SAS® Workshop: Forecasting

Course Notes
Table of Contents

Chapter 1  Forecasting ........................................................................................................ 1-1

  1.1  Introduction ................................................................................................................ 1-3

  1.2  Generating Forecasts Automatically ................................................................. 1-8
      Demonstration: Generating Forecasts ................................................................. 1-1

  1.3  Refining Selected Forecast Models ................................................................. 1-27
      Demonstration: Refining Forecast Models ....................................................... 1-28
      Demonstration: Extending the Forecasts ......................................................... 1-47

To learn more…

For information about other courses in the curriculum, contact the SAS Education Division at 1-800-333-7660, or send e-mail to training@sas.com. You can also find this information on the web at http://support.sas.com/training/ as well as in the Training Course Catalog.

For a list of other SAS books that relate to the topics covered in this course notes, USA customers can contact the SAS Publishing Department at 1-800-727-3228 or send e-mail to sasbook@sas.com. Customers outside the USA, please contact your local SAS office.

Also, see the SAS Bookstore on the web at http://support.sas.com/publishing/ for a complete list of books and a convenient order form.
Chapter 1  Forecasting

1.1  Introduction..................................................................................................................... 1-3

1.2  Generating Forecasts Automatically ............................................................................ 1-8
    Demonstration: Generating Forecasts ................................................................................ 1-10

1.3  Refining Selected Forecast Models ............................................................................ 1-27
    Demonstration: Refining Forecast Models.......................................................................... 1-28
    Demonstration: Extending the Forecasts ........................................................................... 1-47
1.1 Introduction

Introduction to the Large-Scale Forecasting Problem and to SAS Forecast Server

- Define the large-scale forecasting problem.
- Discuss central ideas and key components of the large-scale forecasting process.

Motivation: The Large-Scale Forecasting Problem

Forecasting a Single Time Series

A skilled analyst can forecast a single time series by

- applying good judgment based on knowledge and experience
- using various proven time-series analysis techniques
- utilizing good software based on sound statistical theory.
Large-Scale Forecasting

- Modern businesses require efficient, reliable forecasts for many series. These forecasts usually need to be updated on a regular basis.
- There are not sufficient resources to apply the single series forecasting approach to all series that need to be forecast.
- The series might be hierarchically arranged and require reconciliation of forecasts at different levels.

Large-Scale Forecasting Scenario

- 80% can be forecast automatically.
- 10% requires extra effort.
- 10% cannot be forecast accurately.

Time Series Data

Large-Scale Forecasting Requirements

- Automation
- Speed
- Power
1.1 Introduction

SAS High-Performance Forecasting Software

- **Automation**
  - Best model from customizable list
  - Automatic Box-Jenkins model specification
  - Automatic diagnosis of model features such as trend and seasonality
  - Rich set of model selection criteria
  - Ability to use hold-out data for honest assessment

- **Speed**
  - Fast algorithms
  - Efficient implementation

SAS High-Performance Forecasting Software

- **Power**
  - Aggregates time-stamped data
  - Supports a rich set of models:
    - Exponential smoothing models (ESM)
    - Intermittent demand models (IDM)
    - Box-Jenkins autoregressive integrated moving average (ARIMA) models
    - Unobserved components models (UCM)
    - Variants of univariate models that support causal inputs and calendar effects
  - Produces forecasts
  - Evaluates forecasts
  - Aggregates/disaggregates forecasts

The Large Scale Forecasting Process: Essential Ideas

- Data Hierarchies
- Automatic Forecasting
- Forecast Accuracy
- Hierarchical Reconciliation
Data Hierarchies

SKU or item-level sales can be grouped based on geographic regions or distribution levels.

Data Hierarchies: Aggregation

Data in the middle and upper levels of the hierarchy is constructed from data in the base level of the hierarchy. Above, group-level data is created by adding together department-level series. The top level (for example, total sales) is created by adding up all base-level series.

Data Hierarchies ⇝ Automatic Forecasting

- Every series in the hierarchy is forecast separately using a statistical model.
- In SAS Forecast Studio, statistical models reside in a model selection list. A model selection list usually contains several statistical models.
Automatic Forecasting: One Series at a Time

- Each series is forecast separately.
- All models from the model selection list are fit to each individual series.

Question:
How does the software know which fitted model to use when generating forecasts for a series?

Answer:
It uses the model that best fits the series.

Automatic Forecasting: Choosing the Best Model

- MAPE, or mean absolute percent error, is the default fit statistic in SAS Forecast Studio. It measures the average percentage difference between the actual and predicted (forecast) values over the range of the data.
- RMSE, or root mean square error, is another popular fit statistic. It can be thought of as a measure of the average difference between the actual values and the forecast values over the range of the data.

