Advanced Credit Risk Modeling for Basel II Using SAS®

Course Notes

Advanced Credit Risk Modeling for Basel II Using SAS® Course Notes was developed by Dr. Bart Baesens. Editing and production support was provided by the Curriculum Development and Support Department.

Advanced Credit Risk Modeling for Basel II Using SAS® Course Notes

Book code E1253, course code BASELIIA, prepared date 04Feb2008. BASELIIA_001

Table of Contents

To learn more…

A full curriculum of general and statistical instructor-based training is available at any of the Institute's training facilities. Institute instructors can also provide on-site training.

For information on other courses in the curriculum, contact the SAS Education Division at 1-800-333-7660, or send e-mail to training@sas.com. You can also find this information on the Web at support.sas.com/training/ as well as in the Training Course Catalog.

For a list of other SAS books that relate to the topics covered in this Course Notes, USA customers can contact our SAS Publishing Department at 1-800-727-3228 or send e-mail to sasbook@sas.com. Customers outside the USA, please contact your local SAS office.

Also, see the Publications Catalog on the Web at support.sas.com/pubs for a complete list of books and a convenient order form.

Chapter 1 Advanced Credit Scoring

1.1 Introduction

Lecturer

- Bart Baesens
  - Ph.D. from the Catholic University of Leuven, Department of Applied Economic Sciences (Belgium)
  - Title: Developing Intelligent Systems for Credit Scoring using Machine Learning Techniques
  - Defended: September 24, 2003
  - Assistant professor at the K.U.Leuven, Belgium
  - Assistant professor at Vlerick Leuven Ghent Management School, Belgium
  - Lecturer at the School of Management at the University of Southampton, United Kingdom

3

Software Support

- SAS/STAT
  - PROC MEANS, PROC FREQ, PROC LOGISTIC, …
- SAS/INSIGHT
  - interactive data analysis
- SAS Enterprise Miner
- SAS Credit Scoring Nodes

4

Day Schedule

9:00–10:30	Morning Session 1
10:30–10:45	Morning Break
10:45–12:30	Morning Session 2
12:30–2:00	Lunch
2:00–3:15	Afternoon Session 1
3:15–3:30	Afternoon Break
3:30–5:00	Afternoon Session 2
5:00	End of Day

5

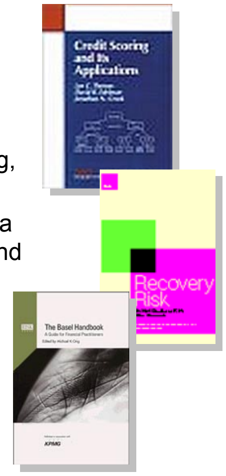
Course Overview

- Review of Basel II
- Review of PD modeling
- LGD modeling
- EAD modeling
- Validation
  - Backtesting
  - Benchmarking
- Low Default Portfolio's
- Stress testing
- New techniques for PD/LGD/EAD modeling
  - Neural networks
  - SVMs
  - Survival analysis

6

Relevant Background: Books

- Credit Scoring and Its Applications, Thomas, Edelman and Crook, Siam Monographs on Mathematical Modeling and Computation, 2002.
- The Basel Handbook, Ong, 2004
- Developing Intelligent Systems for Credit Scoring, Bart Baesens, Ph.D. thesis, K.U.Leuven, 2003.
- Introduction to Modern Credit Scoring: From Data to Basel II Internal Rating System, Van Gestel and Baesens, Oxford University Press, forthcoming
- Recovery Risk: The next challenge in credit risk management, Edward Altman; Andrea Resti; Andrea Sironi (editors), 2005



7

Relevant Background: Journals

- Journal of Credit Risk
- Journal of Banking and Finance
- Risk Magazine
- Journal of the Operational Research Society
- European Journal of Operational Research
- Management Science
- IMA Journal of Mathematics

8

Relevant Background: Web Sites

- www.defaultrisk.com
- http://www.bis.org/
- Websites of regulators:
  - www.fsa.gov.uk (United Kingdom)
  - www.hkma.gov.uk (Hong Kong)
  - www.apra.gov.au (Australia)
  - www.mas.gov.sg (Singapore)

9

Relevant Background: Conferences and Workshops

- Edinburgh conference on Credit Scoring and Credit Control
- Credit Scoring workshops in Belgium/Southampton
- SAS courses
- Data mining conferences
  - KDD, PKDD/ECML, PAKDD

10

1.2 A Review of Basel II

The Basel II Capital Accord

- Basel Committee
  - Central Banks/Bank regulators of major (G10) industrial countries
  - Meet every 3 months at Bank for International Settlements in Basel
- Basel I Capital Accord 1988
  - Aim is to set up minimum regulatory capital requirements in order to ensure that banks are able, at all times, to give back depositor's funds
  - Capital ratio=available capital/risk-weighted assets
  - Capital ratio should be > 8%
  - Capital is shareholder equity and retained earnings (tier 1); other resources (tier 2)
- Need for new Accord
  - Regulatory arbitrage
  - Not sufficient recognition of collateral guarantees
  - Solvency of debtor not taken into account
  - Fixed risk weights per asset class ('standardised approach')

12

Three Pillars of Basel II

Pillar 1: Minimum Capital Requirement

- Credit Risk
  - Standard Approach
  - Internal Ratings Based Approach
    - Foundation
    - Advanced
- Operational Risk
- Market Risk

Pillar 2: Supervisory Review Process

- Sound internal processes to evaluate risk (ICAAP)
- Supervisory monitoring

Pillar 3: Market Discipline and Public Disclosure

- Semi-annual disclosure of:
  - Bank's risk profile
  - Qualitative and Quantitative information
  - Risk management processes
  - Risk management Strategy
- Objective is to inform potential investors

13

Pillar 1: Minimum Capital Requirements

- Credit Risk
- Key components
  - Probability of default (PD) (decimal)
  - Loss given default (LGD) (decimal)
  - Exposure at default (EAD) (currency)
- Expected Loss (EL)=PD x LGD x EAD
- For example, PD=1%, LGD=20%, EAD=1000 Euros➔ EL=2 Euros
- Standardized Approach
- Internal Ratings Based Approach

14

The Standardized Approach

- Risk assessments (for example, AAA, AA, BBB, ...) are based on external credit assessment institution (ECAI)
  (for example, Standard & Poor's, Moody's, ...)
- Eligibility criteria for ECAI given in Accord (objectivity, independence, ...)
- Risk weights given in Accord to compute Risk Weighted Assets (RWA)
- Risk weights for sovereigns, banks, corporates, ...
- Minimum Capital = 0.08 x RWA

15

continued...

The Standardized Approach

Example:

- Corporate/1 mio dollars/maturity 5 years/unsecured/external rating AA
- RW=20%, RWA=0.2 mio, Reg. Cap.=0.016 mio

Credit risk mitigation (collateral)

- guidelines provided in accord (simple versus comprehensive approach)

No LGD, EAD

But:

- Inconsistencies
- Sufficient coverage?
- Need for individual risk profile!

16

Internal Ratings Based Approach

Internal Ratings Based Approach

- Foundation
- Advanced

	PD	LGD	EAD
Foundation approach	Internal estimate	Regulator's estimate	Regulator's estimate
Advanced approach	Internal estimate	Internal estimate	Internal estimate

17

continued...

Internal Ratings Based Approach

- Split exposure into 5 categories:
  - corporate (5 subclasses)
  - sovereign
  - bank
  - retail (residential mortgage, revolving, other retail exposures)
  - equity
- Advanced IRB approach required for retail!
- Credit Scoring=IRB approach to credit risk in retail
- Risk weight functions to derive capital requirements
- Phased rollout of the IRB approach across banking group using implementation plan
- Implementation at the start of 2007 (Europe)

18

Basel II: Retail Specifics

- Retail exposures can be pooled
  - Estimate PD/LGD/EAD per pool!
- Default definition
  - obligor is past due more than 90 days (can be relaxed during transition period)
  - default can be applied at the level of the credit facility
  - PD is the greater of the one-year estimated PD or 0.03%
- Length of underlying historical observation period to estimate loss characteristics must be at least 5 years (can be relaxed during transition period, minimum 2 years at the start)
- No maturity (M) adjustment for retail!

19

1.3 A Review of PD Modeling

PD Modeling: How It Works

Retail

- Develop models that order customers according to risk
- Application scoring: new customers
- Behavioural scoring: existing customers
- Hybrids

Corporate/Sovereign/Bank

- Develop models that perform a mapping to external ratings obtained from external agencies or, for example, experts

21

Application Scoring: Issues and Challenges

- New Customers
- Snapshot to Snapshot
- Two time horizons:
  - 1 year PD (needed for Basel II)
  - Loan term PD (needed for loan approval)
- Data preprocessing issues
  - Outliers/Missing Values
  - Coarse classification (weights-of-evidence coding)
  - Reject Inference
- Modeling issues
  - Logistic regression industry norm!
- Importance of good predictive variables (credit bureaus!)

22

Example Application Scorecard

Characteristic Name	Attribute	Scorecard Points
AGE 1	Up to 26	100
AGE 2	26 - 35	120
AGE 3	35 - 37	185
AGE 4	37+	225
GENDER 1	Male	90
GENDER 2	Female	180
SALARY 1	Up to 500	120
SALARY 2	501-1000	140
SALARY 3	1001-1500	160
SALARY 4	1501-2000	200
SALARY 5	2001+	240

Let cut-off = 500

So, a new customer applies for credit …

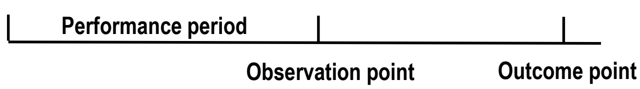
AGE	**32**	**120 points**
GENDER	**Female**	**180 points**
SALARY	**$1,150**	**160 points**
Total		**460 points**

REFUSE CREDIT

23 …

Behavioral Scoring: Issues and Challenges

- Two types of existing customers:
  - Already have loan
  - Already have other products
- Video-clip to snap-shot
- Observation period versus Performance period

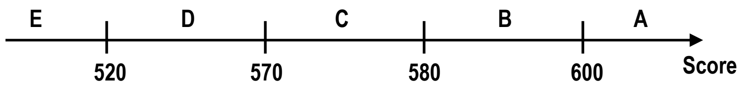


- How to choose observarion point (seasonality)?
- How to define behaviour (role of customer; types of products)?
- Logistic regression
- Input selection!

24

Developing a Rating Based System

- Application and behavioural scoring models provide ranking of customers according to risk
- This was OK in the past (for example, for loan approval), but Basel II requires well-calibrated default probabilities
- Map the scores or probabilities to a number of distinct borrower grades/pools



- Decide on number of classes and their definition
  – Impact on regulatory capital!
- Classes should be sufficiently discriminatory and stable (migration matrix, cf. infra)

25

Developing a Rating System

For corporate, sovereign, and bank exposures

- "*To meet this objective, a bank must have a minimum of seven borrower grades for non-defaulted borrowers and one for those that have defaulted*", paragraph 404 of the Basel II Capital Accord.

For retail

- "*For each pool identified, the bank must be able to provide quantitative measures of loss characteristics (PD, LGD, and EAD) for that pool. The level of differentiation for IRB purposes must ensure that the number of exposures in a given pool is sufficient so as to allow for meaningful quantification and validation of the loss characteristics at the pool level. There must be a meaningful distribution of borrowers and exposures across pools. A single pool must not include an undue concentration of the bank's total retail exposure*", paragraph 409 of the Basel II Capital Accord.

26

Estimating Well-calibrated PDs

- Calculate empirical 1-year realized default rates for borrowers in a specific grade/pool

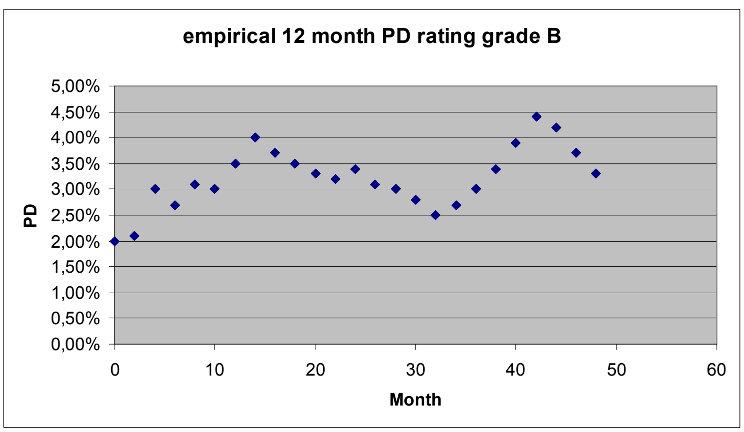
$$\text{1-year PD for rating grade X} = \frac{\text{Number of obligors with rating X at the beginning of the given time period that defaulted during the time period}}{\text{Number of obligors with rating X at the beginning of the given time period}}$$

- Use 5 years of data to estimate future PD (retail)
- Which PD to use?
  - Average
  - Maximum
  - Upper percentile (stressed/conservative PD)
  - Dynamic model (time series)

27

continued...

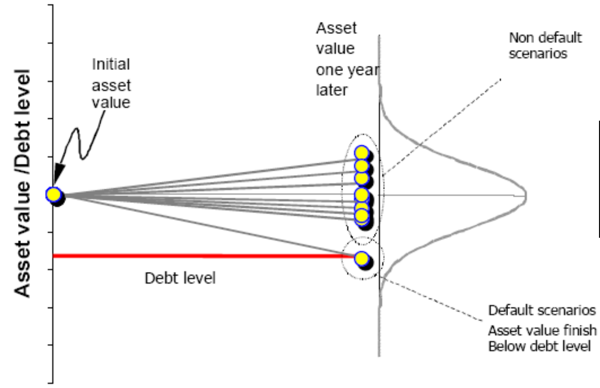
Estimating Well-calibrated PDs



28

The Merton Model (1974)

- Obligor defaults when asset value of the firm falls below the debt level at maturity
- Debt is assumed to be a zero-coupon bond



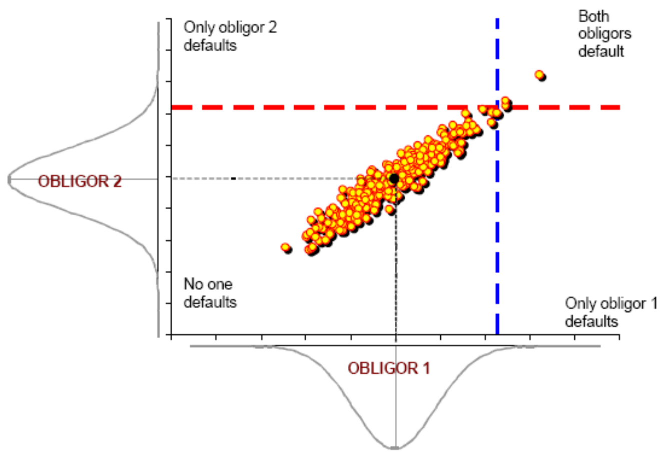
Option Theoretic Model!

29 ***Credit Risk Modelling and Basel II, J.C.G. Céspedes, 2002***

continued...

The Merton Model

The higher the asset correlation, the higher the probability of common default!



30

The Basel II Model

- Assumptions
  - The number of obligors, n, is very high ($n \rightarrow \infty$)
  - The exposures's size is equal to 1/n for all counterparties (goes to 0 for $n \rightarrow \infty$)
  - All obligors have the same probability of default
  - Equal pairwise asset correlations ρ between any two obligors
- Suppose the asset value for all obligors follows a Gaussian process:

$$A_i = \sqrt{\rho} f + \sqrt{1-\rho}\,\varepsilon_i$$

- f is a common factor to all companies
  - for example, economic index
  - one-factor model
- ε_i is an idiosyncratic shock
  - Company specific risk

31

continued...

The Basel II Model

- f, ε_1 ,…, ε_n are mutually independent standard normal variables.
- The asset correlation between all obligors is equal to ρ.
- The probability that the fractional number of defaults X is less than α, P(X≤α) is (Vasicek 1987, 1991):

$$P(X \le \alpha) = N\left[\frac{\sqrt{1-\rho}N^{-1}(\alpha) - N^{-1}(PD)}{\sqrt{\rho}}\right]$$

PD is the probability of default on a single loan of the portfolio.

- Define α* to be the fractional number of defaults that will not be exceeded at the 99.9% confidence level, P(X≤α*)=0.999.

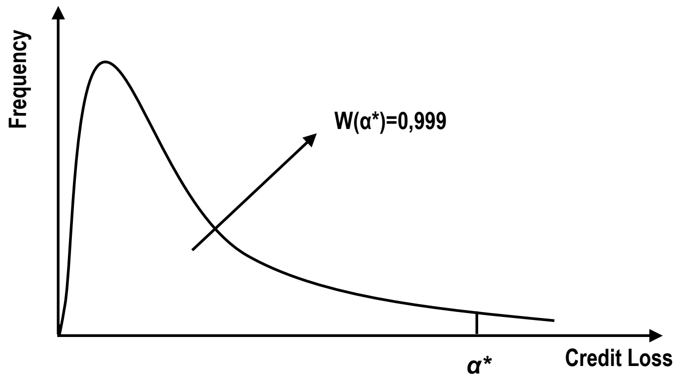
$$\alpha^* = N\left[\frac{N^{-1}(PD) + \sqrt{\rho}N^{-1}(0.999)}{\sqrt{1-\rho}}\right]$$

Value-at-Risk!

32

continued...

The Basel II Model

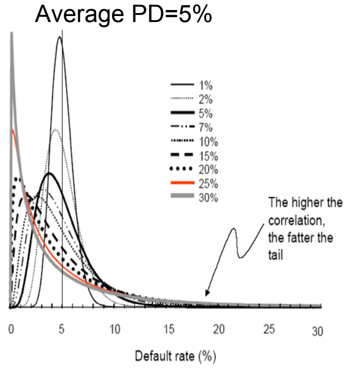


33

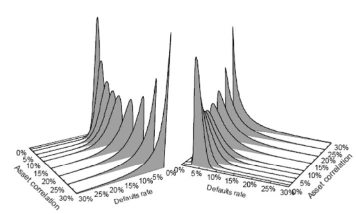
continued...

The Basel II Model

Average PD=5%



Average PD=5%



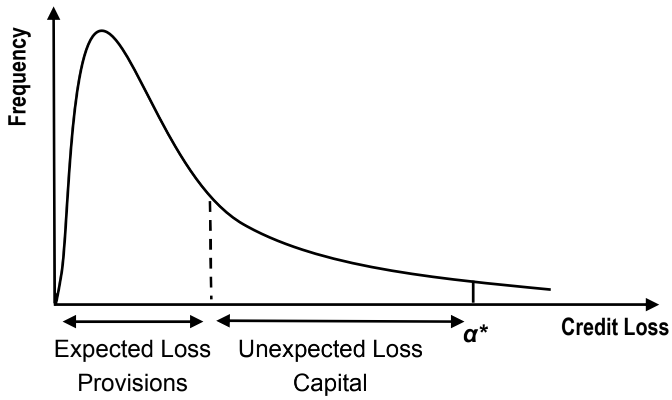
Higher correlation gives fatter tails!

Credit Risk Modelling and Basel II, J.C.G. Céspedes, 2002

34

The Basel II Value at Risk (VAR) Model

Basel sets α* at 99.9%, meaning that there is a 0.1% chance (once in 1000 years) that an institution's capital would fail to absorb the unexpected loss and becomes insolvent!



35

Risk Weight Functions for Retail

$$K = LGD.\left(N\left(\sqrt{\frac{1}{(1-\rho)}}\, N^{-1}(PD) + \sqrt{\left(\frac{\rho}{1-\rho}\right)} N^{-1}(0.999) \right) - PD \right)$$

- N is cumulative standard normal distribution, N^{-1} is inverse cumulative standard normal distribution and ρ is asset correlation
- Residential mortgage exposures $\rho = 0.15$
- Qualifying revolving exposures $\rho = 0.04$
- Other retail exposures

$$\rho = 0.03\left(\frac{1-e^{-35PD}}{1-e^{-35}}\right) + 0.16\left(1 - \frac{1-e^{-35PD}}{1-e^{-35}}\right)$$

36

Risk Weight Functions for Retail

- Only covers unexpected loss (UL)!
  - Expected loss (=LGD.PD) covered by provisions!
- Regulatory Capital = K (PD, LGD). EAD
- Regulatory Capital = 8% . RWA
- RWA=12.50 x Regulatory Capital = 12.50 x K x EAD
- The correlations ρ are chosen depending on the business segment (corporate, sovereign, bank, residential mortgages, revolving, other retail) using some empirical but not published procedure!
  - Basel II correlations have been determined using reverse engineering!

37

LGD/EAD Errors More Expensive Than PD Errors

Consider a credit card portfolio where

- PD = 0.03 ; LGD = 0.5 ; EAD = $10,000

Basel formulae gives capital requirement of K(PD,LGD).EAD

K(0.03,0.50)(10000) = $34.37

10% over estimate on PD means capital required is

K(0.033,0.50)(10,000) = $36.73

10% over estimate on LGD means capital required is

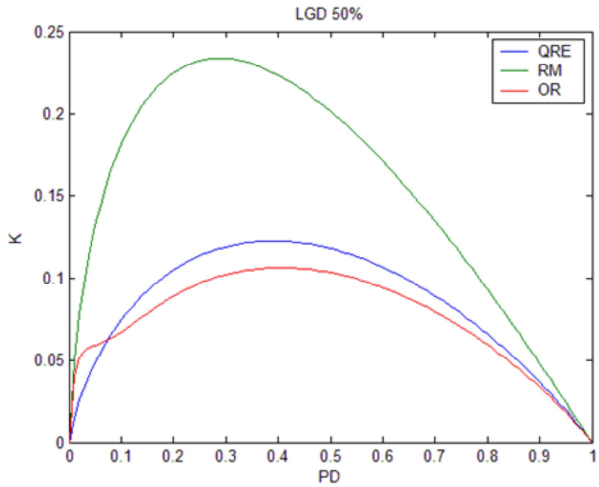
K(0.03,0.55)(10,000) = $37.80

10% over estimate on EAD means capital requirement is

K(0.03,0.50)(11,000) = $37.80

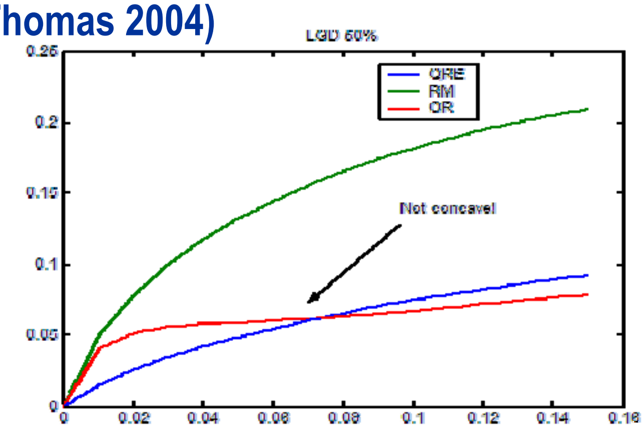
38

The Basel II Risk Weight Functions



39

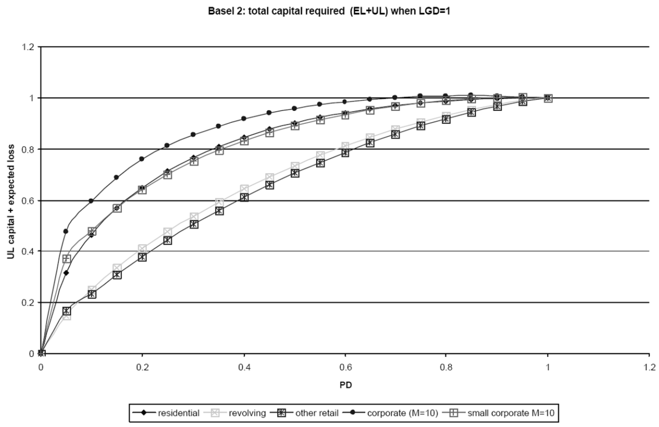
The Basel II Risk Weight Functions (Thomas 2004)



K for revolving and other retail is not concave around PD 0.05:
$K(x.PD_1+1-x)PD_2) \leq x.K(PD_1)+(1-x)K(PD_2)$
Hence, not worth segmenting!

40

The Basel II Risk Weight Functions (Thomas 2004)



Capital set aside for unexpected loss plus the expected loss can exceed the value of the loan!

41

Basel II: Corporates versus Consumer (Thomas 2004)

Corporate Loans

- Well established market
- Market price continuously available
- Bonds only infrequently withdrawn
- Contingent claim model says default is when loan exceeds assets
- Correlation between defaults related to correlation of assets related to correlation of share prices
- Economic conditions built into models

Consumer Loans

- No established market-only occasional securitization sales
- No price available as no public sales
- Consumers often leave lender (attrition)
- Default caused by cash flow (consumer has no idea of assets nor can realize them)
- No share price surrogate for correlation of defaults
- Economic conditions not in models

4

Basel II: Corporates versus Consumer (Thomas 2004)

- No appropriate models for credit risk of portfolios of retail loans (Basel III?)
- Can corporate loans model be re-interpreted to be more acceptable for consumer lending (Wendling Muniz de Andrade, Thomas, 2004)
- Create factor models with factors related to customer default for example. loss of employment, financial naivety, marital or cohabitation status, propensity to illness, loss of income, …)
- One-factor model inappropriate for internationally active banks that want to diversify risk by investing in different countries/industries

43

1.4 LGD Modeling

Overview

- Level 0: data
  - Definition of LGD
  - Ways of measuring LGD
  - Preprocessing data for LGD
- Level 1: predictive model
  - Predictive model for ordinal ranking of LGD
  - Segmentation or Regression
- Level 2: Ratings and calibration
  - Defining LGD ratings and calibration

45

Level 0: Definition of LGD

- LGD is the final loss of an account as a percentage of the exposure, given that the account goes in arrears (=1-recovery rate)
- Economic loss versus Accounting loss
- Problems:
  - Extra Costs
    - Realizing collateral value (haircuts)
    - Administrative costs (letters, telephone calls)
    - Legal costs
    - Time delays in what is recovered (economic loss!)
  - Extra Benefits
    - Interest on arrears
    - Penalties for delays
    - Commissions

46

continued...

Level 0: Ways of Measuring LGD

Market LGD (corporate)

- Use market price of defaulted bonds/loans 1 month after default
- Only for debt securities that trade in the market

Implied market LGD (corporate)

- Use market price of risky but not defaulted bond prices using the asset pricing models (structural/reduced form)
- Spread above risk-free rate reflects EL

47

continued...

Level 0: Ways of Measuring LGD

Workout LGD (corporate/consumer)

- Estimated discounted cash flows from workout/ collections process.
- Direct and indirect costs (for example, operating costs of workout department)
- **Most often used!**

Implied historical LGD

- Use PD estimate and observed total losses to derive implied LGD
- Based on total losses, estimate EL and the LGD=EL/PD
- Only allowed for retail exposure class! (par. 239, CEBS, CP10)

48

continued...

Level 0: LGD According to Basel II

- "The definition of loss used in estimating LGD is economic loss." (par. 460)
- "Owing to the potential of very long-run cycles in house prices which short-term data may not adequately capture, during this **transition period**, LGDs for retail exposures secured by **residential properties cannot be set below 10%** for any sub-segment of exposures …" (par. 266)
- Foundation approach
  - For Corporates/Sovereigns/Banks
    - "Senior claims on corporates, sovereigns and banks not secured by recognised collateral will be assigned a 45% LGD." (par. 287)
    - "All subordinated claims on corporates, sovereigns, and banks will be assigned a 75% LGD." (par. 288)

49

Level 0: Constructing an LGD data set

- See Chapter 4 of Federal Register or paragraph 223, CEBS, CP10
- Data set should cover at least a complete business cycle
- All defaults should be included
- Same default definition as for PD
- Decide on workout period
- Cures and incomplete workouts
- Decide on discount factor
- Censor negative LGD's to zero
- Indirect costs
- Drivers of LGD

50

Level 0: Complete Business Cycle

- For the retail portfolio
  - Minimum 5 years
- For corporate/bank/sovereign
  - Minimum 7 years

51

Level 0: Length of Workout Period

- Non-recovered value is < 5% of EAD (BIS, HKMA)
- One year after default (BIS, HKMA)
- Time of repossession (HKMA)
  - Apply haircut coefficient to the book value of the repossessed asset to convert the associated non-cash recovery into a realized cash recovery
  - Calibrate haircut coefficient based on historical experience (for example, historical volatility of asset value and time required for selling the asset to a third party)
- Time of selling off the debt to a collection agency (for example, United Kingdom)
- 3 years, 5 years?
- Combination

52

Level 0: Cures and Incomplete Workouts

- "*In principle, defaulted exposures that subsequently return to performing status with little or no economic loss to the firm nevertheless form part of the loss history on which LGD estimates should be based.*" (FSA, November 2005)
- FSA suggests cures can be left out provided capital is not underestimated!
- "*The calculation of default-weighted average of realised LGDs, requires the use of all observed defaults in the data sources. Observed defaults include incomplete work-out cases, although they will not have values for the final realisation of LGD because the recovery process has not ended.*" (CEBS, CP10, par. 231, 2005)
- "*Institutions should incorporate the results of incomplete workouts (as data/information) into their LGD estimates, unless they can demonstrate that the incomplete workouts are not relevant.*" (CEBS, CP10, par. 231, 2005)

53

Level 0: Incomplete Workouts

- No prescription as to how incomplete workouts should be treated!
- "The EG seeks recognition from the supervisor that firms should not be required to use incomplete cases in their estimates of LGD." (UK Expert Group on LGD, October 2005)
- If not many incomplete workouts, leaving them out will have no impact anyway
- If many incomplete workouts, use prediction models
  - Survival analysis models whereby recovery amount is considered as censored variable

54

Level 0: Discount Rate

- On-going debate between supervisors and firms
- Contractual interest rate at date of loan origination used a lot
- Subject to discussion (for example, FSA Expert group on LGD)
  - The EG agrees that the use of the contractual rate as the discount rate is conceptually inappropriate.
  - Once a downturn LGD has been estimated, thus accounting for systemic risk, it is assumed that an appropriate level of conservatism has already been built into the estimate.
  - The group proposes that the discount rate for this asset class should be close to the risk free rate, so long as firms can evidence and justify sufficient conservatism in their estimation of the downturn. One potential approach to a discount rate for this asset class could be the risk free rate plus an appropriate premium (for example, 1%)

55

Level 0: Discount Rate

- Discount rate needs to resemble as closely as possible the yield a buyer of distressed assets at the time of default would require to obtain the cash flows projected
- Moody's reported that the long-run average return to defaulted public bonds is 17.4%
- JPMorgan Chase (wholesale portfolio)
  - 18-year old study (1982-1999), 3761 defaulted loans
  - **Discount factor 15%, economic LGD=39.8%**
  - Discount factor 10%, economic LGD=36.2%
  - Discount factor 5%, economic LGD=31.9%
  - Accounting LGD=27%

56

continued...

Level 0: Discount Rate

"*The EG seeks assurance from the supervisor that the UK industry would not be required to adopt overly theoretical approaches which may hinder the evolution of methodologies and their application.*"
(FSA, October 2005)

Industry approaches

- Use contractual rate
- Use risk free rate + a risk premium of 1%
- Use 4%

57

Level 0: Negative LGDs and LGDs > 100%

- Negative LGDs
  - Recovery rate > 100%
  - Reasons
    - EAD measured at moment of default whereas claim on the borrower increases afterwards (fines, fees, …)
    - Result of gain in collateral sales
  - Censor negative LGDs to 0 if not too many!
- LGDs > 100%
  - Recovery rate < 0%
  - Additional costs incurred but nothing recovered
  - Also because of definition of EAD; additional drawings after time of default considered part of LGD
  - Censor to 100% if not too many

58

continued...

Level 0: Indirect Costs

- "*The definition of loss used in estimating LGD is **economic loss**…This must include material discount effects and material direct and indirect costs associated with collecting on the exposure.*" (paragraph 460 of the Accord)
- "Work-out and collection costs should include the costs of running the institution's collection and work-out department, the costs of outsourced services, and an appropriate percentage of other ongoing costs, such as corporate overhead." (par. 205, CEBS, CP10)
- "*Allocated overhead costs, on the other hand, do not vary with defaults. More defaults will simply spread overhead over more loans, lowering the unit 'cost' of such allocations. Allocated overhead costs should therefore not be part of the LGD computation.*" (Federal Reserve, 2005)
- "The EG recommends that costs which would be incurred irrespective of whether there was a work-out should not need to be allocated as part of LGD." (FSA, Expert Group, October 2005)

59

Level 0: Allocating Workout Costs

Calculate a cost rate

Example

year	Total EAD of files in workout (end of year)	Internal workout costs per year	Amount recovered during year
2002	1000	20	250
2003	1500	28	500
2004	800	12	240
2005	1250	27	350

Option 1: use EAD at time of default as denominator

- Assumption: higher costs for higher exposures at default
- Time weighted=1/4*[20/1000+28/1500+12/800+27/1250]=1.8%
- Pooled=[20+28+12+27]/[1000+1500+800+1250]=1.91%
- Use pooled average for more stable cost rate
- Disadvantage: for an individual file, the cost rate has to be multiplied by the number of years the workout lasted

60

Level 0: Allocating Workout Costs

Option 2: Use the amount recovered

- Assumption: higher costs for higher recoveries
- Assumption: higher costs for higher recoveries
- Time weighted=1/4*[20/250+28/500+12/240+27/350]= 6.58%
- Pooled=[20+28+12+27]/[250+500+240+350]=6.49%
- Use pooled average for more stable cost rate
- Advantage is that this is independent of the length of the workout process because each amount was recovered during one year only!
- Much simpler to implement!

61

Level 0: Calculating Individual LGD

Recovery rate=net actualized cash flow (NCF) /EAD

Option 1

- c=direct + indirect cost rate (as a % of amount outstanding, taking into account workout period)

$$NCF = \frac{\sum_{t=1}^{n} CF_t}{(1+i)^t} - c * n * EAD$$

Option 2

- c=direct + indirect cost rate (as a % of recovered amount)
- n=number of years of recovery period

$$NCF = \frac{\sum_{t=1}^{n} CF_t * (1-c)}{(1+i)^t}$$

62

Level 0: Calculating Individual LGD

Option 3

- In case cost information is available for each time period t

$$NCF = \frac{\sum_{t=1}^{n} CF_t - \text{cost}_t}{(1+i)^t}$$

63

Level 0: Drivers for Predictive Modeling of LGD

- Type and value of collateral
  - For example, real estate, cash, inventories, guarantees, …
- Loan to Value (LTV)
  - Ratio of the value of the loan to the fair market value of the underlying asset
  - Estimate and apply haircuts reflecting market value!
  - Initial LTV ratio very important (in the UK)
  - Ratio of collateral value at default/outstanding at default
  - Ratio of collateral value at one year before default/ outstanding at default
- Outstanding balance at default/one year before default
- Authorized balance at default
- Risk rating at default
- Risk rating at one year before default/PD one year before default

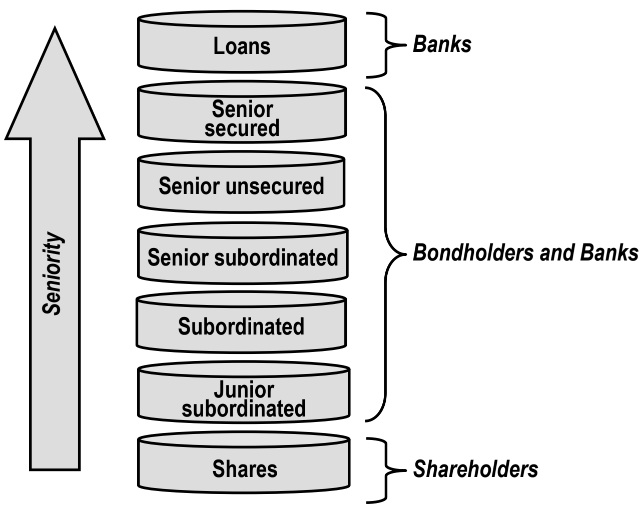
64

Level 0: Drivers for Predictive Modeling of LGD

- Country-related features
  - geographical region (for example, zipcode)
  - For example, how creditor-friendly is the bankruptcy regime?
- Seniority of debt relative to other creditors
- Industry sector (corporate)
- Borrower size
- Type of firm (industry)
- Age of the firm
- Size of loan
- Firm specific characteristics
  - For example, revenue, total assets, net worth, total debt/capital at default, current liability/total liability at one year before default, …
- Economic conditions
  - For example, GDP growth, average default rates, inflation rates, …

65

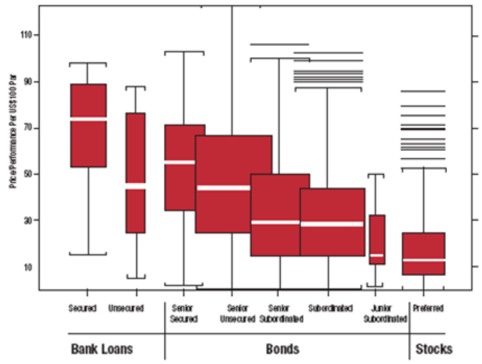
Level 0: LGD Drivers: Seniority



66

Level 0: LGD Drivers: Debt Type and Seniority

Default Recovery by Debt Type and Seniority, 1981-2000



This figure is adapted from Moody's 2001 annual default study; see Exhibit #20 in Hamilton, Gupton & Berthault [2001]. It highlights the wide variability of recoveries even within individual seniority classes. The shaded boxes cover the inter-quartile range with the median marked as a white horizontal line. Squared brackets cover the data range except for outliers that are marked as horizontal lines.

67

Level 0: LGD Drivers: Industry Impact

Industry	Avg. Recovery (cents on dollar)	Industry	Avg. Recovery (cents on dollar)
Utilities	74	High Technology / Office Equipment	47
Insurance & Real Estate	37	Aerospace / Auto / Capital Goods	52
Telecommunications	53	Forest, Building Products / Homebuilders	54
Transportation	39	Consumer / Service	47
Financial Institutions	59	Leisure Time / Media	52
Healthcare / Chemicals	56	Energy & Natural Resources	60

Acharya, Bharath, Srinivasan, 2003

68

Level 0: LGD Drivers: Firm Impact

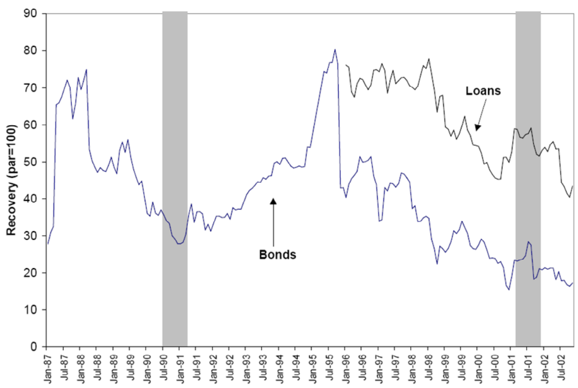
Table 3

LGD by Business Unit
JPMC Resolved Defaults (1Q82-4Q99)

Business Units	Obligor Count	Average Time-to-Resolution (Years)	Net Charge-Offs Mean	Net Charge-Offs Standard Deviation	Discounted LGD Mean	Discounted LGD Standard Deviation
Large Corporates (U.S.)	676	3.33	23.8%	34.2%	41.6%	30.9%
Large Corporates (non-U.S.)	268	2.58	22.9%	33.8%	37.3%	33.2%
Real Estate	719	2.23	29.8%	36.6%	42.0%	33.7%
Emerging Markets	394	3.04	25.8%	39.5%	42.2%	35.6%
Middle Market	1,264	2.15	30.0%	40.4%	40.3%	38.4%
Private Banking	310	1.66	25.4%	40.9%	34.5%	38.3%
Total	3,761	2.43	27.0%	37.9%	39.8%	35.4%

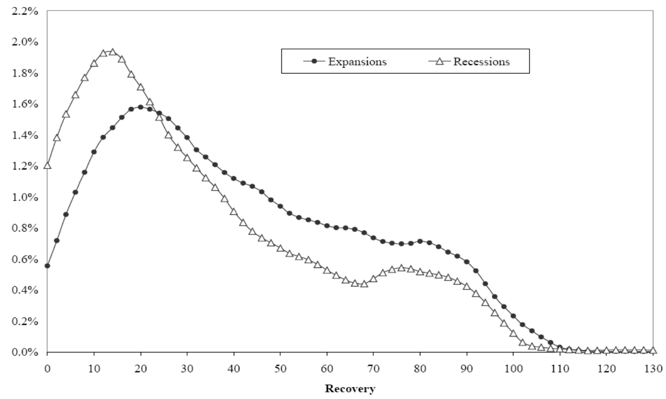
Measuring LGD on Commercial Loans: an 18-year Internal Study, The RMA Journal, May 2004

Level 2: LGD Drivers: Impact of Economy



Altman data set on U.S. firms.
Shaded regions recessions.

Level 0: LGD Drivers: Impact of Economy (1970–2003)



71

Level 0: Data Preprocessing

- Missing values
  - Imputation procedures
- Outlier detection and treatment
  - Truncation procedures
- Coarse classification
  - Decision trees (CART)!

72

Level 1: Challenges in LGD Modeling

- Definition of LGD
- Dependent variable continuous between 0% and 100
- "The CRD does not permit the use of estimates based purely on judgemental considerations."
  (par. 228, CEBS, CP10)
- Use segmentation or regression models!
- Either relatively high or relatively low (bimodal distribution)
- Purpose at level 1 is to provide ordinal ranking of LGD!

73

Level 1: LGD Modeling Approaches Observed in Industry

One-stage

- Segmentation
- Decision trees: expert based or CART
- Linear regression (beta transformed)
- Fractional logit

Two stage:

- Use cumulative logistic regression model with three classes: class 1: 0 LGD, class 2: between 0 and 100% LGD, class 3: 100% LGD. For class 2, estimate linear regression model.
- Use logistic regression for 0 or 100% LGD, then define bins on the score output, each of the bins represents a pool (segment) for which the average historical LGD is calculated.

74

Level 1: Segmentation

- Use historical averages or expert opinions to estimate LGD
- Table look-up
- Long-term versus moving window
- Segmented per
  - Debt type
  - Seniority class (senior versus subordinated)
  - Collateral type (for example, secured versus unsecured)
  - Loan purpose
  - Business segment
- However, one typically observes a wide variability of recovery rates, making averages less suitable to work with.

