Fractal ROC curves - a simple model for impact of Gini coefficient's improvement on credit losses

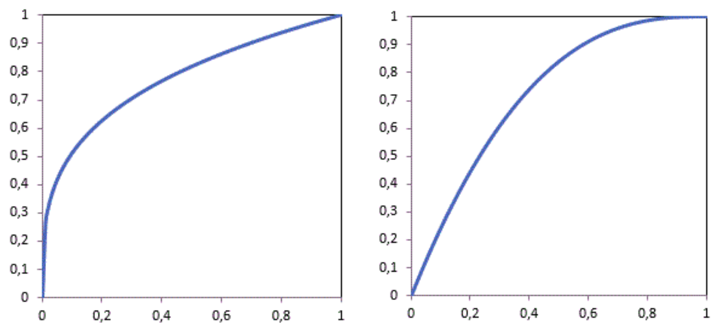
Błażej Kochański,

August 2017

Presentation, updated conference paper, data, R codes, C++ codes, Excel file etc. will be available at scoringology.com

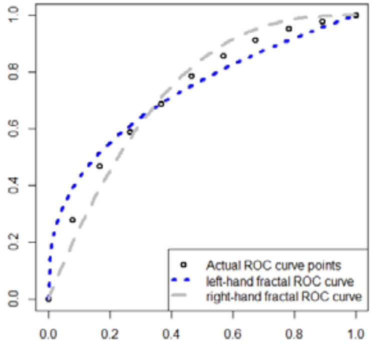
Summary

1



Fractal ROC curves do exist...

