# Lecture 2:

# Further Issues Using OLS with Time Series Data

# Stationary Time Series

- **The assumptions used so far seem to be too restricitive**
  - Strict exogeneity, homoscedasticity, and no serial correlation are very demanding requirements, especially in the time series context
  - Statistical inference rests on the validity of the normality assumption
  - Much weaker assumptions are needed if the sample size is large
  - *A key requirement  for large sample analysis of time series is that the time series in question are stationary and weakly dependent*
- **Stationary time series**
  - Loosely speaking, a time series is stationary if its stochastic properties and its temporal dependence structure do not change over time

# Stationary Time Series (cont.)

- **Stationary stochastic processes (quá trình ngẫu nhiên dừng)**

  A stochastic process $\{x_t : t = 1, 2, \ldots\}$ is stationary, if for every collection of indices $1 \leq t_1 \leq t_2 \leq \cdots \leq t_m$ the joint distribution of $(x_{t_1}, x_{t_2}, \ldots, x_{t_m})$ is the same as that of $(x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_m+h})$ for all integers $h \geq 1$.

- **Covariance stationary processes (quá trình dừng hiệp phương sai)**

  A stochastic process $\{x_t : t = 1, 2, \ldots\}$ is covariance stationary, if its expected value, its variance, and its covariances are constant over time:

  1) $E(x_t) = \mu$, 2) $Var(x_t) = \sigma^2$, and 3) $Cov(x_t, x_{t+h}) = f(h)$.

# Implications of Stationarity

- Mean is stationary: $\mu_y = E(y_t) = E(y_{t+m})$
  → Constant mean (equilibrium) level
  → These time series will exhibit mean reversion

- Variance is stationary: $\sigma^2_y = E[(y_t - \mu_y)^2] = E[(y_{t+m} - \mu_y)^2]$
  → Probability of fluctuation from mean level is same at any point in time

- Covariance (for any lag k) is stationary: $\gamma_k = Cov(y_t, y_{t+k}) = E[(y_t - \mu_y)(y_{t+k} - \mu_y)]$
  $= E[(y_{t+m} - \mu_y)(y_{t+m+k} - \mu_y)]$

  Implies that covariance only depends on lag length, not on point in time

# Remarks

- The $y$ series is stationary when its mean and variance are constant across time and the covariance between $y_t$ and $y_{t-h}$ (and $y_{t+h}$) depends only on the distance between two terms, i.e., $h$, and not on the specific value of $t$. It follows immediately that the correlation also depends only on distance, $h$.

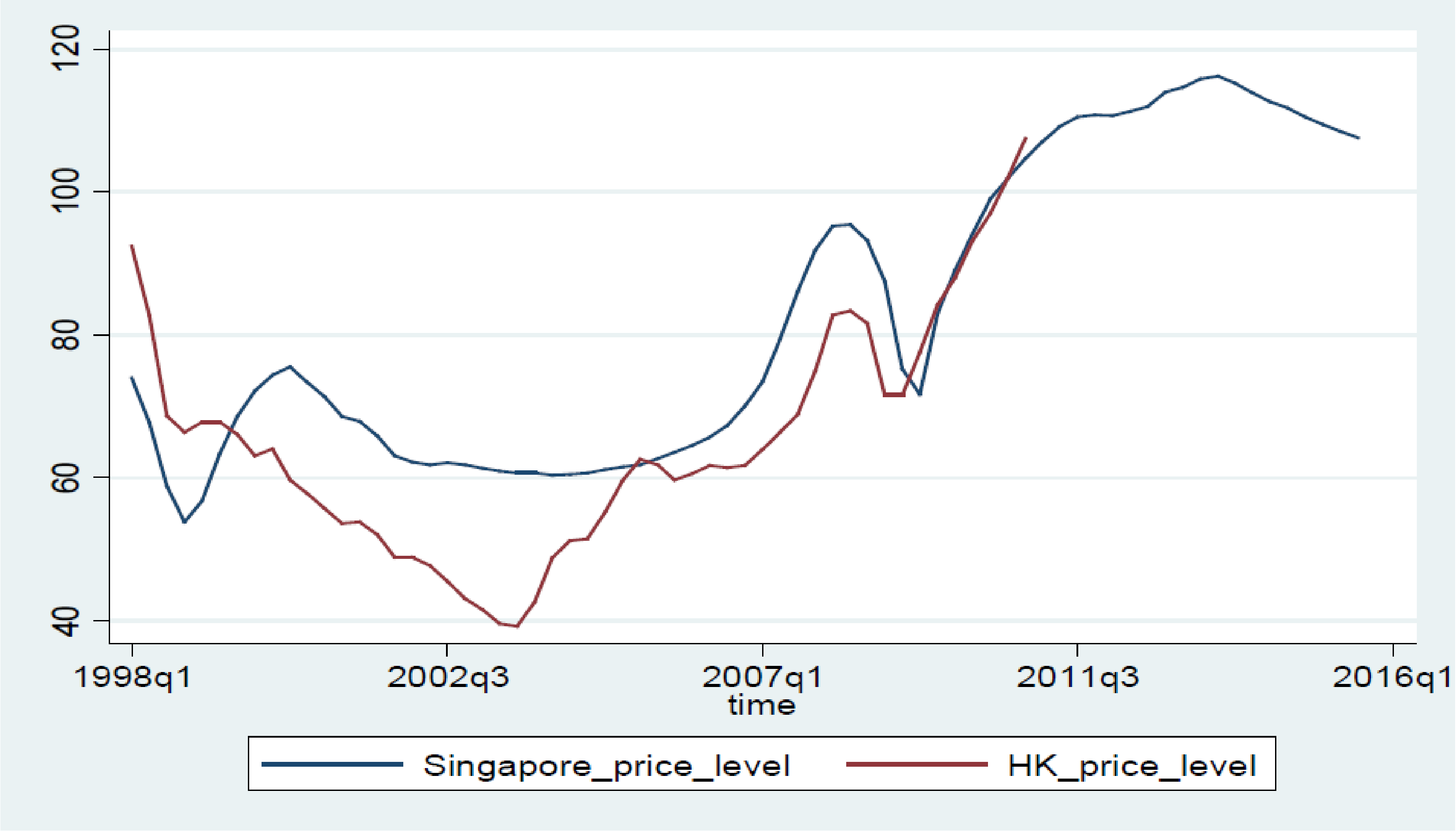
- If the series is not (weakly) stationary, then it is non-stationary series. A NON-stationary series will have a time-varying mean or a time-varying variance or both. For example, exchange rates and housing prices are usually non-stationary.

# Remark: Why do we need to make sure that the time series process is at least covariance stationary?

- Practically speaking, if we want to understand the relationship between two or more variables using regression analysis, we need to assume some sort of stationarity. In other words, if we allow the relationship between two variables, say $yt$ and $xt$, to change arbitrarily over time, then we cannot hope to learn much about a change in one variable affects the other variable.

- If the time series is not stationary, we can study its behavior only for the time period consideration. Each set of time series data will therefore be for a particular episode. Therefore, for the purpose of forecasting time series may be of little practical value.

# Example: Can we compare the "indexes" between Singapore and Hong Kong directly?



# Example: Are the following time series data covariance stationary?

- Vietnam's inflation rate
- Vietnam's consumer price index (CPI)
- Vietnam's GDP

- Take data plot in Stata
- Code: line var time_var

# Weakly dependent time series

- **Weakly dependent time series**

  A stochastic process $\{x_t : t = 1, 2, \ldots\}$ is weakly dependent , if $x_t$ is "almost independent" of $x_{t+h}$ if $h$ grows to infinity (for all $t$ ).

- **Discussion of the weak dependence (phụ thuộc yếu) property**
  - An implication of weak dependence is that the correlation between $x_t$ and $x_{t+h}$ must converge to zero if $h$ grows to infinity
  - (Note that a series may be nonstationary but weakly dependent)

# Stationarity vs. Weak Dependency

- **Stationarity (tính dừng)** deals with joint distribution (phân phối đồng thời) being same over time.
- **Weak dependency (phụ thuộc yếu)** deals with how strongly related xt and xt+h are as distance (h) gets large. As k increases, if xt and xt+h are "almost indepdendent," then is weakly stationary.
- Weak dependency/stationarity: This assumption replaces assumption of random sampling → allows LLN and CLT to hold to get consistent OLS estimates.

# Examples for weakly dependent time series

- **Moving average process of order one (MA(1)): Quá trình trung bình trượt bậc nhất**

$$x_t = e_t + \alpha_1 e_{t-1}$$

The process is a short moving average of an i.i.d. series $e_t$

The process is weakly dependent because observations that are more than one time period apart have nothing in common and are therefore uncorrelated.

