

Paper 328-2011

SAS/ETS® and the Nobel Prize

David A. Dickey, North Carolina State University, Raleigh, NC

Abstract

Techniques for which the 2003 Economics Nobel Prize was awarded are illustrated with examples. These are cointegration and autoregressive conditionally heteroscedastic models. For this second topic, an introductory example without the heteroscedasticity leads into the more complex scenario.

1. Introduction

The 2003 Nobel Prize in Economics, formally the “Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel,” was awarded in 2003 to Robert F. Engle and Sir Clive W. J. Granger who worked together on the two research projects for which the prize was given while they were at the University of California San Diego. The first project, cointegration, is an extension of the unit root testing area in time series. The second, the ARCH model, (AutoRegressive Conditionally Heteroskedastic models) is a way of dealing with variances that change locally over time. Extensions of these have been derived by researchers such as Stock and Watson (1988), Johansen (1988), Bollerslev (1986) and others. The easier of the two to explain is ARCH. Section 2 of this paper gives the underlying idea behind the ARCH model and Bollerslev’s extension called the GARCH model. Section 3 gives an example of interest in economic history. Section 4 introduces the idea of cointegration and explains some of the mathematical ideas behind the technique. Section 5 is a cointegration example and section 6 is a brief conclusion section. Most examples here come from the book SAS® for Forecasting Time Series (Brocklebank and Dickey, 2003).

2. ARCH Models – Motivation and Introductory Example

We begin by looking at a main motivation for ARCH models, namely returns on investments. In particular we think about the stock market. What is usually of most interest to investors is the percent gain that they get for an investment. In that light, a gain of \$10 on a \$10,000 investment in, say, the stock market is a 0.01% gain, the same as a gain of \$1 on a \$1000 investment. Whatever we had to start with got multiplied by 1.001. If the investment period is short, overnight for example, we would expect the gain to be not too far from 0 and the multiplier then to be close to 1. We think of a loss as a negative gain here. For a number X near 1, the (natural) logarithm of X is approximately X-1. The logarithm of 1.001 is very close to 0.001, the logarithm of 0.999 is very close to -0.001. Here are two graphs of log(X) and X-1 as X goes from 0.4 to 1.6 (left) and as X goes from -0.9 to 1.1 (right) .

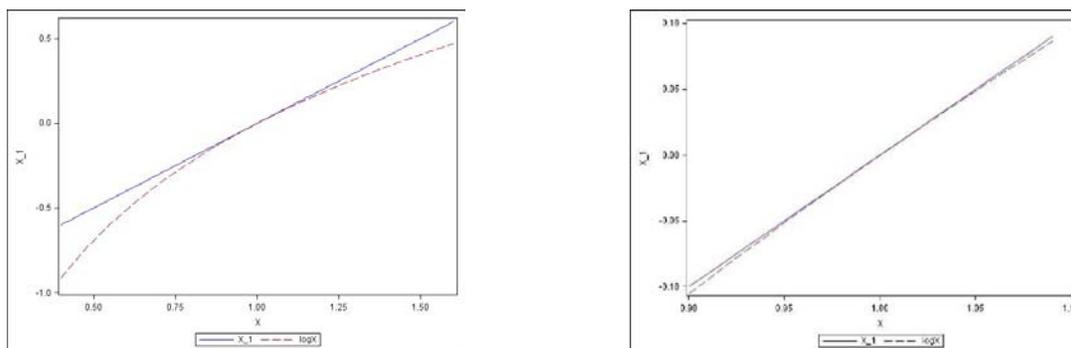


Figure 1: $\log(X)$ and $X-1$ over two domains.

The ratio of what my investment is worth today, X_t to what it was worth yesterday X_{t-1} is unlikely to lie outside the narrower range in the right graph. If X_t/X_{t-1} is 1.01, a 1% overnight return, then the return is approximately

$\ln(X_t / X_{t-1}) = \ln(X_t) - \ln(X_{t-1}) \approx 0.01$, that is, within a reasonable range the return is approximately the first difference of the log transformed series

Taking logarithms and differencing them is often advantageous from a statistical point of view as well as an interpretation point of view. That is, these differences of logarithms seem more stable in terms of their mean and variance as well as seeming to satisfy the conditions for stationarity better than the original measured variables do. The conditions for stationarity are that the mean is constant and the covariance function between responses at two time points is a function only of the difference between those times. Tests for stationarity are available in SAS/ETS PROC ARIMA.

The example in the next section involves the Dow Jones Industrials Average (DJIA). In Figure 2 we see a plot of the returns on the DJIA over an interesting historical period. The vertically paired dots in the figure are for interesting historical dates. The dates are, from left to right, 29OCT1929 (Great Depression begins), 04Mar1933 (FDR takes office), 12Sept1939 (WW II begins), 07Dec1941 (Pearl Harbor), 12Apr1945 (FDR leaves office), 14Aug1945 (VJ day - war ends). The mean return is 0.000070174 which seems small until you realize it is a *daily* percentage return (gain) in a rather early period in US economic history. A striking feature of the graph is the changes in the local variation shown by the data. Local (in time) variation is referred to as “volatility” and we see periods of low volatility and periods of high volatility, especially during the Great Depression.

