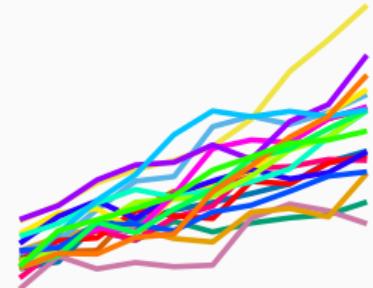


# Uncertain futures: what can we forecast and when should we give up?

Rob J Hyndman



[robjhyndman.com/uncertain\\_futures](http://robjhyndman.com/uncertain_futures)



MONASH University

# Outline

- 1 What can we forecast?
- 2 Forecastability factors
- 3 Forecasting is difficult
- 4 The statistical forecasting perspective
- 5 PBS forecasting
- 6 COVID19 case forecasting
- 7 Are we uncertain enough?

# Outline

- 1 What can we forecast?
- 2 Forecastability factors
- 3 Forecasting is difficult
- 4 The statistical forecasting perspective
- 5 PBS forecasting
- 6 COVID19 case forecasting
- 7 Are we uncertain enough?

# What can we forecast?



# What can we forecast?



# What can we forecast?



# What can we forecast?



# What can we forecast?



# What can we forecast?



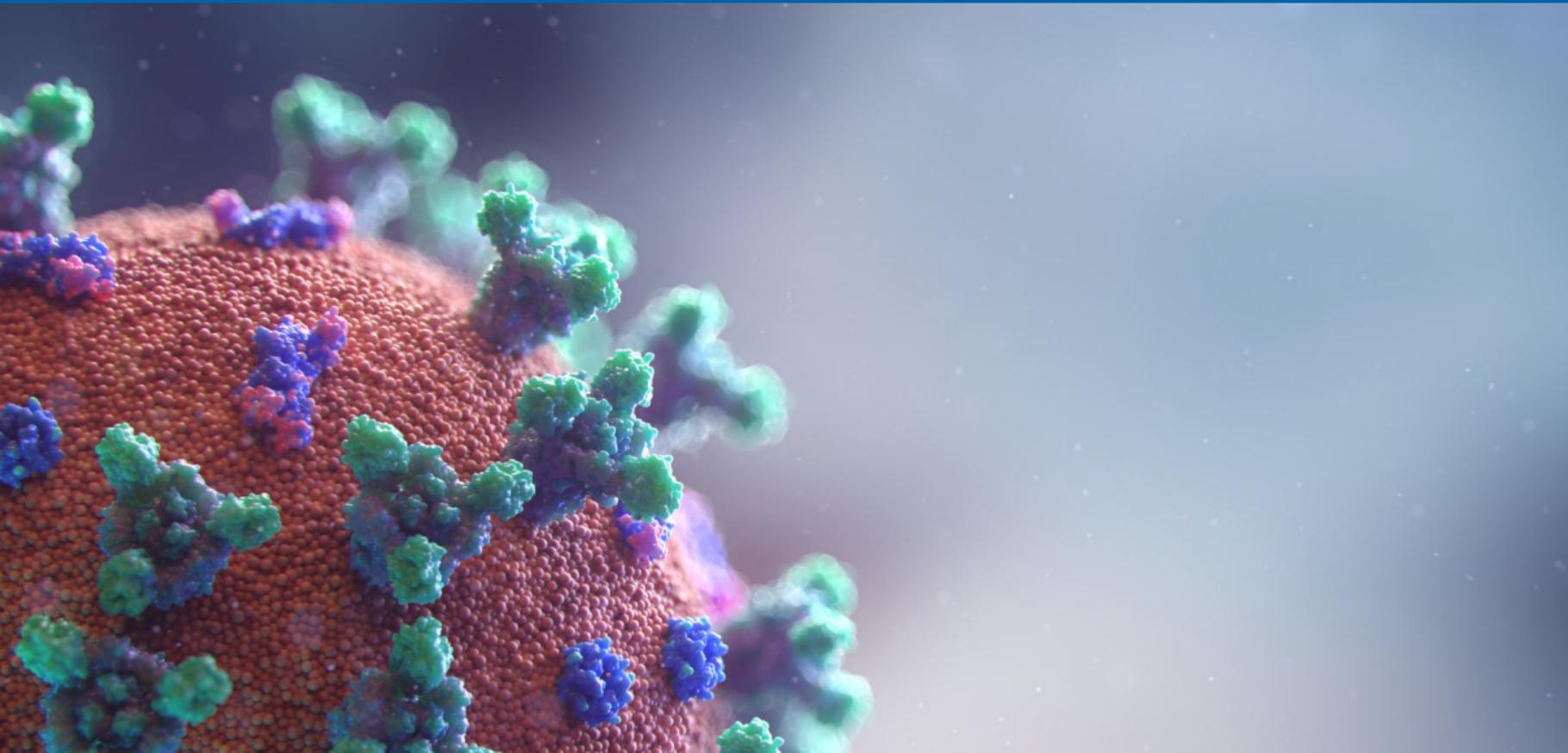
# What can we forecast?



# What can we forecast?



# What can we forecast?



# Outline

- 1 What can we forecast?
- 2 Forecastability factors
- 3 Forecasting is difficult
- 4 The statistical forecasting perspective
- 5 PBS forecasting
- 6 COVID19 case forecasting
- 7 Are we uncertain enough?

# Which is easiest to forecast?

- daily electricity demand in 3 days time
- timing of next Halley's comet appearance
- time of sunrise this day next year
- Google stock price tomorrow
- Google stock price in 6 months time
- maximum temperature tomorrow
- exchange rate of \$US/AUS next week
- total sales of drugs in Australian pharmacies next month

# Which is easiest to forecast?

- 1 time of sunrise this day next year
- 2 timing of next Halley's comet appearance
- 3 maximum temperature tomorrow
- 4 daily electricity demand in 3 days time
- 5 total sales of drugs in Australian pharmacies next month
- 6 Google stock price tomorrow
- 7 exchange rate of \$US/AUS next week
- 8 Google stock price in 6 months time

# Which is easiest to forecast?

- 1 time of sunrise this day next year
  - 2 timing of next Halley's comet appearance
  - 3 maximum temperature tomorrow
  - 4 daily electricity demand in 3 days time
  - 5 total sales of drugs in Australian pharmacies next month
  - 6 Google stock price tomorrow
  - 7 exchange rate of \$US/AUS next week
  - 8 Google stock price in 6 months time
- how do we measure “easiest”?
  - what makes something easy/difficult to forecast?

# Forecastability factors

Something is easier to forecast if:

- 1 we have a good understanding of the factors that contribute to it
- 2 there is lots of data available;
- 3 the future is somewhat similar to the past
- 4 the forecasts cannot affect the thing we are trying to forecast.

# Outline

- 1 What can we forecast?
- 2 Forecastability factors
- 3 Forecasting is difficult
- 4 The statistical forecasting perspective
- 5 PBS forecasting
- 6 COVID19 case forecasting
- 7 Are we uncertain enough?

# Reputations can be made and lost

“I think there is a world market for maybe five computers.”

(Chairman of IBM, 1943)

# Reputations can be made and lost

“I think there is a world market for maybe five computers.”

(Chairman of IBM, 1943)

“There is no reason anyone would want a computer in their home.”

(President, DEC, 1977)

# Reputations can be made and lost

“I think there is a world market for maybe five computers.”

(Chairman of IBM, 1943)

“There is no reason anyone would want a computer in their home.”

(President, DEC, 1977)

“We’re going to be opening relatively soon ... The virus ... will go away in April.”

(Donald Trump, February 2020)

# Reputations can be made and lost

“I think there is a world market for maybe five computers.”

(Chairman of IBM, 1943)

“There is no reason anyone would want a computer in their home.”

(President, DEC, 1977)

“We’re going to be opening relatively soon ... The virus ... will go away in April.”

(Donald Trump, February 2020)

“We expect that Australians will be fully vaccinated by the end of October.”

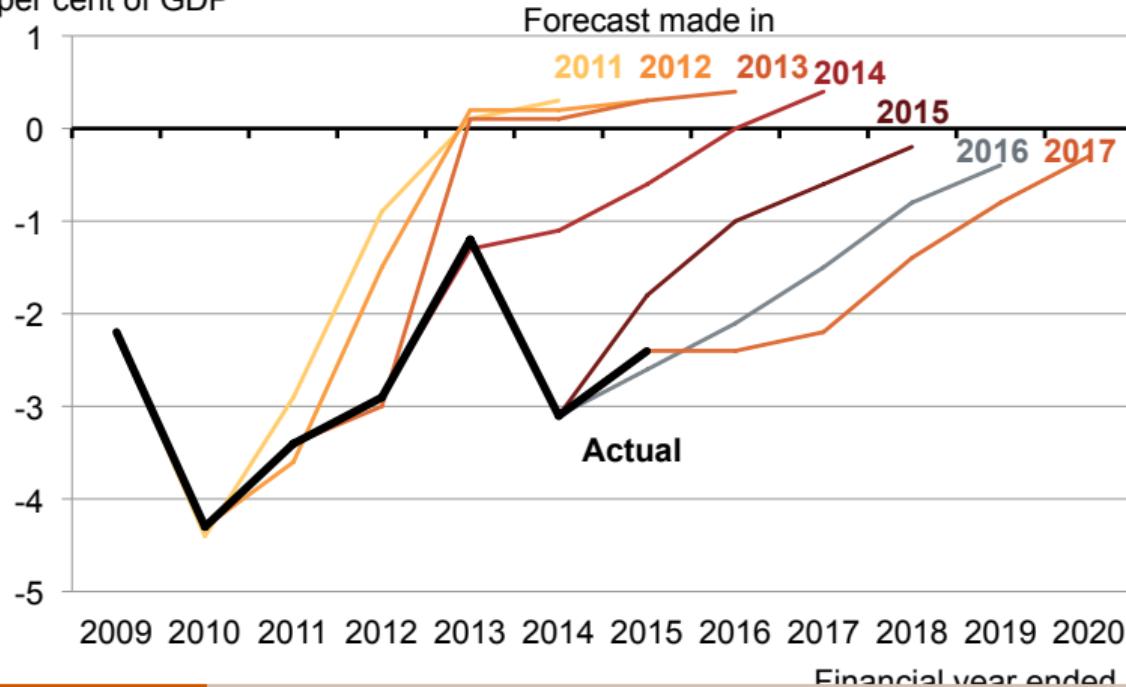
(Scott Morrison, December 2020)

# Forecasting is difficult

Commonwealth plans to drift back to surplus  
show the triumph of hope over experience

GRATTAN  
Institute

Actual and forecast Commonwealth underlying cash balance  
per cent of GDP

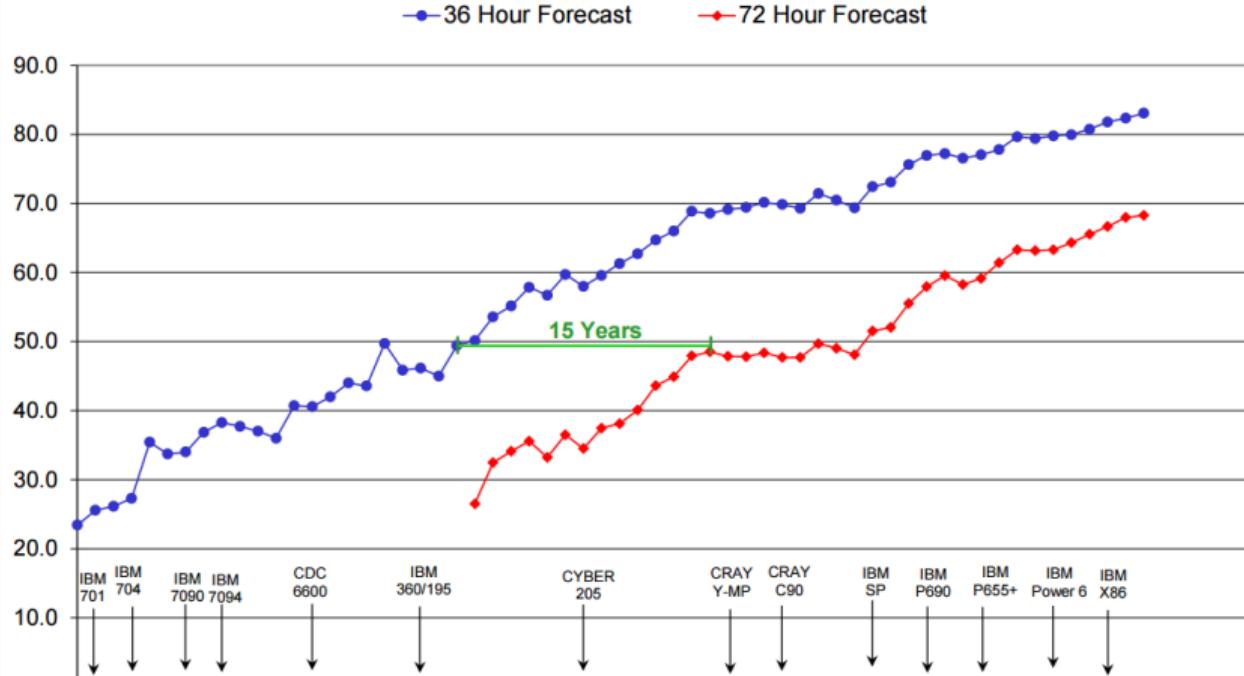


# Improving forecasts



## NCEP Operational Forecast Skill

36 and 72 Hour Forecasts @ 500 MB over North America  
[ $100 * (1 - S1/70)$  Method]



# Outline

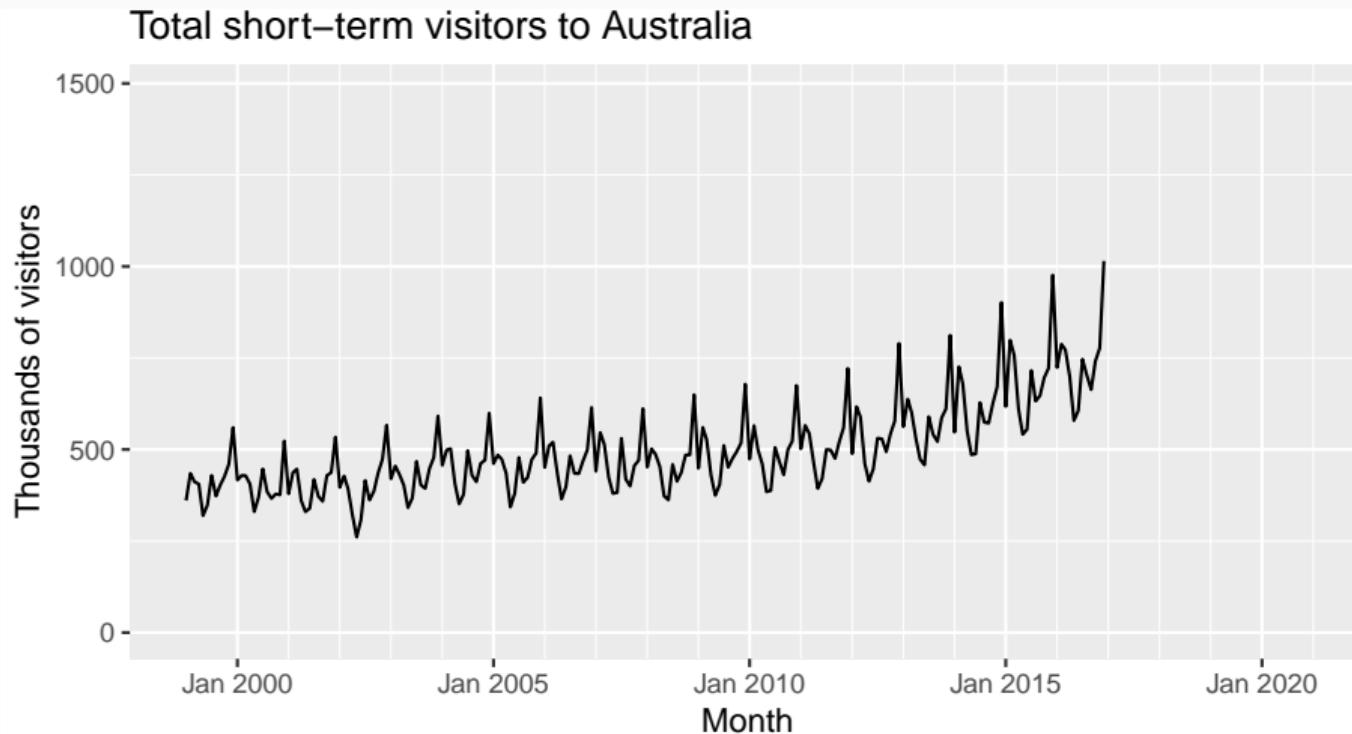
- 1 What can we forecast?
- 2 Forecastability factors
- 3 Forecasting is difficult
- 4 The statistical forecasting perspective
- 5 PBS forecasting
- 6 COVID19 case forecasting
- 7 Are we uncertain enough?

# Random futures

A forecast is an estimate of the probability distribution of a variable to be observed in the future.

