# Forecast reconciliation

## 4. Probabilistic forecast reconciliation

Rob J Hyndman

MONASH University

robjhyndman.com/fr2023

Photo by Edvard Alexander Rølvaag on Unsplash

# Outline

# Notation reminder

- Data: $\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t$ where $\boldsymbol{S}$ is a summing matrix and $\boldsymbol{b}_t$ is a vector of disaggregated time series
- Base forecasts: $\hat{\boldsymbol{y}}_{T+h|T}$
- Reconciled forecasts: $\tilde{\boldsymbol{y}}_{T+h|T} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_{T+h|T}$
- MinT: $\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}$ where $\boldsymbol{W}_h$ is covariance matrix of base forecast errors.

# Probabilistic forecasts

- Gaussian
- Non-parametric
- Count

# Coherent probabilistic forecasts

## Coherent probabilistic forecasts

Given the triple $(\mathbb{R}^m, \mathscr{F}_{\mathbb{R}^m}, \nu)$, a coherent probability triple $(\mathfrak{s}, \mathscr{F}_{\mathfrak{s}}, \breve{\nu})$ is such that

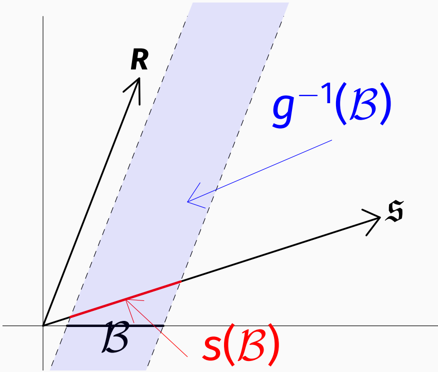
$$\breve{\nu}(s(\mathcal{B})) = \nu(\mathcal{B}) \quad \forall \mathcal{B} \in \mathscr{F}_{\mathbb{R}^m}.$$

## Probabilistic forecast reconciliation

The reconciled probability measure of $\hat{\nu}$ wrt $\psi(.)$ is such that

$$\tilde{\nu}(\mathcal{B}) = \hat{\nu}(\psi^{-1}(\mathcal{B})) \qquad \forall \mathcal{B} \in \mathscr{F}_{\mathfrak{s}},$$

where $\psi^{-1}(\mathcal{B}) := \{\boldsymbol{y} \in \mathbb{R}^n : \psi(\boldsymbol{y}) \in \mathcal{B}\}$ is the



# Construction of reconciled distributions

## Reconciled density of bottom-level

Density of bottom-level series under reconciled distribution is

$$\tilde{f}_{\boldsymbol{b}}(\boldsymbol{b}) = |\boldsymbol{G}^*| \int \hat{f}(\boldsymbol{G}^- \boldsymbol{b} + \boldsymbol{G}_\perp \boldsymbol{a}) d\boldsymbol{a}$$

- $\hat{f}$ is density of incoherent base probabilistic forecast
- $\boldsymbol{G}^-$ is $n \times m$ generalised inverse of $\boldsymbol{G}$ st $\boldsymbol{G}\boldsymbol{G}^- = \boldsymbol{I}$
- $\boldsymbol{G}_\perp$ is $n \times (n-m)$ orthogonal complement to $\boldsymbol{G}$ st $\boldsymbol{G}\boldsymbol{G}_\perp = \boldsymbol{0}$
- $\boldsymbol{G}^* = \left(\boldsymbol{G}^- \vdots \boldsymbol{G}_\perp\right)$, and $\boldsymbol{b}$ and $\boldsymbol{a}$ are obtained via the change of variables $\boldsymbol{y} = \boldsymbol{G}^* \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{pmatrix}$

# Construction of reconciled distributions

## Reconciled density of full hierarchy

Density of full hierarchy under reconciled distribution is

$$\tilde{f}_{\boldsymbol{y}}(\boldsymbol{y}) = |\boldsymbol{S}^*|\tilde{f}_{\boldsymbol{b}}(\boldsymbol{S}^-\boldsymbol{y})\mathbb{1}\{\boldsymbol{y} \in \mathfrak{s}\}$$

- $\boldsymbol{S}^* = \begin{pmatrix} \boldsymbol{S}^- \\ \boldsymbol{S}_\perp' \end{pmatrix}$
- $\boldsymbol{S}^-$ is $m \times n$ generalised inverse of $\boldsymbol{S}$ such that $\boldsymbol{S}^-\boldsymbol{S} = \boldsymbol{I}$,
- $\boldsymbol{S}_\perp$ is $n \times (n - m)$ orthogonal complement to $\boldsymbol{S}$ such that $\boldsymbol{S}_\perp'\boldsymbol{S} = \boldsymbol{0}$.

# Construction of reconciled distributions

## Reconciled density of full hierarchy

Density of full hierarchy under reconciled distribution is

$$\tilde{f}_{\boldsymbol{y}}(\boldsymbol{y}) = |\boldsymbol{S}^*|\tilde{f}_{\boldsymbol{b}}(\boldsymbol{S}^-\boldsymbol{y})\mathbb{1}\{\boldsymbol{y} \in \mathfrak{s}\}$$

- $\boldsymbol{S}^* = \begin{pmatrix} \boldsymbol{S}^- \\ \boldsymbol{S}'_\perp \end{pmatrix}$
- $\boldsymbol{S}^-$ is $m \times n$ generalised inverse of $\boldsymbol{S}$ such that $\boldsymbol{S}^-\boldsymbol{S} = \boldsymbol{I}$,
- $\boldsymbol{S}_\perp$ is $n \times (n-m)$ orthogonal complement to $\boldsymbol{S}$ such that $\boldsymbol{S}'_\perp\boldsymbol{S} = \mathbf{0}$.

## Gaussian reconciliation

If the incoherent base forecasts are $\mathrm{N}(\hat{\mu}, \hat{\Sigma})$,

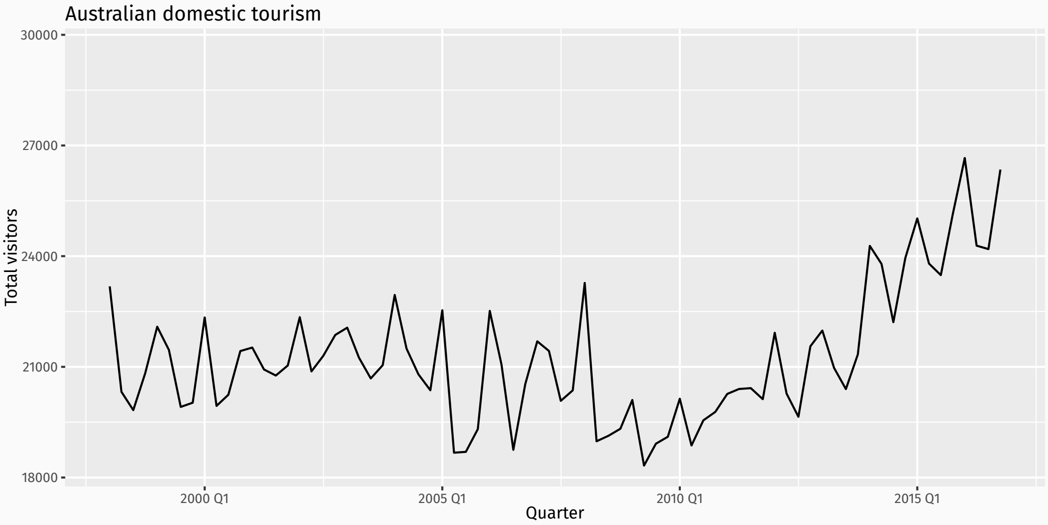
# Simulation from a reconciled distribution

Suppose that $\left(\hat{\boldsymbol{y}}^{[1]}, \dots, \hat{\boldsymbol{y}}^{[L]}\right)$ is a sample drawn from an incoherent probability measure $\hat{\nu}$. Then $\left(\tilde{\boldsymbol{y}}^{[1]}, \dots, \tilde{\boldsymbol{y}}^{[L]}\right)$ where $\tilde{\boldsymbol{y}}^{[\ell]} := \psi(\hat{\boldsymbol{y}}^{[\ell]})$ for $\ell = 1, \dots, L$, is a sample drawn from the reconciled probability measure $\tilde{\nu}$.

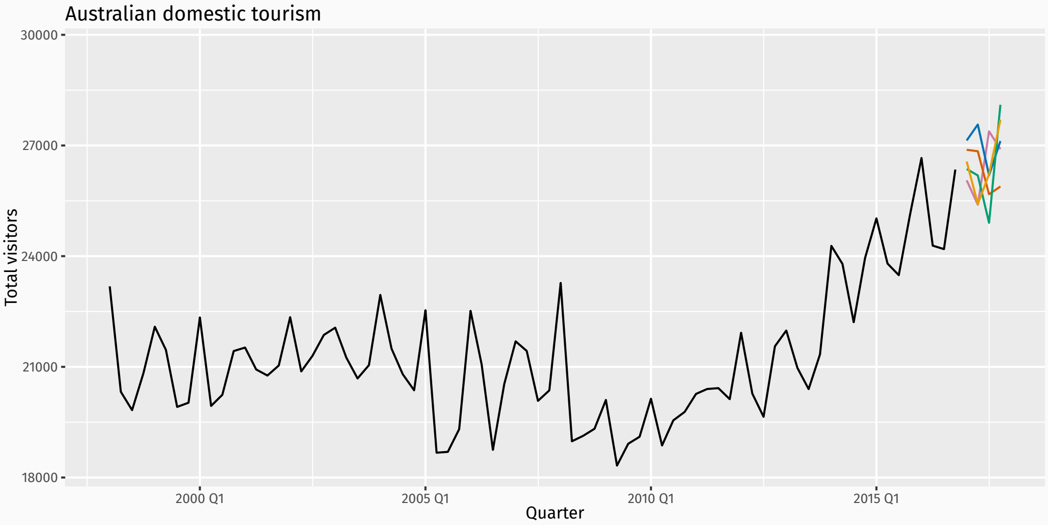
- So reconciling sample paths from incoherent distributions works.

