

## *A Review of Unit Root Testing Results*

**David A. Dickey**

*North Carolina State University*

**Abstract:** This paper is an introduction to unit root testing. This is an area well known to advanced econometricians, and it is not my purpose here to extend the frontiers of knowledge further for this group. Instead, I present an overview, intended for a more general statistical audience. With that in mind, I define the problem, explain why it is of practical interest especially in economic applications, and then review some of the findings that my students and I, as well as others, have made over the past several years. The ideas are illustrated with examples.

**Keywords:** Multivariate analysis, discriminant analysis, supervised classification, region of doubt, detection of outliers, robust estimators, clustering methods, mixture models.

### 1 Introduction

The unit root problem is introduced in statistical and economic contexts in sections 2 and 3. Two examples are introduced in section 4 and the models for testing are introduced in section 5 along with a brief overview of some distributional results. In section 6, we return to analyze the two example series mentioning some extensions of the results in the process. Section 7 gives a third example and 8 briefly reviews a small number of the many research articles that have appeared since the early work on unit roots.

### 2 The statistical context

The class of ARIMA (Autoregressive Integrated Moving Average) models has been found especially useful for explaining the correlation structure of data taken over time and for forecasting. The ARIMA models serve well in forecasting future values based on correlations with their own past. Standard statistical results, such as normal distributions of parameter estimates and test statistics in large samples, have been shown under a condition known as stationarity. Stationarity will be formally defined after a bit more notation is developed. Under the stationarity assumption, an ARIMA model expresses the deviations of some observed series  $Y_t$  from a mean that is constant over time. These deviations  $Y_t - \mu$  are related to past deviations  $Y_{t-j} - \mu$  by a so-called "difference equation," an example of which is

$$Y_t - \mu = \alpha_1 (Y_{t-1} - \mu) + \alpha_2 (Y_{t-2} - \mu) + \alpha_3 (Y_{t-3} - \mu) + \dots + \alpha_p (Y_{t-p} - \mu) + e_t.$$

This particular example is an autoregressive model of order  $p$ , AR( $p$ ). The  $e_t$  are often called "shocks" or "innovations" in economic discussions, and are assumed to be reasonably well behaved, (independent identically distributed with finite moments, for example). Under these or even milder conditions, along with the stationarity assumption, nice large sample results are available which can be summarized by stating

that all of the usual least squares statistics and resulting inferences from a regression of  $Y_t$  on  $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_{t-p}$  with an intercept, are justified in large samples. For example in this regression the overall model F test is an asymptotically valid test of the null hypothesis that no lagged  $Y$  values help in predicting the current  $Y$ . Note that the intercept in the model above is the result of combining all of the  $\mu$  terms, namely  $\mu(1 - \alpha_1 - \alpha_2 - \alpha_3 - \dots - \alpha_p)$ .

When  $e_t$  is replaced by a moving average, say  $e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$ , the former autoregressive model becomes an autoregressive moving average of orders  $p$  and  $q$ , ARMA( $p, q$ ). This and the AR( $p$ ) model have a so-called "characteristic equation" given by

$$m^p - \alpha_1 m^{p-1} - \alpha_2 m^{p-2} - \alpha_3 m^{p-3} - \dots - \alpha_p = 0.$$

It is when the roots  $m$  of this equation all satisfy  $|m| < 1$  that the model is said to satisfy the stationarity condition. In this case, it is possible to express  $Y_t - \mu$  as a convergent infinitely weighted sum of current and past shocks  $e_t$ . From this it follows that the expected value of  $Y_t$  is indeed  $\mu$ , a constant over time, and the covariance between  $Y_t$  and  $Y_{t-h}$  depends only on  $h$ , that is, this covariance is a constant function of time  $t$ . The formal definition of stationarity (technically covariance or second order stationarity) is that the mean is constant and the covariance is a function of  $h$  only, not  $t$ . The covariance at lag  $h = 0$  is the variance which thus must also be constant over time. Within the ARIMA class of models, it is enough to check the roots  $m$  of the characteristic equation to determine stationarity. Of course the statistically challenging problem is that the true parameters themselves are unknown so that we have only an estimate of the characteristic equation.

