Covers Chapter 10-12, some of 16, some of 18 in Wooldridge

Regression Analysis with Time Series Data

Obviously time series data different from cross section in terms of source of variation in x and y—temporal ordering

2nd difference—NOT randomly sampled in same way as cross sectional—each obs not i.i.d
Why?

Data over time is a "stochastic process"—we have one realization of the process from a set of all possible realizations

A number of common features of Time Series Data: (Put up and fill in what problem is later)

- errors correlated over time—high errors today→ high next time→ problem in estimating standard errors (but not a bias problem)

- effects may take a while to show up→ difficult to know how long should wait to see effects (tax cuts—is growth in Clinton years due to Clinton? Reagan?) (specification problem—incorrect specification leads to bias)

- feedback effects (x→ y but after seeing y, people adjust x. Hmk example: interest rates may affect unemployment, but after looking at unemployement, Fed) (bias problem)

- trending data over time→ data series can look like they are related, but really is "spurious" (bias problem)

I. Finite Sample Properties of OLS under Classical Assumptions

Have time series analogs to all Gauss Markov assumptions
Use this to identify common problems in time-series data

TS1 Linear in Parameters—ok here
TS2 No perfect collinearity—ok here

TS3 Zero conditional mean

Two ways to express this: Strongest condition:

- $E(ut|\mathbf{X})=0$, t=1,2,…n
  - error at time t (ut) is uncorrelated with each explanatory variable in EVERY time period
  - known as STRICT exogeneity
  - need this condition for **unbiasedness**

- $E(ut, xt1, xt2, \ldots . xtk) = E(ut|\mathbf{x}t) = 0$
  - if holds for same period (ut uncorrelated with **x**t) that is *contemporaneous exogeneity*
  - → this assumption is sufficient for **consistency**

Why would STRICT EXOGENEITY assumption fail?

1. As before, omitted vars and measurement error
2. Lagged effects of x
   look at model $y_t = \beta_0 + \beta_1 z_t + u_t$

   ut can't be correlated with zt, or with past or future zt
   → z can't have a lagged effect on y (if does, have specified model incorrectly—use distributed lag)
   BUT As noted above, often DO have effects that emerge over time

3. no feedback of y on future values of z—example of this?
   (book: murder rate and police)
   Again, as noted above, often DO have feedback effects

4. Lagged effects of y—will discuss later

TS4 Homoskedasticity

TS5 NO serial correlation

Corr(ut,us|**X**)=0 for all t≠s

If violated, errors exhibit autocorrelation

Problem Autocorrelation (Chapter 12)

See earlier autocorrelation notes $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$

Bottom line: Leads to biased standard errors although coefficients are unbiased and consistent

Test: Durbin-Watson test

Solution: Apply a generalized difference model

$y^*_t = \beta_0(1-\rho) + \beta_1 x^*_{1t} + \beta_2 x^*_{2t} + \ldots . \beta_k X^*_{kt} + v_t$

Where $y^*t = y_t - \rho y_{t-1}$, same for xs

Estimate ρ using Hidreth-Lu or Cochrane-Orcutt procedures
For Both Idea: start with a guess of ρ and iterate to make better and better guesses

Problem Effects are not Instantaneous (Section 10.2)

Many regression models rely on use of lags

1. Static models--Model a contemporaneous change

   $y_t = \beta_0 + \beta_1 z_t + u_t \quad t=1,2,\ldots n$

   What assumptions does this embody? Change in z has an immediate effect—in same period—on y

   Can you give me an example of this? (Books uses Phillips curve—tradeoff between unemployment and inflation)

2. Finite Distributed Lag Models (FDL)

   $y_t = \alpha + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t \quad t=1,2,\ldots n$

What assumptions does this embody? Change in z has a lagged effect—some effects contemporaneous, some effects in next period, period after that

→ dynamic effect—effect changes over time

- Know number of lags

  Or—estimate successive models and test for significant of additional lags
  Problem with this approach: reduces degrees of freedom as add lags, multicollinearity becomes a problem, issues of "data mining"

Digression on Data Mining

Suppose have a model with c potential X variables. Not sure exactly which ones to include. Could start with a parsimonious model and then test significance level of additional x's.