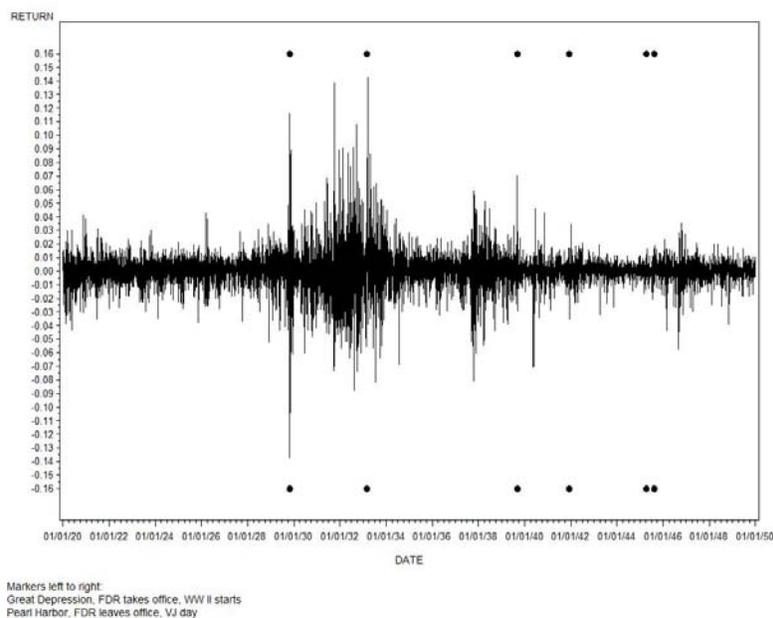


Figure 2: Returns on DJIA

Clearly, the periods of low and high volatility last for days, weeks, or months at a time and a good model should incorporate the fact that a high volatility day is more likely to be followed by another high volatility day than by a low volatility day. Likewise low volatility days seem to occur in runs over time. This suggests autocorrelation in the volatility. We have lots of models for autocorrelated data, most notably the ARIMA models. Here, however, it is the squares of the deviations from the mean that seem autocorrelated rather than the values themselves. Because there is such a strong tradition in statistics of measuring variation by the statistic we know as the variance, we might propose a model in which the variances change over time in an autocorrelated way. If a stationary ARMA (AutoRegressive Moving Average) model can describe the behavior, then we might be able to estimate appropriate parameters.

In general, the series might not have a constant mean but rather may be trending in some way or have deterministic seasonal components, for example. For this series, however, it appears that a constant mean is sufficient for the deterministic part of the returns.

Thinking of the deviation of each day's return R_t from the sample mean as $D_t = R_t - 0.000070174$, and treating the sample mean as though it were the true mean, we could create the square of these D_t deviations D_t^2 . Each D_t^2 would then be an estimate of variance (volatility) for that day. Continuing with this somewhat simple idea, we could

model the D_t^2 series as a time series. In Figure 2, we see a plot of $Y_t = D_t^2$ versus t and its autocorrelation function (ACF, upper right) as well as the partial and inverse autocorrelation functions (lower left and right). These graphics were created with the ODS GRAPHICS ON; statement and PROC ARIMA.

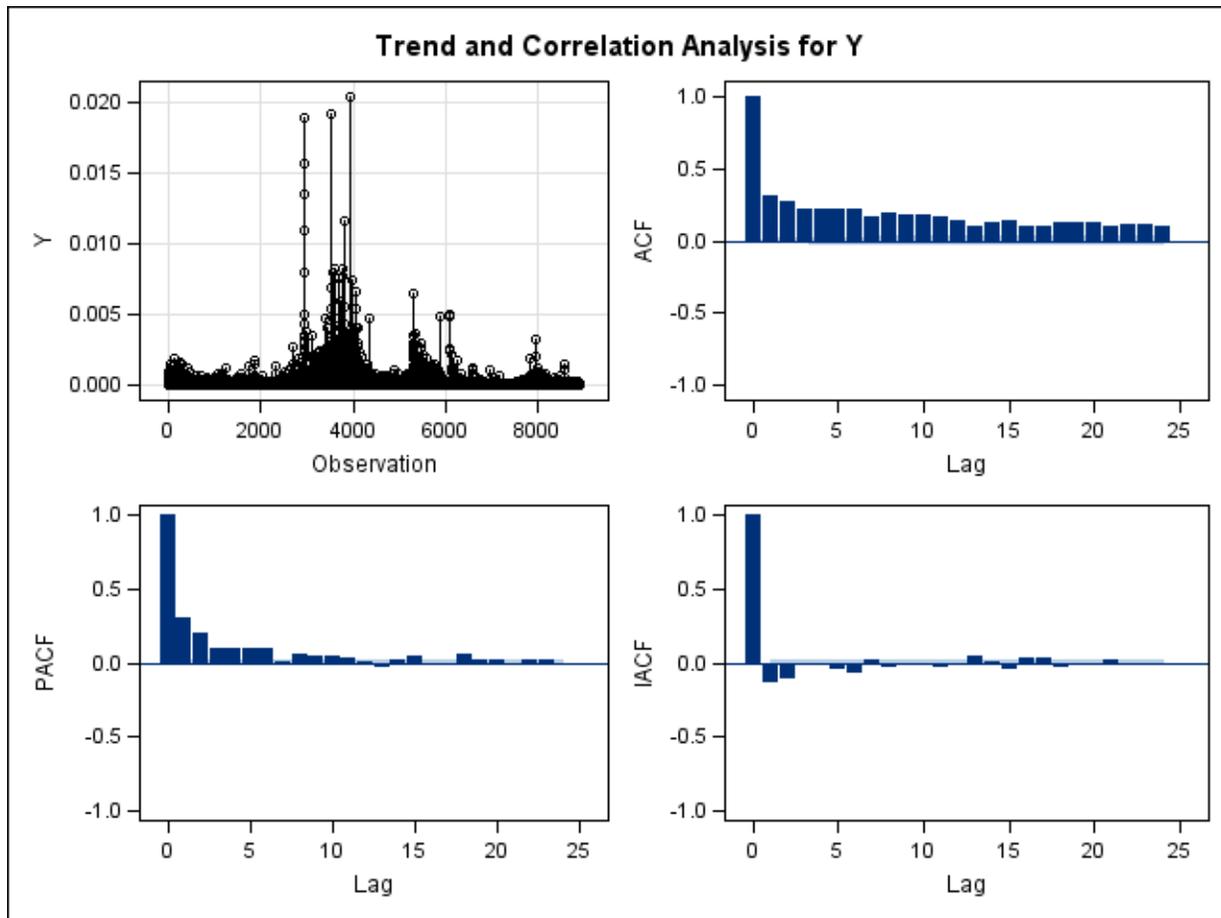


Figure 3. Diagnostic Plots for Dow Jones Industrials Average

Notice especially the ACF (autocorrelation function) in the upper right corner. The rather severe drop in autocorrelation from lag 0 to lag 1, followed by slow decay, is characteristic of autoregressive moving averages of orders 1 and 1, ARMA(1,1). If the decay is extremely slow, the series might even have a unit root and thus be in a class of nonstationary series. It would be an IMA or integrated moving average. The ARCH models and their variants GARCH and IGARCH will provide nice machinery for analysis of this kind of volatility. The models are available in the SAS/ETS package with PROC AUTOREG. The idea here is to propose a model whose error term has theoretical variance σ_t^2 at time t and from there, model σ_t^2 with an ARIMA model. Before demonstrating its ARCH capacity, we begin with a simpler introductory PROC AUTOREG example for a regression with time series errors and no volatility changes.

