# Forecasting and the importance of being uncertain

**Rob J Hyndman**

MONASH University

# 2006 Belz Lecture

**Maurice H Belz (1897–1975)**

Professor of Statistics, University of Melbourne, 1955–1963

# Outline

# Outline

1. **Dangers and difficulties of forecasting**
2. A brief history of forecasting
3. Forecasting the PBS
4. Forecasting Australia's population
5. Conclusions

# Forecasters are to blame!

## News report on 16 August 2006

A Russian woman is suing weather forecasters for wrecking her holiday. A court in Uljanovsk heard that Alyona Gabitova had been promised 28 degrees and sunshine when she planned a camping trip to a local nature reserve, newspaper *Nowyje Iswestija* said.

But it did nothing but pour with rain the whole time, leaving her with a cold. Gabitova has asked the court to order the weather service to pay the cost of her travel.

# Reputations can be made & lost

- "Tell us what the future holds, so we may know that you are gods." (Isaiah 41:23, 700 B.C.)

# Reputations can be made & lost

- "Tell us what the future holds, so we may know that you are gods." (Isaiah 41:23, 700 B.C.)
- "I think there is a world market for maybe five computers." (Chairman of IBM, 1943)

# Reputations can be made & lost

- “Tell us what the future holds, so we may know that you are gods.” (Isaiah 41:23, 700 B.C.)
- “I think there is a world market for maybe five computers.” (Chairman of IBM, 1943)
- “Computers in the future may weigh no more than 1.5 tons.” (*Popular Mechanics*, 1949)

# Reputations can be made & lost

- “Tell us what the future holds, so we may know that you are gods.” (Isaiah 41:23, 700 B.C.)
- “I think there is a world market for maybe five computers.” (Chairman of IBM, 1943)
- “Computers in the future may weigh no more than 1.5 tons.” (*Popular Mechanics*, 1949)
- “There is no reason anyone would want a computer in their home.” (President, DEC, 1977)

# Reputations can be made & lost

- “Tell us what the future holds, so we may know that you are gods.” (Isaiah 41:23, 700 B.C.)
- “I think there is a world market for maybe five computers.” (Chairman of IBM, 1943)
- “Computers in the future may weigh no more than 1.5 tons.” (*Popular Mechanics*, 1949)
- “There is no reason anyone would want a computer in their home.” (President, DEC, 1977)
- “There are four ways economists can lose their reputation. Gambling is the quickest, sex is the most pleasurable and drink the slowest. But forecasting is the surest.” (Max Walsh, *The Age*, 1993)

# Those "unforeseen events"

Precautions should be taken against running into unforeseen occurrences or events. (Horoscope, *New York Times*)

# Those "unforeseen events"

Precautions should be taken against running into unforeseen occurrences or events. (Horoscope, *New York Times*)

We are ready for any unforeseen event which may or may not occur.

(Dan Quayle)

# Outline

# What is it?

# What is it?

**Clay model of sheep's liver**

Used by Babylonian forecasters approximately 600 B.C.

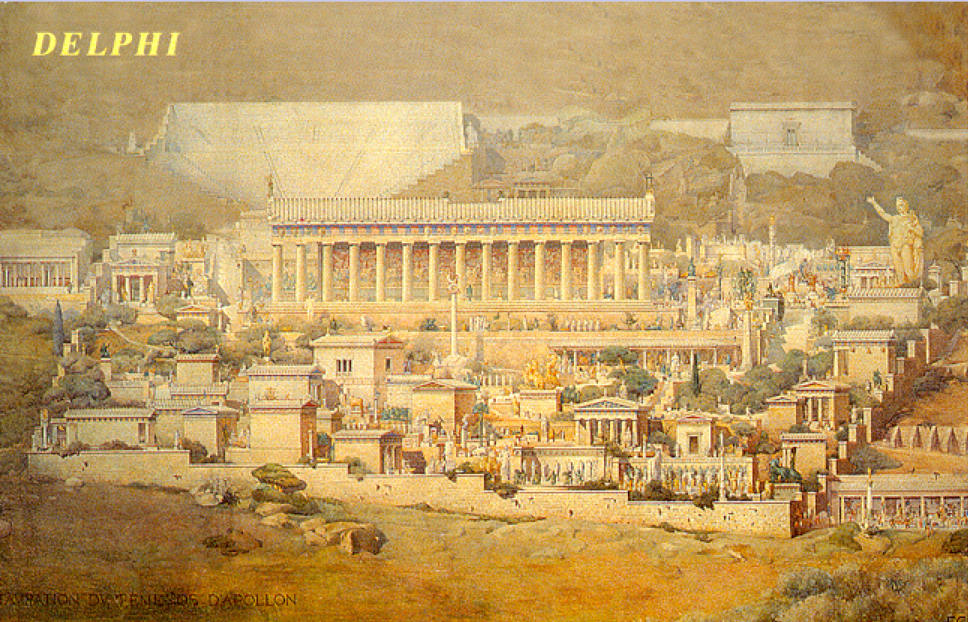
Now in British Museum.

# Delphic oracle

# Delphic oracle

# Delphic oracle

# Temple of Apollo



# Temple of Apollo

# Temple of Apollo

# Vagrant forecasters

The British Vagrancy Act (1736) made it an offence to defraud by charging money for predictions.

# Vagrant forecasters

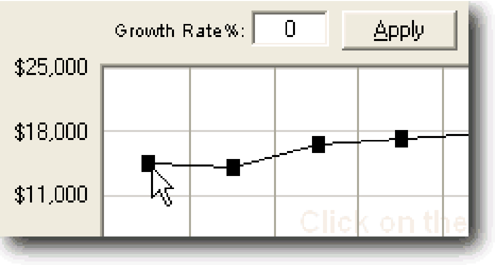
The British Vagrancy Act (1736) made it an offence to defraud by charging money for predictions.

**Punishment:** a fine or three months' imprisonment with hard labour.

# Standard business practice today

## Graphic Forecaster

Create forecasts visually with a "drag and drop" graphic forecaster. The Graphic Forecaster is a simple and powerful tool to streamline the forecasting process. You can change your sales and expenses estimates by simply clicking your mouse button to move the line on your forecast chart or apply a specific growth rate to the whole year. Build forecasts using visual common sense.




View More Detail

# Standard business practice today

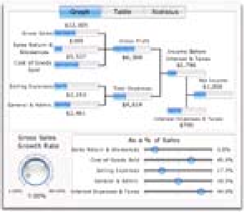
## Crystal Xcelsius Showcase

Examples of what you can build with Crystal Xcelsius.

Free Trial Buy Now

If you cannot open these demos, download the latest version of Macromedia's Flash Player.

### Featured Example: Profitability Analysis



**Profitability Analysis**

This profitability model allows you to create "what-if" scenarios by modifying sales growth rate and all other relevant accounts measured as a percentage of total sales. This example, built with fictitious data, depicts the most relevant accounts of a profit and loss statement, and shows the impact of changes on net income. The results change immediately, allowing you to create endless what-if scenarios.

