DIFFERENCING AND UNIT ROOT TESTS

In the Box-Jenkins approach to analyzing time series, a key question is whether to difference the data, i.e., to replace the raw data $\{x_t\}$ by the differenced series $\{x_t - x_{t-1}\}$. Experience indicates that most economic time series tend to wander and are not stationary, but that differencing often yields a stationary result. A key example, which often provides a fairly good description of actual data, is the random walk, $x_t = x_{t-1} + \varepsilon_t$, where $\{\varepsilon_t\}$ is white noise, assumed here to be independent, each having the same distribution (e.g., normal, t, etc.). The random walk is said to have a **unit root.**

To understand what this means, let's recall the condition for stationarity of an $AR(p)$ model. In Chapter 3, part II, we said that the $AR(p)$ series

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + \varepsilon_t$$

will be stationary if the largest root θ of the equation (in the complex variable z)

$$z^p = \alpha_1 z^{p-1} + \alpha_2 z^{p-2} + \cdots + \alpha_{p-1} z + \alpha_p \tag{1}$$

satisfies $|\theta| < 1$. So stationarity is related to the location of the roots of Equation (1).

We can think of the random walk as an $AR(1)$ process, $x_t = \alpha x_{t-1} + \varepsilon_t$ with $\alpha = 1$. But since it has $\alpha = 1$, the random walk is not stationary. Indeed, for an $AR(1)$ to be stationary, it is necessary that all roots of the equation $z = \alpha$ have "absolute value" less than 1. Since the root of the equation $z = \alpha$ is just α, we see we see that the $AR(1)$ is stationary if and only if $-1 < \alpha < 1$. For the random walk, we have a *unit root,* that is, a root equal to one. The first difference of a random walk is stationary, however, since $x_t - x_{t-1} = \varepsilon_t$, a white noise process.

In general, we say that a time series $\{x_t\}$ is **integrated of order 1,** denoted by $I(1)$, if $\{x_t\}$ is not stationary but the first difference $\{x_t - x_{t-1}\}$ is stationary and invertible. If $\{x_t\}$ is $I(1)$, it is considered important to difference the data, primarily because we can then use all of the methodologies developed for stationary time series to build a model, or to otherwise analyze, the differenced series. This, in turn, improves our understanding (e.g., provides better forecasts) of the original series, $\{x_t\}$. For example, in the Box-Jenkins $ARIMA(p, 1, q)$ model, the differenced series is modeled as a stationary $ARMA(p, q)$ process. In practice, then, we need to decide whether to build a stationary model for the raw data or for the differenced data.

More generally, there is the question of how many times we need to difference the data. In the $ARIMA(p, d, q)$ model, the d'th difference is a stationary $ARMA(p, q)$. The series is **integrated of order d,** denoted by $I(d)$, where d is an integer with $d \geq 1$, if the series and all its differences up to the $d-1$'st are nonstationary, but the d'th difference is stationary. A series is said to be **integrated of order zero,** denoted by $I(0)$, if the series is both stationary and invertible. (The importance of invertibility will be discussed later). If the series $\{x_t\}$ is $I(d)$ with $d \geq 1$, then the differenced series $\{x_t - x_{t-1}\}$ is $I(d-1)$.

For an example of an $I(2)$ process, consider the $AR(2)$ series $x_t = 2x_{t-1} - x_{t-2} + \varepsilon_t$. This process is not stationary. Equation (1) becomes $z^2 = 2z - 1$, that is, $z^2 - 2z + 1 = 0$. Factoring this gives $(z-1)(z-1) = 0$, so the equation has *two* unit roots. Since the largest root (i.e., one) does not have "absolute value" less than one, the process is not stationary. It can be shown that the first difference is not stationary either. The second difference is

$$x_t - x_{t-1} - [x_{t-1} - x_{t-2}] = x_t - 2x_{t-1} + x_{t-2} \ ,$$

which is equal to ε_t by the definition of our $AR(2)$ process. Since the second difference is white noise, $\{x_t\}$ is an $ARIMA(0, 2, 0)$. Since the second difference is stationary, $\{x_t\}$ is $I(2)$. In general, for any $ARIMA$ process which is integrated of order d, Equation (1) will have exactly d unit roots. In practice, however, the only integer values of d which seem to occur frequently are 0 and 1. So here, we will limit our discussion to the question of whether or not to difference the data one time.

