

Count Series Forecasting

Michael Leonard and Bruce Elsheimer, SAS Institute Inc.

ABSTRACT

Many organizations need to forecast large numbers of time series that are discretely valued. These series, called count series, fall approximately between continuously valued time series, for which there are many forecasting techniques (ARIMA, UCM, ESM, and others), and intermittent time series, for which there are few forecasting techniques (Croston's method and others). This paper proposes a technique for large-scale automatic count series forecasting and uses SAS® Forecast Server and SAS/ETS® software to demonstrate this technique.

INTRODUCTION

Most traditional time series analysis techniques assume that the time series values are continuously distributed. For example, autoregressive integrated moving average (ARIMA) models assume that the time series values are generated by continuous white noise passing through various types of filters. When a time series takes on small, discrete values (0, 1, 2, 3, and so on), this assumption of continuity is unrealistic.

By using discrete probability distributions, count series analysis can better predict future values and, most importantly, more realistic confidence intervals. In addition, count series often contain many zero values (a characteristic that is called zero-inflation). Any realistic distribution must account for the "extra" zeros.

Count series analysis includes in the following analyses:

- **Forecasting count series:** From the historical count series values, you can predict its future values and generate confidence intervals that are discretely valued. Count series forecasting is important for inventory replenishment where unit demand is very low. For example, the demand for spare parts in a particular week is small and discretely valued.
- **Monitoring count series:** You can monitor recent values of a count series to detect anomalous values from its historical values. Count series monitoring is important for monitoring processes that generate count data. For example, the number of failures in an industrial process and the many devices in the new world of the Internet of Things generate counts over time (count series data).

The paper combines proven traditional time series analysis techniques with proven discrete probability distribution analysis techniques to propose a novel technique for forecasting and monitoring count series.

MOTIVATION AND SCOPE

Many real-world time series data are not continuously valued. Sometimes these time series values are small and discretely valued. This situation is true for many inventory control problems that are related to "slow moving items." Figures 1 through 4 illustrate the issues of concern.

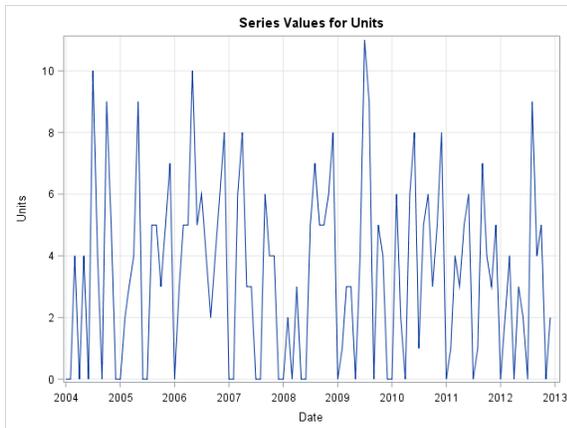


Figure 1. Time Series Plot

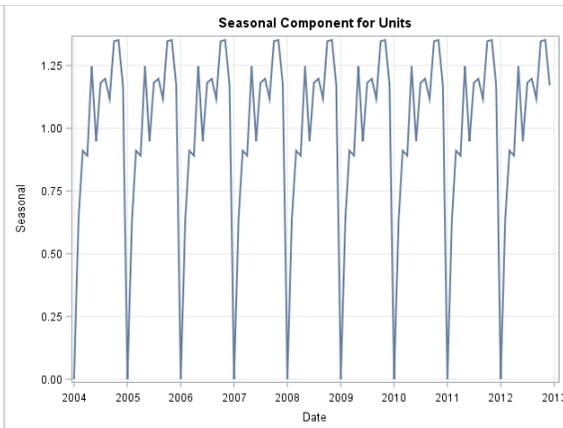


Figure 2. Seasonal Component

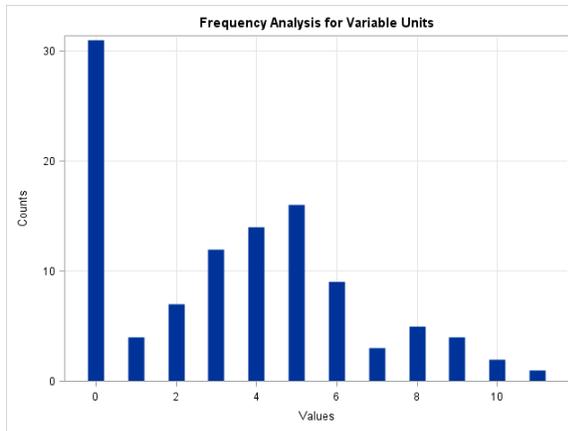


Figure 3. Frequency Plot

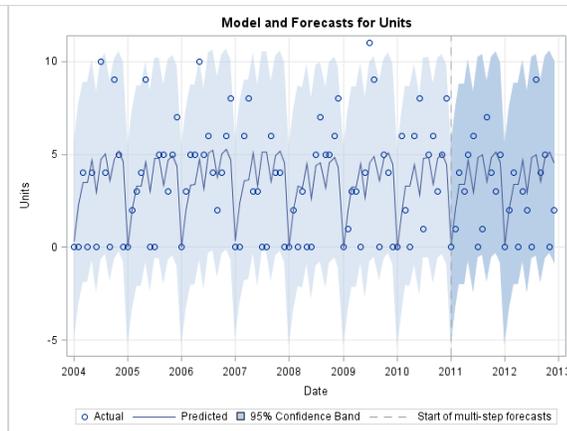


Figure 4. Continuous Time Series Model Forecasts

Figure 1 shows the monthly time series values. Figure 2 shows that the seasonal component of the time series exhibits a strong seasonal pattern. Figure 3 shows that the time series values are discretely valued (0, 1, 2, ..., 11) and that there are many zero values. This time series is a seasonally varying count series with zero inflation. Figure 4 shows the forecasts for seasonal exponential smoothing model. These forecasts have confidence regions that extend to negative values, which are not realistic. In addition, the confidence region is not integer-valued.