3. PROC AUTOREG

The basic model in PROC AUTOREG, in matrix form, is $Y=XB+Z$ which looks just like an ordinary least squares regression model. The difference is that the elements Z_t of the error vector are assumed to have an autoregressive structure, for example it might be that $Z_t = \rho Z_{t-1} + e_t$ where the e_t series is an independent mean 0 and constant variance process, that is, e_t is "white noise." In such a case, the best estimates of the parameters use generalized least squares, that is, the estimates of the β s are given by $\hat{\beta} = (X'V^{-1}X)^{-1}(X'V^{-1}Y)$ where V is the variance-

covariance matrix of the random vector \mathbf{Z} . The problem is that this matrix \mathbf{V} involves unknown parameters, ρ for instance. Nevertheless it is possible to write down the likelihood for the model and maximize it. If one then assumes the estimated autoregressive parameters are the true ones, one can compute a \mathbf{V} and then the generalized least squares estimates. In addition, the errors can have the kind of autocorrelated variances as described in Section 2, but before describing this situation, an example of the discussion thus far, from Brocklebank and Dickey (2003) will be given.

The data here are energy usage numbers for the campus of North Carolina State University during the July 1, 1979 through June 30 1980 academic year. In 1979, the Shah of Iran was evicted and the flow of oil from the Arab world disrupted resulting in the second so-called "oil crisis" the first having been an embargo in 1973 by the Organization of Arab Petroleum Exporting Countries. This heightened concern over energy consumption in the U.S. A sign listing the previous day's energy consumption greeted anyone entering the campus. Temperature is a rather obvious input as is a class variable for type of day, this having three levels. One level (lower circles in Figure 4) is a non workday, meaning that neither faculty nor students were required on campus. This includes weekends of course. Another level (upper stars in Figure 4) is a class day. The third level (middle squares in Figure 4) is a work day that is not also a class day. On such days, faculty and staff are expected to report to work but not students.

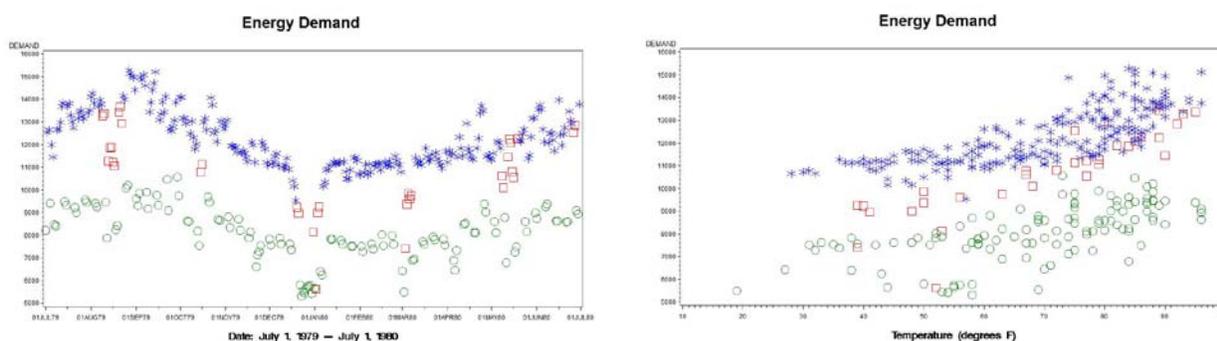


Figure 4: Energy Demand at a University

In the data set we will label the resulting 2 indicator variables as WORK (0 for a non work day, 1 otherwise) and, to avoid confusion with the CLASS statement TEACH (1 for days when classes were taught, 0 otherwise). These variables will have coefficients and it should be understood that on a day when classes are taught, both the WORK and TEACH variables will take on the value 1, that is, the effect will be the sum of those two coefficients.

The left panel in Figure 4 shows energy demand versus time (day, July 1 1979-July 1 1980) and on the right demand versus temperature. On the left, a general pattern of lower demand in the cooler months and higher in the warmer months may result from just temperature effects in the North Carolina area as well as relative costs of cooling versus heating. Temperature effects shown in the right panel are consistent with this idea and seem to show a nonlinear response that will be modeled as a quadratic effect in temperature. The code for this model includes a sine and cosine of 1 year duration, S and C, and is given by

```
PROC AUTOREG data=energy;
MODEL DEMAND = TEMP TEMPSQ TEACH WORK S C
/NLAG=15 BACKSTEP;
```

The procedure starts with an AR(15) model, eliminating insignificant parameters (BACKSTEP) and arriving at a model with all parameters highly significant and a total R^2 over 95%. This does not involve heterogeneous variances and is included simply to show the power and ease of use of PROC AUTOREG. The final coefficients and appropriate standard errors that account for autocorrelation are in the following output:

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	6076	296.5261	20.49	<.0001
TEMP	1	28.1581	3.6773	7.66	<.0001
TEMPSQ	1	0.6592	0.1194	5.52	<.0001
TEACH	1	1159	117.4507	9.87	<.0001
WORK	1	2769	122.5721	22.59	<.0001
S	1	-764.0316	186.0912	-4.11	<.0001
C	1	-520.8604	188.2783	-2.77	0.0060

We see there is a quadratic temperature component (TEMPSQ is the square of (temperature-65)), a sinusoidal trend over and beyond the sinusoidal movement of temperature, and an intercept 6076. Add to this 2769 on work days plus another 1159 if the work day is also a class day. Several autoregressive parameters, dominated by lags at 1, 7, and 8, were retained for the error term as indicated by this part of the output:

Estimates of Autoregressive Parameters				
Lag	Coefficient	Standard Error	t Value	
1	-0.559658	0.043993	-12.72	
5	-0.117824	0.045998	-2.56	
7	-0.220105	0.053999	-4.08	
8	0.188009	0.059577	3.16	
9	-0.108031	0.051219	-2.11	
12	0.110785	0.046068	2.40	
14	-0.094713	0.045942	-2.06	

