A (personal) view of statistical issues in (mainly retail) credit risk assessment

David J. Hand
Imperial College London

Introduction

Risk management starts with risk assessment
that is, risk measurement

Measuring risk is a statistical estimation/inference problem

Statistical estimation/inference rely on adequate data

Different kinds of risk ➔ different kinds of model

Classic ***risk*** taxonomy:

- market risk
- operational risk
- credit risk
- etc

But also a ***statistical*** risk modelling taxonomy:

- structural models
- reduced form default models (int rates; risk ratings)
- reduced form mark-to-market models
- empirical models

Structural models: model based on financial ‘theory’

... iconic, mechanistic, phenomenological ...

Example: Merton structural model for PD

$$PD = \Phi\left(-\frac{\ln\left(A_0/F\right)}{\sigma_A\sqrt{T}} + \frac{\sigma_A\sqrt{T}}{2} \right)$$

Empirical models: based on observed relationships in (large) data sets

... descriptive,

Example: Credit scoring

Logistic regression

Neural networks

Random forests

Logistic regression tree

Broadly speaking

structural models in the corporate sector

empirical models in the retail sector

Note: earliest work in corporate sector was empirical

e.g. Altman’s linear discriminant analysis to predict company default

Good reasons for the difference:

Different sorts of data available

Corporate default models vs consumer default models

Corporate default models	Consumer default models
Small volume, large loans	Large volume, small loans
Hierarchy of approval: Senior managers approve large complex loans	Hierarchy of approval: Senior managers approve credit products
Audited corporate financial statements	Limited financial data on customers
Corporate payment records available	Credit history, usually incomplete
Close account management and review on loan-by-loan basis	Portfolio management by aggregate statistics performance
Market valuation	Monitor behaviour of individuals over time

Retail characterised by large data sets

Possibly billions of data points

And high dimensions

Credit card data:

- Transaction ID
- Transaction type
- Date and time of transaction (to nearest second)
- Amount
- Currency
- Local currency amount
- Merchant category
- Card issuer ID
- ATM ID
- POS type
- Cheque account prefix
- Savings account prefix
- Acquiring institution ID
- Transaction authorisation code
- Online authorisation performed
- New card
- Transaction exceeds floor limit
- Number of times chip has been accessed
- Merchant city name
- Chip terminal capability
- Chip card verification result

Condense past transactions on a credit line into a number of attributes which summarise behaviour

Combine attributes across credit lines

Example: Experian’s over 440 ‘STAGGs’
http://www.experian.com/products/attributetoolbox.html

Another example: fraud detection

US Patent 5,819,226 (see USPTO website) on *Fraud detection and modeling*, (HNC Software in 1992) lists the following variables:

Customer usage pattern profiles representing time-of-day and day-of-week profiles;
Expiration date for the credit card;
Dollar amount spent in each SIC (Standard Industrial Classification) merchant group category during the current day;
Percentage of dollars spent by a customer in each SIC merchant group category during the current day;
Number of transactions in each SIC merchant group category during the current day;
Percentage of number of transactions in each SIC merchant group category during the current day;
Categorization of SIC merchant group categories by fraud rate (high, medium, or low risk);
Categorization of SIC merchant group categories by customer types (groups of customers that most frequently use certain SIC categories);
Categorization of geographic regions by fraud rate (high, medium, or low risk);
Categorization of geographic regions by customer types;
Mean number of days between transactions;
Variance of number of days between transactions;
Mean time between transactions in one day;
Variance of time between transactions in one day;
Number of multiple transaction declines at same merchant;
Number of out-of-state transactions;
Mean number of transaction declines;
Year-to-date high balance;
Transaction amount;
Transaction date and time;
Transaction type.

“Additional fraud-related variables which may also be considered are listed below”

