Forecasting medium- and long-term peak electricity demand

Rob J Hyndman

Business & Economic Forecasting Unit

MONASH University

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 ten 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 ten 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

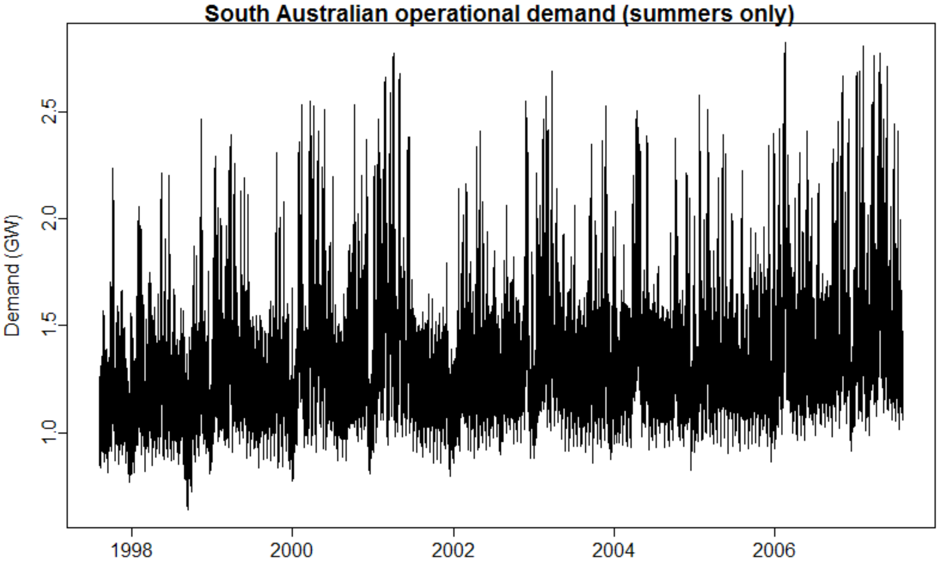
- We want to forecast the peak electricity demand in a half-hour period in ten years time.
- We have ten years of half-hourly electricity data, temperature data and some economic and demographic data.
- The location is South Australia: home to the most volatile electricity demand in the world.

The problem

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

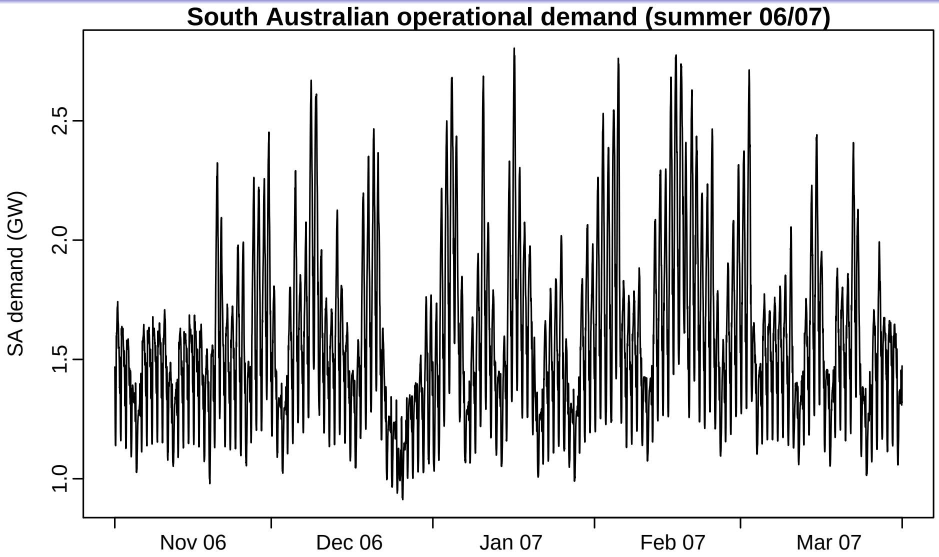
Sounds impossible?