Because MAPE and RMSE measure the difference between actual and predicted values (model error), smaller values are better.

For any series, the best model is the one with the lowest MAPE (or RMSE).

The automatic forecasting process described selects the best model for each series in the hierarchy. The selected models are then used to generate forecasts for each series in the hierarchy.
1.2 Generating Forecasts Automatically

Automatic Forecasting ⇒ Reconciliation

- Because the automatic forecasting process generally uses different models to forecast the individual series up and down the hierarchy, the statistical forecasts usually will not add up or reconcile.
- *Reconciliation* is the process of making the statistical forecasts add up for each time interval in the data.

Forecasting Reconciliation Approaches

- **Top-down**
- **Middle-out**
- **Bottom-up**

Large-Scale Forecasting Scenario

- 80% can be forecast automatically.
- 10% requires extra effort.
- 10% cannot be forecast accurately.

Time Series Data
1.2 Generating Forecasts Automatically

The Data

- The data is weekly case sales of wine from a wine distribution company from January 17, 2004, to May 26, 2007.
- The data hierarchy has three levels: Type (base level), Region, and Total Sales.
- There are four aggregate wine types: tblred (Table Red), tblwt (Table White), Value, and Vintage (limited production).
- There are four Regions: Reg1 – Reg4.

The Business Problem: Maximize Profit

- It is assumed that profit is maximized when the following conditions are met: wine sales are maximized and inventory costs are minimized.
  - Wine sales are maximized when there are no lost sales due to wine being out of stock.
  - Inventory costs are minimized when inventories are kept as small as possible while still satisfying demand.
- Accurate forecasts of wine demand over wine types and distribution regions are essential components in a profit-oriented business strategy.

2. Select Log On.

   Your user name, password, and server name is already provided. You only need to log on.
3. Select the **New...** button to create a new project.

4. In step 1, provide a name and a description for your project. Select **Next >**.
Naming the project something meaningful and providing a reasonably detailed description is considered a best practice.

5. In step 2, expand the FCTOUR library and select the FCDEMO data set.

![Screenshot of SAS software interface](image)

6. Select Next > to move to step 3, BY variable specification.
7. Select **Region** and then **type**. Use the arrow as indicated below. **Bottom Up** reconciliation is used to reconcile the generated statistical forecasts.
The Hierarchy Preview window confirms the appropriate data hierarchy structure.

Region and type are BY variables in the FCDEMO data set. BY variables subset the data set by wine type and region. These variables define the data hierarchy.

8. Select Next > to move to step 4. The Time ID variable is assigned in step 4.

The detected time interval is a one-week interval that begins on Saturday.
9. Select **Next >** to move to step 5. Step 5 assigns dependent and independent variables for the project.

**Sales** is the dependent, or forecast, variable for the project.
**BasePrice** is an independent variable. Assigned independent variables are used as candidate variables in the automatic forecasting process.

10. Select **Next >** to move to step 6. Step 6 lists the data preparation options for the project.

![Data preparation options](image)

Use the default values for these options.

11. Select **Next >** to move to step 7.

Step 7 informs the user that SAS Forecast Studio has enough information to run the project under default settings for model generation, model selection, and forecast generation.

12. Select **Create and import events** to begin the process of bringing these variables into the project.
It is hypothesized that recurring calendar events impact wine demand over some regions and wine types. Event variables that characterize these effects were created and stored in a SAS data set.

New event variables can be created in the Event Repository.

13. To access existing event variables, select **Add events from a data set**. The SAS data set that contains the event variables is in the FCTOUR library.
14. Double-click the **FCTOUR** library to access the **FC_EVENTS** data set.

15. Select **OK** to populate the project’s Event Repository with the event variables contained in the **FC EVENTS** data set.

Notice that the Usage in System-Generated Models status is **Do not use**.

16. The idea is to have the event variables, shown above, considered as candidates in the model-building process. To accomplish this, change their usage status to **Try to use**.
17. Select **Close** to return to step 7.

There is a hypothesized effect on wine demand associated with individual weeks containing Mother’s Day, January 1, and Valentine’s Day for some wine types and regions. This effect is indicated by the Pulse event variable type. The `xmas_4wkbf` event variable is a Temporary Change type. This event variable captures variation in wine demand that starts four weeks before Christmas week. The effect then decays and goes to zero in the week that contains December 25.

It is also hypothesized that there are irregular event effects in wine demand data. **Outlier detection** can be helpful in identifying time intervals in which this type of variation occurs and for building ad hoc event variables to model it.

