The Stata Journal (2009)
9, Number 2, pp. 321–326

Stata tip 76: Separating seasonal time series

Nicholas J. Cox
Department of Geography
Durham University
Durham, UK
n.j.cox@durham.ac.uk

Many researchers in various sciences deal with seasonally varying time series. The part rhythmic, part random character of much seasonal variation poses several graphical challenges for them. People usually want to see both the broad pattern and the fine structure of trends, seasonality, and any other components of variation. The very common practice of using just one plot versus date typically yields a saw-tooth or roller-coaster pattern as the seasons repeat. That method is often good for showing broad trends, but not so good for showing the details of seasonality. I reviewed several alternative graphical methods in a *Speaking Stata* column (Cox 2006). Here is yet another method, which is widely used in economics. Examples of this method can be found in Hylleberg (1986, 1992), Ghysels and Osborn (2001), and Franses and Paap (2004).

The main idea is remarkably simple: plot separate traces for each part of the year. Thus, for each series, there would be 2 traces for half-yearly data, 4 traces for quarterly data, 12 traces for monthly data, and so on. The idea seems unlikely to work well for finer subdivisions of the year, because there would be too many traces to compare. However, quarterly and monthly series in particular are so common in many fields that the idea deserves some exploration.

One of the examples in Franses and Paap (2004) concerns variations in an index of food and tobacco production for the United States for 1947–2000. I downloaded the data from http://people.few.eur.nl/paap/pbook.htm (this URL evidently supersedes those specified by Franses and Paap [2004, 12]) and named it `ftp`. For what follows, year and quarter variables are required, as well as a variable holding quarterly dates.

```
. egen year = seq(), from(1947) to(2000) block(4)
. egen quarter = seq(), to(4)
. gen date = yq(year, quarter)
. format date %tq
. tsset date
. gen growth = D1.ftp/ftp
```

Although a line plot is clearly possible, a scatterplot with marker labels is often worth trying first (figure 1). See an earlier tip by Cox (2005) for more examples.

 gr0037

```
. scatter growth year, ms(none) mla(quarter) mlabpos(0)
```

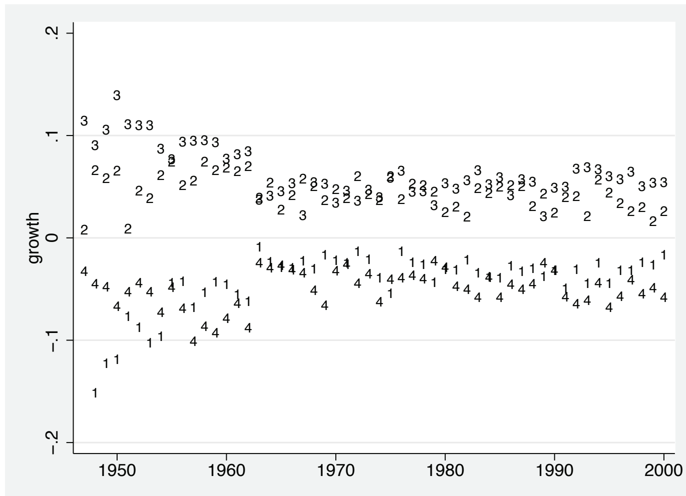

Figure 1. Year-on-year growth by quarter for food and tobacco production in the United States: separate series

Immediately, we see some intriguing features in the data. There seems to be a discontinuity in the early 1960s, which may reflect some change in the basis of calculating the index, rather than a structural shift in the economy or the climate. Note also that the style and the magnitude of seasonality change: look in detail at traces for quarters 1 and 4. No legend is needed for the graph, because the marker labels are self-explanatory. Compare this graph with the corresponding line plot given by Franses and Paap (2004, 15).

In contrast, only some of the same features are evident in more standard graphs. The traditional all-in-one line plot (figure 2) puts seasonality in context but is useless for studying detailed changes in its nature.

```
. tsline ftp
```