Demand data



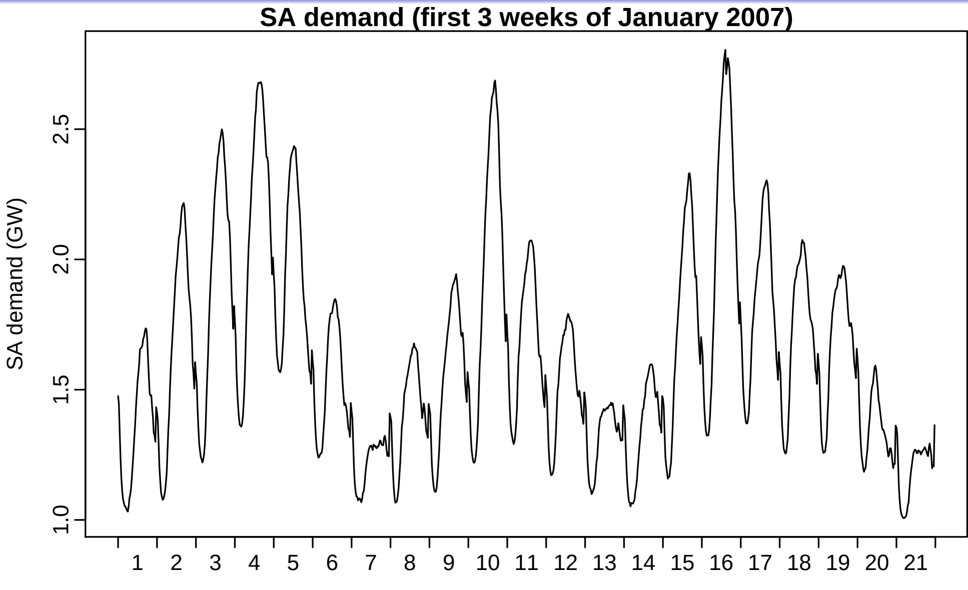
Half-hourly demand data for South Australia from 1 July 1997 to 31 March 2007. Only data from November–March are shown.

Demand data



Half-hourly demand data for South Australia from 1 November 2006 to 31 March 2007.

Demand data



Half-hourly demand data for South Australia from 1–21 January 2007.

Demand boxplots

Temperature data

Demand drivers

- calendar effects

Demand drivers

- calendar effects
- prevailing weather conditions (and the timing of those conditionals

Demand drivers

- calendar effects
- prevailing weather conditions (and the timing of those conditionals
- climate changes

Demand drivers

- calendar effects
- prevailing weather conditions (and the timing of those conditionals
- climate changes
- economic and demographic changes

Demand drivers

- calendar effects
- prevailing weather conditions (and the timing of those conditionals
- climate changes
- economic and demographic changes
- changing technology

Demand drivers

- calendar effects
- prevailing weather conditions (and the timing of those conditionals
- climate changes
- economic and demographic changes
- changing technology

Demand drivers

- calendar effects
- prevailing weather conditions (and the timing of those conditionals
- climate changes
- economic and demographic changes
- changing technology

Modelling framework

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

Demand drivers

- calendar effects
- prevailing weather conditions (and the timing of those conditionals
- 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 weather conditions (and the timing of those conditionals
- 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 for 2005/06 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 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 demand at time t (measured in half-hourly intervals) during period p, $p = 1, \dots, 48$;
- $h_p(t)$ 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 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 Kent Town and $\mathbf{w}_{2,t}$ is a vector of recent temperatures at Adelaide airport;

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 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 Kent Town and $\mathbf{w}_{2,t}$ is a vector of recent temperatures at Adelaide airport;
- $z_{j,t}$ 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 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 Kent Town and $\mathbf{w}_{2,t}$ is a vector of recent temperatures at Adelaide airport;
- $z_{j,t}$ 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:

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

- $\ell_p(t)$ 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:

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

- $\ell_p(t)$ 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:

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

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

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:

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

- $\ell_p(t)$ is "time of summer" effect (a regression spline);
- $\alpha_{t,p}$ is day of week effect;
- $\beta_{t,p}$ is "holiday" effect;
- $\gamma_{t,p}$ New Year's Eve effect;

