MONASH University

MONASH BUSINESS SCHOOL

ETW3420: Principles of Forecasting and Applications

Topic 1: Introduction

Dr. Jason Ng

Outline

1. What can we forecast?
2. Time series data
3. The Forecasting Process
4. Forecasting Models
5. Some case studies
6. The statistical forecasting perspective

Forecasting is difficult

A Timeline of Very Bad Future Predictions

1800

"Rail travel at high speed is not possible because passengers, unable to breathe, would die of asphyxia."

Dr. Dionysys Larder, Professor of Natural Philosophy & Astronomy, University College London

1859

"Drill for oil? You mean drill into the ground to try and find oil? You're crazy!"

Associates of Edwin L. Drake refusing his suggestion to drill for oil in 1859 (Later that year, Drake succeeded in drilling the first oil well.)

1876

"This telephone has too many shortcomings to be seriously considered as a means of communication."

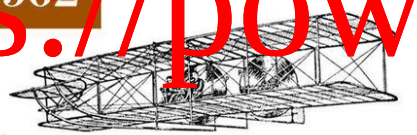
Western Union internal memo

1880

"Everyone acquainted with the subject will recognize it as a conspicuous failure."

Henry Morton, president of the Stevens Institute of Technology, on Edison's light bulb

1902



"Flight by machines heavier than air is unpractical and insignificant, if not utterly impossible."

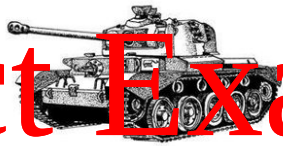
Simon Newcomb, Canadian-American astronomer and mathematician, 18 months before the Wright Brothers' flight at Kitty Hawk

1903

"The horse is here to stay, but the automobile is only a novelty, a fad."

The president of the Michigan Savings Bank, advising Henry Ford's lawyer not to invest in the Ford Motor Company

1916



"The idea that cavalry will be replaced by these iron coaches is absurd. It is little short of treasonous."

Comment of Aide-de-camp to Field Marshal Haig, at tank demonstration

1918

"The cinema is little more than a fad. It's canned drama. What audiences really want to see is flesh and blood on the stage."

Charlie Chaplin, actor, producer, director, and studio founder

1921

"The wireless music box has no imaginable commercial value. Who would pay for a message sent to no one in particular?"

Associates of commercial radio and television pioneer, David Sarnoff, responding to his call for investment in the radio

1946

"Television won't last because people will soon get tired of staring at a plywood box every night."

Darryl Zanuck, movie producer, 20th Century Fox

1977

"There is no reason for any individual to have a computer in his home."

Ken Olsen, president, chairman and founder of Digital Equipment Corporation

1995

"The truth is no online database will replace your daily newspaper…"

Clifford Stoll, Newsweek article entitled ***The Internet? Bah!***

What can we forecast?

What can we forecast?

What can we forecast?

Which is easiest to forecast?

1. daily electricity demand in 3 days time
2. timing of next Hailey's comet appearance
3. time of sunrise this day next year
4. Google stock price tomorrow
5. Google stock price in 6 months time
6. maximum temperature tomorrow
7. exchange rate of $US/RM next week

Which is easiest to forecast?

1. daily electricity demand in 3 days time
2. timing of next Hailey's comet appearance
3. time of sunrise this day next year
4. Google stock price tomorrow
5. Google stock price in 6 months time
6. maximum temperature tomorrow
7. exchange rate of $US/RN1 next week

- how do we measure "easiest"?
- what makes something easy/difficult to forecast?

Factors affecting forecastability

Something is easier to forecast if:

- we have a good understanding of the factors that contribute to it
- there is lots of data available;
- the forecasts cannot affect the thing we are trying to forecast.
- there is relatively low natural/unexplainable random variation.
- the future is somewhat similar to the past

What can be forecast?

- Many people wrongly assume that forecasts are not possible in a changing environment.
- Every environment is changing, and a good forecasting model captures the way things are changing.
- Forecasts rarely assume that the environment is unchanging.

What can be forecast?

- What is normally assumed is that the way the environment is changing will continue into the future - a highly volatile environment will continue to be highly volatile; a business with fluctuating sales will continue to have fluctuating sales; an economy that has gone through booms and busts will continue to go through booms and busts.

- A forecasting model is intended to capture the way things move, not just where things are.

