How Will We Know We Are in Recovery?

Applying Data Science to the Analysis of the Business Cycle

(and the Covid-19 Pandemic)

Michael Boldin, PhD.
Fox School of Business, Temple University
Department of Statistical Science

August 19, 2020
DataPhilly

Outline / todo

–Intro

–Chart 1a, 1b, 1c

–terms

–Charts2

x– SS equations – DFM results

x–MSM slide – MSM unrate prob

x–DFM 2 slide

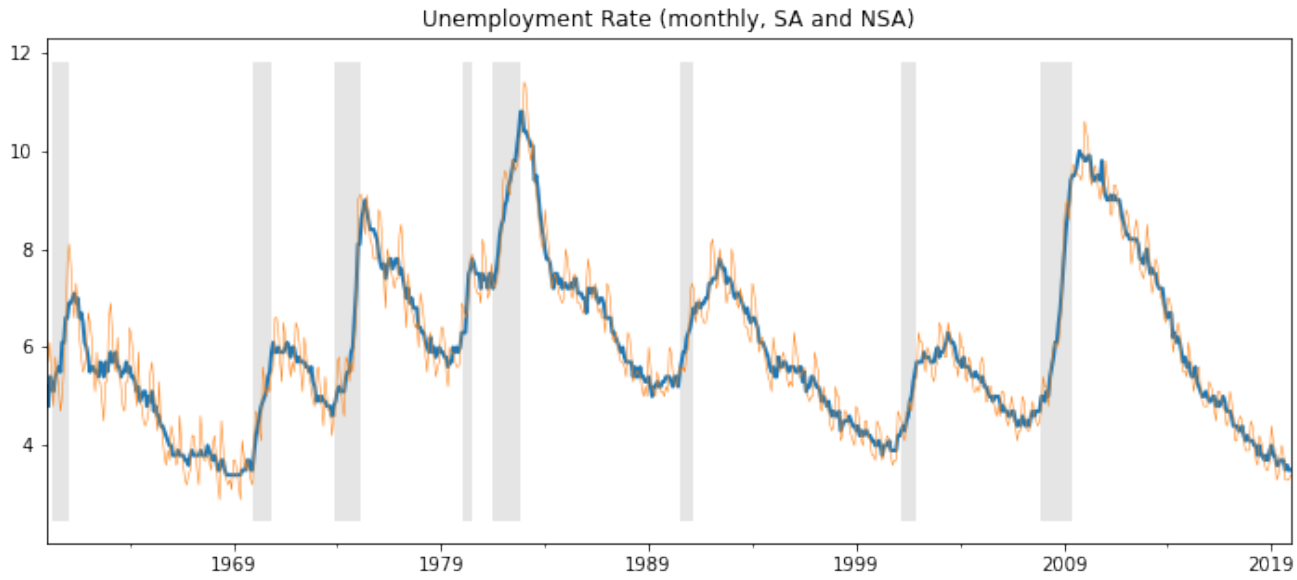
– MSA versions – service sector employ by msa

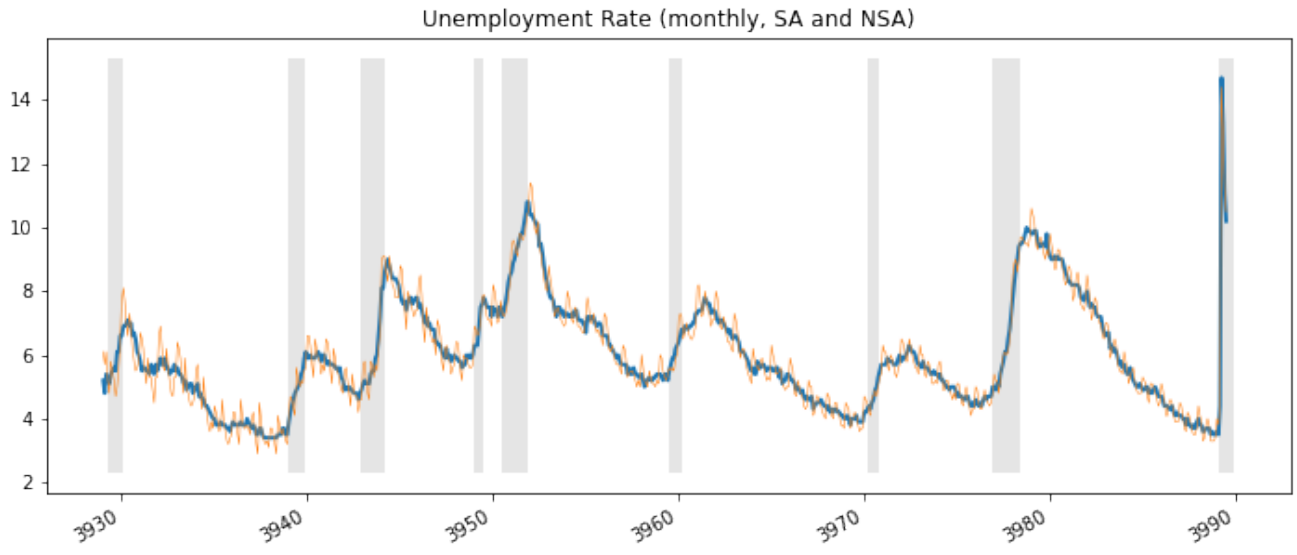
Research Project How Will We Know We Are in Recovery?

