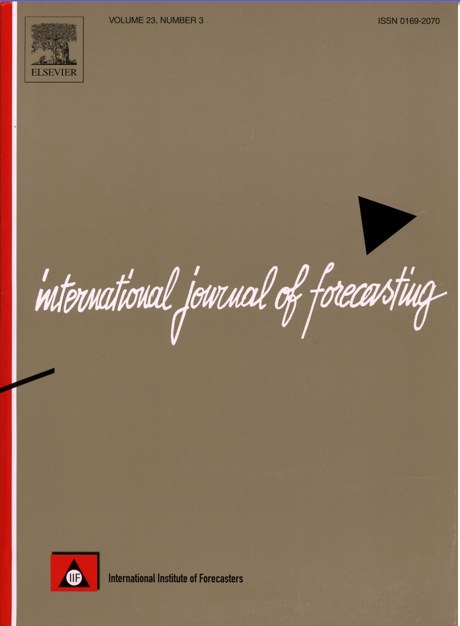
Extreme Forecasting

Rob J Hyndman

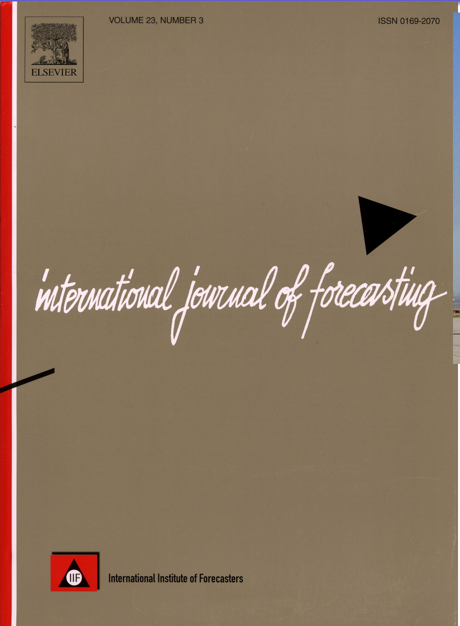
Business & Economic Forecasting Unit

MONASH University

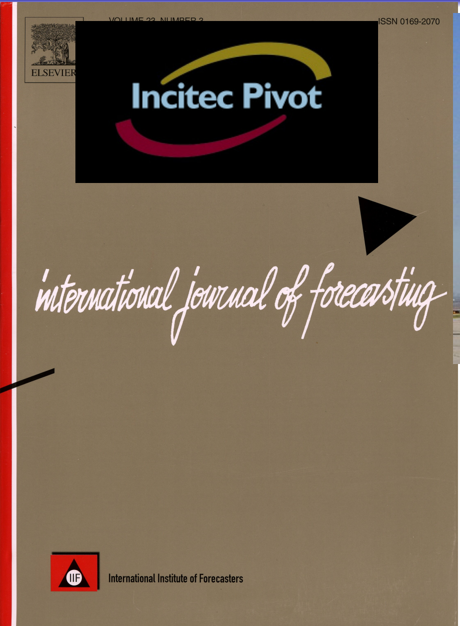
My perspective



My perspective



My perspective



My perspective

My perspective

My perspective

My perspective

My perspective

My perspective

- By focusing on simplistic problems, textbooks give the impression that forecasting is easy, leading to badly trained forecasters.

My perspective

- By focusing on simplistic problems, textbooks give the impression that forecasting is easy, leading to badly trained forecasters.
- Real-world problems are often much more difficult than existing forecasting methods can handle.

My perspective

- By focusing on simplistic problems, textbooks give the impression that forecasting is easy, leading to badly trained forecasters.
- Real-world problems are often much more difficult than existing forecasting methods can handle.
- Many research papers tackle “small” problems rather than the “big” problems that arise in commercial forecasting.

My perspective

- By focusing on simplistic problems, textbooks give the impression that forecasting is easy, leading to badly trained forecasters.
- Real-world problems are often much more difficult than existing forecasting methods can handle.
- Many research papers tackle “small” problems rather than the “big” problems that arise in commercial forecasting.
- Commercial forecasting problems can (and should) motivate research.

Three case studies

1. Weekly airline passenger traffic: Ansett Australia.

Three case studies

1. Weekly airline passenger traffic: Ansett Australia.
2. Monthly government expenditure on pharmaceutical benefits: Federal Government of Australia.

Three case studies

1. Weekly airline passenger traffic: Ansett Australia.
2. Monthly government expenditure on pharmaceutical benefits: Federal Government of Australia.
3. Half-hourly peak electricity demand: Electricity planning council of South Australia.

Three case studies

1. Weekly airline passenger traffic: Ansett Australia.
2. Monthly government expenditure on pharmaceutical benefits: Federal Government of Australia.
3. Half-hourly peak electricity demand: Electricity planning council of South Australia.

Three case studies

1. Weekly airline passenger traffic: Ansett Australia.
2. Monthly government expenditure on pharmaceutical benefits: Federal Government of Australia.
3. Half-hourly peak electricity demand: Electricity planning council of South Australia.

- Highly complex data for which no standard models were suitable.

Three case studies

1. Weekly airline passenger traffic: Ansett Australia.
2. Monthly government expenditure on pharmaceutical benefits: Federal Government of Australia.
3. Half-hourly peak electricity demand: Electricity planning council of South Australia.

- Highly complex data for which no standard models were suitable.
- Real-world complications that textbooks (and often journals) don't cover.

Three case studies

1. Weekly airline passenger traffic: Ansett Australia.
2. Monthly government expenditure on pharmaceutical benefits: Federal Government of Australia.
3. Half-hourly peak electricity demand: Electricity planning council of South Australia.

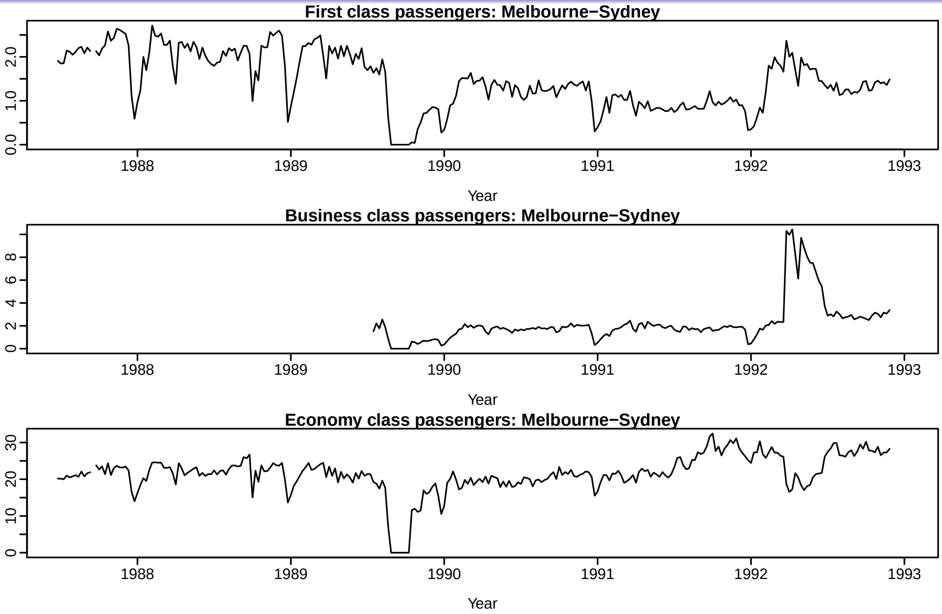
- Highly complex data for which no standard models were suitable.
- Real-world complications that textbooks (and often journals) don't cover.
- Problems motivated some of my research papers.

Outline

1. **Extremely messy data**
2. Extremely expensive forecast errors
3. Extreme electricity demand
4. Extremely useful conclusions