18. In the automatic forecasting process, outlier detection is useful for finding unknown events or outliers. Outliers usually have non-regular or non-recurring effects. To activate outlier detection, select **Change other forecasting settings**.
19. Select Diagnostics. Select the check box next to **Detect outliers** to activate outlier detection.
You will briefly look at other subsets of forecasting options before you proceed with the automatic forecasting process.

Selecting **Model Generation** shows the default model generation options. A system-generated ARIMAX and ESM model will be built for each series in the data by default.

Selecting **Model Selection** shows that MAPE is the default model selection criterion for the project.
20. Select **OK** to return to step 7.

21. Select **Next >** to move to step 8.
The default project action is to run the project and display the results in SAS Forecast Studio.

22. Select \textbf{Finish}.

The Forecast Summary window shows the distribution of fit statistics, generated model types, and families, up and down the data hierarchy.
In the Forecasting view, details about the total sales forecast are shown in the plot and table.

Notice that the aggregated series at the top level of the hierarchy is given the project name. This series represents total wine sales in each time interval across all regions and wine types.

The data in the plot and table corresponds to the highlighted or active series in the view table to the left.

The forecast plot for total sales indicates overall trends in the data.
In the Forecasting view – regional sales, forecast results for region 2 sales are shown.

Sales in regions 1 through 4 differ in scale but have similar characteristics.

In the Forecasting view – base-level sales, forecast results for region 1, table red wine are shown.
Sales of table red and table white wines are trending up across regions.

In the Forecasting view – base-level sales, forecast results for region 1, value wine are shown.

Sales of value wine are trending down across regions.

In the Forecasting view – base-level sales, forecast results for region 1, vintage wine are shown.
Sales of vintage type wines are trending up across regions. Sales for vintage wines have the smallest scale and highest volatility among wine types.

1.3 Refining Selected Forecast Models

Large-Scale Forecasting Scenario

- 80% can be forecast automatically.
- 10% requires extra effort.
- 10% cannot be forecast accurately.
Refining Selected Forecast Models

- Vintage-type wine forecasts have been identified as “high-value” because of a high associated profit margin but “problematic” because of high volatility.
- Management wants good baseline forecasts and also to quantify the effects of own price variation and promotional events on the demand for vintage-type wines.

Strategy for Refining Vintage Forecast Models

- Find vintage forecast models that have outlier variables. Use these as a basis for building custom event variables.
- Improve baseline forecasts by improving the fit of the generated models, if possible. Pay particular attention to large or systematic residuals near the end of the series.
- Quantify the effects of own price variation and promotional events on sales of vintage wines.

Refining Forecast Models

1. Switch to the Modeling View to view details about the models generating the forecasts. The Modeling View tab is shown in the circle below.
2. Make sure that the region 1 vintage series is selected in the Hierarchy view in the left panel, as shown above by the arrow.

3. Select Tables ⇒ Parameter Estimates to view the fitted model.
• Apparently, region 1 vintage wine sales are impacted by a New Year's Day effect. The effects of other recurring events, flagged by other candidate event variables, seem negligible for this series.

• The effect of base price is significant and negative as expected.

• Two outlier variables are included:
  • A level shift that begins on May 5, 2007
  • A pulse that occurs on January 6, 2007
Refining Selected Forecast Models

The marketing director for region 1 furnishes the following details:

- The shift outlier identified for May 5, 2007, actually represents three separate events:
  - a national news magazine story that ran on May 1 about the health benefits of vintage wine
  - a buy-one-get-one-free (BOGOF) promotion on May 5
  - a coupon insert that ran on May 19
- The company offered a selection of green vintage wines in honor of Saint Patrick’s Day. The campaign started on March 17, 2007. The campaign is scheduled to run again next Saint Patrick’s Day.

The arrow indicates the observation that is associated with the week that contains March 17, 2007.

A temporary change event variable seems to be a reasonable way to qualify the variation associated with this event.

All other residuals seem reasonably random.
4. From the Modeling view, select the **Copy** button to begin augmenting the forecast model for region 1 vintage wine.

The dialog box containing the copy of the forecast model specification is shown.

5. Select **Events** in the panel on the left.
6. Select **New…** to begin creating the indicated event variables.
7. Name the event variable something descriptive such as **bogof_may5_2007**.
8. Leave the event variable type as **Pulse**.
9. Select **Add…** to add the date of occurrence.
10. Set the year to **2007**.