# Outline

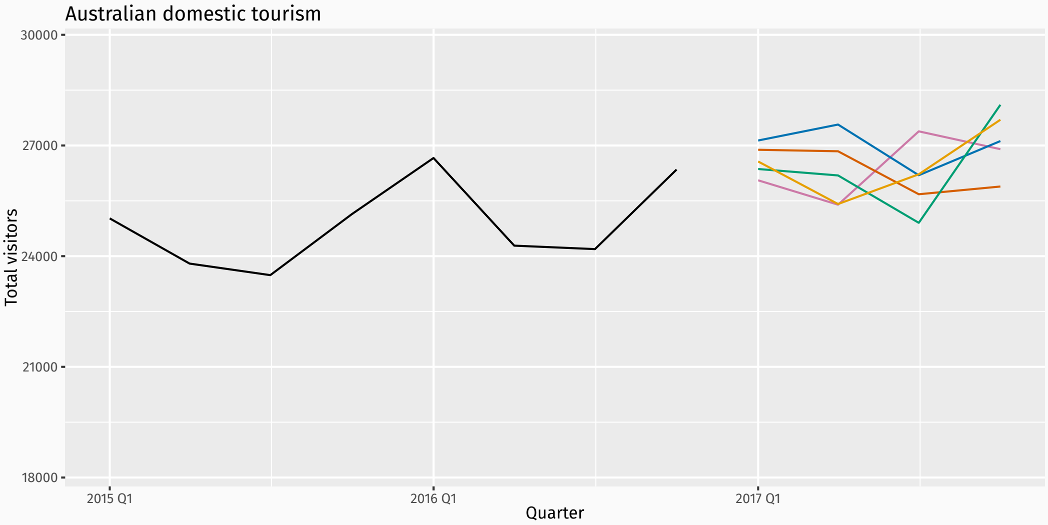
# Evaluating probabilistic forecasts



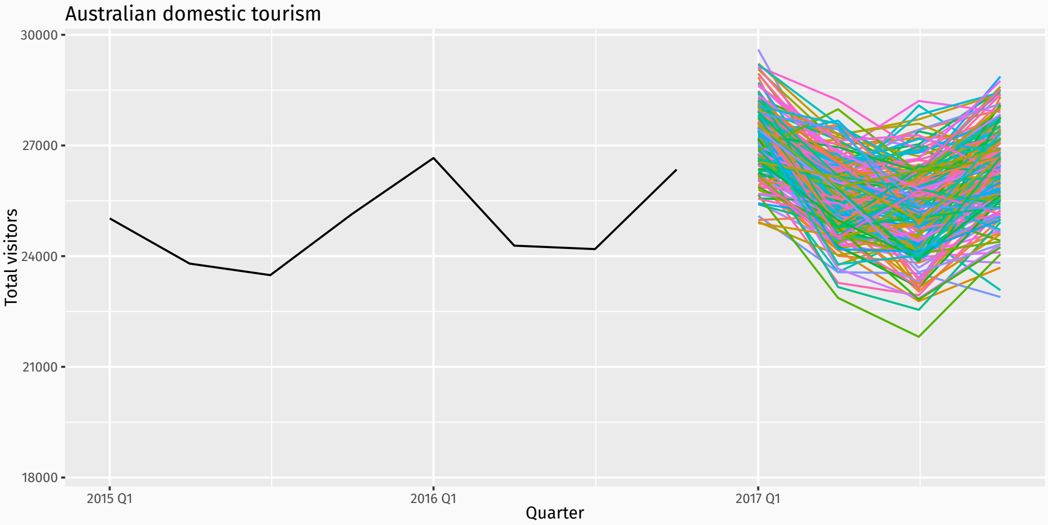
# Evaluating probabilistic forecasts



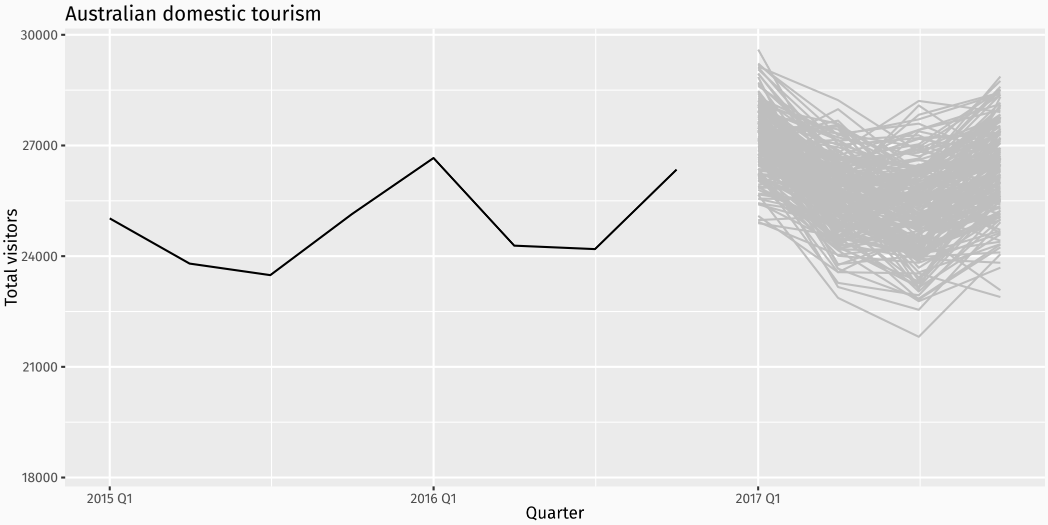
# Evaluating probabilistic forecasts



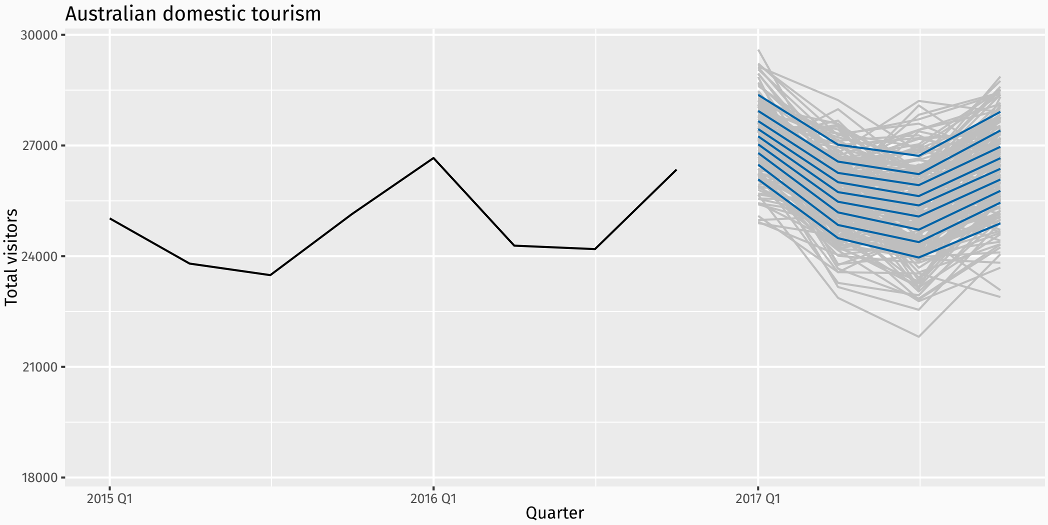
# Evaluating probabilistic forecasts



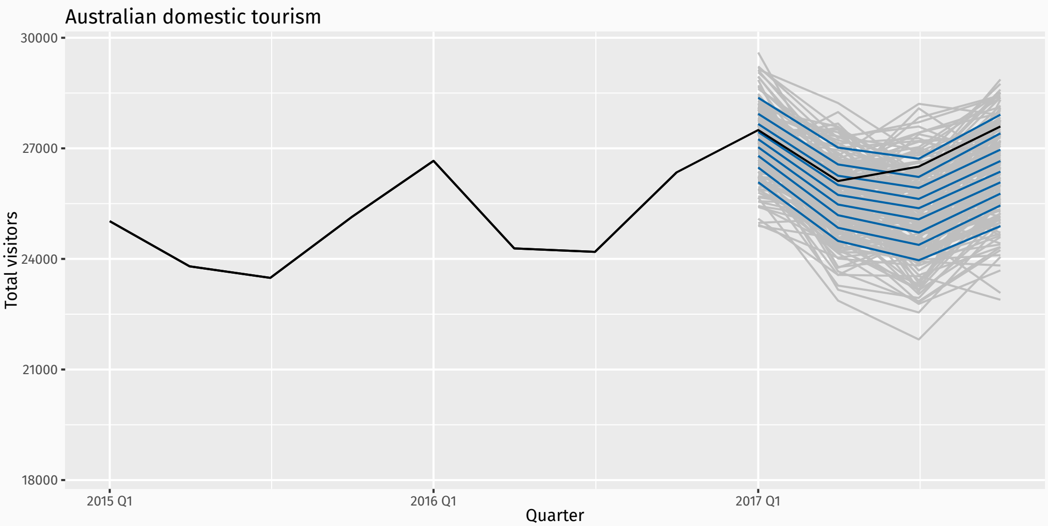
# Evaluating probabilistic forecasts



# Evaluating probabilistic forecasts

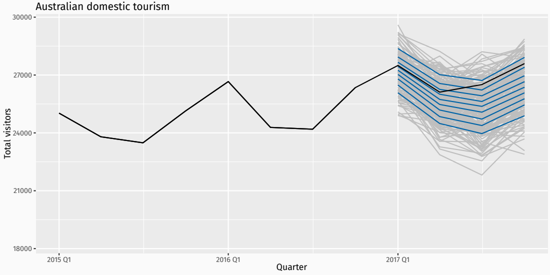


# Evaluating probabilistic forecasts



# Evaluating probabilistic forecasts

$q_{p,t}$ = quantile forecast with prob. $p$ at time $t$.
$y_t$ = observation at time $t$




## Quantile score

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

- Low $S_t$ is good
- Multiplier of 2 often omitted, but useful for interpretation
- $S_t$ like absolute error, weighted to account for likely exceedance
- Average $S_t(p, y)$ over $p$ =

# Evaluating probabilistic forecasts

## Continuous Rank Probability Score (univariate forecsts)

Forecast distribution $F_t$ and observation $y_t$.

$$\text{CRPS}(F_t, y_t) = \int_0^1 S_{p,t}(p, y_t)dp = \text{E}_F|Y - y_t| - \frac{1}{2}\text{E}_F|Y - Y^*|$$

- $Y$ and $Y^*$ are iid draws from $F_t$.
- Optimal when $F_t$ is true distribution (i.e., it is a proper score)

# Evaluating probabilistic forecasts

## Continuous Rank Probability Score (univariate forecsts)

Forecast distribution $F_t$ and observation $y_t$.

$$\text{CRPS}(F_t, y_t) = \int_0^1 S_{p,t}(p, y_t)dp = \text{E}_F|Y - y_t| - \frac{1}{2}\text{E}_F|Y - Y^*|$$

- $Y$ and $Y^*$ are iid draws from $F_t$.
- Optimal when $F_t$ is true distribution (i.e., it is a proper score)

## Energy score (multivariate forecasts)

- $\text{ES}(F_t, \boldsymbol{y}_t) = \text{E}_F||\boldsymbol{Y} - \boldsymbol{y}_t|| - \frac{1}{2}\text{E}_F||\boldsymbol{Y} - \boldsymbol{Y}^*||$

# Evaluating probabilistic forecasts

## Continuous Rank Probability Score (univariate forecsts)

Forecast distribution $F_t$ and observation $y_t$.

$$\text{CRPS}(F_t, y_t) = \int_0^1 S_{p,t}(p, y_t)dp = \text{E}_F|Y - y_t| - \frac{1}{2}\text{E}_F|Y - Y^*|$$

- $Y$ and $Y^*$ are iid draws from $F_t$.
- Optimal when $F_t$ is true distribution (i.e., it is a proper score)

## Energy score (multivariate forecasts)

- $\text{ES}(F_t, \boldsymbol{y}_t) = \text{E}_F||\boldsymbol{Y} - \boldsymbol{y}_t|| - \frac{1}{2}\text{E}_F||\boldsymbol{Y} - \boldsymbol{Y}^*||$

# Evaluating probabilistic forecasts

**Proper scoring rule**

optimized when true forecast distribution is used.

