Paper SD-05

Case Studies in Time Series

D. A. Dickey

Introduction

Data taken over time often exhibit autocorrelation, This is a phenomenon in which positive deviations from a mean are followed by positive and negative by negative or, in the case of negative autocorrelation, positive deviations tend to be followed by negative ones more often that would happen with independent data.

While the analysis of autocorrelated data may not be included in every statistics training program, it is certainly becoming more popular and with the development of software to implement the models, we are likely to see increasing need to understand how to model and forecast such data.

A classic set of models known as ARIMA models can be easily fit to data using the SAS@ procedure PROC ARIMA. In this kind of model, the observations, in deviations from an overall mean, are expressed in terms of an uncorrelated random sequence called white noise.

This paper presents an overview of and introduction to some of the standard time series modeling and forecasting techniques as implemented in SAS with PROC ARIMA and PROC AUTOREG, among others. Examples are presented to illustrate the concepts. In addition to a few initial ARIMA examples, more sophisticated modeling tools will be addressed. Included will be regression models with time series errors, intervention models, a discussion of nonstationarity, and transfer function models.

White noise

The fundamental building block of time series models is a white noise series e(t). This symbol e(t) represents an unanticipated incoming "shock" to the system. The assumption is that the e(t) sequence is an uncorrelated sequence of random variables with constant variance. A simple and yet often reasonable model for observed data is

$$Y(t)-\mu=\rho(Y(t-1)-\mu)+e(t)$$

where it is assumed that ρ is less than 1 in magnitude. In this model an observation at time t, Y(t), deviates from the mean μ in a way that relates to the previous, time t-1, deviation and the incoming white noise shock e(t). As a result of this relationship, the correlation between Y(t) and Y(t-j) is ρ raised to the power |j|, that is, the correlation is an exponentially decaying function of the lag j.

The goal of time series modeling is to capture, with the model parameter estimates, the correlation structure. In the model above, known as an autoregressive order 1 model, the current Y is related to its immediate predecessor in a way reminiscent of a regression model. In fact one way to model this kind of data is to simply regress Y(t) on Y(t-1). This is a type of "conditional least squares" estimation. Once this is done, the residuals from the regression should mimic the behavior of the true errors e(t), that is, the residuals should appear to be an uncorrelated sequence, that is, white noise.

The term "lag" is used often in time series analysis. To understand that term, consider a column of numbers starting with Y(2) and ending with Y(n) where n is the number of observations available. A corresponding column beginning with Y(1) and ending with Y(n-1) would constitute the "lag 1" values of that first column. Similarly lag 2, 3, 4 columns could be created.

To see if a column of residuals, r(t), is a white noise sequence, one might compute the correlations between r(t) and various lag values r(t-j) for j=1,2,…,k. If there is no true autocorrelation, these k estimated autocorrelations will be approximately normal with mean 0 and variance 1/n., Taking n times the sum of their squares will produce a statistic Q having a Chi-square distribution in large samples. A slight modification of this formula is used in PROC ARIMA as a test for white noise. Initially, the test is performed on residuals that are just deviations from the sample mean. If the white noise null hypothesis is not rejected, the analyst goes on to model the series and for each model, another Q is calculated on the model residuals. A good model should produce white noise residuals so Q tests the null hypothesis that the model currently under consideration is adequate.

Autoregressive models

The model presented above is termed "autoregressive" as it appears to be a regression of Y(t) on its own past values. It is of order 1 since only 1 previous Y is used to model Y(t). If additional lags are required, it would be called autoregressive of order p where p is the number of lags in the model. An autoregressive model of order 2, AR(2) would be written, for example, as

$$Y(t)-\mu=\alpha_1(Y(t-1)-\mu)+\alpha_2(Y(t-2)-\mu)+e(t)$$

or as

$$(1-\alpha_1 B-\alpha_2 B^2)(Y(t)-\mu)=e(t)$$

where B represents a "backshift operator" that shifts the time index back by 1 (from Y(t) to Y(t-1)) and thus B squared or B times B would shift t back to t-2.

Recall that with the order 1 autoregressive model, there was a single coefficient, ρ, and yet an infinite number of nonzero autocorrelations, that is, ρ^j is not 0 for any j. For higher order autoregressive models, again there are a finite number, 2 in the example immediately above, of coefficients and yet an infinite number of nonzero autocorrelations. Furthermore the relationship between the autocorrelations and the coefficients is not at all as simple as the exponential decay that we saw for the AR(1) model. A plot of the lag j autocorrelation against the lag number j is called the autocorrelation function or ACF. Clearly, inspection of the ACF will not show how many coefficients are required to adequately model the data.

A function that will identify the number of lags in a pure autoregression is the partial autocorrelation or PACF. Imagine regressing Y(t) on Y(t-1),…,Y(t-k) and recording the lag k coefficient. Call this coefficient $\pi(k)$. In ARIMA modeling in general, and PROC ARIMA in particular, the "regression" is done using correlations between the various lags of Y. In particular, where the matrix **X'X** would usually appear in the regression normal equations, substitute a matrix whose ij entry is the autocorrelation at lag |i-j| and for the usual **X'Y**, substitute a vector with j^{th} entry equal to the lag j autocorrelation. Once the lag number k has passed the number of needed lags p in the model, you would expect $\pi(k)$ to be 0. A standard error of $1/\sqrt{n}$ is appropriate in large samples for an estimated $\pi(k)$.

