Large-Scale Automatic Forecasting: Millions of Forecasts

Michael Leonard, SAS Institute Inc.
Cary, NC, USA

Abstract

Web sites and transactional databases collect large amounts of time-stamped data. Businesses often want to make future predictions (forecasts) based on numerous sets of time-stamped data (sets of transactions). There are many time series analysis techniques related to forecasting, and an experienced analyst can effectively use these techniques to analyze, model, and forecast time series data. However, the number of time series to forecast may be enormous or the forecasts may need to be updated frequently, making human interaction impractical. Additionally, these time series analysis techniques require that the data be recorded on fixed time intervals. This paper proposes the following technique for automatically forecasting sets of transactions. For each set of transactions recorded in the database: The time-stamped data are accumulated to form a time series. The time series is diagnosed to choose an appropriate set of candidate forecasting models. Each of the diagnosed candidate models are fitted (trained) to the time series data with the most recent data excluded (holdout sample or test data). Based on a model selection criterion, the best performing candidate model within the holdout sample is selected to forecast the time series. This automatic forecasting technique can efficiently generate millions of forecasts related to time-stamped data. This paper demonstrates this technique using SAS® High-Performance Forecasting Software.

Keywords

Automatic Forecasting, Forecasting, Time Series Analysis, Time Series Databases, Temporal Data Mining, Data Mining

Introduction

Businesses often want to generate a large number of forecasts based on time-stamped data stored in their transactional or time series databases. Web sites, point-of-sale (POS) systems, call centers, and inventory systems are examples of transactional databases. A skilled analyst can forecast a single time series by applying good judgment based on his or her knowledge and experience, by using various time series analysis techniques, and by utilizing good software based on proven statistical theory. Generating large numbers of forecasts and/or frequently generating forecasts requires some degree of automation. Common problems that a business faces are:

- No skilled analyst is available
- Many forecasts must be generated
- Frequent forecast updates are required
- Time-stamped data must be converted to time series data
- The forecasting model is not known for each time series

This paper hopes to help solve these problems by proposing a technique for large-scale automatic forecasting. This paper provides a background on forecasting, describes an automatic forecasting technique, and demonstrates how SAS High-Performance Forecasting Software can be used for automatic forecasting.

Background

This section provides a brief theoretical background on automatic forecasting. It is intended to provide the analyst with motivation, orientation, and references. An introductory discussion of these topics can be found in Makridakis et al. (1997), Brockwell and Davis (1996), and Chatfield (2000). A more detailed discussion of time series analysis and forecasting can be found in Box et al. (1994), Hamilton (1994), Fuller (1995), and Harvey (1994).

Transactional Data

Transactional data are time stamped data collected over time at *no* particular frequency. Some examples of transactional data are:

- Internet data
- Point of Sales (POS) data
- Inventory data
- Call Center data
- Trading Data

Businesses often want to analyze transactional data for trends and seasonal variation. To analyze transactional data for trends and seasonality, statistics must be computed for each time period and season of concern. The frequency and the season may vary with the business problem. For example, various statistics can be computed on each time period and season

- Web visits by hour and by hour of day
- Sales per month and by month of year
- Inventory draws per week and by week of month
- Calls per day and by day of week
- Trades per weekday and by weekday of week

Time Series Data

Time series data are time-stamped data collected over time at a particular frequency. Some examples of time series data are:

- Web visits per hour
- Sales per month
- Inventory draws per week
- Calls per day
- Trades per weekday

As can be seen, the frequency associated with the time series varies with the problem at hand. The frequency or *time interval* may be hourly, daily, weekly, monthly, quarterly, yearly, or many other variants of the basic time intervals. The choice of frequency is an important modeling decision. This decision is especially true for automatic forecasting. For example, if you want to forecast the next four weeks, it is best to use weekly data rather than daily data. The forecast horizon in the former case is 4, in the latter case is 28.

Associated with each time series is a seasonal cycle or *seasonality*. For example, the length of seasonality for a monthly time series is usually assumed to be 12 because there are 12 months in a year. Likewise, the seasonality of a daily time series is usually assumed to be 7. The usual seasonality assumption may not always hold. For example, if a particular business's seasonal cycle is 14 days long, the seasonality is 14, not 7.

Time series that consist of mostly zero values (or a single value) are called interrupted or *intermittent time series*. These time series are mainly constant valued except for relatively few occasions. Intermittent time series must be forecast differently from non-intermittent time series.