# Evaluating probabilistic forecasts

**Proper scoring rule**

optimized when true forecast distribution is used.

| **Scoring Rule** | **Coherent v Incoherent** | **Coherent v Coherent** |
|---|---|---|
| Log Score | Not proper | • Ordering preserved if compared using bottom-level only |
| Energy Score | Proper | • Full hierarchy should be used. |

# Score optimal reconciliation

Algorithm proposed by Panagiotelis et al (2020) for optimizing $\boldsymbol{G}$ using stochastic gradient descent to optimize Energy Score.

1. Compute base forecasts over a test set.
2. Compute OLS reconciliation: $\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$
3. Iteratively update $\boldsymbol{G}$ using SGD with Adam method and ES objective over a test set

# Outline

# Example: Australian tourism

```
tourism
```

```
# A tsibble: 24,320 x 5 [1Q]
# Key:       Region, State, Purpose [304]
   Quarter Region   State           Purpose  Trips
     <qtr> <chr>    <chr>           <chr>    <dbl>
 1 1998 Q1 Adelaide South Australia Business  135.
 2 1998 Q2 Adelaide South Australia Business  110.
 3 1998 Q3 Adelaide South Australia Business  166.
 4 1998 Q4 Adelaide South Australia Business  127.
 5 1999 Q1 Adelaide South Australia Business  137.
 6 1999 Q2 Adelaide South Australia Business  200.
 7 1999 Q3 Adelaide South Australia Business  169.
 8 1999 Q4 Adelaide South Australia Business  134.
 9 2000 Q1 Adelaide South Australia Business  154.
10 2000 Q2 Adelaide South Australia Business  169.
# i 24,310 more rows
```

# Example: Australian tourism

```r
tourism_agg <- tourism %>%
  aggregate_key(State/Region * Purpose, Trips = sum(Trips))
```

```
# A tsibble: 34,000 x 5 [1Q]
# Key:       State, Purpose, Region [425]
   Quarter State        Purpose      Region        Trips
     <qtr> <chr*>       <chr*>       <chr*>        <dbl>
 1 1998 Q1 <aggregated> <aggregated> <aggregated> 23182.
 2 1998 Q2 <aggregated> <aggregated> <aggregated> 20323.
 3 1998 Q3 <aggregated> <aggregated> <aggregated> 19827.
 4 1998 Q4 <aggregated> <aggregated> <aggregated> 20830.
 5 1999 Q1 <aggregated> <aggregated> <aggregated> 22087.
 6 1999 Q2 <aggregated> <aggregated> <aggregated> 21458.
 7 1999 Q3 <aggregated> <aggregated> <aggregated> 19914.
 8 1999 Q4 <aggregated> <aggregated> <aggregated> 20028.
 9 2000 Q1 <aggregated> <aggregated> <aggregated> 22339.
10 2000 Q2 <aggregated> <aggregated> <aggregated> 19941.
```

# Example: Australian tourism

```
fit <- tourism_agg %>%
  filter(year(Quarter) <= 2015) %>%
  model(ets = ETS(Trips))
```

```
# A mable: 425 x 4
# Key:     State, Purpose, Region [425]
   State  Purpose      Region                 ets
   <chr*> <chr*>       <chr*>             <model>
 1 ACT    Business     Canberra      <ETS(M,N,M)>
 2 ACT    Business     <aggregated>  <ETS(M,N,M)>
 3 ACT    Holiday      Canberra      <ETS(M,N,A)>
 4 ACT    Holiday      <aggregated>  <ETS(M,N,A)>
 5 ACT    Other        Canberra      <ETS(M,N,N)>
 6 ACT    Other        <aggregated>  <ETS(M,N,N)>
 7 ACT    Visiting     Canberra      <ETS(A,N,N)>
 8 ACT    Visiting     <aggregated>  <ETS(A,N,N)>
 9 ACT    <aggregated> Canberra      <ETS(A,N,N)>
```

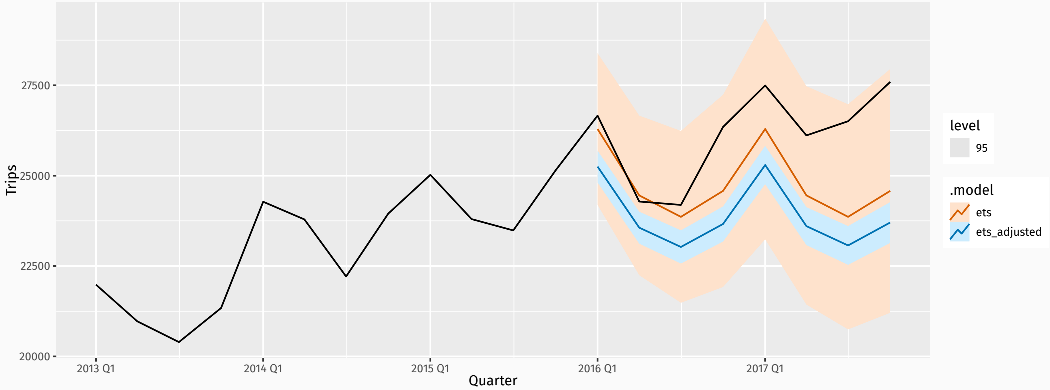
# Example: Australian tourism

```r
fc <- fit %>%
  reconcile(ets_adjusted = min_trace(ets)) %>%
  forecast(h = "2 years")
```

```
# A fable: 6,800 x 7 [1Q]
# Key:     State, Purpose, Region, .model [850]
   State  Purpose  Region   .model       Quarter        Trips .mean
   <chr*> <chr*>   <chr*>   <chr>          <qtr>       <dist> <dbl>
 1 ACT    Business Canberra ets          2016 Q1  N(111, 669)  111.
 2 ACT    Business Canberra ets          2016 Q2 N(156, 1312)  156.
 3 ACT    Business Canberra ets          2016 Q3 N(156, 1320)  156.
 4 ACT    Business Canberra ets          2016 Q4 N(152, 1248)  152.
 5 ACT    Business Canberra ets          2017 Q1  N(111, 669)  111.
 6 ACT    Business Canberra ets          2017 Q2 N(156, 1312)  156.
 7 ACT    Business Canberra ets          2017 Q3 N(156, 1320)  156.
 8 ACT    Business Canberra ets          2017 Q4 N(152, 1248)  152.
 9 ACT    Business Canberra ets_adjusted 2016 Q1  N(116, 328)  116.
```

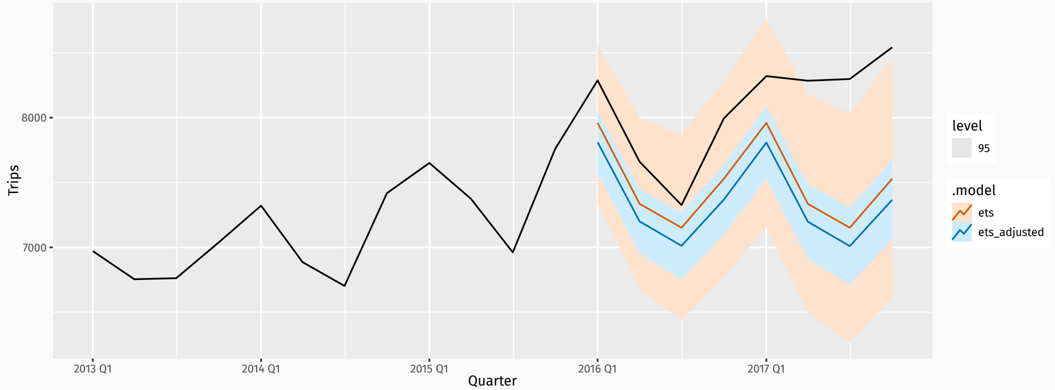
# Example: Australian tourism

```r
fc %>%
  filter(is_aggregated(State) & is_aggregated(Purpose)) %>%
  autoplot(filter(tourism_agg, year(Quarter) > 2012), level = 95)
```



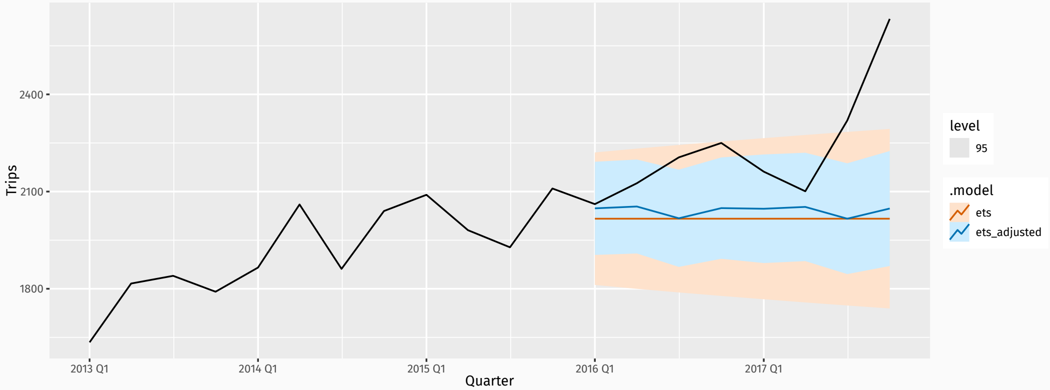
# Example: Australian tourism

```r
fc %>%
  filter(State == "New South Wales" & is_aggregated(Region) & is_aggregated(Purpose)
  autoplot(filter(tourism_agg, year(Quarter) > 2012), level = 95)
```



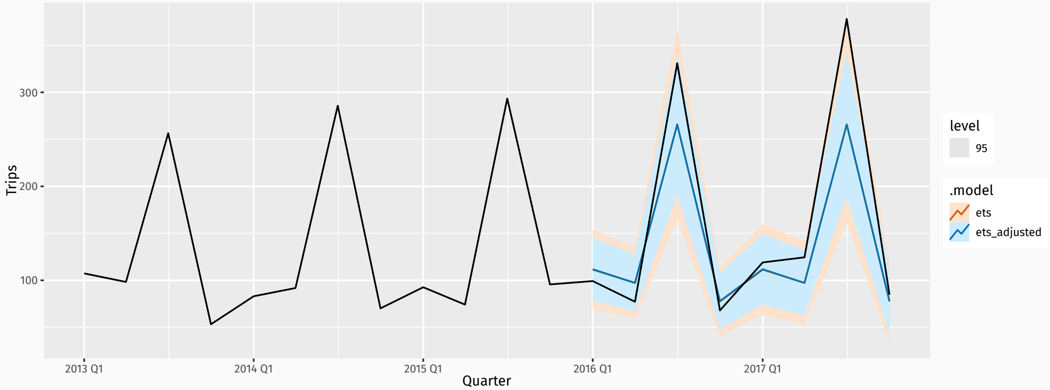
# Example: Australian tourism

```
fc %>%
  filter(Region == "Melbourne" & is_aggregated(Purpose)) %>%
  autoplot(filter(tourism_agg, year(Quarter) > 2012), level = 95)
```