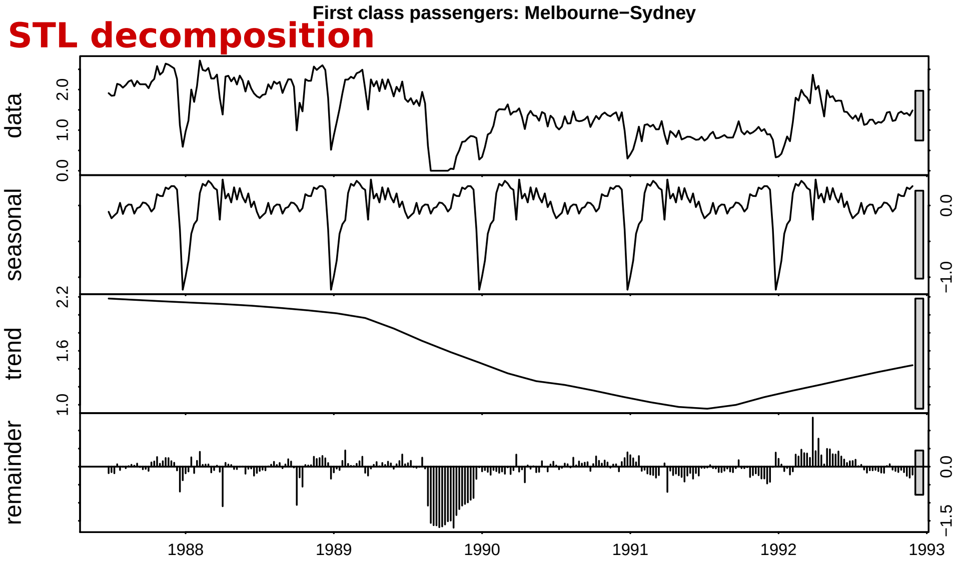
Airline passenger traffic

Airline passenger traffic

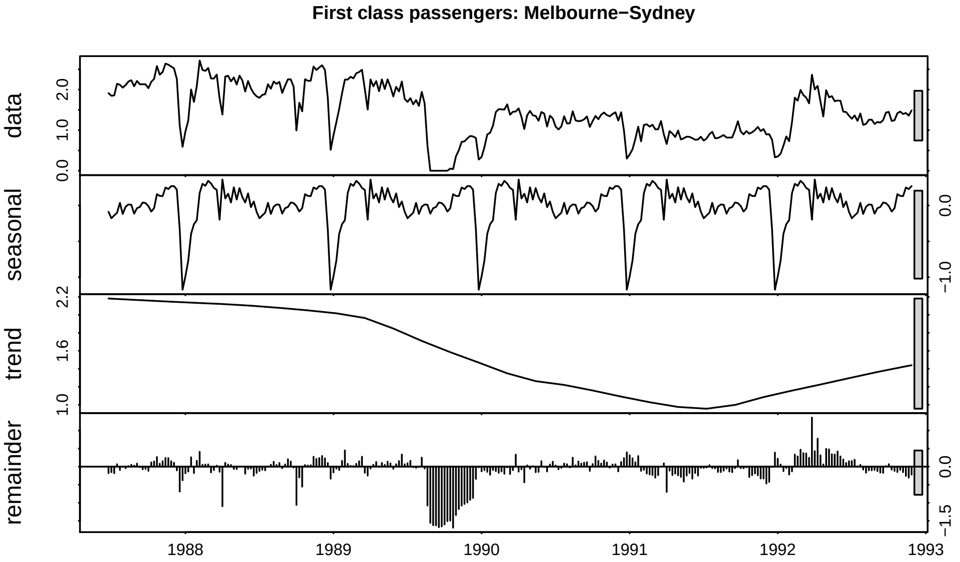


Airline passenger traffic

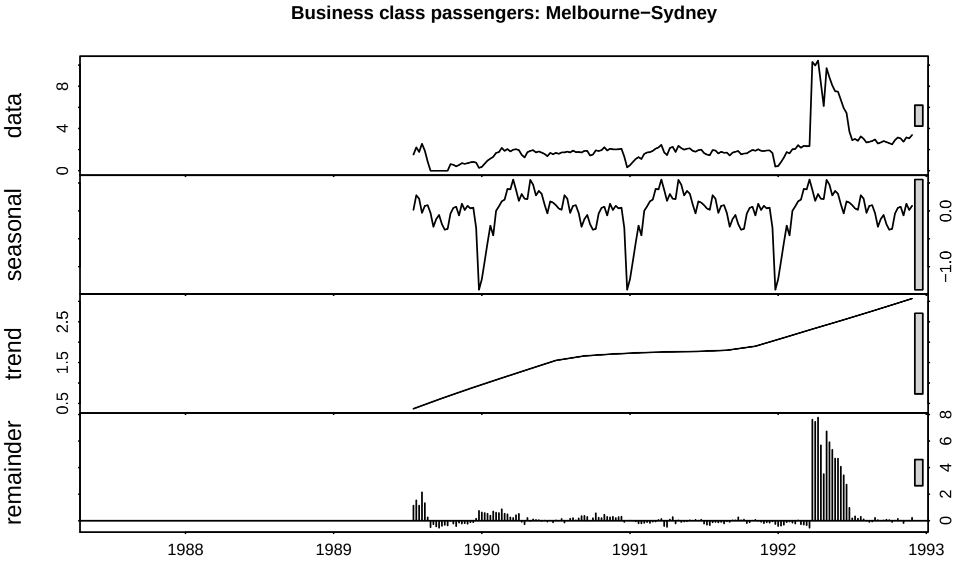
STL decomposition



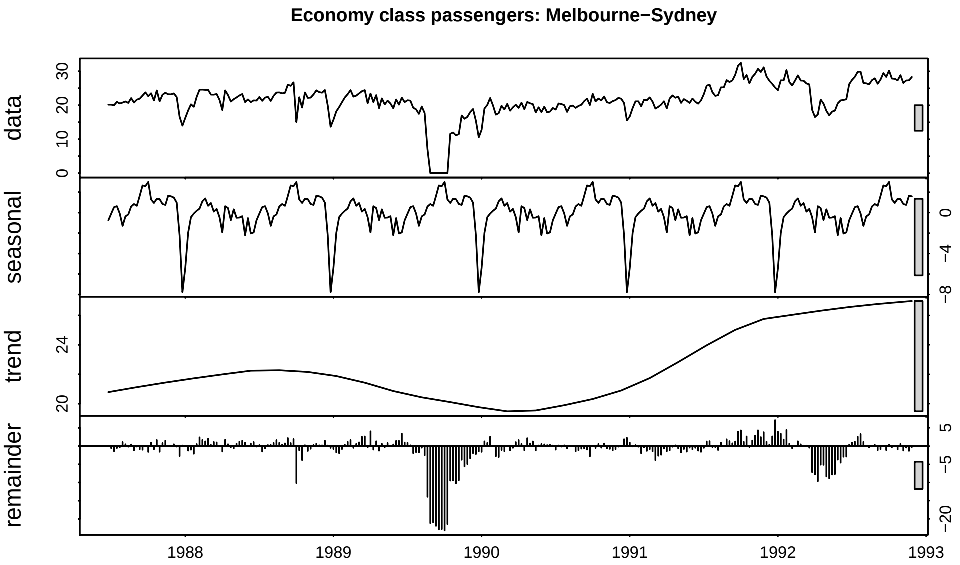
Airline passenger traffic



Airline passenger traffic



Airline passenger traffic



Airline passenger traffic

- Statistical models tend not to be able to allow for the effects seen in these data.

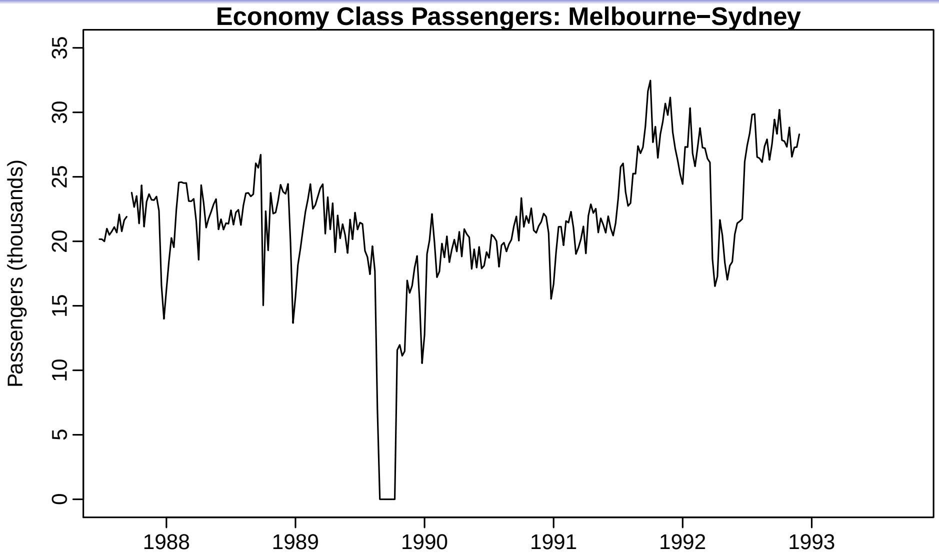
Airline passenger traffic

- Statistical models tend not to be able to allow for the effects seen in these data.
- A common approach is to use “cleaned” data where the time series are reconstructed to look like what might have been if there were no strikes, no changes of definition, etc.

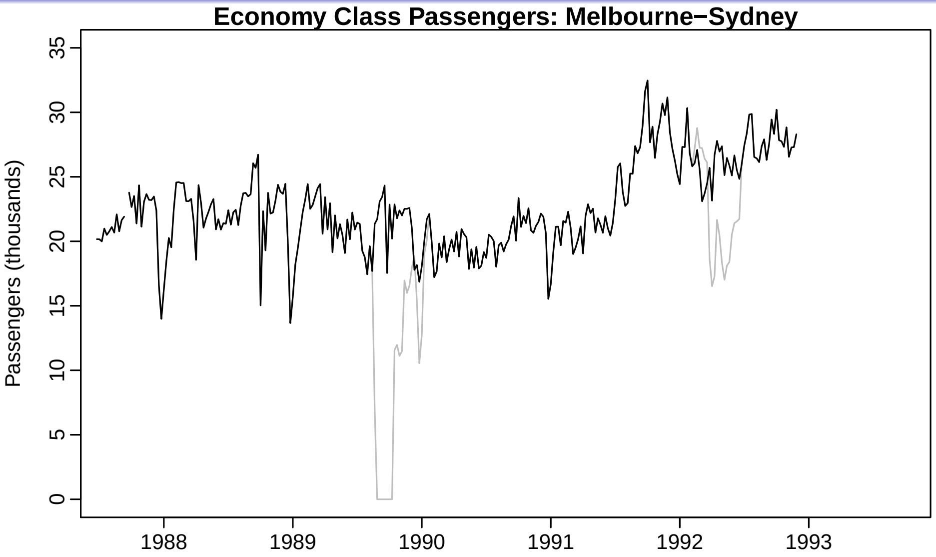
Airline passenger traffic

- Statistical models tend not to be able to allow for the effects seen in these data.
- A common approach is to use “cleaned” data where the time series are reconstructed to look like what might have been if there were no strikes, no changes of definition, etc.
- Probably has small effect on point forecasts, but may have large effect on prediction intervals.

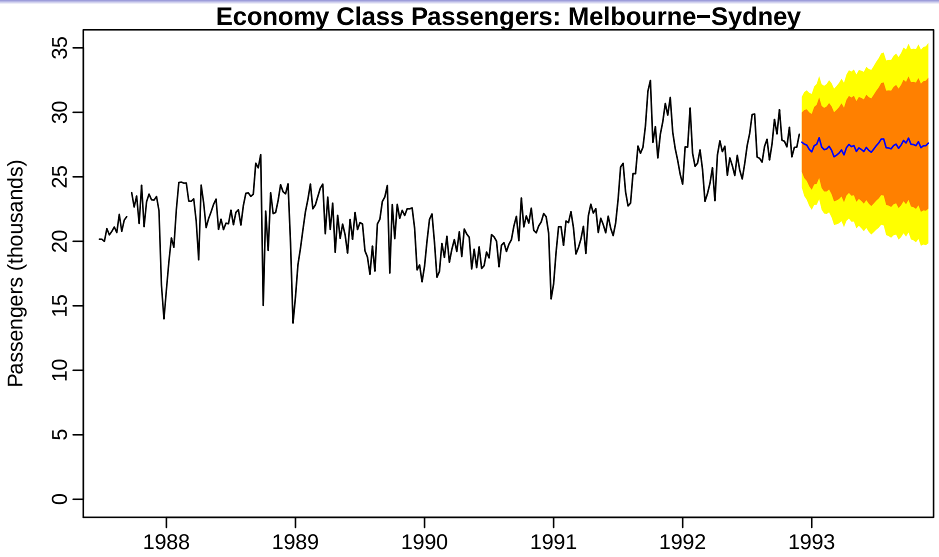
Airline passenger traffic



Airline passenger traffic



Airline passenger traffic



Possible model

$$Y_t = Y_t^* + Z_t$$

$$Y_t^* = \beta_0 + \sum_j \beta_j x_{t,j} + N_t$$

- Y_t = observed data for one passenger class.
- Y_t^* = reconstructed data.
- Z_t = latent process (usually equal to zero).
- $x_{t,j}$ are covariates and dummy variables.
- N_t = seasonal ARIMA process of period 52.

Possible model

$$Y_t = Y_t^* + Z_t$$
$$Y_t^* = \beta_0 + \sum_j \beta_j x_{t,j} + N_t$$

- Y_t = observed data for one passenger class.
- Y_t^* = reconstructed data.
- Z_t = latent process (usually equal to zero).
- $x_{t,j}$ are covariates and dummy variables.
- N_t = seasonal ARIMA process of period 52.

Research issues

How to model Z_t in the presence of industrial action and changing definitions?

Possible model

$$Y_t = Y_t^* + Z_t$$
$$Y_t^* = \beta_0 + \sum_j \beta_j x_{t,j} + N_t$$

- Y_t = observed data for one passenger class.
- Y_t^* = reconstructed data.
- Z_t = latent process (usually equal to zero).
- $x_{t,j}$ are covariates and dummy variables.
- N_t = seasonal ARIMA process of period 52.

Research issues

What to do with the non-integer seasonality? (average 52.19)?

Possible model

$$Y_t = Y_t^* + Z_t$$

$$Y_t^* = \beta_0 + \sum_j \beta_j x_{t,j} + N_t$$

- Y_t = observed data for one passenger class.
- Y_t^* = reconstructed data.
- Z_t = latent process (usually equal to zero).
- $x_{t,j}$ are covariates and dummy variables.
- N_t = seasonal ARIMA process of period 52.

Research issues

How to deal with the correlations between classes and between routes?

Outline

Forecasting the PBS

Forecasting the PBS

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

Forecasting the PBS

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

- Many drugs bought from pharmacies ("drug stores" in the US) are subsidised to allow more equitable access to modern drugs.

Forecasting the PBS

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

- Many drugs bought from pharmacies ("drug stores" in the US) 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.

Forecasting the PBS

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

- Many drugs bought from pharmacies (“drug stores” in the US) 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.

Forecasting the PBS

ABC News Online

AUSTRALIAN BROADCASTING CORPORATION

NewsRadio
Streaming audio news
LISTEN: WMP | Real

Select a Topic from the list below

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

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

- Top Stories
- Just In
- World
- Asia-Pacific
- Business
- Sport
- Arts
- Sci Tech
- Indigenous
- Weather
- Rural
- Local News
- Broadband

Search

SPECIALS

- Federal Election

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 PBS.

the Public Record
For full election coverage

FEATURES

the Public Record
Federal Election 2001

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

Audio News Online

Forecasting the PBS

- In 2001: $4.5 billion budget, under-forecasted by $800 million.

Forecasting the PBS

- In 2001: $4.5 billion budget, under-forecasted by $800 million.
- Thousands of products. Seasonal demand.

Forecasting the PBS

- In 2001: $4.5 billion budget, under-forecasted by $800 million.
- Thousands of products. Seasonal demand.
- Subject to covert marketing, volatile products, uncontrollable expenditure.

Forecasting the PBS

- 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.

Forecasting the PBS

- 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!

