The SAS Time Series Forecasting System
An Overview for Public Health Researchers

Charles DiMaggio, PhD
College of Physicians and Surgeons
Departments of Anesthesiology and Epidemiology
Columbia University
New York, NY
cjd11@columbia.edu

March 23, 2011

1 SAS Time Series Tools

Time series analyses can be useful for evaluating health outcomes over time. You might, for example, be interested in determining if a disaster or other event had an effect on the occurrence of some outcome and whether one could expect future occurrences to change in pattern or frequency. Time series analysis can be conducted by invoking SASs PROC ARIMA and related procedures, but most recent versions of SAS come packaged with the convenience of the menu-driven SAS Time Series Forecasting System (TSFS).[1] The following notes present a very brief overview of an approach to times series data using SASs TSFS. [2]

1.1 Starting the TSFS

Typing dm forecast in the editor window invokes the TSFS and leads to a screen where you specify a SAS data set that includes a time variable. If your data set does not appear as a time series, there is likely a problem with the time variable, such as a missing observation. After specifying the file, click the *Develop Models* button to access most all tools. A *Left Click* of your mouse on a white area will bring up an interactive menu, including an option to fit models automatically which is useful to establish a benchmark against which to measure your own efforts. Before fitting a model, you should view the raw time series with an smoothed overlaid line either in Base SAS or in R.

---

[1] Computing convenience tends to be subjective. While I usually insist on syntax-based analyses, 'consistency is the hobgoblin of little minds', and SAS has done a nice job here. And, all the point-and-click stuff doesn't absolve folks from documenting their work so others can reproduce it.

[2] This is most decidedly *not* not an introduction to time series analysis. I am assuming you already have some basic grounding in this interesting and important statistical topic.

2 Descriptive and Exploratory Analysis

You will need to account for trend and seasonality to fit an appropriate model. Look for trend and seasonality first by plotting the data. A smoothed time plot may be more instructive than some of the diagnostic tools. Use autocorrelation plots and statistics such as the Ljung-Box test for white noise and the Dickey-Fuller unit root test for stationarity to help you diagnose trend and seasonality.

A trend may be evident in a plot of in an autocorrelation function (ACF), partial autocorrelation function (PACF), or inverse autocorrelation function (IACF) with a significant lag at 1. The ACF will decay slowly from lag 1 and will have few significant values after first differencing. A non-significant unit-root test will become significant after differencing.

Seasonality presents as repetitive behavior on a plot at time S. There will be a significant ACF, PAC and/or IACF at lag S. There may be continued significant ACF values at lags that are multiples of S. The ACF will become non-significant after seasonal differencing. The seasonal unit root test becomes significant after seasonal differencing. [3]

3 Fitting a Simple Model

Left click the *Develop Models* button. Review the plot, autocorrelations, and statistical test by clicking on them. Close out of the viewer. Left click a blank space of the Develop Models window. Start with *Fit Model Automatically* to get a benchmark against which to work.

3.1 Linear Trend

To fit a linear trend, left click the blank area again, choose *Fit Custom Model*. Click the down arrow next to *Trend Model* and select *Linear Trend* Click *OK* to fit the model.

3.2 Non-Linear Trend

To specify a model with, say a quadratic trend, proceed as above, except select *Trend Curve* rather than *Linear Trend* Other trend types are available and their choice will be driven by the form of any apparent trend in your data. Fit a number of models to compare.

---

[3]Note that a while a stochastic trend will generally convert to to white noise with simple differencing, a linear trend will not.

3.3 Seasonality

To model seasonality, the two basic choices are seasonal dummy variables or differencing. In a simple model, you will likely choose a seasonal dummy or linear trend with a seasonal term. Select *Develop Models*, select *Fit Models from List* then select *Seasonal Dummy* or *Linear Trend with Seasonal Terms*

3.4 Interrupted Time Series

Interrupted time series analyses using transfer functions to model events such as disasters or other impacts are appealing, but conceptually complex and require knowledge of the underlying process and statistical theory. Interpret them cautiously. To analyze an interrupt term in SASs TSFS select *Fit Custom Model*, then *Add*, then *Interventions*. Specify a date and whether the event has a point, step or ramp, specify decay pattern. The choice is informed by your knowledge of the event and its expected effect as well as how it appears on the time series plot. Choices include an *Abrupt, Temporary Effect* (a point transfer function with exponential decay), an *Abrupt, Permanent Effect* (a step function with no decay) and a *Gradual, Permanent Effect* (a step function with exponential decay thatdecays up).

When reviewing the parameter estimates associated with the event, look at the intercept estimate which reflects the level or average per month prior to the event, the point, step or ramp estimate which is the increase or decrease in the outcome due to event and at all the associated p values to see if any of the coefficients were statistically significant.

3.5 Forecasting

To obtain forecast values, first select your best-fitting model based on indices such as Root Mean Square Error (RMSE), Mean Absolute Percent Error (MAPE) or Akaike Information Criteria (AIC). Then close out of the developer window and click on *Produce Forecast → Run → Output* [4]

4 Box-Jenkins Methodology

Box-Jenkins represents a powerful methodology that addresses trend and seasonality well. ARIMA models have a strong theoretical foundation and can closely approximate any stationary process. The process consists of model identification by using autocorrelation functions, evaluation by assessing the fit of the possible models and forecasting using the best model. [5] See the following section for how to proceed with the usual situation of non-stationary time series.

---

[4]Was previously grayed out, should now be active

[5]Caution, these methods are applied to stationary time series.