- **Autoregressive process of order one (AR(1)): Quá trình tự hồi quy bậc nhất**

$$y_t = \rho_1 y_{t-1} + e_t$$

The process carries over to a certain extent the value of the previous period (plus random shocks from an i.i.d. series $e_t$)

$$\Rightarrow \quad Corr(y_t, y_{t+h}) = \rho_1^h$$

If the stability condition $|\rho_1| < 1$ holds, the process is weakly dependent because serial correlation converges to zero as the distance between observations grows to infinity.

# Why is an MA(1) process weakly dependent?

For example, $x_{t2}$ is independent of $x_t$ because $\{e_t\}$ is independent across $t$.

Due to the identical distribution assumption on the $e_t$, $\{x_t\}$ in (11.1) is actually stationary.

Thus, an MA(1) is a stationary, weakly dependent sequence, and the law of large numbers and the central limit theorem can be applied to $\{x_t\}$.

# Autoregressive process of order one (AR(1)): Quá trình tự hồi quy bậc nhất

Now, we can find the covariance between $y_t$ and $y_{t+h}$ for $h \geq 1$. Using repeated substitution,

$$\begin{aligned} y_{t+h} &= \rho_1 y_{t+h-1} + e_{t+h} = \rho_1(\rho_1 y_{t+h-2} + e_{t+h-1}) + e_{t+h} \\ &= \rho_1^2 y_{t+h-2} + \rho_1 e_{t+h-1} + e_{t+h} = \dots \\ &= \rho_1^h y_t + \rho_1^{h-1} e_{t+1} + \dots + \rho_1 e_{t+h-1} + e_{t+h}. \end{aligned}$$

Because $\mathrm{E}(y_t) = 0$ for all $t$, we can multiply this last equation by $y_t$ and take expectations to obtain $\mathrm{Cov}(y_t, y_{t+h})$. Using the fact that $e_{t+j}$ is uncorrelated with $y_t$ for all $j \geq 1$ gives

$$\begin{aligned} \mathrm{Cov}(y_t, y_{t+h}) = \mathrm{E}(y_t y_{t+h}) &= \rho_1^h \mathrm{E}(y_t^2) + \rho_1^{h-1}\mathrm{E}(y_t e_{t+1}) + \dots + \mathrm{E}(y_t e_{t+h}) \\ &= \rho_1^h \mathrm{E}(y_t^2) = \rho_1^h \sigma_y^2. \end{aligned}$$

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

$$\mathrm{Corr}(y_t, y_{t+h}) = \mathrm{Cov}(y_t, y_{t+h})/(\sigma_y \sigma_y) = \rho_1^h. \qquad \boxed{11.4}$$

In particular, $\mathrm{Corr}(y_t, y_{t+1}) = \rho_1$, so $\rho_1$ is the correlation coefficient between any two adjacent terms in the sequence.

# Asymptotic properties of OLS (tính chất tiệm cận của OLS)

- **Assumption TS.1` (Linear in parameters)**
  - Same as assumption TS.1 but now the dependent and independent variables are assumed to be <u>stationary</u> and <u>weakly dependent</u>
- **Assumption TS.2` (No perfect collinearity)**
  - Same as assumption TS.2
- **Assumption TS.3` (Zero conditional mean)**
  - Now the explanatory variables are assumed to be only ***contempo-raneously*** exogenous rather than strictly exogenous, i.e.

$$E(u_t|\mathbf{x}_t) = 0$$

The explanatory variables <u>of the same period</u> are uninformative about the mean of the error term

# Theorem 11.1 (Consistency of OLS)

- **Theorem 11.1 (Consistency of OLS)**

$$TS.1' - TS.3' \quad \Rightarrow \quad plim\ \widehat{\beta}_j = \beta_j, \quad j = 0, 1, \ldots, k$$

Important note: For consistency it would even suffice to assume that the explanatory variables are merely contemporaneously *uncorrelated* with the error term.

- **Why is it important to relax the strict exogeneity assumption?**
  - Strict exogeneity is a serious restriction because it rules out all kinds of dynamic relationships between explanatory variables and the error term
  - In particular, it rules out feedback from the dep. var. on future values of the explanat. variables (which is very common in economic contexts)
  - Strict exogeneity precludes the use of lagged dep. var. as regressors

# Review: Probability limit

So here is the definition of a probability limit.

*Definition:* Let $X_1, X_2, X_3, \ldots$ be a sequences of random variables and let $X$ be a random variable. $X_n \to X$ in probability if for every $\varepsilon > 0$ we have

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0.$$

# Theorem 11.1 (Consistency of OLS) (cont.)

- **Why do lagged dependent variables violate strict exogeneity?**

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

This is the simplest possible regression model with a lagged dependent variable

<u>Contemporanous exogeneity:</u> $E(u_t | y_{t-1}) = 0$

<u>Strict exogeneity:</u> $E(u_t | y_0, y_1, \ldots, y_{n-1}) = 0$

Strict exogeneity would imply that the error term is uncorrelated with all $y_t$, t=1, … , n-1

This leads to a contradiction because:

$$Cov(y_t, u_t) = \beta_1 Cov(y_{t-1}, u_t) + Var(u_t) > 0$$

- **OLS estimation in the presence of lagged dependent variables**
  - Under contemporaneous exogeneity, OLS is consistent but biased

# Asymptotic properties of OLS (cont.)

- **Assumption TS.4�much (Homoscedasticity): (phương sai thuần nhất)**

$$Var(u_t|\mathbf{x}_t) = Var(u_t) = \sigma^2$$

The errors are contemporaneously homoscedastic

- **Assumption TS.5‵ (No serial correlation): (không có tương quan chuỗi)**

$$Corr(u_t, u_s|\mathbf{x}_t, \mathbf{x}_s) = 0, \;\; t \neq s$$

Conditional on the explanatory variables in periods t and s, the errors are uncorrelated

- **Theorem 11.2 (Asymptotic normality of OLS): (tính tiệm cận chuẩn của OLS)**
  - Under assumptions TS.1‵ – TS.5‵, the OLS estimators are asymptotically normally distributed. Further, the usual OLS standard errors, t-statistics and F-statistics are asymptotically valid.

# Example: Efficient Markets Hypothesis (EMH)

- **Example: Efficient Markets Hypothesis (EMH)**

The EMH in a strict form states that information observable to the market prior to week t should not help to predict the return during week t. A simplification assumes in addition that only past returns are considered as relevant information to predict the return in week t.This implies that

$$E(return_t | return_{t-1}, return_{t-2}, \ldots) = E(return_t)$$

A simple way to test the EMH is to specify an AR(1) model. Under the EMH assumption,TS.3' holds so that an OLS regression can be used to test whether this week's returns depend on last week's.

$$\widehat{return}_t = \underset{(.081)}{.180} + \boxed{\underset{(.038)}{.059}}\, return_{t-1}$$

$$n = 689, R^2 = .0035, \bar{R}^2 = .0020$$

There is no evidence against the EMH. Including more lagged returns yields similar results.

# Example: Efficient Markets Hypothesis (EMH) (cont'd)

In the previous example, using an AR(1) model to test the EMH might not detect correlation between weekly returns that are more than one week apart. It is easy to estimate models with more than one lag. For example, an *autoregressive model of order two*, or AR(2) model, is

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + u_t$$
$$\mathrm{E}(u_t | y_{t-1}, y_{t-2}, \ldots) = 0. \quad \mathbf{11.17}$$

There are stability conditions on $\beta_1$ and $\beta_2$ that are needed to ensure that the AR(2) process is weakly dependent, but this is not an issue here because the null hypothesis states that the EMH holds:

$$\mathrm{H}_0\colon \beta_1 = \beta_2 = 0. \quad \mathbf{11.18}$$

If we add the homoskedasticity assumption $\mathrm{Var}(u_t | y_{t-1}, y_{t-2}) = \sigma^2$, we can use a standard $F$ statistic to test (11.18). If we estimate an AR(2) model for $return_t$, we obtain

$$\widehat{return_t} = \underset{(.081)}{.186} + \underset{(.038)}{.060}\, return_{t-1} - \underset{(.038)}{.038}\, return_{t-2}$$
$$n = 688,\ R^2 = .0048,\ \bar{R}^2 = .0019$$

# Exercise C11.3

(i) In Example 11.4, it may be that the expected value of the return at time $t$, given past returns, is a quadratic function of $return_{t-1}$. To check this possibility, use the data in NYSE.RAW to estimate

$$return_t = \beta_0 + \beta_1 return_{t-1} + \beta_2 return_{t-1}^2 + u_t;$$

report the results in standard form.

(ii) State and test the null hypothesis that $E(return_t|return_{t-1})$ does not depend on $return_{t-1}$. (*Hint:* There are two restrictions to test here.) What do you conclude?

(iii) Drop $return_{t-1}^2$ from the model, but add the interaction term $return_{t-1} \cdot return_{t-2}$. Now test the efficient markets hypothesis.

(iv) What do you conclude about predicting weekly stock returns based on past stock returns?

