

# Non-linear Model을 통한 신규 확진자 예측 및 정책의 영향 분석

2020년 12월 19일



서울대학교 박태성 교수

생물정보통계 연구실

박사과정 구태완 석사과정 한결희, 허규진, 이도은

학부과정 고영현, 김태현, 김학용, 정혜원



# 목차

## 1. 데이터 소개

## 2. 한국 일일 확진자 개요

## 3. 모델링

### 3.1. Polynomial Regression

### 3.2. Cubic Spline

### 3.3. Local Regression

### 3.4. GAM

### 3.5. Segmented Poisson



# 데이터 소개

## Kaggle 데이터

2020년 1월 20일 ~ 6월 30일 \*검역 데이터 부재

<https://www.kaggle.com/kimjihoo/coronavirusdataset/data?select=TimeProvince.csv>

## 공공데이터포털 데이터

2020년 3월 04일 ~ 12월 12일

<https://data.go.kr/data/15043378/openapi.do>

## COVID-19 Daily Cases

Korea, 2020-01-20~2020-11-20

**1st - 2월 17일**

31번 확진자 (신천지)

**2nd - 5월 06일**

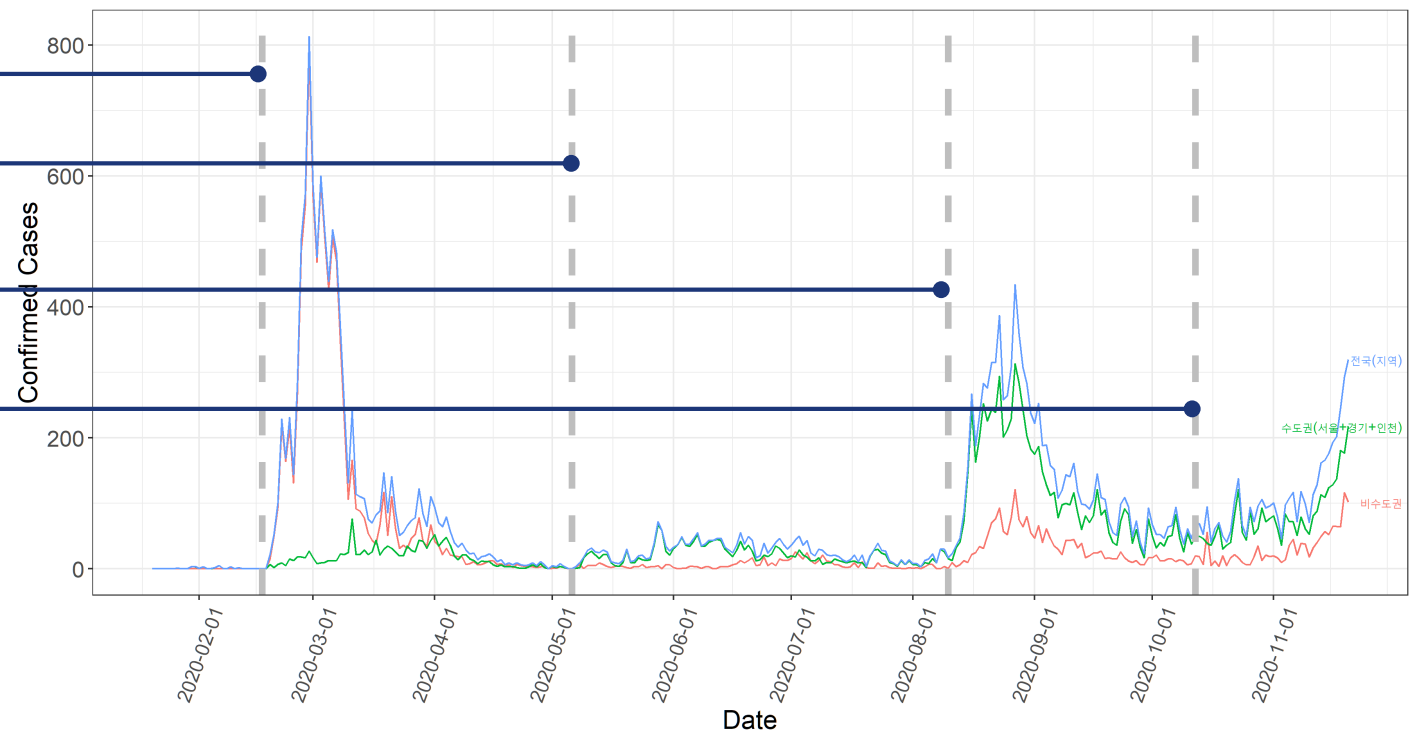
용인 66번 확진자 (이태원)

