Moody's Risk Management Services

Rare Event Modeling and Validation Through Time:

The case of corporate credit analysis

Roger M. Stein

Managing Director

Credit Risk Analytics

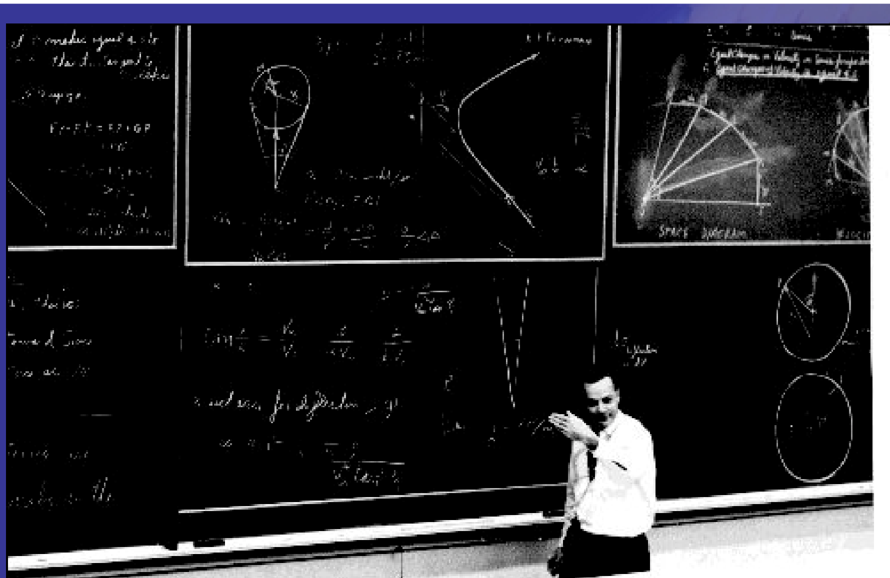
Moody's Risk Management Services

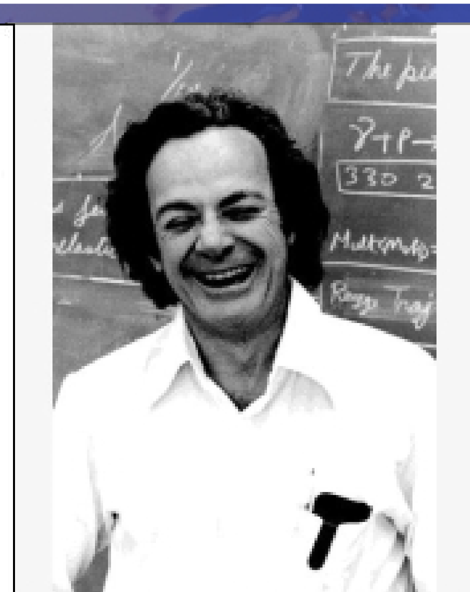
steinr@moodys.com

Moody's Risk Management Services

The first principle is that you must not fool yourself -- and you are the easiest person to fool.

- Richard Feynman





Overview

- Brief introduction to the RiskCalc default model
- Discussion of validation and backtesting in finance
- Differences between validating market- and credit-related models
- A validation approach for sparse data sets
- Examples of problems that arise from violating the approach
- Conclusion
- Research Co-authors
  - V. Dhar
  - S. Keenan
  - J. Sobehart

The corporate credit problem

- What is the probability of “default” (PD) within a specified period of time?
- Uses of PD’s
  - Regulators
    - Basel, National bank regulators
  - Securitization
    - Collateralized Loan Obligations
  - Credit Process
    - Decisioning (yes/no)
    - Monitoring (work-out, remedial action)
    - Provisioning
    - Pricing
    - Incentive compensation
- Related problems
  - Recovery (loss given default)
  - Correlation of default rates and arrival times
  - State transiton

RiskCalc™ modeling approach

- Transform
  - Ratios transformed from unwieldy broad distributions to more uniform and predictive variants
  - Micro-modeling used to capture useful aspects of behavior and to decompose problem
- Model
  - Transformed variables weighted statistically to produce default scoring model
- Map
  - Model output (score) converted to PD by non-parametrically mapping into historical population default estimates

The components of Moody's modeling approach

- Structural model (Merton variant)
  - Distance to distress
- Rating information (where available)
  - Moody's rating or quantitative rating estimate
- Financial statement information
  - Leverage, Liquidity, Size, Profitability, etc.
- Non-linear statistical regression
  - Simple neural network
- Mapping result to empirical probability of default (PD) and adjusting for prior probabilites
- Extensive validation
  - Out-of-sample / out-of-time (walk forward analysis)
  - Multidimensional metrics

Distance to Distress: Equity as a Call Option

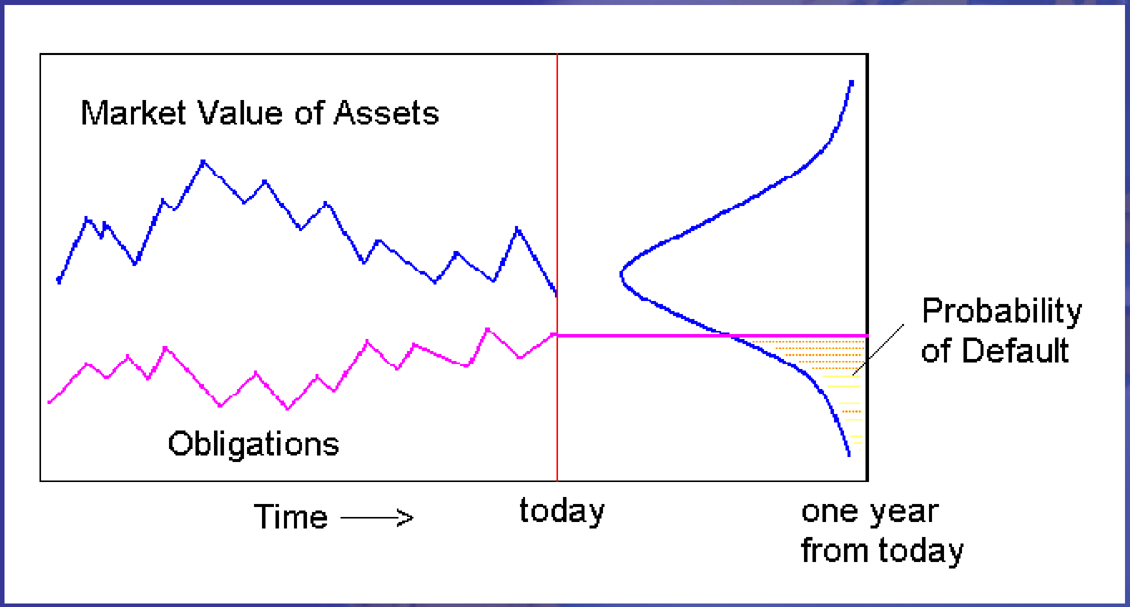
1. Calculate the firm's obligations (CL, LTD)
2. Use equity information to estimate:
   - a) market value of the firm's assets (MVA)
   - b) volatility of assets

   This is done with a variant of the Merton model:

   Market Equity = **Present Value (Residual Value of the Firm)**

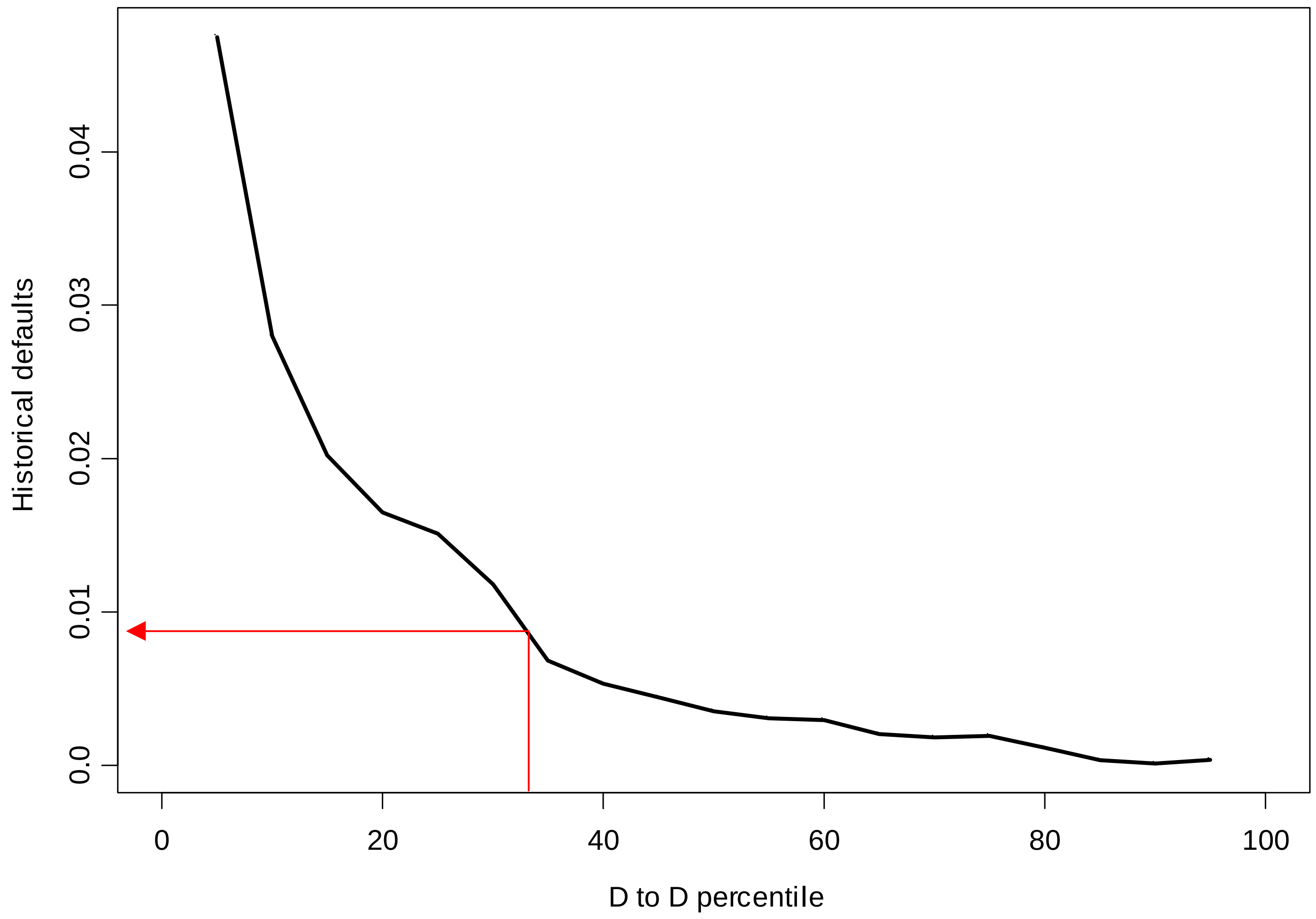
   Stock Volatility = **Leveraged Volatility of Assets**
3. Calculate

   Distance to Distress = (MVA -(1/2 LTD + CL))/(volatility x MVA)