However, this approach is not legitimate.
Lovell (1983) shows that if have c candidate X's, choose k

$\alpha^* = 1 - (1-\alpha)^{c/k}$ where α^* is the true level of significance and α is the nominal level
$\alpha^* \approx (c/k)\alpha$

so, for example, if c=15, k=5, α=5%, α^*=(15/5)5 = 15%

- Assume δ follow a specific pattern
  e.g., Koyck distributed lags follow $\delta_k = \delta_0 \lambda^k$

- May be more interested in **Long run propensity/long run multiplier (LRP)**—sum of all the deltas
  → effect of a permanent increase of one unit in z

Note that there is often substantial correlation in zs over time—so may have issues with multicollinearity→ can't necessarily estimate separate δs, but can use other methods to get LRP

Homework has you work through this

3. <u>Autoregressive models</u> $y_t = \alpha + \delta_0 z_t + \delta_1 y_{t-1} + u_t$ $t=1,2,\ldots n$ is an

These notes from P&R—see Gujarati for another presentation

Problem: Time Trends and Spurious Correlation

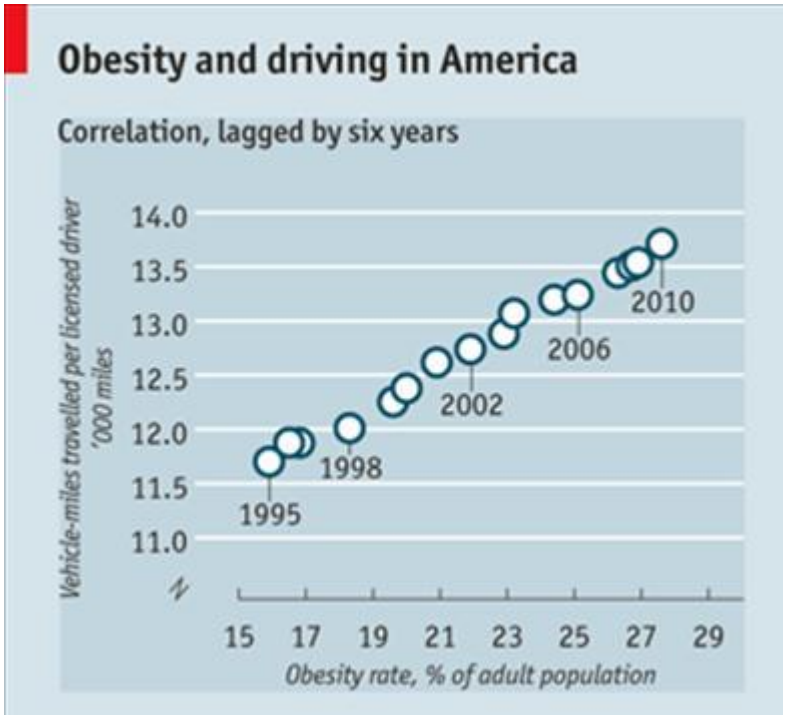
Many economic time series have a common tendency to grow over time

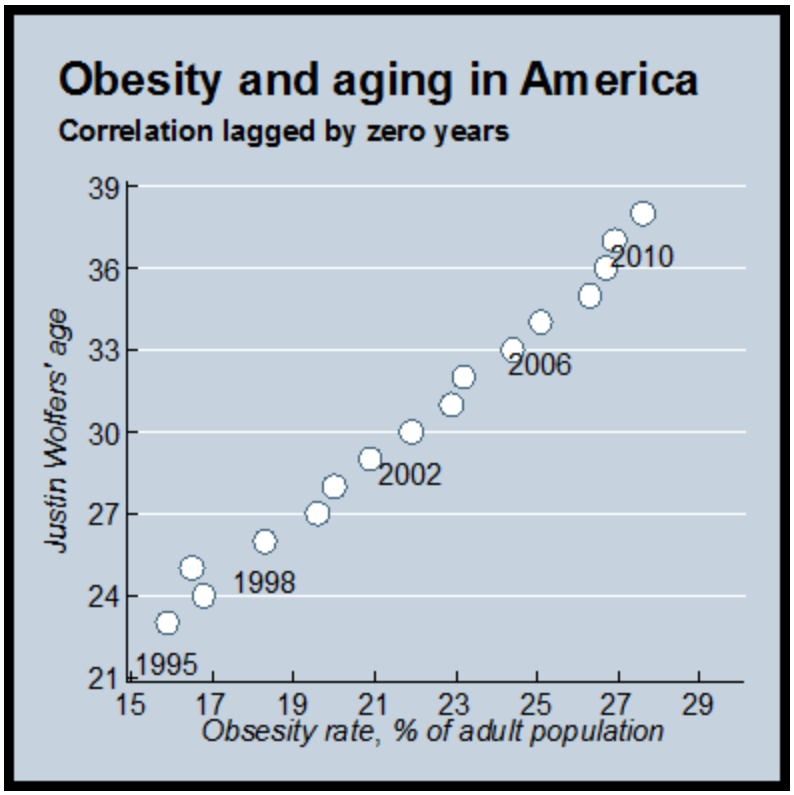
Ex: $y_t = \alpha_0 + \alpha_1 t + e_t$, t= 1,2, ... → linear time trend