Download as PowerPoint
Download as Flash
Download as PDF
Download Source Files
Download as Word

# Standard business practice today

## Budget Maestro by Centage

**Click here for a free demo** FREE!

**Application:** Business Intelligence and Analytics

**Price Range:** Solutions start at $5K

---

A huge advance over spreadsheet-based systems, Budget Maestro is a complete solution for budgeting, forecasting, what-if scenario planning, reporting and analysis. Budget Maestro takes the pain out of the budgeting process (no tedious data entry and formula verification) while providing you a tool to more accurately analyze and measure business performance and profitability. Budget Maestro's capabilities include:

**Budgeting and Forecasting:** Budget Maestro utilizes database technology for real-time data collection and reporting. A common interface for all users fosters collaboration and increases the accuracy of data entry. There are no formulas or macros to create, no tedious re-keying of data and no mystery links to chase down and fix. Budget Maestro's built-in "financial intelligence and business rules" builds the formulas for you ensuring 100% accuracy.

# Standard business practice today

- “What-if scenarios” based on assumed and fixed future conditions.

# Standard business practice today

- "What-if scenarios" based on assumed and fixed future conditions.
- Highly subjective.

# Standard business practice today

- “What-if scenarios” based on assumed and fixed future conditions.
- Highly subjective.
- Not replicable or testable.

# Standard business practice today

- “What-if scenarios” based on assumed and fixed future conditions.
- Highly subjective.
- Not replicable or testable.
- No possible way of quantifying probabilistic uncertainty.

# Standard business practice today

- "What-if scenarios" based on assumed and fixed future conditions.
- Highly subjective.
- Not replicable or testable.
- No possible way of quantifying probabilistic uncertainty.
- Lack of uncertainty statements leads to false sense of accuracy.

# Standard business practice today

- "What-if scenarios" based on assumed and fixed future conditions.
- Highly subjective.
- Not replicable or testable.
- No possible way of quantifying probabilistic uncertainty.
- Lack of uncertainty statements leads to false sense of accuracy.
- Largely guesswork.

# Standard business practice today

- “What-if scenarios” based on assumed and fixed future conditions.
- Highly subjective.
- Not replicable or testable.
- No possible way of quantifying probabilistic uncertainty.
- Lack of uncertainty statements leads to false sense of accuracy.
- Largely guesswork.

# Standard business practice today

- “What-if scenarios” based on assumed and fixed future conditions.
- Highly subjective.
- Not replicable or testable.
- No possible way of quantifying probabilistic uncertainty.
- Lack of uncertainty statements leads to false sense of accuracy.
- Largely guesswork.

**Is this any better than a sheep’s liver or hallucinogens?**

# Data mining prediction

"Data mining" prediction methods include

- Classification and regression trees

# Data mining prediction

## "Data mining" prediction methods include

- Classification and regression trees
- Neural networks

# Data mining prediction

## “Data mining” prediction methods include

- Classification and regression trees
- Neural networks
- Nearest neighbour and naïve Bayes.

# Data mining prediction

**“Data mining” prediction methods include**

- Classification and regression trees
- Neural networks
- Nearest neighbour and naïve Bayes.

## Good points

- Smart algorithms, few assumptions and applied to huge data sets.

# Data mining prediction

**"Data mining" prediction methods include**

- Classification and regression trees
- Neural networks
- Nearest neighbour and naïve Bayes.

## Good points

- Smart algorithms, few assumptions and applied to huge data sets.
- Solve problems which traditional statistical methods can't handle (largely due to size of data sets).

# Data mining prediction

**“Data mining” prediction methods include**

- Classification and regression trees
- Neural networks
- Nearest neighbour and naïve Bayes.

## Good points

- Smart algorithms, few assumptions and applied to huge data sets.
- Solve problems which traditional statistical methods can’t handle (largely due to size of data sets).
- Strong emphasis on out-of-sample predictive performance (the test data).

# Data mining prediction

**"Data mining" prediction methods include**

- Classification and regression trees
- Neural networks
- Nearest neighbour and naïve Bayes.

## Good points

- Smart algorithms, few assumptions and applied to huge data sets.
- Solve problems which traditional statistical methods can't handle (largely due to size of data sets).
- Strong emphasis on out-of-sample predictive performance (the test data).
- Recognition that many problems are about prediction not p-values.

# Data mining prediction

## “Data mining” prediction methods include

- Classification and regression trees
- Neural networks
- Nearest neighbour and naïve Bayes.

**Problems**

# Data mining prediction

**“Data mining” prediction methods include**

- Classification and regression trees
- Neural networks
- Nearest neighbour and naïve Bayes.

## Problems

- Rare to provide uncertainty statements about individual predictions (leading to false sense of accuracy).

# Data mining prediction

## "Data mining" prediction methods include

- Classification and regression trees
- Neural networks
- Nearest neighbour and naïve Bayes.

## Problems

- Rare to provide uncertainty statements about individual predictions (leading to false sense of accuracy).
- Limited interpretability of many models.

# Data mining prediction

## "Data mining" prediction methods include

- Classification and regression trees
- Neural networks
- Nearest neighbour and naïve Bayes.

### Problems

- Rare to provide uncertainty statements about individual predictions (leading to false sense of accuracy).
- Limited interpretability of many models.
- Software can be up to 50 times the cost of comparable statistical software.

# Data mining prediction

**“Data mining” prediction methods include**

- Classification and regression trees
- Neural networks
- Nearest neighbour and naïve Bayes.

## Problems

- Rare to provide uncertainty statements about individual predictions (leading to false sense of accuracy).
- Limited interpretability of many models.
- Software can be up to 50 times the cost of comparable statistical software.
- Ignorance of basic statistical principles and methods.

# The rise of stochastic models

**1959** exponential smoothing (Brown)

**1970** ARIMA models (Box, Jenkins)

**1980** VAR models (Sims, Granger)

**1980** non-linear models (Granger, Tong, Hamilton, Teräsvirta, . . . )

**1982** ARCH/GARCH (Engle, Bollerslev)

**1986** neural networks (Rumelhart)

**1989** state space models (Harvey, West, Harrison)

**1994** nonparametric forecasting (Tjøstheim, Härdle, Tsay,. . . )

# Advantages of stochastic models

- Based on empirical data

# Advantages of stochastic models

- Based on empirical data
- Computable

# Advantages of stochastic models

- Based on empirical data
- Computable
- Replicable

# Advantages of stochastic models

- Based on empirical data
- Computable
- Replicable
- Testable

# Advantages of stochastic models

- Based on empirical data
- Computable
- Replicable
- Testable
- Objective measure of uncertainty

# Advantages of stochastic models

- Based on empirical data
- Computable
- Replicable
- Testable
- Objective measure of uncertainty
- Able to compute prediction intervals

# The skeptics

*It is utterly implausible that a mathematical formula should make the future known to us, and those who think it can would once have believed in witchcraft.*

(Bertrand de Jouvenel
The Art of Conjecture, 1967.)