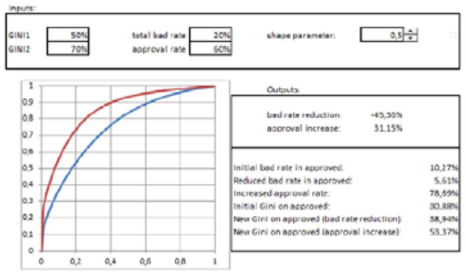
2



and are useful for modelling real-life ROC curves,

...which helps us assess impact of Gini improvement on credit losses

3

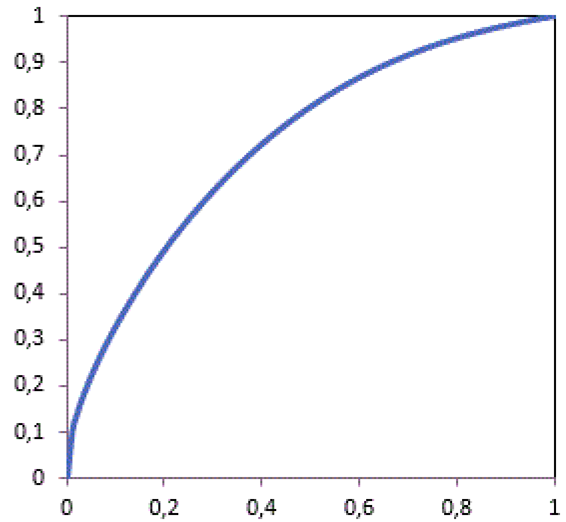


Example

GINI 0.50
↗ 0.52

Scored population bad rate = 25%
Approval rate = 40%
Beta = 0.5

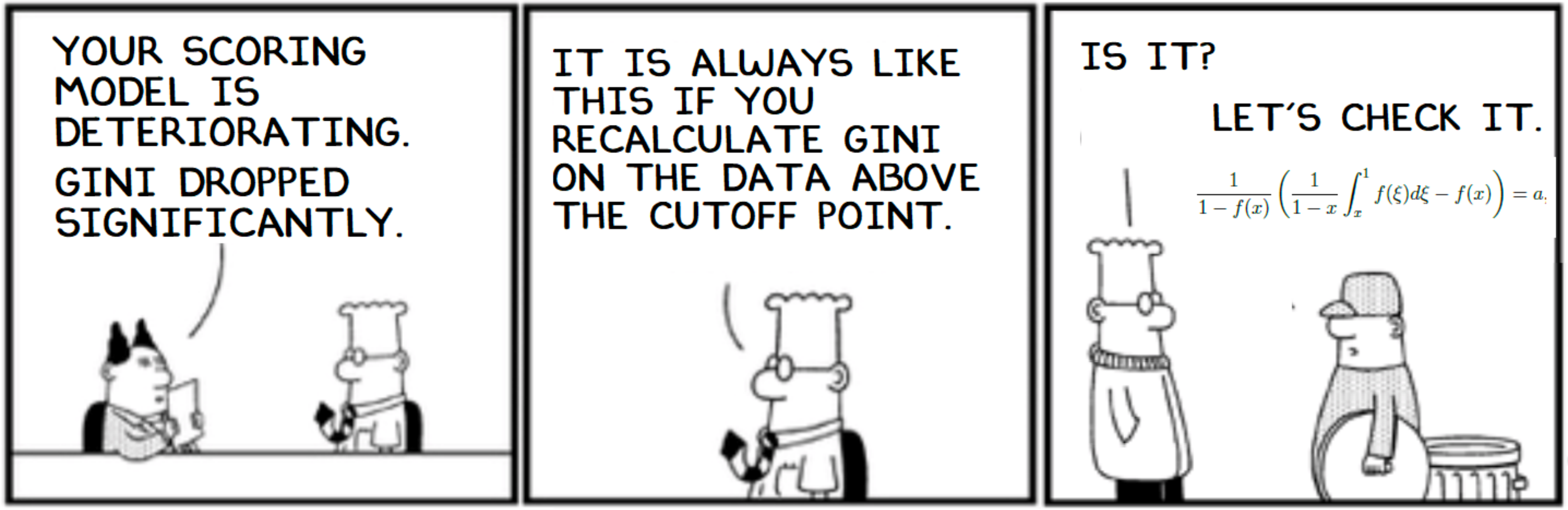
Portfolio bad rate
↘ by 6%



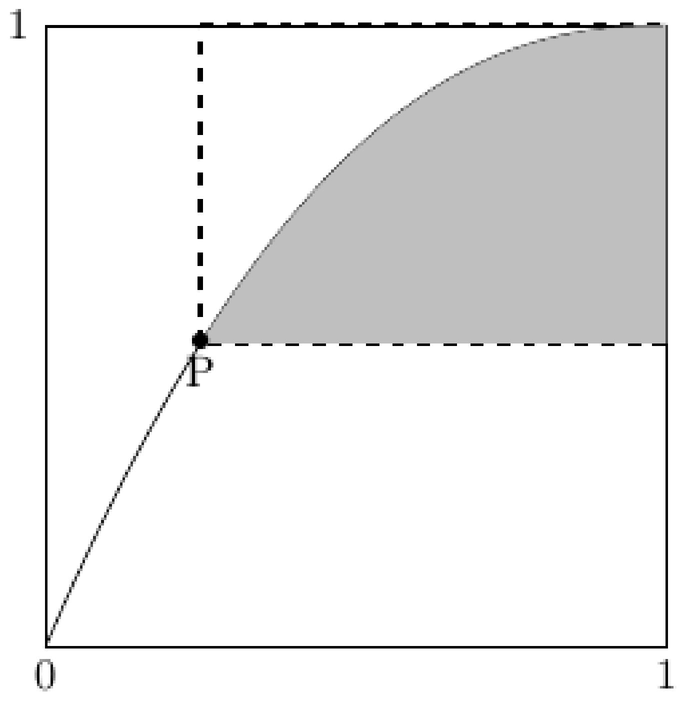
Is it always that Gini drops after cut-off?

The idea of modelling Gini coefficient for scoring models with „fractal ROC curves" came with a question asked during one of the credit scoring workshops held in Poland.

An attendee of the workshop - risk modelling manager complained about recommendation he received from audit department in his bank...



*unfortunately, this Dilbert comic strip is fake



(Right-hand) fractal ROC curve:

Shaded area as a fraction of the dashed line rectangle should be equal to AUC (area under the curve) whatever cut-off point P we take.

In other words, we are looking for a (continuous) function $f : [0,1] \to [0,1]$ such that $f(0) = 0$, $f(1) = 1$, and for all $x \in [0,1)$

$$\frac{1}{1-f(x)}\left(\frac{1}{1-x}\int_x^1 f(\xi)d\xi - f(x)\right) = a, \qquad (1)$$

where a is area under the curve (AUC).

where a is area under the curve (AUC). We can bring this integral equation to a differential equation introducing function F such that $F' = f$. Then equation (1) can be rewritten as follows:

$$F'(x)(1-a) + \frac{F(x)}{1-x} = \frac{F(1)}{1-x} - a. \tag{2}$$

Let $F(1) = c$. Consequently, we have an ordinary differential equation with two parameters a, c

