SESUG Paper 161-2019

Time Series Forecasting using PROC MCMC

SESUG® 2019

Murali Sastry, LGFCU, Dr. Tyler Hicks, University of Kansas

ABSTRACT

Forecasting variables of interest in time series analysis can be quite intriguing. It can also be challenging to identify various factors that predict time series data patterns. For example, time series forecasting has been effectively applied in scheduling nurses in hospital emergency rooms, where the number of patients routinely fluctuates throughout the day based on the time of month and time of year. Similarly, time series forecasting has been helpful in forecasting credit card use by cardholders based on their past card usage history. The tricky part in modeling time series data is to combine variables such as seasonality, trend, and regressor components in one model to develop an accurate forecast. Whereas SAS/STAT® PROC ARIMA (Autoregressive Integrated Moving Average Procedure) offers SAS users a classical approach to estimating models for time series data with autoregressive, differencing, and moving average structures, it is also possible for SAS users to fit these same models to time series data using Bayesian approach with SAS/STAT® PROC MCMC (Markov Chain Monte Carlo procedure). The intent of this paper is to provide SAS users with example code and demonstrations on how to use PROC MCMC to estimate models suitable in time series analysis.

Keywords: PROC MCMC, Bayesian Structural Model, PROC ARIMA, Time Series Forecasting

INTRODUCTION

A remarkable application of Markov Chain Monte Carlo (MCMC) methods in statistical science came about through a statistical publication by Tanner & Wong (1987). They disseminated its usefulness for posterior distribution calculations to statisticians. Till then frequentist statistical methods were used in practice to address unknown parameters by considering them as fixed constants. Since the classical statistical methods consider parameters as fixed constants, probabilistic statements could not logically be made using these methods. In contrast, the Bayesian approach is to address unknown parameters (or uncertainty about parameters) through probabilistic statements and distributions rather than exact value as described earlier with classical statistical methods.

Per SAS (2014), the Bayesian approach to estimating parameters has three fundamental steps. These steps are:

1. Based on previously known information, facts, or subject expertise, a probability distribution is assigned to parameters, which is known as prior distribution (before the review of data of interest); designated by $\pi(\Theta)$.
2. Based on observed data variable of interest (y), a statistical model $p(y|\Theta)$ to describe distribution for y is selected, given a model (likelihood function).
3. Previous knowledge is enhanced by merging prior distribution knowledge with observed data by calculating a posterior distribution, $p(\Theta|y)$.

The Bayesian approach typically includes extensive computations by simulation to derive posterior distributions in all but the most trivial applications. That is, modern Bayesian computation is typically accomplished by randomly sampling values from the target posterior distribution and then inferring the posterior distribution parameters and shape from summary statistics derived from the generated posterior sample.

PROC MCMC offers users a suite of algorithms to implement MCMC methods, such as the self-tuning Metropolis-Hastings algorithm. Different algorithms define rules for taking sequential samples of random variables from the target distribution. In an MCMC process, however, each sample depends on the immediately preceding sample denoted by l1 or rv.1 (rv=random variable, and l1=immediately preceding sample). Hence this linked event is referred to as *Markov Chain*.

Monte Carlo integration is then utilized to estimate the parameters that define a target posterior distribution with the MCMC sample.

This dataset provides monthly international airline passenger counts from Jan 1949 to Dec 1960 in what many airline industry enthusiasts and airline passengers consider as “Golden age of air travel”. SAS documentation on PROC ARIMA provides the worked example of the same dataset.

PURPOSE

The purpose of this paper is to demonstrate the use of PROC MCMC procedure for time series data. Time series forecasting involves prediction of future observations based on a trend in previous data. Descriptive time series modeling involves analysis of existing time series data that dissects level, trend, seasonality, and noise components.

This paper demonstrates the time series application of PROC MCMC through international airline passenger count data. Because of the centrality of this dataset in this paper, a brief description is in order. This is the dataset used by Box and Jenkins as an example to explain time series applications in their seminal work. It provides monthly international airline passenger counts from Jan 1949 to Dec 1960 in what many airline industry enthusiasts and airline passengers consider as the “Golden age of air travel”. We selected this database because SAS documentation on PROC ARIMA provides the worked example of the same dataset for reference when discussing the application of PROC MCMC to this dataset in time series forecasting. We use international airline passenger data from 1949-60 to compare the two SAS procedures mentioned above.

Time series forecasting involves prediction of future observations. Descriptive Time Series modeling involves analysis of existing time series data that dissects level, trend, seasonality, and noise components. Per Scott and Varian (2013), trend and seasonality are captured by time series component of the MCMC model while regression component captures relationship between factors impacting the response variable.

Per Larsen (2016), Bayesian Structural Time Series (BSTS) model with unobserved components provides better transparency than ARIMA models as its method does not rely on differencing, lags, and moving averages. In addition, one can visually inspect the underlying components of the model in the MCMC procedure.

The following points from Larsen (2016) provided further motivation for authors of this paper to simulate the airline model using SAS PROC MCMC procedure:

- MCMC handles uncertainty better as one can quantify posterior uncertainty of individual components;
- Regressors, unobserved trend (local level term), and seasonality can be estimated simultaneously;
- Control the variance of components;
- Make assumptions with prior distributions of parameters; and,
- An ARIMA Time Series model can be recast as a structural model.

The basic Structural Time Series model is represented by the following algorithm per Ghosh, Prajenshu, and Wadhwa (2017):

$$Y_t = T_t + C_t + S_t + \varepsilon_t \tag{1}$$

Equation (1) above shows Y_t is the observed values of response variable, T_t is the trend component, C_t is the cyclical component, S_t is the seasonal component, and ε_t is the irregular component. All these components are of stochastic nature, while Bayesian structural time series shows autoregressive behavior with more weight attributed to the value that is immediately preceding the current value, i.e., and might have correlation with a T_{t-1}, than to T_{t-12}, for example.

