Welcome to instats

The Session Will Begin Shortly

(At the top of the hour, Eastern USA time)

1

START

2

Nonlinear Time Series Analysis, Part II: Modeling and Phenomenology

Barney Ricca
Lyda Hill Institute for Human Resilience
University of Colorado Colorado Springs

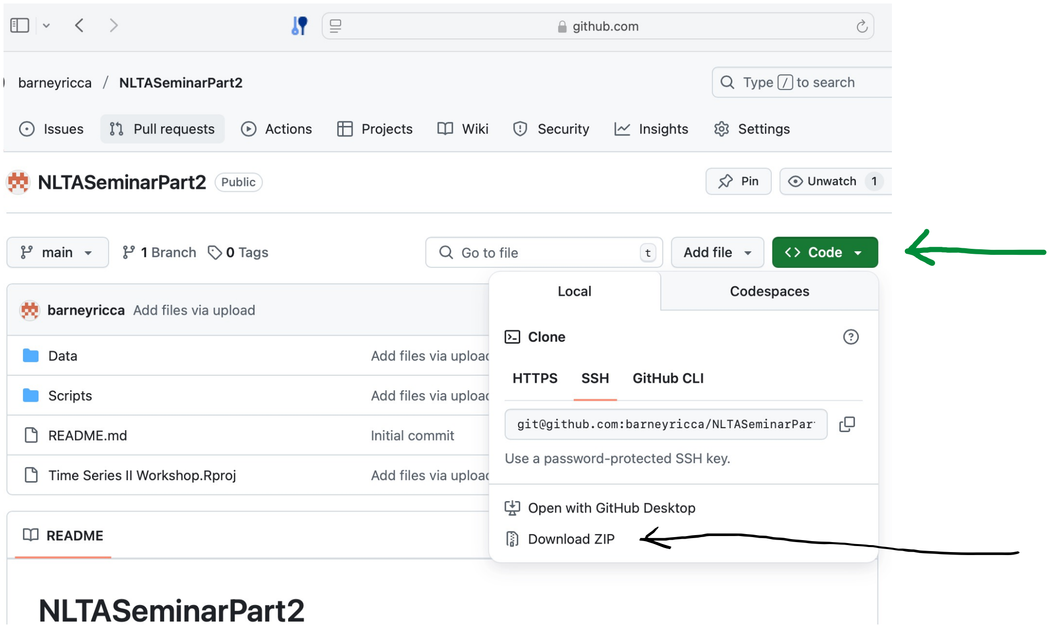
3

Seminar Overview

- Day 1
  - **Session 1 – Overview of Phenomenology and Tools**
  - Session 2 – Dynamical Systems Analysis
- Day 2
  - Session 3 – Sparse Identification of Nonlinear Dynamics
  - Session 4 – Dynamic Mode Decomposition
- Day 3
  - Session 5 – Hidden Markov Models
  - Session 6 – Machine Learning Approaches
- Day 4
  - Session 7 – Putting it All Together: Lorenz
  - Session 8 – Putting it All Together: Infectious Diseases

4

https://github.com/barneyricca/NLTASeminarPart2



5

Session Outline

- Introductions
- Modeling
- Phenomenology
- Data wrangling
- Mathematical helpers
  - Estimating derivatives
  - Multilevel Modeling
    - Latent classes
  - Regularization
- Overview of approaches

6

Introductions

- You
  - Location and field of interest into the chat, please
- Barney Ricca
  - ”Upstate” New York (USA)
  - Physics, computer science, statistics: Data scientist
  - STEM Education
  - Psychology (trauma and resilience)
  - Idiosyncratic R user
- The schedule
  - Should be reasonably close, but time left at the end because it won’t be…

7

Modeling From Theory

- Theory
- Operationalizing
  - Trajectories of theories
  - Being specific
- Coherence of assumptions

8

Phenomenological models mathematically describe relationships among empirically observed phenomena without attempting to explain [sic] the underlying mechanisms....Within the context of NLTS, phenomenological modelling goes beyond phase space reconstruction to extract equations governing real-world system dynamics from....observed time series....Data-driven phenomenological modelling is a valuable intermediary between empirical observation and theory, because reliable explanation depends on accurate descriptions of how things actually work.

HBR, p. 264

9

Phenomenological models mathematically describe relationships among empirically observed phenomena without attempting to explain [sic] the underlying mechanisms....Within the context of NLTS, phenomenological modelling goes beyond phase space reconstruction to extract equations governing real-world system dynamics from....observed time series....Data-driven phenomenological modelling is a valuable intermediary between empirical observation and theory, because reliable explanation depends on accurate descriptions of how things actually work.

HBR, p. 264

10

Phenomenological models mathematically describe relationships among empirically observed phenomena without attempting to explain [sic] the underlying mechanisms....Within the context of NLTS, phenomenological modelling goes beyond phase space reconstruction to extract equations governing real-world system dynamics from....observed time series....Data-driven phenomenological modelling is a valuable intermediary between empirical observation and theory, because reliable explanation depends on accurate descriptions of how things actually work.

HBR, p. 264

11

Phenomenological models mathematically describe relationships among empirically observed phenomena without attempting to explain [sic] the underlying mechanisms....Within the context of NLTS, phenomenological modelling goes beyond phase space reconstruction to extract equations governing real-world system dynamics from....observed time series....Data-driven phenomenological modelling is a valuable intermediary between empirical observation and theory, because reliable explanation depends on accurate descriptions of how things actually work.

