1

Further Issues in Time Series Regressions: Stationarity and Weakly Dependent Time Series

Hüseyin Taştan[1]

[1]Yıldız Technical University
Department of Economics

Econometrics II

2

Further Issues in Using OLS with Time Series

- In our previous classes, we discussed the finite sample properties of OLS for time series data under stronger sets of assumptions (see ch.10 in the text).
- Under the full set of classical linear model assumptions for time series, TS.1 through TS.6, OLS has exactly the same desirable properties that we derived for cross-sectional data. Likewise, statistical inference is carried out in the same way as it was for cross-sectional analysis.
- From our cross-sectional analysis, we know that there are good reasons for studying the large sample properties of OLS.
- For example, if the error terms are not drawn from a normal distribution, then we must rely on the central limit theorem to justify the usual OLS test statistics and confidence intervals.

3

Further Issues in Using OLS with Time Series

- Even though large time series samples can be difficult to come by; but we often have no choice other than to rely on large sample properties.
- As we will show in Section 11.2, models with lagged dependent variables , such as y_{t-1}, must violate the strict exogeneity assumption (TS.2).
- For cross-sectional analysis in Chapter 5, we obtained the large sample properties of OLS in the context of random sampling.
- But in time series, things are more complicated when we allow the observations to be correlated across time (past and future).
- The **crucial point** in time series analysis is **whether the correlation between the variables at different time periods tends to zero quickly enough.**

4

Stationary and Nonstationary Time Series

- We need new concepts in order to examine the large sample properties of OLS estimators in time series regressions.
- These concepts are: stationarity, weak dependence, non-stationarity, high persistence (strong dependence)
- Historically, the notion of a **stationary process** has played an important role in the analysis of time series.
- There are two broad types of stationarity: strict stationarity, and weak (covariance) stationarity.

(Strictly) Stationary Stochastic Process

Definition:
The stochastic process $\{x_t : t = 1, 2, ...\}$ is (strictly) stationary if for every collection of time indices $1 \leq t_1 \leq t_2 \leq ... \leq t_m$, the joint distribution of $(x_{t1}, x_{t2}, ..., x_{tm})$ is the same as the joint distribution of $(x_{t1+h}, x_{t2+h}, ..., x_{tm+h})$ for all integers $h \geq 1$.

- In other words, the sequence $\{x_t : t = 1, 2, ...\}$ is identically distributed.
- A strictly stationary time series process is one whose probability distributions are stable over time.
- If we take any collection of random variables in the sequence and then shift that sequence ahead h time periods, the joint probability distribution must remain unchanged.
- We still allow them to be correlated but the correlation structure does not change over time.

Covariance-Stationary Time Series

It's almost impossible to show strict stationarity. In practice, we use a weaker definition of stationarity that is much easier to work with.

Definition:
If $\{x_t : t = 1, 2, ...\}$ has a finite second moment, ($\mathrm{E}(x_t^2) < \infty$) for all t, provided that the following 3 conditions are satisfied, the stochastic process is called **covariance stationary**.

1. $\mathrm{E}(x_t)$ is constant (independent of t).
2. $\mathrm{Var}(x_t)$ is constant (independent of t).
3. for any t and $h \geq 1$, $\mathrm{Cov}(x_t, x_{t+h})$ depends only on h and not on t.

The Notion of Covariance Stationarity

- **Covariance stationarity** focuses only on the first two moments of a stochastic process: the mean and variance of the process are constant across time.
- And the covariance (of course correlation) between x_t and x_{t+h} depends only on the distance between the two terms, h, and not on t.
- Covariance stationarity is also called weak stationarity.
- **Strict stationarity** is a stronger requirement than covariance stationarity. So we do not focus on the strict stationarity in this course.
- We will call a series stationary if it satisfies the conditions for covariance stationarity.