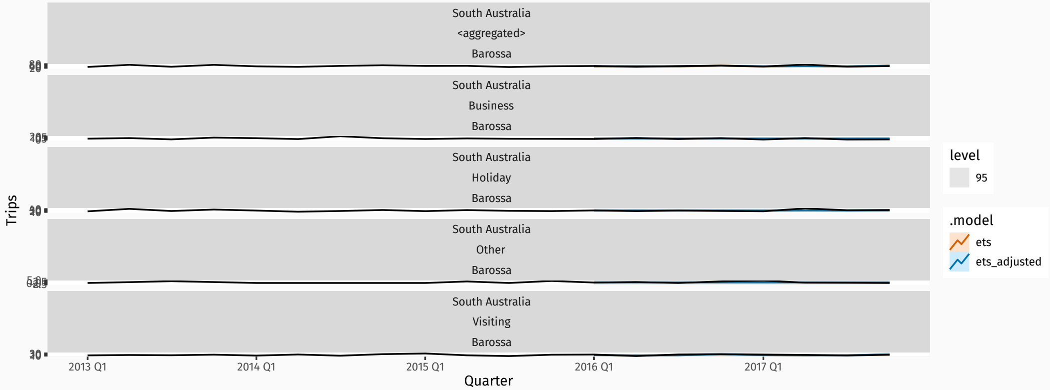
# Example: Australian tourism

```
fc %>%
  filter(Region == "Snowy Mountains", Purpose == "Holiday") %>%
  autoplot(filter(tourism_agg, year(Quarter) > 2012), level = 95)
```



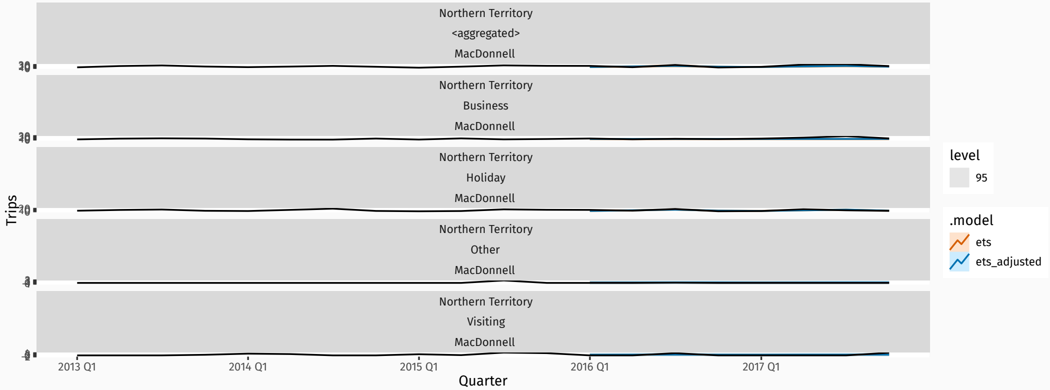
# Example: Australian tourism

```r
fc %>%
  filter(Region == "Barossa") %>%
  autoplot(filter(tourism_agg, year(Quarter) > 2012), level = 95)
```



# Example: Australian tourism

```r
fc %>%
  filter(Region == "MacDonnell") %>%
  autoplot(filter(tourism_agg, year(Quarter) > 2012), level = 95)
```



# Example: Australian tourism

```r
fc <- tourism_agg %>%
  filter(year(Quarter) <= 2015) %>%
  model(
    ets = ETS(Trips),
    arima = ARIMA(Trips)
  ) %>%
  mutate(
    comb = (ets + arima) / 2
  ) %>%
  reconcile(
    ets_adj = min_trace(ets),
    arima_adj = min_trace(arima),
    comb_adj = min_trace(comb)
  ) %>%
  forecast(h = "2 years")
```

# Example: Australian tourism

```r
fc %>%
  accuracy(data = tourism_agg,
           measures = list(crps = CRPS, ss=skill_score(CRPS)))
```

```
# A tibble: 2,550 x 7
   .model State  Purpose      Region       .type  crps      ss
   <chr>  <chr*> <chr*>       <chr*>       <chr> <dbl>   <dbl>
 1 arima  ACT    Business     Canberra     Test   25.2   0.229
 2 arima  ACT    Business     <aggregated> Test   25.2   0.229
 3 arima  ACT    Holiday      Canberra     Test   33.5   0.103
 4 arima  ACT    Holiday      <aggregated> Test   33.5   0.103
 5 arima  ACT    Other        Canberra     Test    9.97  0.0684
 6 arima  ACT    Other        <aggregated> Test    9.97  0.0684
 7 arima  ACT    Visiting     Canberra     Test   34.7  -0.0985
 8 arima  ACT    Visiting     <aggregated> Test   34.7  -0.0985
 9 arima  ACT    <aggregated> Canberra     Test  106.   -0.633
10 arima  ACT    <aggregated> <aggregated> Test  106.   -0.633
```

# Example: Australian tourism

```r
fc %>%
  accuracy(tourism_agg,
           measures = list(crps = CRPS, ss=skill_score(CRPS))) %>%
  group_by(.model) %>%
  summarise(sspc = mean(ss) * 100) %>%
  arrange(sspc)
```

```
# A tibble: 6 x 2
  .model     sspc
  <chr>     <dbl>
1 arima      4.05
2 arima_adj  4.16
3 comb_adj   8.69
4 comb       8.93
5 ets_adj    9.41
6 ets        9.45
```

# Outline

1. Evaluating probabilistic forecasts
2. Example: Australian tourism
3. Example: Australian electricity generation
4. Example: Australian electricity generation
5. Emergency Services Demand
6. Bayesian versions

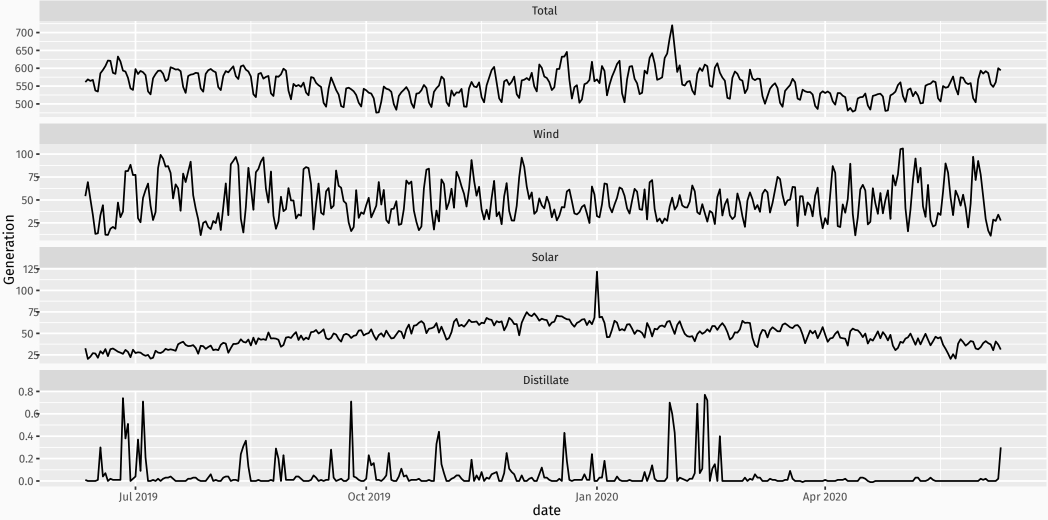
# Example: Australian electricity generation

## Daily time series from opennem.org.au

1. Total = Renewable + Non-renewable
2. Renewable = Batteries + Hydro + Solar + Wind + Biomass
   Non-Renewable = Coal + Gas + Distillate
3. Battery = Battery (Discharging) + Battery (Charging)
   Solar = Solar (Rooftop) + Solar (Utility)
   Coal = Black Coal + Brown Coal
   Gas = Gas (OCGT) + Gas (CCGT) + Gas (Steam) + Gas (Recip)

$n$ = 23 series; $m$ = 15 bottom-level series.

# Example: Australian electricity generation

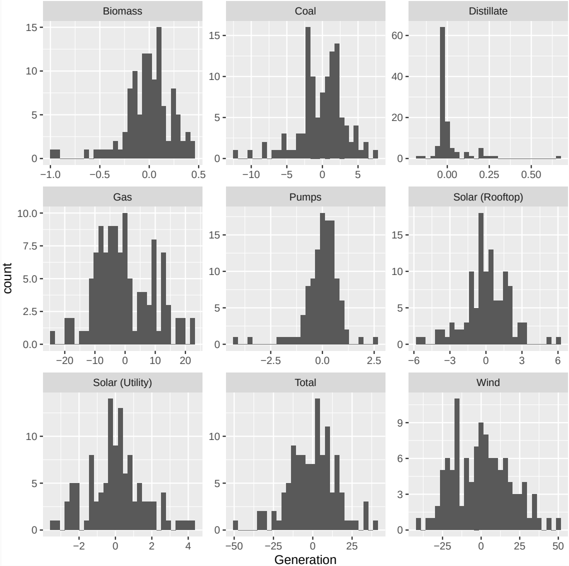


# Example: Australian electricity generation

## Forecast evaluation

- Rolling window of 140 days training data, and one-step-forecasts for 170 days test data.
- One-layer feed-forward neural network with up to 28 lags of target variable as inputs.
- Implemented using `NNETAR()` function in `fable` package.
- Model could be improved with temperature predictor.

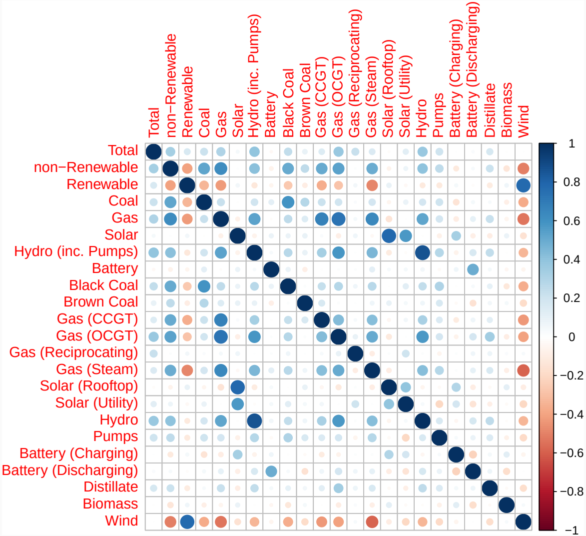
# Example: Australian electricity generation




**Histogram of residuals: 2 Oct 2019 – 21 Jan 2020**

Clearly non-Gaussian

# Example: Australian electricity generation



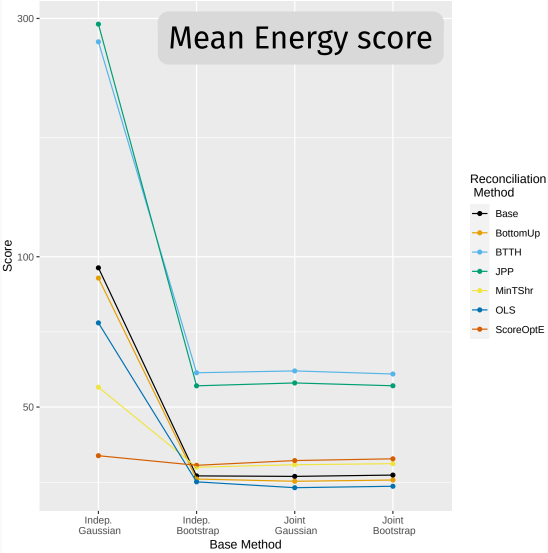

## Correlations of residuals: 2 Oct 2019 – 21 Jan 2020

Blue = positive correlation.

