

# ***Measurement and prediction models in consumer credit***

**David J. Hand  
Imperial College**

d.j.hand@ic.ac.uk

<http://stats.ma.ic.ac.uk/~djhand/>

Presented at:  
*Credit Risk Modelling and Decisioning*  
The Wharton School  
30th May 2002

## ***Structure of this talk:***

Overall CRM models and measurement models

The aims of customer measurement models

Difficulties in building customer measurement models

- what do we want to know?
- on what data are they based?
- local and global
- the future

## ***Overall customer management models***

Customers flow through a branching, looping network, interconnecting customer

- *measurements*
  - *decisions*
  - and *actions*
- 
- some driven by the customer
  - some by the supplier
- 
- and the supplier tries to maximise overall profit

*Each measurement is based on a small model of its own*

## ***Customer measurement models 1: What is the aim?***

The *fundamental conflict*:

There are two distinct and often opposed objectives:

*(1) make good decisions about each individual*

*(2) collect data to build better model*

## ***Example 1: make good decisions about each individual***

Hard accept/reject threshold:

- minimises the chance of misclassifying each individual
- maximises the profit, assuming the model is correct
- but gives poor information about how to improve model

e.g. predictor variable: '*County Court Judgements?*'

- 0 indicating no adverse judgements
- 1 indicating one or more such judgements

Suppose this is highly predictive of creditworthiness

Then the original scorecard may be such that anyone who had a score of 1 on this variable is rejected

⇒ a new scorecard constructed from the accepts alone will not include this variable at all (all accepts score 0 on this variable)

- almost all current systems place all the emphasis on (1)
- leaving you to manage as best you can
- *which means that better models could be built*

A couple of exceptions:

- Champion-Challenger
- Gold sample

## ***Example 2: collect data to build the best model***

- data collection strategy which yields most accurate model
- which you can apply if you haven't gone bust
  - e.g. accept everyone who applies, with the same probability  $p$ 
    - much information about how to improve model
    - but huge losses?
- much scope for experimental design tools

## ***What is needed is a compromise***

- for example: *soft accept/reject thresholds*
  - accept customer with probability  $p(s)$ , where  $s$  is their score
  - choose  $p(s)$  distribution to optimise overall profit
    - *which consists of immediate profit*
    - *and discounted future profit dependent on reliable model*

## ***Customer measurement models 2: Building the model***

### **(1) *We want a model which***

- ***compares*** ‘what would happen if I did X?’
- ***with*** ‘what would happen if I did Y?’

‘I think this customer will behave in this undesirable way if I do not intervene. I therefore intervene. Now I cannot tell if the behaviour is due to my intervention or not. I cannot use the outcome to assess the accuracy of my prediction of what would have happened had I not intervened.’

‘Should I send customer letter A or letter B?’

- I send A and want to compare with *what would have happened had I sent B*

Current models answer

- what would happen if I did X?

*or*

- what would happen if I did Y?

*but not*

- the comparison of what would happen if I did X instead of Y?

This information can only be obtained if  $0 < p(s) < 1$

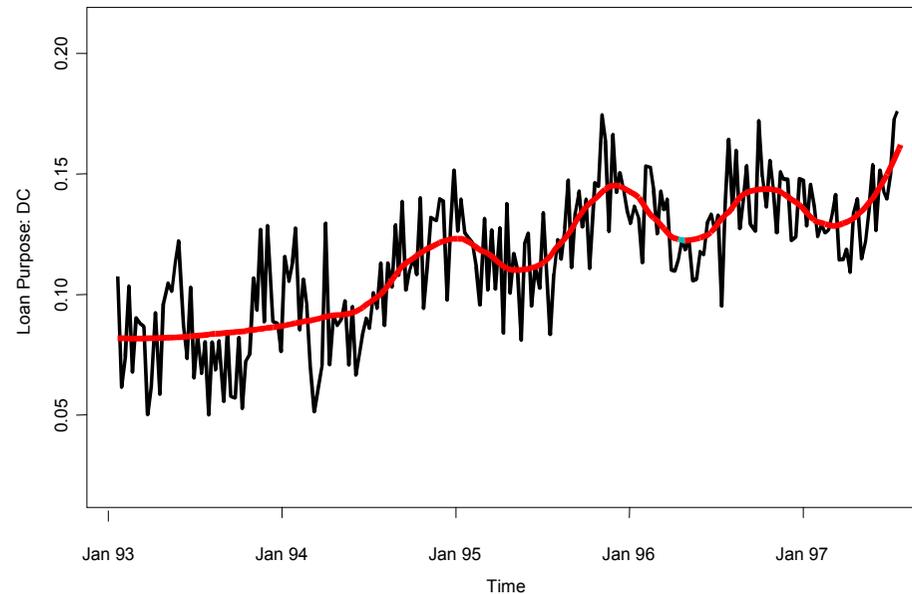
- if everyone has a probability of being assigned to each class