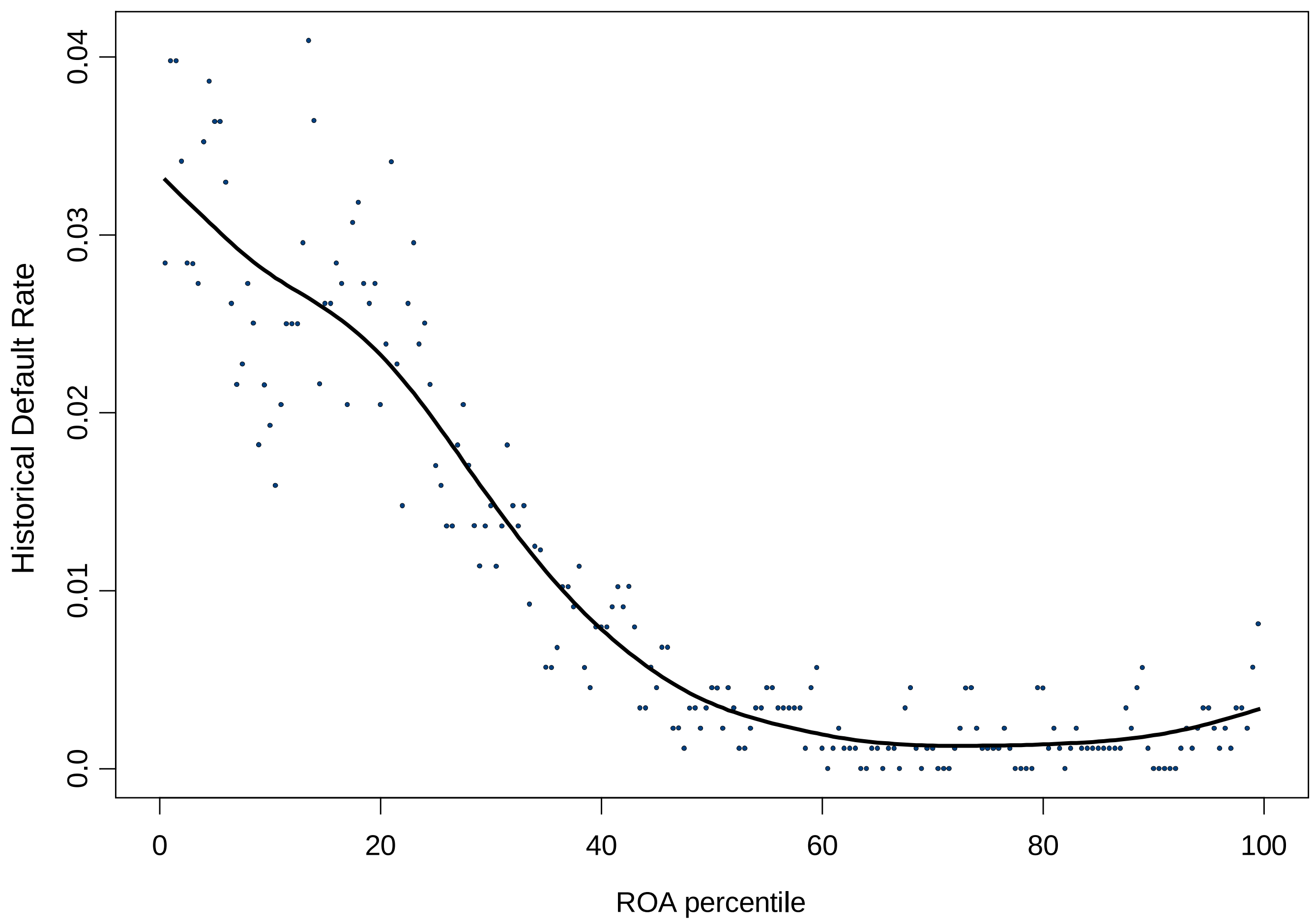


Moody's Risk Management Services

Mapping score to PD

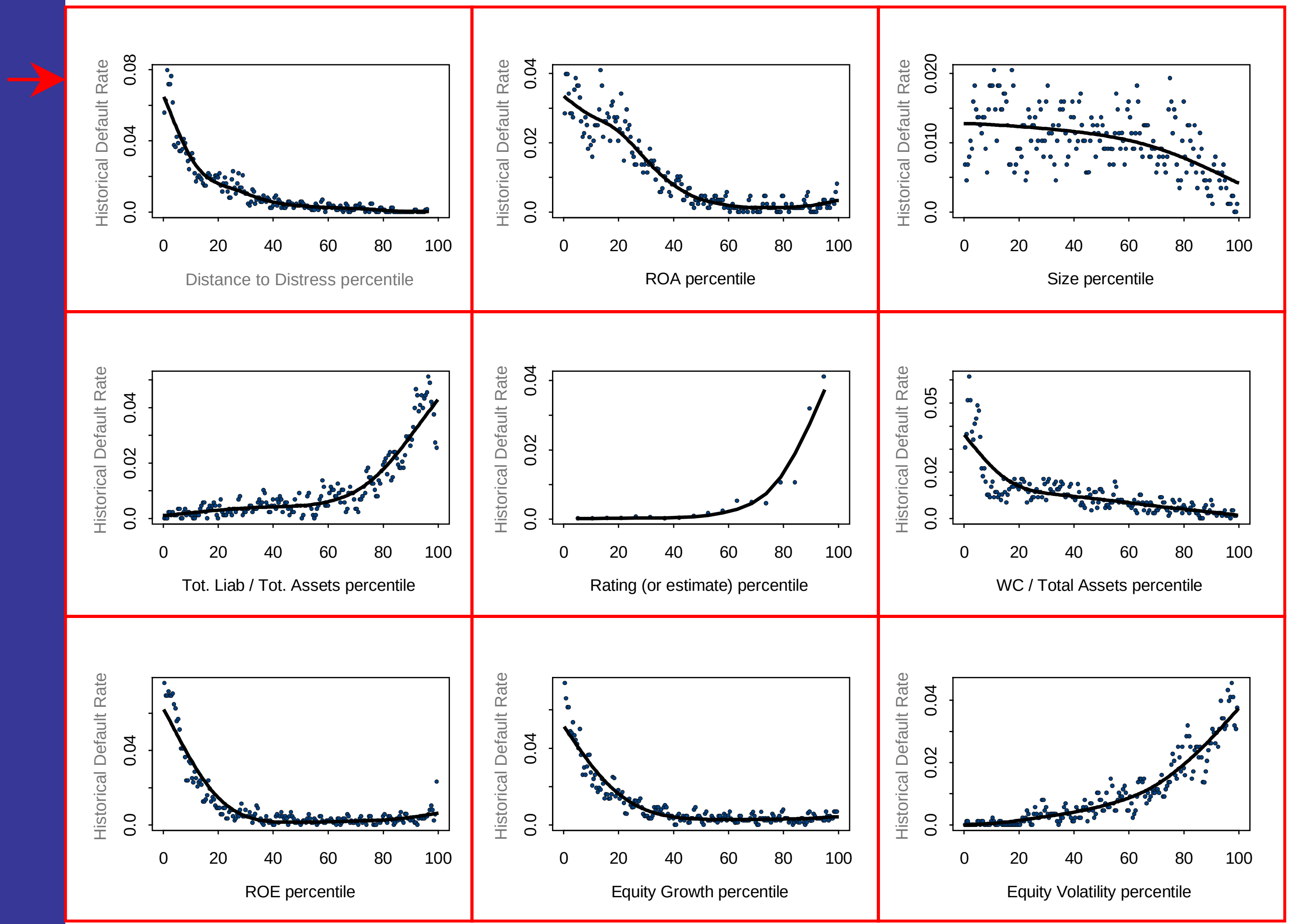


Predictive power of financial



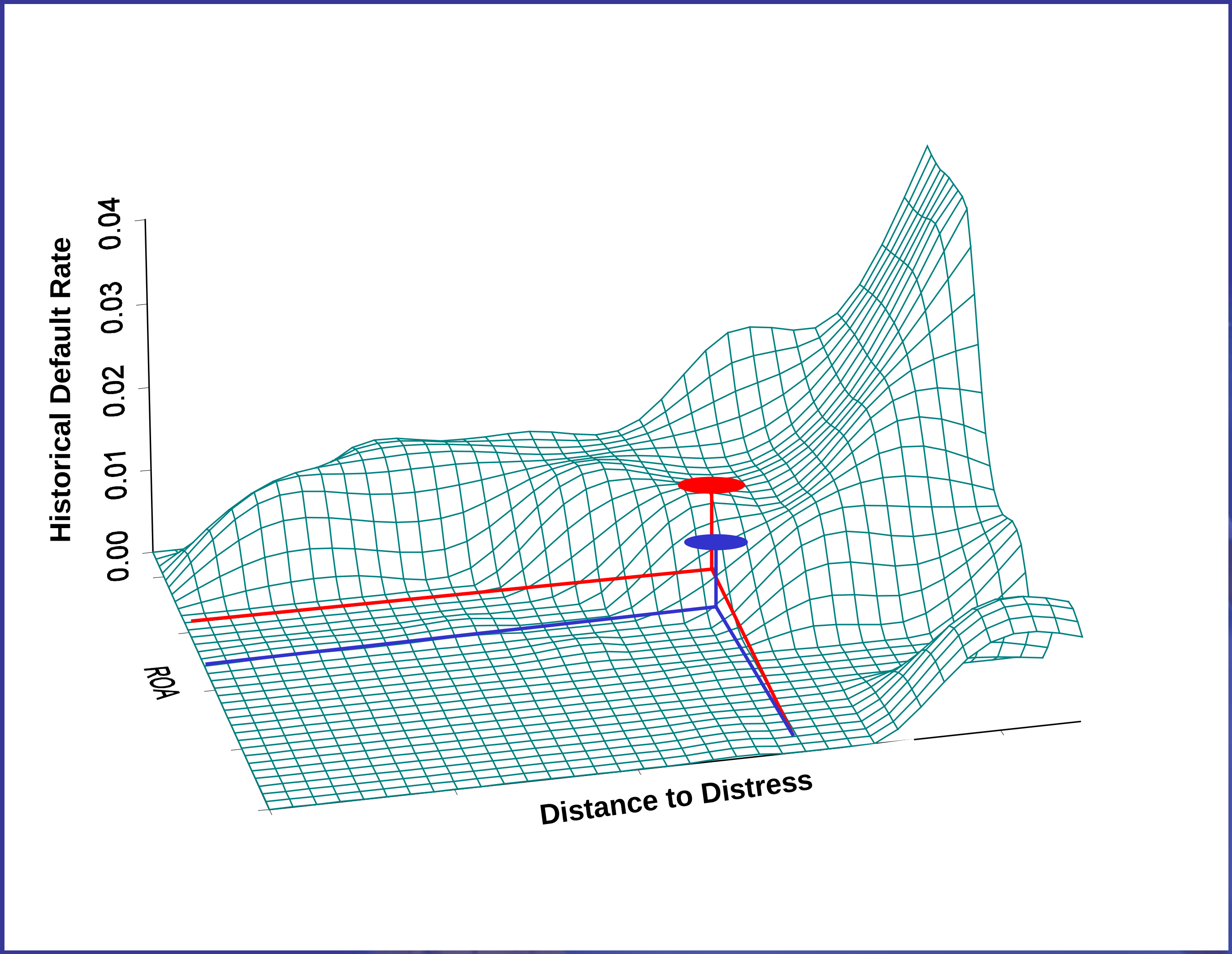
Moody's Risk Management Services

Univariate Performance of Variables

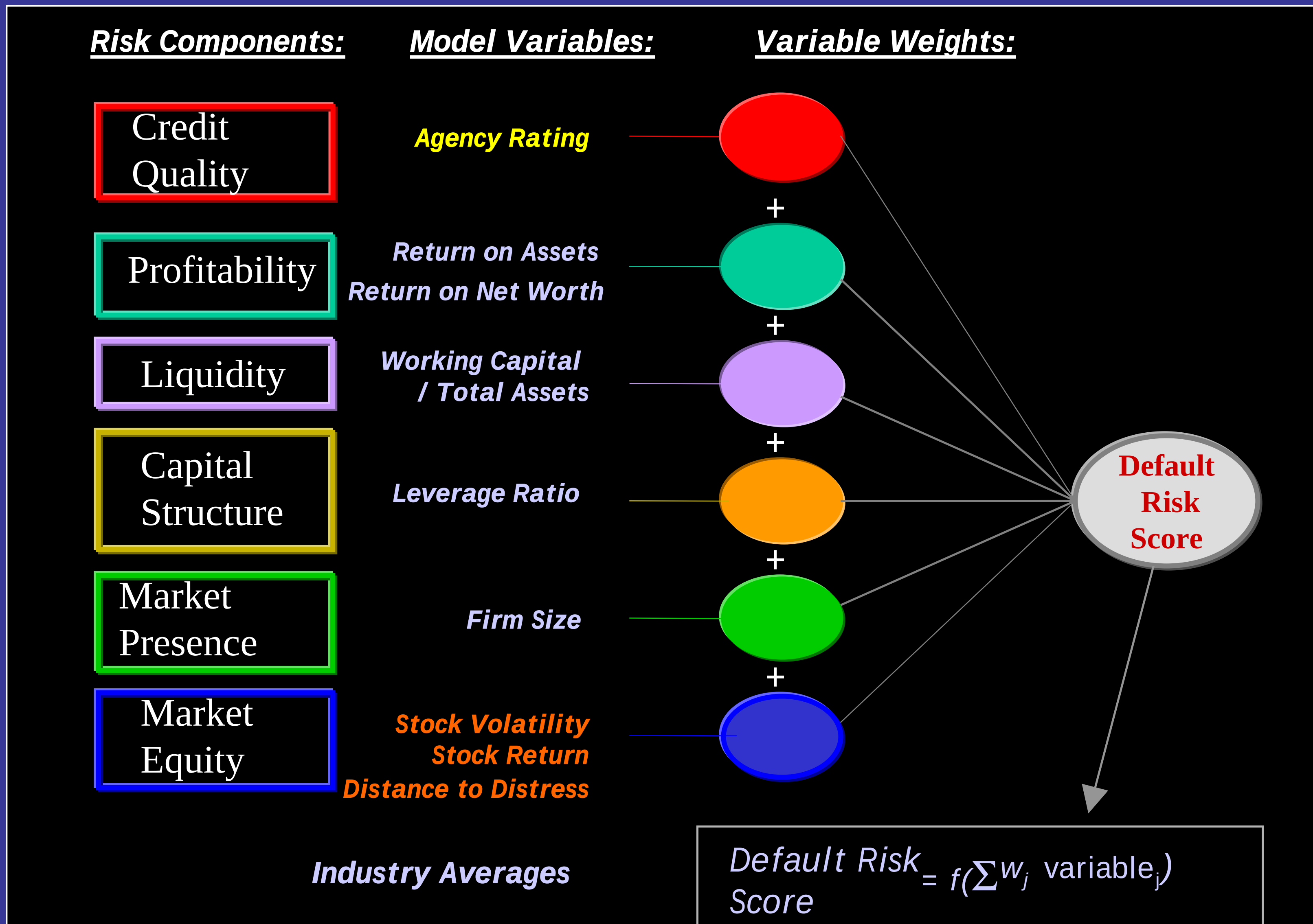


Moody's Risk Management Services

Non-Linear Relationships: ROA vs. Distance to Distress



Heuristic overview of the model



Validation Methodology

Moody's Risk Management Services

Moody’s view of the spectrum of validation

Development
Data Poor

Certification
Data Rich

Anecdotal cases

Validating on small samples
of “training” cases
“number right”

Validating on
out-of-sample data
“number right”

Validating on
out-of-sample
out-of-time data
“number right”

Bootstrapping
out-of-sample
out-of-time data
“number right”

Bootstrapping
out-of-sample
out-of-time data
higher order statistics

Bootstrapping
out-of-sample
out-of-time data
higher order statistics
w/ cost function

Moody’s Q is currently here

Moody’s Risk Management Services

Validation

"...the area of validation will prove to be a key challenge for banking institutions in the foreseeable future."

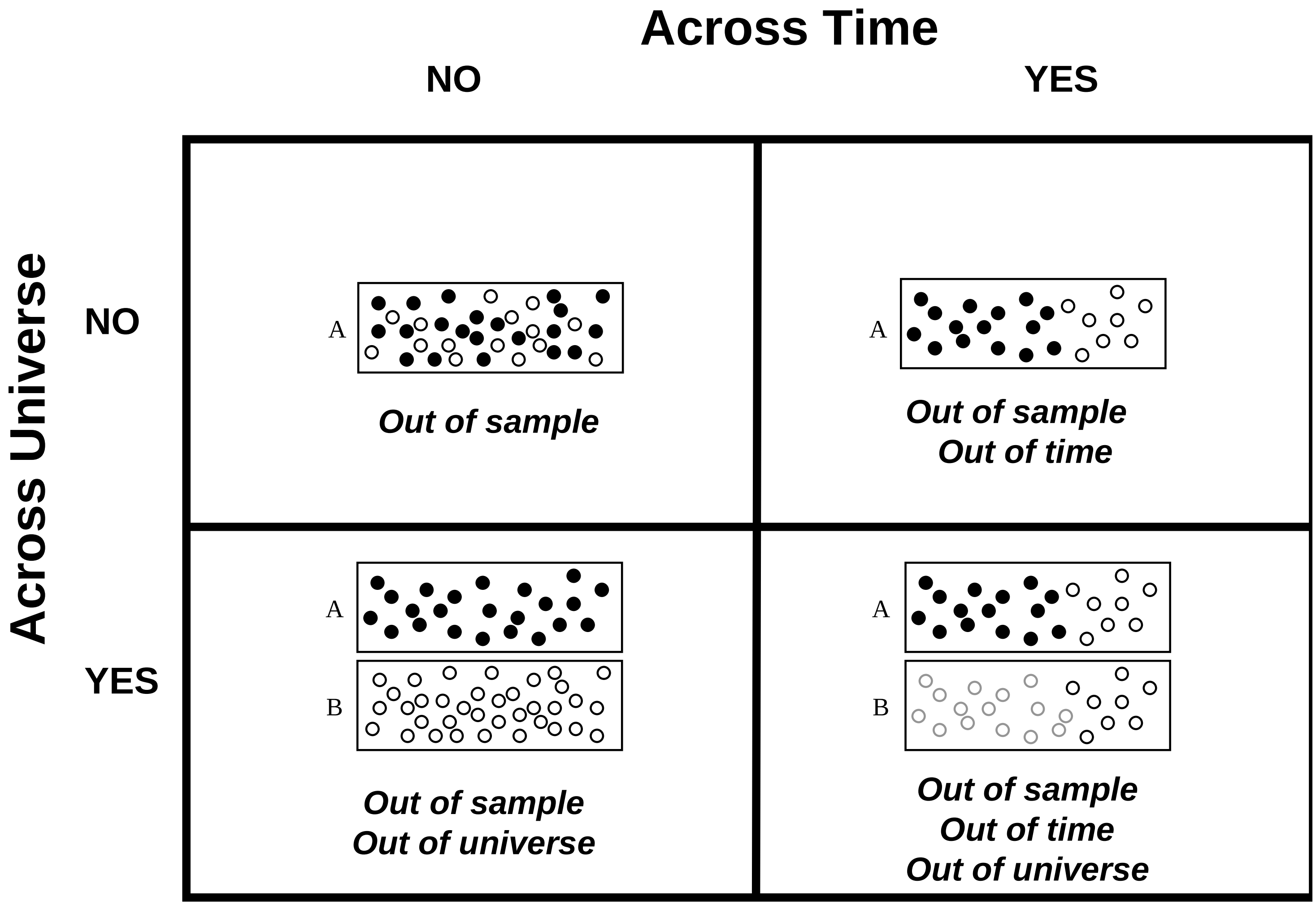
"Credit Risk Modeling Practices and Applications,"