Forecasting the PBS

- New drugs forecasts and new policy impacts estimated using judgemental methods.

Forecasting the PBS

- New drugs forecasts and new policy impacts estimated using judgemental methods.
- Monthly data on thousands of drug groups and 4 concession types available from 1991.

Forecasting the PBS

- New drugs forecasts and new policy impacts estimated using judgemental methods.
- Monthly data on thousands of drug groups and 4 concession types available from 1991.
- Method needs to be automated and implemented within Excel.

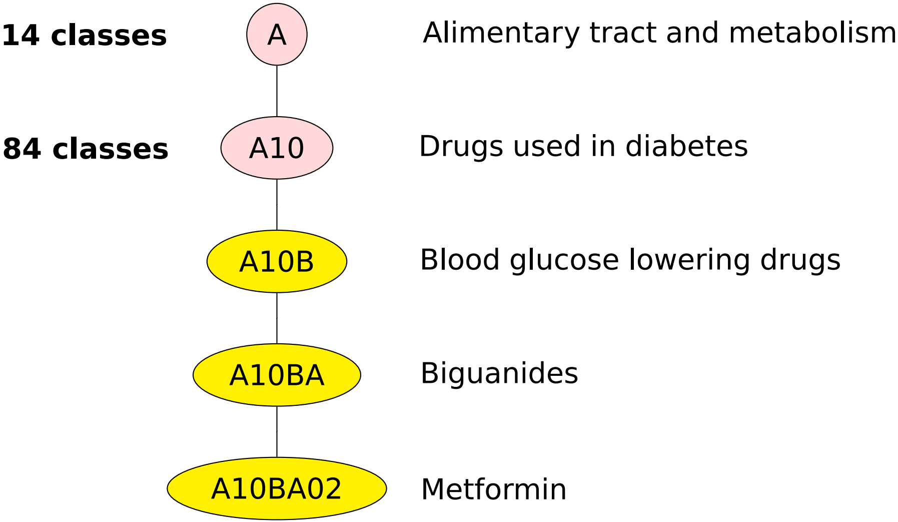
Forecasting the PBS

- New drugs forecasts and new policy impacts estimated using judgemental methods.
- Monthly data on thousands of drug groups and 4 concession types available from 1991.
- Method needs to be automated and implemented within Excel.
- Exponential smoothing seems appropriate (monthly data with changing trends and seasonal patterns), but in 2001, *automated* exponential smoothing was not well-developed.

ATC drug classification

A	Alimentary tract and metabolism
B	Blood and blood forming organs
C	Cardiovascular system
D	Dermatologicals
G	Genito-urinary system and sex hormones
H	Systemic hormonal preparations, excluding sex hormones and insulins
J	Anti-infectives for systemic use
L	Antineoplastic and immunomodulating agents
M	Musculo-skeletal system
N	Nervous system
P	Antiparasitic products, insecticides and repellents
R	Respiratory system
S	Sensory organs
V	Various

ATC drug classification



Forecasting the PBS

Research issues

1. How to automate exponential smoothing to give robust and reliable forecasts for a wide range of time series?

Forecasting the PBS

Research issues

1. How to automate exponential smoothing to give robust and reliable forecasts for a wide range of time series?
2. How to generate prediction intervals for all varieties of exponential smoothing forecasts?

Forecasting the PBS

Research issues

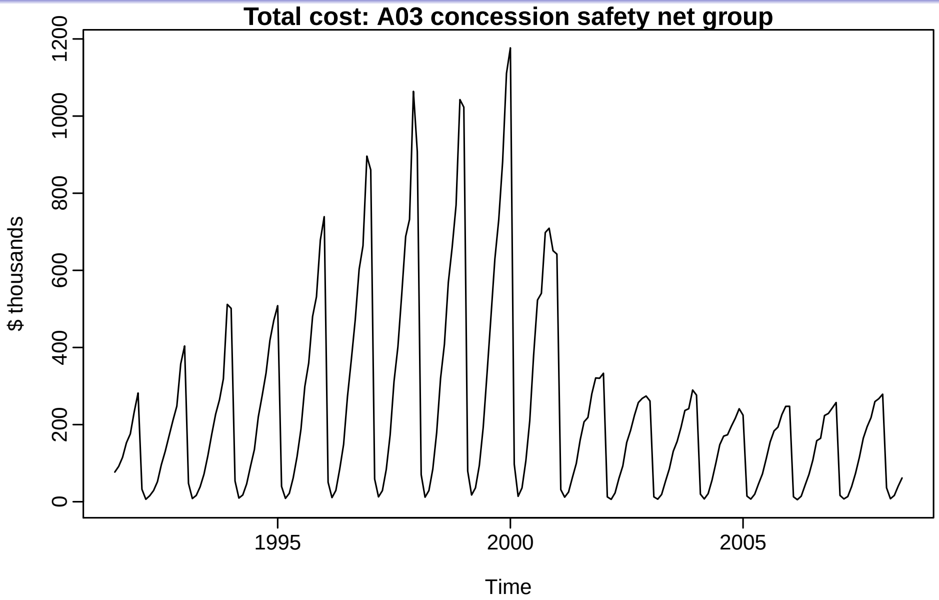
1. How to automate exponential smoothing to give robust and reliable forecasts for a wide range of time series?
2. How to generate prediction intervals for all varieties of exponential smoothing forecasts?
3. Data can be disaggregated at five levels of a drug hierarchy. At what level of disaggregation should we forecast?

Forecasting the PBS

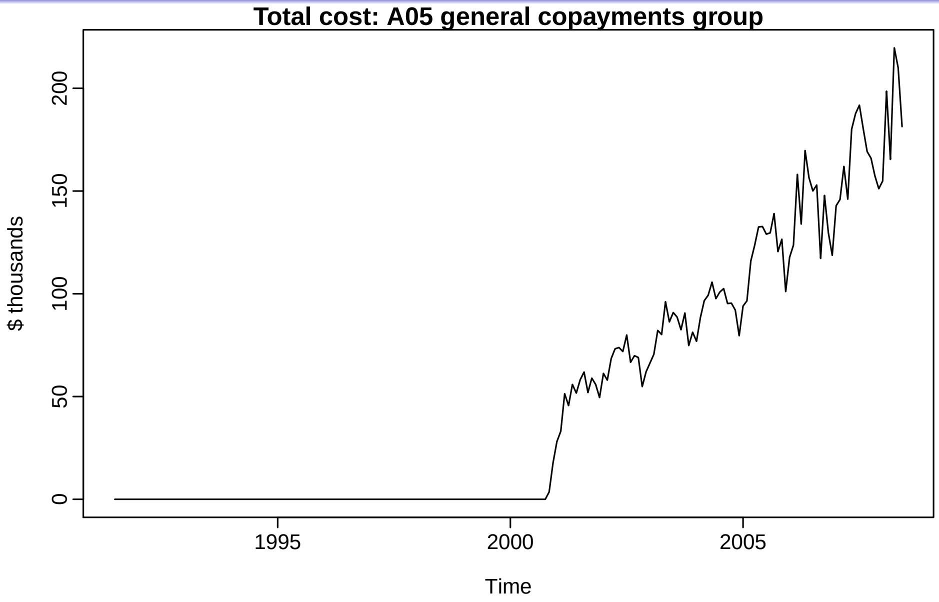
Research issues

1. How to automate exponential smoothing to give robust and reliable forecasts for a wide range of time series?
2. How to generate prediction intervals for all varieties of exponential smoothing forecasts?
3. Data can be disaggregated at five levels of a drug hierarchy. At what level of disaggregation should we forecast?
4. How can correlations and substitutions between drugs be handled across a very large hierarchy of time series?

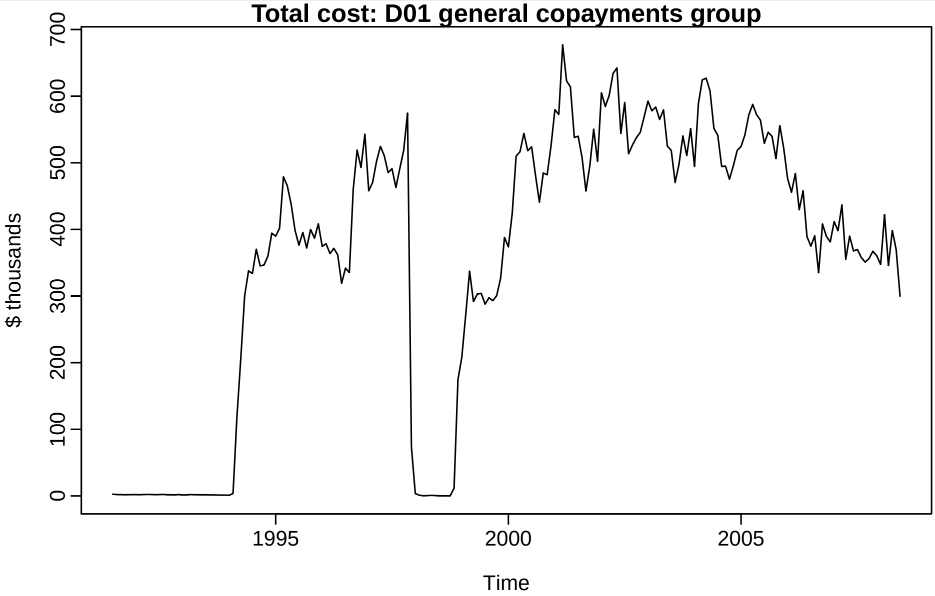
PBS data


Total cost: A03 concession safety net group
$ thousands
0
200
400
600
800
1000
1200
1995
2000
2005
Time

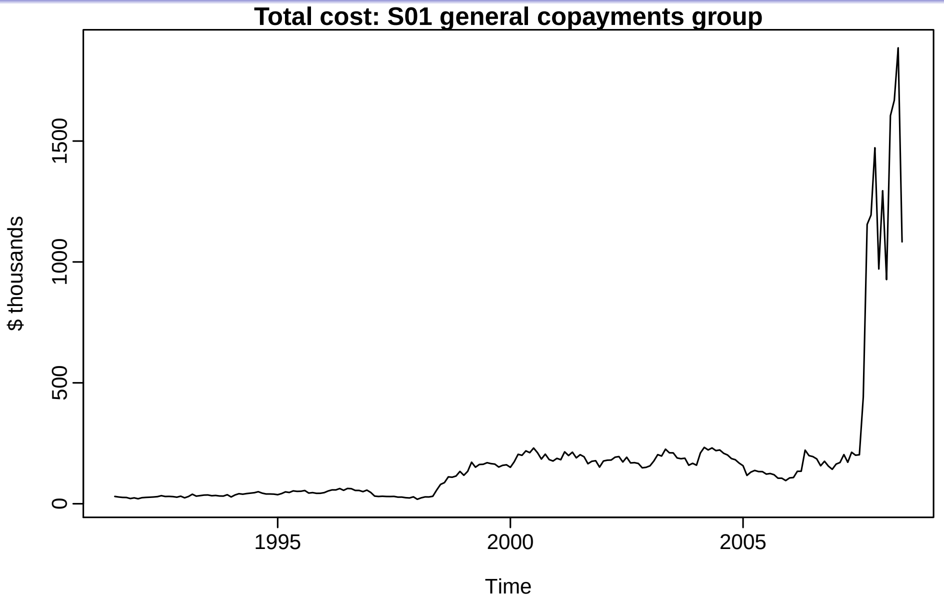
PBS data


Total cost: A05 general copayments group
$ thousands
0
50
100
150
200
1995
2000
2005
Time