$$F'(x)(1-a) + \frac{F(x)}{1-x} = \frac{c}{1-x} - a. \tag{3}$$

A solution to equation (3) is given by

$$F(x) = B(1-x)^{\frac{1}{1-a}} + c - (1-x), \tag{4}$$

where B is a parameter. Since $f = F'$ we obtain

$$f(x) = \frac{B}{1-a}(1-x)^{\frac{a}{1-a}} + 1. \tag{5}$$

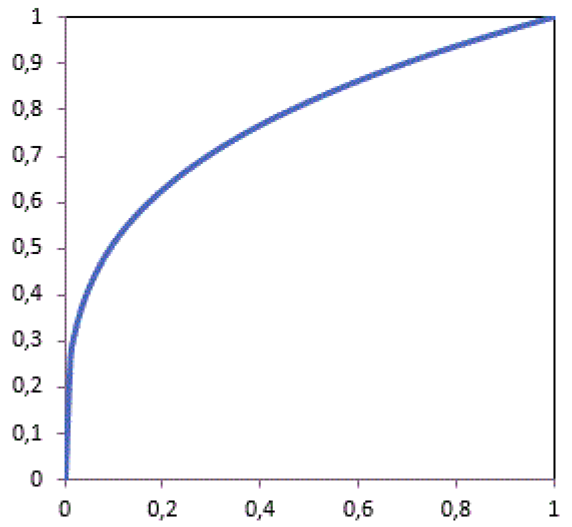
From an assumption $f(0) = 0$ then $= \frac{B}{1-a} = -1$.

$$f(x) = 1 - (1-x)^{\frac{a}{1-a}}, x \in [0,1]. \tag{6}$$

The function meets the initial condition $f(1) = 1$ and has just one parameter a which is AUC[1]. As Gini coefficient is $2 \cdot AUC - 1$, we can rewrite function f so that is a function of one parameter γ being equivalent to the Gini coefficient:

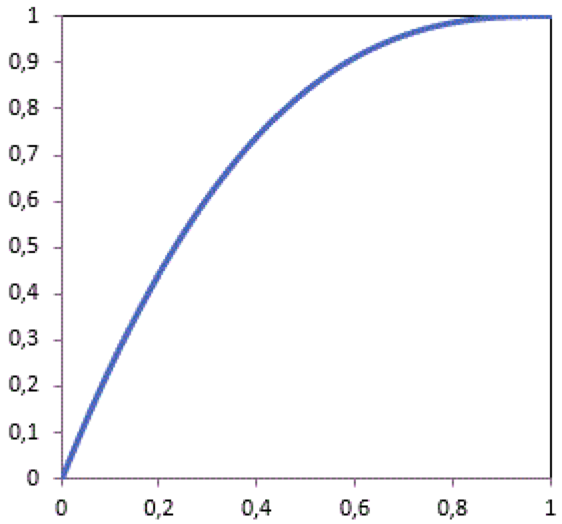
$$f(x) = 1 - (1-x)^{\frac{1+\gamma}{1-\gamma}}, x \in [0,1]. \tag{7}$$

Left-hand fractal ROC curve:



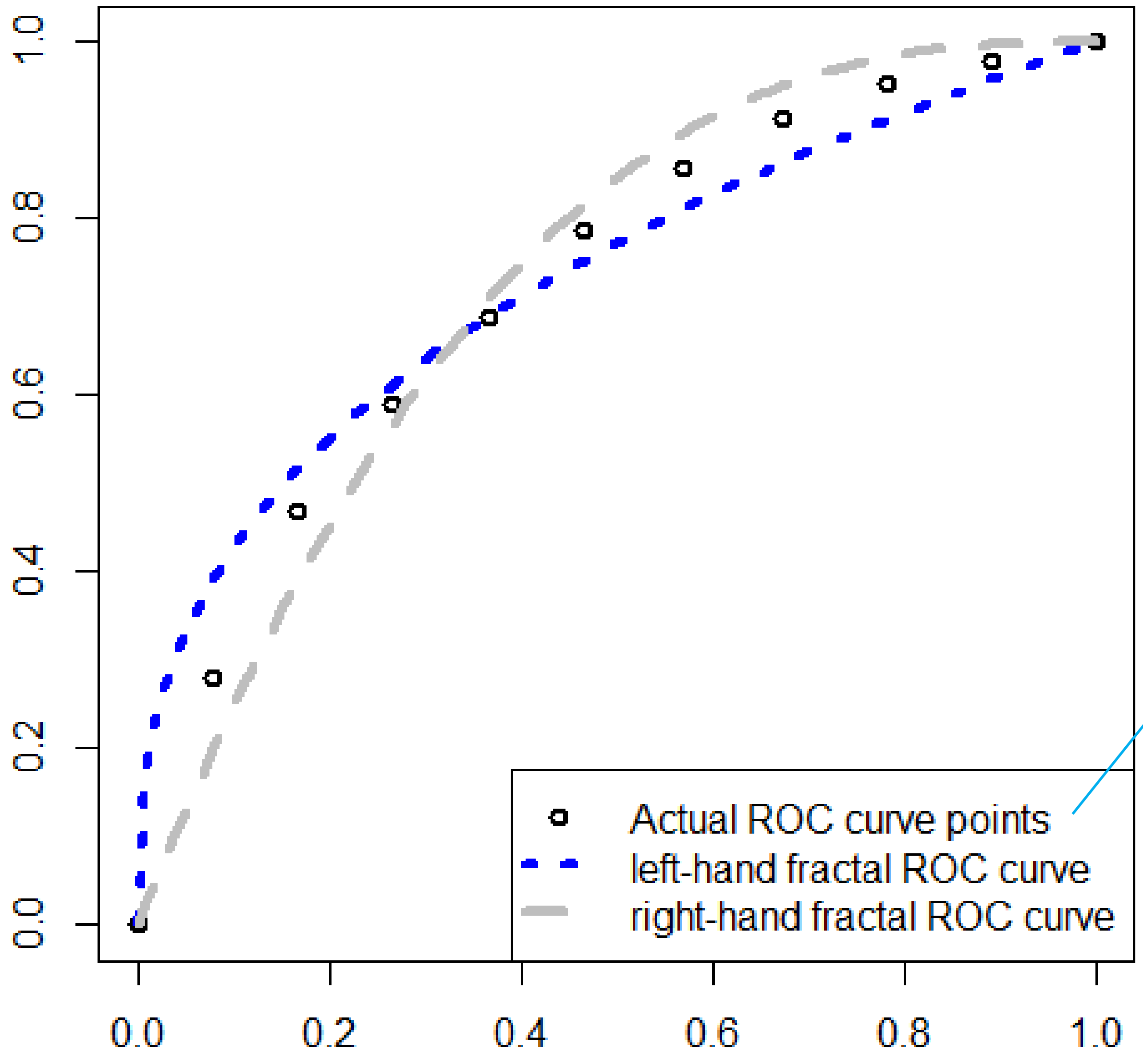
$$g(x) = x^{\frac{1-\gamma}{1+\gamma}}, x \in [0, 1]$$