Equations

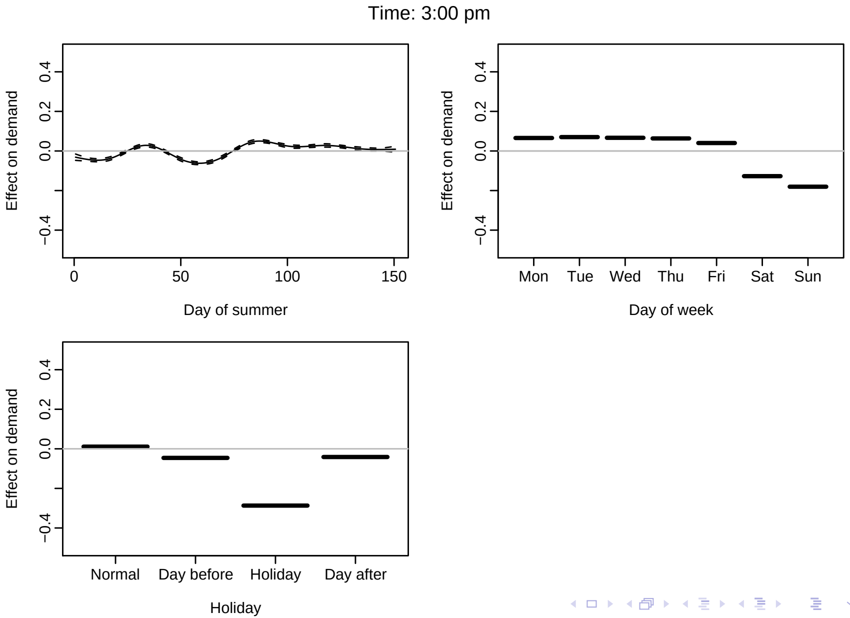
$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

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

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

- $\ell_p(t)$ is “time of summer” effect (a regression spline);
- $\alpha_{t,p}$ is day of week effect;
- $\beta_{t,p}$ is “holiday” effect;
- $\gamma_{t,p}$ New Year’s Eve effect;
- $\delta_{t,p}$ is millennium effect;

Fitted results (3pm)



Equations

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

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

- x_t is ave temp across sites at time t;

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;

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

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

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

- 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.

Equations

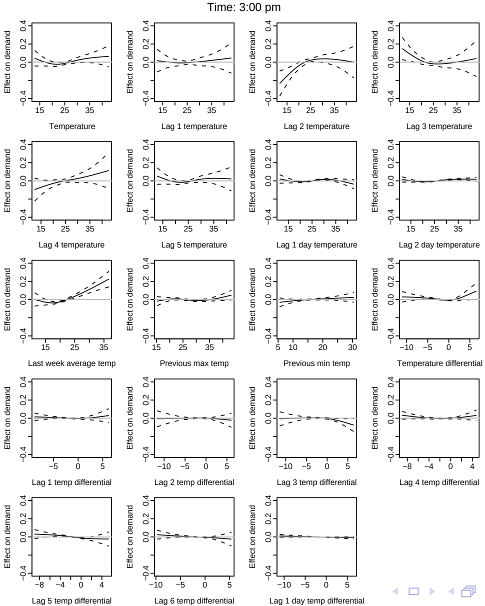
$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

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

- 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;
- 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.

Fitted results (3pm)



Equations

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

- 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.

