Paper 192-30

Stationarity Issues in Time Series Models

David A. Dickey
North Carolina State University

ABSTRACT

The decision on whether analyze a time series in levels or differences is an important aspect of forecasting. Visual methods have been around for a long time. Relatively recently, statistical tests for the null hypothesis that the series is nonstationary, meaning that differencing is required, have been developed. This paper reviews the development of these tests, give motivating examples of why they are needed, demonstrates their use, and shows some other related procedures. The paper will be accessible to beginning level SAS™ programmers, but the reader should have a reasonably strong grounding in statistics.

1. INTRODUCTION

Time series modeling and forecasting became quite popular following the publication of the text Time Series Analysis Forecasting and Control by George Box and Gwilym Jenkins in 1976. In this text they used the symbol ARIMA(p,d,q) to describe a large class of models which could describe the behavior of many observed time series. The d in this symbol indicated the number of times the series needed to be differenced to arrive at a stationary series. I will give the formal definition of stationarity below, but for the moment, you can think of stationarity as a set of conditions that allow estimation of model parameters whose properties are standard, for example t test statistics that have approximately normal distributions in large samples. In their initial work, these authors suggested some properties of time series that would suggest departures from stationarity, that is, these would suggest nonstationarity. The two main properties were visual cues. First, if a series does not seem to have a constant mean (part of the definition of stationarity) when graphed, that is a visible symptom of nonstationarity.

A second symptom had to do with the plot of the lag j correlation against the lag number j. Symbolizing the series value at time t by Y_t, the lag j correlation is the correlation between Y_{t-j} and Y_t. If the plot of the estimated autocorrelations dies down very slowly with increasing j, this provides a second visual symptom of nonstationarity. Using these ideas, Box and Jenkins realized that observed time series often did not appear stationary. However in many of these cases it appeared that the series of first differences $Y_t - Y_{t-1}$ did seem to be stationary. The authors thus suggested modeling the series of changes, or first differences, in these cases. After all, if you can forecast how the series will change a few steps into the future and you know its value now, you can forecast its future values by adding the changes to the current value. Taking differences is a pretty familiar operation. You are used to hearing the first difference in the Dow Jones Industrials Average announced just after the level, for example.

Sections 2 and 3 give a simple introductory model and a motivating example for the test, while 4 and 5 do the same for the general autoregressive model. Section 6 reviews the test results and 7 gives an example. In section 8 results are extended to cover models with trends while 9 and 10 review alternate tests, the normalized bias and F-type tests. Section 11 extends the results from autoregressive models to mixed ARIMA models. Section 12 shows how to check for overdifferencing.

2. SOME MODELS

A simple and yet very useful model for a time series is the model

$$Y_t-\mu = \rho\,(Y_{t-1} - \mu\,) + e_t$$

for all integers t. In this model, μ is the mean or expected value, $E(Y_t)$, of Y_t. The series e_t is assumed to be a series of uncorrelated random variables with mean 0 and constant variance. With these properties, the e_ts

are referred to as "white noise". Often the e_t series is assumed to be normal but that is not required. Independent e_ts rather than just uncorrelated are also sometimes assumed. The e_ts are also referred to as innovations, errors, or shocks. In this model, ρ is the correlation between Y_t and Y_{t-1} while ρ^j is the correlation between Y_t and Y_{t-j}. This model is called an autoregressive order 1 model or AR(1). Notice that if ρ is close to 1, then ρ^j does not die off very quickly as a function of j. Here ρ being the lag 1 correlation implies $|\rho|<1$. With a bit of algebra, one finds that Y_t is a function of the past shocks e_{t-j}. Because the model equation above holds for all t, you see that at time t-1, $Y_{t-1} - \mu = \rho (Y_{t-2} - \mu) + e_{t-1}$ and substituting that into the right side of the original model equation, you find

$$Y_t - \mu = e_t + \rho e_{t-1} + \rho^2 (Y_{t-2} - \mu)$$

and continuing in this fashion you find that

$$Y_t = \mu + e_t + \rho e_{t-1} + \rho^2 e_{t-2} + \ldots.$$

where this infinite weighted sum of past shocks is sometimes referred to as the "Wold" representation of the series. Notice that the sum of the variances on the right side is $\sigma^2(1 + \rho + \rho^2 + \ldots) = \sigma^2/(1- \rho^2)$ where σ^2 is the variance of the shocks. If $\rho=1$ then this sum is infinite suggesting that Y is a random variable with infinite variance. This could not exist. If it did, and you wanted to compute the probability that Y is greater than any number C using, say, the normal distribution, you would divide C-μ by an infinite standard deviation getting 0 as a Z score and ½ as the probability of exceeding any C you could write down. Clearly that is nonsense.

Nevertheless the expression

$$Y_t - \mu = e_t + e_{t-1} + e_{t-2} + \ldots + e_1 + (Y_0 - \mu)$$

would give a reasonable random variable as long as you stipulate that $Y_0 - \mu$ has a finite variance, for example you could assume $Y_0 - \mu$ is 0. In fact this finite sum expression is a quite famous mathematical model called a random walk. In such a model, you note that $Y_t = Y_{t-1} + e_t$, that is, the μ cancels out on both sides and the next Y is the same as the current one plus a random step that is as likely to go up as down. The forecasts, like the series itself, have no tendency to be mean reverting. This model, or a minor variant of it, is often used as a model for stock prices.

In the case of the AR(1) model, the random walk is nonstationary while any series with $|\rho|<1$ (and appropriate starting conditions) is stationary.