Red = negative correlation.

Large = stronger correlations.

# Example: Australian electricity generation



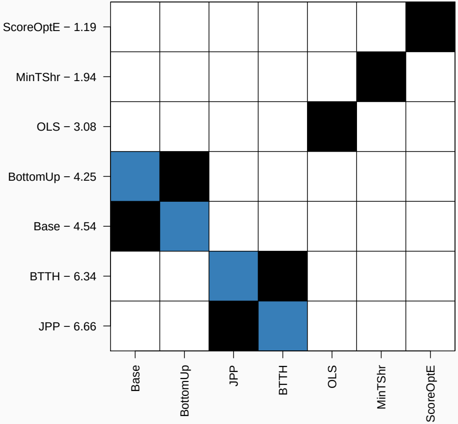

## Base residual assumptions

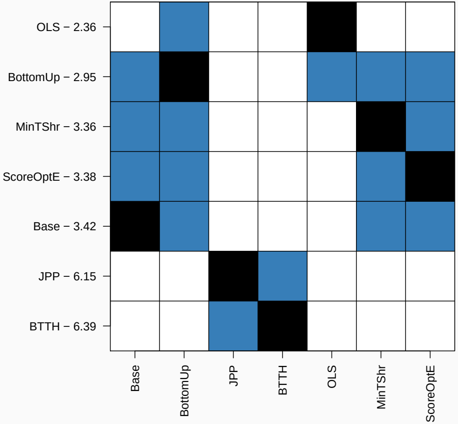
- Gaussian independent
- Gaussian dependent
- Non-Gaussian independent
- Non-Gaussian dependent

## Reconciliation methods

- Base
- BottomUp
- BTTH: Ben Taieb, Taylor, Hyndman
- JPP: Jeon, Panagiotelis, Petropoulos
- OLS

# Example: Australian electricity generation




## Nemenyi test for different scores

Base forecasts are independent and

## Nemenyi test for different scores

Base forecasts are obtained by jointly

# Probabilistic forecast reconciliation

## Key papers

- Ben Taieb, Taylor, Hyndman (*ICML*, 2017)
- Jeon, Panagiotelis, Petropoulos (*EJOR*, 2019)
- Ben Taieb, Taylor, Hyndman (*JASA*, 2020)
- Panagiotelis, Gamakumara, Athanasopoulos, Hyndman (2020). robjhyndman.com/publications/coherentprob/

# Probabilistic forecast reconciliation

## Key papers

- Ben Taieb, Taylor, Hyndman (*ICML*, 2017)
- Jeon, Panagiotelis, Petropoulos (*EJOR*, 2019)
- Ben Taieb, Taylor, Hyndman (*JASA*, 2020)
- Panagiotelis, Gamakumara, Athanasopoulos, Hyndman (2020). robjhyndman.com/publications/coherentprob/

- The reconciled multivariate density must lie on the coherent subspace.
- The univariate density at each node is a convolution of the densities of its children.

# Construction of reconciled distributions

## Reconciled density of bottom-level

Density of bottom-level series under reconciled distribution is

$$\tilde{f}_{\boldsymbol{b}}(\boldsymbol{b}) = |\boldsymbol{G}^{*}| \int \hat{f}(\boldsymbol{G}^{-}\boldsymbol{b} + \boldsymbol{G}_{\perp}\boldsymbol{a})d\boldsymbol{a}$$

- $\hat{f}$ is density of incoherent base probabilistic forecast
- $\boldsymbol{G}^{-}$ is $n \times m$ generalised inverse of $\boldsymbol{G}$ st $\boldsymbol{G}\boldsymbol{G}^{-} = \boldsymbol{I}$
- $\boldsymbol{G}_{\perp}$ is $n \times (n-m)$ orthogonal complement to $\boldsymbol{G}$ st $\boldsymbol{G}\boldsymbol{G}_{\perp} = \boldsymbol{0}$
- $\boldsymbol{G}^{*} = \left(\boldsymbol{G}^{-} \vdots \boldsymbol{G}_{\perp}\right)$, and $\boldsymbol{b}$ and $\boldsymbol{a}$ are obtained via the change of variables $\boldsymbol{y} = \boldsymbol{G}^{*}\begin{pmatrix}\boldsymbol{b}\\ \boldsymbol{a}\end{pmatrix}$

# Construction of reconciled distributions

## Reconciled density of full hierarchy

Density of full hierarchy under reconciled distribution is

$$\tilde{f}_{\boldsymbol{y}}(\boldsymbol{y}) = |\boldsymbol{S}^*|\,\tilde{f}_{\boldsymbol{b}}(\boldsymbol{S}^-\boldsymbol{y})\,\mathbb{1}\{\boldsymbol{y} \in \mathfrak{s}\}$$

- $\boldsymbol{S}^* = \begin{pmatrix}\boldsymbol{S}^{-\prime} & \boldsymbol{S}_\perp\end{pmatrix}'$
- $\boldsymbol{S}^-$ is $m \times n$ generalised inverse of $\boldsymbol{S}$ such that $\boldsymbol{S}^-\boldsymbol{S} = \boldsymbol{I}$,
- $\boldsymbol{S}_\perp$ is $n \times (n-m)$ orthogonal complement to $\boldsymbol{S}$ such that

## Gaussian reconciliation

If the incoherent base forecasts are $\text{N}(\hat{\mu}, \hat{\Sigma})$, then the reconciled density is $\text{N}(\boldsymbol{S}\boldsymbol{G}\hat{\mu}, \boldsymbol{S}\boldsymbol{G}\hat{\Sigma}\boldsymbol{G}'\boldsymbol{S}')$.

## Bootstrap reconciliation

Reconciling sample paths from incoherent distributions works.

# Evaluating probabilistic forecasts

**Proper scoring rule**

optimized when true forecast distribution is used.

# Evaluating probabilistic forecasts

**Proper scoring rule**

optimized when true forecast distribution is used.

| Scoring Rule | Coherent v Incoherent | Coherent v Coherent |
|---|---|---|
| Log Score | Not proper | • Ordering preserved if compared using bottom-level only |
| Energy Score | Proper | • Full hierarchy should be used.<br>• Rankings may |

# Score optimal reconciliation

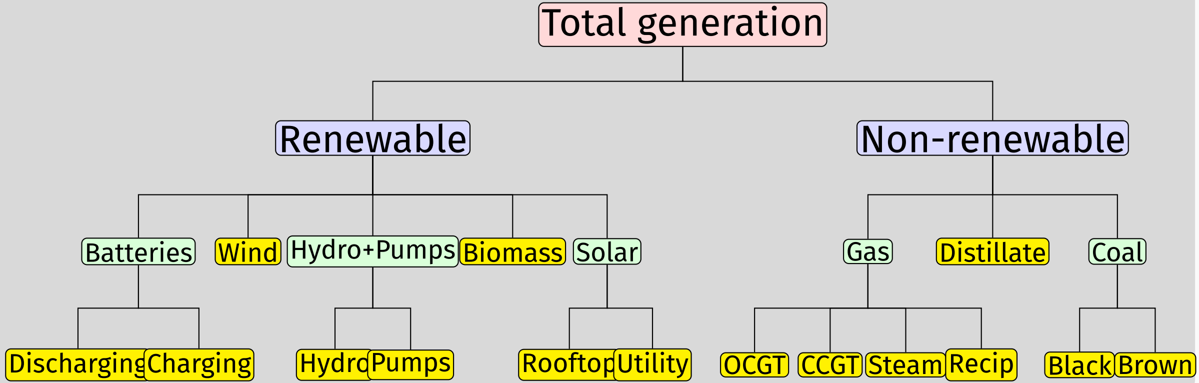
Algorithm proposed by Panagiotelis et al (2020) for optimizing $\boldsymbol{G}$ using stochastic gradient descent to optimize Energy Score.

1. Compute base forecasts over a test set.
2. Compute OLS reconciliation: $\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$
3. Iteratively update $\boldsymbol{G}$ using SGD with Adam method and ES objective over a test set

# Outline

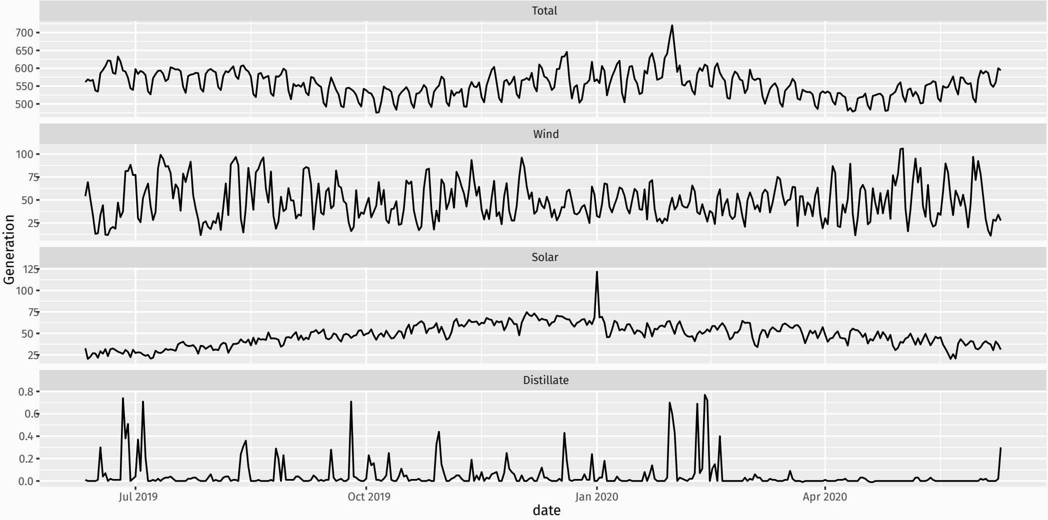
# Example: Australian electricity generation

## Daily time series from opennem.org.au




$n$ = 23 series $\quad$ $m$ = 15 bottom-level series

# Example: Australian electricity generation

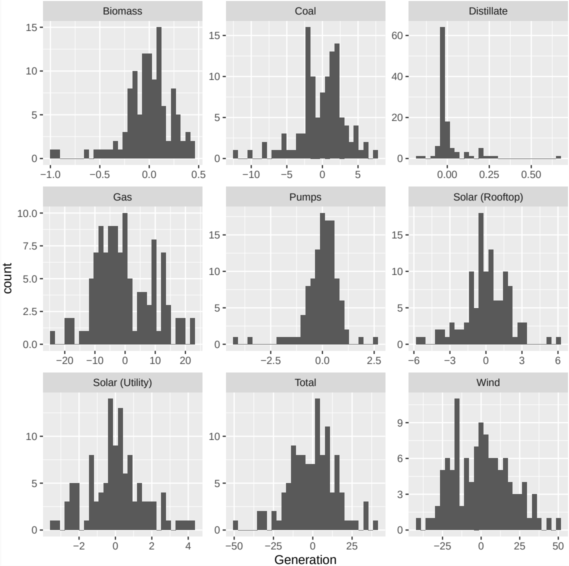


# Example: Australian electricity generation

## Forecast evaluation

- Rolling window of 140 days training data, and one-step-forecasts for 170 days test data.
- One-layer feed-forward neural network with up to 28 lags of target variable as inputs.
- Implemented using `NNETAR()` function in `fable` package.
- Model could be improved with temperature predictor.