*And note:* the complication that we are dealing with a perverse adversary

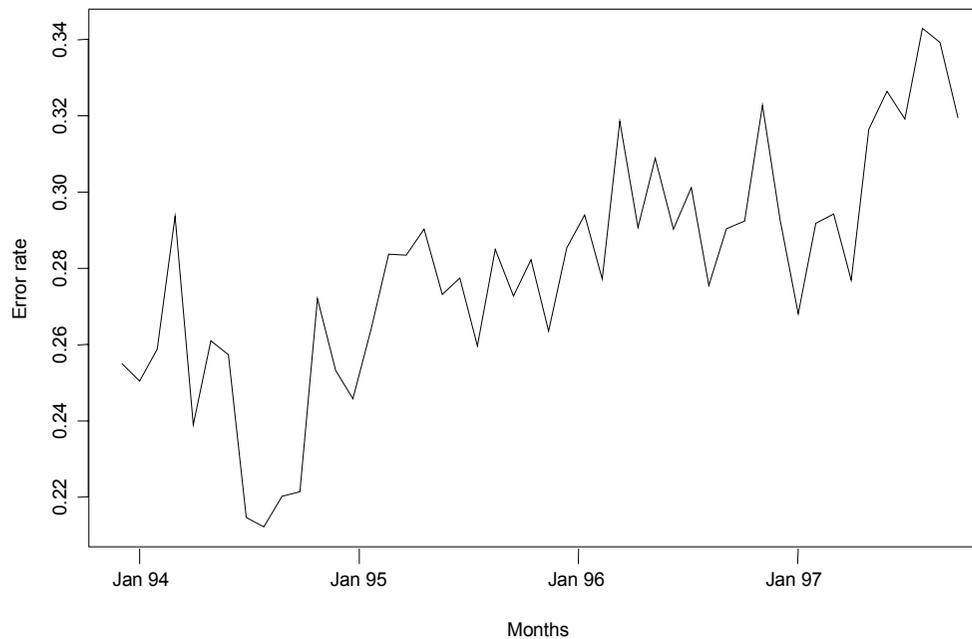
- in normal science, nature does not actively work to frustrate your actions
- in CVM, the customer may

## (2) Design data are out of date

- we want to predict outcome  $X$  years in the future
- so we look back at least  $X$  years in the past, find similar people, and examine their outcomes
- but customer distributions change over time

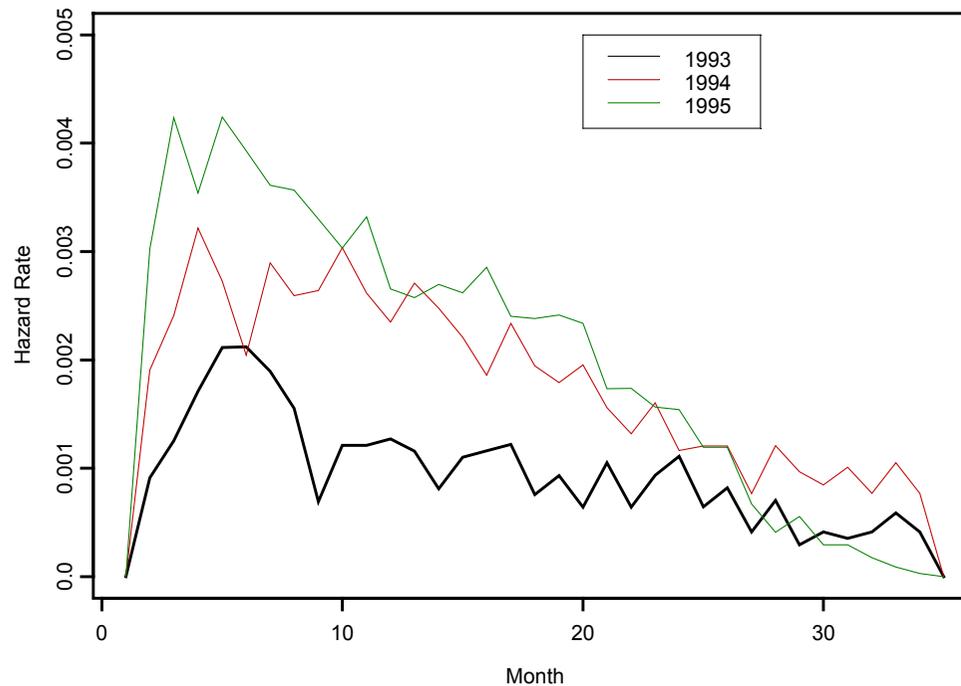


- and scorecard performance changes over time  
(life would be much easier if it didn't)
  - due to changing population
  - due to changing customer behaviour  
(the relative importance of these is an open question)