Moving Average Models

We are almost ready to talk about ARIMA modeling using SAS, but need a few more models in our arsenal. A moving average model again expresses Y(t) in terms of past values, but this time it is past values of the incoming shocks e(t). The general moving average model of order q is written as

$$Y(t) = \mu + e(t) - \theta_1 e(t-1) - ... - \theta_q e(t-q)$$

or in backshift notation as

$$Y(t) = \mu + (1 - \theta_1 B - ... - \theta_q B^q) e(t)$$

Now if Y(t) is lagged back more than q periods, say Y(t-j) with j>q, the e(t) values entering Y(t) and those entering Y(t-j) will not share any common subscripts. The implication is that the autocorrelation function of a moving average of order q will drop to 0, and the corresponding estimated autocorrelations will be estimates of 0, when the lag j exceeds q.

So far, we have seen that inspection of the partial autocorrelation function, PACF, will allow us to identify the appropriate number of model lags p in a purely autoregressive model while inspection of the autocorrelation function, the IACF, will allow us to identify the appropriate number of lags q in a moving average model. The following SAS code will produce the ACF, PACF, and the Q test for white noise for a variable called SALES

```
PROC ARIMA DATA=STORES;
IDENTIFY VAR=SALES NLAG=10;
```

The ACF etc. would be displayed for 10 lags here rather than the default 24.

Mixed (ARMA) Models

The model

$$(1 - \alpha_1 B - ... - \alpha_p B^p)(Y(t) - \mu) =$$

$$(1 - \theta_1 B - ... - \theta_q B^q) e(t)$$

is referred to as an ARMA(p,q), that is, an autoregressive model of orders p (autoregressive lags) and q (moving average lags). Unfortunately, neither simple function ACF or PACF drops to 0 in a way that is useful for identifying both p and q in an ARMA model. Specifically, the ACF and PACF are persistently nonzero.

There are some more complex functions, the Extended Sample Autocorrelation Function ESACF and the SCAN table due to Tsay and Tiao (1984, 1985) that can give some preliminary ideas about what p and q might be. Each of these consists of a table with q listed across the top and p down the side, where the practitioner looks for a pattern of insignificant SCAN or ESACF values in the table. Different results can be obtained depending on the number of user specified rows and columns in the table being searched. In addition, a method called MINIC is available in which every possible series in the aforementioned table is fit to the data and an information criterion computed. The fitting is based on an initial autoregressive approximation and thus avoids some of the nonidentifiability problems normally associated with fitting large numbers of autoregressive and moving average parameters, but still some (p,q) combinations often show failure to converge. More information on these is available in Brocklebank and Dickey (2003). Based on the more complex diagnostics PROC ARIMA will suggest models that are acceptable with regard to the SCAN and/or ESACF tables.

Example 1: River Flows

To illustrate the above ideas, us the log transformed flow rates of the Neuse river at Goldsboro North Carolina

```
PROC ARIMA;
IDENTIFY VAR=LGOLD NLAG=10 MINIC SCAN;
```

(a) The ACF, PACF, and IACF are nonzero for several lags indicating a mixed (ARMA) model. For example here is a portion of the autocorrelation and partial autocorrelation output. The plots have been modified for presentation here:

```
          Autocorrelations
Lag Correlation 0 1 2 3 4 5 6 7 8 9 1
 0 1.000 |     |********************|
 1 0.973 |     |******************* |
 2 0.927 |     |******************* |
 3 0.873 |     |*****************   |
 4 0.819 |     |****************    |
 5 0.772 |     |***************     |
 6 0.730 |     |***************     |
 7 0.696 |     |**************      |
 8 0.668 |     |*************       |
 9 0.645 |     |*************       |
10 0.624 |     |************        |
    "." marks two standard errors

       Partial Autocorrelations
Lag  Corr.  -3 2 1 0 1 2 3 4 5 6 7 8 9 1

 1  0.97 |       . |******************* |
 2 -0.36 |  *******| .                  |
 3 -0.09 |       **| .                  |
 4  0.06 |       . |*.                  |
 5  0.09 |       . |**                  |
 6 -0.00 |       . | .                  |
 7  0.06 |       . |*.                  |
 8  0.02 |       . | .                  |
 9  0.03 |       . |*.                  |
10 -0.02 |       . | .                  |
```

The partial autocorrelations become close to 0 after about 3 lags while the autocorrelations remain strong through all the lags shown. Notice the dots indicating two standard error bands. For the ACF, these use Bartlett's formula and for the PACF (and for the IACF, not shown) the standard error is approximated by the reciprocal of the square root of the sample size. While not a dead giveaway as to the nature of an appropriate model, these do suggest some possible models and seem to indicate that some fairly small number of autoregressive and/or moving average components might fit the data well.

(b) The MINIC chooses p=5, that is, an autoregressive model of order 5 as shown:

```
      Error series model:  AR(6)
 Minimum Table Value: BIC(5,0) = -3.23809
```

The BIC(p,q) uses a Bayesian information criterion to select the number of autoregressive (p) and moving average (q) lags appropriate for the data, based on an initial long autoregressive approximation (in this case a lag 6 autoregression). It is also possible that the data require differencing, or that there is some sort of trend that cannot be accounted for by an ARMA model. In other words, just because the data are taken over time is no guarantee that an ARMA model can successfully capture the structure of the data.

(c) The SCAN table's complex diagnostics are summarized in this table:

```
        ARMA(p+d,q) Tentative
        Order Selection Tests

        ---------SCAN--------
        p+d     q          BIC

          3     1    -3.20803
          2     3    -3.20022
          5     0    -3.23809
       (5% Significance Level)
```

Notice that the BIC information criteria are included and that p+d rather than p is indicated in the autoregressive part of the table. The d refers to differencing. For example, financial reports on the evening news include the level of the Dow Jones Industrials as well as the change, or first difference, of this series. If the changes have an ARMA(p,q) representation the we say the levels have an ARIMA(p,d,q) where d=1 represents a single difference, or first difference, of the data. The BIC for p+d=5 and q=0 is smallest (best) where we take the mathematical definition of smallest, that is, farthest to the left on the number line.