### 3 The economic context

It is typical on the evening news or in a financial report to see the value of some stock index  $Y_t$ , for example the famous Dow-Jones average in the U.S. Usually this is followed by an announcement of the change  $Y_t - Y_{t-1}$  from the previous period. Why are the changes reported? Often these are easier to comprehend than the actual level  $Y_t$  of the series as  $Y_t$  seems to wander quite a bit over time. For most people it is easier to judge whether a gain or drop is unusual than to make such a judgment about the level. Furthermore, there is reason to expect stock prices to follow a random walk model,  $Y_t = Y_{t-1} + e_t$ . This model says that the level  $Y_t$  is the same as yesterday's plus a random shock  $e_t$  uncorrelated with, and thus unpredictable from, the past.

This is where things get interesting from an econometrics point of view. The AR(1) model  $Y_t = \alpha_1(Y_{t-1} - \mu) + e_t$  actually becomes a random walk when  $\alpha_1 = 1$  and we note that when this happens  $\mu$  drops out of the model, that is, it is not identifiable. The best forecast of the future in a random walk, is just the current value and the forecast error grows without bound as we forecast further into the future. In contrast, if  $|\alpha_1| < 1$  then the  $j$  step ahead forecast, that is, the forecast for time  $t + j$  based on observations  $Y$  up through time  $t$ , is  $\mu + (\alpha_1)^j(Y_t - \mu)$ . The forecast converges exponentially fast to the historic mean. The forecast error grows with  $j$ , but approaches a limit given by the series variance. The characteristic equation is  $m - \alpha_1 = 0$  with root easily seen to be  $m = \alpha_1$ . Thus the random walk has  $\alpha_1 = 1$  and so violates the stationarity condition while  $|\alpha_1| < 1$  gives stationarity. Forecasts from a stationary model approach the historic mean so that a stock that is priced over the mean will tend to go down and one priced under the mean will tend to go up. Thus an investor could make money by buying low and selling high if his stock were stationary. A test of the null hypothesis that the series is a random walk, or in general a "unit root" process, is of interest. Notice that when  $\alpha_1 = 1$  in the AR(1) model, the first difference series  $Y_t - Y_{t-1}$  is just  $e_t$  and hence is stationary. In the general notation ARIMA(p,d,q), the d refers to the multiplicity of (number of) unit roots or equivalently, the number of first differences required to render the series stationary. The notation ARMA refers to a case with  $d = 0$ .

As it happens there are many series besides stock prices whose stationarity status is of interest in economics. Consider, for example, the ratio of long to short term interest rates, perhaps on a logarithmic scale. Is this ratio more or less constant, that is, is the ratio at least stationary? If not, then the ratio does not tend to fluctuate around any mean and can wander arbitrarily far from any given level. Similarly, we could look at the ratio of prices in two different countries, perhaps dividing this ratio by the exchange rate, then test to see if the relationship is stable (stationary). Another ratio of interest would be wages in two similar industries. On the logarithmic scale we are asking if  $Y_t = \log(wage1) - \log(wage2)$  forms a stationary time series. If we additionally allow a coefficient  $b$  on  $\log(wage2)$  the combination of estimating  $b$  and testing the resulting difference for stationarity is called "cointegration analysis" and is a currently popular topic in econometrics. Thus if  $b$  is known, a unit root test becomes a test for cointegration.

It is always possible to represent the autoregressive order  $p$  part of an ARIMA model in terms of  $p-1$  differences and a single level at some lag, usually at lag 1. For example, the models

$$\begin{aligned} Y_t - \mu &= 1.5(Y_{t-1} - \mu) - 0.56(Y_{t-2} - \mu) + e_t \\ Y_t - \mu &= 1.5(Y_{t-1} - \mu) - 0.50(Y_{t-2} - \mu) + e_t \\ Y_t - \mu &= \alpha_1(Y_{t-1} - \mu) + \alpha_2(Y_{t-2} - \mu) + e_t \end{aligned}$$

become

$$\begin{aligned} Y_t - Y_{t-1} &= -0.06(Y_{t-1} - \mu) + 0.56(Y_{t-1} - Y_{t-2}) + e_t \\ Y_t - Y_{t-1} &= 0(Y_{t-1} - \mu) + 0.50(Y_{t-1} - Y_{t-2}) + e_t \\ Y_t - Y_{t-1} &= -(1 - \alpha_1 - \alpha_2)(Y_{t-1} - \mu) - \alpha_2(Y_{t-1} - Y_{t-2}) + e_t \end{aligned}$$

