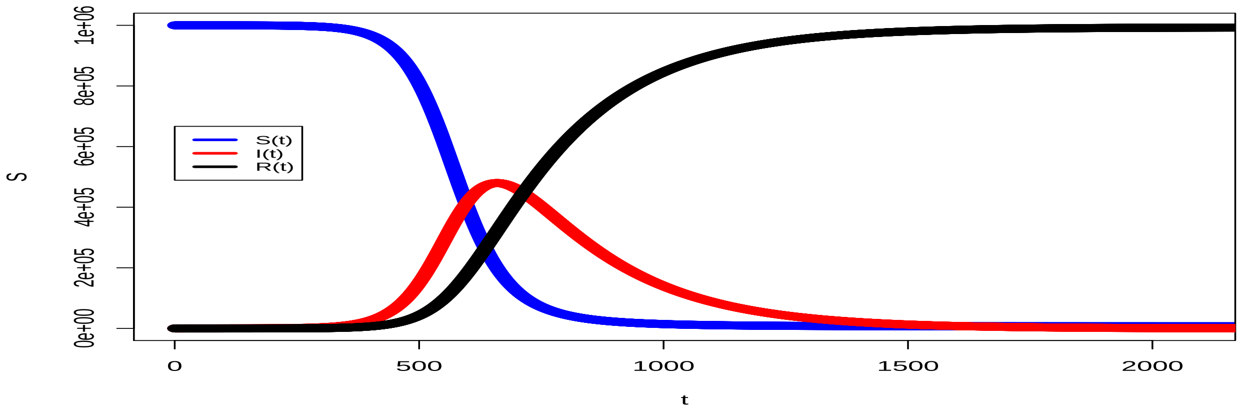
Modeling and Estimation of COVID-19 Under-Reported Counts

Eric Baron[1] and Michael Baron[2]

[1]University of Connecticut, Storrs CT, and [2]American University, Washington DC



Supported by the U.S. National Science Foundation

Data Science Conference on COVID-19 [DSCC-19], August 28, 2020

Plan

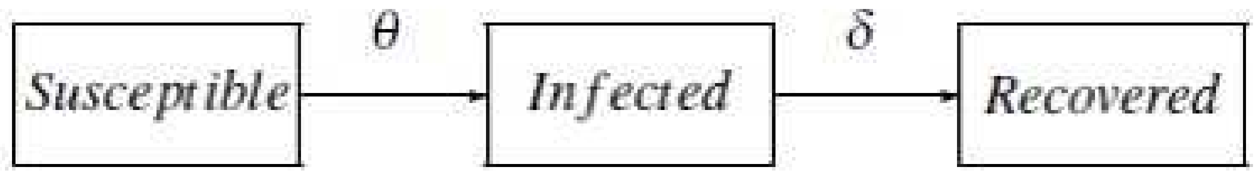
Under-Reported Counts

- Modeling
  - Untested people
  - Unconfirmed transmissions
  - Under-reported recoveries
  - Parameters and their dynamics
- Estimation
  - Bayesian approach
  - MCMC
- Discussion

SIR epidemic model

Population consists of 3 groups: **S**usceptible, **I**nfected, **R**ecovered.

Any individual can move through the states in order, $S \to I \to R$



$S_t = \#$ of individuals not infected at time t

$I_t = \#$ of individuals infected

$R_t = \#$ of individuals infected and then recovered

- Infection rate θ
- Recovery rate δ

Kermack and McKendrick 1927

SIR Model

$$\boxed{\mathbf{S}} \xrightarrow{\theta = \beta r I/N} \boxed{\mathbf{I}} \xrightarrow{\delta} \boxed{\mathbf{R}}$$

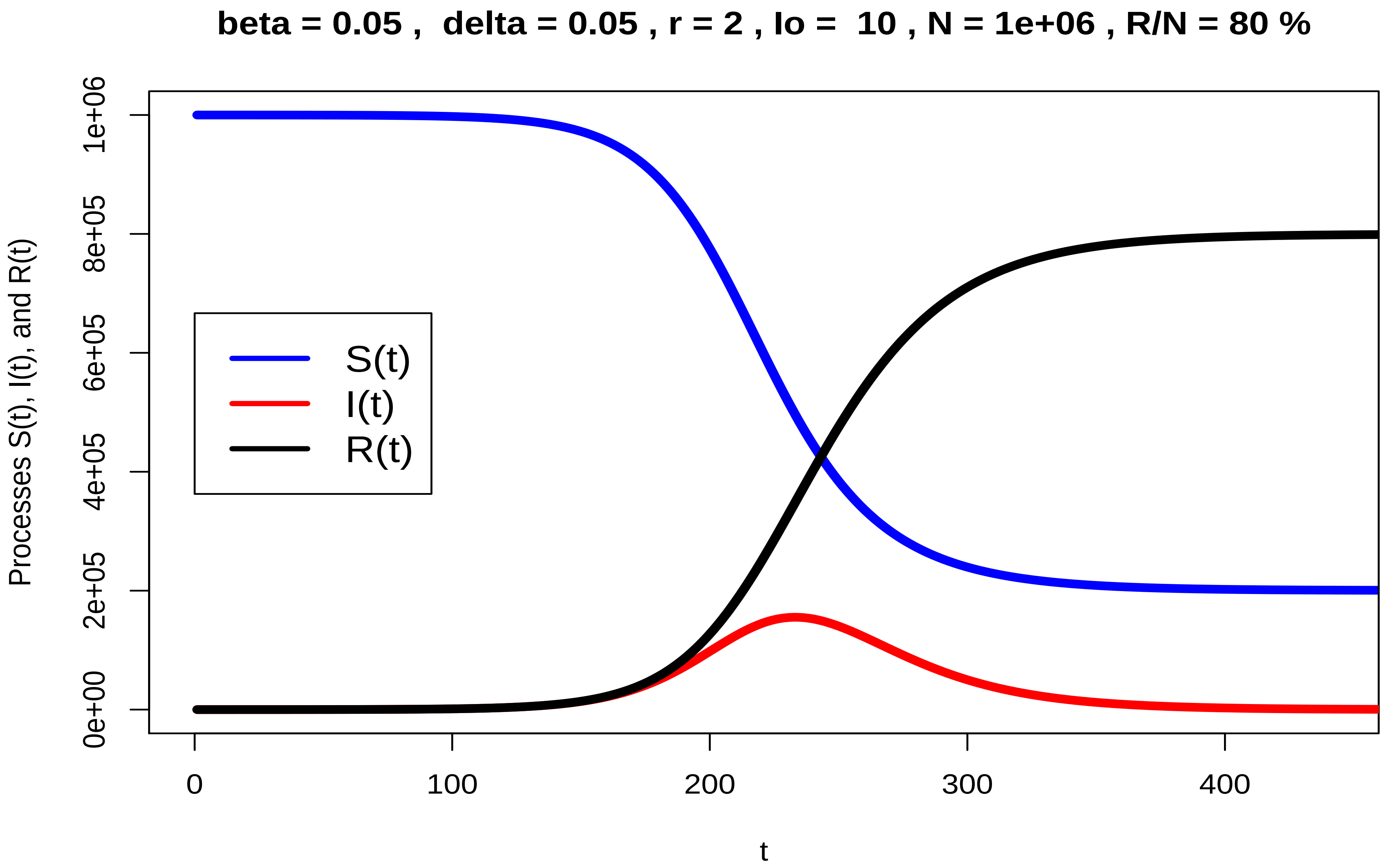
$$\begin{cases} \dfrac{dS(t)}{dt} & = & -\beta r \frac{I(t)}{N} \\ \dfrac{dI(t)}{dt} & = & \beta r \frac{I(t)}{N} - \delta I(t) \\ \dfrac{dR(t)}{dt} & = & \delta I(t) \end{cases} \quad \text{with} \quad \begin{cases} S(0) & = & N \\ I(0) & = & 1 \\ R(0) & = & 0 \end{cases}$$