Competitive with the optimal BIC model would be one with p+d=3 and q=1. This might mean, for example, that the first differences satisfy an ARMA(2,1) model. This topic of differencing plays a major role in time series analysis and is discussed next.

Differencing and unit roots

In the full class of ARIMA(p,d,q) models, from which PROC ARIMA derives its name, the I stands for "integrated." The idea is that one might have a model whose terms are the partial sums, up to time t, of some ARMA model. For example, take the moving average model of order 1, Wtj) = e(j) - .8 e(j-1).

Now if Y(t) is the partial sum up to time t of W(j) it is then easy to see that Y(1) = W(1) = e(1)-.8 e(0), Y(2) = W(1) + W(2) = e(2) + .2 e(1) - .8 e(0),
And in general Y(t) = W(1) + W(2) + … + W(t) = e(t) + .2[e(t-1) + e(t-2) + ….+e(1)] - .8 e(0). Thus Y consists of accumulated past shocks, that is, shocks to the system have a permanent effect. Note also that the variance of Y(t) increases without bound as time passes. A series in which the variance and mean are constant and the covariance between Y(t) and Y(s) is a function only of the time difference |t-s| is called a stationary series. Clearly these integrated series are nonstationary.

Another way to approach this issue is through the model. Again let's consider an autoregressive order 1 model

$$Y(t) - \mu = \rho(Y(t-1) - \mu) + e(t)$$

Now if $|\rho| < 1$ then one can substitute the time t-1 value

$$Y(t-1) - \mu = \rho(Y(t-2) - \mu) + e(t-1)$$

to get

$$Y(t) - \mu = e(t) + \rho e(t-1) + \rho^2 (Y(t-1) - \mu)$$

and with further back substitution arrive at

$$Y(t) - \mu = e(t) + \sum_{i=1}^{\infty} \rho^i \; e(t-i)$$

which is a convergent expression satisfying the stationarity conditions. However, if $|\rho| = 1$ the infinite sum does not converge so one requires a starting value, say $Y(0) = \mu$, in which case $Y(t) = \mu + e(t) + \sum_{i=1}^{t-1} e(t-i)$ that is, Y is just the initial value plus an unweighted sum of e's or "shocks" as economists tend to call them. These shocks then are permanent.

Notice that the variance of Y(t) is t times the variance of the shocks, that is, the variance of Y grows without bound over time. As a result there is no tendency of the series to return to any particular value and the use of μ as a symbol for the starting value is perhaps a poor choice of symbols in that this does not really represent a "mean" of any sort.

The series just discussed is called a random walk. Many stock series are believed to follow this or some closely related model. The failure of forecasts to return to a mean in such a model implies that the best forecast of the future is just today's value and hence the strategy of buying low and selling high is pretty much eliminated in such a model since we do not really know if we are high or low when there is no mean to judge against. Going one step further with this, the model

$$Y(t) - \mu = 1.2(Y(t-1) - \mu) - 0.2(Y(t-2) - \mu) + e(t)$$

can be rewritten in terms of

$\nabla Y(t) = Y(t) - \mu - (Y(t-1) - \mu)$ as

$\nabla Y(t) = 0.2 \nabla Y(t-1) + e(t)$.

Notice that the term μ that usually denotes a mean drops out of the model, again indicating that there is no tendency to return to any particular "mean." On the other hand if one tries to write the model

$$Y(t)-\mu=1.2(Y(t-1)-\mu)-0.32(Y(t-2)-\mu)+e(t)$$

in terms of $\nabla Y(t)$ the closet they can get is

$$\nabla Y(t)=-0.12[Y(t-1)-\mu]+0.32\nabla Y(t-1)+e(t)$$

In general, if one writes the model in terms of $\nabla Y(t)$, its lags, and $[Y(t-1)-\mu]$ then the forecasts will not be attracted toward the mean if the lag Y coefficient is 0.

A test for this kind of nonstationarity (H0) versus the stationary alternative was given by Dickey and Fuller (1979, 1981) for general autoregressive models and extended to the case of ARIMA models by Said and Dickey (1984) . The test simply regresses $\nabla Y(t)$ on the lag level $Y(t-1)$, an intercept (or possibly a trend) and enough lagged values of $\nabla Y(t)$, referred to as "augmenting lags," to make the error series appear uncorrelated. The test has come to be known as ADF for "augmented Dickey-Fuller" test.

While running a regression is not much of a contribution to statistical theory, the particular regression described above does not satisfy the usual regression assumptions in that the regressors or independent variables are not fixed, but in fact are lagged values of the dependent variables so even in extremely large samples, the distribution of the "t test" for the lag Y coefficient has a distribution that is quite distinct from the t or the normal distribution. Thus the contribution of researchers on this topic has been the tabulation of distributions for such test statistics. Of course the beauty of the kind of output that SAS ETS procedures provide is that p-values for the tests are provided. Thus the user needs only to remember that t p-value less than 0.05 causes rejection of the null hypothesis (nonstationarity) .

