

10.7 The optimal reconciliation approach

Optimal forecast reconciliation will occur if we can find the \mathbf{G} matrix which minimises the forecast error of the set of coherent forecasts. We present here a simplified summary of the approach. More details are provided in [Wickramasuriya, Athanasopoulos, & Hyndman \(2019\)](#).

Suppose we generate coherent forecasts using Equation (10.6), repeated here for convenience:

$$\tilde{\mathbf{y}}_h = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_h.$$

First we want to make sure we have unbiased forecasts. If the base forecasts $\hat{\mathbf{y}}_h$ are unbiased, then the coherent forecasts $\tilde{\mathbf{y}}_h$ will be unbiased provided $\mathbf{S}\mathbf{G}\mathbf{S} = \mathbf{S}$ ([Hyndman, Ahmed, Athanasopoulos, & Shang, 2011](#)). This provides a constraint on the matrix \mathbf{G} . Interestingly, no top-down method satisfies this constraint, so all top-down methods are biased.

Next we need to find the error in our forecasts. [Wickramasuriya, Athanasopoulos, & Hyndman \(2019\)](#) show that the variance-covariance matrix of the h -step-ahead coherent forecast errors is given by

$$\mathbf{V}_h = \text{Var}[\mathbf{y}_{T+h} - \tilde{\mathbf{y}}_h] = \mathbf{S}\mathbf{G}\mathbf{W}_h\mathbf{G}'\mathbf{S}'$$

where $\mathbf{W}_h = \text{Var}[(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_h)]$ is the variance-covariance matrix of the corresponding base forecast errors.

The objective is to find a matrix \mathbf{G} that minimises the error variances of the coherent forecasts. These error variances are on the diagonal of the matrix \mathbf{V}_h , and so the sum of all the error variances is given by the trace of the matrix \mathbf{V}_h . [Wickramasuriya, Athanasopoulos, & Hyndman \(2019\)](#) show that the matrix \mathbf{G} which minimises the trace of \mathbf{V}_h such that $\mathbf{S}\mathbf{G}\mathbf{S} = \mathbf{S}$, is given by

$$\mathbf{G} = (\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}.$$

Therefore, the optimal reconciled forecasts are given by

$$\tilde{\mathbf{y}}_h = \mathbf{S}(\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}\hat{\mathbf{y}}_h. \quad (10.8)$$

We refer to this as the “MinT” (or Minimum Trace) estimator.

To use this in practice, we need to estimate \mathbf{W}_h , the forecast error variance of the h -step-ahead base forecasts. This can be difficult, and so we provide four simplifying approximations which have been shown to work well in both simulations and in practice.

1. Set $\mathbf{W}_h = k_h \mathbf{I}$ for all h , where $k_h > 0$.²¹ This is the most simplifying assumption to make, and means that \mathbf{G} is independent of the data, providing substantial computational savings. The disadvantage, however, is that this specification does not account for the differences in scale between the levels of the structure, or for relationships between series. This approach is implemented in the `forecast()` function by setting `method = "comb"` and `weights = "ols"`.

The weights here are referred to as OLS (ordinary least squares) because setting $\mathbf{W}_h = k_h \mathbf{I}$ in (10.8) gives the least squares estimator we introduced in Section 5.7 with $\mathbf{X} = \mathbf{S}$ and $\mathbf{y} = \hat{\mathbf{y}}$.

2. Set $\mathbf{W}_h = k_h \text{diag}(\hat{\mathbf{W}}_1)$ for all h , where $k_h > 0$,

$$\hat{\mathbf{W}}_1 = \frac{1}{T} \sum_{t=1}^T \mathbf{e}_t \mathbf{e}_t',$$

and \mathbf{e}_t is an n -dimensional vector of residuals of the models that generated the base forecasts stacked in the same order as the data. The approach is implemented by setting `method = "comb"` and `weights = "wls"`.

This specification scales the base forecasts using the variance of the residuals and it is therefore referred to as the WLS (weighted least squares) estimator using *variance scaling*.

