Redesigning Ratings: Assessing the Discriminatory Power of Credit Scores under Censoring

21. Juni 2004

Marlene Müller

Holger Kraft, Gerald Kroisandt

Fraunhofer Institut Techno- und Wirtschaftsmathematik

Credit Rating/Scoring

- **rating**
  classification of individuals (private, corporate, sovereign) into groups of equivalent default risk
- **rating score**
  quantitative indicator of individual default risk
- **individual default probability (PD)**
  - ⋆ typically a one-to-one mapping of the score
  - ⋆ basis to construct rating groups

Methods to Estimate Scores and PDs

- discriminance analysis, classification
  $\longrightarrow$ Scores

- categorical regression (logit/probit, panel, ordered categories)
  $\longrightarrow$ Scores + PDs

- Merton approach (stock price as estimate for the market value)
  $\longrightarrow$ Scores by “distance to default”

Censoring

▷ not all credit applicants obtain a loan ⇒ no representative sample

as a consequence:

- estimates are be biased
- problem of bias accumulation over time

possible remedies:

- model the process of credit acceptance/rejection
- find bounds for the estimates that reflect the worst cases (nonparametric solution)

Data Example

sample of private loans

- default indicator: $Y \in \{0, 1\}$, where $1 =$ default
- explanatory variables:
  - ⋆ personal characteristics (age, occupation, telephone, savings)
  - ⋆ credit characteristics (amount, duration)
  - ⋆ credit history (previous credits)
- sample size: 1000 (300 defaults!)

How to evaluate if a subset of the customers (e.g. those which showed "hesitant payment of previous credits") would not have granted a loan?

References: Fahrmeir/Hamerle (1984); Fahrmeir & Tutz (1995)

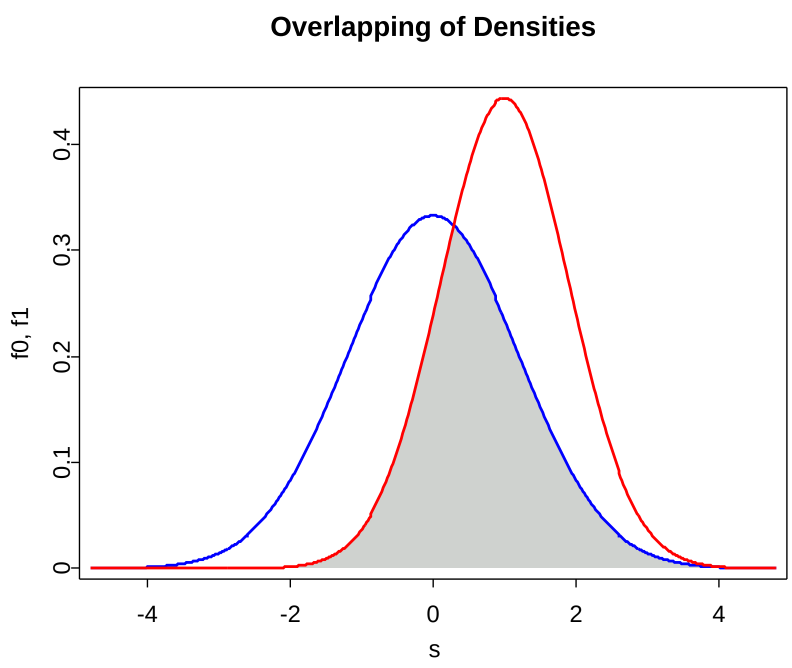
Evaluation of Credit Scores

main objectives:

- discriminatory power:
  $\longrightarrow$ relative assessment of the PDs
- calibration:
  $\longrightarrow$ absolute assessment of the PDs

References: cf. Krämer (2000), Deutsche Bundesbank Monthly Report (Sept. 2003)

Score Distributions of Defaults vs. Non-Defaults



● defaults ● non-defaults

from a statistical point of view

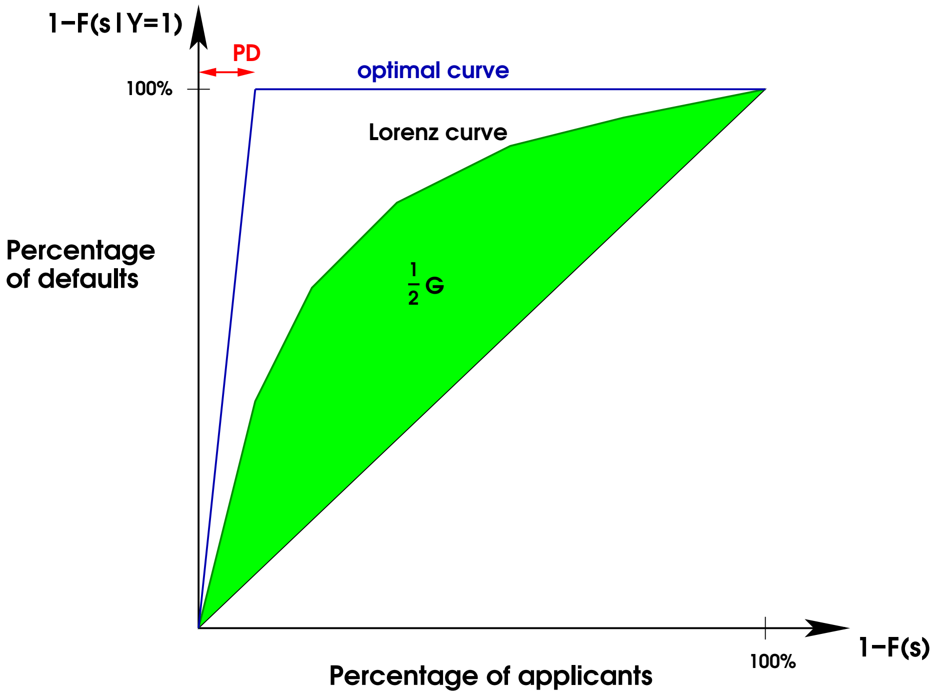
- ⋆ overlapping region $\approx U = \min_s \{F_1(s) + 1 - F_0(s)\}$
  (F_j denote the CDFs)
- ⋆ $T = 1 - U$ corresponds to the Kolmogorov-Smirnov test statistic
  $\Rightarrow$ *maximal deviation between* F_0 *and* F_1