If we fail to take a difference when the process is nonstationary, regressions on time will often yield a spuriously significant linear trend, and our forecast intervals will be much too narrow (optimistic) at long lead times. For an example of the first phenomenon, recall that for the Deflated Dow Jones series, we got a t-statistic for the slope of 5.27 (creating the illusion of a very strong indication of trend), but the mean of the first differences was not significantly different from zero. For an example of the second phenomenon, let's compare a random walk with a stationary $AR(1)$ model. For a random walk, the variance of the h-step forecast error is

$$var\,[x_{n+h} - x_n] = var\,[\varepsilon_{n+h} + \varepsilon_{n+h-1} + \cdots - (\varepsilon_n + \varepsilon_{n-1} + \cdots)]$$

$$= var\,[\varepsilon_{n+h} + \cdots + \varepsilon_{n+1}] = h\ var\,[\varepsilon_t]\ ,$$

which goes to ∞ as h increases. The width of the forecast intervals will be proportional to $\sqrt{h}$, indicating that our uncertainty about the future value of the series grows without bound as the lead time is increased. On the other hand, for the stationary $AR(1)$ process $x_t = \alpha x_{t-1} + \varepsilon_t$ with $-1 < \alpha < 1$, the best linear h-step forecast is $f_{n,h} = \alpha^h x_n$, which goes to zero as h increases. The variance of the forecast error is $var\,[x_{n+h} - \alpha^h x_n]$, which tends to $var\,[x_t]$, a finite constant. So as the lead time h is increased, the width of the h-step prediction intervals grows without bound for a random walk, but remains bounded for a stationary $AR(1)$. Clearly, then, if our series were really a random walk, but we failed to difference it and modeled it instead as a stationary $AR(1)$, then our prediction intervals would give us much more faith in our ability to predict at long lead times than is actually warranted.

It is also undesirable to take a difference when the process is stationary. Problems arise here because the difference of a stationary series is not invertible, i.e., cannot be represented as an $AR(\infty)$. For example, if $x_t = .9x_{t-1} + \varepsilon_t$, so that $\{x_t\}$ is really a stationary $AR(1)$, then the first difference $\{z_t\}$ is the non-invertible $ARMA(1,1)$ process $z_t = .9z_{t-1} + \varepsilon_t - \varepsilon_{t-1}$, which has more parameters than the original process. (Recall that an $ARMA(p,q)$ is invertible if the largest root θ of the equation $z^q + b_1 z^{q-1} + \cdots + b_q = 0$ satisfies $|\theta| < 1$, where $b_1, \ldots, b_q$ are the MA parameters.) Because of the non-invertibility of $\{z_t\}$, its parameters will be difficult to estimate, and it will be difficult to construct a forecast of z_{t+h}. Consequently, taking an unnecessary difference (i.e., **overdifferencing**) will tend to degrade the quality of forecasts.

Ideally, then, what we would like is a way to decide whether the series is stationary, or integrated of order 1. A method in widespread use today is to declare the series nonstationary if the sample autocorrelations decay slowly. If this pattern is observed, then the series is differenced and the autocorrelations of the differenced series are examined to make sure that they decay rapidly, thereby indicating that the differenced series is stationary. This method is somewhat *ad hoc*, however. What is really needed is a more objective way of deciding between the two hypotheses, $I(0)$ and $I(1)$, without making any further assumptions. Unfortunately, each of these hypotheses covers a vast range of possibilities, and any classical approach to discriminate between them seems doomed to failure unless we limit the scope of the hypotheses.

The Dickey-Fuller Test of Random Walk Vs. Stationary AR(1)

A test involving much more narrowly-specified null and alternative hypotheses was proposed by Dickey and Fuller in 1979. In its most basic form, the Dickey-Fuller test compares the null hypothesis

$$H_0 : x_t = x_{t-1} + \varepsilon_t \quad ,$$

i.e., that the series is a random walk without drift, against the alternative hypothesis

$$H_1 : x_t = c + \rho x_{t-1} + \varepsilon_t \quad ,$$

where c and ρ are constants with $|\rho| < 1$. According to H_1, the process is a stationary $AR(1)$ with mean $\mu = c/(1-\rho)$. To see this, note that, under H_1, we can write

$$x_t = \mu(1-\rho) + \rho x_{t-1} + \varepsilon_t \quad ,$$

so that

$$x_t - \mu = \rho(x_{t-1} - \mu) + \varepsilon_t \quad .$$

Note that by making the random walk the *null* hypothesis, Dickey and Fuller are expressing a preference for differencing the data unless a strong case can be made that the raw series is stationary. This is consistent with the conventional wisdom that, most of the time, the data do require differencing. A Type I error corresponds to deciding the process is stationary when it is actually a random walk. In this case, we will fail to recognize that the data should be differenced, and will build a stationary model for our nonstationary series. A Type II error corresponds to deciding the process is a random walk when it is actually stationary. Here, we will be inclined to difference the data, even though differencing is not desirable.