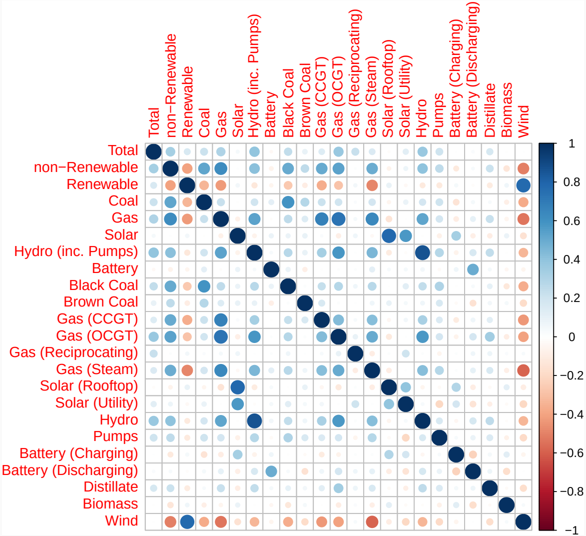
# Example: Australian electricity generation




**Histogram of residuals: 2 Oct 2019 – 21 Jan 2020**

Clearly non-Gaussian

# Example: Australian electricity generation



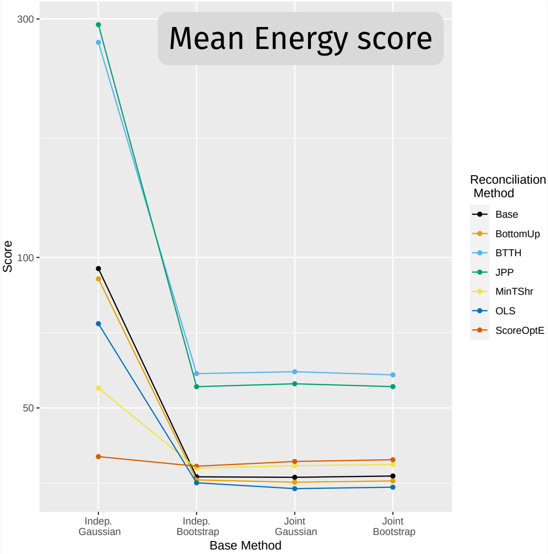

## Correlations of residuals: 2 Oct 2019 – 21 Jan 2020

Blue = positive correlation.

Red = negative correlation.

Large = stronger correlations.

# Example: Australian electricity generation



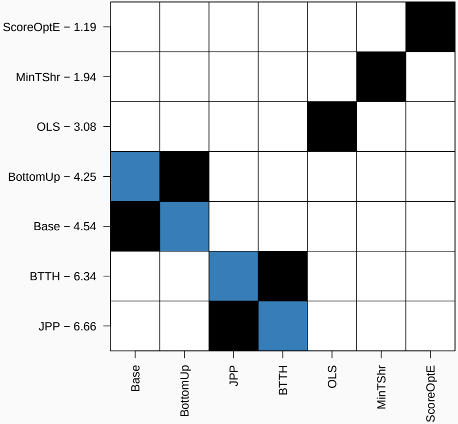

## Base residual assumptions

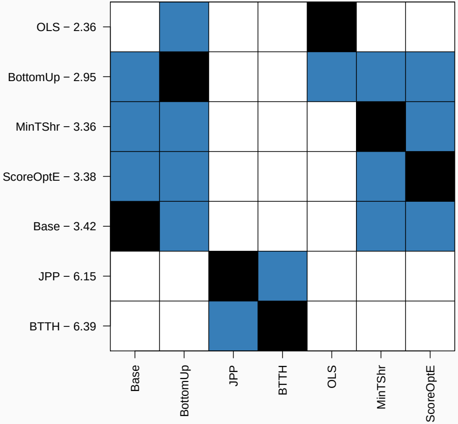
- Gaussian independent
- Gaussian dependent
- Non-Gaussian independent
- Non-Gaussian dependent

## Reconciliation methods

- Base
- BottomUp
- BTTH: Ben Taieb, Taylor, Hyndman
- JPP: Jeon, Panagiotelis, Petropoulos
- OLS

# Example: Australian electricity generation




## Nemenyi test for different scores

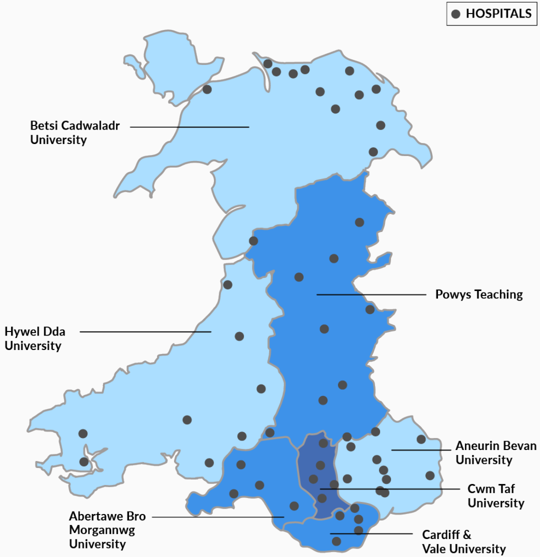
Base forecasts are independent and

## Nemenyi test for different scores

Base forecasts are obtained by jointly

# Outline

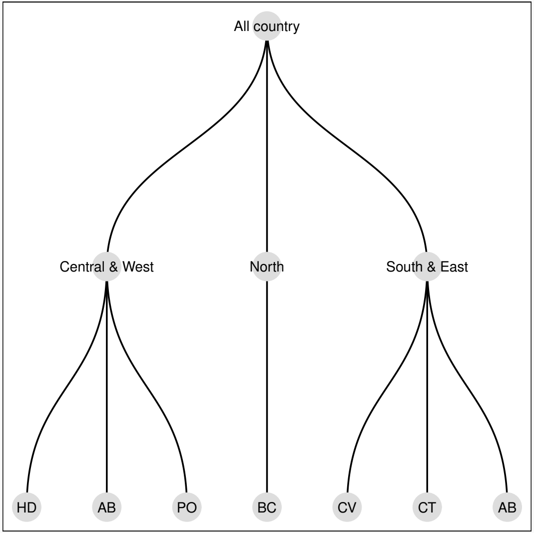
# Wales Health Board Areas



# Data

- Daily number of attended incidents:
  1 October 2015 – 31 July 2019
- Disaggregated by:
  - control area
  - health board
  - priority
  - nature of incidents
- 2,142,000 rows observations from 1,530 time series.

# Data structure




*

**Priority**

| |
|---|
| Red |
| Amber |
| Green |

*

**Nature of incident**

| |
|---|
| Chest pain |
| Stroke |
| Breathing problem |
| ... |
| Abdominal pain |

# Data structure

| Level | Number of series |
|---|---|
| All country | 1 |
| Control | 3 |
| Health board | 7 |
| Priority | 3 |
| Priority * Control | 9 |
| Priority * Health board | 21 |
| Nature of incident | 35 |
| Nature of incident * Control | 105 |
| Nature of incident * Health board | 245 |
| Priority * Nature of incident | 104 |
| Control * Priority * Nature of incident | 306 |
| Control * Health board * Priority * Nature of incident (Bottom level) | 691 |
| Total | 1530 |

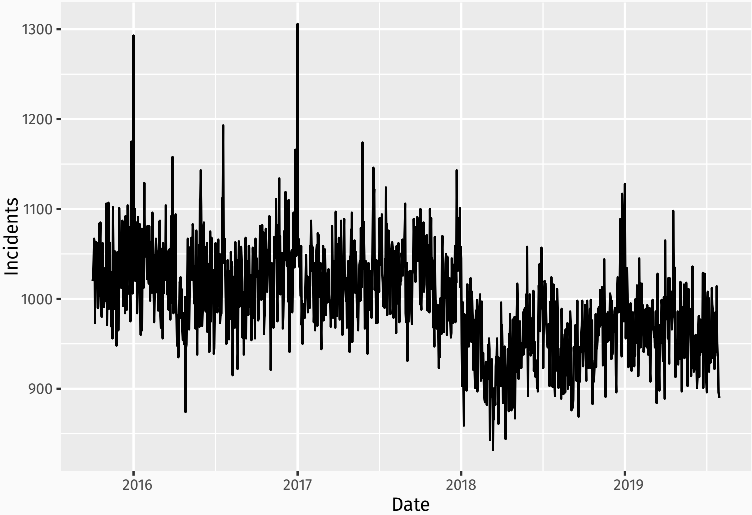
# Data

```
# A tsibble: 2,142,000 x 6 [1D]
# Key:       region, category, nature, lhb [1,530]
   date       region       category     nature       lhb          incident
   <date>     <chr*>       <chr*>       <chr*>       <chr*>          <dbl>
 1 2015-10-01 <aggregated> <aggregated> <aggregated> <aggregated>     1020
 2 2015-10-02 <aggregated> <aggregated> <aggregated> <aggregated>     1021
 3 2015-10-03 <aggregated> <aggregated> <aggregated> <aggregated>     1025
 4 2015-10-04 <aggregated> <aggregated> <aggregated> <aggregated>     1043
 5 2015-10-05 <aggregated> <aggregated> <aggregated> <aggregated>     1067
 6 2015-10-06 <aggregated> <aggregated> <aggregated> <aggregated>     1063
 7 2015-10-07 <aggregated> <aggregated> <aggregated> <aggregated>      973
 8 2015-10-08 <aggregated> <aggregated> <aggregated> <aggregated>     1057
 9 2015-10-09 <aggregated> <aggregated> <aggregated> <aggregated>     1026
10 2015-10-10 <aggregated> <aggregated> <aggregated> <aggregated>     1063
# i 2,141,990 more rows
```

# Data

```
# A tsibble: 2,142,000 x 6 [1D]
# Key:       region, category, nature, lhb [1,530]
   date       region category nature    lhb          incident
   <date>     <chr*> <chr*>   <chr*>    <chr*>          <dbl>
 1 2015-10-01 C      Amber    ABDOMINAL HD                 0
 2 2015-10-01 C      Amber    ABDOMINAL PO                 0
 3 2015-10-01 C      Amber    ABDOMINAL SB                 0
 4 2015-10-01 C      Amber    ABDOMINAL <aggregated>       0
 5 2015-10-01 C      Amber    ALLERGIES HD                 0
 6 2015-10-01 C      Amber    ALLERGIES PO                 1
 7 2015-10-01 C      Amber    ALLERGIES SB                 0
 8 2015-10-01 C      Amber    ALLERGIES <aggregated>       1
 9 2015-10-01 C      Amber    ANIMALBIT HD                 0
10 2015-10-01 C      Amber    ANIMALBIT PO                 0
# i 2,141,990 more rows
```