Outline

1. What can we forecast?
2. **Time series data**
3. The Forecasting Process
4. Forecasting Models
5. Some case studies
6. The statistical forecasting perspective

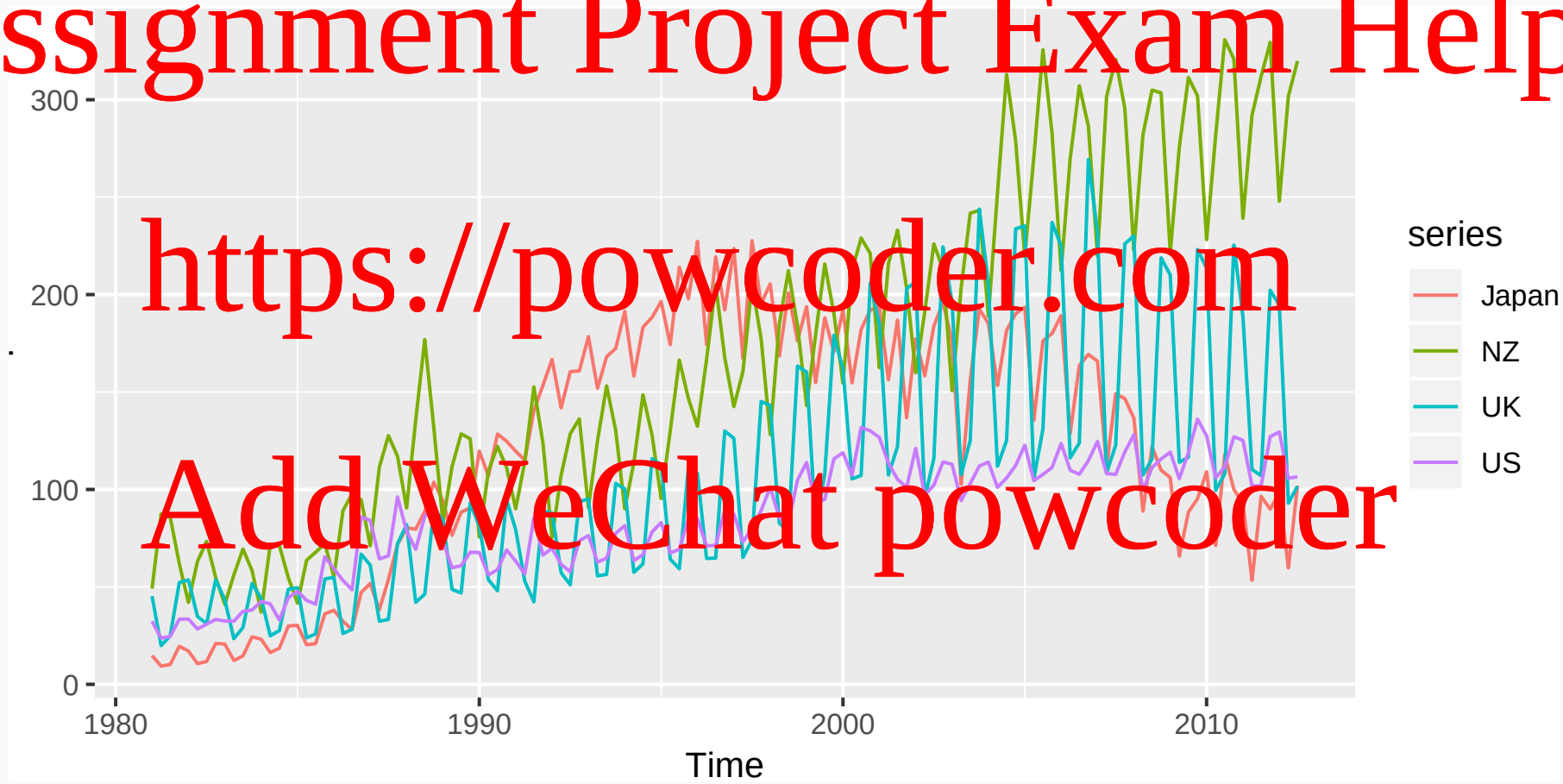
Time series data

Definition

Time series data are observations on some variable over time, **usually** at regular intervals.