# Example 11.5: Expectations Augmented Phillips Curve

A linear version of the *expectations augmented Phillips curve* can be written as

$$inf_t - inf_t^e = \beta_1(unem_t - \mu_0) + e_t,$$

where $\mu_0$ is the *natural rate of unemployment* and $inf_t^e$ is the *expected* rate of inflation formed in year $t - 1$. This model assumes that the natural rate is constant, something that macroeconomists question. The difference between actual unemployment and the natural rate is called *cyclical unemployment*, while the difference between actual and expected inflation is called *unanticipated inflation*. The error term, $e_t$, is called a *supply shock* by macroeconomists. If there is a tradeoff between unanticipated inflation and cyclical unemployment, then $\beta_1 < 0$. [For a detailed discussion of the expectations augmented Phillips curve, see Mankiw (1994, Section 11.2).]

To complete this model, we need to make an assumption about inflationary expectations. Under *adaptive expectations*, the expected value of current inflation depends on recently observed inflation. A particularly simple formulation is that expected inflation this year is last year's inflation: $inf_t^e = inf_{t-1}$. (See Section 18.1 for an alternative formulation of adaptive expectations.) Under this assumption, we can write

$$inf_t - inf_{t-1} = \beta_0 + \beta_1 unem_t + e_t$$

# Example 11.5: Expectations Augmented Phillips Curve

or

$$\Delta inf_t = \beta_0 + \beta_1 unem_t + e_t,$$

where $\Delta inf_t = inf_t - inf_{t-1}$ and $\beta_0 = -\beta_1\mu_0$. ($\beta_0$ is expected to be positive, since $\beta_1 < 0$ and $\mu_0 > 0$.) Therefore, under adaptive expectations, the expectations augmented Phillips curve relates the *change* in inflation to the level of unemployment and a supply shock, $e_t$. If $e_t$ is uncorrelated with $unem_t$, as is typically assumed, then we can consistently estimate $\beta_0$ and $\beta_1$ by OLS. (We do not have to assume that, say, future unemployment rates are unaffected by the current supply shock.) We assume that TS.1′ through TS.5′ hold. Using the data through 1996 in PHILLIPS.RAW we estimate

$$\begin{aligned} \widehat{\Delta inf_t} &= \underset{(1.38)}{3.03} - \underset{(.230)}{.543}\, unem_t \\ n &= 48,\ R^2 = .108,\ \bar{R}^2 = .088. \end{aligned} \tag{11.19}$$

# Exercise C11.4

Use the data in PHILLIPS.RAW for this exercise, but only through 1996.

(i) In Example 11.5, we assumed that the natural rate of unemployment is constant. An alternative form of the expectations augmented Phillips curve allows the natural rate of unemployment to depend on past levels of unemployment. In the simplest case, the natural rate at time $t$ equals $unem_{t-1}$. If we assume adaptive expectations, we obtain a Phillips curve where inflation and unemployment are in first differences:

$$\Delta inf = \beta_0 + \beta_1 \Delta unem + u.$$

Estimate this model, report the results in the usual form, and discuss the sign, size, and statistical significance of $\hat{\beta}_1$.

(ii) Which model fits the data better, (11.19) or the model from part (i)? Explain.

# Exercise C11.8

Use the data in PHILLIPS.RAW for this exercise.

(i) Estimate an AR(1) model for the unemployment rate. Use this equation to predict the unemployment rate for 2004. Compare this with the actual unemployment rate for 2004. (You can find this information in a recent *Economic Report of the President*.)

(ii) Add a lag of inflation to the AR(1) model from part (i). Is $inf_{t-1}$ statistically significant?

(iii) Use the equation from part (ii) to predict the unemployment rate for 2004. Is the result better or worse than in the model from part (i)?

(iv) Use the method from Section 6.4 to construct a 95% prediction interval for the 2004 unemployment rate. Is the 2004 unemployment rate in the interval?

# Exercise C11.7

Use CONSUMP.RAW for this exercise. One version of the *permanent income hypothesis* (PIH) of consumption is that the *growth* in consumption is unpredictable. [Another version is that the change in consumption itself is unpredictable; see Mankiw (1994, Chapter 15) for discussion of the PIH.] Let $gc_t = \log(c_t) - \log(c_{t-1})$ be the growth in real per capita consumption (of nondurables and services). Then the PIH implies that $\mathrm{E}(gc_t|I_{t-1}) = \mathrm{E}(gc_t)$, where $I_{t-1}$ denotes information known at time $(t-1)$; in this case, $t$ denotes a year.

(i) Test the PIH by estimating $gc_t = \beta_0 + \beta_1 gc_{t-1} + u_t$. Clearly state the null and alternative hypotheses. What do you conclude?

(ii) To the regression in part (i), add $gy_{t-1}$ and $i3_{t-1}$, where $gy_t$ is the growth in real per capita disposable income and $i3_t$ is the interest rate on three-month T-bills; note that each must be lagged in the regression. Are these two additional variables jointly significant?

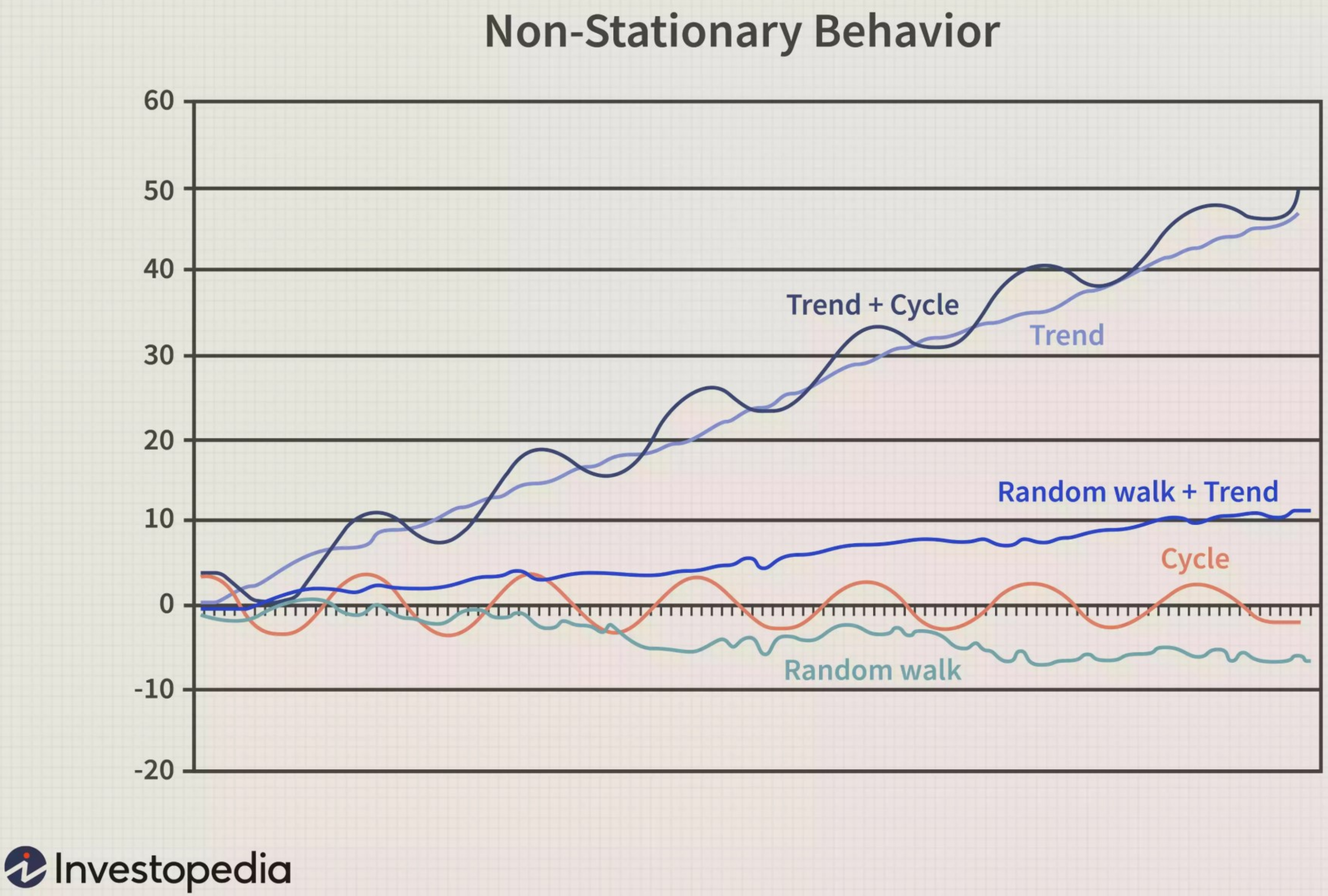
# Using time series in regression analysis