The notion of Stationarity

- In regression analysis using time series data, stationarity is of prime importance.
- Stationarity simplifies statements of the law of large numbers and the central limit theorem. These theorems can only be applied to stationary processes.
- In order to understand the relationship between two or more variables using regression analysis, we need to assume some sort of **stability over time**.
- If the relationship between two variables (say, y_t and x_t) changes arbitrarily in each time period, then it may not be possible to learn much about how a change in one variable affects the other variable if we only have access to a single time series realization.
- Stationarity also implies that the β_j parameters do not change over time.

Weakly Dependent Time Series

- Stationarity is related to the joint distributions of a process as it moves through time.
- Another concept is weak dependence which places restrictions on how strongly related the random variables x_t and x_{t+h} can be as the time distance between them, h, gets large.

Definition: Weakly Dependent Time series

A stationary time series process is said to be weakly dependent if x_t and x_{t+h} are almost independent as h increases without bound, i.e. **asymptotically uncorrelated**:

$$As \quad h \to \infty \ , \quad Corr(x_t, x_{t+h}) \to 0$$

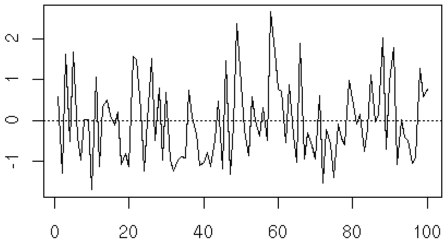
Weakly Dependent Time Series

$$As \quad h \to \infty \quad , \quad Corr(x_t, x_{t+h}) \to 0$$

- In practice, covariance stationary sequences can be characterized in terms of correlations: a covariance stationary time series is weakly dependent if the correlation between x_t and x_{t+h} goes to zero **sufficiently quickly** as h gets large.
- Weak dependence replaces the assumption of random sampling in implying that the law of large numbers (LLN) and the central limit theorem (CLT) hold.
- The central limit theorem for time series data requires stationarity and weak dependence: thus, **stationary, weakly dependent time series are ideal for use in multiple regression analysis.**

Weakly Dependent Time Series

- The simplest example of a weakly dependent time series is an **independent, identically distributed (i.i.d)** sequence : a sequence that is independent is trivially weakly dependent.
- A series randomly drawn from a normal distribution is a an example.
- A time series sequence of 100 drawn from the standard normal distribution is graphed below.



White Noise Process

Definition: White Noise Process

A stochastic process, $\{e_t : t = 1, 2, \ldots\}$, is called **white noise process** if this process satisfies the following conditions.

$$\begin{aligned} \mathrm{E}[e_t] &= 0 \\ \mathrm{Var}(e_t) &= \sigma_e^2 \\ \mathrm{Cov}(e_t, e_s) &= 0, \quad t \neq s \end{aligned}$$

We denote white noise process as $e_t \sim wn(0, \sigma_e^2)$.

- If $\{e_t\}_{t=1}^T$ also follows a Normal (Gaussian) distribution with mean 0, variance σ_e^2 , this process is called Normal (Gaussian) White Noise process. Notation: $e_t \sim GWN(0, \sigma_e^2)$. If this process is i.i.d., the notation is $e_t \sim iid\ N(0, \sigma_e^2)$.

Example: Moving Average(MA) Process

MA(1)

- A more interesting example of a weakly dependent sequence is called a moving average process of order one, **MA(1)**.

$$x_t = e_t + \alpha_1 e_{t-1}, \quad t = 1, 2, \ldots$$

e_t is an iid white noise process.

- x_t is a weighted average of e_t and e_{t-1}.
- Why is an MA(1) process weakly dependent? Adjacent x terms in the sequence are correlated.
- For example, if we rewrite the above model for $t+1$:

$$x_{t+1} = e_{t+1} + \alpha_1 e_t.$$

- The (unconditional) expected value of this process:

$$\begin{aligned} \mathsf{E}(x_t) &= \mathsf{E}(e_t + \alpha_1 e_{t-1}) \\ &= \mathsf{E}(e_t) + \alpha_1 \mathsf{E}(e_{t-1}) = 0 = \mu \end{aligned}$$