$$\Delta y = y_t - y_{t-1} = \alpha_1$$

Problem—could end up with spurious regression if time is driving both x and y→ Biased coefficients

Like an omitted variable bias




Obesity and aging in America
Correlation lagged by zero years
Justin Wolfers' age
39
36
33
30
27
24
21
1995
1998
2002
2006
2010
15
17
19
21
23
25
27
29
Obsesity rate, % of adult population

Solution 1—add a Deterministic time trend

Some relevant Deterministic time trend models:

(Deterministic models make no use of randomness in y, simply describe the time series.)

Example 1:

$yt = a_o + a_1t + a_2t^2 + a_3t^3 + \ldots .+ a_nt^n$
t=0 in base period, increases by 1 in each successive period

n=T-1→ line will pass through every point of the data

Note that could also have a quadratic, cubic—as add more t terms, can capture series more and more exactly, but then none of our explanatory vars matter.

Example 2: Exponential growth curve

$yt = f(t) = Ae^{rt}$

→ $\log y_t = c_1 + c_2t$ (c_1=logA, c_2=r)

c2 is growth rate

Caution—too many t terms will explain the whole series
Caution—interpreting R^2 is more problematic—high R^2 may be because time trend captures most of variation in y, not because xs do.

Solution 2: Detrend the data

$y_t = \alpha_0 + \alpha_1 t + e_t \rightarrow \hat{e}_t = \ddot{y}_t$ which is the "detrended" y

Do the same thing for the xs.
Run $\ddot{y}_t \ldots on \ldots \ddot{x}_{t1}, \ddot{x}_{t2}$

Makes interpretation of R^2 easier.

Both approaches:
Nice thing about including t is that useful even if y_t is not trending and one of the x's is. May be that deviations in x from its trend matter for changes in y.

Chapter 11

Problem: Stationarity and Weakly Dependent Time Series—Is y growing? Or Does it follow a Random Walk?

Suppose y grows over time:
Consider the model

$$y_t = \alpha + \beta t + \rho y_{t-1} + \varepsilon_t$$

Is y growing because there is a trend? $\beta>0$ or because follows a random walk with positive drift ($\alpha>0$, $\beta=0$, $\rho>0$)?

Has important implications for modeling.
Recall: one of GM assumptions was random sampling.

Need random sampling to get consistent estimates.

With time series data, don't have random sampling. When can we use the realization we have to make inferences? What is the time-series counterpart to random sampling assumption?

1. Definition of stationarity
2. Why series need to be stationary for inference/prediction
3. How to test if data series is stationary
4. What to do if it is not
   If stationary, just include t.
   If not, may need to difference the data to get consistent estimates.

Stationary and Nonstationary Time Series

Does the stochastic process vary with time? (If Yes—non-stationary)

Stochastic process: collection of random variables ordered in time.

Example: GDP and temperature are random variables. Temperature today, a random variable. Temp yesterday was a RV. 65 degrees is a particular realization.