- **Using trend-stationary series in regression analysis**
  - Time series with deterministic time trends are nonstationary
  - If they are stationary around the trend and in addition weakly dependent, they are called trend-stationary processes
  - Trend-stationary processes also satisfy assumption TS.1'
- **<u>Using highly persistent time series in regression analysis</u>**
  - Unfortunately many economic time series violate weak dependence because they are highly persistent (= strongly dependent)
  - In this case OLS methods are generally invalid (unless the CLM hold)
  - In some cases transformations to weak dependence are possible

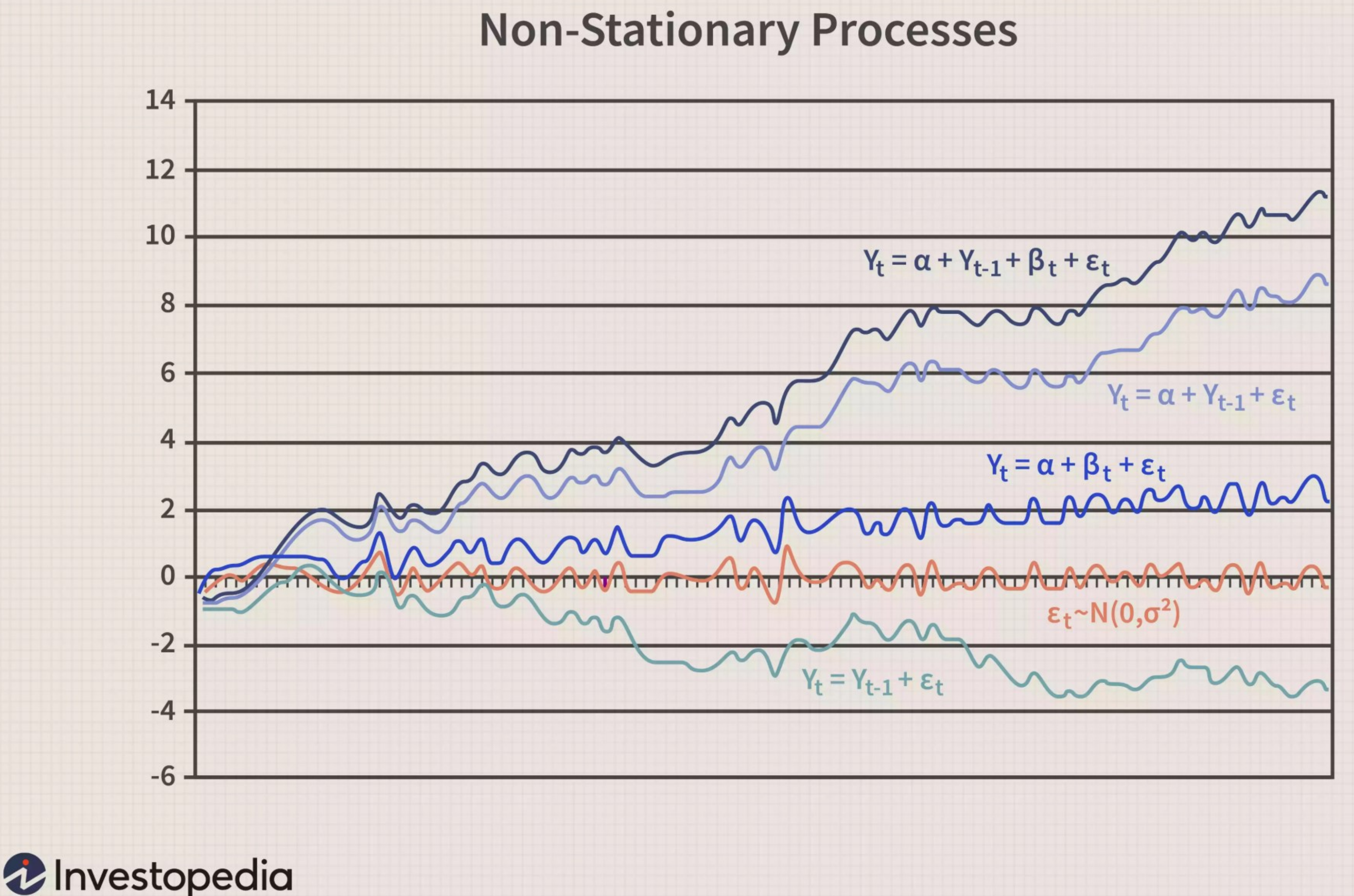
# Trending mean

- A common violation of stationarity
  - *Trend stationary*: The mean trend is deterministic. Once the trend is estimated and removed from the data, the residual series is a stationary stochastic process.
  - *Difference stationary*: The mean trend is stochastic. Differencing the series *D* times yields a stationary stochastic process.