Implications for the AR(1) model

The importance of the distinction made above becomes clear when you start thinking about forecasting. For example, if a stock's closing price satisfies

$$Y_t - 100 = 0.8 (Y_{t-1} - 100) + e_t$$

and the last Y you observe is 120, it should not be surprising that the forecast of the next Y is 100 + 0.8(120-100) = 116. Since that is your best estimate of what will happen next, your best estimate for the number after that is, again not surprisingly, 100 + 0.8(116-100) = 114.4 and continuing, you see that each successive forecast is 100 plus 0.8 times the previous deviation from 100. These decline exponentially to the series mean 100 so if you sold when the stock hit 120, you'd be happy to be out of the declining market if it followed your forecast. Likewise had you observed Y=80 as your last value, the forecasted exponential rise to the series mean 100 would signal you to buy that stock.

This stands in stark contrast to the model with $\rho=1$, namely

$$Y_t - 100 = Y_{t-1} - 100 + e_t$$

Here the 100 on each side cancels out and a last observation 120 results in a forecast equal to 120, the most recently observed value. Estimating the historic mean, then buying low and selling high would not be indicated. In fact there really is no constant "mean" for this series.

3. EXAMPLE: DOW JONES INDUSTRIAL AVERAGE (DJIA)

This series is often quoted on newscasts. Further, there are various "index funds" whose portfolio is assembled so as to mimic the behavior of the DJIA. Below you take a look at the plot of this series after a logarithmic transformation, followed by a plot of its first differences. The SAS code is given as well.

```
PROC GPLOT;
PLOT (ldjia ldjia)*date;
SYMBOL1 V=NONE   I=JOIN;
SYMBOL2 V=NONE   I=RL;
TITLE "Dow Jones"; RUN;
```

Notice how the series does not seem to vary about a constant mean whereas the differences do seem to have a relatively constant mean.

Figure 1: Dow Jones

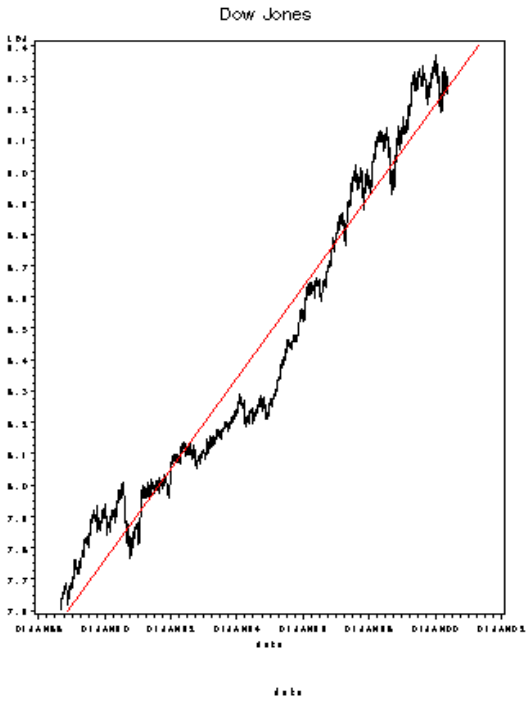
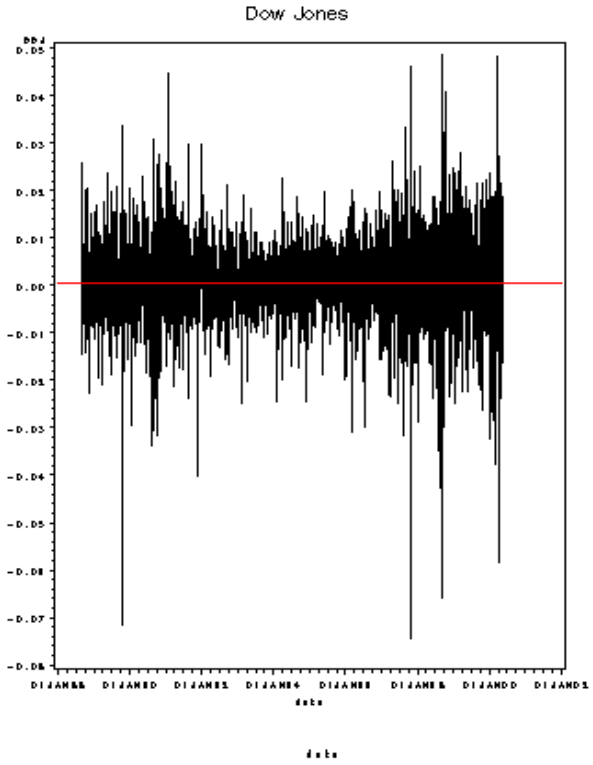


Figure 2: Differences



.

Next the autocorrelation function is produced for 10 lags. The plots are output from the following code.

```
PROC ARIMA;
I var=djia nlag=12;
I var=djia(1) nlag=12;
```

Autocorrelations

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	7612550	1.00000	|********************|
1	7604217	0.99891	|********************|
2	7595529	0.99776	|********************|
3	7586855	0.99662	|********************|
4	7578152	0.99548	|********************|
5	7569481	0.99434	|********************|
6	7560553	0.99317	|********************|
7	7551925	0.99204	|********************|
8	7543869	0.99098	|********************|
9	7535957	0.98994	|********************|
10	7528240	0.98892	|********************|
11	7519890	0.98783	|********************|
12	7511672	0.98675	|********************|

"." marks two standard errors

DIFFERENCES

Autocorrelations

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	4003.285	1.00000	|********************|
1	102.471	0.02560	.|* |
2	-117.368	-.02932	*|. |
3	-235.578	-.05885	*|. |
4	-26.946567	-.00673	.|. |
5	-46.750761	-.01168	.|. |
6	-77.100469	-.01926	.|. |
7	-224.055	-.05597	*|. |
8	-27.874814	-.00696	.|. |
9	132.415	0.03308	.|* |
10	316.534	0.07907	.|** |
11	-254.117	-.06348	*|. |
12	200.979	0.05020	.|* |

"." marks two standard errors

The lines of asterisks graphically represent the autocorrelations at various lags. Note the dramatic difference in behavior between the autocorrelations on the original levels scale and those of the first differences.