***Basle Committee on Banking and Supervision*, Basle,**
April,1999, p. 50.

The components of our current approach to validation

- How to measure and calculate performance statistics
  - How to sample available data
  - How to use the data to achieve robust statistics
- What types of statistics to measure
  - Simple (hits vs. misses)
  - Measures of goodness based on geometry
  - Measures of information content and association based on entropy
  - Other measures (forthcoming)

Validation in Finance

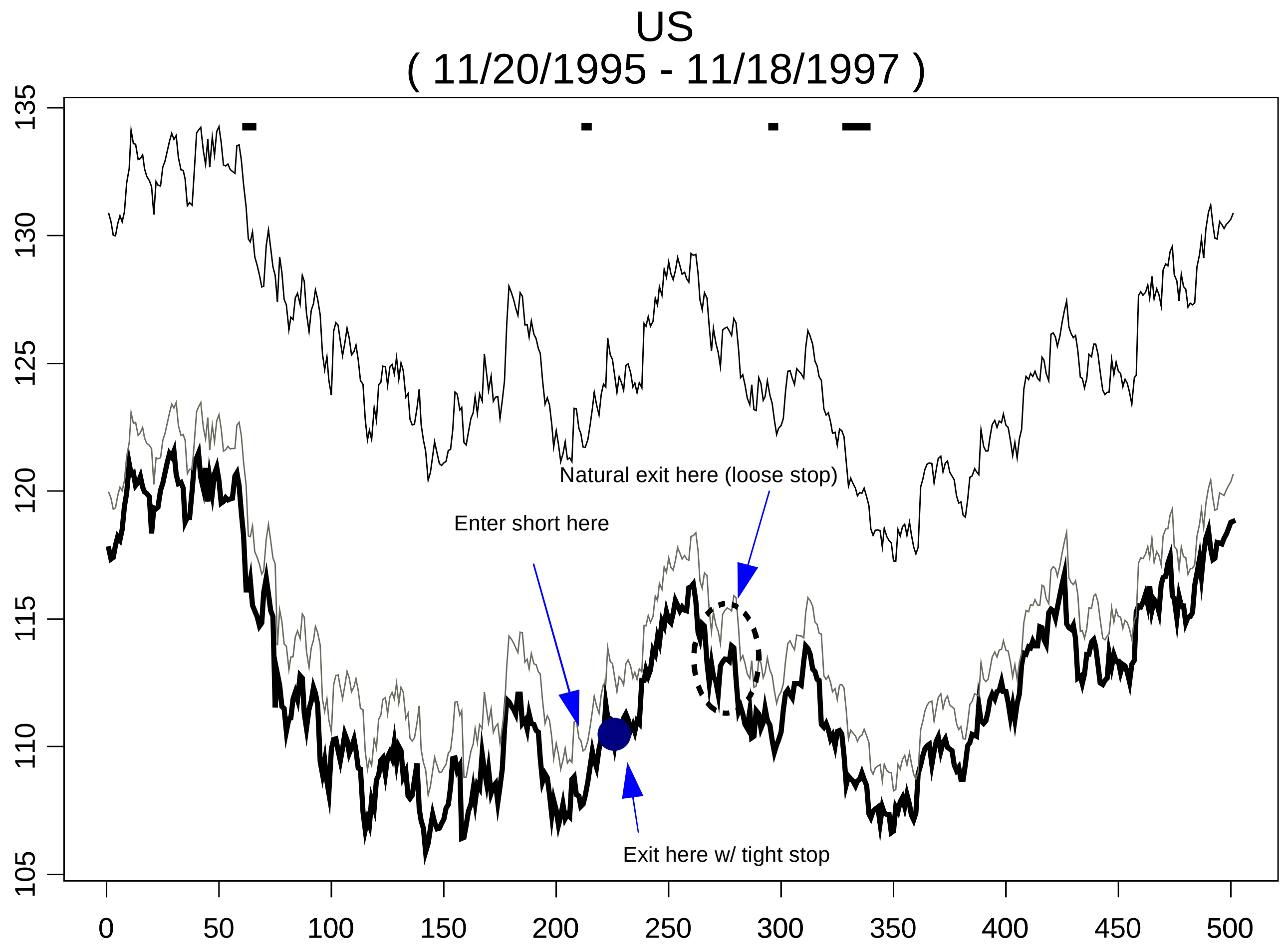
- Backtesting dominates market research
  – Identify interesting relationship
  – Evaluate the (risk-adjusted) “profitability” of the relationship through simulated trading on historical data
- Backtesting requires long time series of relatively high frequency
- Backtesting is often not appropriate for lower frequency data or rare/long term events since not enough data exists to both build a model and test it
  - If more data are saved for testing, models tend to be mis-specified
  - If more data are used to parameterize a model, tests loose power: too few examples exist for meaningful inferences



From: Dhar, V. and Stein, R., “Finding Robust and Usable Models with Data Mining: Examples from Finance,” PCAI, Sept., 1998.