The first of these three involves a mean and the forecasts from this model tend to return to the mean. The second simplifies to a function only of differences and the mean disappears. Like the random walk, the forecasts have no tendency to return to any estimable "mean". The roots of the first characteristic equation are 0.7 and 0.8, and the second has roots 1 and 0.5 where the unit root 1 is the key to the mean dropping out of the model. The third (general) model shows that the coefficient of the lag level term is in fact the negative of the characteristic equation evaluated at 1. Thus a root of 1 implies that the term involving the mean has a coefficient 0. Further this general form suggests a testing strategy: regress the differences on a lagged level term and enough additional lagged differences to get the error term to an uncorrelated sequence (known as "white noise"). Test the null hypothesis that the lag level coefficient is 0.

We have a set of models that address questions of economic interest. We have a way, least squares, to estimate those model coefficients. If the distributions of those coefficients possessed the usual normal distributions, we would be done. The problem here is that our null hypothesis model, the unit root model, does not give normally distributed estimates, even in the limit. Our purpose here is to show how to use these and to say a little about some extensions and improvements on least squares.

#### 4 Two motivating examples

In the left panel of figure 1, we see stocks of silver in the New York commodities exchange, along with two forecasts. The dotted lines with wide bands represent forecasts from a nonstationary model and the solid ones are from a stationary model. Both include forecast error bands. Data prior to the forecast were used to illustrate unit root tests in the first edition (1986) of Brocklebank and Dickey and are the source of the two predicting models. Data through the forecast period are given in the second edition (2003) and it seems clear at this point that the dotted lines are more appropriate. Will the test agree?

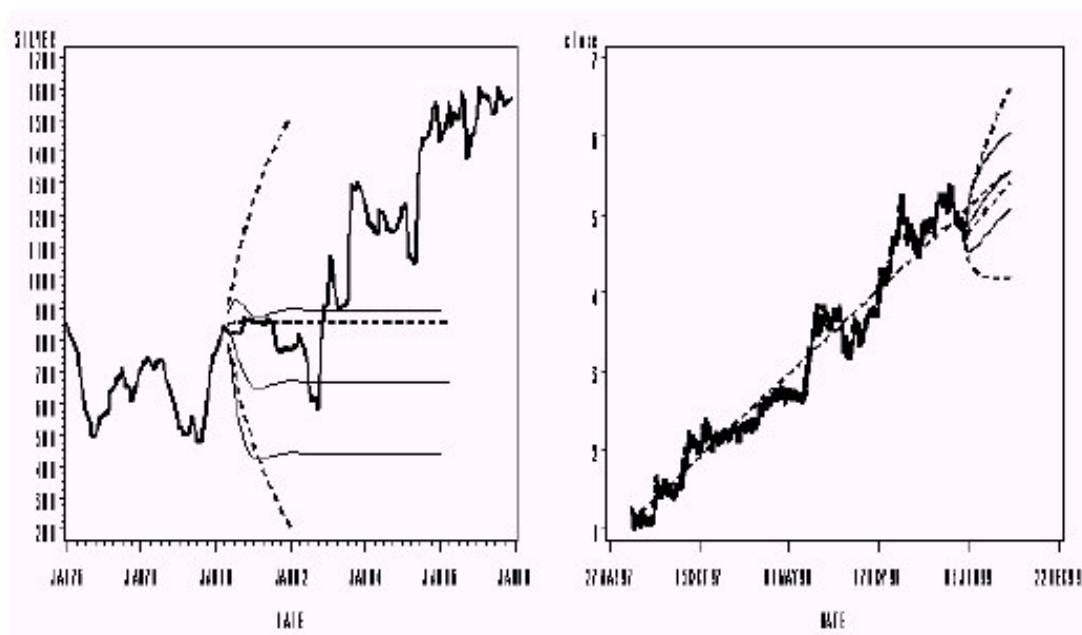


Figure 1: Examples - silver amounts and stock closing prices.