Current Day Cardholder Fraud Related Variables bweekend current day boolean indicating current datetime considered weekend cavapvdl current day mean dollar amount for an approval cavapvdl current day mean dollar amount for an approval cavaudl current day mean dollars per auth across day ccoscdoni current day cosine of the day of month i.e. cos(day ((datepart(cst.sub.-- dt) * &TWOPI)/30)); ccoscdow current day cosine of the day of week i.e. cos(weekday ((datepart(cst.sub.-- dt) * &TWOPI)/7)); ccoscmoy current day cosine of the month of year i.e. cos(month ((datepart(cst.sub.-- dt) * &TWOPI)/12)); cdom current day day of month cdow current day day of week chdzip current cardholder zip chibal current day high balance chidcapv current day highest dollar amt on a single cash approve chidcdec current day highest dollar amt on a single cash decline chidmapv current day highest dollar amt on a single merch approve chidmdec current day highest dollar amt on a single merch decline chidsapv current day highest dollar amount on a single approve chidsau current day highest dollar amount on a single auth chidsdec current day highest dollar amount on a single decline cmoy current day month of year cratdcau current day ratio of declines to auths csincdom current day sine of the day of month i.e. sin(day ((datepart(cst.sub.-- dt) * &TWOPI)/30)); csincdow current day sine of the day of week i.e. sin(weekday ((datepart(cst.sub.-- dt) * &TWOPI)/7)), csincmoy current day sine of the month of year i.e: sin(month ((datepart(cs.sub.-- dt) * &TWOPI)/12)); cst.sub.-- dt current day cst datetime derived from zip code and CST auth time ctdapv current day total dollars of approvals ctdau current day total dollars of auths ctdcsapv current day total dollars of cash advance approvals ctdcsdec current day total dollars of cash advance declines ctddec current day total dollars of declines ctdmrapv current day total dollars of merchandise approvals ctdmrdec current day total dollars of merchandise declines ctnapv current day total number of approves ctnau current day total number of auths ctnau10d current day number of auths in day <=$10 ctnaudy current day total number of auths in a day ctncsapv current day total number of cash advance approvals ctncsapv current day total number of cash approves ctncsdec current day total number of cash advance declines ctndec current day total number of declines cmmrapv current day total number of merchandise approvals ctnmrdec current day total number of merchandise declines ctnsdapv current day total number of approvals on the same day of week as current day ctnwdaft current day total number of weekday afternoon approvals ctnwdapv current day total number of weekday approvals ctnwdeve current day total number of weekday evening approvals ctnwdmor current day total number of weekday morning approvals ctnwdnit current day total number of weekday night approvals ctnweaft current day total number of weekend afternoon approvals ctnweapv current day total number of weekend approvals ctnweeve current day total number of weekend evening approvals ctnwemor current day total number of weekend morning approvals ctnwenit current day total number of weekend night approvals currbal current day current balance cvraud1 current day variance of dollars per auth across day czrate1 current day zip risk group 1 `Zip very high fraud rate` czrate2 current day zip risk group 2 `Zip high fraud rate` czrate3 current day zip risk group 3 `Zip medium high fraud rate` czrate4 current day zip risk group 4 `Zip medium fraud rate` czrate5 current day zip risk group 5 `Zip medium low fraud rate` czrate6 current day zip risk group 6 `Zip low fraud rate` czrate7 current day zip risk group 7 `Zip very low fraud rate` czrate8 current day zip risk group 8 `Zip unknown fraud rate` ctdsfa01 current day total dollars of transactions in SIC factor group 01 ctdsfa02 current day total dollars of transactions in SIC factor group 02 ctdsfa03 current day total dollars of transactions in SIC factor group 03 ctdsfa04 current day total dollars of transactions in SIC factor group 04 ctdsfa05 current day total dollars of transactions in SIC factor group 05 ctdsfa06 current day total dollars of transactions in SIC factor group 06 ctdsfa07 current day total dollars of transactions in SIC factor group 07 ctdsfa08 current day total dollars of transactions in SIC factor group 08 ctdsfa09 current day total dollars of transactions in SIC factor group 09 ctdsfa10 current day total dollars of transactions in SIC factor group 10 ctdsfa11 current day total dollars of transactions in SIC factor group 11 ctdsra01 current day total dollars of transactions in SIC fraud rate group 01 ctdsra02 current day total dollars of transactions in SIC fraud rate group 02 ctdsra03 current day total dollars of transactions in SIC fraud rate group 03 ctdsra04 current day total dollars of transactions in SIC fraud rate group 04 ctdsra05 current day total dollars of transactions in SIC fraud rate group 05 ctdsra06 current day total dollars of transactions in SIC fraud rate group 06 ctdsra07 current day total dollars of transactions in SIC fraud rate group 07 ctdsva01 current day total dollars in SIC VISA group 01 ctdsva02 current day total dollars in SIC VISA group 02 ctdsva03 current day total dollars in SIC VISA group 03 ctdsva04 current day total dollars in SIC VISA group 04 ctdsva05 current day total dollars in SIC VISA group 05 ctdsva06 current day total dollars in SIC VISA group 06 ctdsva07 current day total dollars in SIC VISA group 07 ctdsva08 current day total dollars in SIC VISA group 08 ctdsva09 current day total dollars in SIC VISA group 09 ctdsva10 current day total dollars in SIC VISA group 10 ctdsva11 current day total dollars in SIC VISA group 11 ctnsfa01 current day total number of transactions in SIC factor group 01 ctnsfa02 current day total number of transactions in SIC factor group 02 ctnsfa03 current day total number of transactions in SIC factor group 03 ctnsfa04 current day total number of transactions in SIC factor group 04 ctnsfa05 current day total number of transactions in SIC factor group 05 ctnsfa06 current day total number of transactions in SIC factor group 06 ctnsfa07 current day total number of transactions in SIC factor group 07 ctnsfa08 current day total number of transactions in SIC factor group 08 ctnsfa09 current day total number of transactions in SIC factor group 09 ctnsfa10 current day total number of transactions in SIC factor group 10 ctnsfa11 current day total number of transactions in SIC factor group 11 ctnsra01 current day total number of transactions in SIC fraud rate group 01 ctnsra02 current day total number of transactions in SIC fraud rate group 02 ctnsra03 current day total number of transactions in SIC fraud rate group 03 ctnsra04 current day total number of transactions in SIC fraud rate group 04 ctnsra05 current day total number of transactions in SIC fraud rate group 05 ctnsra06 current day total number of transactions in SIC fraud rate group 06 ctnsra07 current day total number of transactions in SIC fraud rate group 07 ctnsva01 current day total number in SIC VISA group 01 ctnsva02 current day total number in SIC VISA group 02 ctnsva03 current day total number in SIC VISA group 03 ctnsva04 current day total number in SIC VISA group 04 ctnsva05 current day total number in SIC VISA group 05 ctnsva06 current day total number in SIC VISA group 06 ctnsva07 current day total number in SIC VISA group 07 ctnsva08 current day total number in SIC VISA group 08 ctnsva09 current day total number in SIC VISA group 09 ctnsva10 current day total number in SIC VISA group 10 ctnsva11 current day total number in SIC VISA group 11 7 Day Cardholer Fraud Related Variables raudymdy 7 day ratio of auth days over number of days in the window ravapvdl 7 day mean dollar amount for an approval ravaudl 7 day mean dollars per auth across window rddapv 7 day mean dollars per day of approvals rddapv2 7 day mean dollars per day of approvals on days with auths rddau 7 day mean dollars per day of auths on days with auths rddauall 7 day mean dollars per day of auths on all days in window rddcsapv 7 day mean dollars per day of cash approvals rddcsdec 7 day mean dollars per day of cash declines rdddec 7 day mean dollars per day of declines rdddec2 7 day mean dollars per day of declines on days with auths rddmrapv 7 day mean dollars per day of merchandise approvals rddmrdec 7 day mean dollars per day of merchandise declines rdnapv 7 day mean number per day of approvals rdnau 7 day mean number per day of auths on days with auths rdnauall 7 day mean number per day of auths on all days in window rdncsapv 7 day mean number per day of cash approvals rdncsdec 7 day mean number per day of cash declines rdndec 7 day mean number per day of declines rdnmrapv 7 day mean number per day of merchandise approvals rdnmrdec 7 day mean number per day of merchandise declines rdnsdap2 7 day mean number per day of approvals on same day of week calculated only for those days which had approvals rdnsdapv 7 day mean number per day of approvals on same day of week as current day rdnwdaft 7 day mean number per day of weekday afternoon approvals rdnwdapv 7 day mean number per day of weekday approvals rdnwdeve 7 day mean number per day of weekday evening approvals rdnwdmor 7 day mean number per day of weekday morning approvals rdnwdnit 7 day mean number per day of weekday night approvals rdnweaft 7 day mean number per day of weekend afternoon approvals rdnweapv 7 day mean number per day of weekend approvals rdnweeve 7 day mean number per day of weekend evening approvals rdnwemor 7 day mean number per day of weekend morning approvals rdnwenit 7 day mean number per day of weekend night approvals rhibal 7 day highest window balance rhidcapv 7 day highest dollar amt on a single cash approve rhidcdec 7 day highest dollar amt on a single cash decline rhidmapv 7 day highest dollar amt on a single merch approve rhidmdec 7 day highest dollar amt on a single merch decline rhidsapv 7 day highest dollar amount on a single approve rhidsam 7 day highest dollar amount on a single auth rhidsdec 7 day highest dollar amount on a single decline rhidtapv 7 day highest total dollar amount for an approve in a single day rhidtau 7 day highest total dollar amount for any auth in a single day rhidtdec 7 day highest total dollar amount for a decline in a single day rhinapv 7 day highest number of approves in a single day rhinau 7 day highest number of auths in a single day rhindec 7 day highest number of declines in a single day rnaudy 7 day number of days in window with any auths rnausd 7 day number of same day of week with any auths rnauwd 7 day number of weekdays days in window with any auths rnauwe 7 day number of weekend days in window with any auths rncsaudy 7 day number of days in window with cash auths rnmraudy 7 day number of days in window with merchant auths rtdapv 7 day total dollars of approvals rtdau 7 day total dollars of auths rtdcsapv 7 day total dollars of cash advance approvals rtdcsdec 7 day total dollars of cash advance declines rtddec 7 day total dollars of declines rtdmrapv 7 day total dollars of merchandise approvals rtdmrdec 7 day total dollars of merchandise declines rtnapv 7 day total number of approvals rtnapvdy 7 day total number of approvals in a day rtnan 7 day total number of auths rtnau10d 7 day number of auths in window <= $10 rtncsapv 7 day total number of cash advance approvals rtncsdec 7 day total number of cash advance adeclines rtndec 7 day total number of declines rtnmrapv 7 day total number of merchandise approvals rtnmrdec 7 day total number of merchandise declines rtnsdapv 7 day total number of approvals on the same day of week as current day rtnwdaft 7 day total number of weekday afternoon approvals rtnwdapv 7 day total number of weekday approvals rtnwdeve 7 day total number of weekday evening approvals rtnwdmor 7 day total number of weekday morning approvals rtnwdnit 7 day total number of weekday night approvals rtnweaft 7 day total number of weekend afternoon approvals rtnweapv 7 day total number of weekend approvals rtnweeve 7 day total number of weekend evening approvals rtnwemor 7 day total number of weekend morning approvals rtnwenit 7 day total number of weekend night approvals rvraudl 7 day variance of dollars per auth across window Profile Cardholder Fraud Related Variables paudymdy profile ratio of auth days over number of days in the month pavapvdl profile mean dollar amount for an approval pavaudl profile mean dollars per auth across month pchdzip profile the last zip of the cardholder pdbm profile value of `date became member` at time of last profile