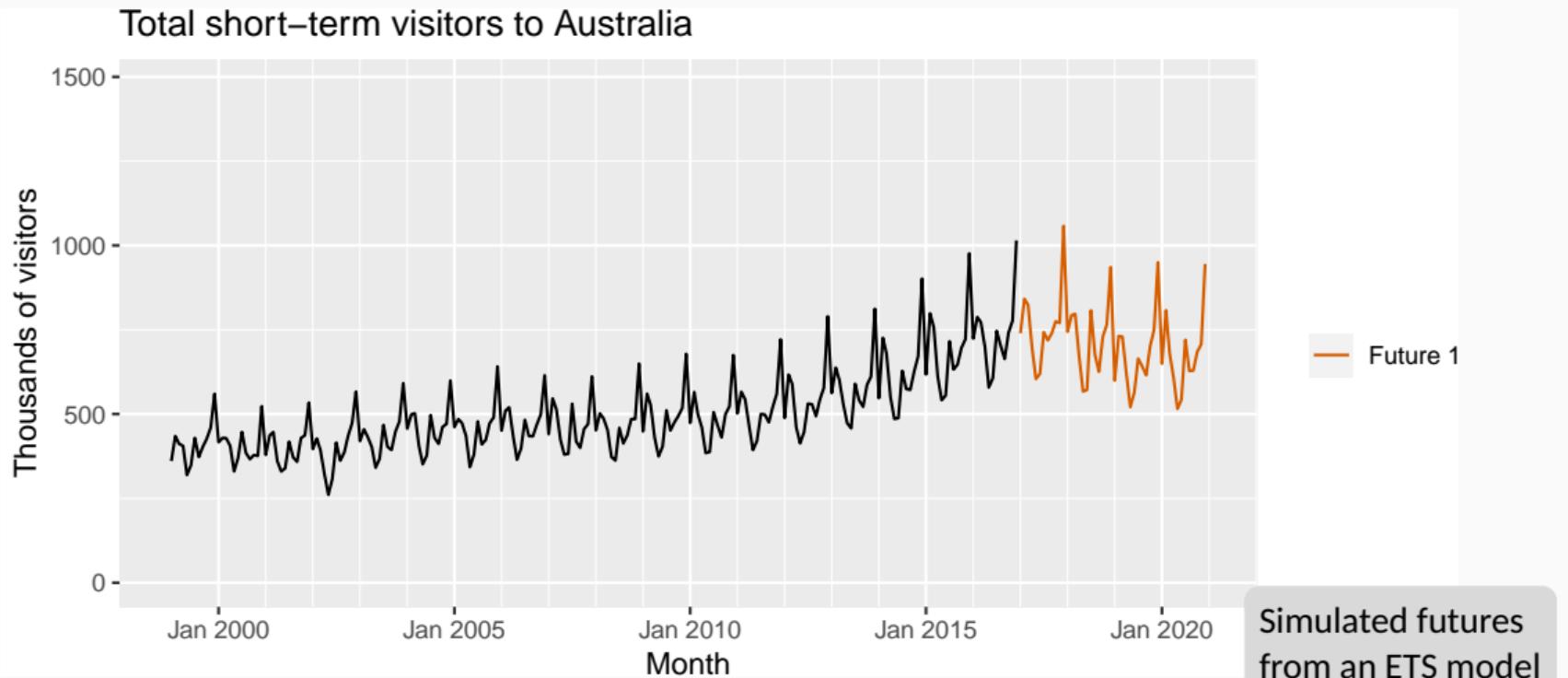
# Random futures

A forecast is an estimate of the probability distribution of a variable to be observed in the future.



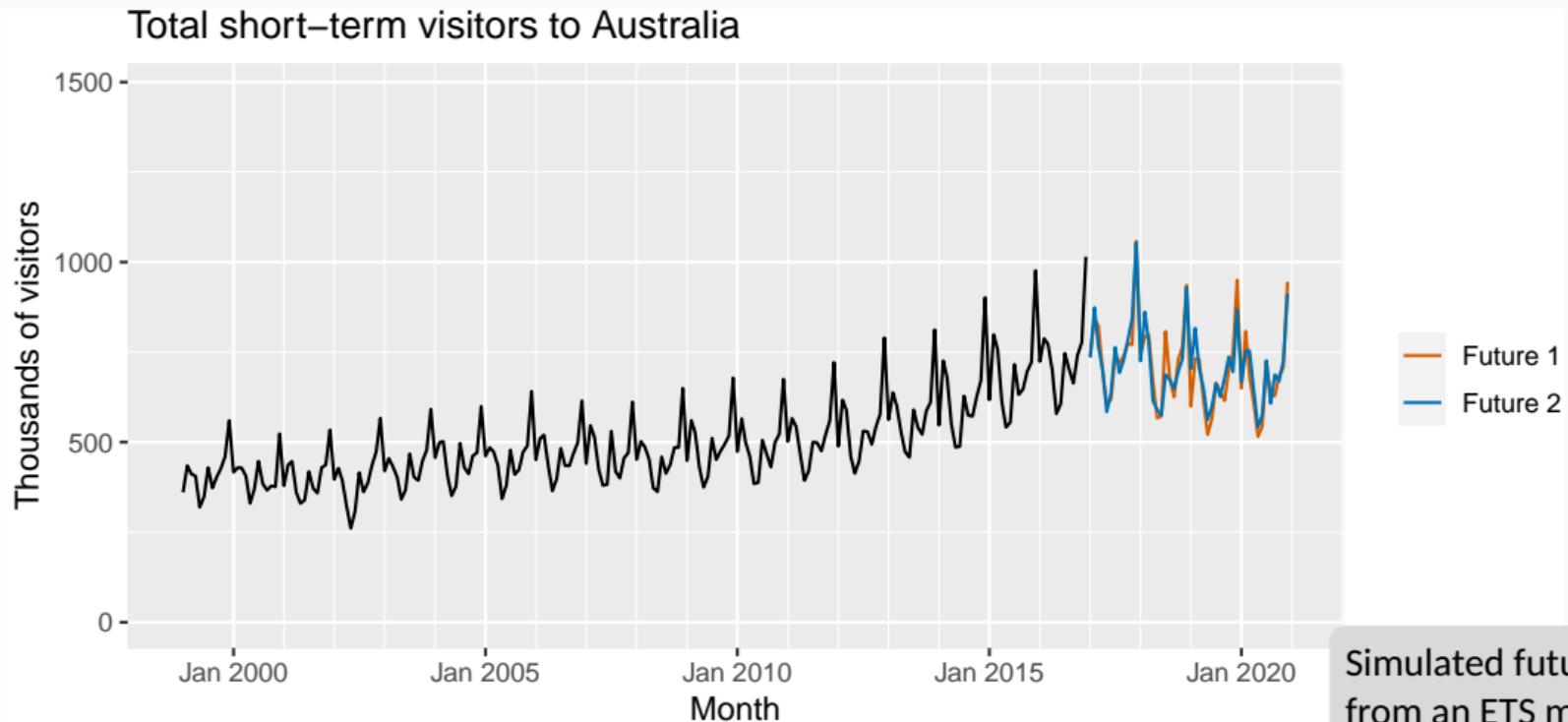
# Random futures

A forecast is an estimate of the probability distribution of a variable to be observed in the future.



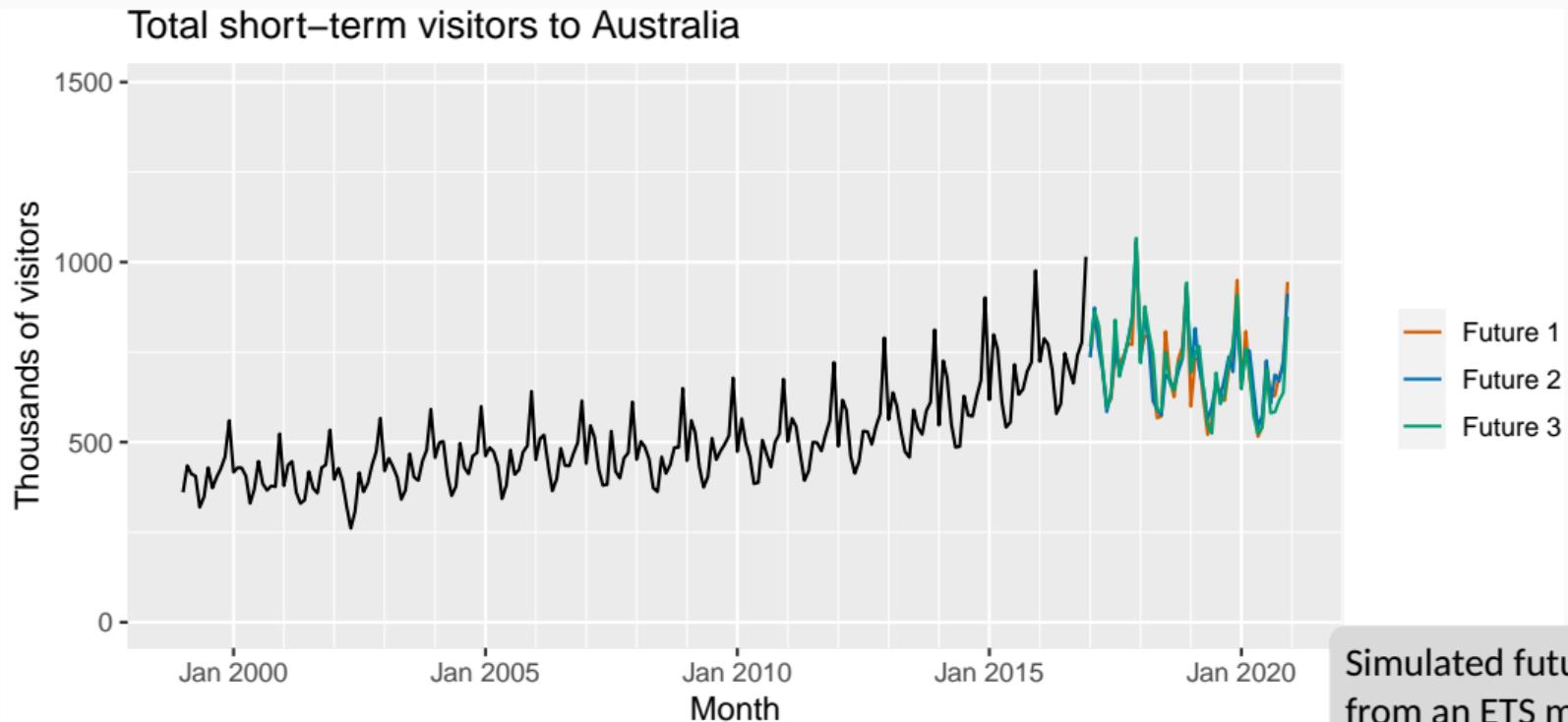
# Random futures

A forecast is an estimate of the probability distribution of a variable to be observed in the future.



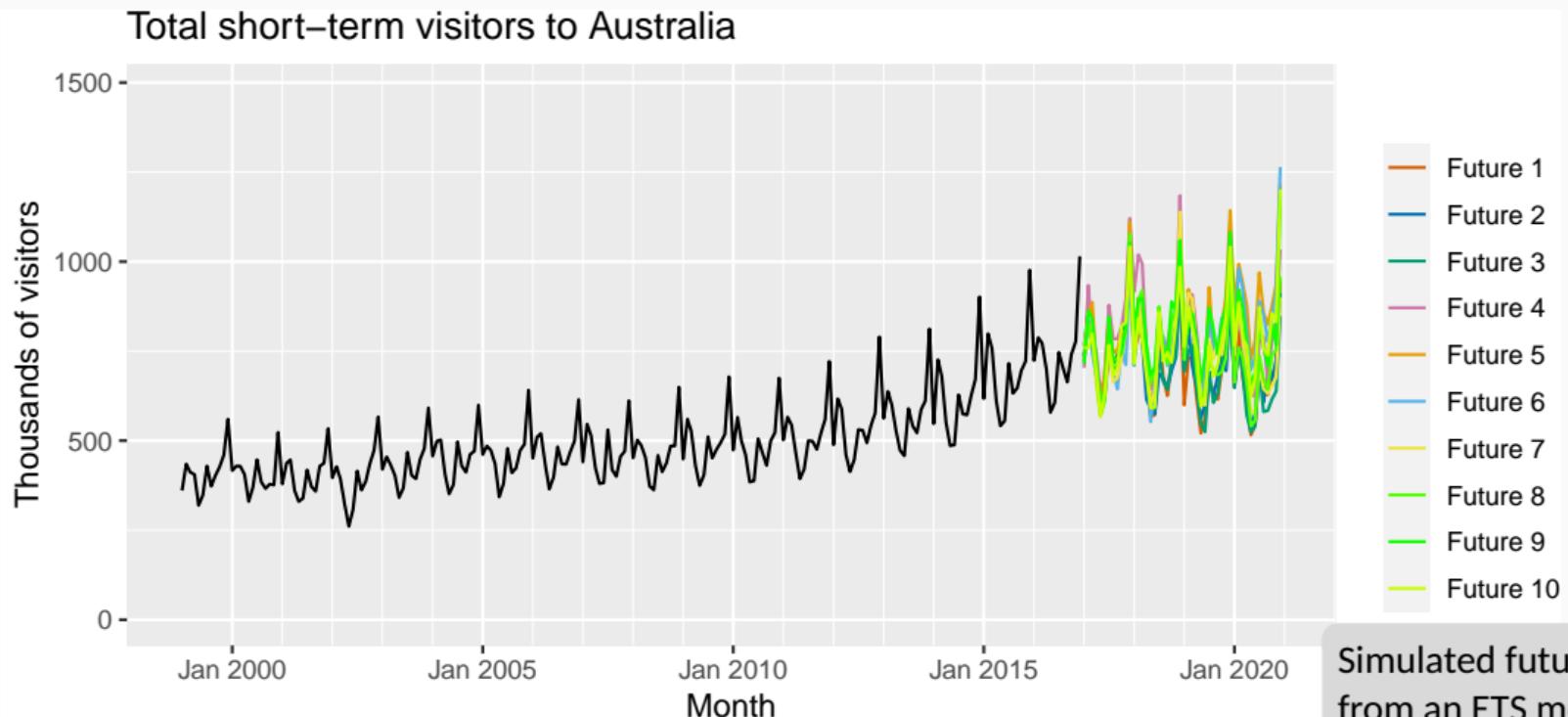
# Random futures

A forecast is an estimate of the probability distribution of a variable to be observed in the future.



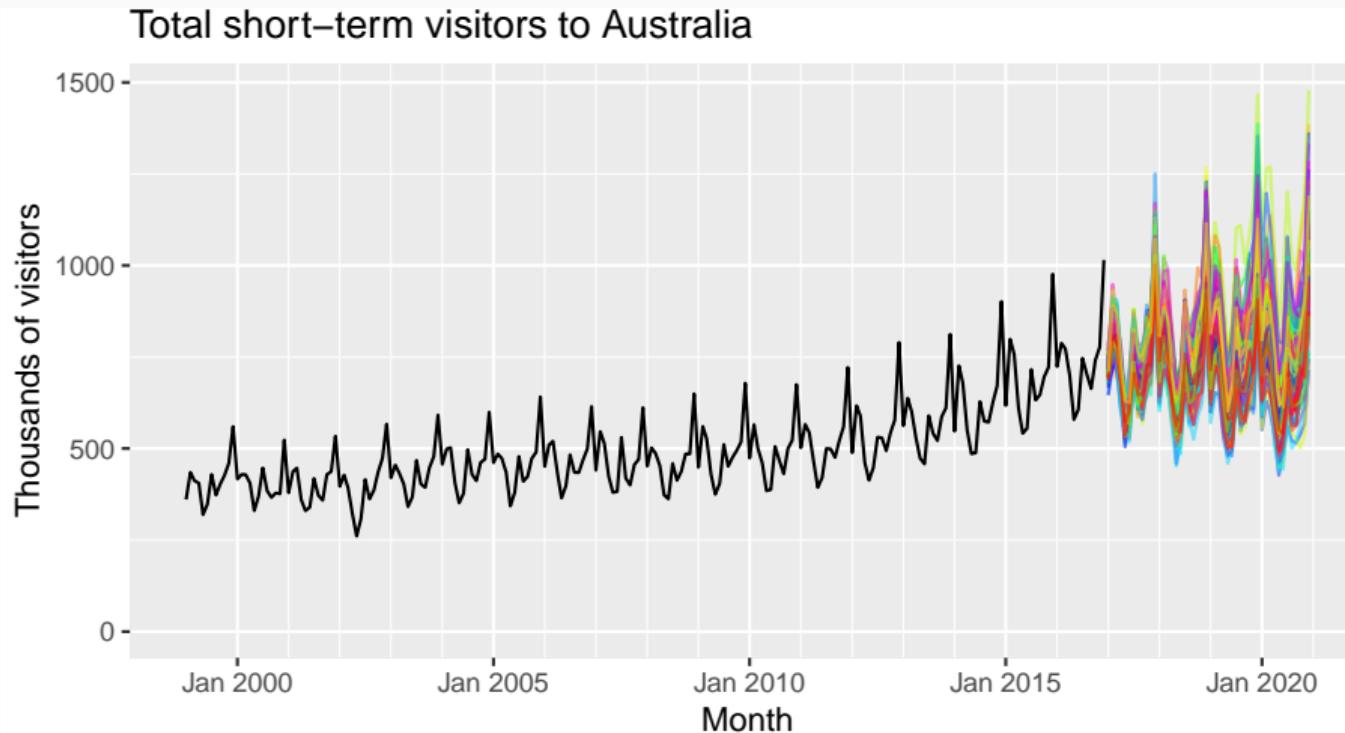
## Random futures

A forecast is an estimate of the probability distribution of a variable to be observed in the future.