r = number of contacts per unit of time

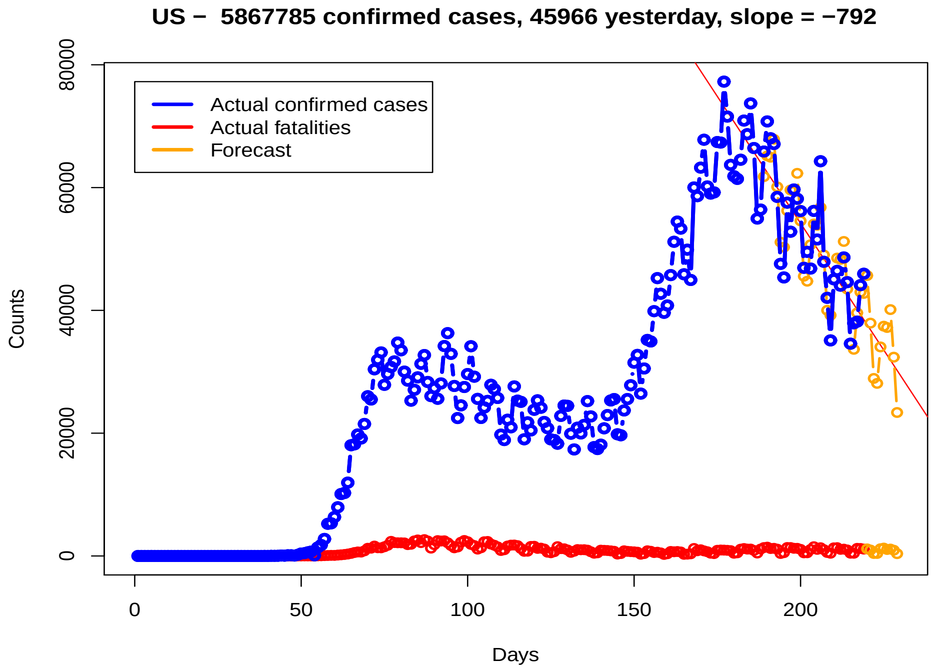
N = population size

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$$

SIR Model


beta = 0.05 , delta = 0.05 , r = 2 , Io = 10 , N = 1e+06 , R/N = 80 %
Processes S(t), I(t), and R(t)
0e+00
2e+05
4e+05
6e+05
8e+05
1e+06
0
100
200
300
400
t
S(t)
I(t)
R(t)

SIR Model. Does not quite fit the USA data...

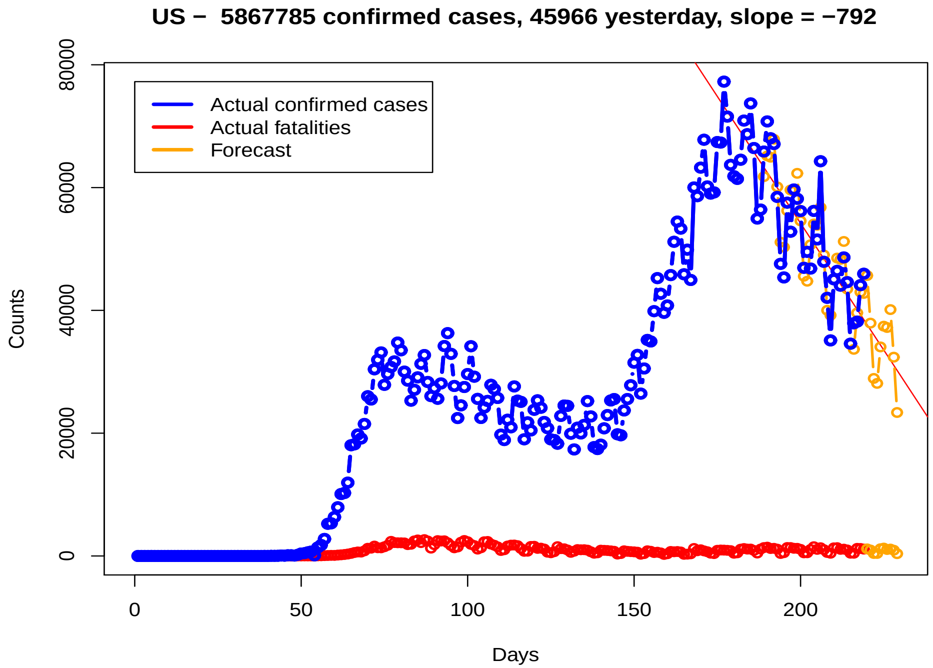


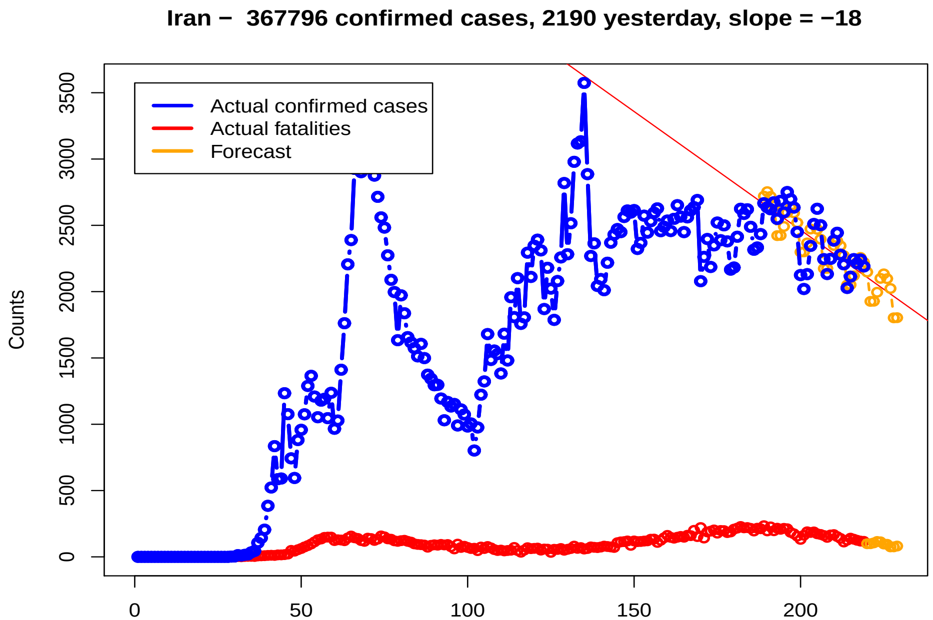
Data from:

Humanitarian Data Exchange (HDX)

https://data.humdata.org/dataset/

SIR Model. Does not quite fit USA, Iran...



