MONASH University

MONASH BUSINESS SCHOOL

Forecasting: principles and practice

Rob J Hyndman

3.4 Extras

Outline

1 Forecast combinations

2 Missing values

3 Outliers

Forecast combinations

Clemen (1989)

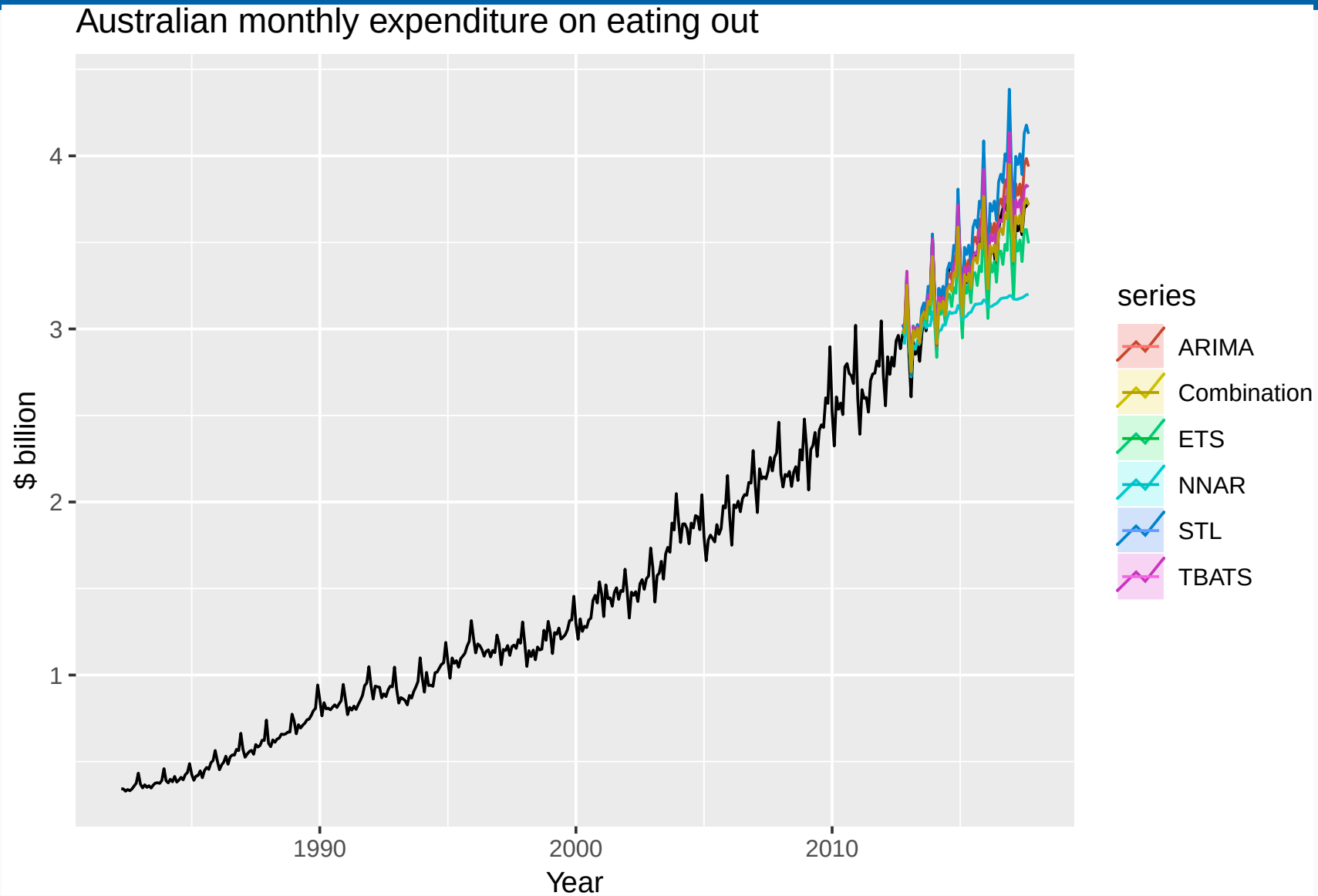
"The results have been virtually unanimous: combining multiple forecasts leads to increased forecast accuracy. . . . In many cases one can make dramatic performance improvements by simply averaging the forecasts."