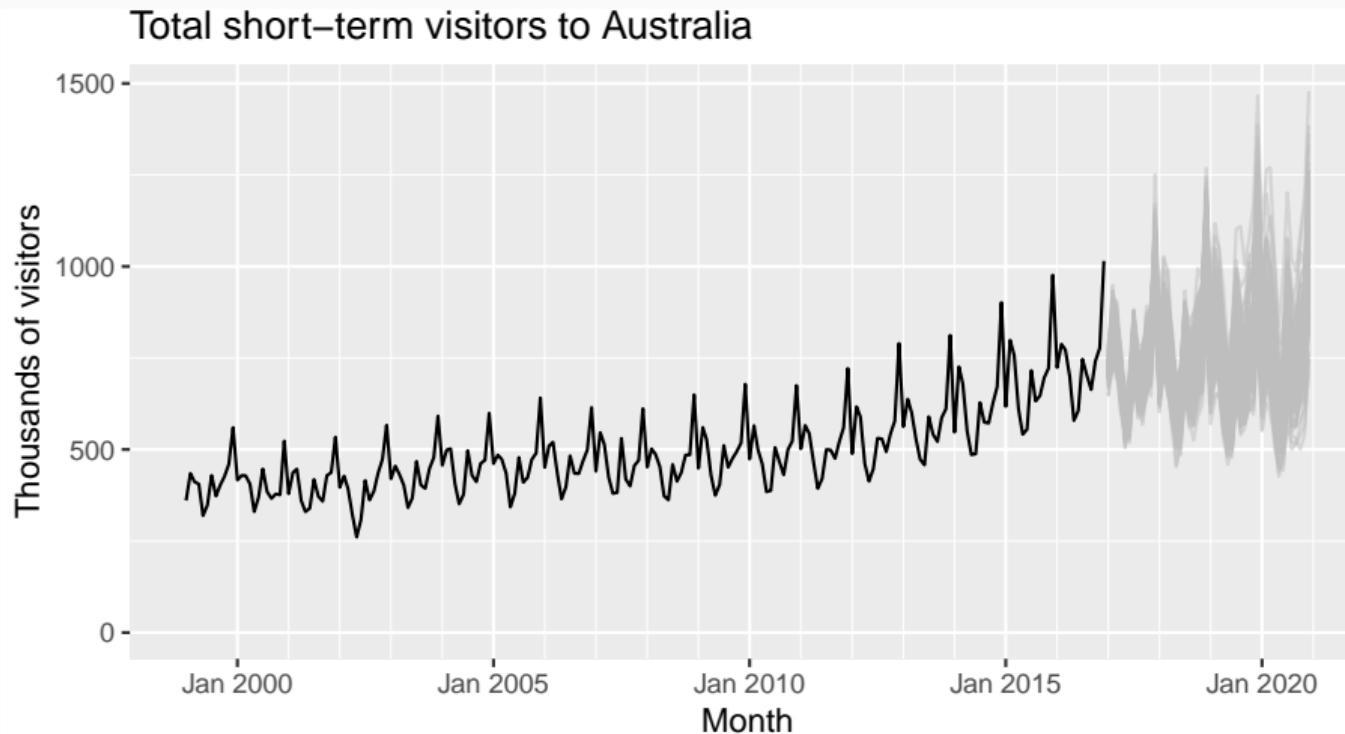
# Random futures

A forecast is an estimate of the probability distribution of a variable to be observed in the future.



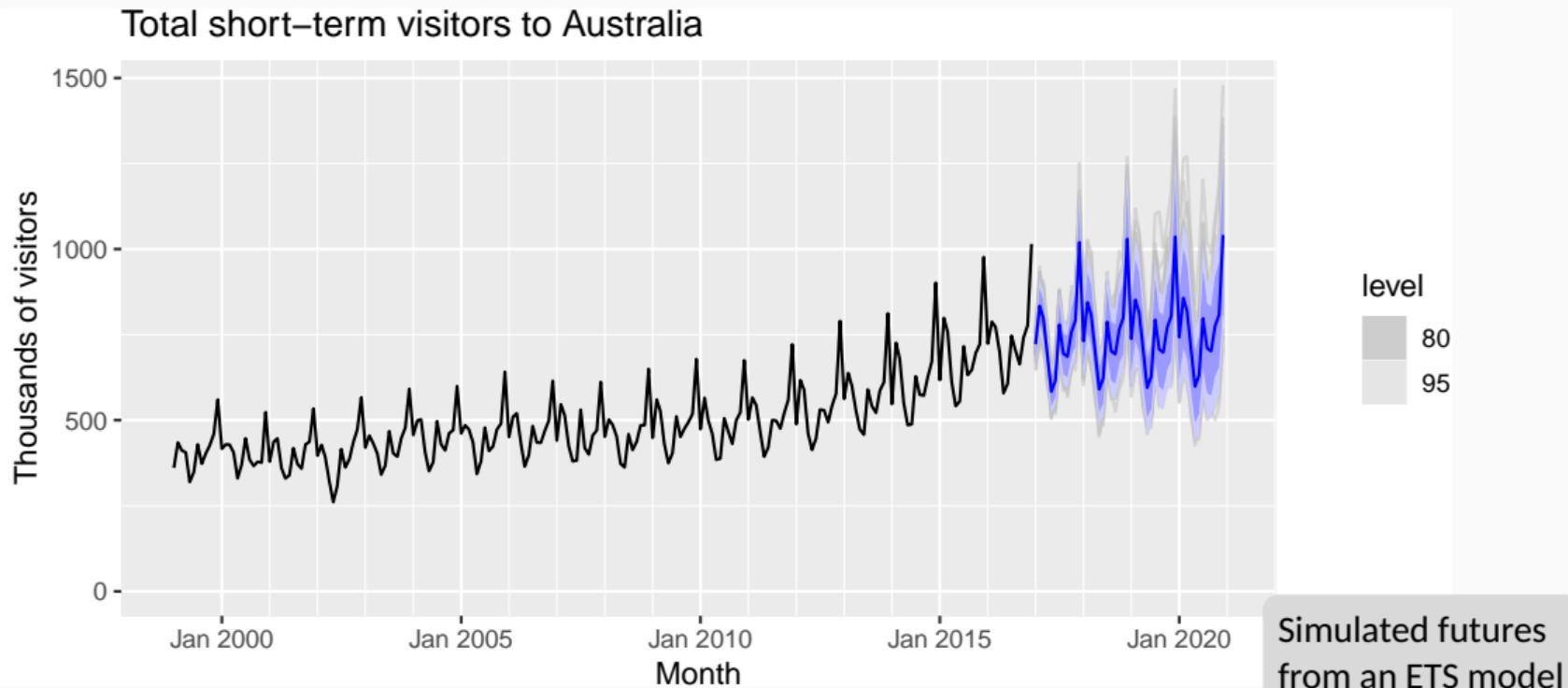
# Random futures

A forecast is an estimate of the probability distribution of a variable to be observed in the future.



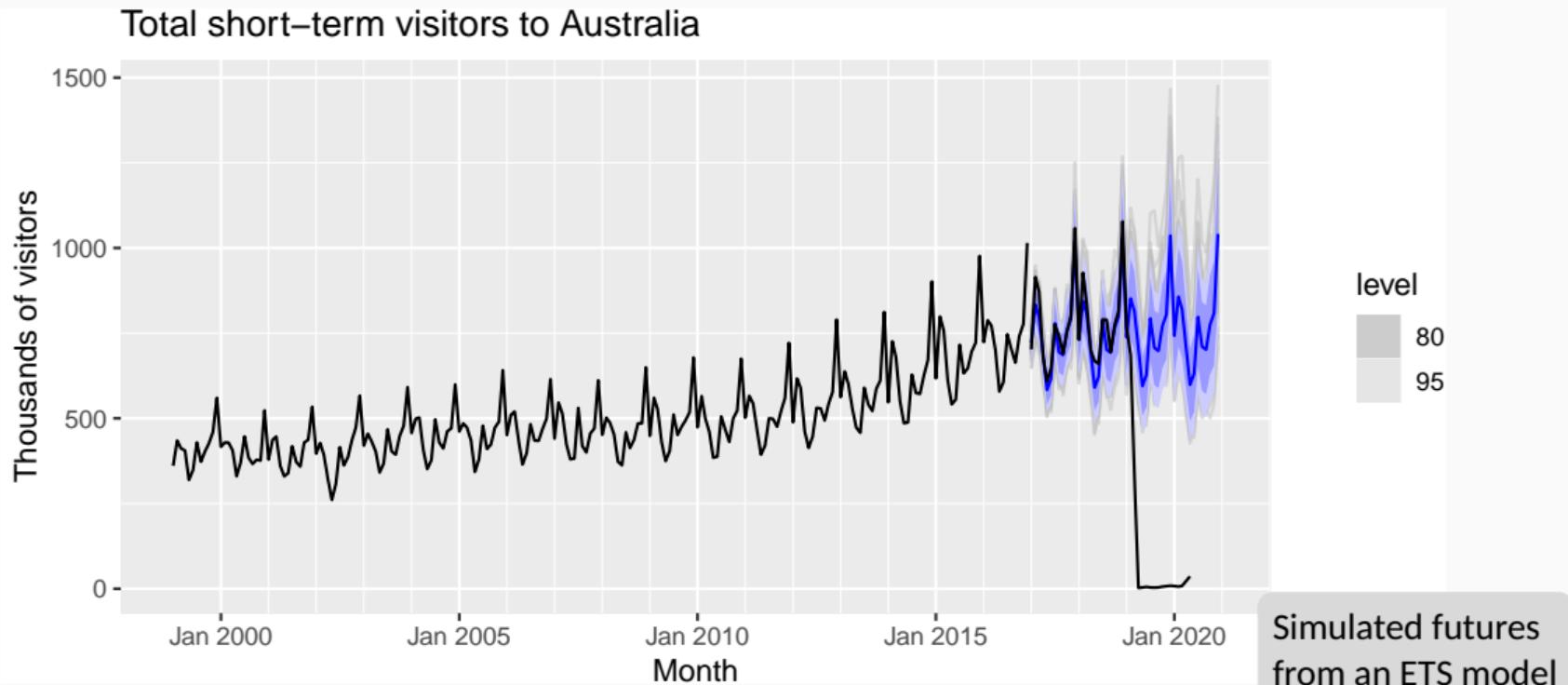
# Random futures

A forecast is an estimate of the probability distribution of a variable to be observed in the future.



# Random futures

A forecast is an estimate of the probability distribution of a variable to be observed in the future.



# Outline

- 1 What can we forecast?
- 2 Forecastability factors
- 3 Forecasting is difficult
- 4 The statistical forecasting perspective
- 5 PBS forecasting
- 6 COVID19 case forecasting
- 7 Are we uncertain enough?

# PBS forecasting

The Pharmaceutical Benefits Scheme (PBS) is the Australian government drugs subsidy scheme.

- Many drugs bought from pharmacies are subsidised to allow more equitable access to modern drugs.
- The cost to government is determined by the number and types of drugs purchased. Currently nearly 1% of GDP.
- The total cost is budgeted based on forecasts of drug usage.

# PBS forecasting

**ABC News Online**  
AUSTRALIAN BROADCASTING CORPORATION

Select a Topic from the list below

- ▶ [Top Stories](#)
- ▶ [Just In](#)
- ▶ [World](#)
- ▶ [Asia-Pacific](#)
- ▶ [Business](#)
- ▶ [Sport](#)
- ▶ [Arts](#)
- ▶ [Sci Tech](#)
- ▶ [Indigenous](#)
- ▶ [Weather](#)
- ▶ [Rural](#)
- ▶ [Local News](#)
- ▶ [Broadband](#)

Search

Click "Refresh" or "Reload" on your browser for the latest edition.

This Bulletin: Wed, May 30 2001 6:22 PM AEST

**POLITICS**

## Opp demands drug price restriction after PBS budget blow-out

The Federal Opposition has called for tighter controls on drug prices after the Pharmaceutical Benefits Scheme (PBS) budget blew out by almost \$800 million.

The money was spent on two new drugs including the controversial anti-smoking aid Zyban, which dropped in price from \$220 to \$22 after it was listed on the

**Windows Media** **NewsRadio**  
Streaming audio news  
LISTEN: [WMP](#) | [Real](#)

**the Public Record**  
For full election coverage

**FEATURES**

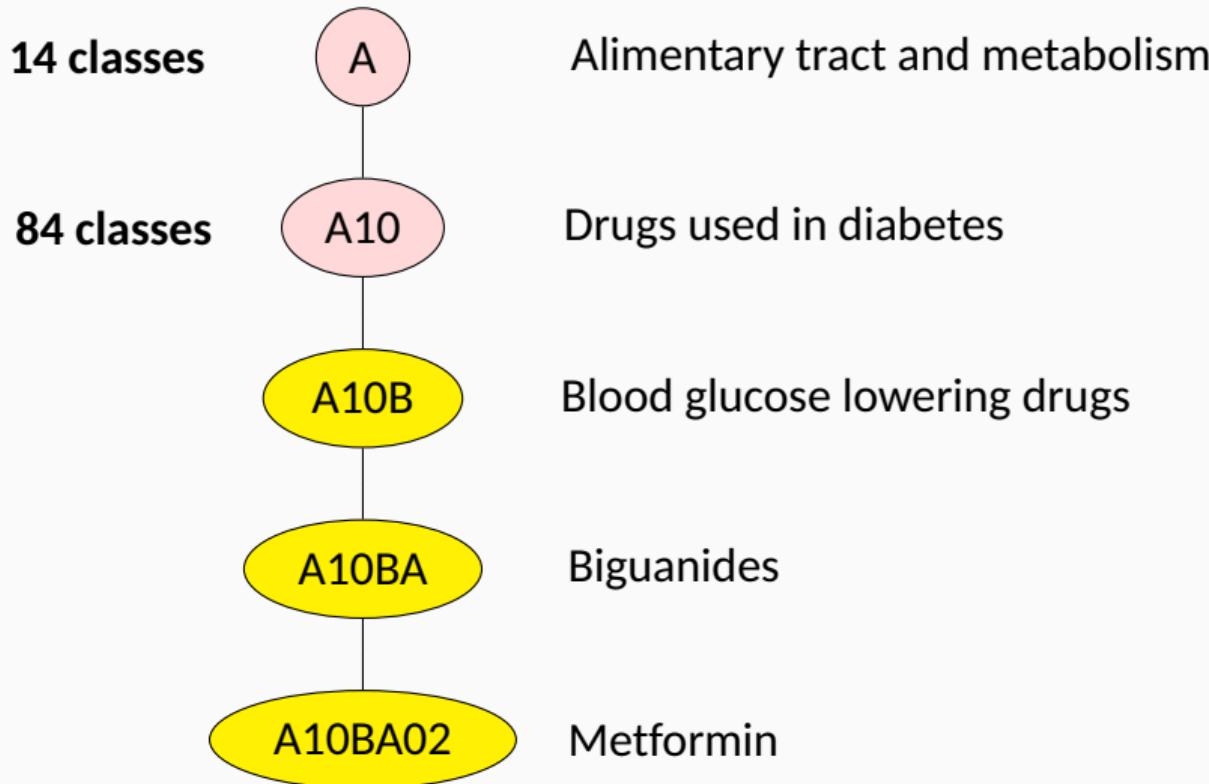
**Public Record**  
Federal Election 2001

For a fresh perspective on the federal election, reach into ABC Online's campaign weblog, [The Poll Vault](#).

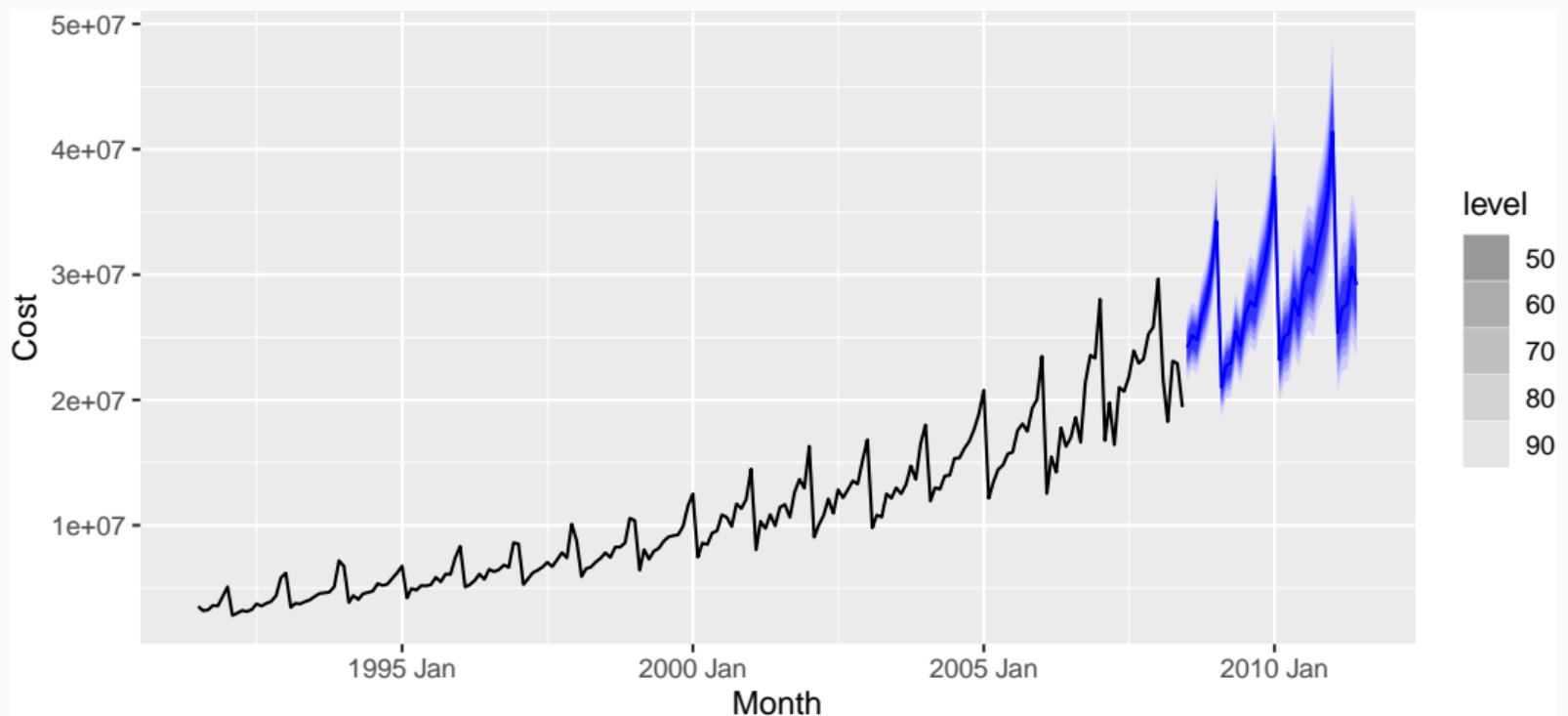
## PBS forecasting

- In 2001: \$4.5 billion budget, under-forecasted by \$800 million.
- Thousands of products. Seasonal demand.
- Subject to covert marketing, volatile products, uncontrollable expenditure.
- Although monthly data available for 10 years, data are aggregated to annual values, and only the first three years are used in estimating the forecasts.
- All forecasts being done with the FORECAST function in MS-Excel!