# Cooking Raw Data



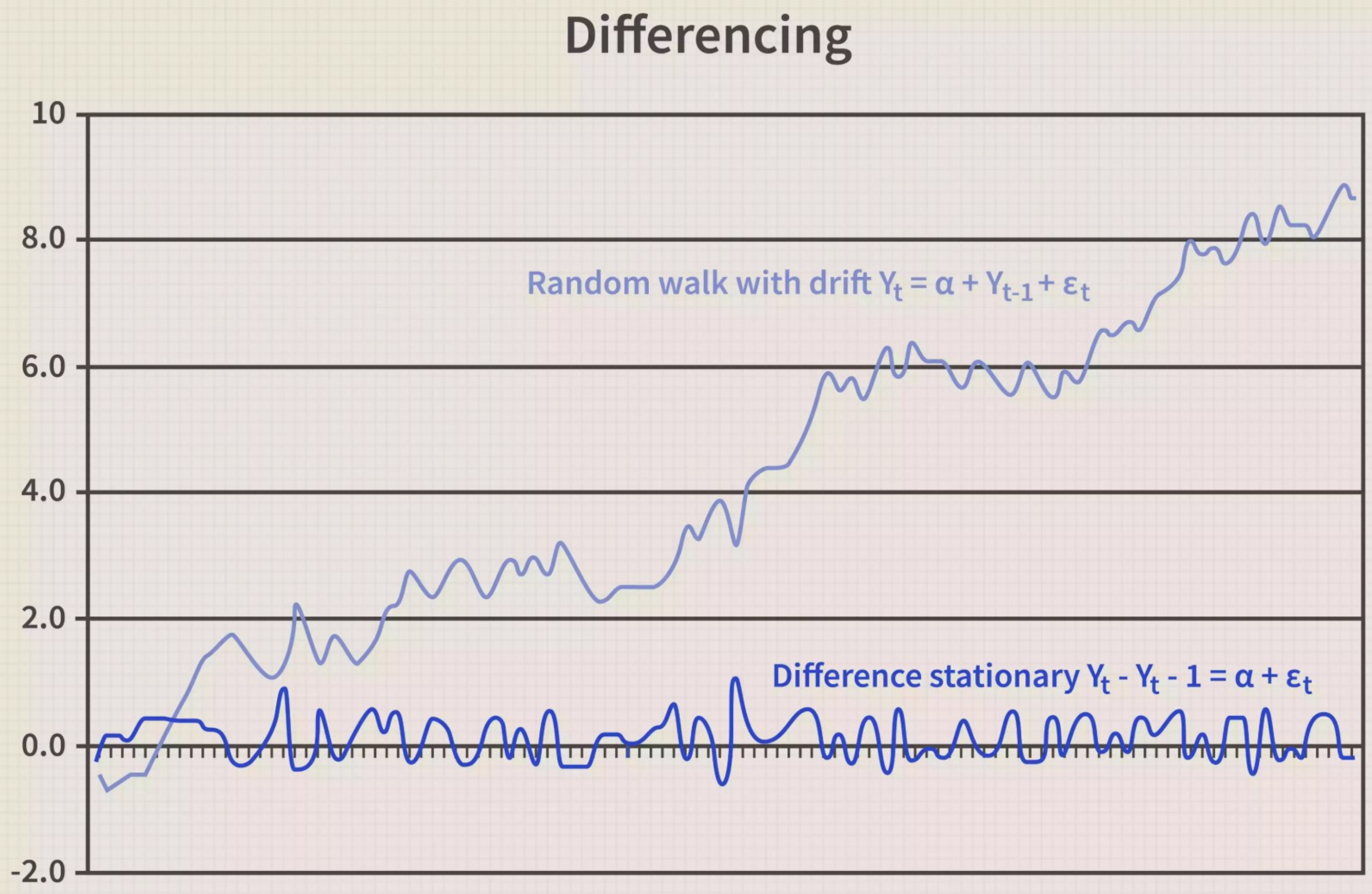
# Types of Non-Stationary Processes



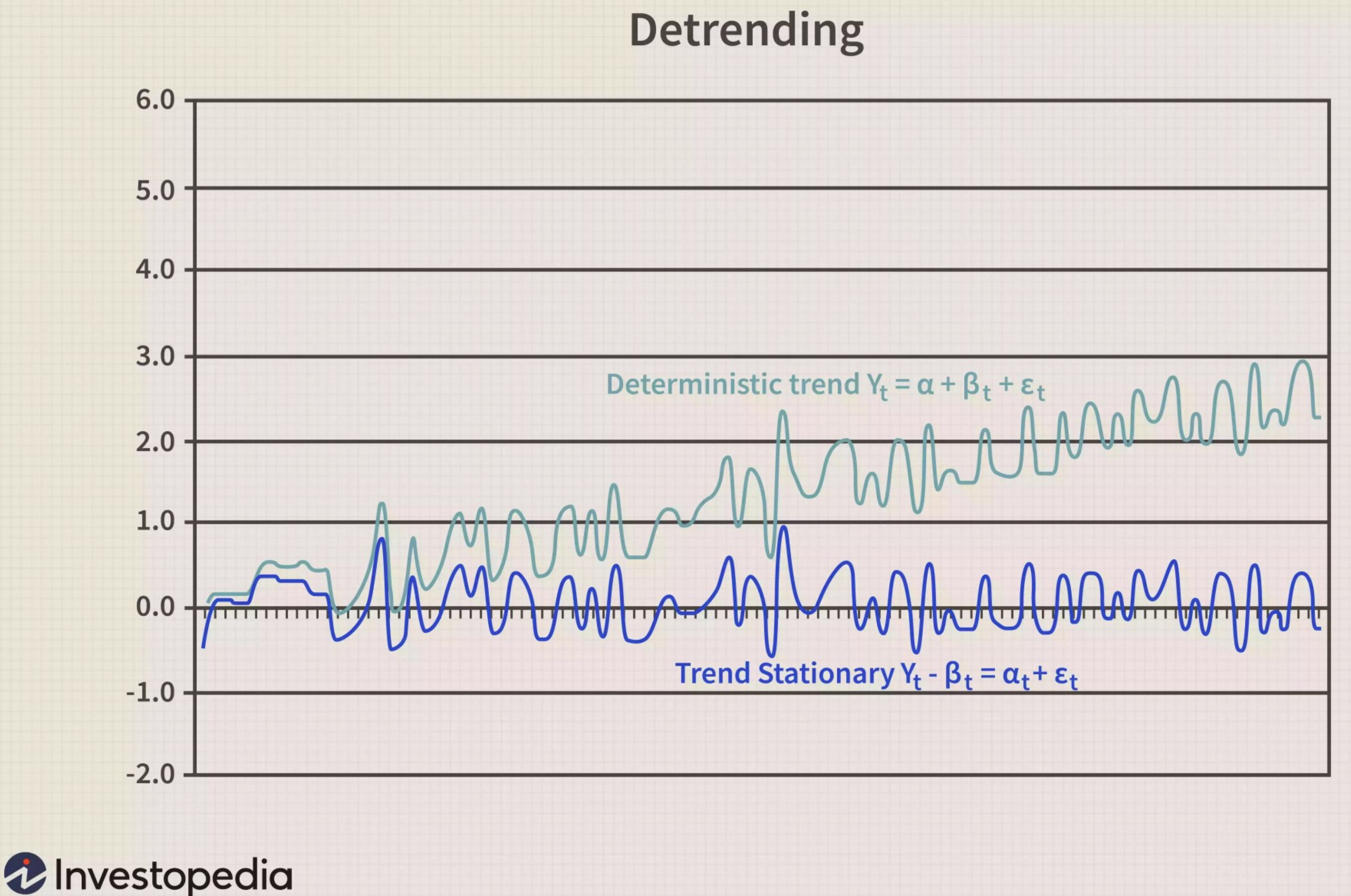
# Types of Non-Stationary Processes

- Pure Random Walk ($Y_t = Y_{t-1} + \varepsilon_t$ )
- Random Walk with Drift ($Y_t = \alpha + Y_{t-1} + \varepsilon_t$ )
- Deterministic Trend ($Y_t = \alpha + \beta t + \varepsilon_t$ )
- Random Walk with Drift and Deterministic Trend ($Y_t = \alpha + Y_{t-1} + \beta t + \varepsilon_t$ )

# Trend and Difference Stationary



# Trend and Difference Stationary (cont'd)



# Conclusion

- Using non-stationary time series data produces unreliable and spurious results and leads to poor understanding and forecasting.
- The solution to the problem is to transform the time series data so that it becomes stationary.
  - If the non-stationary process is a random walk with or without a drift, it is transformed to stationary process by differencing.
  - If the time series data analyzed exhibits a deterministic trend, the spurious results can be avoided by detrending.
  - Sometimes the non-stationary series may combine a stochastic and deterministic trend at the same time and to avoid obtaining misleading results both differencing and detrending should be applied, as differencing will remove the trend in the variance and detrending will remove the deterministic trend.

# Random Walks

- **Random walks**

$$\boxed{y_t = y_{t-1} + e_t}$$

The random walk is called random walk because it wanders from the previous position $y_{t-1}$ by an i.i.d. random amount $e_t$

$$\Rightarrow \; y_t = (y_{t-2} + e_{t-1}) + e_t = \ldots = e_t + e_{t-1} + \ldots + e_1 + y_0$$

The value today is the accumulation of all past shocks plus an initial value. This is the reason why the random walk is highly persistent: The effect of a shock will be contained in the series forever.

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

$$Var(y_t) = \sigma_e^2 t$$

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

The random walk is <u>not covariance stationary</u> because its variance and its covariance depend on time.

It is also <u>not weakly dependent</u> because the correlation between observations vanishes very slowly and this depends on how large t is.

# Random Walks (cont'd)

First, we find the expected value of $y_t$. This is most easily done by using repeated substitution to get

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

Taking the expected value of both sides gives

$$\begin{aligned} E(y_t) &= E(e_t) + E(e_{t-1}) + \dots + E(e_1) + E(y_0) \\ &= E(y_0), \text{ for all } t \geq 1. \end{aligned}$$

Therefore, the expected value of a random walk does *not* depend on $t$. A popular assumption is that $y_0 = 0$—the process begins at zero at time zero—in which case, $E(y_t) = 0$ for all $t$.

# Random Walks (cont'd)

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 $\mathrm{Var}(y_0) = 0$; this does not affect any important conclusions. Then, by the i.i.d. assumption for $\{e_t\}$,

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

In other words, the variance of a random walk increases as a linear function of time. This shows that the process cannot be stationary.

# Random Walks (cont'd)

Even more importantly, a random walk displays highly persistent behavior in the sense that the value of $y$ today is important for determining the value of $y$ in the very distant future. To see this, write for $h$ periods hence,

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

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 $j \geq 1$, we have

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

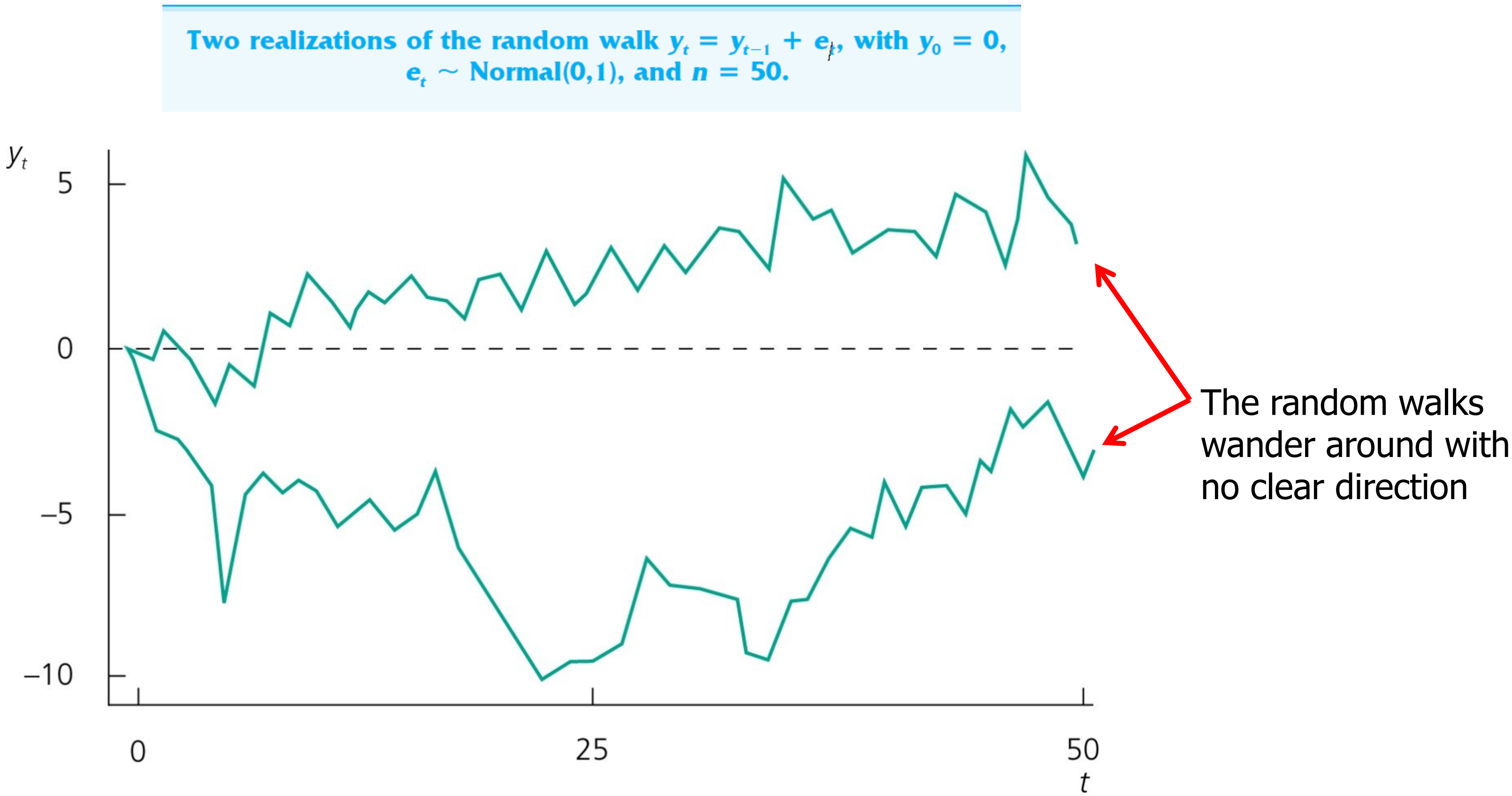
# Random Walks (cont'd)

