HIGH-PERFORMANCE FORECASTING USING SAS® GRID MANAGER

Michael Leonard, Cheryl Doninger, and Udo Sglavo, SAS Institute Inc., Cary, NC

ABSTRACT

Many organizations need to forecast large numbers of time series that are organized in a hierarchical fashion. Good forecasting practices recommend that several hierarchies be used and that each hierarchy contain a homogeneous set of time series with similar statistical properties. Modeling and forecasting homogeneous time series hierarchies provide better out-of-sample forecast performance. Because an organization might have many time series hierarchies, it is often desirable to model and forecast these hierarchical time series in parallel for computational efficiency. Also, it is often desirable to aggregate forecasts from several nonhomogeneous time series hierarchies for report generation. This paper demonstrates these techniques for forecasting time series hierarchies in parallel and aggregating the forecasts by using SAS® Forecast Server and SAS® Grid Manager.

INTRODUCTION

Given a collection of time series without structure, the following steps are needed to create several homogenous time series forecasts:

1. Structure the time series to explore their statistical properties.

2. Partition the collection of time series into homogeneous groups of time series with similar statistical properties.

3. Define a hierarchical structure for each partition that provides the best out-of-sample forecast model performance.

4. Model and forecast each partition in a hierarchical fashion.

5. Aggregate the results of each of the hierarchical forecasts.

Since the modeling and forecasting of many hierarchical time series is computationally intensive, executing these tasks in parallel will greatly reduce the overall time needed to produce these forecasts.

SCOPE

This paper focuses on time series that can be formed into a hierarchical structure. The idea of parallelizing time series modeling and forecasting applies to non-hierarchical time series as well.

BACKGROUND

You can find introductory discussions about time series and automatic forecasting in Makridakis, Wheelwright, and Hyndman (1997); Brockwell and Davis (1996); and Chatfield (2000). You can find a more detailed discussion of time series analysis and forecasting in Box, Jenkins, and Reinsel (1994); Hamilton (1994); Fuller (1995); and Harvey (1994). You can find a more detailed discussion of large-scale automatic forecasting in Leonard (2002) and a more detailed discussion of large-scale automatic forecasting systems with input and calendar events in Leonard (2004). A more detailed discussion of hierarchical reconciliation can be found in Trovero, Joshi, and Leonard (2007). A more detailed discussion of SAS Grid computing can be found in Doninger and Wong (2006).

TIME SERIES INDEXING

The following describes the indexing related to forecasting for a two-level hierarchy.

Series Index

Let N represent the *number of series* recorded in the time series data set and let $i = 1,…, N$ represent the *series index.*

Time Index

Let T represent the length of the series and let $t = 1, …, T$ represent the *time index*. The time index is an ordered set of contiguous integers representing time periods associated with equally spaced time intervals.

In some cases, the beginning and/or ending time index coincide; sometimes they do not. Let $t \in \left\{ t_i^b, \left(t_i^b + 1 \right), ..., \left(t_i^e - 1 \right), t_i^e \right\}$ represent the time index where t_i^b and t_i^e represent the beginning and ending time index for the ith series, respectively.

Time Series

Let $y_{i,t}$ represent the *time series values* where $t \in \left\{ t_i^b, ..., t_i^e \right\}$ is the time index for the ith dependent series and where $i = 1,..., N$. Let $Y_t = \sum_{i=1}^{N} y_{i,t}$ represent the aggregate of the individual time series. In the aggregation, missing values are ignored.

Hierarchical Time Series

In the previous discussion, the time series are described to have individual (disaggregate) and aggregate forms. This situation describes a two-level hierarchy. By continuing to aggregate the aggregates, multilevel hierarchies can be formed. This paper discusses only a two-level hierarchy, but the ideas presented also apply to multilevel hierarchies.

Model Forecasts

The time series may be modeled and forecast independently or jointly using vector time series forecasting techniques. These models may be uniquely specified or generated using automatic forecasting techniques. See Leonard 2002 and Leonard 2004 for more details related to automatic forecasting. For the remaining discussion, all that is assumed is that the series have been forecast using either a statistical model or human judgment.