Applying this test to our river flow data is easy. Here is the code

```
        PROC ARIMA;
        IDENTIFY VAR=LGOLD NLAG=10
                STATIONARITY=(ADF=3,4,5);
```

Recall that sufficient augmenting terms are required to validate this test. Based on our MINIC values and the PACF function, it would appear that an autoregressive order 5 or 6 model is easily sufficient to approximate the model so adding 4 or 5 lagged differences should be plenty. Results modified for space appear below:

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Tau	Pr < Tau
Zero Mean	3	0.11	0.7184
	4	0.20	0.7433
	5	0.38	0.7932
Single Mean	3	-3.07	0.0299
	4	-2.69	0.0768
	5	-2.84	0.0548
Trend	3	-3.08	0.1132
	4	-2.70	0.2378
	5	-2.82	0.1892

These tests are called "unit root tests." The reason for this is that any autoregressive model, like

$$Y(t)=\alpha_1 Y(t-1)+\alpha_2 Y(t-2)+e(t)$$

has associated with it a "characteristic equation" whose roots determine the stationarity or nonstationarity of the series. The characteristic equations is computed from the autoregressive coeffients, for example

$m^2-\alpha_1 m-\alpha_2=0$ which becomes

$m^2-1.2m+0.32=(m-.8)(m-.4)=0$,

$m^2-1.2m+0.2=(m-1)(m-.2)=0$

in our two examples. The first example has roots less than 1 in magnitude while the second has a unit root, m=1. Note that

$$Y(t)-Y(t-1)=-(1-\alpha_1-\alpha_2)Y(t-1)$$
$$+\alpha_2 Y(t-1)-Y(t-2)+e(t)$$

so that the negative of the coefficient of Y(t-1) in the ADF regression is the characteristic polynomial evaluated at m=1 and will be zero in the presence of a unit root. Hence the name unit root test.

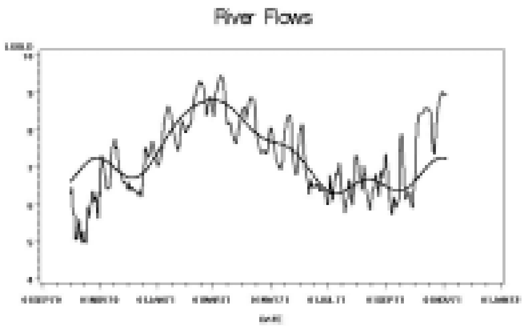
Tests labeled "Zero Mean" run the regression with no interceot and should not be used unless the original data in levels can be assumed to have a zero mean, clearly unacceptable for our flows. The Trend cases add a linear trend term (i.e. time) into the regression and this changes the distribution of the test statistic. These trend tests should be used if one wishes the alternative hypothesis to specify stationary residuals around a linear trend. Using this test when the series really has just a mean results in low power and indeed here we do not reject the nonstationarity hypothesis with these Trend tests. The evidence is borderline using the single mean tests, which would seem the most appropriate in this case.

This is interesting in that the flows are measured daily so that seasonal patterns would appear a long terms oscillations that the test would interpret as failure to revert to the mean (in a reasonably short amount of time) whereas in the long run, a model for flows that is like a random walk would seem unreasonable and could forecast unacceptable flow values in the long run. Perhaps accounting for the seasonality with a few sinusoids would shed some light on the situation.

Using PROC REG, one can regress on the sine and cosine of $2\pi k/365.25$ for a few k values to take out the periodic pattern. Creating these sine and cosine variables as S1, C1, S2, etc. we have

```
        PROC REG;
        MODEL LGOLD=S1 C1 S2 C2
                S3 C3 S4 C4 S5 C5;
```

and a graph of the river flows and the predicting function are shown below



Fuller, Hasza, and Goebel (1981) show that, for large samples, a function bounded between two constants, such as a sinusoid or seasonal dummy variables, can be fitted and the error series tested using the ADF test percentiles that go with a single mean. The test is, of course, approximate and we would feel more comfortable if the yearly wave had gone through several cycles rather than slightly more than 1, but as a rough approximation, the ADF on the residuals is shown below in modified output.

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Tau	Pr < Tau
Zero Mean	2	-4.97	<.0001
	3	-4.49	<.0001
	4	-3.95	<.0001
Single Mean	2	-4.96	<.0001
	3	-4.48	0.0003
	4	-3.94	0.0021
Trend	2	-5.98	<.0001
	3	-5.49	<.0001
	4	-4.94	0.0003

Using sinusoidal inputs to capture seasonality has resulted in rather strong rejections of nonstationarity on the residuals. The series appears to be well modeled with the sinusoids and a stationary error model, even accounting for the approximate nature of the tests.

We next see how to correctly incorporate inputs into a time series model.

Models with deterministic inputs

Our sines and cosines are deterministic in that we know exactly what their values will be out into the infinite future. In contrast, a series like rainfall or unemployment, used as a predictor of some other series, suffers from the problem that it is stochastic, that is, future values of this predictor are not known and must themselves be forecast.

How do we know how many sines and cosines to include in our regression above? One possibility is to overfit then use test statistics to remove unneeded sinusoids. However, we are not on solid statistical ground if we do this with ordinary regression because it appears even in the graph that the residuals, while stationary, are not uncorrelated. Note how the data go above the predicting function and stay for a while with similar excursions below the predicting function. Ordinary regression assumes uncorrelated residuals and our data violate that assumption.

It is possible to correct for autocorrelation. PROC AUTOREG, for example, fits an ordinary least squares model, analyzes the residuals for autocorrelation automatically, fits an autoregressive model to them with unnecessary lags eliminated statistically, and then, based on this error model, refits the regression using generalized least squares, and extension of regression to correlated error structures.

```
PROC AUTOREG;
MODEL LGOLD=S1 C1 S2 C2
    S3 C3 S4 C4 S5 C5 / NLAG=5 BACKSTEP;
```

for example fits our sinusoidal trend with up to 5 autoregressive lags.
Testing these lags, PROC AUTOREG ejects lags 3 through 5 from the model, giving this output

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-1.309659	0.046866	-27.94
2	0.387291	0.046866	8.26

so that the error term Z(t) satisfies the autoregressive order 2 model $Z(t) = 1.31\, Z(t-1) – 0.39\, Z(t-2) + e(t)$. This is stationary, but somewhat close to a unit root model so there could be a rather strong effect on the standard errors, that is, sinusoids that appeared to be important when the errors were incorrectly assumed to be uncorrelated may not seem so important in this correctly specified model. In fact all of the model terms except S1 and C1 can be omitted (one at a time) based on significance of the autoreg-corrected t statistics, resulting in a slight change in the autoregressive coefficients.