3. Set $\mathbf{W}_h = k_h \mathbf{\Lambda}$ for all h , where $k_h > 0$, $\mathbf{\Lambda} = \text{diag}(\mathbf{S}\mathbf{1})$, and $\mathbf{1}$ is a unit vector of dimension m (the number of bottom-level series). This specification assumes that the bottom-level base forecast errors each have variance k_h and are uncorrelated between nodes. Hence each element of the diagonal $\mathbf{\Lambda}$ matrix contains the number of forecast error variances contributing to each node. This estimator only depends on the structure of the aggregations, and not on the actual data. It is therefore referred to as *structural scaling*. Applying the structural scaling specification is

particularly useful in cases where residuals are not available, and so variance scaling cannot be applied; for example, in cases where the base forecasts are generated by judgmental forecasting (Chapter 4). This approach is implemented by setting `method="comb"` and `weights = "nseries"`.

4. Set $\mathbf{W}_h = k_h \mathbf{W}_1$ for all h , where $k_h > 0$. Here we only assume that the error covariance matrices are proportional to each other, and we directly estimate the full one-step covariance matrix \mathbf{W}_1 . The most obvious and simple way would be to use the sample covariance. This is implemented by setting `method = "comb"`, `weights = "mint"`, and `covariance = "sam"`.

However, for cases where the number of bottom-level series m is large compared to the length of the series T , this is not a good estimator. Instead we use a shrinkage estimator which shrinks the sample covariance to a diagonal matrix. This is implemented by setting `method = "comb"`, `weights = "mint"`, and `covariance = "shr"`.

In summary, unlike any other existing approach, the optimal reconciliation forecasts are generated using all the information available within a hierarchical or a grouped structure. This is important, as particular aggregation levels or groupings may reveal features of the data that are of interest to the user and are important to be modelled. These features may be completely hidden or not easily identifiable at other levels.

For example, consider the Australian tourism data introduced in Section 10.1, where the hierarchical structure followed the geographic division of a country into states and zones. Some coastal areas will be largely summer destinations, while some mountain regions may be winter destinations. These differences will be smoothed at the country level due to aggregation.

Example: Forecasting Australian prison population

We compute the forecasts for the Australian prison population, described in Section 10.2. Using the default arguments for the `forecast()` function, we compute coherent forecasts by the optimal reconciliation approach with the WLS estimator using variance scaling.

```
prisonfc <- forecast(prison.gts)
```

To obtain forecasts for each level of aggregation, we can use the `aggrts()` function. For example, to calculate forecasts for the overall total prison population, and for the one-factor groupings (State, Gender and Legal Status), we use:

```
fcsts <- aggrts(prisonfc, levels=0:3)
```

A simple plot is obtained using

```
groups <- aggrts(prison.gts, levels=0:3)
```

```
autoplot(fcsts) + autolayer(groups)
```

A nicer plot is available using the following code. The results are shown in Figure 10.8. The vertical line marks the start of the forecast period.

```
prisonfc <- ts(rbind(groups, fcsts),
  start=start(groups), frequency=4)
p1 <- autoplot(prisonfc[, "Total"]) +
  ggtitle("Australian prison population") +
  xlab("Year") + ylab("Total number of prisoners ('000)") +
  geom_vline(xintercept=2017)
cols <- sample(scales::hue_pal(h=c(15,375),
  c=100,l=65,h.start=0,direction = 1)(NCOL(groups)))
p2 <- as_tibble(prisonfc[, -1]) %>%
  gather(Series) %>%
  mutate(Date = rep(time(prisonfc), NCOL(prisonfc)-1),
    Group = str_extract(Series, "[A-Za-z ]*")) %>%
  ggplot(aes(x=Date, y=value, group=Series, colour=Series)) +
  geom_line() +
  xlab("Year") + ylab("Number of prisoners ('000)") +
  scale_colour_manual(values = cols) +
  facet_grid(. ~ Group, scales="free_y") +
  scale_x_continuous(breaks=seq(2006,2018,by=2)) +
  theme(axis.text.x = element_text(angle=90, hjust=1)) +
  geom_vline(xintercept=2017)
gridExtra::grid.arrange(p1, p2, ncol=1)
```