We should mention two additional important differences between the $AR(1)$ and the random walk. Whereas the innovation ε_t has a temporary (exponentially decaying) effect on the $AR(1)$, it has a permanent effect on the random walk. Whereas the expected length of time between crossings of μ is finite for the $AR(1)$ (so the $AR(1)$ fluctuates around its mean of μ), the expected length of time between crossings of any particular level is *infinite* for the random walk (so the random walk has a tendency to wander in a non-systematic fashion from any given starting point).

The Dickey-Fuller test is easy to perform. Given data $x_1, \ldots, x_n$, we run an ordinary linear regression of the observations $(x_2, \ldots, x_n)$ of the "dependent variable" $\{x_t\}$, against the observations $(x_1, \ldots, x_{n-1})$ of the "independent variable" $\{x_{t-1}\}$, together with a constant term. Under both H_0 and H_1, the data x_t obey the linear regression model

$$x_t = c + \rho x_{t-1} + \varepsilon_t \quad ,$$

and H_0 corresponds to $\rho = 1$, $c = 0$.

Denote the t-statistic for the least squares estimate $\hat{\rho}$ by

$$\tau_\mu = (\hat{\rho} - 1)/s_{\hat{\rho}} \quad ,$$

where $s_{\hat{\rho}}$ is the estimated standard error for $\hat{\rho}$. Note that τ_μ is easy to calculate, since $\hat{\rho}$ and $s_{\hat{\rho}}$ can be obtained directly from the output of the standard computer regression packages.

For the Deflated Dow data, regressing $x_2, \ldots, x_{547}$ on $x_1, \ldots, x_{546}$, we obtain the following regression output:

Residual Standard Error = 0.6933, Multiple R-Square = 0.9907
N = 546, F-statistic = 57991.38 on 1 and 544 df, p-value = 0

	coef	std.err	t.stat	p.value
Intercept	0.1095	0.0832	1.3165	0.1886
X	0.9963	0.0041	240.8140	0.0000

The R-Square statistic is .9907, indicating a very strong linear relationship between $\{x_t\}$ and $\{x_{t-1}\}$. The estimated slope is $\hat{\rho} = .9963$, and $s_{\hat{\rho}} = .0041$. We calculate

$$\tau_\mu = (\hat{\rho} - 1)/s_{\hat{\rho}} = (.9963 - 1)/.0041 = -.9024 \quad .$$

Note that we do NOT use the t statistic (240.8140) from the output, since this was computed relative to a null value of zero, instead of 1.

The statistic τ_μ can be used to test H_0 versus H_1. The percentiles of τ_μ under H_0 are given in the attached table. The null hypothesis is rejected if τ_μ is less than the tabled value. The tabulations for finite n were based on simulation, assuming ε_t are *iid* Gaussian. The tabled values for the asymptotic distribution ($n = \infty$) are valid as long as the ε_t are *iid* with finite variance. (No Gaussian assumption is needed here.) It should be noted that τ_μ does not have a t distribution in finite samples, and does not have a standard normal distribution asymptotically. In fact, the asymptotic distribution is longer-tailed than the standard normal. For example, the asymptotic .01 percentage point of τ_μ is at -3.43, instead of -2.326 for a standard normal. Thus, use of the standard normal table would result in an excess of spurious declarations of stationarity.

For the Deflated Dow data, we obtained $\tau_\mu = -.9024$, which is not significant according to the Table. So we are not able to reject the random walk hypothesis. As usual in statistical hypothesis testing, this does not mean that we should conclude that the series is a random walk. In fact, from our earlier analysis we have strong statistical evidence that the series is *not* a random walk, since the lag-1 autocorrelation for the first differences is highly significant. All we can conclude from the Dickey-Fuller test is that there is no strong evidence to support the hypothesis H_1 that the series is a stationary $AR(1)$. This is the type of alternative that the test was designed to detect. The question of whether the first difference has any autocorrelation is another issue altogether, and the test was not designed to detect this type of failure of the random walk hypothesis. In any case, the results of the test indicate that it would be a good idea to difference the data. We could have come to this same conclusion by examining the ACF of the raw data, but the Dickey-Fuller test provides a more objective basis for making this decision.