Right-hand fractal ROC curve:



$$f(x) = 1 - (1 - x)^{\frac{1+\gamma}{1-\gamma}}, x \in [0, 1].$$

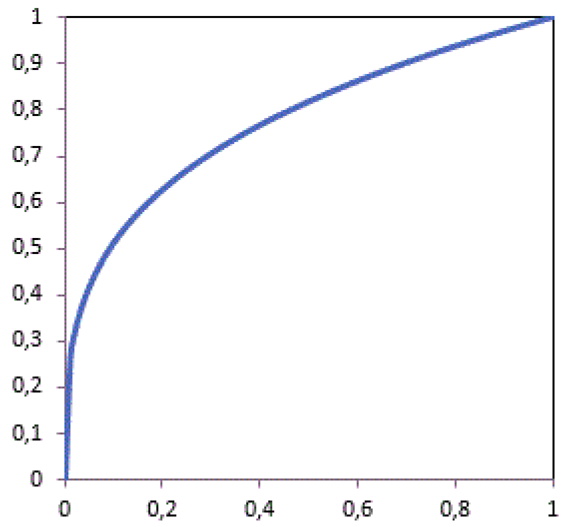
For each Gini coefficient (gamma) from [0,1)
there is exactly <u>one left-hand</u>
and <u>one right-hand</u> fractal ROC curve.



Real-life ROC curves lie between the two fractal curves.

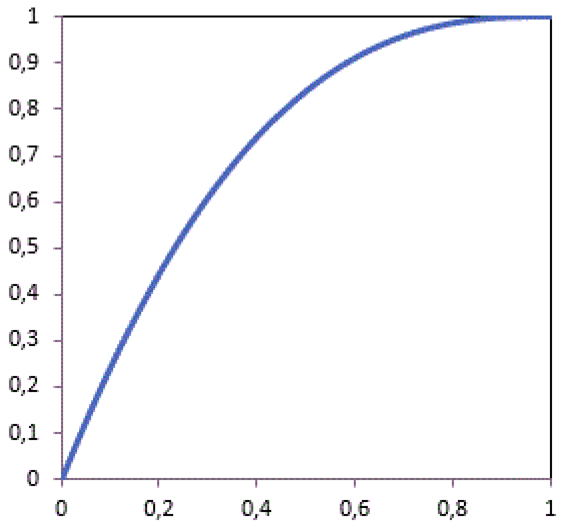
Actual ROC curve points „from an unnamed European consumer finance institution”
Rezac, Rezac [2011]