# Aggregated daily incidents

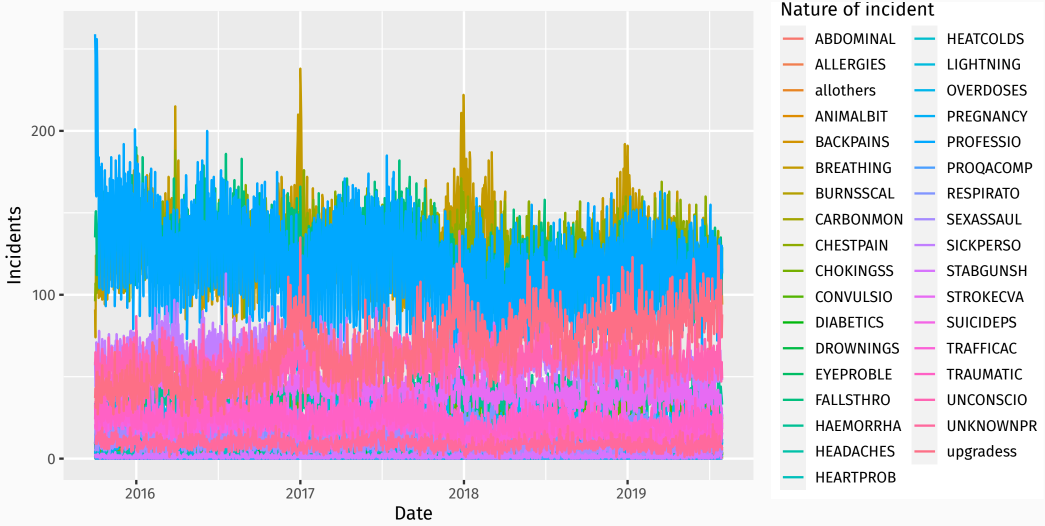


# Daily incidents by control area

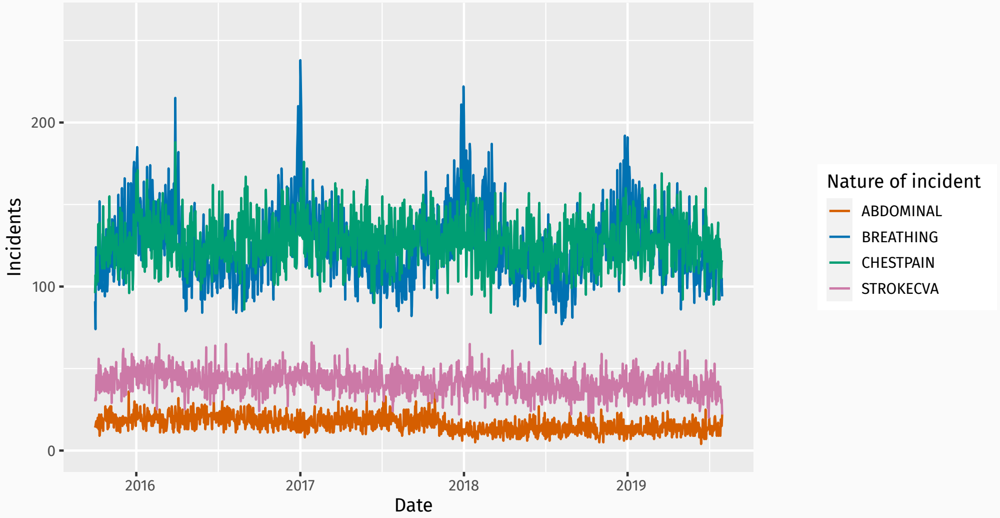
# Data incidents by health board

# Data incidents by priority

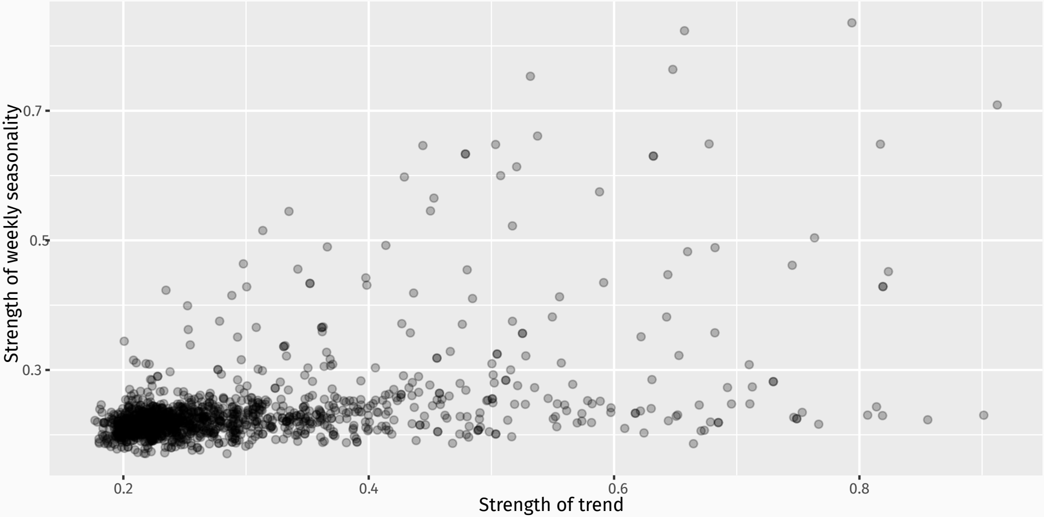
# Data incidents by nature of incident



# Data incidents by nature of incident



# Data features



# Forecasting methods

1. **Naïve**: Empirical distribution of past daily attended incidents.
2. **ETS**: Exponential Smoothing State Space models.
3. **GLM**: Poission Regression with spline trend, day of the week, annual Fourier seasonality, public holidays, school holidays, Christmas Day, New Year's Day.
4. **TSGLM**: Poisson Regression with same covariates plus three autoregressive terms.
5. **Ensemble**: Mixture distribution of 1–4.

# Forecasting methods

1. **Naïve**: Empirical distribution of past daily attended incidents.

$$y_{T+h|T} \sim \text{Empirical}(y_1, \dots, y_T)$$

# Forecasting methods

1. **Naïve**: Empirical distribution of past daily attended incidents.

$$y_{T+h|T} \sim \text{Empirical}(y_1, \dots, y_T)$$

2. **ETS**: Exponential Smoothing State Space models.

$$y_{T+h|T} \sim \text{Normal}(\hat{y}_{T+h|T}, \hat{\sigma}^2_{T+h|T})$$

# Forecasting methods

3 **GLM**: Poission Regression

$$y_{T+h|T} \sim \text{Poisson}(\hat{y}_{T+h|T}) \qquad \text{where} \qquad \hat{y}_{T+h|T} = \exp(\boldsymbol{x}'_{T+h}\beta)$$

and $\boldsymbol{x}_{T+h}$ is a vector of covariates including

- spline trend
- day of the week
- annual Fourier seasonality
- public holidays
- school holidays
- Christmas Day
- New Year's Day

# Forecasting methods

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 6.998511 | 0.017412 | 401.93 | < 2e-16 | *** |
| Spline_1 | 0.027859 | 0.004740 | 5.88 | 4.2e-09 | *** |
| Spline_2 | -0.088244 | 0.006394 | -13.80 | < 2e-16 | *** |
| Spline_3 | -0.075036 | 0.004784 | -15.68 | < 2e-16 | *** |
| Spline_4 | -0.111854 | 0.010202 | -10.96 | < 2e-16 | *** |
| Spline_5 | -0.043009 | 0.004462 | -9.64 | < 2e-16 | *** |
| Monday | 0.019147 | 0.003174 | 6.03 | 1.6e-09 | *** |
| Tuesday | -0.016414 | 0.003180 | -5.16 | 2.4e-07 | *** |
| Wednesday | -0.015479 | 0.003184 | -4.86 | 1.2e-06 | *** |
| Thursday | -0.006804 | 0.003178 | -2.14 | 0.03230 | * |
| Friday | 0.012235 | 0.003156 | 3.88 | 0.00011 | *** |
| Saturday | 0.005293 | 0.003165 | 1.67 | 0.09438 | . |
| Fourier_S1_365 | 0.005365 | 0.001294 | 4.15 | 3.4e-05 | *** |
| Fourier_C1_365 | 0.008263 | 0.001263 | 6.54 | 6.1e-11 | *** |
| Fourier_S2_365 | 0.004235 | 0.001271 | 3.33 | 0.00086 | *** |
| Fourier_C2_365 | -0.010510 | 0.001216 | -8.64 | < 2e-16 | *** |
| Fourier_S3_365 | -0.000556 | 0.001275 | -0.44 | 0.66303 | |
| Fourier_C3_365 | 0.002650 | 0.001243 | 2.13 | 0.03294 | * |
| Public_holiday | 0.033278 | 0.005697 | 5.84 | 5.2e-09 | *** |
| School_holiday | 0.004857 | 0.002346 | 2.07 | 0.03843 | * |
| Xmas | -0.051902 | 0.016772 | -3.09 | 0.00197 | ** |
| New_years_day | 0.120385 | 0.015573 | 7.73 | 1.1e-14 | *** |

**Significance**

| | |
|---|---|
| *** | $p < 0.001$ |
| ** | $p < 0.01$ |
| * | $p < 0.05$ |
| . | $p < 0.1$ |

# Forecasting methods

4 **TSGLM:** Poisson Regression

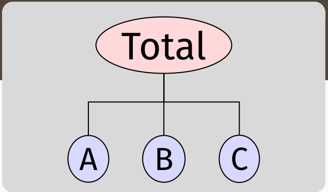
$$y_{T+h|T} \sim \text{Poisson}(\hat{y}_{T+h|T})$$

$$\text{where} \qquad \hat{y}_{T+h|T} = \exp\left(\boldsymbol{x}'_{T+h}\beta + \sum_{k=1}^{3} \alpha_k \log(y_{T+h-k} + 1)\right)$$

and $\boldsymbol{x}_{T+h}$ is a vector of covariates including

- spline trend
- day of the week
- annual Fourier seasonality
- public holidays
- school holidays
- Christmas Day
- New Year's Day

# Notation



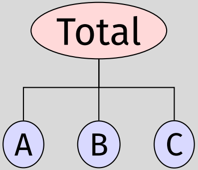

Every collection of time series with linear constraints can be written as

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t$$

- $\boldsymbol{y}_t$ = vector of all series at time $t$
- $y_{\text{Total},t}$ = aggregate of all series at time $t$.
- $y_{X,t}$ = value of series $X$ at time $t$.
- $\boldsymbol{b}_t$ = vector of most disaggregated series at time $t$
- $\boldsymbol{S}$ = "summing matrix" containing the linear constraints.

$$\boldsymbol{y}_t = \begin{pmatrix} y_{\text{Total},t} \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{S}} \underbrace{\begin{pmatrix} y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{pmatrix}}_{\boldsymbol{b}_t}$$

# Notation




Every collection of time series with linear constraints can be written as

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t$$

- $\boldsymbol{y}_t$ = vector of all series at time $t$
- $y_{\text{Total},t}$ = aggregate of all series at time $t$.
- $y_{X,t}$ = value of series $X$ at time $t$.
- $\boldsymbol{b}_t$ = vector of most disaggregated series at time $t$
- $\boldsymbol{S}$ = "summing matrix" containing the linear constraints.