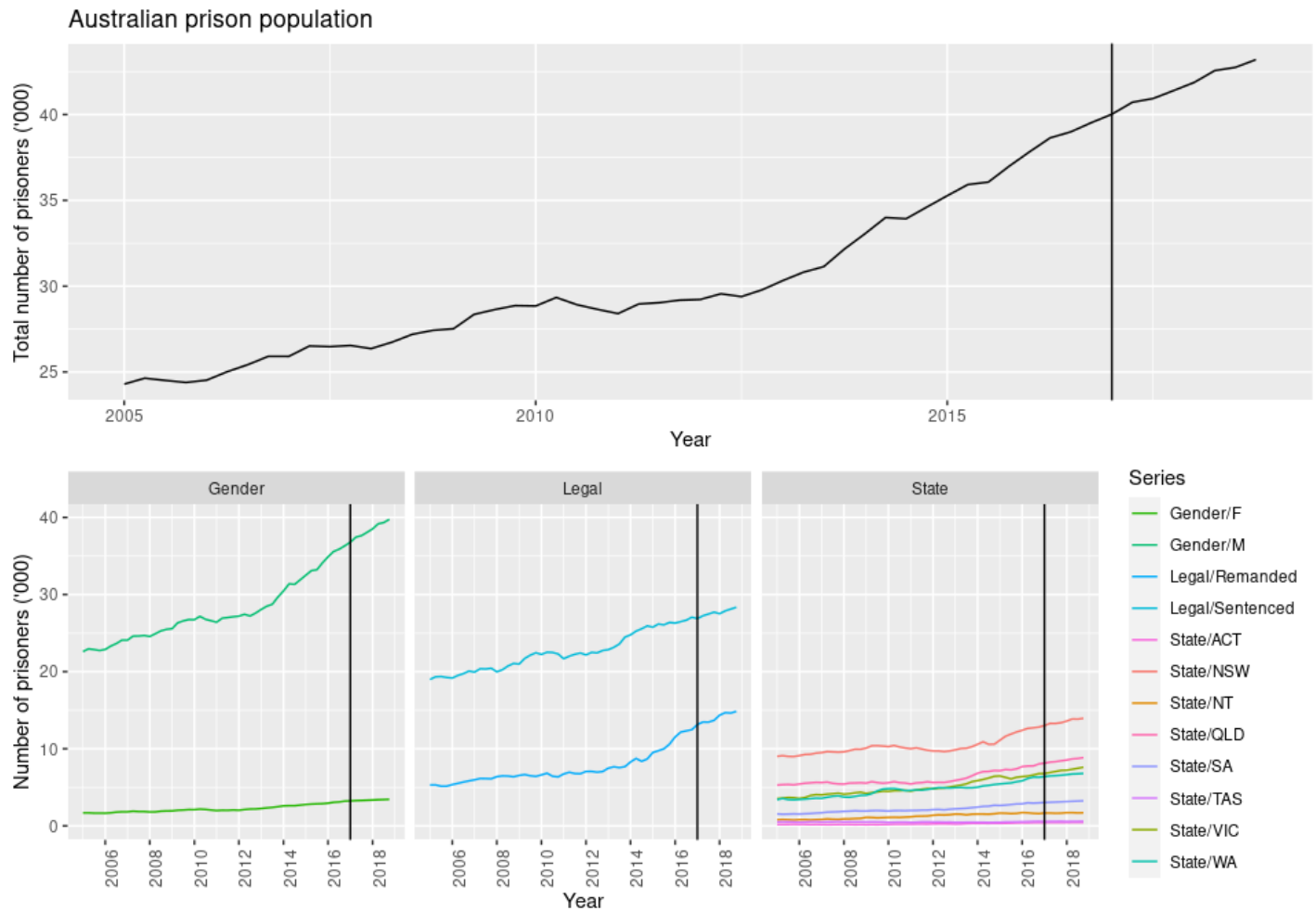


Figure 10.8: Coherent forecasts for the total Australian adult prison population and for the population grouped by state, by legal status and by gender.

Similar code was used to produce Figure 10.9. The left panel plots the coherent forecasts for interactions between states and gender. The right panel shows forecasts for the bottom-level series.