- Why?

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

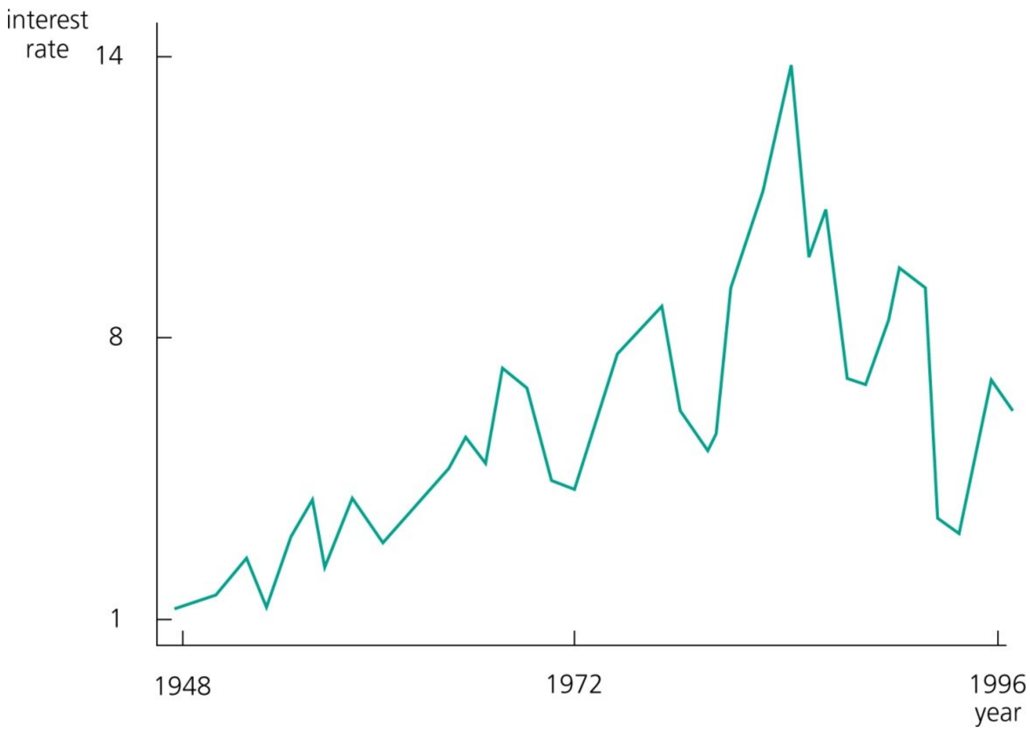
# Examples for random walk realizations

- **Examples for random walk realizations**



# Examples for random walk realizations (cont.)

- **Three-month T-bill rate as a possible example for a random walk**




A random walk is a special case of a unit root process.

Unit root processes are defined as the random walk but $e_t$ may be an arbitrary weakly depen-dent process.

From an economic point of view it is important to know whether a time series is highly persistent. In highly persistent time series, shocks or policy changes have lasting/permanent effects, in weakly dependent processes their effects are transitory.

# Random walks with drift

- **Random walks with drift**

$$y_t = \boxed{\alpha_0} + y_{t-1} + e_t$$

In addition to the usual random walk mechanism, there is a deterministic increase/decrease (= drift) in each period

$$\Rightarrow \quad y_t = \boxed{\alpha_0 t} + e_t + e_{t-1} + \ldots + e_1 + y_0$$

This leads to a linear time trend around which the series follows its random walk behaviour. As there is no clear direction in which the random walk develops, it may also wander away from the trend.

$$E(y_t) = \boxed{\alpha_0 t} + E(y_0)$$

$$Var(y_t) = \sigma_e^2 t$$

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

Otherwise, the random walk with drift has similar properties as the random walk without drift.

Random walks with drift are <u>not covariance stationary</u> and <u>not weakly dependent</u>.

# Random walks with drift (cont'd)

What is new is the parameter $\alpha_0$, which is called the *drift term*. Essentially, to generate $y_t$, the constant $\alpha_0$ is added along with the random noise $e_t$ 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:

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

Therefore, if $y_0 = 0$, $\mathrm{E}(y_t) = \alpha_0 t$: 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 $\mathrm{E}(y_{t+h}|y_t) = \alpha_0 h + y_t$, and 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.

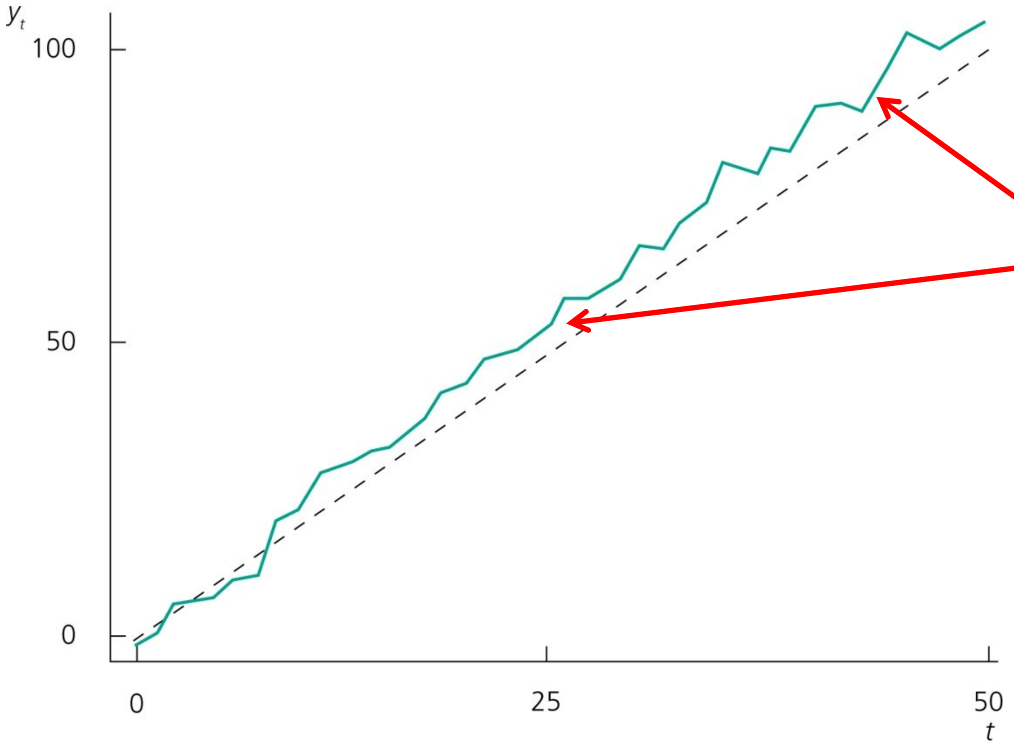
A random walk with drift is another example of a unit root process, because it is the special case $\rho_1 = 1$ in an AR(1) model with an intercept:

$$y_t = \alpha_0 + \rho_1 y_{t-1} + e_t.$$

# Sample path of a random walk with drift

- **Sample path of a random walk with drift**

A realization of the random walk with drift, $y_t = 2 + y_{t-1} + e_t$, with $y_0 = 0$, $e_t \sim$ Normal(0, 9), and $n = 50$. The dashed line is the expected value of $y_t$, $E(y_t) = 2t$.