PBS data

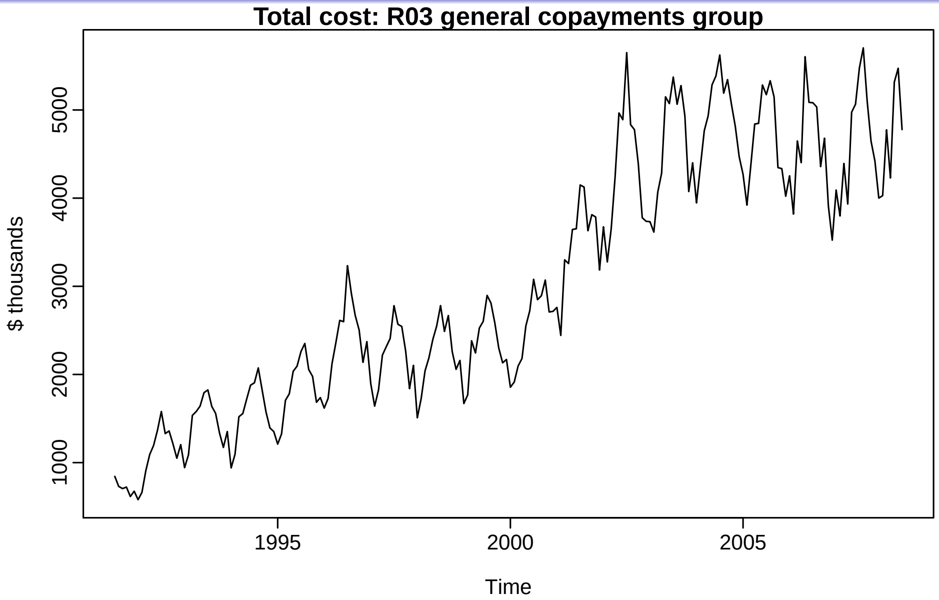

Total cost: D01 general copayments group
$ thousands
0
100
200
300
400
500
600
700
1995
2000
2005
Time

PBS data



Total cost: S01 general copayments group
$ thousands
0
500
1000
1500
1995
2000
2005
Time

PBS data

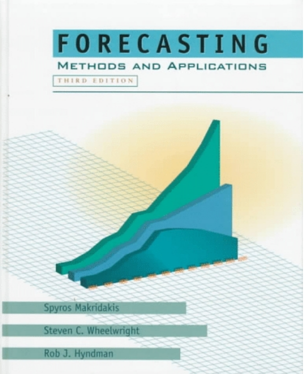

Total cost: R03 general copayments group
$ thousands
1000
2000
3000
4000
5000
1995
2000
2005
Time

Exponential smoothing

Exponential smoothing is extremely popular, simple to implement, and performs well in forecasting competitions.

Exponential smoothing

Exponential smoothing is extremely popular, simple to implement, and performs well in forecasting competitions.



"Unfortunately, exponential smoothing methods do not allow easy calculation of prediction intervals."

Makridakis, Wheelwright and Hyndman, p.177. (Wiley, 3rd ed., 1998)

Exponential smoothing

Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models

J. K. ORD, A. B. KOEHLER, and R. D. SNYDER

A class of nonlinear state-space models, characterized by a single source of randomness, is introduced. A special case, the model underpinning the multiplicative Holt–Winters method of forecasting, is identified. Maximum likelihood estimation based on exponential smoothing instead of a Kalman filter, and with the potential to be applied in contexts involving non-Gaussian disturbances, is considered. A method for computing prediction intervals is proposed and evaluated on both simulated and real data.

KEY WORDS: Forecasting; Holt–Winters method; Maximum likelihood estimation; State-space models.

1. INTRODUCTION

The fact that business and economic time series often possess seasonal cycles with increasing amplitudes together with random fluctuations dependent on movements in underlying levels is a compelling reason to study nonlinear statistical models. Most nonlinear behavior commonly observed in time series plots can be represented by nonlinear state-space models. Being dynamic, the latter have the capacity to reflect many forms of inter temporal dependency encountered in economic time series. Like their linear counterparts (Ansley and Kohn 1985; Harvey and Todd 1983), nonlinear state-space models can be adapted to accommodate the nonstationary features of most economic time series surable quantity y_t in typical period t is governed by the following model:

Model 1:

$$y_t = h(\mathbf{x}_{t-1}, \boldsymbol{\theta}) + k(\mathbf{x}_{t-1}, \boldsymbol{\theta})e_t,$$
$$\text{where} \quad \mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}, \boldsymbol{\theta}) + \mathbf{g}(\mathbf{x}_{t-1}, \boldsymbol{\theta})\mathbf{e}_t.$$

The p vector $\mathbf{x}_{t-1}$ represents the state of the underlying process at the beginning of period t, $\boldsymbol{\theta}$ is a vector of parameters, e_t is from an iid$(0, \sigma^2)$ series of disturbances, h and k are known continuous functions with continuous derivatives mapping from $\Re^p \rightarrow \Re$, and $\mathbf{f}$ and $\mathbf{g}$ are known continuous mappings with continuous derivatives from $\Re^p \rightarrow \Re^p$. It is assumed that the e_t are independent of $\{y_{t-i}, \mathbf{x}_{t-i}, i \geq 1\}$. Although the functions h, k, $\mathbf{f}$, and $\mathbf{g}$ can potentially be in-

Forecasting the PBS

- As part of this project, we developed an automatic forecasting algorithm for exponential smoothing state space models based on the AIC.

Forecasting the PBS

- As part of this project, we 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.

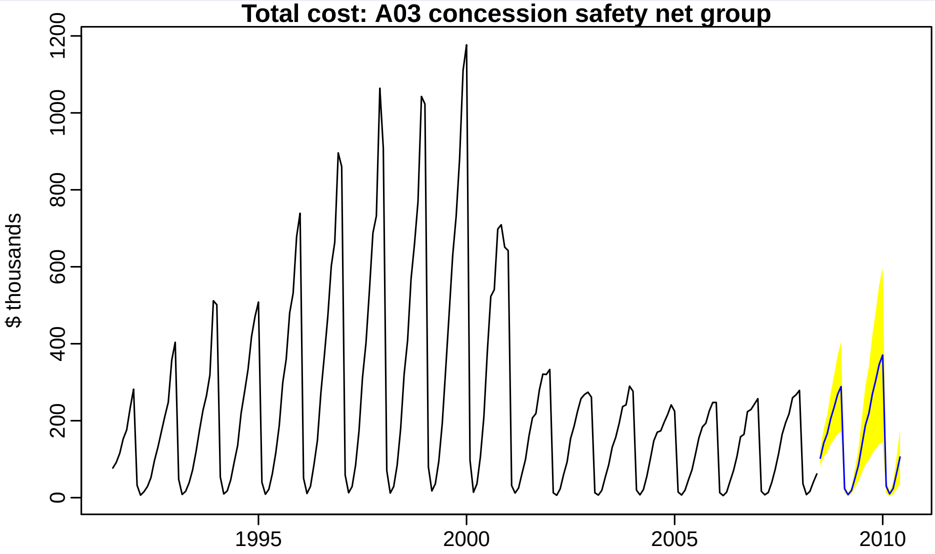
Forecasting the PBS

- As part of this project, we 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.
- State space models provide prediction intervals which give a sense of uncertainty.

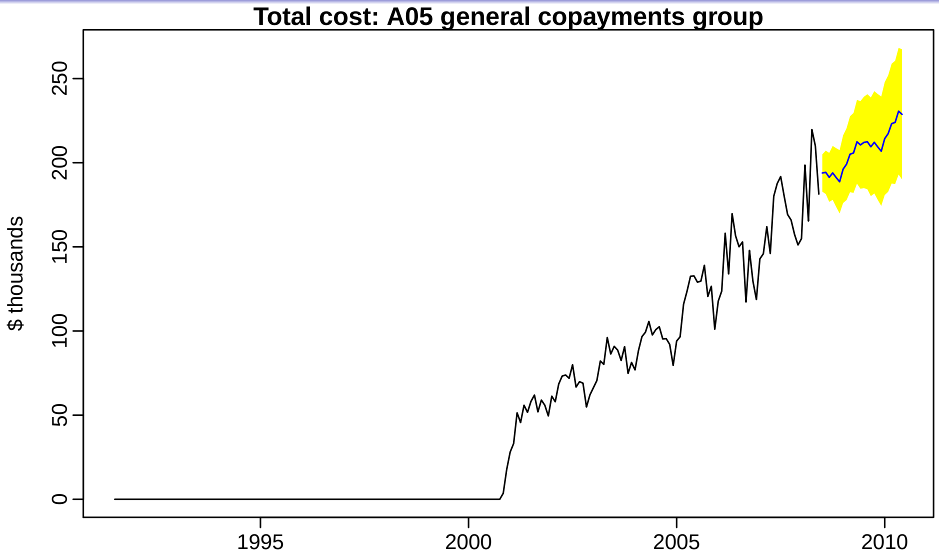
Forecasting the PBS

- As part of this project, we 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.
- State space models provide prediction intervals which give a sense of uncertainty.
- Forecast MAPE reduced from 15–20% to about 0.6%.

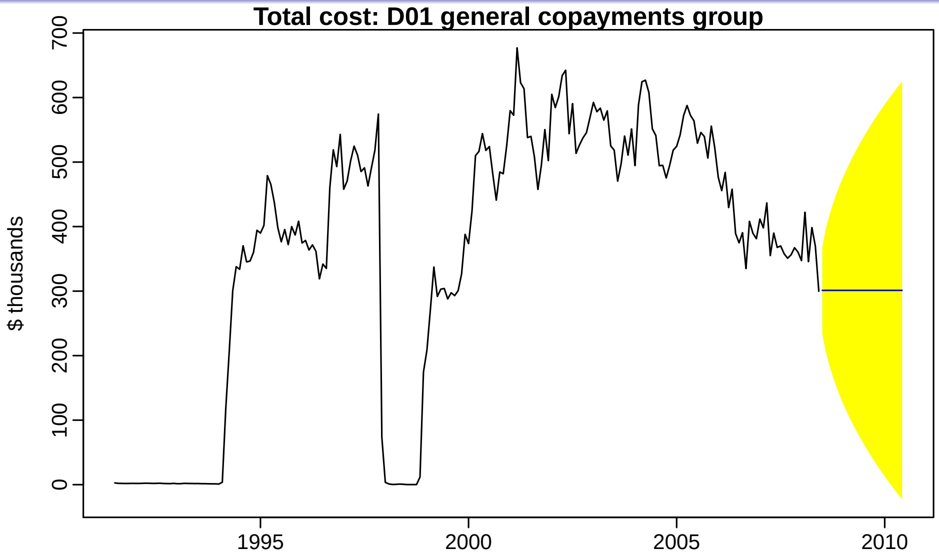
Forecasting the PBS



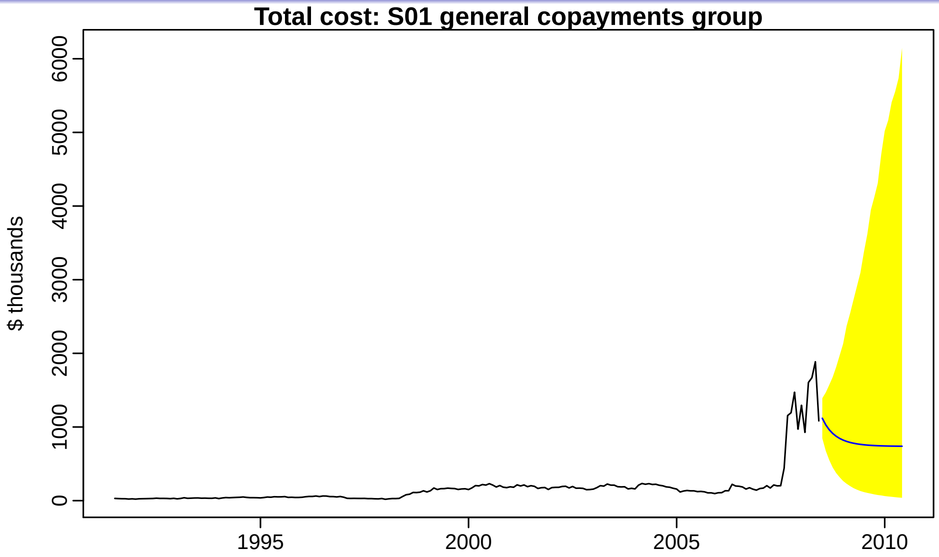
Forecasting the PBS



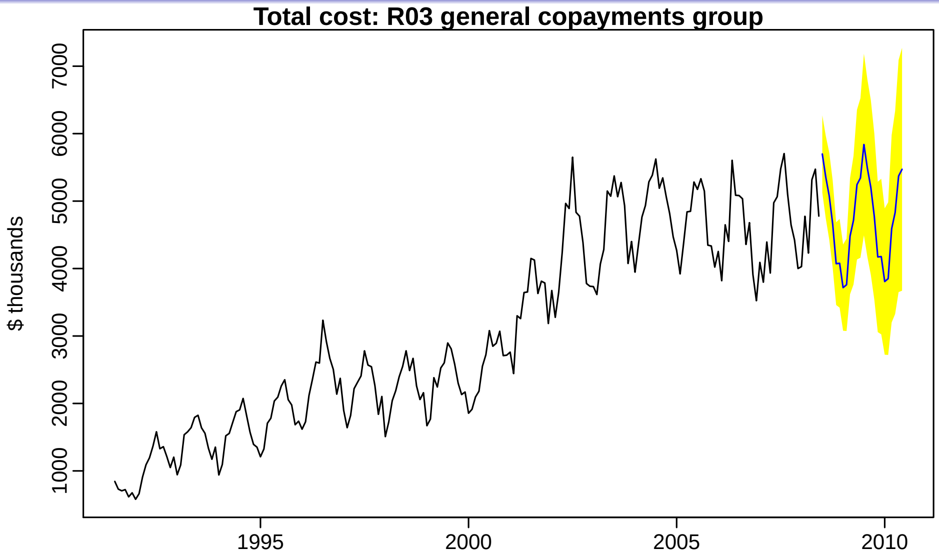
Forecasting the PBS



Forecasting the PBS



Forecasting the PBS



Exponential smoothing

Since 2002. . .