MA(1) Process

- The variance of MA(1) process:

$$\begin{aligned} \mathsf{Var}(x_t) &= \mathsf{E}[(x_t - \mathsf{E}(x_t))^2] = \mathsf{E}(x_t^2) \\ &= \mathsf{E}((e_t + \alpha_1 e_{t-1})^2) \\ &= \mathsf{E}(e_t^2) + \alpha_1^2 \mathsf{E}(e_{t-1}^2) \\ &= \sigma_e^2 + \alpha_1^2 \sigma_e^2 = (1+\alpha_1^2)\sigma_e^2 \end{aligned}$$

- The first autocovariance:

$$\begin{aligned} \mathsf{Cov}(x_t, x_{t-1}) &= \mathsf{E}[(x_t - \mu)(x_{t-1} - \mu)] \\ &= \mathsf{E}[(e_t + \alpha_1 e_{t-1})(e_{t-1} + \alpha_1 e_{t-2})] \\ &= \mathsf{E}[e_t e_{t-1} + \alpha_1 e_t e_{t-2} + \alpha_1 e_{t-1}^2 + \alpha_1^2 e_{t-1} e_{t-2}] \\ &= \mathsf{E}(e_t e_{t-1}) + \alpha_1 \mathsf{E}(e_t e_{t-2}) + \alpha_1 \mathsf{E}(e_{t-1}^2) \\ &\quad + \alpha_1^2 \mathsf{E}(e_{t-1} e_{t-2}) \\ &= 0 + 0 + \alpha_1 \mathsf{E}(e_{t-1}^2) + 0 \\ &= \alpha_1 \sigma_e^2 \end{aligned}$$

MA(1) Process

- The first autocorrelation:

$$\rho_1 = Corr(x_t, x_{t-1}) = \frac{\mathsf{Cov}(x_t, x_{t-1})}{\mathsf{Var}(x_t)} = \frac{\alpha_1}{1+\alpha_1^2}$$

- The sign of the first autocorrelation depends on the sign of α_1.
- For example, if $\alpha_1 = 0.5$, $Cor[x_t, x_{t-1}] = 0.40$.
- The maximum positive correlation occurs when $\alpha_1 = 1$; in which case, $Cor[x_t, x_{t-1}] = 0.50$.

MA(1) Process

- When we look at variables in the sequence that are two or more time periods apart, these variables are uncorrelated because they are independent.
- For example, $x_{t+2} = e_{t+2} + \alpha_1 e_{t+1}$ is independent of x_t
- Because because e_t is independent across t.
- Due to the identical distribution assumption on the e_t, x_t is actually stationary.
- Its mean, variance and autocovariance are independent of t.
- So, an MA(1) is a stationary, weakly dependent sequence, and the law of large numbers and the central limit theorem can be applied.

17

Example: Autoregressive Process of Order One, AR(1)

AR(1)

$$y_t = \rho_1 y_{t-1} + e_t, \quad t = 1, 2, \ldots$$

- The starting point in the sequence is y_0 (at $t = 0$), and $e_t : t = 1, 2, \ldots$ is an i.i.d. sequence with zero mean and variance σ_e^2 (white noise). We also assume that the e_t are independent of y_0 and that $E(y_0) = 0$ 0.
- The crucial assumption for weak dependence of an AR(1) process is the stability condition $|\rho_1| < 1$. Then we say that y_t is a stable AR(1) process.
- For a stable AR(1) process ($|\rho_1| < 1$),

$$\sigma_y^2 = \sigma_e^2/(1 - \rho_1^2)$$

18

AR(1) Process

- Since σ_y is the standard deviation of both y_t and y_{t+h}, we can easily find the correlation between them for any $h \geq 1$:

$$Corr(y_t, y_{t+h}) = Cov(y_t, y_{t+h})/(\sigma_y \sigma_y) = \rho_1^h$$

- In particular, $Corr(y_t, y_{t+1}) = \rho_1$ and so ρ_1 is the correlation coefficient between any two adjacent terms in the sequence.
- It shows that, while y_t ve y_{t+h} are correlated for any $h \geq 1$, this correlation gets very small for large h.

$$Since \ \ |\rho_1| < 1 \ , \ \ \rho_1^h \to 0 \ \text{ as } \ h \to \infty$$

- For example, if $\rho_1 = 0.9$,then $Corr(y_t, y_{t+20}) = 0.122$
- Hence, a stable AR(1) model is weakly dependent.