4.1 Fitting an ARIMA Model

The steps in ARMA(p,q) identification are: (1) Ensure a stationary series by accounting for trend and seasonality (2) Determine the highest possible value of q from the highest significant ACF lag (3) Determine the highest possible value of p from highest significant PACF/IACF lag (4) Determine all ordered pairs (j, k) such that $0 \leq j \leq p$ and $0 \leq k \leq q$ (5) Fit an ARMA model for each ordered pair, and (6) Select the model with smallest value of RMSE on a holdout sample or AIC on a fit sample

In the SAS TSFS left click in an empty section of the *Develop Models* window and choose *Fit ARIMA Model*Specfiy the p and q values. To specify a holdout sample, click the *Set Ranges* button on the *Develop Models* window. Type in the number of periods next to *Hold-Out Sample*.

4.2 Model Evaluation

To evaluate a model, first click the *View Selected Models Graphically* button found below the *Browse* button. You will see a plot of predicted values and original data. Select *Prediction Errors* for a plot of residuals. Take a look at *Prediction Errors Autocorrelation*, too . A good model will not have any spikes significantly different from zero. Under *Prediction Error* you will want to see white noise and significant unit root tests. In the *Parameter Estimates* window, the intercept represent the mean (μ) of the series. Check if any of the parameters are significantly different from zero. Finally, review the *Statistics of Fit* (for a number of different fit statistics), *Forecast Graph* (to plot the forecast and its 95% CI) and *Forecast Table* (to get same information in tabular fashion).

Once you have developed and assessed a model, you can use it to produce forecasts and save the forecasts to a data set. Close out of the *Develop Models* window. Select the *Produce Forecasts* button on the Intro window. The default is to forecast all series in a data set and save the forecast SAS data set as work.forecast. Accept this simple format to forecast 12 time periods into the future. Click on *Select* to forecast only a subset of the series. Format will allow you to change the format from simple to interleaved or concatenated. Horizon will allow you to change how many time periods into the future you want to forecast. Run starts the process. The Output button allows you to view the table.

5 Modeling Non-Stationary Time Series

It is rare to encounter purely stationary time series. Before applying the Box-Jenkins methods above,[6] you will need to (1) diagnose trend and seasonality (2) determine appropriate trend/seasonal components to use for modeling (3) check residuals after

---

[6]OK. Maybe I should switch these sections around

applying trend/seasonal components (4) verify that residuals appear stationary only then (5) apply methods from the previous section.

5.1 Differencing

Box-Jenkins emphasizes differencing to achieve stationarity. You will rarely require greater than 2nd order differencing for non-seasonal trends and 1st order differencing for seasonal trends. The TSFS is designed to make differencing as easy as possible. Begin by looking at both the ACF and PACF as outlined above. Click the differencing buttons. Do the ACF and PACF graphs look better (smaller bars)? If there is no trend or seasonality in the raw data, first differencing will create a trend so you will see bigger bars. If you believe there is a quadratic trend, a 2nd order differencing may be necessary. If first differencing does not work, you can apply linear trend modeling through the custom model window as described above.

The mechanics of seasonal differencing for a ARIMA(p,d,q)(P,D,Q) model proceeds similarly to trend differencing, where P is seasonal AR term, D is the seasonal differencing term and Q is the seasonal MA term.

When you are specifying an ARIMA model in SASs TSFS, specify the model both with and without an intercept term because when TSFS applies differencing the default is to omit the intercept term. You can duplicate and edit a model by left clicking the white are in the *Develop Model* window. This allows you to do things like adding a linear or quadratic trend to an ARIMA model. You can compare models head to head by going to the *Tools* menu, then selecting the *Compare Models* feature. The objective is to select the best model using a robust list of accuracy criteria.

6 Adding Interrupt or Event Terms to ARIMA models

Interventions can be added through the *Custom Model Specification* window or the *ARIMA Model Specification* window if seasonal AR or MA terms are required. The ARIMA window has the most general model specification tools. Click *Add → Interventions* . Proceed as previously described, by specifying the *Date of the onset of the event term*, the *Type* and the *Effect Decay Pattern* (where Exp indicates an order 1 decay and Wave describes an order

7 Conclusion

Time series analysis is a powerful and potentially important tool for public health researchers. SAS's Time Series Forecasting System offers a convenient and often readily accessible tool to apply time series analytic concepts to public health data.

8 Annotated Bibliography

1. **Peter J. Diggle. Time Series: A Biostatistical Introduction. Oxford. 2004.** I first came across Dr. Diggle's work when I started reading into spatial analysis. There are many parallels between time series and spatial data, particularly the issue of non-independence of nearby observations. Diggle is a master of both, and this is a masterful introduction.

2. **Chris Chatfield. The Analysis of Time Series: An Introduction.** CRC Press. 2004. An early classic known for it's broad accessibility. Any book that begins with a passage from Alice in Wonderland deserves recommending.

3. **John C. Brocklebank and David A Dickey. SAS for Forecasting Time Series. SAS Press.2003** The book you want to spend some *time* with (pun intended) if you want to use syntax to conduct your SAS time series.

4. **Terry Woodfield and Bob Lucas. Business Forecasting Using SAS: A Point-and-Click Approach SAS Press. 2006.** This is the course material for SAS's training workshop in time series analysis. Most of the notes in this document are based on this book. I was fortunate enough to take the course with Dr. Woodfield. He almost made we want to give up epidemiology and enter the business world. I continue to opt out of a higher tax bracket.