohler, WZB, Berlin
Jens Lauritsen, Odense University Hospital
Stanley Lemeshow, Ohio State University
J. Scott Long, Indiana University
Thomas Lumley, University of Washington–Seattle
Roger Newson, Imperial College, London
Marcello Pagano, Harvard School of Public Health
Sophia Rabe-Hesketh, University of California–Berkeley
J. Patrick Royston, MRC Clinical Trials Unit, London
Philip Ryan, University of Adelaide
Mark E. Schaffer, Heriot-Watt University, Edinburgh
Jeroen Weesie, Utrecht University
Nicholas J. G. Winter, University of Virginia
Jeffrey Wooldridge, Michigan State University

Stata Press Production Manager Lisa Gilmore
Stata Press Copy Editor Gabe Waggoner

Copyright Statement: The Stata Journal and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the Stata Journal.

The articles appearing in the Stata Journal may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the Stata Journal.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the Stata Journal, in whole or in part, on publicly accessible web sites, fileservers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the Stata Journal or the supporting files understand that such use is made without warranty of any kind, by either the Stata Journal, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the Stata Journal is to promote free communication among Stata users.

The *Stata Journal*, electronic version (ISSN 1536-8734) is a publication of Stata Press. Stata and Mata are registered trademarks of StataCorp LP.

The Stata Journal (2007)
7, Number 1, pp. 137–139

Stata tip 40: Taking care of business

Christopher F. Baum
Department of Economics
Boston College
Chestnut Hill, MA 02467
baum@bc.edu

Daily data are often generated by nondaily processes: for instance, financial markets are closed on weekends and holidays. Stata's time-series date schemes ([U] **24.3 Time-series dates**) allow for daily data, but gaps in time series may be problematic. A model that uses lags or differences will lose several observations every time a gap appears, discarding many of the original data points. Analysis of "business-daily" data often proceeds by assuming that Monday follows Friday, and so on. At the same time, we usually want data to be placed on Stata's time-series calendar so that useful tools such as the `tsline` graph will work and label data points with readable dates; see [TS] **tsline**.

At a recent Stata Users Group presentation in Boston, David Drukker spoke to this point. His solution: generate two date variables, one containing the actual calendar dates, another numbering successive available observations consecutively. The former variable (`caldate`) is `tsset` (see [TS] **tsset**) when the calendar dates are to be used, whereas the latter (`seqdate`) is `tsset` when statistical analyses are to be performed.

We download daily data on the 3-month U.S. Treasury bill rate with Drukker's `freduse` command (Drukker 2006) and retain the August 2005–present data for analysis. (We can also view the data graphically with `tsline`.)

```
. freduse DTB3
(13764 observations read)
. rename daten caldate
. tsset caldate
        time variable:  caldate, 04jan1954 to 05oct2006, but with gaps
. keep if tin(1aug2005,)
(13455 observations deleted)
. label var caldate date
```

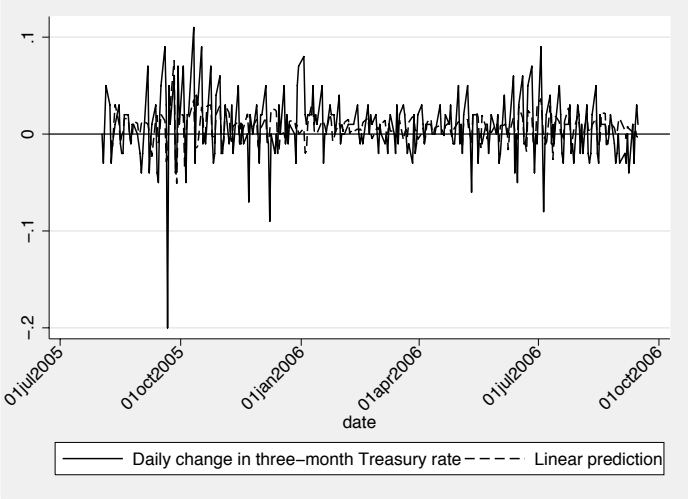
These data do not contain observations for weekends and are missing for U.S. holidays. We may not want to drop the observations containing missing data, though, as we may have complete data for other variables: for instance, exchange rate data are available every day. If there were no missing data in our series—only missing observations—we could use Drukker's suggestion and `generate seqdate = _n`. As we have observations for which `DTB3` is missing, we follow a more complex route:

```
. quietly generate byte notmiss = DTB3 < .
. quietly generate seqdate = cond(notmiss, sum(notmiss),.)
. tsset seqdate
        time variable:  seqdate, 1 to 297
```

© 2007 StataCorp LP

The variable `seqdate` is created as the sequential day number for every nonmissing day and is itself missing when `DTB3` is missing—allowing us to use this variable in `tsset` and then use time-series operators (see [U] **11.1.1 varlist**) in `generate` or estimation commands such as `regress`. We may want to display the transformed data (or results from estimation, such as predicted values) on a time-series graph. We can just revert to the other `tsset`:

```
. quietly generate dDTB3 = D.DTB3
. quietly regress dDTB3 L(1/5).dDTB3
. predict double dDTB3hat, xb
(18 missing values generated)
. label var dDTB3 "Daily change in three-month Treasury rate"
. tsset caldate
        time variable:  caldate, 01aug2005 to 05oct2006, but with gaps
. tsline dDTB3 dDTB3hat, yline(0) xlabel(, angle(45))
```



If we retain both the `caldate` and `seqdate` variables in our saved dataset, we will always be able to view these data either on a time-series calendar or as a sequential series. In my research, I need to know how many calendar days separate each observed point (1 for Thursday–Friday but 3 for Friday–Monday) and then sum `DTB3` by month, weighting each observation by the square root of the days of separation:

```
. tsset seqdate
        time variable:  seqdate, 1 to 297
. quietly generate dcal = D.caldate if seqdate < .
. quietly generate month = mofd(caldate) if seqdate < .
. format %tm month
. sort month (seqdate)
. quietly by month: generate adjchange = sum(dDTB3/sqrt(dcal))
. quietly by month: generate sumchange = adjchange if _n==_N & month < .
```

```
. list month sumchange if sumchange < ., sep(0) noobs
```

month	sumchange
2005m8	-.003812
2005m9	-.0810769
2005m10	.2424316
2005m11	-.063453
2005m12	.096188
2006m1	.2769615
2006m2	.099641
2006m3	.0142265
2006m4	.0938675
2006m5	.0350555
2006m6	.0327906
2006m7	.0304485
2006m8	-.083812
2006m9	-.123094
2006m10	.0442265

References

Drukker, D. M. 2006. Importing Federal Reserve economic data. *Stata Journal* 6: 384–386.