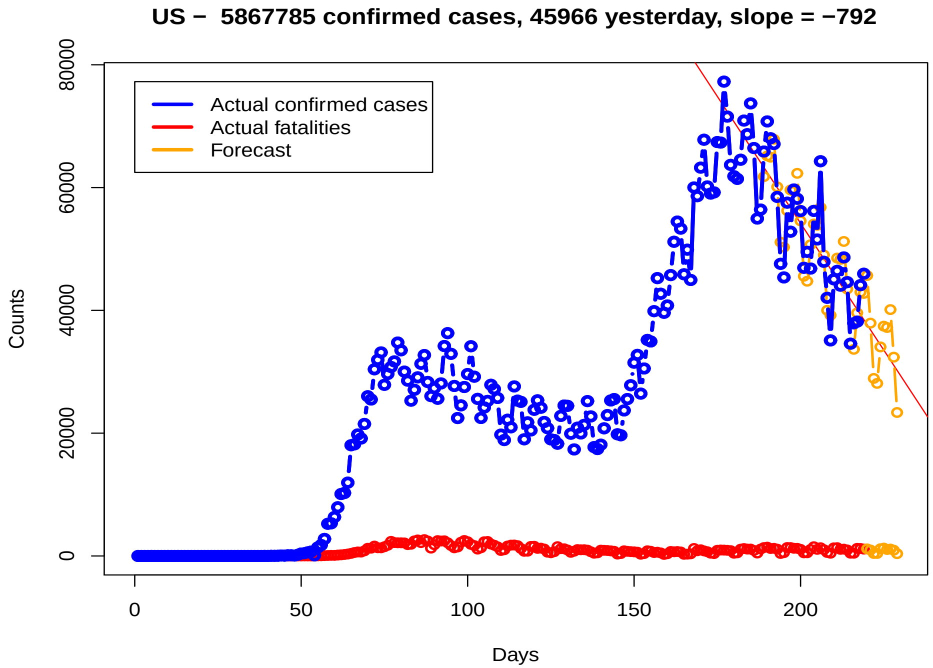


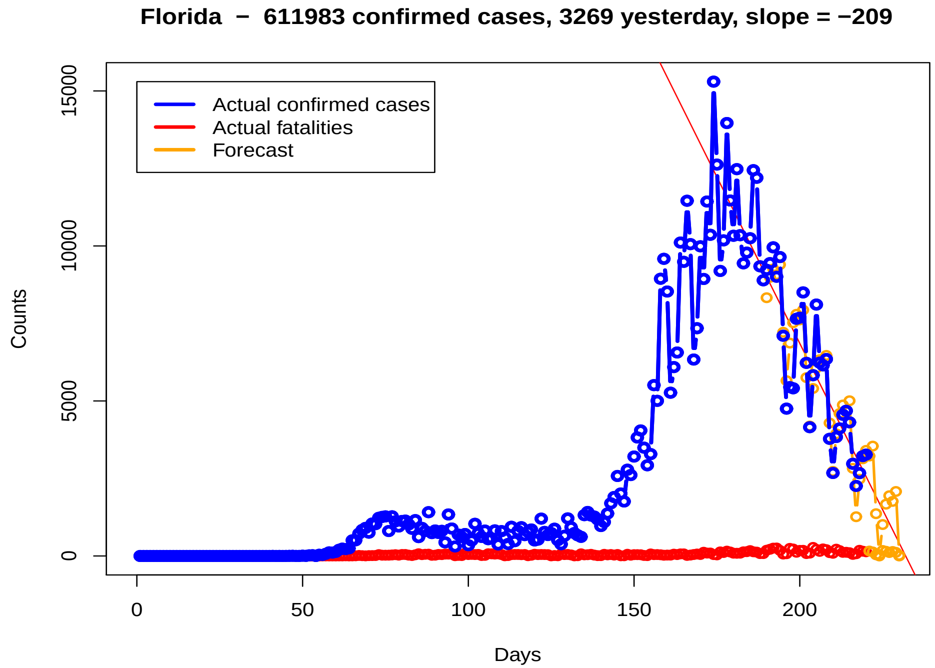
Data from:

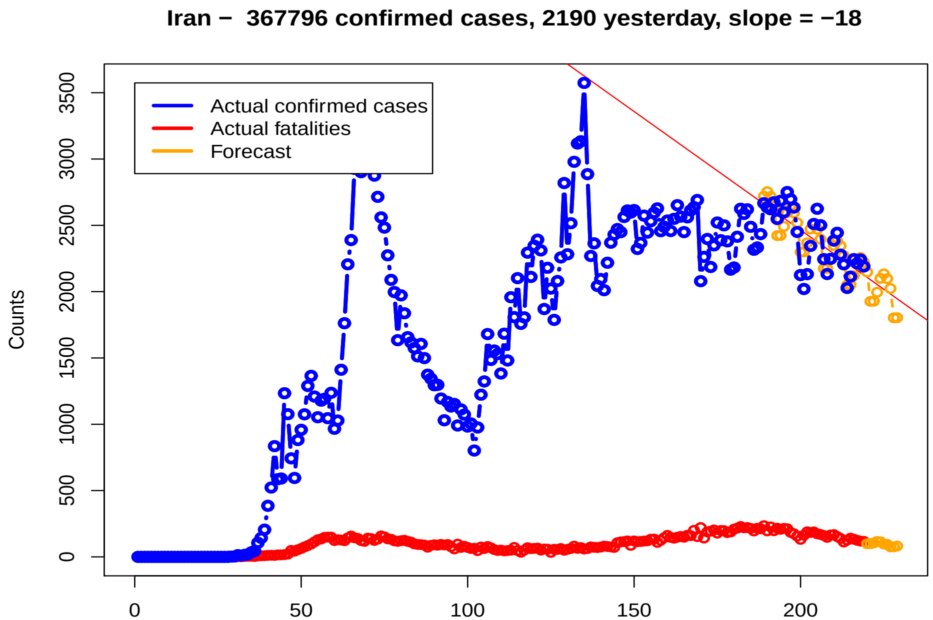
Humanitarian Data Exchange (HDX)

https://data.humdata.org/dataset/

SIR Model. Does not quite fit USA, Iran, Florida...





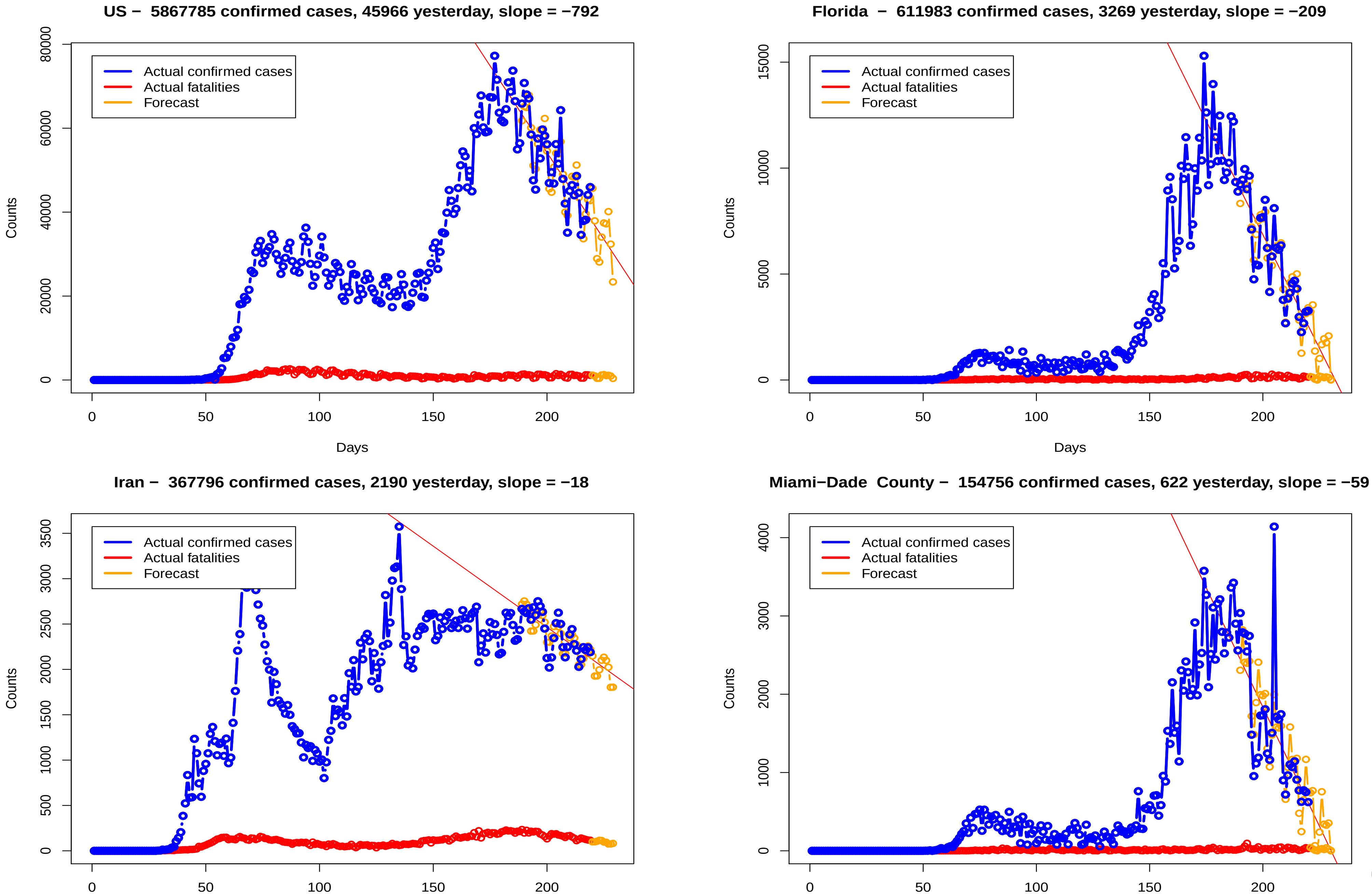


Data from:

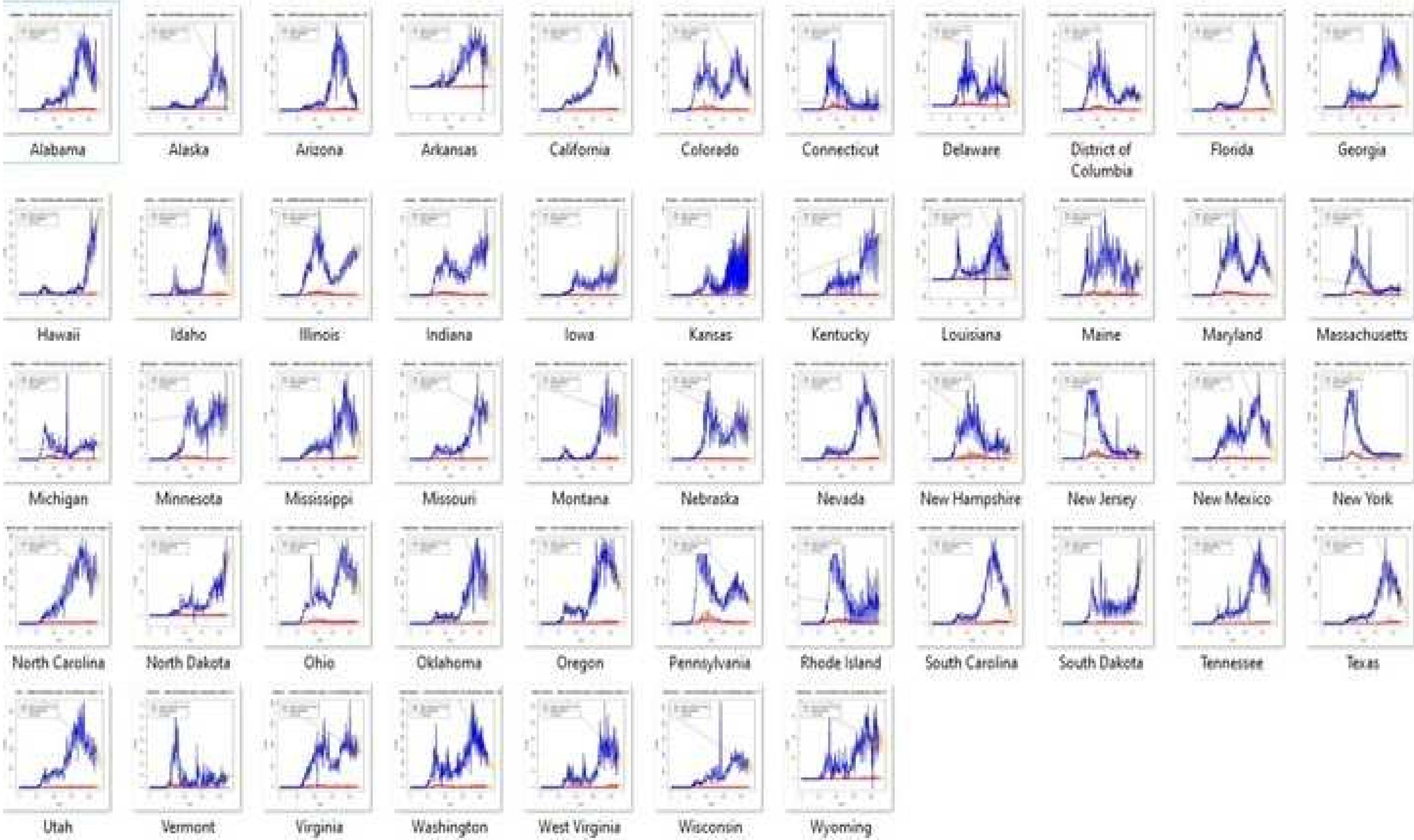
Humanitarian Data Exchange (HDX)

https://data.humdata.org/dataset/

SIR Model. Does not quite fit USA, Iran, FL, M-D county



Why are the states so different?



SIR model

$$\boxed{S} \xrightarrow{\theta = \beta r I / N} \boxed{I} \xrightarrow{\delta} \boxed{R}$$

θ = infection rate

β = transmission rate

δ = recovery rate

r = average number of contacts

N = population size

<u>Counts</u> :

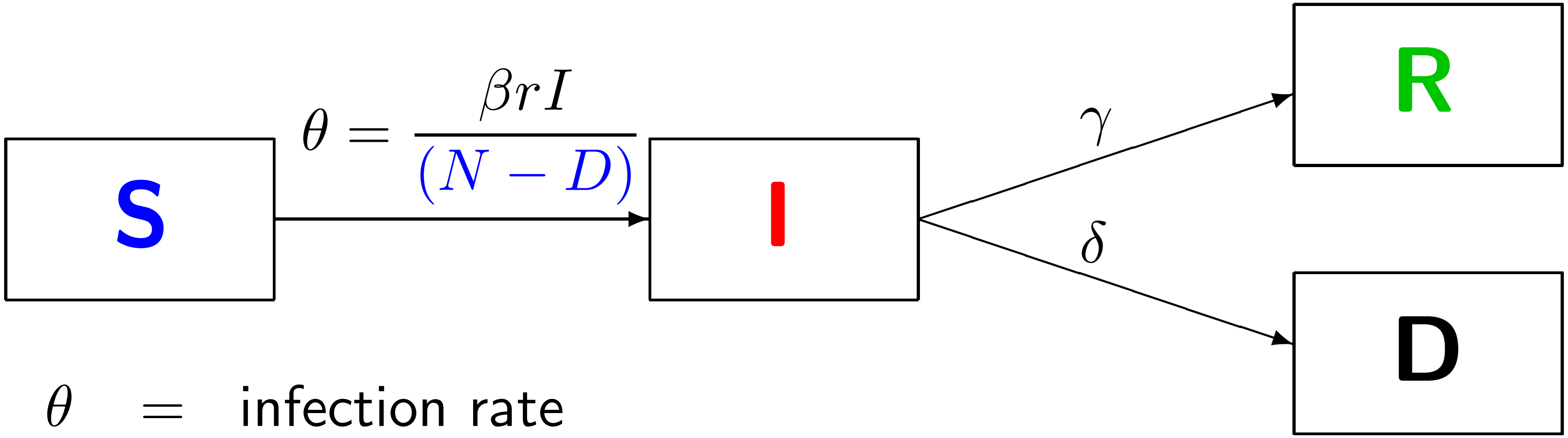
Susceptible

Infected

Recovered

Problem 1: some people do not recover

SIRD model



θ	$=$	infection rate
β	$=$	transmission rate
γ	$=$	recovery rate
δ	$=$	mortality rate
r	$=$	average number of contacts
N	$=$	population size

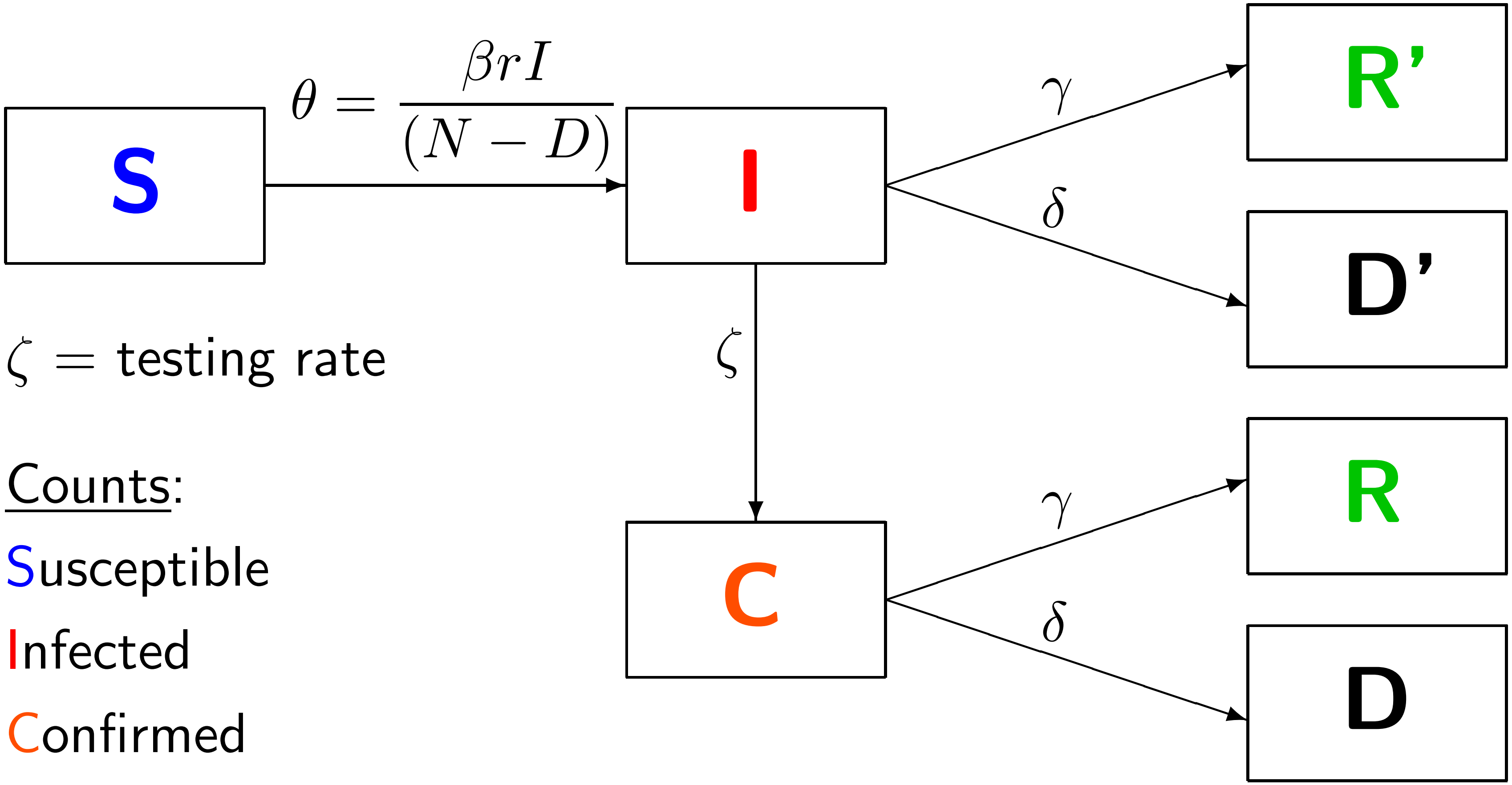
Counts :

Susceptible	Recovered
Infected	Died

Problem 2: many people are not tested.