75

Level 1: Example Segmentation Approach

LGD (percentage of EAD lost in default)			
	Low coverage	Medium coverage	High coverage
Collateral Type 1	25	12	5
Collateral Type 2	5	3	1
Collateral Type 3	16	16	14
Collateral Type 4	90	40	25

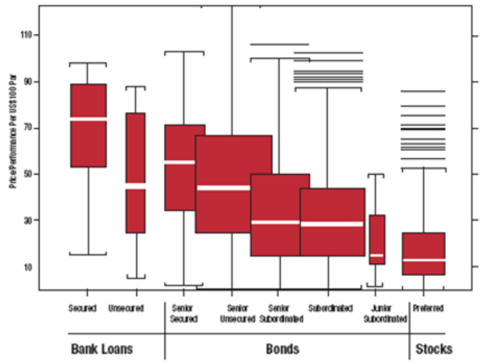
Collateral, for example, real estate, inventories, cash, …
Collateral coverage

Levonian, 2006

76

Level 1: Segmentation Approach to LGD Modeling

Default Recovery by Debt Type and Seniority, 1981-2000



This figure is adapted from Moody's 2001 annual default study; see Exhibit #20 in Hamilton, Gupton & Berthault [2001]. It highlights the wide variability of recoveries even within individual seniority classes. The shaded boxes cover the inter-quartile range with the median marked as a white horizontal line. Squared brackets cover the data range except for outliers that are marked as horizontal lines.

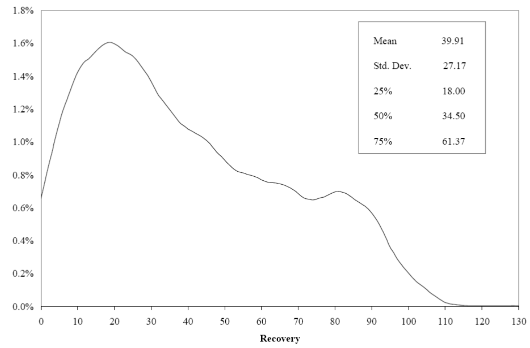
77

Level 1: Regression LGD Modeling

- Commonly used with market LGD data
  - For defaulted bonds and loans which trade in the market
  - Prices reflect the investor's expected recovery
    - For example, Moody's LossCalc observes prices one month after default
- Model recovery rate or LGD as function of loss drivers
- Problem:
  - Recovery rate is not normal distribution.
  - Beta distribution is better fit (Moody's LossCalc assumes this)
  - Regression assumes errors (and hence dependent variable) is normal
- Solution
  - Map Beta distribution onto normal distribution

78

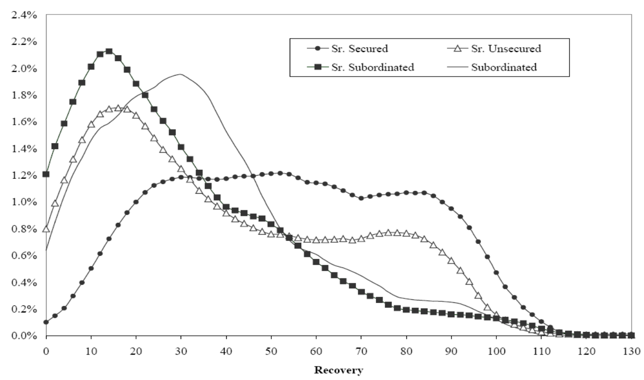
Level 1: Bimodal Distribution of Recovery Rate: All Bonds and Loans (Moody's 1970–2003)



Schuermann 2004

79

Level 1: Probability Density of Recovery by Seniority



This is market LGD and other patterns have been observed for workout LGD!

Schuermann 2004

80

Level1: Linear Regression

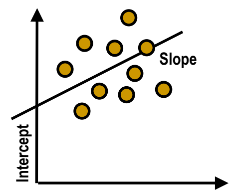
- Linear model

$$Y = w_0 + w_1 x_1 + \ldots w_n x_n = XW$$

- Ordinary Linear Regression (OLS) yields

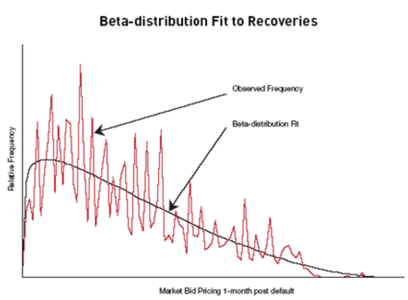
$$W = (X^T X)^{-1} X^T Y$$

- Statistical tests to decide on relevance of variables
- Confidence intervals for LGDs!

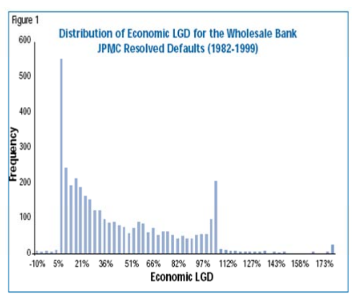


81

Level 1: LGD Regression Modeling



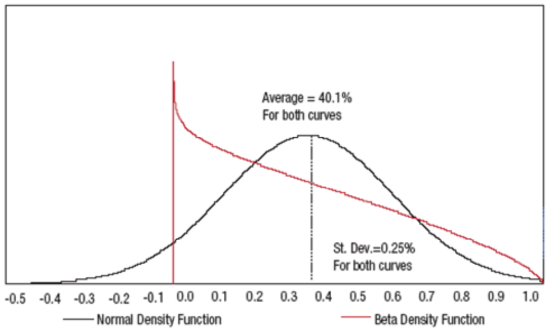
Taken from LossCalc: model for predicting loss given default (LGD), Moody's KMV, February 2002



Measuring LGD on Commercial Loans: an 18-year Internal Study, The RMA Journal, May 2004 (JPMorgan Chase)

82

Level 1: Beta versus Normal Distribution



Taken from LossCalc: model for predicting loss given default (LGD), Moody's KMV, February 2002

83

Level 1:The Beta Distribution, Mean μ, Variance σ2

$$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1},\quad \text{with } \Gamma(\text{x}) \text{ Euler's gamma function}$$

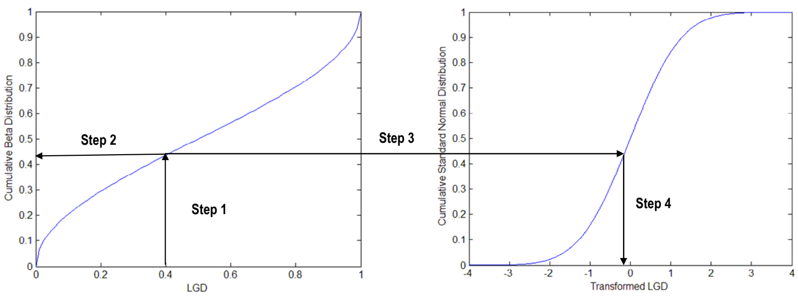
$$\mu=\frac{\alpha}{\alpha+\beta}\quad and\quad \sigma=\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(1+\alpha+\beta)}}$$

- Bounded between 0 and 1! (or any upper bound you like)
- Center parameter α and shape parameter β allow to model a wide variety of distributions
- α and β can be estimated using maximum likelihood or the method of moments

$$\alpha=\left[\mu^2\frac{(1-\mu)}{\sigma^2}\right]-\mu\quad\text{and}\quad\beta=\alpha\left(\frac{1}{\mu}-1\right)$$

84

Level 1: Transforming a Beta Distributed Variable to a Standard Normal Distributed Variable



85

Level 1: Modeling LGD Using Linear Regression

- Step 1: estimate α and β using the method of moments
- Step 2: transform the raw LGDs using the cumulative beta distribution
- Step 3: transform the numbers obtained in step 2 using the inverse standard normal distribution
- Step 4: perform linear regression using the numbers obtained in step 3

Go the other way around to get
LGD predictions based on the regression!

86

Level 1: Two-Stage Models

- Use cumulative logistic regression model with three classes: class 1: 0 LGD, class 2: between 0 and 100% LGD, class 3: 100% LGD. For class 2 estimate linear regression model
- Use logistic regression for 0 or 100% LGD, then define bins on the score output, each of the bins represents a pool (segment) for which the average historical LGD is calculated.

87

Level 1: Summary of Look-up and Regression Approaches

Sophistication Level	Type	Details	Plus	Minus
Low	Contingency or Look-Up Table	A cell might be: *LGD* for Sr. unsecured loans for the automotive industry during a recession	Easy to build and use	Very data intensive to completely fill a possibly very large table
Medium	Basic regression	LGD_{ti} regressed on dummies for Sr/Jr, collateral quality if any (say 3 buckets), industry group (say 6-12), expansion/recession	Relatively easy to build, flexible on data quantity, could easily be converted into a "scorecard"	Grouping/bucketing must be done with care
Medium-high	Advanced regression	As above, but with separate regression models, as warranted, for place in capital structure, collateral quality, expansion/recession; allow for different functional forms (e.g. non-linearity)	Better fit to data	Requires more sophisticated modeling knowledge; somewhat prone to overfitting and datamining
High	Neural nets, tree methods, machine learning	Variety of methods which are often better suited for categorical variables (e.g. place in capital structure, industry) than ordinary regression	Even better fit to data	Even more sophistication; prone overfitting and datamining

88

Level 1: Performance Measures for LGD

y_i – actual LGD; $\hat{y}_i$ – estimated LGD; $\bar{y}$ -average LGD

- R-squared $= \dfrac{\text{Sum of squares regression}}{\text{Sum of squares total}} = \dfrac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

- Mean Squared Error

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}$$

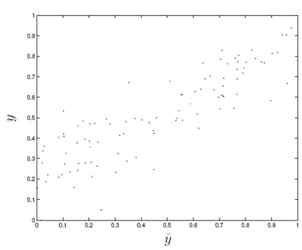
- Mean Absolute Deviation

$$MAD = \frac{\sum_{i=1}^{n}| y_i - \hat{y}_i |}{n}$$

89

Level 1: Performance Measures for LGD

Scatter plot



Correlation

$$corr(\hat{y}, y) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{\hat{y}_i - \bar{\hat{y}}}{s_{\hat{y}}}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$
$$= \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}},$$

90

Level 1: Performance Measures for Estimated LGDs

- CAP plot and accuracy ratio 1
  - binary outcome represents whether observed LGD is higher than the long-term average LGD
  - Indicates how much better model predicts than a long-term average
- CAP plot and accuracy ratio 2
  - Binary outcome represents whether the observed LGD is higher than the long term 75 percentile of the LGD
  - Indicates how much better the model allows to predict high LGDs
  - Illustrates the performance on the right tail of the LGD distribution

91

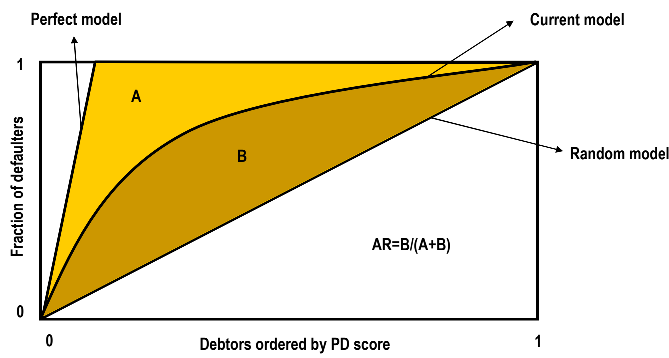
continued...

Level 1: Performance Measures for Estimated LGDs

- CAP plot and accuracy ratio 3
  - Binary outcome represents whether the observed LGD is higher than the long term 25 percentile of the LGD
  - Indicates how much better the model allows to predict low LGDs
  - Illustrates the performance on the left tail of the LGD distribution

92

Level 1: Cumulative Accuracy Profile (CAP) for PD Modeling



Also called Power curve or Lorenz Curve!

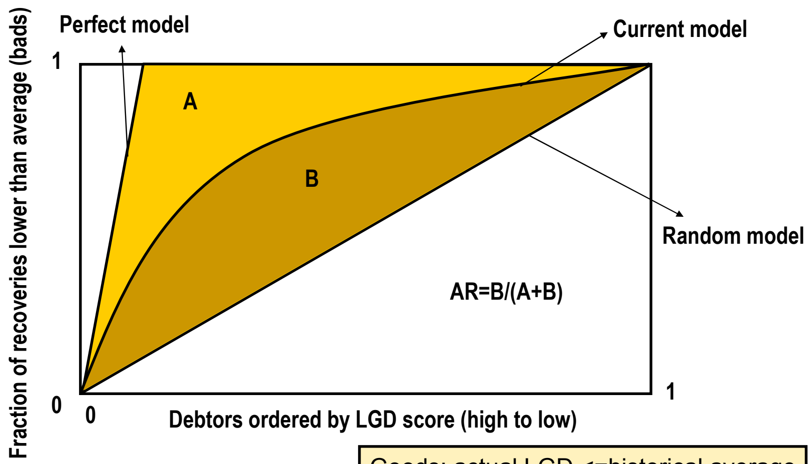
93

Level 1: CAP Profiles

- The accuracy ratio (AR) is defined as:
  (Area below power curve for current model-Area below power curve for random model)/
  (Area below power curve for perfect model-Area below power curve for random model)
- Perfect model has an AR of 1!
- Random model has an AR of 0!
- The AR can be used to compare the performance of several models on the same data but it is not very meaningful as such!

94

Level 1: CAP Curve for LGD Modeling



Goods: actual LGD <=historical average
Bads: actual LGD > historical average

95

Level 1: Efficiency Measures of Estimated LGDs

- Compute confidence intervals for estimated LGDs
- Width of a confidence interval provides information about the precision and efficiency of the estimate
- Narrower confidence intervals are preferred!
  - More certainty regarding required capital to protect against losses
- Reliability can then be measured as the number of times the actual losses fell outside the confidence interval on an independent test set

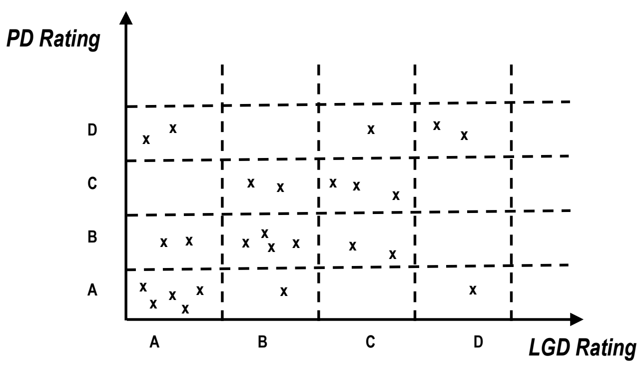
96

Level 2: Mapping to LGD Rating Grades

- Map the output of the regression/segmentation model to LGD facility rating grades
- Compute LGD per grade based on historical data
- Additional standards for corporate, sovereign, and bank exposures
  – "Estimates of LGD must be based on a minimum data observation period that should ideally cover at least one complete economic cycle but must in any case be no shorter than a period of seven years for at least one source." (par. 472)
  – "There is **no specific number of facility grades** for banks using the advanced approach for estimating LGD." (par. 407)
- Additional standards for retail exposures
  – The minimum data observation period for LGD estimates for retail exposures is **five years**. (par 473)

97

Level 2: Mapping to LGD Rating Grades



Check correlations between PD and LGD rating!
Check migration between PD and LGD ratings over time!

98

Level 2: LGD and Rating Agencies

May be useful for calibration or benchmarking!

Moody's		S & P		Fitch	
LGD1	91%-100%	1+	100%	RR1	91%-100%
LGD2	71%-90%	1	100%	RR2	71%-90%
LGD3	51%-70%	2	80%-100%	RR3	51%-70%
LGD4	31%-50%	3	50%-80%	RR4	31%-50%
LGD5	11%-30%	4	25%-50%	RR5	11%-30%
LGD6	0%-10%	5	0-25%	RR6	0%-10%

99

Level 2: Paragraph 468 of the Basel II Accord

Standards for all asset classes (advanced approach)

- "A bank must estimate an LGD for each facility that aims to reflect **economic downturn** conditions where necessary to capture the relevant risks. This LGD **cannot be less** than the **long-run default-weighted average loss rate given default** calculated based on the average economic loss of all observed defaults within the data source for that type of facility… a bank must take into account the potential for the LGD of the facility to be higher than the default-weighted average during a period when credit losses are substantially higher than average…this **cyclical variability** in loss severities may be important and banks will need to incorporate it into their LGD estimates… banks may use averages of loss severities observed during periods of high credit losses, forecasts based on appropriately conservative assumptions, or other similar method…using either internal or external data."

100

Level 2: How to Measure Long-Run LGD

- Historic long-run data
  - Time weighted LGD
    - First calculate LGD for individual years and then average
  - Default weighted LGD
    - Calculated by dividing total losses by the total amount of assets in default
  - Exposure Weighted average
    - Weight each default by EAD
  - Default Count average
    - Each default has equal weighting
- Forward looking LGD
  - Use risk drivers
  - Use historic haircut and adjust for likely future events

101

Level 2: FSA Expert Group's View of LGD Averages

	Default count averaging	Exposure-weighted averaging
Default-weighted averaging	Option 1 Each default has equal weighting, defaults from all years grouped into a single cohort $LGD = \frac{\sum_{y=1}^{m}\sum_{i=1}^{n_y} LR_{i,y}}{\sum_{y=1}^{m} n_y}$	Option 2 Weighting of each default is determined by exposure at default, defaults from all years grouped into a single cohort $LGD = \frac{\sum_{y=1}^{m}\sum_{i=1}^{n_y} EAD_{i,y} \times LR_{i,y}}{\sum_{y=1}^{m}\sum_{i=1}^{n_y} EAD_{i,y}}$
Time-weighted averaging	Option 3 Each default has equal weighting within annual cohort average, average calculated as average of annual averages $LGD = \frac{\sum_{y=1}^{m}\left(\frac{\sum_{i=1}^{n_y} LR_{i,y}}{n_y}\right)}{m}$	Option 4 Weighting of each default within annual cohort average is determined by exposure at default, average calculated as average of annual averages $LGD = \frac{\sum_{y=1}^{m}\left(\frac{\sum_{i=1}^{n_y} EAD_{i,y} \times LR_{i,y}}{\sum_{i=1}^{n_y} EAD_{i,y}}\right)}{m}$

Preferred

Acceptable for Retail

Where,

- i refers to the default observation and y refers to the year/cohort (there are ny defaults in each year y, and a total of m years of observations)
- EAD is the exposure at default
- *LR* is the loss rate (Loss amount / EAD) for each observation

102

Level 2: FSA Expert Group's View of LGD Averages

"A default-weighted average LGD should not be weighted by exposure size. However, historical data that has been collected on this basis may be acceptable for a transitional period, if firms can demonstrate this does not produce significant differences from averages on a 'default count' basis."

(FSA, CP 06/03: Capital Standards 2, February 2006)

103

Level 2: Example of LGD Averages

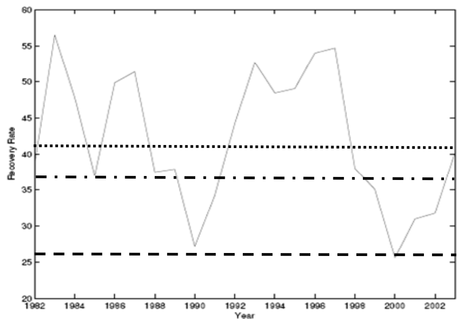
Year 1: 20 defaults of \$40 with average loss of 10%.

Year 2: 50 defaults of \$100 average loss 90% and
30 defaults of \$140 average loss of 60*%.*

Long run LGD	Default count averaging	Exposure weighted averaging
Defaut weighted averaging	$\frac{20.10+50.90+30.60}{20+50+30}=67$	$\frac{20.40.10+50.100.90+30.140.60}{20.40+50.100+30.140}=71$
Time weighted averaging	$\frac{10+\frac{50.90+30.60}{50+30}}{2}=\frac{10+78.75}{2.}=44.4$	$\frac{\frac{50.100.90+30.140.60}{50.100+30.140}+\frac{20.40.10}{20.40}}{2}$ $=\frac{76+10}{2}=43$

104 ...

Level 0: LGD Drivers: Impact of Economy



Time weighted average 42%
Default weighted average 36%
Worst case 26%

: Value Weighted Recovery Rates for Corporate Loans (1982-2003),
Moody's KMV

105

Level 2: Guidance on Paragraph 468 of the Framework Document

- BIS LGD working group
- Problems
  - Potential for realized recovery rates to be lower than average during times of high default rates may be a material source of unexpected credit loss (risk to underestimate capital)
  - Data limitations to estimate LGD (and specify economic downturn conditions)
  - Little consensus in industry on how to incorporate downturn conditions into estimation of LGD
- A principles-based approach is suggested

106

continued...

Level 2: Guidance on Paragraph 468 of the Framework Document

Principle 1: bank must have a rigorous and well documented process for assessing effects, if any, of economic downturn conditions on recovery rates and for producing LGD estimates consistent with downturn conditions

- Identification of downturn conditions (for example, negative GDP growth, elevated unemployment rates, …)
- Identification of adverse dependencies, if any, between default rates and recovery rates
- Incorporation of adverse dependencies, if any, into LGD estimates
  - If adverse dependent: LGD based on averages of loss rates during downturn periods
  - If not adverse dependent: LGD based on long-run default-weighted averages of observed loss rates (neutral conditions)

107

continued...

Level 2: Guidance on Paragraph 468 of the Framework Document

- **Principle 2:** For the estimation of LGDs, measures of recovery rates should reflect the costs of holding defaulted assets over the workout period, including an appropriate risk premium
- For example, discount the stream of recoveries and workout costs by a risk-adjusted discount rate that is the sum of the risk-free rate and a spread appropriate for the risk of the recovery and cost cash flows.
- Still much confusion/debate about paragraph 468!
  - No clear "sound practice"
  - On-going communication between regulators and banks

108

Level 2: Economic Downturn LGD

- Because LGD not based on VAR approximation
- Often not downturn for other business purposes (For example, economic capital)
  - Violation of use-test but allowed by most supervisors provided the reasons are documented!
- Take the average of the 3 worst years of the past 7 years (Germany)!
- Look at downturns in regional sub-portfolios and extrapolate
- If no dependency, long-term average LGD can be used!
- Should it be stress tested (cf. infra)?
  - "When a firm assumes stress and downturn conditions that are similar, the LGD estimates used might also be similar." (FSA, November 2005)

109

Assigning LGD to Non-default Facilities

- Use the mean or median of the empirical distribution of the realized LGD of similar, but defaulted facilities
- For economic downturn conditions:
  - Use average of loss severities observed during periods of high credit losses
  - Use a higher percentile of the distribution appropriate to the degree of adverse dependency
  - In regression, set the predicted LGD to upper limit of confidence interval
- "*Firms that wish to use estimates of LGD that are zero or close to zero should be able to demonstrate that such a figure is appropriate.*" (FSA, November 2005)

110

Exposures in Default

- $K=max(0,LGD-EL_{best})$ whereby LGD is the loss forecast reflecting potential unexpected losses during the recovery period based on downturn conditions, and EL_{best} is the best estimate of expected loss given current economic circumstances and the exposure status
- EL_{best} can be result of regression model estimating expected loss for exposure given current status of collection and economy
- See paragraph 219 of CP10

111

Average LGD Estimates for Non-defaulted Exposures According to QIS5 (June, 2006)

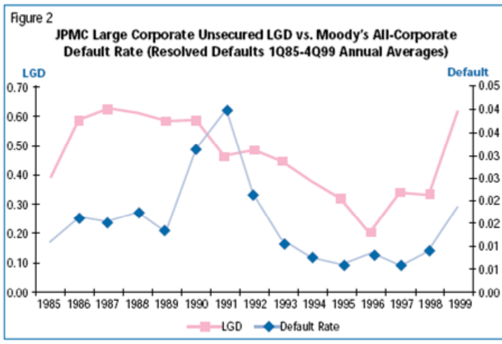
LGD averages for different portfolios in per cent

	IRB Retail				AIRB			
					Wholesale			
	RM	QRE	Other	SME	Corp.	Bank	Sov.	SME Corp.
G10 Group 1 (excl US)					39.8	40.9	33.3	35.0
G10 Group 1 (incl US)	20.3	71.6	48.0	46.2				
G10 Group 2	26.2	57.5	43.0	31.1				
CEBS Group 1	16.1	55.0	47.9	38.8	38.1	37.7	27.7	35.1
CEBS Group 2	21.4	51.9	42.2	31.7	35.2	39.4	38.2	26.7
Other non-G10 Group 1	11.0	67.2	48.3	28.4				
Other non-G10 Group 2	40.4	55.7	45.1	49.6				

This table includes banks which participated in QIS 5, as well as additional data for the US. The figures take account of the 10% LGD floor applicable for exposures in the retail residential mortgage portfolio and include only non-defaulted exposure.

112

LGD versus Business Cycle



Virtually no correlation between LGD of secured exposures and business cycle!

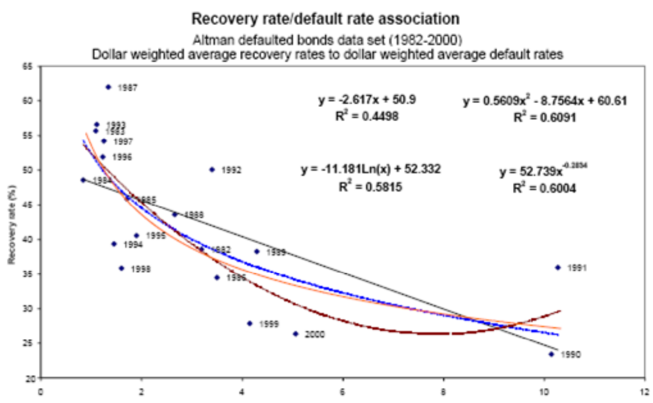
LGD=0.35+ 7.18 x Default Rate, R^2=0,25

LGD=1.16+ 0.16 x ln(Default Rate), R^2=0,40

Measuring LGD on Commercial Loans: an 18-year Internal Study, The RMA Journal, May 2004 (JPMorgan Chase)

113

PD and LGD Are Connected



Altman, Brady, Resti, Sironi 2003

114

Expected Loss versus LGD

- Expected Loss Rate (EL, potential credit loss)=PD*LGD
- Can back out LGD from EL if PD known!
- Combining PD and LGD (Experian, 2002)

PD\LGD	High	Average	Low
High	7.5%	5%	2.5%
Medium	3.75%	2.5%	1.25%
Low	1.875%	1.25%	0.625%

Accept
Reject

Expected Loss Rate!

115

Exposure at Default (EAD) Modeling

- On-balance-sheet (for example, term loan): nominal outstanding balance, net of specific provisions
- E.g. for installment loans
  - EAD is the amount outstanding at the time of capital calculation
- Off-balance-sheet (for example, credit cards, revolving credit): committed but unused loan amount times a credit conversion factor (CCF)
- CCF is needed to take into account additional drawings prior to default
- Also called Loan Equivalency Factor (LEQ).

116

EAD Modeling for Revolving Credit

- EAD=DRAWN+ α × (LIMIT-DRAWN)
- α is the proportion of the committed amount likely to be drawn prior to default also called credit conversion factor (CCF) or loan equivalency factor (LEQ)
- 0≤ α ≤1
- Conservative estimates set to 1
- "*The EG proposes that it should be up to firms to determine and justify their approach for estimation of CCFs and for supervisors to review and agree these approaches.*" (FSA, June 2006)

117

EAD According to Basel II

For corporates/sovereigns/banks:

- For off-balance sheet items, exposure is calculated as the committed but undrawn amount multiplied by a CCF. There are two approaches for the estimation of CCFs: a foundation and an advanced approach." (par. 310).
- Foundation approach:
  - "Commitment with an original maturity up to one year and commitments with an original maturity over one year will receive a CCF of 20% and 50%, respectively. However, any commitments that are unconditionally cancelable at any time by the bank without prior notice, …., will receive a 0% CCF." (par. 83)

118

continued...

EAD According to Basel II

For retail:

- "Both on and off-balance sheet retail exposures are measured gross of specific provisions or partial write-offs. The EAD on drawn amounts should not be less than the sum of (i) the amount by which a bank's regulatory capital would be reduced if the exposure were written-off fully, and (ii) any specific provisions and partial write-offs." (par. 334)
- "For retail off-balance sheet items, banks must use their own estimates of CCFs." (par. 335)

119

continued...

EAD According to Basel II

- "For retail exposures with uncertain future drawdown such as credit cards, banks must take into account their history and/or expectation of additional drawings prior to default in their overall calibration of loss estimates. In particular, where a bank does not reflect conversion factors for undrawn lines in its EAD estimates, it must reflect in its LGD estimates the likelihood of additional drawings prior to default. Conversely, if the bank does not incorporate the possibility of additional drawings in its LGD estimates, it must do so in its EAD estimates." (par. 336)
- "Advanced approach banks must assign an estimate of EAD for each facility." (par. 475)
- Must use margin of conservatism and economic downturn EAD if EAD volatile over economic cycle.
- The minimum data observation period for EAD estimates for retail exposures is five years (seven years for corporate exposures). (par. 478 and 479)

120

Definition of EAD

- Definition 1
  - total exposure at risk on the very moment of default.
  - additional drawings are considered as a cost and enter in the LGD (may cause LGD to be > 100%).
  - EAD fixed at time of default, LGD dependent on length of recovery process
- Definition 2
  - Maximum observed EAD during the default process
  - this definition also takes into account drawings after default (for example, given by bank with the perspective of a cure or reduced loss)
  - LGD almost sure < 100% (except with additional unrecovered costs)
  - Both EAD and LGD dependent on length of recovery process
  - Problem is which time point to choose for discounting
- Include accrued interests and unpaid fees either in EAD or LGD (50/50 according to RMA, 2004 study)

121

EAD Modeling

Need two time points: one to measure EAD, and one to measure DRAWN + risk drivers

Cohort method

- Group defaulted facilities into discrete calendar periods (for example, 12 months unless other time period more conservative and appropriate) according to date of default
- Collect information about risk factors and drawn/undrawn amount at beginning of calendar period and drawn amount at date of default
- Pool data of different calendar periods for estimation
- For example, calendar period is defined as 1 November 2003 to 30 October 2004, then information about risk factors and drawn/undrawn amount on 1 November 2003 should be collected and drawn amounts of facilities upon default.

122

continued...

EAD Modeling

Fixed-horizon method

- Collect information about risk factors and drawn/ undrawn amount for a fixed time interval prior to the date of default (at least 12 months unless other time period more conservative and appropriate) and the drawn amount on date of default, regardless of the actual calendar date on which the default occurred.
- For example, fixed interval is 12 months, if a default occurred on 15 July 2004, then information about risk factors and drawn/undrawn amount of the defaulted facilities on 15 July 2003 is used.

123

continued...

EAD Modeling

Variable time horizon approach

- Variant of fixed time horizon approach using several reference times within the chosen time horizon
- For example, compare drawn amount at time of default with risk factors and drawn/undrawn amounts one month, two months, three months before default

Momentum method

- Express CCF as a percentage of the total limit at time of default
- Drawn amount at time of default is compared to total limit at time of default
- Not allowed because currently drawn amount is not considered (CEBS, CP10, 2006)

No universal guidance yet

What if credit limit changes?

124

Limits for α

- α is not allowed to be negative
  - α can be negative when borrower has paid back portion of the amount prior to default
  - Censor negative α's to zero
- α > 1
  - Reason: off-line transactions
  - Problem: as drawn balance increases, exposure decreases
  - Suppose, limit=2500, CCF=110%
  - Drawn=1000 Euros, ➔ EAD=1000+1.10*1500=2650 Euro
  - Drawn=1500 Euros ➔ EAD=1500+1.10*1000=2600 Euro
  - Solution: convert soft credit limit into hard credit limit based on historical data such that α ≤ 1 (use e.g. 99% confidence level if needed)

125

Drivers for Predictive Modeling of EAD (CCF)

Type of obligor

- Corporates and banks: credit lines often not completely utilized at time of default
- Retail and SMEs: more likely to overdraw (or fully utilize) credit line

A borrower's access to other sources of funding

- Retail and SME fewer access to alternative sources than large corporate obligors and banks (use type of obligor as proxy)

Factors affecting the borrower's demand for funding/facilities

Expected growth in a borrower's business (and accordingly rise in funding requirements)

The nature of the particular facility (for example, industry, geographical region, facility size, convenant protection, credit risk (PD?), …)

126

Developing EAD Models

- Level 0/Level 1/Level 2 (see LGD modeling)
- Construct development data set storing information (risk factors) of the defaulted facilities
- Use same definition of default as for PD and LGD
- Calculate CCF for each defaulted facility
- Develop segmentation or regression model
  - U-shaped distribution observed before; use beta-distribution and same trick as with LGD
- Use model to estimate EAD for non-defaulted facilities
- Expert judgment can be used to fine-tune estimates
  - Should be transparent, well-documented and closely monitored
- Compare estimated CCF with long-run defaulted weighted average average CCF
- For economic downturn CCF:
  - Use average of CCF during periods of high credit losses
  - Use higher percentile of CCF distribution

127

Measuring Performance for EAD Models

Compute correlation between predicted α's and observed α's

CAP plot and accuracy ratio 1

- binary outcome represents whether observed α is higher than the long-term average α
- Indicates how much better model predicts than a long-term average

CAP plot and accuracy ratio 2

- Binary outcome represents whether the observed α is higher than the long term 75 percentile of the α
- Indicates how much better the model allows to predict high α's
- Illustrates the performance on the right tail of the α distribution

128

continued...

Measuring Performance for EAD Models

CAP plot and accuracy ratio 3

- Binary outcome represents whether the observed α is higher than the long term 25 percentile of the α
- Indicates how much better the model allows to predict low α's
- Illustrates the performance on the left tail of the α distribution

Width of confidence intervals for estimated α's (efficiency)

Confidence intervals (efficiency and reliability)

129

EAD Models for Guarantees

- Most respondents measure the exposure of guarantees as the **face value of the guarantee**, or less if there is a permanent diminution of the underlying obligation (CEBS)
- Where firms are extending contingent exposures such as guarantee facilities, EAD could potentially be considered as the **amount of the facility**, **the amount expected to be outstanding under it at the time of default, or the amount expected to be claimed**. We proposed that it would be acceptable to calculate EAD on any of these bases. All respondents supported our approach (FSA)
- If the institution had made a direct loan instead of providing a guarantee, it would not have been required to apply a Conversion Factor. In this situation the institution can assume that the risk is the same as a direct loan and the **appropriate CF is 100 percent**. Therefore, for credit substitutes, supervisors should not require institutions to calculate CF estimates based on seven years of historical data, provided that the credit Conversion Factor used is 100 percent. (CEBS, CP10)

130

Correlation between PD/LGD/EAD

- Correlation of PDs across borrowers
  - Related to asset correlation parameter in Basel II Accord
  - Determined using some empirical but not published procedure
- Positive correlation between PD and LGD
  - PDs usually higher during economic downturns when also asset values get depressed, hence higher than average LGDs
  - As correlation increases, so will level of credit risk
  - Treated as independent in Basel II
- Correlation between PD and EAD
  - For example, revolving credits, if financial distress worsens, a borrower will draw down as much as possible on existing unutilized facilities in order to avoid default
  - When EAD depends on market prices (for example, traded bonds)
- Impact on, for example, stress testing

131

Relations between Basel Parameters

	Some consensus	Open questions
Probability of default (PD)	+ correlation with asset values. Time-varying with systematic risk component.	Relationship between PD correlations and firm credit quality (PD level). Relationship between PD correlations and bankruptcy rules (renegotiation) and macroeconomic shifts.
Loss given default (LGD)	+ correlation with asset and collateral values. Time-varying with systematic risk component.	Relationship between LGD correlations and firm credit quality (PD level). Relationship between LGD correlations and bankruptcy rules (renegotiation) and macroeconomic shifts.
Correlations between LGD and PD		Sign of correlation between LGD and PD. Relationship between PD-LGD correlation and systematic macro effects.
Exposure at default (EAD)	Time-varying with systematic risk component.	Sign of interfirm correlation in EAD. Relationship between EAD correlations and firm credit quality (PD level). Relationship between EAD correlations and bankruptcy rules (renegotiation) and macroeconomic shifts. Integration of market risk and credit risk models.

BIS Working Paper 126 2003

132

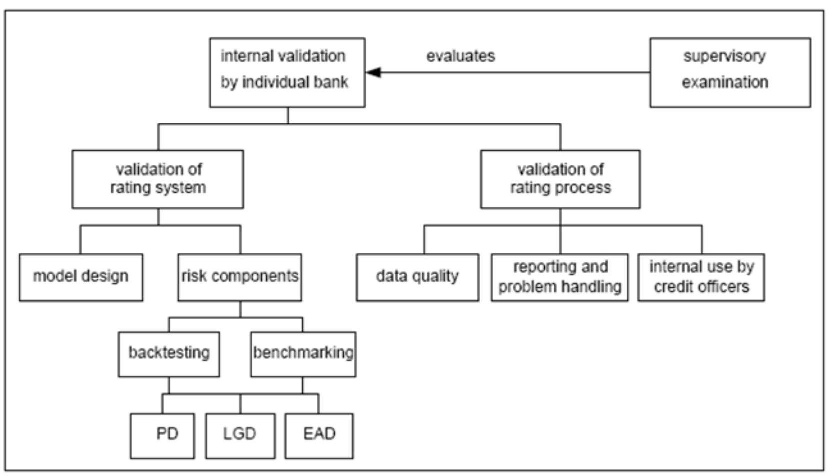
1.5 Validation of Basel II Models

Validation According to Basel II

- "The bank must have a **regular cycle of model validation** that includes monitoring of model performance and stability; review of model relationships; and testing of model outputs against outcomes." (par. 417)
- "Banks must have a robust system in place to validate the **accuracy** and consistency of **rating systems**, **processes** and the estimation of all relevant risk components." (par. 500)
- "Banks must regularly **compare realized default rates** with **estimated PDs** for each grade and be able to demonstrate that the realized default rates are within the expected range for that grade…(similar for LGD and EAD) …This analysis and documentation must be updated at least **annually**." (par. 501)
- "Banks must also use other **quantitative validation tools** and comparisons with relevant **external data sources.**" (par. 502)
- "Banks must have **well-articulated internal standards** for situations where **deviations** in realised PDs, LGDs and EADs from expectations become significant enough to call the validity of the estimates into question." (par. 504)

134

Validation According to Basel II



BIS 14, Working paper

135

Validation Terminology: Backtesting versus Benchmarking

- Backtesting
  - Using statistical (quantitative) methods to compare estimates of PD, LGD, EAD to realized outcomes

Rating Category	Estimated PD	Nr. of observations	Nr. of observed defaults
A	2%	1000	17
B	3%	500	20
C	7%	400	35
D	20%	200	50

- Benchmarking
  - Comparing internal estimates across banks and/or with external benchmarks
- Validation > Backtesting + Benchmarking!

136

Quantitative versus Qualitative Validation

- Quantitative validation
  - Backtesting
  - Benchmarking
  - Data + Statistics!
- Qualitative validation
  - Data quality
  - Use test
  - Model design
  - Documentation
  - Corporate governance and management oversight

137

Global Internal Ratings Validation Survey (2003)

- Joint research sponsored by the International Swaps and Derivatives Association (ISDA), Risk Management Association (RMA), British Banker's Association (BBA)
- Conducted by PricewaterhouseCoopers
- Banks participating in survey represent "cutting edge"
  - Results of survey may overstate actual level of sophistication
- Validation methodologies for internal ratings and PD

138

Global Internal Ratings Validation Survey: Key Findings

- Banks employ a **wide range** of techniques to validate internal ratings. The techniques used to assess corporate and retail ratings are substantially different.
- Ratings validation **is not an exact science**. Absolute performance measures are considered counterproductive by some institutions.
- **Expert judgment** is critical. Data scarcity makes it almost impossible to develop statistically based internal-ratings models in some asset classes.

139

continued...

Global Internal Ratings Validation Survey: Key Findings

- Data issues center around **quantity**, not quality. Default data, in particular, is insufficient to produce robust statistical estimates for some asset classes.
- Definite **regional differences** exist in the structure of ratings systems and validation techniques.
- **Stress-testing standards** need to be defined and the issues surrounding their development need to be debated by the industry. The level of additional stress testing that will be required under Basel II is uncertain.

140

General Validation Principles According to the Basel Committee Validation Subgroup

Principle 1: Validation is fundamentally about assessing the predictive ability of a bank's risk estimates and the use of ratings in credit processes

Principle 2: The bank has the primary responsibility for validation

Principle 3: Validation is an iterative process

Principle 4: There is no single validation method

Principle 5: Validation should encompass both quantitative and qualitative elements

Principle 6: Validation processes and outcomes should be subject to independent review

141

Additional Validation Principles

- Supervisor reviews the bank's validation processes.
- Make validation constructive, not threatening.
- Independent staff included in the validation process.
- Validation does not provide a fixed decision but rather a suggestion for further action and study.
- Develop validation frameworks + action schemes.
- Validation methods not allowed to change with economic cycle unless clearly and thoroughly documented. (par. 503, Basel II Accord)

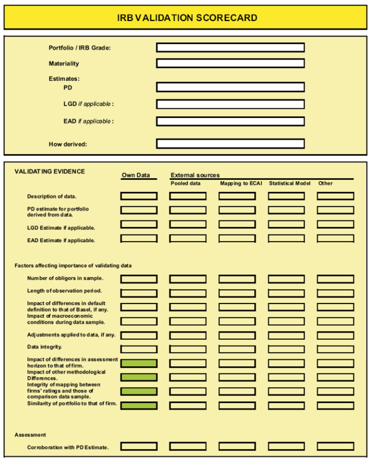
142

Developing a Validation Framework

- Diagnose validation needs
- Work out validation activities
- Timetable for validation activities
- Tests and analyses to be performed
- Actions to be taken in response to findings
- Why/What/Who/How/When
- Should be described in a validation policy!

143

Validation Scorecard Presented in FSA CP189



IRB VALIDATION SCORECARD

Portfolio / IRB Grade:

Materiality

Estimates:

PD

LGD *if applicable* :

EAD *if applicable* :

How derived:

VALIDATING EVIDENCE	Own Data	External sources: Pooled data	Mapping to ECAI	Statistical Model	Other
Description of data.					
PD estimate for portfolio derived from data.					
LGD Estimate if applicable.					
EAD Estimate if applicable.					
Factors affecting importance of validating data					
Number of obligors in sample.					
Length of observation period.					
Impact of differences in default definition to that of Basel, if any.					
Impact of macroeconomic conditions during data sample.					
Adjustments applied to data, if any.					
Data integrity.					
Impact of differences in assessment horizon to that of firm.					
Impact of other methodological Differences.					
Integrity of mapping between firms' ratings and those of comparison data sample.					
Similarity of portfolio to that of firm.					
Assessment					
Corroboration with PD Estimate.					