Using similar plot symbols, the right panel of figure 1 shows the closing stock prices and forecasts for Amazon.com stock over the period 16May97-25May99. There is a fairly clear upward trend. This is not consistent with a simple random walk or a stationary ARMA series. One class of plausible models for this data expresses deviations  $y_t$  of the prices  $Y_t$  from a trend as an ARIMA model. A lag 1 example is

$$Y_t - \beta_0 - \beta_1 t = (Y_{t-1} - \beta_0 - \beta_1 (t-1)) + e_t$$

where, in the case of a unit root  $\alpha_1$  would be 1 and the model would become a so-called random walk with drift  $\beta_1$  :

$$Y_t = Y_{t-1} + \beta_1 + e_t.$$

Forecasts with error bands from a unit root process with drift (dotted lines) and from a regression with stationary ARMA errors (solid lines) are shown in the right panel. Again the question of whether the deviations  $y_t$  follow a unit root process or not is the criterion for deciding between the two scenarios. One step ahead forecasts throughout the historic data are included for both models but they are so close to the actual data that they cannot be distinguished from it or from each other.

## 5 The tests

The lag 1 model with time trend is fully expressed as

$$Y_t - \beta_0 - \beta_1 t = \alpha_1 (Y_{t-1} - \beta_0 - \beta_1 (t-1)) + e_t$$

or equivalently, subtracting  $(Y_{t-1} - \beta_0 - \beta_1 (t-1))$  from both sides, as

$$Y_t - Y_{t-1} = \beta_1 + (\alpha_1 - 1)(Y_{t-1} - \beta_0 - \beta_1 (t-1)) + e_t$$

or, finally, as

$$Y_t - Y_{t-1} = [\beta_1 - (\beta_0 - \beta_1)(\alpha_1 - 1)] - [\beta_1(\alpha_1 - 1)]t + [\alpha_1 - 1]Y_{t-1} + e_t$$

where the terms in square brackets are the intercept, slope, and lag Y coefficient being estimated in a least squares regression (with intercept) of  $Y_t - Y_{t-1}$  on  $t$  and  $Y_{t-1}$ . In the same way as shown above, a model with more than 1 lag can be accommodated by adding lagged differences. For a model with 3 lagged deviations from a trend, the regression would be of the form

$$Y_t - Y_{t-1} = \alpha - \beta t + \gamma_1 Y_{t-1} + \gamma_2 (Y_{t-1} - Y_{t-2}) + \gamma_3 (Y_{t-2} - Y_{t-3}) + e_t$$

where the important points are

- (1) The theoretical coefficient on  $Y_{t-1}$  will be 0 whenever there is a unit root;
- (2) The coefficient on  $t$  will be 0 and the intercept nonzero whenever there is a unit root and  $\beta_1$  is nonzero;
- (3) If  $\beta_1$  is assumed to be 0, the regression should not include time  $t$  and the intercept will become 0 in the presence of unit roots.
- (4) If both  $\beta_0$  and  $\beta_1$  are assumed to be 0 then neither a time trend nor intercept should be used in the regression.

If the series is stationary, all of the least squares regressions mentioned here give asymptotically normal coefficient estimates and asymptotically  $N(0,1)$  studentized statistics. The autoregressive estimates differ from the true coefficients by  $O_p(1/\sqrt{n})$ .

In particular in the lag 1 case where  $\beta_0 = 0 = \beta_1$  is assumed, we find

$$\sqrt{n}(\hat{\alpha}_1 - \alpha_1) = \left[ n^{-1} \sum_{t=2}^n Y_{t-1}^2 \right]^{-1} \left[ n^{-1/2} \sum_{t=2}^n Y_{t-1} e_t \right] \xrightarrow{d} N(0, 1 - \alpha_1^2)$$

and the studentized statistic,  $(\hat{\alpha}_1 - \alpha_1)$  divided by its standard error from a least squares regression, satisfies this distributional limit law:

$$t = \left[ s^2 \sum_{t=2}^n Y_{t-1}^2 \right]^{-1/2} \left[ \sum_{t=2}^n Y_{t-1} e_t \right] \xrightarrow{d} N(0, 1),$$