**3rd - 8월 10일**

여름휴가 성수기 이후 첫 월요일

**4th 10월 12일**

사회적 거리두기 1단계로 완화



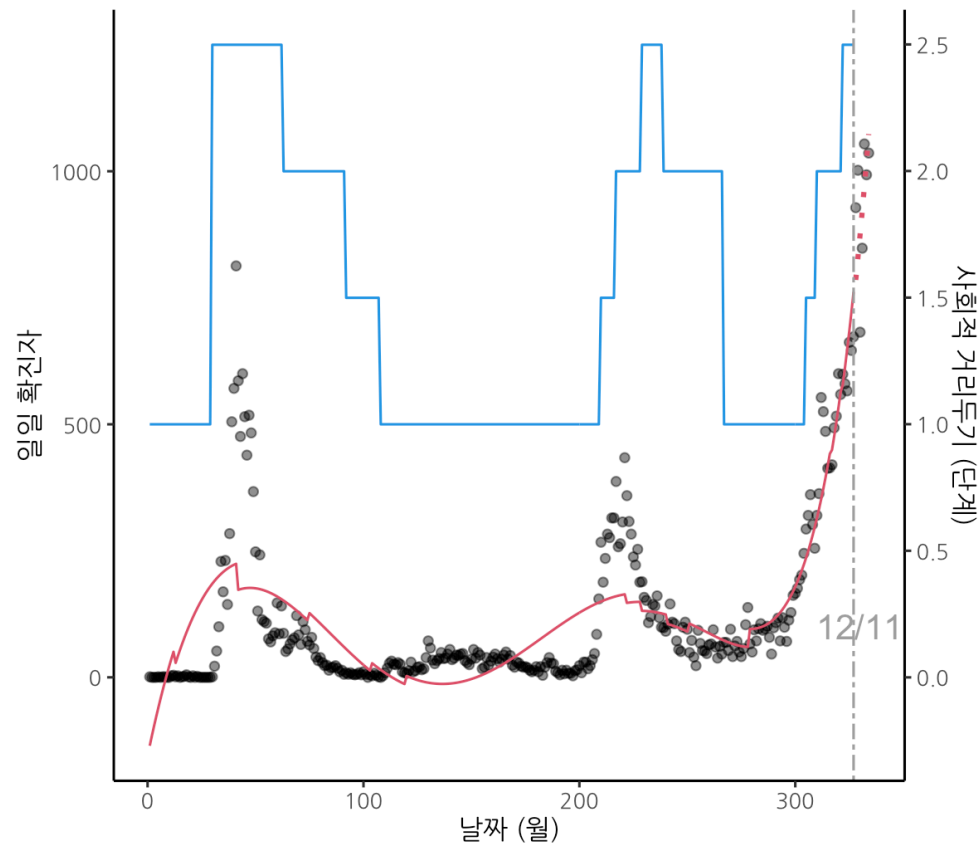
## Polynomial Regression

~ 12/11 데이터로 12/18일까지 예측

Training RMSE = 94.84

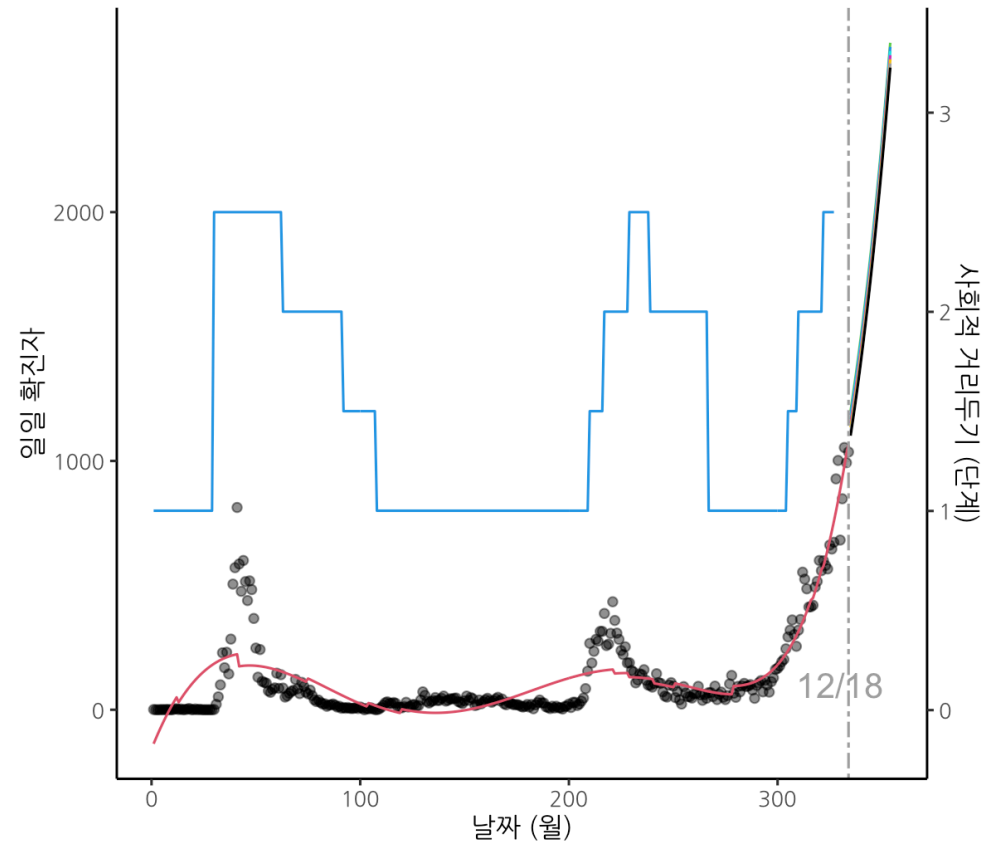
Testing RMSE = 120.56

COVID-19 전국 신규 확진자  
Polynomial Regression (Degree = 6)  
Lag for Government Index = 12



~ 12/18 데이터로 향후 20일 예측

COVID-19 전국 신규 확진자  
Polynomial Regression (Degree = 6)  
Lag for Government Index = 12