Use test scorecards
Data Accuracy Scorecards

"As a result of this consultation we have decided not to implement the user test 'scorecard' as we think it places an unnecessary burden on firms given the range of models being implemented and that we can achieve the same aim by other means." (FSA CP 05/03, par. 7.24)

"At present, we are not planning to proceed with a formal validation scorecard as set out in CP189." (FSA CP 05/03, par. 7.53)

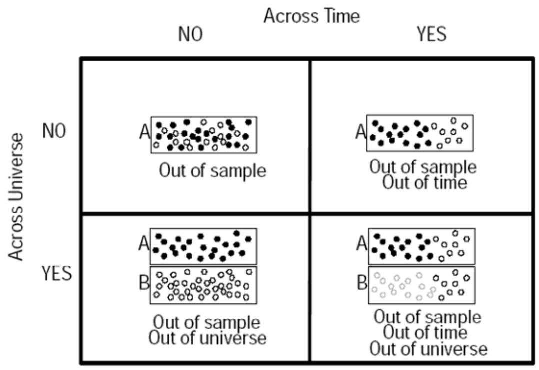
144

Quantitative Validation

- Measure how well ratings predict risk, loss, or exposure
  – PD, LGD, CCF
  – "The burden is on the bank to satisfy its supervisor that a model or procedure has good predictive power and that regulatory capital requirements will not be distorted as a result of its use." (par. 417)
- Compare realized numbers to predicted numbers
- Use appropriate performance metrics and test statistics
- Decide on significance levels!
- Out-of-sample and out-of-time

145

Out of Sample versus Out of time versus Out of Universe Quantitative Validation



Source: Moody's Risk Management Services

146

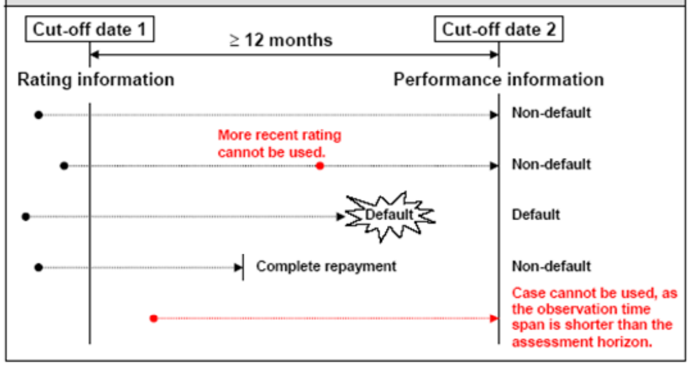
Validation During Model Development versus Validation During Model Usage

- Validation during model development
  - Out of sample
- Validation during model usage
  - Out of sample/Out of time

147

Generating the Data Set for Validation

Figure A1. Generating the data set for validation



HKMA, 2006

148

Problems with Quantitative Validation

- Different sources of variation
  - Sample variation
  - External effects (for example, macro-economy)
  - Internal effects (for example, strategy change)
- Low statistical confidence
  - Suppose we only look at sample variation and the PD for a grade is 100bp, and we want to be 95% confident that actual PD is no more than 20bp off from that estimate

$$n = \left(\frac{1{,}96\sqrt{PD(1-PD)}}{0{,}002} \right)^2$$

  - Would need about 9500 obligors in that grade!
- Statistical independence assumption violated
  - Correlation between defaults
  - Correlation between PD /LGD/EAD
- Data availability!

149

Levels of Backtesting

Calibration	mapping of rating to a quantitative risk measure. A rating system is considered well calibrated if the (ex-ante) estimated risk measures deviate only marginally from what has been observed ex-post.
Discrimination:	measures how well the rating system provides an ordinal ranking of the risk measure considered.
Stability:	measures to what extent the population that was used to construct the rating system is similar to the population which is currently being observed

150

Traffic Light Approach to Backtesting

- Assign traffic light colors to outcomes of tests
- Green color indicates everything is OK
- Yellow color indicates decreasing performance which could be interpreted as an early warning
- Orange color indicates performance difference that should be closely monitored
- Red color indicates severe problem
- More or less colors can be used!

151

Example Traffic Light Implementation

- Lightgreen is ideal
- Darkgreen indicates too conservative PD!

p-value	color
95-100%	red
90-95%	orange
70-90%	yellow
10-70%	lightgreen
0-10%	darkgreen

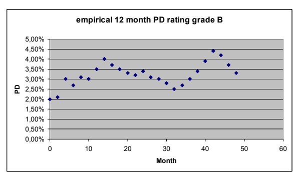
152

Backtesting PD

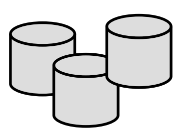
Level 2 — Risk ratings definition / PD calibration

Level 1 — Model / Scorecard / Logistic Regression

Level 0 — Internal Data / External Data / Expert Judgment



Characteristic Name	Attribute	Scorecard Points
AGE 1	Up to 26	100
AGE 2	26 - 35	120
AGE 3	35 - 37	185
AGE 4	37+	225
GENDER 1	Male	90
GENDER 2	Female	180
SALARY 1	Up to 500	120
SALARY 2	501-1000	140
SALARY 3	1001-1500	160
SALARY 4	1501-2000	200
SALARY 5	2001+	240



153

Backtesting PD at Level 0

- Check whether internal or external environmental changes will impact the rating model
  - New developments in economic, political, or legal environment, changes in commercial law or bankruptcy procedures (external)
  - Change of business strategy, exploration of new market segments, changes in organizational structure (internal)
- Two-step approach
  - Step 1: check whether population on which model is currently being used is similar to population that was used to develop model
  - Step 2: if differences occur in Step 1, verify stability of individual variables

154

Backtesting PD at Level 0: Step 1

- Construct system stability index (SSI) across ratings or score ranges

Population Stability Report

Score range	Actual %	Training %	(A-T)	A/T	ln(A/T)	index
0-169	7%	8%	-1%	0,8750	-0,1335	0,0013
170-179	8%	10%	-2%	0,8000	-0,2231	0,0045
180-189	7%	9%	-2%	0,7778	-0,2513	0,0050
190-199	9%	13%	-4%	0,6923	-0,3677	0,0147
200-209	11%	11%	0%	1,0000	0,0000	0,0000
210-219	11%	10%	1%	1,1000	0,0953	0,0010
220-229	10%	9%	1%	1,1111	0,1054	0,0011
230-239	12%	10%	2%	1,2000	0,1823	0,0036
240-249	11%	11%	0%	1,0000	0,0000	0,0000
250+	14%	9%	5%	1,5556	0,4418	0,0221
						0,053278

System Stability Index= $\sum (A-T)\ln\frac{A}{T}$

Rule of Thumb
- < 0.10 : No significant shift (Green flag)
- 0.10 - 0.25 : Minor shift (Yellow flag)
- > 0.25 : Significant shift (Red flag)

155

Backtesting PD at Level 0: Step 1

Segment	Construction Sample	Sample at $t-1$	Sample at t
Class 1	30%	32%	29%
Class 2	25%	28%	27%
Class 3	20%	17%	21%
Class 4	15%	13%	13%
Class 5	10 %	10%	10%
SI reference year		0.012	0.005
SI previous year			0.012

156

Backtesting PD at Level 0: Step 2

Use histograms, t-tests, or SSI to detect shifts in variables

Variable: Name	Variable: Categories	%Reference	% Sample $t-1$	% Sample t
days past due	< 8	85	80	78
	8 to 90	12	15	14
	≥ 90	3	5	8
SI reference			0.020	0.058
SI year before				0.028
savings amount	< 1000	30	38	40
	≥ 1000	70	62	60
SI reference			0.058	0.029
SI year before				0.002
years client	< 3	20	11	10
	3 to 10	30	25	23
	≥ 10	50	64	67
SI reference			0.097	0.138
SI year before				0.003

157

Backtesting PD at Level 0: Step 2

Characteristic analysis

Age	Training	Actual	Points	Index
18-24	12%	21%	10	0,9
25-29	19%	25%	15	0,9
30-37	32%	28%	25	-1
38-45	12%	6%	28	-1,68
46+	25%	20%	35	-1,75
				-2,63

158

Backtesting PD at Level 1

- Scorecard level!
- Validation of the logic behind the model used (for example, assumptions)
- For example, for (logistic) regression: qualitative checks on inputs to confirm whether the signs are as expected
- Inspect p-values, model significance, …
- Input selection (multicollinearity)
- Missing values and outliers
- Coarse classification

159

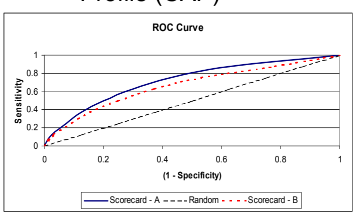
Backtesting at Level 1

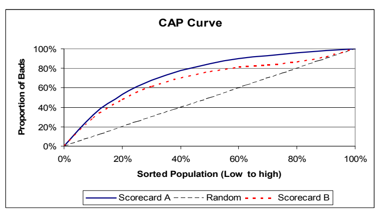
- Training set/Test set (hold out set)
  - Out of sample/out of time
- K-fold cross-validation
  - Split data into k-folds
  - Estimate algorithm k-times
  - Average performance
- Bootstrapping
  - resampling with replacement from the original sample

160

Backtesting at Level 1

- Receiver Operating Curve (ROC) versus Cumulative Accuracy Profile (CAP)





- Area under ROC curve (AUC) versus Accuracy Ratio (AR)
- Accuracy Ratio=Gini coefficient
- AR=2xAUC-1
- Confidence intervals around AR/AUC
- “Need at least 50 defaults to successfully calculate AUC and AR” (B. Engelmann, E. Hayden, and D. Tasche, 2003)
- “Most of the models we tested had ARs in the range of 50% to 75% for (out-of-sample and out-of-time) validation tests.” (Moody’s)

161 ...

Backtesting at Level 1

AUC	AR	Quality
0 < AUC < 0.5	AR < 0	No discrimination
0.5 < AUC < 0.7	0 < AR < 0.4	Poor discrimination
0.7 < AUC < 0.8	0.4 < AR < 0.6	Acceptable discrimination
0.8 < AUC < 0.9	0.6 < AR < 0.8	Excellent discrimination
0.9 < AUC < 1	0.8 < AR < 1	Exceptionnal

162

Backtesting PD at Level 1

	AR	Number of Obs.	Number of defaulters	Traffic Light
AR model				
AR year t AR year $t-1$ AR year $t-2$				
Average AR period 1 Average AR period 2				

163

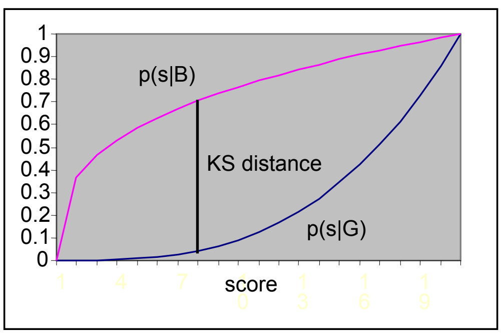
Example Backtesting PD at Level 1

	Number of Obs.	Number of defaulters	AR
AR model			
AR 2006	5677	97	0,81
AR 2005	5462	108	0,80
AR 2004	5234	111	0,83
AR 2003	5260	123	0,79
AR 2002	5365	113	0,79
AR 2001	5354	120	0,75
AR 2000	5306	119	0,82
AR 1999	4970	98	0,78
AR 1998	4501	62	0,80
AR 1997	3983	60	0,83
Average AR	5111,2	101,1	0,80

164

Backtesting at Level 1

- Kolmogorov-Smirnov statistic



- Pietra index, Bayesian error rate, conditional entropy, Divergence, …

165

continued...

Backtesting at Level 1

- "The group has found that the **Accuracy Ratio** (AR) and the **ROC measure** appear to be more meaningful than the other above-mentioned indices because of their statistical properties." (BIS 14 Working Paper)
- "Practical experience shows that the Accuracy Ratio has tendency to take values in the range **50% and 80%**. However, such observations should be interpreted with care as they seem to strongly depend on the composition of the portfolio and the numbers of defaulters in the sample." (BIS 14 Working Paper)
- There is **no absolute KS statistic**, **GINI coefficient**, or **ROC measure** that models need to reach to be considered adequate. (Monika Mars, Global Internal Ratings Validation Survey, 2003).

166

continued...

Backtesting at Level 1

- It is difficult and probably inappropriate to rely on a single measure, such as the widely used Gini coefficient. (FSA CP 05/03, par. 7.57)
- All depends on predictive power of inputs!
- Nice add-ons to consider:
  - Kolmogorov-Smirnov statistic (Pietra index)
  - Entropy measures, Bayesian Error rate, … (see BIS14)
  - Scorecard report (overrulings)
  - Lift curves

167

Overrides

- "For model-based ratings, the bank must have guidelines and processes for monitoring cases where human judgement has overriden the model's rating, variables were excluded, or inputs were altered. Banks must identify overrides and separately track their performance." (par. 428)

168

continued...

Overrides

Override Report

Score Interval	Accepts	Rejects	Total
<100	*1*	40	41
100-120	*2*	30	32
120-140	*2*	25	27
140-150	*3*	20	23
150-160	30	*2*	32
160-170	40	*2*	42
170-180	30	*1*	31
Total	108	120	228

- The black line denotes the cut-off; overrides are denoted in italics.
- Number of low-side overrides = 1+2+2+3=8
- Number of high-side overrides = 2+2+1=5

169

Overrides

Binary targets

Original	Override	Actual
A	A	Good
B	B	Good
A	C	Bad
C	C	Bad
A	B	Good
B	B	Good
B	B	Good
B	D	Good
C	C	Bad

- Compare ROC's
- For example, use test of De Long, De Long, Clarke-Pearson

Benchmark to external party

Original	Override	Benchmark
A	A	A
B	B	B
A	C	B
C	C	D
A	B	B
B	B	B
B	B	C
B	D	C
C	C	C

- Classification accuracy, notch difference graph
- McNemar test

170

Backtesting at Level 1: Validation Scorecard

COMBINED APPROACH TABLE

PROBABILITY OF DEFAULT RATING VALIDATION

Validation Range		Statistics :	1.20	86.38%	43.77%	58.34%	79.17%	1.42	0.71
Lower to Upper Limits		Meaning	Mean Difference	CND% > 50% CD: (1-PH)	K-S Statistic	Accuracy Ratio	ROC Statistic	Information Statistic	Kullback-Leibler
0	1	Random	0.00	50.00%	0.00%	0.00%	50.00%	0.0000	0.0000
1	2	Doubtful	0.25	59.87%	9.95%	14.00%	57.00%	0.0625	0.0313
2	3	Poor	0.50	69.15%	19.74%	27.60%	63.80%	0.2500	0.1250
3	4	Marginal	0.75	77.34%	29.23%	40.40%	70.20%	0.5625	0.2813
4	5	Satisfactory	1.00	84.13%	38.29%	52.00%	76.00%	1.0000	0.5000
5	6	Good	1.25	89.44%	46.80%	62.30%	81.15%	1.5625	0.7813
6	7	Very Good	1.50	93.32%	54.67%	71.10%	85.55%	2.2500	1.1250
7	8	Strong	1.75	95.99%	61.84%	78.40%	89.20%	3.0625	1.5313
8	9	Very Strong	2.00	97.72%	68.27%	84.30%	92.15%	4.0000	2.0000
9	10	Excellent	2.25	98.78%	73.94%	90.08%	95.04%	5.0625	2.5313
10	11	Excellent	2.50	99.38%	78.87%	94.20%	97.10%	6.2500	3.1250
11	12	Excellent	2.75	99.70%	83.09%	97.14%	98.57%	7.5625	3.7813
12	13	Superior	3.00	99.87%	86.64%	98.91%	99.46%	9.0000	4.5000
		Validation Scores :	5.80	5.42	5.64	5.62	5.62	5.74	5.75
					Average Validation Score:		5.66	or Good	

A PD Validation Framework for Basel II Internal Ratings-Based Systems, Maurice P. Joseph, Commonwealth Bank of Australia, 2005

171

Backtesting PD at Level 2

- Is there a sufficient number of rating grades?
  - Relation to masterscale
  - Impact on regulatory capital
- Are credit characteristics of borrowers in the same grade sufficiently homogeneous?
- Enough grades to allow for accurate and consistent estimation of loss characteristics per grade?
- Rating provide **ordinal ranking** of risk
  - Check whether realized PD is properly ranked through the grades
  - PD(A)< PD(B)< PD(C)
- Ratings provide **cardinal measures** of risk
  - For example, "BB" credits have an average PD of about 1.5%; "B" credits have average PD of 1.84%.

172

Backtesting at Level 2

- Compare estimated PDs versus realized PDs
- Test statistics:
  - Binomial test
  - Normal test
  - Hosmer-Lemeshow test
  - Vasicek one-factor model
- Complications:
  - Not enough defaults
  - Correlation between defaults
  - Decide on significance level
- Use as early warning indicators
- Impact of Risk Rating Philosophy
  - TTC versus PIT PDs

173

Brier Score

- The Brier score is defined as:

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{P}D_i-\theta_i)^2$$

whereby n is the number of obligors, $\hat{P}D_i$ the forecast PD and θ_i is 1 if obligor i defaults and 0 otherwise

- The Brier score is always bounded between 0 and 1 and lower values indicate better discrimination ability.

174

The Binomial Test

- Null hypothesis H_0: the PD of a rating category is correct
- Alternative hypothesis H_a: the PD of a rating category is underestimated
- Assumption: default events per rating category are independent!
- Given a confidence level, α (for example, 99%), H_0 is rejected if the number of defaulters k in the rating category is greater than or equal to k* which is obtained as follows:

$$k^* = \min\{k \mid \sum_{\mathrm{i=k}}^{\mathrm{n}}\binom{n}{i}PD^i(1-PD)^{n-i} \le 1-\alpha\}$$

- Use Normal Approximation (CLT): for large n, binomial distribution can be approximated as N(nPD, nPD(1-PD))

175

continued...

The Binomial Test

- Hence, we have:

$$P(z \leq \frac{k^* - nPD}{\sqrt{nPD(1-PD)}}) = \alpha, \text{with z the standard normal distribution}$$

- The critical value can then be obtained as follows:

$$k^* = \Phi^{-1}(\alpha)\sqrt{nPD(1-PD)} + nPD$$

with Φ-1(α) the inverse normal distribution.

- In terms of a maximum observed default rate p*, we have

$$p^* = \Phi^{-1}(\alpha)\sqrt{\frac{PD(1-PD)}{n}} + PD$$

- Summarizing: reject H_0 at significance level α, if the observed PD is higher than p*
- Binomial test assumes defaults are uncorrelated!
- If correlation present, higher probability to erroneously reject H_0 (type I error); use as early-warning system!

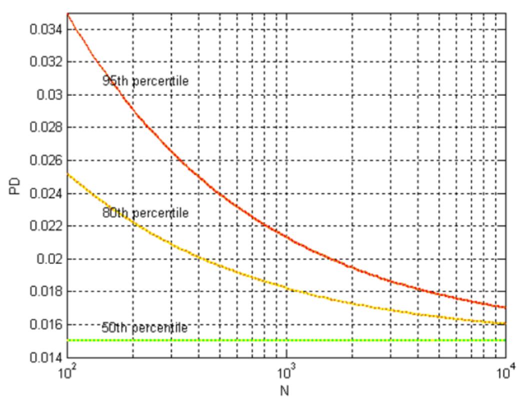
176

continued...

The Binomial Test

Values of the confidence interval as a function of the number of observations N (for a reference PD of 1.5%).

The width of the confidence interval decreases with growing number of observations, which makes it interesting to perform backtests on an aggregated level.



(Van Gestel 2005)

177

Binomial Test: Significance Levels

"*For example, if a Binomial test is used, AIs can set tolerance limits at confidence levels of 95% and 99.9%. Deviations of the forecast PD from the realized default rates below a confidence level of 95% should not be regarded as significant and remedial actions may not be needed. Deviations at a confidence level higher than 99.9% should be regarded as significant and the PD must be revised upward immediately. Deviations which are significant at confidence levels between 95% and 99.9% should be put on a watch list, and upward revisions to the PD should be made if the deviations persist.*"
(Hong Kong Monetary Authority, February 2006)

178

Extensions for the Binomial Test

Take into account default correlation as follows

$$z = \frac{DR - \hat{PD}}{\sqrt{\frac{\hat{PD}(1-\hat{PD})}{n(1-\rho^2)}}} \sim N(0,1)$$

z-statistic becomes smaller and hence less conservative compared to no correlation.

179

Example Traffic Light Implementation (Van Gestel 2005)

PD backtest based on binomial test; DR from Moody's statistics used as example. The PD is estimated as the average DR.

PD	Baa1	Baa2	Baa3	Ba1	Ba2	Ba3	B1	B2	B3	Caa-C	Av
	0.26%	0.17%	0.42%	0.53%	0.54%	1.36%	2.46%	5.76%	8.76%	20.89%	3.05%
DR	**Baa1**	**Baa2**	**Baa3**	**Ba1**	**Ba2**	**Ba3**	**B1**	**B2**	**B3**	**Caa-C**	**Av**
1993	0.00%	0.00%	0.00%	0.83%	0.00%	0.76%	3.24%	5.04%	11.29%	28.57%	**3.24%**
1994	0.00%	0.00%	0.00%	0.00%	0.00%	0.59%	1.88%	3.75%	7.95%	5.13%	**1.88%**
1995	0.00%	0.00%	0.00%	0.00%	0.00%	1.76%	4.35%	6.42%	4.06%	11.57%	**2.51%**
1996	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.17%	0.00%	3.28%	13.99%	**0.78%**
1997	0.00%	0.00%	0.00%	0.00%	0.00%	0.47%	0.00%	1.54%	7.22%	14.67%	**1.41%**
1998	0.00%	0.31%	0.00%	0.00%	0.62%	1.12%	2.11%	7.55%	5.52%	15.09%	**2.83%**
1999	0.00%	0.00%	0.34%	0.47%	0.00%	2.00%	3.28%	6.91%	9.63%	20.44%	**3.35%**
2000	0.28%	0.00%	0.97%	0.94%	0.63%	1.04%	3.24%	4.10%	10.88%	19.65%	**3.01%**
2001	0.27%	0.27%	0.00%	0.51%	1.38%	2.93%	3.19%	11.07%	16.38%	34.45%	**5.48%**
2002	1.26%	0.72%	1.78%	1.58%	1.41%	1.58%	2.00%	6.81%	6.86%	29.45%	**3.70%**
Av	**0.26%**	**0.17%**	**0.42%**	**0.53%**	**0.54%**	**1.36%**	**2.46%**	**5.76%**	**8.76%**	**20.9%**	**3.05%**

180

The Hosmer-Lemeshow Test

- The Hosmer-Lemeshow test can be used to test several rating categories simultaneously
- It also assumes independence of defaults
- The test statistic is defined as follows:

$$T = \sum_{i=0}^{k} \frac{(n_i PD_i - \theta_i)^2}{n_i PD_i (1 - PD_i)}$$

- n_i=number of debtors with rating i, PD_i is the estimated PD of rating i, θ_i is the number of defaulted debtors with rating i
- T converges towards a χ^2 distribution with k+1 degrees of freedom
- Compute p-value and decide on significance
- Type I error underestimated, when correlation is present!

181

The Normal Test

- The normal test is a multi-period test of correctness of a default probability forecast for a single rating category
- H_0: none of the true probabilities of default d_t, in the years t=1,…T, is greater than its corresponding forecast PD_t
- H_a: not H_0
- Reject H_0 at confidence level α if:

$$\frac{\sum_{t=1}^{T}(d_t - PD_t)}{\sqrt{T}\tau} > z_\alpha, \quad \tau^2 = \frac{1}{T-1}\left(\sum_{t=1}^{T}(d_t - PD_t)^2 - \frac{1}{T}\left(\sum_{t=1}^{T}(d_t - PD_t)\right)^2\right)$$

z_α is the standard normal α-quantile, (for example, $z_{0,99} \approx 2.33$)

182

Vasicek One-Factor Model

- Backtest at portfolio level using Vasicek one-factor model
- Define α* to be the fractional number of defaults that will not be exceeded a the 99.9% confidence level, P(X≤α*)=0.999

$$\alpha^* = N\left[\frac{N^{-1}(PD) + \sqrt{\rho}N^{-1}(0.999)}{\sqrt{1-\rho}}\right]$$

- 99.9% confidence interval thus becomes [0 ; α*]
- Allows to take into account correlated defaults (via asset correlation)
- Use Basel II correlations (or half to be more conservative!)
- Assumes infinitely granular portfolio
  - Use Monte Carlo simulation for finite (small) samples

183

Vasicek Test for Small Samples

(1) Generate a random variable $\eta \sim N(0,1)$ representing a factor common to all asset returns (e.g. overall state of economy)
(2) Generate a vector of n random variables $\epsilon_i \sim N(0,1)$

$$\begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix} = \sqrt{\rho}\eta + \sqrt{1-\rho} \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{17}$$

(3) Define the return thresholds that lead to default as $T = \Phi^{-1}(\hat{PD})$.
(4) Calculate the average default rate (DR) in the simulated sample as:

$$DR = \frac{\sum_{i=1}^{n} I(A_i \leq T)}{n}, \tag{18}$$

where the indicator function I is equal to 1 if $A_i \leq T$ and zero otherwise.

- Repeat simulation multiple times to get PD distribution and confidence interval
- Use Basel II correlations (or half)

184

Data Aggregation

- Assume portfolio with N obligors and n risk classes
- Approximately N/n observations per risk class
- More risk classes makes backtesting more difficult
- Aggregate data to improve significance of backtesting
  - Merge risk classes with low number of observations, for example, AA+ & AA & AA- into one overall rating class AA
  - Full portfolio or important segments

185

Risk Rating Philosophy

- Before validation starts, one needs to answer the question: "What is the rating system supposed to do?"
- "Rating philosophy is the term used to describe the assignment horizon of a borrower rating system." (FSA, CP 05/03, par. 7.66)
- Rating philosophy should be clearly articulated in bank rating policy.

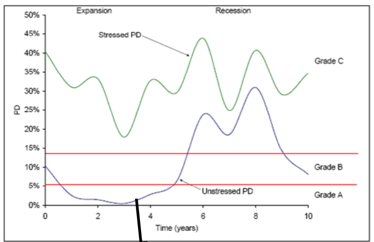
186

Point-in-Time (PIT) versus Through-the-Cycle (TTC)

- Point-in-Time (PIT) Ratings
  - Take into account obligors specific, cyclical and non-cyclical information
  - Obligors with same PIT grade share similar unstressed PDs
  - Rating changes rapidly with macro-economic situation
  - PD is the best estimate of the obligor's default during next 12 months
- Through-the-Cycle (TTC) Ratings
  - Take into account non-cyclical information
  - Obligors with same TTC grade share similar stressed PDs
  - Rating robust with respect to macro-economic situation
  - PD is the best estimate of the obligor's default during a credit cycle
- Many hybrids exist!

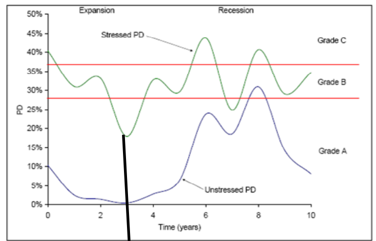
187

Point-in-Time (PIT) versus Through-the-Cycle (TTC)



Year	0	1	2	3	4	5	6	7	8	9	10
Grade	B	A	A	A	A	B	C	C	C	B	B

PIT rating system tied to unstressed PD



Year	0	1	2	3	4	5	6	7	8	9	10
Grade	C	B	B	A	B	B	C	A	C	B	B

TTC rating system tied to stressed PD

Many U.S. Bank rating systems conform more closely to a PIT philosophy (Treacy and Carey (1998))

188

Implications of Risk Rating Philosophy

- High level of rating migration suggests PIT ratings
  - Backtesting should find that realized default rates are close to forecast PD
  - PIT PDs validated against 12-month default rates
- Low level of rating migration suggests TTC ratings
  - Backtesting should find that realized default rates vary around forecast PD (rising in downturns and falling in upturns)
  - TTC PDs against cycle average default rates
- How to quantify migration?
  - Check using migration matrices
  - How much mobility/migration is sufficient (allowed)?
  - Calculate mobility (for example, Euclidean) metric and compare with external benchmarks (for example, S&P) (see above)

189

Rating Migration Analysis

Check stability of ratings using migration matrices:

	AAA-AA	A	BBB	BB	B	CCC	CC	default
AAA-AA	91.30	5.62	0.84	1.03	1.11	0.03	0.00	0.08
A	5.98	85.91	5.71	1.67	0.53	0.09	0.03	0.09
BBB	0.66	7.02	84.31	6.96	0.78	0.11	0.05	0.10
BB	0.08	0.58	3.99	89.28	4.81	0.43	0.26	0.57
B	0.12	0.08	0.26	10.95	84.07	1.61	1.06	1.86
CCC	0.00	0.18	0.09	1.99	15.10	63.47	9.13	10.04
CC	0.10	0.10	0.10	1.40	4.60	1.40	74.57	17.72

Compare with stable Identity matrix using mobility metrics:

$$M_{L1}(P) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}|P_{ij}-I_{ij}|}{N^2}$$

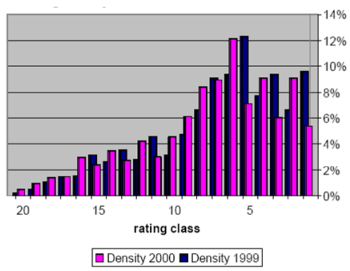
$$M_{L2}(P) = \frac{\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}(P_{ij}-I_{ij})^2}}{N^2}$$

Measurement, Estimation and Comparison of Credit Migration Matrices, Schuermann and Jafry, Journal of Banking and Finance, 2004

190

Backtesting PD at Level 2

- Change in portfolio distribution of PDs



- Is the change due to:
  - Cyclical effect (PIT rating)
  - Systematic changes in population (input variables)

Validating Inputs and Outputs of the Regulatory and Economic Capital Allocation Process, Fritz, Deutsche Bank, 2003

191

Example Traffic Light Indicator Dashboard

Level	Nature	Measure	Green	Yellow	Red
Level 2	Quantitative	Binomial test per rating (year per year)	Not statistically different at 90% level	Statistically different at 90% level but not at 95% level	Statistically different at 95% level
		Hosmer-Lemeshow test over all ratings (year per year)	Not statistically different at 90% level	Statistically different at 90% level but not at 95% level	Statistically different at 95% level
		Normal test per rating (multiple years)	Not statistically different at 90% level	Statistically different at 90% level but not at 95% level	Statistically different at 95% level
		t-test on overall average PD (multiple years)	Not statistically different at 90% level	Statistically different at 90% level but not at 95% level	Statistically different at 95% level
	Qualitative	Portfolio distribution	No shift	Minor shift	Major shift
		Nature of difference between estimated and realized PDs	No difference	Overestimation	Underestimation
		Check ratings migration using migration Matrices (+ rating philosophy!)	No migrations	Moderate migrations	Significant migrations
		Histogram(s) of default rates over all ratings (estimated versus realized PD, multiple years)	No shift	Minor shift	Major shift

192

continued...

Example Traffic Light Indicator Dashboard

Level	Nature	Measure	Green	Yellow	Red
Level 1	Quantitative	Difference in AUC (current sample versus development sample)	Not statistically different at 90% level	Statistically different at 90% level but not at 95% level	Statistically different at 95% level
		Difference in AR (current sample versus development sample)	Change of < 5% in AR	Change between 5% and 10% in AR	Change of > 10% in AR
		Difference in classification accuracy (current sample versus development sample)	Not statistically different at 90% level	Statistically different at 90% level but not at 95% level	Statistically different at 95% level
	Qualitative	Visual Check ROC curve (current sample versus development sample)	No shift	Minor shift	Major shift
		Visual Check CAP curve (current sample versus development sample)	No shift	Minor shift	Major shift
		(Visual) Check Kolmogorov-Smirnov curve and statistic (current sample versus development sample)	No change	Minor change	Major change
		Number of overrides	No change	Minor change	Major change

193

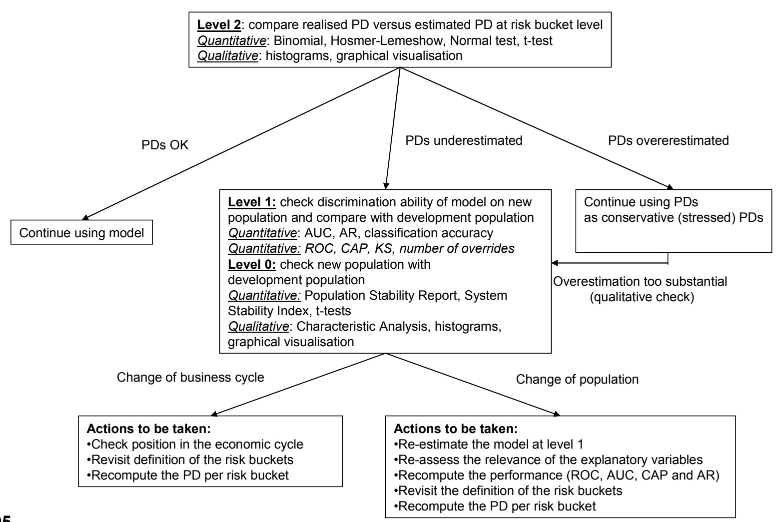
continued...

Example Traffic Light Indicator Dashboard

Level	*Nature*	*Measure*	*Green*	*Yellow*	*Red*
Level 0	*Quantitative*	System Stability Index (current sample versus development sample)	SSI < 0.10	0.10 < SSI < 0.25	SSI > 0.25
		t-test on attribute averages (year per year and attribute per attribute)	Not statistically different at 90% level	Statistically different at 90% level but not at 95% level	Statistically different at 90% level but not at 95% level
	Qualitative	Characteristic analysis (year per year and attribute per attribute)	No shift	Minor shift	Major shift
		Histogram of attributes distribution (year per year and attribute per attribute)	No shift	Minor shift	Major shift
		Compare attribute averages (year per year and attribute per attribute)	No shift	Minor shift	Major shift

194

Example Action Scheme



195

Backtesting LGD and EAD

- Methodologies less well developed than for PD
- Three levels:
  - Calibration
  - Discrimination
  - Stability
- For stability, the same method as for PD backtesting can be used (System Stability index)

196

Backtesting LGD and EAD Discrimination

- Use performance metrics from regression models
  - R-squared, correlation, MSE, MAD, CAP plots and AR, …

	MSE	Number of Obs.	Number of defaulters	Traffic Light
MSE model				
MSE year t MSE year $t-1$ MSE year $t-2$				
Average MSE period 1 Average MSE period 2				

197

Backtesting LGD and EAD Calibration

	Rating 1	Rating 2	...	Rating n	Non-rated	Average
Estimated LGD						
Actual LGD year t Actual LGD year t-1 Actual LGD year t-2 ...						
Average LGD period 1 Average LGD period 2						

Use a paired t-test as follows:

$$t = \frac{\frac{1}{n}\sum_{i=1}^{n} LGD_i - LGD^*}{\frac{s}{\sqrt{n}}}$$

whereby LGD* represents the estimated LGD and t follows a Student's t-distribution with n-1 degrees of freedom.

198

Benchmarking

- Comparison of internal PD/LGD/EAD estimates with external estimates
- Example benchmarking quantities are credit scores, ratings, calibrated risk measurements, or migration matrices
- Benchmarking partners can be credit bureaus, rating agencies, data poolers, or internal experts!

199

Benchmarking Problems

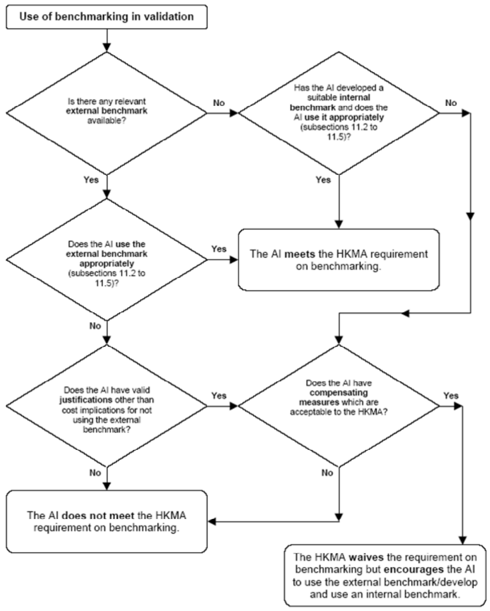
- Unknown quality of external ratings
- Different methodologies/processes/portfolio compositions
- Take into account rating philosophy (PIT versus TTC)
- Different default/loss definition
- Different LGD weighting scheme, discount factor, collection policy, …
- Legal constraints (for example, banking secrecy)
- Endogenity of credit risk (?)
  - Dependent on credit culture/credit process
- Cherrypicking

200

Benchmarking

"*Where a relevant external benchmark is not available (e.g. PD for SME and retail exposures, LGD, and EAD), an AI should develop an internal benchmark. For example, to benchmark against a model-based rating system, an AI might employ internal rating reviewers to re-rate a sample of credit on an expert-judgment basis.*" (HKMA 2006)

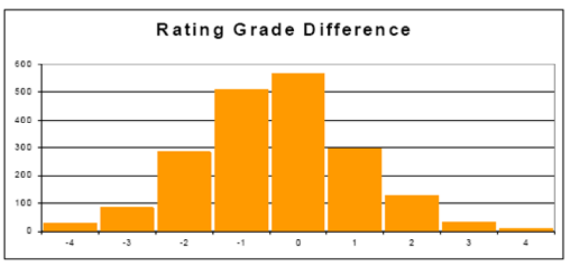
201



HKMA 2006

202

Benchmarking: Use of Histograms



Beaumont 2005

Negative Difference implies over rating by institution relative to benchmark

Van Gestel 2005

Rating comparison	<-2 notches	-2 notches	-1 notch	0 notch	+1 notch	+2 notches	>+2 notches	Nobs
Performance year t								
Performance year t-1								
Performance year t-2								
Performance year t-3								
…								
Av performance period 1								
Av performance period 2								
…								

203

Spearman's Rank-Order Correlation

- Measures the degree to which a monotonic relationship exists between the scores or ratings provided by an internal rating system and those from a benchmark
- Compute numeric rankings by assigning 1 to lowest score or ranking, 2 to second lowest score, …
- Take the average in case of tied scores or ratings
- Spearman's ρ_S can then be computed as:

$$\rho_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)},$$

whereby n is the number of obligors, and d_i the difference between the rankings
- Always ranges between -1 (perfect disagreement) and +1 (perfect agreement)

204

Kendall's Tau

- Assume we want to compare the estimates of institution X to those provided by benchmark Y for n debtors.
- Two debtors are said to be concordant if the debtor who is higher rated by X, is also higher rated (scored) by Y, and are discordant if the debtor higher rated (scored) by X is lower rated by Y. The two cases are neither concordant nor discordant if they are tied on X or Y or both.
- Let A represent the number of concordant pairs and B represent the number of discordant pairs

$$Kendall's\ tau = \frac{A-B}{\frac{1}{2}n(n-1)}$$

the denominator represents the number of possible pairs for n debtors.
- Kendall's tau is 1 for perfect agreement, -1 for perfect disagreement
- Increasing values imply increasing agreement between the ratings or scores

205

Gamma

- Note that Kendall's tau statistic assumes no ties
- Gamma is defined as follows:

$$Gamma = \frac{A - B}{A + B}$$

  where A represents the number of concordant pairs, and B the number of discordant pairs.
- Gamma ignores all pairs of ties, tied on either X or Y.
- Gamma is +1 if there are no discordant pairs, -1 if there are no concordant pairs, and 0 if there are equal numbers of concordant and discordant pairs

206

Example Benchmarking Exercise

By comparing the rating criteria of its internal rating system with those of Moody's, an institution concludes that 50% of the borrowers assigned to its rating grade B would have Moody's ratings "Baa1", 25% "A3", and 25% "Ba1". In the past 5 years, average annual default rates of these Moody's ratings were 3%, 2%, and 4%, respectively. The benchmark PD of rating grade B can then be estimated as:
50% x 3% + 25% x 2% + 25% x 4% = 3%

Compare with own PD and decide!

207

Qualitative Validation

- Use testing
- Data quality
- Model design
- Documentation
- Corporate governance and management oversight

208

Use Test

- "Internal ratings and default and loss estimates must play an essential role in the credit approval, risk management, internal capital allocations, and corporate governance functions of banks using the IRB approach." (par. 444, Basel II Accord)
- Credit pricing, credit approval, economic capital calculation, …

209

continued...

Use Test

- Essential does not necessarily mean exclusive or primary
- Three conditions to meet use test requirement (FSA)
  - **Consistency**: information IRB estimates (PD/LGD/EAD) are based on is consistent with internal lending standards and policies
  - **Use of all relevant information**: any relevant information used in internal lending standards and policies is also used in calculating IRB estimates
  - **Disclosure**: if differences exist between calculation of IRB estimates and internal purposes, it must be documented and reasonableness demonstrated

210

Use Test

- Application scoring PD
  - 1 year time window?
  - 90 days arrears (defaulter versus bad payer)
  - FSA advise to use the Basel scorecard for adjusting cut-off, application of policy rules, and targeting new customer sectors
- Downturn LGD
  - "Firms can use different LGDs for business purposes to those used for regulation and not fail the use test, provided that the rationale for their use and differences/transformation to capital numbers is understood." (FSA, 2005)
  - For example, for economic capital calculation, IFRS provisions (different discount rates), IAS 39 (no indirect costs), …
- Document and demonstrate reasonableness to supervisor!