USA: Population 331M; 75M tested; 5.9M confirmed

SICRD model



ζ = testing rate

Counts:

Susceptible

Infected

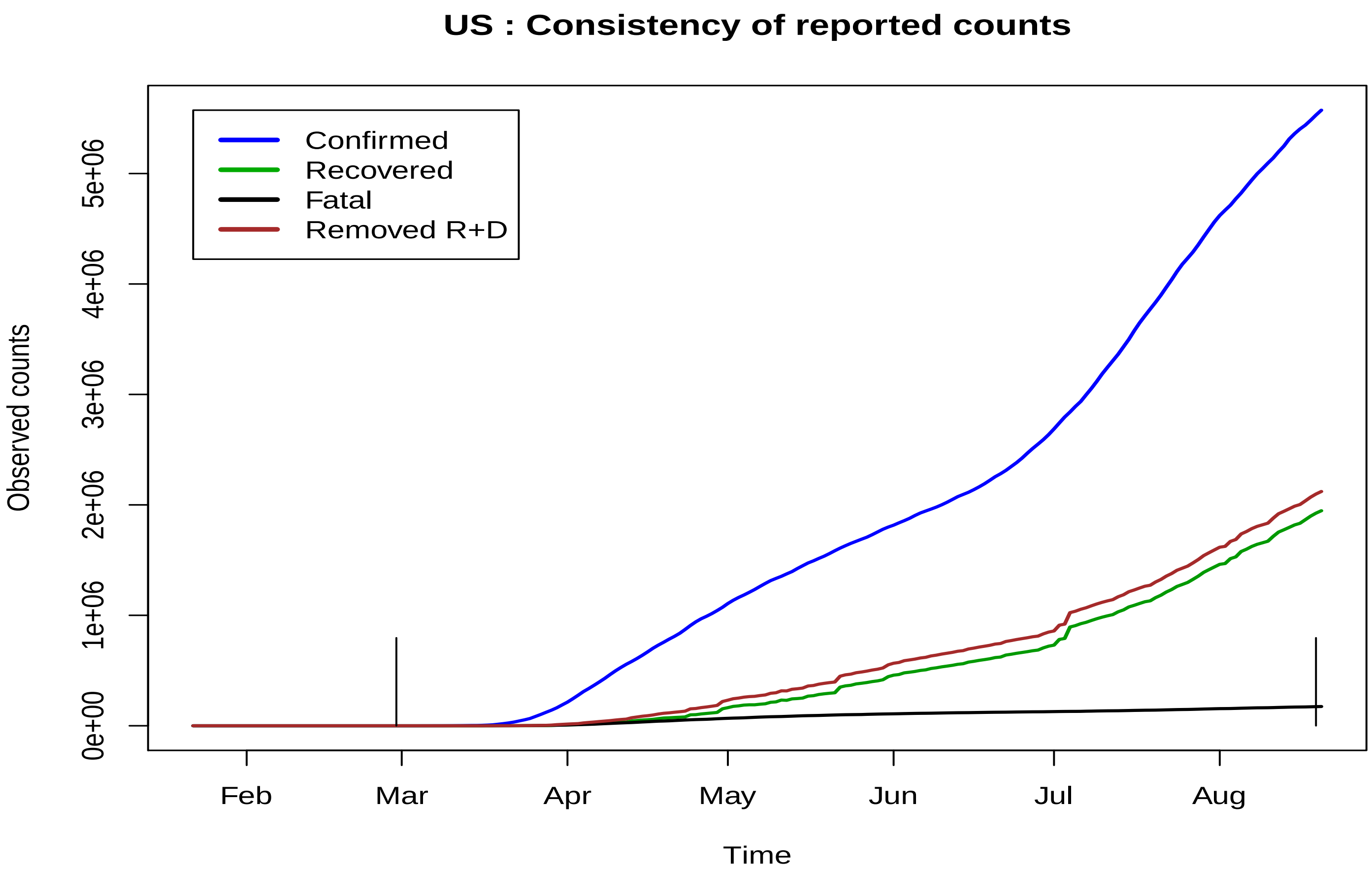
Confirmed

Recovered

Died

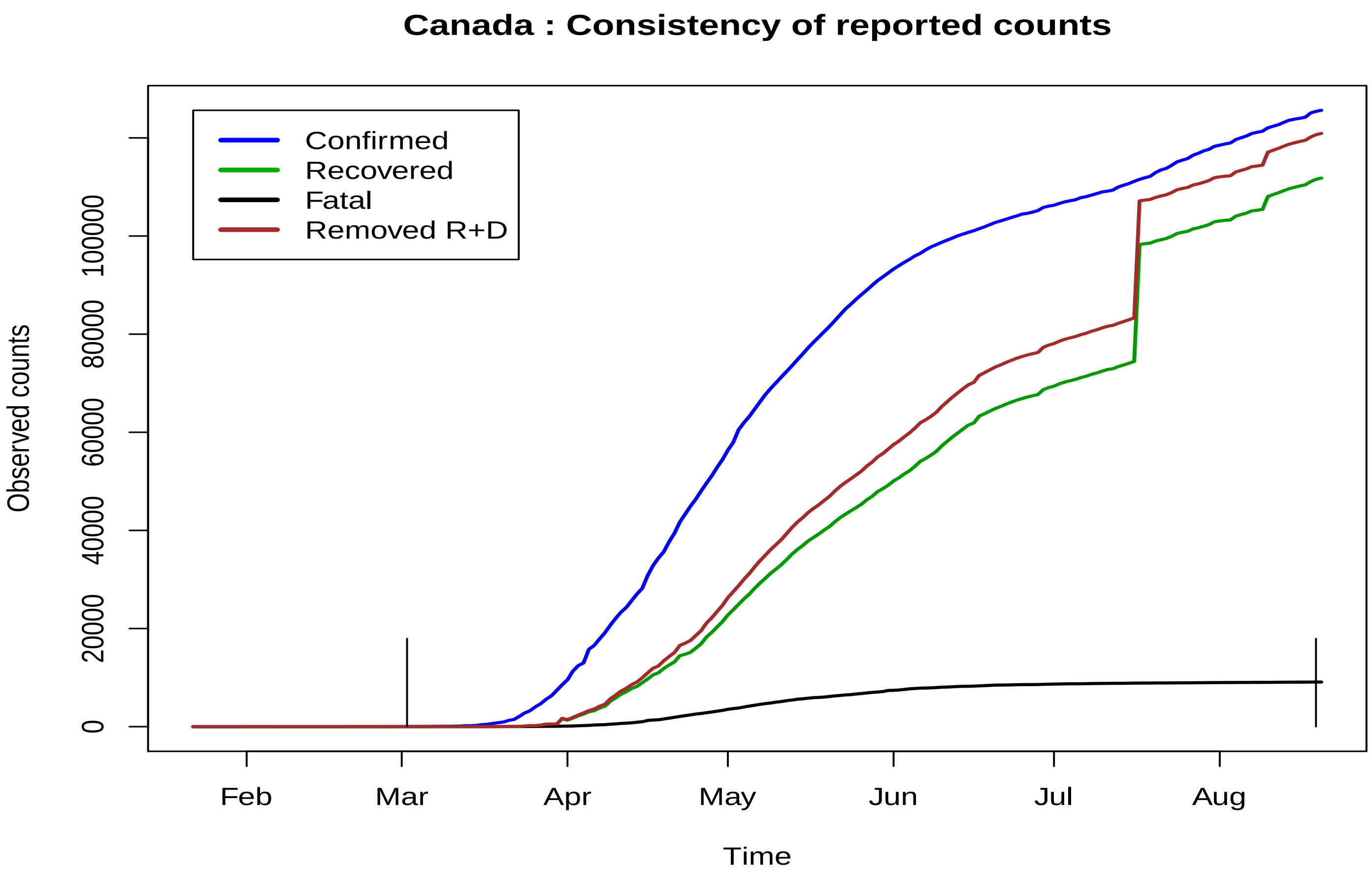
Problem 3: under-reported counts

Inconsistency of reported counts: USA



Officially:	2.69M reported confirmed cases in the USA by July 1st
	1.95M recovered and 0.17M perished by August 20th
	566,298 not reported (actually, many more)