Forecasting Models

There are numerous types of forecasting models that a skilled analyst can use. For automatic forecasting of large numbers of time series, only the most robust models should be used. The goal is not to use the very best model for forecasting each time series. The goal is to provide a list of *candidate models* that will forecast the large majority of the time series well. In general, when an analyst has a large number of time series to forecast, the analyst should use automatic forecasting for the low-valued forecasts; the analyst can then spend a larger portion of his/her time dealing with high-valued forecasts or low-valued forecasts that are problematic.

The candidate models that are used in this paper are considered the most robust in the forecasting literature and these models have proven their effectiveness over time. These models consider the *local* level, trend, and seasonal components of the time series. The term local is used to describe the fact that these components evolve with time. For example, the local trend component may not be a straight line but a trend line that changes with time. In each of these models, there is an error or random component that models the uncertainty.

The components associated with these models are not only useful for forecasting but also for describing how the time series evolves over time. The forecasting model decomposes the series into its various components. For example, the local trend component describes the trend (up or down) at each point in time and the final trend component describes the expected future trend. These forecasting models can also indicate departures from previous behavior or can be used to cluster time series.

The parameter estimates (*weights*) describe how fast the component is changing with time. Weights near zero indicate a relative constant component; weights near one indicate a relatively variable component. For example, a seasonal weight near zero represents a stable seasonal component; a seasonal weight near one represents an unstable seasonal component. Weights should be optimized based on the data for best results.

Local Level Models

The local level models are used to forecast time series whose level (or mean) component varies with time. These models predict the local level for future periods.

(Series) = (Local Level) + (Error)

An example of a local level model is *Simple Exponential Smoothing*. This model has one parameter (level weight), which describes how the local level evolves. The forecasts for the future periods are simply the final local level (a constant).

Local Trend Models

The local trend models are used to forecast time series whose level and/or trend components vary with time. These models predict the local level and trend for future periods.

(Series) = (Local Level) + (Local Trend) + (Error)

Examples of local trend models are *Double* (Brown), *Linear* (Holt), and *Damped-Trend Exponential Smoothing*. The double model has one parameter (level/trend weight), the linear model has two parameters (level and trend weights), and the damp-trend model has three parameters (level, trend, and damping weights). The damping weight dampens the trend over time. The forecasts for the future periods are a combination of the final local level and the final local trend.

Local Seasonal Models

The local seasonal models are used to forecast time series whose level and/or seasonal components vary with time. These models predict the local level and season for future periods.

(Series) = (Local Level) + (Local Season) + (Error)

An example of a local seasonal model is *Seasonal Exponential Smoothing*. The seasonal model has two parameters (level and seasonal weights). The forecasts for the future periods are a combination of the final local level and the final local season.

Local Models

The local models are used to forecast time series whose level, trend, and/or seasonal components vary with time. These models predict the local level, trend, and seasonal component for future periods.

(Series) = (Local Level) + (Local Trend) + (Local Season) + (Error)
(Series) = ((Local Level) + (Local Trend)) x (Local Season) + (Error)

An example of a local level model is *Winters Method* (additive or multiplicative). These models have three parameters (level, trend, and seasonal weights). The forecasts for the future periods are a combination of the final local level, the final local trend, and final local season.

Transformed Models

With the exception of the Winters Method Multiplicative model, the above forecasting models are linear, that is, the components must be added together to re-create the series. Since time series are not always linear with respect to these components, transformed versions of the above forecasting models must be considered when using automatic forecasting. Some useful time series transformations are:

- Logarithmic
- Square-Root
- Logistic
- Box-Cox

For example, suppose the underlying process that generated the series has one of the following non-linear forms:

(Series) = Exp ((Local Level) + (Local Trend) + (Error))   exponential growth model
(Series) = (Local Level) x (Local Season) x (Error)   multiplicative error model

Transforming the above series permits the use of a linear forecasting model:

Log(Series) = (Local Level) + (Local Trend) + (Error)   log local trend model
Log(Series) = Log(Local Level) + Log(Local Seasonal) + Log(Error)
log local seasonal model

The above transformations can only be applied to positive-valued time series.

Intermittent Models

Intermittent or interrupted time series models are used to forecast intermittent time series data. Since intermittent series are mostly constant valued (usually zero) except on relatively few occasions, it is often easier to predict when the series departs and how much the series departs from this constant value rather than the next value. An example of an intermittent time series model is *Croston's Method*.

Intermittent models decompose the time series into two parts: the interval series and the size series. The interval series measure the number of time periods between departures. The size series measures the magnitude of the departures. After this decomposition, each part is modeled and forecast independently. The interval forecast predicts when the next departure will occur. The size forecast predicts the magnitude of the next departure. After the interval and size predictions are computed, they are combined (predicted magnitude divided by predicted number of periods for the next departure) to produce a forecast for the average departure from the constant value for the next time period.