Findings: Market character

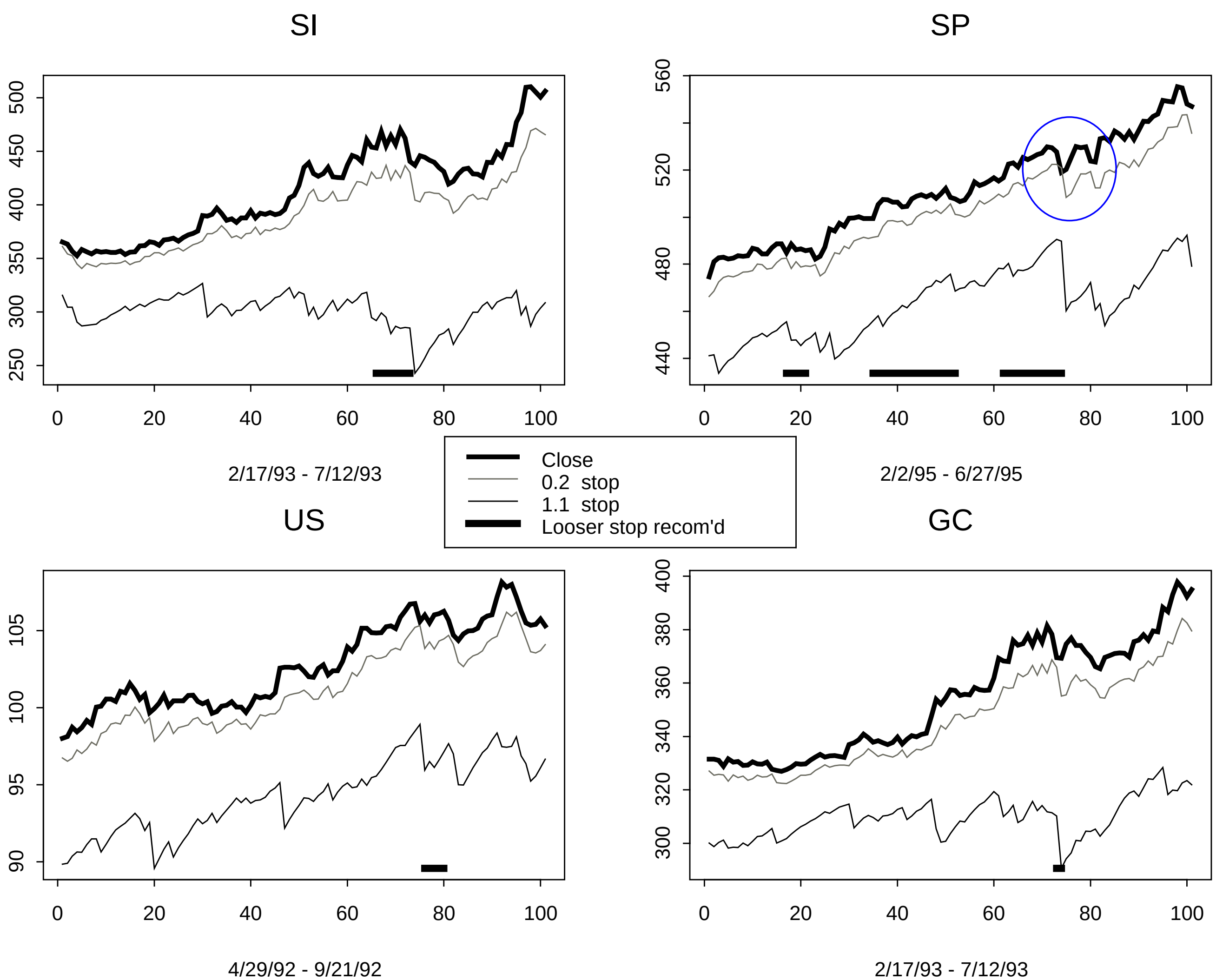
Stop level and market character in favorable systems



Close-up of Cond.omni.60.08.01 for short trades on DM

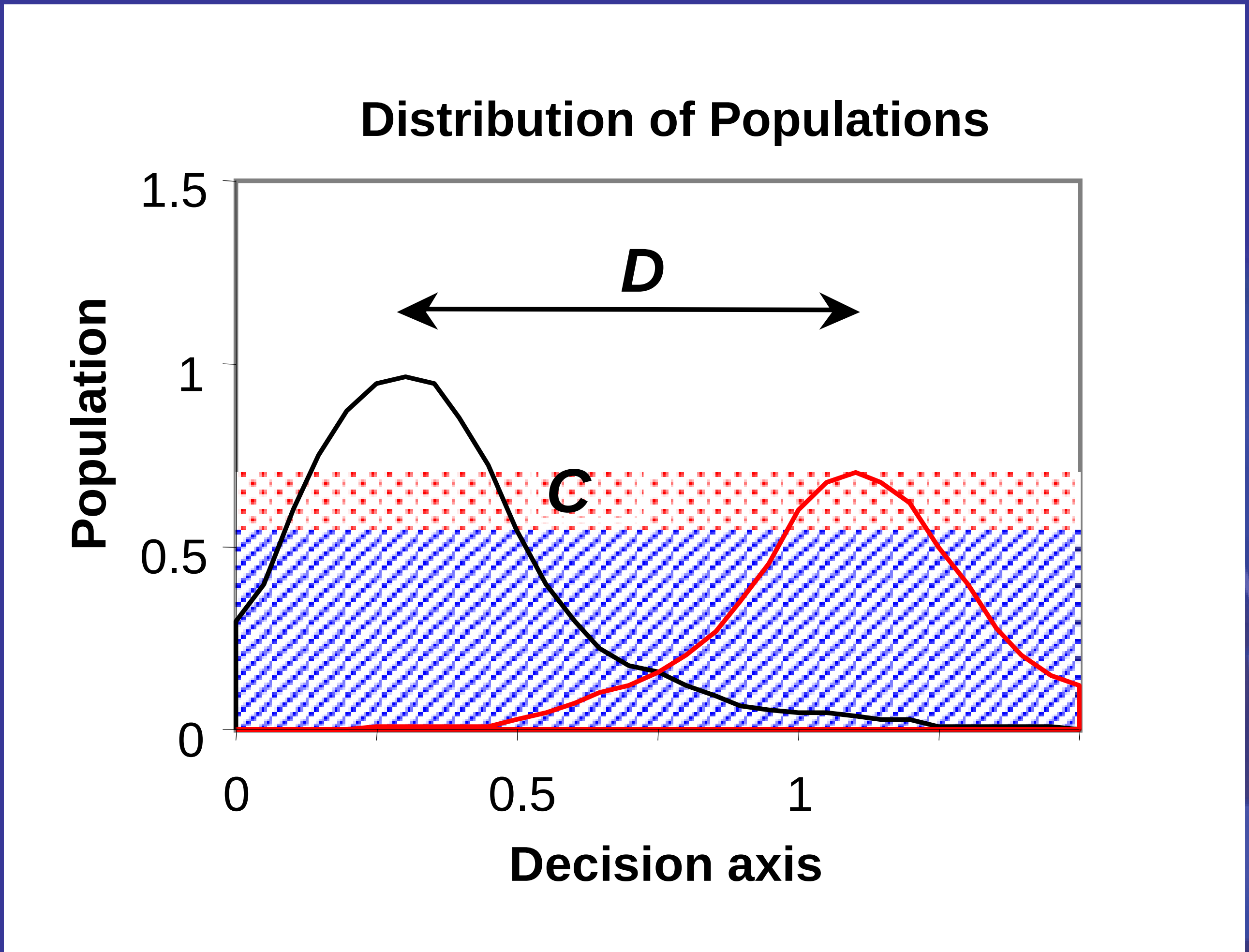
Findings: Market character

Stop level and market character in random systems

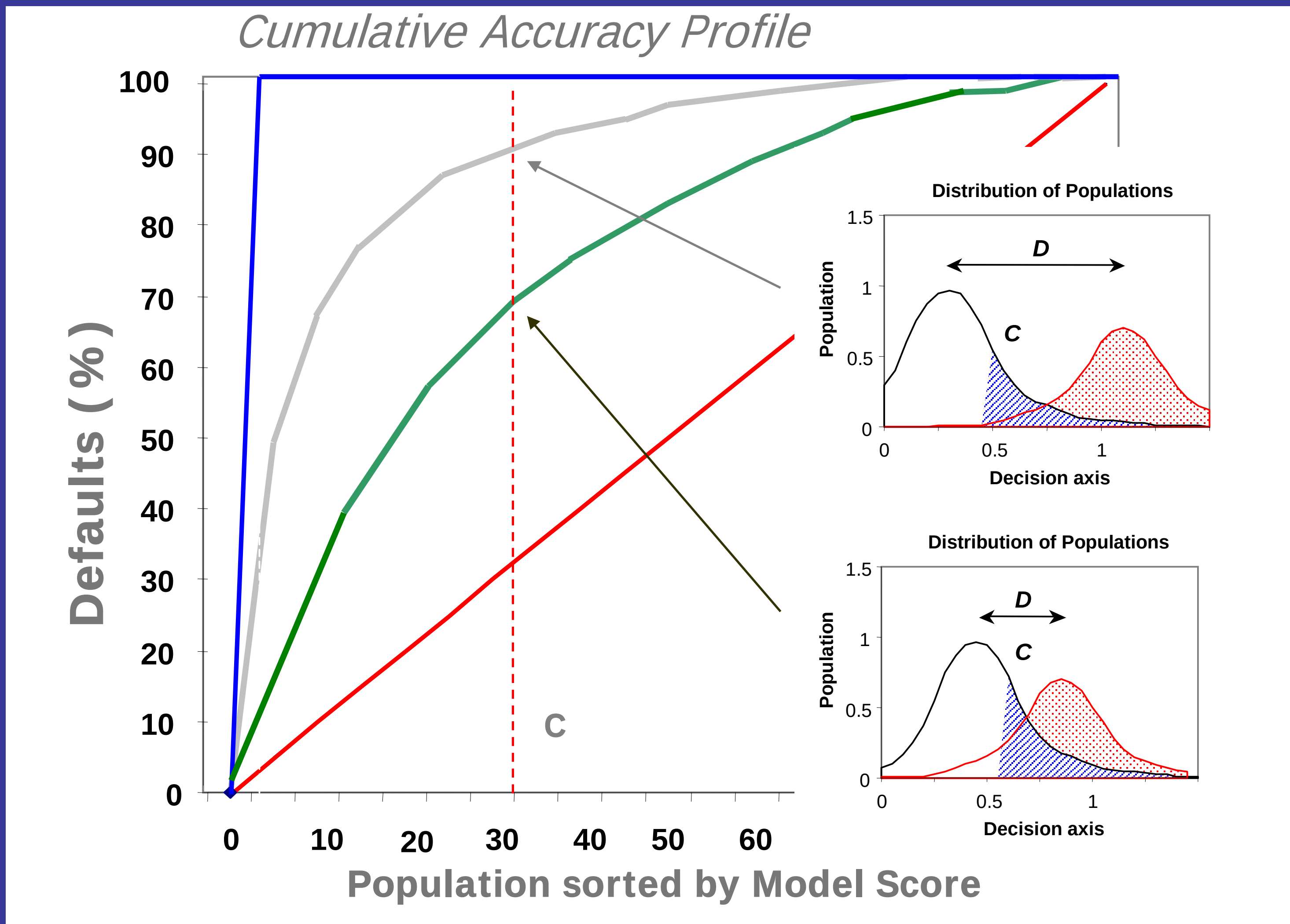


Looser stops avoid minor retracements when Cond.e.v.v.10.9.1 obtains (long)