211

Data Quality

- "Data input into the IRB systems is accurate, complete, and appropriate." (FSA CP 05/03, par. 7.16)
- Accuracy
  - do the inputs measure what they are supposed to measure (for example, data accuracy scorecard in FSA CP189)
  - Data entry errors, missing values, measurement errors, and outliers are all signs of bad data accuracy
- Completeness
  - Observations with missing values can only be removed if sound justifications can be given
  - "While missing data for some fields or record may be inevitable, institutions should attempt to minimize their occurrence and aim to reduce them over time." (CEBS, CP10, par. 297, 2005)

212

Data Quality

- Timeliness
  - Use recent data
  - Data should be updated at least annually
  - Higher updating frequencies for riskier borrowers
- Appropriateness
  - No biases or unjustified data truncation
- Data definition
  - Define data in appropriate way (for example, ratios)
- For internal, external, and expert data!
- Develop a data quality program

213

Model Design

- When was the model designed and by who?
- What is the perimeter of the model?
  - Counterparty types, geographical region, industry sectors, …
- What are the strengths and weaknesses of the model?
- What data was used to build the model? How was the sample constructed? What is the time horizon of the sample? Which default definition was adopted?
- How were the ratings defined?
- Is human judgment used and how?
- ….

214

Documentation

- All steps of the IRB implementation and validation process should be adequately documented:
  - Portfolio differentiation, rating criteria, responsibilities, changes in rating process, organization of rating assignment, override policy, …
  - Detailed outline of theory, assumptions and/or mathematical and empirical basis of the assignment of estimates to grades
  - Stress tests, concentration risk management, …
  - See paragraphs 418-421 of the Accord
- Documentation should be transparent and comprehensive
- Both for internal and external models!
- Use document management systems with appropriate versioning facilities!
- Documentation test: can a new team use existing documentation to continue development or production of an IRB system?

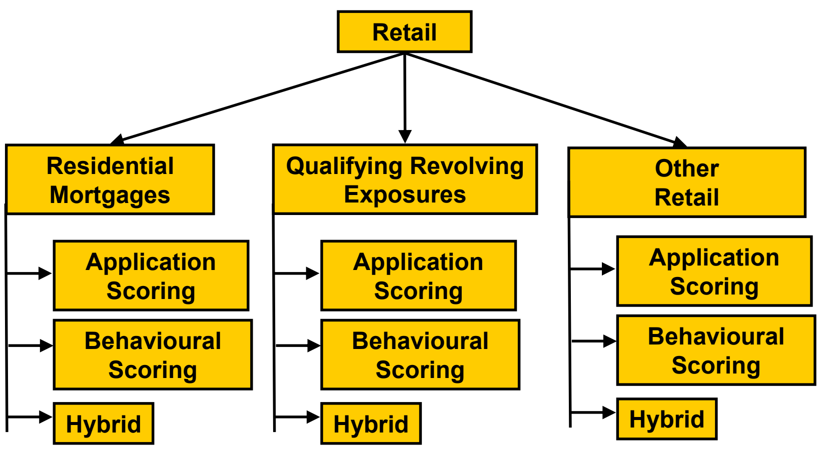
215

Corporate Governance and Management Oversight

- Involvement of the board of directors and senior management in the implementation and validation process
- Senior management is responsible for sound governance of the IRB framework
- Board and senior management should have a general understanding of the IRB systems
- Should demonstrate active involvement on an on-going basis, assign clear responsibilities, and put into place organizational procedures and policies that will allow the proper and sound implementation and validation of the IRB systems
- Outcome of the validation exercise must be communicated to senior management and, if needed, accompanied by appropriate response

216

Impact of Segmentation on Validation



217

continued...

Impact of Segmentation on Validation

- Segmentation usually done for
  - Statistical reasons (variable interactions)
  - Operational (application versus behavioral scoring)
  - Strategic
- Statistical Segmentation versus Expert Segmentation
- Already enforced by Basel II
- Beware not to over-segment!
- Effects
  - Lower number of defaults per segment
  - More efforts needed for validation/backtesting/ benchmarking
- Compare benefit of segmentation with cost!

218

Pillar 3 Reporting

- Encourage market discipline by developing a set of disclosure requirements which will allow market participants to assess key pieces of information on the scope, application, capital, risk exposures, risk assessment processes, and hence the capital adequacy of the institution. (par. 808)
- External reporting of risk management quality

Template 3.III.5 : Commercial and industrial portfolio: number of defaults for all PD grades at the time of default in the foundation approach for period t[39]

	Performing grades PD							Non-performing grades PD		
	grade 1	grade 2	grade 3	grade n				grade x	etc	
Estimated PD										
Actual PD										

Van Gestel, 2005

219

Pillar 3 Reporting

(i) PD Grade (%)	(ii) Exposure	(iii) Weighted Average % LGD(for advanced banks only)	(iv) Weighted Average Maturity (for advanced banks only)	(v) Value of exposures defaulting in the last year	(vi) % Default Rate
PD1					
PD2					
PD3					
PD4					
PD5					
PD6					
Default1					
Default2					

Van Gestel, 2005

220

Low Default Portfolios: References

- Benjamin, N., A. Cathcart, and K. Ryan. "Low Default Portfolios: A Proposal for Conservative Estimation of Default Probabilities." Financial Services Authority. April 2006.
- Forrest, A. "Likelihood Approaches to Low Default Portfolios." Dunfermline Building Society. 2005.
- "Low Default Portfolios." Joint Industry Working Group, BBA, LIBA, ISDA. January 2005.
- Pluto, K. and D. Tasche. "Estimating Probabilities of Default for Low Default Portfolios." 2005.
- "Validation of Low-Default Portfolios in the Basel II Framework." Basel Committee on Banking Supervision Newsletter, no.6, September 2005.

221

Low Default Portfolios (LDP)

- No formal definition in accord of LDP
  - 20 defaults? (Benjamin et al., FSA, 2006)
- Low default versus small in size (low data portfolio)
- For example, exposures to sovereigns, banks, project finance, large corporates, specialised lending, mortgage(?)
- Lack of sufficient statistical data and resulting difficulty in backtesting may exclude LDPs from IRB treatment (for example, paragraph 449 and 501)
- Historical averages are inappropriate
- Credit risk may be underestimated because of data scarcity
- Substantial portion of bank's assets may consist of LDPs!

222

Low Default Portfolios

- Views of the Basel Committee Accord Implementation Group's Validation Subgroup (AIGV)
  - "… LDPs should not, by their very nature, automatically be excluded from IRB treatment"
  - "… an additional set of rules or principles specifically applying to LDPs is **neither** necessary **nor** desirable"
  - "… relatively sparse data might require increased reliance on alternative data sources and data-enhancing tools for quantification and alternative techniques for validation"
  - "…LDPs should not be considered or treated as conceptually different from other portfolios"
- "We believe that it should be possible to include firm's LDPs in the IRB approach." (FSA, CP 05/03, par. 1.35)

223

continued...

Low Default Portfolios

- Data enhancing tools for quantification and validation
  - Pooling of data with other banks or market participants; use of external data
  - Aggregate subportfolios with similar risk characteristics to increase default history
  - Combine rating categories and analyse PDs for the combined category in a manner consistent with paragraphs 404-405 of the Basel Accord (for corporates, sovereigns and banks)
  - Use upper bound on PD estimate as input to capital requirement formula's
  - Infer PD estimates with a horizon of more than 1 year and then annualize the resulting figure
  - Use lowest non-default rating as proxy for default (still need to do calibration of ratings to PD consistent with Basel II definition)

224

continued...

Low Default Portfolios

- Benchmarking tools for validation
  - Compare internal ratings and migration matrices with third parties
  - Benchmark internal ratings against internal or external expert judgments
  - Internal ratings could be compared with market-based proxies (for example, equity prices, bond spreads)
  - Analyze rating characteristics of similarly rated exposures
  - Benchmark at the portfolio level instead of narrowly defined segments
  - Combine rating grades to make benchmarking more meaningful
- Importance of use test
  - Estimated PDs may be too conservative to be used internally
- LGD and EAD: even more difficult
  - smaller data sets/no benchmarks/less well developed methodologies

225

Low Default Portfolios: Pluto and Tasche (2005)

No defaults, assumption of independence

- Rating grades A, B, C with n_A, n_B, n_C obligors
- We assume ordinal borrower ranking to be correct
- $PD_A \leq PD_B \leq PD_C$
- Most prudent estimate of PD_A is obtained under the assumption that $PD_A=PD_C$, or $PD_A=PD_B=PD_C$
- Determine confidence region for PD_A at confidence level α (for example, α=90%)
- Confidence region is the region of all admissible values of PD_A such that the probability of not observing any default is higher than 1- α

226

Low Default Portfolios: Pluto and Tasche (2005)

In other words:

$$1-\alpha \leq (1-PD_A)^{n_A+n_B+n_C}, \quad or \quad PD_A \leq 1-(1-\alpha)^{1/(n_A+n_B+n_C)}$$

For example, suppose n_A=100, n_B=400, n_C=300

α	50%	75%	90%	95%	99%	99.9%
PD_A	0.09%	0.17%	0.29%	0.37%	0.57%	0.86%

227

Low Default Portfolios: Pluto and Tasche (2005)

Most prudent estimate $PD_B=PD_C$

$$1-\alpha \leq (1-PD_B)^{n_B+n_C}, \quad or \quad PD_B \leq 1-(1-\alpha)^{1/(n_B+n_C)}$$

For PD_C

$$1-\alpha \leq (1-PD_C)^{n_C}, \quad or \quad PD_C \leq 1-(1-\alpha)^{1/(n_C)}$$

α	50%	75%	90%	95%	99%	99.9%
PD_B	0.10%	0.20%	0.33%	0.43%	0.66%	0.98%
PD_C	0.23%	0.46%	0.76%	0.99%	1.52%	2.28%

228

Low Default Portfolios: Pluto and Tasche (2005)

Few defaults, assumption of independence

- Assume no default in grade A, two in B, one in C
- Use same principle, but binomial distribution, for example, for PD_A look at probability of observing less than or equal to 3 defaults

$$1-\alpha \leq \sum_{i=0}^{3} \binom{n_A+n_B+n_C}{i} PD_A^{\,i} (1-PD_A)^{n_A+n_B+n_C-i}$$

α	50%	75%	90%	95%	99%	99.9%
PD_A	0.46%	0.65%	0.83%	0.97%	1.25%	1.62%
PD_B	0.52%	0.73%	0.95%	1.10%	1.43%	1.85%
PD_C	0.56%	0.90%	1.29%	1.57%	2.19%	3.04%

229

Low Default Portfolios: Pluto and Tasche (2005)

- Extensions
  - Correlated default events: use Basel II single factor model and asset correlations
  - Calibration by scaling factors: make sure upper confidence bound PD estimates equal average default rate of overall portfolio
  - Multi-period extension

230

continued...

Low Default Portfolios: Pluto and Tasche (2005)

- Open questions
  - Which confidence levels?
    - Pluto and Tasche suggest < 95%
    - Benjamin et al. suggest 50% and 75%
  - At which number of defaults should one change to normal PD estimation methods?
    - Benjamin et al. suggest from 20 defaults (use LDP and non-LDP approach and set PD to maximum)
  - Supervisors may give guidance on this!

231

EAD and LGD for Low-Default Portfolio’s

“*It is accepted that the UK and international community are not well advanced in their thinking on EADs, and as a consequence estimation for LDPs adds further difficulty. As a result the EG fully endorse implementation to be pragmatic, flexible, and principle based. (FSA, 2005)*”

232

1.6 Stress Testing

Stress Testing: Literature References

- "Credit Stress-Testing." 2002. Consultative paper, Monetary Authority of Singapore (MAS).
- "Stress Testing at Major Financial Institutions: Survey Results and Practice." 2005. Report by a working group established by the Committee on the Global Financial System, Bank for International Settlements.
- "We recognize that there is no simple formula for an effective stress testing program: firms need to decide for themselves what is effective, taking into account their own business, structure, etc." (FSA, October 2006)

234

What Is Stress Testing?

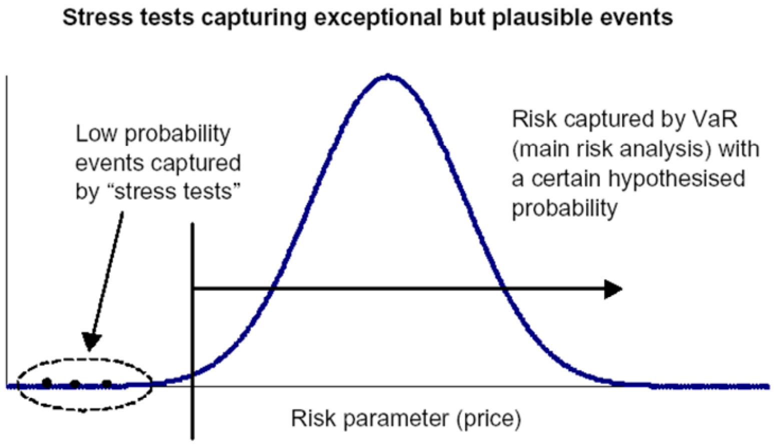
Stress testing is a risk management tool used to evaluate the potential impact of a firm of specific event and/or movement in a set of financial variables.

BIS working group report Jan 2005

- Potential impact
- Specific situations
- Financial variables
- Used as an adjunct to statistical models such as Value at Risk to consider the very unusual events

235

Stress Tests Adjunct to Main Risk Analysis



236

Why Use Stress Testing?

Basel Accord tells you to do so:

Stress testing must involve identifying possible events of future changes in economic conditions that could have unfavorable effects on a bank's credit exposures and assessment of the bank's ability to withstand such changes..

- Three levels of stress testing required in Basel Accord
- Level 0: Pillar 2 supervisory review – are you doing/using stress tests correctly?
  - will review operation and results of Pillar 1 stress testing in deciding how much above minimum capital requirement (MCR) necessary.

237

continued...

Why Use Stress Testing?

- Level 1: General Pillar 1 minimum requirements for stress tests—what happens if risks combine?
  - Objective as above. Examples of scenarios—economy downturns/market risk events/liquidity problems
- Level 2: Specific Pillar 1 credit risk stress tests—do the PD, LGD estimates cover poor economic conditions?
- effect of certain specific conditions on IRB requirements
  - Not worse case
  - Example is 2 quarters of zero growth on PD,LGD,EAD

238

continued...

Why Use Stress Testing?

Business value as well as regulatory value

BIS survey suggested some of the current uses

- In retail, forecasting used to set pricing and product features.
- Capturing impact of exceptional but plausible loss events
  - VaR reflects everyday markets, stress test abnormal markets
- Understanding risk profile of firm
  - Aggregate risks
  - Shocks in non-linear instruments (options)
- Capital allocation or verification of allocation
  - Stress tests used to set “soft limits”
- Evaluating business risks
  - The T of SWOT analysis

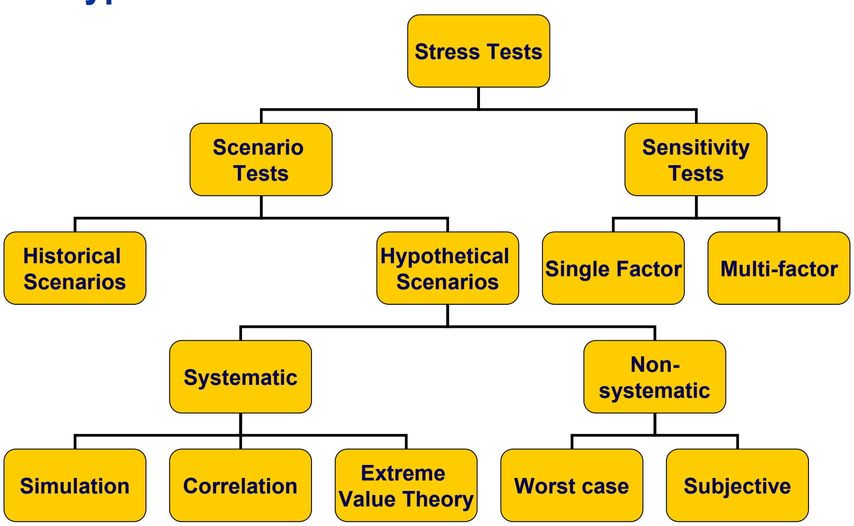
239

Factors That May Cause Stress Scenarios

- Macro-economic downturns
  - Mostly considered by institutions nowadays (FSA, October 2006)
- Deterioration in reputation
- Adverse change in competitive position
- Failure of a market counterparty
- Liquidity conditions

240

Types of Stress Tests



241

Sensitivity Tests

- Single factor stress tests
  - Usually used to check the sensitivity of an operating decision to changes in one risk
    - Yield curve shifts by 100 basis points
    - Decrease in £:$ exchange rate of 5%
  - Mainly used in trading book
- Multi-factor seeks to stress "all factors"
  - Really becomes a scenario type test and critical issue is correlation between factors

242

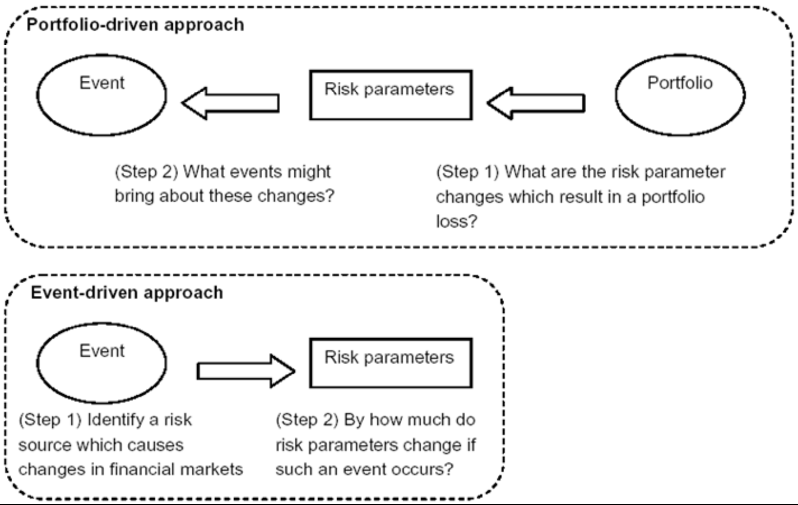
Scenario Tests

- Source of shock well defined as are parameters affected (there may be lags before affected)
  - Portfolio driven or event driven
  - Historical Scenarios
    - actual events so fewer judgments
    - still relevant if changes in portfolio/operations?
  - Hypothetical Scenarios
    - Potentially more relevant
    - Very labor intensive
    - Need lots of expert judgment
  - Always a trade-off between realism and comprehensibility

243

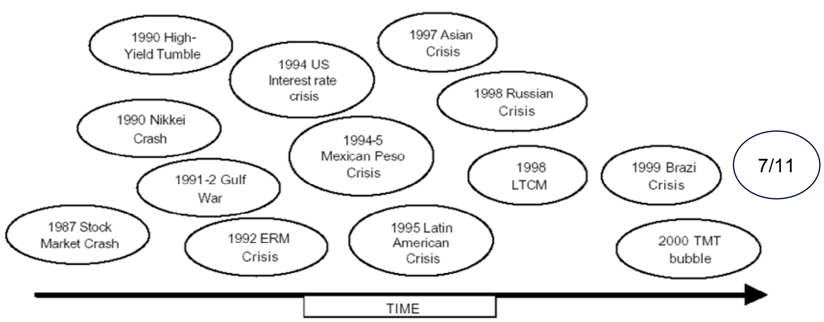
Portfolio Driven versus Event Driven

Approaches to scenario formulation



24

Historical Scenarios



MAS Credit Stress Testing

- Historical scenarios mainly affect trading portfolios

245

Historical Scenarios

- "The bank's stress test in this context should, however, consider at least the effect of mild recession scenarios." (par. 435, Basel Accord)
- "At a minimum, a mildly stressed scenario chosen by the AI should resemble the economic recession in Hong Kong in the second half of 2001 and the first quarter of 2002." (HKMA, 2006)

246

Hypothetical Scenarios

- Use hypothetical when no historic scenario appropriate
  - Portfolio changed or new risks identified
- Ensure no risk factors omitted
- Combined changes in each risk factor make economic sense

247

Types of Hypothetical Scenarios

- Worst-off: look at most adverse movement in each risk factor. Apply all
  - Least plausible—ignores correlations between risk factors
  - Very common
- Subjective: factors stressed using expert input
  - Only as good as the experts
- Simulation: Most approaches run simulations to get loss distribution under stressed conditions, but this reverses idea.
  - Define unacceptable loss and look at “stress” scenarios that cause it

248

continued...

Types of Hypothetical Scenarios

- Correlation: Stress some factors and use correlation matrix to simulate values of rest
  - For example, correlation between PD and LGD
  - Not clear correlations stay the same in times of stress
- Extreme Value Theory: most times loss distribution assumed normal or log normal. But under extreme stress, tail of distribution is “fatter” than normal.
  - Theory of stochastic processes identifies suitable extreme distributions

249

Stress Scenario's According to Basel II

- "*For this purpose, the objective is not to require banks to consider worst-case scenarios. The bank's stress test in this context should, however, consider at least the effect of mild recession scenarios. In this case, one example might be to use two consecutive quarters of zero growth to assess the effect on the bank's PDs, LGDs and EADs, taking account—on a conservative basis—of the bank's international diversification*." (par. 435, Basel II accord)
- According to FSA (October 2006)
  - Mild stress tests included tests that firms characterized as 'once in five years' or 'once in ten years'
  - Severe but plausible scenario: three successive years of GDP contraction (probability of 1 in 350 of occurring)

250

Current Uses of Stress Testing

- Both FSA and BIS have recently done surveys of use of stress testing in financial organisations
- FSA
  - "Stress testing most developed for market risk"
  - "Stress tests for credit risk less advanced"
  - "Surprisingly little evidence firms develop credit risk stress testing methodology"
- BIS
  - "Stress testing more integrated into risk management frameworks"
  - "Loan book stress testing at early stage compared with traded market"
  - "Integration of market risk and credit risk some way off for most firms"
- Throughout, credit risk is thought of in corporate sense and when it comes to retail and untraded markets, little evidence of stress testing of credit risk

251

FSA Survey of Current Use of Stress

Stress Test	Type of Risk					Correlation	
	Market	Liquidity	Credit	Operational (Basel definition)	Other	Market/ Credit	Other
Single Variable ⇩ Multi Variable ⇩ Complete scenarios						Rarely	Never
Aggregated across the firm	Often	Often	Rarely				

FSA DP 05/2

252

BIS Survey – Use of Scenarios

Summary survey statistics

	2004	2000
All firms		
Number of participating central banks	16	10
Number of respondent firms	64	43
Number of stress tests	963	424
Global dealer firms		
Number of respondent firms	21	19
Number of stress tests	262	209

Table 1a

Scenario types - All firms

Number of tests

	Interest rates	Equities	FX	Commodities	Credit	Property	Other	Multiple	Total
Historical	92	50	30	2	32	1	5	na	212
Hypothetical	81	36	26	20	72	18	40	na	293
Sensitivity	184	44	60	15	70	13	21	51	458
Total	357	130	116	37	174	32	66	51	963

BIS Stress Testing
Jan 05

253

continued...

Stress Testing

Level 2: stress test the PD's per risk bucket
Static approaches
- Recalibrate PD towards upper tail of PD distribution
- Recalibrate PD using a multiplication factor
- Downgrade risk ratings by 1 notch

Dynamic correlation approaches
- Correlated trend analysis

Dynamic Model based approach
- Structural form models
- Reduced form models

Level 2: stress test the LGD model
Necessary?
LGD should be downturn LGD anyway

Stress test EAD
Where is this appropriate
Do you have EAD models?

Level 1: stress test the PD/LGD models along with other risk
Market risk – liquidity problems – credit risks
- Scenario analysis
- Outlier analysis

Level 0: Supervisory review
- Are stress tests valid?
- Do their results suggest regulatory capital should be above MCR

254

Static PD Stress Testing

- Static approaches
- Do not take into account external (macro-economic) information
- Change PD for given score buckets
  - Recalibrate the PD towards the upper tail of the PD distribution (e.g. maximum, upper quantile/percentile)
  - Recalibrate the PD by augmenting it with a fixed amount in an absolute way (for example, 10 basis points rise in PD) or in a relative way (for example, 10% increase)
  - Drop all by one score bucket (equivalent to decrease all ratings by 1 notch/grade. For example, AAA becomes AA, AA becomes A)
  - Allow flexibility in score to PD odds transformation and use this to increase PD
- Change Scores
  - Drop fixed number of points or % amount in score
- Relatively easy to implement and understand
- Hard to defend connection with changes in economic conditions

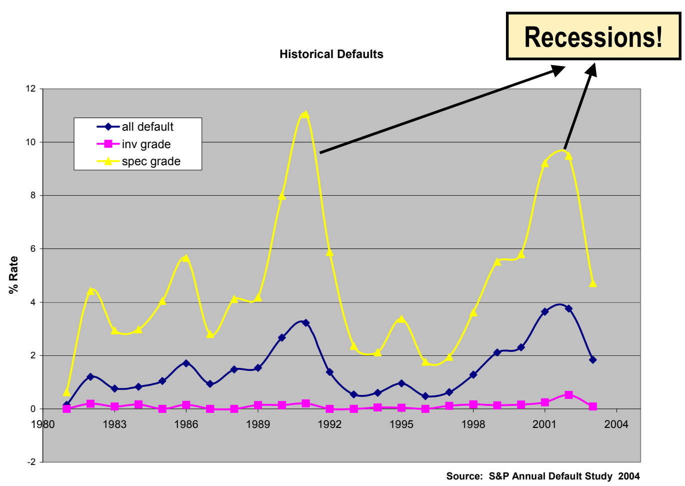
255

Dynamic Correlation Based Approaches to PD Stress Testing

- Look at historical connections between
  - default rates for different types of loan portfolios
  - and different segments within the same portfolio
  - and economic conditions
- Use these connections to connect economic conditions to changes in default rates in different segments/portfolios.

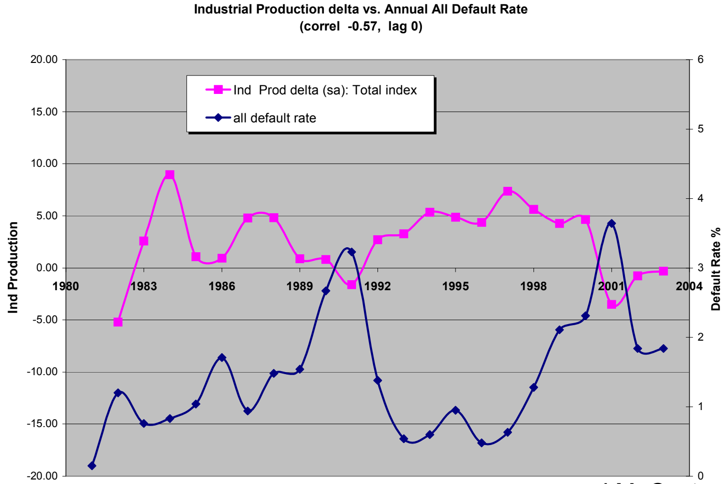
256

Dynamic Correlation Between Segments of a Portfolio



257

Dynamic Correlation Between Default Rate and Economic Conditions



258

J.M. Costa, 2004

Dynamic Correlation-based Approaches to PD Stress Testing

- Look at historical connections between default rates for different types of loan portfolios and different segments within the same portfolio and economic conditions.
- Correlated trend analysis – split into segments A,B,….
  - $PD_A = \beta_0 + \beta_1.GDP$ (single factor model)
  - $PD_A = \beta_0 + \beta_1.GDP + \beta_2.inflation$ (multivariate model)

259

continued...

Dynamic Correlation-based Approaches to PD Stress Testing

- Harder than it looks to get PD_A as segment A keeps changing
- Which economic variables to use
- Concern that "normal" correlations breakdown in stressed times—the very times that need to be modelled
- Is the next recession like the last?

260

Dynamic Model-based Stress Testing; Reduced Form Markov Chain Approach

- Think of ratings in IRB approach as states of Markov Chain
  - So state is score band or default status (0,1,2,3+ overdue)
  - At least one state corresponds to default
- Markov assumption is states of system describes all information concerning credit risk of customer
- Estimate transition probabilities of moving from state i to state j in next time period
- Use logistic regression to get transition probabilities to be functions of economic variables
- To stress test, choose the economic variables for a stressed scenario (again, scenario could last over several periods)

261

Stress Testing at Level 1

- Credit risk as part of other risk models. Are they coherent and portfolio driven?
- First identify the vulnerabilities of the credit portfolio
- Work backwards and identify plausible scenarios where these vulnerabilities become stressed
  - For example, for mortgage portfolios, the housing price may be considered as a major vulnerability, because it directly impacts the value of the collateral. Hence, a drop in house prices represents a stress scenario here.
- Once all vulnerabilities have been identified, one can re-evaluate the models using their stressed, outlying values. Recompute variables like PD and finally gauge the impact on the total regulatory capital.

262

Key Conclusions from FSA Stress Testing Thematic Review (October 2006)

- Close engagement by senior management resulted in the most effective stress testing practices
- Good practice was observed where firms conducted firm-wide stress tests of scenarios which were plausible, yet under which profitability was seriously challenged
- Communicating results clearly and accessibly was important for many firms
- Good practice entailed using group-wide stress testing and scenario analysis to challenge business planning assumptions

263

1.7 New Techniques for PD/LGD Modeling: Neural Networks

Methods for Classification

- Statistical methods
  - Logistic Regression
  - Discriminant Analysis (LDA, QDA, Fisher, regression)
- Machine Learning methods
  - Decision trees and rule extraction
  - K-nearest neighbor
- Genetic Algorithms
- Bayesian Networks
- Fuzzy techniques

265

A Brief Review of Logistic Regression

$$P(Y=1 \mid X_1,\ldots,X_n) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\ldots+\beta_n X_n)}} = \frac{e^{\beta_0+\beta_1 X_1+\ldots+\beta_n X_n}}{1+e^{\beta_0+\beta_1 X_1+\ldots+\beta_n X_n}}$$

- $0 \leq P(Y=1 \mid X_1, \ldots, X_n) \leq 1$
- Logistic transformation
- Maximum likelihood estimation with Newton-Raphson optimization
- Linear in the log-odds (scorecard points)
- Weights-of-Evidence coding

266

Why Is Logistic Regression So Popular?

- Industry Standard in Credit Scoring
- Better than it looks! (Hand, 2003)
- Because of coarse-classifying the input variables, a non-linear model is obtained!

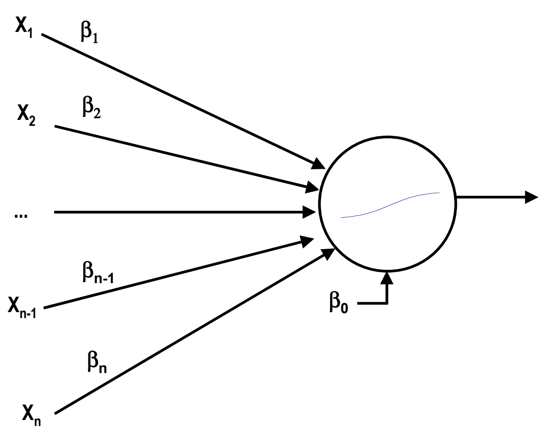
$$\log\left(\frac{P(Y=1 \mid X_1,\ldots,X_n)}{P(Y=0 \mid X_1,\ldots,X_n)}\right) = \beta_0 + \beta_1 f(X_1) + \ldots + \beta_q(X_q)$$

Typically, q > n!

- Generalized Additive Model (GAM)

267

Neural Network Representation of a Logistic Regression Model



$$P(Y \mid X_1,\ldots,X_n) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\ldots+\beta_n X_n)}}$$

268

Decision Boundary of Logistic Regression

- Because,

$$\log\left(\frac{P(Y=1 \mid X_1,\ldots,X_n)}{P(Y=0 \mid X_1,\ldots,X_n)}\right)=\beta_0+\beta_1 X_1+\ldots+\beta_n X_n$$

the logistic regression classifier assumes a linear decision boundary

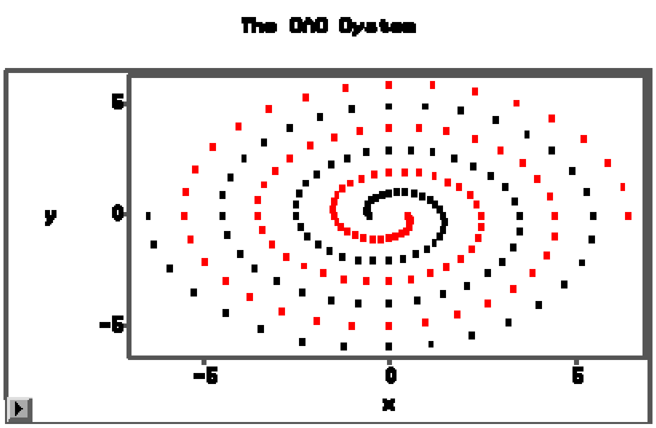
$$\beta_0+\beta_1 X_1+\ldots+\beta_n X_n=0$$

- No distributional assumptions on the independent or dependent variable!

269

Linear versus Non-Linear Decision Boundary

2 spiral data

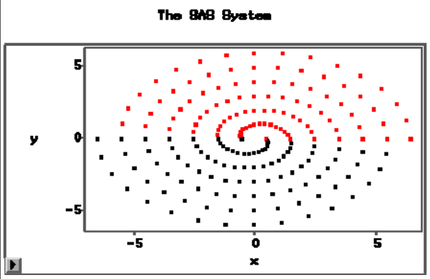


270

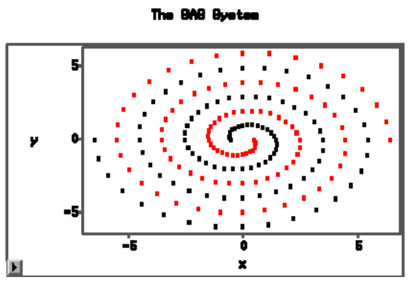
continued...

Linear versus Non-Linear Decision Boundary

Logistic Regression



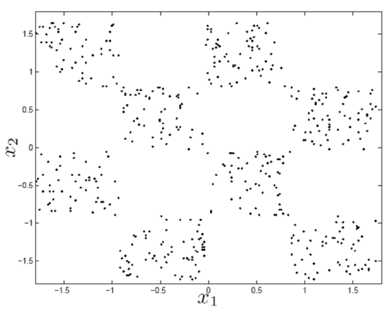
Neural Network with
40 hidden neurons

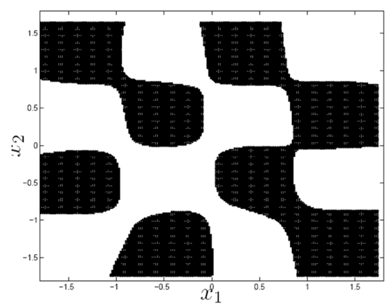


271

continued...

Linear versus Non-Linear Decision Boundary



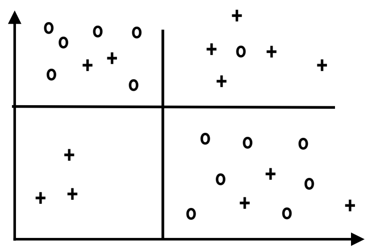


Logistic Regression	53.70
SVM	96.70

272

Decision Boundary of a Decision Tree

Decision boundaries orthogonal to axes



273

A Review of Linear Regression

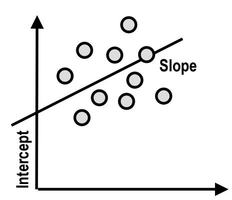
- Linear model

$$Y = \beta_0 + \beta_1 x_1 + \ldots \beta_n x_n = \mathrm{B}W$$

- Ordinary Linear Regression (OLS) yields

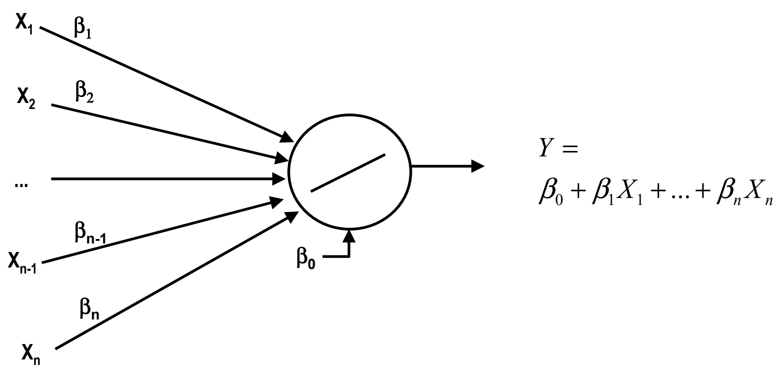
$$\mathrm{B} = (X^T X)^{-1} X^T Y$$

- Statistical tests to decide on relevance of variables
- Confidence intervals



274

Neural Network Representation of Linear Regression

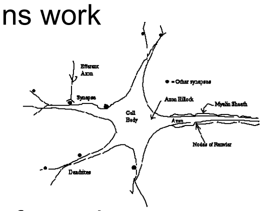


275

1.8 Basic Concepts of Neural Networks

The Biological Perspective

- Neural networks mimic how our brains work
  - Neurons: Brain Cells
  - Nucleus (at the center)
  - Dendrites provide inputs
  - Axons send outputs
- Interconnected Neurons exchange information (signals)
- High parallelization
- Associative memory
  - knowledge distributed over entire brain structure



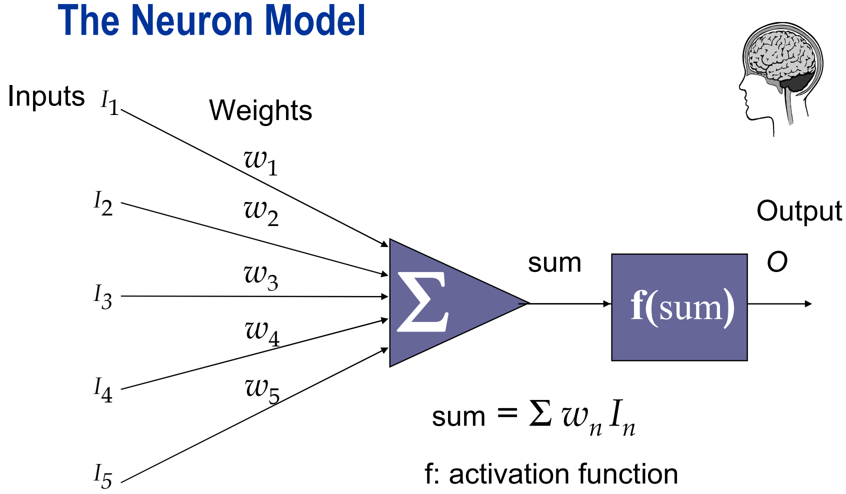
277

The Statistical Perspective

- Neural networks are generalizations of well-known statistical models
- Logistic regression
- Linear regression
- Non-linear extensions of existing statistical models

278

The Neuron Model



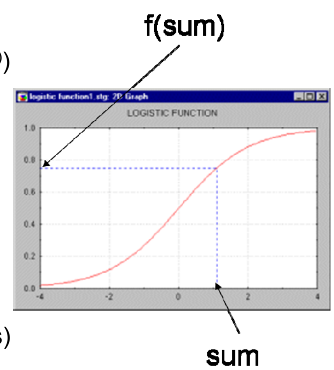
sum = $\Sigma\, w_n I_n$

f: activation function

279

Activation Function f

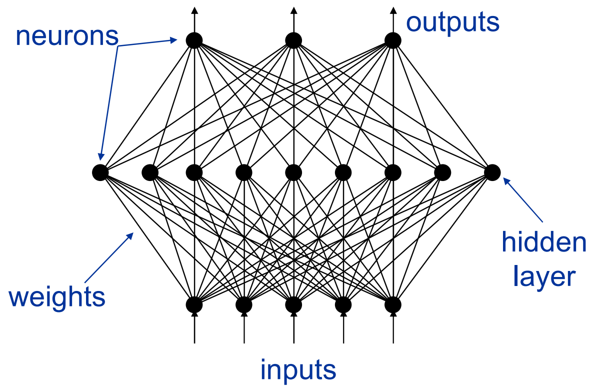
- logistic (sigmoid)
  - f(sum)=1/(1+e(-sum))
  - between 0 and 1
- hyperbolic tangent
  - f(sum)=(e(sum)-e(-sum))/(e(sum)+e(-sum))
  - between -1 and 1
- linear
  - f(sum)=sum
  - between -∞ and +∞
- Exponential
  - f(sum)=e^{sum}
  - between 0 and +∞
- Radial Basis Function (RBF Networks)
  - Gaussian activation functions



280

The Multilayer Perceptron (MLP)

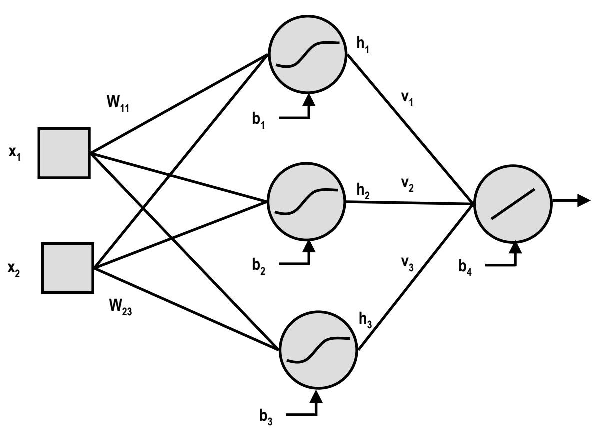
Organize neurons into layers.



281

continued...

The Multilayer Perceptron (MLP)



$$z = \sum_{j=1}^{3} v_i h_i + b_4$$

282

continued...