4. PROC AUTOREG for ARCH, GARCH, and IGARCH models.

The flexibility of PROC AUTOREG extends beyond autocorrelated errors. We now explore the case, as exemplified in the DJIA example, in which the error variance is autocorrelated. The idea originally proposed by Engle (Engle, 1982) is to write the variance of the error term e_t as σ_t^2 and then to model it as a linear combination of past squared errors: $\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \dots + \alpha_q e_{t-q}^2$ where it is assumed that the α s are positive and the sum of α_1 through α_q is less than 1. In this case, there is a long run unconditional variance $\sigma_{LR}^2 = \alpha_0 / (1 - \alpha_1 - \dots - \alpha_q)$. Our ad hoc idea of squaring the errors and estimating autocorrelations is seen to be reasonable when $(e_t^2 - \sigma_t^2)$ is added to both sides of the basic ARCH equation, obtaining $e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \dots + \alpha_q e_{t-q}^2 + (e_t^2 - \sigma_t^2)$. This looks like an autoregressive model in the variable e_t^2 as the error term $(e_t^2 - \sigma_t^2)$ has mean 0. Engle's student Bollerslev (1986) generalized this to include terms analogous to moving average terms. He termed this model "Generalized ARCH" or GARCH. The model can be written $\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \dots + \alpha_q e_{t-q}^2 - \theta_1 \sigma_{t-1}^2 - \dots - \theta_p \sigma_{t-p}^2$. The structure here is similar to an ARMA model. If the variance σ_t^2 is very persistent (long periods of high volatility and long periods of low volatility) then one might entertain the possibility of a unit root in the so-called "characteristic equation" $1 + \theta_1 B + \dots + \theta_p B^p = 0$. See section 5. Analogous to the AutoRegressive Integrated Moving Average (ARIMA) models this is called an integrated GARCH or IGARCH. As expected, this implies that the time to time changes in the variance, $\sigma_t^2 - \sigma_{t-1}^2$ satisfy a GARCH model.

Often, time series analysis is done in order to forecast. If prediction intervals are to be placed around the forecasts, then estimates of the future values of the variance at lead times L , σ_{t+L}^2 are needed. As happens in ARMA models, the GARCH estimates will approach the historic mean of the variances while in an IGARCH, the estimates will not be mean reverting.

The models discussed are available in PROC AUTOREG. An example of the code that illustrates the syntax and a discussion of the resulting output will conclude this section. We look at the returns $Y_t = \ln(X_t) - \ln(X_{t-1})$ where the X_t values are levels of the DJIA. For most procedures any differencing and transformation that is desired must be done outside the procedure and this is true of PROC AUTOREG. For purposes of illustration we will fit both a GARCH and IGARCH model to the data. Arguments in favor of the IGARCH model are given by Brocklebank and Dickey (2003). Code for the IGARCH model is

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

The data set MORE has the Dow Jones series with 500 additional dates with missing Y values concatenated at the end for illustrating forecasts into the future. In this case an AR(2) model (NLAG=2) is needed to account for error autocorrelation. The TYPE=INTEG and NOINT indicate that we have a unit root model for the innovation variance and there is no drift in it, that is, the differences of these time varying variances, $\sigma_t^2 - \sigma_{t-1}^2$, have mean 0. This implies that the forecast of future variances will be just the last observed variance σ_t^2 , plus some local adjustments arising from the autocorrelation in the differences. In particular there is no reversion to the historic average variance. Upper and lower confidence limits along with forecasts and values HT, these being estimates of the local variance σ_t^2 , are output and used to make a graph. In Figure 5, two types of 95% confidence intervals are shown. The PROC AUTOREG default intervals use an average historical variance and plot as nearly parallel lines, the fuzziness arising from the AR(2) errors, not from the use of HT. The more dramatically differing intervals are computed by adding and subtracting 1.96 times the square root of HT to (from) the forecast. The motivation is that HT is the estimate of the time varying innovations variance σ_t^2 , the one step ahead forecast error variance. The unit root (IGARCH) nature of the error variance models manifests itself in the almost horizontal lines associated with prediction intervals into the future. Notice how the HT based intervals reflect variation toward the end of the series and are much narrower than those that equally weigh recent and depression era variation.

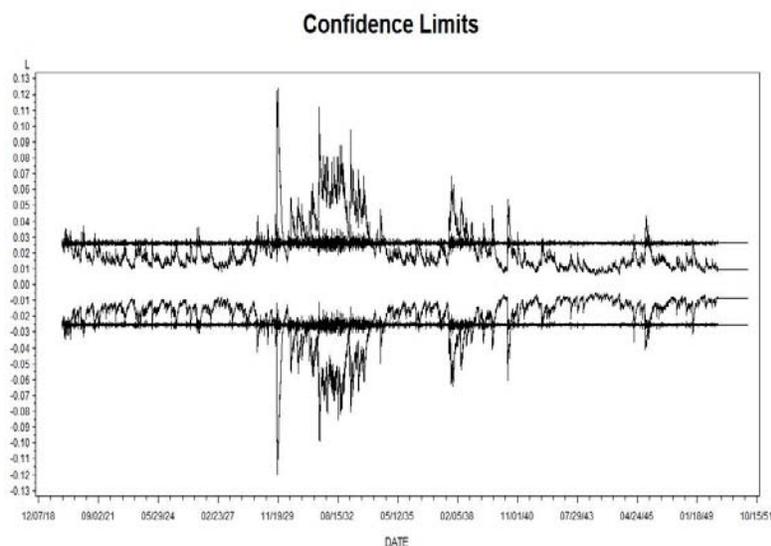


Figure 5. Default and User Defined Confidence Limits for Figure 2 Data.

The output from the above PROC AUTOREG includes a parameter estimate section

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	0.000363	0.0000748	4.85	<.0001
AR1	1	-0.0868	0.009731	-8.92	<.0001
AR2	1	0.0323	0.009576	3.37	0.0008
ARCH1	1	0.0698	0.003963	17.60	<.0001
GARCH1	1	0.7078	0.0609	11.63	<.0001
GARCH2	1	0.2224	0.0573	3.88	0.0001

The interpretation here is that the return Y_t satisfies

$Y_t - 0.000363 = Z_t$ where $Z_t = 0.0868Z_{t-1} - 0.0323Z_{t-2} + e_t$ and where the local variance σ_t^2 of e_t satisfies the unit root model equation $\sigma_t^2 = 0.7078(\sigma_{t-1}^2) + 0.2224(\sigma_{t-2}^2) + 0.0698e_{t-1}^2$.