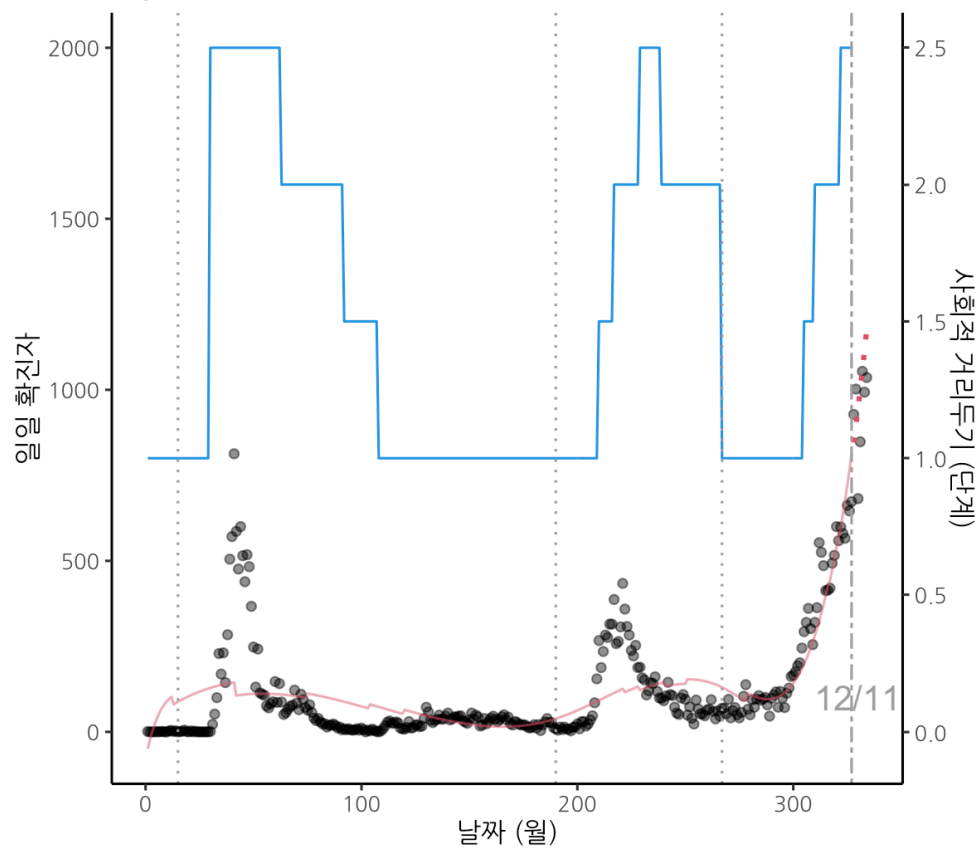
## Cubic Spline

~ 12/11 데이터로 12/18일까지 예측

Training RMSE = 106.29

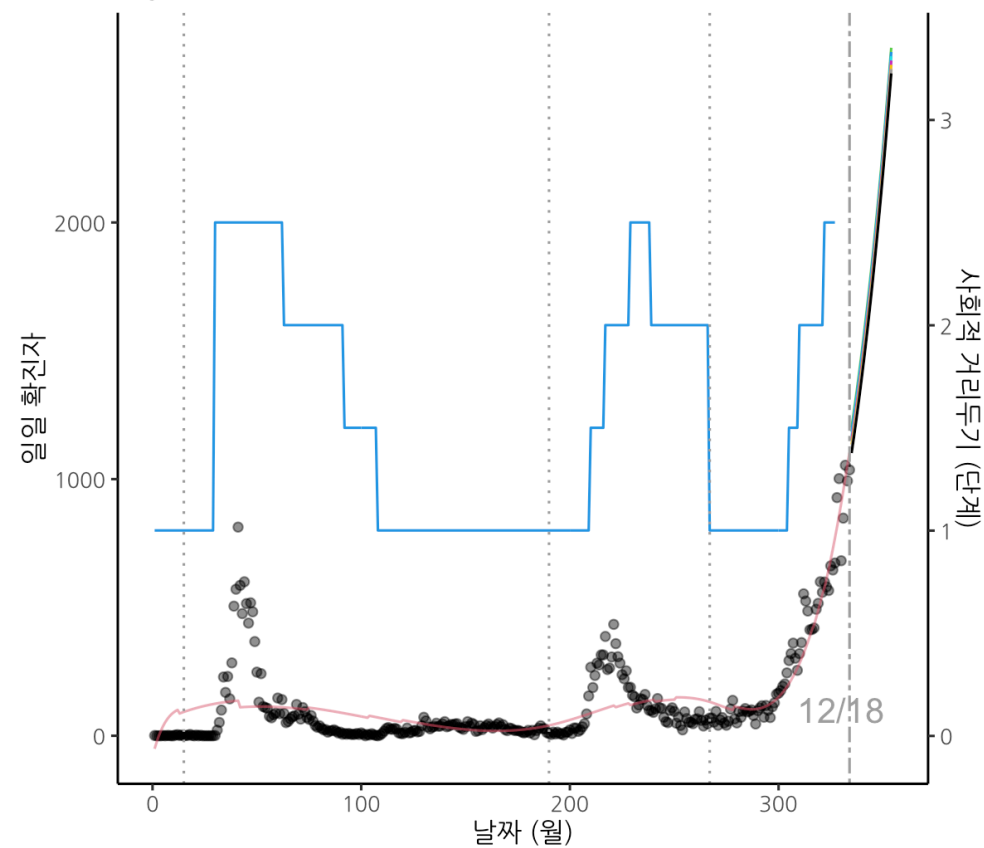
Testing RMSE = 144.84

COVID-19 전국 신규 확진자  
Cubic Spline (Knots = 15, 190, 267)  
Lag for Government Index = 12



~ 12/18 데이터로 향후 20일 예측