As the example chosen to demonstrate in this paper is based on time series with dependencies on random lag variables, Per SAS Institute (n.d.), the equations that follow demonstrate the dependencies and relationships to explain the time series in theoretical constructs that have trend, mean, and seasonality.

Model-Equations	Interpretation	#
$y_t \sim N(x_t, \theta_4)$	Observational Model	(1)
$x_t = \mu_t + s_t$	Trend and seasonal effects	(2)
$\mu_t \sim N(M_{t-1} + \alpha_{t-1}, \theta_1)$	Evolution in mean	(3)
$\alpha_{t-1} \sim N(\phi\alpha_{t-2}, \theta_2)$	Increments in mean	(4)
$S_t \sim N(-S_{t-1} - S_{t-2} - S_{t-3}, \theta_3)$	Seasonality	(5)
$Y_t = Y_{t-1} + (Y_{t-12} - Y_{t-13}) + e_t$	Random variable for seasonality	(6)

The first equation (observational model), states that response variable is normally distributed, and its mean is indexed by time which indicates that the response variable is nonstationary. The second equation (trend and seasonal effects) states that the mean of response variable includes a combination of trend and seasonality. The third equation (evolution in mean), states that the trend follows random walk and drift. The fourth equation (increments in mean), states that drift term follows first order autoregressive process. Fifth equation (seasonality) states normally distributed random shock prevails with mean equal to negated sum of three previous shocks. Per Dickey (2004), the sixth equation (random variable for seasonality) models random seasonality effects. This last feature is further explained in the Method section below.

METHOD

PROC MCMC is a flexible simulation procedure to fit wide range of Bayesian models (SAS 2016). In reference to this procedure, it treats parameters as unknown random variables and makes inferences from posterior distributions of parameters. From Bayes Theorem, posterior distribution is a product of likelihood function and prior parameter distribution. Per SAS (2015), MCMC Methods sample random variables consecutively from a target distribution. Each sample depends on the previous sample. This concept is referred to as Markov Chain. Using predetermined number of samples, simulations introduce additional level of uncertainty to the accuracy of posterior estimates. Monte Carlo is used to approximate an expectation by using Markov Chain samples. Metropolis algorithm is used to generate sequence of samples from joint distribution of multiple variables. This is the foundation for MCMC.

For the four models to obtain posterior distribution the following steps were taken:

- Accessed sas dataset available at SASHELP.AIR that has DATE (Month and year) and AIR (international airline passenger count) variables;
- Added month and year variables to the dataset and took natural logarithm of AIR variable and created a new column “logair”. Per Lütkepohl & Xu (2009), using logarithms of variables help with significant gains in forecast precision if log transformation makes the variance more homogeneous for the entire sample.
- Applied PROC TIMESERIES data to plot the data to graphically view to understand data better before using PROC MCMC.
- Applied PROC MCMC for data (called the updated dataset Series G) and defined parameters (refer to Model equations and interpretation above) to account for trend and seasonality.
- Parameters to account for trend and seasonality were declared;
- Prior distributions were assumed for relevant parameters that account for trend and seasonality; And assumed standard distributions for getting prior distribution of parameters;
- Random statements were developed for declaring random effects of parameters;
- Referred to Example 2: UK Coal Consumption from SAS (n.d.) article for developing PROC MCMC program statements for Model 0, Model 2, and Model 3;

- Referred to Dickey (2004) article for developing the random statement for declaring seasonality parameter for Model 1;
- Trend and Seasonality were simultaneously declared in the likelihood function (assumed Normal distribution) for the response variable (Passenger Count or Log Passenger Count);
- Effective Sample Size table and trace plots for parameters are generated as default.
- PROC FORMAT was used to format X-Axis (Time in years) for posterior distribution plots;
- PROC SGPLOT procedure was used to generate plots of posterior distribution forecast for each model.