Most traditional time series techniques would have difficulty forecasting this count series because of the discrete nature of the data and the zero-inflation. Static discrete distribution analysis would have difficulty capturing the seasonal variations (and possible trends and calendar events that might exist) because these analyses are not dynamic. In addition, static analysis would fail to capture the increasing uncertainty that is associated with long forecast horizons. A hybrid of the two analyses is needed to create realistic forecasts for this count series.

This paper focuses on discrete time series data that are generated by business and economic activity. The techniques described in this paper are less applicable to voice, image, and multimedia data or to continuous-time data.

COUNT SERIES FORECASTING METHOD

To overcome the problems associated with using continuous time series models for count series data, this paper proposes a method of combining automatic time series modeling with automatic discrete probability distribution selection. This proposed method consists of these steps:

1. Automatically select the discrete probability distribution.
2. Automatically forecast the counts series (using traditional exponential smoothing methods).
3. Use information from the selected discrete probability distribution to adjust the forecast.

Step 1: Automatically Select the Discrete Probability Distribution

You can use the TIMESERIES to procedure perform many types of analyses that are related to time series. The COUNT statement, new in SAS/ETS 14.1, enables you to analyze count series. For this example, the CountSeries data set contains monthly count series data from an inventory system. The variable Date records the time ID values, and the variable Units records the time series values for the inventory units.

The following statements analyze the count series in the CountSeries data set:

```
proc timeseries data=countseries out=_NULL_ print=counts countplot=all;
  count / distribution=(zmbinomial zmgeometric zmpoisson)
  criterion=aic alpha=0.05;
  id Date interval=month;
  var Units;
run;
```

The PRINT=COUNTS and COUNTPLOT=ALL options in the PROC TIMESERIES statement request that count series analysis be performed. The DISTRIBUTION= option in the COUNT statement specifies the list of candidate distributions (ZMBINOMIAL, ZMGEOMETRIC, and ZMPOISSON) that are to be considered. The CRITERION= option specifies AIC as the distribution selection criterion. The ALPHA=0.5 specifies the confidence limit size. See Appendix D for more information about the COUNT statement.

Table 1 describes the values PROC TIMESERIES uses to automatically select the discrete probability distribution. PROC TIMESERIES fits the three specified distributions and then evaluates the fit based on the specified selection criterion. Because the zero-modified Poisson distribution has the lowest AIC value, it is the selected distribution. See Appendix B for more information about the automatic discrete probability distribution selection.

Discrete Distribution Selection Criterion = AIC					
Distribution	Zero-Value Log-Likelihood	Log-Likelihood	AIC	BIC	Selected
Zero-Modified Binomial	-60.215168	-217.25543	438.51086	452.93154	
Zero-Modified Geometric	-60.215168	-236.04996	476.09991	490.52059	
Zero-Modified Poisson	-60.215168	-216.33209	436.66418	451.08486	Yes

Table 1. Automatic Distribution Selection

Table 2 describes the selected distribution parameter estimates. The parameter p_0^M is the zero-modification (percentage), and the parameter lambda is the Poisson parameter. Notice that 29% of the data is related to zero-value modification.

ZMPOISSON Parameter Estimates						
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	95% Confidence Limits	
p0M	0.29000	0.04538	6.39	<.0001	0.20106	0.37894
lambda	4.84903	0.02654	182.68	<.0001	4.79700	4.90105

Table 2. Distribution Parameter Estimates

Table 3 describes the distribution estimates, which are based on the parameter estimates in Table 2. Notice that the confidence limits are nonnegative and integer-valued.

ZMPOISSON Distribution Estimates for Units				
Mean	Variance	Standard Error	95% Confidence Limits	
3.47000	8.25522	2.87319	0	8.00000

Table 3. Distribution Estimates

Figure 5 illustrates the selected discrete probability distribution (ZMPOISSON). The blue bars represent the nonzero count series values. The orange bar represents the zero count series values (on a different scale). The blue line represents the estimated discrete probability distribution values. The small circle and the smaller bar represent the zero modification. The dashed vertical line represents the distribution mean, and the blue shading represents the 95% confidence region for the distribution mean.

Figure 6 illustrates the chi-square probabilities (on a log scale) of the selected distribution (ZMPOISSON). The blue bars represent the probabilities. The horizontal lines represent standard significance thresholds of 0.5 and 0.01. The dashed vertical line represents the distribution mean, and the blue shading represents the 95% confidence region. This analysis suggests that the zero-modified Poisson distribution fits the data reasonably well (all the blue bars are well below the thresholds).