COVID-19 전국 신규 확진자  
Cubic Spline (Knots = 15, 190, 267)  
Lag for Government Index = 12



## Local Regression

~ 12/11 데이터로 12/18일까지 예측

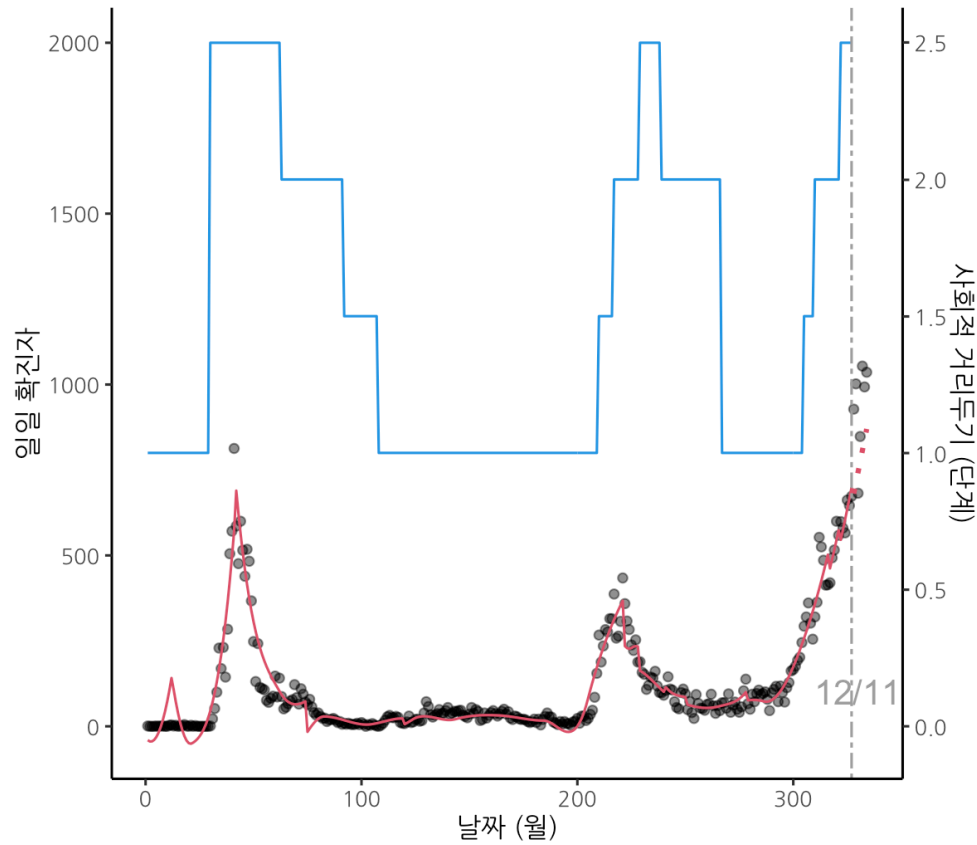
Training RMSE = 41.84

Testing RMSE = 195.16

COVID-19 전국 신규 확진자

Local Regression (alpha = 0.2))