The Multilayer Perceptron (MLP)

- Every neuron has a bias input
  - intercept in linear regression
- Multiple hidden layers possible
  - Universal approximation property with 1 hidden layer
- No feedback connections!
- MLPs are the most common type of neural network used for supervised prediction (also in credit scoring)
- If k inputs, h hidden neurons (one hidden layer), 1 output, then (k+1)h+h+1 parameters need to be estimated

283

Terminology Neural Networks

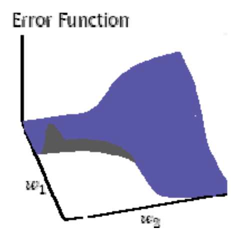
Neural Networks	Statistics
Learning/Training	Optimization, Estimation
Bias	Intercept
Weight	Parameter estimate
Epoch	Iteration Step
Pruning	Model reduction, input selection
Architecture	Model
Feedforward NN	Static nonlinear model

284

1.9 Training Neural Networks

Training Neural Networks

- For simple statistical models (for example, linear regression), there exists closed-form analytical formulas for the optimum parameter estimates
- For nonlinear models like neural networks, the parameter estimates need to be determined numerically, using an iterative algorithm
- The error function defines a surface in weight space
- Algorithm overview
  – Start with weights w^0
  – Recursively update w^0 until convergence



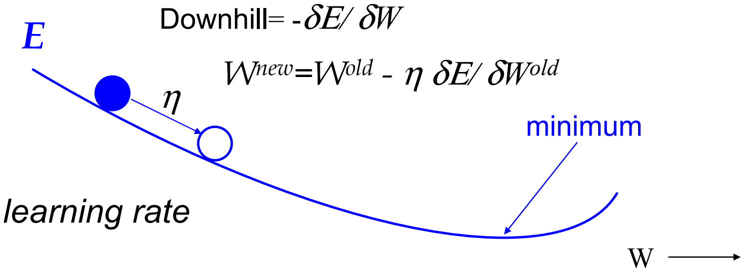
286

Gradient Descent Learning

- The error E is a function of the weights and the training data
- Minimize error E using steepest descent

Downhill= $-\delta E/\delta W$

$W^{new} = W^{old} - \eta\, \delta E/\delta W^{old}$

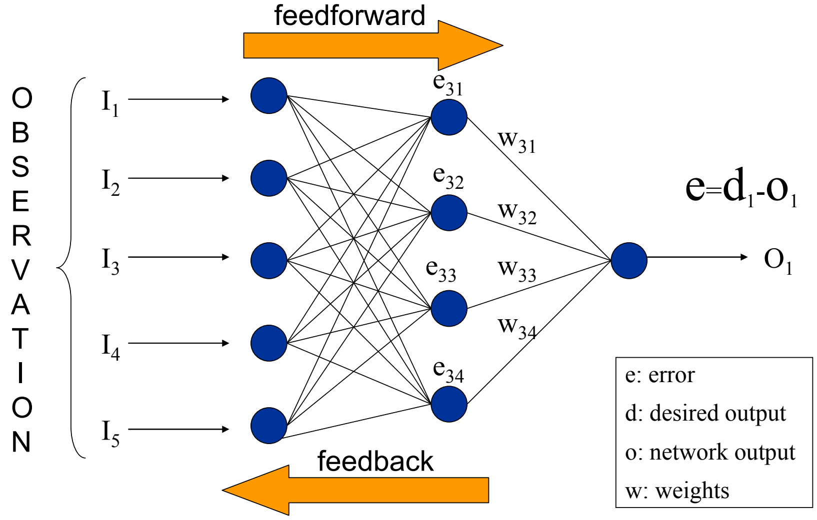


η: *learning rate*

How to evaluate $-\delta E/\delta W$?

287

Backpropagation Learning



288 ...

On-line Learning versus Batch Learning

- On-line (Incremental) learning
  - Update the weights after reading each observation
- Batch learning
  - Update the weights after reading the entire data set

✎ 1 epoch=1 run of all observations through the network

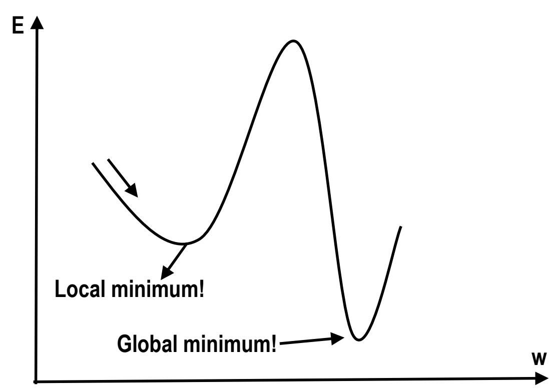
289

Problems with Backpropagation

- Learning rate η
  - Too high: oscillations, algorithm will diverge
  - Too small: slow progress
  - Adaptive learning rate!
- Momentum parameter α
  - $0 \leq \alpha \leq 1$
  - Current step direction is smoothed with the previous step
- Local minima

290

Local Minima



291

Objective Functions

- Most common:

$$E = \frac{1}{2}\sum_{n=1}^{N}(y(\mathbf{x}^n;\mathbf{w}) - t^n)^2$$

  with $y(x^n;w)$ the output of the network for observation n and tn the true output value (target) of observation n
- Variants (for example, Minkowski error)
- Maximum likelihood estimation
- Most error functions can be motivated from the principle of maximum likelihood!

292

Error Functions in SAS Enterprise Miner 5.1

- For interval targets, the default is the normal error function, that is, least squares minimization
- For nominal or ordinal categorical targets, the default is the multiple Bernouilli deviance (multiple when > 2 classes)
- Note that in SAS, the cross-entropy error function is the same as the Bernouilli function when the targets are 0/1! (cross entropy can also be used for interval variables)
- Default works fine!

293

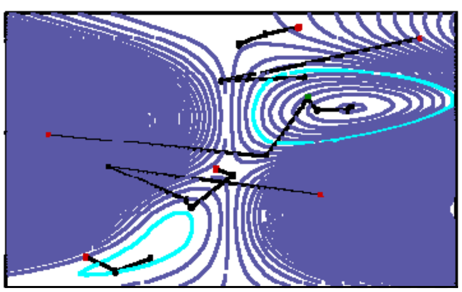
Advanced Training Algorithms

- Advanced non-linear optimization algorithms
- Take into account Hessian
  - Matrix of second-order partial derivatives of the error function with respect to the weights
  - Curvature of the error surface
- Newton based methods
- Conjugate gradient
- Levenberg-Marquardt
- Default in SAS works fine!

294

Preliminary Training

- Use a small number of random starting weights and take a few iterations (20 by default) from each.
- Use the best of the final values as the starting value.



295

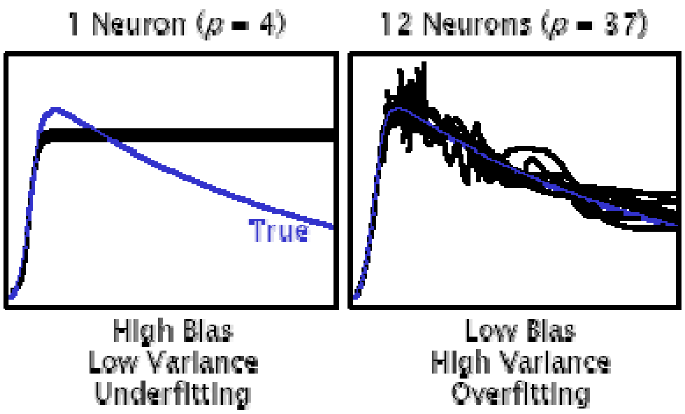
Stop Training

- Objective function shows no progress.
- The parameter estimates stop changing substantially.
- The gradient is close to zero.

296

1.10 Learning versus Outfitting

Bias/Variance Trade-off



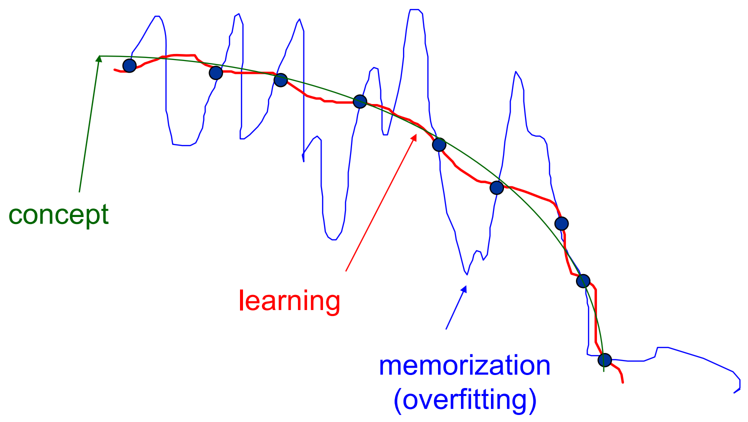
298

Learning versus Overfitting

- Successful Learning:
  – Recognize data outside the training set, that is, data in an independent test set.
- Overfitting ('Memorization')
  – Each data set is characterized by noise (idiosyncrasies) due to incorrect entries, human errors, irrationalities, noisy sensors, ….
  – A model that is too complex (for example, decision tree with too many nodes, neural network with too many neurons) may fit the noise, not just the signal, leading to overfitting.

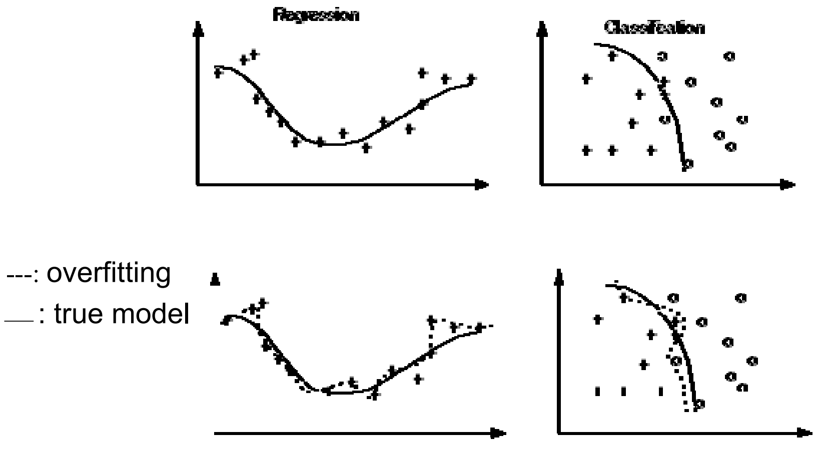
299

Learning versus Overfitting (2)



300

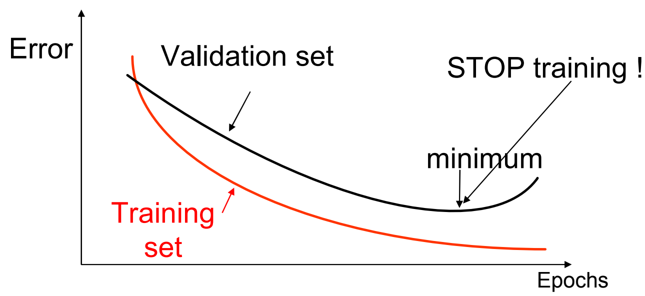
Learning versus Overfitting (3)



301

Avoiding Overfitting Using Early Stopping

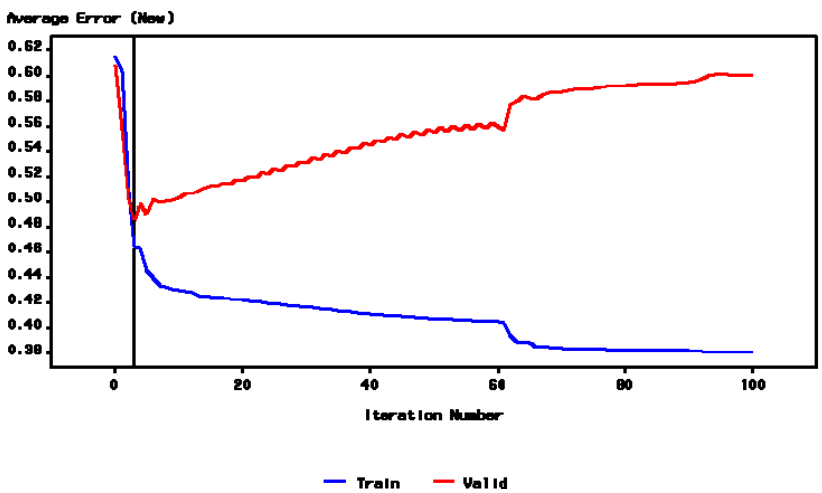
- Set aside validation set (for example, 30% of observations)
- Use training set to train network and validation set to decide when to stop training when overfitting occurs (‘early stopping’)



302

continued...

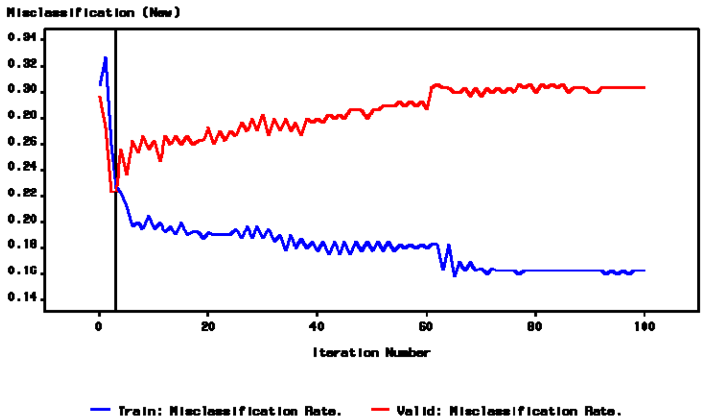
Avoiding Overfitting Using Early Stopping



303

continued...

Avoiding Overfitting Using Early Stopping



Difference between objective function and evaluation metric!

304

Avoiding Overfitting Using Regularization

- Bartlett (1997) demonstrated that generalization depends more on the size of the weights than the number of weights. A large network with small weights acts like a smaller, less complex network.
- Large weights cause the sigmoids to saturate.
- Thus, restraining the weights should prevent a bumpy overfitted surface.
- However, it may also prevent the model from adapting to true features of the data.
- Penalize large weights in the objective function!

305

continued...

Avoiding Overfitting Using Regularization

- Objective function=Error function + λ $||w||^2$
- The decay (shrinkage, smoothing) parameter λ controls the severity of the penalty
- Trade-off
  - Setting λ too low might cause overfitting
  - Setting λ too high might cause underfitting
- Ridge regression in statistics
- Separate weight regularization terms for different weight groups
  - Automatic Relevance Determination (ARD)
- Compared to early stopping
  - No need for validation data set, hence no loss of data

306

1.11 Preprocessing Data for Neural Networks

Normalizing Data

- Motivation
  - For example, neural network requires outputs/inputs between, for example, 0 en 1.
  - Scale variables to the same range (avoid one variable overpowers the other)
  - For example, salary versus age
- Min/Max normalization

$$X_{new} = \frac{X_{old} - \min(X_{old})}{\max(X_{old}) - \min(X_{old})}(\text{newmax} - \text{newmin}) + \text{newmin}$$

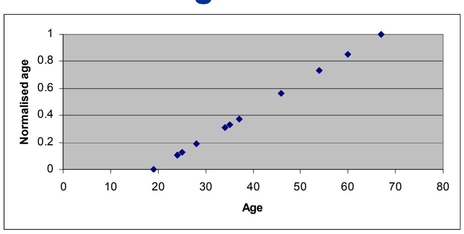
- Statistical normalization (when maximum and minimum are outliers)

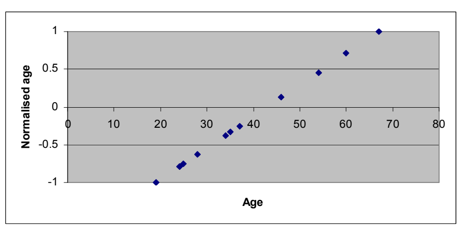
$$X_{new} = \frac{X_{old} - mean(X)}{stdev(X)}$$

- Decimal scaling $X_{new} = \frac{X_{old}}{10^n}$ with n the number of digits of the maximum absolute value

308

Normalizing Data





309

Recoding Nominal/Ordinal Variables

- Nominal variables
  - Dummy encoding
- Ordinal variables
  - Thermometer encoding
- Weights of evidence coding

310

Dummy and Thermometer Encoding

Dummy encoding

	Input 1	Input 2	Input 3
Purpose=car	1	0	0
Purpose=real estate	0	1	0
Purpose=other	0	0	1

Many inputs!
Explosion of data set!

Thermometer encoding

Original Input	Categorical Input	Thermometer Inputs		
		Input 1	Input 2	Input 3
Income ≤ 800 Euro	1	0	0	0
Income > 800 Euro and Income ≤ 1200 Euro	2	0	0	1
Income > 1200 Euro and ≤ 10000 Euro	3	0	1	1
Income > 10000 Euro	4	1	1	1

311

Weights of Evidence

- Measures strength of each (grouped) attribute in separating goods and bads
- Higher the weight of evidence means the less risk the attribute has

$$\text{Weight Of Evidence}_{\text{attribute}} = \log(\text{p\_good}_{\text{attribute}} / \text{p\_bad}_{\text{attribute}}),$$

$$\text{where p\_good}_{\text{attribute}} = \text{number of good}_{\text{attribute}} / \text{number of good}_{\text{total}}$$

$$\text{p\_bad}_{\text{attribute}} = \text{number of bad}_{\text{attribute}} / \text{number of bad}_{\text{total}}$$

If $\text{p\_good}_{\text{attribute}} > \text{p\_bad}_{\text{attribute}}$ then WOE > 0.

If $\text{p\_good}_{\text{attribute}} < \text{p\_bad}_{\text{attribute}}$ then WOE < 0.

312

Weight of Evidence

Age	Count	Distr Count	Goods	Distr Good	Bads	Distr Bad	Bad rate	Weight
Missing	50	2.50%	42	2.32%	8	4.15%	16%	-57.8505
18-22	200	10.00%	152	8.41%	48	24.87%	24%	-108.405
23-26	300	15.00%	246	13.61%	54	27.98%	18%	-72.0386
27-29	450	22.50%	405	22.41%	45	23.32%	10%	-3.95085
30-35	500	25.00%	475	26.29%	25	12.95%	5%	70.77059
35-44	350	17.50%	339	18.76%	11	5.70%	3%	119.1372
44 +	150	7.50%	147	8.14%	3	1.55%	2%	165.5087
Total	**2,000**		**1,807**		**193**		**9.65%**	
Information Value = 0.066								

$$\text{Ln} \left[\text{Distr Good} \; / \; \text{Distr Bad} \right] \times 100$$

313 ...

1.12 Architecture Selection for Neural Networks

Architectural Decisions

- How many hidden layers?
- How many hidden neurons?
- Which transfer functions?
- How to choose the regularization parameter?

315

Trial and Error

- Grid Search
  - Restricted model space
  - Don't use the test set to select best model!
- Sequential Network Contruction (SNC)
- Cascade Correlation
- Optimal Brain Damage/Optimal Brain Surgeon
- Bayesian Methods
- Genetic Algorithms
- Sometimes more art than science!

316

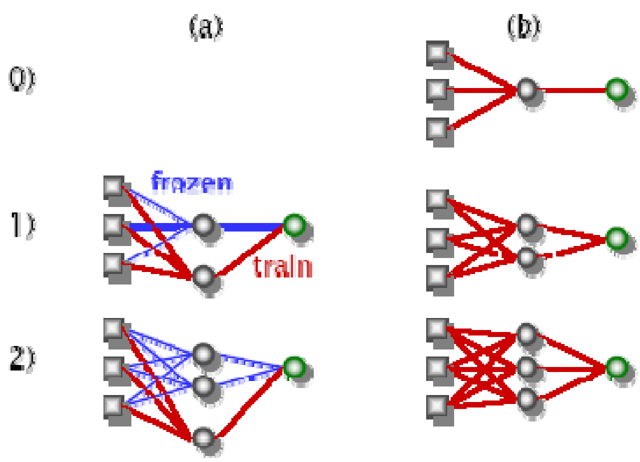
Sequential Network Construction

- Suggested by Moody (1994)
- Procedure
  - Train a 1-neuron MLP. Preliminary runs should be used to initialize the weights.
  - Add another neuron
    - Freeze all the weights involved with the first neuron at their final (current) values and train the connections to and from the second neuron (also retrain the output bias). This is like fitting a 1-neuron MLP to the residuals of the first fit. Preliminary runs could be used to initialize the new weights.
    - Train all the parameters in the full network using the final estimates of the previous step as starting values.
- The process is repeated in the same way by freezing the parameters from the previous step and training each new neuron on the frozen model. SNC alternates between fitting one-neuron networks and full networks that have starting values that are presumably close to a minimum.

317

continued...

Sequential Network Construction



318

Neural Networks in SAS Enterprise Miner 5.1

- Neural network node
- Autoneural node
  - Automatic architecture selection
  - Cascade correlation
  - SNC
- DMneural node
  - Fits an additive non-linear model using principal components

319

1.13 Pruning Neural Networks

Input Pruning

- Redundant and irrelevant inputs can cause overfitting and hinder training
- First select the architecture, then eliminate “unimportant” inputs
  - Update the architecture
- Input selection
  - Filters
    - Independent of learning algorithm
  - Wrappers
    - Integrated with learning algorithm
- Reduce model complexity

321

continued...

Input Pruning

- In statistical methods: use magnitude of the coefficients of the inputs to evaluate their importance (for example, standardized coefficients, significance tests).
- More complicated in neural networks because each input is associated with h input-to-hidden layer weights and h hidden-to-output weights.
- However, if all input-to-hidden weights for an input are zero (or close enough to zero), the input is clearly not important.

322

continued...

Input Pruning

Pruning algorithm

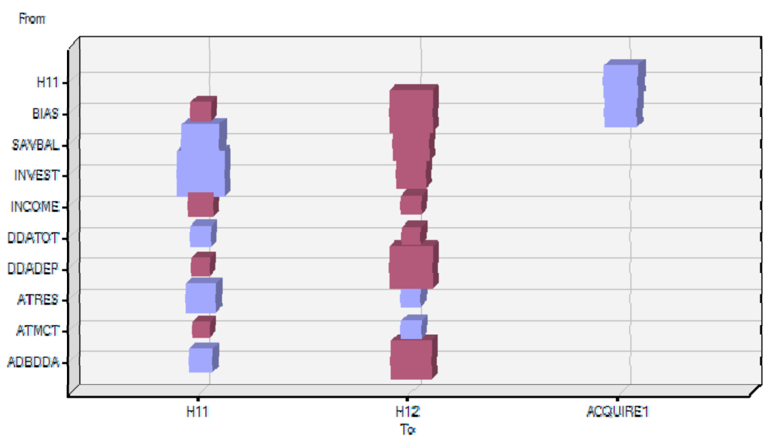
- Train the neural network
- Prune the input where the input-to-hidden weights are closest to zero
- Retrain the network
- If the predictive power increases (or stays the same) then repeat, if not reconnect the input and stop

Remarks

- Unimportant versus irrelevant (correlation!)
- How to measure closeness to zero?

323

Hinton Diagrams



- Size of square indicates size of weight
- Color of square indicates sign of weight

324

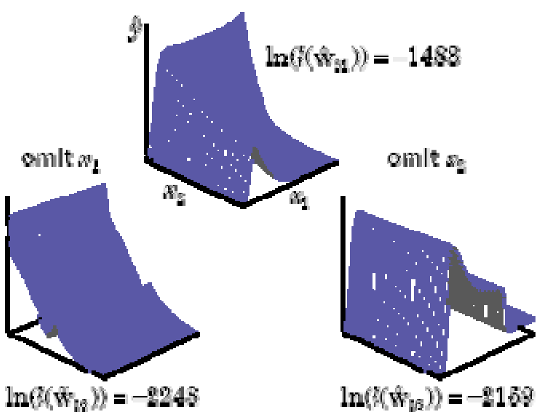
Likelihood Ratio Tests

- A formal statistical test that the input-to-hidden weights are zero can be conducted by removing the input and retraining the network to convergence. The likelihood-ratio (LR) test statistic is $2(\ln(L_{full})-\ln(L_{red}))$ where L_{full} and L_{red} are the likelihood functions for the full model and the reduced model, respectively.
- Under H_o, the test statistic asymptotically has a chi-squared distribution with h degrees of freedom, with h the number of removed parameters.
- The LR test can be used as a heuristic to identify the most likely candidate for pruning (largest p-value).
- Major disadvantage is that it requires retraining, that is, for k inputs, k neural networks need to be trained.

325

continued...

Likelihood Ratio Tests



326

Sensitivity Based Pruning (SBP)

- Moody (1994)
- In combination with SNC
- Set input to constant (for example, mean) and look at impact on error of the neural network
- Sensitivity is difference between the error of the new model and the error of the original model
- No retraining!
- Correlation effects might cause sensitivity to be overestimated
- Interaction effect might cause sensitivity to be underestimated
- Wrapper method

327

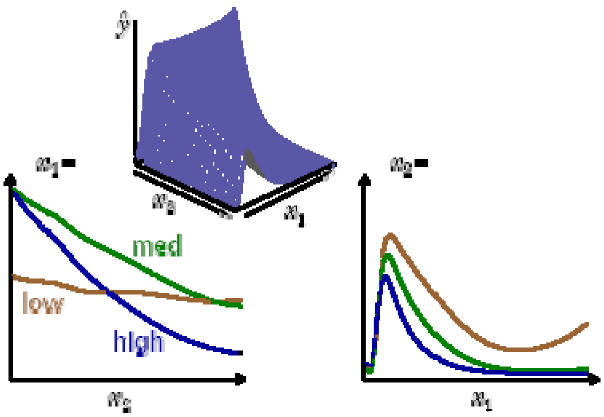
Slices

- Opposite to SBP
- Other inputs are held fixed at a typical value (for example, mean) and the conditional effect of the candidate input is observed as it varies across its range
- Look at a one-dimensional slice of the fitted surface in the input space
- Does not take into account interactions
  - Variable may seem unimportant from univariate perspective but important when considered with others
- Take two-dimensional slices, three-dimensional slices, ...
  - Not scalable

328

continued...

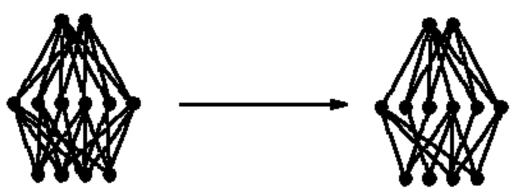
Slices



329

Individual Connection Pruning

- Inspect magnitude of individual weights
- Optimal Brain Damage
- Optimal Brain Surgeon



330

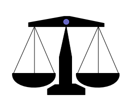
Benefits of Neural Networks

- Networks with one hidden layer are universal approximators
- Very good generalization capability (noise resistant)
- Non-parametric techniques (for example, no normality assumptions)
- Allow to effectively deal with high dimensional, sparse input spaces

331

Drawbacks of Neural Networks

- NNs are black box techniques
  - No easy comprehensible relationship between inputs and outputs
  - Trade-off between model accuracy and model comprehensibility
  - But: Techniques to extract decision rules out of trained Networks (for example, NeuroFuzzy Systems)
- How to choose network topology?
  - For example, hidden layers, hidden neurons, activation functions, …
  - But: New learning algorithms fairly robust w.r.t. network topology



332

More Information on Neural Networks

- Neural Networks for Pattern Recognition,
  Chris Bishop, Oxford University Press, 1999
- Pattern Recognition and Neural Networks,
  Brian D. Ripley, Cambridge University Press, 1996
- Introduction to Artificial Neural Networks,
  Jacek M. Zurada, PWS Publishing Company, 1992
- Journals
  - IEEE Transactions on Neural Networks
  - Neural Computation
  - Neural Networks
  - Neural Processing Letters

333

1.14 Support Vector Machines

Problems with Neural Networks

- Multimodal objective function
  - Multiple local minima
- Highly parameterized
  - How to choose hidden layers?
  - How to choose hidden neurons?
  - How to set regularization parameter?

335

Linear Programming

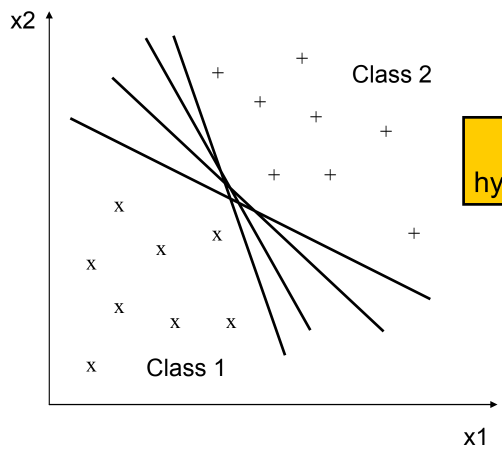
- Mangasarian (1965)
- Minimize the sum of the absolute values of the deviations (MSD)

  min $a_1+a_2+...+a_{ng}+a_{nb}$
  subject to
  $w_1x_{i1}+ w_2x_{i2}+... + w_nx_{in} \geq c\text{-}ai,\ 1 \leq i \leq n_g,$
  $w_1x_{i1}+ w_2x_{i2}+... + w_nx_{in} \leq c+a_i,\ n_g +1 \leq i \leq n_g+n_b,$
  $ai \geq 0.$

- Trivial solution: w_i and c=0
- Use fixed cut-off c, but experiment both with negative and positive c (Freed and Glover, 1986)

336

The Linear Separable Case



Many separating hyperplanes possible.

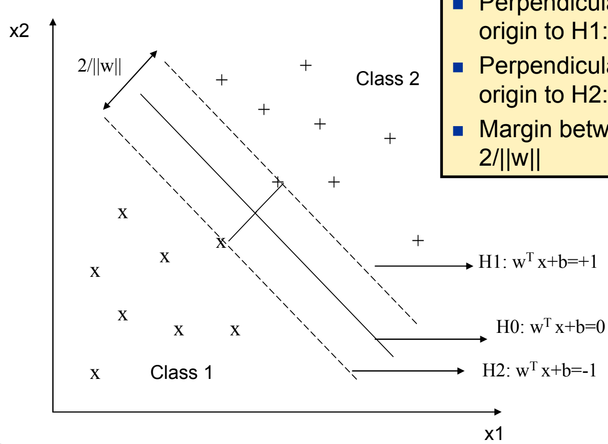
337

continued...

The Linear Separable Case

Consider the hyperplane, which maximizes the distance to the nearest points

- Perpendicular distance from origin to H1: |b-1|/||w||
- Perpendicular distance from origin to H2: |b+1|/||w||
- Margin between H1 and H2: 2/||w||



338

continued...

The Linear Separable Case

- Large margin separating hyperplane (Vapnik 1964)
- Given a training set $\{x_k, y_k\}_{k=1}^{N}, x_k \in \Re^n, y_k \in \Re$, where $y_k \in \{-1,+1\}$
- Assume

$$\begin{cases} w^T x_k + b \geq +1, & \text{if } y_k = +1 \\ w^T x_k + b \leq -1, & \text{if } y_k = -1 \end{cases}$$

- Or,

$$y_k[w^T x_k + b] \geq 1, \quad k = 1,\ldots,N$$

- Maximize the margin, or minimize ½ $||w||^2$
- Optimization problem
  min ½ $||w||^2$

$$y_k[w^T x_k + b] \geq 1, \quad k = 1,\ldots,N$$

- The classifier then becomes: $y(x)=\text{sign}(w^T x+b)$

339

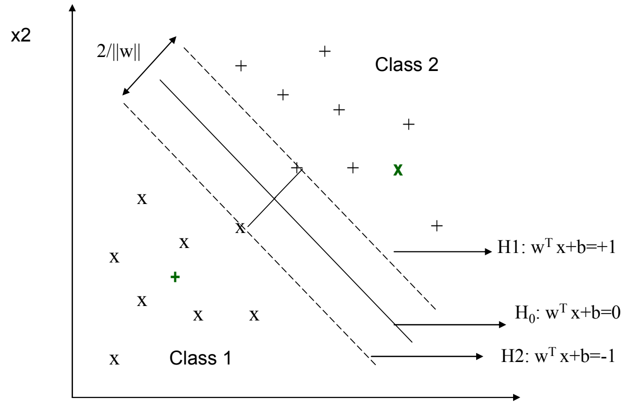
continued...

The Linear Separable Case

- Using Lagrangian optimization, a quadratic programming (QP) problem is obtained
- The solution of the QP problem is global
  - convex optimization
- Training points that lie on one of the hyperplanes H1 or H2 are called support vectors

340

The Nonseparable Case



341

continued...

The Nonseparable Case

- Allow for errors by introducing slack variables in the inequalities

$$y_k[w^T x_k + b] \geq 1 - \varepsilon_k, \quad k = 1,\ldots,N$$
$$\varepsilon_k \geq 0$$

- The optimization problem then becomes

$$\min_{w,\varepsilon} I(w,\varepsilon) = \frac{1}{2} w^T w + C \sum_{k=1}^{N} \varepsilon_k$$

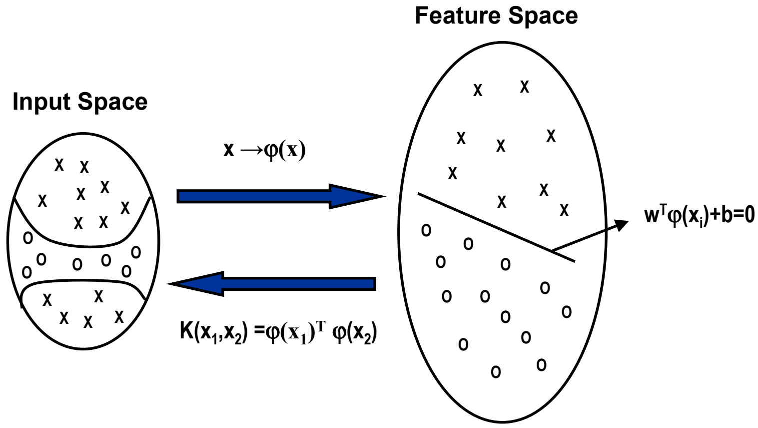
subject to

$$y_k[w^T x_k + b] \geq 1 - \varepsilon_k, \quad k = 1,\ldots,N$$
$$\varepsilon_k \geq 0$$

- C is a user-defined parameter, higher C corresponding to higher penalty for errors
- Again a QP problem is obtained after Lagrangian optimization.

342

The Non-linear SVM Classifier



343

The Non-linear SVM Classifier

- Map data from input space to high-dimensional feature space and construct linear hyperplanes in feature space

$$\begin{cases} w^T\varphi(x_k)+b \geq +1, & \text{if } y_k = +1 \\ w^T\varphi(x_k)+b \leq -1, & \text{if } y_k = -1 \end{cases}$$

- The optimization problem then becomes

$$\min_{w,\varepsilon} I(w,\varepsilon) = \frac{1}{2}w^T w + C\sum_{k=1}^{N}\varepsilon_k$$

$$y_k[w^T\varphi(x_k)+b] \geq 1-\varepsilon_k, \quad k=1,\ldots,N$$

$$\varepsilon_k \geq 0$$

344

continued...

The Non-linear SVM Classifier

- Construct the Lagrangian

$$L(w,b,\varepsilon;\alpha,\upsilon) = I(w,\varepsilon_k) - \sum_{k=1}^{N} \alpha_k \{y_k[w^T\varphi(x_k)+b]-1+\varepsilon_k\} - \sum_{k=1}^{N} \upsilon_k \varepsilon_k$$

with Lagrange multipliers $\alpha_k \geq 0$, $\nu_k \geq 0$

- Solution given by the saddle point of the Lagrangian

$$\max_{\alpha} \min_{w,b,\varepsilon} L(w,b,\varepsilon;\alpha,\upsilon)$$

- This yields,

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{k=1}^{N} \alpha_k y_k \varphi(x_k)$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{k=1}^{N} \alpha_k y_k = 0$$

$$\frac{\partial L}{\partial \varepsilon_k} = 0 \rightarrow 0 \leq \alpha_k \leq C, \qquad k = 1,\ldots N$$

345

continued...

The Non-linear SVM Classifier

- The dual QP problem becomes

$$\max_{\alpha} Q(\alpha) = -\frac{1}{2}\sum_{k,l=1}^{N} y_k y_l K(x_k^T x_l)\alpha_k \alpha_l + \sum_{k=1}^{N} \alpha_k$$

$$\sum_{k=1}^{N} \alpha_k y_k = 0$$

$$0 \leq \alpha_k \leq C, \qquad k = 1,\ldots, N$$

- The kernel function is defined through the Mercer theorem: $K(x_k,x_l)=\varphi(x_k)^T\varphi(x_l)$
- w and $\varphi(x_k)$ are not calculated

346

continued...

The Non-linear SVM Classifier

- The nonlinear SVM classifier becomes:

$$y(x) = sign[\sum_{k=1}^{N} \alpha_k y_k K(x, x_k) + b]$$

with α_k positive real constants, b real constant
- Non-zero α_k are support vectors

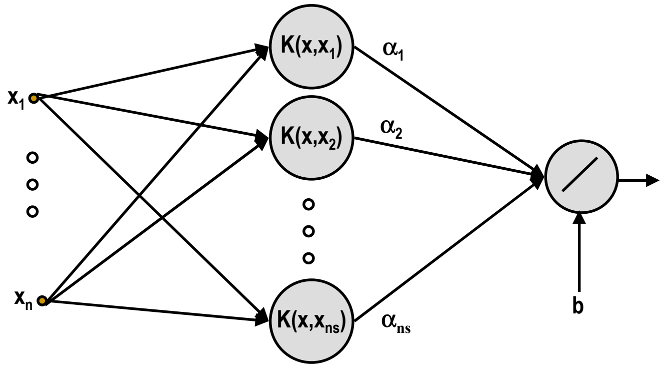
347

Kernel Functions

- $K(x,x_k)=x_k^T x$ (linear SVM)
- $K(x,x_k)=(x_k^T x+1)^d$ (polynomial SVM of degree d)
- $K(x,x_k)=\exp\{-||x-x_k||^2/\sigma^2\}$ (RBF SVM)
- $K(x,x_k)=\tanh(\kappa x_k^T x+\theta)$ (MLP SVM)
- The Mercer condition holds for all σ values in the RBF case but not for all κ and θ values in the MLP case.
- For the RBF and MLP kernels, the number of support vectors corresponds to the number of hidden neurons.

348

Neural Network Interpretation of SVM Classifier



Number of hidden neurons determined automatically

349

The Polynomial Kernel

- Suppose $K(x,y)=(x^T.y)^2$ (x and y two-dimensional vectors)
- Take a mapping to a three-dimensional space

$$\varphi(x) = \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{pmatrix} \quad \varphi(y) = \begin{pmatrix} y_1^2 \\ \sqrt{2}y_1y_2 \\ y_2^2 \end{pmatrix}$$

- Hence

$$\varphi(x)^T.\varphi(y) = x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2 = \left(\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right)^2 = K(x,y)$$

- Note that the mapping is not unique, for example, mapping to four-dimensional space.

$$\varphi(x) = \begin{pmatrix} x_1^2 \\ x_1x_2 \\ x_1x_2 \\ x_2^2 \end{pmatrix}$$

350

Tuning the Hyperparameters

1. Set aside two-thirds of the data for the training/validation set and the remaining one-third for testing.
2. Starting from i=0, perform 10 fold cross-validation on the training/validation set for each (σ,C) combination from the initial candidate tuning sets σ_0={0.5, 5, 10, 15, 25, 50, 100, 250, 500 } and C_0={0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100, 500}.
3. Choose optimal (σ,C) from the tuning sets σ_0 and C_0 by looking at the best cross-validation performance for each (σ,C) combination.
4. If $i=i_{max}$, go to step 5; else i=i+1, construct a locally refined grid σ_i x C_i around the optimal hyperparameters (σ,C) and go to step 3.
5. Construct the SVM classifier using the total training/validation set for the optimal choice of the tuned hyperparameters (σ,C).
6. Assess the test set accuracy by means of the independent test set.

351

Benchmarking Study

- Studied both SVMs and LS-SVMs
- 10 publicly available binary classification data sets and 10 publicly available multiclass classification data sets
- Various domains: medicine, physics, artificial, credit scoring, sociology, ...
- Cross-validation based grid search mechanism to tune the hyperparameters (cf. supra, i_{max}=3)
- RBF kernels, Linear kernels and Polynomial kernels (d=2,...,10)
- Minimum Output Coding and One versus One coding

352

continued...