Time series data - Example

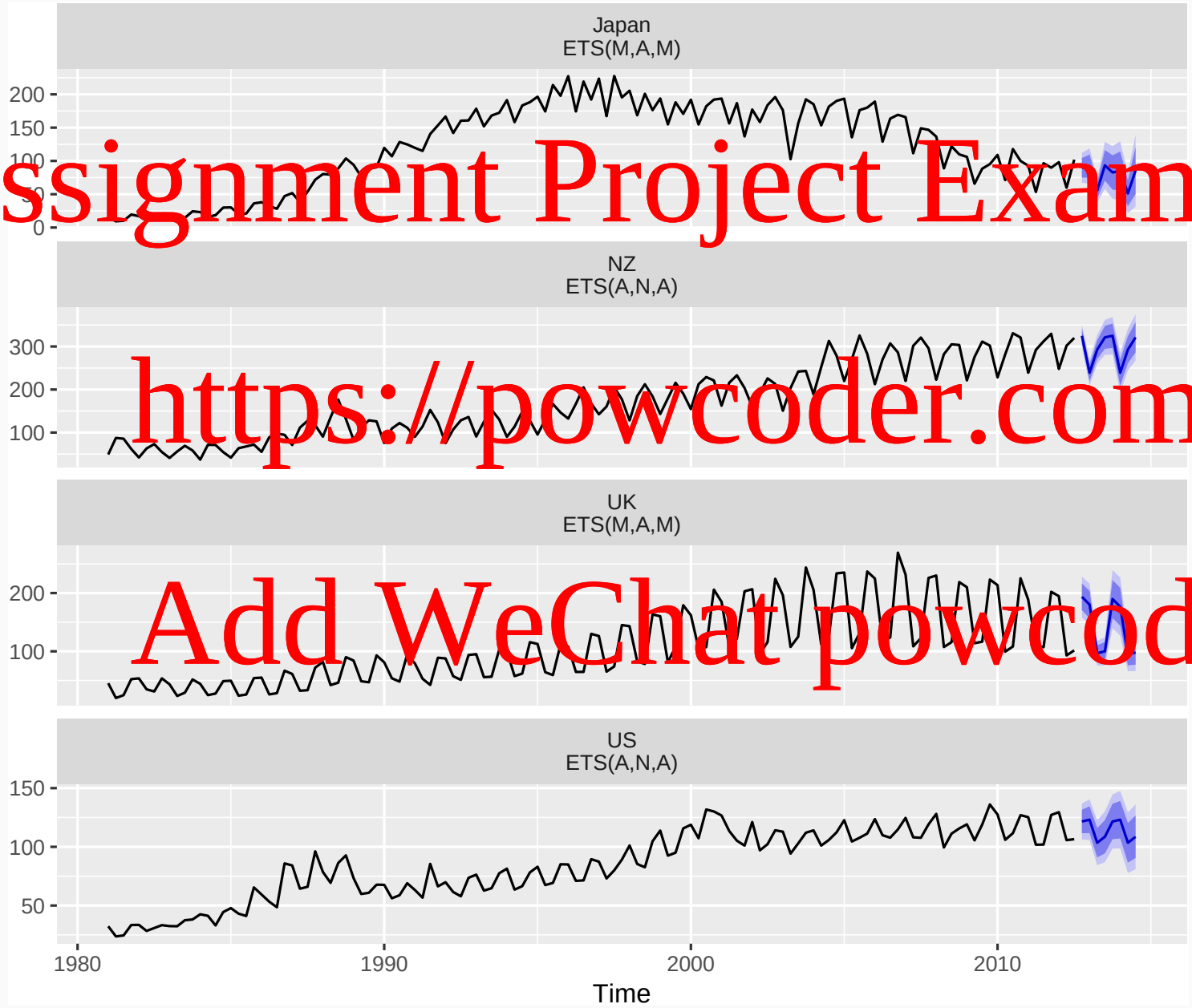
- International arrivals to Australia



Forecasting

Forecasting is estimating how the sequence of observations will continue into the future.

Time series forecasting



Outline

1 What can we forecast?

2 Time series data

3 The Forecasting Process

4 Forecasting Models

5 Some case studies

6 The statistical forecasting perspective

The Forecasting Process

1. Define goal
2. Get data
3. Explore and visualize series
4. Pre-process data
5. Partition series - training/test sets
6. Apply forecasting method(s)/model(s)
7. Evaluate and compare forecasting performance
8. Implement forecasts/systems

Goal Definition

- The first step in the forecasting process is to define the goal of forecasting: Why are we producing forecasts for a particular variable?