CODE

```
LIBNAME AIRLINE '\\<servername>\<Drive>\7.1 EG
Projects\<Dept>\murali.sastry';
%LET extension=xlsx;
%LET folderpath = \\'\\<servername>\<Drive>\7.1 EG
Projects\<Dept>\murali.sastry';
options mlogic mprint;*Import data, create variables, and format
values;
filename datfile "&folderpath\Airline_Passenger_Data";
PROC IMPORT datafile=datfile
                      out=AirData
                      dbms=xlsx
                      replace;
                      getnames=yes;
                 run;
PROC SQL;
Create table work.Air as
      Select t1.DATE
      ,t1.Passenger_Count
      ,month(t1.DATE) as Month
      ,year(t1.DATE) as Year
      ,log(t1.Passenger_Count) as logcount
      from work.AirData t1
      ;
QUIT;
  ODS GRAPHICS ON;
  ODS EXCLUDE NONE;
  ODS RESULTS;
*Plot data;
PROC TIMESERIES data=Air plot=series;
      id DATE interval=month;
      var logcount;
   run;

*Insert sample for prediction check;
DATA seriesG; set Air end=eof;
  output;
  if eof then do;
      do year=1961 to 1962;
           do month=1 to 12;
                Passenger_Count=.; logcount=.; output;
           end;
      end;
```

```
  end; RUN;
*Holdout sample for prediction check;
DATA seriesG; set seriesG;
  holdout=logcount; t=(_N_-1);
  if year(date)>1959 then logcount=.;
RUN;
*Model 0: Bayesian Structural Time Series Analysis;
PROC MCMC data=seriesG NMC=500000 NBI=50000 thin=100 seed=1234
outpost=posterior0;
  PARMS mu0;
  PARMS s2_s s2_mu s2_e;
  prior mu0 ~normal(0,var=2);
  prior s2: ~igamma(shape = 3/10, scale = 10/3);
  random s~normal(0,var=s2_s) subject=month;
  random mu~normal(mu.l1, var=s2_mu) subject=t icond=(mu0);
  model logcount~normal(mu + s,var=s2_e);
  preddist outpred=outpred;
  ODS output PredSumInt=PredSumInt0;
RUN;
*Forecast data Model 0;
DATA forecast0;
  merge seriesG PredSumInt0;
run;
PROC FORMAT;
     value timefmt 13='1950'
                   37='1952'
                   61='1954'
                   85='1956'
                   109='1958'
                   133='1960';
PROC SGPLOT DATA=forecast0;
  title1 "Model 0: Bayesian Time Series Analysis Forecast 0";
  title2 "Monthly Passengers (1949-1960)";
  format t timefmt.;
  series x=t y=logcount / LINEATTRS=(color=red pattern=longdash);
  series x=t y=holdout / LINEATTRS =(color=red pattern=dot);
  series x=t y=mean / LINEATTRS =(color=blue pattern=solid);
  YAXIS label="Count";
  XAXIS values=(13 37 61 85 109 133) ranges=(1-168) label="Year";
  REFLINE 133 / axis=x LINEATTRS=(color=black pattern=dash);
  REFLINE 6.5 / axis=y;
  band x=t upper=HPDUPPER lower=HPDLOWER / transparency=.5;
run; title1; title2;
*/Model 1:Bayesian Structural Time Series Analysis/*;
PROC MCMC data=seriesG NMC=750000 NBI=50000 thin=100 seed=1234
Diagnostics=MCSE;
  PARMS mu0;
  PARMS s2_s s2_mu s2_e;
  prior mu0 ~normal(0,var=2);
/*  prior s:~normal(0,var=s2_s) subject=month;*/
```

```
  prior s2: ~igamma(shape = 3/10, scale = 10/3);

  a=s.l1 +(s.l12-s.l13);
  random s~normal(a,VAR=s2_s) subject=month;

  random mu~normal(mu.l1, var=s2_mu) subject=t icond=(mu0);

  model Passenger_Count~normal(mu + s,var=exp(s2_e));

  preddist outpred=outpred1;
  ODS output PredSumInt=PredSumInt1;
RUN;
*Forecast data;
DATA forecast1;
  merge seriesG PredSumInt1;
RUN;
/*proc format;*/
/*    value timefmt 13='1950' */
/*                  37='1952' */
/*                  61='1954' */
/*                  85='1956' */
/*                  109='1958' */
/*                  133='1960';*/
PROC SGPLOT DATA=forecast1;
  title1 "Model 1: Bayesian Time Series Analysis Forecast 1";
  title2 "Monthly Passengers (1949-1960)";
  format t timefmt.;
  series x=t y=Passenger_Count / LINEATTRS =(color=red
pattern=longdash);
  series x=t y=holdout / LINEATTRS =(color=red pattern=dot);
  series x=t y=mean / LINEATTRS =(color=blue pattern=solid);
  YAXIS label="Count";
  XAXIS values=(13 37 61 85 109 133) ranges=(1-168) label="Year";
  REFLINE 133 / axis=X LINEATTRS=(color=black pattern=dash);
  REFLINE 6.5 / axis=Y;BAND x=t upper= HPDUPPER lower=HPDLOWER /
transparency=.7;
RUN; title1; title2;
```

Code for Model 2 and Model 3 are provided in Appendix A.

DISCUSSION

This section includes the interpretation and our findings in light of what is already known about the PROC MCMC procedure and to explain any new understanding or insights that emerged as a result of our study of time series forecasting in applying PROC MCMC procedure. For the airline dataset, the authors decided to hold out data for the years 1959-60 to observe the accuracy of posterior distribution for this period.

PROC TIMESERIES DATA procedure with the plot statement generated the following graph for visualizing the airline passenger data. The plot shows the unobserved trend and seasonality of the data as mentioned previously. X Axis provides the scale in years between 1949-1960 and the Y Axis shows the natural logarithm of airline passenger data. From the data, it is clear that one can observe an upward trend and seasonal cycles for each year of the date range.

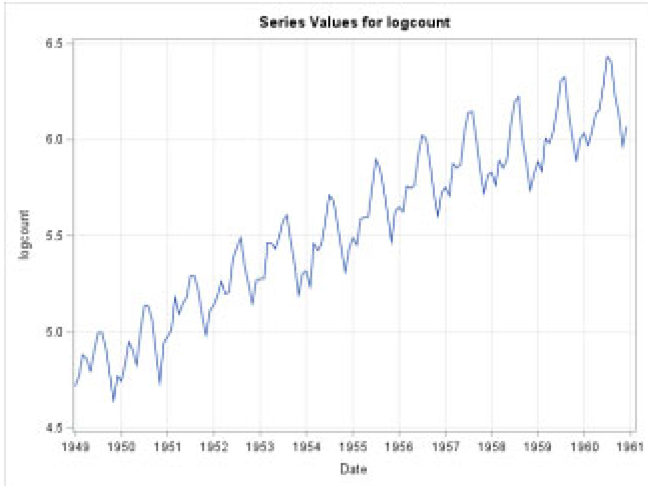


Figure 1. International Airline Passenger Data 1949-60

The highest airline passenger counts were observed during July/August of each year, while the lowest passenger counts were observed during January/November of each year. Per Larsen (2016), this airline data does not have regressors. This makes the model simpler by considering trend and seasonality only.