Lag for Government Index = 12

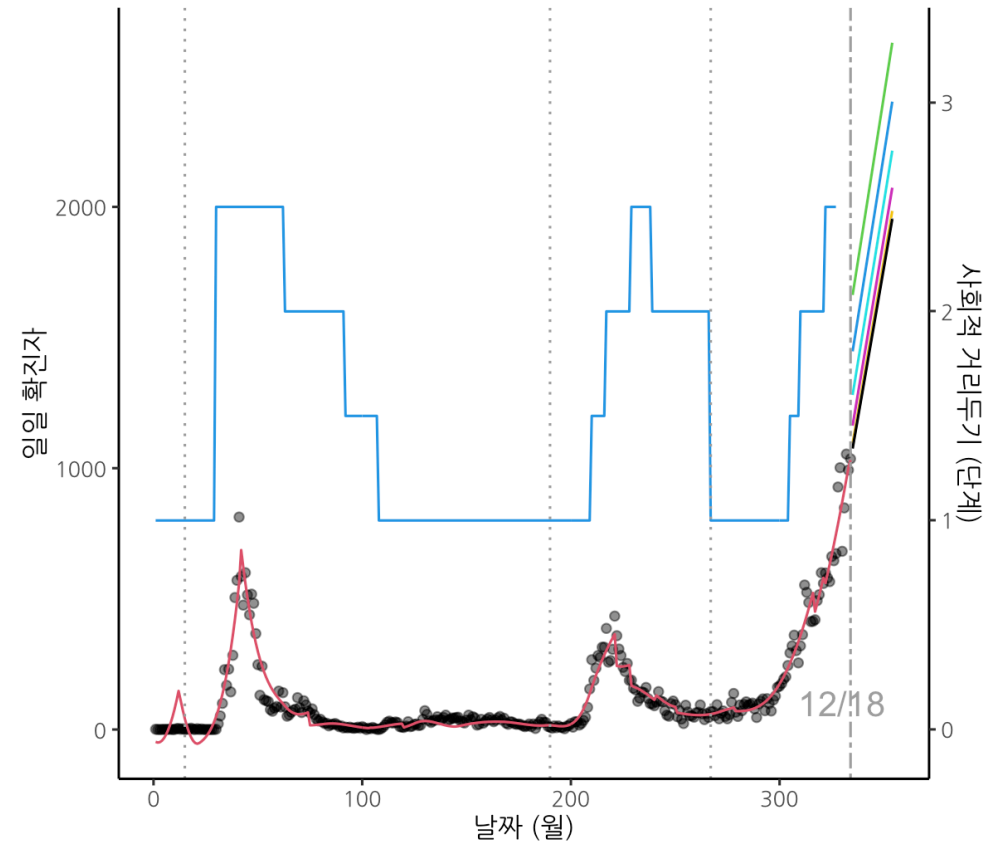


~ 12/18 데이터로 향후 20일 예측

COVID-19 전국 신규 확진자

Local Regression (alpha = 0.2))