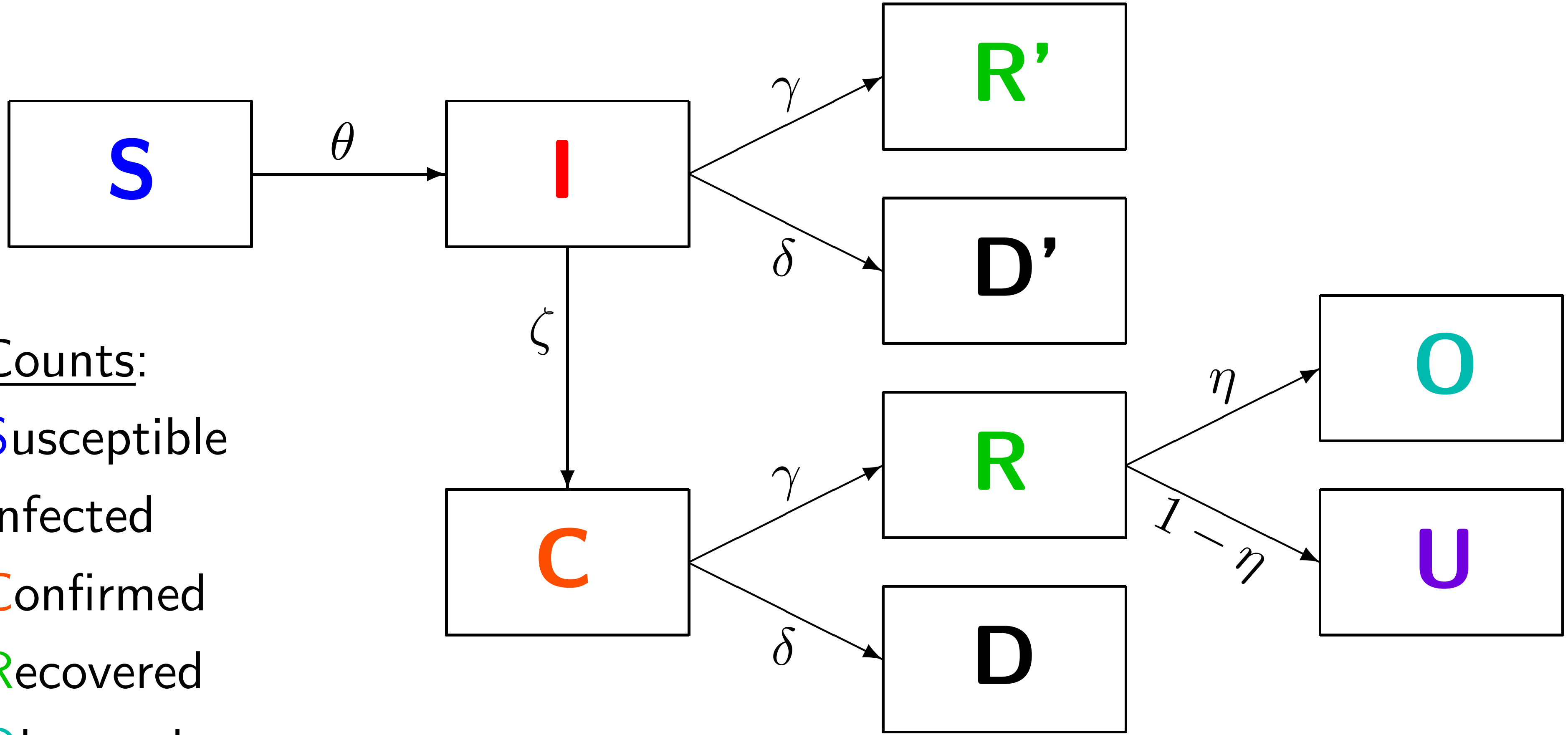
Inconsistency of reported counts: Canada



Officially: 23,848 people recovered on July 17

No more than 2,630 people recovered on any other day

SICROUD model



Counts:

Susceptible

Infected

Confirmed

Recovered

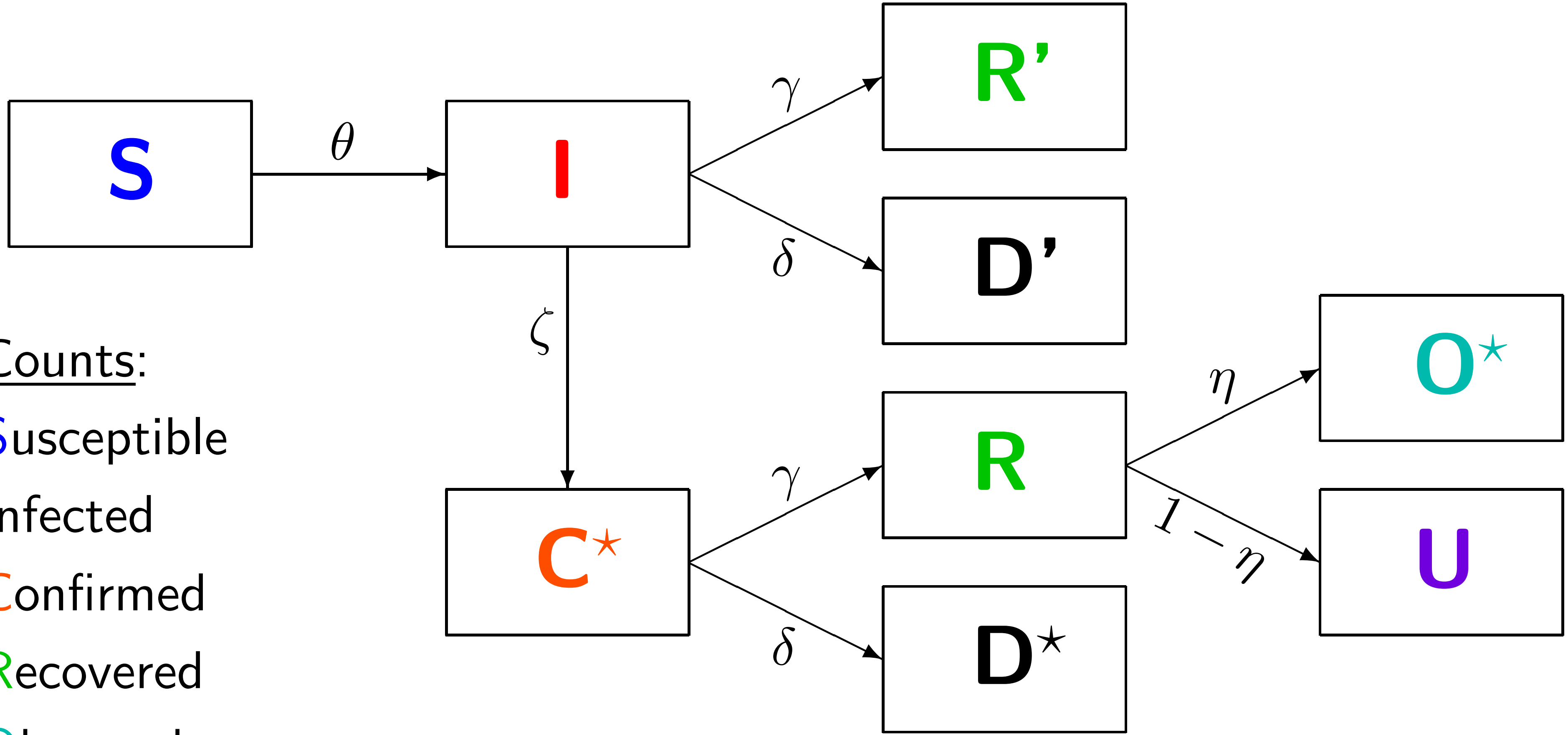
Observed

Unobserved

Died

η = reporting rate

SICROUD model



<u>Counts</u>:

Susceptible

Infected

Confirmed

Recovered

Observed

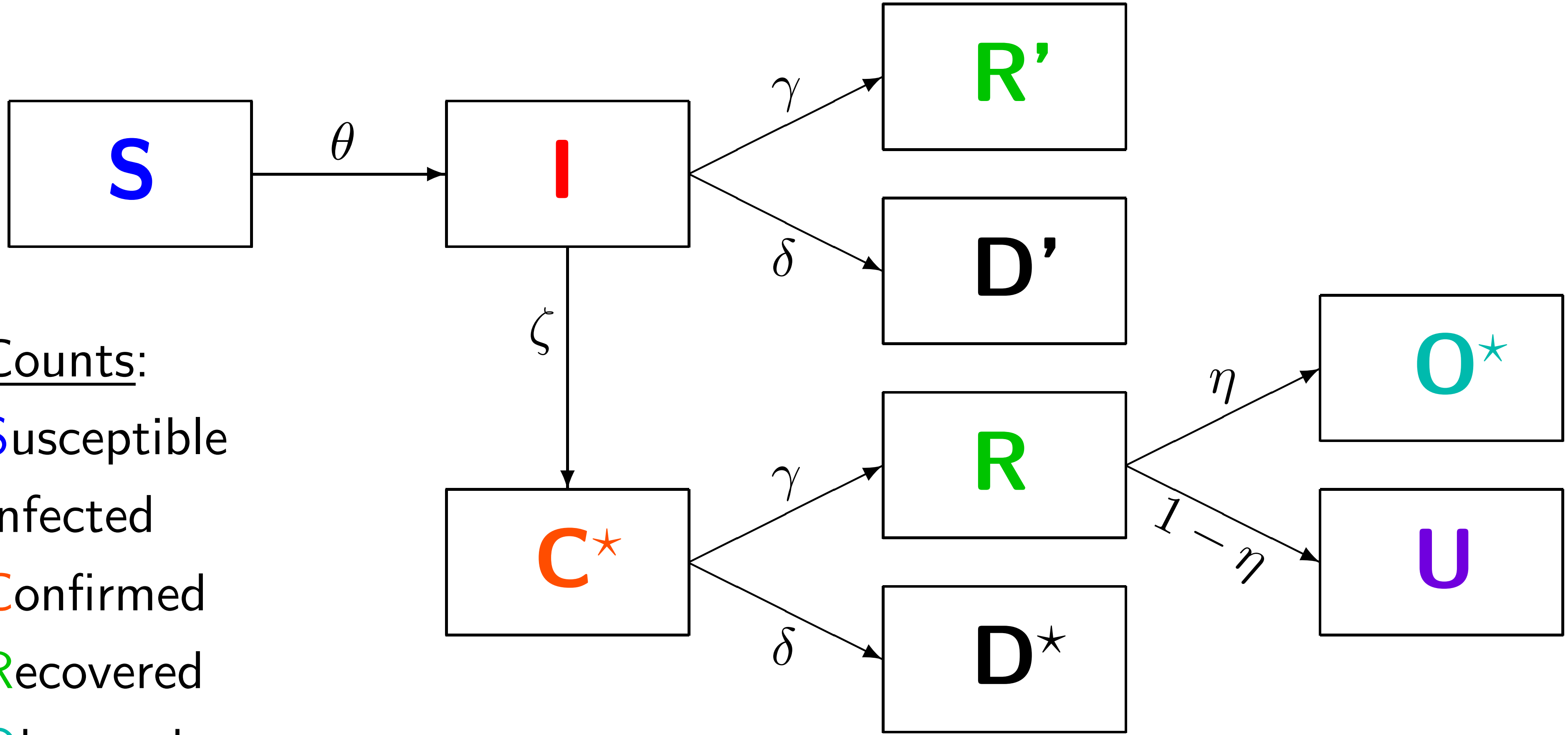
Unobserved

Died

<u>Data</u>:

Confirmed C_t, Observed recoveries O_t, Died D_t

SICROUD model



Counts:

Susceptible

Infected

Confirmed

Recovered

Observed

Unobserved

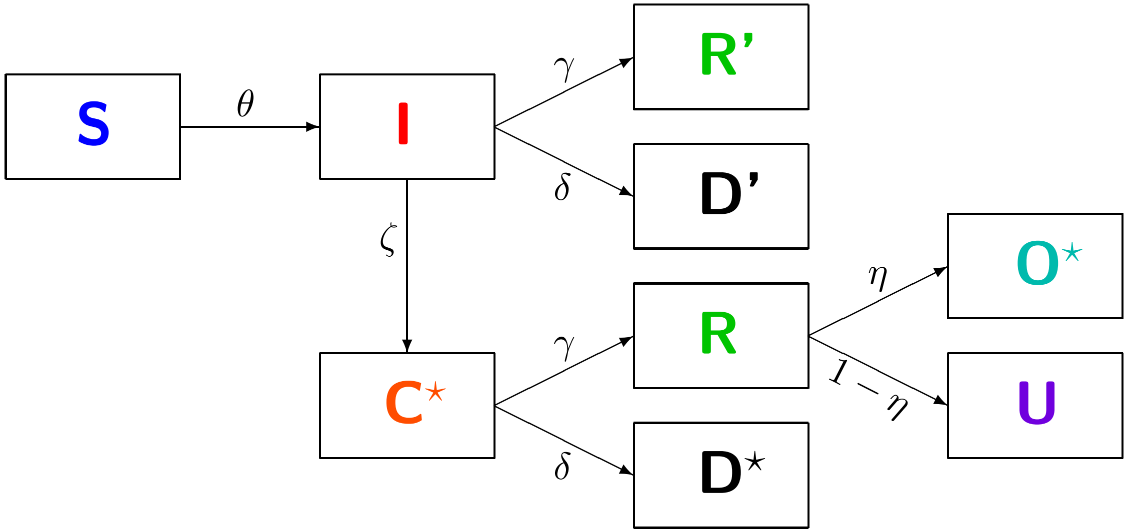
Died

Parameters: θ, γ, δ, ζ, η

Problem 4: SIR model is deterministic.

Bayesian Modeling: Counts and Data

$$
\begin{aligned}
|\Delta S_t| &\sim \text{Binomial}(S_t, \theta_t)\\
\Delta R_t &\sim \text{Binomial}(C_t, \gamma_t)\\
\Delta D_t &\sim \text{Binomial}(C_t, \delta_t)\\
\Delta R'_t &\sim \text{Binomial}(I_t, \gamma_t)\\
\Delta D'_t &\sim \text{Binomial}(I_t, \delta_t)\\
\Delta O_t &\sim \text{Binomial}(R_t, \eta_t)\\
\Delta U_t &= R_t - \Delta O_t\\
\Delta C_t &\sim \text{Binomial}(I_t, \zeta_t) - \Delta R_t - \Delta D_t\\
\Delta I_t &= |\Delta S_t| - \Delta C_t - \Delta R'_t - \Delta D'_t\\
I_t &\sim \text{Poisson}(\rho_t)
\end{aligned}
$$

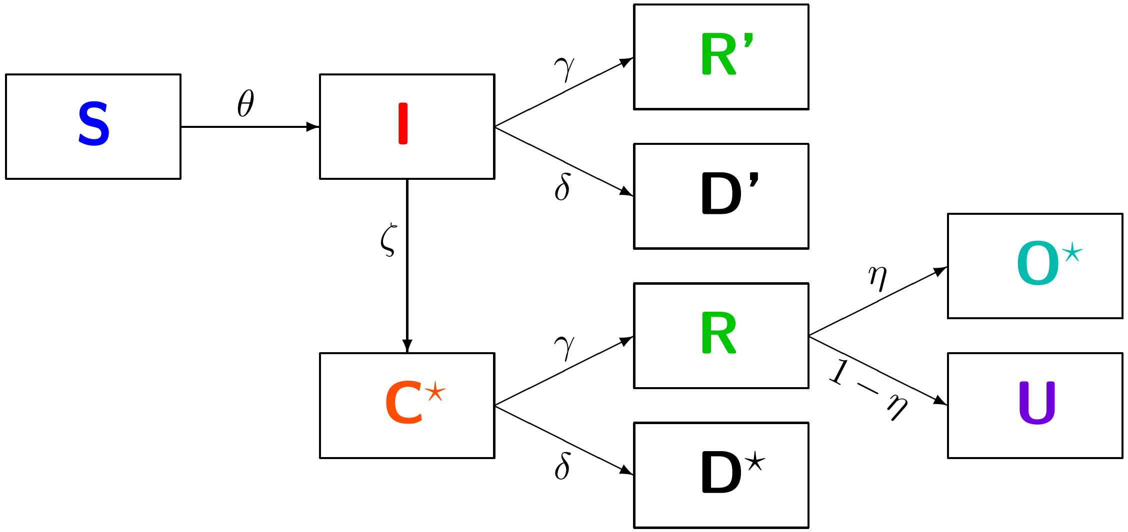


Problem 5:

Parameters are not constant.