Underlying views, facts, assumptions

1. Recession started in March.
2. This recession is different, and current economic data resembles outliers (relative time series history).
3. A true economic recovery will not occur until the public feels the end to the pandemic is insight or over.
4. Need to use the data (Covid infections and economic indicators) we have, which is limited in many ways.


Unemployment Rate (monthly, SA)
11
10
9
8
7
6
5
4
3
1969
1979
1989
1999
2009
2019


Unemployment Rate (monthly, SA and NSA)
12
10
8
6
4
1969
1979
1989
1999
2009
2019


Unemployment Rate (monthly, SA and NSA)
14
12
10
8
6
4
2
3930
3940
3950
3960
3970
3980
3990

Introduction: How Will We Know We Are in Recovery?

What I will cover:

1. Advice – applying data science to a research project.
2. Explain how the economics profession looks at business cycles – not necessarily the same as Wall Street chief economists and other quoted in the business press or on TV.
3. Discuss and show available data: economic indicators and publicly available Covid stats (mainly the Johns Hopkins set).
4. Present modeling ideas and preliminary analysis steps, including why analysis at the County and Metropolitan (MSA) level and not the State level) is appropriate.
5. Discuss insights from theoretical models.

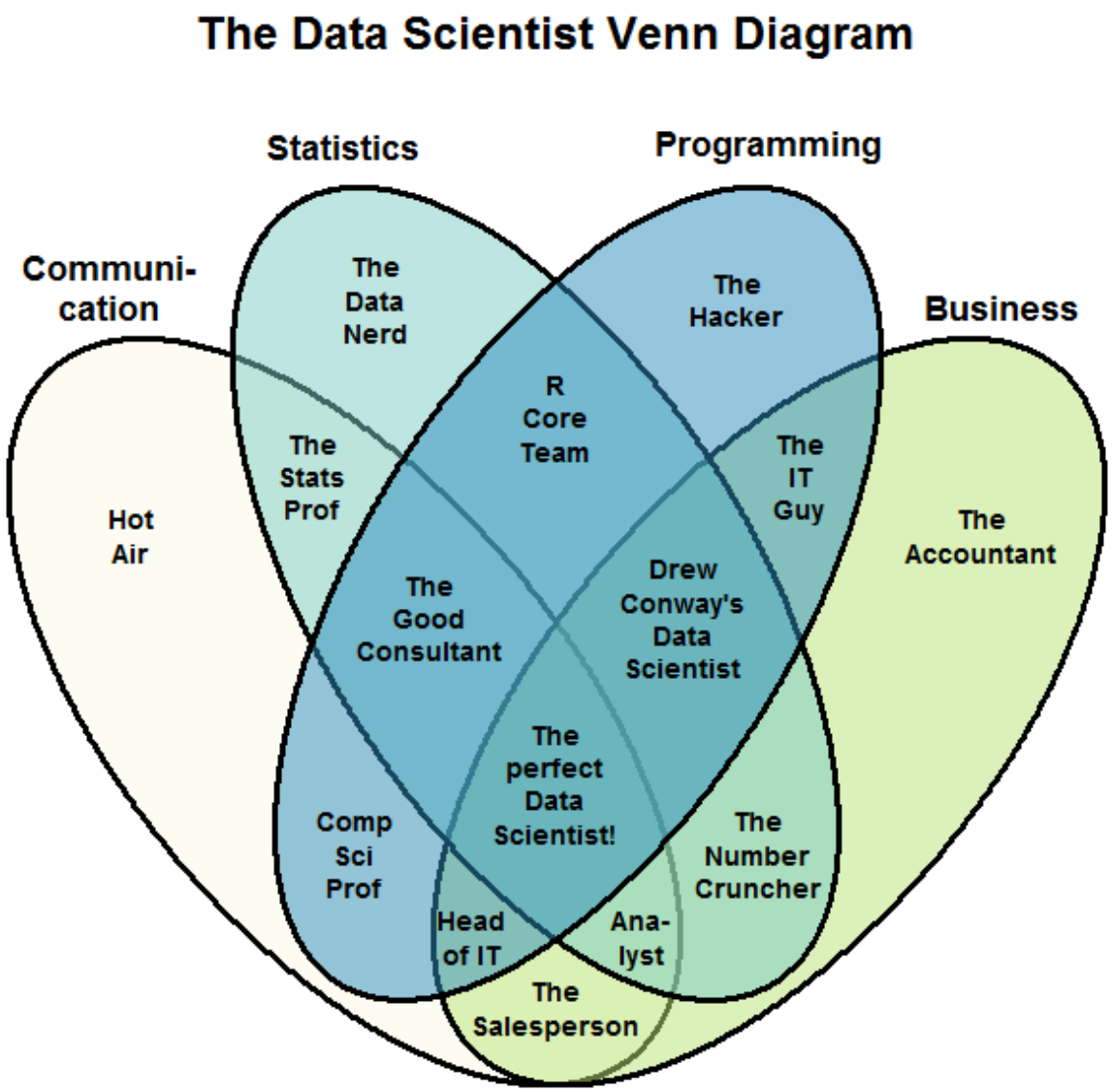