Score Distributions of Defaults vs. Full Sample

- ⋆ Lorenz curve (power curve, cumulated accuracy profile, CAP)

$$1 - F(s) = P(S > s) \quad \text{vs.}$$

$$1 - F_1(s) = P(S > s | Y = 1)$$



- ⋆ Gini coefficient G, accuracy ratio AR

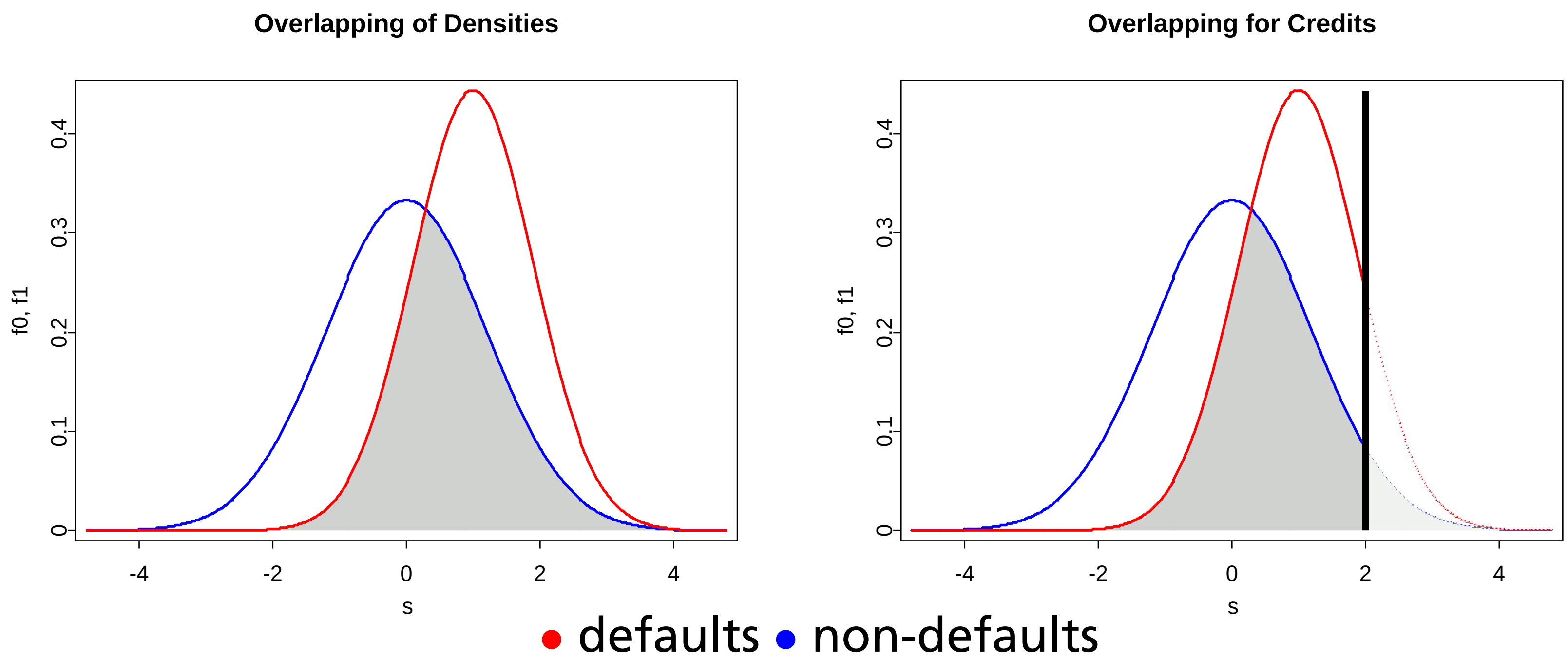
$$AR = \frac{G}{G_{opt}} = \frac{G}{P(Y=0)}$$

from a statistical point of view

- ⋆ AR is a linear function of the Mann-Whitney U test statistic
- ⋆ alternative ROC curve: $AR = 2 \cdot AUC - 1$
  $\Rightarrow$ *average deviation between* F_0 *and* F_1

Censored Sample

- not all credit applicants obtain a loan ⇒ no representative sample



- evaluation criteria are biased
  (bias in both directions, due to conditional probabilities)

- problem of bias accumulation over time

How to Correct for Censoring?

