

Further Issues Using OLS with Time Series Data



Chapter 9 (Ch. 11 of the textbook)

Wooldridge: Introductory Econometrics:
A Modern Approach, 5e

Analyzing Time Series: Further Issues Using OLS

- **The assumptions used so far seem to be too restrictive**
 - **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**

Analyzing Time Series: Further Issues Using OLS

- **Stationary stochastic processes**

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

- **Covariance stationary processes**

A stochastic process $\{x_t : t = 1, 2, \dots\}$ 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)** $Cov(x_t, x_{t+h}) = f(h)$.

Analyzing Time Series: Further Issues Using OLS



- **Weakly dependent time series**

A stochastic process $\{x_t : t = 1, 2, \dots\}$ 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 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
- For the LLN and the CLT to hold, the individual observations must not be too strongly related to each other; in particular their relation must become weaker (and this fast enough) the farther they are apart
- Note that a series may be nonstationary but weakly dependent

Analyzing Time Series: Further Issues Using OLS

- **Examples for weakly dependent time series**
- **Moving average process of order one (MA(1))**

$$x_t = e_t + \alpha_1 e_{t-1} \leftarrow \text{The process is a short moving average of an i.i.d. series } e_t$$

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))**

$$y_t = \rho_1 y_{t-1} + e_t \leftarrow \text{The process carries over to a certain extent the value of the previous period (plus random shocks from an i.i.d. series } e_t)$$

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

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.

- **A trending series can be weakly dependent → trend stationary 4**

Analyzing Time Series: Further Issues Using OLS

- **Asymptotic properties of OLS**
- **Assumption TS.1' (Linear in parameters)**
 - Same as assumption TS.1 but now the dependent and independent variables are assumed to be stationary and weakly dependent
- **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 contemporaneously exogenous rather than strictly exogenous, i.e.

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

← The explanatory variables of the same period are uninformative about the mean of the error term

Analyzing Time Series: Further Issues Using OLS

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

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

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

Analyzing Time Series: Further Issues Using OLS

- **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

Contemporaneous exogeneity: $E(u_t | y_{t-1}) = 0$

Strict exogeneity: $E(u_t | y_0, y_1, \dots, y_{n-1}) = 0$ ← Strict exogeneity would imply that the error term is uncorrelated with all y_t , $t=1, \dots, n-1$

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

Analyzing Time Series: Further Issues Using OLS

- **Assumption TS.4' (Homoscedasticity)**

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

← The errors are contemporaneously homoscedastic

- **Assumption TS.5' (No serial correlation)**

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

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

- **Theorem 11.2 (Asymptotic normality of OLS)**

Under assumptions TS.1' – TS.5',

- the OLS estimators are **asymptotically normally distributed**.
- the usual OLS standard errors, t -statistics and F -statistics are **asymptotically valid**.

Analyzing Time Series: Further Issues Using OLS

■ 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

$$E(\text{return}_t | \text{return}_{t-1}, \text{return}_{t-2}, \dots) = E(\text{return}_t)$$

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.

$$\widehat{\text{return}}_t = .180 + .059 \text{return}_{t-1}$$

(.081) (.038)

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

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

Analyzing Time Series: Further Issues Using OLS

- **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'
- **Using highly persistent time series in regression analysis**
 - 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

Analyzing Time Series: Further Issues Using OLS

■ Random walks

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

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

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

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.