In all the models in this paper, PROC MCMC procedure statement summons the input data set Series G. Outpost option with Posterior generates posterior samples for all parameters. The posterior intervals of prediction table (ODS table name, PredIntervals or PredInt) contains the equal-tail and Highest Posterior Density (HPD) interval estimate for each prediction. The default α value is 0.05.

ODS output PredSumInt contains statistics including parameter (for the airline example, the passenger count or log passenger count), sample size, Mean, Standard Deviation, HPD lower, and HPD upper. The table includes these statistics for all of data and not just for prediction time frame. For this example, the above statistics are generated for the period 1949-62. The true posterior prediction forecast is for 1961-62 time period.

Bayesian inference relies on use of simulated sample draws to summarize posterior distribution of quantities of interest, per SAS (2015). Markov Chain Convergence. After the PROC MCMC procedure runs, one has to decide whether:

- Markov Chain has reached its stationarity or the needed posterior distribution, and
- Decide the number of iterations needed to keep after the posterior distribution has reached stationarity.

Markov Chain convergence diagnostics help us to address the two issues mentioned above. There are no conclusive tests to indicate whether the Markov chain has reached its stationarity. As practitioners, we need to evaluate the stationarity of all parameters, and not just the convergence of response variable or quantities of interest. Convergence diagnostics can be observed with visual trace plots including diagnostics for trend parameters, and the diagnostics for seasonality parameters as well, for this example. Additionally, effective sample sizes provide tabular output diagnostics that show Effective Sample Size (ESS) for each parameter, Autocorrelation time, and the efficiency. The closer the ESS for each parameter is to the sample size of draw, the mixing is good, and the Markov Chain convergence could be observed visually in the diagnostic plots.

RESULTS

The results for Model 0 and Model 1 are discussed in detail to optimize this discussion.

Model 0 Diagnostics:

Posterior Summaries and Intervals					
Parameter	N	Mean	Standard Deviation	95% HPD Interval	
mu0	5000	4.2978	0.4832	3.4226	5.3191
s2_s	5000	0.7398	0.3772	0.2344	1.4536
s2_mu	5000	0.1210	0.0209	0.0835	0.1628
s2_e	5000	0.1070	0.0173	0.0765	0.1428

Table 1. Model 0 Posterior Summaries and Intervals

Effective Sample Sizes			
Parameter	ESS	Autocorrelation Time	Efficiency
mu0	912.3	5.4805	0.1825
s2_s	2449.7	2.0411	0.4899
s2_mu	1736.7	2.8790	0.3473
s2_e	3103.1	1.6113	0.6206

Table 2. Model 0 ESS Data

Table 2 shows ESS for Model 0. Per SAS (2015), ESS relates to autocorrelation and measures mixing of Markov Chain. If there is a significant difference between ESS and the simulation sample size, then it indicates poor mixing. Figure 2 and first data row of Table 2 shows that for mu0, the mixing is slower than for other parameters and is marginally good, and the auto correlation time is not considerably high. When we observe Second row of Table 2 and Figure 3, the ESS is closer to the model sample size and the autocorrelation time is lesser than those for mu0. Similarly, the parameters s2_mu and s2_e indicates better ESS as they are closer to the sample size N (Table 1 column N) and further evidenced by autocorrelation time for these two parameters.

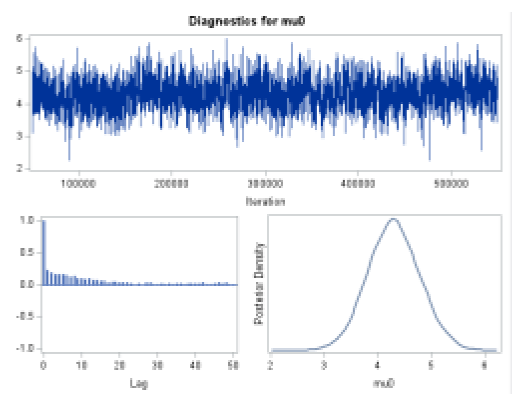


Figure 2: Model 0 mu0 Diagnostics

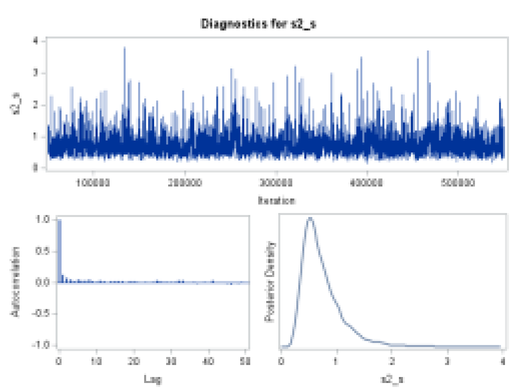
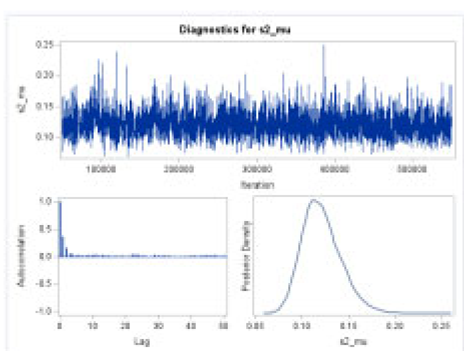


