

# Notes

Sang Woo Park

September 21, 2018

## 1 Discrete time model

We are interested in estimating transmission rate (potentially time varying) across space. First, let's begin with a simple example.

$$\begin{aligned} S_{t+1} &= S_t - \beta_t S_t I_t \\ I_{t+1} &= \beta_t S_t I_t \end{aligned} \tag{1}$$

In this particular case, we have

$$i_{t+1} = \beta S_{t+1} I_{t+1} = \beta S_{t+1} i_t. \tag{2}$$

This is essentially the framework for the TSIR model. The implicit assumption is that all infection occurs in one generation.

### 1.1 Multiple time step

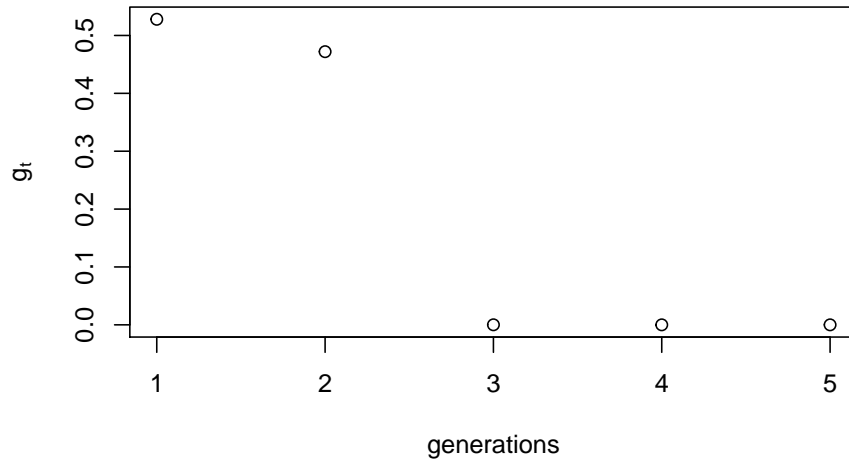
We can extend this framework such that infection lasts for more than one generation (see Fraser (2007)?).

$$S_{t+1} = S_t - S_t \sum_{k=1}^t \beta_{t,k} i_k g_{t+1-k} \tag{3}$$

Incidence is given by

$$i_{t+1} = S_t \sum_{k=1}^t \beta_{t,k} i_k g_{t+1-k}. \tag{4}$$

Measles has mean generation interval of 15 days and CV of 0.21. If we approximate that with a gamma distribution, we can approximate  $g_t$ .



However, there is probably some censoring that we have to take into account and more infection must occur than 52%. In fact, we have to calculate the probability that a person infected within the first two weeks will infected another person during the next two weeks. Given that the distribution of infection time follows some distribution  $f(t)$ , where  $0 \leq t \leq 14$  and generation time  $g(t)$ , we want to compute the following:

$$\int_0^1 4f(t) \int_{14-t}^{28-t} g(s) ds dt$$

Assume that  $f(t)$  is a uniform distribution...

```
library(pracma)
integral2(function(x, y) dunif(x, min=0, max=14) *
          dgamma(y, shape=1/gcv^2, scale=gbar*gcv^2),
          xmin=0,
          xmax=14,
          ymin=function(x) 14-x,
          ymax=function(x) 28-x)

## $Q
## [1] 0.8160168
##
## $error
## [1] 1.693841e-11
```

We obtain 81.6%.

Assume that most of infection occur within the first two generations:

$$i_{t+1} = S_t(\beta_{t,1}i_t g_1 + \beta_{t-1,2}i_{t-1}g_2). \quad (5)$$

If we know  $S_t$ , we can solve for  $\beta$  by minimizing sum of squares. In particular, we can write the above equation as matrix form:

$$\begin{bmatrix} i_3 \\ i_4 \\ \vdots \\ i_{28} \\ i_{29} \\ \vdots \end{bmatrix} = \begin{bmatrix} S_2 i_1 g_2 & S_2 i_2 g_1 & 0 & \cdots & 0 \\ 0 & S_3 i_2 g_2 & S_3 i_3 g_1 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ S_2 i_7 g_1 & 0 & \cdots & 0 & S_2 i_2 g_2 \\ S_2 i_8 g_2 & S_2 i_8 g_1 & 0 \cdots & 0 & \\ \vdots & & & & \vdots \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{25} \\ \beta_{26} \end{bmatrix} \quad (6)$$

We can probably even use tensors to make this cleaner (or not)...