update pddapv profile daily mean dollars of approvals pddapv2 profile daily mean dollars of approvals on days with auths pddau profile daily mean dollars of auths on days with auths pddau30 profile daily mean dollars of auths on all days in month pddcsapv profile daily mean dollars of cash approvals pddcsdec profile daily mean dollars of cash declines pdddec profile daily mean dollars of declines pdddec2 profile daily mean dollars of declines on days with auths pddmrapv profile daily mean dollars of merchandise approvals pddmrdec profile daily mean dollars of merchandise declines pdnapv profile daily mean number of approvals pdnau profile daily mean number of auths on days with auths pdnau30 profile daily mean number of auths on all days in month pdncsapv profile daily mean number of cash approvals pdncsdec profile daily mean number of cash declines pdndec profile daily mean number of declines pdnmrapv profile daily mean number of merchandise approvals pdnmrdec profile daily mean number of merchandise declines pdnw1ap2 profile mean number of approvals on Sundays which had auths pdnw1apv profile mean number of approvals on Sundays (day 1 of week) pdnw2ap2 profile mean number of approvals on Mondays which had auths pdnw2apv profile mean number of approvals on Mondays (day 2 of week) pdnw3ap2 profile mean number of approvals on Tuesdays which had auths pdnw3apv profile mean number of approvals on Tuesdays (day 3 of week) pdnw4ap2 profile mean number of approvals on Wednesdays which had auths pdnw4apv profile mean number of approvals on Wednesdays (day 4 of week) pdnw5ap2 profile mean number of approvals on Thursdays which had auths pdnw5apv profile mean number of approvals on Thursdays (day 5 of week) pdnw6ap2 prdfile mean number of approvals on Fridays which had auths pdnw6apv profile mean number of approvals on Fridays (day 6 of week) pdnw7ap2 profile mean number of approvals on Saturdays which had auths pdnw7apv profile mean number of approvals on Saturdays (day 7 of week) pdnwdaft profile daily mean number of weekday afternoon approvals pdnwdapv profile daily mean number of weekday approvals pdnwdeve profile daily mean number of weekday evening approvals pdnwdmor profile daily mean number of weekday morning approvals pdnwdnit profile daily mean number of weekday night approvals pdnweaft profile daily mean number of weekend afternoon approvals pdnweapv profile daily mean number of weekend approvals pdnweeve profile daily mean number of weekend evening approvals pdnwemor profile daily mean number of weekend morning approvals pdnwenit profile daily mean number of weekend night approvals pexpir profile expiry date stored in profile; update if curr date>pexpir phibal profile highest monthly balance phidcapv profile highest dollar amt on a single cash approve in a month phidcdec profile highest dollar amt on a single cash decline in a month phidmapv profile highest dollar amt on a single merch approve in a month phidmdec profile highest dollar amt on a single merch decline in a month phidsapv profile highest dollar amount on a single approve in a month phidsau profile highest dollar amount on a single auth in a month phidsdec profile highest dollar amount on a single decline in a month phidtapv profile highest total dollar amount for an approve in a single day phidtau profile highest total dollar amount for any auth in a single day phidtdec profile highest total dollar amount for a decline in a single day phinapv profile highest number of approves in a single day phinau profile highest number of auths in a single day phindec profile highest number of declines in a single day pm1avbal profile average bal. during 1st 10 days of mo. pm1nauths profile number of auths in the 1st 10 days of mo. pm2avbal profile average bal. during 2nd 10 days of mo. pm2nauths profile number of auths in the 2nd 10 days of mo. pm3avbal profile average bal. during remaining days pm3nauths profile number of auths in the last part of the month. pmovewt profile uses last zip to determine recent residence move; pmovewt =2 for a move within the previous calendar month; pmovew pnaudy profile number of days with auths pnauw1 profile number of Sundays in month with any auths pnauw2 profile number of Mondays in month with any auths pnauw3 profile number of Tuesdays in month with any auths pnauw4 profile number of Wednesdays in month with any auths pnauw5 profile number of Thursdays in month with any auths pnauw6 profile number of Fridays in month with any auths pnauw7 profile number of Saturdays in month with any auths pnauwd profile number of weekday days in month with any auths pnauwe profile number of weekend days in month with any auths pncsaudy profile number of days in month with cash auths pnmraudy profile number of days in month with merchant auths pnweekday profile number of weekday days in the month pnweekend profile number of weekend days in the month pratdcau profile ratio of declines to auths profage profile number of months this account has had a profile (up to 6 mo.) psdaudy profile standard dev. of # days between transactions in a month psddau profile standard dev. of $ per auth in a month ptdapv profile total dollars of approvals in a month ptdau profile total dollars of auths in a month ptdaudy profile total dollars of auths in a day ptdcsapv profile total dollars of eash advance approvals in a month ptdcsdec profile total dollars of cash advance declines in a month ptddec profile total dollars of declines in a month ptdmrapv profile total dollars of merchandise approvals in a month ptdmrdec profile total dollars of merchandise declines in a month ptdsfa01 profile total dollars of transactions in SIC factor group 01 ptdsfa02 profile total dollars of transactions in SIC factor group 02 ptdsfa03 profile total dollars of transactions in SIC factor group 03 ptdsfa04 profile total dollars of transactions in SIC factor group 04 ptdsfa05 profile total dollars of transactions in SIC factor group 05 ptdsfa06 profile total dollars of transactions in SIC factor group 06 ptdsfa07 profile total dollars of transactions in SIC factor group 07 ptdsfa08 profile total dollars of transactions in SIC factor group 08 ptdsfa09 profile total dollars of transactions in SIC factor group 09 ptdsfa10 profile total dollars of transactions in SIC factor group 10 ptdsfa11 profile total dollars of transactions in SIC factor group 11 ptdsra01 profile total dollars of transactions in SIC fraud rate group 01 ptdsra02 profile total dollars of transactions in SIC fraud rate group 02 ptdsra03 proflle total dollars of transactions in SIC fraud rate group 03 ptdsra04 profile total dollars of transactions in SIC fraud rate group 04 ptdsra05 profile total dollars of transactions in SIC fraud rate group 05 ptdsra06 profile total dollars of transactions in SIC fraud rate group 06 ptdsra07 profile total dollars of transactions in SIC fraud rate group 07 ptdsva01 profile total dollars in SIC VISA group 01 ptdsva02 profile total dollars in SIC VISA group 02 ptdsva03 profile total dollars in SIC VISA group 03 ptdsva04 profile total dollars in SIC VISA group 04 ptdsva05 profile total dollars in SIC VISA group 05 ptdsva06 profile total dollars in SIC VISA group 06 ptdsva07 profile total dollars in SIC VISA group 07 ptdsva08 profile total dollars in SIC VISA group 08 ptdsva09 profile total dollars in SIC VISA group 09 ptdsva10 profile total dollars in SIC VISA group 10 ptdsva11 profile total dollars in SIC VISA group 11 ptnapv profile total number of approvals in a month ptnapvdy profile total number of approves a day ptnau profile total number of auths in a month ptnau10d profile number of auths in month <= $10 ptnaudy profile total number of auths in a day ptncsapv profile total number of cash advance approvals in a month ptncsdec profile total number of cash advance declines in a month ptndec profile total number of declines in a month ptndecdy profile total number of declines in a day ptnmrapv profile total nurnher of merchandise approvals in a month ptnmrdec profile total number of merchandise declines in a month ptnsfa01 profile total number of transactions in SIC factor group 01 ptnsfa02 profile total number of transactions in SIC factor group 02 ptnsfa03 profile total number of transactions in SIC factor group 03 ptnsfa04 profile total number of transactions in SIC factor group 04 ptnsfa05 profile total number of transactions in SIC factor group 05 ptnsfa06 profile total number of transactions in SIC factor group 06 ptnsfa07 profile total number of transactions in SIC factor group 07 ptnsfa08 profile total number of transactions in SIC factor group 08 ptnsfa09 profile total number of taansactions in SIC factor group 09 ptnsfa10 profile total number of transactions in SIC factor group 10 ptnsfa11 profile total number of transactions in SIC factor group 11 ptnsra01 profile total number of transactions in SIC fraud rate group 01 ptnsra02 profile total number of transactions in SIC fraud rate group 02 ptnsra03 profile total number of transactions in SIC fraud rate group 03 ptnsra04 profile total number of transactions in SIC fraud rate group 04 ptnsra05 profile total number of taansactions in SIC fraud rate group 05 ptnsra06 profile total number of transactions in SIC fraud rate group 06 ptnsra07 profile total number of transactions in SIC fraud rate group 07 ptnsva01 profile total number in SIC VISA group 01 ptnsva02 profile total number in SIC VISA group 02 ptnsva03 profile total number in SIC VISA group 03 ptnsva04 profile total number in SIC VISA group 04 ptnsva05 profile total number in SIC VISA group 05 ptnsva06 profile total number in SIC VISA group 06 ptnsva07 profile total number in SIC VISA group 07 ptnsva08 profile total number in SIC VISA group 08 ptnsva09 profile total number in SIC VISA group 09 ptnsva10 profile total number in SIC VISA group 10 ptnsva11 profile total number in SIC VISA group 11 ptnw1apv profile total number of approvals on Sundays (day 1 of week) ptnw2apv profile total number of approvals on Mondays (day 2 of week) ptnw3apv profile total number of approvals on Tuesdays (day 3 of week) ptnw4apv profile total number of approvals on Wednesdays (day 4 of week) ptnw5apv profile total number of approvals on Thursdays (day 5 of week) ptnw6apv profile total number of approvals on Fridays (day 6 of week) ptnw7apv profile total number of approvals on Saturdays (day 7 of week) ptnwdaft profile total number of weekday afternoon approvals in a month ptnwdapv profile total number of weekday approvals in a month ptnwdeve profile total number of weekday evening approvals in a month ptnwdmor profile total number of weekday morning approvals in a month ptnwdnit profile total number of weekday night approvals in a month ptnweaft profile total number of weekend afternoon approvals in a month ptnweapv profile total number of weekend approvals in a month ptnweeve profile total number of weekend evening approvals in a month ptnwemor profile total number of weekend morning approvals in a month ptnwenit profile total number of weekend night approvals in a month pvdaybtwn profile variance in number of days between trx's (min of 3 trx) pvraudl profile variance of dollars per auth accoss month MERCHANT FRAUD VARIABLES mtotturn Merchant Total turnover for this specific merchant msicturn Merchant Cumulative SIC code turnover mctrtage Merchant Contract age for specific merchant maagsic Merchant Average contract age for this SIC code mavgnbtc Merchant Average number of transactions in a batch maamttrx Merchant Average amount per transaction (average amount per authorizations) mvaramt Merchant Variance of amount per transaction mavgtbtc Merchant Average time between batches mavgtaut Merchant Average time between authorizations for this merchant mratks Merchant Ratio of keyed versus swiped transactions mnidclac Merchant Number of identical customer accounts mnidcham Merchant Number of identical charge amounts mtrxsrc Merchant What is the source of transaction (ATM, merchant, etc.) mtrxtrsp Merchant How is the transaction transported to the source (terminal, non-terminal, voice authorization) mfloor Merchant Floor limit mchgbks Merchant Charge-backs received mrtrvs Merchant Retrievals received (per SIC, merchant, etc.). The issuer pays for a retrieval. macqrat Merchant Acquirer risk managment rate (in Europe one merchant can have multiple acquires, but they dont have records about how many or who.) mprevrsk Merchant Previous risk management at this merchant? Yes or No mtyprsk Merchant Type of previous risk management (counterfeit, multiple imprint, lost/stolen/not received) msicrat Merchant SIC risk management rate mpctaut Merchant Percent of transactions authorized