- **econometric models**
  - ⋆ censored (bivariate) probit (Greene, 1998; Boyes/Hoffman/Low, 1989)
- **"reject inference"**
  - ⋆ give loans to all applicants for a certain period (e.g. Hand, 2002)
  - ⋆ re-classification (Ash/Meester, 2002)
  - ⋆ re-weighting (Ash/Meester, 2002; Crook/Banasik, 2002)
  - ⋆ extrapolation (Ash/Meester, 2002; Crook/Banasik, 2002)
- **bounds**
  - ⋆ identification and PDs (Horowitz/Manski, 1998)
  - ⋆ discriminatory power (Kraft/Kroisandt/Müller, 2002+2003)

Discriminatory Power under Censoring

notation:

- default $Y \in \{0, 1\}$
- score S (function of the explanatory variables $X_1, \ldots, X_p$)
- condition for acceptance: $\mathcal{A}$

we can estimate:

- all values given $\mathcal{A}$ (such as $P(Y = j|\mathcal{A})$, $F(s|Y = j, \mathcal{A})$)

unknown are:

- all unconditional terms (such as $P(Y = j)$, $F(s|Y = j)$)

Idea

- numbers of accepted and rejected loans are known:

n	accepted loans
N	accepted+rejected loans

  (realistic requirement, since often details on rejected loans are not available)

- advantages:
  - ⋆ universal approach since no parametric assumptions on the selection mechanism are needed
  - ⋆ verification of parametric assumptions is possible

Technical Details

we search for the relation between

- the unobservable CDF $F_j(s) = P(S \leq s|Y = j)$ and
- the observable CDF $\widetilde{F}_j(s) = P(S \leq s|Y = j, \mathcal{A})$

due to the theorem on the total probability it is simple to derive lower and upper bounds:

$$
\begin{aligned}
F_j(s) &\leq \widetilde{F}_j(s)P(\mathcal{A}|Y = j) + P(\overline{\mathcal{A}}|Y = j) \\
F_j(s) &\geq \widetilde{F}_j(s)P(\mathcal{A}|Y = j)
\end{aligned}
$$

Reference: similar to Horowitz & Manski (1998)

Lemma 1

Using the notation $\alpha_j = P(\mathcal{A}|Y = j)$ *it holds*

$$\alpha_j \widetilde{F}_j(s) \leq F_j(s) \leq 1 - \alpha_j \{1 - \alpha_j \widetilde{F}_j(s)\}.$$

Lemma 2

For α_j *we have*

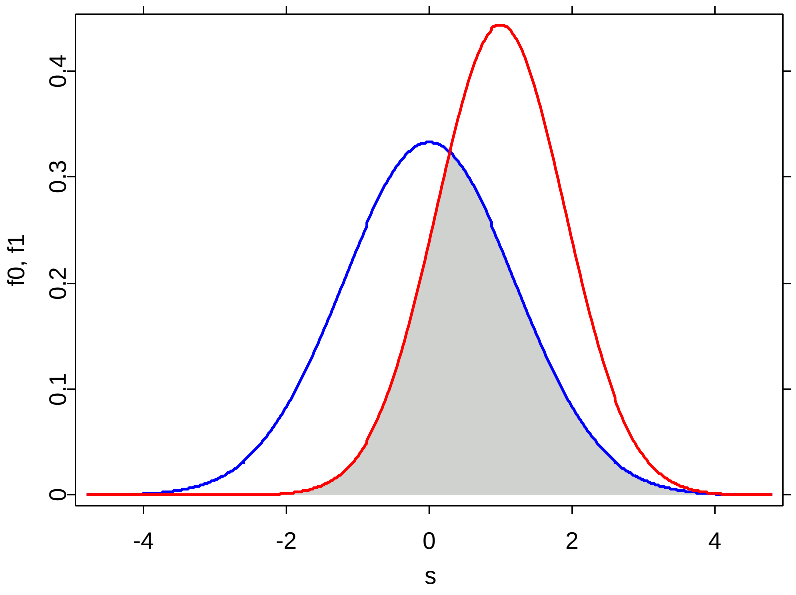
$$\alpha_j^{low} \leq \alpha_j \leq 1 \quad \textit{with} \quad \alpha_j^{low} = \frac{P(Y = j|\mathcal{A})P(\mathcal{A})}{P(Y = j|\mathcal{A})P(\mathcal{A}) + P(\overline{\mathcal{A}})} \,.$$