Suggestions, comments, and help are welcome GITHUB:
https://github.com/mboldin/DataPhillyAug2020

More?

Joint research with Prof. Marc Sobel (Temple Univ.)
– Dirichlet Process Clustering and Bayesian Perspective

If I had more time:

- Talk about how seasonal effects
  and seasonal adjustments to
  economic data matter.
- Discuss estimation of statistical clustering
  and the bayesian advantages.
- Discuss current research by others on the topic.
- Nitty gritty of government data sources
  and alternatives.
- Rant on how the BLS and Census should join
  the 21st century to produce better
  and more timely labor force statistics.



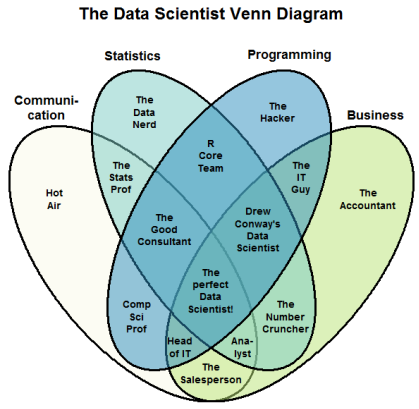
Graphic by Steven Kolassa,

compare to Drew Conway diagram of Math and Statistics, Hacking, Substantive Expertise

Project Planning and Data Science Skills

Some Advice

- Bad data management and misunderstanding the data often ruins good projects.
- Learn at least two programming languages. EXCEL does not count.
- Area expertise
  and understanding 'theory' matters–especially in understanding data limitations.
- To find a modeling sweet spot between overly simple models and overfitting, think hard about the applicable math and statistic methods.
- Data visualization helps at all stages, but don't pretend fancy charts alone yield good communication.