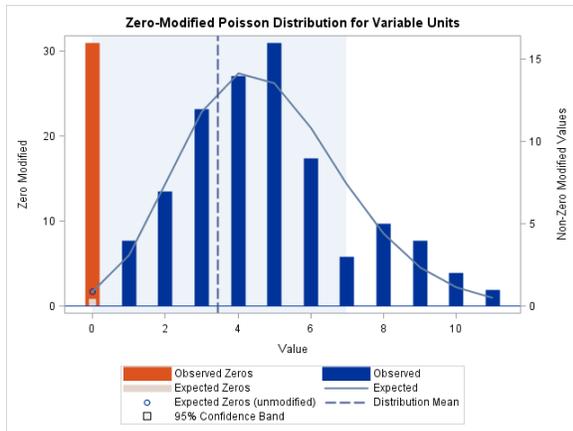


Figure 5. Estimated Distribution

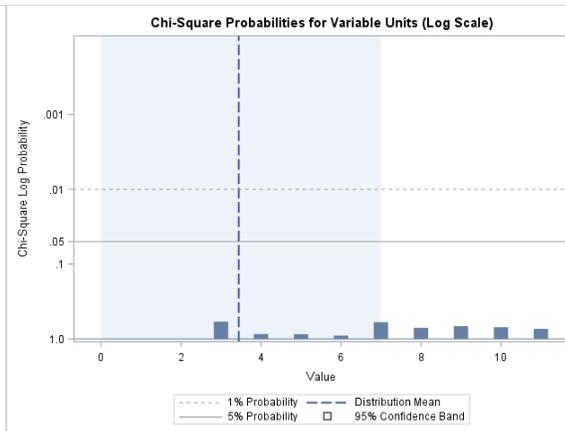


Figure 6. Chi-Square Probabilities for the Distribution

Step 2: Automatically Forecasting the Count Series

SAS Forecast Server Procedures automatically forecasts time series by using input variables and calendar events. The HPF procedure performs automatic forecasting by using a list of exponential smoothing models.

The following statements perform automatic time series forecasting for the CountSeries data set:

```
proc hpf data=countseries out=_NULL_ outfor=forecasts(rename=(actual=Units))
    plot=modelforecasts lead=24 back=24;
    id date interval=month;
    forecast units / model=bests alpha=0.05 select=aic;
run;
```

The LEAD=24 and BACK=24 options in the PROC HPF statement specify the size of the forecast region and the out-of-sample region, respectively. The FORECAST statement specifies the desired automation. The MODEL=BESTS option requests that seasonal models be considered. The SELECT=AIC option specifies the selection criterion for the time series models. No HOLDOUT= option is specified, so the models are evaluated based on the fit region. The ALPHA=0.5 option specifies the confidence limit size. See Appendix A for more information about automatic time series forecasting.

Table 4 describes the automatic time series model selection results; the seasonal exponential smoothing model was selected because it has the lowest AIC value. The multiplicative Winter's method was rejected because the count series contains zero values.

Model Selection Criterion = AIC		
Model	Statistic	Selected
Seasonal Exponential Smoothing	171.71847	Yes
Winters Method (Additive)	173.61427	
Winters Method (Multiplicative)	.	

Table 4. Automatic Time Series Model Selection

Figure 7 illustrates the time series forecasts *without* adjustments for the count series distribution. The blue circles represent the actual time series values. The blue line represents the predictions, and the blue shading represents the 95% confidence region. The vertical line separates the in-sample and out-sample regions.

Step 3: Adjusting for Counts Series

Because the zero-modified Poisson was selected for the count series, the time series forecasts need to be adjusted accordingly. See Appendix C for more for more information about count series forecasting.

You can use the TIMEDATA procedure in SAS/ETS to perform preprocessing and postprocessing of the time series data, including forecasts.

The following statements perform the count series adjustment of the time series forecasts:

```
proc timedata data=forecasts(drop=lower upper) out=_NULL_
    outarray=adjusted(drop=predict);
var units predict std error;
id date interval=month;
outarray lower upper ppred;
do t=1 to _LENGTH_;
    ppred[t] = max(predict[t],0.0001);
    lower[t] = quantile('POISSON',1 - 0.5/2, ppred[t]);
    upper[t] = quantile('POISSON',    0.5/2, ppred[t]);
end;
run;
```

Figure 8 illustrates the time series forecasts *with* adjustments for the count series distribution. The blue circles represent the actual time series values. The blue line represents the predictions, and the blue shading represents the 95% confidence region. The vertical line separates the in-sample and out-sample regions.