Lag for Government Index = 12

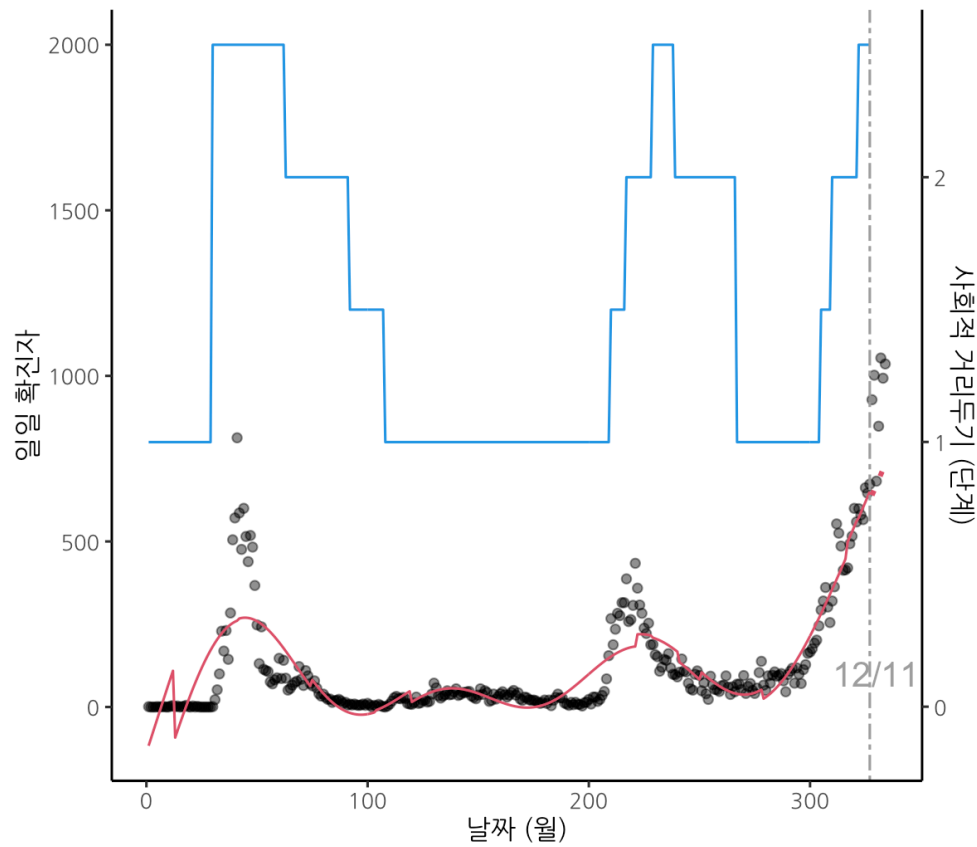


## GAM

~ 12/11 데이터로 12/18일까지 예측

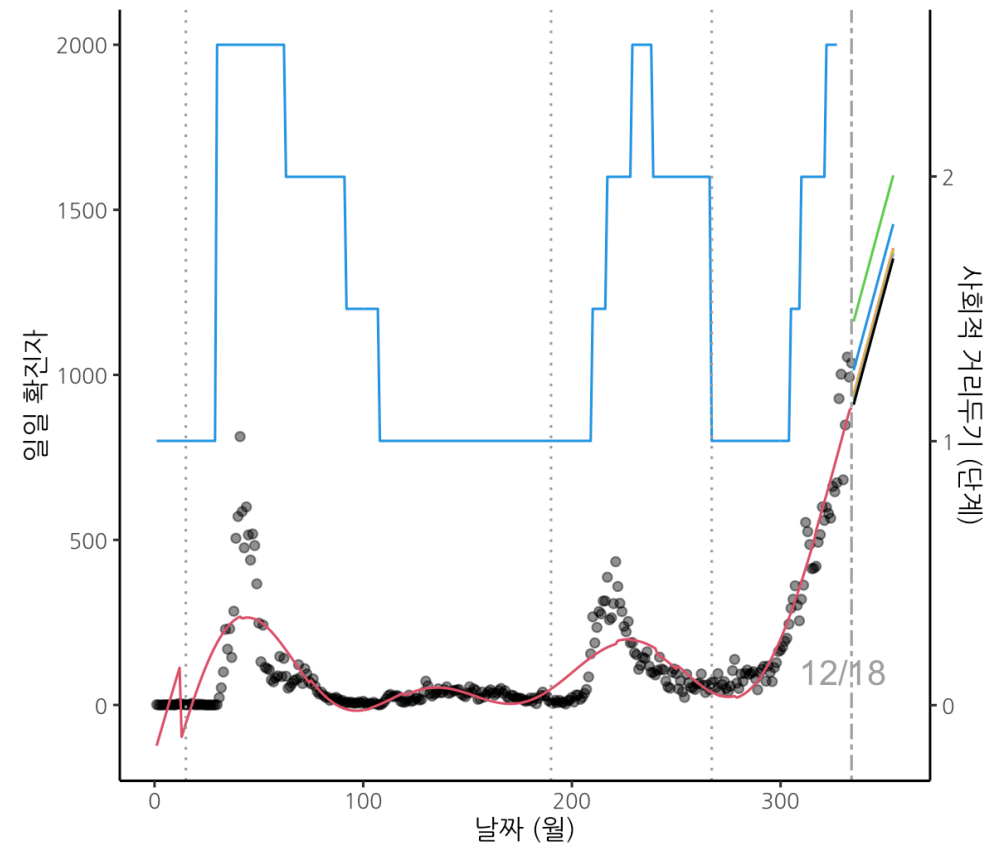
Training RMSE = 79.71  
Testing RMSE = 272.78