Unlike in the trading problem, corporate credit often involves separating “goods” from “bads”



Model Performance: Cumulative Accuracy



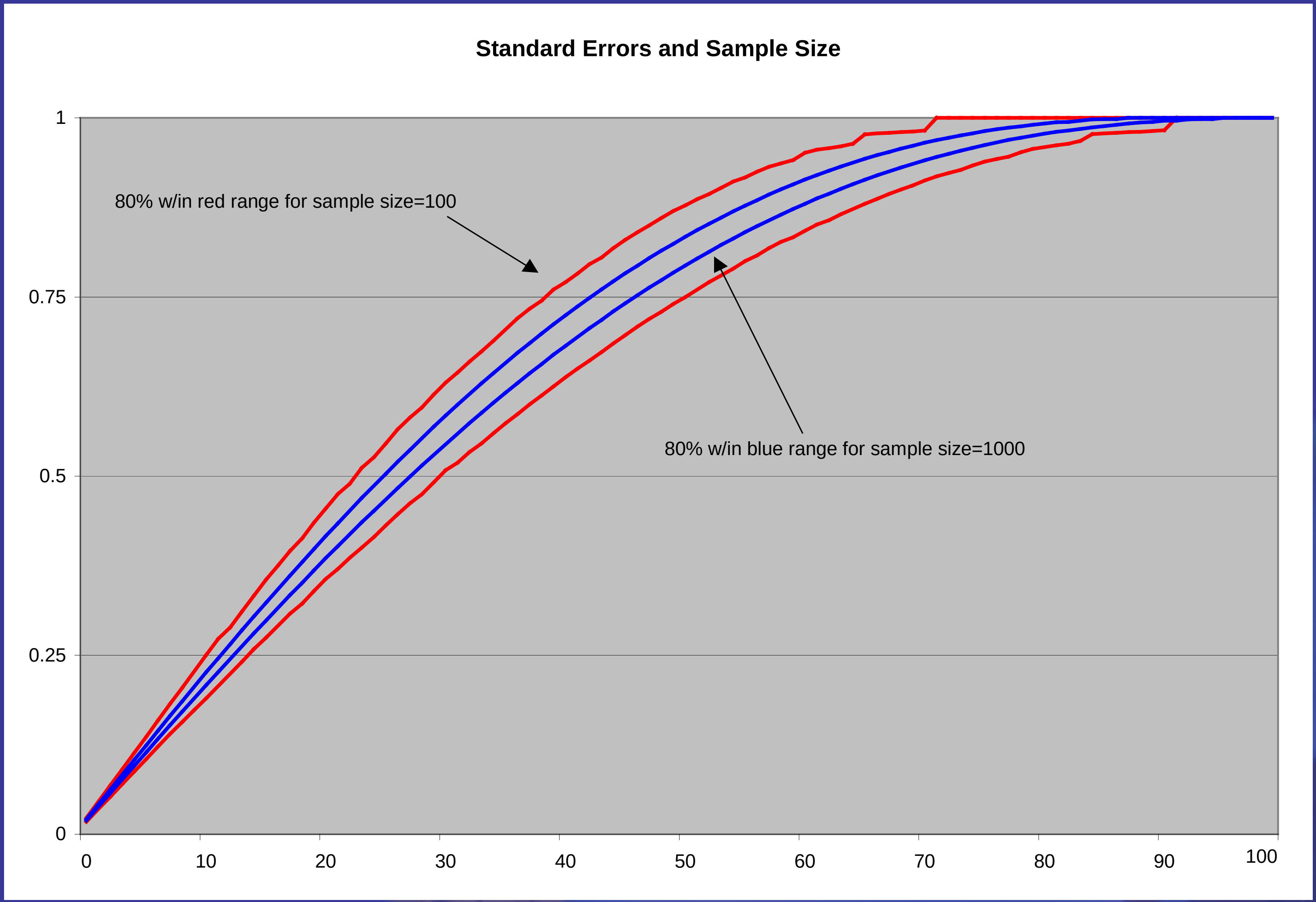
High Risk Low Risk

Moody's Risk Management Services

But the distribution of “interesting” cases for the default problem is sparse

- Data Set
  - Moody’s Default and Ratings databases, Compustat, IDC
  - over 14,000 U.S. non-financial corporations
  - over 1,400 defaults
  - 1980 through 1999
- Firm years
  - Model Fitting: ~ 100,000
  - Validation: ~ 65,000
- Population default rate:
  - Actual: 1.6%
  - Sample: 1.1%

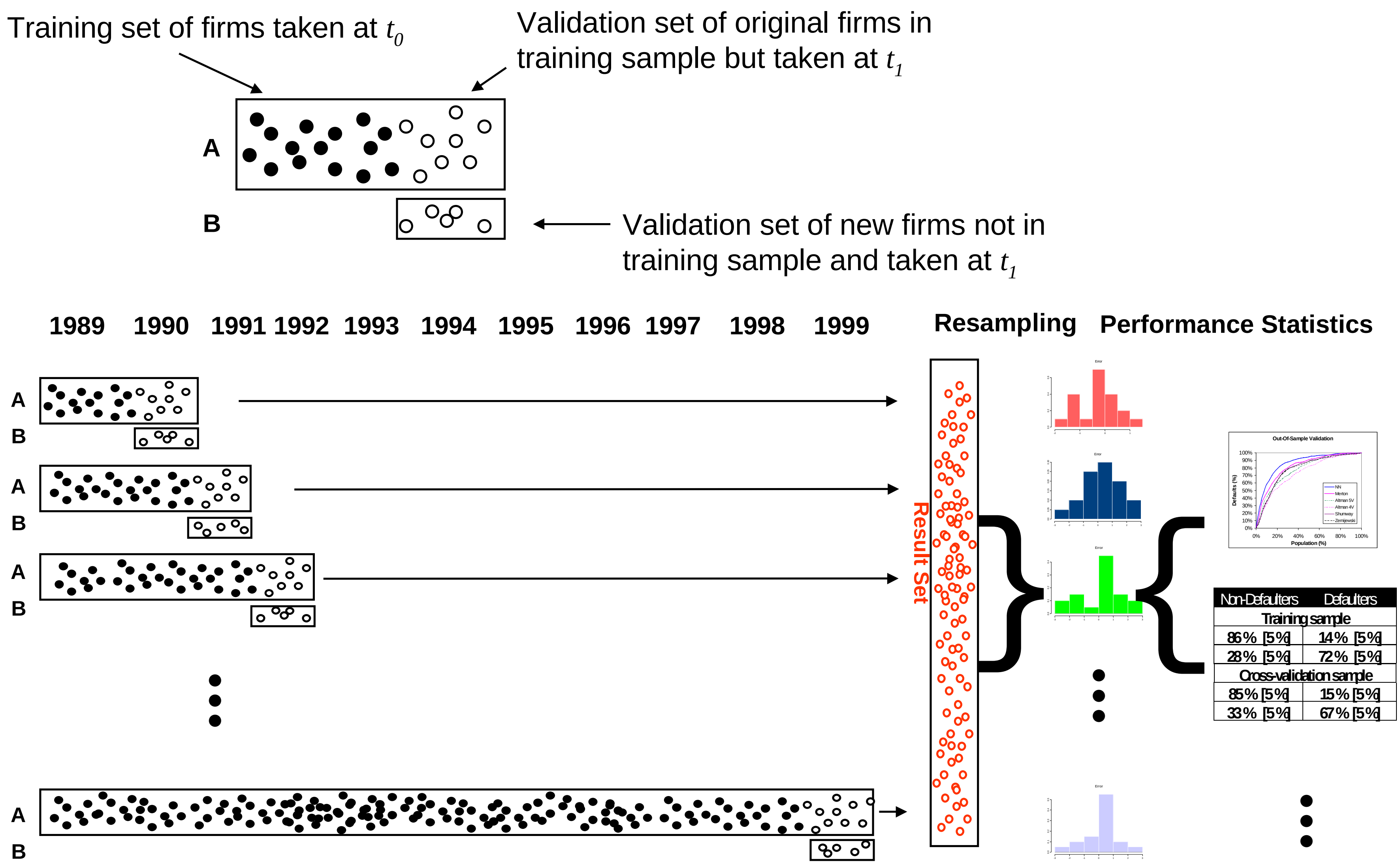
Power and Sample Size Related



Moody's Risk Management Services

Model Validation and Performance

- **Walk forward and K-fold methods**
  - Training sample versus validation sample
  - Out-of-sample and out-of-time validation
- **Empirical validation versus comparable tools**
  - Power statistics are sample biased
  - Performance can be truly assessed relative to a benchmark
  - Muti-dimensional performance measures
- **Use of large datasets**
  - Documented performance on large out-of-sample datasets
  - Testing that the model is not “overfitted”

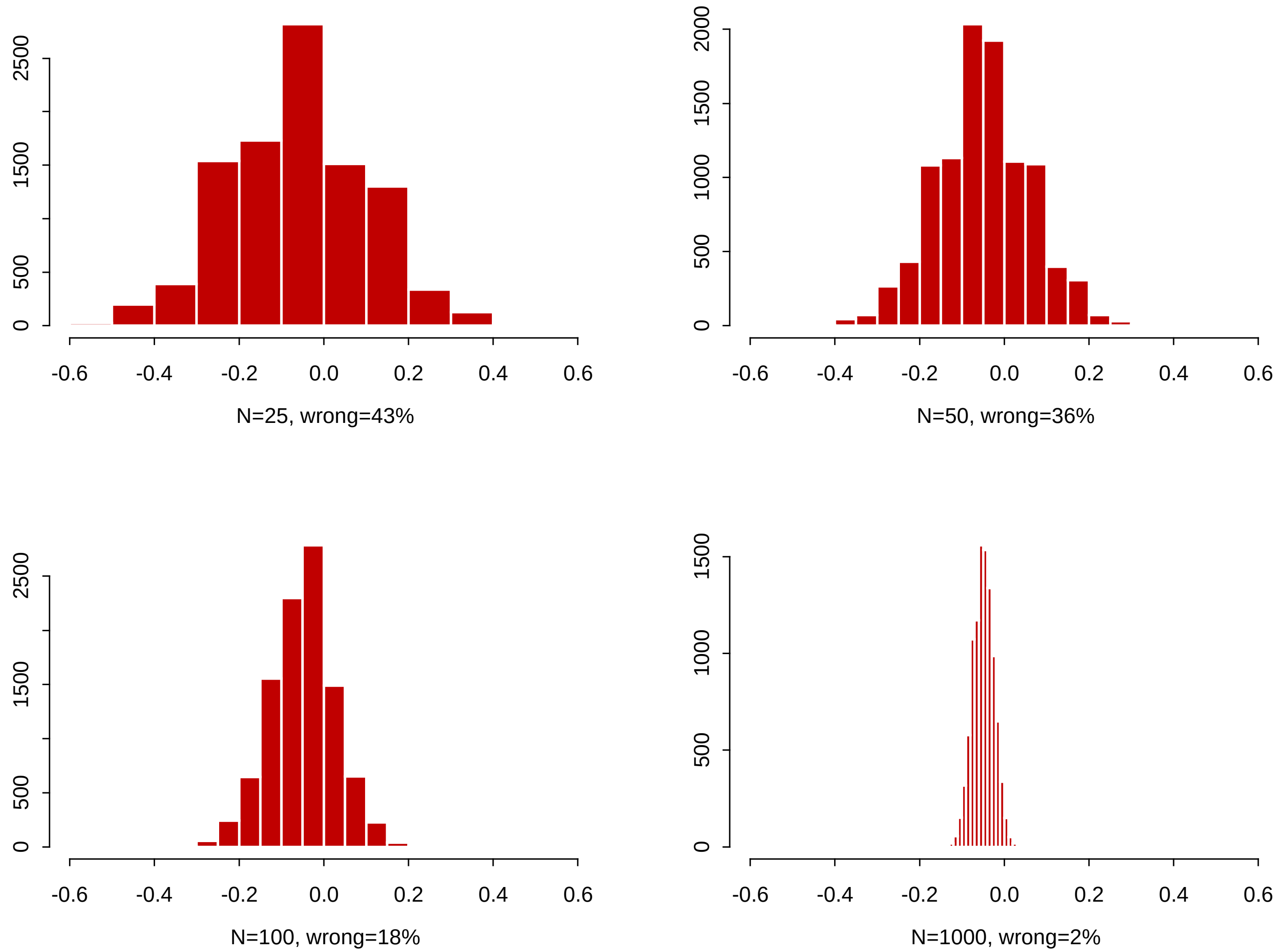


Training set of firms taken at t_0
Validation set of original firms in training sample but taken at t_1
A
B
Validation set of new firms not in training sample and taken at t_1
1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999
Resampling
Performance Statistics
Result Set
Out-Of-Sample Validation
Population (%)
Non-Defaulters Defaulters
Training sample
86% [5%] 14% [5%]
28% [5%] 72% [5%]
Cross-validation sample
85% [5%] 15% [5%]
33% [5%] 67% [5%]