Note that the series does not regularly return to the trend line.

Random walks with drift may be good models for time series that have an obvious trend but are not weakly dependent.

# Transformations on highly persistent time series

- **<u>Transformations on highly persistent time series</u>**
- **Order of integration**
  - Weakly dependent time series are integrated of order zero (= I(0))
  - If a time series has to be <u>differenced</u> one time in order to obtain a weakly dependent series, it is called integrated of order one (= I(1))
- **Examples for I(1) processes**

$$y_t = y_{t-1} + e_t \;\Rightarrow\; \Delta y_t = y_t - y_{t-1} = e_t$$

$$\log(y_t) = \log(y_{t-1}) + e_t \;\Rightarrow\; \Delta \log(y_t) = e_t$$

After differencing, the resulting series are weakly dependent (because $e_t$ is weakly dependent).

- **Differencing is often a way to achieve weak dependence**

# Transformations on highly persistent time series (cont.)

- **Deciding whether a time series is I(1)**
  - There are statistical tests for testing whether a time series is I(1) (= unit root tests) (not covered in this Chapter)
  - Alternatively, look at the sample first order autocorrelation:

$$\hat{\rho}_1 = \widehat{Corr}(y_t, y_{t-1})$$

Measures how strongly adjacent times series observations are related to each other.

  - If the sample first order autocorrelation is close to one, this suggests that the time series may be highly persistent (= contains a unit root)
  - Alternatively, the series may have a deterministic trend
  - Both unit root and trend may be eliminated by differencing

# Transformations on highly persistent time series (cont.)

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

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

where $v_t$ has a zero mean. Then, $\Delta y_t = \gamma_1 + \Delta v_t$, and so $\mathrm{E}(\Delta y_t) = \gamma_1 + \mathrm{E}(\Delta v_t) = \gamma_1$. In other words, $\mathrm{E}(\Delta y_t)$ is constant. The same argument works for $\Delta\log(y_t)$ when $\log(y_t)$ follows a linear time trend. Therefore, rather than including a time trend in a regression, we can instead difference those variables that show obvious trends.

# Example: Fertility equation

**[Fertility Equation]**

In Example 10.4, we explained the general fertility rate, *gfr*, in terms of the value of the personal exemption, *pe*. The first order autocorrelations for these series are very large: $\hat{\rho}_1 = .977$ for *gfr* and $\hat{\rho}_1 = .964$ for *pe*. These autocorrelations are highly suggestive of unit root behavior, and they raise serious questions about our use of the usual OLS $t$ statistics for this example back in Chapter 10. Remember, the $t$ statistics only have exact $t$ distributions under the full set of classical linear model assumptions. To relax those assumptions in any way and apply asymptotics, we generally need the underlying series to be I(0) processes.

We now estimate the equation using first differences (and drop the dummy variable, for simplicity):

$$\begin{aligned}\Delta\widehat{gfr} &= -.785 - .043\,\Delta pe \\ &\quad\;\;(.502)\quad(.028) \\ n &= 71,\ R^2 = .032,\ \bar{R}^2 = .018.\end{aligned} \tag{11.26}$$

Now, an increase in *pe* is estimated to lower *gfr* contemporaneously, although the estimate is not statistically different from zero at the 5% level. This gives very different results than when we estimated the model in levels, and it casts doubt on our earlier analysis.

# Example: Fertility equation

- **Example: Fertility equation**

$$gfr_t = \alpha_0 + \delta_0 pe_t + \delta_1 pe_{t-1} + \delta_2 pe_{t-2} + u_t$$

This equation could be estimated by OLS if the CLM assumptions hold. These may be questionable, so that one would have to resort to large sample analysis. For large sample analysis, the fertility series and the series of the personal tax exemption have to be stationary and weakly dependent. This is questionable because the two series are highly persistent:

$$\widehat{\rho}_{gfr} = .977, \ \ \widehat{\rho}_{pe} = .964$$

It is therefore better to estimate the equation in first differences. This makes sense because if the equation holds in levels, it also has to hold in first differences:

$$\Delta\widehat{gfr} = \underset{(.468)}{-\ .964} \ \underset{(.027)}{-\ .036} \ \Delta pe \ \underset{(.028)}{-\ .014} \ \Delta pe_{-1} + \underset{(.027)}{.110} \ \Delta pe_{-2}$$

$$n = 69, R^2 = .233, \bar{R}^2 = .197$$

Estimate of $\delta_2$

# Example: Wages and productivity

**Example 11.7**

**[Wages and Productivity]**

The variable *hrwage* is average hourly wage in the U.S. economy, and *outphr* is output per hour. One way to estimate the elasticity of hourly wage with respect to output per hour is to estimate the equation,

$$\log(hrwage_t) = \beta_0 + \beta_1 \log(outphr_t) + \beta_2 t + u_t,$$

where the time trend is included because $\log(hrwage_t)$ and $\log(outphr_t)$ both display clear, upward, linear trends. Using the data in EARNS.RAW for the years 1947 through 1987, we obtain

$$\begin{aligned} \widehat{\log(hrwage_t)} &= \underset{(.37)}{-5.33} + \underset{(.09)}{1.64} \log(outphr_t) - \underset{(.002)}{.018}\, t \\ n &= 41,\ R^2 = .971,\ \bar{R}^2 = .970. \end{aligned} \tag{11.28}$$

# Example: Wages and productivity

- **Example: Wages and productivity**

Include trend because both series display clear trends.

$$\widehat{\log}(hrwage) = \underset{(.37)}{-5.33} + \underset{(.09)}{1.64}\log(outphr) - \underset{(.002)}{.018}\,t$$

$$n = 41, R^2 = .971, \bar{R}^2 = .970$$

The elasticity of hourly wage with respect to output per hour (=productivity) seems implausibly large.

It turns out that even after detrending, both series display sample autocorrelations close to one so that estimating the equation in first differences seems more adequate:

$$\Delta\widehat{\log}(hrwage) = \underset{(.0042)}{-.0036} + \underset{(.173)}{.809}\,\Delta\log(outphr)$$

$$n = 40, R^2 = .364, \bar{R}^2 = .348$$

This estimate of the elasticity of hourly wage with respect to productivity makes much more sense.

# Dynamically complete models

- **<u>Dynamically complete models</u>**
  - A model is said to be dynamically complete if enough lagged variab-les have been included as explanatory variables so that further lags do not help to explain the dependent variable:

$$E(y_t|\mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, y_{t-2}, \ldots) = E(y_t|\mathbf{x}_t)$$

- **Dynamic completeness implies absence of serial correlation**
  - If further lags actually belong in the regression, their omission will cause serial correlation (if the variables are serially correlated)
- **One can easily test for dynamic completeness**
  - If lags cannot be excluded, this suggests there is serial correlation

# Dynamically complete models (cont'd)

Consider the simple static regression model

$$y_t = \beta_0 + \beta_1 z_t + u_t, \qquad \text{11.30}$$

where $y_t$ and $z_t$ are contemporaneously dated. For consistency of OLS, we only need $\mathrm{E}(u_t|z_t) = 0$. Generally, the $\{u_t\}$ will be serially correlated. However, if we *assume* that

$$\mathrm{E}(u_t|z_t, y_{t-1}, z_{t-1}, \ldots) = 0, \qquad \text{11.31}$$

then (as we will show generally later) Assumption TS.5′ holds. In particular, the $\{u_t\}$ are serially uncorrelated. Naturally, assumption (11.31) implies that $z_t$ is contemporaneously exogenous, that is, $\mathrm{E}(u_t|z_t) = 0$.

To gain insight into the meaning of (11.31), we can write (11.30) and (11.31) equivalently as

$$\mathrm{E}(y_t|z_t, y_{t-1}, z_{t-1}, \ldots) = \mathrm{E}(y_t|z_t) = \beta_0 + \beta_1 z_t, \qquad \text{11.32}$$

# Dynamically complete models (cont'd)

Next, consider a finite distributed lag model with two lags:

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 z_{t-1} + \beta_3 z_{t-2} + u_t. \quad \boxed{11.33}$$

Since we are hoping to capture the lagged effects that $z$ has on $y$, we would naturally assume that (11.33) captures the *distributed lag dynamics*:

$$\mathrm{E}(y_t | z_t, z_{t-1}, z_{t-2}, z_{t-3}, \ldots) = \mathrm{E}(y_t | z_t, z_{t-1}, z_{t-2}); \quad \boxed{11.34}$$

that is, at most two lags of $z$ matter. If (11.31) holds, we can make further statements: once we have controlled for $z$ and its two lags, no lags of $y$ or additional lags of $z$ affect current $y$:

$$\mathrm{E}(y_t | z_t, y_{t-1}, z_{t-1}, \ldots) = \mathrm{E}(y_t | z_t, z_{t-1}, z_{t-2}). \quad \boxed{11.35}$$

Equation (11.35) is more likely than (11.32), but it still rules out lagged $y$ affecting current $y$.