Figure 3: Model 0 s2_s Diagnostics



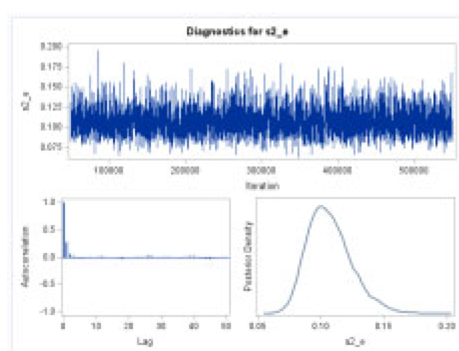


Figure 4: Model 0 s2_mu Diagnostics Figure 5: Model 0 s2_e Diagnostics

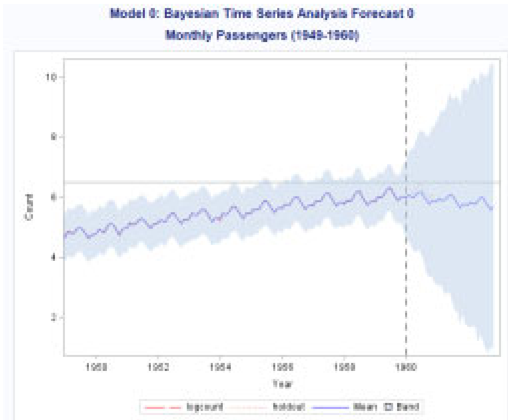


Figure 6: Model 0 Forecast 0

Figure 6 indicates that the band around mean (light blue shaded area indicates the region of HPD interval range around actual log count indicated by red dashed line and the mean values of posterior distribution are indicated by blue line, and the hold out samples are indicated by red dots) of natural log of passenger count posterior distribution during the period 1949-60, while prediction for the forecast years 1961-62 shows a range that is much larger than the previous year data. This indicates that the model could be improved as reliable forecast cannot be inferred from this model. As the model statement and random statements of parameters in the code approaches the true distribution of data, the prediction band in the forecast years 1961-62 improves i.e., the prediction band becomes narrower. Highest Posterior Density (HPD) interval. HPD is an interval in which most of the posterior distribution is contained. The mean values as seen in Table 3 are closer to the actual count for previous years, however the HPD interval range is very wide as shown in Figure 6 and in Table 3.

Date	Description	Mean	HPD Lower	HPD Upper
Jan-1961	Jan Passenger_Count	362.39	23.68	4361.16
Feb-1961	Feb Passenger_Count	351.57	22.03	5180.45
Mar-1961	Mar Passenger_Count	389.98	22.00	6388.57
Apr-1961	Apr Passenger_Count	363.36	20.90	7443.10
May-1961	May Passenger_Count	363.21	15.24	7086.72
Jun-1961	Jun Passenger_Count	402.75	17.17	9444.15
Jul-1961	Jul Passenger_Count	437.39	17.26	13361.20
Aug-1961	Aug Passenger_Count	427.30	12.31	10610.43
Sep-1961	Sep Passenger_Count	378.15	9.46	10908.13
Oct-1961	Oct Passenger_Count	328.35	9.26	11962.24
Nov-1961	Nov Passenger_Count	286.06	5.86	8629.21
Dec-1961	Dec Passenger_Count	317.28	8.64	15278.54

Table 3: HPD Interval for forecast year 1961 (Model 0)

Model 1 Diagnostics:

Posterior Summaries and Intervals					
Parameter	N	Mean	Standard Deviation	95% HPD Interval	
mu0	7500	0.0948	1.4173	-2.6526	2.8656
s2_s	7500	1855.3	977.9	536.9	3717.6
s2_mu	7500	309.4	40.8320	230.1	390.3
s2_e	7500	1.8974	0.6926	0.6592	3.2080

Table 4: Model 1 Posterior Summaries and Intervals

Effective Sample Sizes			
Parameter	ESS	Autocorrelation Time	Efficiency
mu0	7500.0	1.0000	1.0000
s2_s	130.4	57.5224	0.0174
s2_mu	1726.5	4.3442	0.2302
s2_e	881.2	8.5107	0.1175

Table 5: Model 1 ESS Data

For Model 1, the ESS Data table shows that for parameter mu0, ESS is equal to the sample size as indicated in Table 4. However, the ESS for s2_s could be further improved as the ESS for this parameter is 130.4 with an autocorrelation time of 57.5224. Efficiency in the ESS data table (Table 5) is calculated as the ratio of ESS for a given parameter to the sample size. The closer the efficiency to 1, the better the ESS approximation to the sample size, and lesser the autocorrelation time.

