

Paper 417-2011

Expanding PROC EXPAND

Bruce Gilson, Federal Reserve Board, Washington, DC

ABSTRACT

The EXPAND procedure, which is part of SAS/ETS® software, provides many types of data transformation for time series data. This paper shows two examples where getting the desired results requires some manipulation of the data before or after use of PROC EXPAND: calculating an un-weighted quarterly average from monthly data and converting weekly data to a monthly average when the weekly data values represent values for seven days.

INTRODUCTION

The EXPAND procedure, which is part of SAS/ETS software, provides many types of data transformation for time series data, including the following:

- aggregating time series frequency from higher to lower (e.g., monthly to quarterly)
- interpolating time series frequency from lower to higher (e.g., quarterly to monthly)
- interpolating missing values
- performing various transformations that can be applied before or after PROC EXPAND (e.g., centered and backward moving time window operators such as moving averages and moving products, logarithmic transformations, and many others)

This paper shows two examples where the direct use of PROC EXPAND did not yield the desired result. Some manipulation of the data before or after use of PROC EXPAND was required to do the following:

- calculate an un-weighted quarterly average (the average of the three monthly values) from monthly data
- convert weekly data to a monthly average when the weekly values represent values for seven days

Calculate an un-weighted quarterly average from monthly data

The following is a small subset of data set ONE. DATE is a SAS date value and ID and INCOME are numeric. The data set is sorted by ID.

Obs	date	id	income
1	20070101	1	1
2	20070201	1	2
3	20070301	1	15
4	20070401	1	4
5	20070501	1	5
6	20070601	1	12
7	20070701	1	7
8	20070801	1	8
9	20070901	1	9

We want to convert monthly values to quarterly by taking the un-weighted average of the three monthly values. For example, for the third quarter of 2007 we want $(7+8+9) / 3 = 8$. Here is initial PROC EXPAND code that does not yield the desired result.

```
proc expand data=one out=two from=month to=quarter ;
  id date;
  by id;
  convert income / method=aggregate observed=average;
run;
```

Obs	id	date	income	income (expected result)
1	1	2007:1	6.1333	6
2	1	2007:2	6.9780	7
3	1	2007:3	7.9891	8

The problem is that PROC EXPAND weights the monthly values by the number of days in each month. For example, for the third quarter of 2007, PROC EXPAND does the following calculation. Note that 31 or 30 are the number of days in each month and 92 is the number of days in the quarter.

$$((31 * 7) + (31 * 8) + (30 * 9)) / 92 = 7.9891$$

PROC EXPAND cannot calculate an un-weighted quarterly average, but one simple solution is to calculate quarterly totals with PROC EXPAND and then divide by three in a DATA step to get quarterly averages, as in the following code.

```
proc expand data=one out=two from=month to=quarter ;
  id date;
  by id;
  convert income / method=aggregate observed=total;
run;

data three;
  set two;
  income=income/3; /* convert from total to quarterly average */
run;
```

Convert weekly data to a monthly average when the weekly values represent values for seven days

The following is a small subset of data set ONE. DATE is a SAS date value and ID and INCOME are numeric. The data set is sorted by ID. The data are recorded on Wednesdays and each value represents the value for 7 days up to and including the current day. For example, INCOME is 30 on May 31 - June 6, 2007.

Obs	date	id	income
1	20070530	1	11
2	20070606	1	30
3	20070613	1	50
4	20070620	1	70
5	20070627	1	90
6	20070704	1	50
7	20070711	1	40
8	20070718	1	120
9	20070725	1	80
10	20070801	1	100
11	20070808	1	5
12	20070815	1	7
13	20070822	1	9
14	20070829	1	11
15	20070905	1	24

We want to take monthly averages of the weekly data, but the initial PROC EXPAND code does not yield the desired result.

```
proc expand data=one out=two from=week to=month ;
  id date;
  by id;
  convert income / observed=average;
run;
```

Obs	id	date	income	income (expected result)
1	1	JUN2007	57.093	60
2	1	JUL2007	75.905	80
3	1	AUG2007	18.852	12

The problem is that there is no way to specify that the data values represent 7 consecutive days, and PROC EXPAND uses interpolation to calculate the average.

One solution is to transform weekly data to daily data in a DATA step, and then convert daily data to monthly averages with PROC EXPAND, as in the following code.

```
data two;
  set one;
  do dailydate = date-6 to date;
    output; /* transform each weekly observation to 7 daily observations */
  end;
run;

proc expand data=two out=three from=day to=month ;
  id dailydate;
  by id;
  convert income / observed=average;
run;
```

Note that the DATA step transforms the 15 weekly values in data set ONE into 105 daily values in data set TWO. Here are the first 25 values in data set TWO.

Obs	dailydate	id	income
1	20070524	1	11
2	20070525	1	11
3	20070526	1	11
4	20070527	1	11
5	20070528	1	11
6	20070529	1	11
7	20070530	1	11
8	20070531	1	30
9	20070601	1	30
10	20070602	1	30
11	20070603	1	30
12	20070604	1	30
13	20070605	1	30
14	20070606	1	30
15	20070607	1	50
16	20070608	1	50
17	20070609	1	50
18	20070610	1	50
19	20070611	1	50
20	20070612	1	50
21	20070613	1	50
22	20070614	1	70
23	20070615	1	70
24	20070616	1	70
25	20070617	1	70

Alternate solutions

Two alternate ways to solve the two problems discussed in this paper are with a DATA step or with PROC SUMMARY. Disadvantages of these methods are as follows.

- DATA step code would be somewhat more complicated than PROC EXPAND.
- Unlike PROC EXPAND, PROC SUMMARY would include output for partial periods (i.e., May 2007 in the second example) that must be removed in a subsequent step. Also, for the second example, DATA step code to create daily data similar to the code used for the solution shown above would be needed.

CONCLUSION

PROC EXPAND provides many types of useful data transformations for time series data. This paper showed two cases where some modest manipulation of the data before or after the use of PROC EXPAND allowed us to transform data in ways not directly provided by PROC EXPAND.

For more information, contact the author, Bruce Gilson, by mail at Federal Reserve Board, Mail Stop 157, Washington, DC 20551; by e-mail at bruce.gilson@frb.gov; or by phone at 202-452-2494.

REFERENCES

Gilson, Bruce (2009), "Tales from the Help Desk 4: Still Solutions for Common SAS Mistakes," *Proceedings of the SAS Global Forum 2010 Conference*. <<http://support.sas.com/resources/papers/proceedings09/137-2009.pdf>>

SAS Institute Inc. (2004), "SAS/ETS 9.1.3 User's Guide," Cary, NC: SAS Institute Inc.

ACKNOWLEDGMENTS

The following people contributed extensively to the development of this paper: Heidi Markovitz and Donna Hill at the Federal Reserve Board. Their support is greatly appreciated.

TRADEMARK INFORMATION

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.