# Outline

1. Dangers and difficulties of forecasting
2. A brief history of forecasting
3. **Forecasting the PBS**
4. Forecasting Australia's population
5. Conclusions

# 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
- US Attack
- ACT Election
- Middle East

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

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
Windows Media | Real Audio

ABC News Online Bulletin
VIDEO ON DEMAND

# Forecasting the PBS

**Estimation of forward estimates for the Pharmaceutical Benefit Scheme**

Department of Health and Aging

- \$5 billion budget. Underforecasted by \$500–\$800 million in 2000 and 2001.

# Forecasting the PBS

**Estimation of forward estimates for the Pharmaceutical Benefit Scheme**

Department of Health and Aging

- $5 billion budget. Underforecasted by $500–$800 million in 2000 and 2001.
- Thousands of products. Seasonal demand.

# Forecasting the PBS

**Estimation of forward estimates for the Pharmaceutical Benefit Scheme**

Department of Health and Aging

- $5 billion budget. Underforecasted by $500–$800 million in 2000 and 2001.
- Thousands of products. Seasonal demand.
- Subject to covert marketing, volatile products, uncontrollable expenditure.

# Forecasting the PBS

## Estimation of forward estimates for the Pharmaceutical Benefit Scheme

Department of Health and Aging

- $5 billion budget. Underforecasted by $500–$800 million in 2000 and 2001.
- Thousands of products. Seasonal demand.
- Subject to covert marketing, volatile products, uncontrollable expenditure.
- All forecasts being done with the `FORECAST` function in MS-Excel applied to 10 year old data!

# Forecasting the PBS

**Estimation of forward estimates for the Pharmaceutical Benefit Scheme**

Department of Health and Aging

- We used **time series models** — automated exponential smoothing state space modelling applied to about 100 product groups.

# Forecasting the PBS

## Estimation of forward estimates for the Pharmaceutical Benefit Scheme

Department of Health and Aging

- We used **time series models** — automated exponential smoothing state space modelling applied to about 100 product groups.
- Methodological tools developed in 2002 and published in the *International Journal of Forecasting*

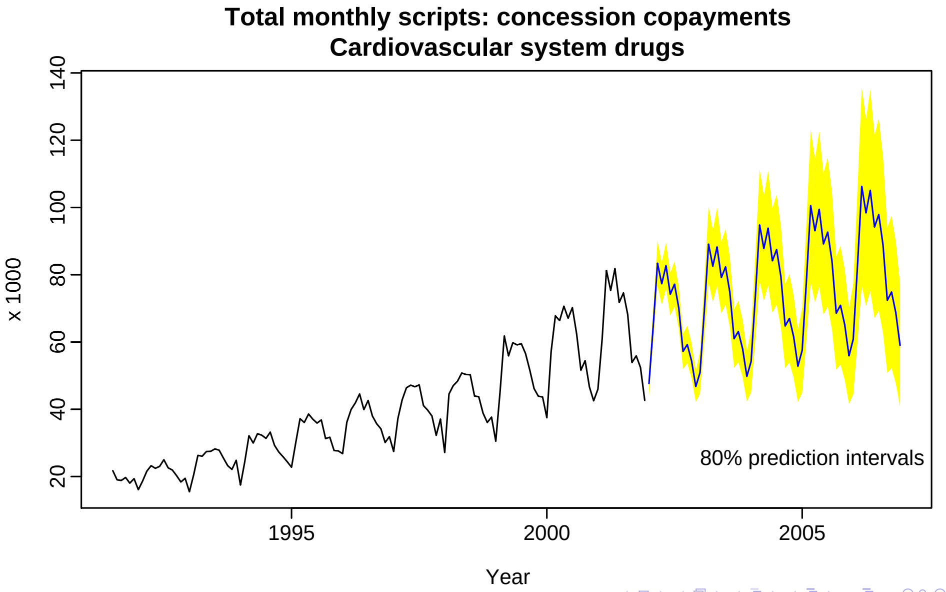
# Forecasting the PBS

**Estimation of forward estimates for the Pharmaceutical Benefit Scheme**

Department of Health and Aging

- We used **time series models** — automated exponential smoothing state space modelling applied to about 100 product groups.
- Methodological tools developed in 2002 and published in the *International Journal of Forecasting*
- Forecast error now a few $million per year.

# Forecasting the PBS



# Forecasting the PBS

- Used stochastic models to describe evolution of sales over time.

# Forecasting the PBS

- Used stochastic models to describe evolution of sales over time.
- Models allowed for time-changing trend and seasonal patterns.

# Forecasting the PBS

- Used stochastic models to describe evolution of sales over time.
- Models allowed for time-changing trend and seasonal patterns.
- Stochastic models provide prediction intervals which give a sense of uncertainty.

# Forecasting the PBS

- Used stochastic models to describe evolution of sales over time.
- Models allowed for time-changing trend and seasonal patterns.
- Stochastic models provide prediction intervals which give a sense of uncertainty.
- Class of models was based on exponential smoothing.

# Forecasting the PBS

- Used stochastic models to describe evolution of sales over time.
- Models allowed for time-changing trend and seasonal patterns.
- Stochastic models provide prediction intervals which give a sense of uncertainty.
- Class of models was based on exponential smoothing.
- At the time, exponential smoothing methods were not thought to be based on stochastic models.

# 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

## 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 the admissible parameter space.

# Exponential smoothing state space models

All exponential smoothing methods can be written in the following form:

$$
\begin{aligned}
Y_t &= h(\boldsymbol{x}_{t-1}) + k(\boldsymbol{x}_{t-1})\varepsilon_t \\
\boldsymbol{x}_t &= f(\boldsymbol{x}_{t-1}) + g(\boldsymbol{x}_{t-1})\varepsilon_t
\end{aligned}
$$

where $\{\varepsilon_t\}$ is a Gaussian white noise process with mean zero and variance $\sigma^2$, and $\boldsymbol{x}_t$ is an unobserved state.

# Exponential smoothing state space models

All exponential smoothing methods can be written in the following form:

$$
\begin{aligned}
Y_t &= h(\boldsymbol{x}_{t-1}) + k(\boldsymbol{x}_{t-1})\varepsilon_t \\
\boldsymbol{x}_t &= f(\boldsymbol{x}_{t-1}) + g(\boldsymbol{x}_{t-1})\varepsilon_t
\end{aligned}
$$

where $\{\varepsilon_t\}$ is a Gaussian white noise process with mean zero and variance $\sigma^2$, and $\boldsymbol{x}_t$ is an unobserved state.

Forecasts are given by $\mathrm{E}(Y_{n+h} \mid \boldsymbol{x}_n)$.

# Exponential smoothing state space models

## Holt-Winters' additive seasonal method

$$
\begin{aligned}
\hat{Y}_{t+h|t} &= \ell_t + hb_t + s_{t+h-m}. \\
\ell_t &= \alpha(Y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\
s_t &= \gamma(Y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}
\end{aligned}
$$

where $0 \le \alpha \le 1$, $0 \le \beta \le 1$, $0 \le \gamma \le 1$ and $m$ is the period of seasonality.