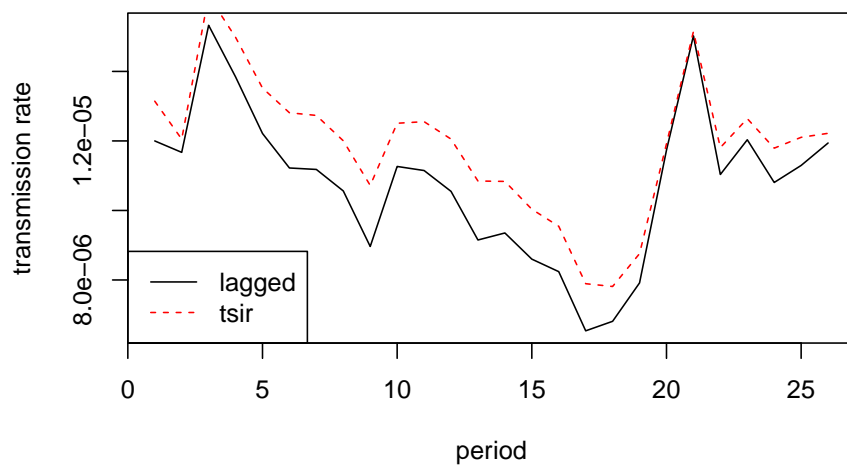
We can compare transmission rates inferred from this model vs. those inferred from the tsir model:

```
london <- twenty meas[["London"]]

g_t_adj <- c(0.816, 1-0.816)

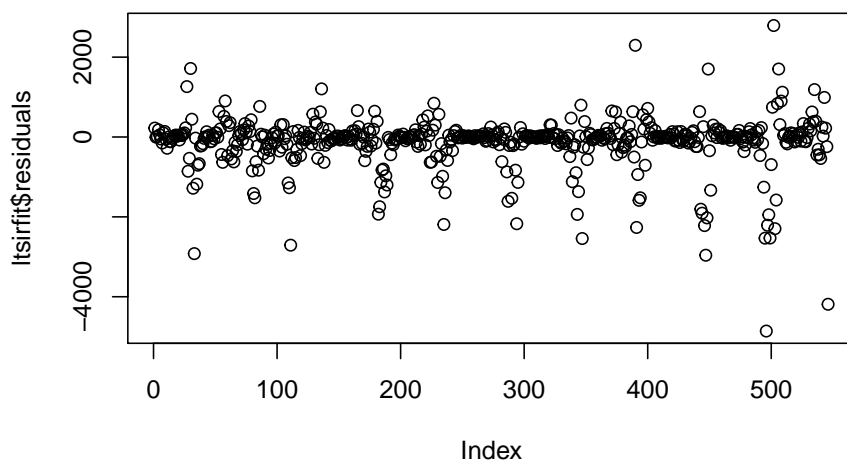
ltsirfit <- runltsir(london, g_t_adj)
tsirfit <- runtsir(data=london, method="pois", nsim=100, IP=2, inits.fit=FALSE)

##          alpha          mean beta          mean rho          mean sus
##          9.60e-01          1.19e-05          4.57e-01          1.14e+05
## prop. init. sus. prop. init. inf.
##          3.01e-02          6.12e-05
```



However, We run into residual issues.

```
plot(ltsirfit$residuals)
```



For the case of measles, changes in transmission rates are mostly driven by changes in contact rates. Hence, we can simplify the model such that

$$i_{t+1} = \beta_{t,1} S_t ((1 - \epsilon) i_t + \epsilon i_{t-1}). \quad (7)$$

Then, we can even include the heterogeneity parameter and perform a log transformation to obtain a linear problem:

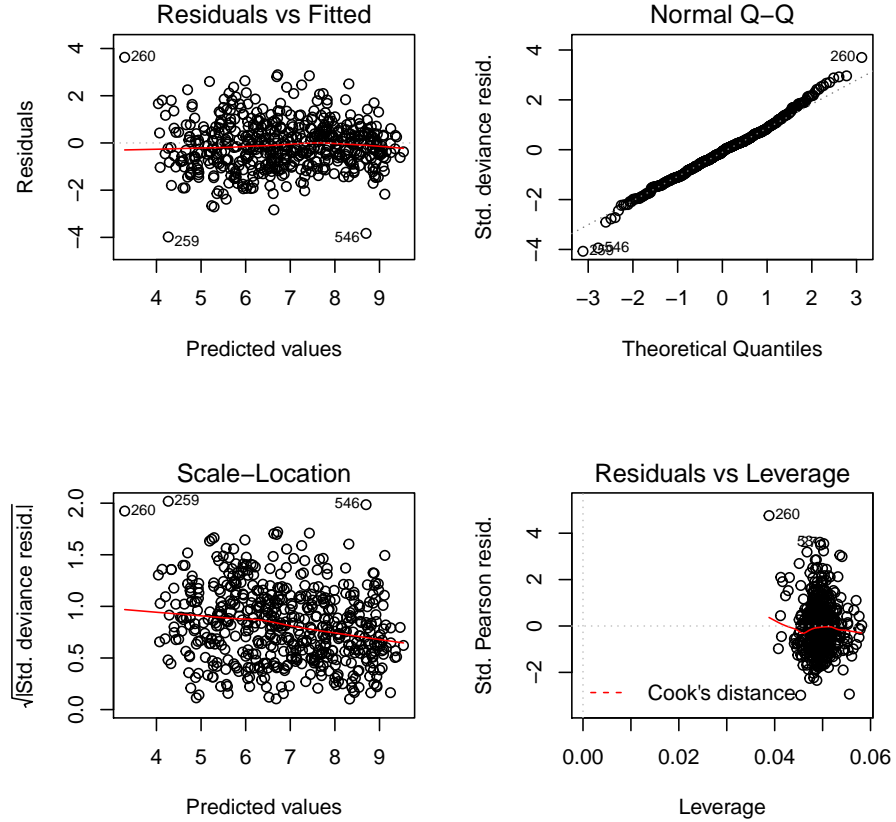
$$\log i_{t+1} = \log \beta_t + \log S_t + \alpha \log((1 - \epsilon)i_t + \epsilon i_{t-1}). \quad (8)$$

We can assume Gaussian error but we can also use negative binomial with log link:

$$i_{t+1} = \text{NegBin}(\exp(\log \beta_t + \log S_t + \alpha \log((1 - \epsilon)i_t + \epsilon i_{t-1}))) \quad (9)$$