Forecasts

Forecasts are time series predictions made for future periods in time. They are random variables and therefore have an associated probability distribution. For example, assuming a normal distribution, the forecasts for the next three months can be viewed as three "bell-curves" that are progressively flatter (or wider). The mean or median of each forecast is called the *prediction*. The variance of each forecast is called the *prediction error variance* and the square root of the variance is called the *prediction standard error*. The variance is computed from the forecast model parameter estimates and the model residual variance.

The forecast for the next future period is called the *one-step ahead forecast*. The forecast for *h* periods in the future is called the *h-step ahead forecast*. The *forecast horizon* or *forecast lead* is the number of periods into the future for which predictions are made (one-step, two-step,..., h-step). The larger the forecast horizon, the larger the prediction error variance at the end of the horizon. For example, forecasting daily data four weeks into the future implies a forecast horizon of 28 whereas forecasting weekly data four weeks into the future implies a forecast horizon of only four. The prediction standard error at the end of the horizon in the former case will be relatively larger than the prediction standard error in the latter case.

The *confidence limits* are based on the prediction standard errors and a chosen confidence limit size. A confidence limit size of 0.05 results in 95% confidence limits. The confidence limits are often computed assuming a normal distribution, but others could be used. As with the prediction standard errors, the width of the confidence limits increases with the forecast horizon. Once again, the forecast horizon of 28 will have wide confidence limits at the end of the horizon.

The *prediction error* is the difference between the predicted value and the actual value when the actual value is known. The prediction errors are used to calculate the statistics of fit that are describe later. For transformed models, it is important to understand the difference between the model errors (or residuals) and the prediction errors. The residuals measure the departure from the model in the transformed metric (Log, Square Root, etc.). The prediction errors measure the departure from the original series. You should not directly compare the model residuals of a transformed model and a non-transformed model when evaluating the model fit. You can compare the prediction errors between any two models because prediction errors are computed on the same metric.

Taken together, the predictions, prediction standard errors, and confidence limits at each period in the forecast horizon are the *forecasts*. Although many people use the word "forecast" to imply only prediction, a forecast is not one number for each future time period.

Using a transformed forecasting model requires the following steps:

- The time series data are transformed.
- The transformed time series data are fit using the forecasting model.
- The forecasts are computed using the parameter estimates and the transformed time series data.
- The forecasts (predictions, prediction standard errors, and confidence limits) are inverse transformed.

The naïve inverse transformation results in *median forecasts*. To obtain *mean forecasts* requires that the prediction and the prediction error variance both are adjusted based on the transformation. Additionally, the model residuals will be different from the prediction errors due to this inverse transformation. If no transformation is used the model residual and the prediction error will be the same, and likewise the mean and median forecast will be the same (assuming a symmetric disturbance distribution).

Statistics of Fit

The *statistics of fit* evaluate how well a forecasting model performs by comparing the actual data to the predictions. For a given forecast model that has been fitted to the time series data, the model should be checked or evaluated to see how well it fits or forecasts the data. Commonly used statistics of fit are Mean

Square Error (MSE), Mean Absolute Percentage Error (MAPE), Akaike Information Criteria (AIC), and many others. The statistics of fit can be computed from the model residuals or the prediction errors.

When the full range of data is used to both fit and evaluate the model, this is referred to as *in-sample* evaluation. When the end of the data (holdout) is excluded for parameter estimation and the holdout sample is used for evaluation, this is referred to as *out-of-sample* or *holdout sample* evaluation. Holdout sample analysis is similar to *training* and *testing* of neural networks. You withhold a portion of the data for training (fit) and test how well it performs (holdout).

When a particular statistic of fit is used for forecast model selection, it is referred to as the *model selection criterion*. For example, if the MAPE (an often recommended choice) is used as a model selection criterion, the forecast model with smallest MAPE in the evaluation region (in-sample or out-of-sample) is chosen as the *best model*.

When using model selection criteria to rank forecasting models, it is important to compare the errors on the same metric, that is, you should not compare transformed model residuals with non-transformed model residuals. You should first inverse transform the forecasts from the transformed model prior to computing the prediction errors and then compute the model selection criterion based on the prediction errors.

Automatic Forecasting Technique

Automatic forecasting is usually defined as forecasting without the aid of an analyst skilled in time series analysis techniques or as forecasting when the number of forecasts is too numerous for an analyst to investigate. Automatic forecasting is usually performed on each time series independently. For each time series and for each candidate model, the parameter estimates (weights) should be optimized for best results. This means that several optimizations may be required for each time series.