## Improved model 1:

- survival analysis to predict X years outcome from application characteristics and behaviour up to now
- based on families of survival functions



## **Improved model 2:**

- split model into two parts

A: application data to predict short term behaviour (application model)

B: short term behaviour to predict outcome (behavioural model)

A can be based on very recent data (don't have to wait X years)

B not affected by changing populations

### **(3) Standard models are average models**

#### ***Front end models:***

- potential interaction with customer
  - e.g. initial loan decision
- potential interaction with management
  - interpretable
  - easy to understand
  - easy to implement
  - predictive accuracy but one performance measure

#### ***Back end models:***

- internal decision making
  - e.g. fraud detection
  - as complicated as you like
  - predictive performance is all

## **Different criteria $\Rightarrow$ different methods**

### ***Front end***

- discriminant analysis
- logistic regression
- linear programming
- small trees

### ***Back end***

- neural networks
- projection pursuit
- nearest neighbour
- support vector machines

Fundamental assumption typically made in statistical modelling:

***The inference and prediction stages are separate***

*A statistical inference carries us from observations to conclusions about the populations sampled ...No considerations of losses [consequences] is usually involved directly...The theory of statistical decision deals with the action to take on the basis of statistical information. Decisions are based on not only the considerations listed for inferences, but also on an assessment of the losses resulting from wrong decisions...*

David Cox

*...the inference problem is basic to any decision problem, because the latter can only be solved when the knowledge of [the probability model] ... is completely specified ... The person making the inference need not have any particular decision problem in mind. The scientist in his laboratory does not consider the decisions that may subsequently have to be made concerning his discoveries. His task is to describe accurately what is known about the parameters in question.*

Dennis Lindley

***But this is only true if the models are properly specified***

Which they almost never are

Why does this matter in a credit scoring context?

- because we are not interested in accuracy of  $\hat{p}(0 | \mathbf{x})$  for all  $\mathbf{x}$
- But only in the vicinity of  $p(0 | \mathbf{x}) = t$   
where  $t = c_0 / (c_0 + c_1)$   
with  $c_i$  the cost of misclassifying a class  $i$  customer

Standard statistical approaches, such as logistic regression, *average* over models which are optimal for different  $t$

The model they produce is not best for the  $t$  we want

# Solutions

## 1) Support vector machines

- expand feature space by including many transformations of variables
- fit linear decision surface in transformed space (using math kernel trick)
- to minimise error rate

## 2) Boosting

- a generalised additive model, a weighted sum of base classifiers

## 3) Local nonparametric methods

- kernel smoothing
- loess smoothing
- k-nearest neighbour

But these yield complex models which are fine as back end scorecards

But cannot be used as (interpretable) front end scorecards

We need a local model which is also linear in the variables

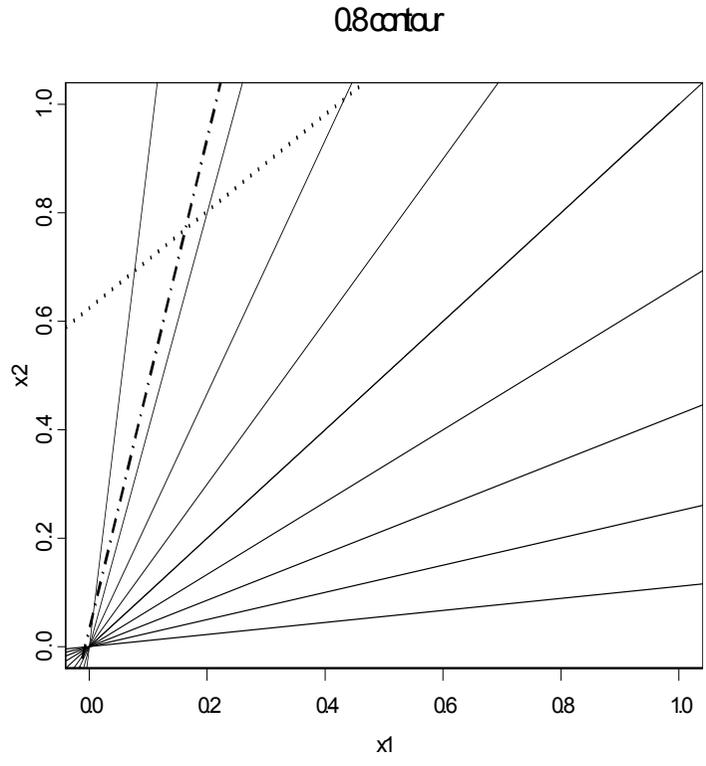
⇒ *locally weighted logistic regression*