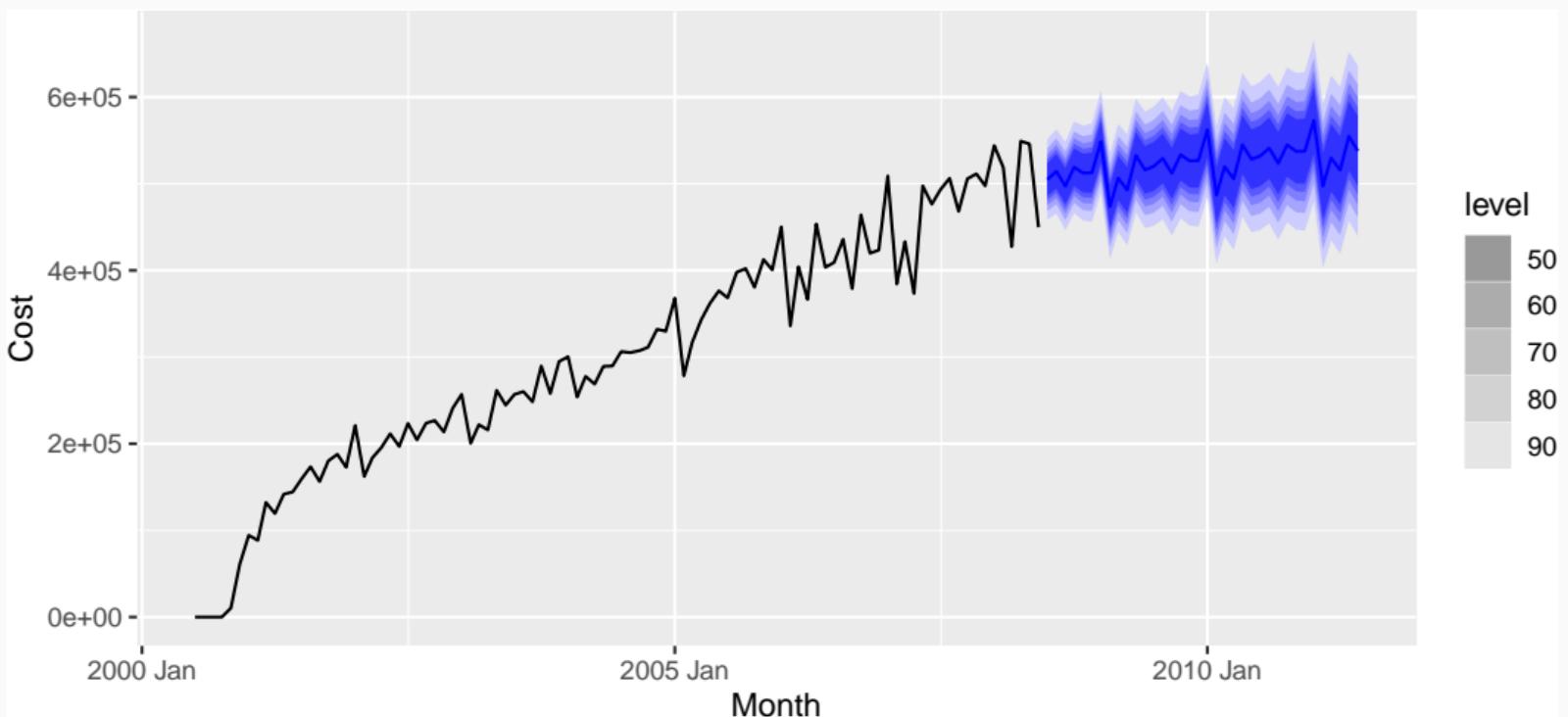
# ATC drug classification



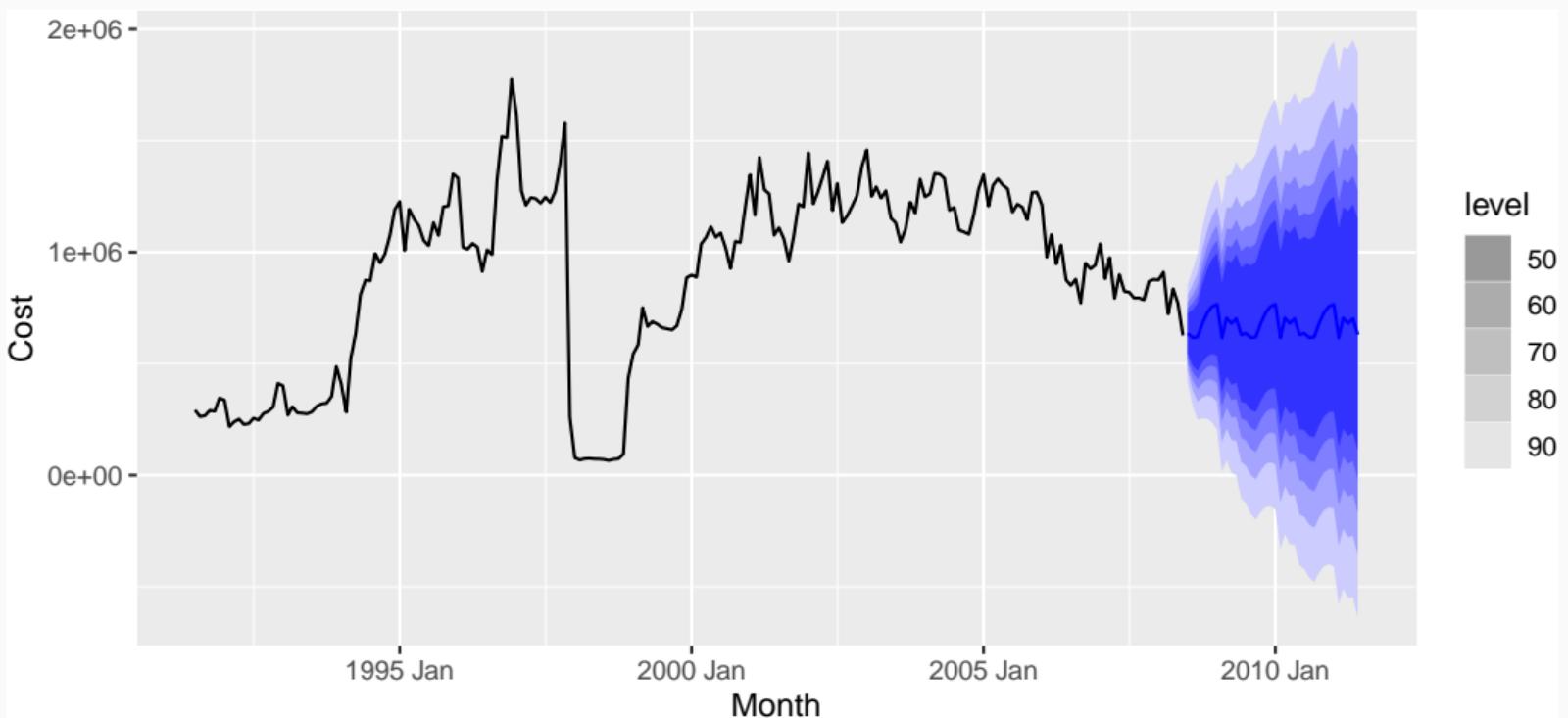
# ETS forecasts of PBS data



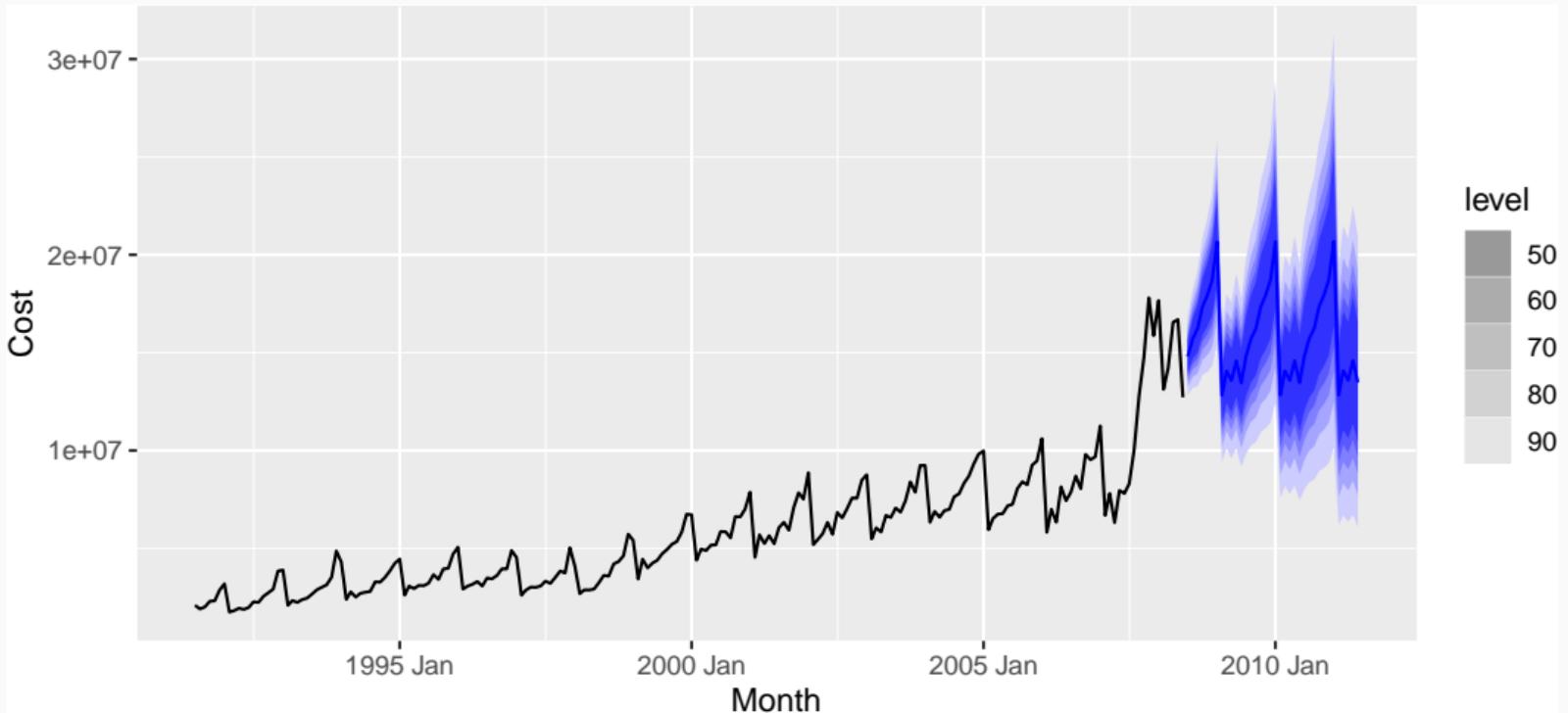
# ETS forecasts of PBS data



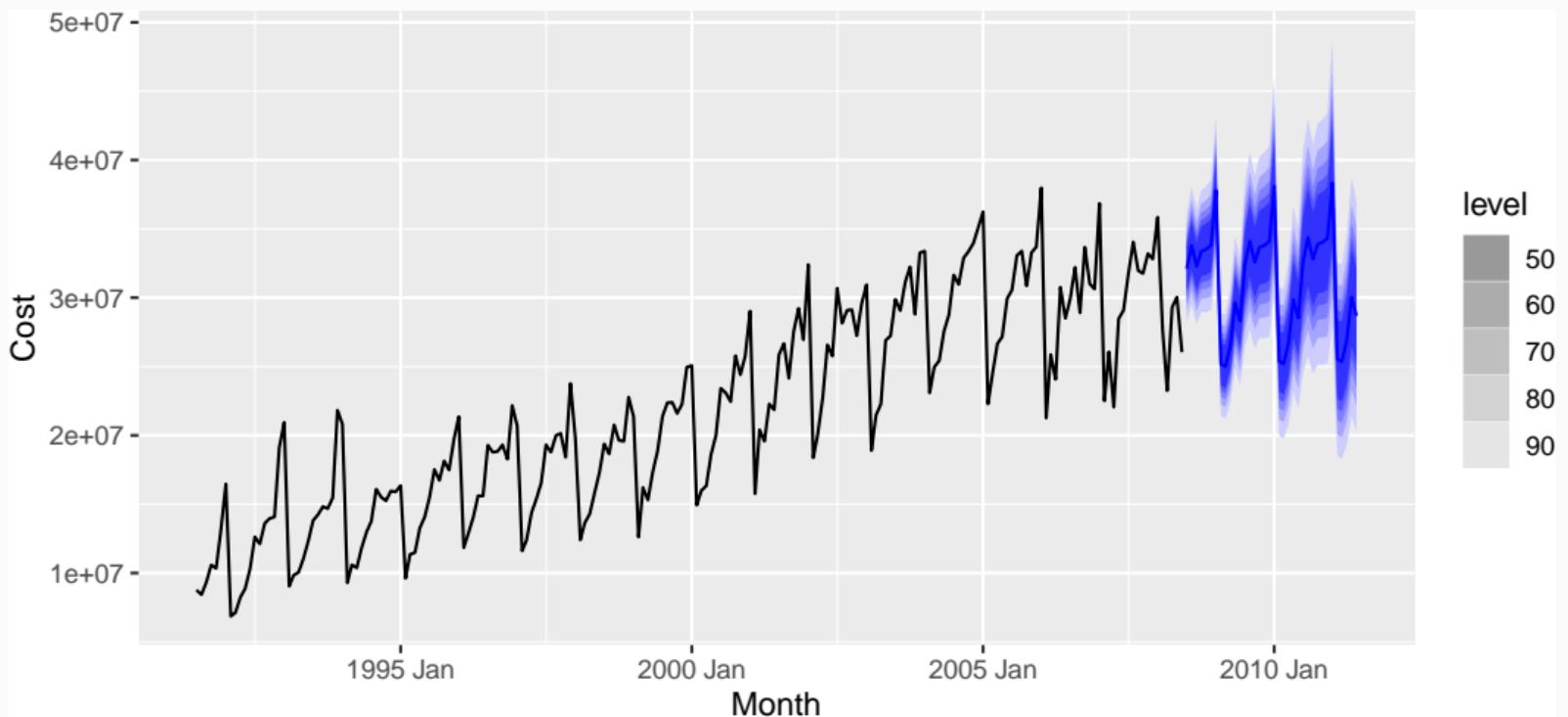
# ETS forecasts of PBS data



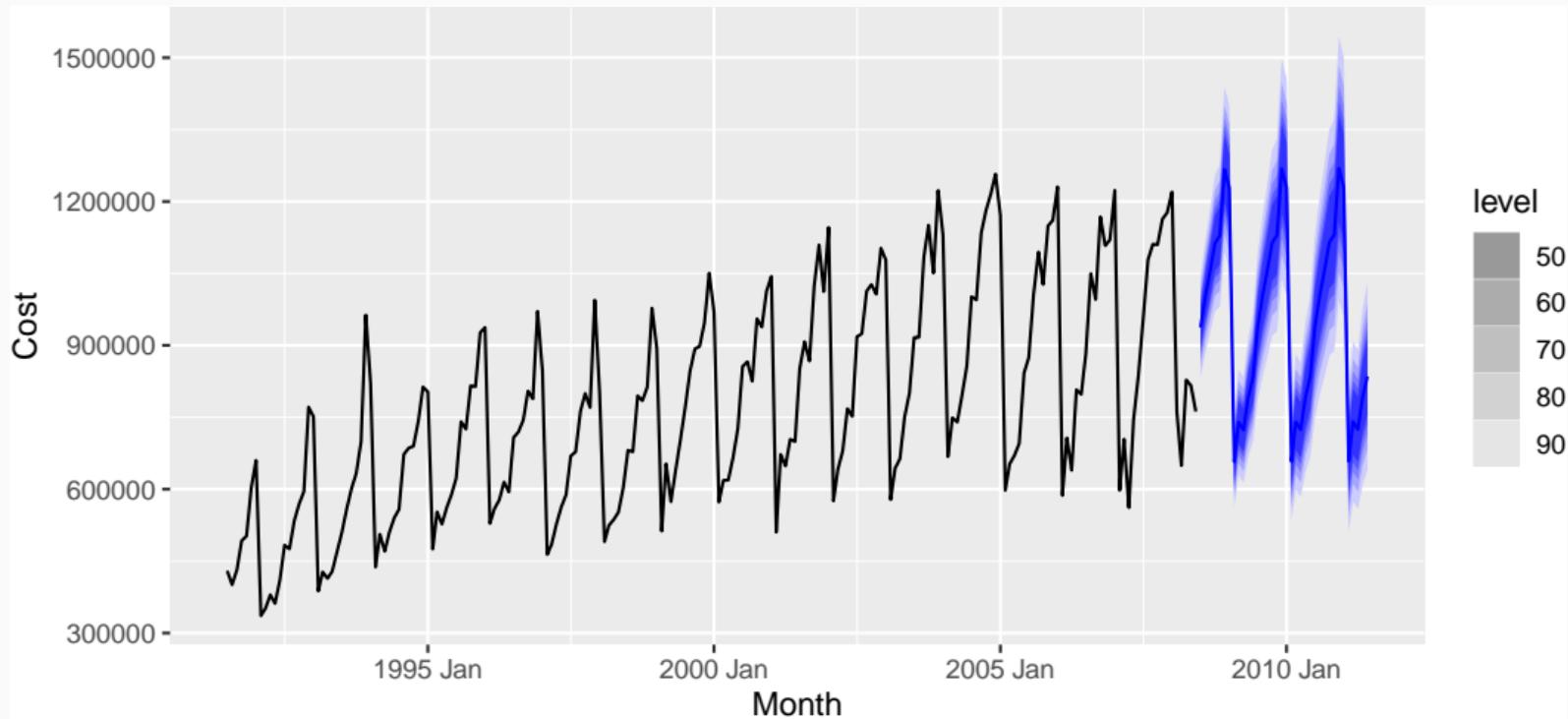
# ETS forecasts of PBS data



# ETS forecasts of PBS data



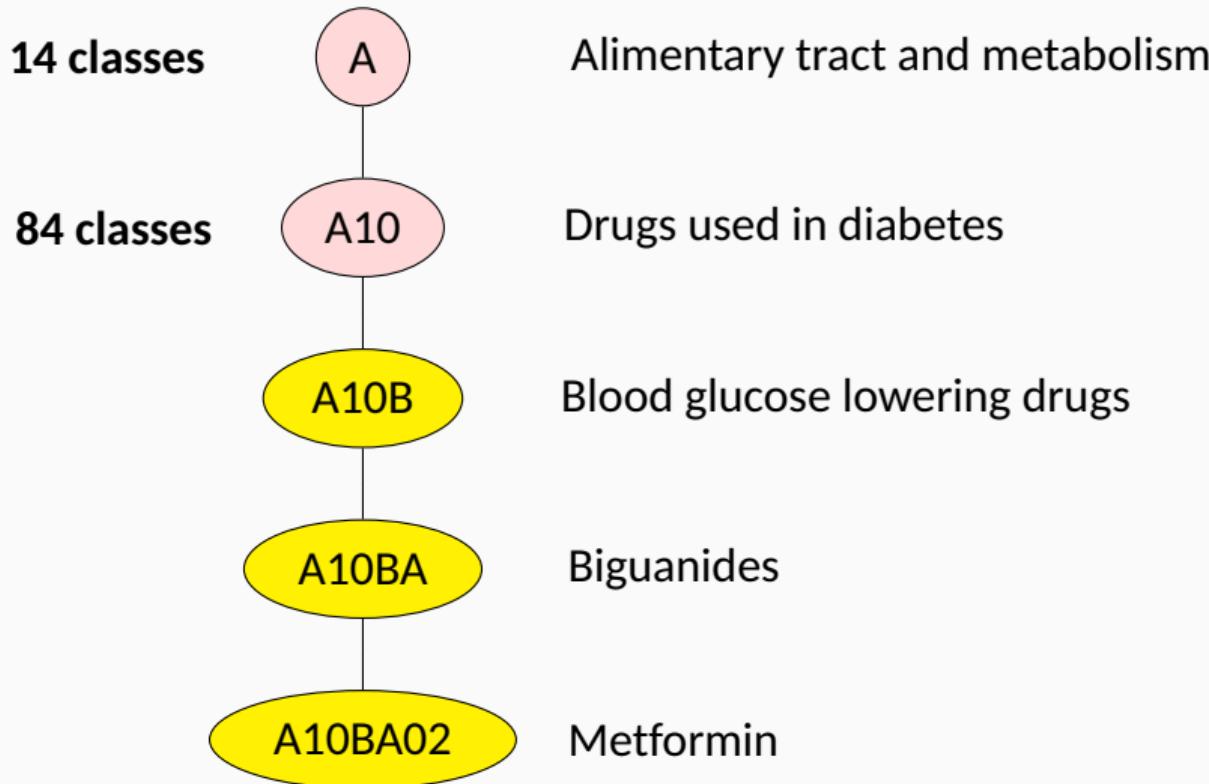
# ETS forecasts of PBS data



# Forecasting the PBS

- Developed an automatic forecasting algorithm for exponential smoothing state space models based on the AIC.
- Exponential smoothing models allowed for time-changing trend and seasonal patterns.
- Forecast MAPE reduced from 15–20% to 0.6%.
- State space models provide prediction intervals which give a sense of uncertainty.
- Theory and algorithm published as Hyndman et al (IJF, 2002).
- Now implemented in R as `ets()` in `forecast` package, as `ETS()` function in `fable` package, and in Tableau and elsewhere.
- NOT implemented in FORECAST.ETS function in MS-Excel.

# Hierarchical forecasting



# Hierarchical forecasting

- Nearly ten years later, Hyndman et al (CSDA, 2011) proposed forecast reconciliation to handle hierarchical time series.
- Now widely used in business and industry and implemented in the `hts` and `fable` packages.

# Outline

- 1 What can we forecast?
- 2 Forecastability factors
- 3 Forecasting is difficult
- 4 The statistical forecasting perspective
- 5 PBS forecasting
- 6 COVID19 case forecasting
- 7 Are we uncertain enough?

# Data sources

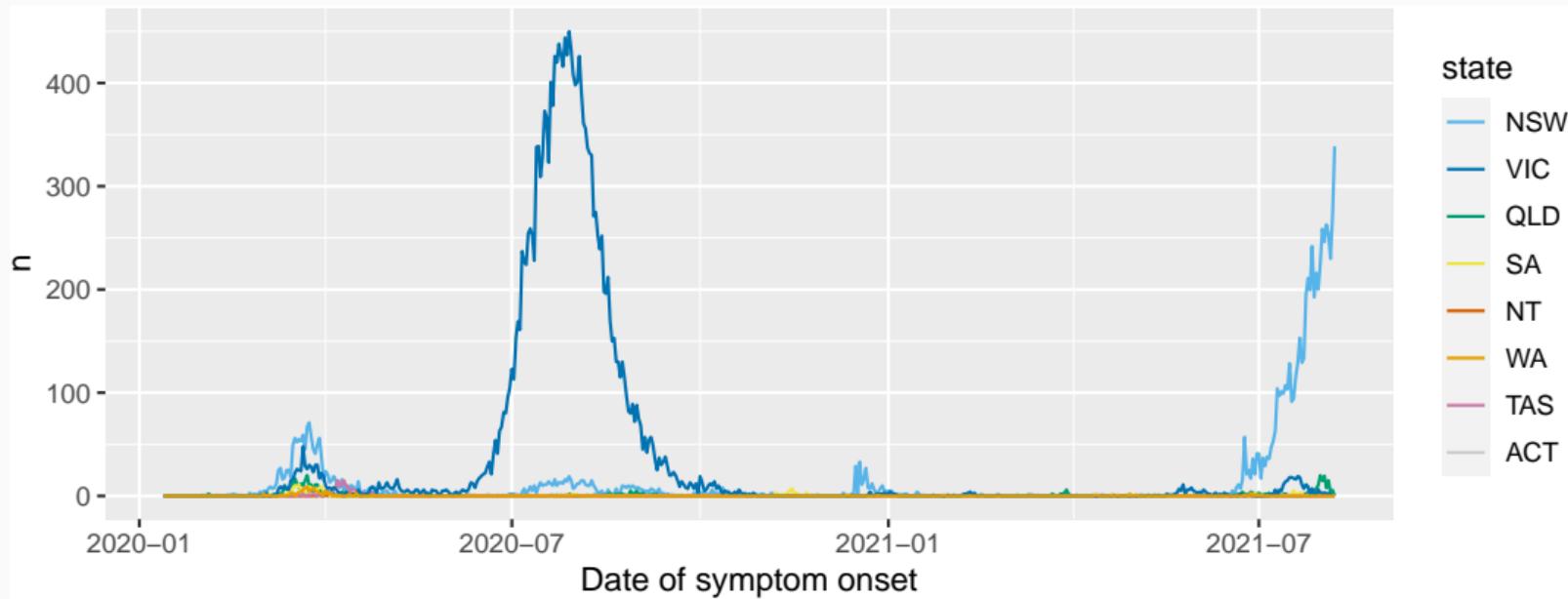
- Case-level data of all positive COVID-19 tests: onset and detection times.