Comparison of Score Distributions

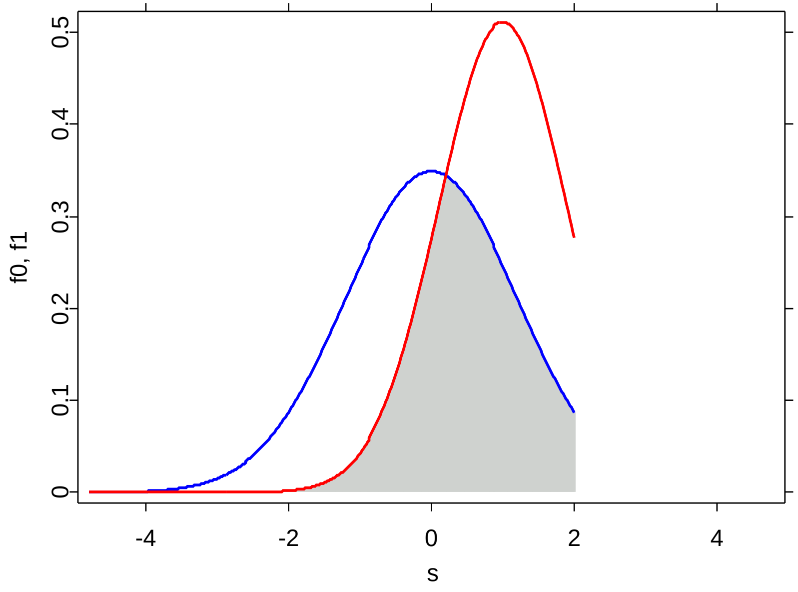
discriminatory power criterion

$$T = 1 - U = \max_s \{F_0(s) - F_1(s)\}$$

U



$\widetilde{U} = (U|\mathcal{A})$ (here $\mathcal{A} = \{S \leq c\}$)



Proposition 1

Bounds for T are given by

$$\max_s \left[\alpha_0^{low}\,\widetilde{F}_0(s) + \alpha_1^{low}\left\{1-\widetilde{F}_1(s)\right\}\right] - 1 \\ \leq T \leq 1 - \min_s \left[\alpha_0^{low}\{1-\widetilde{F}_0(s)\} + \alpha_1^{low}\,\widetilde{F}_1(s)\right]\ ,$$

Proof: Lemmas 1+2

BUT:

$P(Y=1|\mathcal{A})$ and $P(Y=0|\mathcal{A})$ can (of course) not vary freely.

$\Rightarrow$ improved bounds can be found

Proposition 2

Improved bounds for T are given by

$$\max_s \left[\frac{\beta_0}{p_s^{up}} \widetilde{F}_0(s) + \frac{\beta_1}{1 - p_s^{up}} \{1 - \widetilde{F}_1(s)\} \right] - 1$$
$$\leq T \leq 1 - \min_s \left[\frac{\beta_0}{p_s^{low}} \{1 - \widetilde{F}_0(s)\} + \frac{\beta_1}{1 - p_s^{low}} \widetilde{F}_1(s) \right] ,$$

where $\beta_j = P(Y = j, \mathcal{A})$ *and*

$$p_s^{low} \textit{ resp. } p^{up} = \begin{cases} \beta_0 & \textit{if } \gamma_s < \beta_0, \\ \beta_0 + P(\overline{\mathcal{A}}) & \textit{if } \gamma_s > \beta_0 + P(\overline{\mathcal{A}}), \\ \gamma_s^{low} \textit{ resp. } \gamma^{up}, & \textit{otherwise}, \end{cases}$$

$$\gamma_s^{low} = \frac{\sqrt{\beta_0 \{1 - \widetilde{F}_0(s)\}}}{\sqrt{\beta_0 \{1 - \widetilde{F}_0(s)\}} + \sqrt{\beta_1 \widetilde{F}_1(s)}}, \quad \gamma_s^{up} = \frac{\sqrt{\beta_0 \widetilde{F}_0(s)}}{\sqrt{\beta_0 \widetilde{F}_0(s)} + \sqrt{\beta_1 \{1 - \widetilde{F}_1(s)\}}} .$$

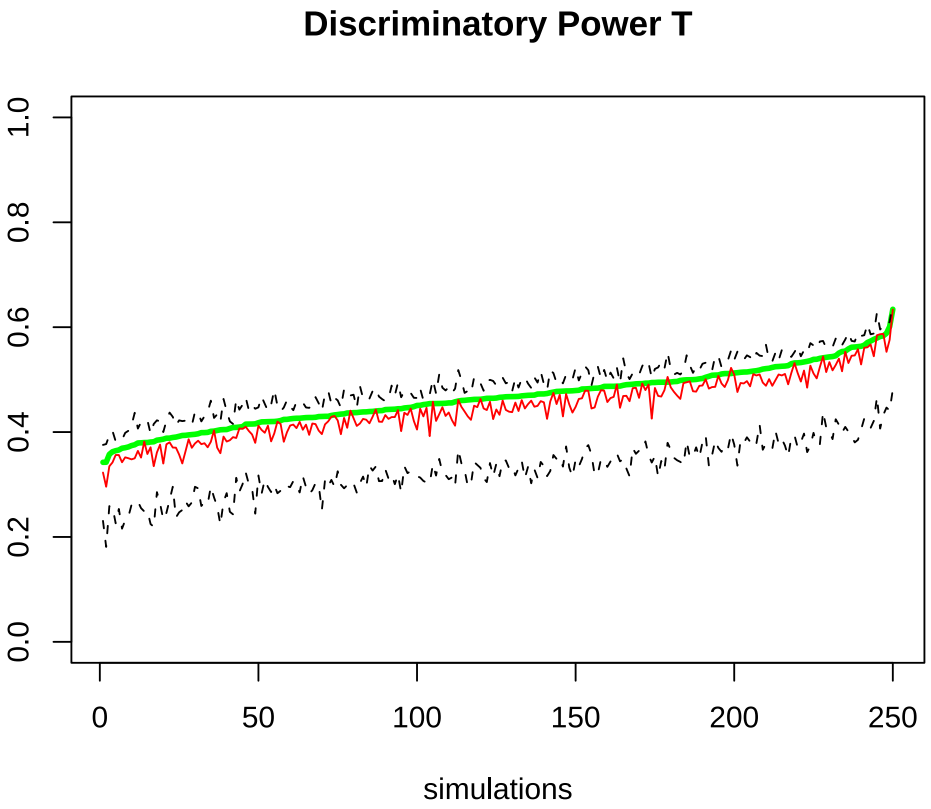
Monte Carlo Simulation

all terms in the propositions are estimable if N is known, in particular:

$$\widehat{P}(\overline{\mathcal{A}}) = 1 - \frac{n}{N}$$

we estimate all unknown terms by relative frequencies or empirical CDFs

- simulation sample size: 250
- credit applicants: $N = 500$, rejected loans: 2%
- simulated model: S normal, $P(Y = 1|S)$ logit