Properties of the different approaches

Structural

- good if theory is right
    - needs less data to estimate
- bias if theory is wrong
    - special danger at extremes, or beyond data

Empirical

- unbiased if sufficiently flexible (e.g. ANNs)
- approximations if less flexible (e.g. logreg)
- cliff-edge effect? (models built in benign conditions?)
- low variance if large data sets

Built on years of experience!

Empirical model performance criteria

- likelihood
- AUC
- sum of squared EAD prediction errors?
- . . . .

Mismatch with business objectives?

A dynamic area

- the whole area is undergoing constant development

Examples:

- better theory in structural models

  e.g. replace log normal distributions

- dynamic adaptive empirical models
- include other variables in empirical models
  - econometric data
  - product data
  - vintage data
  - try to allow for changes in policy

But will never be perfect: competitive environment

How do we avoid the model misspecification problem?

Some cautionary comments about statistical risk models

1) population drift
   - and reactive population drift

2) poor choice of performance criterion

3) selectivity bias

4) data quality

5) non-statistical constraints on statistical models

1) Population drift

Characteristics of distributions of customers change:

- technological changes
- changing customer expectation
- changing financial products
- competitive changes
  - debt consolidation companies
- economic changes
  - higher default risks
  - people less willing to take on large loans

Example:

'*All lenders know that new customers are riskier than old ones*.'

David Lawrence, Citicorp

The old ones have undergone a selection process, with the poorer ones being dropped

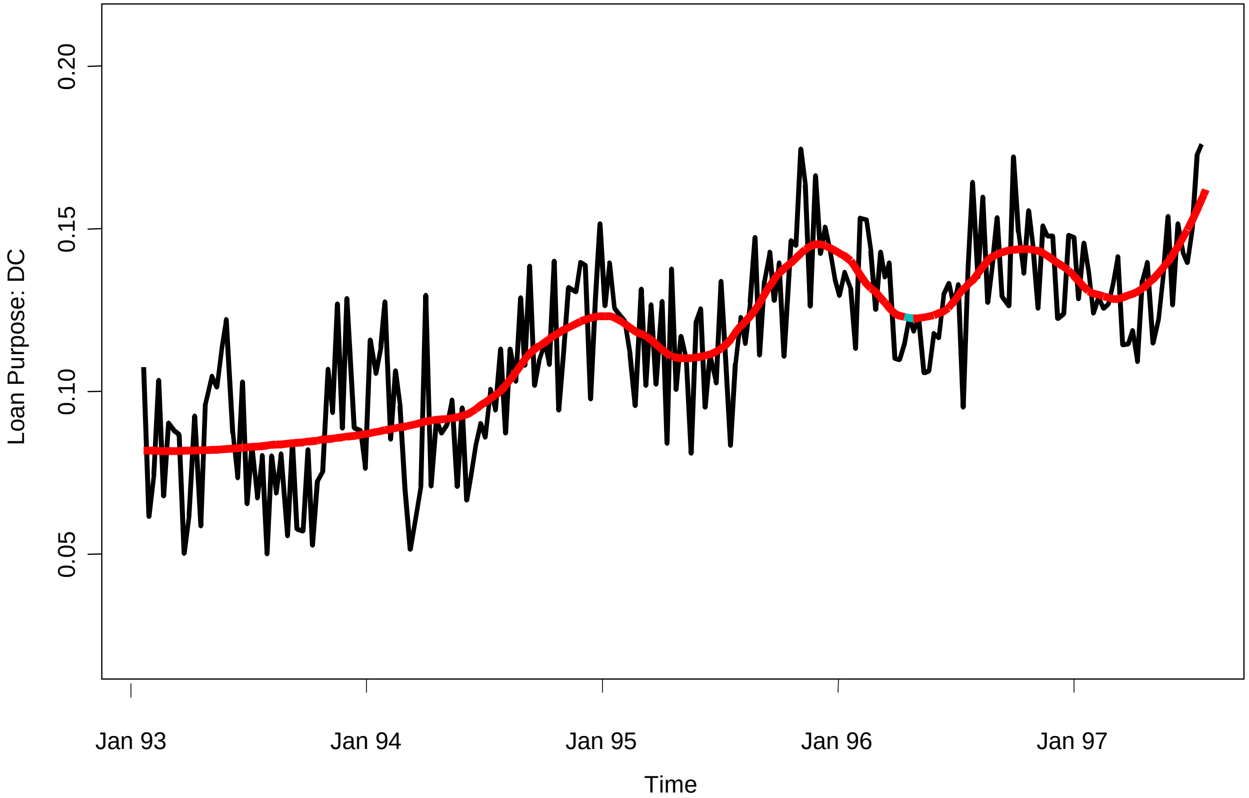
e.g. Citibank had a normal credit card write-off percentage of about 2%. But, when it increased its credit card business by over 500% in two years, the write-off level jumped to 7%