- Daily population mobility data from Google, Apple & Facebook
- Weekly non-household contact surveys
- Weekly behavioural surveys
- Daily case numbers from many countries and regions via the Johns Hopkins COVID-19 repository

# Case numbers

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

```
## # A tsibble: 402 x 3 [1D]
## # Key:       state [1]
##   date      state     n
##   <date>    <chr> <dbl>
## 1 2020-07-01 VIC     123
## 2 2020-07-02 VIC     113
## 3 2020-07-03 VIC     153
## 4 2020-07-04 VIC     169
## 5 2020-07-05 VIC     161
## 6 2020-07-06 VIC     237
## 7 2020-07-07 VIC     227
## 8 2020-07-08 VIC     224
## 9 2020-07-09 VIC     254
```

# Case numbers



- Recent case numbers are uncertain and incomplete as date of onset is not known until symptoms show and a test is obtained.

# Global daily cases by region from Johns Hopkins

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



# A model ensemble

## Model 1: SEIIR (Uni Melbourne/Doherty Institute)

- Stochastic compartmental model with time-varying effective reproduction number.

## Model 2: Generative model (Uni Adelaide)

- Simulation with three types of infectious individuals: imported, asymptomatic, symptomatic

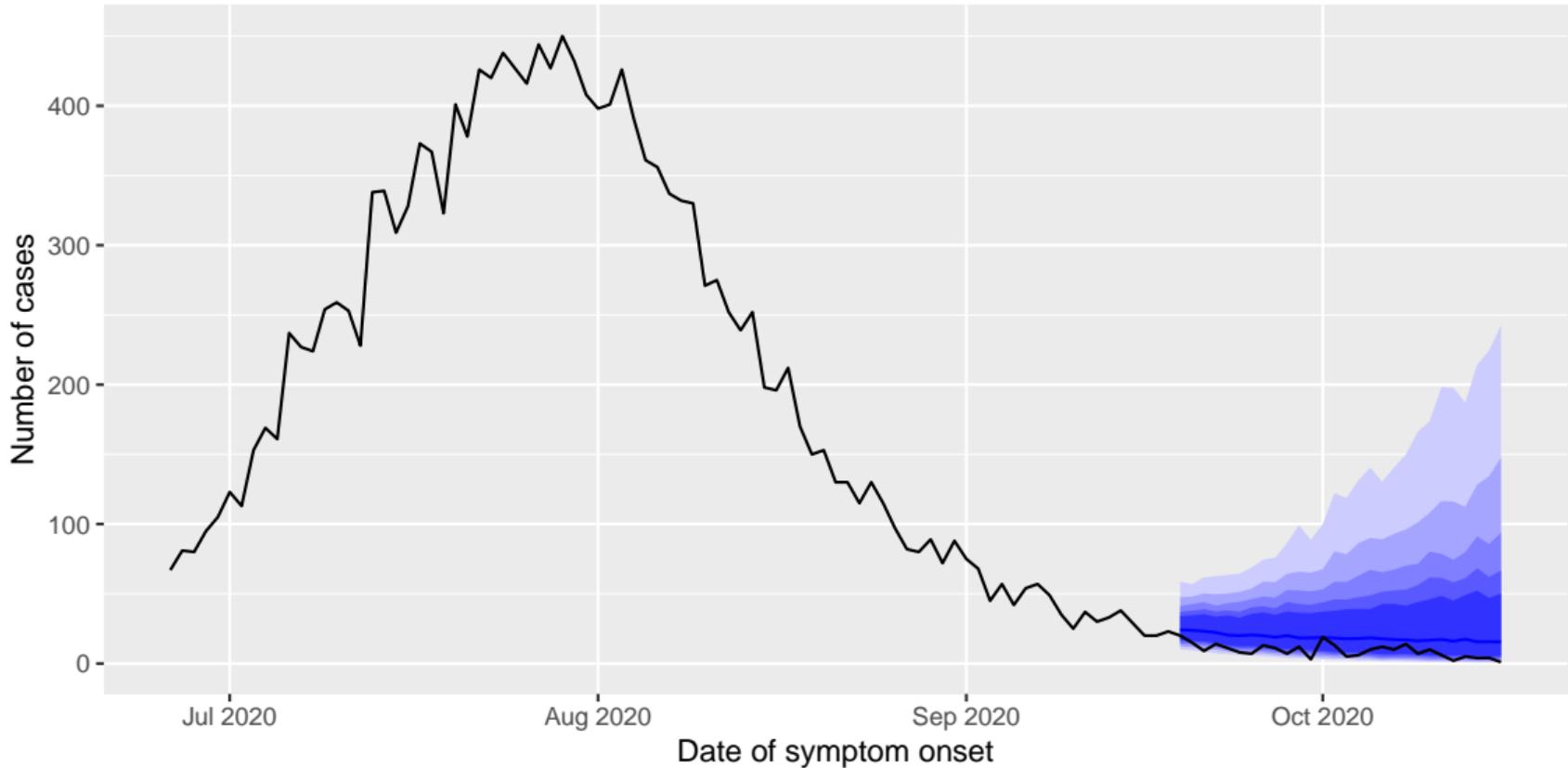
## Model 3: Global AR model (Monash)

- Single model fitted to all Johns Hopkins data from countries and regions with sufficient data.
- Series with obvious anomalies removed.

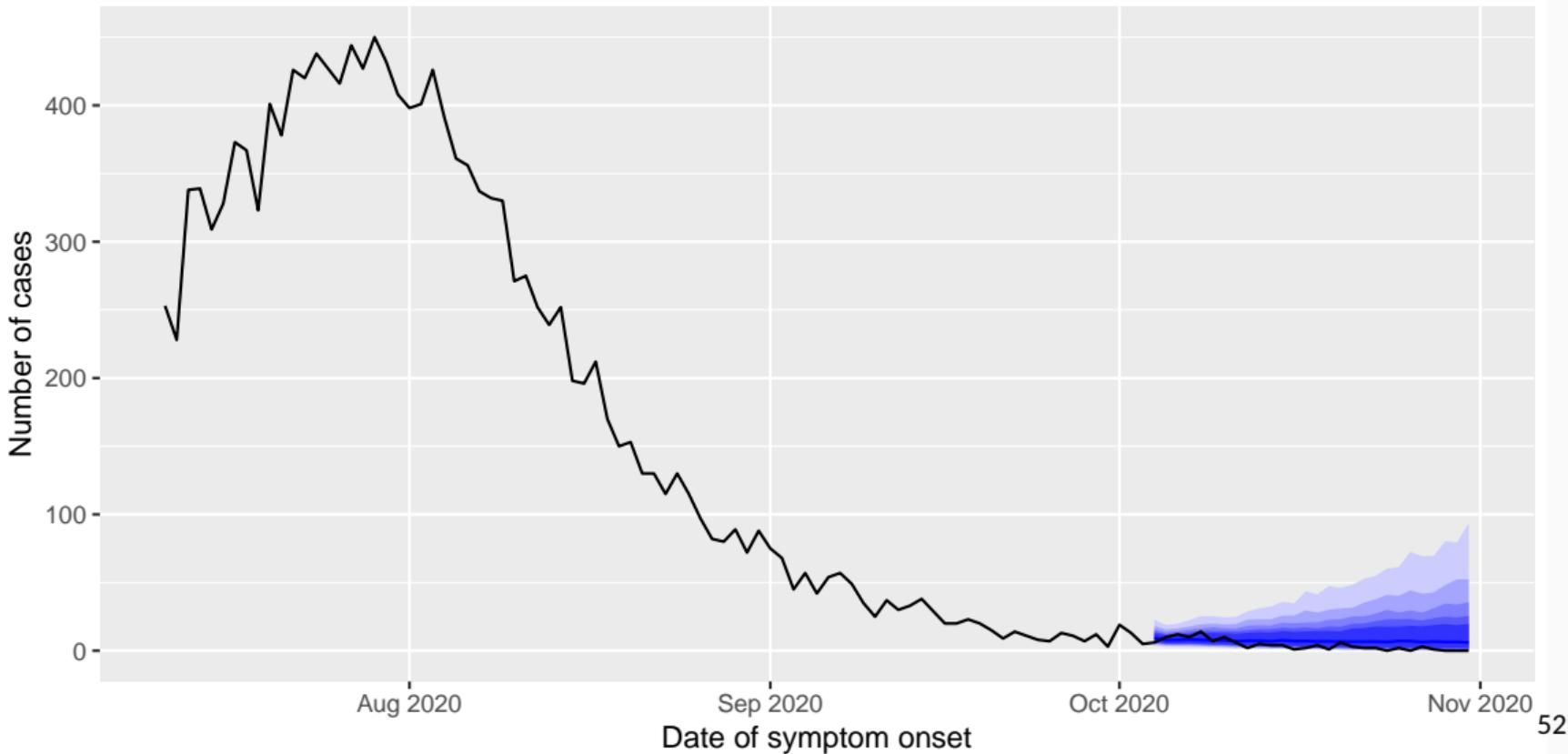
# Forecasting ensemble

- Forecasts obtained from a equally-weighted mixture distribution of the component forecasts.
- Also known as “linear pooling”
- Works best when individual models are over-confident and use different data sources.