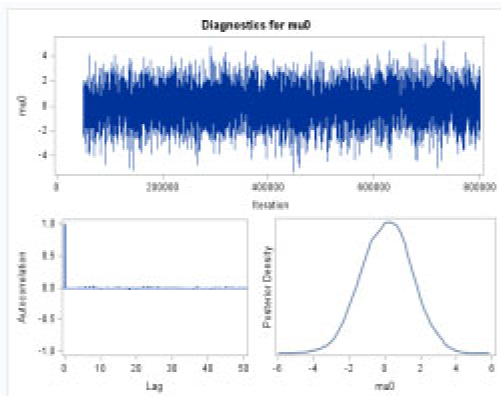


Figure 7: Model 1 mu0 Diagnostics

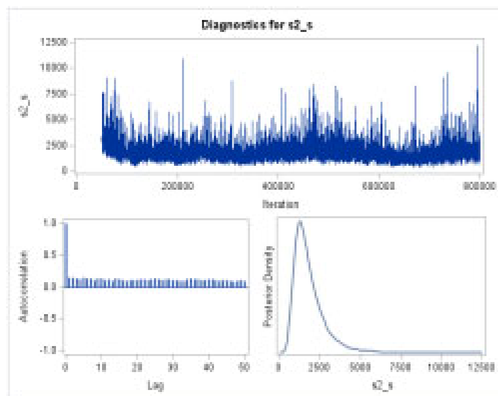


Figure 8: Model 1 s2_s Diagnostics

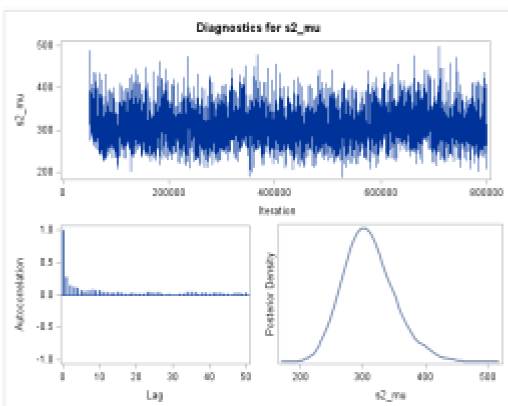


Figure 9: Model 1 s2_mu Diagnostics

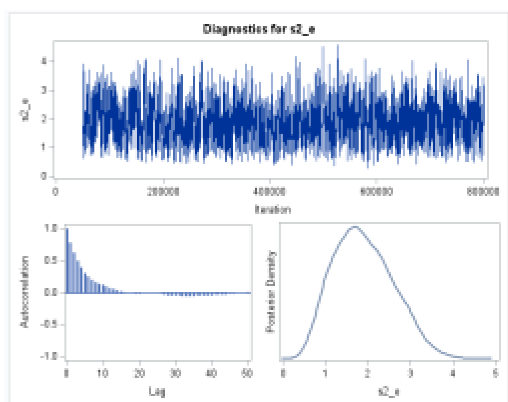


Figure 10: Model 1 s2_e Diagnostics

Similar to the discussion for Model 0, the trace plot for mu0 for Model 1 (Figure 1) shows a "perfect" trace plot, and looks similar to the trace plot identified in SAS (2015) (in the section on Visual Analysis via Trace Plots (Figure 7.1)) It can be observed that the center of the chain appears to be around the value 0, with small fluctuations. This could be construed as the chain might have reached the right distribution. The chain might be mixing well and is exploring the distribution by traversing to areas where its density is low. Figure 8, Figure 9, and Figure 10 show that the Markov Chain for these parameters could be improved and are further evidenced by the ESS, autocorrelation times, and efficiency for these parameters (Table 5).

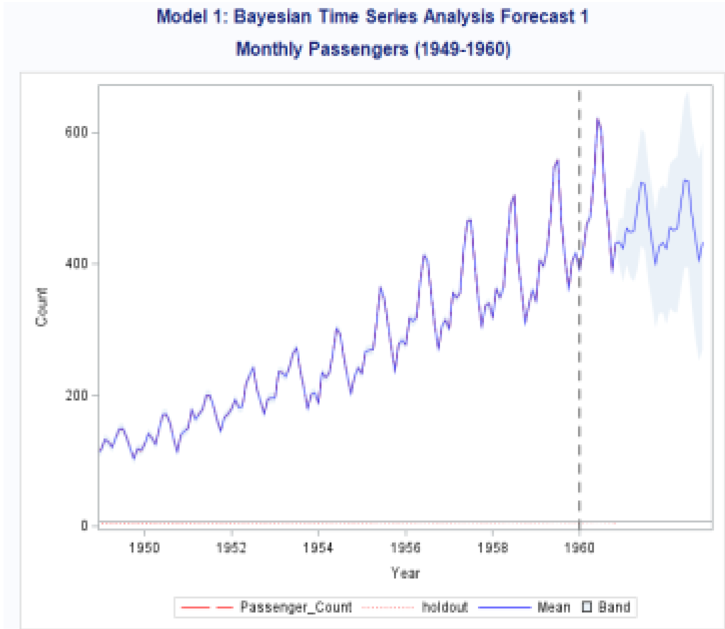


Figure 11: Model 1 Forecast 1

Figure 11 indicates the passenger count, holdout samples (1959-60), and the shaded regions for 1961-62. The mean values of posterior distribution and actual passenger count are very close and the hold out samples for 1959-60 indicate that the mean values of posterior distribution and the airline passenger count are close to the actual passenger count values.

The light blue shaded region around mean values of posterior distribution (forecast 1) of passenger count for 1961-62 appear to follow the trend for the previous years with a slight increase in the peak values as shown in Figure 11.

Month-Year	Parameter	Mean	HPD Lower	HPD Upper
Jan-1961	Jan Passenger Count	433.40	398.19	469.43
Feb-1961	Feb Passenger Count	423.46	372.06	468.75
Mar-1961	Mar Passenger Count	454.14	398.71	516.42
Apr-1961	Apr Passenger Count	448.06	379.98	514.15
May-1961	May Passenger Count	450.01	372.76	530.09
Jun-1961	Jun Passenger Count	487.08	391.76	564.00
Jul-1961	Jul Passenger Count	524.03	428.08	606.54
Aug-1961	Aug Passenger Count	521.45	420.46	600.19
Sep-1961	Sep Passenger Count	471.30	374.35	551.42
Oct-1961	Oct Passenger Count	435.08	340.39	516.66
Nov-1961	Nov Passenger Count	399.98	302.61	480.79
Dec-1961	Dec Passenger Count	426.62	324.46	513.65

Table 6: HPD Interval for forecast year 1961 (Model 1)

Table 6 shows the mean, HPD Lower, and HPD Upper values for forecast year 1961 for Model 1, and Figure 11 visually represents these values. As mentioned in a previous section, the accuracy of the model depends on the accuracy of random variable function approximation and the model density function equation.