Forecast combinations

```
train <- window(auscafe, end=c(2012,9))
h <- length(auscafe) - length(train)
ETS <- forecast(ets(train), h=h)
ARIMA <- forecast(auto.arima(train, lambda=0, biasadj=TRUE),
  h=h)
STL <- stlf(train, lambda=0, h=h, biasadj=TRUE)
NNAR <- forecast(nnetar(train), h=h)
TBATS <- forecast(tbats(train, biasadj=TRUE), h=h)
Combination <- (ETS[["mean"]] + ARIMA[["mean"]] +
  STL[["mean"]] + NNAR[["mean"]] + TBATS[["mean"]])/5

autoplot(auscafe) +
  autolayer(ETS, series="ETS", PI=FALSE) +
  autolayer(ARIMA, series="ARIMA", PI=FALSE) +
  autolayer(STL, series="STL", PI=FALSE) +
  autolayer(NNAR, series="NNAR", PI=FALSE) +
  autolayer(TBATS, series="TBATS", PI=FALSE) +
  autolayer(Combination, series="Combination") +
  xlab("Year") + ylab("$ billion") +
  ggtitle("Australian monthly expenditure on eating out")
```

Forecast combinations



Forecast combinations

```
c(ETS = accuracy(ETS, auscafe)["Test set","RMSE"],
  ARIMA = accuracy(ARIMA, auscafe)["Test set","RMSE"],
  STL-ETS = accuracy(STL, auscafe)["Test set","RMSE"],
  NNAR = accuracy(NNAR, auscafe)["Test set","RMSE"],
  TBATS = accuracy(TBATS, auscafe)["Test set","RMSE"],
  Combination =
    accuracy(Combination, auscafe)["Test set","RMSE"])
```

```
##          ETS        ARIMA      STL-ETS         NNAR
##      0.13700      0.12146      0.21446      0.32619
##        TBATS Combination
##      0.09406      0.07185
```

Outline

1 Forecast combinations

2 Missing values

3 Outliers

Missing values

Functions which can handle missing values

- `auto.arima()`, `Arima()`
- `tslm()`
- `nnetar()`

Models which cannot handle missing values

- `ets()`
- `stl()`
- `stlf()`
- `tbats()`

Missing values

Functions which can handle missing values

- `auto.arima()`, `Arima()`
- `tslm()`
- `nnetar()`

Models which cannot handle missing values

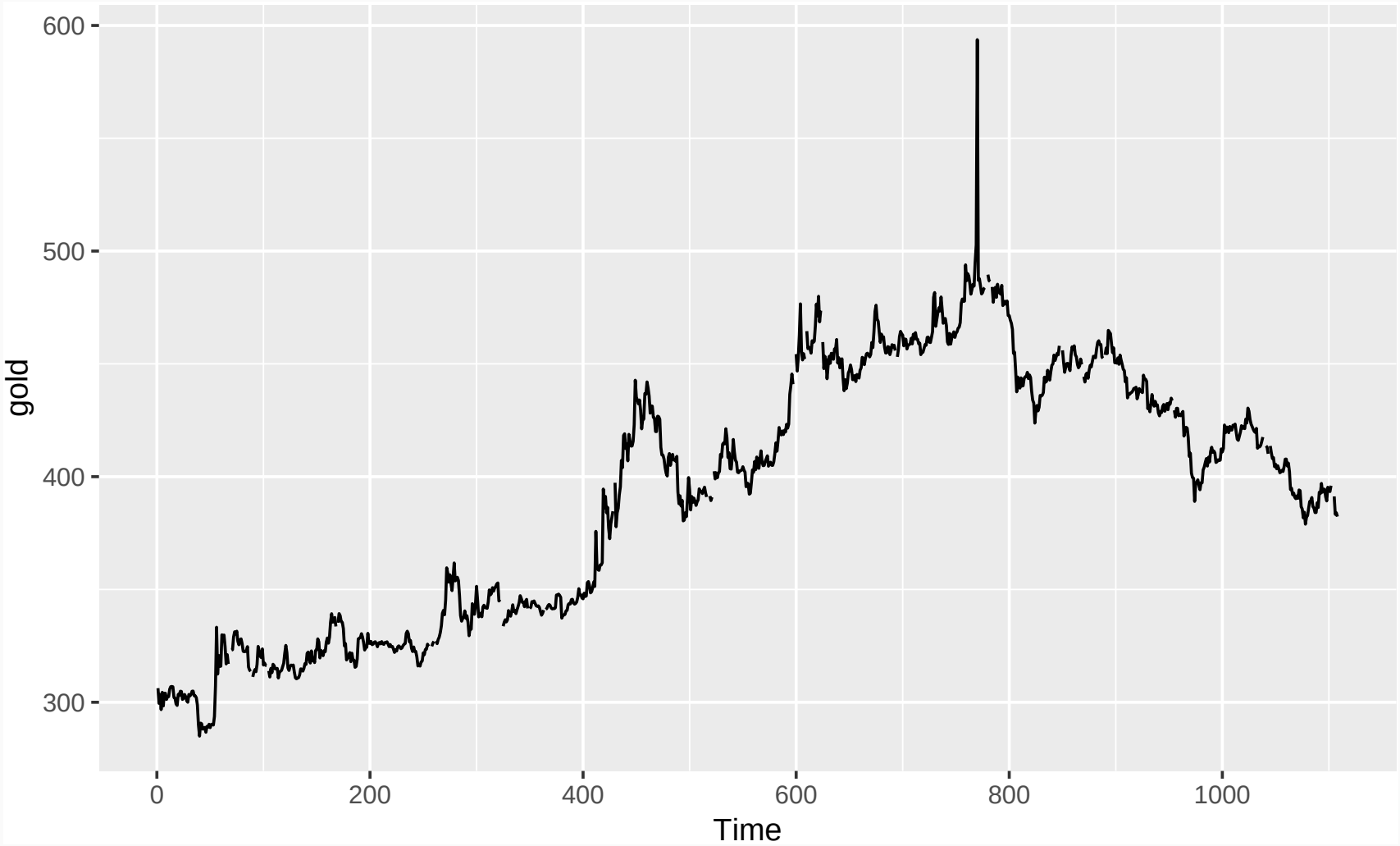
- `ets()`
- `stl()`
- `stlf()`
- `tbats()`