As an illustration of the long tails in the τ_μ distribution, consider the random walk data ($n = 547$) which was used in the last handout for comparison with the Dow and Deflated Dow series. For this random walk data set, we obtain the following regression output.

```
Residual Standard Error = 0.9917,  Multiple R-Square = 0.9823
N = 546,  F-statistic = 30111.83 on 1 and 544 df, p-value = 0

             coef  std.err    t.stat  p.value
Intercept 0.0771  0.0550    1.4028  0.1612
        X 0.9913  0.0057 173.5276  0.0000
```

We therefore get $\tau_\mu = (.9913 - 1)/.0057 = -1.53$. If τ_μ had a standard normal distribution, we would obtain a p-value of .063 (one-sided), indicating some evidence in favor of the alternative hypothesis (that the series is a stationary $AR(1)$). Of course, we know that this series was in fact a random walk, and so it is somewhat distressing that we are almost being led to commit a Type I error. But when we use the true distribution of τ_μ under the null hypothesis (see table) we find that the actual significance level is substantially greater than .10, although the table is not precise enough to allow us to find the exact p-value.

Of course, the null and alternative hypotheses H_0 and H_1 described above are too narrow to be very useful in a wide variety of situations. Often, we will want to consider differencing the data because we hope the difference may be stationary, but we do not want to commit ourselves to the assumption that the series is either a random walk or a stationary $AR(1)$. Fortunately, although we will not describe the details here, there is a similar test known as the Augmented Dickey-Fuller test, which allows us to test an $ARIMA(p, 1, 0)$ null hypothesis versus an $ARIMA(p+1, 0, 0)$ alternative, where $p \geq 0$ is known. If $p = 1$, for example, the null hypothesis would be that the series is nonstationary, but its first difference is a stationary $AR(1)$; the alternative hypothesis would be that the series is a stationary $AR(2)$. In retrospect, it seems that the Deflated Dow series is better described by the above null hypothesis than by the one which was actually tested, i.e., the random walk. But it is never a good idea to change a statistical hypothesis after looking at the data; it can destroy the validity of the test. Furthermore, the use of the random walk as a null hypothesis for financial time series seems wise as a general rule.

Difference Stationarity Vs. Trend Stationarity

In the ordinary Dickey-Fuller (τ_μ) test, the series is assumed to be free of deterministic trend, under both the null and alternative hypotheses. Many actual series do have trend, however, and it is of interest to study the nature of this trend. Perhaps the most important issue is the way in which the trend is combined with the random aspects of the series. In the case of a random walk with drift $x_t = c + x_{t-1} + \varepsilon_t$ where $\{\varepsilon_t\}$ is zero mean white noise, there is a mixture of deterministic and stochastic trend, the process has a unit root, and the forecast intervals grow without bound as the lead time increases. Differencing $\{x_t\}$ yields a stationary series, so $\{x_t\}$ is said to be **difference stationary.** (This is the same as $I(1)$).

Another way to combine trend and randomness is to start with a deterministic linear trend and bury it in white noise: $x_t = \alpha_0 + \alpha_1 t + \varepsilon_t$. This is a standard linear regression (trend-line) model, which can be analyzed without using time series methods. If the parameters $(\alpha_0, \alpha_1, var\,[\varepsilon_t])$ are known, then the forecast of x_{n+h} is simply $f_{n,h} = \alpha_0 + \alpha_1(n+h)$. If the ε_t are normally distributed, a forecast interval for x_{n+h} is given for large h by $f_{n,h} \pm z_{\alpha/2}\sqrt{var\ \varepsilon_t}$. The width of this forecast interval *does not* tend to infinity as the lead time increases.

More generally, any series

$$x_t = \alpha_0 + \alpha_1 t + y_t$$

formed by adding a deterministic linear trend to a stationary, invertible, zero mean "noise" series $\{y_t\}$ is said to be **trend stationary.** Trend stationary series do not contain a unit root. The width of their forecast intervals for large h is $2 z_{\alpha/2}\sqrt{var\ y_t}$, which does not tend to infinity. Trend stationary series are not difference stationary, since it can be shown that the difference of $\{y_t\}$ is not invertible. Since the trend stationary series obeys a regression model with autocorrelated errors, we can use generalized least squares (a popular linear regression technique) to estimate the trend and assess its statistical significance.