$C_t, O_t, D_t, \Delta C_t, \Delta O_t, \Delta D_t$ are observed

Bayesian Modeling: Parameters

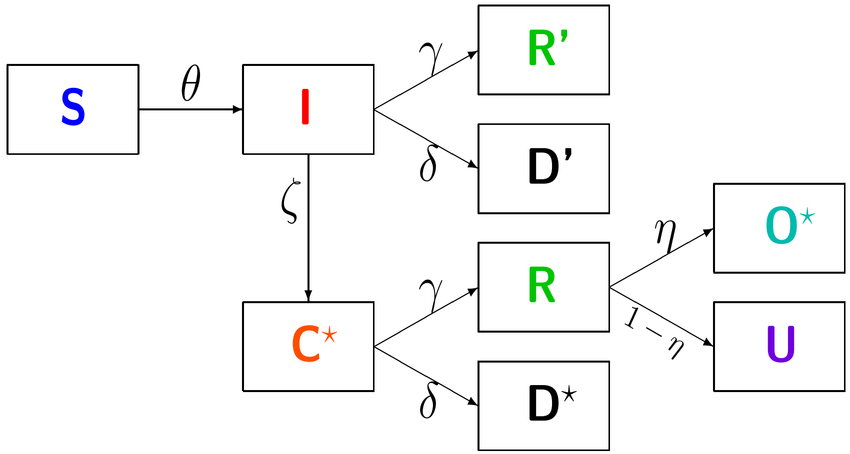


Daily rates:

$$\left.\begin{array}{lccl}
\text{Infection rate} & \theta_t & \sim & \text{Beta}(\alpha^{(\theta)}, \beta^{(\theta)}) \\
\text{Recovery rate} & \gamma_t & \sim & \text{Beta}(\alpha^{(\gamma)}, \beta^{(\gamma)}) \\
\text{Mortality rate} & \delta_t & \sim & \text{Beta}(\alpha^{(\delta)}, \beta^{(\delta)}) \\
\text{Testing rate} & \zeta_t & \sim & \text{Beta}(\alpha^{(\zeta)}, \beta^{(\zeta)}) \\
\text{Reporting rate} & \eta_t & \sim & \text{Beta}(\alpha^{(\eta)}, \beta^{(\eta)}) \\
\text{Currently infected} & \rho_t & \sim & \text{Gamma}(\alpha^{(\rho)}, \lambda^{(\rho)})
\end{array}\right\} \quad \text{Prior distributions}$$

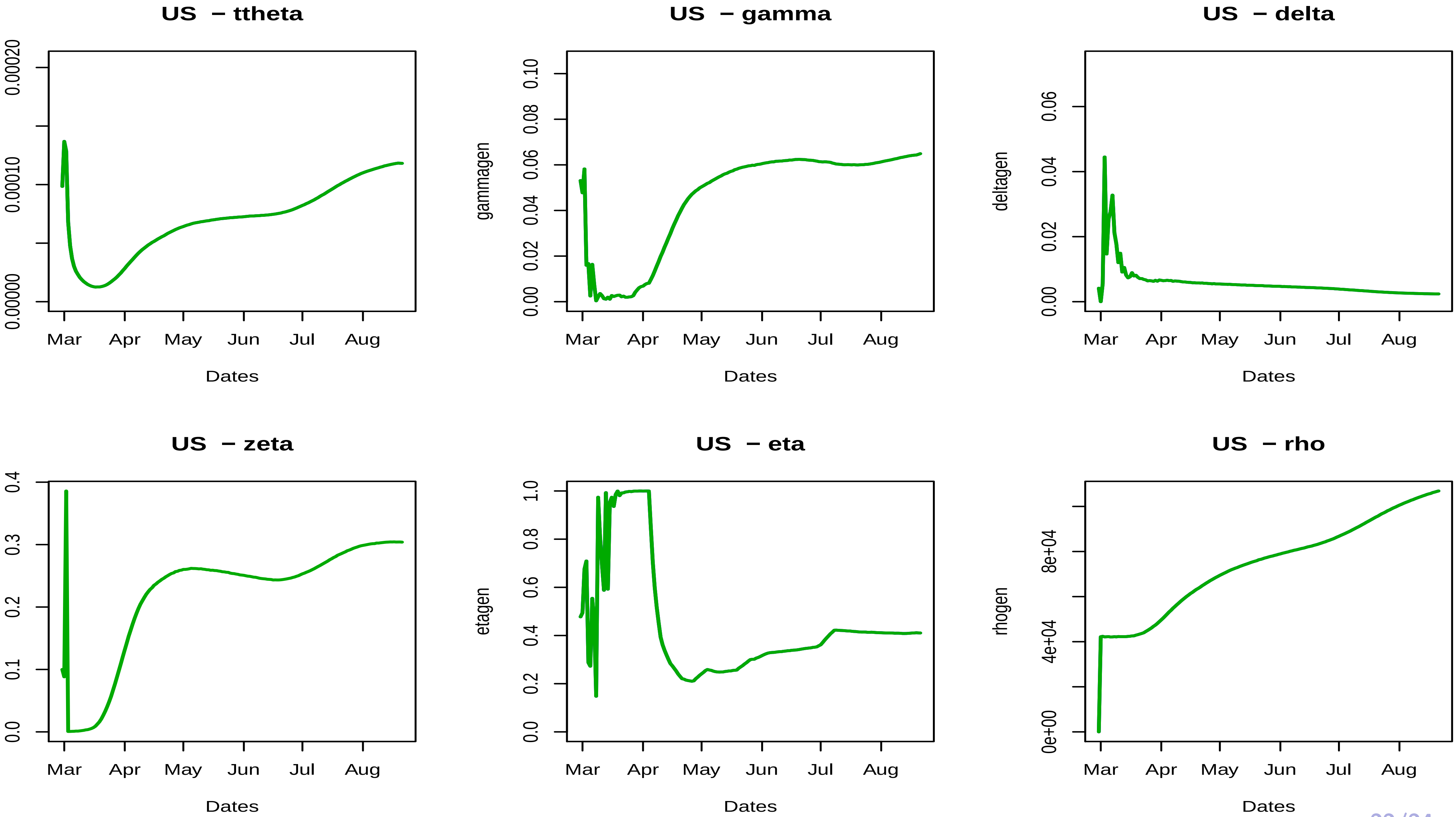
Posterior Distributions: Update of Hyper-Parameters

$$\begin{cases} \theta \,|\, I & \sim & \text{Beta}(\alpha^{(\theta)} + I,\ \beta^{(\theta)} + S - I) \\ \gamma \,|\, R & \sim & \text{Beta}(\alpha^{(\gamma)} + R,\ \beta^{(\gamma)} + C - R) \\ \delta \,|\, D & \sim & \text{Beta}(\alpha^{(\delta)} + D,\ \beta^{(\delta)} + C - D) \\ \zeta \,|\, C & \sim & \text{Beta}(\alpha^{(\zeta)} + C,\ \beta^{(\zeta)} + I - C) \\ \eta \,|\, O & \sim & \text{Beta}(\alpha^{(\eta)} + O,\ \beta^{(\eta)} + R - O) \\ \rho \,|\, I & \sim & \text{Gamma}(\alpha^{(\theta)} + I,\ \beta^{(\theta)} + 1) \\ I \,|\, C & \sim & \text{Poisson}(\rho), \text{shifted by } C \end{cases}$$



- Dynamics of the COVID-19 pandemic:
- Observed counts induce daily updates of hyper-parameters
- The Beta-Binomial model is ready for forecasting

Parameter Estimates: USA



Latest Estimates: USA (08/28/2020)

Infection rate	θ	$=$	0.00012
Transmission rate	β	$=$	0.056
Recovery rate	γ	$=$	0.066
Mortality rate	δ	$=$	$0.0.0023$
Testing rate	ζ	$=$	0.369
Currently infected, not tested	ρ	$=$	$88,944$
Basic reproduction number	R_0	$=$	0.831
Mean disease duration	$\frac{1}{\gamma+\delta}$	$=$	14.7 days
Total infected	$\sum \lvert\Delta S_t\rvert$	$=$	$6,964,213$
Confirmed	C	$=$	$5,801,712$
Total recovered	$R + R'$	$=$	$6,014,168$
Reported recoveries	O	$=$	$2,065,066$
Total casualties	$D + D'$	$=$	$258,993$
Reported casualties	D	$=$	$179,066$

Discussion

- This approach fills *some* gaps in official reports
- It brings consistency among observed counts
- But it may not recover *everything* that is hidden
- It allows parameters to change, reflecting the epidemic dynamics
- It provides natural forecasting
- It can be extended accounting for new types of data

Any questions?
Thank you!