# Dynamically complete models (cont'd)

Next, consider a model with one lag of both $y$ and $z$:

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1} + u_t.$$

Since this model includes a lagged dependent variable, (11.31) is a natural assumption, as it implies that

$$\mathrm{E}(y_t | z_t, y_{t-1}, z_{t-1}, y_{t-2}, \ldots) = \mathrm{E}(y_t | z_t, y_{t-1}, z_{t-1});$$

in other words, once $z_t$, $y_{t-1}$, and $z_{t-1}$ have been controlled for, no further lags of either $y$ or $z$ affect current $y$.

# Dynamically complete models (cont'd)

In the general model

$$y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t, \qquad \boxed{11.36}$$

where the explanatory variables $\mathbf{x}_t = (x_{t1}, \ldots, x_{tk})$ may or may not contain lags of $y$ or $z$, (11.31) becomes

$$\mathrm{E}(u_t | \mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, \ldots) = 0. \qquad \boxed{11.37}$$

Written in terms of $y_t$,

$$\mathrm{E}(y_t | \mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, \ldots) = \mathrm{E}(y_t | \mathbf{x}_t). \qquad \boxed{11.38}$$

In other words, whatever is in $\mathbf{x}_t$, enough lags have been included so that further lags of $y$ and the explanatory variables do not matter for explaining $y_t$. When this condition holds, we have a **dynamically complete model**. As we saw earlier, dynamic completeness can be a very strong assumption for static and finite distributed lag models.

# Example

## Example 11.8

### [Fertility Equation]

In equation (11.27), we estimated a distributed lag model for $\Delta gfr$ on $\Delta pe$, allowing for two lags of $\Delta pe$. For this model to be dynamically complete in the sense of (11.38), neither lags of $\Delta gfr$ nor further lags of $\Delta pe$ should appear in the equation. We can easily see that this is false by adding $\Delta gfr_{-1}$: the coefficient estimate is .300, and its $t$ statistic is 2.84. Thus, the model is not dynamically complete in the sense of (11.38).

$$\begin{aligned}\Delta\widehat{gfr} = &-.964 - .036\ \Delta pe - .014\ \Delta pe_{-1} + .110\ \Delta pe_{-2} \\ &(.468)\quad (.027)\qquad\quad (.028)\qquad\quad (.027) \\ n = 69&,\ R^2 = .233,\ \bar{R}^2 = .197.\end{aligned} \tag{11.27}$$

# Sequential exogeneity

- <u>**Sequential exogeneity**</u>
  - A set of explanatory variables is said to be sequentially exogenous if "enough" lagged explanatory variables have been included:

$$E(u_t|\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots) = E(u_t) = 0$$

  - Sequential exogeneity is weaker than strict exogeneity
  - Sequential exogeneity is equivalent to dynamic completeness if the explanatory variables contain a lagged dependent variable
- **Should all regression models be dynamically complete?**
  - Not necessarily: If sequential exogeneity holds, causal effects will be correctly estimated; absence of serial correlation is not crucial

# Remarks

- Sequential exogeneity is implied by strict exogeneity and sequential exogeneity implies contemporaneous exogeneity.
- Because ($\mathbf{x}_t$, $\mathbf{x}_{t-1}$, …) is a subset of ($\mathbf{x}_t$, $\mathbf{y}_{t-1}$, $\mathbf{x}_{t-1}$, …), sequential exogeneity is implied by dynamic completeness.
- If $\mathbf{x}_t$ contains $\mathbf{y}_{t-1}$, the dynamic completeness and sequential exogeneity are the same condition.
- When $\mathbf{x}_t$ does not contain $\mathbf{y}_{t-1}$, sequential exogeneity allows for the possibility that the dynamics are not complete in the sense of capturing the relationship between $\mathbf{y}_t$ and all past values of y and other explanatory variables.

# Example 11.5: Expectations Augmented Phillips Curve

A linear version of the *expectations augmented Phillips curve* can be written as

$$inf_t - inf_t^e = \beta_1(unem_t - \mu_0) + e_t,$$

where $\mu_0$ is the *natural rate of unemployment* and $inf_t^e$ is the *expected* rate of inflation formed in year $t - 1$. This model assumes that the natural rate is constant, something that macroeconomists question. The difference between actual unemployment and the natural rate is called *cyclical unemployment*, while the difference between actual and expected inflation is called *unanticipated inflation*. The error term, $e_t$, is called a *supply shock* by macroeconomists. If there is a tradeoff between unanticipated inflation and cyclical unemployment, then $\beta_1 < 0$. [For a detailed discussion of the expectations augmented Phillips curve, see Mankiw (1994, Section 11.2).]

To complete this model, we need to make an assumption about inflationary expectations. Under *adaptive expectations*, the expected value of current inflation depends on recently observed inflation. A particularly simple formulation is that expected inflation this year is last year's inflation: $inf_t^e = inf_{t-1}$. (See Section 18.1 for an alternative formulation of adaptive expectations.) Under this assumption, we can write

$$inf_t - inf_{t-1} = \beta_0 + \beta_1 unem_t + e_t$$

# Example 11.5: Expectations Augmented Phillips Curve

or

$$\Delta inf_t = \beta_0 + \beta_1 unem_t + e_t,$$

where $\Delta inf_t = inf_t - inf_{t-1}$ and $\beta_0 = -\beta_1\mu_0$. ($\beta_0$ is expected to be positive, since $\beta_1 < 0$ and $\mu_0 > 0$.) Therefore, under adaptive expectations, the expectations augmented Phillips curve relates the *change* in inflation to the level of unemployment and a supply shock, $e_t$. If $e_t$ is uncorrelated with $unem_t$, as is typically assumed, then we can consistently estimate $\beta_0$ and $\beta_1$ by OLS. (We do not have to assume that, say, future unemployment rates are unaffected by the current supply shock.) We assume that TS.1′ through TS.5′ hold. Using the data through 1996 in PHILLIPS.RAW we estimate

$$\begin{aligned} \widehat{\Delta inf_t} &= 3.03 - .543\ unem_t \\ &\quad\ (1.38)\quad (.230) \\ n &= 48,\ R^2 = .108,\ \bar{R}^2 = .088. \end{aligned} \tag{11.19}$$

# Exercise C11.4

Use the data in PHILLIPS.RAW for this exercise, but only through 1996.

(i) In Example 11.5, we assumed that the natural rate of unemployment is constant. An alternative form of the expectations augmented Phillips curve allows the natural rate of unemployment to depend on past levels of unemployment. In the simplest case, the natural rate at time $t$ equals $unem_{t-1}$. If we assume adaptive expectations, we obtain a Phillips curve where inflation and unemployment are in first differences:

$$\Delta inf = \beta_0 + \beta_1 \Delta unem + u.$$

Estimate this model, report the results in the usual form, and discuss the sign, size, and statistical significance of $\hat{\beta}_1$.

(ii) Which model fits the data better, (11.19) or the model from part (i)? Explain.

# Exercise C11.8

Use the data in PHILLIPS.RAW for this exercise.

(i) Estimate an AR(1) model for the unemployment rate. Use this equation to predict the unemployment rate for 2004. Compare this with the actual unemployment rate for 2004. (You can find this information in a recent *Economic Report of the President*.)

(ii) Add a lag of inflation to the AR(1) model from part (i). Is $inf_{t-1}$ statistically significant?

(iii) Use the equation from part (ii) to predict the unemployment rate for 2004. Is the result better or worse than in the model from part (i)?

(iv) Use the method from Section 6.4 to construct a 95% prediction interval for the 2004 unemployment rate. Is the 2004 unemployment rate in the interval?

# Exercise C11.7

Use CONSUMP.RAW for this exercise. One version of the *permanent income hypothesis* (PIH) of consumption is that the *growth* in consumption is unpredictable. [Another version is that the change in consumption itself is unpredictable; see Mankiw (1994, Chapter 15) for discussion of the PIH.] Let $gc_t = \log(c_t) - \log(c_{t-1})$ be the growth in real per capita consumption (of nondurables and services). Then the PIH implies that $\mathrm{E}(gc_t|I_{t-1}) = \mathrm{E}(gc_t)$, where $I_{t-1}$ denotes information known at time $(t-1)$; in this case, $t$ denotes a year.

(i) Test the PIH by estimating $gc_t = \beta_0 + \beta_1 gc_{t-1} + u_t$. Clearly state the null and alternative hypotheses. What do you conclude?

(ii) To the regression in part (i), add $gy_{t-1}$ and $i3_{t-1}$, where $gy_t$ is the growth in real per capita disposable income and $i3_t$ is the interest rate on three-month T-bills; note that each must be lagged in the regression. Are these two additional variables jointly significant?

# Reference

Chapter 11: Further Issues Using OLS with Time Series Data.

In: Wooldridge: Introductory Econometrics: A Modern Approach, 5e