By omitting the two options that distinguish IGARCH from GARCH we have the code

```
PROC AUTOREG DATA=MORE;
  MODEL DDOW = / NLAG=2 GARCH=(P=2,Q=1);
  OUTPUT OUT=OUT3 HT=HT PREDICTED=F LCLI=L UCLI=U;
RUN;
```

The resulting one step ahead prediction limits in Figure 6 look quite similar through the historic data to those in Figure 5 while their forecasts into the future are quite different in that the HT based intervals approach the long run default intervals as we forecast further out.

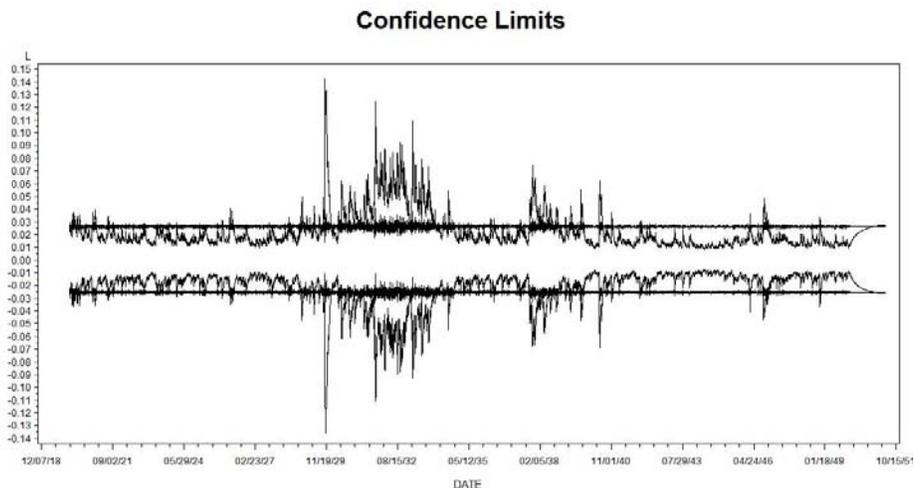


Figure 6. GARCH Prediction Limits Differ From IGARCH.

The maximum likelihood method used to estimate these models is quite sensitive to the assumption of a normal distribution. The output includes the Jarque-Bera test for normality which is $n(b_1^2/6 + (b_2 - 3)^2/24)$ where n is the sample size and the b_s are third and fourth moments (0 and 3 in theory for a normal distribution) and are defined in terms of residuals r_t , as

$$b_1 = \frac{\sum_{t=1}^n r_t^3 / n}{\left(\sum_{t=1}^n r_t^2 / n\right)^{3/2}} \quad \text{and} \quad b_2 = \frac{\sum_{t=1}^n r_t^4 / n}{\left(\sum_{t=1}^n r_t^2 / n\right)^2}.$$

The test has approximately a 2 degree of freedom Chi-square distribution for normal data. The output shows a test statistic $\chi^2 = 3886$ as the “normality test” (Jarque-Bera) which is way too large to accept the normal assumption. Further details in Brocklebank and Dickey (2003) and references therein indicate that the estimates of parameters on the regressors (just the mean in this example) can be biased as a result, but that the confidence limits are still valid for large samples. Further analysis of the data appears in that book.

5. Cointegration.

The 2003 Nobel Prize citation also mentions cointegration which is closely related to unit root tests. Here we briefly review some background on unit root tests as can be found in many time series texts. The idea of unit root tests can be seen in the model $Y_t - \mu = 1.2(Y_{t-1} - \mu) - c(Y_{t-2} - \mu) + e_t$ which is rewritten in differences and one lagged level as $Y_t - Y_{t-1} = -(1-1.2+c)(Y_{t-1} - \mu) + c(Y_{t-1} - Y_{t-2}) + e_t$, using the fact that, for example,

$(Y_t - \mu) - (Y_{t-1} - \mu) = (Y_t - Y_{t-1})$. We see that if $c=0.2$ there is no dependence on μ , and in fact the whole model

is expressed in terms of differences as an AR(1): $Y_t - Y_{t-1} = .2(Y_{t-1} - Y_{t-2}) + e_t$. There is no tendency to return

to the mean. If $c>.2$, say .32 for example, then we have $Y_t - Y_{t-1} = -0.12(Y_{t-1} - \mu) + c(Y_{t-1} - Y_{t-2}) + e_t$ which in turn says that if last period's Y is above the mean (positive deviation) then, ignoring the lagged difference, the next change will tend to be negative, specifically - 0.12 times the deviation. Likewise a negative deviation will tend to be followed by a positive change. This is the idea of mean reversion and is a simple case of an error correction model, one in which an “error” (deviation from the mean) tends to be followed by movement back toward the mean. The model $Y_t - Y_{t-1} = -(1-1.2+c)(Y_{t-1} - \mu) + c(Y_{t-1} - Y_{t-2}) + e_t$ can be fit by simply regressing the differences on a lagged level and enough lagged differences so that the error term appears to be uncorrelated noise, one difference in our case. A test of the hypothesis of no mean reversion is a test that the lagged Y coefficient is 0. The t test can be used, but due to violations of the standard regression assumptions (in particular the assumption of fixed regressors, not lagged values of the response) the t test does not have a t *distribution* and a special distribution, the Dickey-Fuller test distribution, must be used. It is available in PROC ARIMA. The term “unit root” comes from the so-called “characteristic equation” of the process, in our case $1-1.2B+cB^2=0$ with roots 1 and 5 when $c=0.2$, our “unit root” case with no mean reversion, and roots 1.25 and 2.5 when $c=0.32$, our stationary mean reverting case. Roots larger than 1 in magnitude are associated with mean reversion, that is, with stationarity.