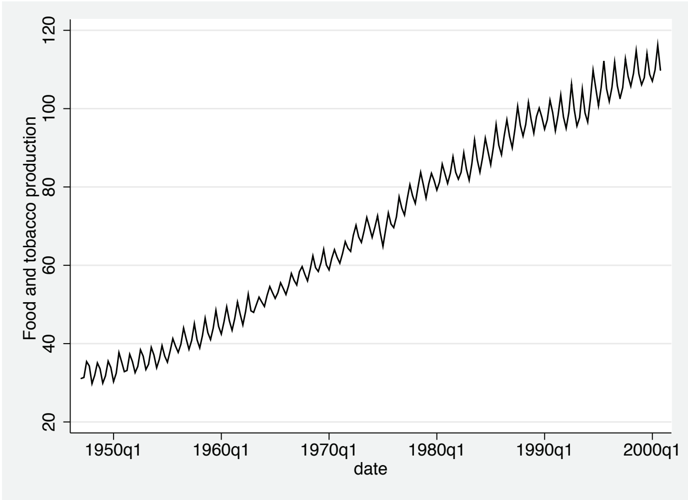

Figure 2. Quarterly food and tobacco production in the United States

The apparent discontinuity in the early 1960s is, however, clear in a plot of growth rate versus date (figure 3).

```
. tsline growth
```

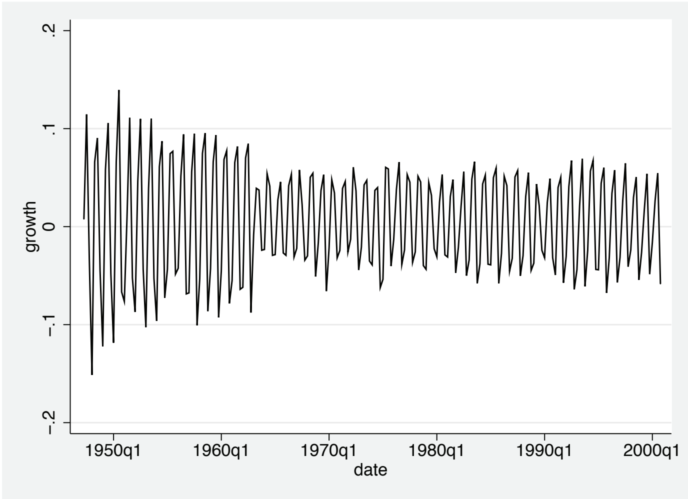

Figure 3. Year-on-year growth by quarter for food and tobacco production in the United States: combined series

An example with monthly data will push harder at the limits of this device. Grubb and Mason (2001) examined monthly data on air passengers in the United Kingdom for 1947–1999. The data can be found at http://people.bath.ac.uk/mascc/Grubb.TS; also see Chatfield (2004, 289–290). We will look at seasonality as expressed in monthly shares of annual totals (figure 4). The graph clearly shows how seasonality is steadily becoming more subdued.

```
. egen total = total(passengers), by(year)
. gen percent = 100 * passengers / total
. gen symbol = substr("123456789OND", month, 1)
. scatter percent year, ms(none) mla(symbol) mlabpos(0) mlabsize(*.8) xtitle("")
> ytitle(% in each month) yla(5(5)15)
```

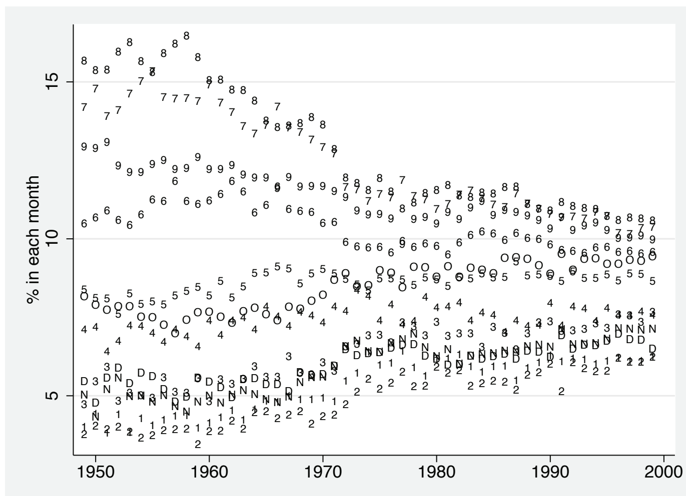

Figure 4. Monthly shares of UK air passengers, 1947–1999 (digits 1–9 indicate January–September; O, N, and D indicate October–December)

Because some users will undoubtedly want line plots, how is that to be done? The `separate` command is useful here: see Cox (2005), [D] **separate**, or the online help. Once we have separate variables, they can be used with the `line` command (figure 5).

```
. separate percent, by(month) veryshortlabel
. line percent1-percent12 year, xtitle("") ytitle(% in each month) yla(5(5)15)
> legend(pos(3) col(1))
```

