

## Forecasting Short Time Series Using SAS® Forecast Server

Cobey Abramowski, Chris Houck, Christian Haxholdt, Benjamin Perryman, SAS Institute Inc.,  
Cary, NC

### ABSTRACT

SAS® Forecast Server provides an excellent tool towards forecasting of time series across a variety of scenarios and industries. Given sufficient data, SAS Forecast Server can provide insights into seasonality, trend, and effects of input variables and events. In some business cases, there might not be sufficient data within individual series to produce these complex models. This paper presents an approach to forecasting these short time series. Using sufficiently long time series as a training data set, forecasted shape and volumes are created through clustering and attribute-based regressions, allowing for new series to have forecasts based on their characteristics. These forecasts are passed into SAS Forecast Server through preprocessing. Production of artificial history and input variables, matched with a custom model addition to the standard model repository, allows for future values of the artificial variable to be realized as the forecast. This process not only relieves the need for manual entry of overrides, but also allows SAS Forecast Server to choose alternative models as additional observations are provided through incremental loads.

### INTRODUCTION

As companies process more and more data, the ability to manually touch the data becomes more and more difficult. Forecasting engines (like SAS Forecast Server) are quite good at forecasting a large majority of data. However, new product introductions are a segment of data that is not easy to automate. This paper describes a process to automatically forecast new product data based on similar products.

### PROBLEM

New product forecasting is the process of predicting demand associated with a time series that has little to no previous data available for use of standard time series forecasting techniques. Currently, a number of techniques, such as chaining items together (one product is discontinued and a similar item is being introduced) or copying the forecast of a like item, can be used to forecast new items. These techniques require manual choice by clients and are often too costly and too subjective to implement on a large scale.

Another common methodology used is the creation of forecasts through external methods, such as using these forecasts as overrides in SAS® Forecast Studio. This method can be automated, but using forecasting overrides is static, with incremental loads not affecting a forecast.

Clearly, these approaches demonstrate two key difficulties that need to be addressed for a comprehensive new product forecasting implementation: automation and dynamic forecasts (with respect to incremental updates).

The approach in this paper addresses both of these difficulties. Using the similarity of attributes of a new series to that of established series, an initial time zero forecast is created in an automated fashion. In this paper, this forecast is referred to as the Initial New Product Forecast, INPF. Using the INPF to fabricate history and to construct artificial, but logically supporting, input variables spliced into the active forecasting data set, SAS Forecast Studio can progressively predict the INPF beyond time zero as new actual records arrive. As these actual records are received for a new series during incremental loads, this approach allows Forecast Studio to use these values to update the models, using an ARIMAX modeling framework. This creates a dynamic new product forecasting approach which is fully automated, removing the need for manual chaining or overrides.

## SOLUTION

The approach used can be broken into 2 stages:

1. Initial (or time zero) new product forecasting.
2. Progressive forecasting, updating the INPF when new observations become available.

### INITIAL NEW PRODUCT FORECASTING (INPF)

When there is no available history and no logical chaining can be performed on the variable of interest, prediction of future behavior is based on a family of regression type models. The approach combines traditional predictive modeling (in other words, data mining techniques that uses time dimensions) and predictive analytics to generate a forecast for new products.

This process assumes there is some history that is similar to the new products (for example, history of the same segment). This history is used to create a training data set. From the training data set, shape patterns and volume possibilities are determined using clustering analysis. A shape pattern is simply the demand for a given product over time (Figure 1. Demand across Time).

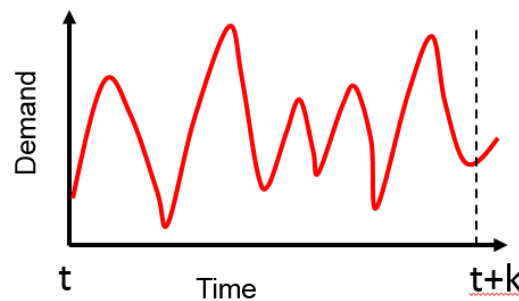


Figure 1. Demand across Time

Volume and shape classifications are made according to attributes to create models that can then be used to score new products. That is, the models are trained using historical demand data, then new products with no demand history are scored.