- a general class of state space models proposed underlying all the common exponential smoothing methods.

Exponential smoothing

Since 2002. . .

- a general class of state space models proposed underlying all the common exponential smoothing methods.
- analytical results for prediction intervals.

Exponential smoothing

Since 2002...

- a general class of state space models proposed underlying all the common exponential smoothing methods.
- analytical results for prediction intervals.
- likelihood calculation for estimation.

Exponential smoothing

Since 2002. . .

- a general class of state space models proposed underlying all the common exponential smoothing methods.
- analytical results for prediction intervals.
- likelihood calculation for estimation.
- AIC for model selection.

Exponential smoothing

Since 2002...

- a general class of state space models proposed underlying all the common exponential smoothing methods.
- analytical results for prediction intervals.
- likelihood calculation for estimation.
- AIC for model selection.
- an algorithm for automatic forecasting using the new class of models.

Exponential smoothing

Since 2002. . .

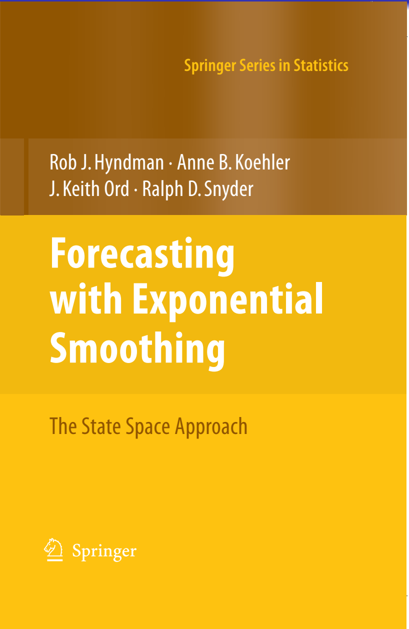
- a general class of state space models proposed underlying all the common exponential smoothing methods.
- analytical results for prediction intervals.
- likelihood calculation for estimation.
- AIC for model selection.
- an algorithm for automatic forecasting using the new class of models.
- new results on admissible parameter space.

Exponential smoothing

Since 2002. . .

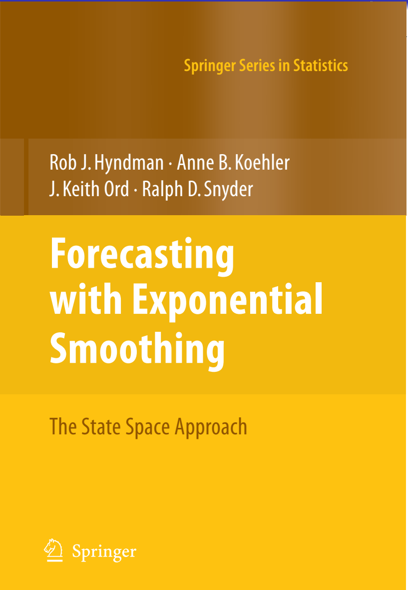
- a general class of state space models proposed underlying all the common exponential smoothing methods.
- analytical results for prediction intervals.
- likelihood calculation for estimation.
- AIC for model selection.
- an algorithm for automatic forecasting using the new class of models.
- new results on admissible parameter space.
- algorithm now part of **forecast** package in **R**.

Recent book



- State space modeling framework
- Prediction intervals
- Model selection
- Maximum likelihood estimation
- All the important research results in one place with consistent notation
- Many new results
- 375 pages but only US$54.95 / £33.99 / €38.88

Recent book



- State space modeling framework
- Prediction intervals
- Model selection
- Maximum likelihood estimation
- All the important research results in one place with consistent notation
- Many new results
- 375 pages but only US$54.95 / £33.99 / €38.88

www.exponentialsmoothing.net

Forecasting the PBS

Research issues

1. How to automate exponential smoothing to give robust and reliable forecasts for a wide range of time series?
2. How to generate prediction intervals for all varieties of exponential smoothing forecasts?
3. Data can be disaggregated at five levels of a drug hierarchy. At what level of disaggregation should we forecast?
4. How can correlations and substitutions between drugs be handled across a very large hierarchy of time series?

Forecasting the PBS

Research issues

1. How to automate exponential smoothing to give robust and reliable forecasts for a wide range of time series?
2. How to generate prediction intervals for all varieties of exponential smoothing forecasts?
3. Data can be disaggregated at five levels of a drug hierarchy. At what level of disaggregation should we forecast?
4. How can correlations and substitutions between drugs be handled across a very large hierarchy of time series?

Forecasting the PBS

Research issues

1. How to automate exponential smoothing to give robust and reliable forecasts for a wide range of time series?
2. How to generate prediction intervals for all varieties of exponential smoothing forecasts?
3. Data can be disaggregated at five levels of a drug hierarchy. At what level of disaggregation should we forecast?
4. How can correlations and substitutions between drugs be handled across a very large hierarchy of time series?

Forecasting the PBS

Research issues

✓ 1. How to automate exponential smoothing to give robust and reliable forecasts for a wide range of time series?

✓ 2. How to generate prediction intervals for all varieties of exponential smoothing forecasts?

? 3. Data can be disaggregated at five levels of a drug hierarchy. At what level of disaggregation should we forecast?

4. How can correlations and substitutions between drugs be handled across a very large hierarchy of time series?

Forecasting the PBS

Research issues

1. ✓ How to automate exponential smoothing to give robust and reliable forecasts for a wide range of time series?
2. ✓ How to generate prediction intervals for all varieties of exponential smoothing forecasts?
3. ? Data can be disaggregated at five levels of a drug hierarchy. At what level of disaggregation should we forecast?
4. ? How can correlations and substitutions between drugs be handled across a very large hierarchy of time series?

Outline

1. Extremely messy data
2. Extremely expensive forecast errors
3. **Extreme electricity demand**
4. Extremely useful conclusions

Extreme electricity demand

Joint work with
Dr Shu Fan

The problem

- We want to forecast the peak electricity demand in a half-hour period in ten years time.

The problem

- We want to forecast the peak electricity demand in a half-hour period in ten years time.
- We have twelve years of half-hourly electricity data, temperature data and some economic and demographic data.

The problem

- We want to forecast the peak electricity demand in a half-hour period in ten years time.
- We have twelve years of half-hourly electricity data, temperature data and some economic and demographic data.
- The location is South Australia: home to the most volatile electricity demand in the world.

The problem

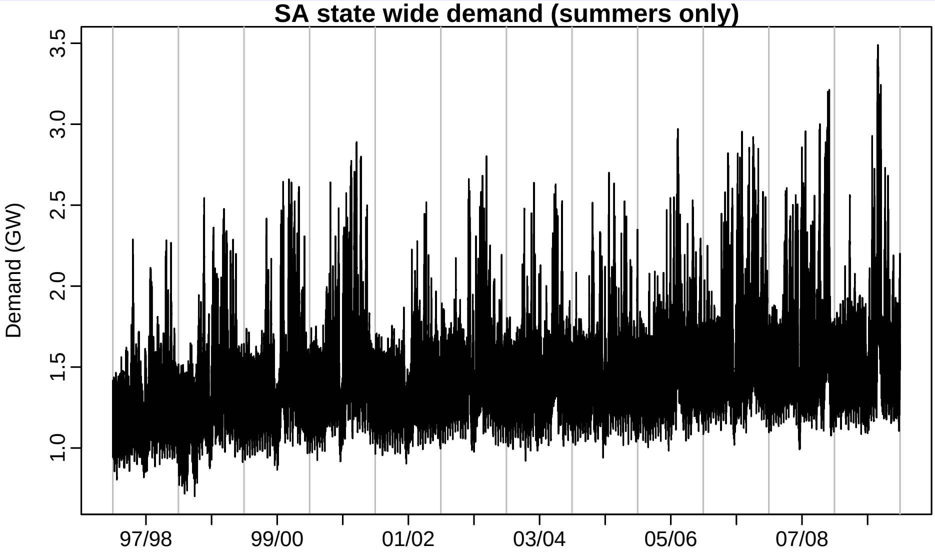
- We want to forecast the peak electricity demand in a half-hour period in ten years time.
- We have twelve years of half-hourly electricity data, temperature data and some economic and demographic data.
- The location is South Australia: home to the most volatile electricity demand in the world.

The problem

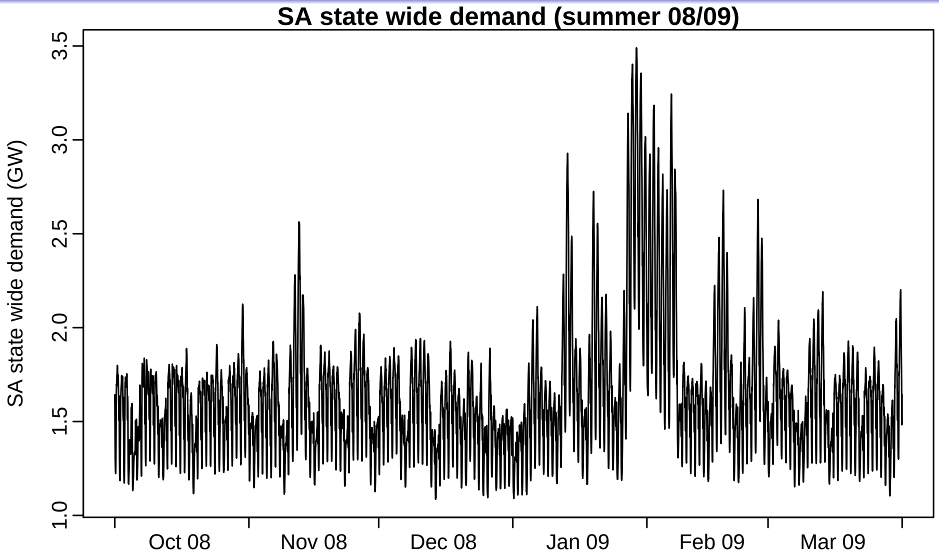
- We want to forecast the peak electricity demand in a half-hour period in ten years time.
- We have twelve years of half-hourly electricity data, temperature data and some economic and demographic data.
- The location is South Australia: home to the most volatile electricity demand in the world.

Sounds impossible?