where  $s^2$  denotes the regression mean square. These results also hold for regressions with intercept and time trend terms. The key to these results is that  $n^{-1} \sum_{t=2}^n Y_{t-1}^2$  converges to a constant. In contrast, for the unit root case with one lag,  $\alpha_1$  is 1 and we have

$$n(\hat{\alpha}_1 - 1) = \left[ n^{-2} \sum_{t=2}^n Y_{t-1}^2 \right]^{-1} \left[ n^{-1} \sum_{t=2}^n Y_{t-1} e_t \right] \xrightarrow{d} \left[ \int_0^1 W^2(t) dt \right]^{-1} [0.5(W^2(1) - 1)]$$

and

$$\tau = \left[ s^2 \sum_{t=2}^n Y_{t-1}^2 \right]^{-1/2} \left[ \sum_{t=2}^n Y_{t-1} e_t \right] \xrightarrow{d} \left[ \int_0^1 W^2(t) dt \right]^{-1/2} [0.5(W^2(1) - 1)]$$

where  $W(t)$  denotes a standard Wiener process on  $[0,1]$ . Both the numerators and denominators of these fractions are random variables in the limit, the coefficient is now normalized by  $n$ , and we have changed the  $t$  symbol to  $\tau$  to explicitly remind the reader that this does not satisfy the distributional properties of Student's  $t$ , even in the limit.

Further, the addition of an intercept or time trend changes the distribution, even in the limit, implying that tables of percentiles for all three cases (trend, no trend, and no trend or intercept in the model) must be computed.

One nice thing does happen here, namely the addition of lagged differences to these regressions does not affect the limit distribution of  $\tau$  in any of the three cases so once the three tables of  $\tau$  are computed the limit percentiles can be used to test for a unit root in an autoregressive model of arbitrary order as long as the sample size is large. Tables of the limit distribution appear in several textbooks, for example Fuller (1996). Negative values of  $\tau$ ,  $\tau_\mu$ ,  $\tau_\tau$  as the studentized statistics are called in the three cases, indicate stationarity and the 5% critical values for these in the limit are -1.95, -2.86, and -3.41. Clearly these are more extreme than the -1.645 that would have been used had the studentized statistics been normal.

The expression  $\left[ \int_0^1 W^2(t) dt \right]^{-1/2} [0.5(W^2(1) - 1)]$  is a neat way to summarize the limit distribution of  $\tau$ , but it does not give a simple way to analytically

find percentiles or even to simulate them except by picking some large  $n$ , simulating and hoping the results are close to the limit. Dickey and Fuller (1979, 1981) computed the spectral decomposition of the denominator quadratic form

$[n^2\sigma^2]^{-1} \sum_{t=2}^n Y_{t-1}^2$  and its limit, expressing the limit as  $\sum_{j=1}^{\infty} w_j^2 Z_j^2$ , an infinite weighted sum of independent standard normal variates where the weights were found to be  $w_j = [(2j-1)\pi]^{-1} 2(-1)^{j+1}$ . The numerator  $[n\sigma^2]^{-1} \sum_{t=2}^n Y_{t-1}e_t$  was

found to have limit  $0.5 \left[ \sum_{j=1}^{\infty} (w_j Z_j)^2 - 1 \right]$ . The infinite sums could be approxi-

mated nicely by the sum of the first several terms. An even better approximation resulted from adding a few additional terms with very small coefficients such that the moments up through the second of the numerator and denominator were matched almost exactly. Percentiles approximated by Monte Carlo simulation with this method matched those generated straightforwardly using series with large  $n$  (e.g. 500) which served as a check on the programming as well as the adequacy of the "large" sample size.

Quadratic form expressions as well as Wiener process representations of limit distributions in the models with intercept and trend terms are available in the literature. Fuller (1996) summarizes the results.

Of course now that the percentiles have been calculated and the results are well known, it is enough to show that a test statistic has the same limit representation as our least squares estimators, e.g.