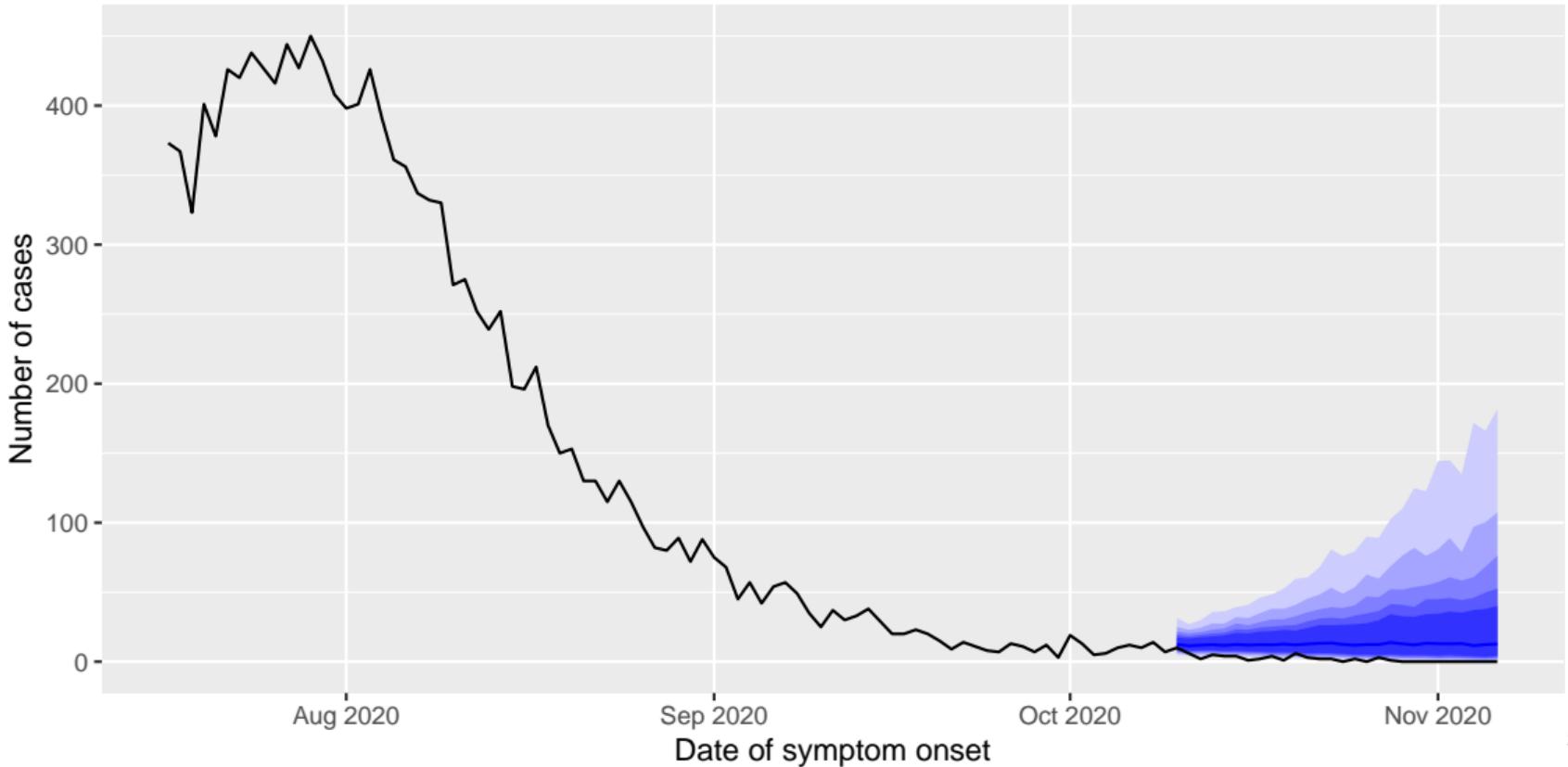
# Ensemble forecasts: Victoria



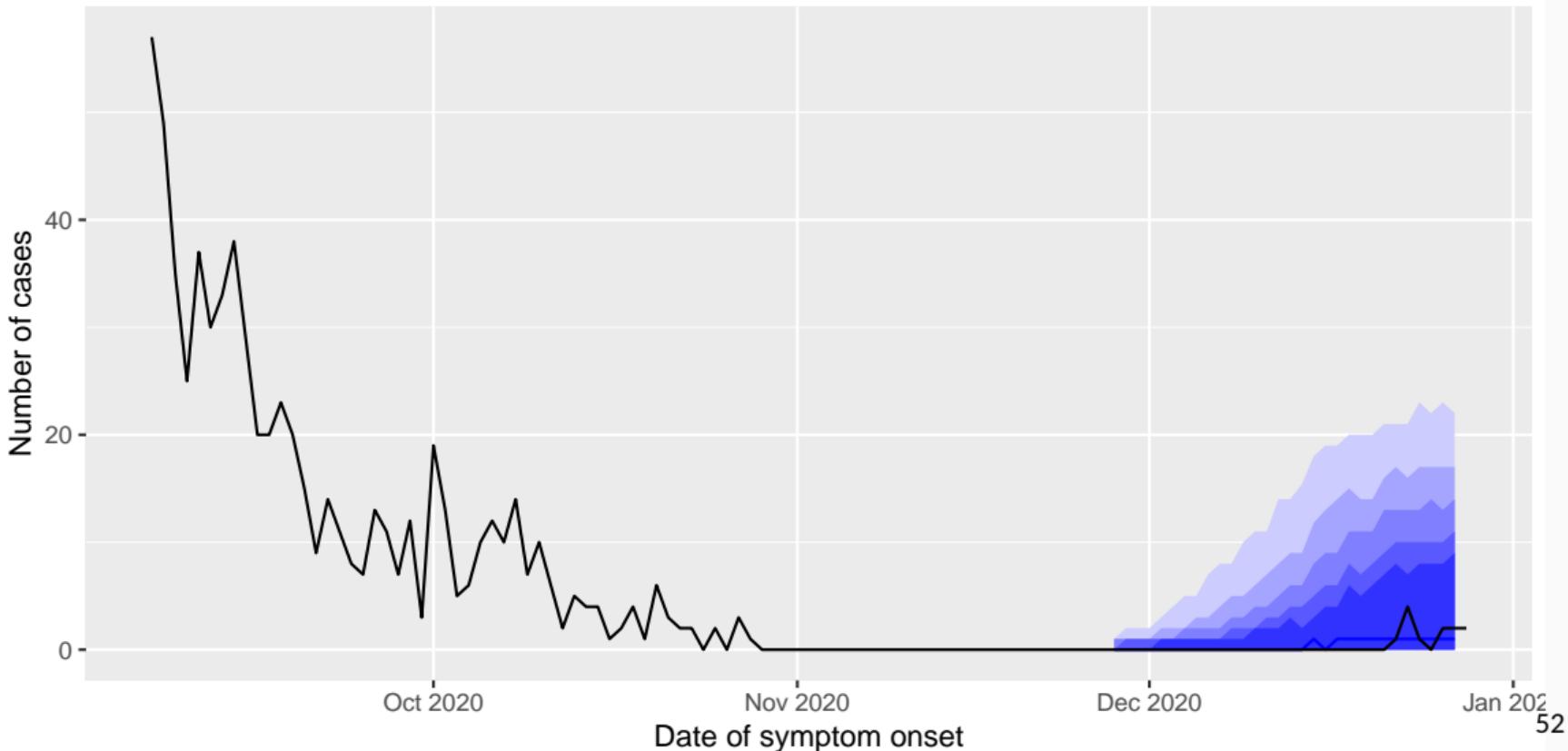
# Ensemble forecasts: Victoria



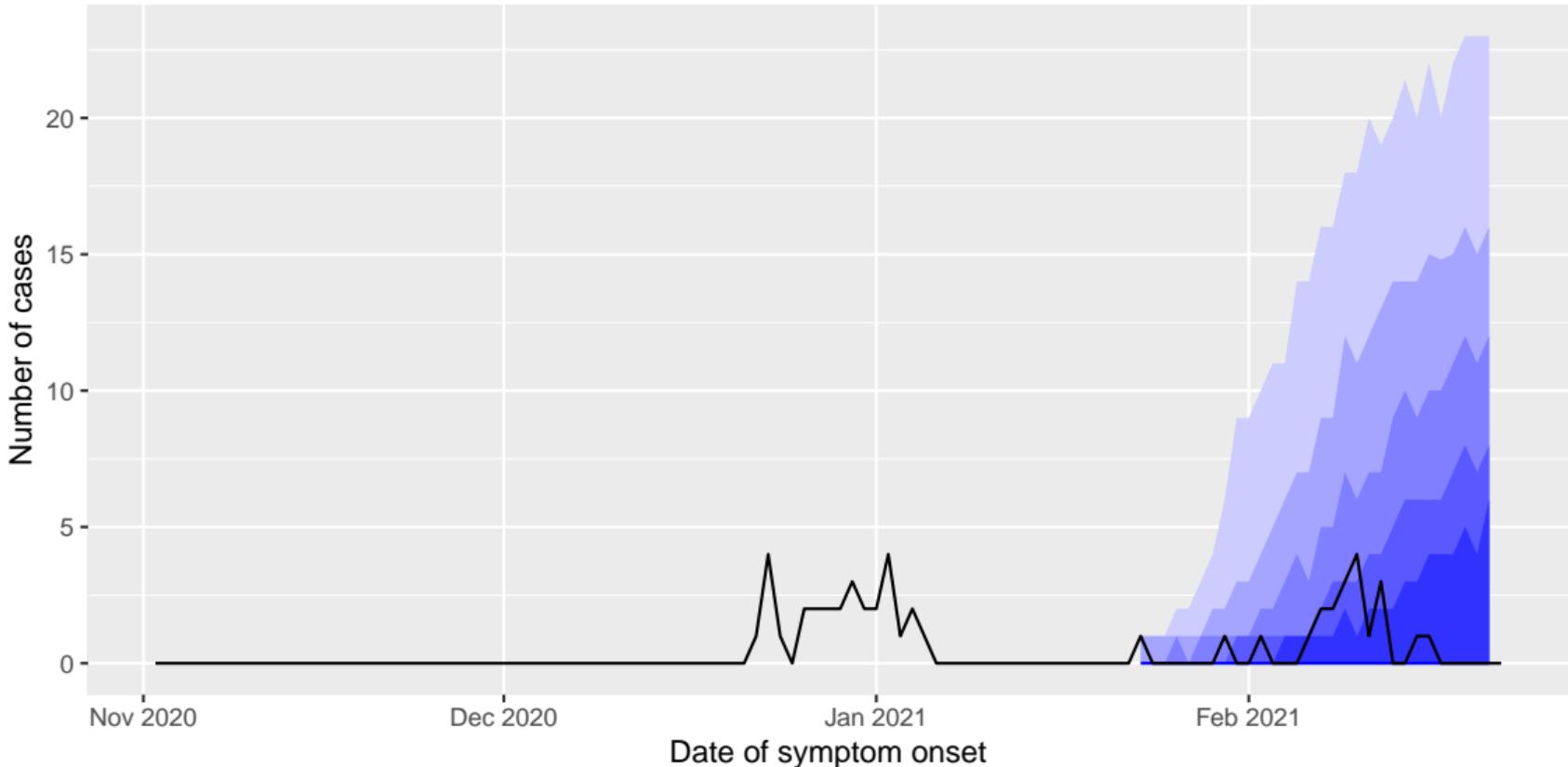
# Ensemble forecasts: Victoria



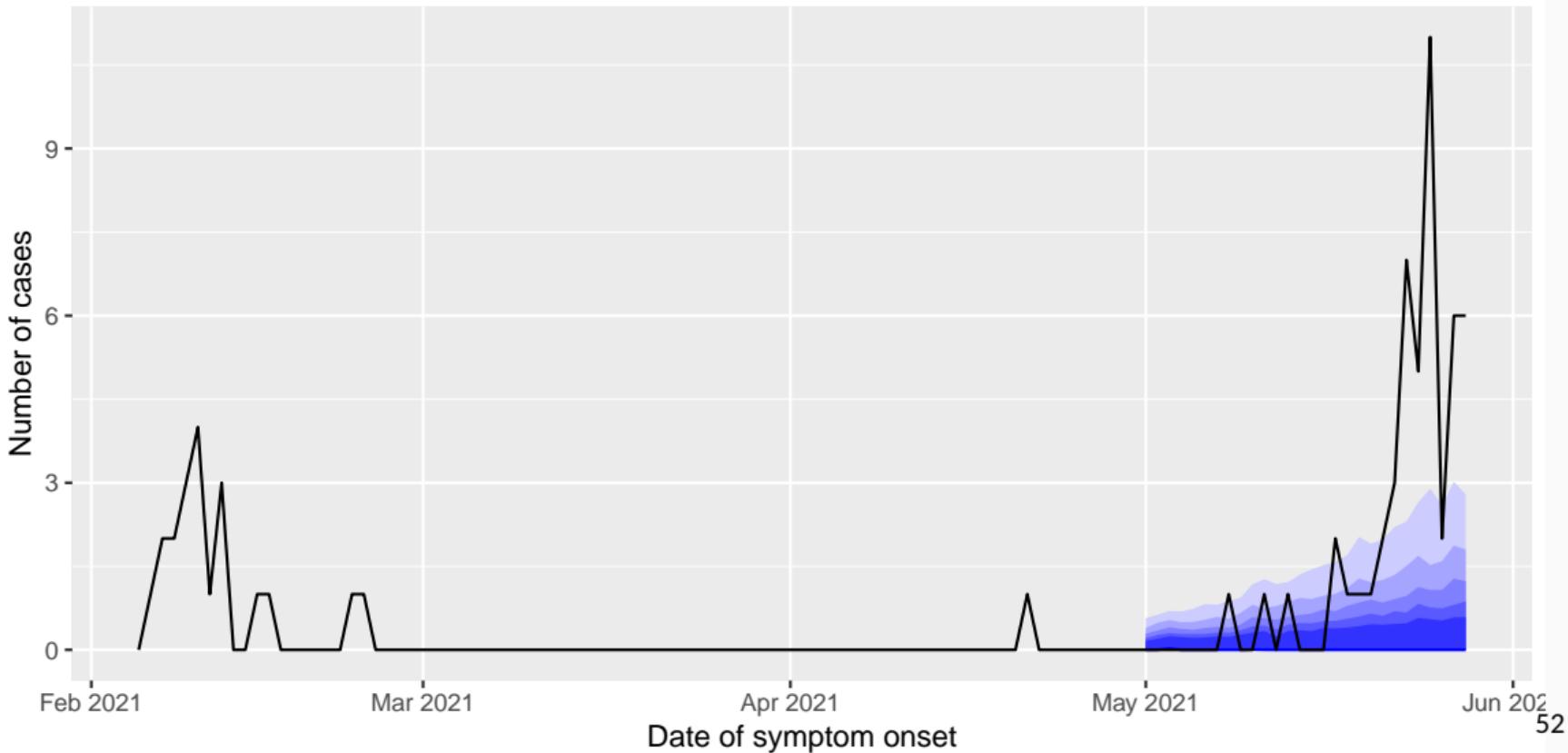
# Ensemble forecasts: Victoria



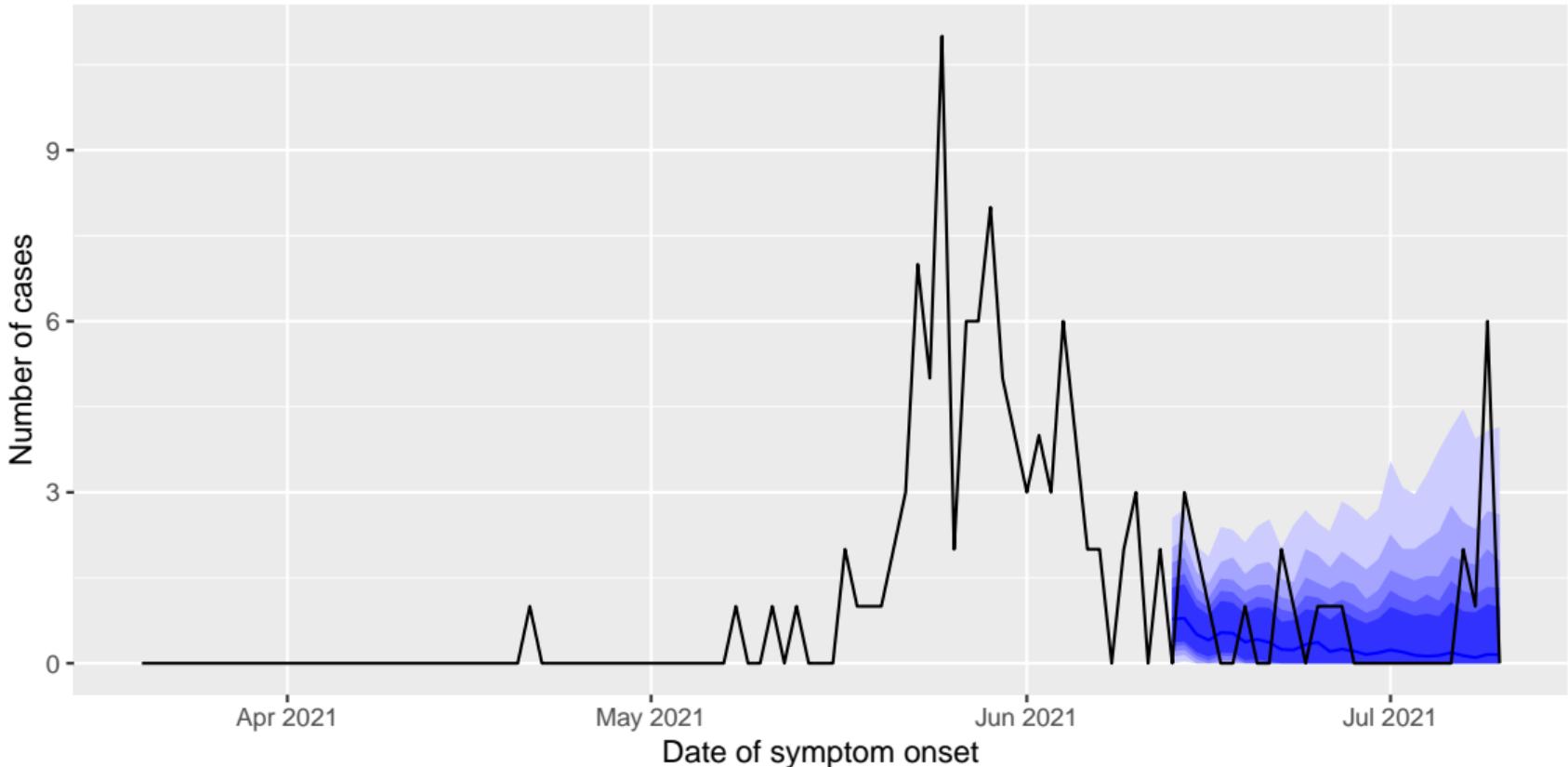
# Ensemble forecasts: Victoria



# Ensemble forecasts: Victoria



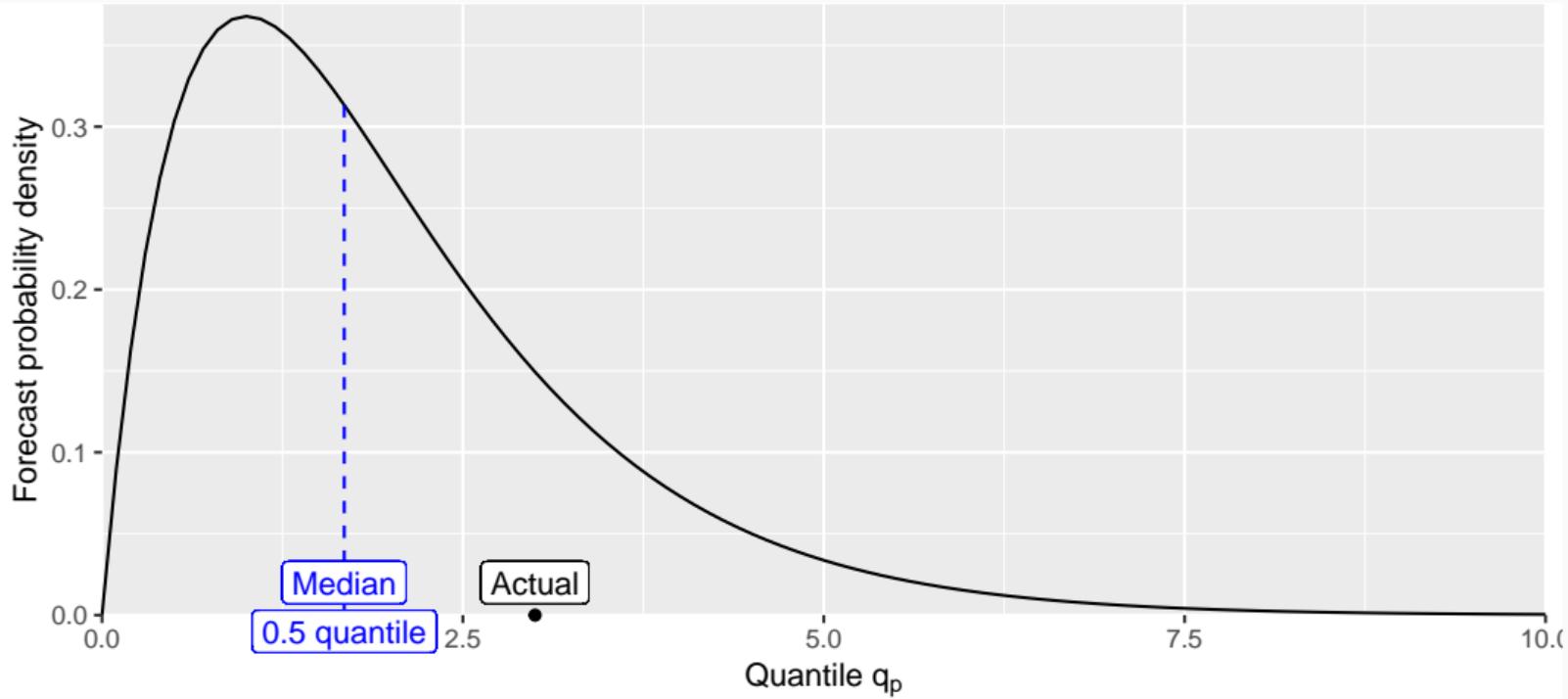
# Ensemble forecasts: Victoria



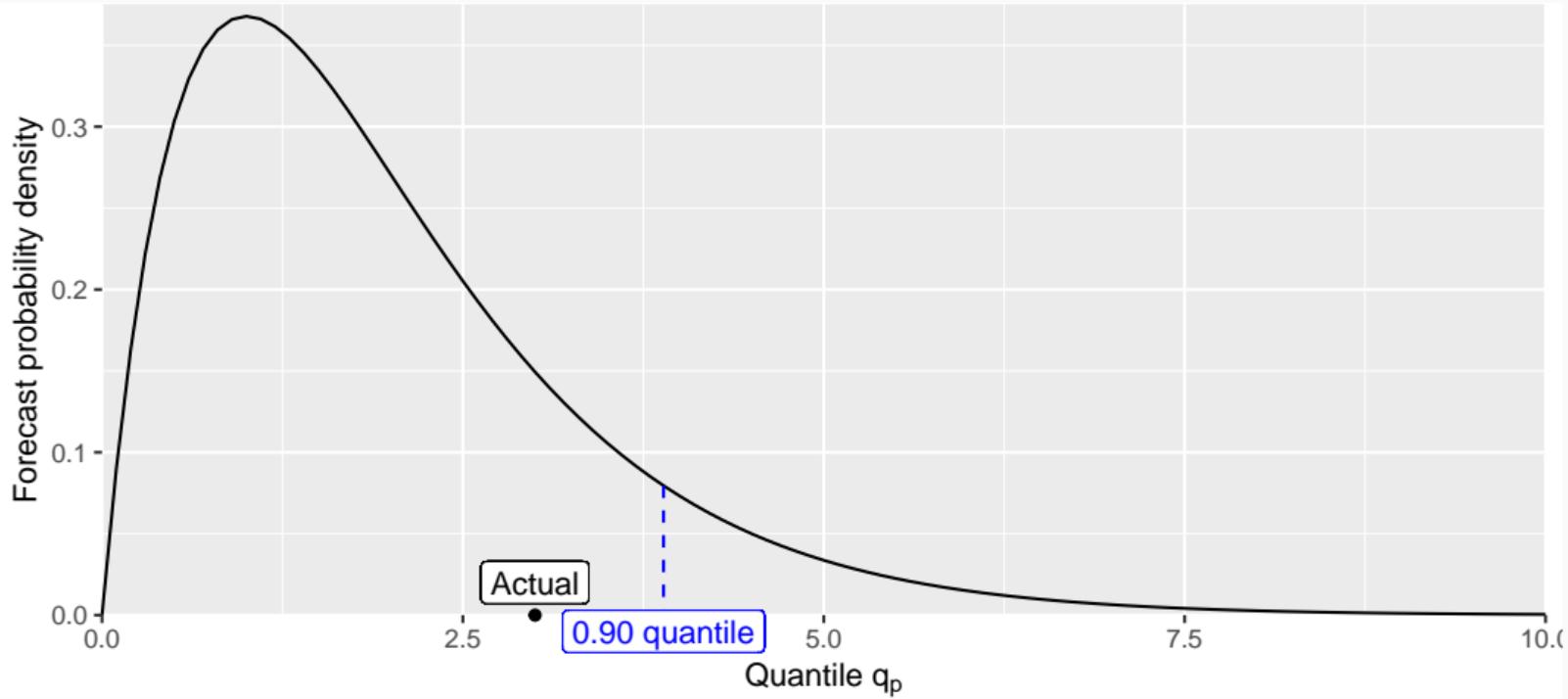
# Outline

- 1 What can we forecast?