19

Trend-Stationary Process

- A trending series can not be weakly dependent.
- A series that is stationary about its time trend, as well as weakly dependent, is often called a trend-stationary process.
- After detrending, if the remaining part of the trending series is stationary, these series are trend-stationary, weekly dependent.
- Deterministic linear trend process:

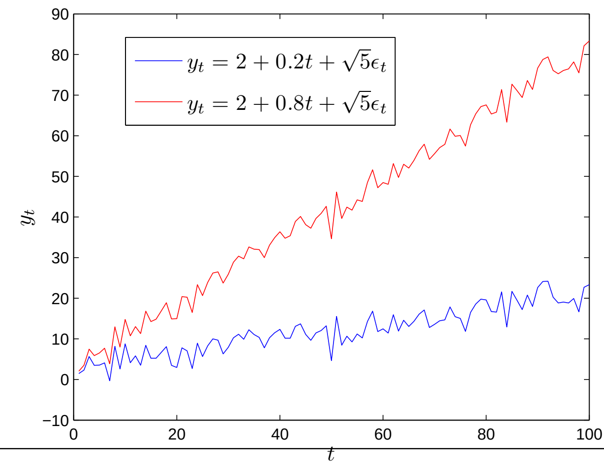
$$y_t = \beta_0 + \beta_1 t + \epsilon_t, \quad \epsilon_t \sim \ wn(0, \sigma^2)$$

20

Example: Deterministic linear trend process

$$y_t = \beta_0 + \beta_1 t + \epsilon_t, \quad \epsilon_t \sim \ wn(0, \sigma^2)$$

Graph: Two linearly trending series with different slopes



21

Highly Persistent Time Series

- We need **weakly dependent** variables for OLS to be consistent. In that case, the standard inference procedures will still be valid.
- However, many economic time series **cannot** be characterized by weak dependence, but **strong dependence**.
- **Highly persistent or strongly dependent** time series display high correlation with its past values.
- Let's see some examples of highly persistent or strongly dependent time series.

22

Highly Persistent Time Series

- Many economic time series are **strongly dependent** or **highly persistent** series. For example, inflation rate, budget deficit etc.
- Using time series with strong dependence in regression analysis poses no problem, if the TS assumptions hold (see ch.10 in text).
- But the usual inference procedures are not valid when the data are not weakly dependent. The reason is that CLT and LLN do not apply.
- In the AR(1) model , the assumption $|\rho_1| < 1$ is crucial for the series to be weakly dependent.
- It turns out that many economic time series are better characterized by the AR(1) model with $\rho_1 = 1$. As ρ_1 approaches 1, the more the time series gets persistent.
- When $\rho_1 = 1$, AR(1) process is called **Random Walk**.

23

Random Walk

- When $\rho_1 = 1$, AR(1) model can be written

$$y_t = y_{t-1} + e_t, t = 1, 2, ...$$

- $\{e_t : t = 1, 2, ...\}$ is independent and identically distributed with mean zero and variance σ_e^2. We assume that the initial value, y_0, is independent of e_t for all $t \geq 1$.
- The expected value of y_t can be found by the method of repeated substitution:

$$y_t = e_t + e_{t-1} + e_{t-2} + ... + e_1 + y_0$$

- Taking the expectation of both sides gives

$$E(y_t) = E(e_t) + E(e_{t-1}) + E(e_{t-2}) + ... + E(e_1) + E(y_0)$$

$$E(y_t) = E(y_0)$$

- The expected value of random walk process does not depend on t (independent of time). If the initial value is assumed as $y_0 = 0$, $E(y_t) = 0$ is true for all $t \geq 1$.