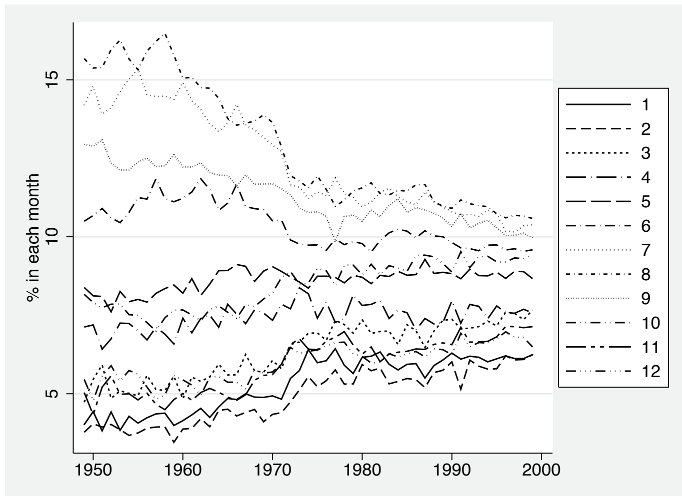

Figure 5. Monthly shares of UK air passengers, 1947–1999

You may think that the graph needs more work on the line patterns (and thus the legend), although perhaps now the scatterplot with marker labels seems a better possibility.

If graphs with 12 monthly traces seem too busy, one trick worth exploring is subdividing the year into two, three, or four parts and using separate panels in a `by()` option. Then each panel would have only six, four, or three traces.

References

Chatfield, C. 2004. *The Analysis of Time Series: An Introduction*. 6th ed. Boca Raton, FL: Chapman & Hall/CRC.

Cox, N. J. 2005. Stata tip 27: Classifying data points on scatter plots. *Stata Journal* 5: 604–606.

———. 2006. Speaking Stata: Graphs for all seasons. *Stata Journal* 6: 397–419.

Franses, P. H., and R. Paap. 2004. *Periodic Time Series Models*. Oxford: Oxford University Press.

Ghysels, E., and D. R. Osborn. 2001. *The Econometric Analysis of Seasonal Time Series*. Cambridge: Cambridge University Press.

Grubb, H., and A. Mason. 2001. Long lead-time forecasting of UK air passengers by Holt–Winters methods with damped trend. *International Journal of Forecasting* 17: 71–82.

Hylleberg, S. 1986. *Seasonality in Regression*. Orlando, FL: Academic Press.

Hylleberg, S., ed. 1992. *Modelling Seasonality*. Oxford: Oxford University Press.

The Stata Journal (2009)
9, Number 2, p. 327

Software Updates

gr0033_1: Contour-enhanced funnel plots for meta-analysis T. M. Palmer, J. L. Peters, A. J. Sutton, and S. G. Moreno. *Stata Journal* 8: 242–254.

The `confunnel` command to draw contour-enhanced funnel plots has been updated to produce plots with shaded regions of statistical significance. By default, the plots now resemble those in Peters et al., Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry, *Journal of Clinical Epidemiology*, 2008, 61: 991–996.

sbe24_3: metan: fixed- and random-effects meta-analysis. R. Harris, M. Bradburn, J. Deeks, R. Harbord, D. Altman, and J. Sterne. *Stata Journal* 8: 3–28; *Stata Technical Bulletin* 45: 21; *Stata Technical Bulletin* 44: 4–15. Reprinted in *Stata Technical Bulletin Reprints* vol. 8, pp. 86–100.

This contains two fixes:

- `by()` variables were being converted to string format
- the predictive interval option `rfdist` had an error when used for nonratio measures

st0096_2: GLS for trend estimation of summarized dose–response data. N. Orsini, R. Bellocco, and S. Greenland. *Stata Journal* 9: 173; *Stata Journal* 6: 40–57.

The updated version of `glst` makes sure the dataset is sorted by the study identification variable when pooling multiple studies.

st0143_2: The Stata command felsdvreg to fit a linear model with two high-dimensional fixed effects. T. Cornelissen. *Stata Journal* 9: 173; *Stata Journal* 8: 170–189.

A bug in the `grouponly` option of the program has been fixed. This option was previously sensitive to the sort order of the data and produced a correct grouping variable only if the user had sorted the data by the two fixed effect identifiers. The option now works regardless of the sort order of the data.

st0152_1: The Blinder–Oaxaca decomposition for nonlinear regression models. M. Sinning, M. Hahn, and T. K. Bauer. *Stata Journal* 8: 480–492.

Unweighted sample means were calculated, even after the specification of weights or the use of `svy` commands. This affected the estimates of the raw gap of the decomposition and the resulting contribution of its components. The calculation of sample means has now been adjusted for the use of weights and `svy` commands.

 up0025