The dotted lines represent two standard errors around 0. Spikes in the autocorrelation plot that do not penetrate these bands are not considered statistically significant at the usual 5%level. Taking the first differences has essentially reduced the autocorrelation to 0. A random walk model seems reasonable for this stock index. The model $\log(S_t) = \log(S_{t-1}) + e_t$ where S_t is the stock price is the same as $\log(S_t/S_{t-1}) = e_t$, and because the log of a ratio near 1, for example log(1.01), is about the same as that ratio minus 1 (0.01 in our case) the quantity $\log(S_t/S_{t-1})$ is referred to as the *return* after 1 time period on an investment S_{t-1}. The model implies a random return.

4. THE GENERAL AR(p) MODEL

The model

$$Y_t-\mu = \alpha_1 (Y_{t-1}-\mu) + \alpha_2(Y_{t-2}-\mu) +\ldots+ \alpha_p(Y_{t-p}- \mu) + e_t$$

is referred to as autoregressive of order p. An order 2 example is

$$Y_t – 100 =1.2 (Y_{t-1}-100) – 0.35 (Y_{t-2}-100) + e_t$$

Letting B denote the backshift operator$B(Y_t) = Y_{t-1}$, it is seen that

$$(1-1.2B – 0.35 B^2)(Y_t-100) = e_t,$$

or

$$(1-0.7B)(1-0.5B)(Y_t-100) = e_t.$$

Viewing the operator polynomial as a regular algebraic polynomial, you see it has roots B=10/7 and B=2 both of which exceed 1 in magnitude. This is called the “characteristic polynomial” of the time series. Note also that the model can be written in terms of first differences and a single lagged level

$$Y_t –Y_{t-1} = (-0.15)(Y_{t-1}-100) +0.35 (Y_{t-1}-Y_{t-2}) + e_t$$

and that the polynomial (1-0.7B)(1-0.5B) evaluated at B=1 gives 0.15. It is no coincidence that the lag level coefficient is the negative of the characteristic polynomial evaluated at B=1. This will always happen.

The model

$Y_t - 100 = 1.2\ (Y_{t-1}-100) - 0.20\ (Y_{t-2}-100) + e_t$

tells quite a different story. In backshift notation it becomes

$(1-1.2B - 0.20\ B^2)(Y_t-100) = e_t,$

or

$(1-0.2B)[(1-B)(Y_t-100)] = e_t$

which in turn is

$(1-0.2B)[Y_t - Y_{t-1}] = e_t.$

The characteristic polynomial has a unit root and as a result, the mean drops out of the model and the first difference series satisfies a lag 1 autoregressive model with coefficient 0.20. Again here the model can be written in terms of first differences and a single lagged level

$Y_t - Y_{t-1} = (0)(Y_{t-1}-100) + 0.35(Y_{t-1}-Y_{t-2}) + e_t$

which is just another algebraic expression indicating that the mean term has dropped out.

The general autoregressive model of order 2 becomes

$Y_t - Y_{t-1} = -(1-\alpha_1-\alpha_2)(Y_{t-1}-\mu) + \alpha_2\ (Y_{t-1}-Y_{t-2}) + e_t$

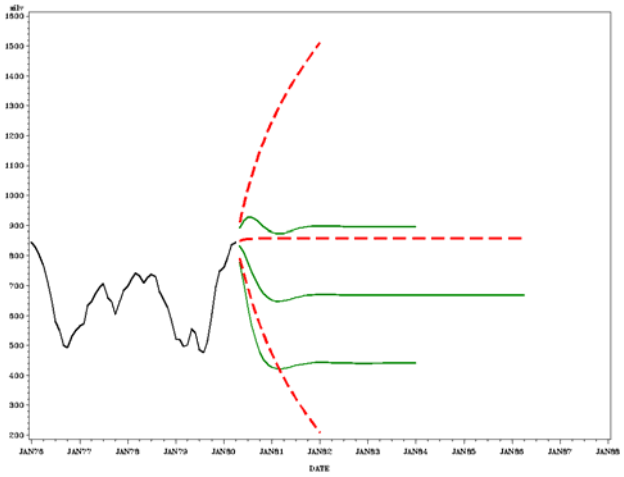
If the characteristic equation $(1-\alpha_1 B-\alpha_2 B^2) = 0$ has a unit root, this means that B=1 is a solution so then $(1-\alpha_1 1-\alpha_2 1^2) = 0$ and the term involving the mean drops out and the series is nonstationary. For ARIMA models, nonstationarity is equivalent to one or more unit roots in the characteristic polynomial.

5. EXAMPLE: SILVER SERIES

The figure below shows the stocks of silver in the New York Commodities Exchange monthly from January 1976 through April 1980. This series seems to satisfy an AR(2) model, however the parameter estimates provide an estimated model with a root of the characteristic equation that is quite close to 1. Extending beyond the data are two forecasts and forecast intervals. The upper forecast with wider bands is from a model with a unit root while the forecast that approaches the historic average value and has narrower bands is from a stationary AR(2) model fit to the data. The differences in long run forecasts are rather dramatic. The fitted AR(2) model, written as above, with t test statistics below the estimates is

$Y_t - Y_{t-1} = -0.11703(Y_{t-1}-\mu) + 0.67115(Y_{t-1}-Y_{t-2}) + e_t$
(t= *- 2.78*) (t = *6.21*)

Figure 3: Stocks of Silver



Recall that if the series has a unit root, the coefficient (-0.11703) on the lagged level is an estimate of 0 and is expected to be negative when the series is stationary. Is our estimate close enough to 0? The t statistic, -2.78, is far enough below 0 to strongly reject the unit root hypothesis *if* you compare it to a t or normal distribution. However, the theory saying that a regression statistic calculated by the t formula has a t distribution rests on some assumptions that are *not* met here. In fact, you will see that a t statistic of -2.78 or less is in fact not so rare when the data have a unit root. It happens by chance more than 5% of the time under that hypothesis and so you cannot reject the unit root null hypothesis.

