Probabilistic ensemble forecasting of Australian COVID-19 cases

Rob J Hyndman

robjhyndman.com/covidtalk

MONASH University

Australian Health Protection Principal Committee

The **Australian Health Protection Principal Committee** is the key decision-making committee for national health emergencies. It comprises all state and territory Chief Health Officers and is chaired by the Australian Chief Medical Officer.

COVID-19 forecasting group

- Peter Dawson
- Nick Golding
- Rob J Hyndman
- Dennis Liu
- James M McCaw
- Jodie McVernon
- Pablo Montero-Manso
- Robert Moss
- Mitchell O'Hara-Wild
- David J Price
- Joshua V Ross
- Gerry Ryan
- Freya M Shearer
- Tobin South
- Ruarai Tobin

Data sources

- Case-level data of all positive COVID-19 tests: onset and detection times.
- Daily population mobility data from Google, Apple & Facebook
- Weekly non-household contact surveys
- Weekly behavioural surveys
- Daily case data from ?? countries via the Johns Hopkins COVID-19 repository

Case numbers

```
localcases %>% filter(state=="VIC", date >= "2020-07-01")
```

```
## # A tsibble: 353 x 3 [1D]
## # Key:        state [1]
##    date       state     n
##    <date>     <chr> <dbl>
##  1 2020-07-01 VIC     116
##  2 2020-07-02 VIC     113
##  3 2020-07-03 VIC     161
##  4 2020-07-04 VIC     161
##  5 2020-07-05 VIC     156
##  6 2020-07-06 VIC     237
##  7 2020-07-07 VIC     235
##  8 2020-07-08 VIC     221
##  9 2020-07-09 VIC     272
```
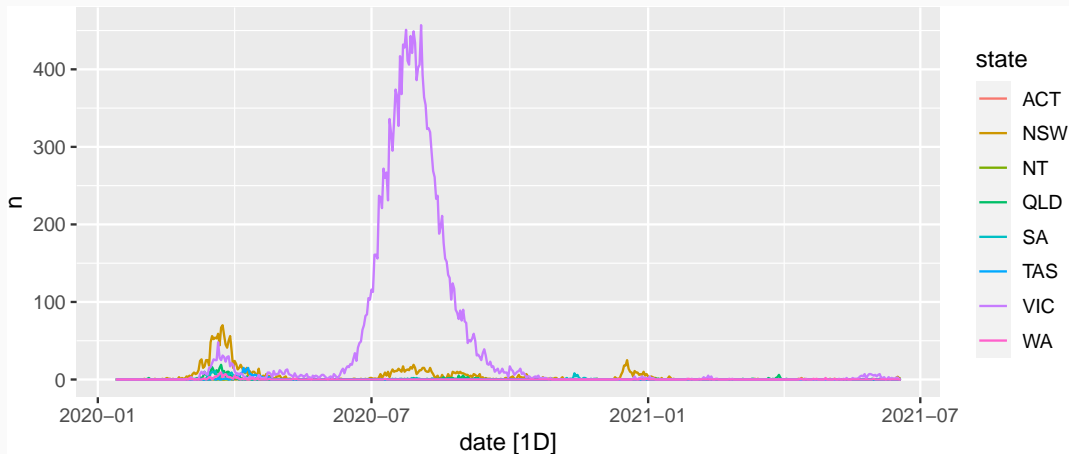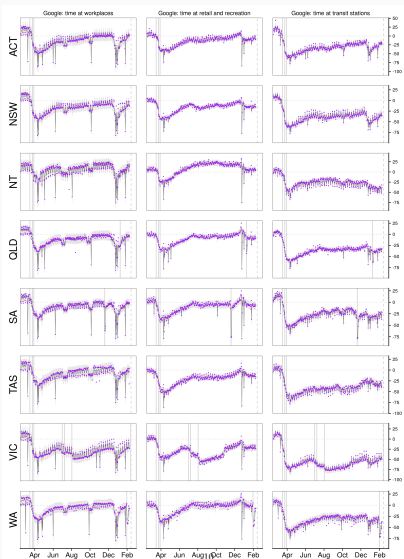
Case numbers
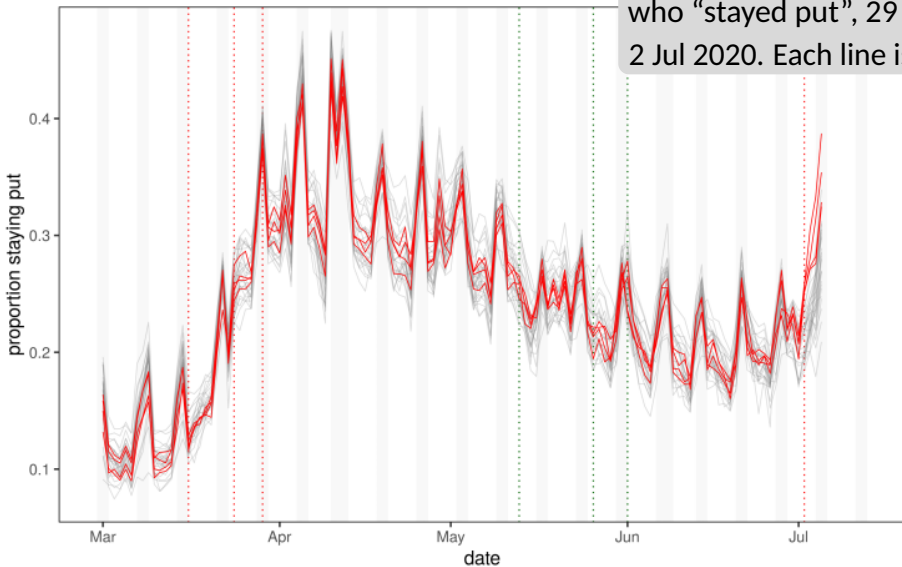
```
localcases %>% autoplot(n)
```