The idea of cointegration is that two or more series, say H_t and L_t , may have unit roots, but some linear combination of them, like $Y_t = H_t - L_t$ is stationary, meaning that its roots are larger than 1 in magnitude and thus that the difference is mean reverting. Thinking about H_t as the weekly high price for a stock and L_t as the low price, we know that Y_t has a nonzero mean, the long run average spread between high and low, and we expect the high and low to track each other pretty well so we would then expect the spread Y_t to be mean reverting around the long run difference in these prices, that is, you would expect the difference variable Y to reject unit roots even though the individual H and L series appear to have unit roots. Thus our plan is to simply calculate the difference Y in a data step then test Y to see if it is stationary.

Figure 7 is a plot of weekly IBM high and low stock prices from January 3, 2005 through February 19, 2011. The picture seems to indicate that the high and low prices rise and fall together but it is clear from the graph (and consistent with the unit root tests with p-values near 0.9) that the two components are nonstationary, likely unit root processes.

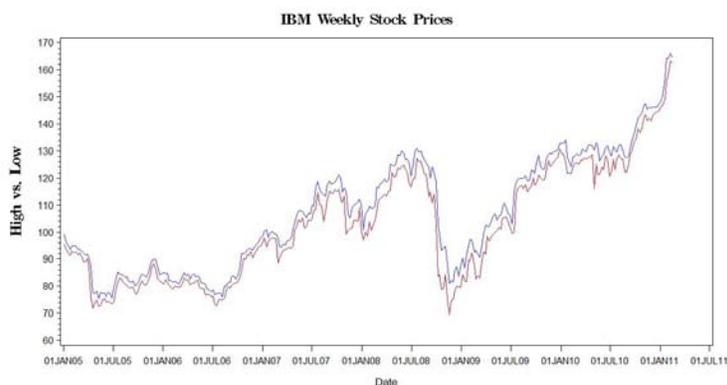


Figure 7. IBM high and low stock prices.

Define the spread to be the difference between the high and low prices, a measure of profit for a perfectly timed trade. If that spread, $\text{Spread} = (1)\text{High} + (-1)\text{Low}$, is stationary, then it is very unlikely that the high and low prices will become far apart. We say that the high and low prices are then cointegrated and the vector of weights $(1, -1)$ is called the “cointegrating vector”. The idea applies to stationary weighted averages of several nonstationary series and with, say, 3 such series one could have 0, 1, or 2 linearly independent cointegrating vectors. In many applications, there is just one such vector. For the IBM data, the spread can be tested for stationarity in PROC ARIMA using

```
proc arima;
  i var=spread stationarity=(ADF=(0 1 2 3 4 5 6 7 8));
  e p=1 q=1 ml; run;
```

Here the use of up to 8 lagged differences may be overkill, but even so, all of the unit root test p-values were less than 0.0125 so the evidence against unit roots is reasonably strong. The fitted model, ARMA(1,1) for the error terms had lack of fit p-values up through 48 lags all of which exceeded 0.09 so the fit seemed reasonable and the stationary model for the spread Y_t was estimated as $Y_t - 4.43 = 0.93(Y_{t-1} - 4.43) + e_t - 0.64e_{t-1}$. A plot of the spread is shown in Figure 8.

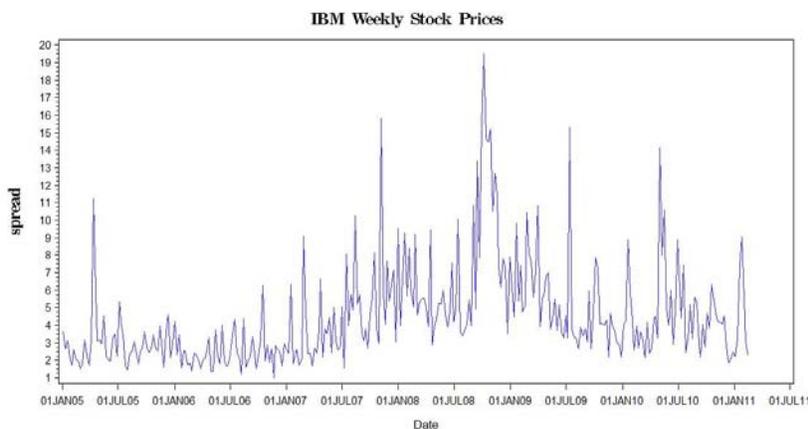


Figure 8. High-low spread, IBM stock.

When the cointegrating vector, $(1, -1)$ in our case, is known, the problem is simple given the many software packages that perform unit root tests. What if the cointegrating vector is unknown? Engle and Granger (1987) originally suggested regressing one series on the other and testing the residuals for unit roots. That way the cointegrating vector would be $(1, -b)$ where b is the regression coefficient when Y is regressed on X – but which variable is X and

which Y? It does make a difference and several researchers looked into multivariate methods of estimation with a method due to Johansen (1988) being implemented in PROC VARMAX. Before illustrating VARMAX, let us take a simple but well justified approach to estimation. We have a bivariate vector $V_t = (H_t - \mu_H, L_t - \mu_L)'$ and will postulate a vector autoregressive model in which this vector is related to lags of itself by

$V_t = A_1 V_{t-1} + \dots + A_p V_{t-p} + E_t$ with the A s being matrices and the E a vector of possibly cross correlated white noise series. The elements of each A can be found, row by row, by regressing each element of V on the lags of all the elements in V . Further, mimicking the unit root tests, the model can be rewritten in terms of differences and just one lagged level as $\nabla V_t = \pi V_{t-1} + B_1 \nabla V_{t-1} \dots + B_p \nabla V_{t-p+1} + E_t$ where ∇ symbolizes a difference. Regressing the differenced low and high prices on the lagged levels and 2 lagged differences of both (after using the multivariate MTEST option in PROC REG to determine that $p=3$ lags were sufficient) we obtain an estimate

$$\hat{\pi} = \begin{pmatrix} -0.226 & 0.235 \\ 0.076 & -0.081 \end{pmatrix} \approx \begin{pmatrix} -0.23 & 0.23 \\ 0.08 & -0.08 \end{pmatrix} = \begin{pmatrix} -0.23 \\ 0.08 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix}$$