Accumulation Step

The *accumulation* of time-stamped data into time series data is based on a particular frequency. For example, time-stamped data can be accumulated to form hourly, daily, weekly, monthly, or yearly time series. Additionally, the method for accumulating the transactions within each time period is based on a particular statistic. For example, the sum, mean, median, minimum, maximum, standard deviation, and other statistics can be used to accumulate the transactions within a particular time period.

There may be no data recorded for certain time periods resulting in missing values in the accumulated time series. These missing values can be represent unknown values and thus left missing or they can represent no activity in which case in which case they should be set to zero or some other appropriate value.

For automatic forecasting, accumulation is the most important decision because the software makes most of the remaining decisions. If weekly forecasts of the average of the transactions are needed, then the accumulation frequency should be weekly and the accumulation statistic should be the average.

Accumulating the transactional data on a relatively small time interval may require long forecast horizon. For example, if the data are accumulated on an hourly basis and if it is desired to forecast one month into the future, the forecast horizon is enormous and the width of the confidence limits will be very wide at the end of the horizon.

Diagnostic Step

The time series diagnostics subset the potential list of candidate models to those that are judged appropriate to a particular time series. Time series that have trends (deterministic or stochastic) should be forecast with models that have a trend component. Time series with seasonal trends (deterministic or stochastic) should be forecast with models that have a seasonal component. Time series that are non-linear should be transformed for use with linear models. Time series that are intermittent should be forecast with intermittent models.

The importance of the diagnostics should not be underestimated. Applying a seasonal model to a non-seasonal time series, particularly one with a short history, can lead to over-parameterization or false seasonality. Applying a linear model to a nonlinear time series can lead to under-estimation of the growth (or decline). Applying a non-intermittent model to an intermittent series will result in predictions biased toward zero.

If it is known, *a priori*, that a time series has a particular characteristic, then the diagnostics should be overridden and the appropriate model should be used. For example, if the time series is known to be seasonal, the diagnostics should be overridden to always choose a seasonal model.

Model Selection Step

After the candidate models have been subsetted by the diagnostics, each model is fit to the data with the holdout sample excluded. After model fitting, the one-step-ahead forecasts are made in the fit region (in-sample) or the multi-step-ahead forecasts are made in the holdout sample region (out-of-sample). The model selection criterion is used to select the best performing model from the appropriate subset of the candidate models. As described above, the model selection criteria are statistics of fit.

If the time series is short in length, holdout sample analysis may not be possible due to a lack of data. In this situation, the full-range of the data should be used for fitting and evaluation. Otherwise, holdout sample analysis is often recommended.

Forecasting Step

Once the best forecasting model is selected from the candidate models, the selected model is fit to the full range of the data. If we excluded the holdout sample in this step, we would be ignoring the most recent and influential observations. Most univariate forecasting models are weighted averages of the past data, with the most recent having the greatest weight. Once the model is selected, excluding the holdout sample can result in poor forecasts. Holdout sample analysis should only be used for forecast model selection, not forecasting.

Forecasts (predictions, prediction standard errors, prediction errors, and confidence limits) are made using the model parameter estimates, the model residual variance, and the full-range of data. If a model transformation was used, the forecasts are inverse transformed on a mean or median basis.

When it comes to decision-making based on the forecasts, the analyst must decide whether to base the decision on the predictions, lower confidence limits, upper confidence limits or the distribution (predictions and prediction standard errors). If there is a greater penalty for over-predicting, the lower confidence limit should be used. If there is a greater penalty for under-predicting, the upper confidence limit should be used. Often for inventory control decisions, the distribution (mean and variance) is important.

Evaluation Step

Once the forecasts are made, the in-sample statistics of fit are computed based on the one-step ahead forecasts and the actual data. These statistics can be used to identify poorly fitting models prior to making business decisions based on these forecasts. If forecasts do not predict the actual data well, they can be flagged to signal the need for more detailed investigation by the analyst.

If holdout sample analysis was used to select the model, it is possible that the selected model performs less than other candidate models because the more recent data may depart from the previous underlying process.

Performance Step

The previous steps are used to forecast the future. This step judges the performance of the forecasts or the *ex-post* forecast evaluation. After forecasting future periods, the actual data becomes available as time passes. For example, suppose that monthly forecasts are computed for next three months into the future.

After three months pass, the actual data are available. The forecasts made three months ago can now be compared to the actual data of the last three months.

The availability of the new data begs the following questions:

- How well are we forecasting?
- Why are we forecasting poorly?
- If we were forecasting well before, what went wrong?