Google mobility data

Percentage change compared to pre-COVID-19 baseline for:

(a) time at workplace;
(b) time at retail/recreation;
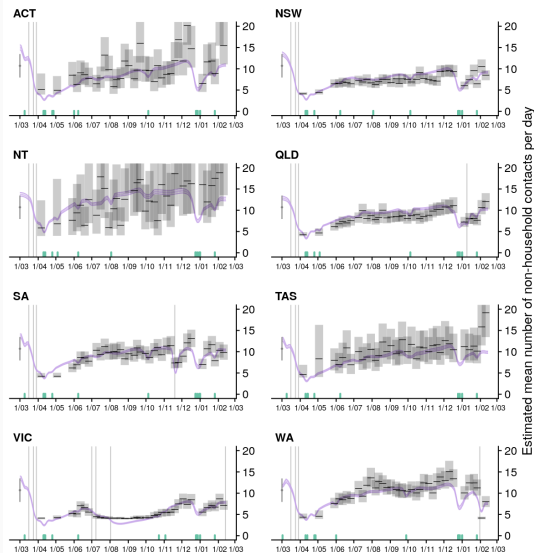(c) time at transit stations.

Vertical lines: physical distancing measures implemented.

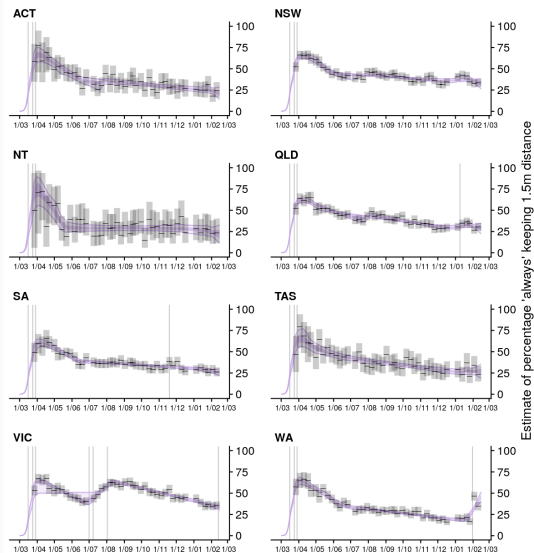# Facebook mobility data

Proportion of Facebook users who "stayed put", 29 Feb – 2 Jul 2020. Each line is one LGA.