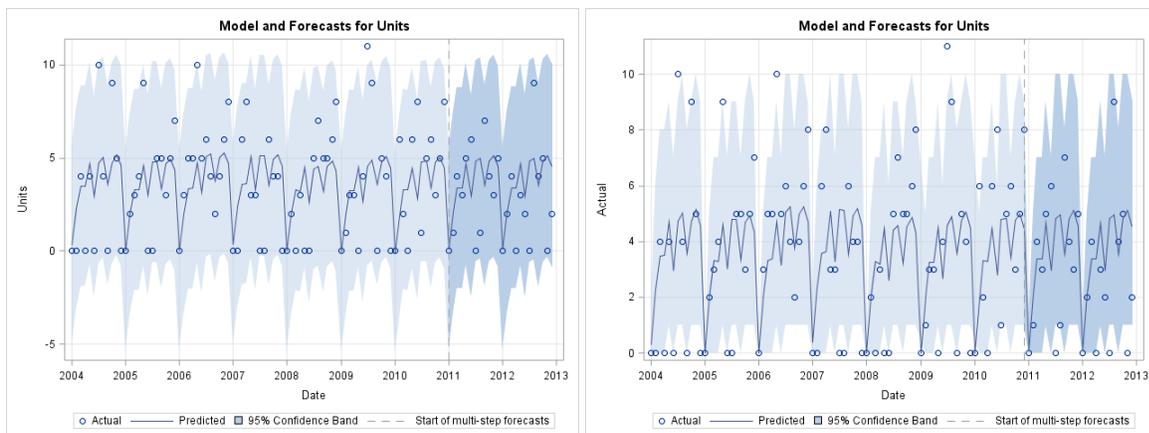


Figure 7. Count Series Forecasts (No Adjustment) Figure 8. Count Series Forecasts (with Adjustment)

When you compare Figure 7 and 8, you can see that Figure 7 has confidence regions that extend to negative values, which are not realistic. In addition, the confidence values in Figure 7 are not integer-valued. In Figure 8, the confidence regions are nonnegative and integer-valued, and the confidence region is narrower.

LIMITATIONS

The method that is proposed in this paper will not work when the majority of time series values are zero. Intermittent demand models, such as Croston's Method, are more appropriate. In addition, there are more complex techniques (such as integer-valued autoregression models) that are data intensive—for example, Poisson autoregression (PAR) models.

CONCLUSION

This paper illustrates a technique that combines automatic discrete probability distribution selection with automatic time series model selection to produce count series forecasts and shows how you can use SAS Forecast Server Procedures software to forecast count series. This technique is useful for count series forecasting and count series monitoring.

RECOMMENDED READING

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). You can find introductory discussions about discrete probability distributions in Stuart, Panjer, and Willmot (2012).

Appendices A, B, and C explain the mathematics of the proposed method for combining automatic time series forecasting with automatic count series analysis. Appendix D explains the syntax related to the TIMESERIES procedure that is relevant to count series analysis.

APPENDIX A: TIME SERIES MATHEMATICAL CONCEPTS

This appendix explains the mathematical concepts that are associated with time series analysis, modeling, and forecasting. If you already understand the mathematical notation that is used in most time series analysis, you might choose to skip this appendix.

Time Indexing

Time series analysis and forecasting use the following time indices:

- A (discrete) time index is represented by $t = 1, \dots, T$, where T represents the length of the historical time series.
- The index of future time periods (which are predicted by forecasting) is represented by $l = 1, \dots, L$, where L represents the forecast horizon (also called the lead).
- The holdout region index is represented by $h = 1, \dots, H$, where h represents the length of the out-of-sample selection region (also called the holdout region). You use this index when you want to use out-of-sample evaluation techniques to select time series models. When $H = 0$, in-sample fit selection is used.
- The performance region index is represented by $b = 1, \dots, B$, where B represents the length of the out-of-sample performance region (also called the holdback region). When $B = 0$, in-sample performance analysis is used.

Various combinations of $T > 0$, $B \geq 0$, $H \geq 0$, and $L \geq 0$ produce results. When $B > 0$, B is usually equal to L , but this is not necessarily so.

Figure A1 illustrates how the time series indices can divide the time line. Note that the forecast horizon can extend beyond the out-of-sample and performance regions.

Time Regions ($H > 0$, $B > 0$)

Historical Time Series (1,...,T)		
In-Sample Region (1,...,T - B)		Out-Sample Region (T - B + 1,..., T)
Initial Fit Region (1,...,T - B - H)	Selection Region (T - B - H + 1,...,T - B)	Performance Region (T - B + 1,..., T)
Fit Region (1,...,T - B)		Forecast Horizon (T - B + 1,..., T - B + L)

Figure A1. Time Indices Divide the Time Line

Figure A2 illustrates how the time series indices can divide an example time series data set.