But this depends on the financial product and the market

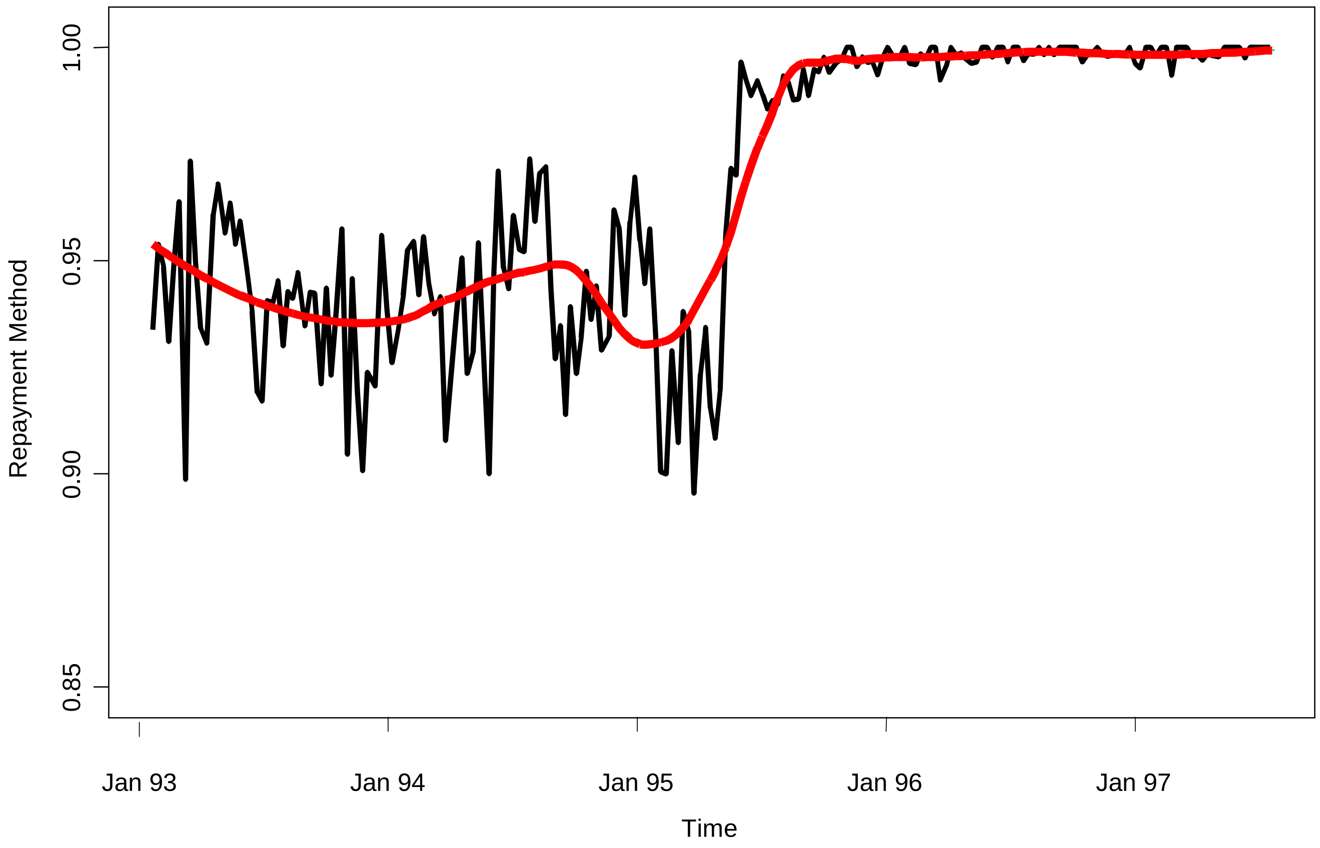
Nowadays, perhaps the ‘old’ customers are simply those who cannot get a better deal elsewhere

This has been observed in the mortgage market

Seasonal changes in types of customer
Trend in types of customer



Marketing initiatives changing type of customer

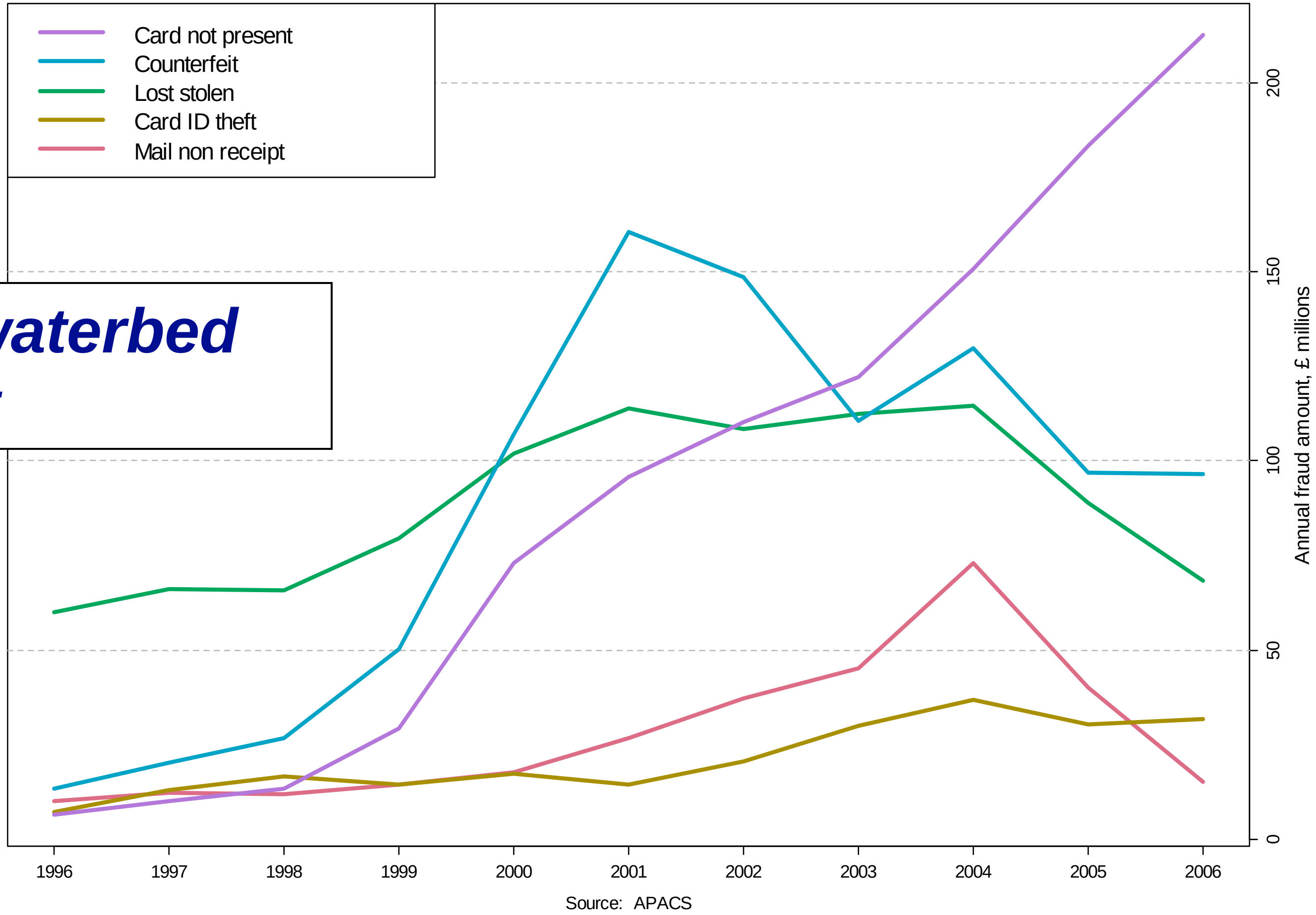


Reactive population drift

In all areas of retail risk modelling

- After all, the aim is not to predict for the fun of it
  - But in order to choose an optimal action

Example: fraud detection



The waterbed effect

Reactive population drift example: *Chip and PIN*

Chip and PIN intended/predicted to end card fraud

After UK rollout on 14 Feb 06, UK CC fraud declined
How much was a consequence of the publicity?

but

- Lloyds TSB observed increase in fraudulent use of UK cards in Europe (no C&P – mag stripe counterfeited)
- observed increase in ATM and cardholder not present fraud

2) Poor choice of performance criterion

Evaluation of models

- statistical metrics, quick and easy, miss the point?
- business metrics, hard to define

Choice of appropriate metrics

e.g. Basel II: PDxPGDxEAD

i) estimate components separately

ii) justified using assumption of independence. Is this valid?

iii) even if it is, is it the right thing to do? The metric ought to be sum of losses (not sum of squared losses, etc).