6. THE MAIN RESULT

Dickey and Fuller (1979, 1981) show that the t statistic as calculated above does not have a t distribution. This does not mean that t does not have any distribution or that it is an invalid test, it simply means that the usual percentiles of the t distribution cannot be used as comparison numbers for the calculated t. The above authors, through some theory and simulations, develop the distribution of the t statistic under the null hypothesis of a unit root and give percentiles for testing. The main results are these

(1) In the lag 1 case, the lag 1 coefficient estimator $\hat{\rho}$ is such that $n(\hat{\rho}-1)$ has a large sample distribution, but it is not standard. They give tables. This is sometimes called the "*normalized bias*" test.

(2) The t statistic for $(\hat{\rho}-1)$, which is the "t-test" on the lag level coefficient in the regression of the first difference on the lag level, is also nonstandard even in large samples. They give tables and, to emphasize that the distribution is not the usual t, they change the name of the statistic from t to τ.

(3) In higher order models (more lags) with a single unit root, the lag 1 coefficient has a distribution different from that in the lag 1 case.

(4) The t statistic for the lag level coefficient has the same limit distribution as that in the lag 1 model.

(5) In higher order models (more than 1 lag) the lagged difference terms have coefficients that behave as in the stationary case, that is, you can trust the usual t tests and their p-values.

(6) Changes in the model will have an effect on the distributions of the normalized bias and τ tests, specifically, different distributions are found depending on whether

(a) μ is known
(b) μ is estimated
(c) μ is replaced by $\alpha+\beta t$ where α and β are estimated (trend model)

Points 1 and 2 imply the need for special tables in doing these tests. Points 3 and 4 tell you that adding lagged differences to your model implies that you must use the t statistic (and its nonstandard distribution) only – not the normalized bias – unless you are prepared to do some additional adjustments.

7. SILVER EXAMPLE REVISITED

Armed with this information, you revisit the Silver series example. You first check to see if augmenting lags are needed in your model. To that end, you regress the differenced series on one lagged level (LSILVER) and 4 lagged differences (DEL1 through DEL4) in PROC REG.

```
PROC REG data=silver;
MODEL del=lsilver del1 del2 del3 del4;
TEST del2=0, del3=0, del4=0;
```

Here is part of the output :

```
Parameter
Variable DF   Estimate t Value  Pr > |t|

Intercept 1  121.02544    3.10    0.0035
LSILVER   1   -0.18796   -3.07    0.0038
DEL1      1    0.63923    4.59    <.0001
DEL2      1    0.04992    0.30    0.7691
DEL3      1 0.00030885    0.00    0.9985
DEL4      1    0.26323    1.72    0.0924
```

Can you just use the p-values as given? Recall that for the lagged differences (DEL variables) you can trust the p-values in the sense that they are justified in large samples. The t tests seem to indicate that no lagged differences are needed, but recall that joint decisions such as this cannot be made by looking at the t statistics. Because the lag difference estimates have, in large samples, the usual multivariate normal distributions, standard tests such as F tests also have the standard distributional properties. Thus it is only tests involving the intercept, trend, and/or LSILVER variable that require special tables for testing.

The implication of the above is that the F test requested in the TEST statement can be trusted, again meaning by this expression that there is some theoretical justification, at least in large samples, of the fact that these test statistics have the usual distributions under the null hypothesis.

```
The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable DEL

                 Mean
Source  DF     Square  F Value   Pr > F

Num      3 1152.19711     1.32   0.2803
Den     41  871.51780
```

The output from the TEST statement shows that DEL2 through DEL 4 are unnecessary. One might consider this as a preliminary test, to determine which test should be used to test the hypothesis of interest. Some practitioners like to use a more liberal level, like 0.25, for such preliminary tests. Even so, there is no evidence of a need for lagged differences at lags 2 through 4.

Refitting using

```
PROC REG data=silver;
MODEL del=silver del1;
```

You now obtain this output.

```
          Parameter Estimates

             Parameter
Variable  DF  Estimate  t Value Pr > |t|

Intercept 1   75.58073     2.76   0.0082
LSILVER   1   -0.11703    -2.78   0.0079
DEL1      1    0.67115     6.21   <.0001
```

Notice the rather extreme change in the p-value for DEL1 now that the other lagged differences have been removed. This underlines the point made earlier that t-test statistics cannot be considered together to give a joint inference. Once one of the variables is removed, the t test for the others will usually change as happened here. Of course the p-value for LSILVER, the lagged level of silver, has changed as well. This is of no interest here, having been calculated based on a t distribution rather than the appropriate τ distribution.

Now that you seem to have a reasonable model for the data, you turn to the test of main interest. Recall that if the series should be differenced, the theoretical coefficient on LSILVER would be 0, as it is the characteristic polynomial evaluated at 1. A zero value here implies a unit root, but the distribution is nonstandard so that p-values computed in PROC REG would not be appropriate for this particular test. Had the p-value 0.0079 been correct, it would rather strongly suggest that the series is stationary and no differencing is needed. This then would imply that the mean reverting forecast in Figure 3 would suffice.

One option is to find a table of critical values for this test statistic. Fuller (1996) gives such a table for τ_μ indicating that τ_μ would have to be less than -2.86 to achieve significance at the 5% level. The subscript μ is used to indicate that the mean (intercept) has been estimated. The calculated τ_μ of -2.78 does not meet that criterion and hence the correct test does not suggest stationarity and thus does suggest a need for differencing and the corollary implication that the mean reverting forecast is not justified. In this particular example, knowledge of the correct distributional results seems quite important.

Because stationarity is an issue of some importance in time series analysis, the tests described above have been incorporated in SAS PROC ARIMA as well as in the SAS Time Series Forecasting System accessible using the Tools option in the program editor menu bar. Decisions must be made about the number of lagged differences to incorporate so, for a careful analysis, calculations such as those shown above can be quite useful.