Want a model where

- Stochastic properties are time invariant
- Can model process with a single equation with fixed coefficients—equation is invariant with respect to time

Note that many time series do not satisfy these characteristics. BUT—can often transform data into a series that DOES satisfy these properties

Need to work with stationary models to make inference/predictions.

Definition of a Stationary process:

- $y_1, y_2, \ldots . y_t$ one realization (one outcome) drawn from joint probability distribution function: $P(y_1, y_2, \ldots . y_t)$

- Future observation y_{t+1} drawn from conditional probability distribution:
  $P(y_{t+1}|y_1, \ldots , y_t)$

- Stationary process: joint and conditional distributions are invariant with respect to time

  More formally $P(y_t, \ldots . y_{t+k}) = P(y_{t+m}, \ldots . y_{t+k+m})$

  $P(y_t) = P(y_{t+m})$

  - One example: White noise process. Stochastic process is purely random if has zero mean , constant variance, and is serially uncorrelated. Usual assumptions for error term in classic model.

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

- Since $P(y_t)$ is same for all t, **observations $y_1, y_2, \ldots . y_t$ can be used for inference**

  - Shape approximated by histogram
  - Sample mean ($\bar{y}$) an estimator of population mean μ_y
  - Sample variance an estimator for σ^2_y

Need stationarity for unbiased estimates. Wooldridge 11.1 discusses a weaker condition for OLS Estimates to be consistent: **Weak dependency**

Definition is loose. **Stationarity** deals with joint distribution being same over time. **Weak dependency** deals with how strongly related x_t and x_{t+k} are as distance (h) gets large. As k increases, if x_t and x_{t+k} are "almost indepdendent," then is weakly stationary. Wooldridge states is no real formal definition because can't cover all cases.

Weak dependency/stationarity: This assumption replaces assumption of random sampling→ allows LLN and CLT to hold to get consistent OLS estimates.

Show graphs of stationary and non-stationary series.

16.2.1 Homogenous Non-stationary Processes
Let's see what problems have if series is not stationary:

1. **Definition of Random walk—A non-stationary series**

Example: in efficient capital mkt hypothesis, stock prices are a random walk and there is no scope for speculation

$y_t = y_{t-1} + \varepsilon_t$ $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_s) = 0$ for $t \neq s$
[Random walk with drift: $y_t = \alpha + y_{t-1} + \varepsilon_t$]

Example: coin flips—tails = -1, heads = +1

Random walk (with or without drift) known as a unit root process

$y_t = \rho y_{t-1} + \varepsilon_t$ where ρ lies between -1 and 1 AR(1) model
If $\rho = 1$, this is a Random walk

(Nonstationarity, random walk, unit root, stochastic trend all interrelated concepts)

2. **Forecasting with Random Walk:**
Show figure 16.1—block out forecast

What would forecast be one period ahead?

$$\hat{y}_{t+1} = E(y_{t+1}|y_1, \ldots, y_t) = y_t + E(\varepsilon_{t+1}) = y_t$$

What about 2 periods ahead?

$$\hat{y}_{t+2} = E(y_{t+2}|y_1, \ldots, y_t) = y_{t+1} + E(\varepsilon_{t+2}) = E(y_t + \varepsilon_{t+1} + \varepsilon_{t+2}) = y_t$$

→So no matter how far in the future look, best forecast of y_{t+k} is y_t
Idea is that with a stationary series, best guess of y_{t+1} is
$y=0$ or $y=\beta_0+\beta_1 x$ or $y = \beta_0+\beta_1 x +\beta_2 t$

With a non-stationary series, best guess is y_t

Contrast with AR(1): $y_t = \rho_1 y_{t-1} + \varepsilon_t$ *where* $|\rho_1| < 1$
→ *Further and further into future, best forecast is* $y_t = 0$

What about variance of forecast error?

one period: $E(\varepsilon^2_{t+1}) = \sigma^2_\varepsilon$

two periods: $E[(\varepsilon_{t+1} + \varepsilon_{t+2})^2] = E(\varepsilon^2_{t+1}) + E(\varepsilon^2_{t+2}) + 2E(\varepsilon_{t+1} + \varepsilon_{t+2})$
$= \sigma^2_\varepsilon$

Show figure 16.1—with forecast

What is variance of the series?