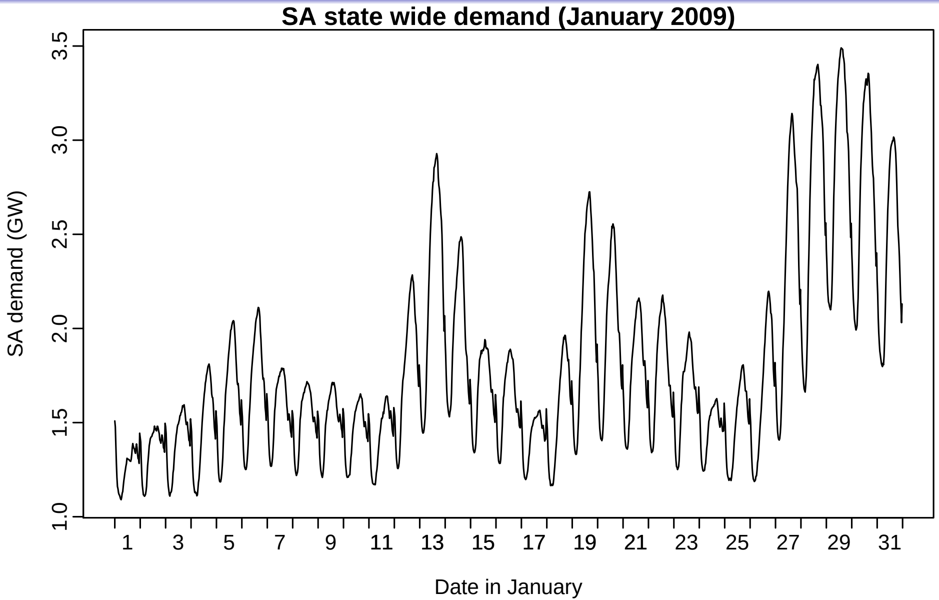
Demand data


SA state wide demand (summers only)
Demand (GW)
1.0
1.5
2.0
2.5
3.0
3.5
97/98
99/00
01/02
03/04
05/06
07/08

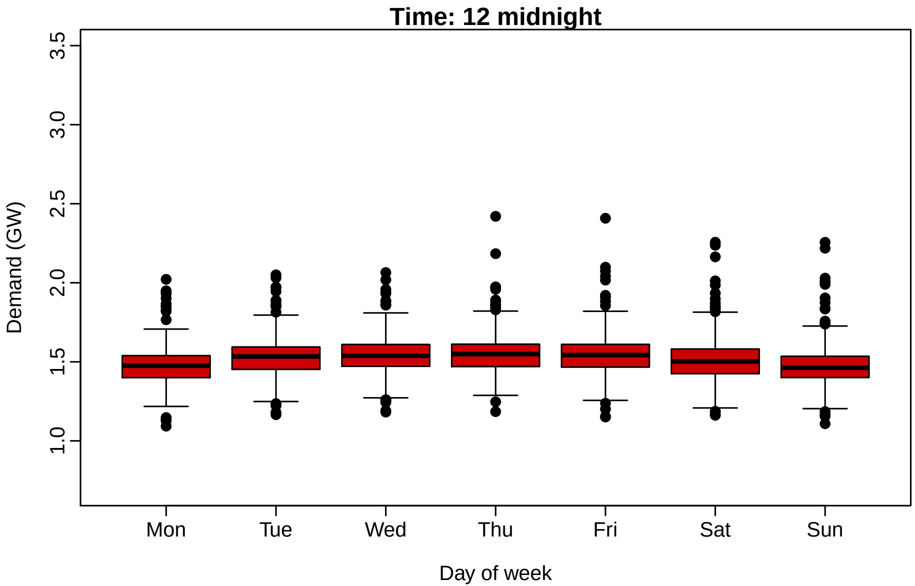
Demand data


SA state wide demand (summer 08/09)
SA state wide demand (GW)
1.0
1.5
2.0
2.5
3.0
3.5
Oct 08
Nov 08
Dec 08
Jan 09
Feb 09
Mar 09

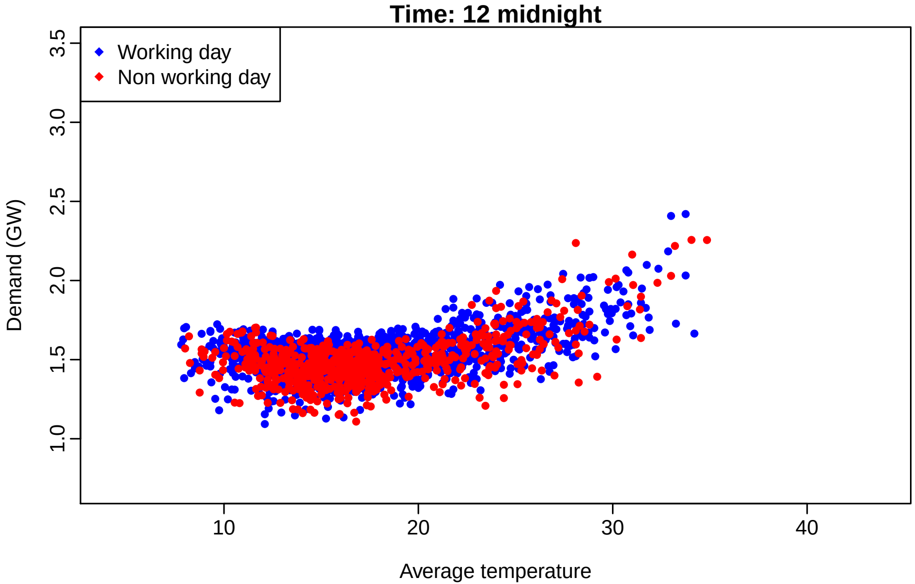
Demand data


SA state wide demand (January 2009)
SA demand (GW)
1.0
1.5
2.0
2.5
3.0
3.5
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
Date in January

Demand boxplots



Temperature data



Demand drivers

- calendar effects

Demand drivers

- calendar effects
- prevailing and recent weather conditions

Demand drivers

- calendar effects
- prevailing and recent weather conditions
- climate changes

Demand drivers

- calendar effects
- prevailing and recent weather conditions
- climate changes
- economic and demographic changes

Demand drivers

- calendar effects
- prevailing and recent weather conditions
- climate changes
- economic and demographic changes
- changing technology

Demand drivers

- calendar effects
- prevailing and recent weather conditions
- climate changes
- economic and demographic changes
- changing technology

Demand drivers

- calendar effects
- prevailing and recent weather conditions
- climate changes
- economic and demographic changes
- changing technology

Modelling framework

- **Semi-parametric additive models** with correlated errors.

Demand drivers

- calendar effects
- prevailing and recent weather conditions
- climate changes
- economic and demographic changes
- changing technology

Modelling framework

- **Semi-parametric additive models** with correlated errors.
- Each half-hour period modelled separately.

Demand drivers

- calendar effects
- prevailing and recent weather conditions
- climate changes
- economic and demographic changes
- changing technology

Modelling framework

- **Semi-parametric additive models** with correlated errors.
- Each half-hour period modelled separately.
- Variables selected to provide best out-of-sample predictions using cross-validation on each summer.

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

- $y_{t,p}$ denotes per capita demand at time t (measured in half-hourly intervals) during period p, $p = 1, \dots, 48$;

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

- $y_{t,p}$ denotes per capita demand at time t (measured in half-hourly intervals) during period p, $p = 1, \ldots, 48$;
- $h_p(t)$ models all calendar effects;

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

- $y_{t,p}$ denotes per capita demand at time t (measured in half-hourly intervals) during period p, $p = 1, \dots, 48$;
- $h_p(t)$ models all calendar effects;
- $f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t})$ models all temperature effects where $\mathbf{w}_{1,t}$ is a vector of recent temperatures at location 1 and $\mathbf{w}_{2,t}$ is a vector of recent temperatures at location 2;

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

- $y_{t,p}$ denotes per capita demand at time t (measured in half-hourly intervals) during period p, $p = 1, \dots, 48$;
- $h_p(t)$ models all calendar effects;
- $f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t})$ models all temperature effects where $\mathbf{w}_{1,t}$ is a vector of recent temperatures at location 1 and $\mathbf{w}_{2,t}$ is a vector of recent temperatures at location 2;
- $z_{j,t}$ is a demographic or economic variable at time t

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

- $y_{t,p}$ denotes per capita demand at time t (measured in half-hourly intervals) during period p, $p = 1, \dots, 48$;
- $h_p(t)$ models all calendar effects;
- $f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t})$ models all temperature effects where $\mathbf{w}_{1,t}$ is a vector of recent temperatures at location 1 and $\mathbf{w}_{2,t}$ is a vector of recent temperatures at location 2;
- $z_{j,t}$ is a demographic or economic variable at time t
- n_t denotes the model error at time t.

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

$h_p(t)$ includes handle annual, weekly and daily seasonal patterns as well as public holidays:

$$h_p(t) = \ell_p(t) + \alpha_{t,p} + \beta_{t,p}$$

- $\ell_p(t)$ is “time of summer” effect (a regression spline);

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

$h_p(t)$ includes handle annual, weekly and daily seasonal patterns as well as public holidays:

$$h_p(t) = \ell_p(t) + \alpha_{t,p} + \beta_{t,p}$$

- $\ell_p(t)$ is "time of summer" effect (a regression spline);
- $\alpha_{t,p}$ is day of week effect;

Equations

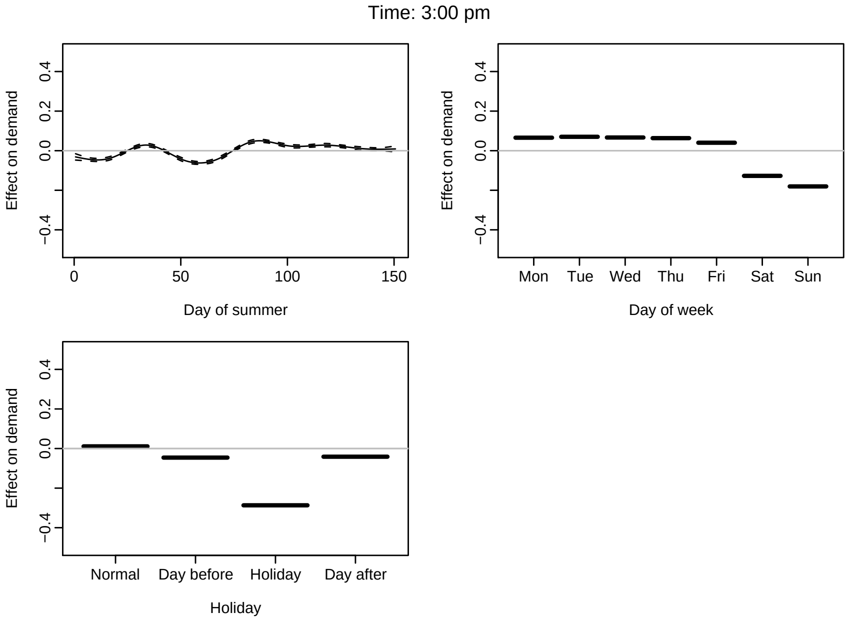
$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

$h_p(t)$ includes handle annual, weekly and daily seasonal patterns as well as public holidays:

$$h_p(t) = \ell_p(t) + \alpha_{t,p} + \beta_{t,p}$$

- $\ell_p(t)$ is “time of summer” effect (a regression spline);
- $\alpha_{t,p}$ is day of week effect;
- $\beta_{t,p}$ is “holiday” effect;

Fitted results (3pm)



Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

$$\begin{aligned} f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) = &\sum_{k=0}^{6} \left[f_{k,p}(x_{t-k}) + g_{k,p}(d_{t-k}) \right] + q_p(x_t^+) + r_p(x_t^-) + s_p(\bar{x}_t) \\ &+ \sum_{j=1}^{6} \left[F_{j,p}(x_{t-48j}) + G_{j,p}(d_{t-48j}) \right] \end{aligned}$$

- x_t is ave temp across sites at time t;

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

$$f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) = \sum_{k=0}^{6} \left[f_{k,p}(x_{t-k}) + g_{k,p}(d_{t-k}) \right] + q_p(x_t^+) + r_p(x_t^-) + s_p(\bar{x}_t) + \sum_{j=1}^{6} \left[F_{j,p}(x_{t-48j}) + G_{j,p}(d_{t-48j}) \right]$$