COVID-19 전국 신규 확진자  
GAM (method = 'RELM')  
Lag for Government Index = 12



~ 12/18 데이터로 향후 20일 예측

COVID-19 전국 신규 확진자  
GAM (method = 'RELM')  
Lag for Government Index = 12





## Segmented Poisson

~ 12/11 데이터로 12/18일까지 예측

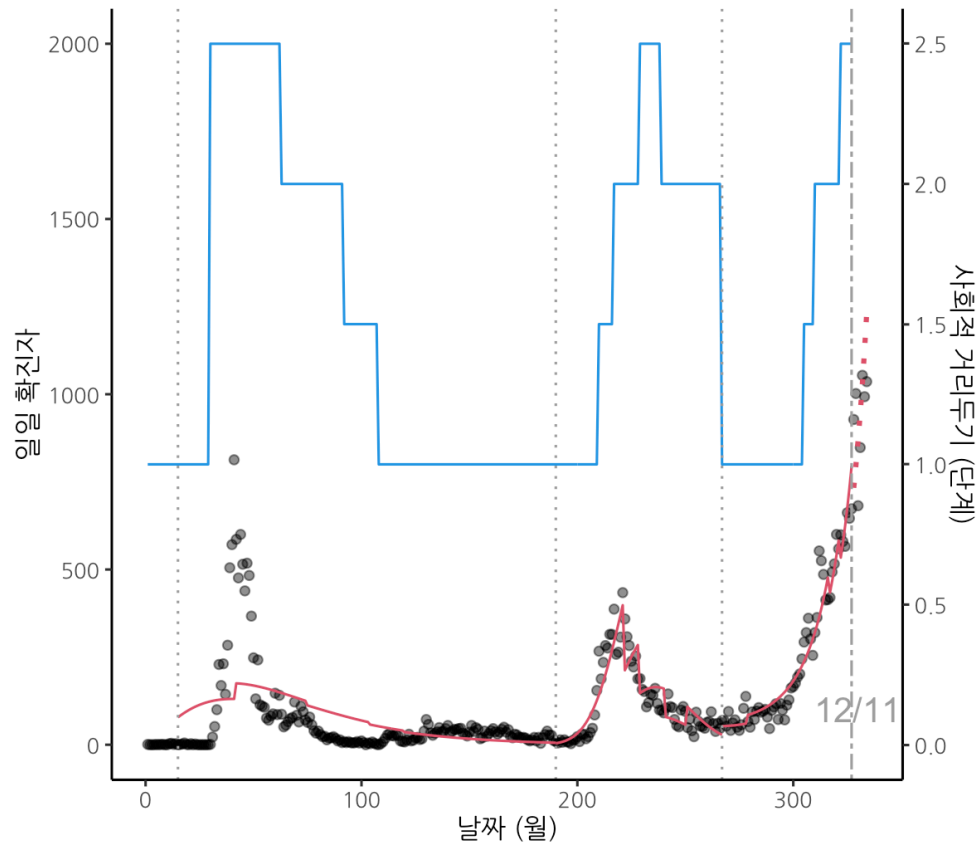
Training RMSE = 137.79

Testing RMSE = 160.42

COVID-19 전국 신규 확진자

Segmented Poisson (Breakpoint = 15, 190, 267)

Lag for Government Index = 12



~ 12/18 데이터로 향후 20일 예측

COVID-19 전국 신규 확진자

Segmented Poisson (Breakpoint = 15, 190, 267)

Lag for Government Index = 12