24

Method of Repeated Substitution): $y_t = y_{t-1} + e_t$

- $y_1 = y_0 + e_1$
- $y_2 = y_1 + e_2 = y_0 + e_1 + e_2$
- $y_3 = y_2 + e_3 = y_0 + e_1 + e_2 + e_3$
- ....
- $y_t = y_0 + e_1 + e_2 + ... + e_{t-1} + e_t = y_0 + \sum_{t=1}^{t} e_t$

Random Walk

- By contrast, the variance of a random walk does change with t. To compute the variance of a random walk, for simplicity we assume that y_0 is nonrandom so that $Var(y_0) = 0$. This does not affect any important conclusions. Then, by the i.i.d. assumption for e_t:

$$Var(y_t) = Var(e_t) + Var(e_{t-1}) + ... + Var(e_1) = \sigma_e^2 t$$

- A random walk displays highly persistent behavior in the sense that the value of y today is significant for determining the value of y in the very distant future. To see this, write for h periods hence

$$y_{t+h} = e_{t+h} + e_{t+h-1} + ... + e_{t+1} + y_t$$

- Now, suppose at time t, we want to compute the expected value of y_{t+h} given the current value y_t. Since the expected value of e_{t+j}, given y_t, is zero for all $j \geq 1$,

$$E(y_{t+h}|y_t) = y_t \ for \ all \ h \geq 1$$

Random Walk

- This means that, no matter how far in the future we look, our best prediction of y_{t+h} is today's value, y_t.
- We can contrast this with the stable AR(1) case, where a similar argument can be used to show that

$$E(y_{t+h}|y_t) = \rho_1^h y_t, \ for \ all \ h \geq 1$$

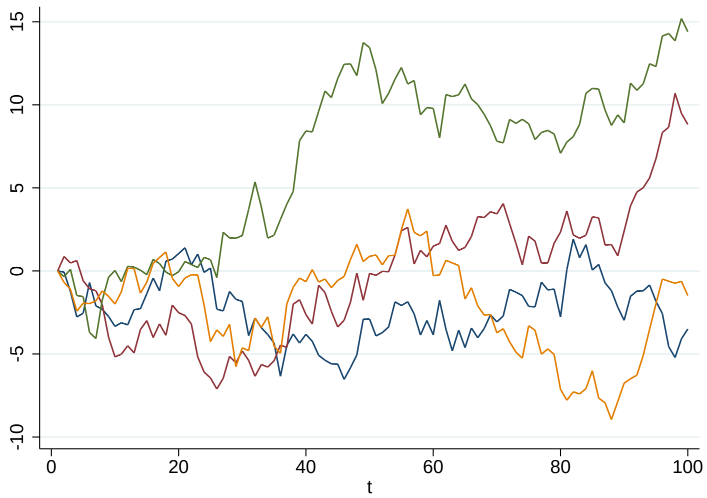
- If $|\rho_1| < 1$, this expectation approaches to zero as $h \to \infty$.
- We can also see that the correlation between y_t and y_{t+h} is close to 1 for large t when y_t follows a random walk. If $Var(y_0) = 0$, it can be shown that,

$$Corr(y_t, y_{t+h}) = \sqrt{t/(t+h)}$$

- So, a **random walk** process is a **nonstationary** process.