Some useful measures of forecast performance are the statistics of fit described above. When the statistics of fit are used for performance measures, the statistics are computed from the previous predictions and the newly available actual data in the forecast horizon. For example, the MAPE can be computed from the previous predictions and the newly available actual data in the three-month forecast horizon.

Another useful measure of forecast performance is determining whether the newly available data falls within the previous forecasts' confidence limits. For example, performance could be measured by whether or not the newly available actual data falls outside the previous forecasts' confidence limits in the three-month forecast horizon.

If the forecasts were judged to be accurate in the past, a poor performance measure, such as actual data outside the confidence limits, could also be indicative of a change in the underlying process. For example, a change in behavior, an unusual event, or other departure from the past may have occurred since the forecasts were made. This change in forecast performance may indicate fraud, a change in fashion, loss of market share, or other departure from the past; or it simply may be indicative of a poor forecast model choice.

Implementation

The automatic forecasting technique describe above is efficiently implemented in the SAS® High-Performance Forecasting software (HPF). As input, HPF requires a SAS data set or data view containing the time-stamped data. If the time-stamped data is contained in a common commercial database, SAS/ACCESS® software can be used to create a SAS data view that can access the data.

For each set of transactions or time series, HPF accumulates the data to form a time series, diagnoses the time series, selects an appropriate forecasting model using in-sample or holdout sample analysis, forecasts the time series, evaluates the forecast, and/or analyzes the forecast performance. HPF performs these tasks with a single pass through the data set or data view.

For someone familiar with basic SAS programming, HPF is fairly simple to use. The following example illustrates some of the fundamental capabilities of the HPF.

Suppose that a SAS data set or data view named TRANSACTIONS contains the time-stamped data. The data set contains the time-stamp variable called DATE and numerous analysis variables. The data set is sorted by the time-stamp and the categorical variables: COUNTRY, PROVINCE, and CITY. To forecast each of the analysis variables by the categorical variables on a median monthly basis using a holdout sample of 3 and a model selection criterion of MAPE, the following SAS code can be used:

```
proc hpf data=transactions out=forecasts;
     by country province city;
     id date interval=month accumulate=median;
     forecast _all_ / holdout=3 select=mape;
run;
```

The above SAS code creates a data set named FORECASTS that contains the forecasts for each analysis variable by each of the categorical grouping. For each analysis variable and for each categorical grouping, a separate time series is formed and an appropriate forecasting model is chosen to forecast the time series. Once the time series have been analyzed and the forecasts have been stored in a SAS data set, the results can be used for a wide variety of software provided by SAS.

Results

Since each time series is analyzed and forecast separately, the processing time of the implementation described above is the sum of the processing times of each accumulated time series. The processing time for each set of transactions depends upon the number of transactions to be accumulated to form a time series and the length of the accumulated time series. Additionally, the processing time is heavily dependent on the access speeds to disk space.

As of the writing of this paper, one thousand time series can be forecast in less than one minute on commonly sold laptop computers. One thousand per minute roughly equals one million overnight. Listed are some examples of commonly used computing environments and the processing times for transactional series of length 1000 that result in time series of length 100:

CPU	Memory	Operating System	CPU Time per 10000 in minutes
400 MHz Intel Pentium II	128 MB	Windows NT 4.0	11
700 MHz Intel Pentium III	2 GB	Windows NT Terminal Server	3

Note that the CPU processing times listed above are per processor. Since each time series is processed separately, installations with multiple processors and/or distributed processing can take advantage of SAS/CONNECT® software. This software can greatly reduce the real-time processing needed to forecast millions of time series by taking advantage of additional processors.

Conclusion

This paper described a technique for large-scale automatic forecasting. This technique was demonstrated to be both efficient and effective in forecasting millions of time series. Past empirical evidence suggests that the statistical forecasting models used with this technique have been proven to be robust and effective in forecasting a variety of time series encountered by businesses. By using holdout sample analysis with a model selection criterion, an appropriate forecasting model can be chosen without the aid of a skilled time series analyst. This technique will not produce the very best forecast for each time series, but it can produce good forecasts for the large majority. This technique can be used on both time series data and time-stamped transactional data after accumulation.

References

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

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

Chatfield, C. (2001), *Time Series Models*, Boca Raton, FL: Chapman & Hall/CRC.

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. (1994), *Time Series Models*, Cambridge, MA: MIT Press.

Makridakis, S. G., Wheelwright, S. C., and Hyndman, R. J. (1997), *Forecasting: Methods and Applications*, New York: John Wiley & Sons, Inc.