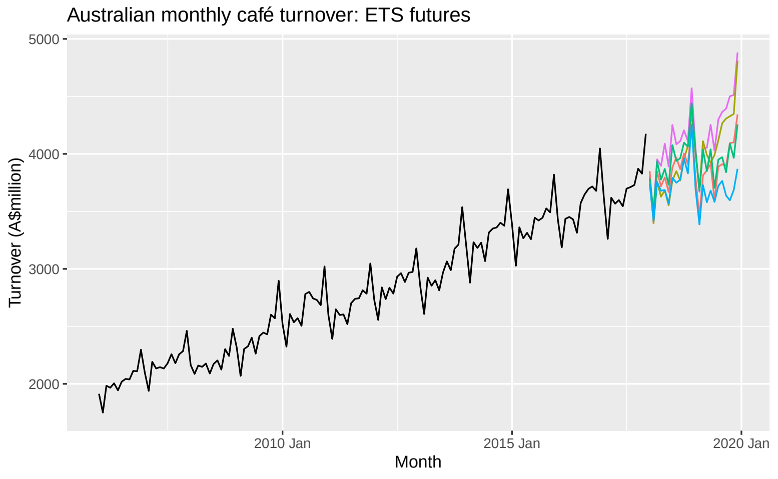
Ensemble forecasts with fable

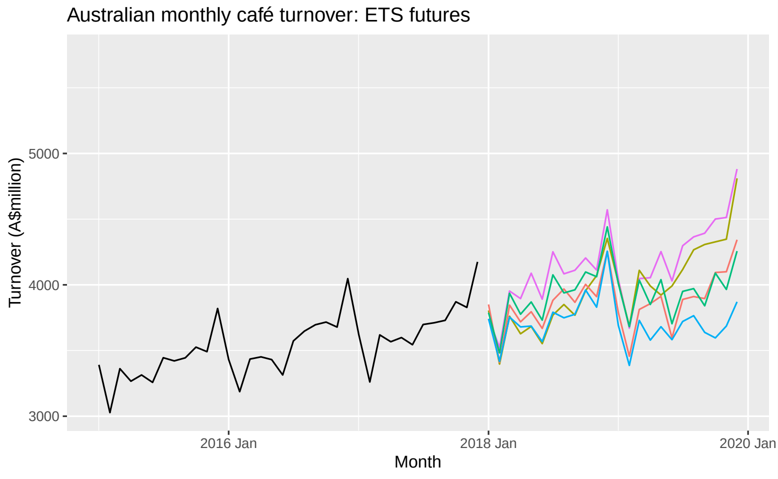
Rob J Hyndman

14 August 2020

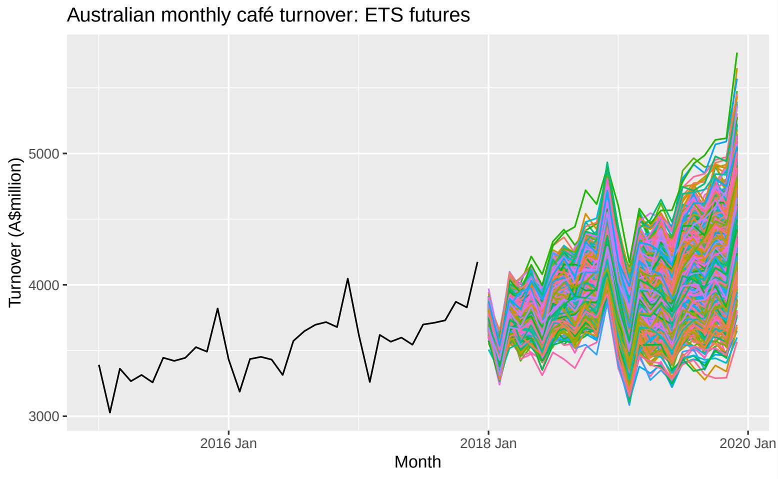
Forecasting using possible futures



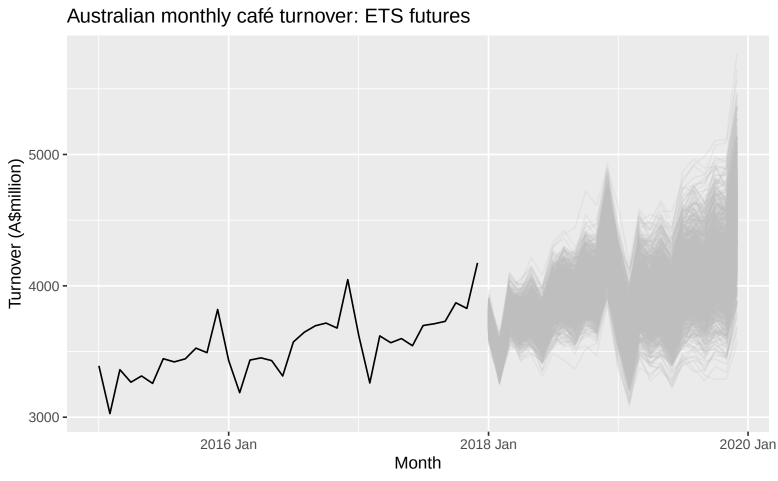
Forecasting using possible futures



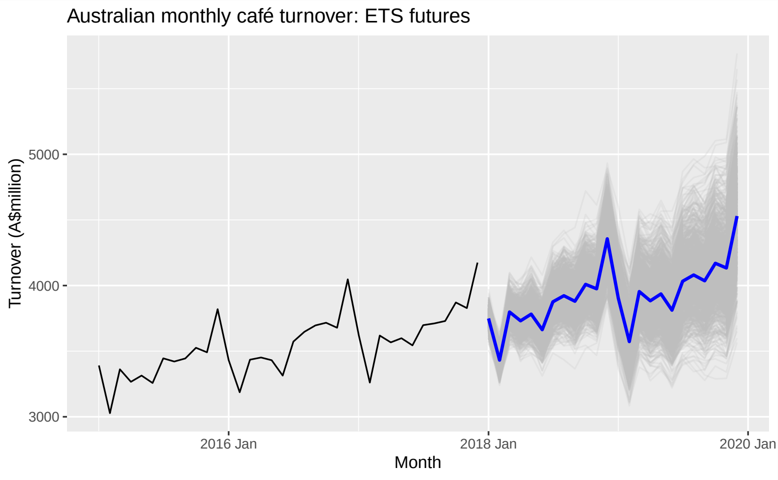