- Must be able to motivate the need to forecast a variable of interest.
- There is a subtle difference as to why a variable of interest might be important, and why forecasting that variable might be important.

Goal Definition

- Goals can be categorised as either *descriptive* or *predictive*.
- Descriptive goal: impact assessment; causal arguments
- Predictive goal: purely forecasting

Descriptive Goal Example

Japan Tourism Recovery Strategies Post 2011 Earthquake

Andrea Kwok Chin Jui (26380863) Supervisor: Dr Elaine Chew Yin Teng

Background of the research

On March 11, 2011, a 9.0 magnitude earthquake, better known as the Great East Japan Earthquake (GEJE) struck Tohoku, Japan. It was the deadliest and largest in the country's history. This was followed by a 40 metre high tsunami which devastated three reactors of Fukushima Daiichi nuclear power plant resulting in high radiation leakage into the sea, soil and air. Unprecedented fear was rife amongst the locals as well as international tourists that such leakage would affect the whole ecosystem.

Subsequent to this triple disaster are:

- The harmful impact from the radiation leakage adding complications to the challenge of disaster management for both earthquake and tsunami despite Japan being renowned for its leadership in disaster management.
- Tourism is one of the top six income generating industries in Japan (MLIT, 2012). It is the most impacted by the disaster.

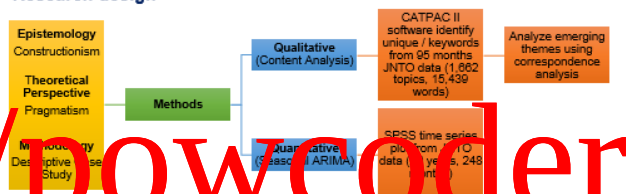
Research problem and gap

- Japan is renowned for disaster management leadership. However, not much is known about Japan's disaster management strategies.
- Lack of study on Japan's disaster management framework and its contribution to tourism recovery in Japan. Existing frameworks are Western-centric and limited.
- The unique case of GEJE – triple disaster challenge (earthquake, tsunami, nuclear radiation leakage).
- Japan has successfully developed different disaster management strategies to recover its tourism industry. This success with valuable lessons has not been studied.

Research objectives

1. To understand in great depth the approaches undertaken (WHAT and HOW) by the Japanese government to address tourism-related concerns.
2. To understand how the Japanese government reform Japan's destination image as a means of wooing back inbound tourists in order to recover its tourism industry.
3. To investigate the extent to which the Japanese government's efforts in disaster management and promotion strategies of Japan tourism attract international tourists (tourist arrival statistics).

Research design



Methodology and methods

Content analysis as a qualitative method

- Content analysis of the information on Japan National Tourism Organization (JNTO) website provides understanding of how the destination image of Japan is projected to spur post-disaster recovery efforts.

Seasonal ARIMA as a quantitative method

- Seasonal Autoregressive Integrated Moving Average (Seasonal ARIMA) is used as a triangulation to validate the extent of the efforts by the Japanese government in contributing to the tourism recovery. This is based on 10 years data of inbound tourists to Japan obtained from Japan National Tourism Organization. Forecasts are compared with the actual number of visitor arrivals (Huang, Min, 2002, Tsui, Balli, Gilbey & Gow, 20
- ARIMA $(p, d, q) \times (P, D, Q)S$ (Non-seasonal part of the model) × (Seasonal part of the model)
- where p = non-seasonal AR order, d = non-seasonal differencing, q = non-seasonal MA order, P = seasonal AR order, D = seasonal differencing, Q = seasonal MA order, and S = time span of repeating seasonal pattern.

Theoretical framework

Generic Framework

Japanese Framework

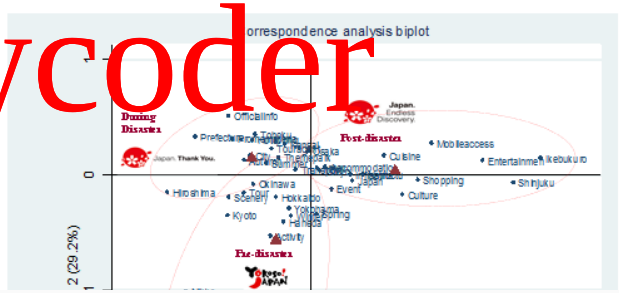
Data Analysis

Content Analysis with CATPAC II shows the most frequent keywords identified by rank (highest to lowest frequency % according to three different periods of promotion strategy (Stepchenkova & Morrison, 2006; Choi, Lehto & Morrison, 2007).