# Exponential smoothing state space models

## Additive error state space model

Let $\boldsymbol{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})'$. Then

$$\begin{aligned} Y_t &= h(\boldsymbol{x}_{t-1}) + \varepsilon_t \\ \boldsymbol{x}_t &= f(\boldsymbol{x}_{t-1}) + g(\boldsymbol{x}_{t-1})\varepsilon_t \end{aligned}$$

where $h(\boldsymbol{x}_{t-1}) = \ell_{t-1} + b_{t-1} + s_{t-m}$

$$f(\boldsymbol{x}_{t-1}) = \begin{bmatrix} \ell_{t-1} + b_{t-1} \\ b_{t-1} \\ s_{t-m} \\ s_{t-1} \\ \vdots \\ s_{t-m+1} \end{bmatrix}, \quad g(\boldsymbol{x}_{t-1}) = \begin{bmatrix} \alpha \\ \alpha\beta \\ \gamma \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

# Exponential smoothing state space models

Then $\hat{Y}_{n+h|n} \equiv \mathrm{E}[Y_{n+h} \mid \boldsymbol{x}_n] = \ell_n + hb_n + s_{n+h-m}.$

- So Holt-Winter forecasts are optimal for this state space model.

# Exponential smoothing state space models

Then $\hat{Y}_{n+h|n} \equiv \mathrm{E}[Y_{n+h} \mid \boldsymbol{x}_n] = \ell_n + hb_n + s_{n+h-m}$.

- So Holt-Winter forecasts are optimal for this state space model.
- The likelihood can be optimized to obtain estimates of smoothing parameters and initial states.

# Exponential smoothing state space models

Then $\hat{Y}_{n+h|n} \equiv \mathrm{E}[Y_{n+h} \mid \boldsymbol{x}_n] = \ell_n + hb_n + s_{n+h-m}$.

- So Holt-Winter forecasts are optimal for this state space model.
- The likelihood can be optimized to obtain estimates of smoothing parameters and initial states.
- The model can be used to obtain prediction intervals.

# Exponential smoothing state space models

| Trend Component | Seasonal Component: **N** (none) | Seasonal Component: **A** (additive) | Seasonal Component: **M** (multiplicative) |
|---|---|---|---|
| **N** (none) | NN | NA | NM |
| **A** (additive) | AN | AA | AM |
| **M** (multiplicative) | MN | MA | MM |

**General notation**

$$\underset{\text{Error}}{E}\ \underset{\text{Trend}}{T}\ \underset{\text{Seasonal}}{S}$$

# Exponential smoothing state space models

| Trend Component | Seasonal Component N (none) | A (additive) | M (multiplicative) |
|---|---|---|---|
| **N** (none) | NN | NA | NM |
| **A** (additive) | AN | AA | AM |
| **M** (multiplicative) | MN | MA | MM |

**General notation** E T S

Error Trend Seasonal

ETS(A,N,N): Simple exponential smoothing with additive errors

# Exponential smoothing state space models

| **Trend Component** | **Seasonal Component**<br>**N** (none) | <br>**A** (additive) | <br>**M** (multiplicative) |
|---|---|---|---|
| **N** (none) | NN | NA | NM |
| **A** (additive) | AN | AA | AM |
| **M** (multiplicative) | MN | MA | MM |

**General notation** E T S

Error Trend Seasonal

ETS(A,N,N): Simple exponential smoothing with additive errors

ETS(A,A,N): Holt's linear method with additive errors

# Exponential smoothing state space models

| Trend Component | Seasonal Component: N (none) | Seasonal Component: A (additive) | Seasonal Component: M (multiplicative) |
|---|---|---|---|
| **N** (none) | NN | NA | NM |
| **A** (additive) | AN | AA | AM |
| **M** (multiplicative) | MN | MA | MM |

**General notation** E T S

Error Trend Seasonal

ETS(A,N,N): Simple exponential smoothing with additive errors

ETS(A,A,N): Holt's linear method with additive errors

ETS(A,A,A): Additive Holt-Winters' method with additive errors

# Exponential smoothing state space models

| Trend Component | Seasonal Component N (none) | Seasonal Component A (additive) | Seasonal Component M (multiplicative) |
|---|---|---|---|
| **N** (none) | NN | NA | NM |
| **A** (additive) | AN | AA | AM |
| **M** (multiplicative) | MN | MA | MM |

## General notation

E T S

Error Trend Seasonal

ETS(A,N,N): Simple exponential smoothing with additive errors
ETS(A,A,N): Holt's linear method with additive errors
ETS(A,A,A): Additive Holt-Winters' method with additive errors
ETS(M,A,M): Multiplicative Holt-Winters' method with multiplicative errors

# Exponential smoothing state space models

- enables easy calculation of the likelihood

# Exponential smoothing state space models

- enables easy calculation of the likelihood
- provides facilities to compute (analytical) prediction intervals

# Exponential smoothing state space models

- enables easy calculation of the likelihood
- provides facilities to compute (analytical) prediction intervals
- allows the state equations to be expressed in a form which coincides with the error-correction form of the usual smoothing equations.

# Exponential smoothing state space models

- enables easy calculation of the likelihood
- provides facilities to compute (analytical) prediction intervals
- allows the state equations to be expressed in a form which coincides with the error-correction form of the usual smoothing equations.
- two possible state space models for each method (additive error and multiplicative error). Equivalent point forecasts, different prediction intervals.

# Automatic forecasting

For each method, we derive an equivalent state space formulation

- Apply each of the methods that are appropriate to the data. Optimize parameters using MLE (or some other criteria).

# Automatic forecasting

For each method, we derive an equivalent state space formulation

- Apply each of the methods that are appropriate to the data. Optimize parameters using MLE (or some other criteria).
- Select best method using AIC.

# Automatic forecasting

For each method, we derive an equivalent state space formulation

- Apply each of the methods that are appropriate to the data. Optimize parameters using MLE (or some other criteria).
- Select best method using AIC.
- Produce forecasts using best method.

# Automatic forecasting

For each method, we derive an equivalent state space formulation

- Apply each of the methods that are appropriate to the data. Optimize parameters using MLE (or some other criteria).
- Select best method using AIC.
- Produce forecasts using best method.

# Automatic forecasting

For each method, we derive an equivalent state space formulation

- Apply each of the methods that are appropriate to the data. Optimize parameters using MLE (or some other criteria).
- Select best method using AIC.
- Produce forecasts using best method.

This methodology is available as

- ➥ An R package (**forecast**)
- ➥ An Excel add-in (**PhiCast**)

# Exponential smoothing

## Key References

- Hyndman, R.J., Koehler, A.B., Snyder, R.D., & Grose, S. (2002) A state space framework for automatic forecasting using exponential smoothing methods. *International J. Forecasting*, **18**(3), 439–454.