The model for Y(t) = LGOLD at time t becomes

$$Y(t) = 7.36 + .78\, S1 + .60\, C1 + Z(t)$$

where

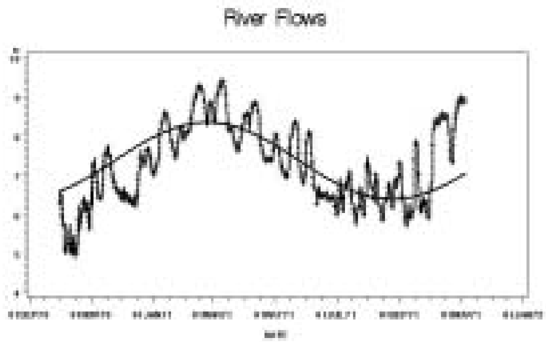
$$Z(t) = 1.31\, Z(t-1) - .37\, Z(t-2) + e(t)$$

Another interesting piece of output is this

Regress	R-Square	0.0497
Total	R-Square	0.9636

What this means is that the predictors in the model (S1 C1) explain only 4.97 percent of the variation in Y(t) whereas if they are combined with lagged residuals, that is if you use the previous 2 days' residuals to help predict today's log flow rate, you can explain 96% of the variability. Of course as you predict farther into the future, the correlation with the most recent residuals dies off and the proportion of variation in future log flows that can be predicted by the sinusoids and residuals from the last couple of days drops toward 4.97%

A plot of the predicted values using the sine wave only and using the sine wave plus 2 previous errors, are overlaid as lines along with the data plotted as plus signs. Note how the plus signs (data) are almost exactly predicted by the model that includes previous residuals while the sine wave only captures the long term seasonal pattern.



The same kind of modeling can be done in PROC ARIMA, however there you have nice syntax for handling differencing if needed, you are not restricted to purely autoregressive models for the errors, and in the case of stochastic inputs PROC ARIMA has the ability to incorporate the prediction errors in those inputs into the uncertainty (standard errors) for the forecasts of the target variable. Here is the syntax that fits the same model as PROC AUTOREG chose.

```
PROC ARIMA;
IDENTIFY VAR=LGOLD CROSSCOR=(S1 C1);
ESTIMATE INPUT = (S1 C1) P=2;
```

IF you replace the INPUT statement with

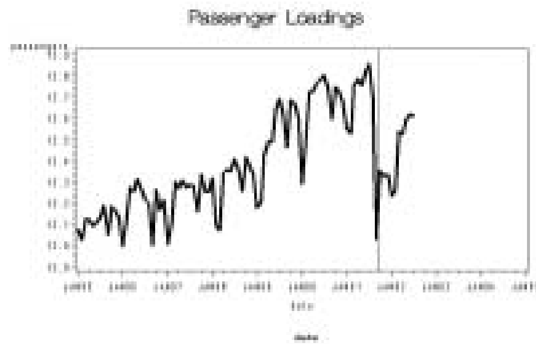
```
ESTIMATE INPUT = (S1 C1) PLOT;
```

You will get the ACF, PACF etc. for the residuals from the sinusoidal fit, thus allowing you to make your own diagnosis of the proper error structure for your data.

A second interesting example involves passenger loadings at RDU airport. This is the airport that serves the Research Triangle area of North Carolina. Data are taken form the RDU web page and represent passenger loadings for the airport. This is an area with several universities and research oriented companies, including the SAS Institute world headquarters. Thus we anticipate a seasonal travel pattern. Either sinusoidal inputs or seasonal dummy variables might be appropriate. A monthly seasonal dummy variable is 0 except when the date represents the particular month for that dummy variable, say January. There the dummy variable is 1. Multiplied by its coefficient, we see that this gives a January shift from some baseline pattern. With 12 months, we use 11 dummy leaving one month (often December) as the baseline.

In order to account for a long run increase or decrease in passenger traffic, a variable DATE will also be included as an input. Finally, imagine another dummy variable WTC that is 0 prior to 9/11/01 and 1 thereafter. Its coefficient would represent a permanent shift, we suppose downward, associated with the terrorist attacks on that date. This model gives little flexibility to the attack response. It can only be represented as a permanent shift in passenger loadings. A bit of additional flexibility is added with a variable IMPACT that is 1 only for September 2001. This allows a temporary one time only additional impact of the attacks. It could be positive, in that 10 days of business as usual occurred, but more likely it will be negative as the airport was actually closed for several days and there was an immediate negative reaction on the part of the flying public.

Here is a plot of the data with a vertical reference line at 9/11/01. The passenger loadings were log transformed because of an apparent increase in variation as the level went up over time. The effect of the World Trade Center attacks is evident in the plot.



We analyze this data as before, using the monthly dummy variables and an ARMA(1,2) error structure obtained by diagnosing the ACF etc. of the residuals along with some trial and error fitting. Here are the estimated coefficients, showing a dropoff in air travel. Again only partial output is shown:

Parm	Estimate	Pr>|t|	Lag	Variable
MU	9.298	<.0001	0	passengers
MA1,1	0.233	0.1485	1	passengers
MA1,2	0.317	0.0300	2	passengers
AR1,1	0.904	<.0001	1	passengers
NUM1	-0.135	<.0001	0	Jan
NUM2	-0.121	0.0007	0	Feb
NUM3	0.041	0.2586	0	Mar
NUM4	0.071	0.0566	0	Apr
NUM5	0.082	0.0298	0	May
NUM6	0.096	0.0113	0	Jun
NUM7	0.106	0.0054	0	Jul
NUM8	0.055	0.1580	0	Aug
NUM9	-0.099	0.0108	0	Sep
NUM10	0.072	0.0492	0	Oct
NUM11	0.022	0.4274	0	Nov
NUM12	0.000	<.0001	0	date
NUM13	-0.145	0.0363	0	impact
NUM14	-0.374	<.0001	0	WTC