Let $\hat{y}_{i,t}$ represent the model forecasts for the individual time series. Let $\hat{Y}_t$ represent the model forecasts for the aggregate time series. Note that except for unusual cases: $\hat{Y}_t \neq \sum_{i=1}^{N} \hat{y}_{i,t}$.

Reconciled Forecasts

Hierarchical time series forecasting requires some form of reconciliation of the forecasts. Typical techniques are top-down, middle-out, bottom-up, and others.

Let $\hat{y}_{i,t}^R$ represent the reconciled forecasts for the individual time series. Let $\hat{Y}_t^R$ represent the reconciled forecasts for the aggregate time series. Let $\omega_{i,t}$ represent the reconciliation weights. Reconciliation is achieved with the following equations:

$$\hat{Y}_t^R = \sum_{i=1}^{N} \omega_{i,t} \hat{y}_{i,t} \text{ (bottom-up) or } \hat{Y}_t = \sum_{i=1}^{N} \omega_{i,t} \hat{y}_{i,t}^R \text{ (top-down).}$$

Multilevel reconciliation can be achieved repeating this two-level reconciliation iteratively. See Trovero, Joshi, and Leonard 2007 for more details related to reconciling forecasts.

MULTIPLE HIERARCHIES

To achieve better out-of-sample *forecast performance*, homogeneous groups of time series (those that share similar statistical properties) should be modeled and forecast in the same hierarchy. Often, several hierarchies are needed for large amounts of time series data that do not share similar statistical properties. To achieve better computational performance, each hierarchy should be processed on separate processors in parallel. For reporting purposes, it is often desirable to aggregate these hierarchies to summarize the final results.

Figure 1 illustrates three hierarchies being aggregating to one single forecast. Three time series hierarchies are formed from homogeneous groups of time series. Each hierarchy is formed (BY variable ordering) individually to achieve the best out-of-sample forecast performance. Each hierarchy is modeled and forecast individually to achieve the best out-of-sample forecast performance. Each hierarchy is reconciled (top-down, middle-out, or bottom-up) individually to achieve the best out-of-sample forecast performance. Each hierarchy is processed on separate

processors in parallel. Afterwards, the top-level reconciled forecasts can be aggregated upward for reporting purposes (using bottom-up reconciliation).
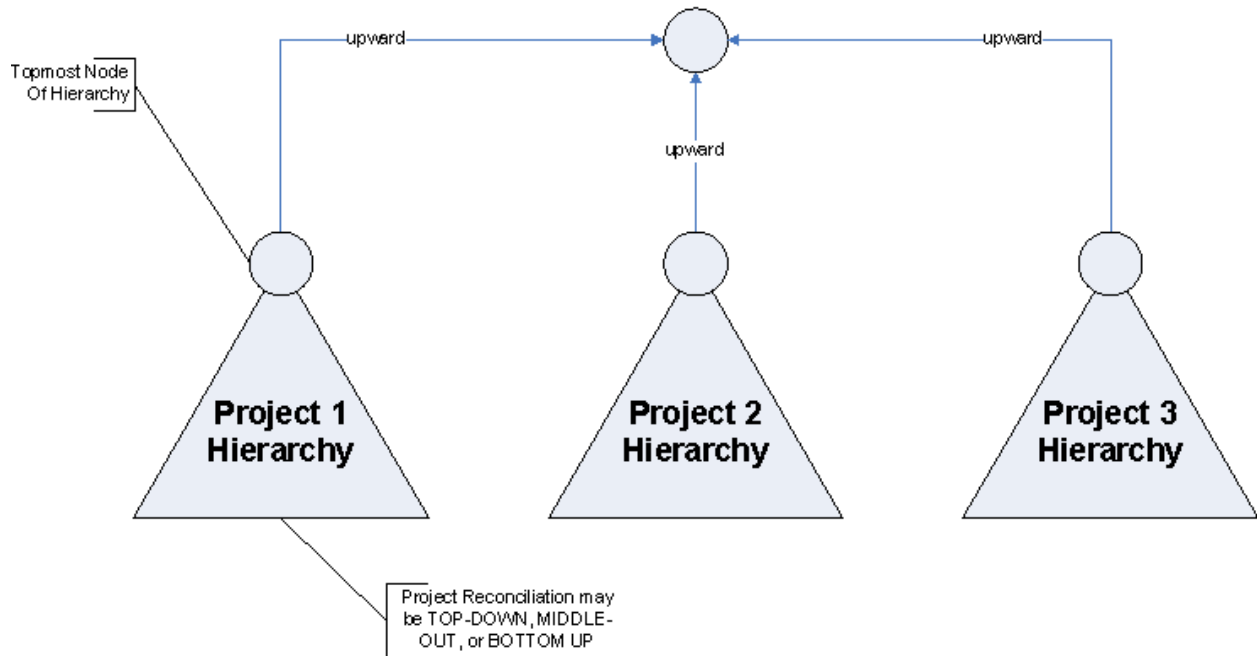
Reconciling Hierarchies