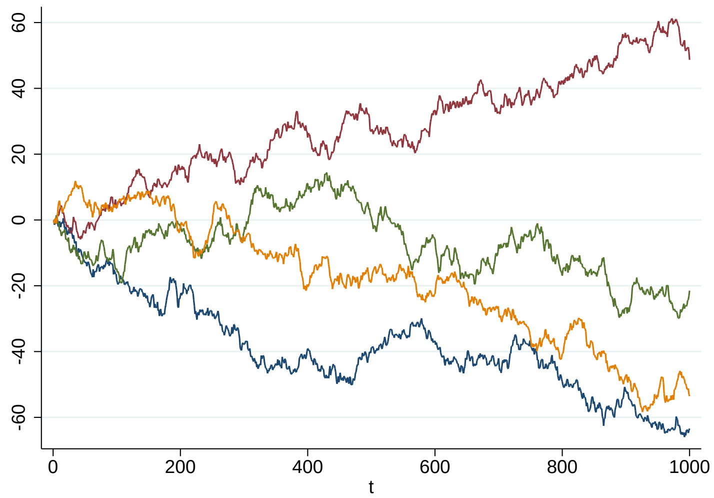
Random Walk

Four realizations of random walk process with $y_0 = 0$, $T = 100$



Random Walk

Four realizations of random walk process with $y_0 = 0$, $T = 1000$



Random Walk

- A random walk is a special case of what is known as a unit root process. The name comes from the fact that $\rho_1 = 1$ in the AR(1) model.
- A more general class of unit root processes can be generated by defining various e_t processes.
- But, e_t is now allowed to be a general, weakly dependent series. For example, e_t could itself follow an MA(1) or a stable AR(1) process.
- When e_t is not an i.i.d. sequence, the properties of the random walk we derived earlier no longer hold.

Random Walk

- From a policy perspective, it is often important to know whether an economic time series is highly persistent or not. Consider the case of gross domestic product in Turkey. If GDP is asymptotically uncorrelated, then the level of GDP in the coming year is at best weakly related to what GDP was, say, thirty years ago. This means a policy that affected GDP long ago has very little lasting impact.
- GDP is strongly dependent, then next year's GDP can be highly correlated with the GDP from many years ago. Then, we should recognize that a policy which causes a discrete change in GDP can have persisting and long-lasting effects.
- Effect of shocks are very persistent and lasting in random walk models.

Random Walk with Drift

- It is often the case that a highly persistent series also contains a clear trend.
- One model that leads to this behavior is called the **random walk with drift**.

$$y_t = \alpha_0 + y_{t-1} + e_t, t = 1, 2, ...$$

- where $\{e_t : t = 1, 2, ...\}$ and y_0 satisfy the same properties as in the random walk model. What is new is the parameter α_0, which is called the drift term. Essentially, to generate y_t, the constant α_0 is added along with the random noise et to the previous value y_{t-1}
- We can show that the expected value of y_t follows a linear time trend by using repeated substitution

$$y_t = \alpha_0 t + e_t + e_{t-1} + ... + e_1 + y_0$$

Random Walk with Drift

- Therefore, if $y_0 = 0$, $E(y_t) = \alpha_0 t$: the expected value of y_t is growing over time if $\alpha_0 > 0$ and shrinking over time if $\alpha_0 < 0$.
- By reasoning as we did in the pure random walk case, we can show that

$$E(y_{t+h}|y_t) = \alpha_0 h + y_t$$

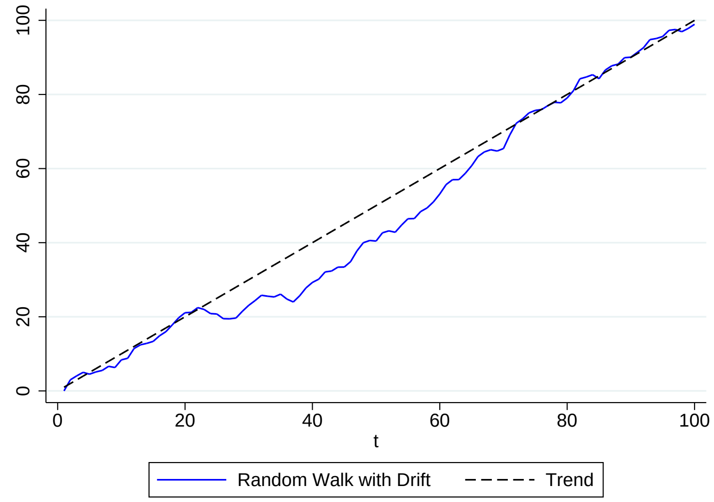
- So the best prediction of y_{t+h} at time t is y_t plus the drift $\alpha_0 h$. The variance of y_t is the same as it was in the pure random walk case.

Figure 11.2

The U.S. three-month T-bill rate, for the years 1948–1996.