# Exponential smoothing

## Key References


- Hyndman, R.J., Koehler, A.B., Snyder, R.D., & Grose, S. (2002) A state space framework for automatic forecasting using exponential smoothing methods. *International J. Forecasting*, **18**(3), 439–454.
- Hyndman, R.J., Koehler, A.B., Ord, J.K., & Snyder, R.D. (2005) Prediction intervals for exponential smoothing using two new classes of state space models. *Journal of Forecasting*, **24**(1), 17–37.

# Exponential smoothing

## Key References

- Hyndman, R.J., Koehler, A.B., Snyder, R.D., & Grose, S. (2002) A state space framework for automatic forecasting using exponential smoothing methods. *International J. Forecasting*, **18**(3), 439–454.
- Hyndman, R.J., Koehler, A.B., Ord, J.K., & Snyder, R.D. (2005) Prediction intervals for exponential smoothing using two new classes of state space models. *Journal of Forecasting*, **24**(1), 17–37.
- Snyder, R.D., Koehler, A.B., Hyndman, R.J., & Ord, J.K. (2003) Exponential smoothing models: Means and variances for lead-time demand. *European Journal of Operational Research*, **158**(2) 444–455.

# Outline

# ABS population projections

The Australian Bureau of Statistics provide population "projections".

*"The projections are not intended as predictions or forecasts, but are illustrations of growth and change in the population that would occur if assumptions made about future demographic trends were to prevail over the projection period.*
*While the assumptions are formulated on the basis of an assessment of past demographic trends, both in Australia and overseas, there is no certainty that any of the assumptions will be realised. In addition, no assessment has been made of changes in non-demographic conditions."*

ABS 3222.0 - Population Projections, Australia, 2004 to 2101

# Projections vs forecasts

- **Forecast:** best estimate of future outcomes

# Projections vs forecasts

- **Forecast:** best estimate of future outcomes
- **Projection:** calculation of future outcome under fixed assumptions.

# Projections vs forecasts

- **Forecast:** best estimate of future outcomes
- **Projection:** calculation of future outcome under fixed assumptions.

# Projections vs forecasts

- **Forecast:** best estimate of future outcomes
- **Projection:** calculation of future outcome under fixed assumptions.

> "Demographers have in theory insisted that, given the inherent unpredictability of human behaviour, they can make only projections."
>
> Dictionary of Demography, Pressat and Wilson, 1985

# Projections vs forecasts

- **Forecast:** best estimate of future outcomes
- **Projection:** calculation of future outcome under fixed assumptions.

> "Demographers have in theory insisted that, given the inherent unpredictability of human behaviour, they can make only projections."
>
> Dictionary of Demography, Pressat and Wilson, 1985

➥ No-one who uses the ABS projections appreciates the distinction between projections and forecasts.

# ABS population projections

The ABS provides three projection scenarios labelled "High", "Medium" and "Low".

- Based on assumed mortality, fertility and migration rates

# ABS population projections

The ABS provides three projection scenarios labelled "High", "Medium" and "Low".

- Based on assumed mortality, fertility and migration rates
- No objectivity.

# ABS population projections

The ABS provides three projection scenarios labelled "High", "Medium" and "Low".

- Based on assumed mortality, fertility and migration rates
- No objectivity.
- No dynamic changes in rates allowed

# ABS population projections

The ABS provides three projection scenarios labelled "High", "Medium" and "Low".

- Based on assumed mortality, fertility and migration rates
- No objectivity.
- No dynamic changes in rates allowed
- No variation allowed across ages.

# ABS population projections

The ABS provides three projection scenarios labelled "High", "Medium" and "Low".

- Based on assumed mortality, fertility and migration rates
- No objectivity.
- No dynamic changes in rates allowed
- No variation allowed across ages.
- No probabilistic basis.

# ABS population projections

The ABS provides three projection scenarios labelled "High", "Medium" and "Low".

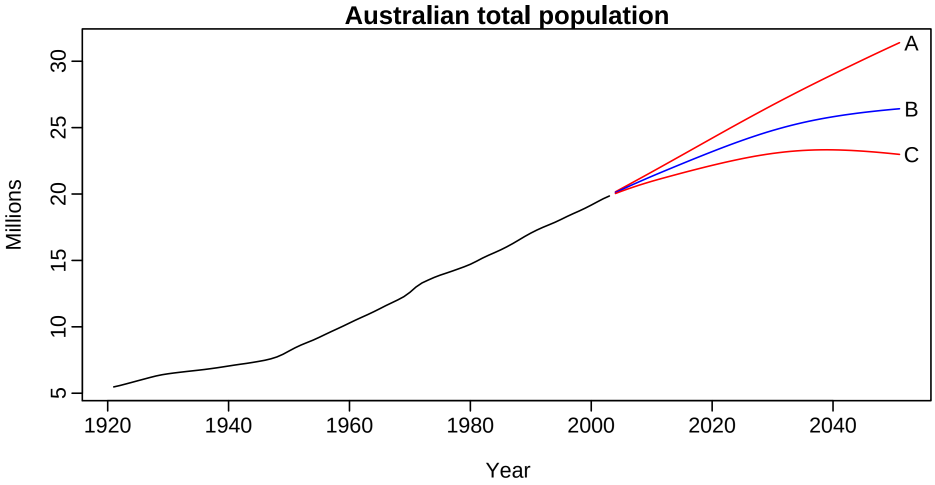
- Based on assumed mortality, fertility and migration rates
- No objectivity.
- No dynamic changes in rates allowed
- No variation allowed across ages.
- No probabilistic basis.
- Not prediction intervals.

# ABS population projections

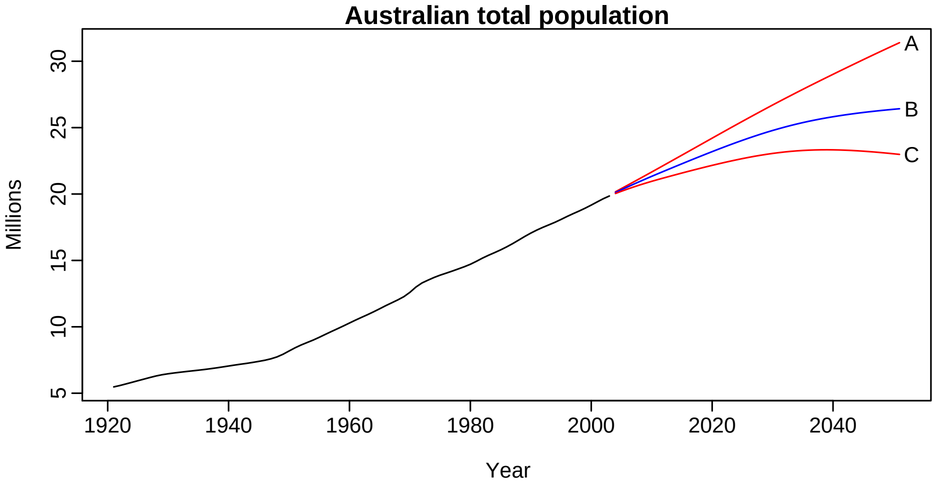
The ABS provides three projection scenarios labelled “High”, “Medium” and “Low”.