$\left[ \int_0^1 W^2(t) dt \right]^{-1/2} [0.5(W^2(1) - 1)]$

using the methods in Billingsley (1968) or otherwise. That ensures the appropriateness of the tables for performing any such test. This, along with Donsker's theorem which shows invariance of this limit distribution under fairly weak assumptions on the errors, is the beauty of the Wiener process representation.

## 6 Analysis and Extensions

The tests described above have been incorporated in several software packages. Here we use the SAS ETS package (SAS Institute, Cary, N.C., USA) and base SAS software to analyze the silver and stock price series introduced earlier. Looking at the silver series, there is no apparent trend over the initial time period used to estimate the model. We decide to regress the differences on a lag level and lagged differences. These lagged differences are called "augmenting terms" and the resulting test an "augmented Dickey-Fuller test" in econometrics. Getting the right number of augmenting terms is important. In that light, an interesting result is that in the regression, under the unit root null hypothesis, the coefficients on the lagged differences satisfy the usual limit theory - they

are jointly multivariate normal. Only the lag Y coefficient, intercept and time trend coefficients (if any) have nonstandard limits. A reasonable strategy is to run a least squares regression using a fairly large number of augmenting terms and standard tests to cut down on this number then performing the unit root test on the final model.

Using 4 augmenting differences on the silver series, the F test for omitting the last 3 is  $F=1.32$  ( $\Pr>F$  is .2803) so we can omit all but the first augmenting lag difference. Fitting that model we obtain

$$Y_t - Y_{t-1} = \underset{(27.4)}{75.58} - \underset{(0.042)}{0.117}Y_{t-1} + \underset{(0.108)}{0.6712}(Y_{t-1} - Y_{t-2}) + e_t$$

Numbers in parentheses are standard errors and the estimation was done by an ordinary regression program (SAS PROC REG). The lag Y coefficient is negative and the regression program reports a p-value 0.0079 which would become 0.0040 (one-sided test) when stationarity is the alternative. Because this is just a regression program, the test is performed by comparing  $t=-2.78$ , which is really  $\tau_\mu$ , to a t distribution whereas we saw earlier that the 5th percentile of the appropriate distribution is approximately -2.86 so our test really is not significant! From the graph it is clear that a decision of stationarity, as would have been made had we trusted the t distribution, would have been wrong. Of course we can only make that comment having the luxury of several subsequent years of observations. Had we used a time series package with the test built in, we'd guard against this potential error. Using SAS PROC ARIMA, the test statistic -2.78 is again computed but the correct p-value (0.0689) is reported and it is seen that there is not enough evidence at the usual level to reject a unit root. The wide band forecast errors would be chosen in the graph.

Said and Dickey (1984) show that even if a series has moving average terms, the addition of enough lagged differences, where this number increases with sample size, gives a test on lag Y in the augmented regression whose studentized statistic converges to the tabulated distribution. In other words, all ARIMA models can be treated as though they are autoregressive.

Turning to the stock price series, it seems clear that if a stationary alternative to the unit root model is to have any chance of being chosen, it would have to account for the upward trend. Thus the regression would have to be the one with a linear trend term and  $\tau_\tau$  would be the test. Practitioners sometimes mistakenly fit the model  $Y_t = \lambda + \alpha Y_{t-1} + e_t$  which has a linear trend with slope when  $\alpha = 1$  and has mean  $\mu = \lambda/(1 - \alpha)$  but no trend when  $|\alpha| < 1$ . Because it does not allow a trend unless  $\alpha = 1$ , this model intertwines the unit root issue with the trend issue. This confusion is the price paid for not explicitly writing the model in deviations form  $(Y_t - \mu) = \alpha(Y_{t-1} - \mu) + e_t$  i.e.  $Y_t = \mu(1 - \alpha) + \alpha Y_{t-1} + e_t$ . We saw that the model  $Y_t - (\beta_0 + \beta_1 t) = \alpha(Y_{t-1} - (\beta_0 + \beta_1(t-1))) + e_t$  has a drift  $\lambda = \beta_1$  when  $\alpha = 1$  and trend with slope  $\beta_1$  when  $|\alpha| < 1$  thus separating the trend question from the unit root question. This model, in which both an intercept and time term are included in the regression, is the appropriate model here.