- x_t is ave temp across sites at time t;
- d_t is the temp difference between sites at time t;

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

$$\begin{aligned} f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) = \sum_{k=0}^{6} \left[f_{k,p}(x_{t-k}) + g_{k,p}(d_{t-k}) \right] + q_p(x_t^+) + r_p(x_t^-) + s_p(\bar{x}_t) \\ + \sum_{j=1}^{6} \left[F_{j,p}(x_{t-48j}) + G_{j,p}(d_{t-48j}) \right] \end{aligned}$$

- x_t is ave temp across sites at time t;
- d_t is the temp difference between sites at time t;
- x_t^+ is max of x_t values in past 24 hours;

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

$$f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) = \sum_{k=0}^{6} \left[f_{k,p}(x_{t-k}) + g_{k,p}(d_{t-k}) \right] + q_p(x_t^+) + r_p(x_t^-) + s_p(\bar{x}_t) + \sum_{j=1}^{6} \left[F_{j,p}(x_{t-48j}) + G_{j,p}(d_{t-48j}) \right]$$

- x_t is ave temp across sites at time t;
- d_t is the temp difference between sites at time t;
- x_t^+ is max of x_t values in past 24 hours;
- x_t^- is min of x_t values in past 24 hours;

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

$$\begin{aligned} f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) = \sum_{k=0}^{6} \big[f_{k,p}(x_{t-k}) + g_{k,p}(d_{t-k})\big] + q_p(x_t^+) + r_p(x_t^-) + s_p(\bar{x}_t) \\ + \sum_{j=1}^{6} \big[F_{j,p}(x_{t-48j}) + G_{j,p}(d_{t-48j})\big] \end{aligned}$$

- x_t is ave temp across sites at time t;
- d_t is the temp difference between sites at time t;
- x_t^+ is max of x_t values in past 24 hours;
- x_t^- is min of x_t values in past 24 hours;
- $\bar{x}_t$ is ave temp in past seven days.

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

$$f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) = \sum_{k=0}^{6} \left[f_{k,p}(x_{t-k}) + g_{k,p}(d_{t-k})\right] + q_p(x_t^+) + r_p(x_t^-) + s_p(\bar{x}_t) + \sum_{j=1}^{6} \left[F_{j,p}(x_{t-48j}) + G_{j,p}(d_{t-48j})\right]$$

- x_t is ave temp across sites at time t;
- d_t is the temp difference between sites at time t;
- x_t^+ is max of x_t values in past 24 hours;
- x_t^- is min of x_t values in past 24 hours;
- $\bar{x}_t$ is ave temp in past seven days.

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

$$f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) = \sum_{k=0}^{6} \left[f_{k,p}(x_{t-k}) + g_{k,p}(d_{t-k})\right] + q_p(x_t^+) + r_p(x_t^-) + s_p(\bar{x}_t) + \sum_{j=1}^{6} \left[F_{j,p}(x_{t-48j}) + G_{j,p}(d_{t-48j})\right]$$

- x_t is ave temp across sites at time t;
- d_t is the temp difference between sites at time t;
- x_t^+ is max of x_t values in past 24 hours;
- x_t^- is min of x_t values in past 24 hours;
- $\bar{x}_t$ is ave temp in past seven days.

Each function is smooth and estimated using regression splines.

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

- Other variables described by linear relationships with coefficients $c_1, \dots, c_J$.

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

- Other variables described by linear relationships with coefficients $c_1, \dots, c_J$.
- Estimation based on annual data.

Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

- Other variables described by linear relationships with coefficients $c_1, \dots, c_J$.
- Estimation based on annual data.

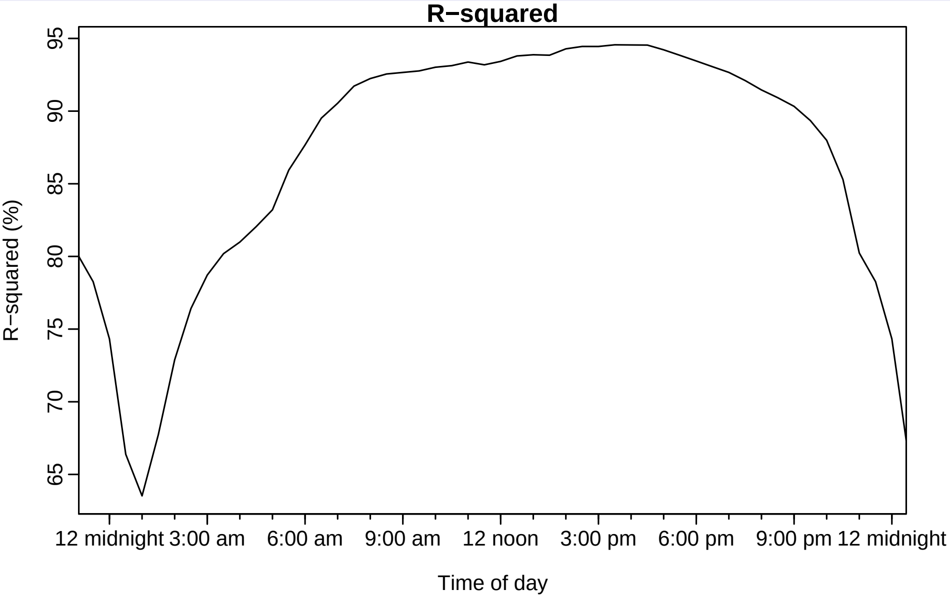
Equations

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

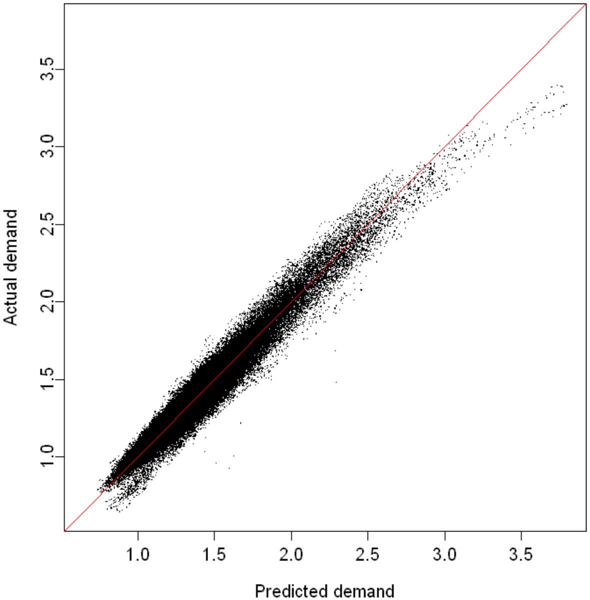
- Other variables described by linear relationships with coefficients $c_1, \dots, c_J$.
- Estimation based on annual data.

Variable	Coefficient	Std. Error	P value
Per-capita GSP	18.56	1.174	0.000
Lag Price	-0.0179	0.00491	0.002
Cooling Degree Days ($\times$1000)	0.1002	0.00456	0.043