Forecasting using possible futures



Forecasting using possible futures

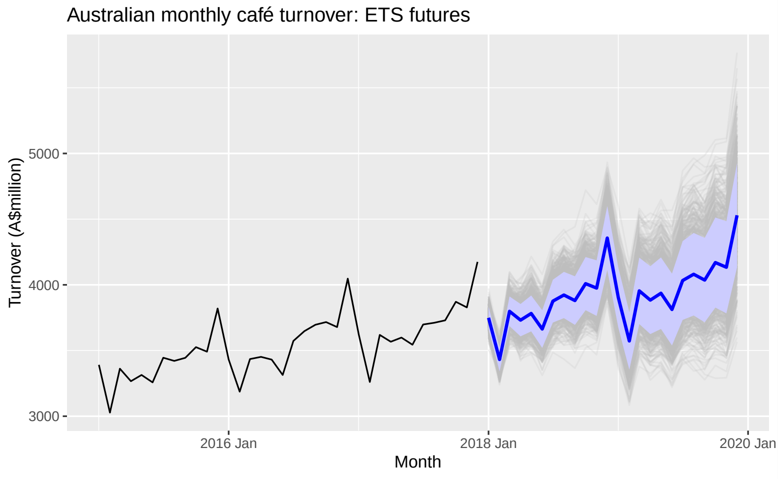


Forecasting using possible futures



Point forecasts: means of the sample paths.

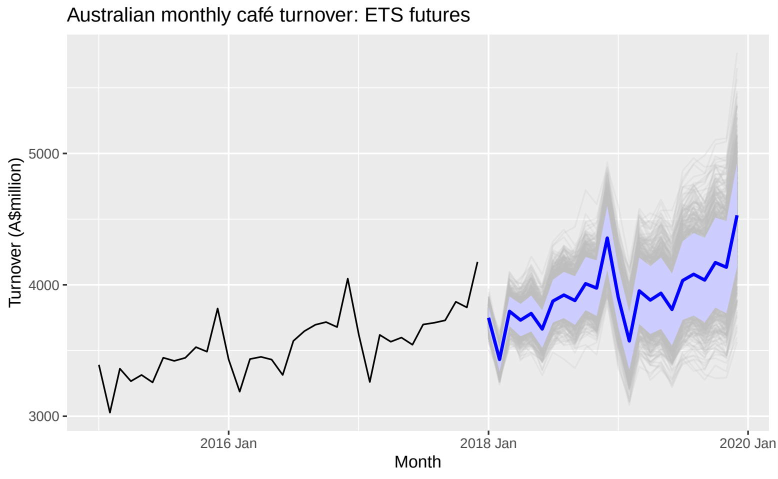
Forecasting using possible futures



Point forecasts: means of the sample paths.