Note that the autocorrelation function is another statistic that intertwines trend and unit root questions. That is, if  $Y_t$  is a linear trend plus stationary errors, the autocorrelations of  $Y_t$  will die down slowly as though a unit root were present. The first differences would have quickly decaying autocorrelations. They would not pick up the unit root that was induced on the moving average side of the model by the differencing (e.g.  $Y_t = \beta_0 + \beta_1 t + e_t$  implies that  $Y_t - Y_{t-1} = \beta_1 + e_t - \theta e_{t-1}$  where  $\theta = 1$ ). Chang and Dickey (1994) show that the inverse autocorrelation of such an overdifferenced series displays the same slow decay behavior as is characteristic of the autocorrelation function of a unit root process. For the stock price series on the right in figure 1, the inverse autocorrelation function of the differences in fact dies off quite rapidly, thus suggesting that differencing was appropriate, despite the fact that the data appear to hug the displayed linear trend line quite closely. To decide the issue more formally, a regression of the first difference on an intercept, time, lagged level and several lagged differences was run. No lagged differences were significant and the regression was refit without them. The "t test" from the least squares regression printout (SAS PROC REG) was -2.59 with a reported p-value  $< 0.01$  based (incorrectly) on Student's t distribution. As reported above, the correct 5% critical value for the studentized test for lagged Y in the trend model is -3.41 and the test is not even close to being significant.

## 7 A trend stationary example

In addition to closing prices, most stock reports show daily high, low, and volume numbers. For the Amazon stock data studied here,  $Y = \log(\text{volume})$  is somewhat interesting. In Figure 2, the  $Y_t$  is graphed and below it, the autocorrelation, inverse autocorrelation, and partial autocorrelation of the series followed by the same three functions on the differences.

The autocorrelation does not die off very fast, suggesting a unit root. However the middle plot in the bottom row, the inverse autocorrelation of the differences, dies off very slowly. The results of Chang and Dickey (1994) suggest this could be due to over differencing. These mixed signals, autocorrelations implying differencing and inverse autocorrelations implying that differencing is a mistake, result from the fact that the data show a strong trend. The original series' autocorrelations do not account for this. Notice that a trend in the original series reduces to a constant in the differenced series as we have noted several times, so that the message in the inverse autocorrelations of differences is still worthy of trust.

Again using a formal test to make the decision, a regression of the differenced series on lagged level, two augmenting lagged differences and a time trend was used. The test statistic on lagged  $\log(\text{volume})$  was -6.388, large enough in magnitude to strongly reject the unit root hypothesis. This condition, a linear trend with stationary errors, is often called "trend stationarity".