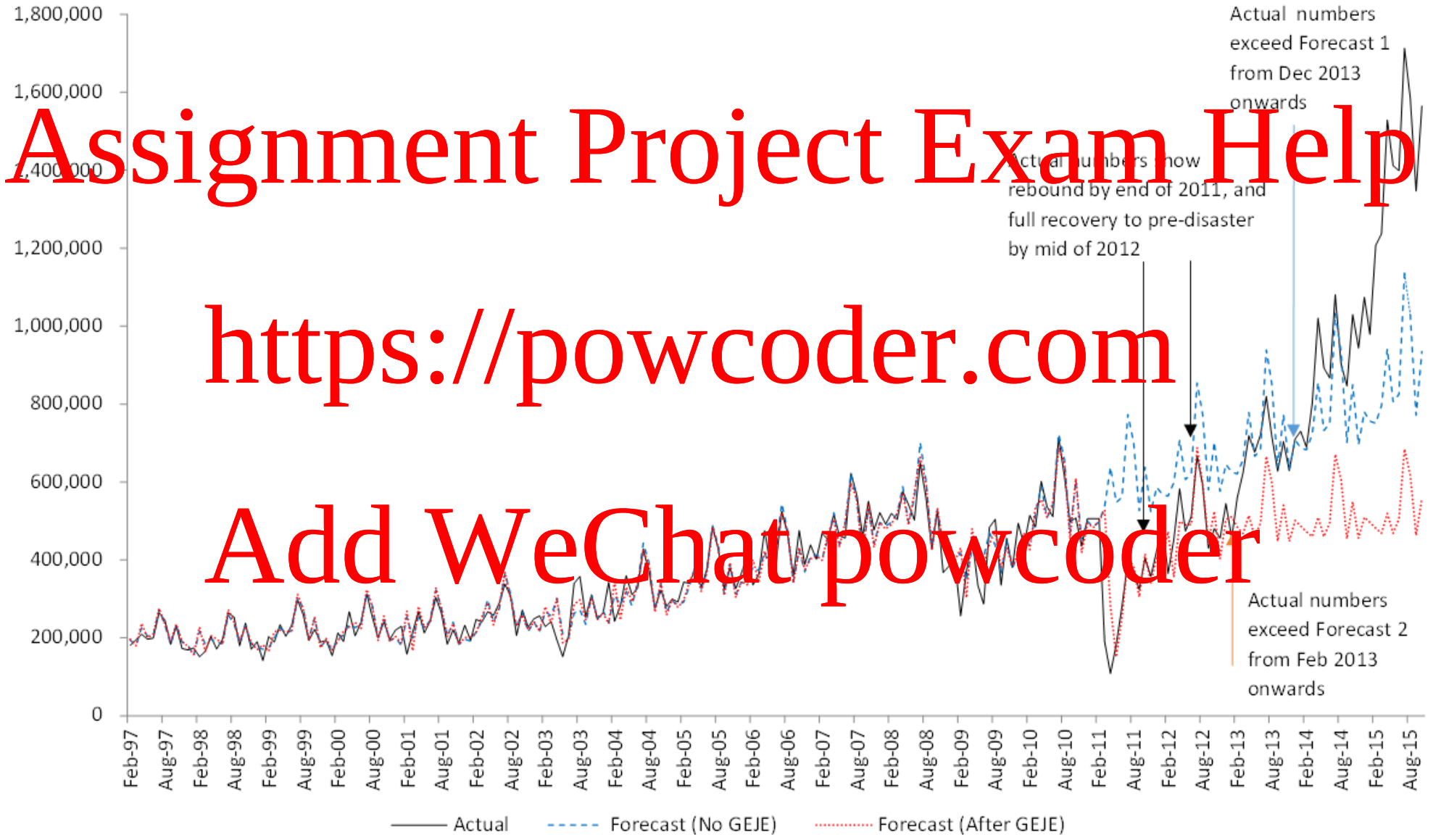
Rank	Pre-disaster		Disaster		Post Disaster	
1	Japan	2.406	Transportation	2.392	Japan	2.008
2	Event	1.5	Prefecture	2.262	Event	1.173
3	Tour	1.84	Japan	2.040	Transportation	1.173
4	Transportation	1.84	Tour Agent	1.873	Entertainment	1.119
5	Spring	1.343	Tour	1.595	Shopping	1.105
6	Activity	1.287	Event	1.446	Infrastructure	0.795

*this list is not exhaustive due to limitation of space.