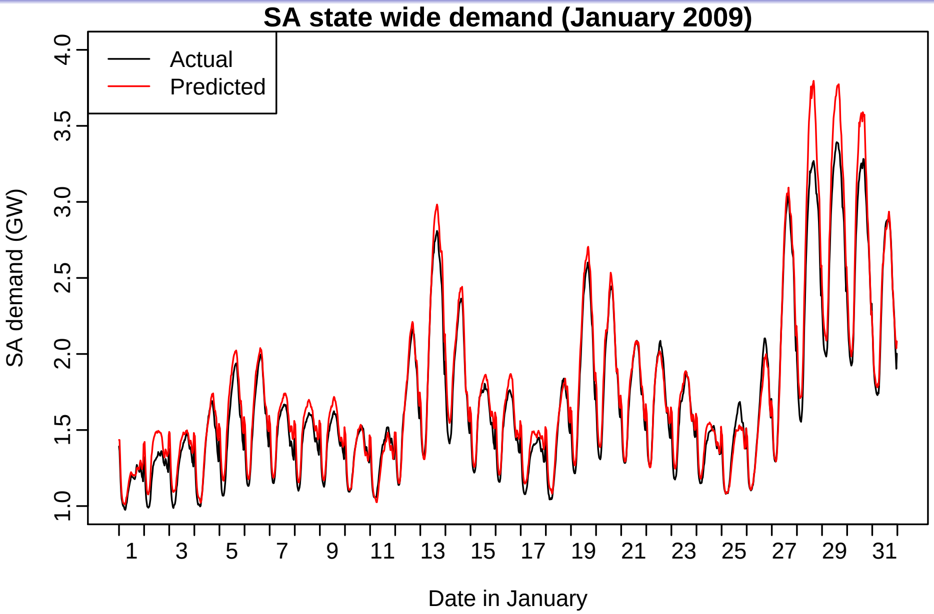
Predictions



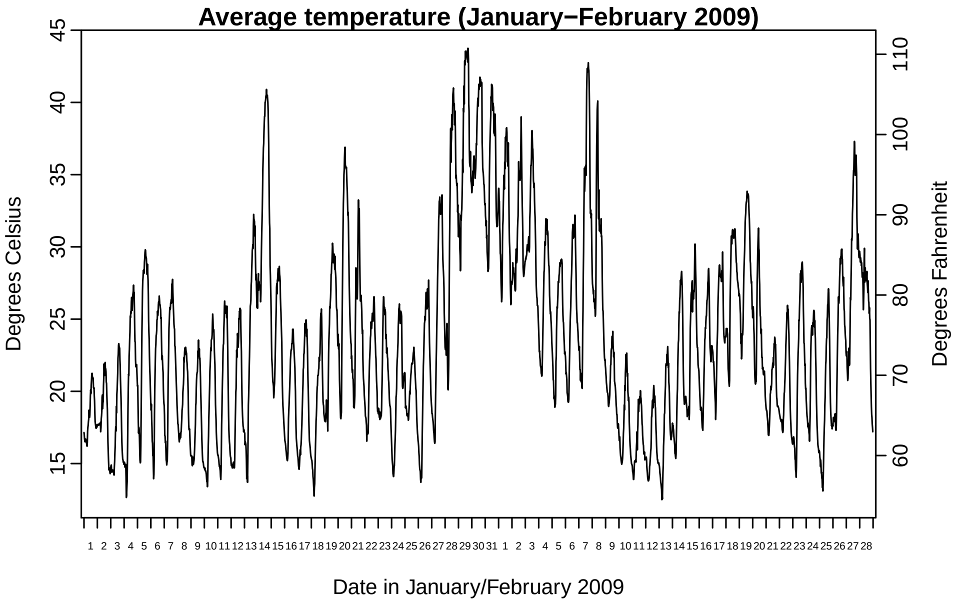
Predictions



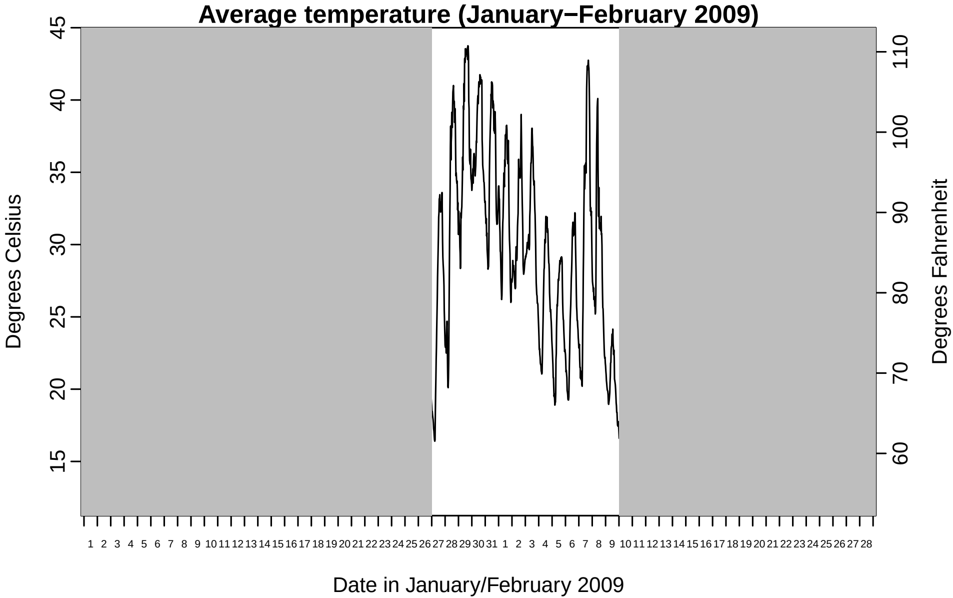
Predictions



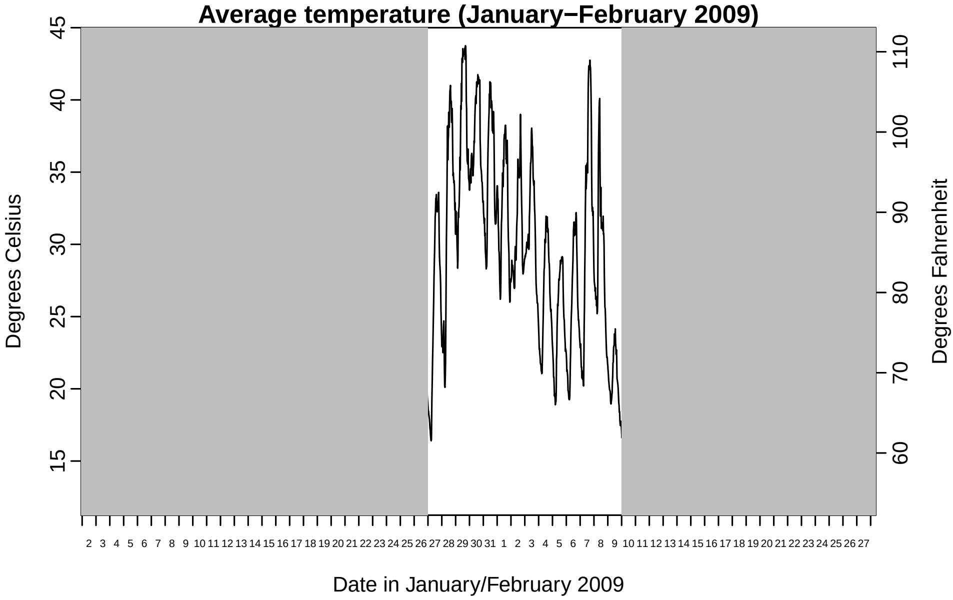
The heatwave



The heatwave



The heatwave



The heatwave

The heatwave

The heatwave

Modified predictions

Original model

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

Modified predictions

Original model

$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

Model allowing saturated usage

$$q_{t,p} = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

$$\log(y_{t,p}) = \begin{cases} q_{t,p} & \text{if } q_{t,p} \le \tau; \\ \tau + k(q_{t,p} - \tau) & \text{if } q_{t,p} > \tau. \end{cases}$$

Peak demand forecasting

$$q_{t,p} = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

$$\log(y_{t,p}) = \begin{cases} q_{t,p} & \text{if } q_{t,p} \le \tau; \\ \tau + k(q_{t,p} - \tau) & \text{if } q_{t,p} > \tau. \end{cases}$$

Multiple alternative futures created:

- resample residuals using seasonal bootstrap;

Peak demand forecasting

$$q_{t,p} = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

$$\log(y_{t,p}) = \begin{cases} q_{t,p} & \text{if } q_{t,p} \le \tau; \\ \tau + k(q_{t,p} - \tau) & \text{if } q_{t,p} > \tau. \end{cases}$$

Multiple alternative futures created:

- resample residuals using seasonal bootstrap;
- simulate future temperatures using seasonal bootstrap (with adjustment for climate change);

Peak demand forecasting

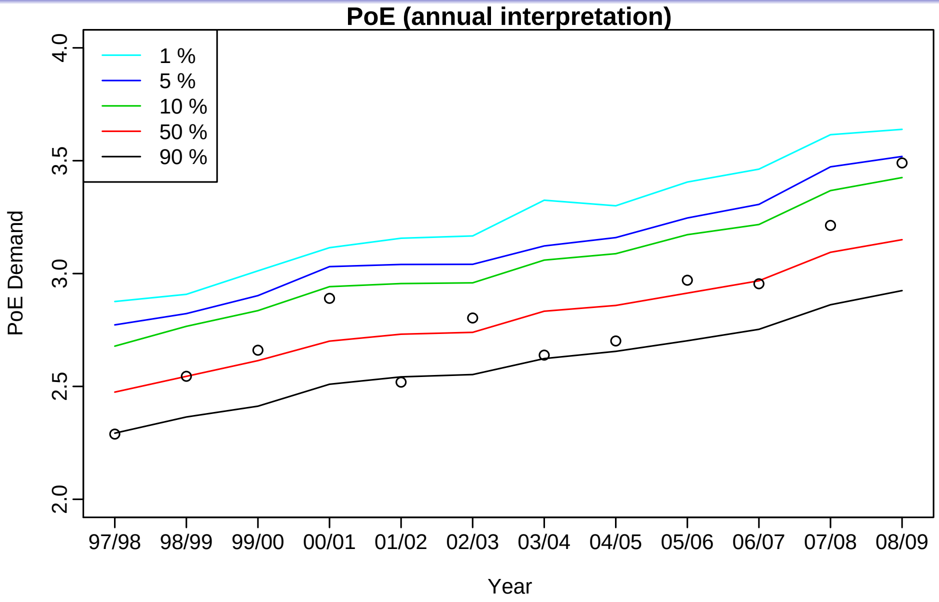
$$q_{t,p} = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

$$\log(y_{t,p}) = \begin{cases} q_{t,p} & \text{if } q_{t,p} \le \tau; \\ \tau + k(q_{t,p} - \tau) & \text{if } q_{t,p} > \tau. \end{cases}$$

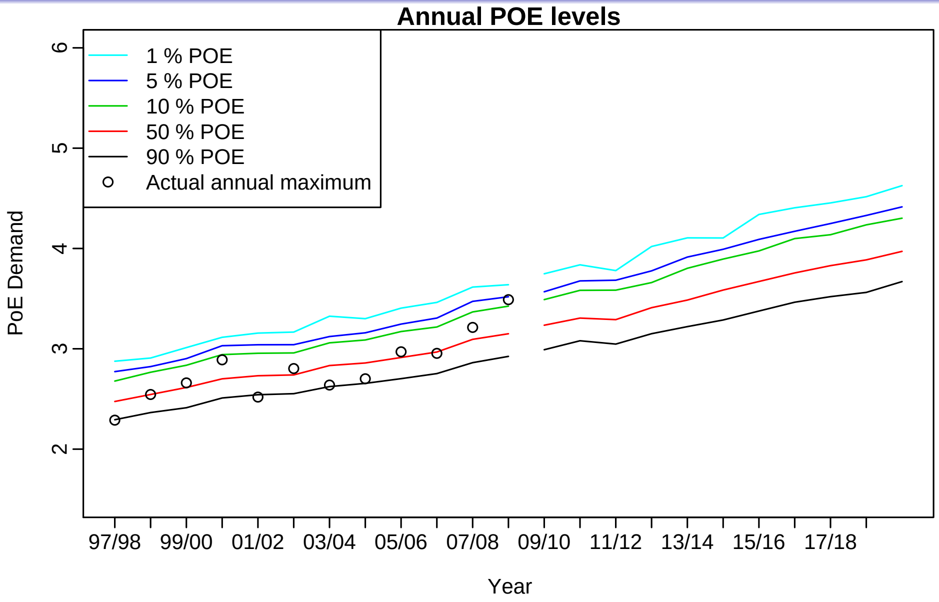
Multiple alternative futures created:

- resample residuals using seasonal bootstrap;
- simulate future temperatures using seasonal bootstrap (with adjustment for climate change);
- use assumed values for GSP and Price.

Peak demand distribution



Peak demand distribution



Results

- We have successfully forecast the extreme upper tail in ten years time using only twelve years of data!

Results

- We have successfully forecast the extreme upper tail in ten years time using only twelve years of data!
- This method has now been adopted for the official South Australian and Victorian long-term peak electricity demand forecasts.

Results

- We have successfully forecast the extreme upper tail in ten years time using only twelve years of data!
- This method has now been adopted for the official South Australian and Victorian long-term peak electricity demand forecasts.

Results

- We have successfully forecast the extreme upper tail in ten years time using only twelve years of data!
- This method has now been adopted for the official South Australian and Victorian long-term peak electricity demand forecasts.

Research issues

- Extend method for short-term demand forecasting.

Results

- We have successfully forecast the extreme upper tail in ten years time using only twelve years of data!
- This method has now been adopted for the official South Australian and Victorian long-term peak electricity demand forecasts.

Research issues

- Extend method for short-term demand forecasting.
- Generalize model to allow coincident peaks across different interconnected markets.

Results

- We have successfully forecast the extreme upper tail in ten years time using only twelve years of data!
- This method has now been adopted for the official South Australian and Victorian long-term peak electricity demand forecasts.

Research issues

- Extend method for short-term demand forecasting.
- Generalize model to allow coincident peaks across different interconnected markets.
- How to measure forecast accuracy for a 1-in-10 year event?

Outline

Research and reality

- A lot of published research deals with minor variations on existing problems.

Research and reality

- A lot of published research deals with minor variations on existing problems.
- There are lots of tough research problems that few people are addressing.

Research and reality

- A lot of published research deals with minor variations on existing problems.
- There are lots of tough research problems that few people are addressing.
- If only academics and practitioners talked more . . .

Complexity

- Real data is complex, and we need methods to handle this.

Complexity

- Real data is complex, and we need methods to handle this.
- Textbook approaches are building blocks, not the final word.

Complexity

- Real data is complex, and we need methods to handle this.
- Textbook approaches are building blocks, not the final word.
- Understand what is driving your data, and try to build it into a model.

Complexity

- Real data is complex, and we need methods to handle this.
- Textbook approaches are building blocks, not the final word.
- Understand what is driving your data, and try to build it into a model.
- Complex methods work well when there is a lot of information the data.

Complexity

- Real data is complex, and we need methods to handle this.
- Textbook approaches are building blocks, not the final word.
- Understand what is driving your data, and try to build it into a model.
- Complex methods work well when there is a lot of information the data.
- Simple methods work better when there is mostly noise in the data.

Complexity

- Real data is complex, and we need methods to handle this.
- Textbook approaches are building blocks, not the final word.
- Understand what is driving your data, and try to build it into a model.
- Complex methods work well when there is a lot of information the data.
- Simple methods work better when there is mostly noise in the data.
- Choose methods to model all the information available, but no more.

Stationarity and structural change

Herakleitos (c.500BC)

All is flux, nothing is stationary.

Stationarity and structural change

Herakleitos (c.500BC)

All is flux, nothing is stationary.

- Change happens, but the rate at which things change is often constant.

Stationarity and structural change

Herakleitos (c.500BC)

All is flux, nothing is stationary.

- Change happens, but the rate at which things change is often constant.
- Don't throw away data just because things have changed.

Stationarity and structural change

Herakleitos (c.500BC)

All is flux, nothing is stationary.

- Change happens, but the rate at which things change is often constant.
- Don't throw away data just because things have changed.
- Methods need to be adaptive but history is important for measuring the rate of change.

Stationarity and structural change

Herakleitos (c.500BC)

All is flux, nothing is stationary.

- Change happens, but the rate at which things change is often constant.
- Don’t throw away data just because things have changed.
- Methods need to be adaptive but history is important for measuring the rate of change.
- Uncertainty is not always easy to measure, is not always easy to compute, but is often the thing of most interest.

Go forth and forecast

A good forecaster is not smarter than everyone else, he merely has his ignorance better organised.

(Anonymous)

Go forth and forecast

A good forecaster is not smarter than everyone else, he merely has his ignorance better organised.

(Anonymous)

Slides available from

www.robjhyndman.com