Some PROC ARIMA code to compute the tests is as follows.

```
PROC ARIMA;
IDENTIFY VAR=SILVER
STATIOANRITY=(ADF=(1,4));
```

Partial output from this code follows:

```
Augmented Dickey-Fuller Unit Root Tests

Type      Lags       Tau   Pr < Tau

Single Mean  1     -2.78     0.0689
             4     -3.07     0.0356
```

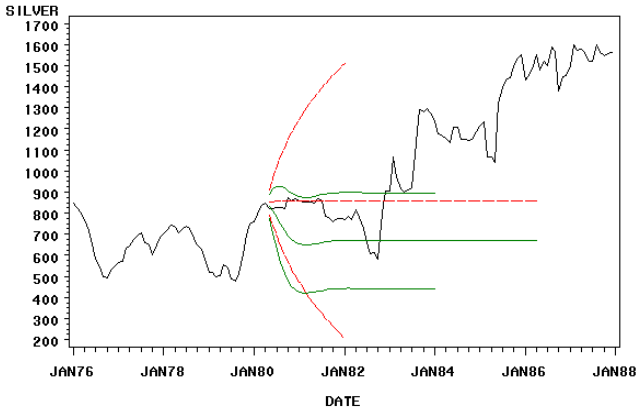
Notice that the Tau test statistics have the same values as the corresponding t statistics from PROC REG. Unlike that output, however, these statistics have the correct p-values so without recourse to a table of critical values you see that the data, under the model you decided on, do not quite give strong enough evidence to reject unit roots, that is, the series is not shown to be stationary. The reason for using the letter τ rather than t is to emphasize that the t distribution is not appropriate for computing p-values.

A down side to this example is that one would expect a test in the context of a model with extraneous parameters to have less power to detect the alternative than one with a good fitting model. However in our case it is the model with seemingly extraneous parameters, the lag 4 model, that has p-value less than 0.05.

This may in part be due to the loss of observations by lagging or because the sample size is really not extremely large and the expected large sample behavior is thus not realized.

The silver series shown in Figure 6 was used in the first edition of SAS System for Forecasting Times Series (1981) by Brocklebank and myself. In a recent revision, subsequent data were added to the series. Figure 7 gives the extended data and from this it is quite clear that using a forecast based on the stationary model, the one that would erroneously have been chosen had you used the p-values coming from PROC REG, would be a mistake.

Figure 7: Extended Silver Data



8. EXTENSION: TRENDING DATA

Notice that the PROC ARIMA unit root test was labeled “Single Mean”. That is because your model, under the alternative, has the single mean μ. Two variations on this test are available. In the first, the “no mean” model, it is assumed that the series mean is 0 whether or not there are unit roots. Of course anytime the mean is known rather than estimated, you can assume a 0 mean by simply subtracting this known mean from all observations to get a mean 0 data set.

In the single mean case, the mean drops out under the unit root null hypothesis but not under the alternative hypothesis. Imagine a time series taking values between 80 and 110. Figure 8 shows such data where the time series is a mean value, 100, plus a white noise error term, that is, the series is not only stationary, the lag correlation is 0. Figure 8 is a graph of Y_t versus Y_{t-1}. The fitted regression line with no intercept is also shown. The “no intercept” option forces the line through the origin so that its slope clearly must be quite close to 1 to come anywhere near the observed data.

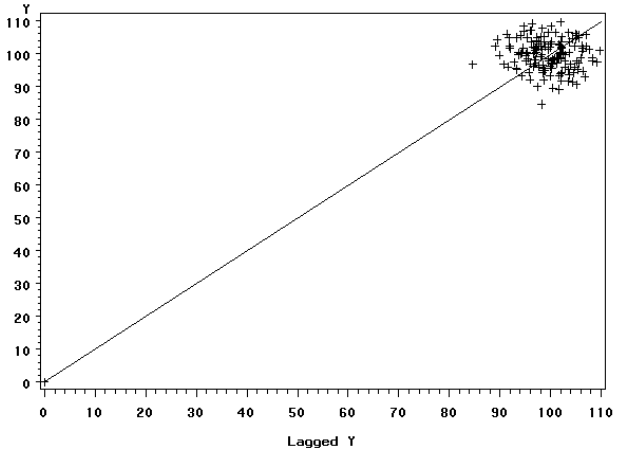
Just before Figure 8 you see the result of regressing the first difference on the lag level of Y with the NOINT option in PROC REG. The τ test is –0.33, nowhere near significance and thus if the mean is (incorrectly) assumed to be 0, the conclusion is exactly what is suggested by Figure 8, that is, a unit root.

```
                Parameter Estimates

                  Parameter
Variable    DF    Estimate t Value Pr > |t|

Y1           1    -0.00188   -0.33   0.7382
```

Figure 8: Mean plus White Noise.



The model incorrectly assumes a mean 0 so that the unit root test is asking whether forecasting the next value with the current one is as good as moving toward the assumed mean 0. The moral of this is that one should not fit the zero mean model unless there is a-priori reason to think that the mean is 0 whether or not there are unit roots. An example in which this might be appropriate is testing for an additional unit root in data that have already been differenced.

In addition to the zero mean and single mean models, a model is needed for data like those shown in Figure 1. These data are clearly not stationary around any single mean, 0 or otherwise. The data exhibit a clear upward trend. Rather than stationary deviations from a mean, the only reasonable alternative involving stationary deviations would be stationary deviations around a trend. If data are generated as a trend line plus white noise, and if the slope of that trend is not 0, the tests discussed previously will not be able to reject unit roots even though the deviations from trend are white noise.