Benchmarking Study

- Compared their performance with:
  - Linear discriminant analysis, quadratic discriminant analysis, logistic regression, C4.5, oneR, 1-nearest neighbor, 10-nearest neighbor, Naive Bayes (with and without kernel approximation), majority rule
- Performance was compared using paired t-test
- Average rankings were computed and compared using sign test
- 4 months of computer time + 2 additional months

353

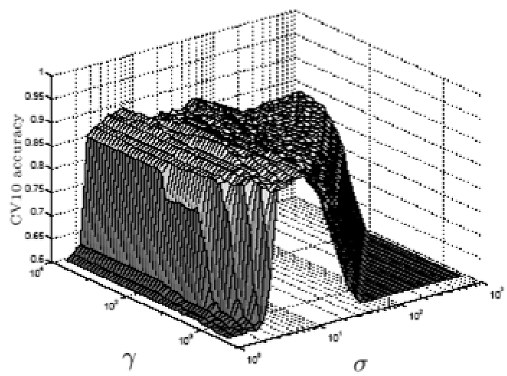
Characteristics of Data Sets

	acr	bld	gcr	hea	ion	pid	snr	ttt	wbc	adu
N_{CV}	460	230	666	180	234	512	138	638	455	33000
N_{test}	230	115	334	90	117	256	70	320	228	12222
N	690	345	1000	270	351	768	208	958	683	45222
n_{num}	6	6	7	7	33	8	60	0	9	6
n_{cat}	8	0	13	6	0	0	0	9	0	8
n	14	6	20	13	33	8	60	9	9	14

	bal	cmc	ims	iri	led	thy	usp	veh	wav	win
N_{CV}	416	982	1540	100	2000	4800	6000	564	2400	118
N_{test}	209	491	770	50	1000	2400	3298	282	1200	60
N	625	1473	2310	150	3000	7200	9298	846	3600	178
n_{num}	4	2	18	4	0	6	256	18	19	13
n_{cat}	0	7	0	0	7	15	0	0	0	0
n	4	9	18	4	7	21	256	18	19	13
M	3	3	7	3	10	3	10	4	3	3
L_{MOC}	2	2	3	2	4	2	4	2	2	2
L_{1vs1}	3	3	21	3	45	3	45	6	2	3

354

Flat Maximum Effect



355

Performance for Binary Classification Data

	acr	bld	gcr	hea	ion	pid	snr	ttt	wbc	adu	AA	AR	P_{ST}
N_{test}	230	115	334	90	117	256	70	320	228	12222			
n	14	6	20	13	33	8	60	9	9	14			
RBF LS-SVM	**87.0**(2.1)	**70.2**(4.1)	**76.3**(1.4)	**84.7**(4.8)	**96.0**(2.1)	**76.8**(1.7)	73.1(4.2)	99.0(0.3)	96.4(1.0)	84.7(0.3)	**84.4**	**3.5**	**0.727**
RBF LS-SVM$_F$	**86.4**(1.9)	65.1(2.9)	70.8(2.4)	83.2(5.0)	93.4(2.7)	72.9(2.0)	73.6(4.6)	97.9(0.7)	96.8(0.7)	77.6(1.3)	81.8	8.8	**0.109**
Lin LS-SVM	**86.8**(2.2)	65.6(3.2)	75.4(2.3)	**84.9**(4.5)	87.9(2.0)	76.8(1.8)	72.6(3.7)	66.8(3.9)	95.8(1.0)	81.8(0.3)	79.4	7.7	**0.109**
Lin LS-SVM$_F$	**86.5**(2.1)	61.8(3.3)	68.6(2.3)	82.8(4.4)	85.0(3.5)	73.1(1.7)	73.3(3.4)	57.6(1.9)	**96.9**(0.7)	71.3(0.3)	75.7	12.1	**0.109**
Pol LS-SVM	**86.5**(2.2)	**70.4**(3.7)	**76.3**(1.4)	**83.7**(3.9)	91.0(2.5)	**77.0**(1.8)	**76.9**(4.7)	**99.5**(0.5)	96.4(0.9)	84.6(0.3)	**84.2**	**4.1**	**0.727**
Pol LS-SVM$_F$	**86.6**(2.2)	65.3(2.9)	70.3(2.3)	82.4(4.6)	91.7(2.6)	73.0(1.8)	**77.3**(2.6)	98.1(0.8)	**96.9**(0.7)	77.9(0.2)	82.0	**8.2**	**0.344**
RBF SVM	86.3(1.8)	**70.4**(3.2)	**75.9**(1.4)	**84.7**(4.8)	95.4(1.7)	**77.3**(2.2)	**75.0**(6.6)	98.6(0.5)	96.4(1.0)	84.4(0.3)	**84.4**	**4.0**	**1.000**
Lin SVM	**86.7**(2.4)	67.7(2.6)	75.4(1.7)	83.2(4.2)	87.1(3.4)	77.0(2.4)	74.1(4.2)	66.2(3.6)	96.3(1.0)	83.9(0.2)	79.8	7.5	0.021
LDA	85.9(2.2)	65.4(3.2)	**75.9**(2.0)	**83.9**(4.3)	87.1(2.3)	76.7(2.0)	67.9(4.9)	68.0(3.0)	95.6(1.1)	82.2(0.3)	78.9	9.6	0.004
QDA	80.1(1.9)	62.2(3.6)	72.5(1.4)	78.4(4.0)	90.6(2.2)	74.2(3.3)	53.6(7.4)	75.1(4.0)	94.5(0.6)	80.7(0.3)	76.2	12.6	0.002
Logit	**86.8**(2.4)	66.3(3.1)	**76.3**(2.1)	82.9(4.0)	86.2(3.5)	**77.2**(1.8)	68.4(5.2)	68.3(2.9)	96.1(1.0)	83.7(0.2)	79.2	7.8	**0.109**
C4.5	85.5(2.1)	63.1(3.8)	71.4(2.0)	78.0(4.2)	90.6(2.2)	73.5(3.0)	72.1(2.5)	84.2(1.6)	94.7(1.0)	**85.6**(0.3)	79.9	10.2	0.021
oneR	85.4(2.1)	56.3(4.4)	66.0(3.0)	71.7(3.6)	83.6(4.8)	71.3(2.7)	62.6(5.5)	70.7(1.5)	91.8(1.4)	80.4(0.3)	74.0	15.5	0.002
IB1	81.1(1.9)	61.3(6.2)	69.3(2.6)	74.3(4.2)	87.2(2.8)	69.6(2.4)	**77.7**(4.4)	82.3(3.3)	95.3(1.1)	78.9(0.2)	77.7	12.5	0.021
IB10	**86.4**(1.3)	60.5(4.4)	72.6(1.7)	80.0(4.3)	85.9(2.5)	73.6(2.4)	69.4(4.3)	94.8(2.0)	96.4(1.2)	82.7(0.3)	80.2	10.4	0.039
NB$_k$	81.4(1.9)	63.7(4.5)	74.7(2.1)	83.9(4.5)	92.1(2.5)	75.5(1.7)	71.6(3.5)	71.7(3.1)	**97.1**(0.9)	84.8(0.2)	79.7	7.3	**0.109**
NB$_n$	76.9(1.7)	56.0(6.9)	74.6(2.8)	**83.8**(4.5)	82.8(3.8)	75.1(2.1)	66.6(3.2)	71.7(3.1)	95.5(0.5)	82.7(0.2)	76.6	12.3	0.002
Maj. Rule	56.2(2.0)	56.5(3.1)	69.7(2.3)	56.3(3.8)	64.4(2.9)	66.8(2.1)	54.4(4.7)	66.2(3.6)	66.2(2.4)	75.3(0.3)	63.2	17.1	0.002

356

Performance for Multiclass Classification Data

LS-SVM	bal	cmc	ims	iri	led	thy	usp	veh	wav	win	AA	AR	P_{ST}
N_{test}	209	491	770	50	1000	2400	3298	282	1200	60			
n	4	9	18	4	7	21	256	18	19	13			
RBF LS-SVM (MOC)	92.7(1.0)	**54.1**(1.8)	95.5(0.6)	96.6(2.8)	70.8(1.4)	96.6(0.4)	95.3(0.5)	81.9(2.6)	**99.8**(0.2)	**98.7**(1.3)	**88.2**	**7.1**	**0.344**
RBF LS-SVM$_F$ (MOC)	86.8(2.4)	43.5(2.6)	69.6(3.2)	**98.4**(2.1)	36.1(2.4)	22.0(4.7)	86.5(1.0)	66.5(6.1)	99.5(0.2)	93.2(3.4)	70.2	17.8	**0.109**
Lin LS-SVM (MOC)	90.4(0.8)	46.9(3.0)	72.1(1.2)	89.6(5.6)	52.1(2.2)	93.2(0.6)	76.5(0.6)	69.4(2.3)	90.4(1.1)	**97.3**(2.0)	77.8	17.8	0.002
Lin LS-SVM$_F$ (MOC)	86.6(1.7)	42.7(2.0)	69.8(1.2)	77.0(3.8)	35.1(2.6)	54.1(1.3)	58.2(0.9)	69.1(2.0)	55.7(1.3)	85.5(5.1)	63.4	22.4	0.002
Pol LS-SVM (MOC)	94.0(0.8)	53.5(2.3)	87.2(2.6)	**96.4**(3.7)	70.9(1.5)	94.7(0.2)	95.0(0.8)	81.8(1.2)	99.6(0.3)	**97.8**(1.9)	87.1	9.8	**0.109**
Pol LS-SVM$_F$ (MOC)	93.2(1.9)	47.4(1.6)	86.2(3.2)	96.0(3.7)	67.7(0.8)	69.9(2.8)	87.2(0.9)	81.9(1.3)	96.1(0.7)	92.2(3.2)	81.8	15.7	0.002
RBF LS-SVM (1vs1)	94.2(2.2)	**55.7**(2.2)	**96.5**(0.5)	**97.6**(2.3)	**74.1**(1.3)	96.8(0.3)	94.8(2.5)	83.6(1.3)	99.3(0.4)	**98.2**(1.8)	**89.1**	**5.9**	**1.000**
RBF LS-SVM$_F$ (1vs1)	71.4(15.5)	42.7(3.7)	46.2(6.5)	79.8(10.3)	58.9(8.5)	92.6(0.2)	30.7(2.4)	24.9(2.5)	97.3(1.7)	67.3(14.6)	61.2	22.3	0.002
Lin LS-SVM (1vs1)	87.8(2.2)	50.8(2.4)	93.4(1.0)	**98.4**(1.8)	**74.5**(1.0)	93.2(0.3)	95.4(0.3)	79.8(2.1)	97.6(0.9)	**98.3**(2.5)	**86.9**	**9.7**	**0.754**
Lin LS-SVM$_F$ (1vs1)	87.7(1.8)	49.6(1.8)	93.4(0.9)	**98.6**(1.3)	**74.5**(1.0)	74.9(0.8)	95.3(0.3)	79.8(2.2)	98.2(0.6)	**97.7**(1.8)	85.0	**11.1**	**0.344**
Pol LS-SVM (1vs1)	95.4(1.0)	53.2(2.2)	95.2(0.6)	96.8(2.3)	72.8(2.6)	88.8(14.6)	**96.0**(2.1)	82.8(1.8)	99.0(0.4)	**99.0**(1.4)	**87.9**	**8.9**	**0.344**
Pol LS-SVM$_F$ (1vs1)	56.5(16.7)	41.8(1.8)	30.1(3.8)	71.4(12.4)	32.6(10.9)	92.6(0.7)	95.8(1.7)	20.3(6.7)	77.5(4.9)	82.3(12.2)	60.1	21.9	0.021
RBF SVM (MOC)	**99.2**(0.5)	51.0(1.4)	94.9(0.9)	96.6(3.4)	69.9(1.0)	96.6(0.2)	95.5(0.4)	77.6(1.7)	**99.7**(0.1)	**97.8**(2.1)	**87.9**	**8.6**	**0.344**
Lin SVM (MOC)	98.3(1.2)	45.8(1.6)	74.1(1.4)	**95.0**(10.5)	50.9(3.2)	92.5(0.3)	81.9(0.3)	70.3(2.5)	99.2(0.2)	97.3(2.6)	80.5	16.1	0.021
RBF SVM (1vs1)	98.3(1.2)	**54.7**(2.4)	96.0(0.4)	97.0(3.0)	64.6(5.6)	98.3(0.3)	**97.2**(0.2)	83.8(1.6)	99.6(0.2)	**96.8**(5.7)	**88.6**	**6.5**	**1.000**
Lin SVM (1vs1)	91.0(2.3)	50.8(1.6)	95.2(0.7)	**98.0**(1.9)	**74.4**(1.2)	97.1(0.3)	95.1(0.3)	78.1(2.4)	99.6(0.2)	**98.3**(3.1)	**87.8**	**7.3**	**0.754**
LDA	86.9(2.1)	51.8(2.2)	91.2(1.1)	**98.6**(1.0)	73.7(0.8)	93.7(0.3)	91.5(0.5)	77.4(2.7)	94.6(1.2)	**98.7**(1.5)	85.8	11.0	**0.109**
QDA	90.5(1.1)	50.6(2.1)	81.8(9.6)	**98.2**(1.8)	**73.6**(1.1)	93.4(0.3)	74.7(0.7)	**84.8**(1.5)	60.9(9.5)	**99.2**(1.2)	**80.8**	11.8	**0.344**
Logit	88.5(2.0)	51.6(2.4)	95.4(0.6)	**97.0**(3.9)	**73.9**(1.0)	95.8(0.5)	91.5(0.5)	78.3(2.3)	**99.9**(0.1)	95.0(3.2)	86.7	**9.8**	0.021
C4.5	66.0(3.6)	50.9(1.7)	96.1(0.7)	96.0(3.1)	73.6(1.3)	**99.7**(0.1)	88.7(0.3)	71.1(2.6)	**99.8**(0.1)	87.0(5.0)	82.9	**11.8**	**0.109**
oneR	59.5(3.1)	43.2(3.5)	62.9(2.4)	95.2(2.5)	17.8(0.8)	96.3(0.5)	32.9(1.1)	52.9(1.9)	67.4(1.1)	76.2(4.6)	60.4	21.6	0.002
IB1	81.5(2.7)	43.3(1.1)	**96.8**(0.6)	95.6(3.6)	**74.0**(1.3)	92.2(0.4)	97.0(0.2)	70.1(2.9)	99.7(0.1)	95.2(2.0)	**84.5**	12.9	**0.344**
IB10	83.6(2.3)	44.3(2.4)	94.3(0.7)	97.2(1.9)	**74.2**(1.3)	93.7(0.3)	96.1(0.3)	67.1(2.1)	99.4(0.1)	96.2(1.9)	84.6	**12.4**	**0.344**
NB_k	89.9(2.0)	51.2(2.3)	84.9(1.4)	**97.0**(2.5)	**74.0**(1.2)	96.4(0.2)	79.3(0.9)	60.0(2.3)	99.5(0.1)	**97.7**(1.6)	83.0	12.2	0.021
NB_n	89.9(2.0)	48.9(1.8)	80.1(1.0)	**97.2**(2.7)	**74.0**(1.2)	95.5(0.4)	78.2(0.6)	44.9(2.8)	99.5(0.1)	97.5(1.8)	80.6	13.6	0.021
Maj. Rule	48.7(2.3)	43.2(1.8)	15.5(0.6)	38.6(2.8)	11.4(0.0)	92.5(0.3)	16.8(0.4)	27.7(1.5)	34.2(0.8)	39.7(2.8)	36.8	24.8	0.002

357

Conclusions

- RBF SVMs and RBF LS-SVMs yield very good classification performances compared to the other algorithms
- For the multiclass case, the One versus One coding scheme yielded better performance than the minimum output coding scheme
- Simple classification algorithms (for example, linear discriminant analysis and logistic regression) also yield satisfactory results
- Most data sets are only weakly non-linear
- But: importance of marginal performance benefits (for example, credit scoring)

358

Support Vector Machines in SAS

- PROC DMDB (data mining database)
- A data mining database is a SAS data set that is designed to optimize the performance of the analytical nodes by reducing the number of passes through the data required by the analytical engine
- Also generates catalog of metadata information
- The catalog contains important information (for example, range of variables, number of missing values of each variable, moments of variables)
- Writes interval scaled variables and class variables in a specific form
- In SAS Enterprise Miner, if a node requires a DMDB, then PROC DMDB is automatically run.

359

continued...

Support Vector Machines in SAS

```
PROC SVM options;      Required statement
VAR variables;
TARGET variables;
FREQ variables;
WEIGHT variables;      Optional statements
C values;
DECISION options;
```

SVM Options

DATA= *SASdataset*	Specifies an input data set generated by PROC DMDB.
DMDBCAT= SAScatalog	Specifies an input catalog of meta information generated by PROC DMDB associated with the *SASdataset.*

360

continued...

Support Vector Machines in SAS

TESTDATA= *SASdataset*	Specifies a test data set (not generated by PROC DMDB).
OUT= *SASdataset*	■ Data set with predictions relating to the data specified in the DATA=*SASdataset* option. ■ _Y_: observed value of target as used for modeling, _P_: predicted value, _R_: residual.
TESTOUT= *SASdataset*	Data set with predictions relating to the data specified in the TESTDATA=*SASdataset option.*

361

continued...

Support Vector Machines in SAS

C	Specifies the regularization parameter of the objective function.
KERNEL= LINEAR/ POLYNOM/ RBF/ SIGMOID	
K_PAR=r	Specifies the first parameter of the kernel function, that is, the degree of the polynomial or the scale of the sigmoid function.
K_PAR2=r	Specifies the second parameter of the kernel function.

362

continued...

Support Vector Machines in SAS

NOMONITOR	Supresses the output of the status monitor indicating the progress made in the computations.
PALL	Prints all output except that specified by PPRED.
PPRED	Print observational statistics (observed and predicted values and residuals).

363

continued...

Support Vector Machines in SAS

C statement	A list of C values can be specified. For each C>0 value, ordered from small to large, the SVM training is performed. At the end the best fitted result, normally that with the largest C value, is scored. If C is too large, this solution is overfitted. Note the C statement overrides the C specification in the proc statement.
TARGET statement	Specifies the target. If a target is specified in the proc dmdb run, it must not be specified in the proc svm call.
VAR statement	Specifies the independent variables.

364

Example

```
proc dmdb data=neural.dmlspir batch out=dmdb dmdbcat=meta;
var x y;
class c(desc);
target c;
run;

proc svm data=dmdb dmdbcat=meta nomonitor out=spiralout kernel=RBF
K_PAR=0.5 C=100;
var x y;
target c;
run;
```

365

More Information on Support Vector Machines

- An introduction to support vector machines, Nello Cristianini and John Shawe-Taylor, Cambridge University Press, 2000
- Kernel Methods for Pattern Analysis, John Shawe-Taylor and Nello Cristianini, Cambridge University Press, forthcoming 2004
- The nature of statistical learning theory, Vladimir Vapnik, Springer, 1995
- Learning with Kernels, Bernard Schölkopf and Alex Smola, MIT Press, 2002
- www.kernel-machines.org

366

1.15 Survival Analysis

Survival Analysis for Credit Scoring

- Traditionally, credit scoring models aim at distinguishing good customers from bad customers
- However, the timing of the problem is also important!
- The advantages of having models that estimate when customers default are these (from Banasik, Crook, and Thomas 1999; Thomas, Edelman and Crook 2002):
  - the ability to compute the profitability over a customer's lifetime and perform profit scoring
  - these models may provide the bank with an estimate of the default levels over time which is useful for debt provisioning
  - the estimates might help to decide upon the term of the loan
  - changes in economic conditions can be incorporated more easily

368

continued...

Survival Analysis for Credit Scoring

- Basel II defines default as 90 days overdue or not recoverable
- Some regulators require more strict definitions (for example, 180 days overdue)
- Some current scoring systems use 60, 90, 120, 180 days
- How to recalculate PD(180) from PD(90)?
  - Build a separate scorecard for each horizon
  - Use survival analysis to get estimates for all PD(x)
- Moody's KMV RiskCalc® v3.1 Model

369

Literature Overview: General References

- Allison, P. D. 1995. *Survival Analysis Using SAS: A Practical Guide.* Cary, NC: SAS Institute Inc.
- Cantor, A. 1997.*SAS Survival Analysis Techniques for Medical Research.* Cary, NC: SAS Institute Inc.
- Cox, D. R. and D. Oakes. 1984. *Analysis of Survival Data (Monographs on Statistics and Applied Probability).* London: Chapman and Hall/CRC.
- Kalbfleisch, J. D. and R. L. Prentice. 1980. *The Statistical Analysis of Failure Time Data (Wiley Series in Probability and Statistics).* New York: Wiley.
- Kleinbaum, D. G. 1996. *Survival Analysis: A Self-Learning Text.* New York: Springer.

370

Literature Overview: Credit Scoring References

- Baesens B., T. Van Gestel, M. Stepanova, and J. Vanthienen. 2003. "Neural Network Survival Analysis for Personal Loan Data," *Proceedings of the Eighth Conference on Credit Scoring and Credit Control (CSCCVII'2003).* Edinburgh, Scotland.
- Banasik J., J. N. Crook., and L.C. Thomas. 1999. "Not If but When Will Borrowers Default." *Journal of the Operational Research Society* 50:1185–1190.
- Narain B. 1992. "Survival Analysis and the Credit Granting Decision." In *Credit Scoring and Credit Control,* eds. L.C. Thomas, J.N. Crook, and D.B. Edelman, 109-121, Oxford University Press.
- Stepanova M. and L.C Thomas. 2001. "PHAB Scores: Proportional Hazards Analysis Behavioural Scores." *Journal of the Operational Research* Society 52(9): 1007–1016.
- Stepanova M. and L.C. Thomas. 2002. "Survival Analysis Methods for Personal Loan Data." *Operations Research* 50(2):277–289.

371

Basic Concepts of Survival Analysis

- Aim is to study occurance and timing of events
- Event is a qualitative change that can be situated in time
  - For example, death, marriage, promotion, default, early repayment
- Originated from medicine (study of deaths)
- Two typical problems:
  - Censoring
  - Time-dependent covariates

372

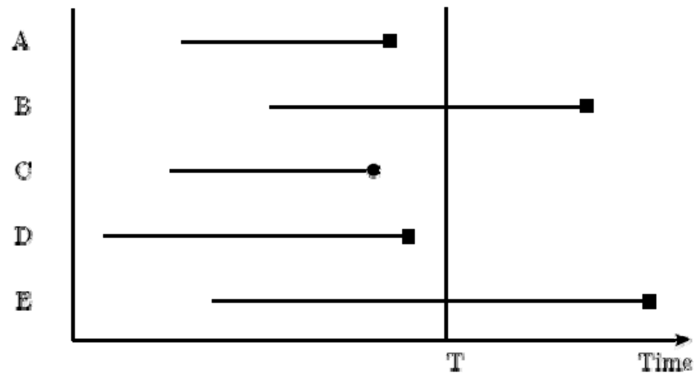
Censoring

- Left-censoring versus right-censoring.
- An observation on a variable T is right censored, if all you know about T is that it is greater than some value c.
  - For example, suppose T is person's age at death, and you only know that T is > 50, hence, the observation is right censored at age 50.
- An observation on a variable T is left censored if all you know about T is that it is smaller than some value c.
  - For example, study menarche (start of menstruation) and start by girls of 12 years whereas some already begun.
  - Less common
- Interval censoring
  - You know that a < T < b.

373

continued...

Censoring



374

Survival Distributions

- Cumulative distribution: F(t)=P(T≤t)
- Survival function:

$$S(t) = 1 - F(t) = P(T > t) = \int_t^\infty f(u)du$$

with probability density function f(u)

- S(t) is monotone decreasing with S(0)=1 and S(∞)=0
- The probability density function f(t) is defined as:

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}$$

375

continued...

Survival Distributions

- The hazard function is defined as follows

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

- The hazard function tries to quantify the instantaneous risk that an event will occur at time t, given that the individual has survived to t
- Probability of event at time t is 0, hence look at interval between t and $t+\Delta t$
- Only consider individuals surviving to time t, because others have already died and are no longer at risk
- The probability is a non-decreasing function of Δt, hence, divide by Δt
- We want the probability at exactly time t, hence, take limit for $\Delta t \to 0$

Can be thought of as number of events per interval of time

376

The Hazard Function

- Remember, a probability distribution can also be defined as follows:

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}$$

- The hazard function is sometimes described as a conditional density.
- The survivor function S(t), the probability density function f(t), and the hazard function h(t) are mathematically equivalent ways of describing a continuous probability distribution.
- The following relationships hold:

$$h(t) = \frac{f(t)}{S(t)} \qquad h(t) = -\frac{d \log S(t)}{dt}$$

$$S(t) = \exp\{-\int_0^t h(u)du\}$$

377

Kaplan Meier Method

- Also known as the product-limit estimator.
- Non-parametric maximum likelihood estimator for S(t).
- If no censoring, the KM estimator $\hat{S}(t)$ is just the sample proportion with event times greater than t.
- If censoring, start with ordering the event times in ascending order $t_1 < t_2 < ... < t_k$. At each time t_j, there are n_j individuals who are at risk of the event. At risk means that they have not undergone the event, nor have they been censored prior to t_j. Let d_j be the number of individuals who die at t_j.
- The KM estimator is then defined as follows:

$$\hat{S}(t) = \prod_{j:t_j \leq t} [1 - \frac{d_j}{n_j}] = \hat{S}(t-1).(1 - \frac{d_t}{n_t})$$

for $t_1 \leq t \leq t_k$
- The term between brackets is the conditional probability of surviving to time t_{j+1}, given that the subject has survived to time t_j.

378

Kaplan Meier Analysis in SAS

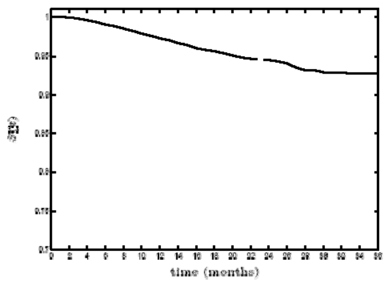
```
proc lifetest data=credit;
time dur*status(0);
run;
```

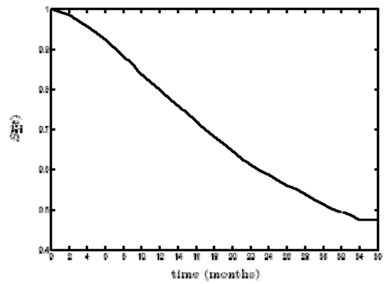
- PROC LIFETEST
- KM estimator is default (to be explicit: use METHOD=KM)
- The **dur** variable is the time of the event
- The **status** variable is the censoring indicator; value of 0 corresponds to censored observations

379

continued...

Kaplan Meier Analysis





380

continued...

Kaplan Meier Analysis

Comparing survival curves

- H_o: the survival curves are statistically the same
- H_a: the survival curves are statistically different
- log-rank test (also known as the Mantel-Haenzel test), the Wilcoxon test, and the likelihood-ratio statistic

If many unique event times, use life-table (also known as actuarial) method to group event times into intervals.

KM estimator does not account for covariates.

Test for the effect of covariates.

381

Parametric Survival Analysis Models: Exponentially Distributed Event Times

- The probability function f(t) is

$$f(t) = \lambda e^{-\lambda t}$$

- The survival function then becomes

$$S(t) = e^{-\lambda t}$$

- The hazard is constant over time

$$h(t) = \lambda$$

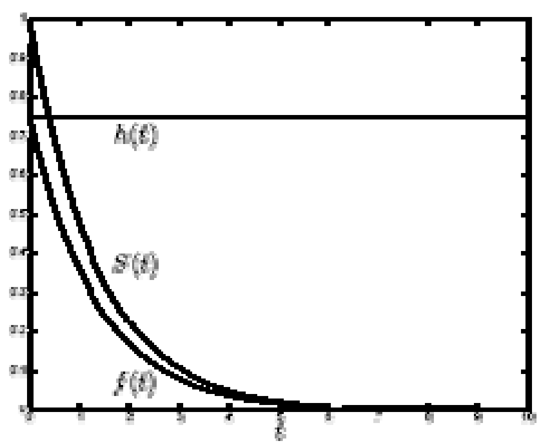
- When covariates are present

$$\log h(t) = \mu + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

382

continued...

Parametric Survival Analysis Models: Exponentially Distributed Event Times



383

Parametric Survival Analysis Models: Weibull Distributed Event Times

- The probability function f(t) is

$$f(t) = \kappa\rho(\rho t)^{\kappa-1}\exp[-(\rho t)^{\kappa}]$$

- The survival function then becomes

$$S(t) = \exp[-(\rho t)^{\kappa}]$$

- The hazard is (not constant over time)

$$h(t) = \kappa\rho(\rho t)^{\kappa-1}$$

- When covariates are present:

$$\log\ h(t) = \mu + \alpha\log(t) + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

384

General Parametric Survival Analysis Model

- The general model is

$$\log\ T_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in} + \varepsilon_i$$

with T_i the event time for the i^{t}h individual.
- The log transformation is to ensure that predicted values of T are positive.

$$T_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in} + \varepsilon_i)$$

- Also known as accelerated failure time (AFT) model.
- If no censored data, estimate with ordinary least squares (OLS).
- If censored data, use maximum likelihood estimation.
- PROC LIFEREG

385

Example: Exponential Distribution

- Exponential distribution for t corresponds to constant hazard.

$$T_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in} + \varepsilon_i)$$

$$\log\ h(t) = \beta_0^{\bullet} + \beta_1^{\bullet} x_1 + \beta_2^{\bullet} x_2 + \ldots + \beta_n^{\bullet} x_n$$

- Both models are equivalent; it can be shown mathematically that $\beta_j = -\beta_j^{\bullet}$ for all j.
- If hazard is high (low), survival times are short (long).

386

Distributions for T

Distribution of T	Distribution of e
Weibull	Gumbel distribution (2 par.)
Exponential	Gumbel distribution (1 par.)
Gamma	Log-gamma
Log-logistic	Logistic
Log-normal	normal

387

Maximum Likelihood Estimation

- Maximize the probability of getting the sample at hand
- Two steps
  - Construct the likelihood function
  - Maximize the likelihood function
- Construct the likelihood function as follows:

$$L = \prod_{i=1}^{N} f(t_i)$$

$$L = \prod_{i=1}^{N} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}$$

δ_i is 0 if observation is censored at t_i; δ_i is 1 if observation dies at t_i.

388

Maximum Likelihood Estimation

- For example, consider exponentially distributed failure times (no covariates).

$$L = \prod_{i=1}^{N} [\lambda e^{-\lambda t_i}]^{\delta_i} [e^{-\lambda t_i}]^{1-\delta_i}$$

- If covariates, replace λ by exp{**-β$\mathbf{x}_i$**}
- Take logarithm and maximize using Newton Raphson.

389

Evaluating Model Fit Graphically

- Because $h(t) = -\frac{d \log S(t)}{dt}$, we have

$$-\log(S(t)) = \int_0^t h(u)du$$

- Because of this relationship, the log survivor function is commonly referred to as the cumulative hazard function, denoted as Λ(t).
- Λ(t) can be interpreted as the sum of risks that are faced when going from time 0 to time t.
- If the survival times are exponentially distributed, then the hazard is constant h(t)=λ, hence Λ(t) =λt, and a plot of –log(S(t)) versus t should yield a straight line with an origin at 0.
- If the survival times are Weibull distributed, then a plot of log(-log(S(t))) is a straight line (not through the origin) with a slope of κ.

390

continued...

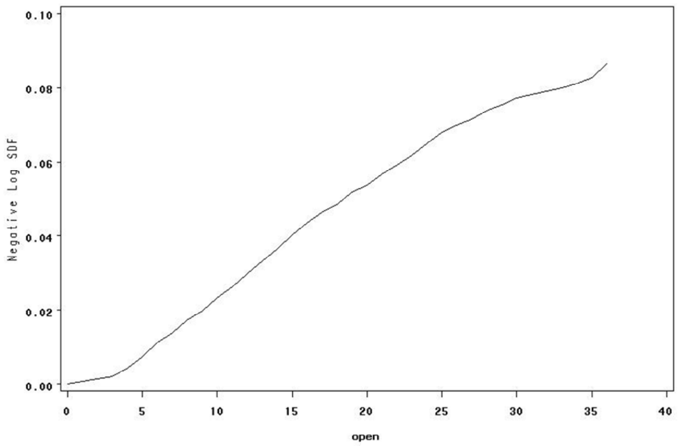
Evaluating Model Fit Graphically

- The plots of –log(S(t)) and log(–log(S(t))) can be asked using PROC LIFETEST with the plots(ls,lls) option
- A similar approach can be followed to evaluate the log-normal and log-logistic distributions (see Allison 1995).

391

continued...

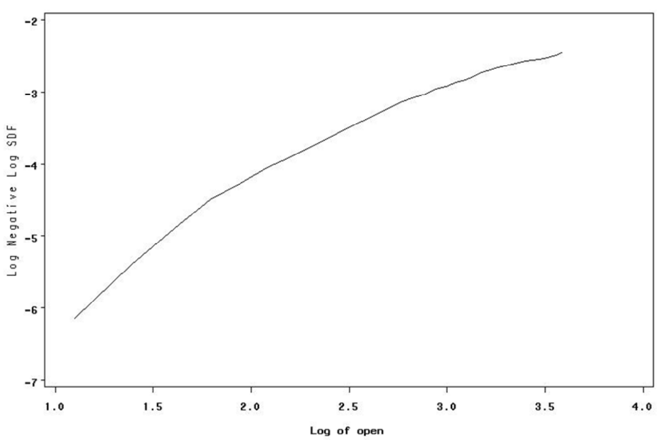
Evaluating Model Fit Graphically



392

continued...

Evaluating Model Fit Graphically



393

Evaluating Model Fit Statistically

- The likelihood ratio test statistic can be used to compare models if one model is a special case of another (nested models).
- The generalized gamma distribution is defined as follows:

$$f(t) = \frac{\beta}{\Gamma(k).\theta}\left(\frac{t}{\theta}\right)^{k\beta-1} e^{-(\frac{t}{\theta})^{\beta}}$$

- Suppose $\sigma = \frac{1}{\beta\sqrt{k}}$ and $\delta = \frac{1}{\sqrt{k}}$, then the Weibull, exponential, standard gamma, and log-normal models are all special versions of the generalized gamma model as follows:

  σ=δ: standard gamma

  δ=1: Weibull

  σ=1, δ=1: exponential

  δ=0: lognormal

394

continued...

Evaluating Model Fit Statistically

- Let L_{full} be the likelihood of the full model and L_{red} be the likelihood of the reduced (specialized) model.
- The chi-squared test statistic can then be computed as twice the positive difference in the log likelihood for the two models.
- The degrees of freedom corresponds to the number of reduced parameters.
  - Exponential versus Weibull: 1 degree of freedom
  - Exponential versus standard gamma: 1 degree of freedom
  - Exponential versus generalized gamma: 2 degrees of freedom
  - Weibull versus generalized gamma: 1 degree of freedom
  - Log-normal versus generalized gamma: 1 degree of freedom
  - Standard gamma versus generalized gamma: 1 degree of freedom

395

PROC LIFEREG for Parametric Survival Analysis

```
proc lifereg data=creditsurv;
class custgend freqpaid homephon loantype marstat homeowns;
model open*censor(0)=age amount curradd curremp custgend depchild
freqpaid homephon insprem loantype marstat term homeowns
/dist=exponential;
run;
```

- The option DIST= exponential / weibull / lognormal / gamma / loglogistic.
- The CLASS statement automatically creates dummy variables for categorical variables (not in PROC PHREG).

396

The Proportional Hazards Model

- Basic model:

$$h(t,\mathbf{x}_i) = h_0(t)\exp\{\beta_1 x_{i1} + \ldots + \beta_n x_{in}\} = h_0(t)\exp\{\boldsymbol{\beta}^T x_i\}$$

- Hazard of individual i at time t is product of:
  - A baseline hazard function $h_0(t)$ that is left unspecified (except that it can't be negative)
  - A linear function of a set of fixed covariates which is exponentiated.
- $h_0(t)$ can be considered as hazard for individual with all covariates equal to 0.
- If x_i increases with 1, then the hazards for all t increase with $\exp(\beta_i)$ which is called the hazard ratio (HR).
- If $\beta > 0$ then HR > 1, $\beta < 0$ then HR < 1; $\beta = 0$ then HR = 1.
- David Cox, Regression Models and Life Tables, Journal of the Royal Statistical Society, Series B, 1972.
- Very popular.

397

continued...

The Proportional Hazards Model

- Taking logarithms yields

$$\log h(t, \mathbf{x}_i) = \alpha(t) + \beta_1 x_{i1} + \ldots + \beta_n x_{in}$$

with $\alpha(t)=\log h_0(t)$

- Hence, if $\alpha(t)=\alpha$, we get the exponential model, if $\alpha(t)=\alpha\log(t)$, we get the Weibull model
- However, no need to specify $\alpha(t)$ (semiparametric model)
- Taking the ratios of the hazards for individuals i and j

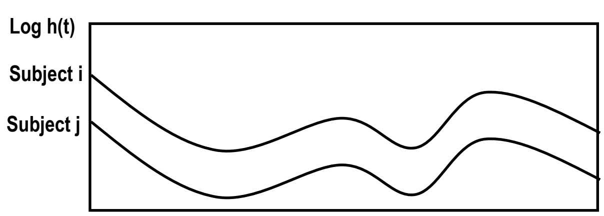
$$\frac{h_i(t)}{h_j(t)} = \exp\{\beta_1(x_{i1} - x_{j1}) + \ldots + \beta_n(x_{i1} - x_{j1})\} = \exp[\boldsymbol{\beta}^T(\mathbf{x}_i - \mathbf{x}_j)]$$

398

continued...

The Proportional Hazards Model

- Hazard of any individual is a fixed proportion of the hazard of any other individual (proportional hazards).
- The subjects most at risk at any one time remain the subjects most at risk at any one other time.



399

Partial Likelihood Estimation

- Suppose n individuals (i=1,...,n)
- For each individual: $\mathbf{x}_i$, t_i, δ_i, t_i is time of event or censoring, δ_i is 1 if uncensored, δ_i is 0 if censored observation
- Start by ranking all events of the non-censored subjects $t_1,\ldots, t_k$
- Given the fact that one subject has event time t_i, the probability that this subject has inputs $\mathbf{x}_j$ is then given by:

$$\frac{h(t_i,\mathbf{x}_j)\Delta t}{\sum_{l\in R(t_i)} h(t_i,\mathbf{x}_l)\Delta t} = \frac{\exp(\boldsymbol{\beta}^T\mathbf{x}_j).h_0(t_i)}{\sum_{l\in R(t_i)} \exp(\boldsymbol{\beta}^T\mathbf{x}_l).h_0(t_i)} = \frac{\exp(\boldsymbol{\beta}^T\mathbf{x}_j)}{\sum_{l\in R(t_i)} \exp(\boldsymbol{\beta}^T\mathbf{x}_l)}$$

R(t_i) represents the subjects that are at risk at t_i

400

continued...

Partial Likelihood Estimation

- The likelihood function then becomes

$$\prod_{j=1}^{k} \frac{\exp(\boldsymbol{\beta}^T\mathbf{x}_j)}{\sum_{l\in R(t_j)} \exp(\boldsymbol{\beta}^T\mathbf{x}_l)}$$

(Note that for ease of notation, we assumed that individual $\mathbf{x}_j$ has event time t_j.)

- The β parameters are then optimized using the Newton-Raphson algorithm.
- Observe how the censored observations do enter the partial likelihood function; they will be included in the risk sets R(t_j) until their censoring time.

401

continued...

Partial Likelihood Estimation

- The β parameters can be estimated without having to specify the baseline hazard $H_0(t)$.
- Partial likelihood estimates are
  - consistent (converge to true values when sample gets larger)
  - asymptotically normal
- Partial likelihood estimates only depend on the ranks of the event times not on the numerical values.
- But important assumption: no tied event times.

402

Partial Likelihood for Tied Event Times

- The exact method (TIES=EXACT in PROC PHREG)
  - Assumes ties occur because of imprecise time measurements
  - Considers all possible orderings of the event times and constructs a likelihood term for each ordering (for example, when 3 ties, 3 possible orderings)
  - Very time-consuming for heavily tied data.
- Approximations
  - The Breslow likelihood (default in PROC PHREG) and the Efron likelihood (TIES=EFRON in PROC PHREG)
- The discrete method (TIES=DISCRETE in PROC PHREG)

403

Tied Event Times: Guidelines

- Allison (1995)
- When no ties, all four methods give identical results
- When few ties, use exact method
- If exact method is too time consuming, use the Efron approximation

404

Estimating Survivor Functions

- Because

  $$S(t,\mathbf{x}) = \exp[-\int_0^t h_0(u)\exp(\boldsymbol{\beta}^T\mathbf{x})du]$$

  we have,

  $$S(t,\mathbf{x}) = S_0(t)^{\exp(\boldsymbol{\beta}^T\mathbf{x})}$$

  with

  $$S_0(t) = \exp(-\int_0^t h_0(u)du) = \exp(-\Lambda_0(t)))$$

- $S_0(t)$ is the baseline survivor function (that is, survivor function for an individual whose covariates are all 0).

405

continued...

Estimating Survivor Functions

- If x_i increases with 1, the survival probabilities are raised to the power $\exp(\beta_i)$ which is the hazard ratio (HR).
  - For example, if the HR for a group with a covariate value of 1 relative to the group having the covariate value of 0 is 2.0, then for any time t, the survival probability of group 1 at t is the square of the corresponding survival probability for group 0.
- Once the β parameters have been estimated, $S_0(t)$ can be estimated using a nonparametric maximum likelihood method.
- Individual survival functions can then be computed.
- Use the BASELINE statement in PROC PHREG.

406

Estimating Survivor Functions in SAS

```
data testset;
 input age amount curradd curremp;
 datalines;
 29 3000 1 2
run;

proc phreg data=creditsurv;
model open*censor(0)=age amount curradd curremp /ties=efron;
baseline out=a covariates=testset survival=s lower=lcl upper=ucl
/nomean;
run;

proc print data=a;
run;
```

407

Time Dependent Covariates

- Covariates that change in value over the course of the observation (for example, behavioral scoring).

$$h(t, \mathbf{x}_i) = h_0(t)\exp\{\boldsymbol{\beta}^T x_i(t)\}$$

- Note that the proportional hazards assumption no longer holds because the time-dependent inputs will change at different rates for different subjects, so the ratios of their hazards cannot remain constant.
- We can also let the β parameters vary over time.

$$h(t, \mathbf{x}_i) = h_0(t)\exp\{\boldsymbol{\beta}^T(t) x_i(t)\}$$

- The partial likelihood estimation method can easily be extended to accomodate these changes in the model formulation.

408

Drawbacks of Statistical Survival Analysis Models

- Functional relationship remains linear or some mild extension thereof.
- Interaction and non-linear terms have to be specified ad hoc.
- Extreme hazards for outlying observations.
- Proportional hazards assumption.