The list generated from CATPAC II is translated to a correspondence bi-plot in order to provide clarity on the emphasis of distinct or unique keywords used extensively during each promotion period, thereby demonstrating the thematic characteristics of each promotion.



Descriptive Goal Example



Outline

1. What can we forecast?
2. Time series data
3. The Forecasting Process
4. **Forecasting Models**
5. Some case studies
6. The statistical forecasting perspective

Time Series Models

Definition

Time series models use only information on the variable to be forecast.

$$y_{t+1} = f(y_t, y_{t-1}, y_{t-2}, \ldots, error)$$

where t is time and y is the variable of interest.

e.g. ARIMA models and Exponential Smoothing models

- useful when predictor variables not known or measured.
- doesn't lead to much understanding of system.

Cross-Sectional Models

Definition

Cross-sectional models assume that variable to be forecast is affected by one or more predictor variables.

$$y = f(x_1, x_2, \ldots, x_k, error)$$

Example:

Electricity demand = f(. . .)

- useful when predictor variables are known or measured.
- regression models

Mixed Models

Definition

Mixed models combine features of both the time series and cross-sectional models.

$$y_{t+1} = f(y_t, y_{t-1}, \ldots, x_1, x_2, \ldots, x_k, \text{error})$$

Example:

- Dynamic regression models
- Autoregressive distributed lag models
- Panel data models

Time series models vs Cross-sectional models

Several reasons why a forecaster might select a time series model rather than an explanatory mode :

1. System may not be understood, and even if it was understood, it may be extremely difficult to measure the relationships that are assumed to govern its behaviour.
2. It is necessary to know or forecast the various predictors in order to be able to forecast the variable of interest.
3. Main concern may be only to predict what will happen, not to know why it happens.

Outline

1. What can we forecast?
2. Time series data
3. The Forecasting Process
4. Forecasting Models
5. **Some case studies**
6. The statistical forecasting perspective

CASE STUDY 1: Paperware company

Problem: Want forecasts of each of hundreds of items. Series can be stationary, trended or seasonal. They currently have a large forecasting program written in-house but it doesn't seem to produce sensible forecasts. They want to know what is wrong and fix it.

Additional information

- Program written in COBOL making numerical calculations limited. It is not possible to do any optimisation.
- Their programmer has little experience in numerical computing.
- They employ no statisticians and want the program to produce forecasts automatically.

CASE STUDY 1: Paperware company

Methods currently used

- **A** 12 month average
- **C** 6 month average
- **F** straight line regression over last 12 months
- **G** straight line regression over last 6 months
- **H** average slope between last year's and this year's values. (Equivalent to differencing at lag 12 and taking mean.)
- **I** Same as H except over 6 months.