Prediction intervals: middle 80% of the sample paths at each forecast horizon.

Forecasting using possible futures

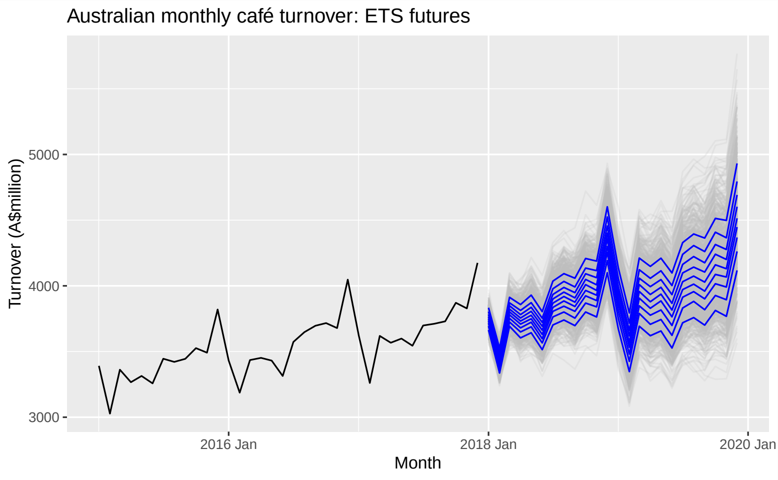


Point forecasts: means of the sample paths.

Prediction intervals: middle 80% of the sample paths at each forecast horizon.

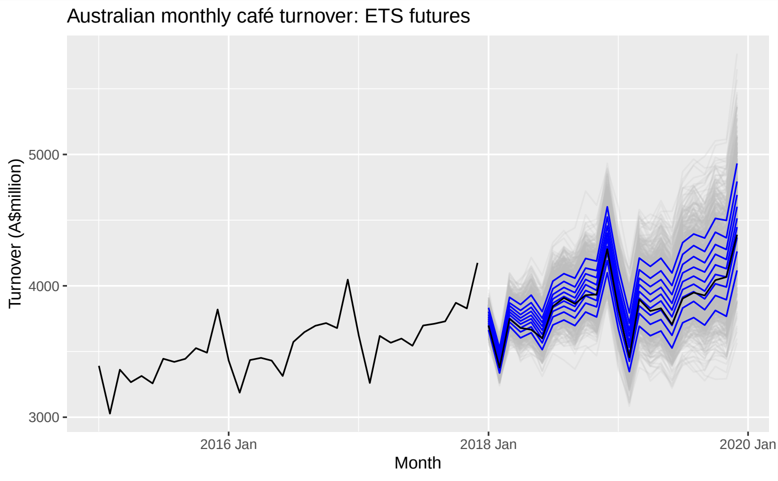
Quantile forecasts: Quantiles of the sample paths at each forecast horizon.

Quantile forecasts



Blue: Deciles for the ETS forecasts for the Australian monthly café turnover.

Quantile forecasts



Blue: Deciles for the ETS forecasts for the Australian monthly café turnover.
Black: Observed values.

Evaluating quantile forecasts

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

y_t = observation at time t

Evaluating quantile forecasts

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

y_t = observation at time t

Quantile score

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

Evaluating quantile forecasts

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

y_t = observation at time t

Quantile score

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

Evaluating quantile forecasts

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

y_t = observation at time t

Quantile score

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

- Low Q_p is good
- Multiplier of 2 often omitted, but useful for interpretation
- Q_p like absolute error (weighted to account for likely exceedance)
- Average Q_p = CRPS (Continuous Rank Probability Score)

Evaluating quantile forecasts

```r
cafe %>%
  filter(year(date) <= 2017)
```

```
## # A tsibble: 144 x 2 [1M]
##         date turnover
##        <mth>    <dbl>
##  1 2006 Jan     1914.
##  2 2006 Feb     1750.
##  3 2006 Mar     1984.
##  4 2006 Apr     1966.
##  5 2006 May     2005.
##  6 2006 Jun     1944.
##  7 2006 Jul     2019.
##  8 2006 Aug     2043.
##  9 2006 Sep     2039.
## 10 2006 Oct     2113.
## # ... with 134 more rows
```