- 2 Forecastability factors
- 3 Forecasting is difficult
- 4 The statistical forecasting perspective
- 5 PBS forecasting
- 6 COVID19 case forecasting
- 7 Are we uncertain enough?

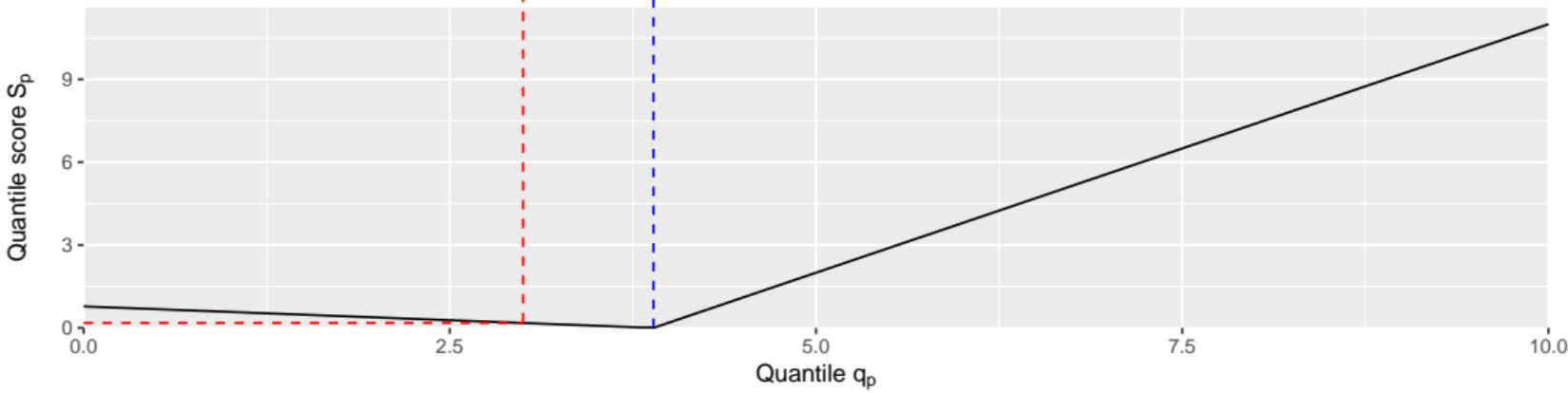
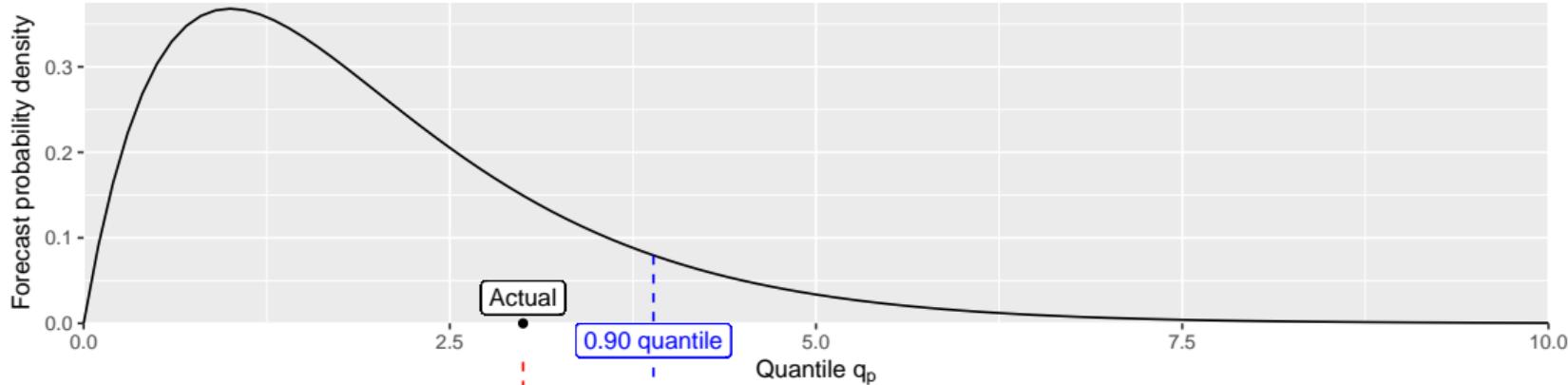
# Evaluating probabilistic forecasts



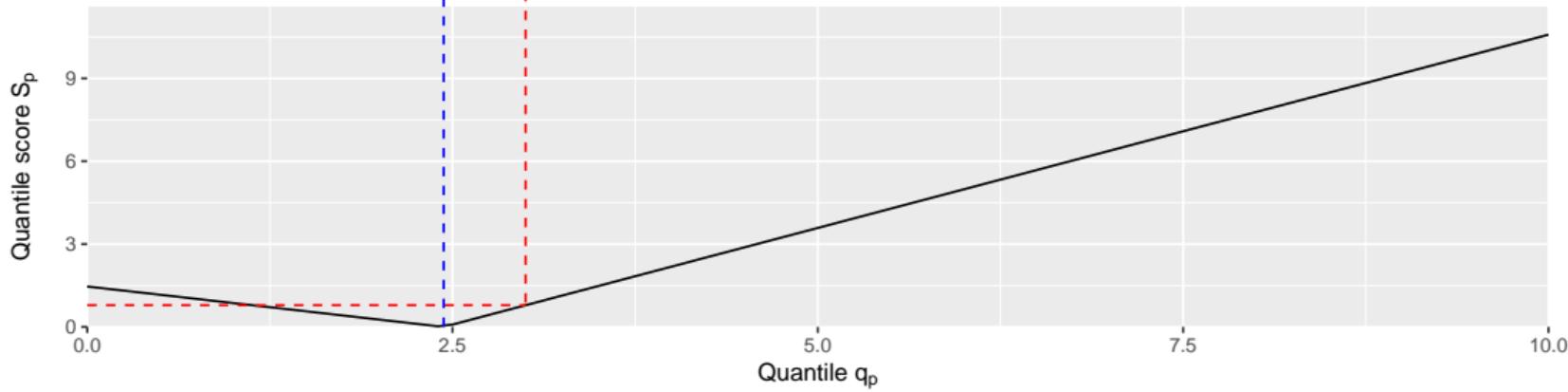
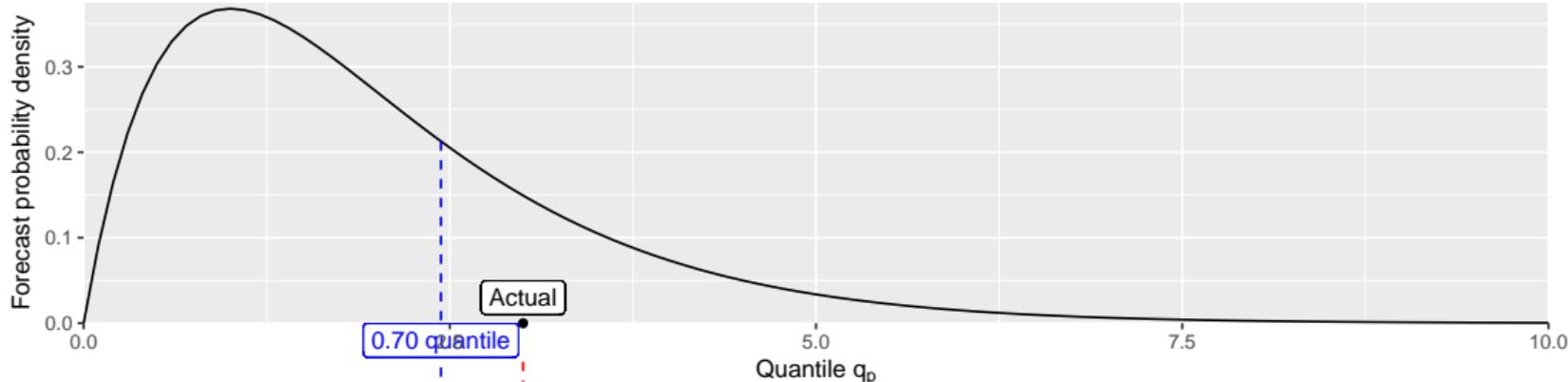
# Evaluating probabilistic forecasts



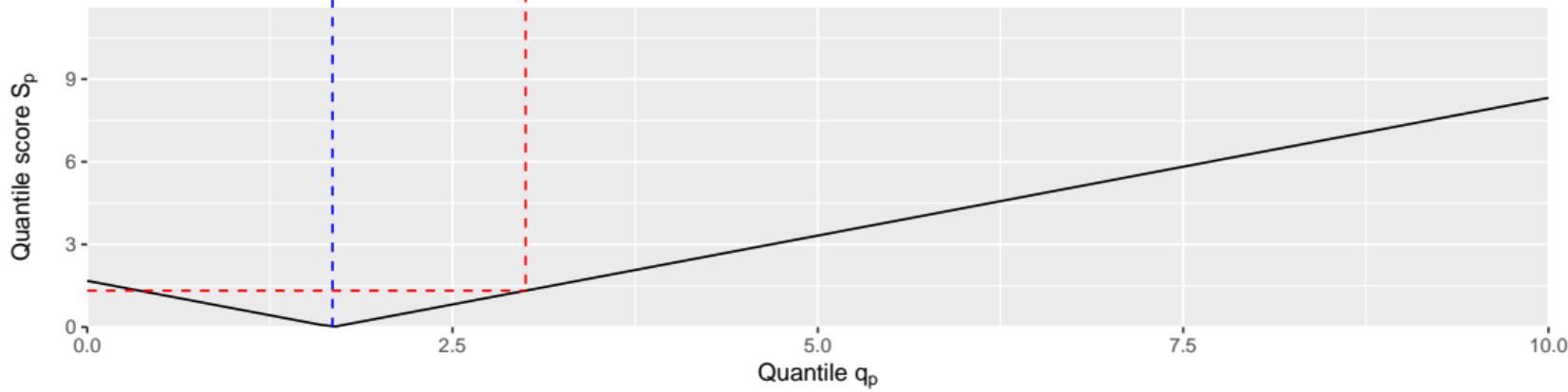
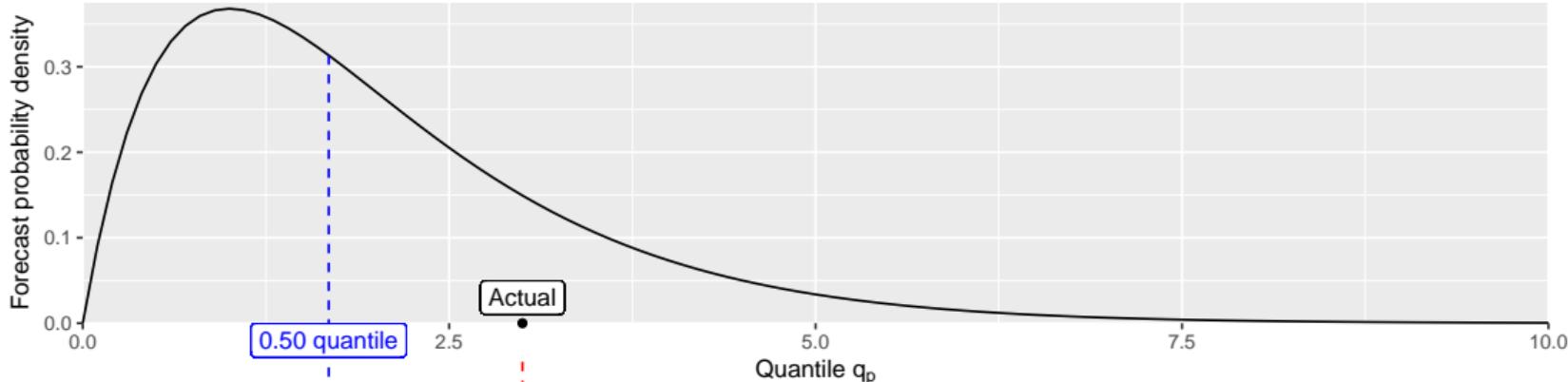
# Evaluating probabilistic forecasts