Notice that the initial impact of the terrorist attack includes a drop -0.145 that is temporary and a permanent component -0.374 in the log transformed passenger loadings. Both are significant. The standard error associated with the permanent effect is .06871. Thus the log transformed passenger loadings after the attack were 0.374 +/- 1.96(0.06871) less than before the attacks. This gives a confidence interval [-0.2393, -0.5087]. Now if the logarithms of two numbers differ by x, the ratio fo the two number is e^x . Therefore the attacks cause a permanent dropoff in air travel down to between exp(-0.2392)= 78% and exp(-0.5087)= 60% of its former volume, according to the model.

As one might anticipate from the nature of this region of the country, there is always a drop in travel in September as evidenced by the negative September dummy variable coefficient -0.099. Because the variable DATE counts in days, its coefficient is small but yet significant (.0002911 per day increase in log(passengers) corresponding to exp(365(.0002911))=1.112, that is about 11% growth per year) Note the importance of multiplying a regression coefficient times its X variable to assess the effect. Recall that the seasonal dummy variable coefficients represent the difference between the month listed and December. Note the low traffic volume in January and February and the high summer volume.

PROC ARIMA provides some very nice diagnostics. The Q statistics described in the white noise section of this paper are shown next for this model. Replacing q=2 with q=(2) in the ARIMA ESTIMATE statement would put a moving average parameter only at lag 2, not 1 and 2. Some researchers prefer not to skip over lags except in certain seasonal models. We have followed that philosophy.

```
  Autocorrelation Check of Residuals

 To   Chi-        Pr >
Lag Square  DF ChiSq  -Autocorrelations-

 6  3.07   3  0.3808  -0.062 ... -0.071
12  6.59   9  0.6797  -0.041 ... -0.074
18  9.40  15  0.8558   0.074 ... -0.001
24 22.66  21  0.3622   0.007 ... -0.296
```

Notice that there is no evidence of autocorrelation in the residuals. That means our ARMA(1,2) model was sufficient to explain the autocorrelation in the residuals.

Dynamic response to inputs

Another possibility for responding to an input falls under the heading of distributed lags, that is, the effect is spread over several periods. Take for example, an X variable that is 0 except at time t=8. Now specify a model component

$$Y(t) = 5\,X(t) / (1-.8\,B)$$

where as before, B is a backshift operator. What is the meaning of this expression? Algebraically we can write

$$(1-.8\,B)\,Y(t) = 5\,X(t)$$

or

$$Y(t) = 5\,X(t) + .8\,Y(t-1)$$

Now this Y(t) component starts as a string of 0s. At time t=8 X is 1 so Y(8) = 5. Next Y(9) is 0 + .8(5) = 4, Y(10) = 3.2, Y(11) = 2.56 etc. Thus there is an immediate impact of X (perhaps at time 8 there was a major storm, an advertising blitz, etc.) and a lingering but exponentially decreasing effect for several periods thereafter.

If instead X(t) is 1 at time t=8 *and thereafter* then Y again is 7 0s and a 5. This time Y(9) = 5 + .8 Y(7) = 9 and Y(10) = 5 + .8(9) = 12.2 etc. Eventually the Y values approach 5/(1-.8) = 25. This might be the case with a new regulation like an anti-pollution law that takes effect at time 8. The response is only partially felt immediately and accumulates over time. If the coefficient 5 is replaced by -5, this would represent an ultimate decrease of 25 in pollutants.

We can specify such a component for our passenger loadings. Using a dynamic backshift multiplier for the IMPACT variable would allow an immediate negative impact on passenger travel which becomes smaller and smaller over time. If this causes the permanent WTC variable to become insignificant then there is no permanent effect of the attacks on passenger loadings. The syntax is

```
ESTIMATE INPUT=
  (JAN … NOV DATE  /(1) IMPACT   WTC) P=1 Q=2;
```

Notice the syntax /(1) IMPACT. This indicates a denominator lag, because of the "/" and just one lag (1).
Here is part of the output.

```
Parm      Estimate Pr>|t|  Lag Variable
NUM13   -0.30318  0.0219    0 impact
DEN1,1   0.76686  0.0001    1 impact
NUM14   -0.26236  0.0439    0 WTC
```

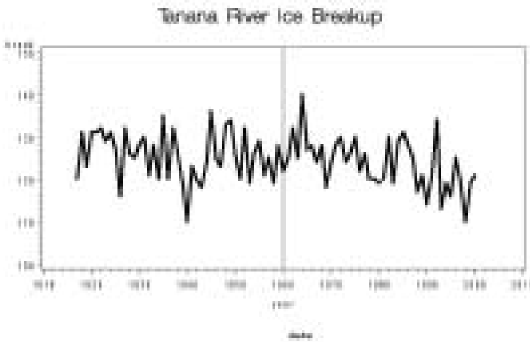
The three values of interest are significant. The rate of "forgetting" as we might call it is 0.77. That is, 77% o fthe initial impact of 9/11/01 remains in October, 59% in November, etc. That is above and beyond the permanent effect -0.26236. Now that the initial impact is allowed to dissipate slowly over time rather than having to disappear instantly, the permanent effect is not quite as pronounced and its statistical significance not quite so strong, although at the standard 5% level we do have evidence of a permanent effect.