Figure 1: Reconciling Multiple Hierarchies

SAS IMPLEMENTATION

The following section describes how to achieve the results in the previous section using SAS® Forecast Server and SAS® Grid Manager. For this example, the data set Sashelp.Pricedata contains sales data organized by three BY variables (REGION, LINE, and PRODUCT) and time ID variable (DATE).

NOTE: The code colored BLUE is required to submit jobs to SAS® Grid Manager

Partition the input time series data set into homogenous groups.

The %HPFPART macros contain various techniques for partitioning time series data sets. The %HPF_PART_SPLIT macro partitions the data sets based on the variable values associated with a split variable contained in the input data set. See the *SAS® Forecast Server Administrator's Guide* for a description of the %HPF_PART_SPLIT macro.

```
%HPFPART();

%HPF_PART_SPLIT( dataset=sashelp.pricedata,  splitvar=region, validvalues=NO,
partset=parts, view=yes, libref=testlib);
```

The DATASET= macro parameter specifies the input data set to be partitioned. The SPLITVAR= macro parameter specifies a variable in the input data set whose values will be used to partition the data set. The VALIDVALUES= macro parameter specifies that the SPLITVAR= variable contains valid SAS names that can be used to name the partitions. The PARTSET= macro parameter specifies the output data set name that describes the partition. The VIEW= macro parameter specifies whether DATA VIEWs or DATA SETs are created. The LIBREF= macro parameter specifies the SAS library reference to store the partition.

Request three SAS Grid sessions

In this example, there are three projects to be run (proj1, proj2, and proj3).

```
options metaserver='metadata-server-address';

options metaport=metadata-server-port;

options metauser=username;

options metapass="password";

%let rc = %sysfunc(grdsvc_enable(_all_, server=SASApp);

signon proj1;

signon proj2;

signon proj3;
```

Create a separate SAS Forecast Studio Project for each partition of the input time series data set.

The %FSCREATE macro creates a SAS Forecast Studio project. Each project can be modeled and forecast independently and reconciled in different ways. Therefore, each project can be submitted to SAS Grid Manager, which will determine the best node in the grid for running each project. See the *SAS® Forecast Server Administrator's Guide* for a description of the %FSCREATE macro.

The first project uses a hierarchy organized by LINE, PRODUCT.

```
rsubmit proj1 wait=no cmacvar=done;

%fslogin(desktop=0, user=sasdemo, password=Password1, sasenvironment=default);

%fscreate( projectname=region1, data=testlib.region1, by=line product, id=date,
interval=month, var=sale);

%fslogout();

endrsubmit;
```

The second project uses a hierarchy organized by PRODUCT, LINE.

```
rsubmit proj2 wait=no cmacvar=done;

%fslogin(user=sasdemo, password=Password1, sasenvironment=default, desktop=no);

%fscreate( projectname=region2, data=testlib.region2, by=line product, id=date,
interval=month, var=sale);

%fslogout();

endrsubmit;
```

The third project uses a hierarchy that is reconciled BOTTOM-UP.

```
rsubmit proj3 wait=no cmacvar=done;

%fslogin(user=sasdemo, password=Password1, sasenvironment=default, desktop=no);

%fscreate( projectname=region3, data=testlib.region3, by=line product, id=date,
interval=month, reconciliation=bottomup, var=sale);

%fslogout();

endrsubmit;
```

Each project could have been created with many more differing modeling options associated with the %FSCREATE macro. Each project was created using the same Forecasting Environment (ENVIRONMENT=DEFAULT). However, each project could have been created in a different Forecasting Environment, which could be assigned to different processors.