- estimated T
- estimated $\widetilde{T} = (T|\mathcal{A})$
- estimated bounds

- $\widetilde{T}$ over- and underestimates T
- lower bound is "far away" as consequence of the very general selection condition $\mathcal{A}$
- narrower bounds can be found if the selection condition is more precisely specified, such as $\mathcal{A} = \{S \leq c\}$
- if $N = n$ (no censoring), all curves are identical

Accuracy Ratio

Lorenz curve

$$\{1 - F(s), 1 - F_1(s)\}$$

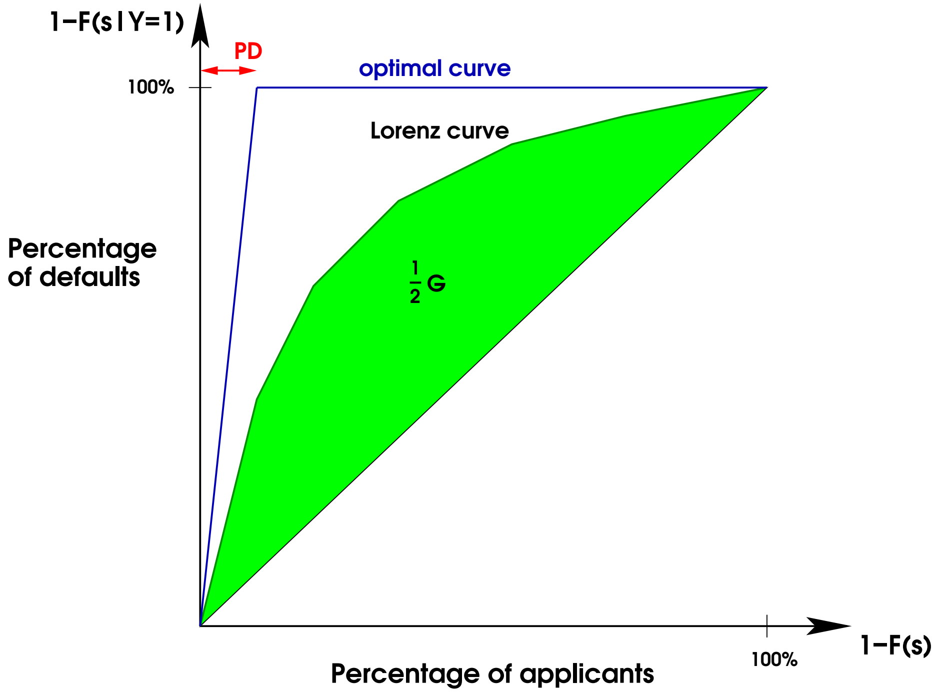
Gini coefficient

$$G = 2\int_{+\infty}^{-\infty} (1 - F_1)\, d(1 - F) - 1$$

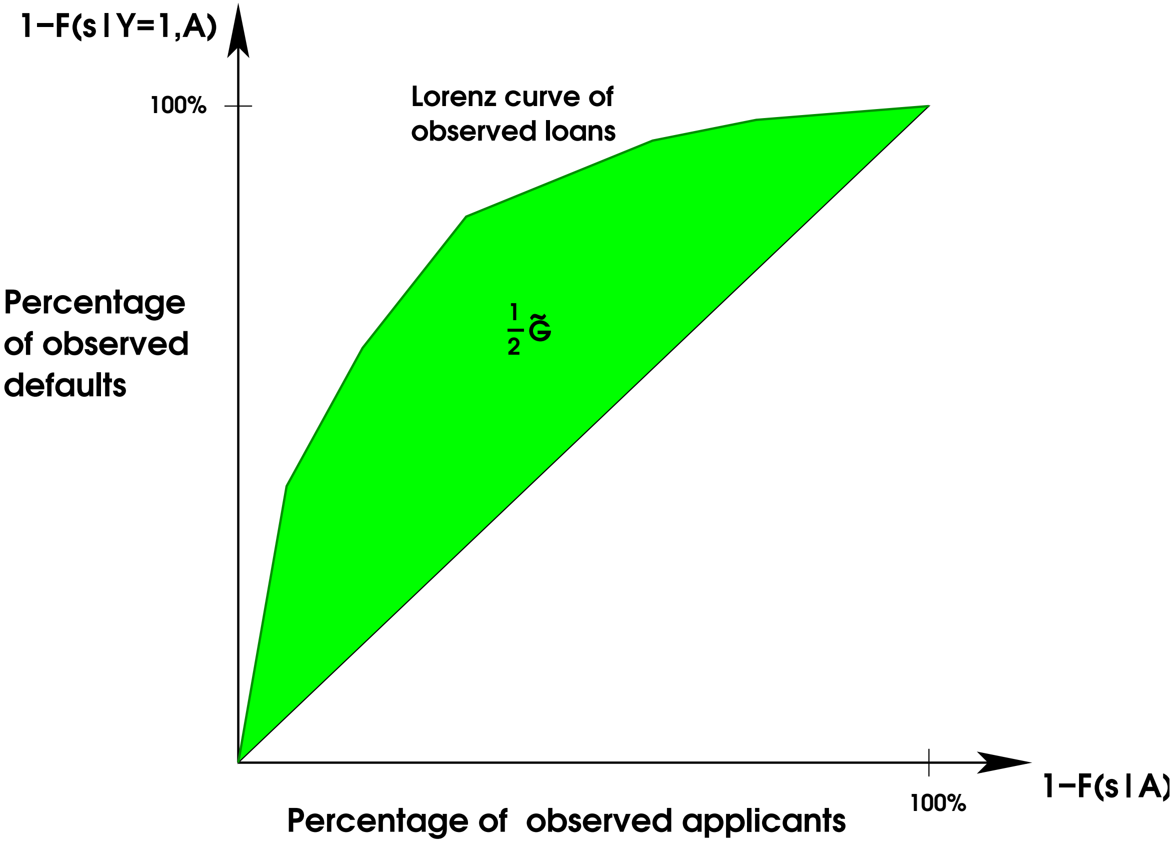
$$= 1 - 2\int_{-\infty}^{+\infty} F_1\, dF$$