Number of Defaults Historically for Model Development

	Year	Defaults	Non-Defaults
Fitzpatrick	(32)	19	19
Beaver	(67)	79	79
Altman	(68)	33	33
Lev	(71)	37	37
Wilcox	(71)	52	52
Deakin	(72)	32	32
Edmister	(72)	42	42
Blum	(74)	115	115
Taffler	(74)	23	45
Libby	(75)	30	30
Diamond	(76)	75	75
Altman, Haldeman and Narayanan	(77)	53	58
Marais	(79)	38	53
Dambolena and Khoury	(80)	23	23
Ohlson	(80)	105	2,000
Taffler	(82, 83)	46	46
El Hennawy and Morris	(83a)	22	22
Moyer	(84)	35	35
Taffler	(84)	22	49
Zmijewski	(84)	40	800
Zavgren	(85)	45	45
Casey and Bartczak	(85)	60	230
Peel and Peel	(88)	35	44
Barniv and Raveh	(89)	58	142
Boothe and Hutchninson	(89)	33	33
Gupta, Rao, and Bagchi	(90)	60	60
Kease and McGuiness	(90)	43	43
Keasey, McGuiness and Short	(90)	40	40
Shumway	(96)	300	1,822
Moody’s RiskCalc Public Firm	(00)	1,406	13,041
Moody’s RiskCalc Private Firm	(00)	1,621	23,089
Median		40	45