Evaluating quantile forecasts

```
cafe %>%
  filter(year(date) <= 2017) %>%
  model(
    ETS = ETS(turnover),
    ARIMA = ARIMA(turnover ~ pdq(d=1) + PDQ(D=1))
  )
```

```
## # A mable: 1 x 2
##             ETS                          ARIMA
##         <model>                        <model>
## 1 <ETS(M,A,M)> <ARIMA(0,1,1)(0,1,1)[12]>
```

Evaluating quantile forecasts

```r
cafe %>%
  filter(year(date) <= 2017) %>%
  model(
    ETS = ETS(turnover),
    ARIMA = ARIMA(turnover ~ pdq(d=1) + PDQ(D=1))
  ) %>%
  forecast(h = "2 years")
```

```
## # A fable: 48 x 4 [1M]
## # Key:     .model [2]
##    .model     date        turnover .mean
##    <chr>     <mth>          <dist> <dbl>
##  1 ETS    2018 Jan  N(3749, 4324) 3749.
##  2 ETS    2018 Feb  N(3432, 4943) 3432.
##  3 ETS    2018 Mar  N(3799, 7766) 3799.
##  4 ETS    2018 Apr  N(3731, 9229) 3731.
##  5 ETS    2018 May N(3782, 11359) 3782.
##  6 ETS    2018 Jun N(3663, 12505) 3663.
```

Evaluating quantile forecasts

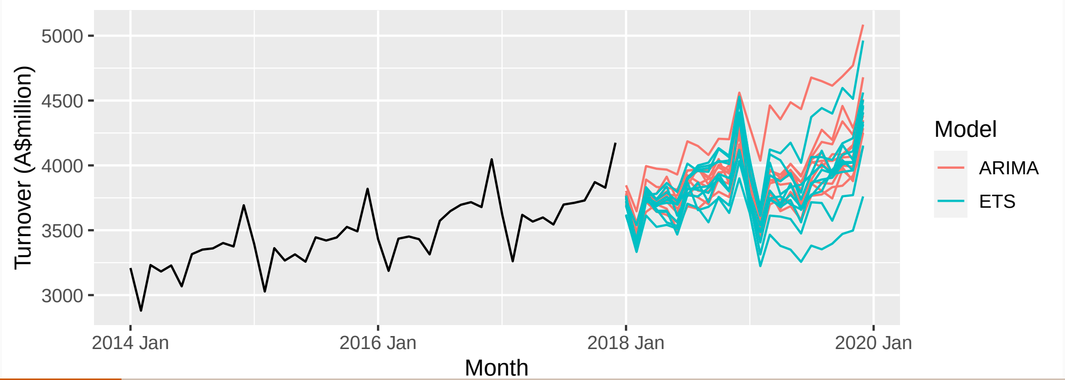
```r
cafe %>%
  filter(year(date) <= 2017) %>%
  model(
    ETS = ETS(turnover),
    ARIMA = ARIMA(turnover ~ pdq(d=1) + PDQ(D=1))
  ) %>%
  forecast(h = "2 years") %>%
  accuracy(cafe, measures = list(CRPS = CRPS))
```

```
## # A tibble: 2 x 3
##   .model .type  CRPS
##   <chr>  <chr> <dbl>
## 1 ARIMA  Test   64.7
## 2 ETS    Test   56.9
```

Ensemble forecasting

Ensemble forecasting involves combining the forecast distributions from multiple models.

- "All models are wrong, but some are useful" (George Box, 1976)
- Allows diverse models to be included, while reducing impact of any specific model.
- Allows uncertainty of model selection to be incorporated.