To create a training data set, the first step is to load a time stamped data set, for example, transaction history data, and convert this data into fixed-interval time series. In other words, define a time interval that is suitable for analysis (for example, hours, days, weeks) and accumulate the data for the chosen interval. A cycle index is then added to each time series; cycle 1 is equal to the introduction date of a product. Order (not dates) are important here. The cycle index becomes useful for modeling and comparing time series data with different introduction times. All time series are now aligned according to the cycle index.

The structured automated process for new product forecasting will emulate, instead of directly incorporate human judgment. The process has four main steps:

- I. Cluster time series according to their demand pattern over a predefined time interval.
  - a. The clustering could be on the normalized version of the actual time series. The normalized version could be the distribution function or cumulative, statistical properties, or anything that will make a similarity measure.
  - b. The demand pattern will be a distribution over the predefined time interval. This distribution sums to one.

- II. Predict demand pattern cluster membership based on product attributes over a predefined time interval.
  - a. The membership is estimated using one or more models in combination and then optimized.
  - b. In this approach, we use two related models: decision trees (PROC HPSPPLIT) and random forest (PROC HPFOREST). The final membership is then optimized by entering the two results into a neural network (PROC HPNEURAL).
- III. Predict total demand volume over a predefined time interval.
  - a. The volume is predicted using one or more models in combination and then optimized.
  - b. In this approach, we use three models: random forecast trees (PROC HPFOREST) where we predict volume cluster membership, regression (PROC HPREG), and a neural network (HPNEURAL). The final volume is then optimized by entering the three results into a neural network (PROC HPNEURAL).
- IV. Run a score data set through the process to get a demand pattern and a volume estimate for each individual time series. Final prediction is generated by multiplying the cluster demand pattern with the predicted total volume over the time horizon.

This part of the New Product Forecast is illustrated in Figure 2 below.