Left-hand fractal ROC curve:



$$g(x) = x^{\frac{1-\gamma}{1+\gamma}}, x \in [0, 1]$$

Right-hand fractal ROC curve:



$$f(x) = 1 - (1 - x)^{\frac{1+\gamma}{1-\gamma}}, x \in [0, 1].$$

linear combination (weighted with beta parameter)

$$r(x) = \beta f(x) + (1 - \beta) g(x) = \beta \left(1 - (1 - x)^{\frac{1+\gamma}{1-\gamma}}\right) + (1 - \beta) x^{\frac{1-\gamma}{1+\gamma}} \qquad (10)$$

In this setup, a real-life ROC curve can be modelled with just two parameters. One of them (γ) represents Gini coefficient, another, (β) is responsible for the shape of the curve. Formula (10) returns a left-hand fractal ROC curve for $\beta = 0$ and a right-hand fractal ROC curve for $\beta = 1$.

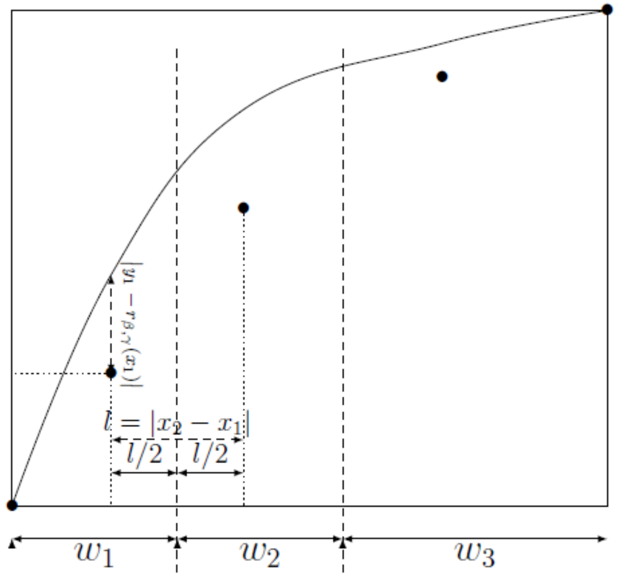


Figure 5: Illustration of setting up weights for the BOBYQA objective function.

In order to use the combination of fractal curves for real-life ROC curves modelling, we need to find Beta and Gamma values that best match real life data.

Once we introduced equation (10), we can fit the curve to the empirical data points. To fit the curve we are using function bobyqa (described by [6]) from the R package minqa [1]. We choose the following objective function to be minimized:

$$f_{obj}(\boldsymbol{x}, \boldsymbol{y}, \beta, \gamma) = \sum_{i=1}^{n} |y_i - r_{\beta,\gamma}(x_i)| \cdot w_i, \tag{11}$$

where x_i, y_i are coordinates of points from the empirical ROC curve and $\boldsymbol{x}, \boldsymbol{y}$ are vectors containing those coordinates; $r_{\beta,\gamma}(x_i)$ is a vertical coordinate of x_i point calculated with equation (10) and w_i are weights established based on vector $\boldsymbol{x}$ in a way illustrated in figure 5.