This model displays excellent Q statistics just as the one before. In comparing models, one possibility is to use one of the many information criteria in the literature. One of the earliest and most popular is AIC, commonly referred to as Akaike's information criterion. It trades off the fit, as measured by -2L, twice the negative of the log transformed likelihood function versus the number of parameters k in the model. We have AIC = -2L + 2k and the model with the smallest AIC is to be selected. The AIC for the one time only impact model was -215.309 and for the dynamic impact response model it was a smaller (better) number AIC =
-217.98. Other information criteria are available such as the Schwartz Bayesian criterion SBC which has a penalty for parameters related to the sample size, thus a heavier penalty than AIC in large data sets. SBC is smaller in the dynamic model thus agreeing with AIC that the dynamic response is preferred.

Ramps

Some analysts use a ramp function to describe a breaking trend in their data, that is, a linear trend that suddenly changes slope at some point in time. We illustrate with data from the Nenana Ice Classic. This is a very interesting data set. In 1917 workers assembled to build a bridge across the Tanana River near Nenana Alaska, however work could not begin because the river was frozen over. To kill time, the workers began betting on exactly when the ice would break up. A large tripod was erected on the ice and it was decided that when the ice began to break up so that the tripod moved a certain distance downstream, that would be the breakup time and the winner would be the one with the closes time. This occurred at 11:30 a.m. April 30, 1917, but this was only the beginning. It appears that every year, this same routine is followed and the betting has become more intense. The tripod is hooked by a rope to an official clock which is tripped to shut off when the rope is extended by the tripod floating downstream. There is even a Nenana ice classic web site. The data have some ecological value. One wonders if this official start of spring, as the residents interpret it, is coming earlier in the year and if so, could it be some indicator of long term temperature change?

A graph of the breakup time is shown with a vertical line appearing at a point where there seems to be a change from a somewhat level series to one with a decreasing trend, that is, earlier and earlier ice breakup.



Now imagine a variable X(t) that is 0 up to year 1960 then is (year-1960), that is, it is 0, 0 , …, 0 , 1, 2, 3 … Now this variable would have a coefficient in our model. If the model has an intercept and X(t) then the intercept is the level before 1960 and the coefficient of X(t) is the slope of the line after 1960. This variable is sometimes referred to as a "ramp" because if we plot X(t) against t it looks like a ramp that might lead up to a building.

If, in addition, year were entered as an explanatory variable, its coefficient would represent the slope prior to 1960 and its coefficient plus that of X(t) would represent the slope after 1960. In both cases the coefficient of X(t) represents a slope change, and in both cases the reader should understand that the modeler is pretending to have known beforehand that the year of change was 1960, that is, no accounting for the fact that the year chosen was really picked because it looked the most like a place where change occurred. This data derived testing has the same kinds of statistical problems as are encountered in the study of multiple comparison tests.

For the Nenana Ice Classic data, the ramp can be created in the data step by issuing the code

```
RAMP = (YEAR-1960)*(YEAR>1960);
```

We can incorporate RAMP into a time series model by using the following PROC ARIMA code where BREAK is the variable measuring the Julian date of the ice breakup:

```
PROC ARIMA;
IDENTIFY VAR=BREAK CROSSCOR=(RAMP);
ESTIMATE INPUT=(RAMP) PLOT;
```

Some modified output from the partial autocorrelation function produced by the PLOT option follows. No values larger than 0.5 in magnitude occurred and no correlations beyond lag 8 were significant. The labeling of the plot uses 5, for example, to denote 0.5. Note that there is some evidence of correlation at around a 6 year lag.

```
      Partial Autocorrelations

Lag  Corr.   -5 4 3 2 1 0 1 2 3 4 5
  1 -0.029    |       .  *|   .       |
  2  0.097    |       .   |** .       |
  3  0.033    |       .   |*  .       |
  4  0.025    |       .   |*  .       |
  5 -0.187    |       ****|   .       |
  6 -0.250    |      *****|   .       |
  7  0.127    |       .   |***.       |
  8  0.000    |       .   |   .       |
```

The Q statistics test the null hypothesis that the errors are uncorrelated. Here is the (modified) output:

```
     Autocorrelation Check of Residuals

 To   Chi-       Pr >
Lag Square DF   ChiSq  -Autocorrelations-

 6   8.60  6  0.1972  -0.029 ... -0.221
12  15.06 12  0.2380   0.102 ... -0.054
18  18.81 18  0.4037  -0.033 ... -0.018
24  20.81 24  0.6502  -0.037 ... -0.097
```

The Q statistics indicate that the ramp alone is sufficient to model the structure of the ice breakup date. Recall that Q is calculated as n times a weighted sum of the squared autocorrelations. Like the partial autocorrelations, the lag 6 autocorrelation -0.221 appears form the graph (not shown) to be significant. test the null hypothesis that the errors are uncorrelated. I conclude that this one somewhat strong correlation, when averaged in with 5 other small ones, is not large enough to show up whereas on its own, it does seem to be nonzero. There are two courses of action possible. One is to state that there is no particular reason to think of a 6 year lag as sensible and to attribute the statistical result to the one in 20 false rejections that are implied by the significance level 0.05. In that case we are finished – this is just a regression and we would go with the results we have right now:

```
Parm.   Est.  t Value  Pr > |t|

MU    126.24   164.59    <.0001
NUM1  -0.159    -3.36    0.0012
```

The slope is negative and significant (p=0.0012) indicating an earlier ice breakup in recent years.

A second choice would be to put a lag 6 autoregressive term in the model. To do this, issue the code

```
ESTIMATE INPUT=(RAMP) P=(6);
```

Note the parenthesized "lag list" format that is used when you want to include just some lags and skip over others. With several lags in the list, comma separators are required.
The results here are about the same. The autoregressive coefficient is about -.24 and its t statistic -2.11 has a p-value between .03 and .04 so accounting for the fact that looking at lag 6 is a data derived decision, there is not a very strong argument against the original simpler model shown above. The Q statistics, of course, still look good – slightly better than those of the simpler model.

