Unit Roots and Dickey-Fuller test.

Technical definition of stationarity:

1. Constant theoretical mean ($\mu$)
2. Constant variance $\gamma(0)$
3. The covariance between $Y_t$ and $Y_{t-j}$ depends only on $j$ (call it $\gamma(j)$)

Review:
(A) The expected value of something is the long run population average.

$$E\{Y\} = \mu$$

(B) The covariance between $X$ and $Y$ is

$$\text{Cov}(X,Y) = E\{ (X-\mu_x)(Y-\mu_y) \} \text{ so } \text{Cov}(Y_t,Y_{t-j}) = \gamma(j) = E\{ (Y_t-\mu)(Y_{t-j}-\mu) \}$$

(C) The variance of $Y$ is $E\{ (Y-\mu)^2 \}$ so it is $\gamma(0)$ for stationary time series

(D) The correlation between $X$ and $Y$ is $\frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$$

So autocorrelation is $\frac{\gamma(j)}{\sqrt{\gamma(0)^2\gamma(0)}} = \frac{\gamma(j)}{\gamma(0)}$

With stationarity, standard results like normality of parameter estimates and test statistics can be proved for large samples. Without it, all bets are off.

AR(1) model: \begin{align*}
(Y_t-\mu) &= \rho(Y_{t-1}-\mu) + e_t \iff (Y_t-Y_{t-1}) &= (\rho-1)(Y_{t-1}-\mu) + e_t \\
(Y_{t-1}-\mu) &= (Y_{t-1}-\mu) \quad \text{[subtract $(Y_{t-1}-\mu)$ from both sides]} \end{align*}

If $|\rho|<1$ series is stationary, forecasts are “mean reverting”.

If $\rho=1$ forecasts are not mean reverting and $Y_t=Y_{t-1}+e_t$

(random walk, not stationary)
Test: $H_0: \rho = 1$ versus $H_1: \rho < 1$
Fit $(Y_t - Y_{t-1}) = (\rho - 1)(Y_{t-1} - \mu) + \epsilon_t$ by regression

(1) Regressions:

Case 1: $(Y_t - Y_{t-1})$ on $(Y_{t-1} - \mu)$ (i.e. known mean subtracted off)

Case 2: $(Y_t - Y_{t-1})$ on intercept and $Y_{t-1}$ if $\mu$ is unknown

and if $Y_t - \alpha - \beta t = \rho(Y_{t-1} - \alpha - \beta(t-1)) + \epsilon_t$ ....

Case 3: $(Y_t - Y_{t-1})$ on intercept, $t$, and $Y_{t-1}$ ($\mu$ is replaced by $\alpha + \beta t$)

(2) Look at student t test on $Y_{t-1}$ to test if $(\rho - 1) = 0$.

(3) **Major problem**: t test does NOT have t distribution under $H_0$.

(why? Predictor variable $Y_{t-1}$ is random, not fixed and known without error as regression requires. This is a nonstationary case when $\rho = 1$)

(4) Our contribution: Figure out the right distribution for this t statistic which we now call $\tau$. There are 3 of these, one for each case. Limit distribution theory beyond scope of this course. Limit is not normal.

<table>
<thead>
<tr>
<th>n</th>
<th>P1</th>
<th>P5</th>
<th>P10</th>
<th>P25</th>
<th>P50</th>
<th>mean</th>
<th>P75</th>
<th>P90</th>
<th>P95</th>
<th>P99</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>-2.67</td>
<td>-1.96</td>
<td>-1.61</td>
<td>-1.07</td>
<td>-0.47</td>
<td>-0.40</td>
<td>0.25</td>
<td>0.93</td>
<td>1.34</td>
<td>2.14</td>
</tr>
<tr>
<td>50</td>
<td>-2.61</td>
<td>-1.94</td>
<td>-1.61</td>
<td>-1.08</td>
<td>-0.49</td>
<td>-0.41</td>
<td>0.23</td>
<td>0.91</td>
<td>1.31</td>
<td>2.08</td>
</tr>
<tr>
<td>100</td>
<td>-2.59</td>
<td>-1.94</td>
<td>-1.62</td>
<td>-1.09</td>
<td>-0.49</td>
<td>-0.42</td>
<td>0.23</td>
<td>0.90</td>
<td>1.30</td>
<td>2.04</td>
</tr>
<tr>
<td>250</td>
<td>-2.57</td>
<td>-1.94</td>
<td>-1.61</td>
<td>-1.09</td>
<td>-0.50</td>
<td>-0.42</td>
<td>0.22</td>
<td>0.89</td>
<td>1.29</td>
<td>2.02</td>
</tr>
<tr>
<td>500</td>
<td>-2.57</td>
<td>-1.94</td>
<td>-1.61</td>
<td>-1.09</td>
<td>-0.50</td>
<td>-0.42</td>
<td>0.22</td>
<td>0.89</td>
<td>1.29</td>
<td>2.02</td>
</tr>
<tr>
<td>1000</td>
<td>-2.57</td>
<td>-1.94</td>
<td>-1.62</td>
<td>-1.09</td>
<td>-0.50</td>
<td>-0.42</td>
<td>0.22</td>
<td>0.89</td>
<td>1.28</td>
<td>2.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>f(t) = \mu</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>250</td>
</tr>
<tr>
<td>500</td>
</tr>
<tr>
<td>1000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>f(t) = \alpha + \beta t</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>250</td>
</tr>
<tr>
<td>500</td>
</tr>
<tr>
<td>1000</td>
</tr>
</tbody>
</table>

(2 million simulation runs for each sample size, n=1000 ran in < 3 min.)
What about AR(p) with more lags \( p>1 \) ?