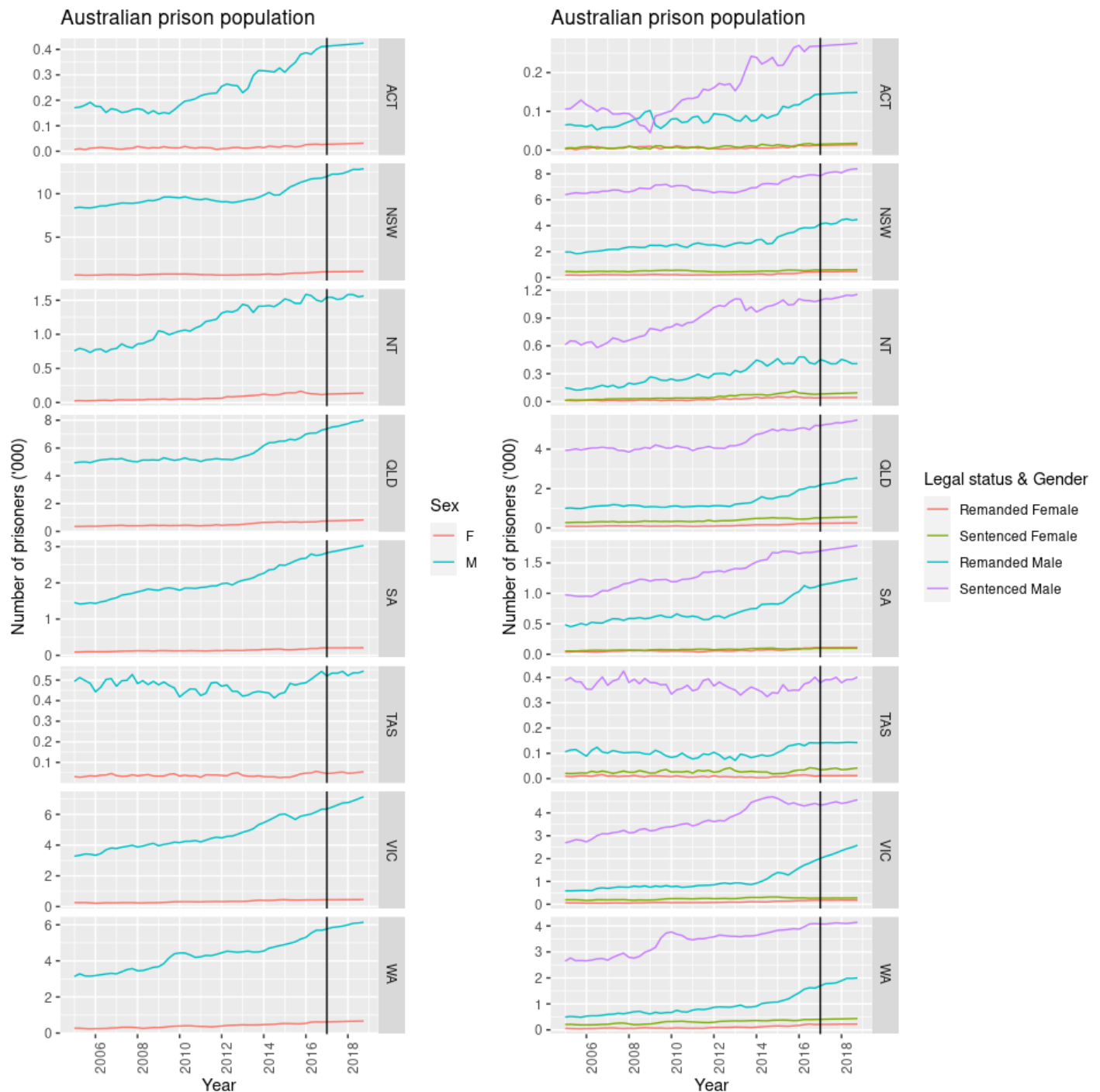


Figure 10.9: Coherent forecasts for the Australian adult prison population grouped by all interactions of attributes.

The `accuracy()` command is useful for evaluating the forecast accuracy across hierarchical or grouped structures. The following table summarises the accuracy of the bottom-up and the optimal reconciliation approaches, forecasting 2015 Q1 to 2016 Q4 as a test period.

The results show that the optimal reconciliation approach generates more accurate forecasts especially for the top level. In general, we find that as the optimal reconciliation approach uses information from all levels in the structure, it generates more accurate coherent forecasts than the other traditional alternatives which use limited information.

Table 10.1: Accuracy of Australian prison population forecasts for different groups of series.

	Bottom-up		Optimal	
	MAPE	MASE	MAPE	MASE
Total	5.32	1.84	3.08	1.06
State	7.59	1.88	7.62	1.85
Legal status	6.40	1.76	4.32	1.14
Gender	8.62	2.68	8.72	2.74
Bottom	15.82	2.23	15.25	2.16
All series	12.41	2.16	12.02	2.08

21. Note that k_h is a proportionality constant. It does not need to be estimated or specified here as it gets cancelled out in (10.8).↵