Here, we show how to test a specific form of difference stationarity against a specific form of trend stationarity, using a variant (τ_τ) of the Dickey-Fuller test. The null hypothesis is

$$H_0 : x_t = c + x_{t-1} + \varepsilon_t \quad ,$$

a random walk with drift (which is difference stationary), versus

$$H_1 : x_t = \alpha_0 + \alpha_1 t + y_t \ \ ; \ \ y_t = \rho y_{t-1} + \varepsilon_t \quad .$$

Under H_1, $\{x_t\}$ is trend stationary, and the "noise" term is $AR(1)$. If we put $\rho = 0$, then we get the trend-line model. It can be shown that under H_1, $\{x_t\}$ can be expressed as

$$x_t = \beta_0 + \beta_1 t + \rho x_{t-1} + \varepsilon_t \quad , \tag{2}$$

where β_0 and β_1 are constants. (Specifically, $\beta_0 = \alpha_0(1-\rho) + \rho\alpha_1$ and $\beta_1 = \alpha_1(1-\rho)$.) If we put $\rho = 1$, then Equation (2) reduces to

$$x_t = \alpha_1 + x_{t-1} + \varepsilon_t \quad ,$$

i.e., a random walk with drift. Thus, we want to test the null hypothesis that $\rho = 1$ versus the alternative that $\rho < 1$ in Equation (2).

To perform the test, we run an ordinary linear regression of the "dependent variable" $\{x_t\}$ against the explanatory variables time (t) and $\{x_{t-1}\}$, together with a constant term. The observations on $\{x_t\}$ are $(x_2, \ldots, x_n)$, the observations on t are $(2, \ldots, n)$, and the observations on $\{x_{t-1}\}$ are $(x_1, \ldots, x_{n-1})$. The test statistic is the standardized estimate of ρ in Equation (2),

$$\tau_\tau = \frac{\hat{\rho} - 1}{s_{\hat{\rho}}} \quad .$$

Although this may appear to be the same as the ordinary Dickey-Fuller statistic τ_μ, it is actually different because of the presence of time as an explanatory variable. The percentiles of τ_τ under the null hypothesis $(\rho = 1)$ are given in the attached table. The null hypothesis is rejected if τ_τ is less than the tabled value. The percentiles of τ_τ are considerably less than the corresponding percentiles of τ_μ, indicating the effects of including time as an explanatory variable. For example, the asymptotic .01 percentage point of τ_τ is at –3.96 for τ_τ, compared with –3.43 for τ_μ.

The log10 Dow data seems to contain a trend, but what is the nature of this trend? Would it be more appropriate to model this data as a random walk with drift, or as a trend line plus stationary $AR(1)$ errors? In our original analysis of this data, we first tried an ordinary trend-line model, and found a highly significant trend. We then questioned the validity of this finding, since the Durbin-Watson statistic showed strong error autocorrelation. We could have pursued the use of a trend stationary model (i.e., linear trend plus autocorrelated errors) for this series, by re-estimating the trend line using generalized least squares. This still would not have answered the question as to whether such a model is more appropriate than a random walk with drift, however. To address this question, we now run the τ_τ test. The regression described above yielded

Residual Standard Error = 0.0147, Multiple R-Square = 0.9977
N = 546, F-statistic = 119357.4 on 2 and 543 df, p-value = 0

	coef	std.err	t.stat	p.value
Intercept	0.0207	0.0130	1.5885	0.1128
Time	0.0000	0.0000	1.3508	0.1773
x.lag	0.9923	0.0054	182.1997	0.0000

where Time denotes $(2, \ldots, 547)$, x.lag is $(x_1, \ldots, x_{546})$ and the dependent variable is $(x_2, \ldots, x_{547})$. The estimated coefficient of x.lag is $\hat{\rho} = .9923$, and $s_{\hat{\rho}} = .0054$. We calculate

$$\tau_\tau = (\hat{\rho} - 1)/s_{\hat{\rho}} = (.9923 - 1)/.0054 = -1.43 \ .$$

Since this is not less than the tabled value of –3.42, we do not reject the null hypothesis of random walk with drift at level .05. In fact, examination of the table reveals that our observed τ_τ is not small at all, with a p-value around .9, indicating that there is virtually no evidence in favor of trend stationarity for this series. This does not mean that the log10 Dow data is actually a random walk with drift. (Indeed, we previously found strong evidence that the differences of this data are not uncorrelated, even though they seem to have a nonzero expectation.) It just means that we cannot reject the random walk with drift hypothesis in favor of trend stationarity.