Random Walk

A single realization of random walk with drift process with $y_0 = 0$, $T = 100$



Transformations on Highly Persistent Time Series

- Using time series with **strong persistence** of the type displayed by a **unit root process** in a regression equation can lead to very misleading results if the CLM assumptions are violated.
- We will study the **spurious regression** problem in more detail in Chapter 18.
- But for now we must be aware of potential problems. Fortunately, simple transformations are available that render a unit root process weakly dependent.
- Weakly dependent processes are said to be **integrated of order zero**, denoted by **I(0)**. Practically, this means that nothing needs to be done to such series before using them in regression analysis: averages of such sequences already satisfy the standard limit theorems.

Transformations on Highly Persistent Time Series

- Unit root processes, such as a random walk (with or without drift), are said to be **integrated of order one**, or **I(1)**. This means that the **first difference** of the process is weakly dependent (and often stationary).
- Let us consider a random walk(RW) process: $y_t = y_{t-1} + e_t$. Subtracting y_{t-1} from both sides gives the first difference:

$$\Delta y_t = y_t - y_{t-1} = e_t, t = 2, 3, ...$$

- Δy_t is weakly dependent, I(0).
- Many economic time series y_t that are strictly positive are such that $log(y_t)$ is I(1).
- In this case, we can use the first difference in the logs in regression analysis,

$$\Delta \log y_t = \log y_t - \log y_{t-1}$$

- $log(y_t)$ is I(1) and the growth rate of y_t, Δy_t, is I(0).

Transformations on Highly Persistent Time Series

- Differencing time series before using them in regression analysis has another benefit: it removes any linear time trend. This is easily seen by writing a linearly trending variable as:

$$y_t = \gamma_0 + \gamma_1 t + v_t$$

- Writing y_t for $t-1$ gives

$$y_{t-1} = \gamma_0 + \gamma_1(t-1) + v_{t-1}$$

- The first difference is

$$\Delta y_t = y_t - y_{t-1} = \gamma_1 + \Delta v_t$$

- The expected value of the final expression

$$E[\Delta y_t] = \gamma_1 + E[\Delta v_t] = \gamma_1$$

- As seen, $E[\Delta y_t]$ is equal to a constant. It implies that the mean of the first difference of the trending series is a constant and we can use this stationary series in regression analysis.

Deciding Whether a Time Series Is I(1) or I(0)

- Some statistical tests (unit root tests) can be used for this purpose, we provide an introductory treatment in Chapter 18.
- There are informal methods that provide useful guidance about whether a time series process is roughly characterized by weak dependence.
- A very simple tool is motivated by the AR(1) model: if ρ_1 is absolutely less than 1, then the process is I(0), but it is I(1) if ρ_1 is 1.
- we know that $\rho_1 = corr(y_t, y_{t-1})$ and the correlation, ρ_1, between y_t and y_{t-1} can be estimated.
- This sample correlation coefficient is called the **first order autocorrelation** of y_t; we denote this by $\hat{\rho}_1$.
- When $|\rho_1| < 1$, $\hat{\rho}_1$ is a consistent but biased estimator of ρ_1.
- We can use the value of $\hat{\rho}_1$ to help decide whether the process is I(1) or I(0).

Deciding Whether a Time Series Is I(1) or I(0)

- Ideally, we could compute a confidence interval for ρ_1 to see if it excludes the value $\rho_1 = 1$, but this turns out to be rather difficult: the sampling distributions of the estimator of $\hat{\rho}_1$ are extremely different when ρ_1 is close to one and when ρ_1 is much less than one.
- In fact, when ρ_1 is close to one, ρ_1 can have a severe downward bias.
- We can only use $\hat{\rho}_1$ as a rough guide for determining whether a series needs to be differenced.
- Most economists think that differencing is warranted if $\hat{\rho}_1$ is greater than 0.90. Some would difference when $\hat{\rho}_1$ is greater than 0.80.
- When the series has an trend, it makes more sense to obtain the first order autocorrelation after detrending. If the data are not detrended, the autoregressive correlation tends to be overestimated, which biases toward finding a unit root in a trending process.