

Paper 268-2012

PROC EXPAND and SGPLOT: The Dynamic Duo of Time Series Data Display

Andrew H. Karp

Sierra Information Services
Sonoma, California USA

ABSTRACT

Many reporting and analysis applications require preparation and display of data values across or along a time horizon. These representations are often referred to as a “time series” or a “realization of a process in the time domain.” The displays inform the reader about changes in the data values, suggest trends in the series over time, highlight seasonal changes in the series and might also show the impact of abrupt changes to the process under analysis, often referred to as “shocks” or “interventions.” Effective display of time series data is not only an essential aspect of a reporting or analytics process, but is also the first step in developing statistical models to forecast future values of the series as a function of the prior observed data. But, many available time series need substantial manipulation before they are ready for display. The series may have missing data points, require aggregation or interpolation of values, or have moving time window statistics applied to them to create the final displayed series. Regardless of the project’s ultimate aim, the display must “tell the story” in an effective, compelling and easy to understand way. This paper shows how to combine the powerful data preparation/manipulation tools available in SAS® Software’s EXPAND Procedure and the data display features in PROC SGPLOT to overcome many problems faced in the development and display of time series data.

INTRODUCTION

Effective time series data display is an integral part of many reporting and analysis projects. An effective graphic showing the changes of data values over time can “tell the story” of your data to others in your organization. But, the “real world” of data presents many challenges to creating a compelling time series data display. Values may be missing from the series, or data must be aggregated from a lower collection frequency (e.g., hourly) to an estimated higher frequency (daily). In some instances your available data must be interpolated from a lower collection frequency (e.g., monthly) to a higher estimated frequency (weekly). And, some applications require application of a “moving time window” statistic to calculate a moving average, moving sum or other value that is derived from multiple adjacent values in the observed series.

PROC EXPAND, in the SAS/ETS® module, is uniquely suited to perform these and other tasks on your time series data. As this paper demonstrates, it offers a powerful array of data preparation tools that would otherwise be unavailable or require extensive programming in a Data Step to implement. As a “Stat Graph enabled” procedure as of the release of SAS 9.2 Software, EXPAND can now create a series of graphics/displays that can be delivered to Output Delivery System (ODS) “destinations” in addition to, or instead of, its default creation of a SAS data set containing the results of its “work” on your time series data.

PROC SGPLOT, added to SAS/GRAPH® Software in SAS 9.2 and as of 9.3 in the BASE Module, offers a wider array of tools, options and features to generate and customize the display of time series data values than are available from the suite of ODS statistical graphics available in PROC EXPAND. Combining SGPLOT’s data display capabilities, then, with the unparalleled data preparation/manipulation tools in PROC EXPAND provide you with the “dynamic duo” of tools to prepare your displays of your time series data quickly, effectively, and with limited coding/syntax.

EXAMPLES

The presentation scheduled for delivery at SAS Global Forum (SGF) 2012 will show a range of ways to combine PROC EXPAND and PROC SGPLOT capabilities that address many different types of data preparation and display challenges. Please visit <http://www.sierrainformation.com> to download a PDF of the SGF presentation, which will include SAS code samples and example output. The examples to be presented include:

- Replacement of missing values in a time series using PROC EXPAND, followed by display of the “corrected” series using the SERIES statement in PROC SGPLOT
- Overlay of a SERIES and SCATTER plot on the same SGPLOT-generated display showing the actual, non-missing values of the series and the interpolated replacement values generated by PROC EXPAND

PROC EXPAND and SGPLOT: The Dynamic Duo of Time Series Data Display, *continued*

- Aggregation of a time series with missing data from month to calendar quarter using PROC EXPAND, followed by its display using PROC SGPLOT
- Interpolation of weekly values from a time series containing observed monthly data (with some missing values) with estimated values “rounded up” to the nearest whole number via the TRANSFORMOUT option in PROC EXPAND’s CONVERT Statement
- Override of the default interpolation of missing values in an hourly observed series using PROC EXPAND’s TRANSFORMIN option so that the resulting aggregation to daily data contains only the observed data values
- Application of SAS date alignment operators in PROC EXPAND to aggregate observed data to the US federal fiscal year (1 Oct to 30 Sept) and to the State of California fiscal year (1 Jul to 30 June) from a monthly series
- Calculation in PROC EXPAND of “moving time window statistics” such as the moving sum and moving average of a long series and the overlaid display of both the actual and transformed series’ using PROC SGPLOT
- Rapid calculation of lags and other transformations of time series using PROC EXPAND.
- How to impose the mean and other descriptive statistics on time series plots generated by PROC SGPLOT
- Methods to display multiple time series plots on the same display using the Y2 AXIS option in PROC SGPLOT
- Examples of how to use PROC SGSCATTER to create multiple time series plot on one page.

CONCLUSION

Display of time series data is a critical aspect of most analytic projects. PROC EXPAND provides a wide range of powerful, easy to apply, tools to prepare/transform your time series and to then save its work in a SAS data set that can then be displayed using PROC SGPLOT’s extensive data display capabilities. Combining the capabilities of both PROCs in one project makes it easier than ever to “tell the story” in your time series. I hope the examples I present in this paper will provide you with some useful ways to apply these tools and techniques in your own projects.

ACKNOWLEDGEMENTS

The author thanks Anatoli Sznoluch of the US Department of Labor/Employment and Training Administration for providing the weekly unemployment claims time series data and Paul Smilanick of the State of California Department of Social Services for providing the CalWorks time series data used as examples in this paper. The author also thanks the District of Columbia SAS Users Group (DCSUG) for allowing him to present an earlier version of this paper at their December 2011 meeting.

CONTACT

Your comments and questions are both valued and encouraged. Contact the author at:

Andrew H. Karp
Sierra Information Services
19229 Sonoma Highway PMB 264
Sonoma, CA 95476 USA +1 707 996 7380
Email: Andrew@sierrainformation.com
Web: www.SierralInformation.com

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 trademarks of their respective companies.