Business Cycle Tracking

Dating Recessions

- Rule of thumb: GDP 2-Quarter rule (for recessions), Unemployment rate movements, Coincident Index patterns.
- NBER Committee – decides on peak and trough dates with no formal rules.
- Dynamic Factor Model (linear)
- Switching Model (nonlinear)
- Combined: Factor(s) + Switching

Dynamic Factor Model

Stock-Watson Coincident Index

State Equation – Common Unobserved Dynamic Factor

$$c_t = a * c_{t-1} + v_t \qquad v \sim \mathcal{N}(0, 1)$$

Measurement Equations

$$y_{i,t} = b_i * c_t + e_{i,t} \qquad e \sim \mathcal{N}(0, \Sigma)$$

for $i = 1, 2, ..K$ indicators

Parameters estimated using

- Kalman Filter
- Maximum Likelihood Estimation (MLE)

I used the Python statsmodels module
and the DynamicFactor() method

Switching Model

Measurement Equation – GDP or unemployment rate

$$un_{i,t} = a(s_t) + b(s_t) * un_{t-1} + e_t \qquad e \sim \mathcal{N}(0, \sigma^2)$$

s = 1 Recession, unemployment rate tends to rise
s = 2 Expansion, unemployment rate tends to fall

Markov process probability of regime switching:

$$p(s_{t+1} = 1 | s_t = 2), \quad p(s_{t+1} = 2 | s_t = 1)$$

expansion-to-recession or *recession-to-expansion*
Only depends on the current 'state'

More than 2 states or regimes are possible:

3: recession-recovery-expansion
4: recession-recovery-expansion + stagnation

Dynamic Factor Model + Regime Switching

State Equation – Common Unobserved Dynamic Factor

$$c_t = a_0(s) + a_1(s) * c_{t-1} + v_t \qquad v \sim \mathcal{N}(0, \sigma_s^2)$$

Measurement Equations

$$y_{i,t} = b_0(s) + b_1(s) * c_t + e_{i,t} \qquad e \sim \mathcal{N}(0, \Sigma(s))$$

for $i = 1, 2, ..K$ indicators and s = 1 recession or 2 expansion,

Random innovations in the state equation (v_t)
are subtle (relatively small), while regime effects (through s) can yield large 'breaks'.

Parameters estimation requires alternative Kalman Filter or a Particle Filter

S-I-R Virus Model

S-I-R is the simplest case

3 containers for N individuals in total population

- **S**usceptible (not immune)
- **I**nfectious
- **R**ecovered (and immune)

Other versions add

- Exposed before Infectious (S-E-I-R)
- Quarantined and Hospitalized containers
- Death besides Recovered
- Return to Susceptible from Recovered is possible

S-I-R Virus Model

Equations for S, I, & R

Laws of Motion:

$dS = a * I * (S/N)$

$dI = a * I * (S/N) - b * I$

$dR = b * I$

Fixed parameters

a = number of contacts of N per day
    * probability a contract of an S individual results in an infection

b = recovery rate = 1 / (days infectious)

Easiest to solve the continuous time version.