Macrodistancing

Estimated # non-household contacts per day based on nationwide weekly surveys (gray) and Google mobility data. Green: public holidays.

Microdistancing

Estimated % keeping 1.5m distance from non-household contacts based on nationwide weekly surveys (gray).

Global daily case data from Johns Hopkins

https://github.com/CSSEGISandData/COVID-19

Model 1: SEEIIR (Uni Melbourne/Doherty Institute)

- Stochastic susceptible-exposed-infectious-recovered compartmental model that incorporates changes in local transmission potential via a time-varying effective reproduction number.
- Uses mobility and survey data and case onset and detection times.
- Daily counts $\sim$ Negative Binomial.
- Time in class $\sim$ Gamma.
- Forecasts obtained using a bootstrap particle filter.

Model 2: Generative model (Uni Adelaide)

- Uses mobility and survey data and case onset and detection times.
- Three types of infectious individuals: imported, asymptomatic, symptomatic
- Class counts $\sim$ Negative Binomial.
- Incubation times $\sim$ Gamma.
- Estimation via Hamilton Monte Carlo
- Forecasts obtained via simulation

Model 3: Global AR model (Monash)

- Uses Johns Hopkins data from countries and regions with sufficient data.
- Series with obvious anomalies (negative cases and large step changes) removed.
- $n_{t,i}$ = daily cases on day t in country/region i (scaled so all data have same mean and variance).
- $y_{t,i} = \phi_1 y_{t-1,i} + \cdots + \phi_p y_{t-p,i} + \varepsilon_{t,i}$
  where $y_{t,i} = \log(n_{t,i} + 0.5)$ and $\varepsilon_{t,i} \sim N(0, \sigma_i^2)$.
- No stationarity constraints. Common coefficients.
- Current model has $p = 24$ (selected to minimize the 7-day-ahead MAE on recent Australian data).

Forecasting ensemble

- Forecasts obtained from a mixture distribution of the component models.

$$\tilde{p}(y_{t+h}|I_t) = \sum_{k=1}^{3} w_{t+h|t,k} p_k(y_{t+h}|I_t)$$

where $p_k(y_{t+h}|I_t)$ is the forecast distribution from model k, I_t denotes the data available at time t and the weights $w_{t+h|t,k} > 0$ sum to one.

- Also known as "linear pooling"
- Works best when individual models are over-confident and use different data sources.

14

Evaluating probabilistic forecasts

$f_{p,t}$ = quantile forecast with prob. p at time t.

y_t = observation at time t

Evaluating probabilistic forecasts

$f_{p,t}$ = quantile forecast with prob. p at time t.

y_t = observation at time t

Quantile score

$$Q_{p,t} = \begin{cases} 2(1-p)|y_t - f_{p,t}|, & \text{if } y_t < f_{p,t} \\ 2p|y_t - f_{p,t}|, & \text{if } y_t \geq f_{p,t} \end{cases}$$

Evaluating probabilistic forecasts

$f_{p,t}$ = quantile forecast with prob. p at time t.

y_t = observation at time t

Quantile score

$$Q_{p,t} = \begin{cases} 2(1-p)|y_t - f_{p,t}|, & \text{if } y_t < f_{p,t} \\ 2p|y_t - f_{p,t}|, & \text{if } y_t \geq f_{p,t} \end{cases}$$

Evaluating probabilistic forecasts

$f_{p,t}$ = quantile forecast with prob. p at time t.

y_t = observation at time t

Quantile score

$$Q_{p,t} = \begin{cases} 2(1-p)|y_t - f_{p,t}|, & \text{if } y_t < f_{p,t} \\ 2p|y_t - f_{p,t}|, & \text{if } y_t \geq f_{p,t} \end{cases}$$

- Low $Q_{p,t}$ is good
- Multiplier of 2 often omitted, but useful for interpretation
- $Q_{p,t}$ like absolute error (weighted to account for likely exceedance)
- Average $Q_{p,t}$ over p = CRPS (Continuous Rank Probability Score)

More information

🏠 robjhyndman.com

🐦 @robjhyndman

🐙 @robjhyndman

✉ rob.hyndman@monash.edu