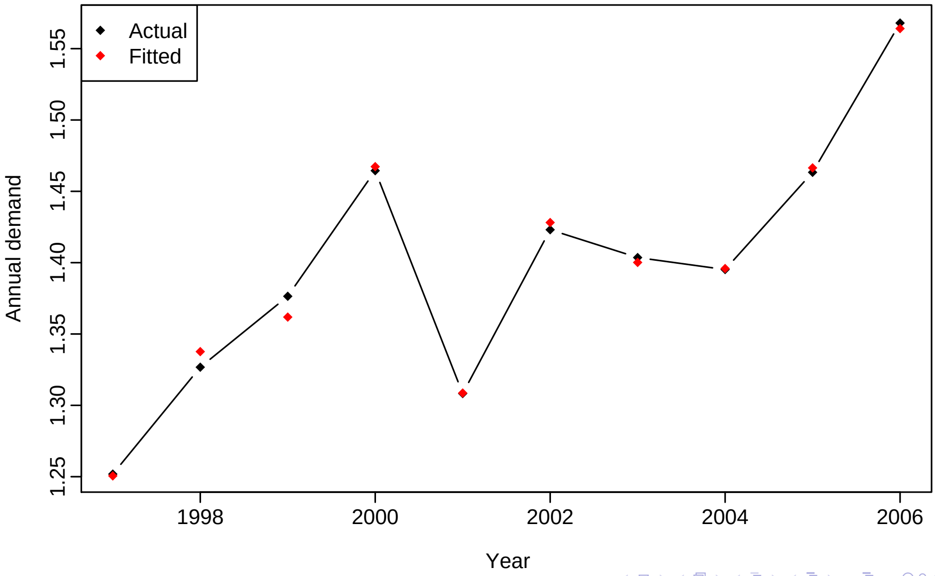
Equations

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

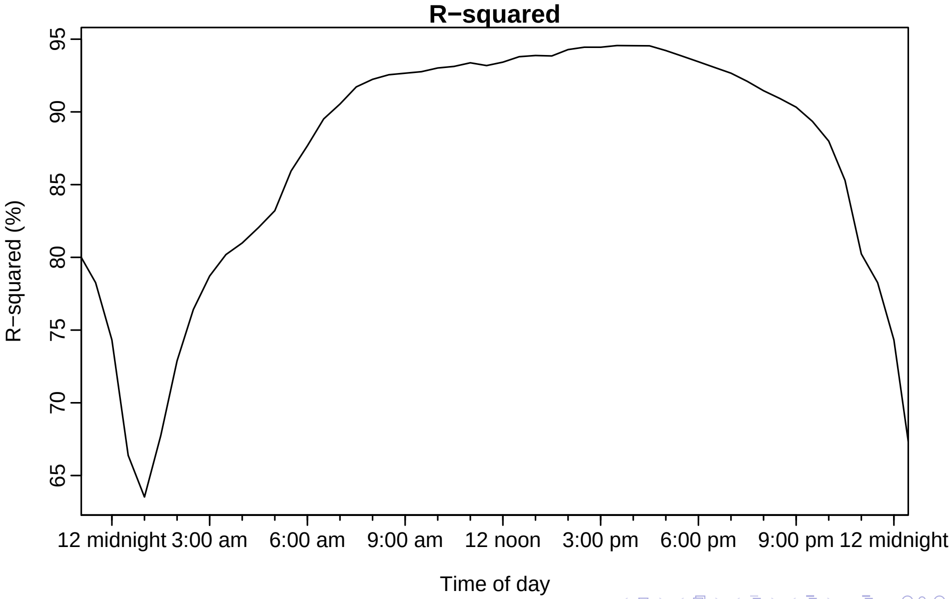
- Other variables described by linear relationships with coefficients $c_1, \dots, c_J$.
- Estimation based on annual data.

Variable	Coefficient	Std. Error	t value	P value
Intercept	-0.13981	0.04338	-3.222	0.018094
Gross State Product	0.01684	0.00108	15.649	0.000004
Lag Price	-0.04957	0.00727	-6.818	0.000488
Cooling Degree Days	0.36300	0.01716	21.157	0.000001