In the previous two cases, deviations from a possibly 0 mean were studied and, under the unit root null hypothesis, that mean disappeared. In the case of a trend of the form $Y_t = \alpha + \beta t + Z_t$ where Z_t is a zero mean time series, the first differences $Y_t - Y_{t-1}$ are seen to have the form $\alpha + \beta t - (\alpha + \beta(t-1)) + Z_t - Z_{t-1} = \beta + Z_t - Z_{t-1}$. If the Z_t series were stationary, a regression with time series errors as can be fit in PROC ARIMA or PROC AUTOREG would be appropriate. Differencing would not be appropriate as it would induce a unit root moving average term. On the other hand if Z_t is a unit root process then differencing is appropriate so that $Y_t = Y_{t-1} + \beta$ plus the stationary time series error term $Z_t - Z_{t-1}$. This is called a unit root process with drift β.

The series Y_t in that case is seen to be a series whose first differences average out to β in the long run. Thus β represents a time trend, whether as a slope in the stationary deviations case or as a drift in the unit root case.

At this point you see that there are three regression cases to consider. The cases can be laid out in a table

Table 1: Regression Estimators

Case	Regression
No mean, H_0 or H_1	$Y_t - Y_{t-1}$ on t, Y_{t-1}, and lagged differences (NOINT)
Mean under H_1	$Y_t - Y_{t-1}$ on t, Y_{t-1}, and lagged differences
Trend under H_1	$Y_t - Y_{t-1}$ on t, Y_{t-1}, and lagged differences

In each case it is the coefficient of Y_{t-1} that indicates whether differencing needs to be done with a 0 coefficient indicating differencing. In each case, the usual t formula applies but the distribution is nonstandard. The remaining question is whether the nonstandard distribution is the same in each case. Unfortunately the answer is no, but fortunately the distributional results needed are available. For the SAS user, it is enough to know that the test computations in PROC ARIMA use the proper distributions so their p-values can be trusted.

9. NORMALIZED BIAS TEST

Thus far you have used the t statistic in the regression of $Y_t - Y_{t-1}$ on Y_{t-1} and an appropriate distribution to formulate a test called "tau" for unit roots. The model was written as

$$Y_\tau - \mu = \rho\,(Y_{\tau-1} - \mu) + e_\tau$$

or, as used in constructing the t test,

$$Y_t - Y_{t-1} = (\rho - 1)\,(Y_{t-1} - \mu) + e_\tau.$$

Dickey and Fuller (1979) give tables not only for the tau statistics, but also for the so called "normalized bias" test $n(\hat{\rho} - 1)$. Dickey (1984) shows that the power of this test in the lag 1 model is a bit higher than that of tau, but this test does not extend as well to higher order models and thus has not achieved the popularity of the tau tests, as was mentioned earlier. Nevertheless, SAS PROC ARIMA provides this kind of test as well and it will be described in this section.

Under the null hypothesis of a unit root, the autoregressive backshift operator has a factor (1-B) as was seen in an earlier example where $(1 - 1.2B - 0.20\,B^2)$ factored as $(1 - 0.2B)(1 - B)$. Recall that the coefficients on lag Y in the regressions of Table 1 were estimates of the backshift operator with 1 substituted for B and thus were estimates of 0 under the unit root null hypothesis.

In anticipation of what is about to be seen, recall that in a regression of people's weight on height in inches, the conversion of height from inches to feet leaves the t test as it was before, but the coefficient on height is now 12 times its former value.

In a unit root regression with no lagged differences the series is represented as

$$Y_t - Y_{t-1} = (1 - \rho)(Y_{t-1} - \mu) + e_t.$$

Taking the AR(2) model as an example, the higher order model

$$(1 - \alpha_1 B - \alpha_2 B^2)\,(Y_t - \mu) = e_t$$

can be rewritten in factored form as

$$(1 - \rho B)(1 - \gamma B)\,(Y_t - \mu) = e_t$$

where the backshift operator has been factored and our interest will be in testing $H_0{:}\rho = 1$ assuming that $|\gamma| < 1$.

Another algebraic representation of the series, relevant to the unit root regression, is

$$(Y_t - Y_{t-1}) = -(1 - \rho)(1 - \gamma)\,(Y_{t-1} - \mu) + \rho\gamma\,(Y_{t-1} - Y_{t-2}) + e_t.$$

Writing out the coefficients of Y at lag 1 and 2 in the two preceding expression will allow the interested reader to check the equivalence of these expressions.

The coefficient of lagged Y in the lag 1 regression is $-(1 - \rho)$ and in the lag 2 regression it is $-(1 - \rho)(1 - \gamma)$. It thus is not surprising, in light of the height and weight regression discussed earlier, that the large sample distribution of the τ test statistic is not affected by the presence of the $(1 - \gamma)$ multiplier whereas the normalized bias test now involves an estimate of $n(\rho - 1)(1 - \gamma)$ rather than $n(\rho - 1)$.

This suggests that the normalized bias should be divided by an estimate of $(1-\gamma)$ in order to use the normalized bias distribution that was developed for the lag 1 model. Indeed, it can be proved that such an adjustment allows the use of this distribution in higher order models.

Returning to the first example of this paper, the Dow Jones Industrials Average series, the PROC ARIMA commands

```
PROC ARIMA data=dow;
i var=DJIA nlag=12 stationarity=(ADF=(0,1));
```

result in output that includes the tests just described.

```
Augmented Dickey-Fuller Unit Root Tests

Type             Lags        Rho    Pr < Rho

Zero Mean           0     1.3567      0.9565
                    1     1.3481      0.9557
Single Mean         0     0.4065      0.9744
                    1     0.3500      0.9725
Trend               0    -6.3073      0.7203
                    1    -6.5833      0.6981
```

Previous analysis of this data suggested that a lag 1 model would suffice so the lines associated with 0 lags (that is, no lagged differences in the model) are appropriate. The Zero Mean and Single Mean types of tests are not relevant here because of the apparent trend in the data.

The normalized bias test, called “Rho” in the output, is -6.3073 for the model with no lagged differences and -6.5833 for the model with one lagged difference. In that second case, the adjustment described earlier has been applied. The output included the τ tests as well (not shown) and these gave the same conclusion. With the p-value 0.7203, that conclusion is that there is a unit root in the trend residuals. You might say that the data are not hugging that trend line tightly enough to be considered stationary.