Example:

\[
Y_t - \mu = 1.2(Y_{t-1} - \mu) - .2(Y_{t-2} - \mu) + e_t
\]

or

\[
(1-1.2B+0.2B^2)(Y_t - \mu) = e_t
\]

or

\[
(1-0.2B)(1-B)(Y_t - \mu) = e_t
\]

which is

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

so, no mean reversion. It is like a random walk in this regard. First difference is AR(1). The polynomial \( 1-0.2B)(1-B) = 0 \) has roots \( B=5 \) and \( B=1 \) so this is called a “unit root process” and such processes are not mean reverting (they are nonstationary) just like a random walk. The differences form an AR(1) process \( (Y_t - Y_{t-1}) = 0.2(Y_{t-1} - Y_{t-2}) + e_t \) .
**Important point:** For ARIMA models only, stationarity is determined by the roots of the autoregressive backshift operator. If the roots are all greater than 1 in magnitude (some may be complex) then the series will be stationary. If any root is 1, the series needs to be differenced to get standard distributional results and tests. Roots exceeding 1 are rarely if ever encountered in practice – possibly in developing countries experiencing hyper-inflation.

More on the term “Unit Root”

As above, consider a backshift operator on the autoregressive side as though it were a polynomial whose roots you were finding back in Middle School. For example, $(1 - 1.3 B + 0.4B^2)=0$ is solved by setting $B=2$ (1-2.6+1.6) or $B=1.25$ (1-1.625+0.625). You can also see this by factoring: $(1-.5B)(1-.8B)$ becomes 0 if either factor is 0 ($B=1/.5$ or $B=1/.8$). These two values of $B$ are called “roots” of the backshift polynomial. This case requires no differencing – it is already stationary. If you differed anyway, you would induce a unit root in the moving average operator. This results in nonoptimal forecasts.

To test $\text{H}_0$: unit roots versus $\text{H}_1$: stationary in AR(2) model,

1. Regress $(Y_t-Y_{t-1})$ on 1, $Y_{t-1}$, and $(Y_{t-1}-Y_{t-2})$ (and maybe t) with ordinary least squares (PROC REG)
2. Compare student t test on $Y_{t-1}$ to the $\tau$ tables above (same as for one lag model)
3. If test is insignificant ($p>0.05$), difference the data.
   If test is significant ($p<0.05$) the series is already stationary.

For AR($p$) add $(Y_{t-2}-Y_{t-3})$, $(Y_{t-3}-Y_{t-4})$, ..., $(Y_{t-p+1}-Y_{t-p})$ to the regression

Fortunately all of this is built into PROC ARIMA so you can bypass PROC REG and the tables unless you are unsure how many of these “augmenting lagged differences” $Y_{t-j}-Y_{t-j-1}$ are needed.
(1) Lagged differences are called augmenting lags or augmenting lagged differences. Test is ADF: “Augmented Dickey-Fuller” test.

(2) Unsure of p? Use t tests from PROC REG on augmenting lags and compare to usual t distribution! 😊

Example: Monthly stocks of silver in the NY commodities exchange (historical data). Two models: AR(2), and AR(1) on the first differences.

Step 1: Decide visually if linear trend (case 3) is needed. (probably not here)
Step 2: Decide on p (overfit with several augmenting lagged differences and delete insignificant ones using usual t and F tests)
Step 3: Add the appropriate number of augmenting lagged differences and compare the $Y_{t-1}$ student t statistic to the correct $\tau$ distribution.

Run the demo silver.sas here.

What actually happened next?
A commonly held misconception is that if a series is increasing or decreasing over time, you need to difference it. For example, if you have

\[ Y_t = \alpha + \beta t + e_t \]

then the best estimates and forecast are given by just regressing \( Y \) on an intercept and \( t \) because \( e_t \) is just white noise. This is just a standard simple linear regression from a beginning stat course. There is no need to difference. If you difference anyway, what happens?

\[ Y_t = \alpha + \beta t + e_t \]
\[ Y_{t-1} = \alpha + \beta (t-1) + e_{t-1} \]

so

\[ Y_t - Y_{t-1} = \beta + e_t - \theta e_{t-1} \]

and you have eliminated the trend but have introduced a unit root (because clearly \( \theta = 1 \)) on the moving average side, resulting in nonoptimal forecasts. It is like using a sledge hammer to kill a fly on your window. Yes, it does kill the fly but do you really want to do that? Differencing does reduce a linear trend to a constant but do you always want to do that? You decide whether to difference by using the case 3 trend model unit root test above.
A real data example is the average yields of corn in the U.S. since 1942 when hybrid corn and modern herbicides came into use. Here are the data from 1866 onward (from NASS)