11. Set the month to **May**.

12. Set the week to the week that begins on Saturday the 5th.
13. Select **OK**. The defined event variable to flag the May 5, 2007, BOGOF event is shown.

14. Select **OK**.
The created event populates the project Event Repository.

15. Select **New...** to begin the creation of the coupon event variable.
16. Follow similar steps to create an event variable that flags the variation associated with the coupon promotion that ran on May 19, 2007.
### Select Occurrences

#### Months
- May
- Year: 2007

#### Week:

<table>
<thead>
<tr>
<th>Sun</th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
<th>Fri</th>
<th>Sat</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>30</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

#### Holidays:
- Abraham Lincoln’s and George Washington’s birthdays observed
- Boxing Day
- Canada Day
- Canada Observed Day
- Christmas Day
- Columbus Day
- Easter Day
- Father’s Day

[Image of a software interface showing a calendar with selected dates and holidays listed]
17. Select **New...** to begin the creation of the Saint Patrick’s Day event variable.
Recall that variation in the data associated with the Saint Patrick’s Day event looks like an abrupt, temporary change that begins on March 17, 2007.
18. Name the event variable something descriptive such as StPats_2007.
19. Select the Change… button next to the event variable Type heading.
20. Set the event variable type to Temporary Change.
21. Select **OK**.
22. Select the **Add…** button next to Occurrences to set the occurrence date.
23. Select the date **March 17, 2007**.
24. Select **OK**.
Recall that the Saint Patrick’s Day promotion is set to occur again next year.

25. Select the **Edit…** button next to Recurrence.
Because Saint Patrick’s Day is an annual occurrence, the correct frequency is **Year**.

Because the first instance of this promotion occurred last March, and one more is planned, the appropriate number of occurrences is **2**.
26. Select **OK**.

The temporary change, Saint Patrick’s Day 2007 event variable, is added to the Event Repository for the project.

The three new events are automatically set to be included in the model.
27. Select **OK** to run the model.

The fit of the refined forecast model is better, \( \text{MAPE} = 7.73 \), and it is selected as the forecast model for this series.

28. Select **Tables** ⇒ **Parameter Estimates** to view the fitted model.

| Component      | Parameter  | Estimate | Standard Error | t Value | Approx Fr > |t| |
|----------------|------------|----------|----------------|---------|-------------|---|
| sales          | CONSTANT   | 4041.0   | 67.18164       | 60.17   | < 0.0001    |
| sales          | AR1_1      | 0.20628  | 0.07887        | 2.58    | 0.0107      |
| BasePrice      | SCALE      | -25.03514| 0.04346        | -29.64  | < 0.0001    |
| NEW_YEAR_WEEK  | SCALE      | 748.19578| 124.15694      | 6.03    | < 0.0001    |
| BOGOF_MAR2007  | SCALE      | 525.67342| 240.91773      | 2.18    | 0.0304      |
| COUPON_MAY2007 | SCALE      | 503.43121| 229.92290      | 2.22    | 0.0274      |
| STPTS_2007     | SCALE      | 512.56771| 202.57686      | 2.53    | 0.0123      |
| L505MAY2007    | SCALE      | 357.36381| 159.18337      | 2.23    | 0.0344      |
| A026JAN2007    | SCALE      | 1169.8   | 217.53366      | 5.38    | < 0.0001    |

The effects of the variables characterizing the news story effect, the BOGOF event, and the coupon promotion are positive and significant.
Interpreting the Refined Model

- The BOGOF promotion generated a temporary increase of 526 cases of vintage wine sales in the week beginning May 5, 2007.
- The coupon promotion generated a temporary increase of 503 cases of vintage wine sales in the week beginning May 19, 2007.
- The media event generated a permanent increase of 337 cases of vintage wine sales beginning in the first week of May 2007.

- The estimated parameter on the base price indicates that a $1.00 increase in the vintage case price leads to a decrease of about 25 cases/week in vintage wine sales over the range of the data.
- The Saint Patrick’s Day promotion generated an abrupt, temporary, and positive effect on vintage case sales. The effect in the week ending March 17, 2007, is 513 cases. This effect persisted with decay for approximately five weeks.

Extending the Forecasts

- The final step in this demonstration extends lead forecast values over a one-year forecast horizon.
The option to set the lead forecast horizon is in the project’s forecasting settings.

1. Select **Project ⇢ Forecasting Settings…**

2. Select the **Forecast** options.

3. Change the horizon to **52** periods.
4. Select **OK**.

The 52-week lead forecast shows lifts associated with New Year’s Day and the Saint Patrick’s Day promotion.

The lead forecast for sales also reflects a trend forecast for the **BasePrice** input variable.