Is it possible that the differenced series needs to be differenced once again? To check this out, a unit root test on the differenced data can easily be accomplished. Using the specification DJIA(1) to indicate first differenced data used as input, you submit the code

```
PROC ARIMA data=dow;
i var=DJIA(1) nlag=12 stationarity = (ADF=0);
```

All of the printed tests strongly reject a unit root in the already differenced series. If the original data had a linear trend, the differences would have a non-zero mean indicating that the “Single Mean” test would be of interest. The output follows:

```
The ARIMA Procedure

Dickey-Fuller Unit Root Tests

Type          Lags       Tau  Pr < Tau
Zero Mean        0    -52.90    <.0001
Single Mean      0    -53.00    <.0001
Trend            0    -53.01    <.0001
```

An interesting technical point here is that if the test on the differenced series indicates a unit root this means the original series had two (or more) unit roots. Because all of the tests discussed here are developed under the assumption of at most a single unit root, this means that the distribution used for the test on the original series did not use the correct p-values. Sen and Dickey (1985) and the references therein address the magnitude of this problem and Dickey and Pantula (2002) point out that the test for a unit root in the differenced series should be done first as a validation for the test on the original levels. Of course if both tests are done, as was done here, the order in which they are done is not an issue.

10. F-TYPE TESTS

One last collection of tests appears in unit root testing output. Using the model

$$Y_t - Y_{t-1} = (1-\rho)(Y_{t-1} - \mu) + e_t$$

as a basis for discussion, note that the model can be re-expressed as

$$Y_t - Y_{t-1} = (\rho-1)\mu + (1-\rho)Y_{t-1} + e_t$$

which presents the model in a form with an intercept and a slope. Because this has been carefully developed from the lag 1 model, it is clear that both the intercept and slope will be driven to 0 under the null hypothesis that $\rho=1$. This suggests yet another form of test in which the null hypothesis reduced model

$$Y_t - Y_{t-1} = (\rho-1)\mu + (1-\rho)Y_{t-1} + e_t$$

is compared to the full model

$$Y_t - Y_{t-1} = \lambda + \delta Y_{t-1} + e_t$$

where $\lambda = (\rho-1)\mu$ and $\delta=(1-\rho)$. This is calculated in the usual way as an F test whose numerator sum of squares is the difference in error sums of squares for the two models and whose denominator is the error mean square from the full model. The numerator degrees of freedom here is 2. As before, even though the statistic is calculated by the usual F formula, the assumptions on which the F distribution rests are not met in unit root processes so the F statistic does not have the standard F distribution. Dickey and Fuller (1981) compute the distribution of this test as well and Dickey (1984) shows that if the goal is to test for a unit root, the power of the τ_μ test exceeds that of this F.

Nevertheless PROC ARIMA gives this test as part of the output as well. In the case of trending data, the F test reported compares the random walk with drift as a reduced model to the full model having stationary deviations around a linear trend. Again here the two models differ by two degrees of freedom.

```
Augmented Dickey-Fuller Unit Root Tests

Type          Lags          F      Pr > F

Zero Mean        0
                 1
Single Mean      0       3.02      0.2958
                 1       2.87      0.3350
Trend            0       2.30      0.7147
                 1       2.31      0.7131
```

The reason that no F test is given for the zero mean case is that F is just the square of τ in that case. All tests shown thus far, the τ tests, the normalized bias tests, and these F tests, agree in their assessment that there is a unit root in this series.

Most economists argue that in a stock market with intelligent traders and equal access to the market, stock prices would have to follow a unit root process. If there were a tendency to return to some linear trend function or constant mean, traders would realize that and bid the prices up or down in such a way as to remove that predictability. The results here are consistent with that idea.

11. ARIMA (p,1,q) MODELS

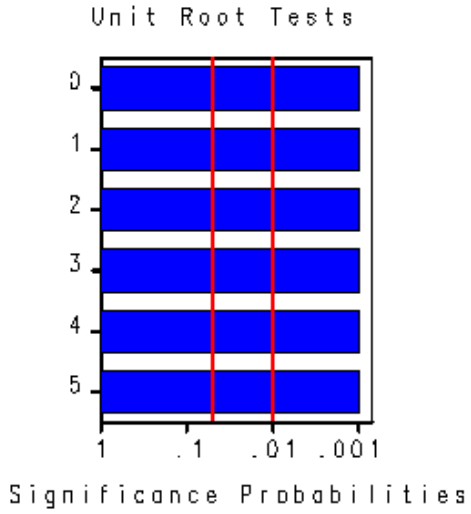
Not all time series are modeled as autoregressive forms. A more general model, expressed using backshift operators, is the autoregressive moving average model. If differencing is required to achieve stationarity, the model is called an autoregressive integrated moving average or ARIMA(p,d,q) where d denotes the amount of differencing required, p denotes the number of autoregressive lags, and q the number of moving average lags. Using $\pi(B) = (1-\alpha_1 B - \ldots - \alpha_p B^p)$ and $\theta(B) = (1-\theta_1 B - \ldots - \theta_q B^q)$ the model becomes and the process has a unit root if $\pi(B)$ has a factor (1-B). The model is $\pi(B)(Y_t-\mu)= \theta(B)e_t$.

Said and Dickey (1985) give a way of estimating an ARIMA model using nonlinear least squares such that a test of unit roots can be done using existing tables in large samples. The estimation is somewhat complicated and has not been much used in practice.

An approximate test in which the ARIMA series is analyzed as though it were just autoregressive is justified in Said and Dickey (1984) under the condition that the sample size gets large and as it does, the number of lagged differences in the unit root regression increases as well. The practical implication of this is that all unit root testing can be done as detailed above, assuming that the underlying series is autoregressive, a very useful result for practitioners. It is assumed that the moving average operator $\theta(B)$ has no unit roots.