Ensemble forecasting

```
cafe %>% filter(year(date) <= 2017) %>%
  model(
    ETS = ETS(turnover),
    ARIMA = ARIMA(turnover ~ pdq(d=1) + PDQ(D=1))
  ) %>%
  forecast(h = "2 years")
```

```
## # A fable: 48 x 4 [1M]
## # Key:      .model [2]
##    .model      date        turnover .mean
##    <chr>      <mth>          <dist> <dbl>
##  1 ETS     2018 Jan  N(3749, 4324) 3749.
##  2 ETS     2018 Feb  N(3432, 4943) 3432.
##  3 ETS     2018 Mar  N(3799, 7766) 3799.
##  4 ETS     2018 Apr  N(3731, 9229) 3731.
##  5 ETS     2018 May N(3782, 11359) 3782.
##  6 ETS     2018 Jun N(3663, 12505) 3663.
##  7 ETS     2018 Jul N(3876, 16166) 3876.
```

Ensemble forecasting

```
cafe %>% filter(year(date) <= 2017) %>%
  model(
    ETS = ETS(turnover),
    ARIMA = ARIMA(turnover ~ pdq(d=1) + PDQ(D=1))
  ) %>%
  forecast(h = "2 years") -> fc
```

```
## # A fable: 48 x 4 [1M]
## # Key:      .model [2]
##    .model      date         turnover .mean
##    <chr>      <mth>           <dist> <dbl>
##  1 ETS     2018 Jan  N(3749, 4324) 3749.
##  2 ETS     2018 Feb  N(3432, 4943) 3432.
##  3 ETS     2018 Mar  N(3799, 7766) 3799.
##  4 ETS     2018 Apr  N(3731, 9229) 3731.
##  5 ETS     2018 May N(3782, 11359) 3782.
##  6 ETS     2018 Jun N(3663, 12505) 3663.
##  7 ETS     2018 Jul N(3876, 16166) 3876.
```

Ensemble forecasting

```
fc %>%
  summarise(
    turnover = dist_mixture(turnover[1], turnover[2], weights=c(0.5,0.5)),
    .mean = mean(turnover)
  ) %>%
  as_fable(response = "turnover", distribution = turnover)
```

```
## # A fable: 24 x 3 [1M]
##         date      turnover .mean
##        <mth>        <dist> <dbl>
##  1 2018 Jan mixture(n=2) 3770.
##  2 2018 Feb mixture(n=2) 3457.
##  3 2018 Mar mixture(n=2) 3799.
##  4 2018 Apr mixture(n=2) 3743.
##  5 2018 May mixture(n=2) 3782.
##  6 2018 Jun mixture(n=2) 3681.
##  7 2018 Jul mixture(n=2) 3884.
##  8 2018 Aug mixture(n=2) 3923.
```

Ensemble forecasting

```
fc %>%
  summarise(
    turnover = dist_mixture(turnover[1], turnover[2], weights=c(0.5,0.5)),
    .mean = mean(turnover)
  ) %>%
  as_fable(response = "turnover", distribution = turnover) -> ensemble
```

```
## # A fable: 24 x 3 [1M]
##         date      turnover .mean
##        <mth>        <dist> <dbl>
##  1 2018 Jan mixture(n=2) 3770.
##  2 2018 Feb mixture(n=2) 3457.
##  3 2018 Mar mixture(n=2) 3799.
##  4 2018 Apr mixture(n=2) 3743.
##  5 2018 May mixture(n=2) 3782.
##  6 2018 Jun mixture(n=2) 3681.
##  7 2018 Jul mixture(n=2) 3884.
##  8 2018 Aug mixture(n=2) 3923.
```

Ensemble forecasting

```r
ensemble %>%
  accuracy(cafe, measures = list(CRPS = CRPS))
```

```
## # A tibble: 1 x 2
##   .type  CRPS
##   <chr> <dbl>
## 1 Test   59.8
```

- In this case, the ensemble forecasts are slightly worse than the ETS forecasts.
-

Combination forecasting

Combination forecasting is a related idea that is more widely used in the general forecasting community. This involves taking a weighted average of the forecasts produced from the component models. Often a simple average is used. For more than 50 years we have known that combination forecasting improves forecast accuracy [@Bates1969-dp;@Clemen1989-fz]. One of the reasons for this is that the combination decreases the variance of the forecasts [@Hibon2005-cv] by reducing the uncertainty associated with selecting a particular model.

Combinations are almost always used to produce point forecasts, not

Conclusions

I have described several tools for forecasting that are likely to be increasingly used in business forecasting in the future.

- Simulated future sample paths allow us to study how the future might evolve, and allow us to answer more complicated forecasting questions than is possible with analytical methods.
- Quantile forecasts can be produced from these simulated future sample paths and provide a way of quantifying the forecast distributions.
- Quantile scores allow us to evaluate quantile forecasts. Averaging quantile scores gives the CRPS which allows us to

Supplements

All the forecasts and calculations produced in this chapter were obtained with the `fable` package for R. The code used is available at `https://github.com/robjhyndman/quantile_ensembles`.