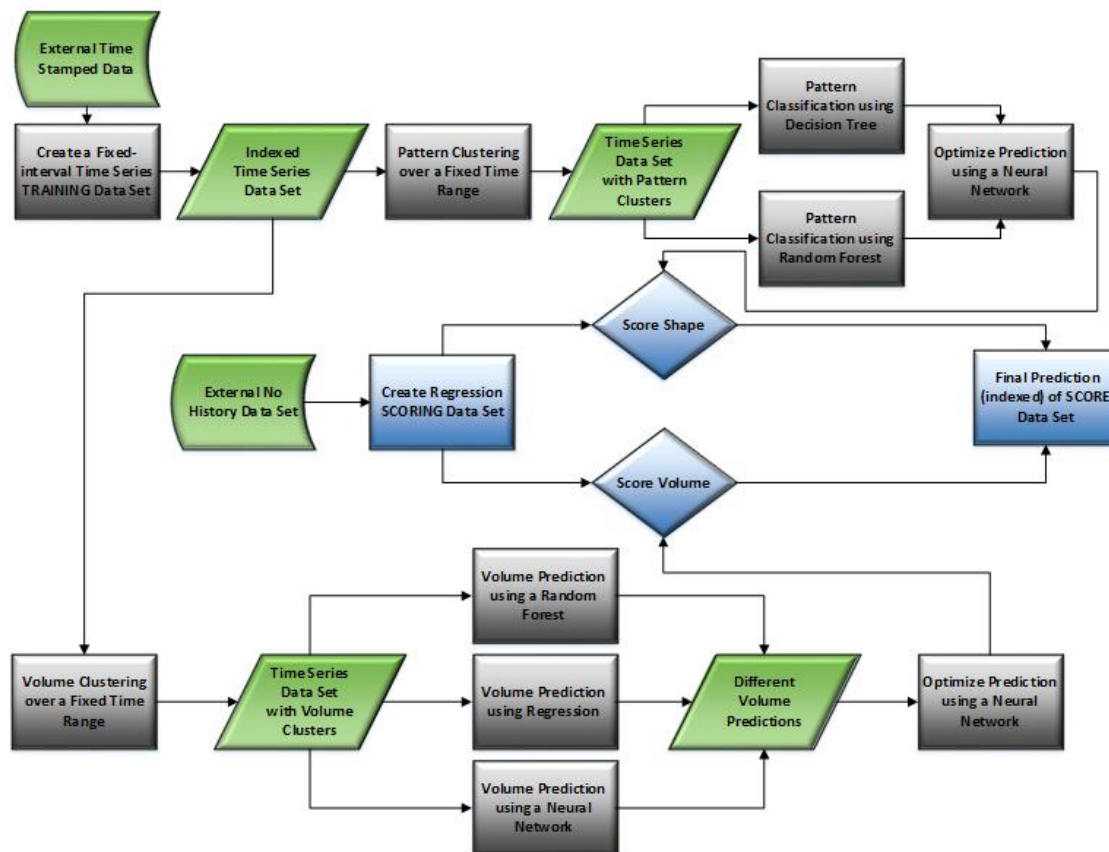


Figure 2. Flow Chart Describing Initial New Process Forecast Creation

## PROGRESSIVE FORECASTING

For each new product, an initial forecast is created (INPF, final prediction of SCORE data set in Figure 2). This forecast is used as the “fake” history for the product. In Table 1, the actual history is set to equal the fake history, and a causal variable called FHIV (fake history indicator variable) is also set to the fake history. FHIV is set equal to the fake history in the future. This is just a repeat of the fake history in the past. Before a product goes live, the forecast will match the FHIV variable since FHIV is a perfect predictor for actuals up to this point: Forecast = FHIV (NPF model).

Time	-5	-4	-3	-2	-1	0	1	2	3	4	5
Score Data Set / INPF						50	100	0	1000	30	50
Actuals	50	100	0	1000	30	.	.	.	.	.	.
Forecast	50	100	0	1000	30	50	100	0	1000	30	50
FHIV	50	100	0	1000	30	50	100	0	1000	30	50

**Table 1. Example of Loaded Data Based on the Score Data Set**

As a product goes live, real data is added incrementally (the missing actuals in Table 1 are populated). SAS Forecast Server is able to reassess the model chosen. When the data begins to differ from the forecast, the NPF model might no longer be the champion model. With each incremental load, the NPF model is compared to other models. The NPF model could immediately or in the future cease to be the best model depending on the accuracy of the new product forecast. Thus, NPF is immediately dynamic.

## CONCLUSION

New product forecasts are created via clustering methodologies. These forecasts are then used as a causal variable to create forecasts for product introduction. Using the new product forecast as a causal variable allows the forecast to change dynamically as the product goes live. Thus, new products are forecast automatically and change dynamically as data is incrementally added.

## ACKNOWLEDGMENTS

We would like to thank Alex Chien, Director of R&D at SAS Institute Inc. for his contributions.

## CONTACT INFORMATION

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

Cobey Abramowski  
100 SAS Campus Drive  
Cary, NC 27513  
SAS Institute Inc.  
[cobey.abramowski@sas.com](mailto:cobey.abramowski@sas.com)  
<http://www.sas.com>

Chris Houck  
100 SAS Campus Drive  
Cary, NC 27513  
SAS Institute Inc.

[chris.houck@sas.com](mailto:chris.houck@sas.com)  
<http://www.sas.com>

Christian Haxholdt  
100 SAS Campus Drive  
Cary, NC 27513  
SAS Institute Inc.  
[christian.haxholdt@sas.com](mailto:christian.haxholdt@sas.com)  
<http://www.sas.com>

Benjamin Perryman  
100 SAS Campus Drive  
Cary, NC 27513  
SAS Institute Inc.  
[ben.perryman@sas.com](mailto:ben.perryman@sas.com)  
<http://www.sas.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.