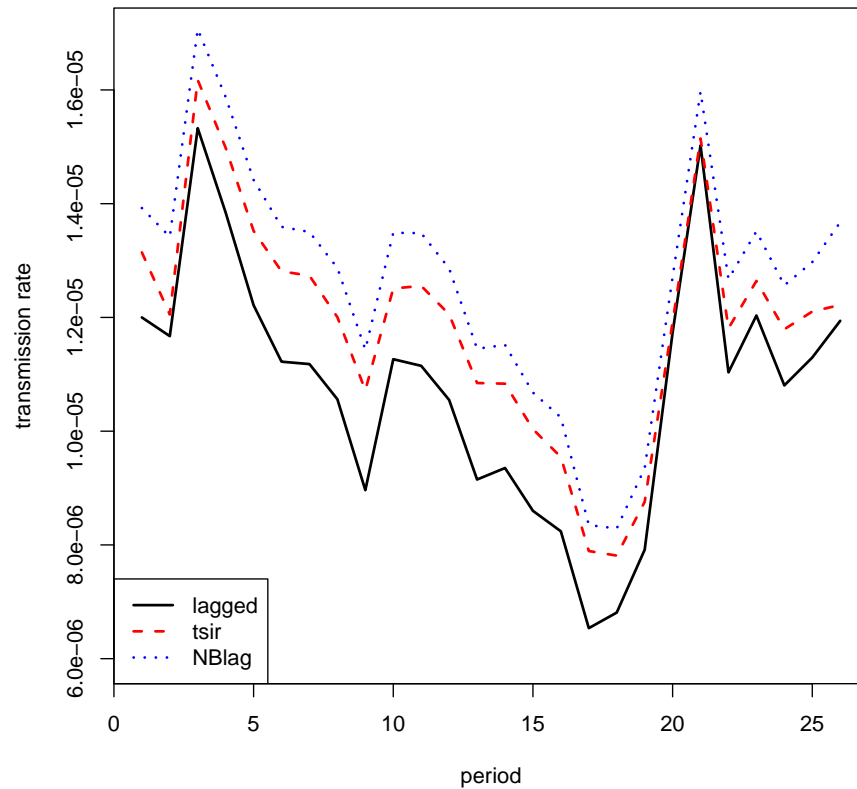
and perform a two dimensional optimization on  $\epsilon$  and  $\bar{S}$ . Then, our residuals are much better:

```
ltsirfit_NB <- runltsir_NB(london, predict=TRUE, nsim=100)
par(mfrow=c(2, 2))
plot(ltsirfit_NB$fit)
```



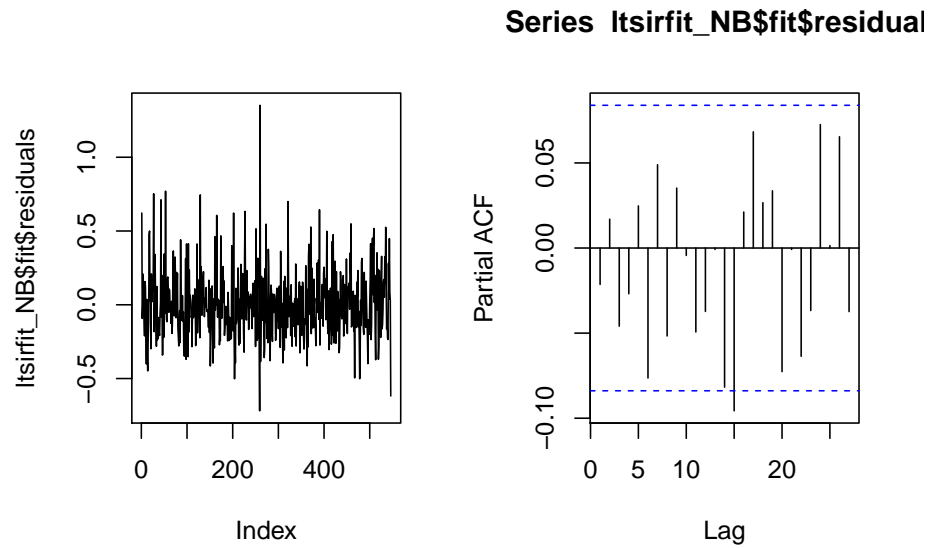
```
par(mfrow=c(1,1))
```

Looking at the estimates of the transmission rates:



However, it appears that there might be weak underlying structures in the residuals that we might have to deal with? Might be too weak to care about?

```
par(mfrow=c(1, 2))
plot(ltsirfit_NB$fit$residuals, type="l")
pacf(ltsirfit_NB$fit$residuals)
```