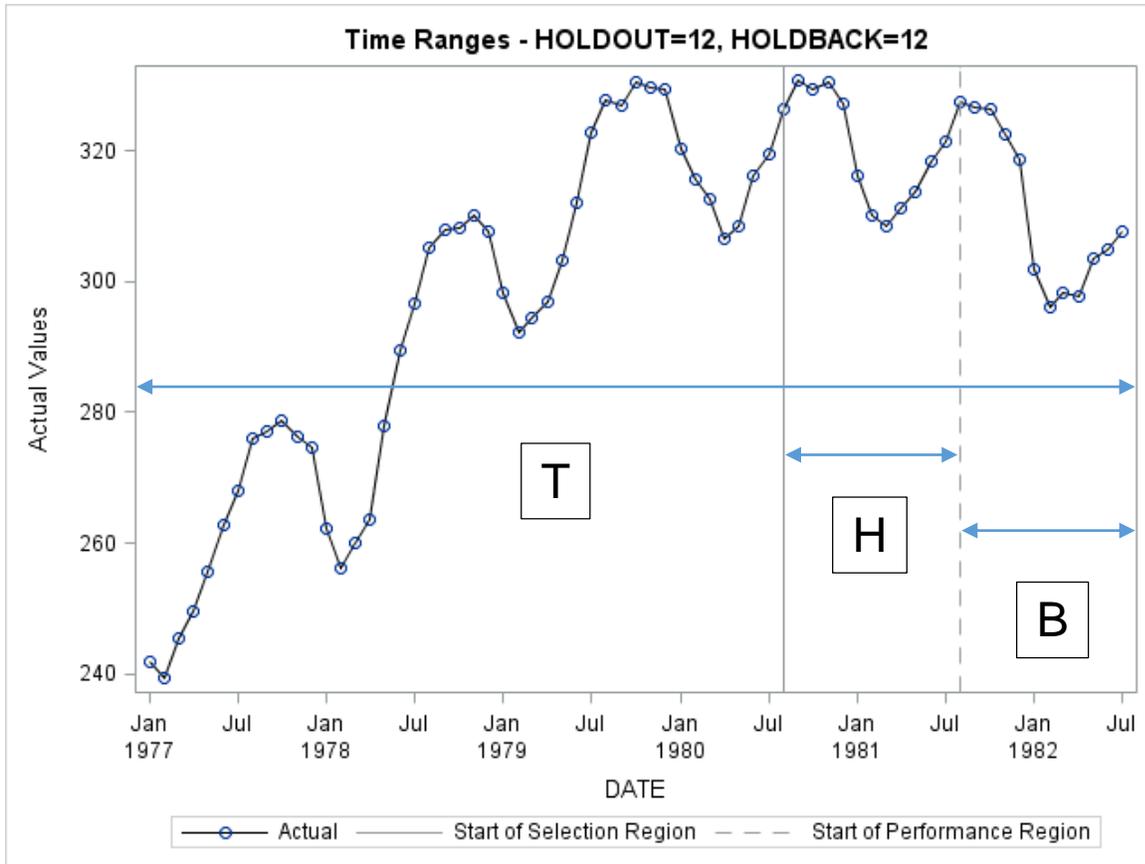


Figure A2. Time Indices Divide a Time Series Data Set

Time Series Data

A continuous time series value at time index t is represented by y_t . The *dependent series* (the historical time series vector) that you want to model and forecast is represented by $Y_T = \{y_t\}_{t=1}^T$.

The historical time series data can also contain *independent series*, such as inputs and calendar events that help model and forecast the dependent series. The historical and future predictor series vectors are represented by $\vec{X}_T = \{\vec{x}_t\}_{t=1}^{T+L}$.

Although this paper focuses only on extrapolation techniques, the concepts generalize to more complicated time series models.

Model Indices

You often analyze time series data by comparing different time series models. Automatic time series modeling selects among several competing models or uses combinations of models. The m th time series model is represented by $F_m(\cdot)$, where $m = 1, \dots, M$ and M represents the number of candidate models under consideration.

Theoretical Model for Any Time Series

You can use a theoretical time series model to model the historical time series values of any time series. When you have historical time series data, $Y_T = \{y_t\}_{t=1}^T$ and $\vec{X}_T = \{\vec{x}_t\}_{t=1}^{T+L}$, you can construct a time series model to generate predictions of future time periods. The theoretical time series model is represented by

$y_{t+l|t}^{(m)} = F_m(Y_t, \vec{X}_{t+l} : \theta_m)$, where $y_{t+l|t}^{(m)}$ represents the prediction for y_{t+l} that uses only the historical time series

data that end at time index t , for the l th future value for the m th time series model, $F_m(\cdot)$, and where θ_m represents the parameter vector to be estimated. The theoretical model can also be a combination of two or more models.

Fitted Model for a Specific Time Series

You can use a fitted time series model to predict the future time series values of a specific time series. Given specific historical time series data, $Y_T = \{y_t\}_{t=1}^T$ and $\bar{X}_T = \{\bar{x}_t\}_{t=1}^{T+L}$, and a theoretical time series model,

$y_{t+l|t}^{(m)} = F_m(Y_t, \bar{X}_{t+l} : \theta_m)$, you can estimate a fitted model to predict future time periods. The fitted model is represented by $\hat{y}_{t+l|t}^{(m)} = F_m(Y_t, \bar{X}_{t+l} : \hat{\theta}_m)$, where $\hat{y}_{t+l|t}^{(m)}$ represents the prediction for y_{t+l} and $\hat{\theta}_m$ represents the estimated parameter vector.

Model Predictions

You can use a fitted time series model to generate model predictions. The model predictions for the m th time series model for the l th time period are represented by $\hat{y}_{t+l|t}^{(m)} = F_m(Y_t, \bar{X}_{t+l} : \hat{\theta}_m)$, where the historical time series data up to the $(t-l)$ time period are used. The prediction vector is represented by $\hat{Y}_T^{(m)} = \{\hat{y}_{t+l|t}^{(m)}\}_{t=1}^T$.

For $l = 1$, the *one-step-ahead prediction* for the l th time period and for the m th time series model is represented by $\hat{y}_{t|t-1}^{(m)}$. The one-step-ahead predictions are used for *in-sample* evaluation. They are commonly used for model fit evaluation, in-sample model selection, and in-sample performance analysis.

For $l \geq 1$, the *multi-step-ahead prediction* for the l th time period and for the m th time series model is represented by $\hat{y}_{t+l|t}^{(m)}$. The multi-step-ahead predictions are used for *out-of-sample* evaluation. They are commonly used for holdout selection, performance analysis, and forecasting.