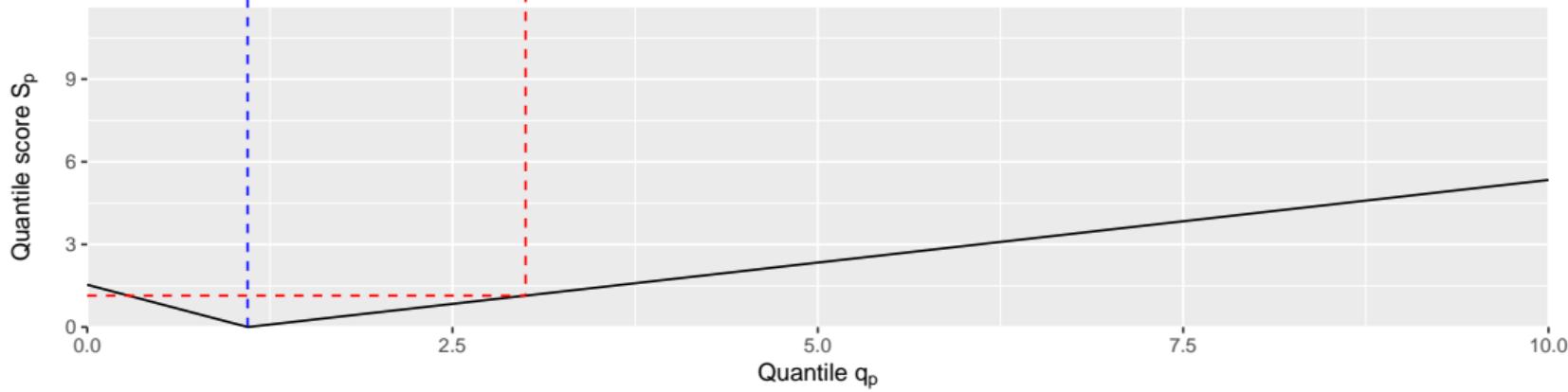
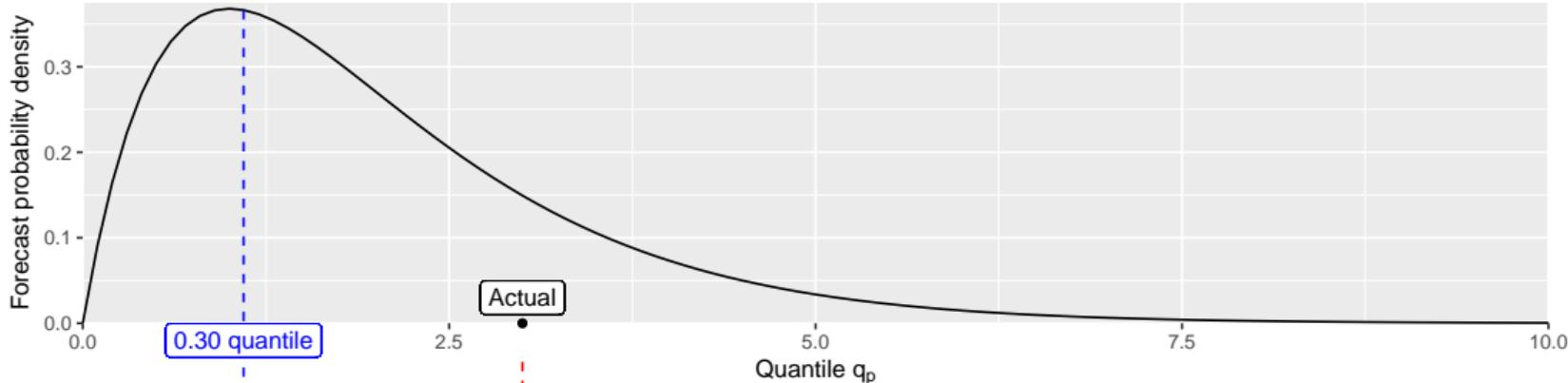
# Evaluating probabilistic forecasts



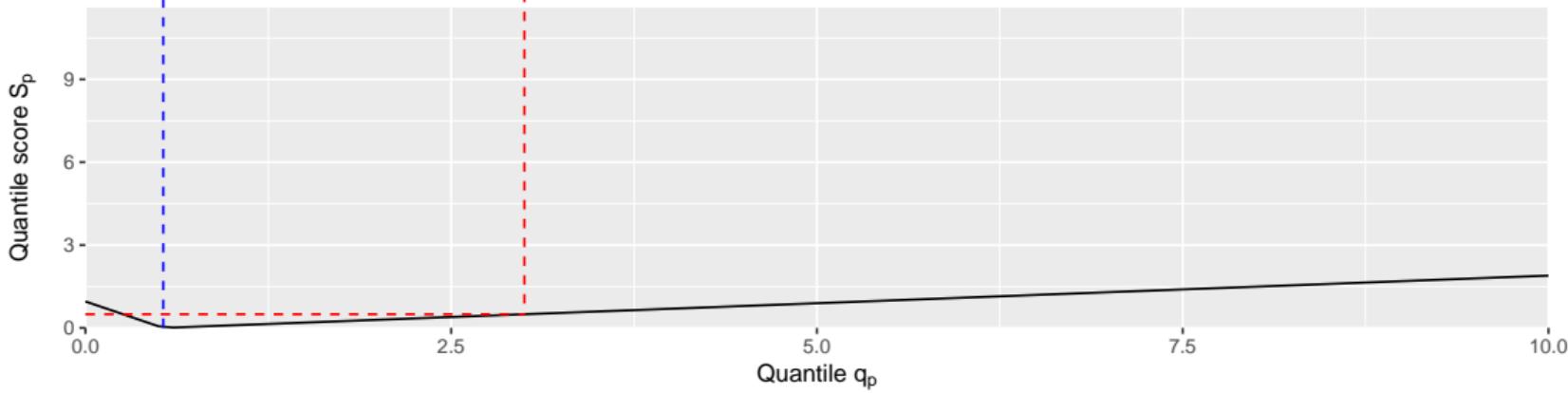
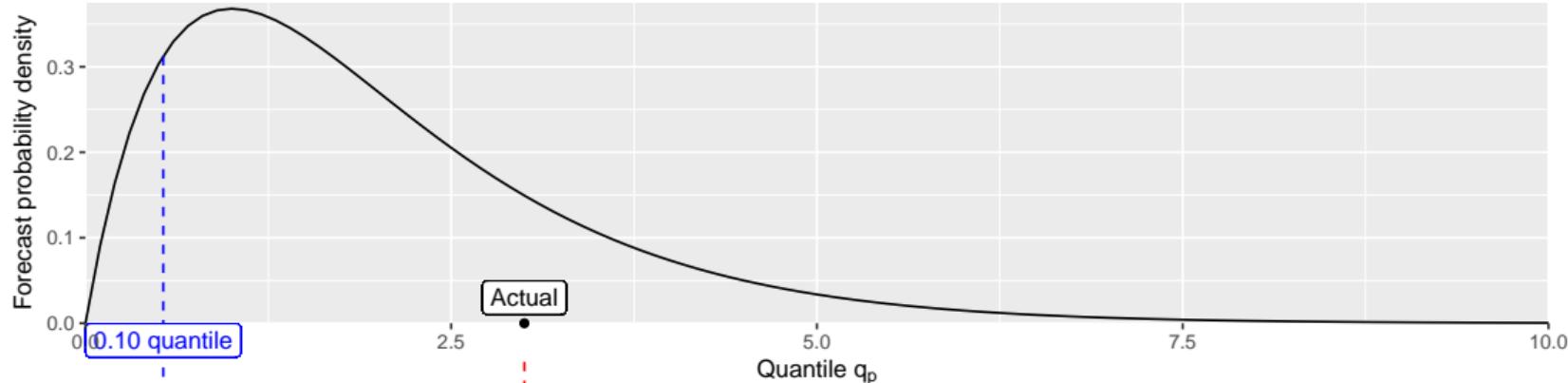
# Evaluating probabilistic forecasts



# Evaluating probabilistic forecasts

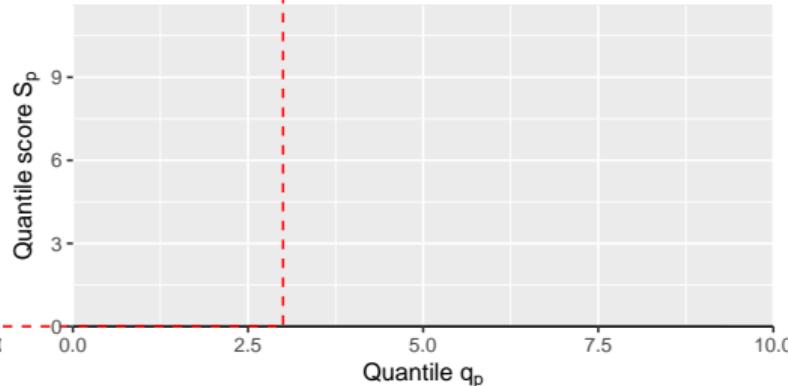
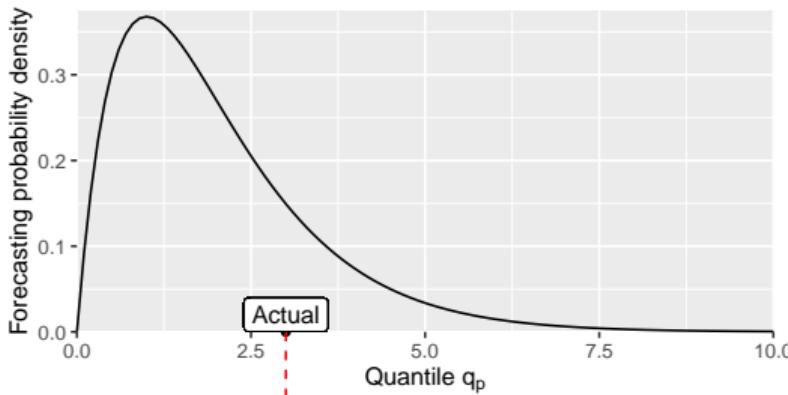
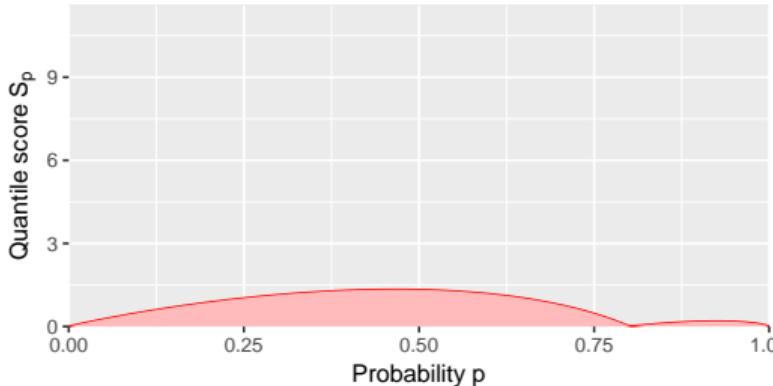


# Evaluating probabilistic forecasts

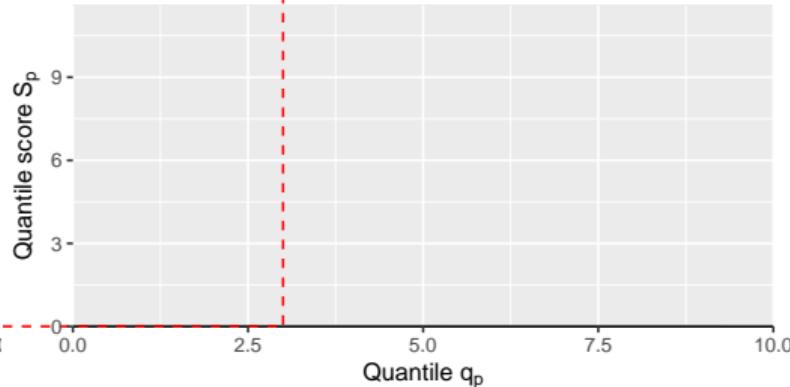
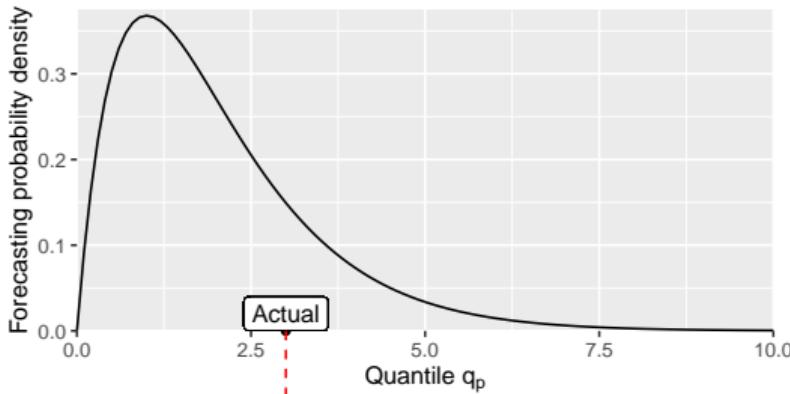
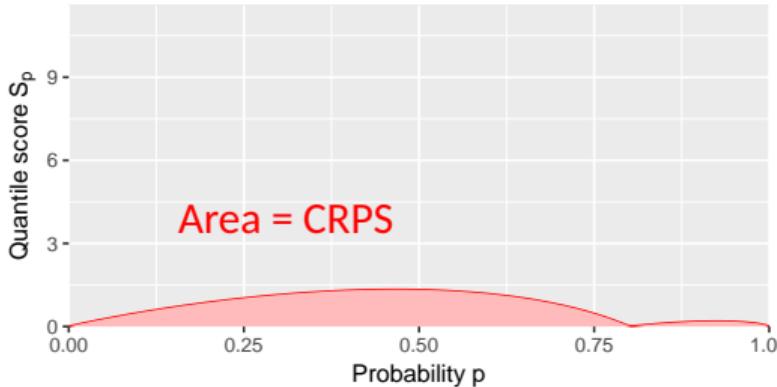


# Evaluating probabilistic forecasts

# Evaluating probabilistic forecasts



# Evaluating probabilistic forecasts



# 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 Ranked Probability Score)

# CRPS: Continuous Ranked Probability Score

$y_t$  = observation at time  $t$

$F_t(u) = \Pr(Y_t \leq u)$  = forecast distribution

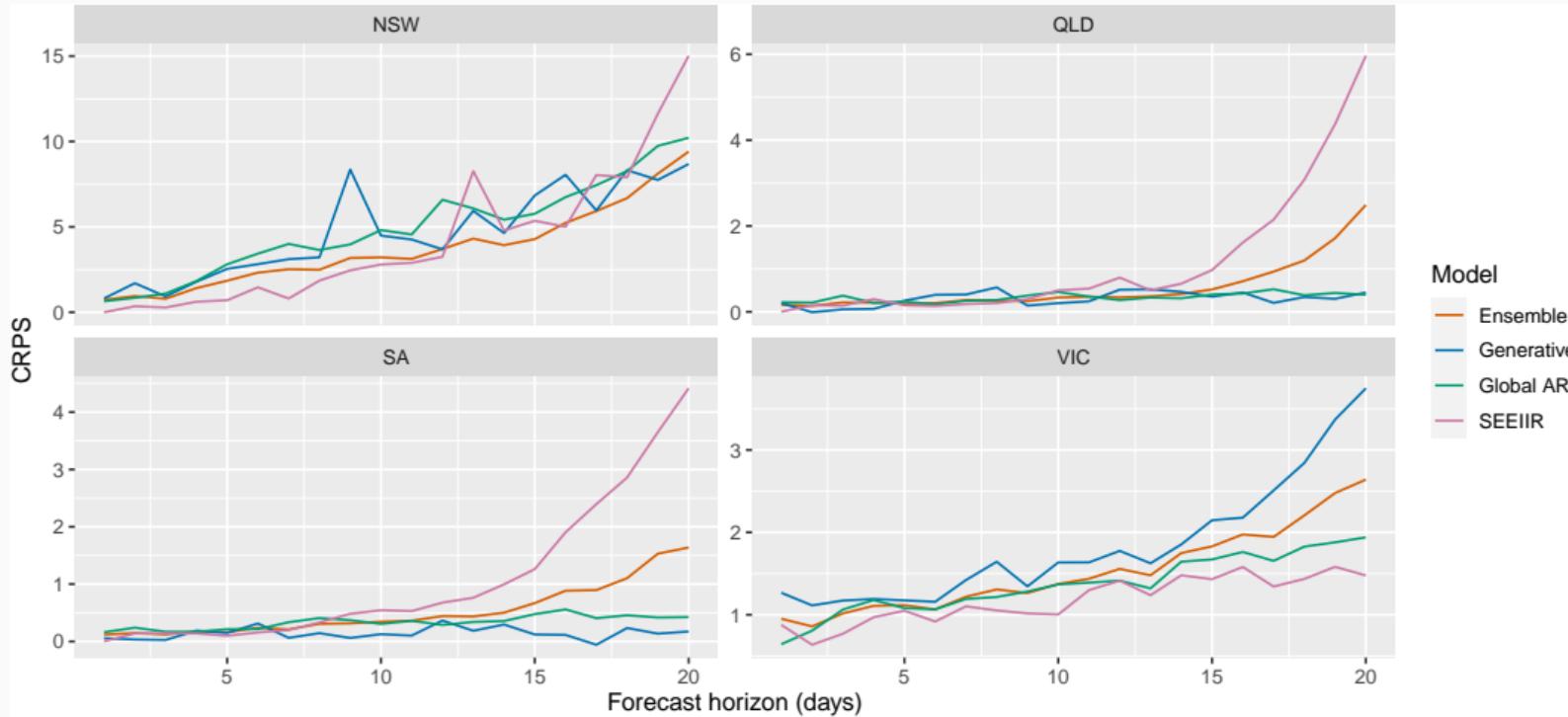
$f_{p,t} = F_t^{-1}(p)$  = quantile forecast with prob.  $p$

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

$Y_t$  and  $Y_t^* \sim \text{iid}$  with distribution  $F_t$ .

$$\begin{aligned}\text{CRPS}_t &= \int_0^1 Q_{p,t} dp \\ &= \int_{-\infty}^{\infty} [F_t(u) - \mathbf{1}_{y_t \leq u}]^2 du \\ &= \mathbb{E}|Y_t - y_t| - \frac{1}{2}\mathbb{E}|Y_t - Y_t^*|\end{aligned}$$

# CRPS: Continuous Ranked Probability Score



For weekly forecasts created from 17 September 2020 to 15 June 2021

# When should we give up?

- When there is insufficient data?
- When the models give implausible forecasts?
- When the forecast uncertainty is too large to assist decision making?

## More information

 [robjhyndman.com](http://robjhyndman.com)

 [@robjhyndman](https://twitter.com/robjhyndman)

 [@robjhyndman](https://github.com/robjhyndman)

 [rob.hyndman@monash.edu](mailto:rob.hyndman@monash.edu)