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

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

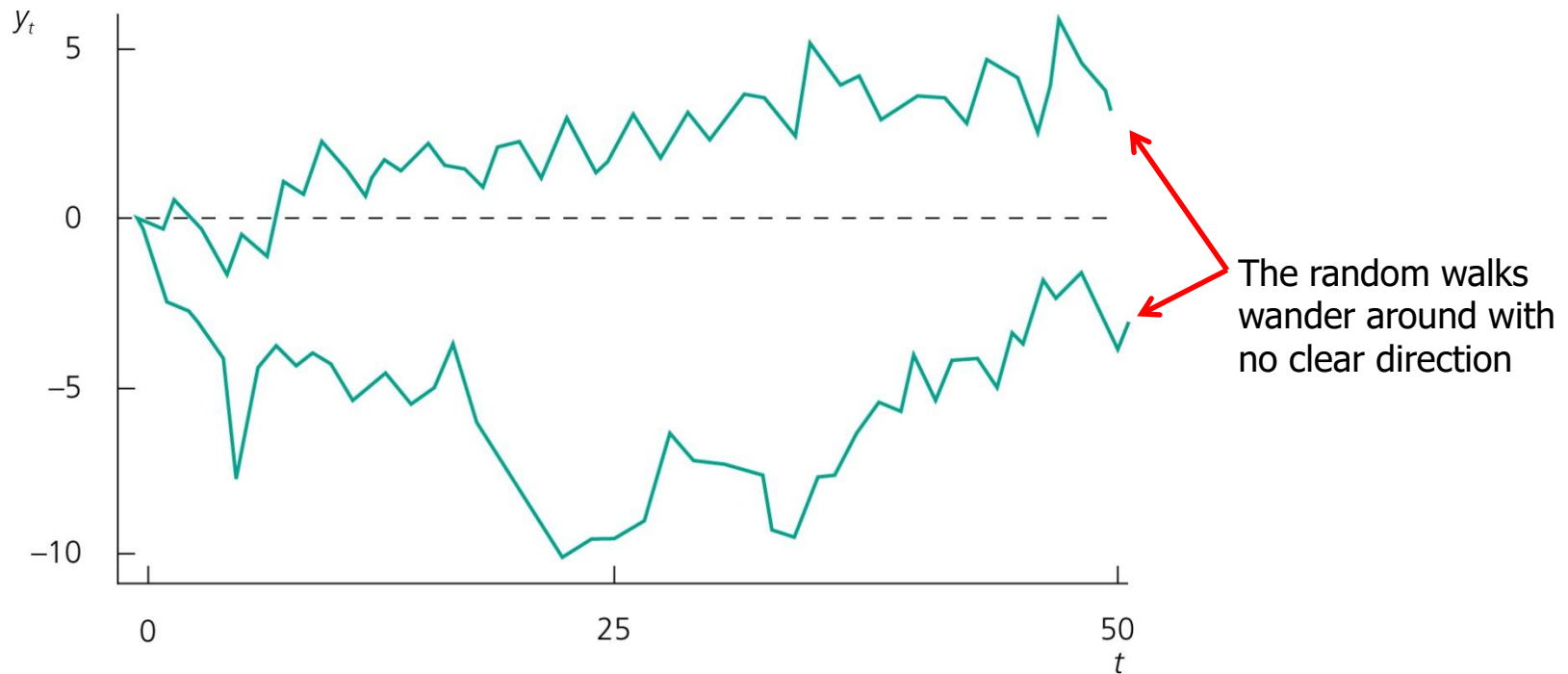
$$Corr(y_t, y_{t+h}) = \sqrt{t/(t+h)}$$

The random walk is not covariance stationary because its variance and its covariance depend on time.

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

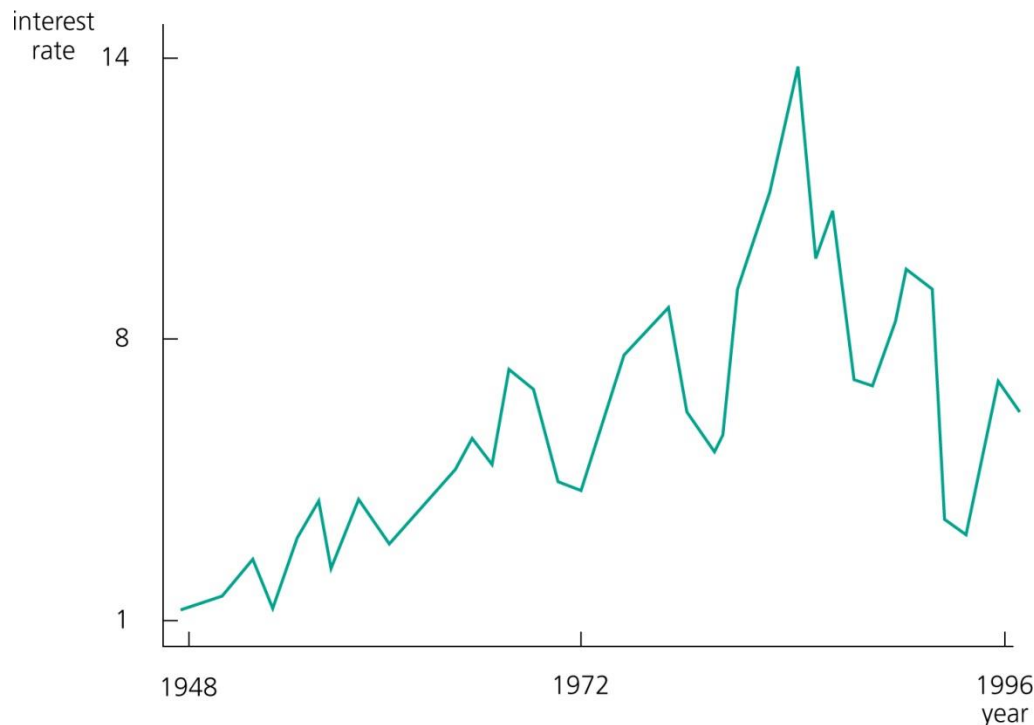
Analyzing Time Series: Further Issues Using OLS

- **Examples for random walk realizations**



Analyzing Time Series: Further Issues Using OLS

■ 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 dependent 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.