Prediction Errors

There are many ways to compare predictions that are associated with time series data. These techniques usually compare the historical time series values, y_t , with their time series model predictions, $\hat{y}_{t+l|t}^{(m)} = F_m(Y_t, \bar{X}_{t+l} : \hat{\theta}_m)$.

The *prediction error* for the m th time series model for the l th time period is represented by $\hat{e}_{t+l|t}^{(m)} = y_t - \hat{y}_{t+l|t}^{(m)}$, where the historical time series data up to the $(t-l)$ th time period are used.

For $l = 1$, the *one-step-ahead prediction error* for the l th time period and for the m th time series model is represented by $\hat{e}_{t|t-1}^{(m)} = y_t - \hat{y}_{t|t-1}^{(m)}$. The one-step-ahead prediction errors are used for *in-sample* evaluation. They are commonly used for evaluating model fit, in-sample model selection, and in-sample performance analysis.

For $l > 1$, the *multi-step-ahead prediction error* for the l th time period and for the m th time series model is represented by $\hat{e}_{t+l|t}^{(m)} = y_t - \hat{y}_{t+l|t}^{(m)}$. The multi-step-ahead prediction errors are used for *out-of-sample* evaluation. They are commonly used for holdout selection and performance analysis.

Statistics of Fit

A statistic of fit measures how well the predictions match the actual time series data. (A statistic of fit is sometimes called a goodness of fit.) There are many statistics of fit: RMSE, MAPE, MASE, AIC, and so on. Given a time series

vector, $Y_T = \{y_t\}_{t=1}^T$, and its associated prediction vector, $\hat{Y}_T^{(m)} = \{\hat{y}_{t+l|t}^{(m)}\}_{t=1}^T$, the relevant statistic of fit is represented by $SOF(Y_T, \hat{Y}_T^{(m)})$. The statistic of fit can be based on the one-step-ahead or multi-step-ahead predictions.

For $Y_T = \{y_t\}_{t=1}^T$ and $\hat{Y}_T^{(m)} = \{\hat{y}_{t+l|t}^{(m)}\}_{t=1}^T$, the *fit statistics* for the m th time series model are represented by $SOF(Y_T, \hat{Y}_T^{(m)})$. The one-step-ahead prediction errors are used for *in-sample* evaluation. They are commonly used

for model fit evaluation, in-sample model selection, and in-sample performance analysis.

For $Y_H = \{y_t\}_{t=T-H+1}^T$ and $\hat{Y}_H^{(m)} = \{\hat{y}_{t|T-H}^{(m)}\}_{t=T-H+1}^T$, the *performance statistics* for the m th time series model are represented by $SOF(Y_H, \hat{Y}_H^{(m)})$. The multi-step-ahead prediction errors are used for *out-of-sample* evaluation. They are commonly used for holdout selection and out-of-sample performance analysis.

Model Selection List

No one modeling technique works for all time series data. A model selection list contains a list of candidate models and describes how to select among the models. The model selection list is represented by $\{F_m(\cdot)\}_{m=1}^M$.

The model selection list is subsetted by using the selection diagnostics (intermittency, trend, seasonality, and so on) and a model selection criterion (such as RMSE, MAPE, and AIC). The selected model is represented by

$$F_*(\cdot) \in \{F_m(\cdot)\}_{m=1}^M.$$

After a model is selected, the selection region is dropped. The model is refitted to all the data except the performance region. See Figure A1 and Figure A2 for a graphical description.

Model Generation

The list of candidate models can be specified by the analyst or automatically generated by the series diagnostics (or both). The diagnostics consist of various analyses that are related to intermittency, trend, seasonality, autocorrelation, and, when predictor variables are specified, cross-correlation analysis.

The *series diagnostics* are used to generate models, whereas the *selection diagnostics* are used to select models.

APPENDIX B: COUNT SERIES MATHEMATICAL CONCEPTS

This appendix explains the mathematical concepts associated with count series analysis and forecasting. In particular, this appendix explains how count series analysis differs from continuous time series analysis. Count series are time series whose values take on small nonnegative integer values (counts). Most of this notation closely follows Klugman et al. (2012), whose analyses are summarized for convenience.

Count Series Data

A time series value at time index t is represented by y_t , where y_t takes on small nonnegative integer values (counts). Such time series are often called *count series*. In many applications, the count series takes on the value of zero either more often or less often than usual. Such count series are called *zero-modified* (or *zero-inflated*) count series.

Count Series Indexing, Values, and Counts

The time series value of time index t , y_t , can take on the count values, $k=0, \dots, K$, where K represents the maximum (observed or possible) count value. The number of times that the value k occurs in the historical time series is represented by n_k . In other words, $y_t \in \{0, \dots, K\}$ and $n_k = \|\{t : y_t = k\}\|$, where n_k is the count series sample distribution when the time index is ignored. For example, n_0 represents the number of observed zero values in the historical time series, y_t .

Count Series Sample Statistics

Given a count series, y_t , the following sample statistics related to the counts can be computed:

- The total counts observed throughout the time series is represented by

$$n = \sum_{k=0}^K n_k$$

- The sample mean is represented by

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^K k n_k$$

- The sample variance is represented by

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^K k^2 n_k - \hat{\mu}^2$$

Note that there is no denominator correction.

The number of zero values are unusually large or small in many applications. It is important to measure the zero component of the count series. When the number of zero values is n_0 , the number of nonzero values is $(n - n_0)$.