Accuracy Ratio

$$AR = \frac{G}{G_{opt}} = \frac{G}{P(Y = 0)}$$

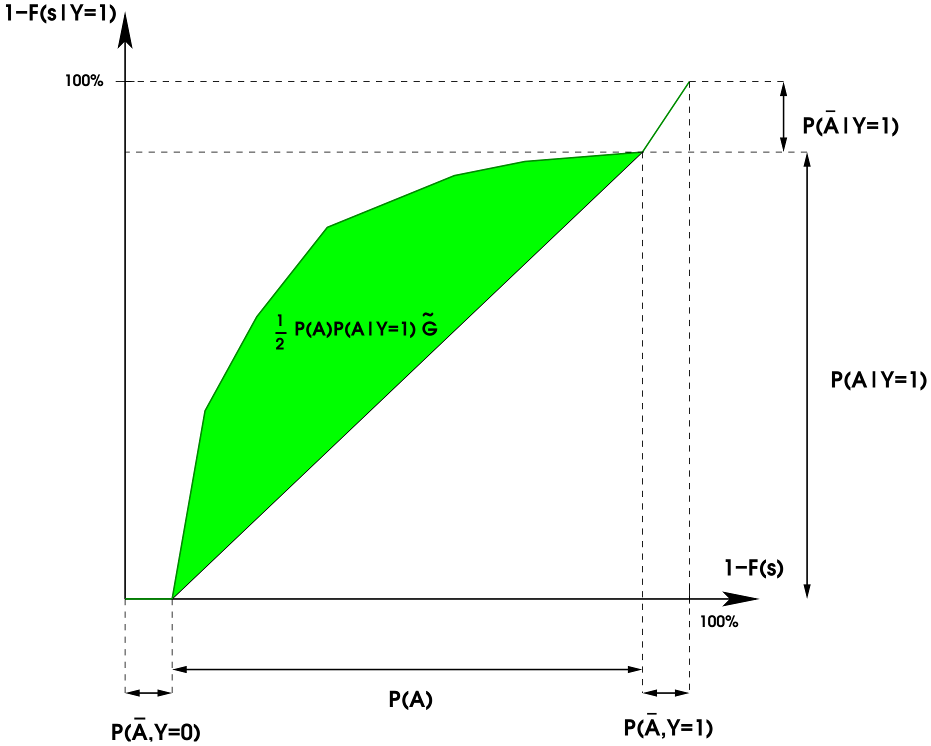


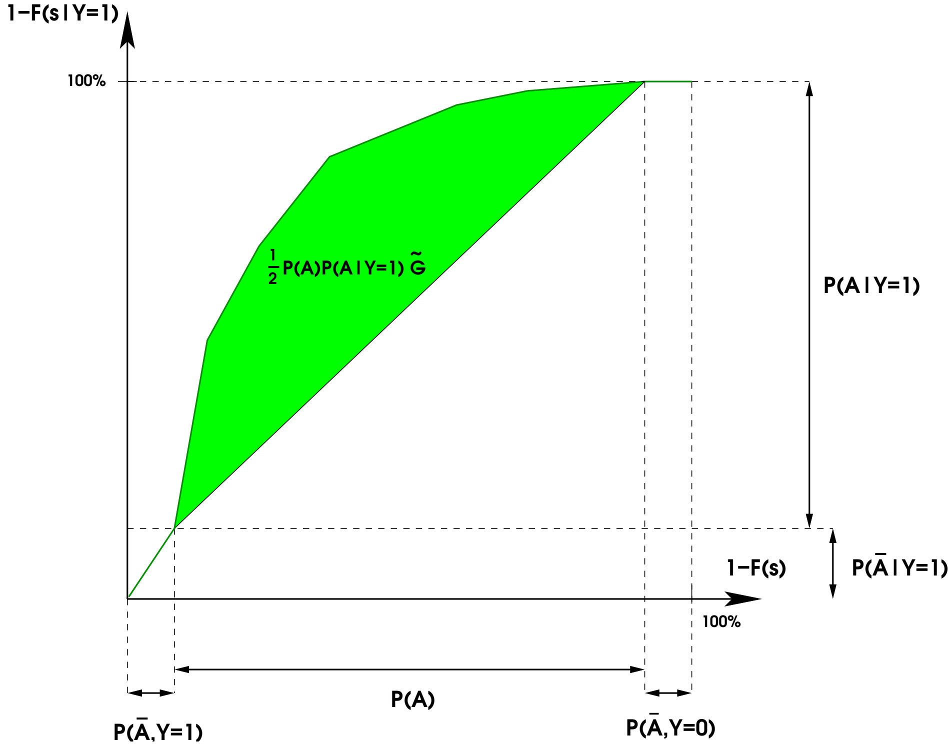
"Censored Lorenz Curve"