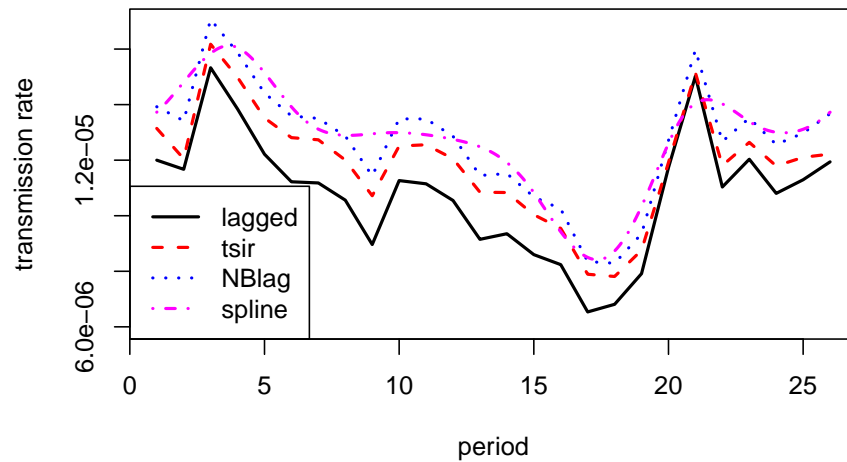
```
par(mfrow=c(1,1))
```

## 1.2 Splines

We can try to smooth transmission rate over time using cyclic splines (`s(period, bs='cc')`) with GAM.

```
ltsirfit_spline <- runltsir_NB(london, spline=TRUE, predict=TRUE, nsim=100)
```

Transmission rates?



Note that  $\epsilon$  estimated by the model is fairly small (1.4%):

```
ltsirfit_spline$epsilon
```

```
##      epsilon
## 0.01434702
```

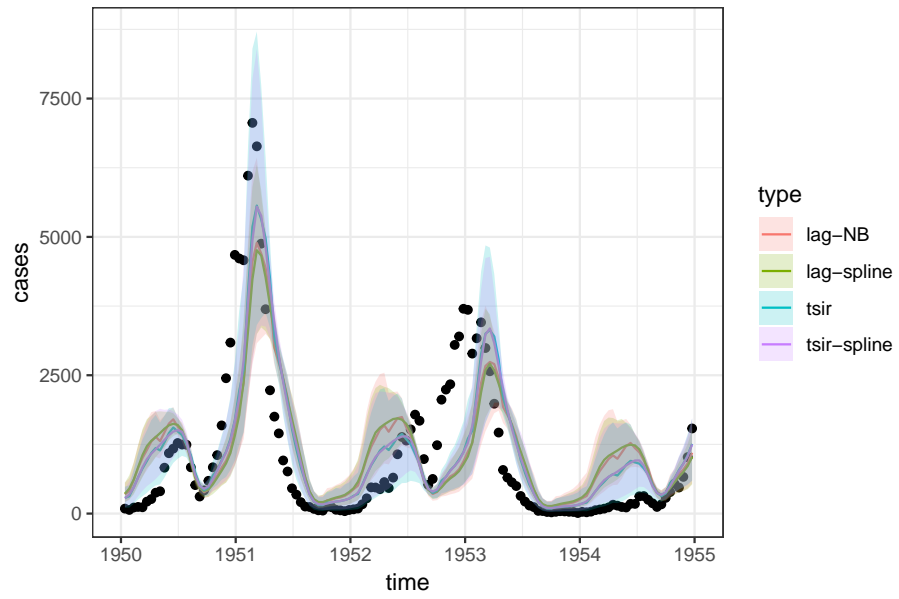
There's another spline fit that I built inside tsiR (without a lag term)...

```
tsirfit2 <- runtsir.spline(london, method="pois", nsim=100, IP=2, inits.fit=FALSE)

##          alpha      mean beta      mean rho      mean sus
##      9.60e-01      1.19e-05      4.57e-01      1.14e+05
## prop. init. sus. prop. init. inf.
##      3.01e-02      6.12e-05
```