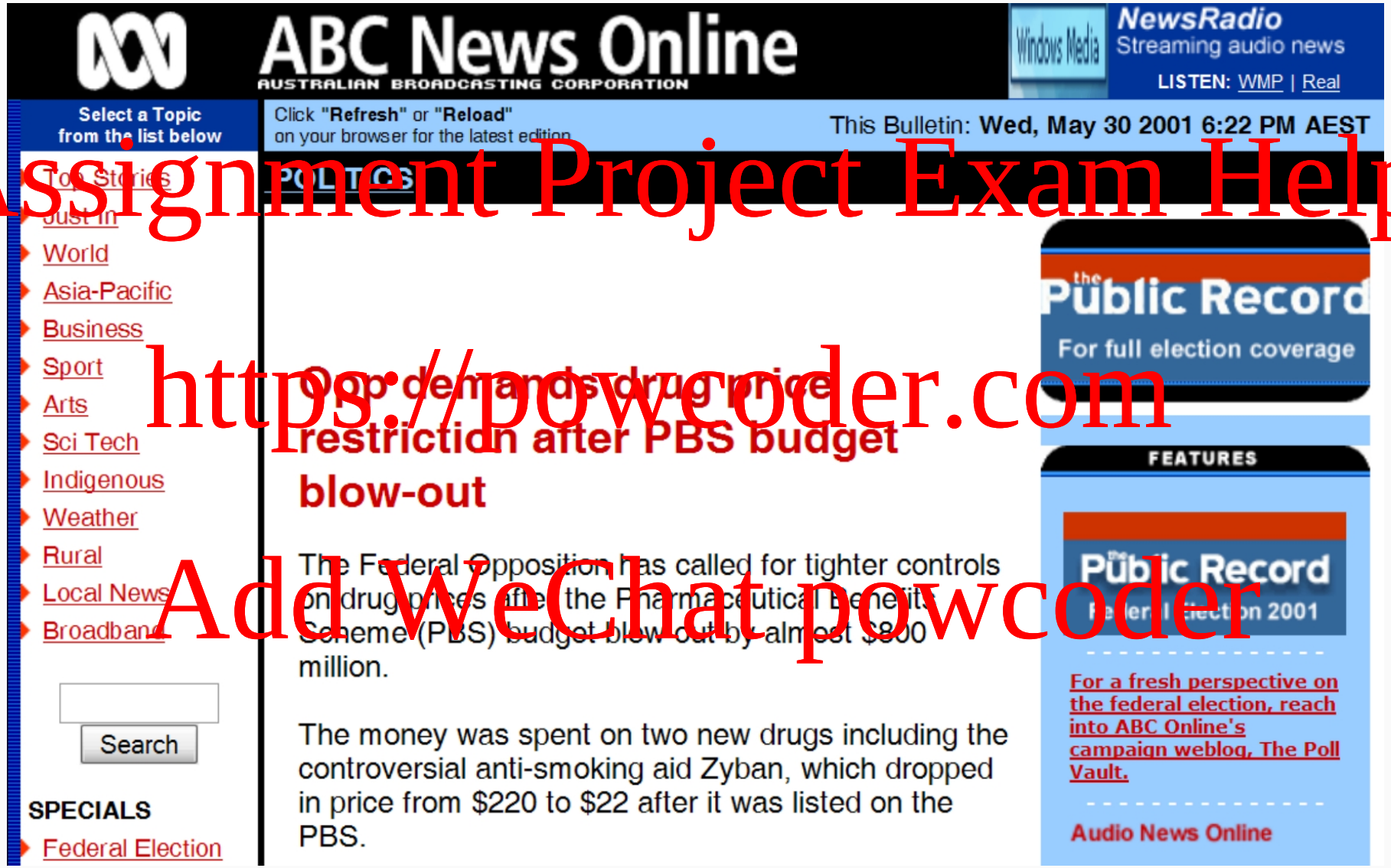
CASE STUDY 2: PBS

CASE STUDY 2: PBS

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

- Many drugs bought from pharmacies 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.

CASE STUDY 2: PBS



CASE STUDY 2: 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!

Outline

1. What can we forecast?
2. Time series data
3. The Forecasting Process
4. Forecasting Models
5. Some case studies
6. **The statistical forecasting perspective**

Basic Notation

$t = 1, 2, \ldots, T$

An index denoting the time period of interest. $t = 1$ is the first period in a series; $t = T$ is the last period in the series.

$y_1, y_2, \ldots, y_T$

A series of T values measured over T periods, where y_t denotes the value of the series at time period t. Also used to denote the random variable.

$\hat{y}_{t+h|1:t}, F_{t+h|1:t}$

The h-step-ahead forecast at time t, conditional on observing $y_1, y_2, \ldots, y_t$. In other words, it is the forecast value of y_{t+h} obtained at time t, given that we observe $y_1, y_2, \ldots, y_t$.

Basic Forecasting Concepts

- A **point forecast** for y_{T+h} is a statement that y_{T+h} will assume a particular value.
- A **forecast interval** is a statement that y_{T+h} will lie in a specified interval with a specified probability.
- Forecast intervals are generally much more useful than point forecasts since they incorporate a statement about how confident we are about the accuracy of our forecast.

Basic Forecasting Concepts

- To distinguish between a **point prediction** and a **point forecast**, assume we have a sample of T observations, $(y_1, y_2, \ldots, y_T)$
- When the predicted value of y is generated for $t = 2, 3, \ldots, T$, using a model, these predictions pertain to **within-sample observations**.
- A forecast of y_{T+j} for $j >= 1$ is a statement about the value of the time series in period $T + j$. These forecasts pertain to **out-of-sample observations**.