What to do?

1. Model section of data after last missing value.
2. Estimate missing values with `na.interp()`.

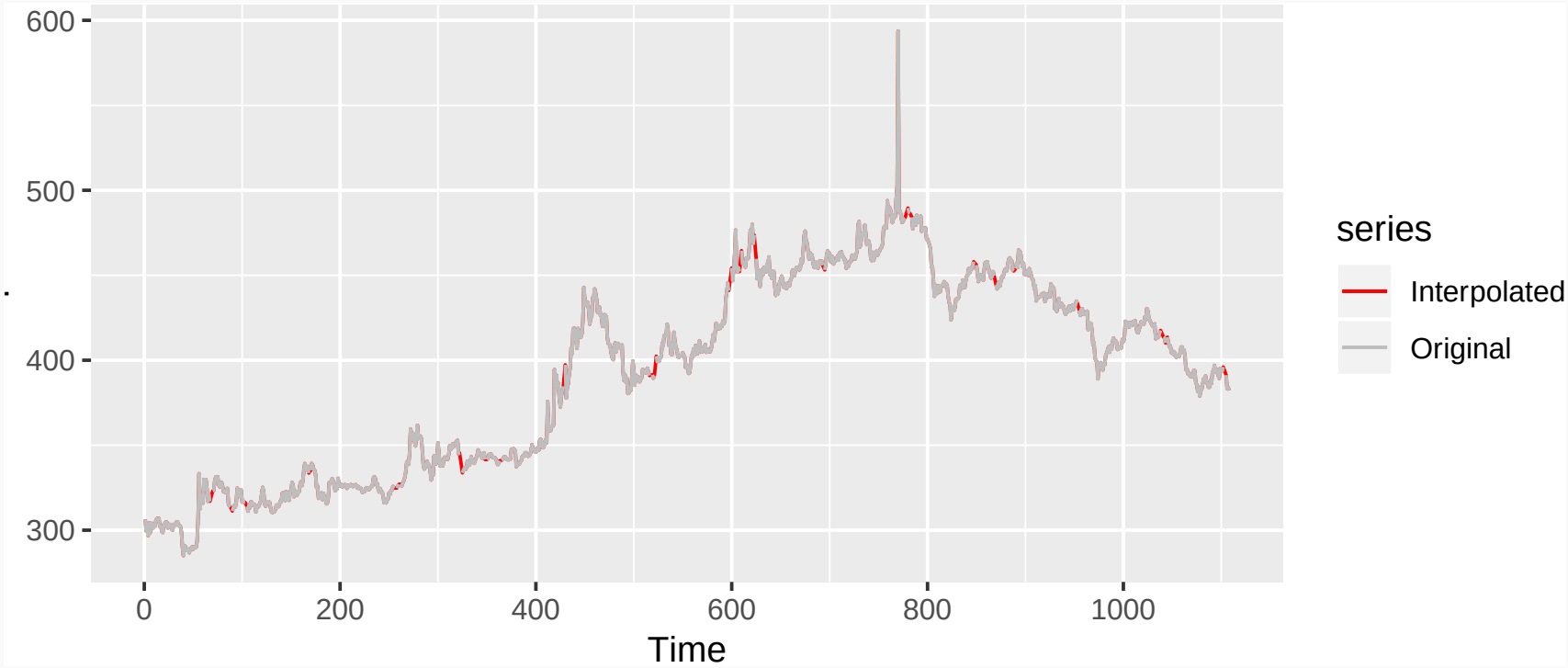
Missing values

```
autoplot(gold)
```



Missing values

```
gold %>% na.interp() %>%
  autoplot(series="Interpolated") +
    autolayer(gold, series="Original") +
    scale_color_manual(
      values=c(Interpolated="red",Original="gray"))
```