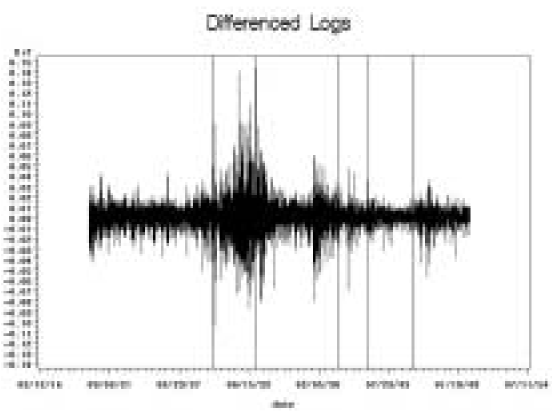
```
 To      Chi-         Pr >
Lag    Square   DF   ChiSq

  6      6.22    5  0.2854
 12     12.56   11  0.3230
 18     15.95   17  0.5274
 24     18.80   23  0.7126
```

In our modeling we have treated the trend break at 1960 as though we had known beforehand that this was the date of some change. In fact, this was selected by inspecting the data.

The data that has been shown thus far are up through 2000 and were popularized in the October 26, 2001 issue of Science, in an article by Raphael Sagarin and Fiorenza Micheli. The next 2 years of data are now available over the web and rather than breaking up early, the last couple of years the ice has stayed frozen until May 8 and 7. With that, the slope becomes -0.128, closer to 0 but still significant.

ARCH models

Some financial data appears to have variance that changes locally. For example, the "returns" on the Dow Jones Industrials Average Y(t) seem to have this property. Letting Y(t) be the day t value of this series, the logarithm of the percentage gain over the previous day R(t) = log(Y(t)/Y(t-1)) represents the overnight return on a dollar. A graph of these returns over a long period of time is shown below. The vertical reference lines are: Black Friday, FDR assumes office, WW II starts, Pearl Harbor, and WW II ends.



Note the extreme volatility during the great depression and the relative stability between Pearl Harbor and the end of WW II.

Engle (1982) and Bollerslev (1986) introduce ARCH and GARCH models. In these models. the innovations variance h(t) at time t is assumed to follow an autoregressive moving average model, with squared residuals where the uncorrelated shocks usually go. The variance model is

$$h(t) = \omega + \sum_{i=1}^{q} \alpha_i e^2(t-i) + \sum_{j=1}^{p} \gamma_j h(t-j)$$

where e(t) is the residual at time t. The model can even be fit with unit roots in the "autoregressive" part in which case the models are called Integrated GARCH or IGARCH models.

We fit an integrated GARCH model (TYPE = INTEG) to these data in PROC AUTOREG as follows:

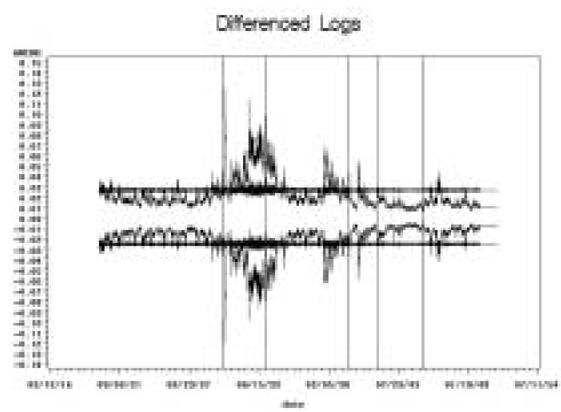
```
PROC AUTOREG DATA=DOW;
  MODEL DDOW = / NLAG=2
  GARCH = (P=2, Q=1, TYPE=INTEG, NOINT );
  OUTPUT OUT=OUT2 HT=HT
  PREDICTED=F LCLI=L UCLI=U;
```

Now the default prediction intervals use an average variance but by outputting the time varying variance HT as shown, the user can construct prediction intervals that incorporate the volatility changes. The output shows the 2 autoregressive parameters and the GARCH parameters, all of which are significant.

Variable	DF	Estimate	Standard t Value	Approx Pr>|t|
Intercept	1	0.000363	4.85	<.0001
AR1	1	-0.0868	-8.92	<.0001
AR2	1	0.0323	3.37	0.0008
ARCH1	1	0.0698	17.60	<.0001
GARCH1	1	0.7078	11.63	<.0001
GARCH2	1	0.2224	3.88	0.0001

Thus it is seen that the variance is indeed changing over time, the ARCH and GARCH parameters being significant.

In closing, here is a graph of the default prediction intervals that use the average variance overlaid with user constructed intervals computed using the square root of HT from the OUT2 data set.



References

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics* , 31, 307-327.

Brocklebank, J. C. and Dickey, D. A. (2003) SAS System for Forecasting Time Series, SAS Institute.

Dickey, D. A. and Fuller, W. A. (1979) "Distribution of the Estimates for Autoregressive Time Series with a Unit Root," *,J. American Stat. Assn*., 74, 427-431.

Dickey, D. A. and Fuller, W. A. (1981) "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica,* 49, 1057-1072.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50, 987-1007.

Fuller, W. A., Hasza, D. P., and Goebel, J. J. (1981) "Estimation of the Parameters of Stocahstic Difference Equations," *Annals of Statistics* 9, 531-543.

Said, S. E. and D. A. Dickey (1984) "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," Biometrika , 71, 599-607.

Tsay, R. S. and Tiao, G. C. (1984) “Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models,” *Journal of the American Statistical Association*, 79, 84-76.

Tsay, R. S. and Tiao, G. C. (1985) “Use of Canonical Analysis in Time Series Identification,” *Biometrika*, 72, 299-315