Hall (1994) studies various methods of choosing the number of augmenting lagged differences in practice. His conclusion was that overfitting, then using standard t or F tests to eliminate insignificant lagged differences prior to the unit root test, was a reasonable way to approach this task. This is the tactic used earlier in the silver series example.

These unit root tests are also available in the SAS/ETS time series forecasting system. A series of bars is used, as an example these bars appear for the differenced Dow Jones series. Each test is a test for a single unit root in the model. The numbers down the side refer to the number of augmenting lagged differences in the model. For this data, based on regression F tests like those used in the silver series, it appears that no lagged differences are necessary. Note that the p-values, listed below the graph, are on a descending logarithmic scale. A bar extending, as all of these do, to the right of the .01 mark indicates rejection of the unit root null hypothesis at the 0.01 level of significance. Clicking on a bar produces a pop-up window that states this conclusion and its implication for further differencing. In this Dow Jones case, no more differencing is required.



12. INVERSE AUTOCORRELATIONS

SAS PROC ARIMA provides three identifying plots for time series the autocorrelation function (ACF) the partial autocorrelation function (PACF) and the inverse autocorrelation function (IACF). Prior to the development of formal unit root tests, nonstationarity was diagnosed when the autocorrelation sequence died down very slowly as shown just after Figure 2.

The inverse autocorrelation function is the autocorrelation function of the model obtained by switching the autoregressive and moving average backshift operators. The inverse autocorrelation function for the model $Y_t = e_t - .8\ e_{t-1}$ is thus the autocorrelation function of the model $Y_t = .8\ Y_{t-1} + e_t$, that is, the inverse autocorrelation function dies off exponentially at a rate 0.8. Estimators of the inverse autocorrelation function in SAS PROC ARIMA are computed by fitting a long autoregressive approximation to the data, moving the autoregressive coefficients over to the moving average side then, treating this estimated moving average as though it were a true moving average model, then algebraically calculating the autocorrelations for it. Chang and Dickey (1993) show that this method of calculation, when applied to the first differences of

a stationary series, gives a series of slowly decaying estimated inverse autocorrelations. The predicted decay is toward 0 at the maximum lag level so that slow decay in the inverse autocorrelation of an overdifferenced series should look somewhat like a right triangle as opposed to the rectangular appearance in the autocorrelations of a unit root process. Thus if you mistakenly difference an already stationary series, a slowly decaying, roughly triangular shaped inverse autocorrelation function should result. This provides a test for overdifferencing.

For example, suppose you observe a time series that is increasing, like the Dow Jones Average in Figure 1. It could be a random walk with drift or it might be a linear trend with stationary residuals. In the second case, the differenced series would have a constant, or drift term, equal to the slope in the original plot but the error term in the differenced series would be a differenced stationary process, hence having a unit root on the moving average side of the model. A slowly decaying inverse autocorrelation, then, would suggest that differencing was inappropriate. Of course without running a regression to eliminate the trend, the autocorrelations on the original levels would also die off slowly, suggesting the differencing that you did in the first place. The presence of that trend in the original data is what is causing the slowly decaying autocorrelations in the original levels, not a unit root.

Here are the inverse autocorrelations for the differenced Dow Jones data. Because this function drops immediately to 0, it appears that differencing was appropriate.

Inverse Autocorrelations

Lag	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1	-0.51119	**********|. |
2	0.01380	.|. |
3	-0.00533	.|. |
4	0.01061	.|. |
5	-0.02324	.|. |
6	0.00722	.|. |
7	0.02122	.|. |
8	-0.01617	.|. |
9	0.02831	.|* |
10	-0.04860	*|. |
11	0.02759	.|* |
12	-0.00422	.|. |

Just for illustration, here is the inverse autocorrelation you would get if you differenced these differences once more, that is, if you took the second difference of the original series. Note the roughly triangular appearance, suggesting that you should have stopped after the first difference

Inverse Autocorrelations

Lag	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1	0.89720	.|****************** |
2	0.80302	.|**************** |
3	0.70785	.|************** |
4	0.60466	.|************ |
5	0.50498	.|********** |
6	0.41173	.|******** |
7	0.32523	.|******* |
8	0.23836	.|***** |
9	0.15871	.|*** |
10	0.09447	.|** |
11	0.05758	.|* |
12	0.01735	.|. |

™: 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 registered trademarks or trademarks of their respective companies.

References:

Box, G. E. P. and Jenkins, G. M. (1976) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Brocklebank, J. C. and Dickey, D. A. (1981, 2004) SAS System for Forecasting Time Series 1st & 2nd ed.

Chang, M.C. and Dickey, D. A. (1993). "Recognizing Overdifferenced Time Series", *Journal of Time Series Analysis,* **15**, 1-18.

Dickey, D. A. and Fuller, W. A. (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root". *Journal of the American Statistical Association*, **74**, p. 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.

Dickey, D. A. (1984). "Power of Unit Root Tests," *Proceedings of Business and Economic Statistics Section*, American Statistical Assn., 489-493.

Dickey, D. A., and Pantula, S. G. (2002). "Determining the Order of Differencing in AR Processes." *Journal of Business Economics Statistics,* **20**: 18-24.

Hall, A. (1994). "Testing for a Unit Root with Pretest Based Model Selection," *Journal of Business Economics Statistics,* **12**: 461-470.

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

Said, S. E. and Dickey, D. A. (1985). "Hypothesis Testing in ARIMA (p, 1, q) Models", *Journal of the American Statistical Association,* **80**, 369-374.

Sen, D.L. and Dickey, D. A. (1987). "Symmetric Test for Second Differencing in Univariate Time Series", *Journal of Business and Economic Statistics,* **5**, 463-473.

Contact Information:

David A. Dickey
Department of Statistics, Box 8203
N.C. State University
Raleigh, N.C. 27695-8203
dickey@stat.ncsu.edu