- Based on assumed mortality, fertility and migration rates
- No objectivity.
- No dynamic changes in rates allowed
- No variation allowed across ages.
- No probabilistic basis.
- Not prediction intervals.
- Most users use the “Medium” projection, but it is unrelated to the mean, median or mode of the future distribution.

# ABS population projections



# ABS population projections




**What do these projections mean?**

# Stochastic population forecasts

- Forecasts represent the median of the future distribution.

# Stochastic population forecasts

- Forecasts represent the median of the future distribution.
- Percentiles of distribution allow information about uncertainty

# Stochastic population forecasts

- Forecasts represent the median of the future distribution.
- Percentiles of distribution allow information about uncertainty
- Prediction intervals with specified probability coverage for population size and all derived variables (total fertility rate, life expectancy, old-age dependencies, etc.)

# Stochastic population forecasts

- Forecasts represent the median of the future distribution.
- Percentiles of distribution allow information about uncertainty
- Prediction intervals with specified probability coverage for population size and all derived variables (total fertility rate, life expectancy, old-age dependencies, etc.)
- The probability of future events can be estimated.

# Stochastic population forecasts

- Forecasts represent the median of the future distribution.
- Percentiles of distribution allow information about uncertainty
- Prediction intervals with specified probability coverage for population size and all derived variables (total fertility rate, life expectancy, old-age dependencies, etc.)
- The probability of future events can be estimated.
- Economic planning is better based on prediction intervals rather than mean or median forecasts.

# Stochastic population forecasts

- Forecasts represent the median of the future distribution.
- Percentiles of distribution allow information about uncertainty
- Prediction intervals with specified probability coverage for population size and all derived variables (total fertility rate, life expectancy, old-age dependencies, etc.)
- The probability of future events can be estimated.
- Economic planning is better based on prediction intervals rather than mean or median forecasts.
- Stochastic models allow true policy analysis to be carried out.

# Functional time series

# Functional time series model

$y_t(x)$ = log mortality rate, log fertility rate,
or net migration numbers for age $x$ in year $t$.

We observe

$$y_t(x) = f_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

$$x = 0, \ldots, 100;$$
$$t = 1 \ldots, n$$

# Functional time series model

$y_t(x)$ = log mortality rate, log fertility rate, or net migration numbers for age $x$ in year $t$.

We observe

$$\boxed{y_t(x) = f_t(x) + \sigma_t(x)\varepsilon_{t,x}} \qquad \begin{array}{l} x = 0, \dots, 100; \\ t = 1 \dots, n \end{array}$$

- $f_t(x)$ is underlying smooth function of $x$ (age)

# Functional time series model

$y_t(x)$ = log mortality rate, log fertility rate,
or net migration numbers for age $x$ in year $t$.

We observe

$$\boxed{y_t(x) = f_t(x) + \sigma_t(x)\varepsilon_{t,x}} \qquad \begin{array}{l} x = 0, \dots, 100; \\ t = 1 \dots, n \end{array}$$

- $f_t(x)$ is underlying smooth function of $x$ (age)
- $\varepsilon_{t,x} \overset{\text{iid}}{\sim} \mathrm{N}(0, 1)$

# Functional time series model

$y_t(x)$ = log mortality rate, log fertility rate, or net migration numbers for age $x$ in year $t$.

We observe

$$\boxed{y_t(x) = f_t(x) + \sigma_t(x)\varepsilon_{t,x}} \qquad \begin{aligned} x &= 0, \dots, 100; \\ t &= 1 \dots, n \end{aligned}$$

- $f_t(x)$ is underlying smooth function of $x$ (age)
- $\varepsilon_{t,x} \overset{\text{iid}}{\sim} \mathrm{N}(0, 1)$
- We want to forecast **whole curve** $y_t(x)$ for $t = n + 1, \dots, n + h$.

# Functional time series model

$$
\begin{aligned}
y_t(x) &= f_t(x) + \sigma_t(x)\varepsilon_{t,x} \\
f_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)
\end{aligned}
$$

where $e_t(x) \sim \mathrm{N}(0, v(x))$.

1. Estimate smooth functions $f_t(x)$ using nonparametric regression.

# Functional time series model

$$\begin{aligned} y_t(x) &= f_t(x) + \sigma_t(x)\varepsilon_{t,x} \\ f_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x) \end{aligned}$$

where $e_t(x) \sim \mathrm{N}(0, v(x))$.

1. Estimate smooth functions $f_t(x)$ using nonparametric regression.
2. Estimate $\mu(x)$ as mean $f_t(x)$ across years.

# Functional time series model

$$
\begin{aligned}
y_t(x) &= f_t(x) + \sigma_t(x)\varepsilon_{t,x} \\
f_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)
\end{aligned}
$$

where $e_t(x) \sim \mathrm{N}(0, v(x))$.

1. Estimate smooth functions $f_t(x)$ using nonparametric regression.
2. Estimate $\mu(x)$ as mean $f_t(x)$ across years.
3. Estimate $\beta_{t,k}$ and $\phi_k(x)$ using principal components.

# Functional time series model

$$
\begin{aligned}
y_t(x) &= f_t(x) + \sigma_t(x)\varepsilon_{t,x} \\
f_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)
\end{aligned}
$$

where $e_t(x) \sim \mathrm{N}(0, v(x))$.

1. Estimate smooth functions $f_t(x)$ using nonparametric regression.
2. Estimate $\mu(x)$ as mean $f_t(x)$ across years.
3. Estimate $\beta_{t,k}$ and $\phi_k(x)$ using principal components.
4. Forecast $\beta_{t,k}$ using time series models.

# Functional time series model

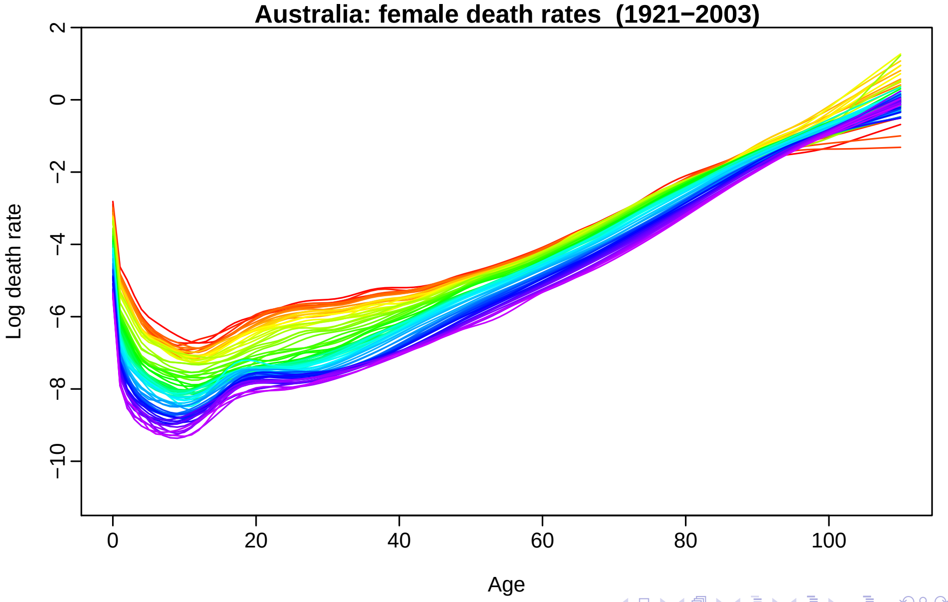
$$
\begin{aligned}
y_t(x) &= f_t(x) + \sigma_t(x)\varepsilon_{t,x} \\
f_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)
\end{aligned}
$$

where $e_t(x) \sim \mathrm{N}(0, v(x))$.