Different measures suit different objectives

Many objectives, many performance criteria:
- recipe for confusion

Many standard measures:

- $Gini = 2 \times \int B(s)\, g(s)\, ds - 1$
- $KS = \max_{s} \left| B(s) - G(s) \right|$
- $t = (\mu_G - \mu_B)/\sigma$
- $IV = \int \{ g(s) - b(s) \} \log g(s)/b(s)\, ds$

Different situations require different peformance measures

Application scorecards

Initial decision about an applicant

e.g. score 'creditworthiness' using application form information (perhaps supplemented by other, e.g. bureau)

Behavioural scorecards

Modelling customer behaviour over time

e.g. transaction patterns, repayment patterns, etc.

Behavioural scoring as a component of adaptive control systems

Example 1: Timeliness in fraud detection

		True class	
Predicted class		Fraud	Legitimate
	Fraud	a	b
	Legitimate	c	d

A very well known consumer credit organisation evaluates fraud using the two ratios

$$R_1 = a/(a+c)$$

$$R_2 = b/(a+b)$$

In itself, this would appear to be fine

But in fact, the units of assessment used are ***accounts***

An account is flagged as potentially fraudulent if ***at least one transaction is so flagged***

Problem 1: This means that one can make the probability of flagging an account as fraudulent as near to 1 as one wishes by examining enough transactions

Problem 2: Fails to include *timeliness* in the measure

A superior measure

An ***epoch*** is a transaction sequence ending with either

(i) a *fraud flag* on a true fraud

Or

(ii) or end of observed sequence

n n n n f n n f n n **n** n n f n n n n n **n** n n n n n n **f**

<table>
<tr><td></td><td></td><td colspan="2">True class</td></tr>
<tr><td rowspan="3">Predicted class</td><td></td><td>Fraud</td><td>Legitimate</td></tr>
<tr><td>Fraud</td><td>a</td><td>b</td></tr>
<tr><td>Legitimate</td><td>c</td><td>d</td></tr>
</table>

n n n n f n n f n n **n** n n f n n n n n **n** n n n n n n **f**

		True class	
Predicted class		Fraud	Legitimate
	Fraud	1	2
	Legitimate	3	21

This matrix includes ***timeliness*** in the count *c*

		True class	
Predicted class		Fraud	Legitimate
	Fraud	a	b
	Legitimate	c	d

Overall performance measure for given threshold:

$$T_1 = (a+b+kc) \Big/ \big(k(a+c)+b+d\big)$$

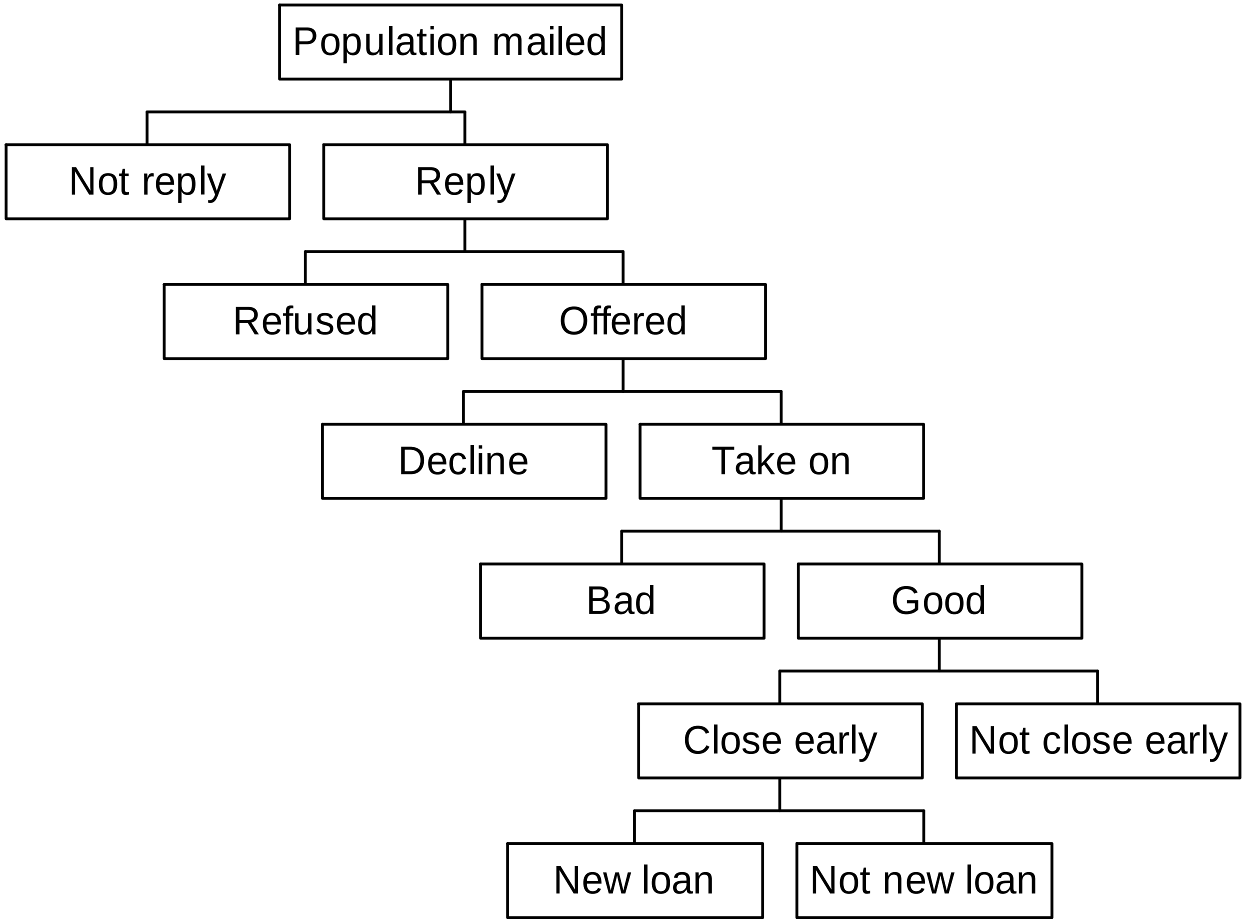
where *k* is the estimated relative cost of misclassifying a fraud as legitimate compared to converse

Or, if the bank can afford to investigate *C* cases

T_2: minimise *c* subject to $a+b=C$

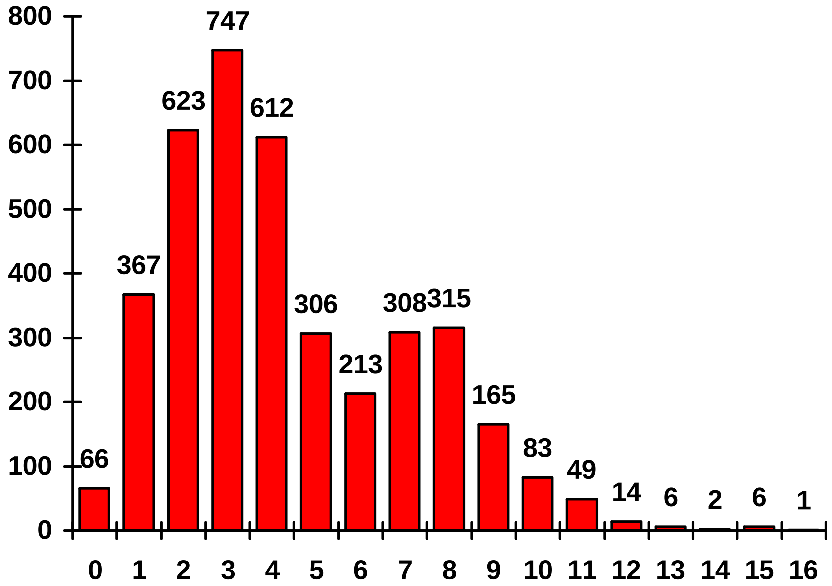
3) Selectivity bias

Sample selectivity problems:



4) Data quality

Missing fields   e.g. *n*=3884, *p*=25



Number of missing values

5 characteristics had no missing values and 2 had over 2000 missing values

In just one data UPL set:

- tiny outstanding amounts unpaid - coded as ‘bad debt’
- negative values in amount owed
- 12 month loans still active after 24 months
- outstanding balance drops to zero, then goes positive
- balances which are always zero
- numbers of months in arrears increasing by more than a single month in one step

One characteristic not being recorded for *any* of the applicants in a particular year (recorded as 'not given')

Leading to a difference of several points on the scorecard values for *all* applicants in that year

5) Non-statistical constraints on statistical models

Excluded characteristics

Aim: avoid discrimination on grounds unrelated to risk?

Approach: exclude gender, religion, age, marital status etc from retail risk models

Is the regulation missing the point?

1) If aim is to obtain most accurate estimates of probability of default

- include all characteristics in models

2) If aim is to obtain most accurate estimates of probability of default which can be obtained from information unrelated to excluded characteristics

- include all in model, and make decisions in subspace orthogonal to excluded characteristics

Predict *y* from

allowed characteristic x_1 and *excluded characteristic* x_2

Take truth to be $y = \alpha + \beta_1 x_1 + \beta_2 x_2$

(interactions can easily be added)

Current legislation:

$$y = \alpha + \gamma_1 x_1$$

Treat x_2 classes ***equally***:

$$y = \alpha + \beta_1 x_1$$

Treat x_2 classes ***fairly*** (best estimate of *y*):

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2$$

$$\gamma_1 = \beta_1 + \beta_2 \delta_{12}$$

Interpretability

Front end

- direct interface with customer
- legal and other requirements for ease of explanation
- feeling of ownership

$\Rightarrow$ - models linear in the customer characteristics

- recursive partitioning models

Back end

- classification performance is all

$\Rightarrow$ highly sophisticated and advanced methods

- neural networks
- SVMs

Special problems in particular areas

e.g. fraud

- the economic imperative
- the Pareto principle
- invisibility of fraud; visibility of detection
- delay in labelling
- mislabelled classes

An example of a possible initial structural alternative to empirical models for retail risk assessment

It's not obvious that structural models for portfolios or for the corporate sector are relevant to retail

But other structural ideas?

Insights from behavioural finance?

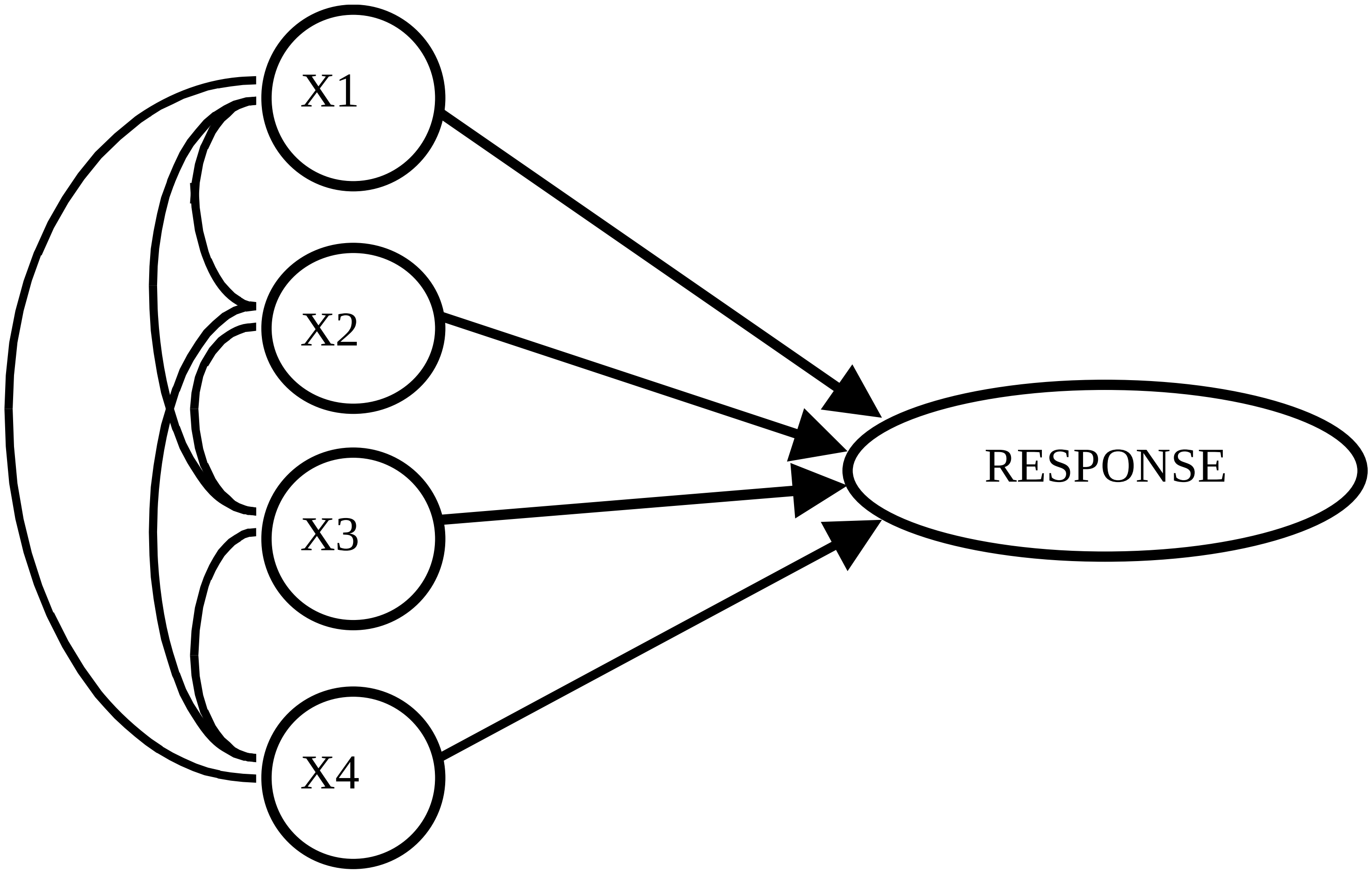
Collaborations with psychologists?