Let p_0^M represent the initial zero value probability estimate:

$$p_0^M = n_0 / n$$

Several discrete probability distributions can be estimated from the sample statistics n , $\hat{\mu}$, $\hat{\sigma}^2$, and p_0^M .

Count Series Theoretical Distributions

Many time series analyses assume continuous time series values (also called disturbance series values); hence, these analyses assume continuous probability distributions such as the normal (Gaussian) distribution. Count series analysis uses discrete probability distributions.

Table B1 shows distributions that are used in count series analysis.

Distribution	Formula	Parameter Estimates	Distribution Estimates	Distribution Predictions
Poisson	$N \sim f(k : \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}$ <p>where $\lambda > 0$ and $k = 0, \dots, K$. Because λ is unknown, it must be estimated from the data.</p>	$\hat{\lambda} = \hat{\mu}$ $Var[\hat{\lambda}] = \hat{\lambda} / n$	$E[N] = \lambda$ $Var[N] = \lambda$ Note: The variance equals the mean.	$a = 0$ $b = \lambda$ $p_0 = e^{-\lambda}$ $p_k = (a + b/k)p_{k-1}$
Geometric	$N \sim f(k : \beta) = \frac{\beta^k}{(1 + \beta)^{k+1}}$ <p>where N is the number of trials until first success and $k = 0, \dots, K$. Because β is unknown, it must be estimated from the data.</p>	$\hat{\beta} = \hat{\mu}$ $Var[\hat{\beta}] = \hat{\beta}(1 + \hat{\beta}) / n$	$E[N] = \beta$ $Var[N] = \beta(1 + \beta)$	$a = \frac{\beta}{(1 + \beta)}$ $b = 0$ $p_0 = \frac{1}{(1 + \beta)}$ $p_k = (a + b/k)p_{k-1}$
Binomial	$N \sim f(k : K, q) = \binom{K}{k} q^k (1 - q)^{K-k}$ <p>where $K > 0$ is the maximum possible value, $(0 < q < 1)$ is probability of "success," and $k = 0, \dots, K$. Because q is unknown, it must be estimated from the data. See Note 1 for K.</p>	$\hat{q} = \hat{\mu} / \hat{K}$ $Var[\hat{q}] = \hat{q}(1 - \hat{q}) / (\hat{K}n)$	$E[N] = Kq$ $Var[N] = Kq(1 - q)$ Note: The variance less than the mean.	$a = -\frac{q}{(1 - q)}$ $b = (K + 1) \frac{q}{(1 - q)}$ $p_0 = (1 - q)^K$ $p_k = (a + b/k)p_{k-1}$
Negative binomial	$N \sim f(k : \beta, r) = \frac{r(r+1) \dots (r+k-1) \beta^k}{k! (\beta+1)^{r+k}}$ <p>where $k = 1, \dots, K$. Because β and r are unknown, they must be estimated from the data.</p>	$\hat{\beta} = \frac{\hat{\sigma}^2}{\hat{\mu}} - 1$ $Var[\hat{\beta}] = \beta(1 + \beta) / (rn)$ $\hat{r} = \frac{\hat{\mu}^2}{\hat{\sigma}^2 - \hat{\mu}}$ $Var[\hat{r}] = r / n$	$E[N] = r\beta$ $Var[N] = r\beta(1 + \beta)$ Note: The variance greater than the mean.	$a = \frac{\beta}{(1 + \beta)}$ $b = \frac{(r-1)\beta}{(1 + \beta)}$ $p_0 = (1 + \beta)^{-r}$ $p_k = (a + b/k)p_{k-1}$

Table B1: Discrete Probability Distributions

Generic Discrete Probability Distribution

As demonstrated in the Table B1, a discrete probability distribution can be estimated from the sample statistics n , $\hat{\mu}$, $\hat{\sigma}^2$, and p_0^M . A function that estimates the distribution parameter vector from the sample statistics is represented by $\hat{\phi} = g(n, \hat{\mu}, \hat{\sigma}^2, p_0^M)$, where the function $g(\)$ depends on the discrete probability distribution, N :

$$N \sim f(k : \phi) \quad \text{where } k = 0, \dots, K$$

Discrete Probability Distribution Automatic Selection

There might be several discrete probability distributions under consideration. Suppose $d = 1, \dots, D$ represents the distribution index from which to select. In this case, $\hat{\phi}$ and N are indexed as follows:

$$\hat{\phi}^{(d)} = g^{(d)}(n, \hat{\mu}, \hat{\sigma}^2, p_0^M)$$

$$N \sim f^{(d)}(k : \phi^{(d)}) \quad \text{where } k = 0, \dots, K$$

By computing the log likelihood, $L^{(d)}$, that is associated with each distribution, an appropriate distribution can be selected by using likelihood statistics such as Akaike's information criterion (AIC) or the Bayesian information criterion (BIC), where the AIC and BIC are computed as

$$AIC^{(d)} = -2L^{(d)} + 2r$$

$$BIC^{(d)} = -2L^{(d)} + r \log(n)$$

where n represents the total number of values and r represents the number of parameters that are associated with each distribution.