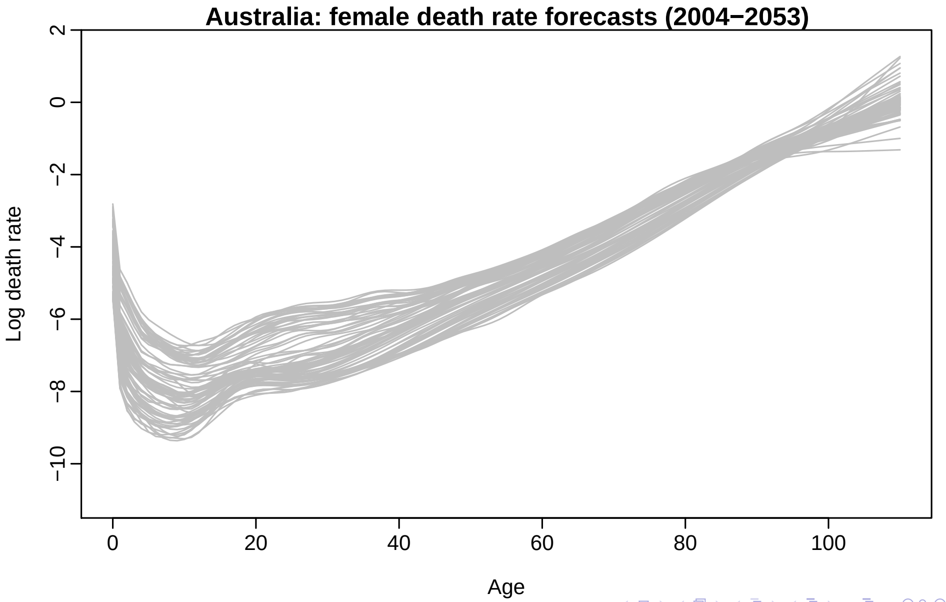
1. Estimate smooth functions $f_t(x)$ using nonparametric regression.
2. Estimate $\mu(x)$ as mean $f_t(x)$ across years.
3. Estimate $\beta_{t,k}$ and $\phi_k(x)$ using principal components.
4. Forecast $\beta_{t,k}$ using time series models.
5. Put it all together to get forecasts of $y_t(x)$.

# Functional time series

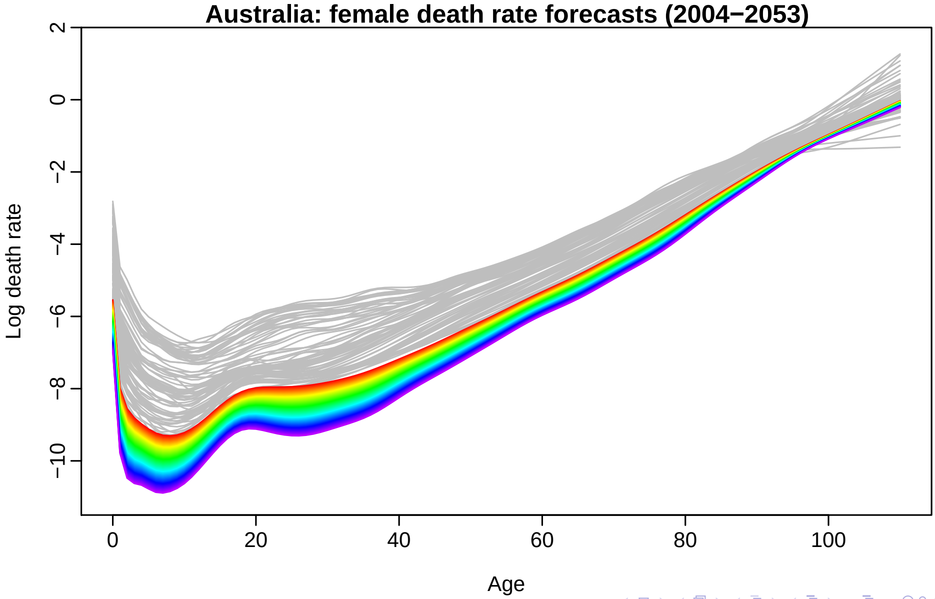
# Functional time series



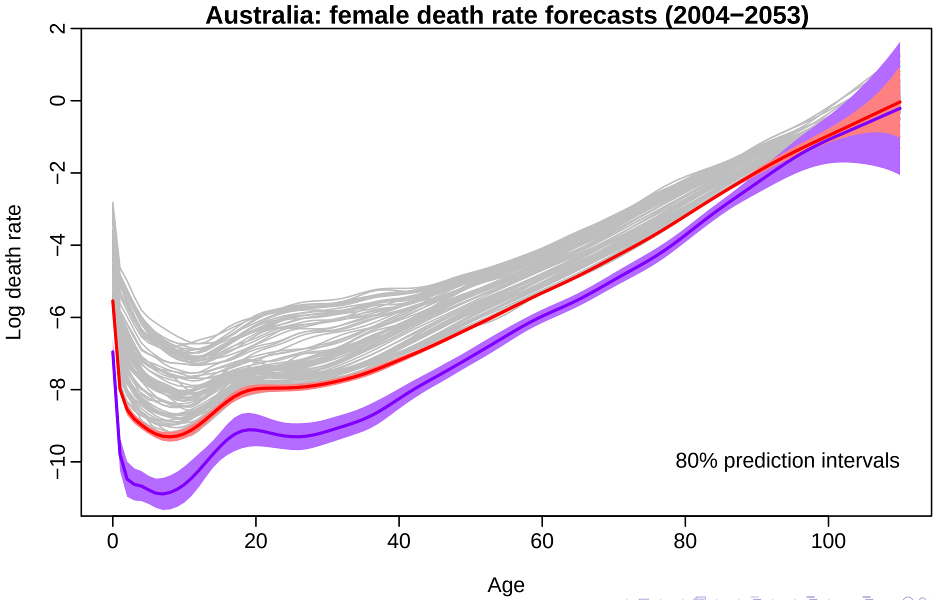
# Functional time series



# Functional forecasts



# Functional forecasts



# Stochastic population forecasts

## Component models

- Data: age/sex-specific mortality rates, fertility rates and net migration. Treat data as **functions of age**.