HBR, p. 264

12

Setting It Up: Data Structures

- Same data structure: Vectors
  - Long data structures (pivoting)
  - Normalization and centering
  - Time column

13

Setting It Up: Imputation

- Missingness & imputation
  - Ugh
  - Various types (MCAR, MNAR, etc.)
  - TL;DR: Jäger et al. (2021) find that k-nearest neighbors and random forest imputation methods best

14

Is There a Nonlinear Signal?

- Shadow state space
- Tests for alternatives
  - Noise
  - Linearity
  - Stationarity
- Singular Spectrum Analysis (SSA)
- Repeat tests
- Convergent Cross-Mapping (CCM)

15

Mathematical Helpers

- Systems of first order equations
  - The *derivative of a derivative* trick to reduce the order to first order
  - Linear or nonlinear
- Estimating derivatives (in detail)
- Linear regression
  - Multilevel modeling
  - Regularization
- Iterative approaches
  - Expectation – Maximization

16

Derivatives

- Slopes -> rates of change -> derivative
  - Need animation (Mechanical Universe Episode 3 has derivatives, but not correct for here.)
  - Accel = d(velocity)/dt. Hence, define the slope of a slope, and make things longer. That's tough here, but do it.
- Need to spend some time with derivatives

17

Estimating Derivatives: The Problems

- TANSTAAFL
  - "there ain't no such thing as a free lunch"
- Overfitting
- Increase in noise or over-smoothing
- Bias
  - Match levels with derivatives

18

Estimating Derivatives

- Finite differences
  - Higher order
  - LLA and GLLA
- Functional data analysis
- Regularized Derivatives
- Empirical Bayes
- Multilevel-based models
  - GOLD

19

Finite Differences: Standard

- Standard
  - Good for: frequently sampled data; EDA
  - Pro: Really simple
  - Con: Shifted velocity points
- LLA and GLLA

20

Finite Differences: Higher Order

- Higher order
  - More data points
  - Savitzky-Golay, loess
- HBR 9.3 recommends fourth-order centered finite differences:

$$\dot{x} = \frac{8(x_{t+1} - x_{t-1}) - (x_{t+2} - x_{t-2})}{12\delta}$$

  - Delta is an integration step. Should be much less than 1/(sampling frequency)
  - Lose some data

21

Finite Differences: LLA and GLLA

- LLA
  - Good for: frequently sampled data; EDA
  - Pro: More flexible smoothing standard finite differences
  - Con: Loss of data (at the ends)
- GLLA
  - More general than LLA
  - Pro: Suitable for end use
  - Con: Must use filtered, not original, level data

22

Functional Data Analysis

- Nonparametric approach
  - A la splines
  - Good for: Noise reduction
  - Pro: Very flexible
  - Pro: Many levels of noise reduction possible
  - Con: Must use filtered, not original, data level
  - Pro/Con: Extra researcher degree of freedom

23

Some Statistics: Regularization

- Ordinary Least Squares errors
  - Minimize these
- Problems
  - Overfitting
  - Collinearity
- Solution:
  - Add a *regularized* term: Penalize model parameters in some way
  - Don’t care about the details, but will show one example on next slide

24

Some Statistics: Regularization

$$L = \sum_{i=1}^{n} \left(y_i + \beta_o - \sum_{j=1}^{p} \beta_j x_{ij}\right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

- Notice: Increases the loss due to every parameter!
  - Tune λ
    - λ=0 provides no help
    - λ small gets a good model
    - λ too high leads to large impact on parameters
  - *Finds solutions with smaller parameters*
- The example uses the L_2 *norm* and is known as *ridge regression*
  - L_1 *norm* is *LASSO* regression (uses absolute value, not square)
    - Can find solutions with parameters pushed to zero (i.e., dropped terms)

25

Regularized Derivatives

- OK, this is calculus
  - *Integration* is the inverse of *differentiation*.
  - Regularization penalty: How far the integral of the estimated derivative is from the original function
- Regularized derivatives:
  - Good for: Works best for long data sequences
  - Pro: Better noise characteristics
  - Con: Not currently implemented in *R*
    - But *Python* and *Matlab*
    - Can run *Python* from *R* using *package:reticulate*

26

More Statistics: Multilevel Modeling

- Nested data are not i.i.d.
  - Measurements (across time) within a person
  - Students within classrooms within schools
- Goldilocks and the Three Approaches
  - No pooling (ignore the groups)
  - Complete pooling (ignore the individual)
  - Partial pooling (just right)
- Why does pooling matter?
  - Parameter estimates are good regardless
  - Inferential statistics are terrible*, except in partial pooling*

27

Generalized Orthogonal Local Derivative

- Is GOLD a derivative?
  - Yeah, but perhaps an odd one
- GOLD
  - Good for: Multilevel modeling
  - Pro: reduces collinearity problems
  - Con: Harder to interpret

28

Emprical Bayes

- Use one subject as basis for another
  - Good for: Mixed models (e.g., across subjects)
  - Pro: Keeps noise down
  - Pro: Seems better than GOLD or GLLA or FDA
  - Con: Multivariate normal derivative distribution

29

Iterative Approach

- Derivative to model to better idea of derivative to use and rework it.
  - As you learn more about the model, you might choose a better derivative
- Just a thought, though…

30

Overview of Modeling Approaches

Method	What it does	When to use it
DSA	Estimate parameters of a model involving derivatives	Some theoretical knowledge built on derivatives
SINDy	Choose a set of functional terms from a large library	Long-term (fixed-points) and trajectories
DMD	Identify the most notable modes of a system	Response to perturbations (modes)
ML	Search widely for possible dynamic relationships	Anytime, but interpretability (fails at phenomenology)
HMM	Search for a ”hidden” model underlying data	Classifying via underlying (hidden) dynamics

31

Questions

32

STOP

33

Next session @ UTC 1900

34