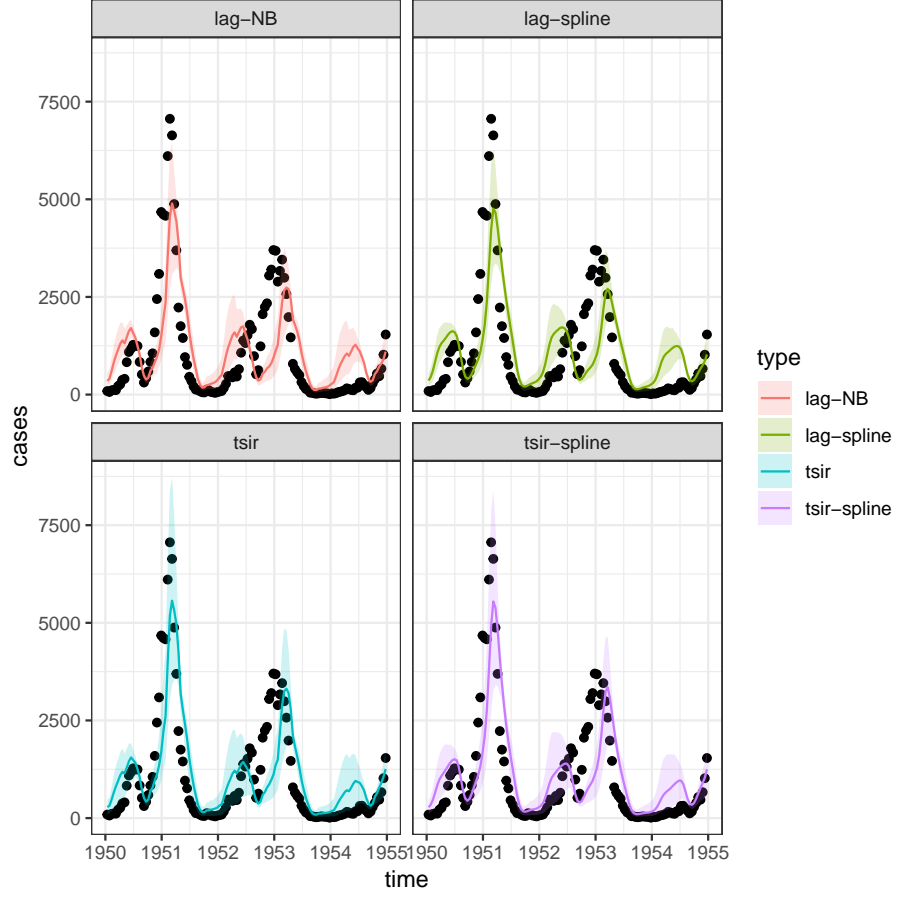
Now, we want to compare how each model predicts the time series:





facet?

```
gcomp +  
  facet_wrap(~type)
```



## 2 Spatial spread

Ultimately, we want to understand the spatial spread of an epidemic.

$$S_{t+1,k} = B_{t,k} + S_{t,k} - I_{t+1,k}$$

$$I_{t+1,k} = \beta_{t,k} \frac{S_{t,k}}{N_{t,k}} \left( \left( 1 - \epsilon \sum_{j \neq k} f(d(k,j)) \right) I_{t,k} + \epsilon \sum_{j \neq k} f(d(j,k)) I_{t,j} \right)^\alpha \quad (10)$$

For convenience, we write

$$\hat{I}_{t,k} = \left( \left( 1 - \epsilon \sum_{j \neq k} f(d(k,j)) \right) I_{t,k} + \epsilon \sum_{j \neq k} f(d(j,k)) I_{t,j} \right) \quad (11)$$

Then, we can linearize the problem:

$$I_{t+1,k} = \text{NegBin} \left( \exp \left( \log \beta_{t,k} + \log S_{t,k} + \alpha \log \hat{I}_{t,k} - \log N_{t,k} \right) \right) \quad (12)$$

## Misc

See Vehtari - Slick time series... (Andrew Gelman); see the cover of third edition of Bayesian analysis.

```
# MASS::glm.fit(family=negative.binomial(theta=10, link="identity"))
```

## References

Fraser, C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic. *PloS one* 2(8), e758.