# Stochastic population forecasts

## Component models

- Data: age/sex-specific mortality rates, fertility rates and net migration. Treat data as **functions of age**.
- Models: Five functional time series models for mortality (M/F), fertility and net migration (M/F) assuming independence between components.

# Stochastic population forecasts

## Component models

- Data: age/sex-specific mortality rates, fertility rates and net migration. Treat data as **functions of age**.
- Models: Five functional time series models for mortality (M/F), fertility and net migration (M/F) assuming independence between components.
- Generate many different future sample paths giving the entire age distribution at every year into the future.

# Stochastic population forecasts

## Component models

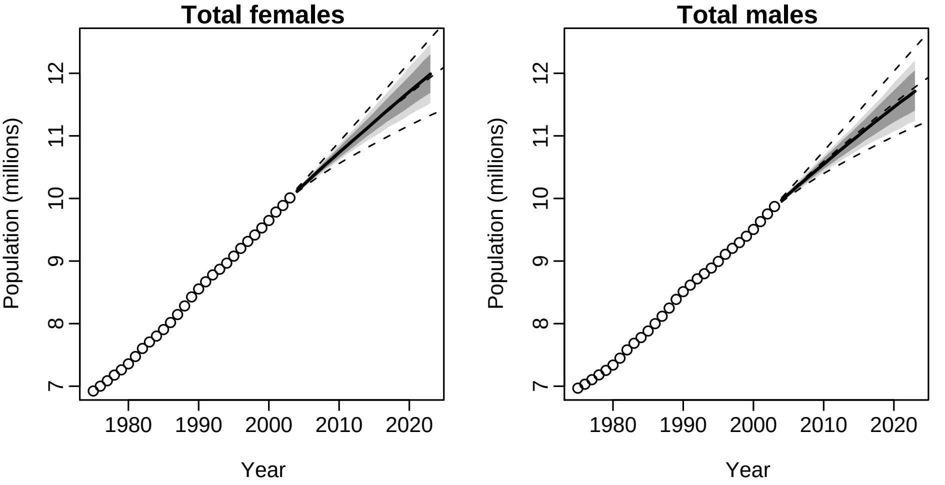
- Data: age/sex-specific mortality rates, fertility rates and net migration. Treat data as **functions of age**.
- Models: Five functional time series models for mortality (M/F), fertility and net migration (M/F) assuming independence between components.
- Generate many different future sample paths giving the entire age distribution at every year into the future.
- Compute future births, deaths, net migrants and populations from simulated rates.

# Stochastic population forecasts

## Component models

- Data: age/sex-specific mortality rates, fertility rates and net migration. Treat data as **functions of age**.
- Models: Five functional time series models for mortality (M/F), fertility and net migration (M/F) assuming independence between components.
- Generate many different future sample paths giving the entire age distribution at every year into the future.
- Compute future births, deaths, net migrants and populations from simulated rates.
- Approach depends on good data and good models for fertility, mortality, and net migration.

# Stochastic population forecasts



# Stochastic population forecasts

## Some recent references

- Booth (2004) On the importance of being uncertain: forecasting population futures for Australia. *People and Place*, **12**(2), 1–11.

# Stochastic population forecasts

## Some recent references


- Booth (2004) On the importance of being uncertain: forecasting population futures for Australia. *People and Place*, **12**(2), 1–11.
- Hyndman and Booth (2006) Stochastic population forecasts using functional data models for mortality, fertility and migration. Monash working paper.

# Outline

# Conclusions

- Uncertainty statements are **essential** when making predictions and should be provided whether they are asked for or not.

# Conclusions

- Uncertainty statements are **essential** when making predictions and should be provided whether they are asked for or not.
- Uncertainty statements can take the form of prediction intervals or prediction densities.

# Conclusions

- Uncertainty statements are **essential** when making predictions and should be provided whether they are asked for or not.
- Uncertainty statements can take the form of prediction intervals or prediction densities.
- Data miners have understood some of the big issues but have misunderstood the importance of uncertainty statements.

# Conclusions

- Uncertainty statements are **essential** when making predictions and should be provided whether they are asked for or not.
- Uncertainty statements can take the form of prediction intervals or prediction densities.
- Data miners have understood some of the big issues but have misunderstood the importance of uncertainty statements.
- Data mining has been enthusiastically adopted in business, but statistical methods have not.

# Some further thoughts

- Our traditional emphasis on p-values has been distracting and unhelpful.

# Some further thoughts

- Our traditional emphasis on p-values has been distracting and unhelpful.
- Most statistical problems are better expressed as prediction problems.

# Some further thoughts

- Our traditional emphasis on p-values has been distracting and unhelpful.
- Most statistical problems are better expressed as prediction problems.

# Some further thoughts

- Our traditional emphasis on p-values has been distracting and unhelpful.
- Most statistical problems are better expressed as prediction problems.

## Forecasts about statistics

- Data mining and statistics will converge.

# Some further thoughts

- Our traditional emphasis on p-values has been distracting and unhelpful.
- Most statistical problems are better expressed as prediction problems.

## Forecasts about statistics

- Data mining and statistics will converge.
- It will become common to compute prediction intervals for data mining methods.

# Some further thoughts

- Our traditional emphasis on p-values has been distracting and unhelpful.
- Most statistical problems are better expressed as prediction problems.

## Forecasts about statistics

- Data mining and statistics will converge.
- It will become common to compute prediction intervals for data mining methods.
- Statisticians will place less emphasis on p-values and more emphasis on predictive ability of models.

# Some further thoughts

- Our traditional emphasis on p-values has been distracting and unhelpful.
- Most statistical problems are better expressed as prediction problems.

## Forecasts about statistics

- Data mining and statistics will converge.
- It will become common to compute prediction intervals for data mining methods.
- Statisticians will place less emphasis on p-values and more emphasis on predictive ability of models.

# Some further thoughts

- Our traditional emphasis on p-values has been distracting and unhelpful.
- Most statistical problems are better expressed as prediction problems.

## Forecasts about statistics

- Data mining and statistics will converge.
- It will become common to compute prediction intervals for data mining methods.
- Statisticians will place less emphasis on p-values and more emphasis on predictive ability of models.

***Uncertainty limits: within 10 years $\pm$5 years.***

# Some further thoughts

## Implications for statisticians

- We need to learn what data miners are doing and contribute to the development of methods.

# Some further thoughts

## Implications for statisticians

- We need to learn what data miners are doing and contribute to the development of methods.
- We need to know (and teach) a lot more computer science.

# Some further thoughts

## Implications for statisticians

- We need to learn what data miners are doing and contribute to the development of methods.
- We need to know (and teach) a lot more computer science.
- We need to teach data mining methods alongside statistical methods in universities.

# Final comments

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

*(Anonymous)*

# Final comments

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

*(Anonymous)*

Slides available from
**http://www.robhyndman.info/**