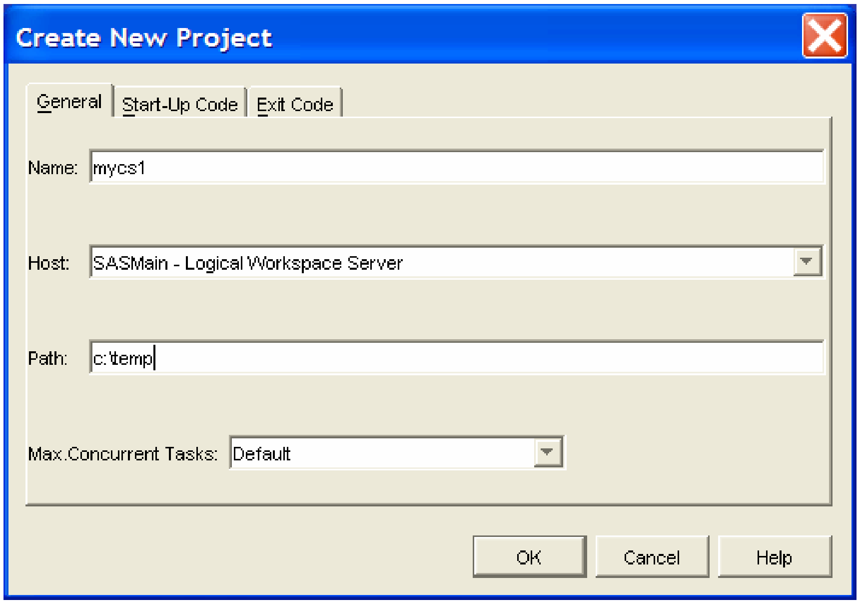
409

Appendix A Exercises

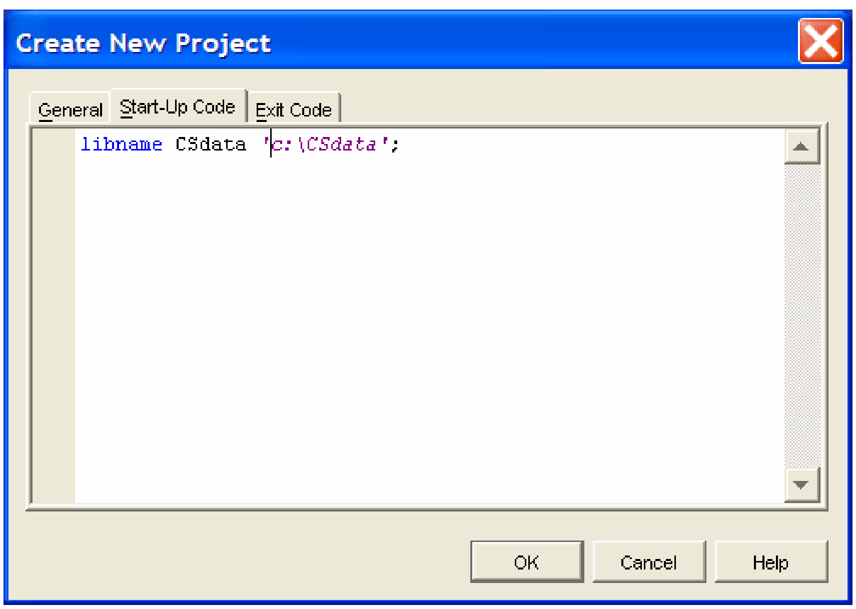
A.1 Exercises

1. Exploring SAS Enterprise Miner

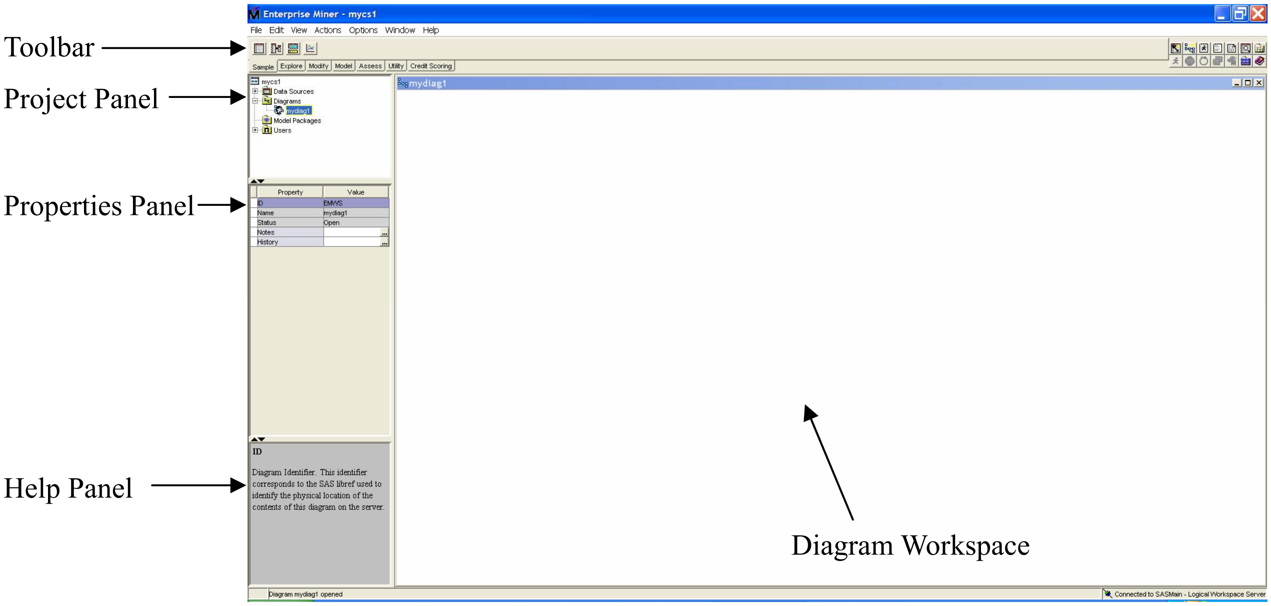
a. Start Enterprise Miner 5.1 and create a new project named **`mycs1`**.. Set the path to **`C:\temp`** (or another directory where you have write access).



b. In the Start-Up code tab, enter **`libname CSdata 'c:\CSdata'`**, where **`c:\CSdata`** is the directory with the data files.



The following window opens:

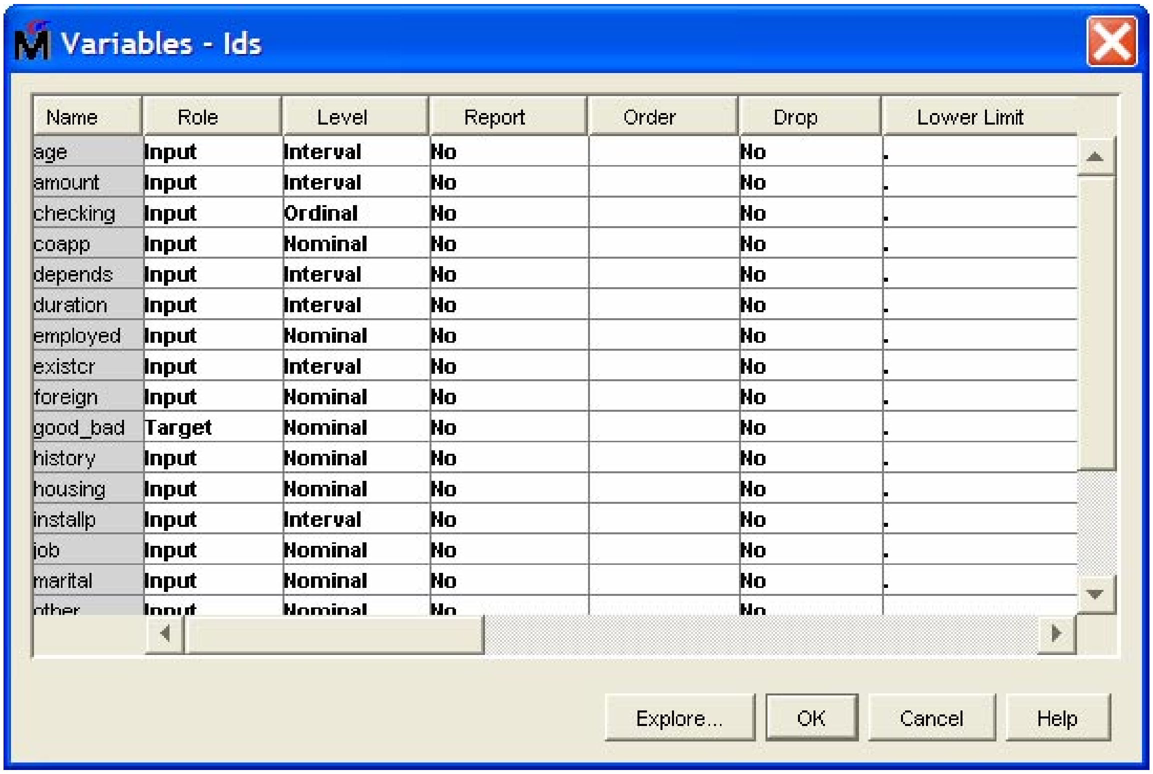


The interface is divided into five components:

- **Toolbar** – The toolbar in SAS Enterprise Miner is a graphical set of node icons and tools that you use to build process flow diagrams in the diagram workspace. To display the text name of any node or tool icon, position your mouse pointer over the icon.
- **Project panel** – Use the Project panel to manage and view data sources, diagrams, results, and project users.
- **Properties panel** – Use the Properties panel to view and edit the settings of data sources, diagrams, nodes, results, and users. Choose **View**⇨ **Property Sheet** ⇨ **Advanced** from the menu to show the Advanced Properties Sheet.
- **Diagram workspace** – Use the diagram workspace to build, edit, run, and save process flow diagrams. This is where you graphically build, order, and sequence the nodes that you use to mine your data and generate reports.
- **Help panel** – The Help panel displays a short description of the property that you select in the Properties panel. Extended help can be found in the **Help Topics** selection from the **Help** main menu.

c. Right-click on the **Data Sources** folder in the Project Panel (or select **File** ⇨ **New** ⇨ **Data Sources**) to open the Data Source Wizard. Create a new data source for the DMAGECR data set, which is located in the `CSdata` library. Set the level of the following variables to ordinal: CHECKING and SAVINGS. Set the level of the following variables to nominal: COAPP, EMPLOYED, FOREIGN, HISTORY, HOUSING, JOB MARITAL, OTHER, PROPERTY, PURPOSE, RESIDENT, and TELEPHON. Accept the Interval (=continuous) level suggested for the other variables.

d. Create a new diagram **`mydiag1`**. Add an Input data node to the diagram workspace. Set its data source property in the property sheet left below to **`DMAGECR`**. Open the Variables window in the property sheet and set the role of the variable **`good_bad`** to **`target`**. Explore some of the characteristics of the variables using the **Explore** button. Create a pie chart for the variable **`savings`**.



Name	Role	Level	Report	Order	Drop	Lower Limit
age	Input	Interval	No		No	.
amount	Input	Interval	No		No	.
checking	Input	Ordinal	No		No	.
coapp	Input	Nominal	No		No	.
depends	Input	Interval	No		No	.
duration	Input	Interval	No		No	.
employed	Input	Nominal	No		No	.
existcr	Input	Interval	No		No	.
foreign	Input	Nominal	No		No	.
good_bad	Target	Nominal	No		No	.
history	Input	Nominal	No		No	.
housing	Input	Nominal	No		No	.
installp	Input	Interval	No		No	.
job	Input	Nominal	No		No	.
marital	Input	Nominal	No		No	.
other	Input	Nominal	No		No	

e. Add a multiplot node to the diagram workspace and connect it to the Input data component. Right-click the Multiplot node, and select **Run**. After the run has finished, right-click the node again, and select **Results**. Inspect the graphs and output generated by this node.

f. Add a Sample node to the diagram workspace and connect it with the input data component. Set the Sample method to stratify so as to create a stratified sample based on the target variable (**`good_bad`**). Run the node and inspect the generated sample.

g. Add a Filter node to the diagram workspace and connect it with the input data component. Eliminate class values that occur < 2 times for class variables with < 20 different values. Eliminate interval variables that are more than 2.5 standard deviations away from the mean. Run the filter node and inspect the results. Inspect the limits for the interval variables and check which class values were excluded. Check how many observations were removed in total.

h. Add a StatExplore node to the diagram workspace and connect it to the input data component. Run the StatExplore node and inspect the results. A chi-square plot is presented reporting Cramer's V statistic. Cramer's V statistic is computed as follows:

$$\text{Cramer's} \quad \text{V} = \sqrt{\frac{\chi^2}{N \min(r-1, c-1)}}$$

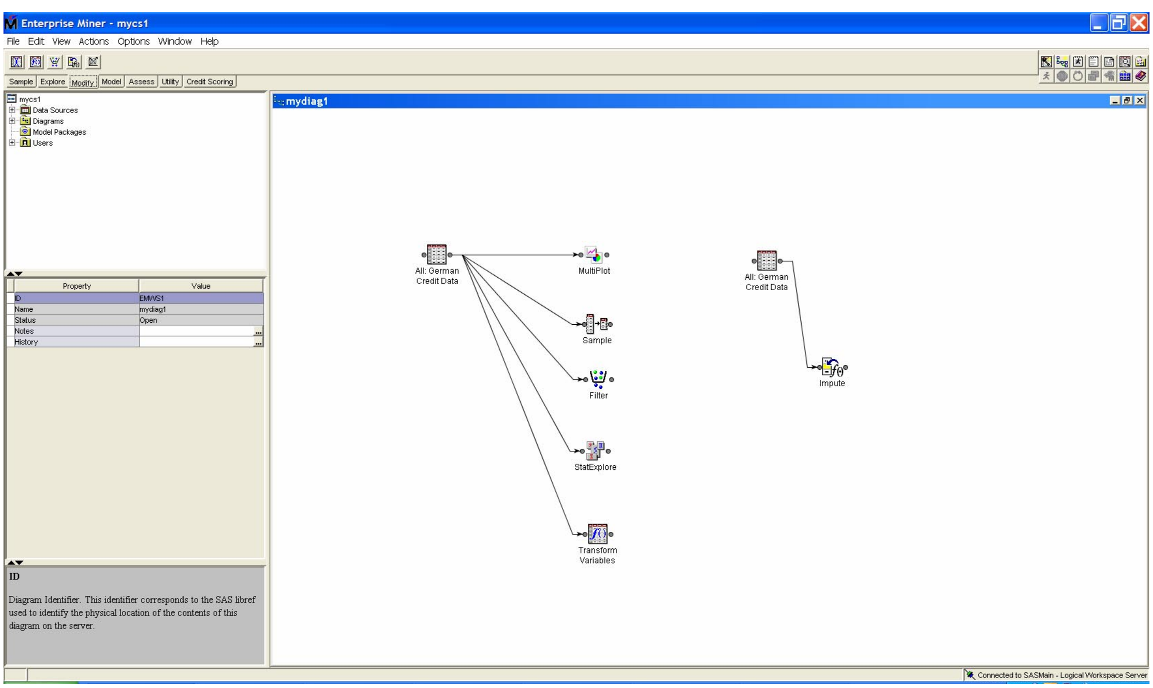
N is the number of observations, r is the number of rows and c is the number of columns of the contingency table. The statistic varies between 1 (perfectly predictive input) and 0 (useless input). Inspect the other output generated by this node.

i. Add a Transform Variables node to the diagram workspace and connect it to the input data component. Standardize the continuous variables AGE and AMOUNT to zero mean and unit standard deviation. Run the node and check whether the mean of the new variables is 0 and the standard deviation is 1.

j. Create a new Input data source component for the data set DMAGECRMISSING. Set the level of the following variables to ordinal: CHECKING and SAVINGS. Set the level of the following variables to nominal: COAPP, EMPLOYED, FOREIGN, HISTORY, HOUSING, JOB MARITAL, OTHER, PROPERTY, PURPOSE, RESIDENT, and TELEPHON. Accept the Interval (=continuous) level suggested for the other variables.

k. Add an Impute node to the diagram workspace and connect it to the input data component corresponding to the DMAGECRMISSING data set. Accept the default settings that replace missing interval variables with the mean and missing class variables with the mode. Run the node and inspect the results.

Your screen should now look as follows:



2. Developing PD scorecards using the Credit Scoring nodes in Enterprise Miner

a. Create a new diagram **`CSnodes`**. Create a data source for the GERMANCREDIT (not DMAGECR!) data set. Set the level of the following variables to ordinal: CHECKING and SAVINGS. Set the level of the following variables to nominal: COAPP, EMPLOYED, FOREIGN, HISTORY, HOUSING, JOB MARITAL, OTHER, PROPERTY, PURPOSE, RESIDENT, and TELEPHON. Set the role of the variable GOOD_BAD to target and its level to binary. Accept the Interval (=continuous) level suggested for the other variables.

b. Add an input node to the diagram workspace and link it with the GERMANCREDIT data source.

c. Add an Interactive Grouping node from the **Credit Scoring** tab to the diagram workspace and connect it to the input data node. Inspect the properties of the Interactive Grouping node.

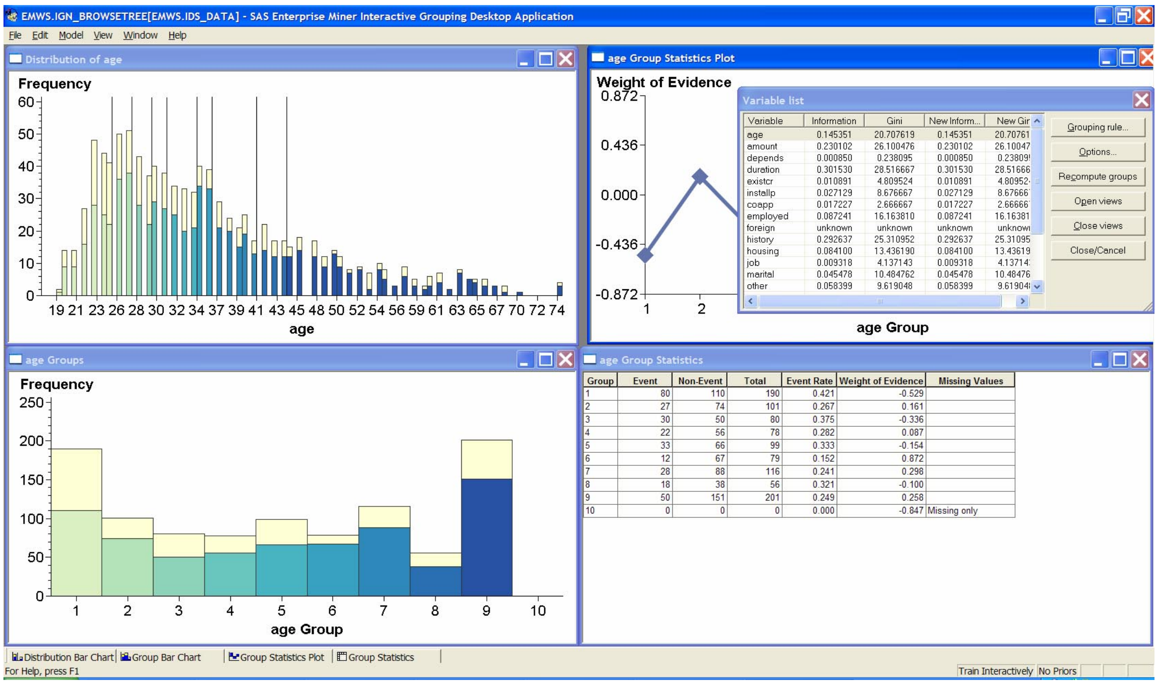
The option "Adjust WOE" allows you to deal with groups in which all observations have the same target value (for example, all good or all bad). The probabilities then become $p\_good_{attribute}$= (number of $good_{attribute}$+ adjustment factor)/ number of $good_{total}$ and $p\_bad_{attribute}$= (number of $bad_{attribute}$+ adjustment factor)/ number of bad_{total}. The adjustment factor can also be specified in the Properties panel. The option "Apply minimum distribution" allows you to specify a minimum size of each group. The minimum size can be set using the "Minimum percent distribution" option. The option "Use Frozen Groupings" can be used to prevent automatic grouping from being performed when the node is run. The "Criterion" option allows to specify which criterion will be used to do the grouping: entropy (a one-level decision tree) or based on chi-squared analysis. The "interval grouping method" option allows you to specify the method to do the grouping:

- Optimal: based on the criterion option
- Equal: creates equally spaced groups between the min and max value of the input ('equal interval binning')
- Quantile: creates groups with an equal number of observations in each group ('equal frequency binning')
- Monotonic event: creates groups that are monotonic in the event rate

The "Maximum Branch" option allows you to restrict the number of subsets that a splitting rule can produce. Set the value of this option to 10. The "Significant Digits" option allows you to specify the precision up to which interval variables are grouped. If this value is 0, then the lower and upper value of an interval group are integers. The "Variable Selection Method" specifies the statistic used for variable selection.

d. Run the Interactive Grouping node and inspect the results. Inspect the different output screens. The event rate plots screen presents event rate plots for each variable. An event rate plot presents for each group of the variable the number of events in that group divided by the total number of events (300 in our case). The output variables screen presents which variables were removed and which were retained based on the specified statistic (i.e. information value > 0.1) in our case. The statistics plot presents the information values of all variables in a bar chart. Choose View, Output data sets, Train, to inspect the columns that were added to the Train data set by the Interactive Grouping node.

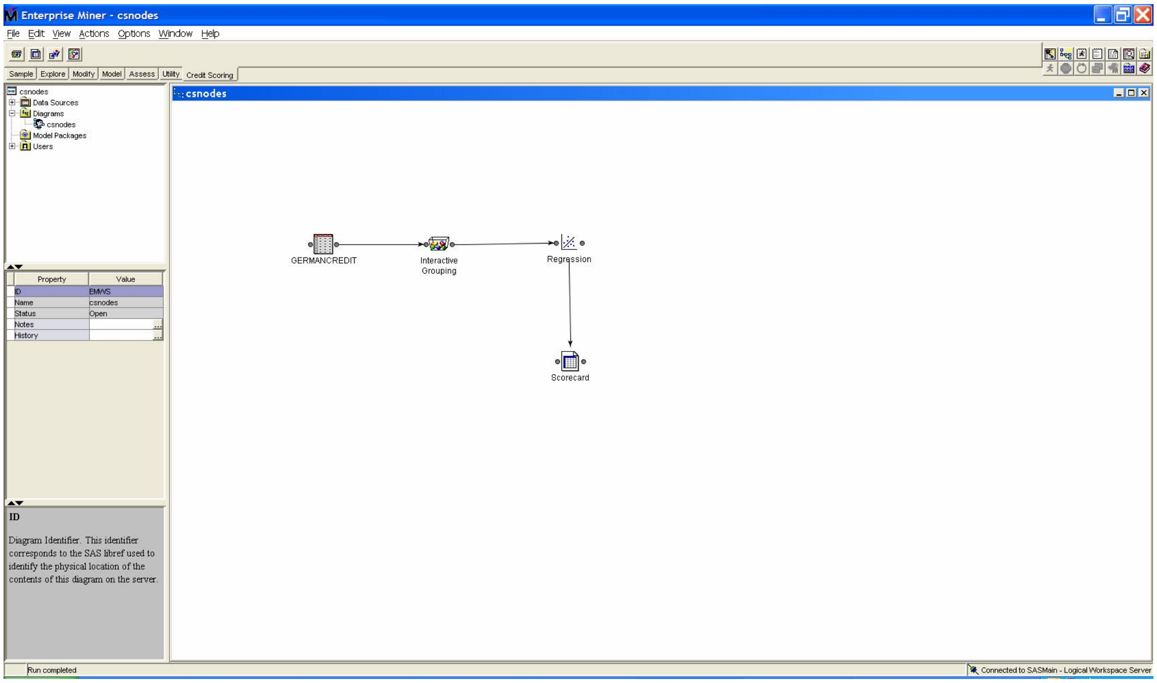
e. The Interactive Grouping node can also be run in interactive mode. Close the output screens and select the Interactive Grouping node again. Click the button next to the Interactive option in the Properties panel. The following screen appears:



f. The Variable list screen in the upper right corner can be used to navigate between the different variables hereby each time showing the corresponding plots and statistics table. Inspect the different plots and the statistics table. Select the variable **age** and click the **Grouping rule** button. Select group 4 and remove the group by clicking the corresponding button. Select **group 6** and add a new split point at 42. Click **OK** and close the window. Inspect the changes on the plots and the statistics table. The **Recompute Groups** button can be used to restore the original groups suggested by SAS. Select the variable **age** again and click the **Options** button. Split the variable into 4 quantiles each containing 25% of the observations. Select the **Recompute Groups** button to recompute the groups. Close the views for the variable **`age`** and select a nominal variable (for example, PURPOSE). Use the **Recompute Group** button to assign value 3 to group 1. Inspect the impact of this on the plots and the statistics table. Now choose an ordinal variable (for example, CHECKING) and experiment with it in the same way. Note that ordinal variables can only be assigned to neighboring groups.

g. Add a regression node to the diagram workspace. Connect the regression node to the interactive grouping node. Run the regression node and inspect the results. See how the regression node used only the weight of evidence type of variables. Inspect the *p*-values of these variables and see whether they are significant or not.

h. Add a scorecard node to the diagram workspace and connect it to the regression node. Your screen should now look as follows:



i. Inspect the properties of the Scorecard node. The Odds option is the event/non-event odds that corresponds to the score value that you specify in the Scorecard Points option, which is the sum of the score points across the attributes of a customer. Set the Odds to 50 (default), the Scorecard Points to 600 and the Points to Double Odds to 20 (default). Set the Scorecard Type option to detailed. A summary scorecard displays the attribute names and the scorecard points for each attribute. A detailed scorecard contains additional information such as group number, weight of evidence, event rate, percentage of population, and regression coefficient for each attribute. Set the number of buckets to 20 (this is the number of intervals in which the score range will be divided in a Gains table or Trade-off chart), the revenue of an accepted good to 1,000, the cost of an accepted bad to 5,000, the Current approval rate to 70 and the Current event rate to 2.5. Set the option Generate Characteristic Analysis to Yes.

In the adverse characteristics option section, you specify one of the following scores that is used to identify adverse characteristics:

Weighted average score

The weighted average score of a characteristic (variable) is the weighted average of the scorecard points of observations in the groups. For example, suppose that there are three groups of the variable `age`, and the average score points of the groups are 36, 42, and 45, with 20, 30, and 40 observations in each of the groups, respectively. The weighted average score of age is (36*20+ 42*30+45*40)/90=42.

Neutral score

The neutral score is the score points when weights of evidence is equal to 0 (that is, odds=1, where the probabilities of good and bad are equal for each attribute). Because the equation for calculating the score points is score=log(odds)*factor + offset, the value of the score equals the offset. Hence, the neutral score equals the offset and is the same for each characteristic (variable).

It is required by law to explain why an applicant is rejected for credit application. You compare the actual value of a variable to the weighted average score or neutral score in order to identify characteristics that are deemed adverse. For example, the following table lists the weighted average score and actual value of variables of an applicant who has been rejected, and the characteristics are listed based on the values of difference in ascending order. In this example, the characteristics that are deemed adverse in the order from most severe to least severe are OWN_RENT, INCOME, AGE, and then EDUCATION.

Characteristic	Weighted Average Score	Actual Value	Difference
OWN_RENT	64	57	-7
INCOME	54	52	-2
AGE	33	35	2
EDUCATION	59	65	6

j. Set the option to Weighted Average Score.

1) Run the Scorecard node and inspect the output. Inspect the KS-Statistic and the AUC of the Scorecard in the Strength Table screen. Note that the Gini coefficient equals 1-2*(1-AUC). Inspect the scorecard in the Scorecard screen. Inspect the Gains table. Inspect the output screen, which contains the characteristic analysis for the various variables. Inspect the other output, which is available from the **View** menu.

Strategy Curve	displays a Strategy Curve chart representing event odds (calculated as *P*(non-event)/*P*(event)) versus scorecard points.
Event Frequency Charts	depicting event frequencies versus cut-off score (cumulative or not, percentage-wise or not).
Strength Statistics	includes Kolmogorov-Smirnov plot, ROC plot, Captured event plot (Lorenz curve), and Strength Table.
Trade-Off Plots	displays Cumulative Event Rate, Average Marginal Profit, Average Total Profit, and Average Predicted Probability.

✎ The Scorecard node can also be run in interactive mode. Click the button next to the Scorecard Points option in the Properties panel. Here, you can manually overwrite the scorecard points that the scorecard has generated by editing the Scorecard Points column. You can also manually adapt the score ranges by clicking the button next to the Score Ranges option in the Properties panel.

3. Weighted average LGD

Suppose you are given the following data:

Year 1: 30 defaults of $50 with average loss of 10%.

Year 2: 20 defaults of $80 with average loss 70% and 40 defaults of $100 with average loss of 60%.

a. Complete the following table:

Long run LGD	Default count averaging	Exposure weighted averaging
Defaut weighted averaging		
Time weighted averaging		

1) Which weighted LGD would you choose according to Basel II?

4. Modeling LGD using the Beta distribution

a. Load the data LGD.txt into SAS as follows:

```
proc import OUT= WORK.LGD
            DATAFILE= "C:\temp\LGD.txt"
            DBMS=TAB REPLACE;
     GETNAMES=YES;
     DATAROW=2;
run;
```

b. The first 13 columns of the data set are ratios that will be used to predict the LGD, which is the last column. Inspect the distribution of the variable **LGD** using SAS/INSIGHT (Select **Solutions** ⇨ **Analysis** ⇨ **Interactive Data Analysis** from the menu above). Does the distribution look skew? Does it look Gaussian?

c. Split the 506 observations into a training set of 337 and a holdout set of 169 observations, as follows:

```
data training holdout;
SET LGD;
IF 1 < = _N_ < = 337 THEN OUTPUT training;
ELSE OUTPUT holdout;
run;
```

d. Inspect both the training and the holdout data sets to see if they were successfully created. We will start by estimating an ordinary linear regression model without transforming the target variable, as follows:

```
proc reg data=training OUTEST= outests ;
model LGD= ratio1 ratio2 ratio3 ratio4 ratio5 ratio6 ratio7
ratio8 ratio9 ratio10 ratio11 ratio12 ratio13;
run;
```

e. Look at the *p*-values of the regression to see the most important inputs. Compute the predictions on the holdout set, as follows:

```
data preds (KEEP = LGD LGDpred);
if _N_=1 then set outests (RENAME = (ratio1=beta1 ratio2=beta2
ratio3=beta3 ratio4=beta4 ratio5=beta5 ratio6=beta6 ratio7=beta7
ratio8=beta8 ratio9=beta9 ratio10=beta10 ratio11=beta11
ratio12=beta12 ratio13=beta13));
set holdout;
LGDpred =
SUM(intercept,ratio1*beta1,ratio2*beta2,ratio3*beta3,ratio4*beta4
,ratio5*beta5,
ratio6*beta6,ratio7*beta7,ratio8*beta8,ratio9*beta9,ratio10*beta1
0, ratio11*beta11, ratio12*beta12, ratio13*beta13);
run;
```

f. Inspect the PREDS data set. Visualise the performance of the linear regression model using a scatter plot in SAS/INSIGHT (Select **Solutions** ⇨ **Analysis** ⇨ **Interactive Data Analysis** from the menu above). Compute the correlation between the actual and predicted LGD, as follows:

```
proc corr data=preds;
var LGD LGDpred;
run;
```

g. Report the correlation in the table at the end of this exercise.

h. Compute the mean squared error (MSE) on the holdout set, as follows:

```
data MSEtemp;
set preds;
MSEterm=(LGD-LGDpred)**2;
run;
proc means data=MSEtemp;
var MSEterm;
run;
```

i. Report the MSE in the table at the end of this exercise.

j. We are now ready to do the second regression, where we are going to assume that the variable **LGD** has a beta distribution and transform it.

1) Compute the mean and the variance of the variable **LGD**, as follows:

```
proc means data=training;
var LGD;
output out=meanstats mean=mu var=sigmasq;
run;
```

k. Estimate the alpha and beta parameters of the beta distribution, as follows:

```
data betaparams;
set meanstats;
alpha=(mu*mu*(1-mu)/sigmasq)-mu;
beta=alpha*(1/mu -1);
run;
```

l. Inspect the BETAPARAMS data set.

1) Transform the variable **LGD** to a normal distributed variable, as follows:

```
data transformedtraining;
if _N_=1 then set betaparams;
set training;
newLGD=probit(cdf('BETA',LGD,alpha,beta));
run;
```

m. Inspect the distribution of the new variable **LGD** using SAS/INSIGHT (**Choose Solutions ⇨ Analysis ⇨ Interactive Data Analysis** from the menu above). Does it look more symmetric and Gaussian now?

n. Estimate a linear regression model on the transformed variable, as follows:

```
proc reg data=transformedtraining OUTEST= outests2 ;
model newLGD= ratio1 ratio2 ratio3 ratio4 ratio5 ratio6 ratio7
ratio8 ratio9 ratio10 ratio11 ratio12 ratio13;
run;
```

5. Look at the *p*-values of the regression to see the most important inputs. Are these the same as in the previous model? Compute the predictions on the HOLDOUT data set, as follows:

```
data transpreds (KEEP = LGD transLGDpred);
if _N_=1 then set outests2 (RENAME = (ratio1=beta1 ratio2=beta2
ratio3=beta3 ratio4=beta4 ratio5=beta5 ratio6=beta6 ratio7=beta7
ratio8=beta8 ratio9=beta9 ratio10=beta10 ratio11=beta11
ratio12=beta12 ratio13=beta13));
set holdout;
transLGDpred =
SUM(intercept,ratio1*beta1,ratio2*beta2,ratio3*beta3,ratio4*beta4,ra
tio5*beta5,
ratio6*beta6,ratio7*beta7,ratio8*beta8,ratio9*beta9,ratio10*beta10,
ratio11*beta11, ratio12*beta12, ratio13*beta13);
run;
```

a. Transform the predictions back, as follows:

```
data preds2;
if _N_=1 then set betaparams;
set transpreds;
LGDpred=betainv(cdf('NORMAL',transLGDpred),alpha,beta);
run;
```

b. Visualise the performance of the transformed linear regression model using a scatter plot in SAS/INSIGHT (**Choose Solutions ⇨ Analysis ⇨ Interactive Data Analysis** from the menu above). Compute the correlation between the actual and predicted LGD, as follows:

```
proc corr data=preds2;
var LGD LGDpred;
run;
```

c. Report the correlation in the table at the end of this exercise.

d. Compute the MSE of the transformed linear regression model, as follows:

```
data MSEtemp2;
set preds2;
MSEterm=(LGD-LGDpred)**2;
run;
proc means data=MSEtemp2;
var MSEterm;
run;
```

e. Report the MSE in the table below.

	Correlation	MSE
Linear regression		
Beta linear regression		

f. According to your findings, which model gives the best performance?

6. The Basel II Capital requirement formulas

a. Create a SAS macro, as follows:

```
%macro BaselCap(PD, LGD, corr);
temp1=(((1/(1-&corr))**0.5)*probit(&PD) + ((&corr/(1-
&corr))**0.5)*probit(0.999));
temp2=CDF('Normal',temp1);
Cap=&LGD*(temp2-&PD);
%mend BaselCap;
```

This macro will compute the Basel II capital requirements for a given PD, LGD, and correlation factor.

b. Create the following SAS data sets computing the Basel II capital requirements for residential mortgages and qualifying revolving exposures for low PD's:

```
data resmortgage;
DO i=1 to 200;
PD=i/1000;
%BaselCap(PD=i/1000,LGD=0.50,corr=0.15);
output;
end;
data QRE;
DO i=1 to 200;
PD=i/1000;
%BaselCap(PD=i/1000,LGD=0.50,corr=0.04);
output;
end;
```

c. Now plot the Basel II capital requirements for residential mortgages and qualifying revolving exposures, as follows:

```
proc gplot data=resmortgage;
plot Cap*PD;
run;
proc gplot data=QRE;
plot Cap*PD;
run;
```

d. Consider a credit card portfolio with the following true characteristics:

PD=0.05; LGD=0.20; and EAD=$10.000,00

1) Compute the capital requirement, assuming the following estimates:

Scenario	PD	LGD	EAD	Capital requirement
Everything correct	0.05	0.20	10.000	
10% overestimate PD	0.055	0.20	10.000	
10% overestimate LGD	0.050	0.22	10.000	
10% overestimate EAD	0.050	0.20	10.100	

2) Which scenario(s) have the biggest impact on the Capital Requirement? Why?

7. EAD modeling

Consider a credit card portfolio with the following characteristics for five customers that defaulted in the past:

Drawn one month prior to default	EAD at time of default
1800	2300
1500	2100
1000	1800
1700	1750
500	2300

a. Each credit card has a credit limit of $2.500. Compute the credit conversion factor (CCF) for each customer in the portfolio.

8. Backtesting at level 0

a. Compute a system stability index (SSI) for the following data:

Score range	Actual%	Training%
0–169	7%	8%
170–179	8%	10%
180–189	7%	9%
190–199	9%	13%
200–209	11%	11%
210–219	11%	10%
220–229	10%	9%
230–239	12%	10%
240–249	11%	11%
250+	14%	9%

b. Enter the data in SAS as follows:

```
data populationstab;
input range$ actual training;
lines;
0-169 0.07 0.08
170-179 0.08 0.10
180-189 0.07 0.09
190-199 0.09 0.13
200-209 0.11 0.11
210-219 0.11 0.10
220-229 0.10 0.09
230-239 0.12 0.10
240-249 0.11 0.11
250+  0.14 0.09
;
run;
```

c. Compute the intermediate values as follows:

```
data SSIdata;
set populationstab;
temp1=actual-training;
temp2=log(actual/training);
index=temp1*temp2;
run;
```

d. Now compute the SSI as follows:

```
proc means data=SSIdata sum;
run;
```

1) Do you observe a difference? What traffic light color would you assign? Play with the numbers in the table and observe the impact on the SSI.

9. Backtesting at level 1

Consider the following scorecard performance data:

Score	Actual Good/Bad
100	Bad
110	Bad
120	Good
130	Bad
140	Bad
150	Good
160	Bad
170	Good
180	Good
190	Bad
200	Good
210	Good
220	Bad
230	Good
240	Good
250	Bad
260	Good
270	Good
280	Good
290	Bad
300	Good
310	Bad
320	Good
330	Good
340	Good

a. Draw the ROC curve.

b. Check that the scorecard performs better than a random scorecard.

c. Draw the CAP curve.

d. Again, check that the scorecard performs better than a random scorecard.

e. Draw the Kolmogorov-Smirnov Curve.

f. Calculate the Kolmogorov-Smirnov statistic.

10. Backtesting at level 2: Binomial test

Consider the following data from an IRB rating system:

Rating Category	Estimated PD	Nr. of observations	Nr. of observed defaults
A	2%	1000	17
B	3%	500	20
C	7%	400	35
D	20%	200	50

a. Compute the critical values according to the normal approximation of the binomial test and see whether there are any significant differences between realized PD rates and estimated PD rates.

b. Enter the data in SAS/STAT as follows:

```
data ratings;
input rating$ estimatedPD nrobs nrdefaults;
lines;
A 0.02 1000 17
B 0.03 500 20
C 0.07 400 35
D 0.20 200  50
;
run;
```

c. Now compute the actual default rates and the critical values as follows:

```
data binomialtest;
set ratings;
actualdefaultrate=nrdefaults/nrobs;
criticalvalue=probit(0.99)*((estimatedPD*(1-
estimatedPD)/nrobs)**0.5)+estimatedPD;
run;
```

d. Inspect the BINOMIALTEST data set. Check every category and see whether the actual default rate exceeds the critical value or not. Alter the variable **`nrdefaults`** and investigate the impact on the results.

11. Backtesting at level 2: Hosmer-Lemeshow test

Consider the following data from an IRB rating system (same as in previous exercise):

Rating Category	Estimated PD	Nr. of observations	Nr. of observed defaults
A	2%	1000	17
B	3%	500	20
C	7%	400	35
D	20%	200	50

a. Conduct a Hosmer-Lemeshow test to see whether the observed data agree with the predicted PDs. Note that the Hosmer-Lemeshow test is available in PROC LOGISTIC. However, because this implementation is too rigid for our purpose, we will program the test ourselves.

b. First, create a SAS data set in SAS/STAT as follows:

```
data ratings;
input rating$ estimatedPD nrobs nrdefaults;
lines;
A 0.02 1000 17
B 0.03 500 20
C 0.07 400 35
D 0.20 200  50
;
run;
```

c. Next, we will create a new data set containing all summation terms of the Hosmer-Lemeshow statistic, as follows:

```
data chisquaredvaluetemp;
set ratings;
chisquaredterm=(nrobs*estimatedPD-
nrdefaults)**2/(nrobs*estimatedPD*(1-estimatedPD));
run;
```

d. Inspect the CHISQUAREDVALUETEMP data set. We will now sum all these terms to arrive at the value of the test statistic using PROC MEANS:

```
proc means data=chisquaredvaluetemp;
var chisquaredterm;
output out=chisquaredvalue sum=sum_chisquaredterm;
run;
Inspect the chisquaredvalue data set. We will now compute the p-
value of the test statistic as follows:
data HLstatistic;
set chisquaredvalue;
pvalue=1-CDF('chisquared',sum_chisquaredterm,_FREQ_);
run;
```

e. Note that the variable `_FREQ_` represents the degrees of freedom. Inspect the **`HLstatistic`** data set. What does the *p*-value tell you? Do the data observed significantly differ from the estimated PDs? How does your conclusion relate to the outcome of the Binomial test? Change the data so as to alter the answer to the previous question.

12. Backtesting at level 2: Traffic light indicator approach

Suppose you are given the following output from a Traffic Light Indicator approach to backtesting:

PD	Baa1	Baa2	Baa3	Ba1	Ba2	Ba3	B1	B2	B3	Caa-C	Av
	0,15%	0,09%	0,37%	0,74%	0,72%	1,91%	3,32%	5,84%	7,40%	17,07%	2,98%
DR	**Baa1**	**Baa2**	**Baa3**	**Ba1**	**Ba2**	**Ba3**	**B1**	**B2**	**B3**	**Caa-C**	**Av**
1993	0,00%	0,00%	0,00%	0,83%	0,00%	0,76%	3,24%	5,04%	11,29%	28,57%	3,24%
1994	0,00%	0,00%	0,00%	0,00%	0,00%	0,59%	1,88%	3,75%	7,95%	5,13%	1,88%
1995	0,00%	0,00%	0,00%	0,00%	0,00%	1,76%	4,35%	6,42%	4,06%	11,57%	2,51%
1996	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	1,17%	0,00%	3,28%	13,99%	0,78%
1997	0,00%	0,00%	0,00%	0,00%	0,00%	0,47%	0,00%	1,54%	7,22%	14,67%	1,41%
1998	0,00%	0,31%	0,00%	0,00%	0,62%	1,12%	2,11%	7,55%	5,52%	15,09%	2,83%
1999	0,00%	0,00%	0,34%	0,47%	0,00%	2,00%	3,28%	6,91%	9,63%	20,44%	3,35%
2000	0,28%	0,00%	0,97%	0,94%	0,63%	1,04%	3,24%	4,10%	10,88%	19,65%	3,01%
2001	0,27%	0,27%	0,00%	0,51%	1,38%	2,93%	3,19%	11,07%	16,38%	34,45%	5,48%
2002	1,26%	0,72%	1,78%	1,58%	1,41%	1,58%	2,00%	6,81%	6,86%	29,45%	3,70%
Av	0,26%	0,17%	0,42%	0,53%	0,54%	1,36%	2,46%	5,76%	8,76%	20,9%	3,05%

The backtesting was done according to a one-tailed Binomial test with respect to the reference PD given in the second row of the table. The colours are coded as follows:

Green	No difference at 10% level
Yellow	Difference at 10% level but not at 5%
Orange	Difference at 5% level but not at 1%
Red	Difference at 1% level

a. How do you interpret the results from this backtesting exercise?

b. What action would you suggest?