Sign off the SAS Grid sessions

Wait for all three project jobs to complete and then sign off.

```
waitfor _all_ proj1 proj2 proj3;
signoff _all_;
```

Aggregate the projects to form a single top-level forecast.

The %FSPRJAGG macro aggregates the top-level forecasts for several SAS Forecast Studio projects. See the *SAS Forecast Server Administrator's Guide* for a description of the %FSPRJAGG macro.

```
%fslogin(user=sasdemo, password=Password1, sasenvironment=default, desktop=no);
%fsprjagg( projects=region1 region2 region3, outfor=forecasts, aggfor=aggregate);
%fslogout();
```

The PROJECTS= macro parameter specifies the list of SAS Forecast Studio projects to aggregate. The OUTFOR= macro parameter specifies the SAS data set name that contains all of the top-level forecasts from each project. The AGGFOR= macro parameter specifies the data set name that will contain the aggregate forecasts of all of the projects.

CONCLUSION

This paper describes how to execute SAS Forecast Studio projects in parallel. Since the modeling and forecasting of many hierarchical time series is computationally intensive, executing these tasks in parallel using SAS Grid Manager will greatly reduce the overall time needed to produce these forecasts.

REFERENCES

Box, G. E. P., G. M.Jenkins, and G. C. Reinsel. 1994. *Time Series Analysis: Forecasting and Control*. Englewood Cliffs, NJ: Prentice Hall, Inc.

Brockwell, P. J., and R. A. Davis. 1996. *Introduction to Time Series and Forecasting*. New York: Springer-Verlag.

Chatfield, C. 2000. *Time Series Models*. Boca Raton, FL: Chapman & Hall/CRC.

Doninger, C. and Wong, A. 2006. "SAS® Goes Grid – Managing the Workload across Your Enterprise." *Proceedings of the Thirty-First Annual SAS Users Group International Conference*. Cary, NC. SAS Institute Inc. Available at http://www2.sas.com/proceedings/sugi31/211-31.pdf.

Fuller, W. A. 1995. *Introduction to Statistical Time Series*. New York: John Wiley & Sons, Inc.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Harvey, A. C. 1993. *Time Series Models*. Cambridge, MA: MIT Press.

Leonard, M. J. 2002. "Large-Scale Automatic Forecasting: Millions of Forecasts*."* International Symposium of Forecasting. Dublin.

Leonard, M. J. 2004. "Large-Scale Automatic Forecasting with Calendar Events and Inputs.*"* International Symposium of Forecasting. Sydney.

Makridakis, S. G., S. C. Wheelwright, and R. J. Hyndman. 1997. *Forecasting: Methods and Applications*. New York: Wiley.

Trovero, M. A., M. Joshi, and M. J. Leonard. 2007. "Efficient Reconciliation of a Hierarchy of Forecasts in Presence of Constraints." International Symposium of Forecasting, Santander.

FOR MORE INFORMATION

SAS Institute Inc. Introduction to Grid Computing. Available http://support.sas.com/rnd/scalability/grid

SAS Institute Inc. Sample Code for Load Balancing SAS Jobs from Multiple Users. Syntax for Grid Enablement. Available support.sas.com/rnd/scalability/grid/gridfunc.html.

SAS Institute Inc. SAS® Grid Manager information. Available www.sas.com/grid.

SAS Institute Inc. Scalability and Performance Community. Available http://support.sas.com/rnd/scalability/index.html

SAS Institute Inc. Syntax for SAS/CONNECT Grid Functions. Syntax for Grid Enablement. Available support.sas.com/rnd/scalability/grid/gridfunc.html

Tran, A., and R. Williams, 2002. "Implementing Site Policies for SAS Scheduling with Platform JobScheduler." Available support.sas.com/documentation/whitepaper/technical/JobScheduler.pdf.

CONTACT INFORMATION

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

Michael Leonard
100 SAS Campus Drive
Cary, NC 27513
SAS Institute Inc.
Michael.Leonard@sas.com
http://www.sas.com