Lower and Upper Bounds

extreme cases are $Y = 1$ resp. $Y = 0$ for all rejected applicants (in $\overline{A}$)





Proposition 3

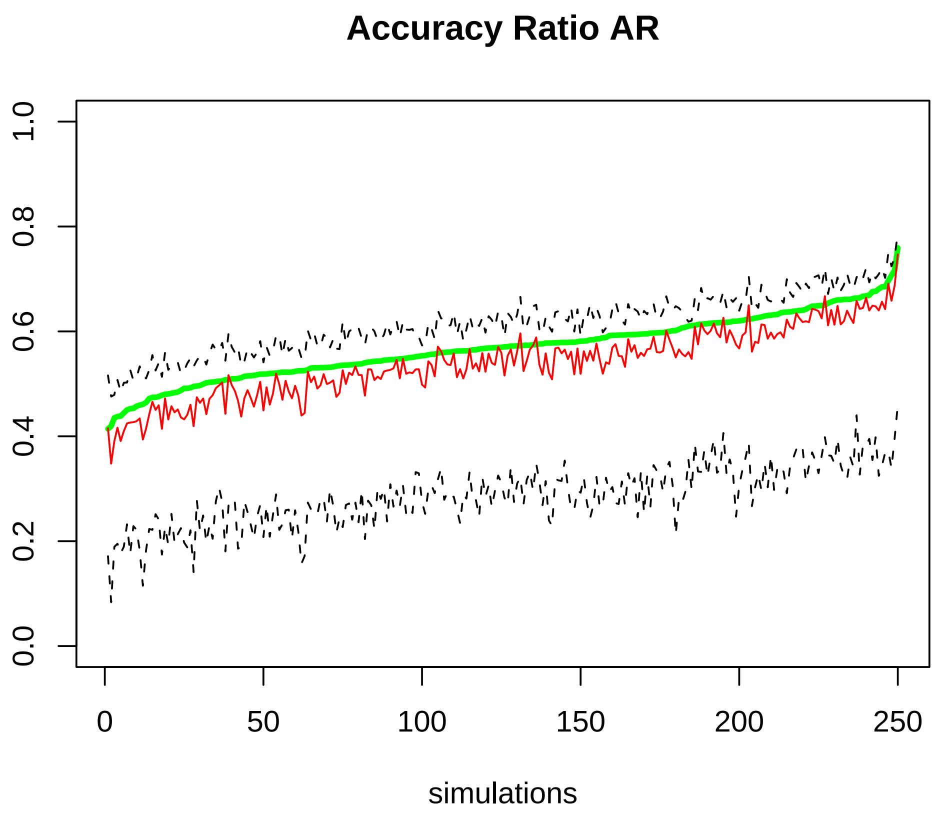
We denote again $\beta_j = P(\mathcal{A}, Y = j)$*. Bounds for* AR *are given by:*

$$\left(\widetilde{AR} + 1\right) \frac{\beta_0 \beta_1}{p_0^\star (1 - p_0^\star)} - 1 \leq AR \leq \left(\widetilde{AR} - 1\right) \frac{\beta_0 \beta_1}{p_0^\star (1 - p_0^\star)} + 1$$

where

$$p_0^\star = \begin{cases} \beta_0 & \textit{if } \beta_0 > \frac{1}{2}\,, \\ \frac{1}{2} & \textit{if } \beta_0 \leq \frac{1}{2} \leq \beta_0 + P(\overline{\mathcal{A}})\,, \\ \beta_0 + P(\overline{\mathcal{A}}) & \textit{if } \beta_0 + P(\overline{\mathcal{A}}) < \frac{1}{2}\,. \end{cases}$$

Monte Carlo Simulation



- estimated T
- estimated $\widetilde{T} = (T|\mathcal{A})$
- estimated bounds

- $\widetilde{AR}$ over- and underestimates AR
- lower bound is "far away" as consequence of the very general selection condition $\mathcal{A}$
- narrower bounds can be found if the selection condition is more precisely specified, such as $\mathcal{A} = \{S \leq c\}$
- if $N = n$ (no censoring), all curves are identical
- bounds are wider (relative to size of AR)

Application to Data

recall

- sample of $N = 1000$ private loans (300 defaults)

“artificial” censoring

- assume that customers with a negative credit history (those which showed a “hesitant payment of previous credits”) would not have granted a loan
- sample size of observed $n = 960$ (275 defaults)
- estimate 2 model specifications (logit)