Outline

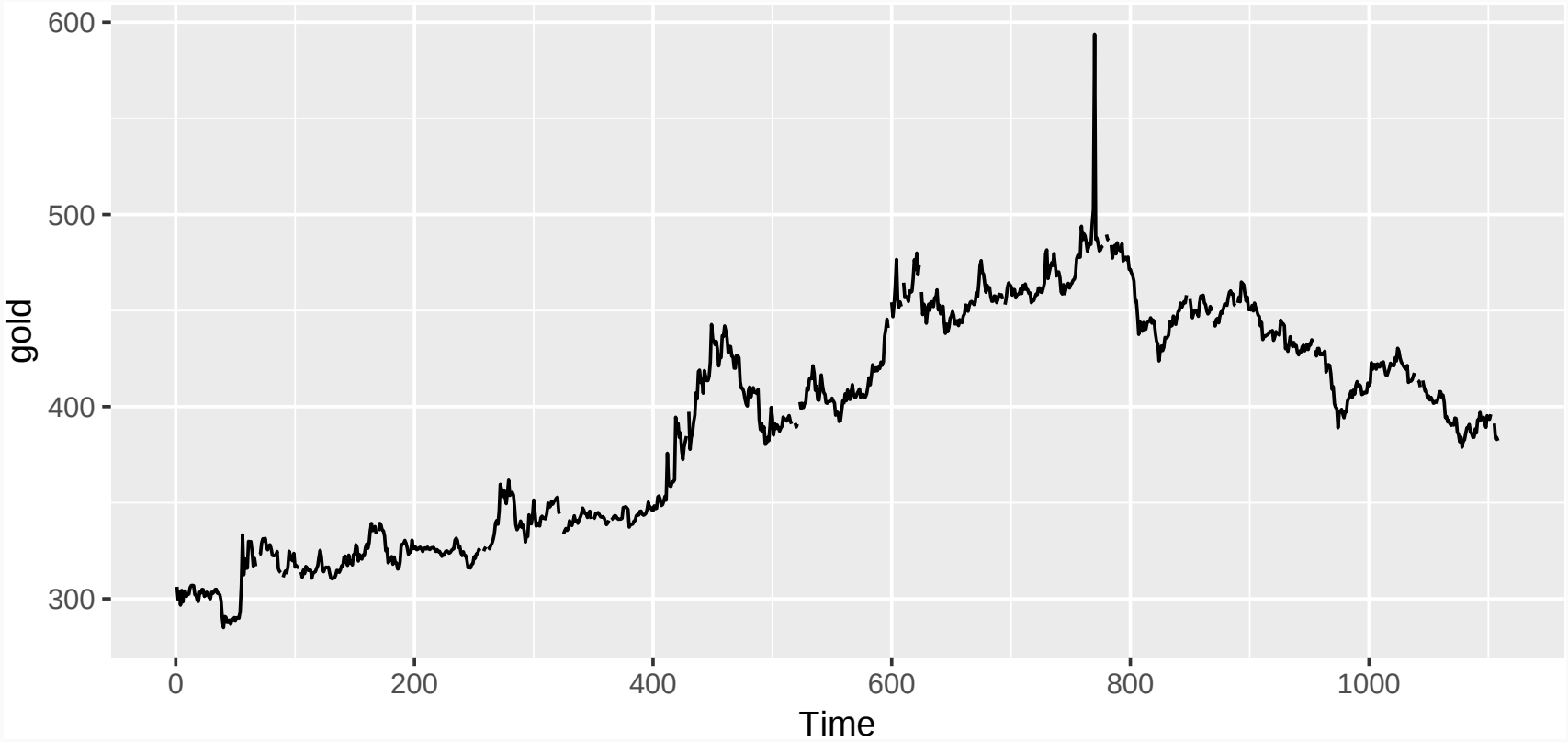
1 Forecast combinations

2 Missing values

3 Outliers

Outliers

```
autoplot(gold)
```



Outliers

```
tsoutliers(gold)
```

```
## $index
## [1] 770
##
## $replacements
## [1] 494.9
```

Outliers

```
gold %>% tsclean() %>% autoplot()
```