Example: The CQS model

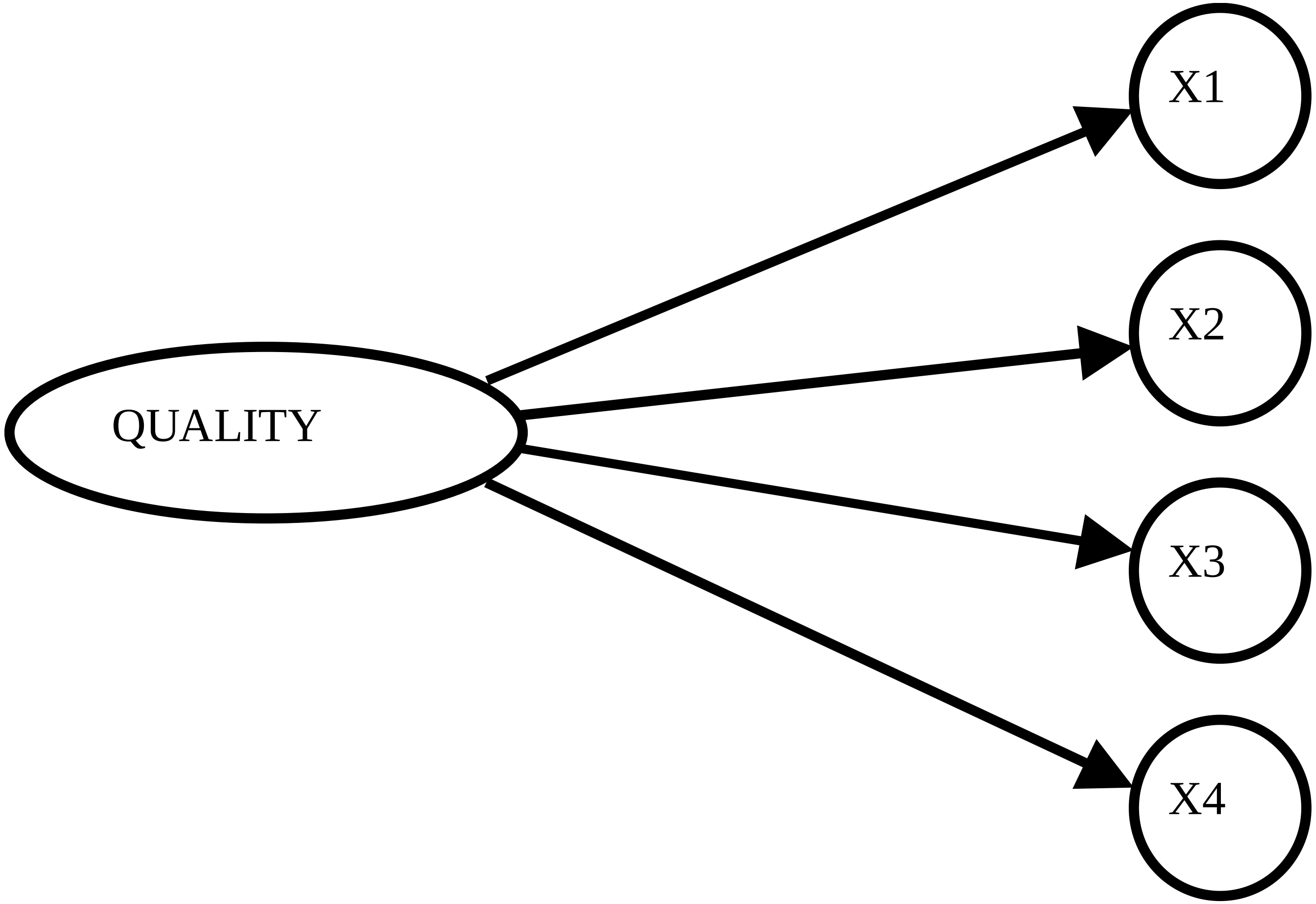
Model based on latent concepts in retail credit

- Character
- Collateral
- Capacity
- Capital
- Condition

Or create others according to expertise or formal model simplification

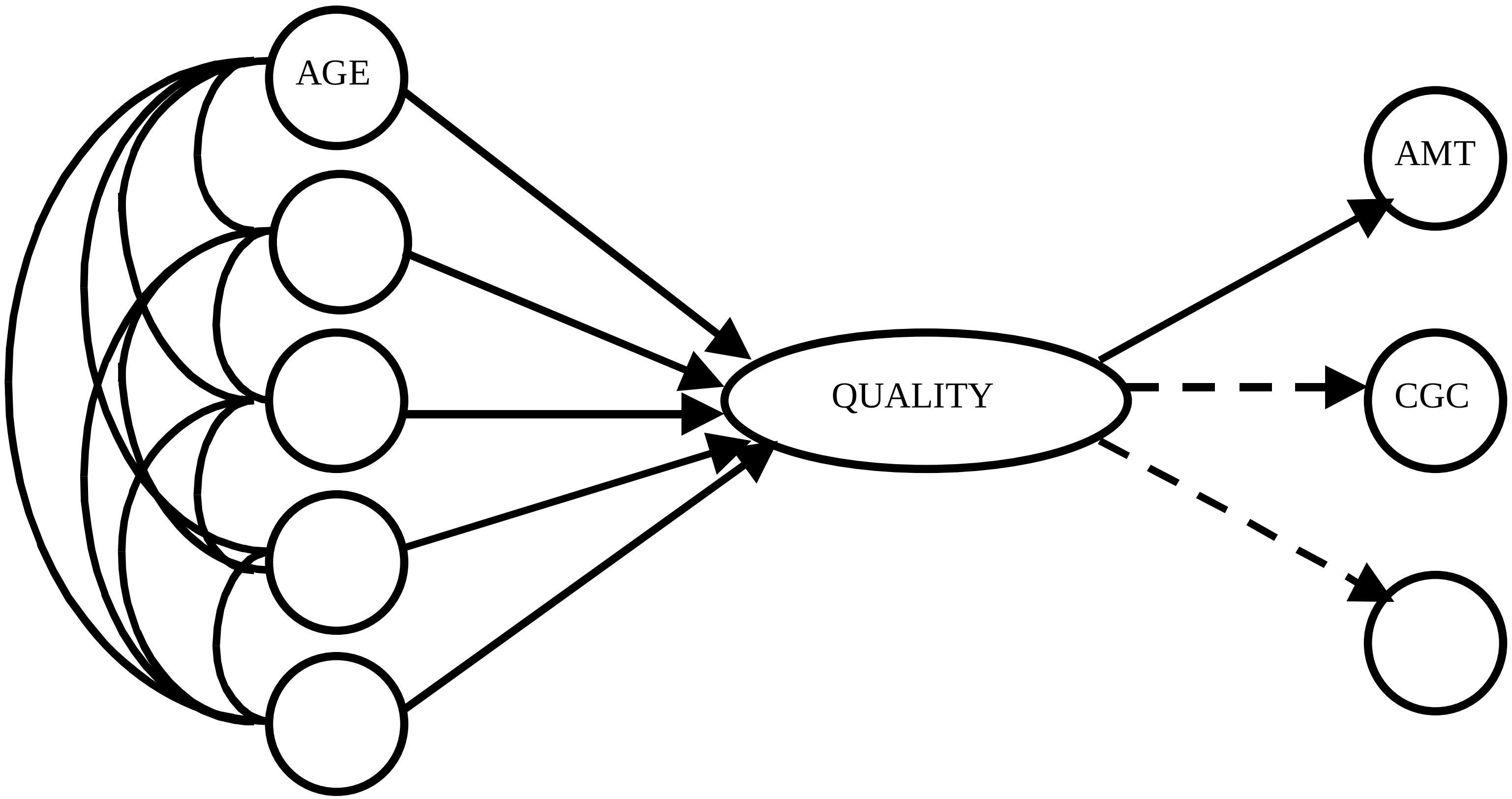


Age, income, other loans,....



CCJs, defaults, months in arrears, parking tickets, speeding fines,....

CQS model



Conclusions

Low level statistical credit risk models are good
But this does not protect against systemic problems

Clever risk evaluation models won't help if embedded in a process of disguising risk and handing it on

If banks shift focus from credit to equity growth

No matter how accurately one can assess the probability that someone or some corporation will default, if circumstances stay as they are, this does not help when:

- a new deal mortgage programme aimed at subprime is introduced
- regulations change making it tougher for Fannie May and Freddie Mac, so that
- banks grasp opportunity, invade Freddie and Fannie territory
- banks shift from credit to equity, growth driven models
- increase in subprime leverage
- Basel II creates an arbitrage opportunity and encourages off balance sheet vehicles
- SEC policy shift encourages greater debt to equity ratio

Future challenges and opportunities

- modern data capture; file merging
- very large data sets
- data streaming, dynamic, adaptive
- adverse selection bias and reject inference
- population drift
- reactive drift

END