HPD Lower and HPD Upper for model 1 are closer in range and the mean values appear to be following the trend from previous year (1960).

CONCLUSION

This section summarizes the paper's findings and the importance of PROC MCMC in light of time series forecasting.

Time Series data offers some of the trickiest analytical challenges. We typically have minimum amount of data to work with, however analytics professionals are expected to provide insights for some of the most important decisions, especially with accurate forecasts. ARIMA has been a popular time series forecasting method. In the last few decades, time series forecasting using Bayesian Structural Model has been gaining popularity.

In the examples provided in this paper, International Airline Data provides passenger data change month over month through 1949-60. This dataset was used to demonstrate the power of PROC MCMC in making accurate forecasts. As the accuracy of defining parameter random variables improve, and the model distribution accurately approximates the actual distribution of input data, the accuracy of forecast improves, especially with PROC MCMC as the parameters can be simultaneously be used in a single block as shown by considering trend and seasonality together in the model equation of airline data example in this paper. So, it is ever more important to consider the author Larsen (2016) statement "Sorry ARIMA, but I'm going Bayesian" in order to improve forecast accuracy. By efficiently accessing lag and lead variables across an index (Month-Year) in the airline data example, as in time series analysis, the likelihood function can depend on lag values. PROC MCMC allows the construction of equations with the use of lag and lead variables across an index in RANDOM statement, as shown in Model 1 and other models in this paper.

The models discussed in this paper could be further improved by fine tuning the parameters, random statements, and model equations further to cater to the time series forecasting needs of organizations.

APPENDIX A

```
*/Model 2: Bayesian Structural Time Series Analysis, Model 2 with
alpha, mu, theta, and phi*/;
PROC MCMC data=seriesG NMC=750000 NBI=75000 seed=123456 thin=100
PROPCOV=QUANEW;
PARMS ALPHA0;
PARMS MU0;
PARMS S0 S1 S2;
PARMS THETA1;
PARMS THETA 2;
PARMS THETA 3;
PARMS THETA 4;
PARMS THETA_PHI;
PARMS PHI;
prior PHI~NORMAL(0,var=exp(THETA_PHI));
prior ALPHA0~normal(0,var=theta2);
prior mu0~normal(0,var=100);
prior s:~normal(0,var=theta3);
prior theta:~igamma(shape = 3/10, scale = 10/3);
```

```
random ALPHA~NORMAL(phi*alpha.l1,var=exp(theta2)) subject=t
icond=(alpha0);
random s~normal(-s.l1-s.l2-s.l3,var=exp(theta3)) subject=Date
icond=(s2 s1 s0);
random mu~normal(mu.l1 + alpha.l1,var=exp(theta1)) subject=t
icond=(mu0);
x=mu + s;
model Passenger_Count~normal(X,VAR=exp(theta4));
preddist outpred=outpred2 statistics=brief;
ODS output PredSumInt=PredSumInt2;
RUN;
*Forecast data;
data forecast2;
  merge seriesG PredSumInt2;
run;
proc format;
     value timefmt 13='1950'
                   37='1952'
                   61='1954'
                   85='1956'
                   109='1958'
                   133='1960';
PROC SGPLOT DATA=forecast2;
  title1 "Model 2: Bayesian Time Series Analysis Forecast 2";
  title2 "Monthly Passengers (1949-1960)";
  format t timefmt.;
  series x=t y=Passenger_Count / LINEATTRS=(color=red
pattern=longdash);
  series x=t y=holdout / LINEATTRS =(color=red pattern=dot);
  series x=t y=mean / LINEATTRS =(color=blue pattern=solid);
  YAXIS label="Count";
  XAXIS values=(13 37 61 85 109 133) ranges=(1-168) label="Year";
  REFLINE 133 / axis=x LINEATTRS=(color=black pattern=dash);
  REFLINE 6.5 / axis=y;
  band x=t upper=HPDUPPER lower=HPDLOWER / transparency=.7;
run; title1; title2;
*/Model 3: Bayesian Structural Time Series Analysis, Model 3 with
alpha, mu, theta, and phi and logcount*/;
PROC MCMC data=seriesG nmc=750000 NBI=75000 seed=123456 thin=100
outpost=posterior3 PROPCOV=quanew;
PARMS alpha0;
PARMS mu0;
PARMS s0 s1 s2;
PARMS theta1;
PARMS theta2;
PARMS theta3;
PARMS theta4;
PARMS THETA_PHI;
PARMS phi;
prior PHI~NORMAL(0,var=exp(THETA_PHI));
prior alpha0~normal(0,var=theta2);
prior mu0~normal(0,var=100);
```

```
prior s:~normal(0,var=theta3);
prior theta:~igamma(shape = 3/10, scale = 10/3);
random ALPHA~NORMAL(phi*alpha.l1,var=exp(theta2)) subject=t
icond=(alpha0);
random s~normal(-s.l1-s.l2-s.l3,var=exp(theta3)) subject=Date
icond=(s2 s1 s0);
random mu~normal(mu.l1 + alpha.l1,var=exp(theta1)) subject=t
icond=(mu0);
x=mu + s;
model logcount~normal(X,VAR=exp(theta4));
preddist outpred=outpred3 statistics=brief;
ODS output PredSumInt=PredSumInt3;
RUN;
*Forecast data;
DATA FORECAST3;
  merge seriesG PredSumInt3;
run;
PROC SGPLOT DATA=forecast3;
  title1 "Model 3: Bayesian Time Series Analysis Forecast 3";
  title2 "Monthly Passengers (1949-1960)";
  format t timefmt.;
  series x=t y=logcount / LINEATTRS =(color=red pattern=longdash);
  series x=t y=holdout / LINEATTRS =(color=red pattern=dot);
  series x=t y=mean / LINEATTRS =(color=blue pattern=solid);
  YAXIS label="Count";
  XAXIS values=(13 37 61 85 109 133) ranges=(1-168) label="Year";
  REFLINE 133 / axis=x LINEATTRS=(color=black pattern=dash);
  REFLINE 6.5 / axis=y;
  band x=t upper=HPDUPPER lower=HPDLOWER / transparency=.5;
RUN; title1; title2;
```