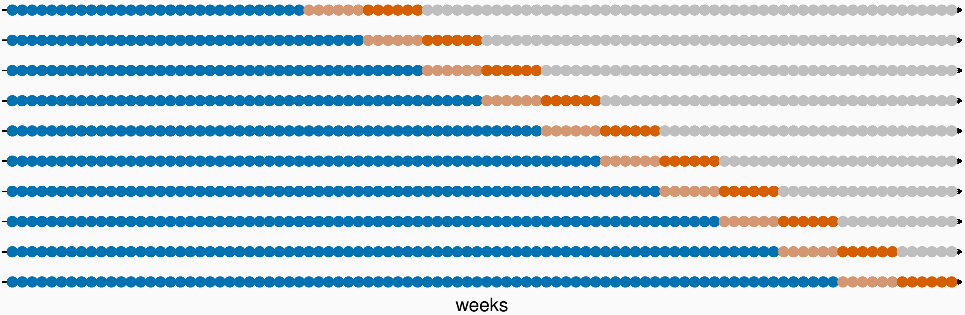
- Base forecasts: $\hat{\boldsymbol{y}}_{T+h|T}$
- Reconciled forecasts: $\tilde{\boldsymbol{y}}_{T+h|T} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_{T+h|T}$
- MinT: $\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}$ where $\boldsymbol{W}_h$ is covariance matrix of base forecast errors.

# Nonparametric bootstrap reconciliation

- Fit model to all series and store the residuals as $\underset{\sim}{\varepsilon}_t$.
- These should be serially uncorrelated but cross-sectionally correlated.
- Draw iid samples from $\underset{\sim}{\varepsilon}_1, \dots, \underset{\sim}{\varepsilon}_T$ with replacement.
- Simulate future sample paths for model using the bootstrapped residuals.
- Reconcile each sample path using MinT.
- Combine the reconciled sample paths to form a mixture distribution at each forecast horizon.

# Performance evaluation

- Ten-fold time series cross-validation
- Forecast horizon of 1–84 days
- Each training set contains an additional 42 days.
- Forecasts at 43–84 days correspond to planning horizon.



# Performance evaluation

$$\text{MASE} = \text{mean}(|q_j|)$$

$$q_j = \frac{e_j}{\dfrac{1}{T-m}\displaystyle\sum_{t=m+1}^{T}|y_t - y_{t-m}|}$$

- $y_t$ = observation for period $t$
- $e_j$ = forecast error for forecast horizon $j$
- $T$ = size of training set
- $m = 7$

# Performance evaluation

$$\text{MSSE} = \text{mean}(q_j^2)$$

$$q_j^2 = \frac{e_j^2}{\dfrac{1}{T-m}\displaystyle\sum_{t=m+1}^{T}(y_t - y_{t-m})^2}$$

- $y_t$ = observation for period $t$
- $e_j$ = forecast error for forecast horizon $j$
- $T$ = size of training set
- $m = 7$

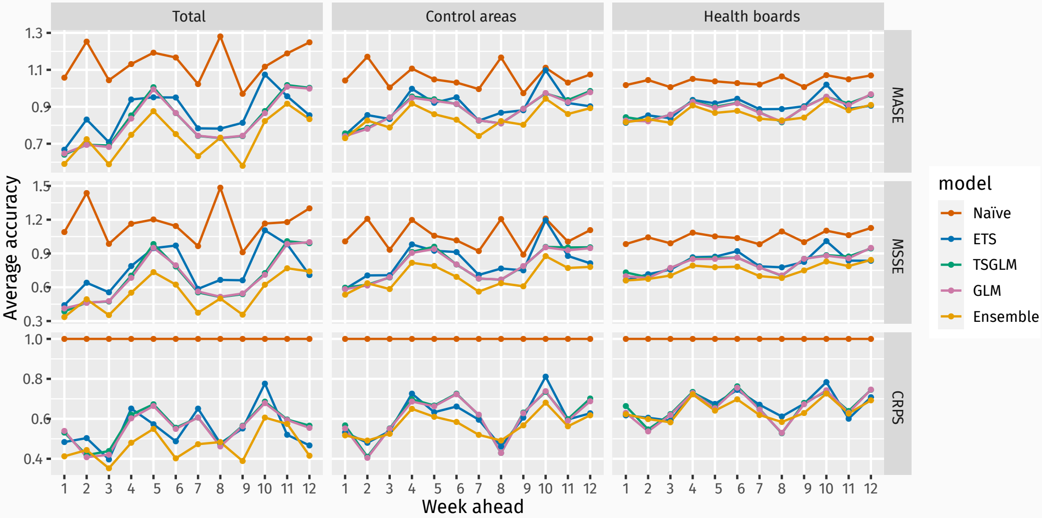
# Performance evaluation

$$\text{CRPS} = \text{mean}(p_j)$$

$$p_j = \int_{-\infty}^{\infty} \left(G_j(x) - F_j(x)\right)^2 dx,$$

- $G_j(x)$ = forecast distribution for forecast horizon $j$
- $F_j(x)$ = true distribution for same period

# Forecast accuracy



# Forecast accuracy: 43–84 days ahead

| | | MSSE | | | |
|---|---|---|---|---|---|
| Method | Model | Total | Control areas | Health boards | Bottom |
| Base | Naïve | 1.169 | 1.056 | 1.062 | 1.031 |
| Base | ETS | 0.979 | 0.875 | 0.816 | **0.975** |
| Base | GLM | 0.813 | 0.897 | 0.875 | 1.009 |
| Base | TSGLM | 0.822 | 0.901 | 0.875 | 1.050 |
| Base | Ensemble | 0.599 | 0.729 | 0.774 | 0.993 |
| MinT | Naïve | 1.168 | 1.057 | 1.062 | 2.095 |
| MinT | ETS | 0.785 | 0.852 | 0.845 | 0.994 |
| MinT | GLM | 0.720 | 0.827 | 0.837 | 1.803 |
| MinT | TSGLM | 0.722 | 0.833 | 0.839 | 1.851 |
| MinT | Ensemble | **0.560** | **0.706** | **0.765** | 1.557 |

# Forecast accuracy: 43–84 days ahead

| | | MASE | | | |
|---|---|---|---|---|---|
| Method | Model | Total | Control areas | Health boards | Bottom |
| Base | Naïve | 1.139 | 1.059 | 1.047 | 1.019 |
| Base | ETS | 0.963 | 0.930 | 0.899 | 1.038 |
| Base | GLM | 0.910 | 0.940 | 0.923 | **1.002** |
| Base | TSGLM | 0.911 | 0.939 | 0.924 | 1.005 |
| Base | Ensemble | 0.782 | 0.856 | 0.876 | 1.008 |
| MinT | Naïve | 1.138 | 1.059 | 1.047 | 2.651 |
| MinT | ETS | 0.877 | 0.916 | 0.915 | 1.289 |
| MinT | GLM | 0.848 | 0.901 | 0.902 | 2.493 |
| MinT | TSGLM | 0.852 | 0.903 | 0.903 | 2.513 |
| MinT | Ensemble | **0.753** | **0.844** | **0.872** | 2.260 |

# Forecast accuracy: 43–84 days ahead

| | | CRPS | | | |
|---|---|---|---|---|---|
| Method | Model | Total | Control areas | Health boards | Bottom |
| Base | Naïve | 30.387 | 10.882 | 5.500 | 0.302 |
| Base | ETS | 14.309 | 6.074 | 3.476 | 0.244 |
| Base | GLM | 15.396 | 6.253 | 3.576 | 0.244 |
| Base | TSGLM | 15.316 | 6.227 | 3.575 | 0.245 |
| Base | Ensemble | 12.978 | **5.727** | 3.430 | 0.243 |
| MinT | Naïve | 30.368 | 10.902 | 5.498 | 0.313 |
| MinT | ETS | 13.515 | 5.967 | 3.547 | **0.243** |
| MinT | GLM | 13.839 | 5.917 | 3.453 | 0.246 |
| MinT | TSGLM | 14.000 | 5.947 | 3.455 | 0.248 |
| MinT | Ensemble | **12.585** | 5.728 | **3.426** | 0.247 |

# Conclusions

- Ensemble mixture distributions give better forecasts than any component methods.
- Forecast reconciliation improves forecast accuracy, even when some component methods are quite poor.
- Forecast reconciliation allows coordinated planning and resource allocation.

# Outline

# Bayesian versions

Novak, McGarvie, and Garcia (2017)

Another strain of the literature brings a Bayesian approach to the regression model interpretation of forecast reconciliation. Novak, McGarvie, and Garcia (2017) recognise that the posterior of $\beta_h$ can act as a probabilistic forecast for the bottom-level series. Using Markov chain Monte Carlo to obtain a sample from this posterior, and then aggregating, gives a probabilistic forecast for the entire hierarchy.

# Bayesian versions

Eckert, Hyndman, and Panagiotelis (2021) also obtain a posterior on $\beta_h$, but their focus is on augmenting the reconciliation regression equation with a vector of intercepts that allow for base forecasts to be biased and evolve according to a state space representation.

Judgement can be incorporated via the prior, in the latter case via an explicit empirical example where prior information about a structural break in data classification can be exploited. Also, while both papers recognise the potential of Bayesian inference to obtain probabilistic forecasts, neither paper

# Bayesian versions

Corani, Azzimonti, Augusto, and Zaffalon (2021) In particular, a prior is placed on the bottom-level series with the mean set to point forecasts obtained in the first step of forecast reconciliation and a variance given by the variance-covariance matrix of one-step ahead errors. This prior is updated using the top-level forecasts obtained in the first stage of forecast reconciliation via Bayes' rule. The method generalises MinT in the sense that the posterior mean is equivalent to the usual MinT approach. The necessary updates via Bayes' rule have parallels with the Kalman filter since the reconciliation problem is recast as a linear Gaussian model. The empirical

# Thanks!

# More information

- Slides and papers: **robjhyndman.com**
- Packages: **tidyverts.org**
- Forecasting textbook using fable package: **OTexts.com/fpp3**


**Find me at ...**

@robjhyndman

@robjhyndman

robjhyndman.com

# References


- Ben Taieb, S, JW Taylor, and RJ Hyndman (2021). Hierarchical Probabilistic Forecasting of Electricity Demand with Smart Meter Data. *J American Statistical Association* **116**(533), 27–43.
- Corani, G, D Azzimonti, JPSC Augusto, and M Zaffalon (2021). Probabilistic reconciliation of hierarchical forecast via Bayes' rule. In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by F Hutter, K Kersting, J Lijffijt, and I Valera. Cham: Springer International Publishing, pp.211–226.
- Corani, G, N Rubattu, D Azzimonti, and A Antonucci (2022). "Probabilistic reconciliation of count time series". https://arxiv.org/abs/2207.09322.
- Eckert, F, RJ Hyndman, and A Panagiotelis (2021). Forecasting Swiss exports using Bayesian forecast reconciliation. *European J Operational Research* **291**(2), 693–710.

# References


- Girolimetto, D, G Athanasopoulos, T Di Fonzo, and RJ Hyndman (2023). "Cross-temporal probabilistic forecast reconciliation". robjhyndman.com/publications/ctprob.html.
- Novak, J, S McGarvie, and BE Garcia (2017). "A Bayesian model for forecasting hierarchically structured time series". https://arxiv.org/abs/1711.04738.
- Panagiotelis, A, P Gamakumara, G Athanasopoulos, and RJ Hyndman (2023). Probabilistic forecast reconciliation: properties, evaluation and score optimisation. *European J Operational Research* **306**(2), 693–706.