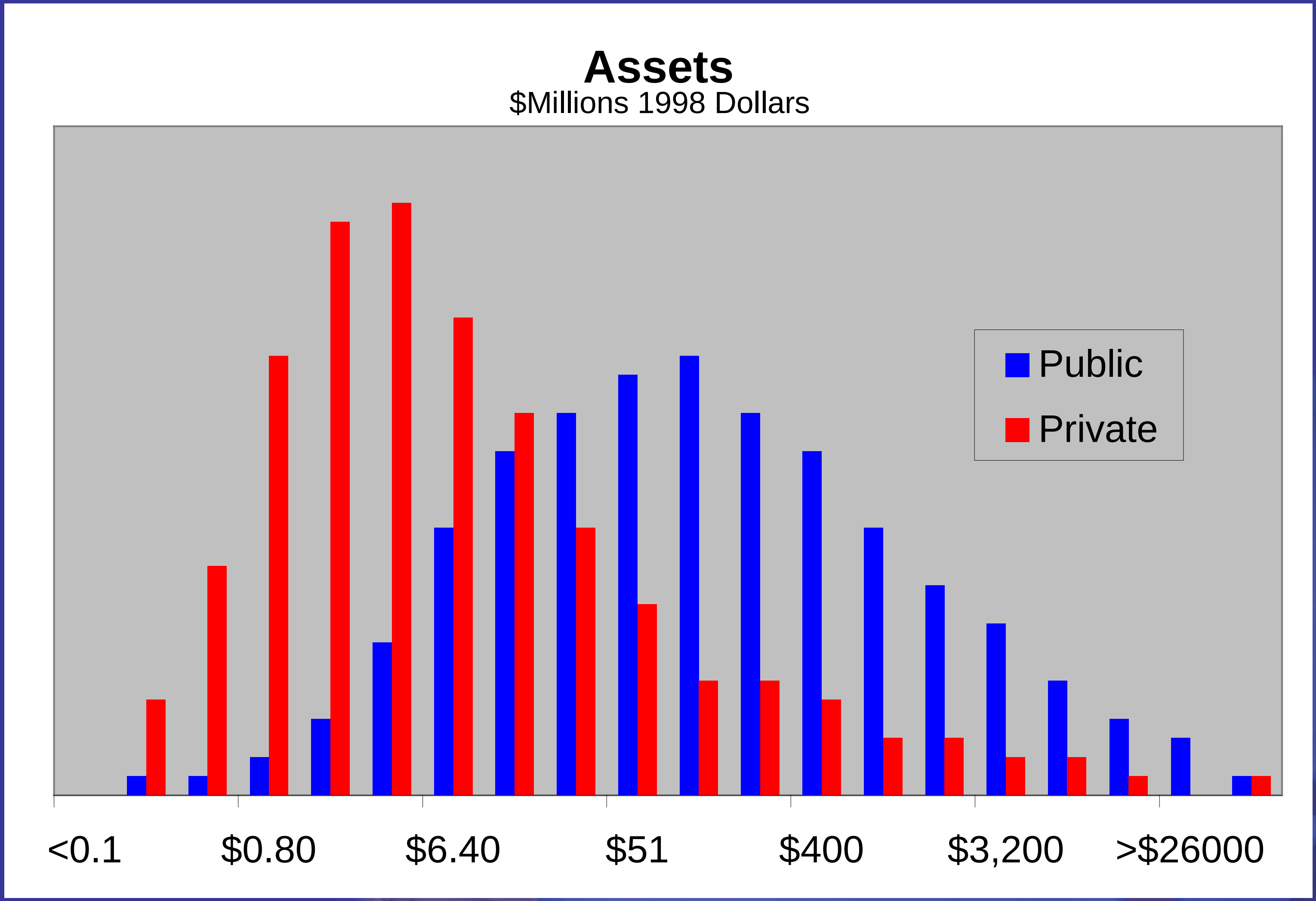
Actual accuracy rate: 65% vs. 70%



Examples of faulty inferences due to violations of universe

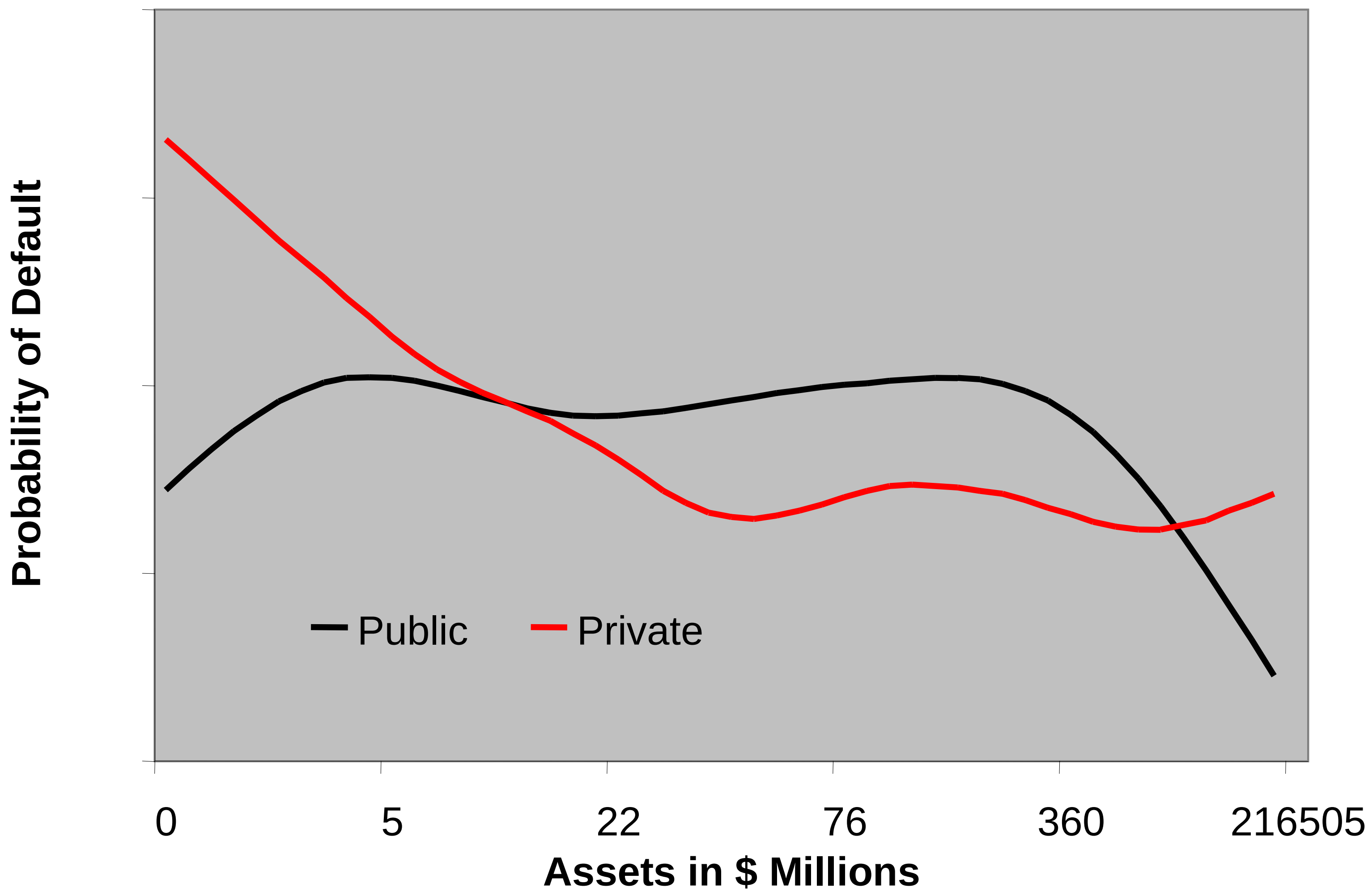
Moody's Risk Management Services

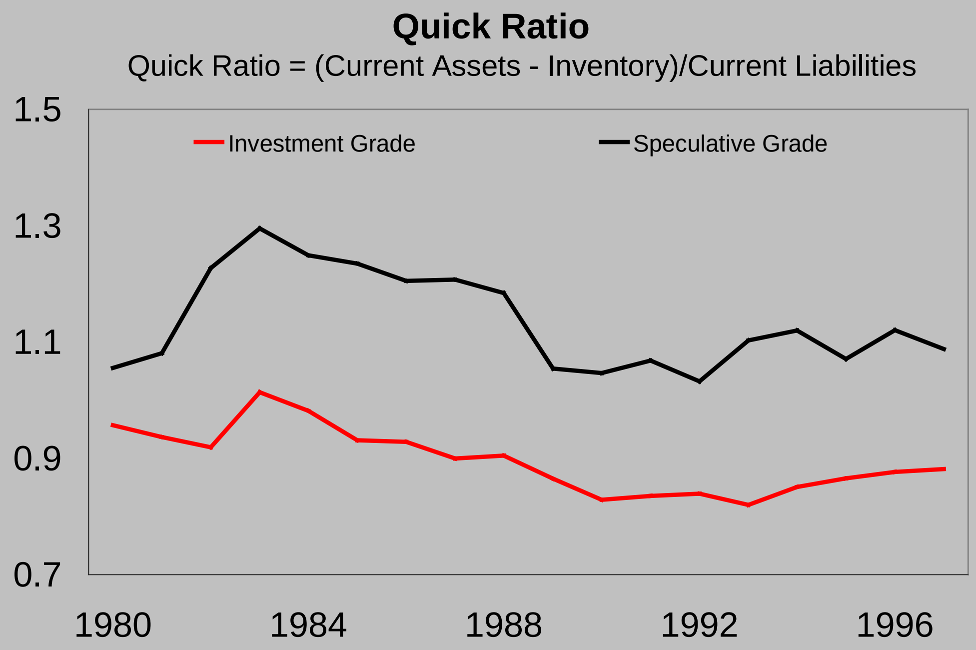
Public and Private Size Groupings



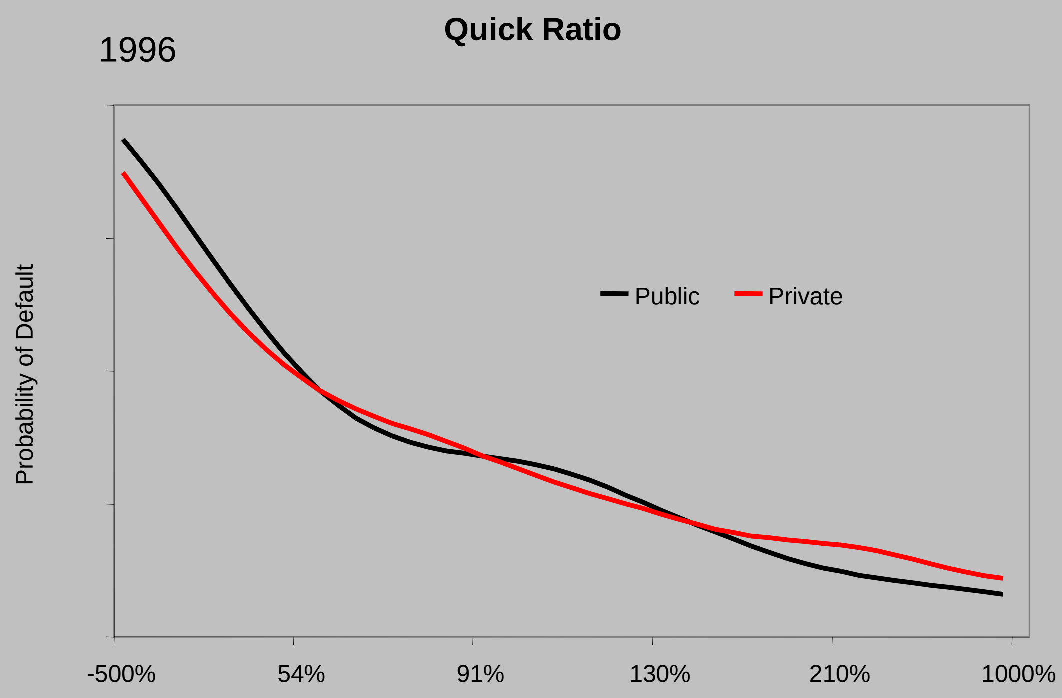
Moody's Risk Management Services

Size Bias Makes Model Estimation, Testing, Difficult



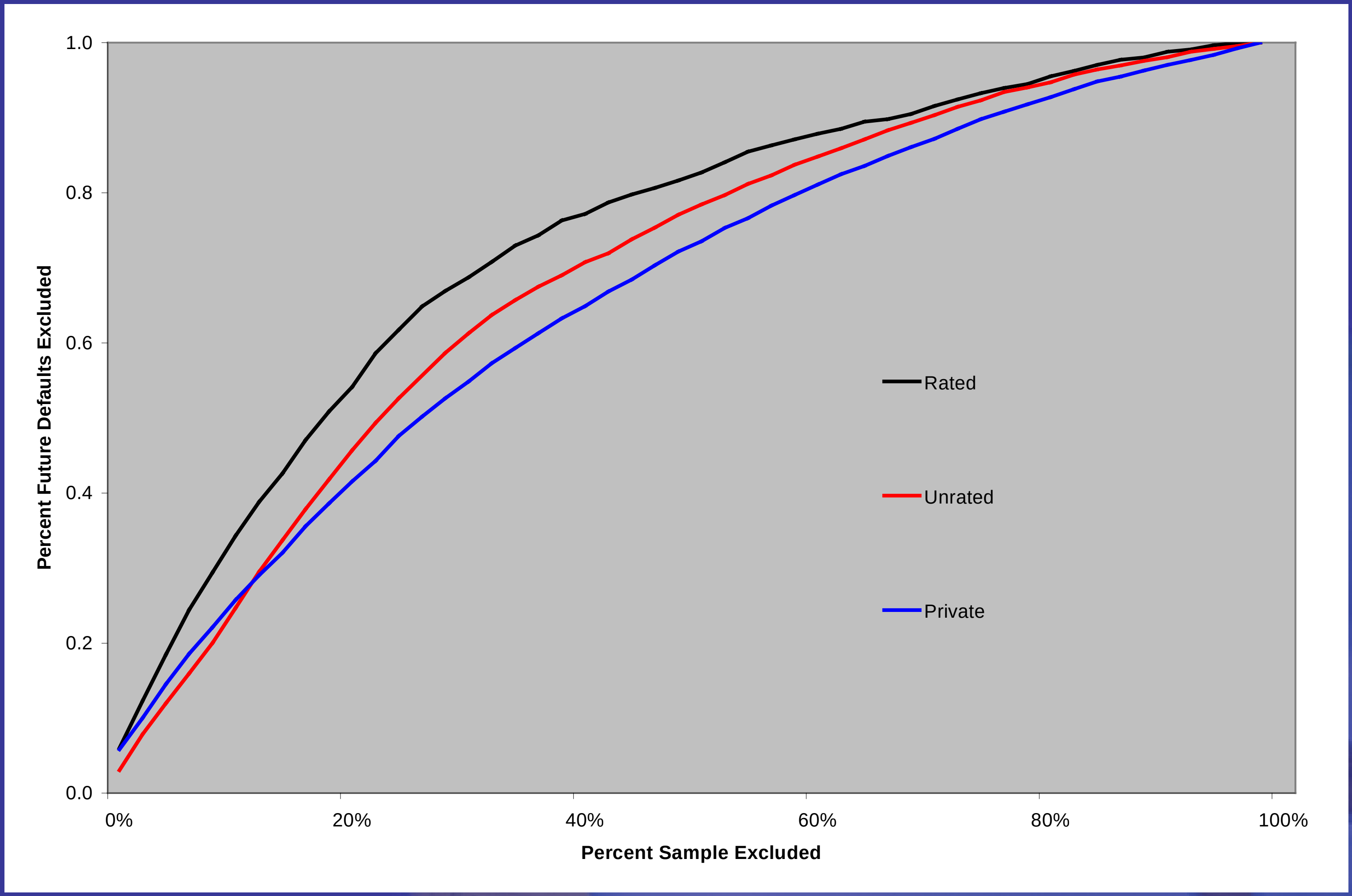


Higher rated companies have lower liquidity...



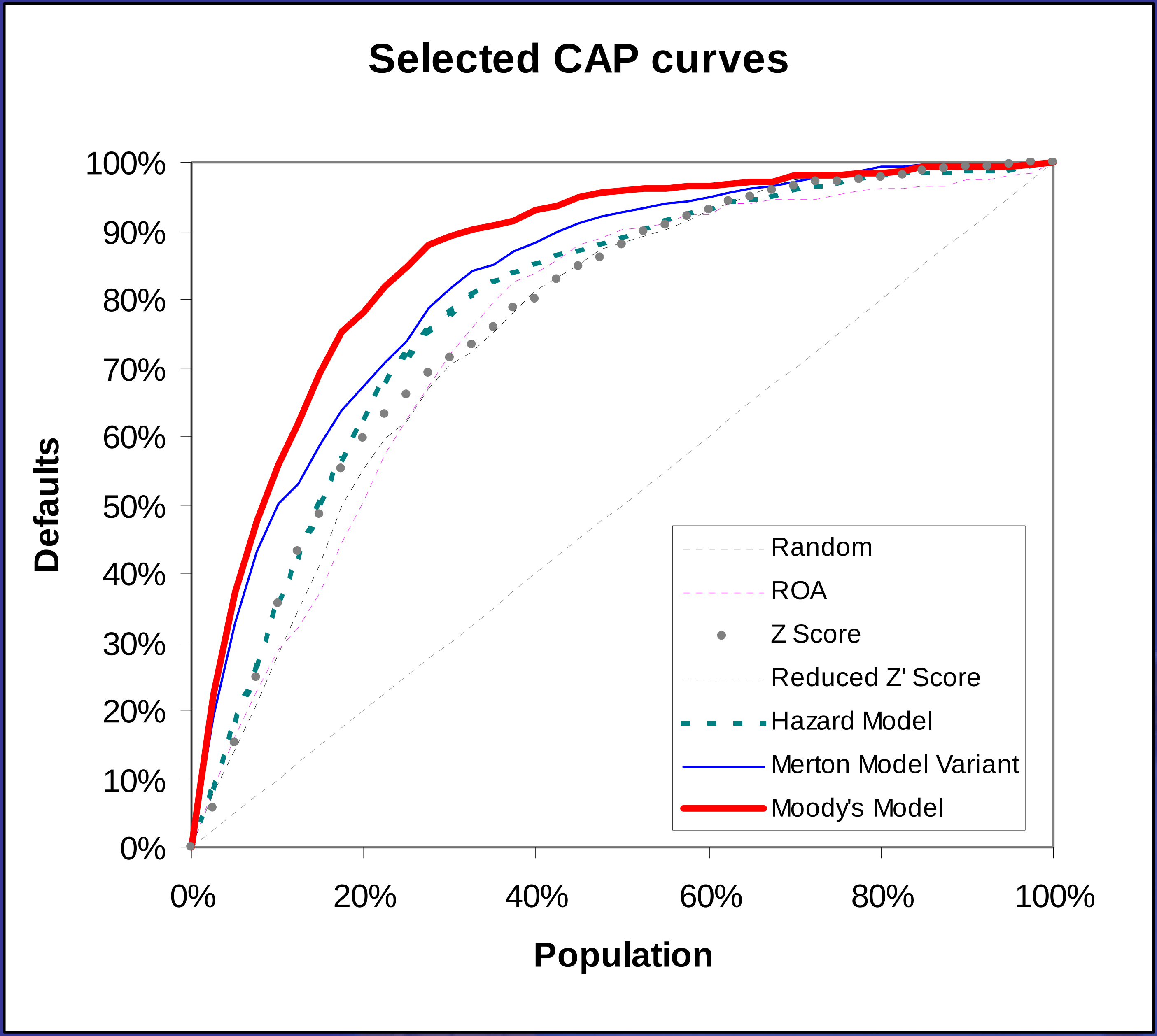
... yet lower liquidity implies higher default rates for both public and private companies