13. Benchmarking

Consider the following data set, which contains ratings of ten customers provided by three institutions:

	Institution 1	Institution 2	Institution 3
Customer 1	B	D	A
Customer 2	A	C	A
Customer 3	A	A	C
Customer 4	D	B	B
Customer 5	C	B	D
Customer 6	A	C	B
Customer 7	C	D	A
Customer 8	B	A	B
Customer 9	B	A	C
Customer 10	A	D	B

a. Which two institutions provide the most similar ranking of customers? Use Spearman's rho and Kendall's tau to determine this.

b. Enter the data in SAS as follows:

```
data ratings;
input inst1$ inst2$ inst3$;
lines;
B D A
A C A
A A A
D B C
C B C
A C B
C D B
B A B
B A C
A D B
;
run;
```

c. Transform the data to numerical values as follows:

```
data ratings2;
set ratings;
if inst1='A' then do
inst1num=1;
end;
if inst1='B' then do
inst1num=2;
end;
if inst1='C' then do
inst1num=3;
end;
if inst1='D' then do
inst1num=4;
end;
if inst2='A' then do
inst2num=1;
end;
if inst2='B' then do
inst2num=2;
end;
if inst2='C' then do
inst2num=3;
end;
if inst2='D' then do
inst2num=4;
end;
if inst3='A' then do
inst3num=1;
end;
if inst3='B' then do
inst3num=2;
end;
if inst3='C' then do
inst3num=3;
end;
if inst3='D' then do
inst3num=4;
end;
run;
Compute Spearman’s rho and Kendall’s tau as follows:
proc corr data=ratings2 spearman kendall;
var inst1num inst3num;
run;
```

d. Now complete the following table:

	Spearman's rho	Kendall's tau-b
Institution 1 versus Institution 2		
Institution 2 versus Institution 3		
Institution 1 versus Institution 3		

e. Which two institutions give the most similar ratings?

f. Now draw a histogram as follows:

```
data histo;
set ratings2;
hist1vs2=inst1num-inst2num;
hist1vs3=inst1num-inst3num;
hist2vs3=inst2num-inst3num;
run;
proc univariate data=histo noprint;
histogram hist1vs3 /cfill=blue midpoints=-4 to 4 by 1;
run;
```

g. Using a histogram, which two institutions give the most similar ratings?

h. Play with the ratings at the start of the exercise and look at the impact on Spearman's rho and Kendall's tau and on the histograms.

14. Low default portfolios

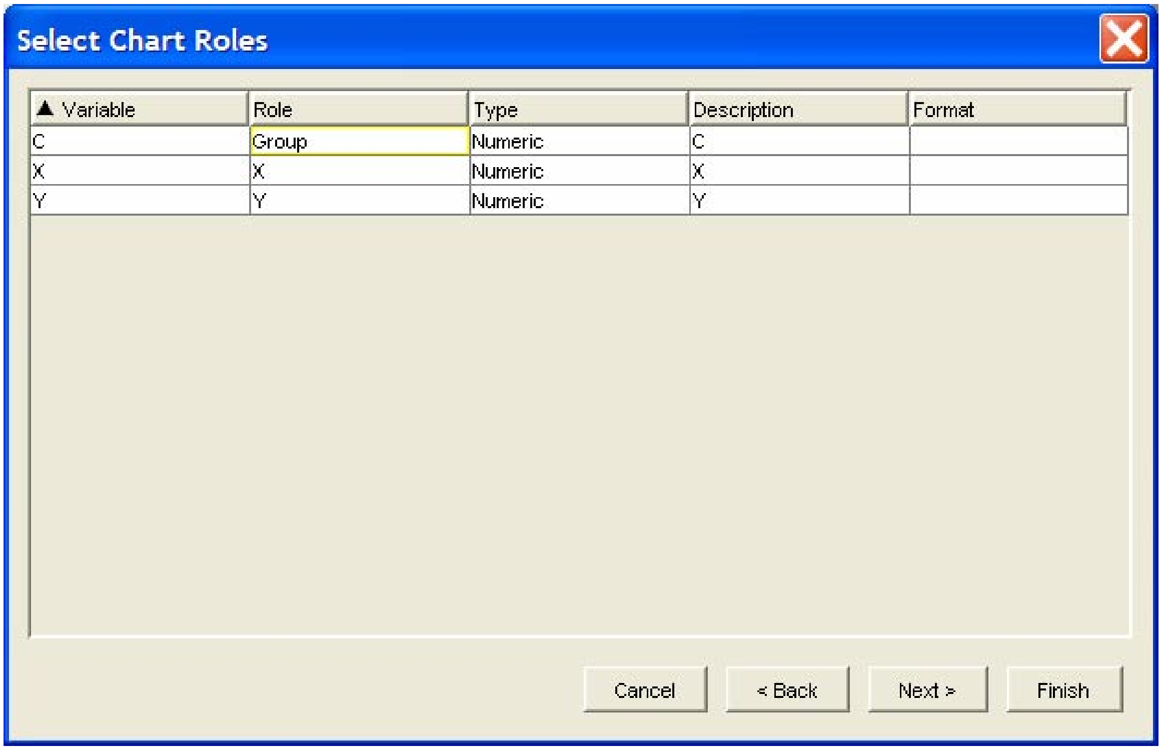
a. Assume you have a portfolio with zero defaults and three rating grades: A (100 obligors), B (150 obligors), and C (80 obligors). You may assume that defaults are independent. Calculate the most prudent estimate of PD_A, PD_B, PD_C, assuming the following significance levels:

	α=0.90	α=0.99
PD_A		
PD_B		
PD_C		

15. Linear versus non-linear decision boundary

a. Create a new project called twospiral (or keep working in your current project). Create a data source for the DMLSPIR data set. Set the role of the variable `C` to target and its level to binary.

b. Create a new diagram called **`twospiral`**. Add an Input data node to the diagram workspace. Set its data source property in the property sheet left below to DMLSPIR. Run the input node, right-click it, and choose **Results…**. From the menu, choose **View** ⇨ **Output Data Sets** ⇨ **Data**. From the new menu, choose **Action** ⇨ **Plot...**. Create a scatter plot. Set the role of the variable X to **`X`**, the variable Y to **`Y`**, and the variable C to **`group`**.



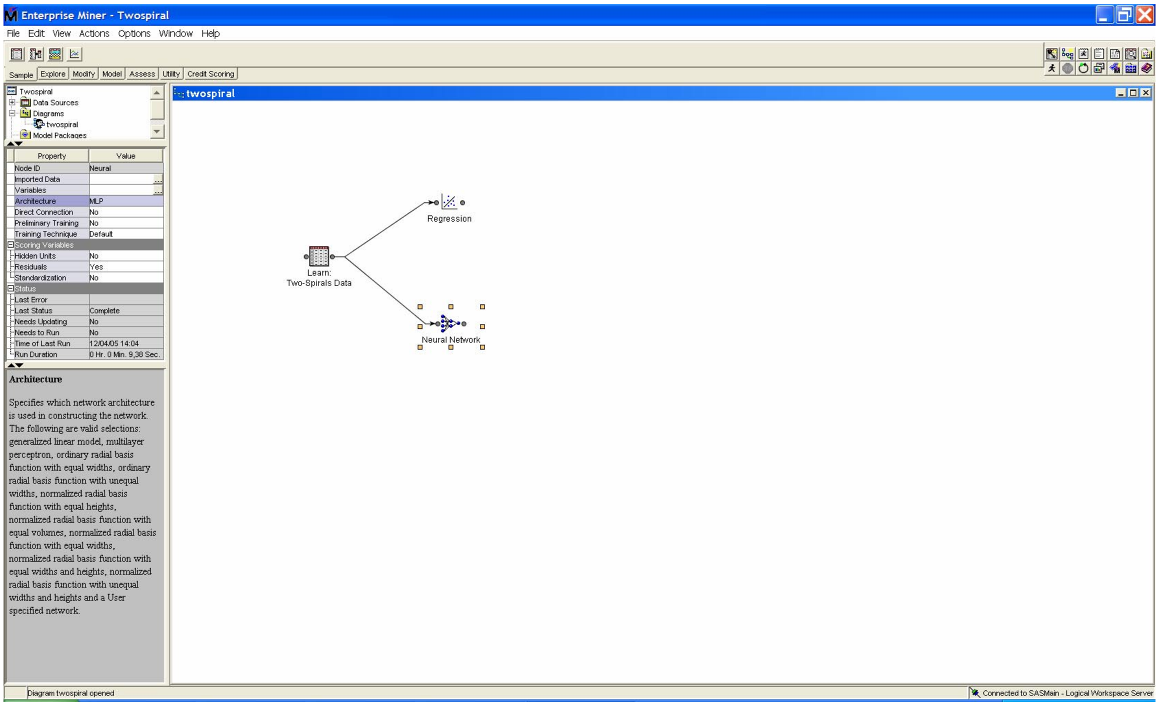
c. Inspect the generated scatter plot.

d. Go back to the diagram workspace. Add a regression node to the diagram workspace and connect it with the input data node. Run the regression node by right-clicking it and choosing **Run**. Inspect the results of the regression node by right-clicking it and choosing **Results**. From the menu, choose **View** ⇨ **Output Data Sets** ⇨ **Train**. From the new menu, choose **Action** ⇨ **Plot**, and create a scatter plot. Set the role of the variable X to **`X`**, the role of the variable Y to **`Y`**, and the role of the variable I_C to **`group`**. Inspect the scatter plot. What type of decision boundary has been modeled? What is the classification accuracy?

e. Go back to the diagram workspace. Add a Neural Network node to the diagram workspace and connect it with the Input Data node. From the menu above, choose **View** ⇨ **Property Sheet** ⇨ **Advanced**. Select the **neural network node**. In the Property panel, set the Number of Hidden Units to `40` and the maximum iterations to `500`. Run the Neural Network node by right-clicking it and selecting **Run.** Inspect the results of the regression node by right-clicking it and choosing **Results…**.

f. From the menu, choose **View** ⇨ **Output Data Sets** ⇨ **Train**. In the new menu, choose **Action** ⇨ **Plot**, and create a scatter plot. Set the role of the variable X to **`X`**, the role of the variable Y to **`Y`**, and the role of the variable I_C to **`group`**. Inspect the scatter plot. What type of decision boundary has been modeled? What is the classification accuracy?

g. **Extra:** Estimate a decision tree using the Gini-splitting criterion and inspect the decision boundaries.



16. Comparing PD models using McNemar's test

a. Create a new diagram called **`McNemar`**. Create a data source for the DMAGECR data set. Set the level of the following variables to ordinal: CHECKING and SAVINGS. Set the level of the following variables to nominal: COAPP, EMPLOYED, FOREIGN, HISTORY, HOUSING, JOB MARITAL, OTHER, PROPERTY, PURPOSE, RESIDENT, TELEPHON. Set the role of the variable GOOD_BAD to target and its level to binary. Accept the Interval (=continuous) level suggested for the other variables.

b. Add an input node to the diagram workspace and link it with the DMAGECR data source.

c. Add an Interactive Grouping node to the diagram workspace and connect it with the input data source node. This node will automatically coarse classify all variables of the credit scoring data set. Run the Interactive Grouping node using the default settings (for more information on this, see Credit Scoring course I).

d. Add a Data Partition node to the diagram workspace and connect it with the Interactive Grouping node. Accept the suggested percentages for the training, validation and test set. Note that by default the GOOD_BAD target variable is used for stratification.

e. Add a linear regression node and connect it with the Data Partition node. Check that only the weight of evidence type of variables (suffix _WOE) are used in the estimation process. Run the Linear Regression node.

f. Add a tree node to the diagram workspace and connect it with the Data Partition node. Run the tree node.

g. Add a neural network node to the diagram workspace and connect it with Data Partition node. Create a multi-layer neural network with 3 hidden neurons (default). Run the Neural Network node.

h. Add a SAS Code node to the diagram workspace and connect it with the Tree, Neural Network, and Regression nodes. Look up the libraries and names of the scored test data of the logistic regression, tree, and neural network models. Select the SAS Code node and push the button next to the SAS Code option in the Property panel. Enter the following code:

```
data linrreg (keep=predlin);
set EMWS.Reg_Test;
rename I_good_bad=predlin;
run;
data treepred (keep=predtree);
set EMWS.Tree_test;
rename I_good_bad=predtree;
run;
data NNpred (keep=predNN good_bad);
set EMWS.Neural_Test;
rename I_good_bad=predNN;
run;
data allpred;
merge linrreg treepred NNpred;
run;
PROC FREQ data=allpred;
tables predlin*good_bad;
run;
PROC FREQ data=allpred;
tables predNN*good_bad;
run;
PROC FREQ data=allpred;
tables predlin*predNN /agree;
run;
```

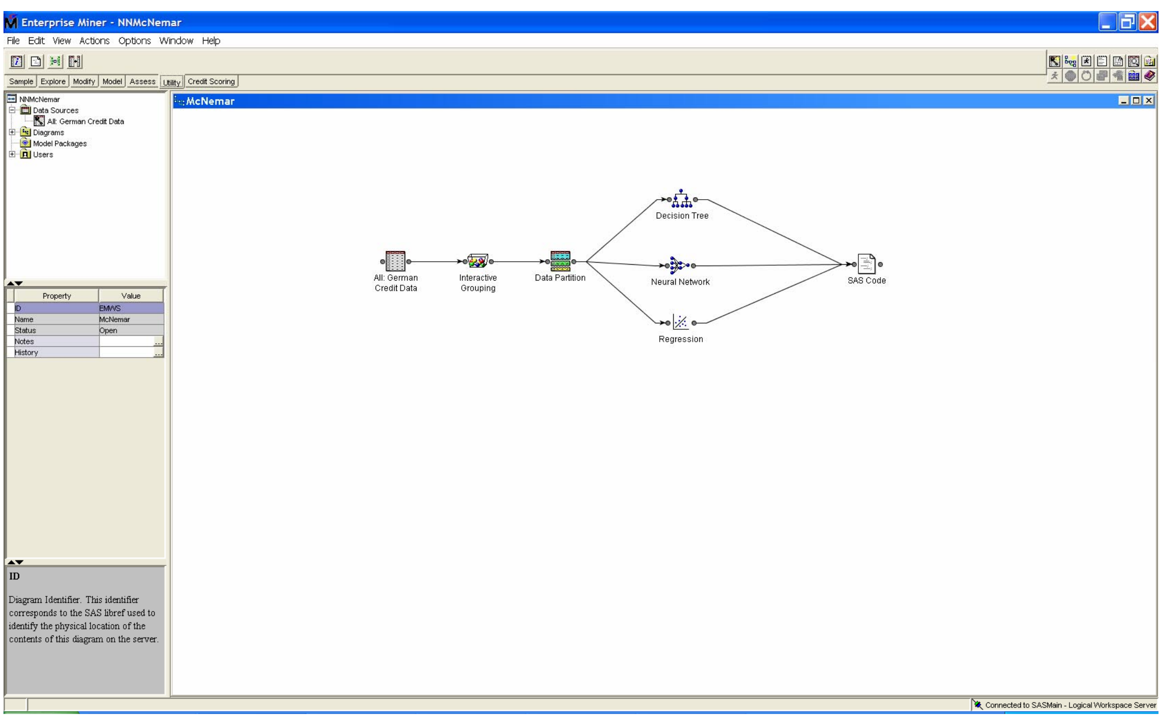
17. The first PROC FREQ statement computes a confusion matrix for the logistic regression model. The second PROC FREQ statement computes a confusion matrix for the neural network. The third PROC FREQ statement compares the classification performance of the logistic regression model with the classification performance of the neural network using the McNemar's test (note the **`/agree`** option).

a. Using the code above, complete the following performance tables:

Classification technique	Classification accuracy (on test set)
Logistic regression	
Decision tree	
Neural network	

Comparison	*p*-value McNemar test	Decision (same or different)
Logistic regression vs Decision tree		
Logistic regression vs. Neural network		
Neural network vs. Decision tree		

b. **Extra:** Vary the number of hidden neurons of the neural network and look at the impact on the performance and the significance of the tests.



18. Neural networks for modeling LGD

a. Create a new diagram called **`NeuralLGD`**. Create a data source for the LGDSAS data set. Set the role of the variable LGD to target. Explore the distribution of this variable.

b. Add an input node to the diagram workspace and link it with the **`LGDSAS`** data source.

c. Add a Data Partition node to the diagram workspace and connect it with the Input node. Accept the suggested training set, validation set, and test set percentages.

d. Add four neural network nodes to the diagram workspace having, respectively, 2, 3, 5, and 10 hidden units. For each neural network node, set the Model Selection Criterion to Average Error. Connect the Neural Network nodes to the Data Partition node.

e. Add a Model Comparison node to the diagram workspace and connect it with the four Neural Network nodes. Set the Selection Statistic property of the Model Comparison node to Validation: Average Squared Error. This node will now select the neural network with the lowest average squared error on the validation set. Run the Model Comparison node. Which neural network was selected by the model comparison node? Inspect the various types of errors on both the validation and the test set. What is the mean squared error on the test set? Compare this with what we got in the first exercise.

f. Add a SAS Code node to the neural network with the lowest validation set error. Compute the correlation between the actual and predicted LGDs as follows:

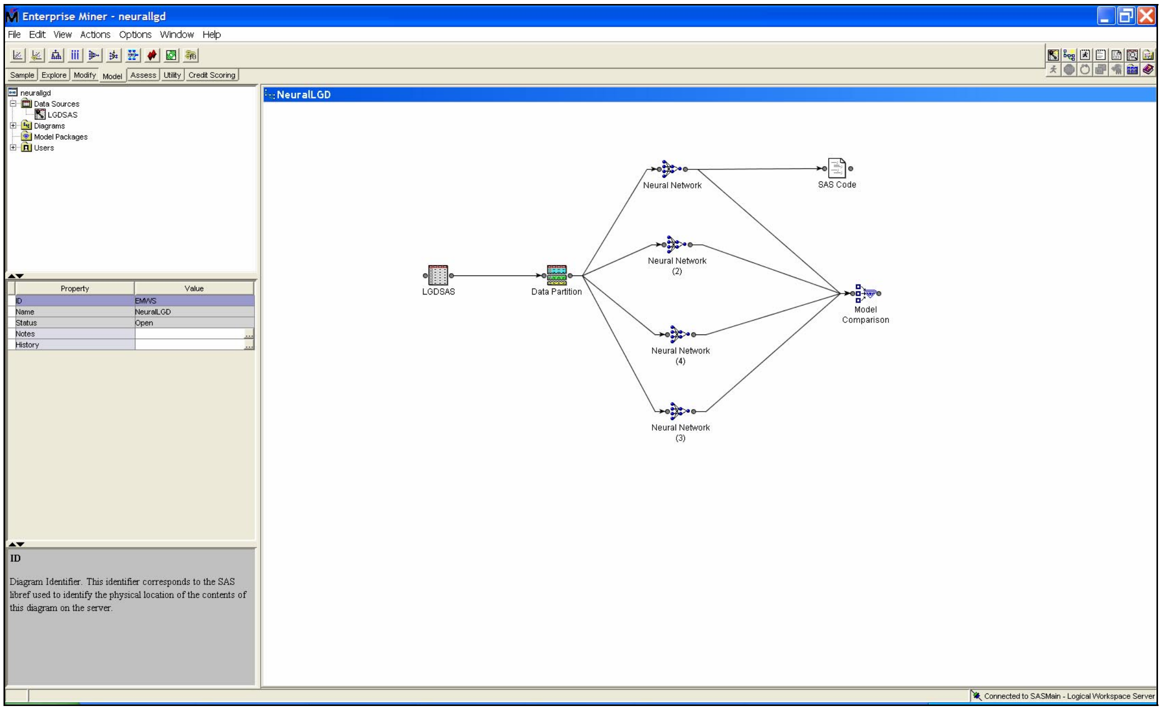
```
proc corr data=EMWS.Neural_Test;
var LGD P_LGD;
run;
```

g. Contrast the reported correlation with the one reported in the first exercise using the linear regression approaches.

h. Create a scatter plot depicting the actual LGD versus the predicted LGD.

i. At this point, we have tried three approaches to modeling LGD: using linear regression, using linear regression with a transformed target variable, and using neural networks. Overall, which one gives you the best performance?

Your screen should now look as follows:



19. Measuring the performance of rating systems

Suppose we developed a rating system for mimicking country ratings provided by Moody's. We used both multi-class neural networks and One-versus-One logistic regression. The classifications on an independent hold-out data set are as follows:

True rating	Multiclass neural network	OvsO logistic regression
A	C	A
C	B	A
A	A	D
D	B	C
A	A	A
A	B	B
D	D	D
B	B	B
C	A	A
A	D	C
D	A	D
A	B	A
C	B	B
C	C	A
B	B	B

a. Based on the above table, complete the following performance table:

	Multiclass neural network	OvsO logistic regression
0-notch accuracy		
1-notch accuracy		
3-notch accuracy		
4-notch accuracy		

b. Represent these numbers in a notch difference graph. Which is the best classification technique?

20. Support vector machines

The two-spiral data

Go to the program editor in SAS/STAT.

a. Add a library called Neural, referring to the directory with the DMLSPIR.SAS7BDAT data set.

b. Make a DMDB database from this data set using PROC DMDB.

```
proc dmdb data=neural.dmlspir batch out=dmdb dmdbcat=meta;
var x y;
class c;
target c;
run;
```

c. Create an SVM classifier using PROC SVM. Choose an RBF kernel with σ=0.5 and set C=100. Save the predictions in a new data set called SPIRALOUT.

```
proc svm data=dmdb dmdbcat=meta nomonitor out=spiralout
kernel=RBF K_PAR=0.5 C=100;
var x y;
target c;
run;
```

d. Inspect the classification accuracy of the estimated SVM classifier. Inspect the SPIRALOUT data set. Note that the column _P_ denotes the output of the SVM classifier before applying the sign operator, the column _Yo_ is the level name of the class target as given in the original data set, _Y_ is the observed value of the target as used for modeling (the same as _Yo_ in our case, because this is already coded as 0/1), _R_ is the residual and _ALPHA_ gives the Lagrange multipliers. How many support vectors are used?

e. Re-estimate the SVM classifier by varying the regularization parameter as follows: 0.1, 10, 20, and 30. Use the C-statement. Investigate the impact of the regularization parameter on the classification performance.

```
proc svm data=dmdb dmdbcat=meta nomonitor out=spiralout
kernel=RBF K_PAR=0.5 C=100;
var x y;
target c;
C 0.1 10 20 30;
run;
```

f. Re-estimate the SVM classifier using a linear kernel with regularization parameter =100. What is the impact on the classification accuracy?

```
proc svm data=dmdb dmdbcat=meta nomonitor out=spiralout
kernel=linear C=100;
var x y;
target c;
run;
```

g. Re-estimate the SVM classifier using a polynomial kernel of degree 3 with regularization parameter =100. What is the impact on the classification accuracy?

```
proc svm data=dmdb dmdbcat=meta nomonitor out=spiralout
kernel=polynom K_PAR=3 C=100;
var x y;
target c;
run;
```

21. Multiclass SVM

a. Open the file **`MulticlassSVMstart.sas`** in the SAS program editor in SAS/STAT.

1) Merge the rating into three classes as follows:

```
data bondrate2;
 set bondrate;
 if rating='AAA' or rating='AA' or rating='A' then rating=1;
 if rating='BAA' or rating='BA' or rating='B' then rating=2;
 if rating='C' then rating=3;
run;
```

b. Create a multiclass SVM classifier using the One-versus-All scheme. Use RBF kernels with σ=0.5 and a regularization parameter of 10.

```
proc dmdb data=bondrate2 batch out=dmdb dmdbcat=meta;
var  LOPMAR     LFIXCHAR     LGEARRAT     LTDCAP     LLEVER
LCASHRAT     LACIDRAT     LCURRAT LRECTURN LASSLTD;
class rating;
target rating;
run;
proc svm data=dmdb dmdbcat=meta kernel=RBF K_PAR=0.5 C=10;
var  LOPMAR     LFIXCHAR     LGEARRAT     LTDCAP     LLEVER
LCASHRAT     LACIDRAT     LCURRAT LRECTURN LASSLTD;
target rating;
run;
```

c. Inspect the generated output. How many classifiers were estimated? What is the accuracy per estimated classifier? What is the overall accuracy on the training set? If One-versus-One coding or Minimum Output Coding was used, how many classifiers would then have been estimated?

d. Change the kernel to linear and investigate the impact on the classification accuracy.

22. Multiclass discriminant analysis using One-versus-One coding

Since the TESTDATA option is not yet fully functional in PROC SVM, this exercise illustrates the use of One-versus-One coding using discriminant analysis.

Open the file **`OneversusOneDiscriminantAnalysisstart.sas`**.

a. Recode the multiclass problem to a three-class problem as follows:

```
data bondrate2;
 set bondrate;
 if rating='AAA' or rating='AA' or rating='A' then rating=1;
 if rating='BAA' or rating='BA' or rating='B' then rating=2;
 if rating='C' then rating=3;
run;
Create a training set and test set as follows:
data training holdout;
 set bondrate2;
 if obs in (3,11,48,71,85,95) then output holdout;
 else output training;
run;
```

b. Create a data set OVSO1VS2, which contains only training set observations of class 1 or class 2 as follows:

```
data OvsO1vs2;
 set training;
 if rating=1 or rating=2;
run;
```

c. Create also the data sets OVSO1VS3 and OVSO2VS3.

1) Estimate a linear discriminant analysis classifier using the data set OVSO1VS2 as training set and the **`holdout set`** as test set. Write the output predictions to a new data set called PRED1.

```
proc discrim data=OvsO1vs2 testdata=holdout testout=pred1;
class rating;
var  LOPMAR     LFIXCHAR     LGEARRAT     LTDCAP     LLEVER
LCASHRAT     LACIDRAT     LCURRAT LRECTURN LASSLTD;
run;
```

2) Also estimate the other discriminant analysis classifiers for the data set OVSO1VS3 and OVSO2VS3, and write the test set predictions to the data sets PRED2 and PRED3, respectively.

d. Create a new data set PRED1VS2 as follows:

```
data pred1vs2 (keep=rating pred1vs2);
set pred1;
rename _INTO_=pred1vs2;
run;
```

e. Also create the data sets PRED1VS3 and PRED2VS3 in a similar manner.

f. Merge these data sets to a new data set ALLPREDS as follows:

```
data allpreds;
merge pred1vs2 pred1vs3 pred2vs3;
run;
```

1) Inspect the data set ALLPREDS to see the predictions of the One-versus-One classifiers. What is the classification accuracy?

g. **Extra:** Adapt the program you just wrote in order to implement a One-versus-All scheme

23. Rule extraction from support vector machines

Think about a method to turn a support vector machine classifier into a white-box model by extracting propositional rules mimicking its behavior.

24. Kaplan-Meier Analysis 1

Consider the following credit scoring data set:

Customer	Month	Default
1	17	1
2	27	1
3	30	0
4	32	1
5	43	1
6	43	1
7	47	1
8	52	1
9	52	0
10	52	0

a. Compute the Kaplan-Meier estimates manually by completing the following table:

Month	Customers at risk	Nr. of defaults	S(t)
<17			
17			
27			
30			
32			
43			
47			
52			

b. Also compute the Kaplan-Meier estimates in SAS and contrast them with your results. Plot the Kaplan-Meier curve.

25. Kaplan-Meier analysis 2

Consider the tab delimited data set INPUT.TXT. This is a credit scoring survival analysis data set with the following attributes:

Nr	Name	Explanation
1	Censor	1: customer defaulted 0: good customer
2	Censore	1: customer paid loan back early 0: customer did not pay back loan early
3	Open	For good customers: time of last observation; For bad customers: time of default; For customers who paid back early: time of early pay back
4	Age	Age of applicant
5	Amount	Amount of Loan
6	Curradd	Years at current address
7	Curremp	Years with current employer
8	Custgend	Gender of customer
9	Depchild	Number of dependent children
10	Freqpaid	Frequency of salary payments (for example, weekly, monthly, …)
11	Homephon	Has home phone or not
12	Insprem	Amount of insurance premium
13	Loantype	Type of Loan (single or joint)
14	Marstat	Marital status
15	Term	Term of loan
16	Homeowns	Has home or not
17	Purpose	Purpose of loan

The data set can thus be used for predicting both customer default and early repayment.

a. Enter the data in SAS using the following SAS code:

```
data creditsurv;
infile 'C:\CS\input.txt'   delimiter='';
input censor censore open age amount curradd curremp custgend
depchild freqpaid homephon insprem loantype marstat term homeowns
purpose;
run;
```

b. Estimate a Kaplan-Meier survival curve for predicting default using the following SAS code:

```
proc lifetest data=creditsurv plots=(s)graphics;
time open*censor(0);
symbol1 v=none;
run;
```

The PLOTS=(S) option also provides a graphical representation of the KM curve. The output contains a line for each observation. Censored observations are starred. The second column gives the KM estimates. No KM estimates are reported for the censored times. The column labeled Failure is just 1 minus the KM estimate, which is the probability of default prior to the specified time. The fourth column is an estimate of the standard error of the KM estimate, obtained by Greenwood's formula (Collet, 1994). This estimate can be used to construct confidence intervals. The column labeled Number Failed, is just the cumulative number of customers that defaulted before or on the time point considered. The last column labeled Number left is the number of customers at risk during the considered time point. Below the main table, you find the estimated 75th, 50th, and 25th percentiles (labeled Quantiles). For example., the 25th percentile is the smallest event time such that the probability of defaulting earlier is bigger than 0.25. Because, in our case the survival probabilities are never lower than 0.9, no values are reported for these percentiles.

1) Take a look at the KM graph. Why is the graph flat during the first three months?

c. If you want to have the KM estimates at specific timepoints, you can use the `timelist` option as follows:

```
proc lifetest data=creditsurv plots=(s) graphics
timelist=1,2,3,4,5,6,7,8,9,10;
time open*censor(0);
symbol1 v=none;
run;
```

d. Suppose you want to test whether customers owing a home or not have the same survival curve. The null hypothesis then becomes Ho: $S_1(t)=S_2(t)$ for all t. PROC LIFETEST provides three test statistics for doing this: the log-rank test (also known as the Mantel-Haenzel test), the Wilcoxon test, and the likelihood-ratio statistic, which is based on the assumption of exponentially distributed event times. You can ask for these tests as follows:

```
proc lifetest data=creditsurv plots=(s) graphics;
time open*censor(0);
symbol1 v=none;
strata homeowns;
run;
```

1) Is the difference significant or not? Also look at the graph with the 2 survival curves.

e. Conduct the same test statistics to check whether gender has an impact on the survival curve. Also check whether the fact that a person is younger or older than 30 has an impact on the survival curve (tip: use `strata age (30);`).

f. Repeat the Kaplan-Meier analysis for predicting early repayment using the censore indicator.

26. Parametric survival analysis

a. Read the data set INPUT.TXT into SAS.

1) Create a plot of $-\log(S(t))$ versus t for predicting default. Also create a plot of $\log[-\log(S(t))]$ versus t for predicting default. Do the plots support the assumption of an exponential or Weibull distribution of the survival times?

2) Use PROC LIFEREG to estimate a parametric survival analysis model (for predicting default) assuming exponentially distributed survival times. The independent variables are AGE, AMOUNT, CURRADD, CURREMP, CUSTGEND, DEPCHILD, FREQPAID, HOMEPHON, INSPREM, LOANTYPE, MARSTAT, TERM, HOMEONS. Be sure to make dummies for the variables CUSTGEND, FREQPAID, HOMEPHON, LOANTYPE, MARSTAT, and HOMEOWNS using the CLASS statement. Inspect the generated output. For a quantitative variable (for example, AGE, AMOUNT, etc.), we can use the formula $100(e^{\beta}-1)$ to estimate the percent increase in the expected survival time for each one-unit increase in the variable, holding the other variables constant. Calculate this number for the variables AGE and DEPCHILD. Is the time at current address important for predicting the survival time? Is the gender important for predicting the survival time?

3) Use PROC LIFEREG to estimate a parametric survival analysis model assuming Weibull, Log-normal, and Generalized gamma distributed survival times. Fill in the log-likelihood of each of these models in the following table:

Distribution	Log-likelihood
Exponential	
Weibull	
Log-normal	
Generalized gamma	

a) Which model gives the best likelihood?

4) Compute the likelihood-ratio statistic to compare the models by filling in the following table

Models	Likelihood ratio chi-square statistic	Degrees of freedom
Exponential versus Weibull		
Exponential versus generalized gamma		
Weibull versus generalised gamma		
Log-normal versus generalised gamma		

5) Using the OUTPUT option, predicted survival times can be generated for all observations in the data set. If you want to have a point estimate for the survival time of an individual observation, it makes sense to use the median survival time, i.e. t for which $S(t)=0.5$. This can be coded as follows:

```
proc lifereg data=creditsurv;
class custgend freqpaid homephon loantype marstat homeowns;
model open*censor(0)=age amount curradd curremp custgend depchild
freqpaid homephon insprem loantype marstat term homeowns
/dist=weibull;
output out=a p=median std=s;
run;
proc print data=a;
var open censor _prob_ median s;
run;
```

When inspecting the data set A, it becomes obvious that most of the predicted survival times are much bigger than the open variable. This is due to

- Heavy censoring of the data (no misclassification cost mechanism)
- Limited predictive power of the attributes
- The choice of the median survival time S(t) (compare with setting cut-off in classification models)

6) Repeat the parametric survival analysis for predicting early repayment using the censore indicator. Are the findings the same?

27. Proportional hazards regression

a. Estimate a proportional hazards model for the credit scoring data set for predicting default.

b. The first part of the output gives information on testing the Null hypothesis: Beta=0. The null hypothesis is that all β coefficients are 0. Three chi-square test statistics are given. Remember, the likelihood ratio statistic is based on the difference between –2 times the (partial) log-likelihood for the model with all covariates and the model with no covariates. Is the model significant?

c. The individual parameter estimates are then depicted. The Wald statistic can be computed by squaring the ratio of each coefficient to its estimated standard error: $(\beta/s(\beta))^2$ and has a chi-squared distribution with 1 degree of freedom. The risk ratio can be computed as exp(β). With the RISKLIMITS option, you can request confidence limits for the estimated hazard ratios.

1) If age increases with 1, what is the percentage increase or decrease of the hazard?

2) If age increases with 1, what is the impact on the survival probability?

3) By default, SAS used the Breslow approximation to estimate the parameters. Contrast the estimated parameters with those obtained from using the Efron approximation (use: ties=efron). Is there a big difference?

✐ PROC PHREG (in contrast to PROC LIFEREG) does note allow a class statement to create dummies for the categorical variables. You have to create the categorical variables yourself in an extra data step.

4) You can also use PROC PHREG to generate predictions for individual observations. Start with creating a test set as follows:

```
data testset;
 input age amount curradd curremp;
 datalines;
 29 3000 1 2
run;
```

5) Estimate a proportional hazards model as follows:

```
proc phreg data=creditsurv;
model open*censor(0)=age amount curradd curremp /ties=efron;
baseline out=a covariates=testset survival=s lower=lcl upper=ucl
/nomean;
run;
proc print data=a;
run;
```

6) The BASELINE statement allows to ask for survival predictions of individual observations. The NOMEAN option suppresses the output of survivor estimates evaluated at the mean values of the covariates, which are otherwise included by default. The LOWER= and UPPER= options give 95-percent confidence intervals around the survival probability. Plot the survivor function for this specific observation.

7) PROC PHREG also allows to do input selection using the SELECTION=backward/forward/stepwise option. Perform all three types of input selection and look at the impact on the results (use `slentry=0.01 slstay=0.01`).

8) Repeat the analysis above for predicting early repayment.

Appendix B References

B.1 References

Allison, P. D. 1995. *Survival Analysis Using SAS: A Practical Guide.* Cary, NC: SAS Institute Inc.

Allison, P. D. 1999. *Logistic Regression Using the SAS System: Theory and Application*. Cary, NC: SAS Institute.

Araten M, M. Jacobs, and P. Varshney. 2004. "Measuring LGD on Commercial Loans: An 18-Year Internal Study." JP Morgan Chase.

Baesens, B. 2003. "Developing Intelligent Systems for Credit Scoring Using Machine Learning Techniques." Ph.D. diss., K.U.Leuven, Belgium.

Baesens, B., R. Setiono, C. Mues, and J. Vanthienen. "Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation." *Management Science* 49, no. 3(March 2003): 312–29.

Baesens, B., T. Van Gestel, S. Viaene., M. Stepanova, J. Suykens, and J. Vanthienen.. "Benchmarking State of the Art Classification Algorithms for Credit Scoring." *Journal of the Operational Research Society* 54, no. 6(2003): 627–35.

Baesens, B., T. Van Gestel, Stepanova M., Van den Poel D., Vanthienen J. "Neural Network Survival Analysis for Personal Loan Data." *Journal of the Operational Research Society* (Special Issue on Credit Scoring) 59, no. 9(2005): 1089–98.

Banasik, J., J.N. Crook., and L.C. Thomas. "Sample Selection Bias in Credit Scoring Models." *Proceedings of the Seventh Conference on Credit Scoring and Credit Control (CSCCVII'2001)*, Edinburgh, Scotland (2001).

Basel Committee on Banking Supervision. "Stress Testing at Major Financial Institutions: Survey Results and Practice." BIS. 2005.

Basel Committee on Banking Supervision. "Studies on the Validation of Internal Rating Systems." BIS Working Paper 14. 2006.

Basel Committee on Banking Supervision. "International Convergence of Capital Measurement and Capital Standards: A Revised Framework Comprehensive." June 2006.

Basel Committee on Banking Supervision Newsletter. "Validation of Low-Default Portfolios in the Basel II Framework." no.6: September 2005.

Benjamin, N., A. Cathcart, and K. Ryan. "Low Default Portfolios: A Proposal for Conservative Estimation of Default Probabilities." Financial Services Authority. April 2006.

Breiman, L., J. Friedman, R. Olsen, and C. Stone. 1984. *Classification and Regression Trees.* CA, USA Wadsworth International.

Cespedes, Credit Risk Modeling and Basel II, *Algo Research Quarterly* 5(no. 1): Spring 2002. Dwyer D.W., The Moody's KMV EDF RiskCalc. 3.1: Model Next-Generation Technology for Predicting Private Firm Credit Risk. Moody's KMV. 2004.

Committee of European Banking Supervisors (CEBS). CP10: Guidelines on the Implementation, Validation, and Assessment of Advanced Measurement (AMA) and Internal Ratings Based (IRB) Approaches. October 2005.

"Credit Stress-Testing." 2002. Consultative paper, Monetary Authority of Singapore (MAS).

Edward, A., A. Resti, and A. Sironi. 2005. *Recovery Risk: The Next Challenge in Credit Risk Management*. Risk Books.

Engelmann, Hayden, and Tasche. "Measuring the Discriminative Power of Rating Systems." Deutsche Bundesbank. 2003.

Fawcett T. 2003. "ROC Graphs: Notes and Practical Considerations for Researchers." HP Labs Tech Report HPL-2003-4. 2003.

Financial Services Authorithy (FSA), CP06/3: Strengthening Capital Standards 2. February 2006.

Financial Services Authorithy (FSA), CP05/3: Strengthening Capital Standards 2. January 2005.

Freed, N. and F. Glover. "Resolving Certain Difficulties and Improving the Classification Power of LP Discriminant Analysis Formulations." *Decision Sciences* 17(1986): 589–595.

Fritz. "Validating Inputs and Outputs of the Regulatory and Economic Capital Allocation Process." Deutsche Bank. 2003.

Grablowsky, B.J. and W.K. Talley. "Probit and Discriminant Factors for Classifying Credit Applicants: A Comparison." *Journal of Economics and Business* 33(1981): 254–61.

Gupton and Stein. LossCalc. 2: Dynamic Prediction of LGD. Moody's KMV. 2005.

Hall, M. and L. Smith. "Practical Feature Subset Selection for Machine Learning," *Proceedings of the Australian Computer Science Conference* (University of Western Australia). 1996.

Hand, D. J. 2003. "Crime, Statistics, and Behaviour." *Statistics, Science, and Public Policy* VIII: Science, Ethics, and the Law, eds. A.M.Herzberg and R.W.Oldford. April 23–26(2003):181–187. Queen's University, Canada.

Hand, D. J. and S. Jacka, eds. 1998. Statistics in Finance, Edward Arnold.

Hand, D. J. and W. E. Henley. "Statistical Classification Methods in Consumer Credit Scoring: A Review." *Journal of the Royal Statistical Society*, Series A 160 3 (1997):523–41.

Hand, D. J., H. Mannila, and P. Smyth. 2001. *Principles of Data Mining*, MIT Press.

Hand, D. and R. J. Till. "A Simple Generalization of the Area under the ROC Curve to Multiple Class Classification Problems." *Machine Learning* 45, no.2(2001):171–86.

Hanley, J. A. and B. J. McNeil. "The Meaning and Use of Area under the ROC Curve." *Radiology* 143(1982):29–36.

Hartigan, J. A. 1975. *Clustering Algorithms*. New York: Wiley.

Henley, W. E. and D. J. Hand. 1997. "Construction of a *k*-nearest Neighbour Credit Scoring System." *IMA Journal of Mathematics Applied in Business and Industry* 8:305–21.

Hong Kong Monetary Authority (HKMA). "Validating Risk Rating Systems under the IRB Approaches." February 2006.

Joseph, M. P. "A PD Validation Framework for Basel II Internal Ratings-Based Systems." Commonwealth Bank of Australia. 2005.

Kelly, M. G. 1998. "Tackling Change and Uncertainty in Credit Scoring." Unpublished Ph.D. diss., The Open University, Milton Keynes, UK.

Mangasarian, O. L. "Linear and Non-linear Separation of Patterns by Linear Programming." *Operations Research*. 13(May–June 1965):444–52.

Merton, R. C. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29:449–70.

Pluto, K. and D. Tasche. "Estimating Probabilities of Default for Low Default Portfolios." 2005.

Quinlan, J. R. 1993. C4.5: *Programs for Machine Learning.* Morgan Kauffman.

Schuermann and Jaffry. "Measurement, Estimation, and Comparison of Credit Migration Matrices." *Journal of Banking and Finance*. 2004.

Thomas, L., D. Edelman, and J. Crook. "Credit Scoring and Credit Control." Philadelphia: Society for Industrial and Applied Mathematics (SIAM). 2002.

Thomas, L. C. "A Survey of Credit and Behavioural Scoring; Forecasting Financial Risk of Lending to Consumers." *International Journal of Forecasting*, 16(2000):149–172.

Thomas, L. C., J. Ho, and W.T. Scherer. "Time Will Tell: Behavioural Scoring and the Dynamics of Consumer Risk Assessment," IMA J. of Management Mathematics 12(2001):89-103.

Van Gestel, T. "From Linear to Kernel Based Methods in Classification, Modelling, and Prediction." ESAT-SCD-SISTA, Katholieke Universiteit Leuven, Belgium. May 3, 2002.

Van Gestel T., B. Baesens, P. Van Dijcke, J. Suykens, J.Garcia, and T. Alderweireld. "Linear and Nonlinear Credit Scoring by Combining Logistic Regression and Support Vector Machines." *Journal of Credit Risk* 1, no. 4(2005).

Van Gestel T., B. Baesens, P. Van Dijcke, J.Garcia, J. A. K. Suykens, and J. Vanthienen. " A Process Model to Develop an Internal Eating System: Sovereign Credit Ratings." Decision Support Systems 42, no. 2(2006):1131–51.

Vasicek O. 1987. "Probability of Loss on Loan Portfolio." KMV Corporation.

Vasicek, O. 1991. "Limiting Loan Loss Probability Distribution." KMV Corporation.