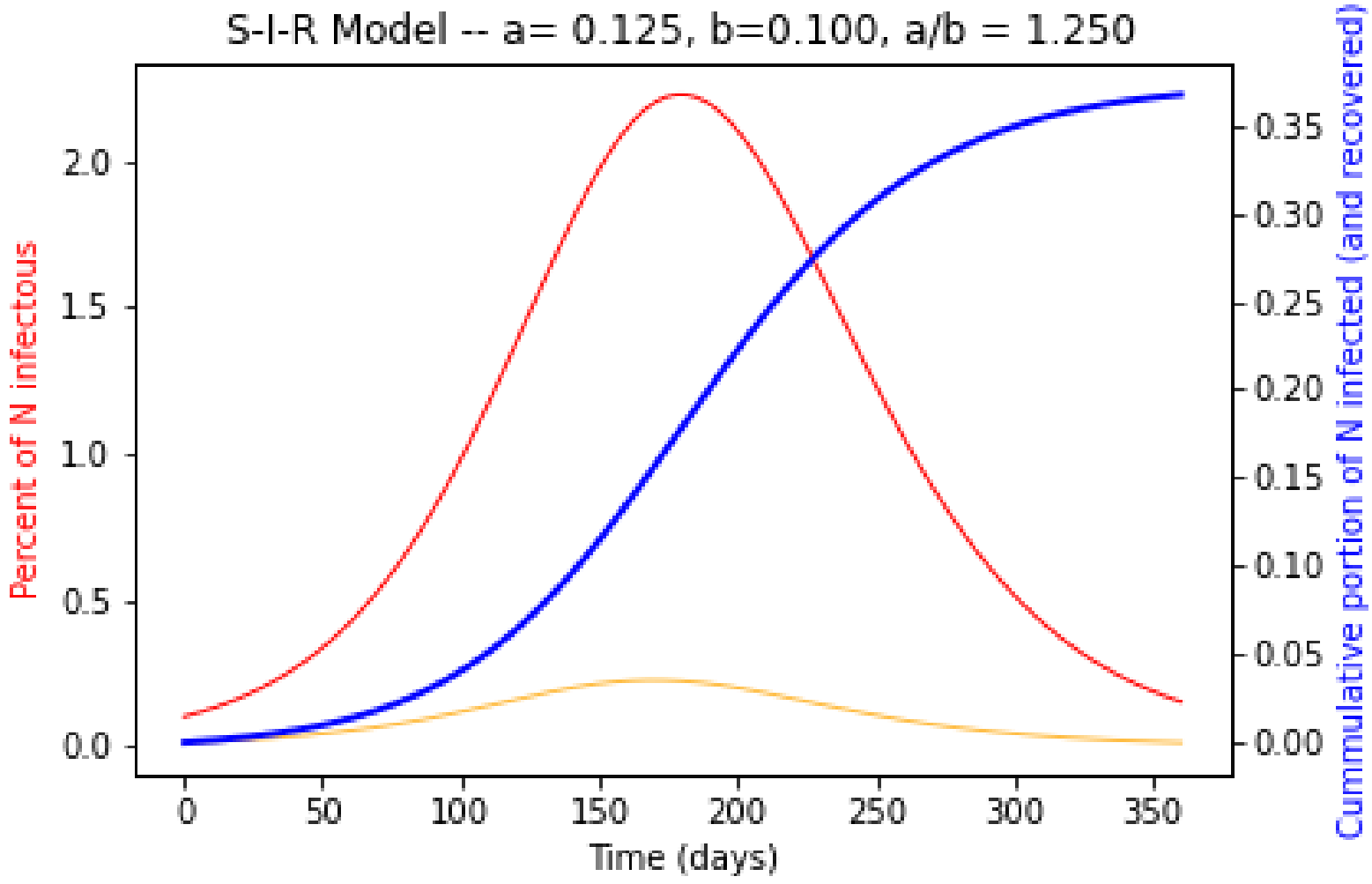
Need to discreetize to 'apply' to data,

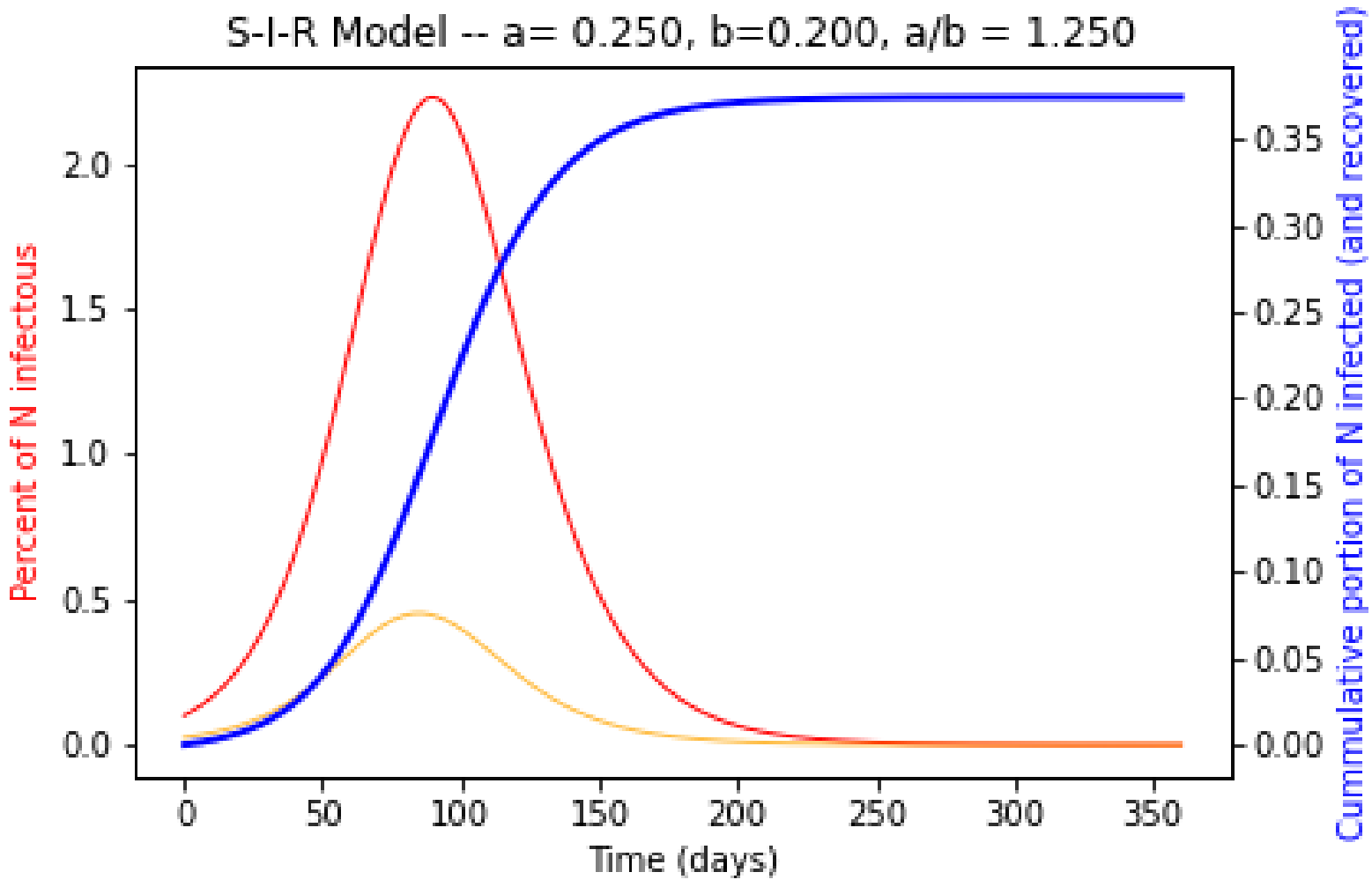
but data is not in the S, I, R form

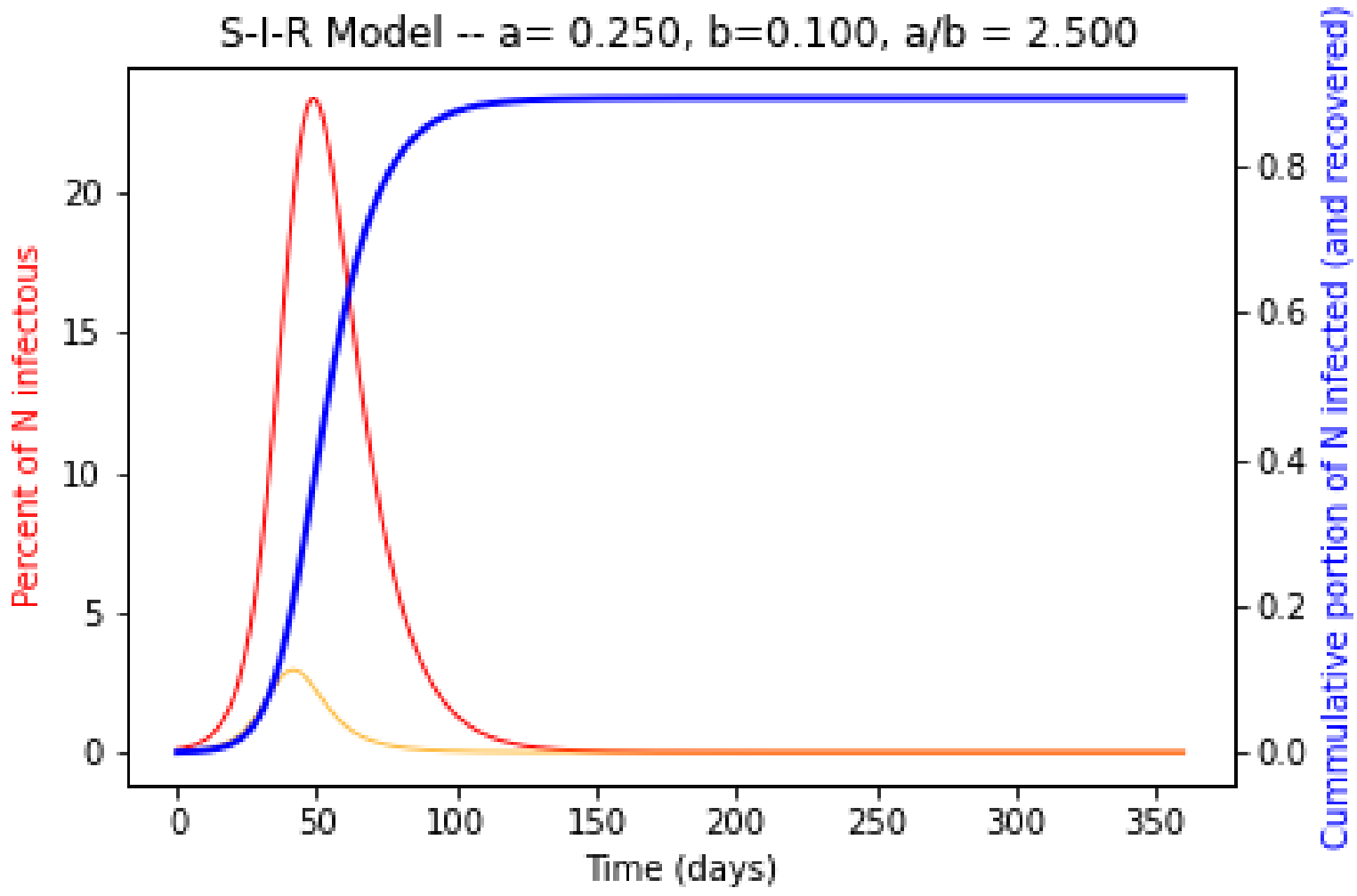
S-I-R Virus Model

Insights

- Ratio a/b is the key
  A high infection rate & long infection time is bad.
- Herd immunity effects: I falls to 0 when $(S/N) * (a/b) < 1$.
- Hard to change number infected
  – but can lower the maximum I/N,
  which elongates the infection time (flattening the curve)
- Can let the a parameter vary as peiopel change their behavior (social distancing) so R(t) = a(t)/b can be estimated as a time-series.


S-I-R Model -- a= 0.125, b=0.100, a/b = 1.250
Percent of N infectous
Cummulative portion of N infected (and recovered)
Time (days)
0.0
0.5
1.0
1.5
2.0
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
50
100
150
200
250
300
350


S-I-R Model -- a= 0.250, b=0.200, a/b = 1.250