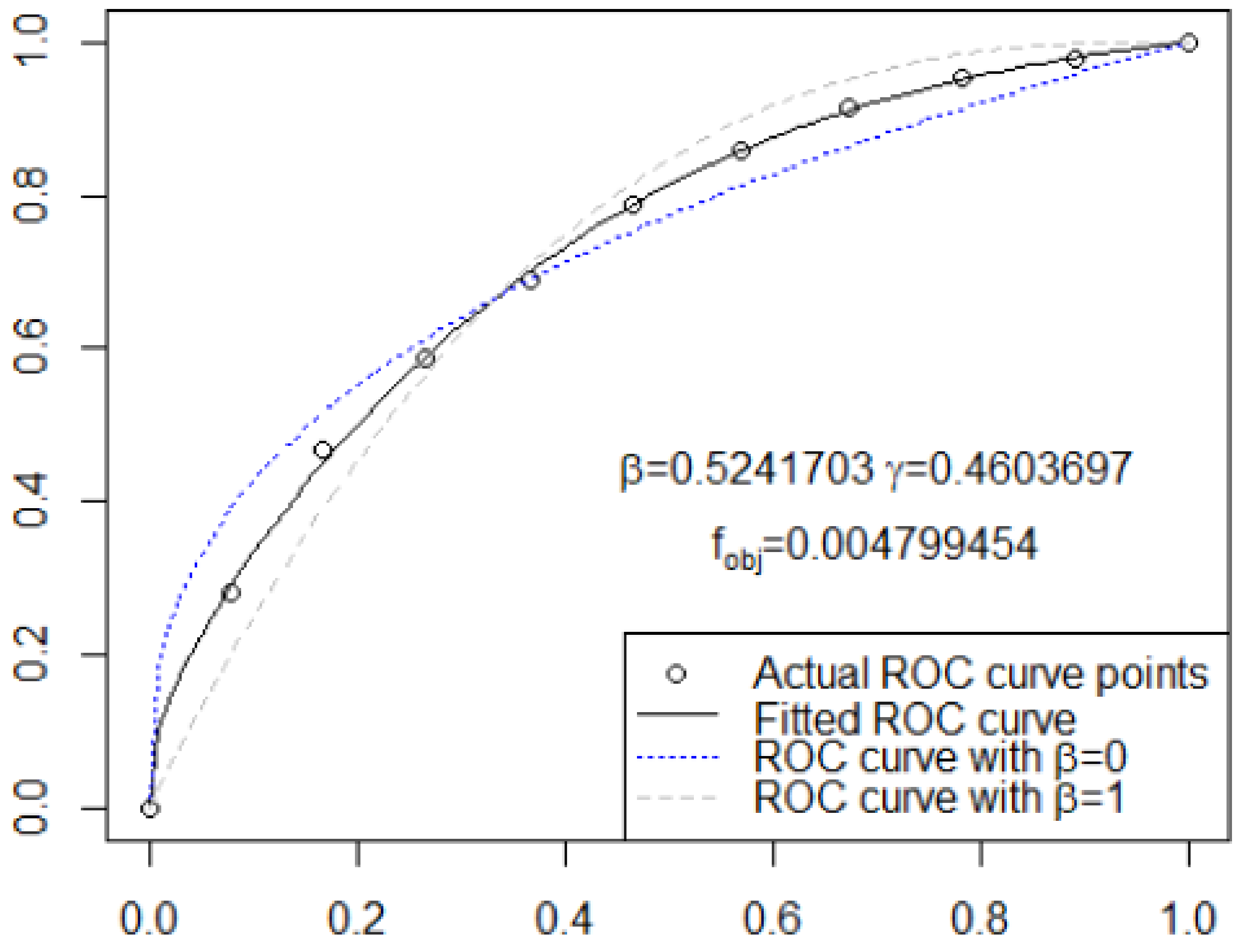


Figure 6: A model ROC curve from equation (10) fitted to the data from [7].

[7] M. Rezac; F. Rezac (2011): How to measure the quality of credit scoring models, Finance a Uver, Volume 61(5), p. 486, Charles University, Faculty of Social Sciences.