Analyzing Time Series: Further Issues Using OLS

■ Random walks with drift

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

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

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

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.

$$E(y_t) = \alpha_0 t + E(y_0)$$

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

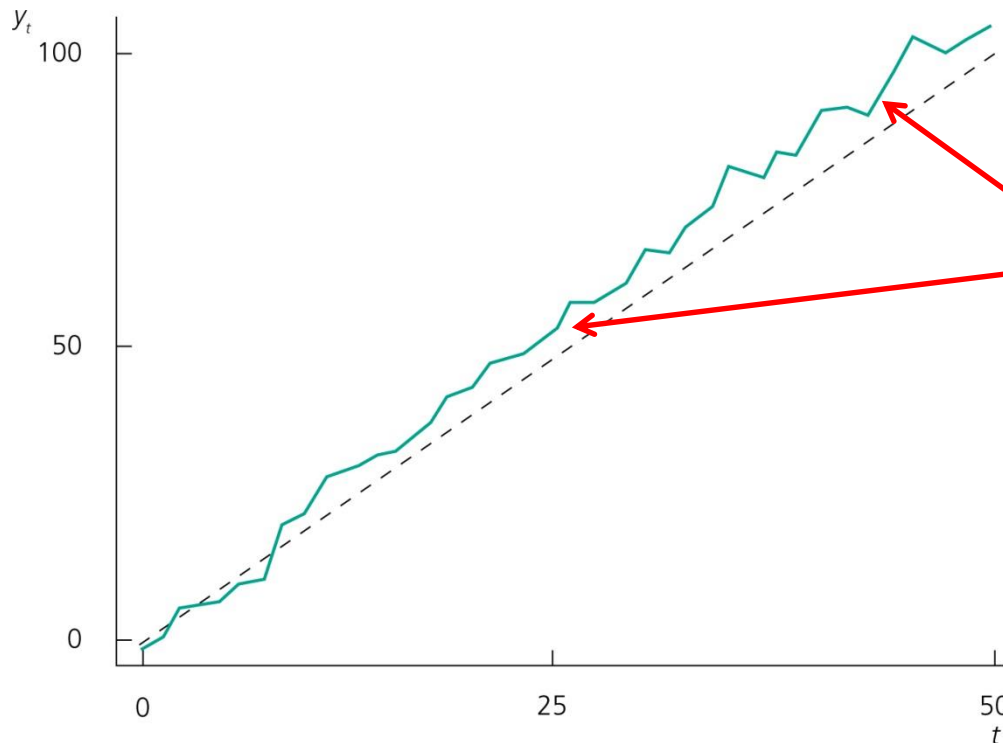
$$Corr(y_t, y_{t+h}) = \sqrt{t/(t+h)}$$

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

Random walks with drift are not covariance stationary and not weakly dependent.

Analyzing Time Series: Further Issues Using OLS

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



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.

Analyzing Time Series: Further Issues Using OLS

- **Transformations on highly persistent time series**
- **Order of integration**
 - Weakly dependent time series are integrated of order zero (= I(0))
 - If a time series has to be differenced 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**

Analyzing Time Series: Further Issues Using OLS

- **Deciding whether a time series is I(1)**

- There are statistical tests for testing whether a time series is I(1) (= unit root tests); these will be covered in later chapters
- Alternatively, look at the sample first order autocorrelation:

$$\hat{\rho}_1 = \widehat{Corr}(y_t, y_{t-1}) \leftarrow \text{Measures how strongly adjacent times series observations are related to each other.}$$

- 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

Analyzing Time Series: Further Issues Using OLS

■ 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:

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

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:

$$\Delta \widehat{gfr} = - .964 - .036 \Delta pe - .014 \Delta pe_{-1} + .110 \Delta pe_{-2}$$

(.468) (.027) (.028) (.027)

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

Estimate of δ_2

Analyzing Time Series: Further Issues Using OLS

■ Example: Wages and productivity

Include trend because both series display clear trends.

$$\widehat{\log}(hrwage) = - 5.33 + 1.64 \log(outphr) - .018 t$$

(.37) (.09) (.002)

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

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:

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

(.0042) (.173)

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

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

Analyzing Time Series: Further Issues Using OLS

- **Dynamically complete models**

- A model is said to be dynamically complete if enough lagged variables have been included as explanatory variables so that further lags do not help to explain the dependent variable:

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



$$E(u_t | \mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, y_{t-2}, \dots) = 0$$

- **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

Analyzing Time Series: Further Issues Using OLS

- **Sequential exogeneity**

- A set of explanatory variables is said to be sequentially exogenous if „enough“ lagged explanatory variables have been included:

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

- Sequential exogeneity is weaker than strict exogeneity
- Sequential exogeneity is equivalent to dynamic completeness if the explanatory variables contain a lagged dependent variable
- Dynamic completeness implies sequential exogeneity