$$E[(y_t)^2] = E[(y_{t-1} + \varepsilon_t)^2] = E(y_{t-1})^2 + \sigma^2_\varepsilon$$
$$= E(y_{t-2})^2 + 2\sigma^2_\varepsilon$$
$$= E(y_{t-n})^2 + n\sigma^2_\varepsilon$$

Note that according to this, the variance is infinite and therefore undefined! (Recall that this violates on of the G-M

3. Phenomenon of Spurious Regression

If have two time series that are random walks

$y_t = y_{t-1} + u_t$
$x_t = x_{t-1} + v_t$ and regress one on the other, usually get highly significant results.

--when R2 is low and DW d value suggests strong autocorrelation, likely have a spurious relationship

--Rule of thumb from Granger and Newbold: R2>d → suspect spurious regression

Summary of Problems Caused by Non-stationarity:

1. OLS estimates will not be consistent
2. Nonstationarity can lead to autocorrelation—transforming to get a stationary series can (sometimes) correct the problem of autocorrelation
3. Non-stationary series regressed on each other can lead to spurious correlation
4. Series does not revert to some mean—temporary shocks lead to permanent effects
5. Time series data are specific to particular period—can't be used to generalize to other time periods. Forecasting is futile if series exhibits a random walk.
6. Causality tests (like Granger causality) assume stationarit**y**

→ Stationarity means can use sample to make inferences about probability distribution of the ys!!! This is the counterpart to random sampling in a time series context. If not stationary, can't

So our goal is to (1) identify when series are NOT stationary (2) transform non-stationary series into stationary series

Solutions for Non-Stationarity

Very few time series are actually stationary
However, if difference (differentiate) 1 or more times, most become stationary

→ If true, Homogenous

→ Number of times need to difference (differentiate) is "order of homogeneity" (Or **integrated of degree ***)**

Here $\Delta y_t = y_t - y_{t-1} = \varepsilon_t$ (First difference)

Since ε_t are independent over time, this is stationary

→ Random walk is 1st order homogenous or Integrated of order 1--I(1)

$\rho_0 = 1,\ \rho_k = 0$ for all k

→ Often use 1st difference of data in OLS regressions

So how can we tell the difference between stationary and non-stationary series? Or weakly and strongly dependent series?

16.2 Characterizing Time Series: The Autocorrelation Function

1. Autocorrelation Function and Correlogram

   One of the key properties we would like to model is the correlation over time. Turns out to be key to identifying stationary, weakly dependent series.

Autocorrelation function will provide a description of correlation over time:

$$\rho_k = \frac{E[(y_t - \mu_Y)(y_{t+k} - \mu_Y)]}{\sqrt{E[(y_t - \mu_Y)^2]E[(y_{t+k} - \mu_Y)^2]}} = \frac{Cov(y_t, y_{t+k})}{\sigma_{y_t}\sigma_{y_{t+k}}}$$

If process is stationary,

$$\rho_k = \frac{E[(y_t - \mu_Y)(y_{t+k} - \mu_Y)]}{\sigma_y^2} = \frac{\gamma_k}{\gamma_0}$$

In practice, estimate this using sample averages for means

Note that this is symmetric → $\rho_k = \rho_{-k}$

Note that if $\rho_k = 0$ for all k>0 → In that case, $y_t = \varepsilon_t$ → "white noise"
→ No value in using a time series model to forecast the series

Wooldridge: Weak dependency means this function goes to zero "sufficiently quickly" as k→∞

Show Figures 16.3 and 16.4

Note that 16.3 appears to be stationary. Autocorrelation function in 16.4 falls off quickly—can use the function to test for stationarity

Show Figures 16.6—autocorrelation function of a non-stationary series

Difference until autocorrelation function exhibits this "dropping off" pattern

Wooldridge: Weak dependency means this function goes to zero "sufficiently quickly" as k→∞

Can use Q statistic or Ljung-Box (LB) statistic to examine joint significance that ρ up to certain lag are equal to zero

$Q = n\sum_{k=1}^{m} \rho_k 2$ (should have hat) m is lag length~chi squared m

Examples of stationary, weakly dependent series:

- MA(1) $x_t = e_t + \alpha_1 e_{t-1}$ t=1,2,. . . .T

Note that adjacent terms are correlated.
However, if are 2 or more time periods apart, are independent

- Stable AR(1) $y_t = \rho_1 y_{t-1} + \varepsilon_t$ where $|\rho_1| < 1$

Wooldridge shows that this process has finite variance, correlation gets smaller and smaller as lags get large

16.3 Testing for random walks:

Is the series a random walk? (Difference once)

Note that

(1) If it is, G-M theorem won't hold
will get spurious results when regress one against the other

(2) First differencing will lead to stationary series

(3) Has important policy implications—means that series does not revert to some mean, but that temporary shocks lead to permanent effects

How determine number of times need to differentiate?
How do we decide when a series is stationary?

2. Dickey-Fuller unit-root test—test for random walk

In Wooldridge, this is in chapter 18

Consider the model

$$y_t = \alpha + \beta t + \rho y_{t-1} + \varepsilon_t$$

Is y growing because there is a deterministic trend? (Model it with t) $\beta>0$ OR does is follow a random walk with drift (difference it)?

Can we just estimate that model and test if $\rho=1$? No—turns out sampling distribution is very different when is close to 1 than if is far from 1

But—do first difference

$Yt - yt\text{-}1 = \rho Yt\text{-}1 - Yt\text{-}1 + ut$
$= (\rho\text{-}1)Yt\text{-}1 + ut$
So basic idea is to take difference, regress on lagged y

Dickey Fuller test of $\beta = 0$, $\rho = 1$:

- Run unrestricted regression: $y_t - y_{t-1} = \alpha + \beta t + (\rho-1) y_{t-1}$
- Run restricted regression: $y_t - y_{t-1} = \alpha$
- Like F test: $(N-k)(ESS_R - ESS_{UR})/q(ESS_{UR})$
- But not distributed as an F stat—have to use Dickey-Fuller table for critical values

A number of variations—Yt is random walk, random walk with drift, random walk with drift around deterministic time trend—all different test

Augmented Dickey-Fuller test—allows for autocorrelation in ε_t

- Run unrestricted regression:

$$y_t - y_{t-1} = \alpha + \beta t + (1-\rho) y_{t-1} + \sum_{j=1}^{p} \lambda_j \Delta y_{t-j}$$

  where number of j lags is selected by econometrician
- Run restricted regression

$$y_t - y_{t-1} = \alpha + \sum_{j=1}^{p} \lambda_j \Delta y_{t-j}$$

- Calculate test statistic as above, use D-F tables

A NUMBER of other unit root tests—see Gujarati for details

Why? Due to size and power of tests. SIZE of test mean level of significance—probability of committing a type I error. What is that? Reject true hypothesis. POWER of test is 1- prob of committing a type II error—reject null, when null is false. EX: power = .8 means 20% chance of committing a type II error.

Most unit root tests have low power—allow us to reject hypothesis that variable is not a random walk, but also means accept null of a unit root more frequently than is warranted (see gujarait 759 for more info on this)

16.4 Co-integrated Time Series

As have said, regressing one random walk on another can lead to spurious correlation→ Difference the time series in question before running the regression

Note however, that this loses information—suppose trend up in x really is related to trend up in y—now only have information from deviations in trend to identify the relationship

Can we ever regress y on x without differencing if they are random walks?

→Is linear combination of the two variables stationary?
If so, can run OLS and will get consistent estimates

$z_t = x_t - \lambda y_t$

Examples:

- aggregate consumption and disposable income
  - Both are random walks
  - expect that should move together over time
  - linear combination should be stationary
- Stock prices and dividends
- Exchange rates and interest differentials

(1) Test whether x and y are random walks—Dickey-Fuller tests
(2) Run a co-integrating regression

$x_t = \alpha + \beta y_t + \varepsilon_t$

(3) Test whether the residuals from this regression are stationary
   a. Can run Dickey-Fuller test on residuals
   b. Or can look at Durbin-Watson statistic from co-integrating regression

$$DW = \frac{\sum_{t=2}^{T} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{T} (\hat{\varepsilon}_t)^2}$$ if ε is a random walk, numerator should be zero—test DW=0