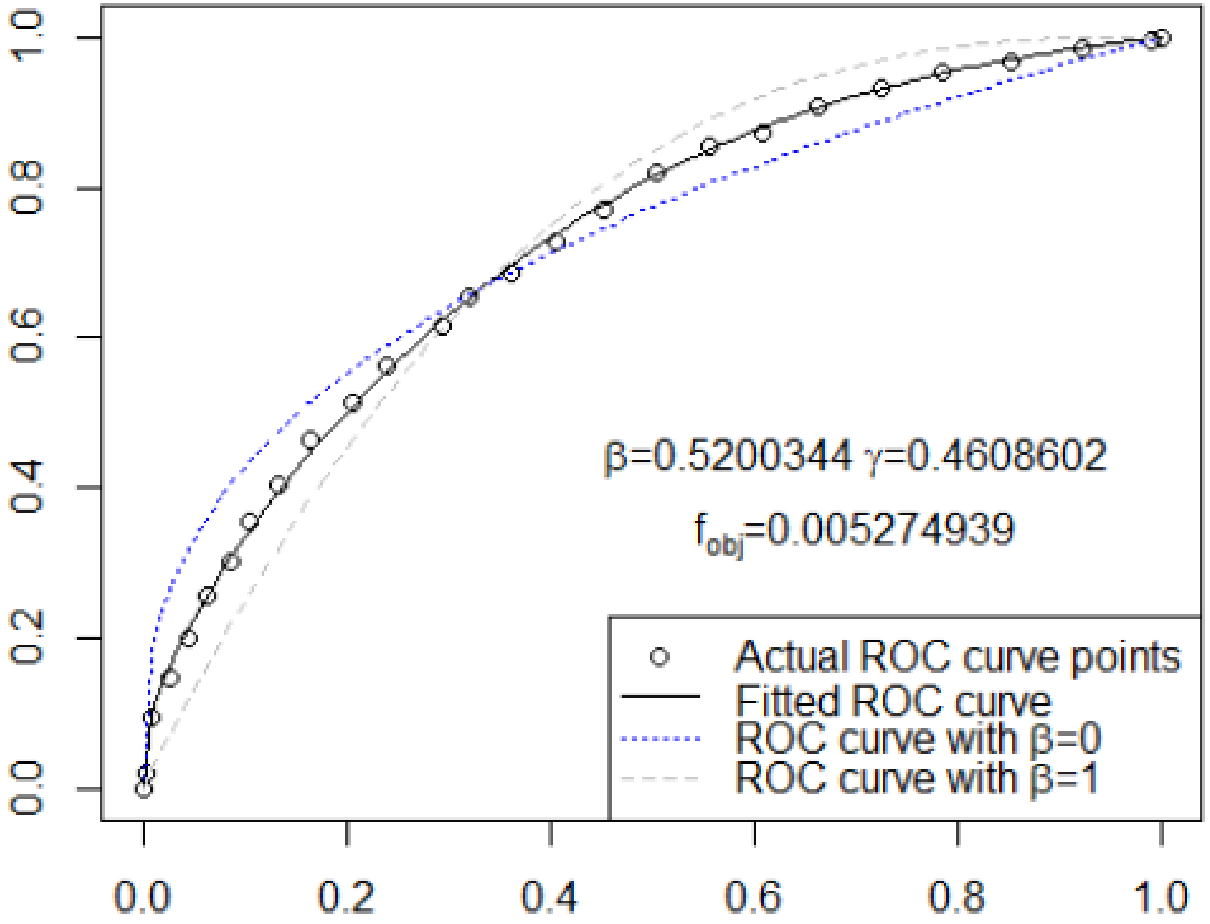


Figure 7: A curve fitted to more data points read from Lorenz curve in [7].

[7] M. Rezac; F. Rezac (2011): How to measure the quality of credit scoring models, Finance a Uver, Volume 61(5), p. 486, Charles University, Faculty of Social Sciences.