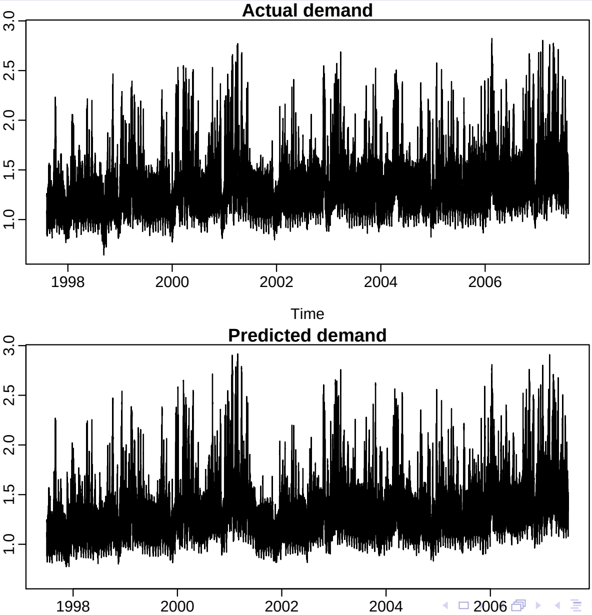
Predictions



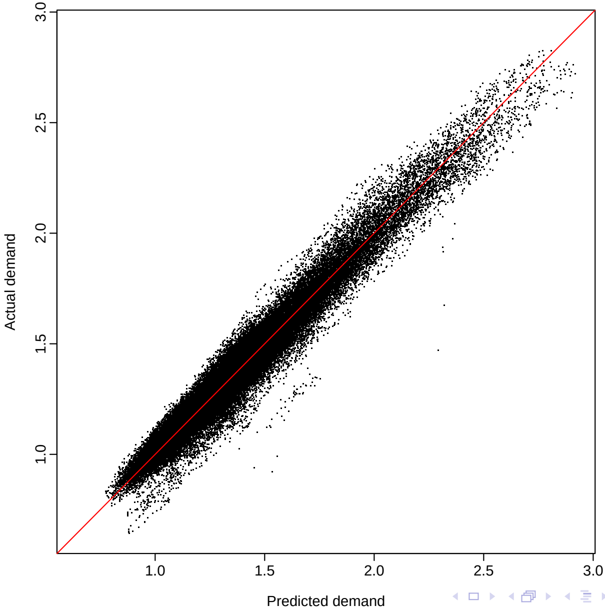
Predictions



Predictions



Predictions



Peak demand forecasting

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

Multiple alternative futures created by

- resampling residuals using a seasonal bootstrap;

Peak demand forecasting

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

Multiple alternative futures created by

- resampling residuals using a seasonal bootstrap;
- generating simulations of future temperature patterns based on seasonally bootstrapping past temperatures (with some adjustment for extremes and climate change);

Peak demand forecasting

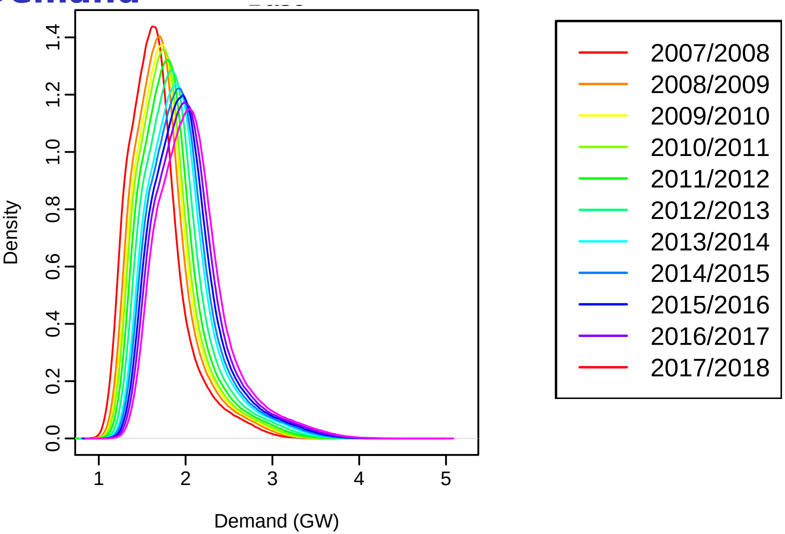
$$\log(y_{t,p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + \sum_{j=1}^{J} c_j z_{j,t} + n_t$$

Multiple alternative futures created by

- resampling residuals using a seasonal bootstrap;
- generating simulations of future temperature patterns based on seasonally bootstrapping past temperatures (with some adjustment for extremes and climate change);
- using assumed values for GSP and Price.