Notice that with some minor rounding, the so-called "impact matrix" π is a column vector α times our previously assumed cointegrating vector $\beta' = (1, -1)$. The matrix is said to be rank one since each of its columns is a scalar multiple of one column, α . We then have $\pi = \alpha\beta'$ from which

$$\nabla V_t = \begin{pmatrix} \nabla H_t \\ \nabla L_t \end{pmatrix} = \begin{pmatrix} -0.23 \\ 0.08 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} H_{t-1} - \mu_H \\ L_{t-1} - \mu_L \end{pmatrix} + \text{lagged differences} \quad (1)$$

Thus when $H_{t-1} - L_{t-1}$ is larger than 4.43, the estimated long run mean spread, say C units above, then (ignoring lagged differences) the next change in H tends to be $-0.23C$ and in L tends to be $0.08C$, that is, the low moves up and the high down. If the difference is less than 4.43 the next changes in L and H tend to separate them. There is continual pressure to move H and L so that the spread between them is close to $\mu_H - \mu_L$ which we estimated as 4.43. This is the general idea of cointegration. Another interesting fact is that if we fill out a full rank square matrix by listing the cointegrating vector(s) β' as its first row(s) then adding a row(s) α'_\perp orthogonal to α in the factored impact matrix π (for example $\alpha'_\perp = (0.08, 0.23)$ is orthogonal to α because $\alpha'_\perp \alpha = (0.08)(-0.23) + (0.23)(0.08) = 0$) then we have a transformation matrix T which, when multiplied by the vector of observations, gives a transformed vector time series whose first row(s) are stationary series, the stationary linear combinations for the cointegrating vectors, and the remaining row(s) are unit root processes referred to as stochastic trends in the series. As an analogy, think of two lifeboats adrift in a choppy sea. They move around tossed by the waves, but if they are tied together by a bungee cord, they cannot get too far apart. A drunken person walking a dog on a leash is another analogy.

In the vector representation (1) above, notice that multiplying both sides by $\alpha'_\perp = (0.08 \quad 0.23)$, eliminates the lagged levels of the variables *and their means*, leaving a driftless unit root process, a process specified strictly in terms of differences with no constant term. This is so because we assumed we could write things in terms of deviations from the mean and that in doing so α'_\perp removes not only the lagged levels but also the two means, leaving a driftless unit root process (no constant). If the two separate means are not in the right ratio then the unit root process will have an up or down drift over time, a stochastic trend. The absence of such a drift will be called the "restriction" in the upcoming VARMAX output.

We still face a testing problem. While we were able to use relatively basic procedures for this data, ending up with a matrix that, after mild rounding, could be factored into a column times a row, it is not clear in general how much "rounding" should be allowed, that is, we need a test to see if we are statistically close to an impact matrix that factors as $\pi = \alpha\beta'$, and if so, what are the factors – especially the rows of β' that show us the cointegrating vector.

Johansen gave such a test and it is implemented in PROC VARMAX. We close by running the procedure on the IBM data.

6. A Cointegration Example – Stock Prices.

Here is the code for applying PROC VARMAX to our IBM data:

```
PROC VARMAX DATA=IBM;
MODEL HIGH LOW/P=3 LAGMAX=4 ECM=(RANK=1 NORMALIZE=HIGH ECTREND) COINTTEST;
COINTEG RANK=1 H=(1 0, -1 0, 0 1);
ID DATE INTERVAL = WEEK; RUN;
```

Here P=3 refers to the 3 lag vector autoregressive model, LAGMAX is a maximum lag for diagnostics, and COINTTEST asks for tests of 0 vs.1 and 1 vs. 2 cointegrating vectors. The ECM (Error Correction Mechanism) option has several parts. The term “error correction” refers to the existence of a stationary linear combination, an equilibrium relationship between the variables, $H_t - L_t = 4.43$ for example, the violation of which is an error which is then at least partially corrected by movement back toward equilibrium. The RANK=1 specifies that we want a solution for which the impact matrix involves only 1 column vector α in its factored form. Thus 1 is the “rank” of the impact matrix π . The ECTREND specifies that the only nonzero constant allowed is in the error correction mechanism itself, that is, it imposes the restriction that the nonstationary transformed series, $\alpha'_{\perp} V_t$, has no drift. The COINT statement tests the hypothesis that in a rank 1 model the cointegrating vector can be taken as (1,-1) without statistically degrading the fit. If weights (1,-1) produce a stationary linear combination then so do (-2,2), (3,-3) and any (c,-c) combination. It is therefore possible to specify that one of the variables have a coefficient 1 and here it is high (NORMALIZE=HIGH).

With these items in mind, we present relevant portions of the output, starting with the “no stochastic trend” restriction

Hypothesis Test of the Restriction

Rank	Eigenvalue	Restricted Eigenvalue	DF	Chi-Square	Pr > ChiSq
0	0.1144	0.1157	2	1.64	0.4395
1	0.0002	0.0039	1	1.19	0.2749

In the rank 1 model we propose, there is thus no evidence of a drift in the nonstationary part ($p=0.2749$). The ECTREND option is justified. We next check for the number of cointegrating vectors.

Cointegration Rank Test Using Trace Under Restriction

H0:	H1:	Eigenvalue	Trace	5% Critical Value	Drift in ECM	Drift in Process
Rank=r	Rank>r					
0	0	0.1157	40.2134	19.99	Constant	Constant
1	1	0.0039	1.2433	9.13		

In the model with no drift, we reject the hypothesis of no cointegration ($40.2134 > 19.99$) but cannot reject the hypotheses of 1 in favor of 2 (or more - here only 2 are possible). Our test is labeled “Trace” as it is Johansen’s trace test. We have one cointegrating vector and hence a rank 1 impact matrix. We turn to the form of that matrix, first looking at the beta then the alpha matrix. Some output has been highlighted in bold for presentation here.

Long-Run Coefficient Beta Based on the Restricted Trend

Variable	1	2
High	1.00000	1.00000
Low	-1.04515	-1.42493
1	-0.23428	74.47035

Adjustment Coefficient Alpha
Based on the Restricted Trend

Variable	1	2
High	-0.22974	0.00325
Low	0.07046	0.00513

Our impact matrix factors as $\pi = \alpha\beta'$ but we are given 2 columns in each output matrix. The leftmost columns are used to construct the rank 1 impact matrix we desire. The rank 2 version, not of interest here, would be obtained by using both columns of each display.

