Time Series Analysis

Why Time Series? Let the Unlearning Begin

Michael Colaresi

Thinking is required

- Split into pairs
- write down the assumed variance-covariance matrix of the errors for regression (OLS)
- If you don't know it
    - think about the errors for OLS
    - what are the mean?
    - what is the variance of the errors?
    - how do the errors covary?

Regression Assumption

- look at $Y = X\beta + \epsilon$
- $\epsilon \sim N_n(0, \Sigma_{\epsilon,\epsilon})$
- $\Sigma_{\epsilon,\epsilon}$ has zeros as off-diagnal elements.

Mommy, where do errors come from?

$$\begin{bmatrix} \sigma^2 & 0 & 0 & \cdots \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ \vdots & 0 & 0 & \ddots \end{bmatrix} = \Sigma_{\epsilon,\epsilon} = \sigma^2 I$$

Truthiness

- $corr(\epsilon_i, \epsilon_j) = 0, \forall i \neq j$
- True or not? – Who cares.
- Useful or not?

A Hinderence to Inference

- "Wrong" Idea from data (mean $\neq$ equilibrium)
- Very certain about incorrect inference: Std. Errors too small
- autocorrelation interacts with endogeneity to make very messy inferences

What Causes Dynamics

- $corr(\epsilon_i, \epsilon_j) \neq 0, \forall i \neq j$
  - Memory (just can't let go)
  - Rules/institutions (formal, informal)
  - Expectations/Conventions
  - Persistent values/preferences
  - Policy/Political Stickiness (can not adapt fast enough)

Slicing and Dicing

- Pair up with someone different this time (3 is fine)
- can only be maximum of 5 groups
- each group gets two dice
- First roll one, right down the value, then roll that one again. Do this 30 times. Then plot the values on the board.

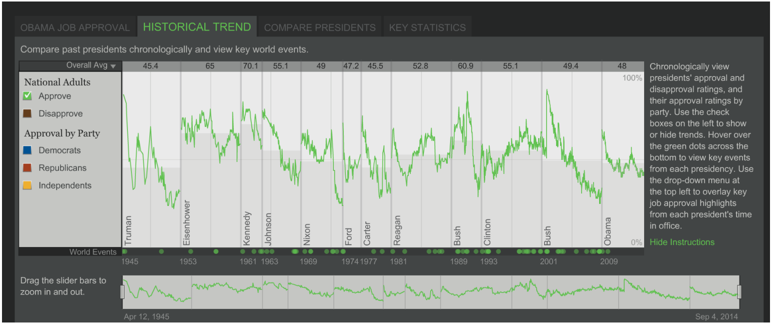
Dice-namic

- Next, roll both dice togther, write down the value.
- Then roll one of them, and add the two values you see together.
- Then roll the other (just 1) and add the values you see.
- keep switching back and forth, writing the sums, until you have 30 values.
- plot on the board next to your first graph

Examples

- Presidential Approval
- International Events data
- Growth, other economic data

Example


OBAMA JOB APPROVAL
HISTORICAL TREND
COMPARE PRESIDENTS
KEY STATISTICS
Compare past presidents chronologically and view key world events.
Overall Avg
45.4
65
70.1
55.1
49
47.2
45.5
52.8
60.9
55.1
49.4
48
100%
0%
National Adults
Approve
Disapprove
Approval by Party
Democrats
Republicans
Independents
Truman
Eisenhower
Kennedy
Johnson
Nixon
Ford
Carter
Reagan
Bush
Clinton
Bush
Obama
World Events
1945
1953
1961
1963
1969
1974
1977
1981
1989
1993
2001
2009
Chronologically view presidents' approval and disapproval ratings, and their approval ratings by party. Use the check boxes on the left to show or hide trends. Hover over the green dots across the bottom to view key events from each presidency. Use the drop-down menu at the top left to overlay key job approval highlights from each president's time in office.
Hide Instructions
Drag the slider bars to zoom in and out.
Apr 12, 1945
Sep 4, 2014

Traits of "Normal" Times Series Data

- 50 observations over time (Information<N)
- On same unit
- Continuous (or nearly so) dependent variable(s)

$\Rightarrow$ Always have to think about the DGP.

Eventually, you can relax/extend these traits, but these are the basics

Difficult Cases/Non-“Normal” Time Series Data

- low counts (eg number of coups)
- Democracy scales (like Polity)
- Repression scales (like Poe and Tate)
- Anything with an underlying continuous latent variable and a limited observable measurement.