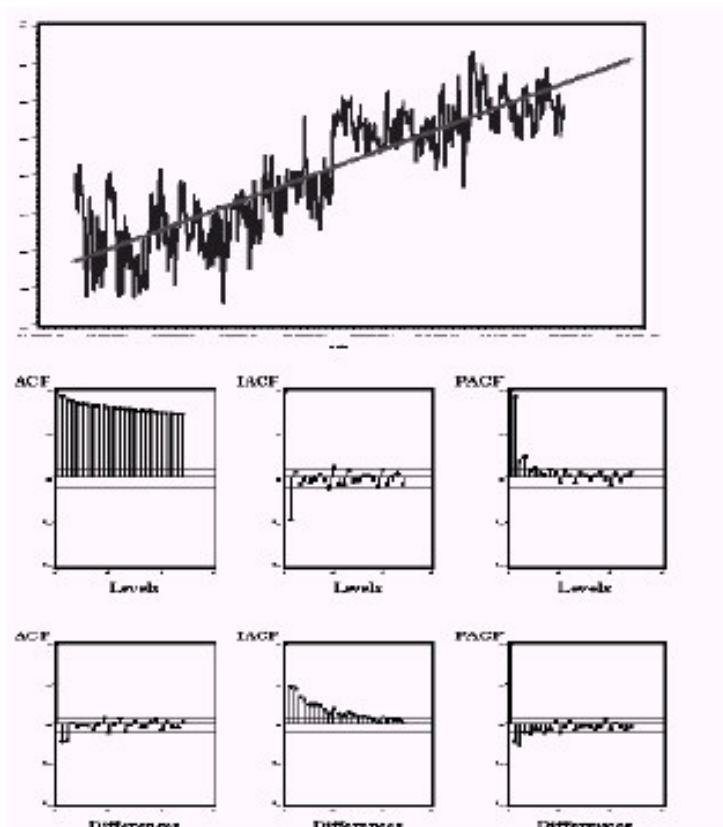


Figure 2: Stock volumes

## 8 Variations

One focus of research has been on alternate estimators. Phillips and Perron (1988) give a different method than augmenting terms for dealing with the higher order models. Their estimator has Wiener process representation that justifies the use of the same distributions as we have been discussing. Dickey, Hasza, and Fuller (1984) suggest a symmetric estimator and Park and Fuller (1995) improve this with a weighted symmetric estimator. Elliott, Rothenberg and Stock (1996) propose doing a generalized least squares estimate of the intercept and slope assuming a nearly nonstationary autoregressive structure, then treating the residuals as input for the kind of regression proposed here. Gonzalez-Farias (1992) in her thesis proposed an "exact maximum likelihood" test based on maximizing the stationary likelihood function. Least squares estimators maximize a likelihood conditional on the first observation. Even when unit root data are fed into the stationary log likelihood function, it is well behaved. This test

is somewhat complicated, involving nonlinear estimation, but appears to have quite good power properties, especially compared to the original least squares based tests. Park and Fuller's weighted symmetric test and Elliott et al's GLS detrended tests are easier to compute and similar in their power properties. Schwert (1989) compares some of these approaches under different true model scenarios and Pantula, Gonzalez-Farias and Fuller (1994) study powers of some of these estimators.

## 9 Summary

Tests for unit roots based on regression estimators are commonly used, but their distributions are not standard even in the limit. A regression of first differences on a lagged level and lagged differences, along with possibly an intercept and time trend term provides a convenient tool for calculating the test statistics. The lagged differences have estimated coefficients with standard limit distributions while the unit root test, the lagged level variable, does not. The inverse autocorrelation and autocorrelation can also be informative unit root diagnostics. Several examples have been given to illustrate the tests and a few recent improvements reviewed.

## Referências

- [1] Billingsley, P (1968). *Convergence of Probability Measures*, Wiley, New York.
- [2] Brocklebank, J. C. and Dickey, D. A. (1986, 2003) *SAS System for Forecasting Time Series*, 1st and 2nd editions. SAS Institute, Cary, N. Carolina.
- [3] Dickey, D. A. and Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. , 427-431. *J. American Statistical Assn.*, 74.
- [4] Dickey, D. A. and Fuller, W. A. (1981). Likelihood Ratio Tests for Autoregressive Time Series with a Unit Root. , 1057-1072. *Econometrica*, 49.
- [5] Dickey, D. A., Hasza, D. P., and Fuller, W. A. (1984). Testing for Unit Roots in Seasonal Time Series. , 355-367. *J. American Statistical Assn.*, 79.
- [6] Elliott, G., Rothenberg, T. J., and Stock, J. A. (1996) Efficient Tests for an Autoregressive Unit Root. , 813-836 *Econometrica*, 64.
- [7] Fuller, W. A. (1996). , Wiley, New York. *Introduction to Statistical Time Series*.

- [8] Gonzalez-Farias, G. (1992) A New Unit Root Test for Autoregressive Time Series, PhD Thesis, North Carolina State University.
- [9] Pantula, S. G., Gonzalez-Farias, G. and Fuller, W. A. (1994). A Comparison of Unit Root Test Criteria. *J. Business and Economic Statistics*, 12, 1-11.
- [10] Park, H. J. and Fuller, W. A. (1995). Alternative Estimators and Unit Root Tests for the Autoregressive Process. , 415-429. *J. Time Series Analysis*, 16.
- [11] Phillips, P.C.B. and Perron, P. (1988). Testing for a Unit Root in Time Series Regression. , 335-346. *Biometrika*, 75.
- [12] Said, S. E. and Dickey, D. A. (1984). Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. , 599-607. *Biometrika*, 71.
- [13] Schwert, G. W. (1989). Tests for Unit Roots: A Monte Carlo Investigation. *J. Business and Economic Statistics* 7, 147-160.