Basic Forecasting Concepts

- The forecasting rule that minimizes the mean squared error of the forecast

$$E(FE(h)^2) = E[(y_{T+h} - F_{T+h})^2]$$

is by setting

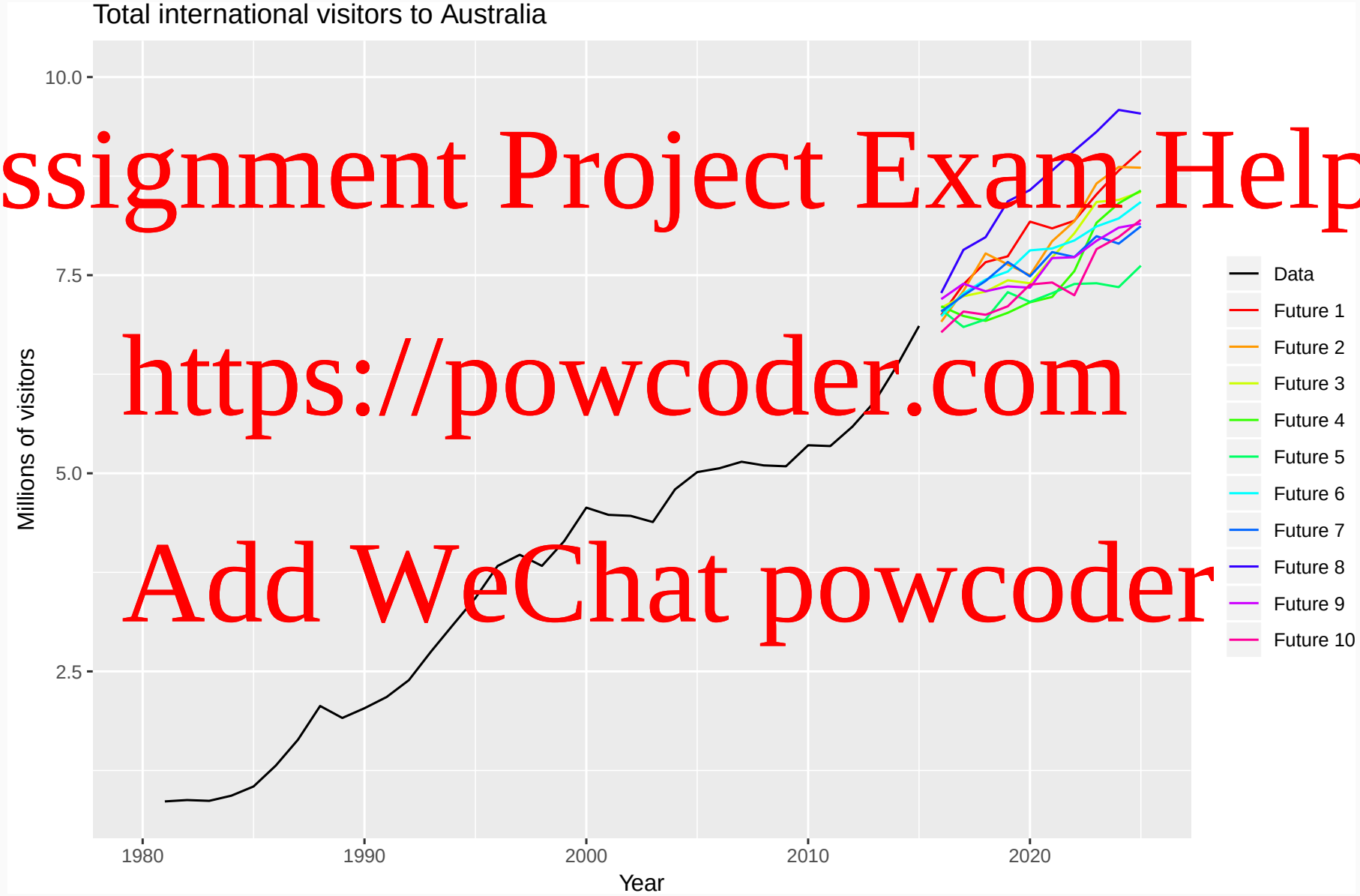
$$F_{T+h} = E(y_{T+h}|y_1, y_2, \ldots, y_T)$$

- However, we don't observe $E(y_{T+h}|y_1, y_2, \ldots, y_T)$, therefore in practise,

$$F_{T+h} = \hat{E}(y_{T+h}|y_1, y_2, \ldots, y_T)$$

where the latter denotes the estimated value of $E(y_{T+h}|y_1, y_2, \ldots, y_T)$.

Sample futures



Forecast intervals