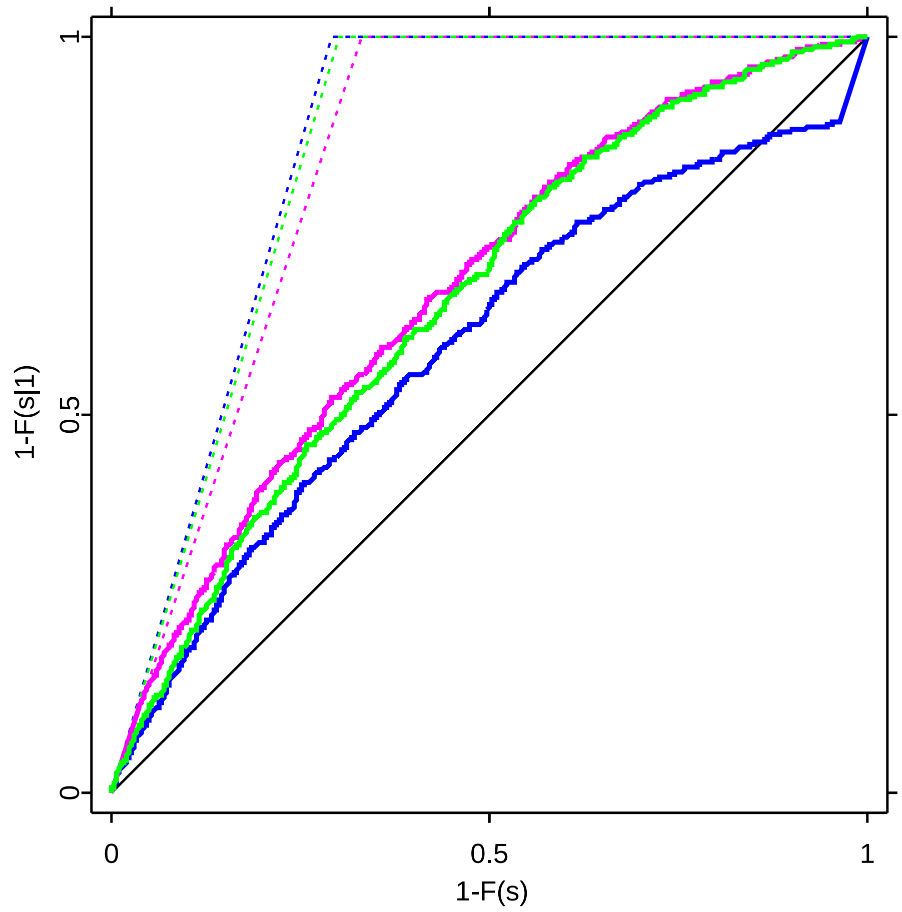
Variable	Specification 1	Specification 2
previous loans (1 for OK, 0 for unknown)	×	
employed (1 for more than one year, 0 otherwise)		×
duration of the loan (discretized with dummies for 10–12, 13–18, 19–24 and more than 24 months)	×	
amount of the loan (+ amount squared)	×	×
age of the borrower (+ age squared)	×	
interaction term for amount and age	×	
savings (1 for more than 1000 DM, 0 otherwise)	×	
foreigner (1 if yes, 0 otherwise)	×	
purpose (1 if loan is used to buy a car, 0 otherwise)	×	
house owner (1 if yes, 0 otherwise)	×	

Estimated Scores

$$\text{Score 1} = 0.162 - 0.696^{\star\star\star} \cdot \text{previous} + 0.496^{\star} \cdot \text{(d9-12)} + 0.818^{\star\star\star} \cdot \text{(d12-18)}$$
$$+0.919^{\star\star\star} \cdot \text{(d18-24)} + 1.502^{\star\star\star} \cdot \text{(d>24)} - 0.91^{\star\star\star} \cdot \text{savings}$$
$$+0.976 \cdot \text{foreign} - 0.339^{\star} \cdot \text{purpose} + 0.614^{\star\star\star} \cdot \text{house}$$
$$-0.000277^{\star\star} \cdot \text{amount} - 0.0971^{\star\star} \cdot \text{age}$$
$$+0.0000000185^{\star\star} \cdot \text{amount}^2 + 0.00086^{\star} \cdot \text{age}^2$$
$$+0.00000272 \cdot (\text{amount} \cdot \text{age})$$

$$\text{Score 2} = -0.807^{\star\star} - 0.244 \cdot \text{employed}$$
$$-0.0000279 \cdot \text{amount} + 0.0000000114^{\star} \cdot \text{amount}^2$$

Bounds of the Lorenz curve (Specification 1)



- upper bound
- lower bound
- Lorenz curve for all applicants

Estimated criterion	Specification 1	Specification 2
$\widetilde{T}$	0.292	0.159
maximal range of T	[0.222,0.349]	[0.108,0.235]
$\widetilde{AR}$	0.419	0.125
maximal range of AR	[0.238,0.492]	[-0.018,0.236]

The quality of the bounds is determined by ...

- sample size
  - ⋆ precision of estimates
  - ⋆ sensitivity to outliers
- number of defaults
- ratio of rejected to all applicants
- macroeconomic changes

Reference: Parnitzke (diploma thesis, 2003)

Summary

- censoring leads to bias in (any) evaluation criterion
- no systematic bias (under- and overestimation may occur)
- lower and upper bounds can be estimated even in the case that only the number of all credit applicants is known