$\Rightarrow$ All lessons in this class apply to these difficult cases too, just that convention says to ignore the problems.

Costs and Benefits

- *Benefits*
  - Useful set of tools to make inference
  - Solve some Causality/Chicken and the Egg Problems
  - Control for expectations/look at surprises
  - Fits numerous types of DGPs
- *Costs*
  - Lose some certainty
  - Observational equivalence
  - Easier to prove you are wrong
  - Ask more of data (sometimes it won't deliver)

How OLS works

- Fitting a line that minimizes the sum of squares $(\Sigma_1^n(y_i - \hat{y}_i)^2)$
- $\hat{\beta} = (X'X)^{-1}X'Y$
- Through assumptions, defines uncertainty around that line.
- $\sqrt{(diag((X'X)^{-1}) \times \sigma_\epsilon^2)}$

OLS

Non-intuitive implications:

- Can scatter rows of data and inference does not change.
  - Sequence plays no role.
- “Weirdness” disappears instantly.
  - A large change (eg 9/11) only effect one observation.

Enter our friend subscript t.

- $y_{it} = \alpha_i + \phi_{i1} y_{it-1} + \beta_{i0} x_{it} + \beta_{i1} x_{it-1} + \epsilon_{it} + \psi_{i1} \epsilon_{it-1}$
  - Separate models of economies/countries/states by countries over time.
  - this is know as an ARMAX(1,1,1) model.
- Lots and lots of simplifications possible:
  - Structural model: $y_i = \alpha + \beta x_i + \epsilon_{it}$
  - Fixed Effects (Intercepts): $y = \alpha_i + \beta x_i + \epsilon_{it}$
  - Random Effects in intercepts and coefficients, etc.
- Lesson: Almost any model you see in the literature that has cross-national time series data is a special case of a time series model.

What our theories don't tell us, but the data can.

- Dynamic structure: $y_t \sim f(y_{t-i})$
- Lag structure: $y_t \sim f(x_{t-i})$
- Error structure: $\epsilon_t \sim f(\epsilon_{t-i} | y_{t-i}, \forall i > 0)$

Sampling

- We will no longer think of sampling individual observations.
- Cross-sectional: $y_i \sim N(\mu, \sigma^2)$
- We draw a block of a time series: $y_t \sim f(y_{t-i}, \epsilon_t)$
  - This is a realization that we have less information on the underlying process than we would given complete independence. How much less $\approx ((1-\rho)/(1+\rho))n$ (Quenouille 1952)
  - Might think of this as a window rather than a point.

Notation

1. Lag Operators("L")/Backshift Operator("B"):

$$BY_t = Y_{t-1}$$
$$B^2 Y_t = B(BY_t) = BY_{t-1} = Y_{t-2}$$
$$B^3 Y_t = B(B^2 Y_t) = BY_{t-2} = Y_{t-3}$$
$$B^j Y_t = Y_{t-j}$$

2. Differencing (Δ)

$$\Delta = 1 - B$$
$$\Delta Y_t = (1 - B) Y_t = Y_t - Y_{t-1}$$
$$\Delta Y_{t-1} = (1 - B) Y_{t-1} = Y_{t-1} - Y_{t-2}$$
$$\Delta_2 Y_t = Y_t - Y_{t-2}$$
$$\Delta_{1,4} Y_t = Y_{t-1} - Y_{t-4}$$

Types of Patterns

If we simplify things and look at one series, we can see different patterns:

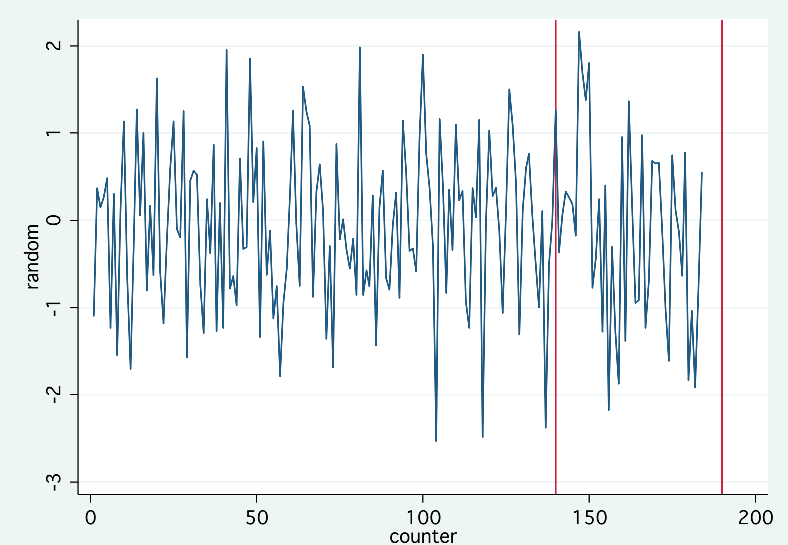
- Seasonal/Cyclical Variation (Business Cycles/Election Cycles) $= y_t \sim f(y_{t-j})$ where $j \gg 2$
- Deterministic Trend (My son's height) $= y_t \sim f(t)$
- Drift (Literacy) $= \Delta y_t \sim f(c)$
- Irregual/Stochastic fluctuations (everywhere) $= y_t \sim f(y_{t-i}, \epsilon_{t-i})$