APPENDIX B

The PROC MCMC statement options considered in this paper are provided in the table below for reference with brief explanations. (Reference: SAS (2015))

PROC MCMC Option	Option Description
Data	Names the input dataset
Icond	Icond specifies the initial values of lag or lead variables for the response variable when observation indices are out of range.
Model	Model statement specifies the conditional distribution of data given the parameters (likelihood function).
NBI	Specifies the number of burn-in iterations
NMC	Specifies the number of iterations, excluding the burn-in iterations
Outpost	Names the output dataset for posterior samples of parameters
PARMS	PARMS statements declare parameters in the model.
Prior	Prior statements declare prior distributions of parameters.
PROPCOV	Controls options for constructing initial proposal covariance matrix
RANDOM	Random statements specify random effects and their prior distributions
Seed	Specifies the random seed for simulation
Subject	Subject argument declares group index in the model
Thin	Specifies the thinning rate

Table x. PROC MCMC Option & Option Description Table

REFERENCES

Brownlee, Jason (2016) "What is Time Series Forecasting?" December 2, 2016.

https://machinelearningmastery.com/time-series-forecasting/

Chen, Fang (2009) Bayesian Modeling using the MCMC Procedure

https://support.sas.com/resources/papers/proceedings09/257-2009.pdf

Chib, Siddhartha & Greenberg, Edward (1995) Understanding the Metropolis-Hastings Algorithm

http://web1.sph.emory.edu/users/hwu30/teaching/statcomp/papers/chibGreenbergMH.pdf

Cook, John D (2011) Markov Chains don't converge

https://www.johndcook.com/blog/2011/08/10/markov-chains-dont-converge/

Dickey, David A. (2004) Case Studies in Time Series II: Periodic Behavior and Related Issue
https://support.sas.com/resources/papers/proceedings/proceedings/sugi29/201-29.pdf

Larsen, Kim (2016). "Sorry ARIMA, but I'm going Bayesian". April 21, 2016.
https://multithreaded.stitchfix.com/blog/2016/04/21/forget-arima/

Lütkepohl, Helmut & Xu, Fang (2009) The role of log transformation in forecasting economic variables

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.596.2695&rep=rep1&type=pdf

Ghosh, Himadri (2017) Structural Time Series Models for describing trend in all India sunflower yield using sas http://www.iasri.res.in/sscnars/socialsci/17-Structural%20Time%20Series%20Models%20for%20Describing%20Trend%20in%20All%20India%20Sunflower%20Yield%20Using%20SAS.pdf

Lütkepohl, Helmut & Xu, Fang (2009) The role of log transformation in forecasting economic variables

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.596.2695&rep=rep1&type=pdf

Nisbet, Robert; Elder, John; & Miner, Gary (2009) Statistical Analysis & Data Mining Applications. Burlington, MA: Elsevier

SAS Institute (2014) The ARIMA Procedure. SAS/ETS® 13.2 User's Guide. Cary, NC: SAS Institute Inc. https://support.sas.com/documentation/onlinedoc/ets/132/arima.pdf

SAS Institute (2015) SAS/STAT® 14.1 User's Guide: Introduction to Bayesian Analysis Procedures. Cary, NC: SAS Institute Inc. https://support.sas.com/documentation/onlinedoc/stat/141/introbayes.pdf

SAS Institute (2016) The MCMC Procedure. SAS/STAT® 14.2 User's Guide. Cary, NC: SAS Institute Inc. https://support.sas.com/documentation/onlinedoc/stat/142/mcmc.pdf

SAS Institute (n.d.) Sample 59685: Bayesian Autoregressive and Time-Varying Coefficients Time Series Models http://support.sas.com/kb/59/685.html

Book

Scott, Steven L & Varian, Hal (2013) Predicting the present with Bayesian Structural Time Series. June 28, 2013 https://pdfs.semanticscholar.org/8471/54972523a849907fdbb4e35b1715cc6e3a4a.pdf

Tanner, Martin A. & Wong, Wing Hung (1987) "The calculation of posterior distributions by data augmentation" *Journal of American Statistical Association*, 82: 528-540.

ACKNOWLEDGMENTS

We appreciate SESUG for considering our paper and the motivation for the paper is to demonstrate the power of PROC MCMC in accurately forecasting time series models using Bayesian Structural Time Series.

RECOMMENDED READING

Congdon, P. (2006), Bayesian Statistical Modeling, 2nd ed. New York: John Wiley & Sons

Gelman, A., Carlin, J.B., Stern, H. S., and Rubin, D. B. (2004), Bayesian Data Analysis, 3rd ed. London: Chapman & Hall.

CONTACT INFORMATION

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

Murali Sastry
Local Government Federal Credit Union (LGFCU)
919-747-4408
Murali.Sastry@lgfcu.org

Tyler Aaron Hicks PhD
University of Kansas
785-864-0782
tahicks@ku.edu