$$L = \prod_{i=1}^n \hat{P}(c_i | x_i)$$

$$L = \prod_{i=1}^n \hat{P}(c_i | x_i)^{w_i}$$

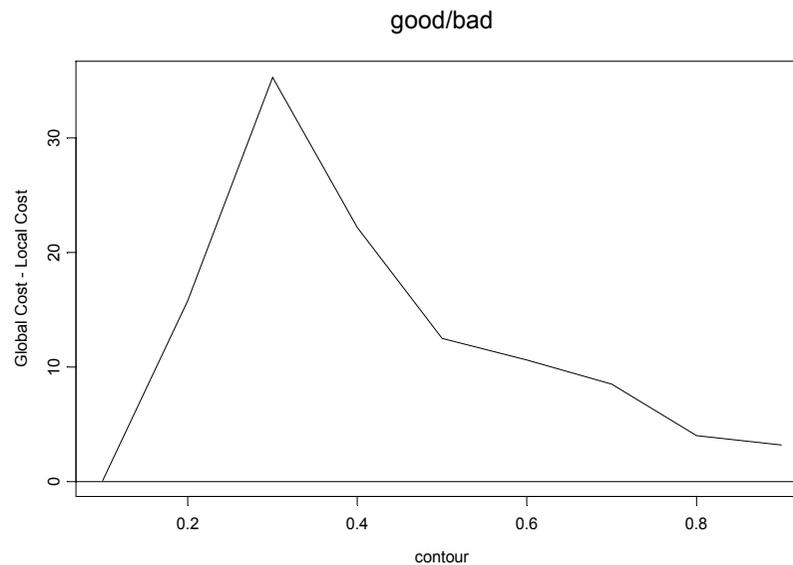
$$w_i = \alpha \exp \left\{ -\lambda (\hat{P}(0 | x_i) - t)^2 \right\}$$

# Example 1



**Example 2:** Data from a major UK credit card company  
Customers divided into 'good' (47%) and 'bad' (53%) users of card  
1490 cases measured on 8 variables

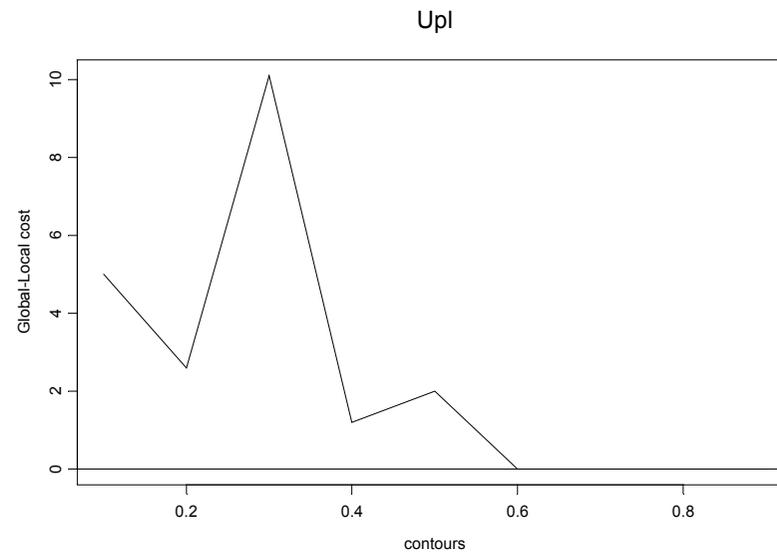
Cost of global-local



## Example 3: Unbalanced classes:

Data supplied by a major UK bank

- 21,618 UPL, 24-month term, two year period Jan 1995 to Dec 1996
- 16 variables describe the application for the loan.
- an account is defined as bad if it is at least three months in arrears
- 11% bads



## The future

- the retail banking environment: decisions by customer and by supplier, in a stochastic environment
- overall models of increasing importance
  - possible, thanks to computers
- quantification and measurement of increasing importance
  - more care about what is measured
  - more sophisticated measurement tools
- dynamic risk management via dynamic stochastic programming: *simulate* possible future decision sets
- risk management as dynamic optimisation