Different universes: All Models Do Better on Bigger Firms



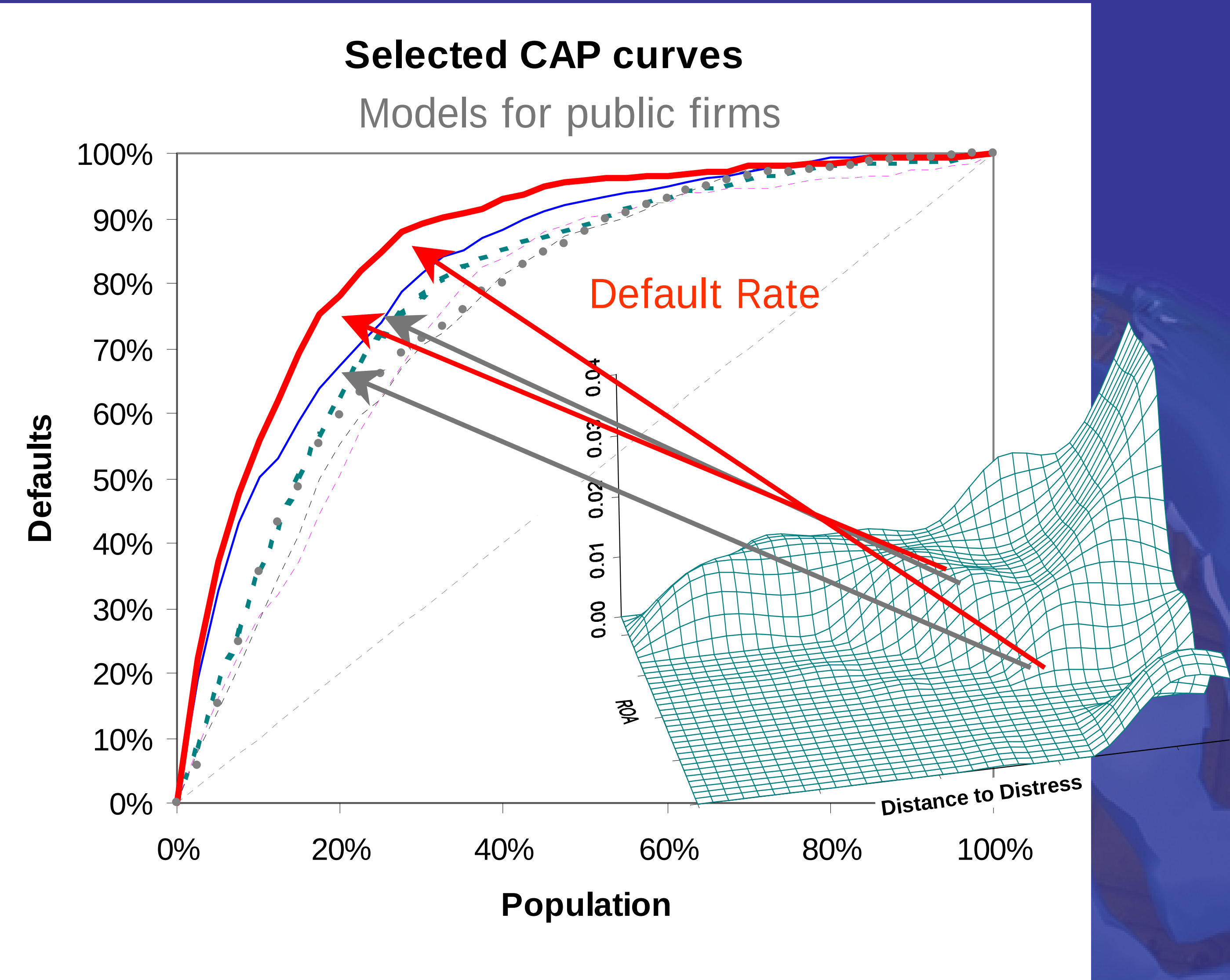
RiskCalc Validation Results

Moody's Risk Management Services

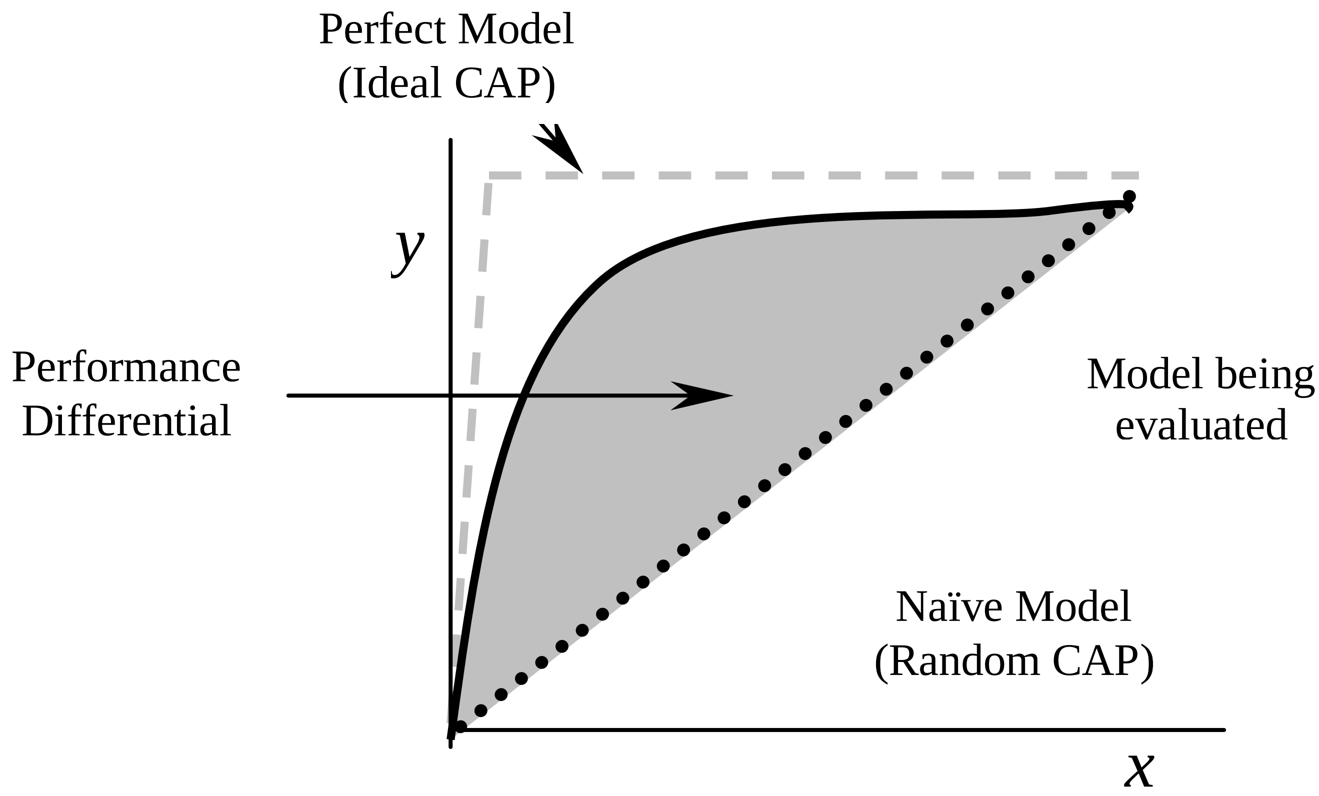


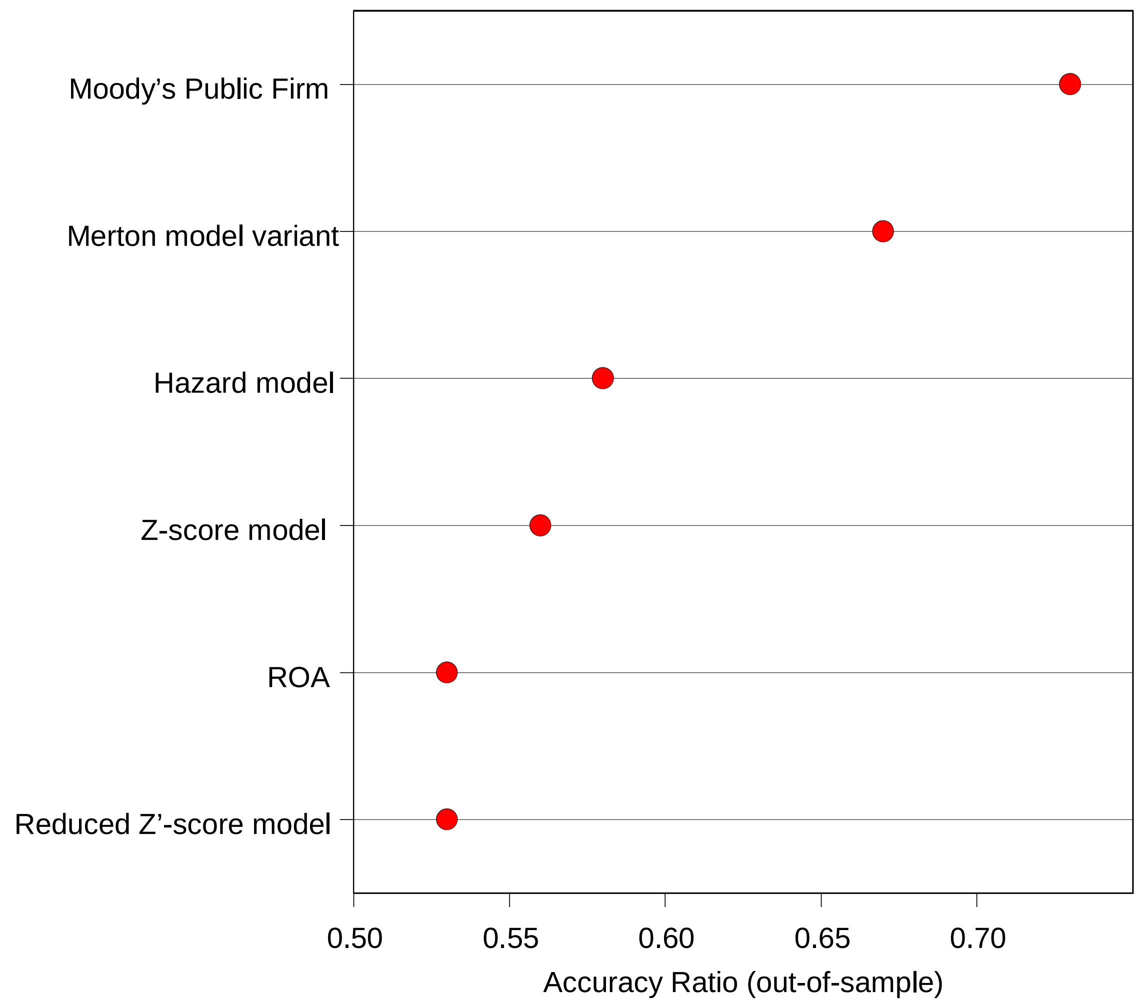
Moody's Risk Management Services

Relative Model Performance



Moody's Risk Management Services


Perfect Model
(Ideal CAP)
y
Performance
Differential
Model being
evaluated
Naïve Model
(Random CAP)
x



Moody's Risk Management Services

Some reading

- RiskCalc documents Available at www.moodysrms.com to download Adobe Acrobat files
  - Navigate to "research"
- Some validation readings

  - Burnham, K.P. and Anderson, D.R., *Model Selection and Inference*, New York, Springer, 1998.
  - Dhar, V. and Stein, R.,"Finding Robust and Usable Models with Data Mining: Examples from Finance," PCAI, Sept., 1998.
  - Hoadley, B. and Oliver, R. M., (1998), "Business measures of scorecard benefit," *IMI Journal of Mathematics Applied in Business & Industry*, 9, pp. 55-64.
  - Sobehart, J., Keenan, S., Stein, R. *Benchmarking quantitative default risk models: A Validation Methodology*, Moody's Special Comment, March 2000.
  - Provost, F. and Fawcett, T.,"Analysis and Visualization of Classifier Performance: Comparison Under Imprecise Class and Cost Distributions," *Proceedings Third International Conference on KDD*, Newport Beach, CA, August 1997.

Conclusion

- We have found that validation can be done *even* with sparse data but is difficult *particularly* with sparse data
- It is useful to carefully design validation experiments that test a model in simulated real-world environments controlling for time and universe
- Meaningful benchmarks (not straw-men) are usually necessary for reference
- Many validation tests are sensitive to the exact sample chosen: observed performance differences may be due to sampling issues particularly with rare events
- There is little that we can do to increase the power in sparse data for validation. The best we can do is to acknowledge limitations and understand bounds

Rare Event Modeling and Validation Through Time:

The case of corporate credit analysis

Roger M. Stein
Managing Director
Credit Risk Analytics
Moody's Risk Management Services
steinr@moodys.com

Moody's Risk Management Services