The extra row in "beta" is due to the intercept entering (only) the error correction mechanism. Recall that this is the "restriction" investigated at the outset. It increases the cointegrating vector by 1 parameter which in turn multiplies a 1 that has been added to the lagged level vector. Ignoring lagged differences and substituting numbers from the display, the leading part of our solution is

$$\begin{pmatrix} H_t - H_{t-1} \\ L_t - L_{t-1} \end{pmatrix} = \begin{pmatrix} -0.2241 \\ 0.07929 \end{pmatrix} \begin{pmatrix} 1.00000 & -1.04515 & -0.23428 \end{pmatrix} \begin{pmatrix} H_{t-1} \\ L_{t-1} \\ 1 \end{pmatrix} =$$

$$\begin{pmatrix} H_t - H_{t-1} \\ L_t - L_{t-1} \end{pmatrix} = \begin{pmatrix} -0.22974 & 0.24011 & 0.05382 \\ 0.07046 & -0.07364 & -0.01651 \end{pmatrix} \begin{pmatrix} H_{t-1} \\ L_{t-1} \\ 1 \end{pmatrix}$$

which is seen to be the product $\alpha\beta'$ using the

"Variable 1" columns (in bold) of the two displays above.

The next part of our interest lies in the table of estimates. The coefficients in the matrix product above, with the intercept coefficient listed first, comprise the first 3 (in bold) entries in the two sections of the "Model Parameter Estimates" output. The other elements are entries of the two 2x2 matrices associated with the 2 lagged difference vectors.

The VARMAX Procedure

Type of Model	VECM(3)
Estimation Method	with a Restriction on the Deterministic Term Maximum Likelihood Estimation
Cointegrated Rank	1

Model Parameter Estimates

Equation	Parameter	Estimate	Standard Error	t Value	Pr > t	Variable
D_High	CONST1	0.05382	0.01350			1, EC
	AR1_1_1	-0.22974	0.05762			High(t-1)
	AR1_1_2	0.24011	0.06023			Low(t-1)
	AR2_1_1	0.05311	0.07608	0.70	0.4856	D_High(t-1)
	AR2_1_2	0.16823	0.07024	2.39	0.0172	D_Low(t-1)
	AR3_1_1	-0.11786	0.06769	-1.74	0.0827	D_High(t-2)
	AR3_1_2	-0.00332	0.06212	-0.05	0.9574	D_Low(t-2)

D_Low	CONST2	-0.01651	0.01693			1, EC
	AR1_2_1	0.07046	0.07228			High(t-1)
	AR1_2_2	-0.07364	0.07554			Low(t-1)
	AR2_2_1	0.53719	0.09543	5.63	0.0001	D_High(t-1)
	AR2_2_2	-0.25195	0.08811	-2.86	0.0045	D_Low(t-1)
	AR3_2_1	0.18072	0.08491	2.13	0.0341	D_High(t-2)
	AR3_2_2	-0.17446	0.07792	-2.24	0.0259	D_Low(t-2)

Covariances of Innovations

Variable	High	Low
High	5.80640	4.97097
Low	4.97097	9.13591

The last matrix above indicates that the shocks (errors) in the high and low prices have variances 5.8 and 9.1 and a positive correlation. One final interesting question is investigated. Looking at the beta listed above, the estimate of the cointegrating relationship is HIGH-1.04(LOW) which differs a bit from HIGH-LOW. Our H=(1 0, -1 0, 0 1) option

asks for a test that the beta vector can be taken to be $\beta = H\gamma = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} c \\ -c \\ d \end{pmatrix}$ which would mean that

the cointegrating vector can be taken to be (1,-1), scaled by c, as we proposed. Recall the arbitrariness of the scaling factor. The relevant output begins with

Restriction Matrix H
with Respect to Beta

Variable	1	2
High	1.00000	0.00000
Low	-1.00000	0.00000
1	0.00000	1.00000

Long-Run Coefficient
Beta with Respect to
Hypothesis on Beta

Variable	1
High	1.00000
Low	-1.00000
1	-0.12635

Adjustment Coefficient
Alpha with Respect to
Hypothesis on Beta

Variable	1
High	-0.01439
Low	0.04353

The last element of beta and more dramatically, the elements of alpha, have been changed by this small adjustment to the second beta entry. The test that this change is statistically acceptable gives a significant (p=0.0154) lack of fit test, that is, the (1,-1) structure is rejected at the 5% but not at the 1% level. This is an interesting (but not

unprecedented in this author's experience) phenomenon. We rejected unit roots for HIGH-LOW and yet we also reject the hypothesis that the stationary linear combination here has that equal but opposite weights form. This is not likely a result of power as both results were rejections of null hypotheses. The output is below.

Test for Restricted Long-Run Coefficient Beta					
Index	Eigenvalue	Restricted Eigenvalue	DF	Chi-Square	Pr > ChiSq
1	0.1157	0.0992	1	5.87	0.0154

7. Summary.

Two Nobel Prize winning ideas have motivated technology available to SAS users in the ETS package. The first is a family of models related to the ARCH idea in which volatility in a series rather than in its level is of interest. This technology is available in PROC AUTOREG along with other tools such as regression with autocorrelated errors both of which were illustrated here. The second is cointegration, available in PROC VARMAX, in which two nonstationary series are linked together by a "cointegrating vector," a stationary linear combination of the series which, because it is stationary, is a relationship that cannot be too badly violated in a statistical sense.

References:

- Bolerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31:307-327.
- Brocklebank, J. C., and D. A. Dickey (2003). *SAS System for Forecasting Time Series 2 ed.*, SAS Institute, Cary, NC.
- Engle, R. F. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50:987-1007.
- Engle, R. F. and C. W. J. Granger (1987). Cointegration and Error Corrections: Representation, Estimation, and Testing *Econometrica* 55:251-276.
- Johansen, S. (1988). "Statistical Analysis of Cointegrating Vectors." *Journal of Economic Dynamics and Control* 12:312-254.
- Stock, J. H. and M. W. Watson (1988). "Testing for Common Trends." *Journal of the American Statistical Association* 83:1097-1107.

David A. Dickey
dickey@stat.ncsu.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.