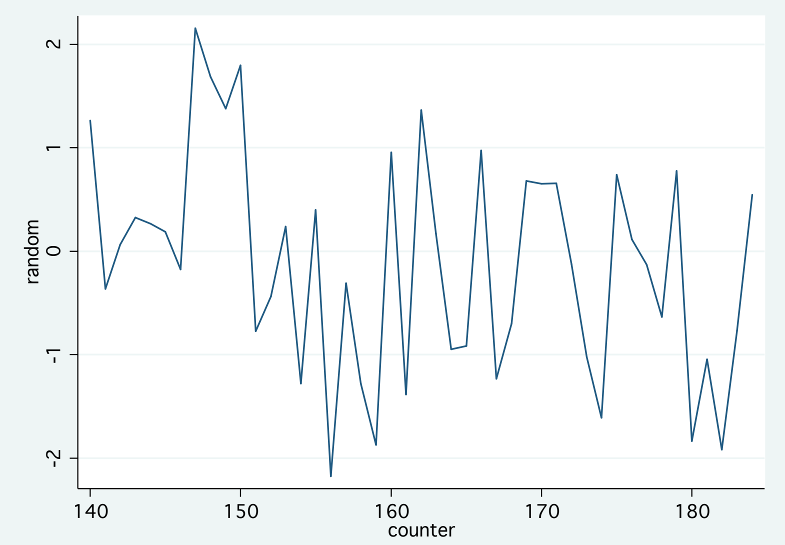
Absence of a Pattern

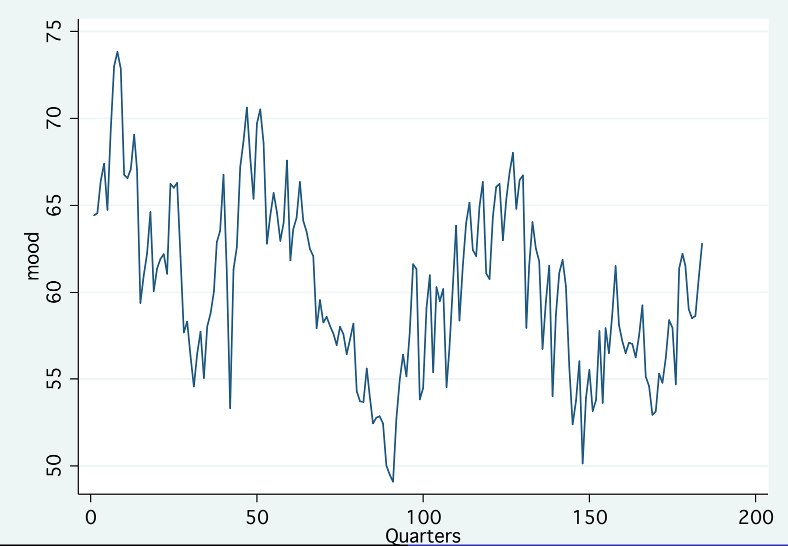
Definition

White Noise: If a series or some transformation of a series has no sequential pattern, we call that series or that transformed series WHITE NOISE.

- that is the assumed error process of most applied work – now you have a name for it.
- Formally, $y_t = \epsilon_t$ where $\epsilon_t \sim N(0, \sigma^2)$
- *So we are back where we started.*


random
2
1
0
-1
-2
-3
0
50
100
150
200
counter


random
2
1
0
-1
-2
140
150
160
170
180
counter


75
70
65
60
55
50
mood
0
50
100
150
200
Quarters