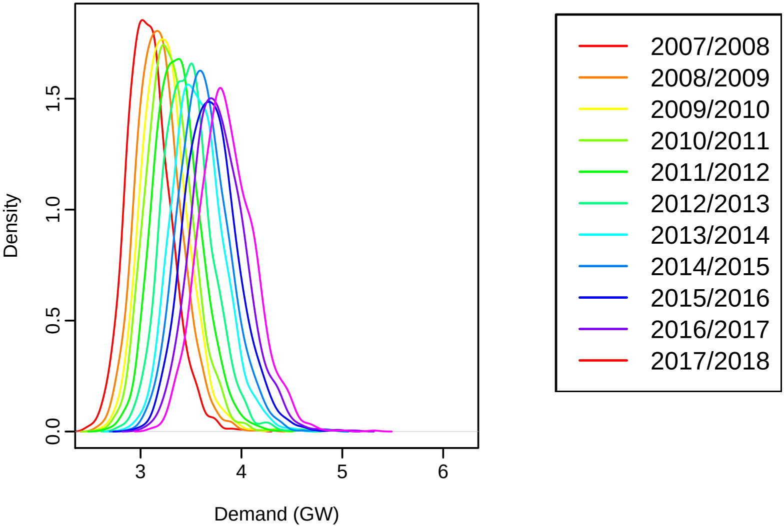
Peak demand distribution

Demand

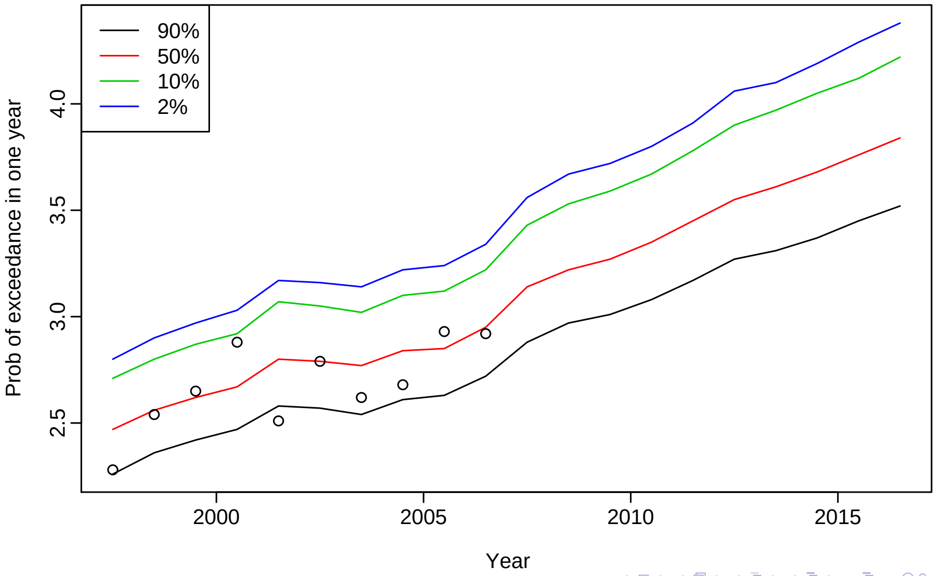


Peak demand distribution

Annual maximum demand



Peak demand distribution



Conclusions

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

Conclusions

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

Conclusions

- We have forecast the extreme upper tail in ten years time using only ten years of data!
- This method has now been adopted for the official South Australian peak electricity demand forecasts.
- We hope to extend method to Victoria in next year.

Conclusions

- We have forecast the extreme upper tail in ten years time using only ten years of data!
- This method has now been adopted for the official South Australian peak electricity demand forecasts.
- We hope to extend method to Victoria in next year.
- Method could also be used for short-term demand forecasting, if we add a model for correlated residuals.

Conclusions

- We have forecast the extreme upper tail in ten years time using only ten years of data!
- This method has now been adopted for the official South Australian peak electricity demand forecasts.
- We hope to extend method to Victoria in next year.
- Method could also be used for short-term demand forecasting, if we add a model for correlated residuals.
- Provides way to analyse probability of coincident peaks across different interconnected markets.

Conclusions

- We have forecast the extreme upper tail in ten years time using only ten years of data!
- This method has now been adopted for the official South Australian peak electricity demand forecasts.
- We hope to extend method to Victoria in next year.
- Method could also be used for short-term demand forecasting, if we add a model for correlated residuals.
- Provides way to analyse probability of coincident peaks across different interconnected markets.
- Could be extended to whole year, providing probabilistic forecasts of total energy requirements.

Conclusions

- We have forecast the extreme upper tail in ten years time using only ten years of data!
- This method has now been adopted for the official South Australian peak electricity demand forecasts.
- We hope to extend method to Victoria in next year.
- Method could also be used for short-term demand forecasting, if we add a model for correlated residuals.
- Provides way to analyse probability of coincident peaks across different interconnected markets.
- Could be extended to whole year, providing probabilistic forecasts of total energy requirements.
- An R package and a paper will (eventually) appear at **www.robhyndman.info**