APPENDIX C: COUNT SERIES FORECASTING MATHEMATICAL CONCEPTS

This appendix explains the mathematical concepts that are associated with count series forecasting. In particular, this appendix explains how count series forecasts can be generated from continuous time series forecasts. Some continuous time series models—in particular, exponential smoothing models (ESM)—have rather weak distributional assumptions. There are more complicated models that use exact distributions. For example, see Koopman (2005) for count series state space models. However, this paper's proposed method is robust, simple to implement, and easier to understand and evaluate.

Count Series Forecast

As described in Appendix A, given a count series, y_t , a continuous time series model can be used to generate a forecast, $\hat{y}_{t|t-l}^{(m)} = F_m(Y_t, \bar{X}_{t+l} : \hat{\theta}_m)$. These forecasts represent estimates of the time-varying *mean* and time-varying *variance*.

When the continuous time series notation and count series notation are combined, the time-varying count series *mean* is represented by $\hat{\mu}_t = \hat{y}_{t|t-l}^{(m)}$ and the time-varying count series *variance* is represented by $\hat{\sigma}_t^2 = \text{Var}[\hat{y}_{t|t-l}^{(m)}]$.

The feasibility of this simplification is the essential assumption of this paper. In essence, this simplification is similar to a moving window (similar to a moving average), where distribution parameters are estimated for each window. (For example, an ESM can be viewed as a clever, weighted moving window.)

As stated earlier, several discrete probability distributions can be estimated from the sample statistics n , $\hat{\mu}$, $\hat{\sigma}^2$, and p_0^M . If you consider n and p_0^M to be fixed with respect to time and allow $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ to vary with respect to time, you can compute time-varying distribution estimates for a desired discrete probability distribution.

Sometimes, both y_t and y_t^2 are modeled independently. The forecasts for y_t are used for the time-varying mean estimates, $\hat{\mu}_t$. The forecasts for y_t^2 are used for the time-varying variance estimates, $\hat{\sigma}_t^2$.

More formally, the estimated parameter vector at time index t is represented by $\hat{\phi}_t = g(n, \hat{\mu}_t, \hat{\sigma}_t^2, p_0^M)$ and the selected discrete probability distribution is represented by N_t :

$$N_t \sim f^*(k : \phi_t) \quad \text{where } k = 0, \dots, K$$

The forecasts for N_t are the predicted values $E[N_t]$. The prediction standard variance $Var[N_t]$ and the discrete-valued confidence limits can be computed from the quantiles of the estimated distribution.

Discrete-valued confidence limits are important for inventory control for “slow moving” items and monitoring count series.

APPENDIX D: SAS IMPLEMENTATION

The TIMESERIES procedure performs numerous time analyses, including count series analysis count series analysis. The following describes the statements and options that are related to count series analysis.

COUNT < / options > ;

Starting with SAS/ETS 14.1, you can use a COUNT statement in a TIMESERIES procedure call to specify options that are related to the discrete distribution analysis of the accumulated time series. Only one COUNT statement is allowed.

You can specify the following *options* in the COUNT statement after the slash (/):

ALPHA=number

specifies the confidence limit size. The value of *number* must be between 0 and 1. By default, ALPHA=0.5.

CRITERION=LOGLIK | AIC | BIC

specifies the discrete probability distribution selection criterion. You can specify the following selection criteria:

AIC	Akaike’s information criterion
BIC	Bayesian information criterion
LOGLIK	log-likelihood

By default, CRITERION=LOGLIK.

DISTRIBUTION=option | (options)

specifies one or more discrete probability distributions for automatic selection. You can specify one or more of the following distributions (if you specify more than one distribution, use spaces to separate the list items and enclose the list in parentheses):

BINOMIAL	binomial distribution
ZMBINOMIAL	zero-modified binomial distribution
GEOMETRIC	geometric distribution
ZMGEOMETRIC	zero-modified geometric distribution
POISSON	Poisson distribution
ZMPOISSON	zero-modified Poisson distribution
NEGBINOMIAL	negative binomial distribution

By default, all distributions are considered.

COUNTPLOT=option | (options)

specifies the type of count series graphical output. You can specify one or more of the following options (if you specify more than one graph option, use spaces to separate the list items and enclose the list in parentheses):

- COUNTS** plots the counts of the discrete values of the time series.
- CHISQPROB** plots the chi-square probabilities.
- DISTRIBUTION** plots the discrete probability distribution.
- VALUES** plots the distinct values of the time series.
- ALL** is the same as specifying PLOTS=(COUNTS CHISQPROB DISTRIBUTION VALUES).

By default, there is no graphical output.

PRINT=option | (options)

specifies the type of printed output. The PRINT= option enables you to specify many types of time series analysis. The following option is related to count series analysis (if you specify more than one printing option, use spaces to separate the list items and enclose the list in parentheses):

- COUNTS** prints the discrete distribution analysis.

By default, there is no printed output.

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. 2000. *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.
- Klugman, S. A., Panjer, H. H., and Willmot, G.E. 2012. *Loss Models: From Data to Decisions*. New York: John Wiley & Sons, Inc.
- 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., Wheelwright, S. C., and Hyndman, R. J. 1997. *Forecasting: Methods and Applications*. New York: Wiley.

CONTACT INFORMATION

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

Michael Leonard
SAS
SAS Campus Drive
State ZIP: 27513
919-531-6967

Michael.Leonard@sas.com

Bruce Elsheimer
SAS
SAS Campus Drive
State ZIP: 27513
919-531-5959

Bruce.Elsheimer@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.