Percent of N infectous
Cummulative portion of N infected (and recovered)
Time (days)
0.0
0.5
1.0
1.5
2.0
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
50
100
150
200
250
300
350


S-I-R Model -- a= 0.250, b=0.100, a/b = 2.500
Percent of N infectous
Cummulative portion of N infected (and recovered)
Time (days)
0
5
10
15
20
0.0
0.2
0.4
0.6
0.8
0
50
100
150
200
250
300
350

Infection Model Fitting at County Level

Dependent variable:

- Change in confirmed covid cases per 1000

Explanatory variables:

- lag of change in confirmed covid cases per 1000 (+)
- Lag of confirmed covid cases (to date) per 1000 (-?)
- Population density per square mile of land (+)
- Percent of population with HS diploma (-)
- Median household income (-)
- Unemployment rate average in 2019 (?)

...

In a simple linear panel model, almost all of the county level explanatory variables seem statistically significant, but R^2 is only around 0.25.

Infection Model Fitting at County Level

Dependent variable:

- Change in confirmed covid cases per 1000

Allowing for 'regimes' that represent county groupings or clusters is attractive for a variety of reasons.

- Unsure about the proper number of regimes
- So a Dirichlet Process (DP) model with infinite potential clusters is being developed.
- Using Bayesian priors to solve estimation problems with the unconstrained model.

Full Model – Economic Indicators + Covid Cases by County

Possible County level clusters (regimes)

3 x 3 x 3 x 3 regime clusters

Open	Low cases	Rising infection	Economic recovery
Partially open	Modest cases	Steady infection	Stagnation
Closed	High cases	Rising infection	Contraction

The degrees of Open, Infection cases,
and Economic contraction or recovery
will depend on what works in making the 'best' cluster sets.

Conclusion

- first item
- second bullet point
- last item