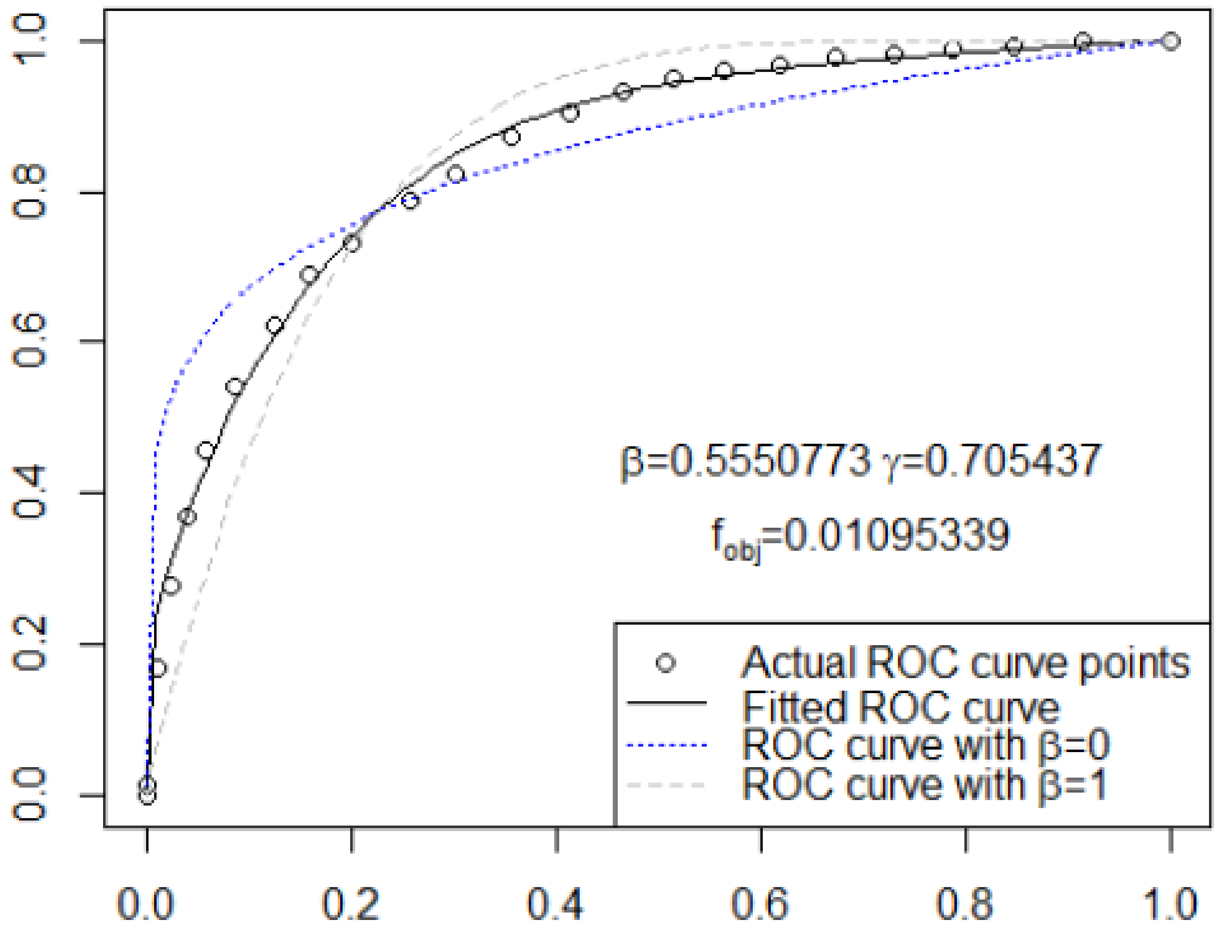


Figure 8: A curve fitted to the empirical ROC curve found in [9].

[9] Bartosz Wójcicki, Grzegorz Migut (2010): Wykorzystanie skoringu do przewidywwania wyłudzeń w Invest Banku. w: Scoring w Zarzadzaniu Ryzykiem, Statsoft, p. 47.

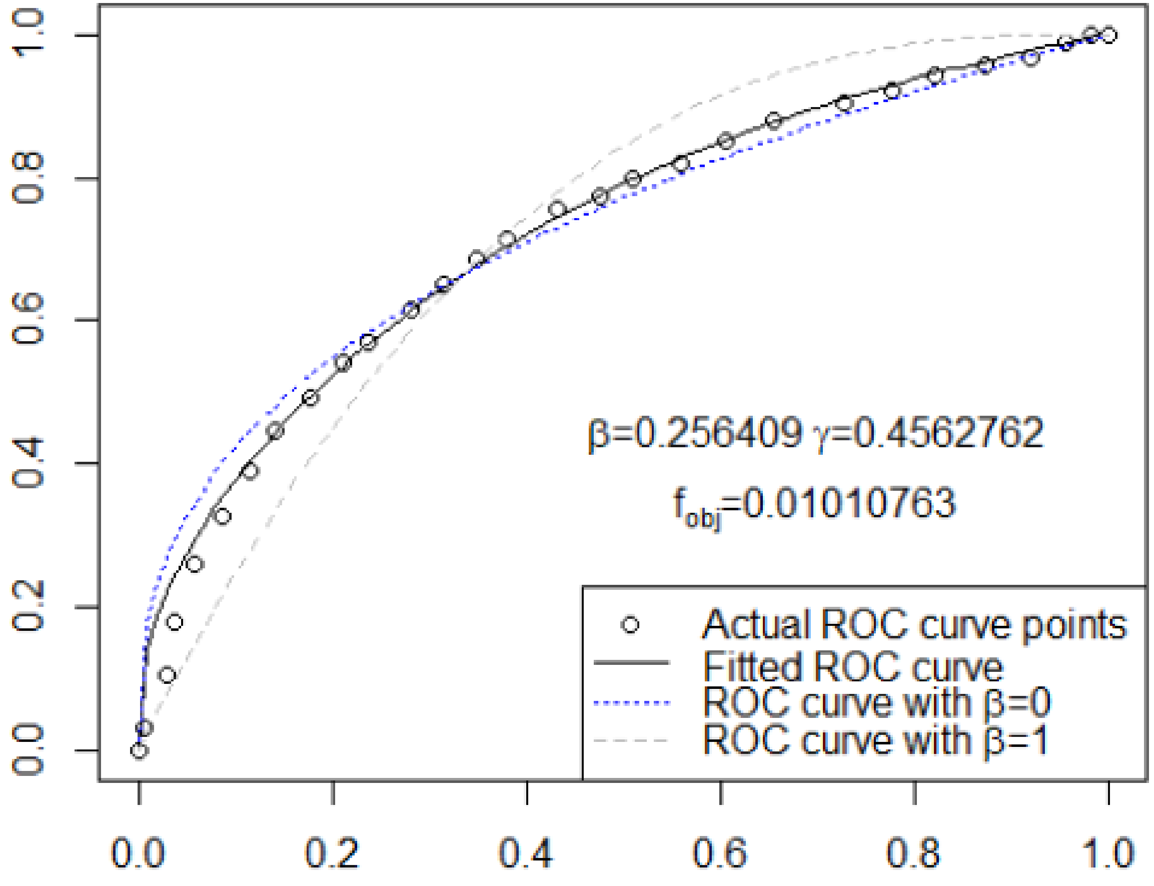


Figure 9: A curve fitted to the empirical data on Altman Z-score from [4].

[4] B. Engelmann, D. Tasche (2003): Testing Rating Accuracy, "Risk Magazine", 1/2003.

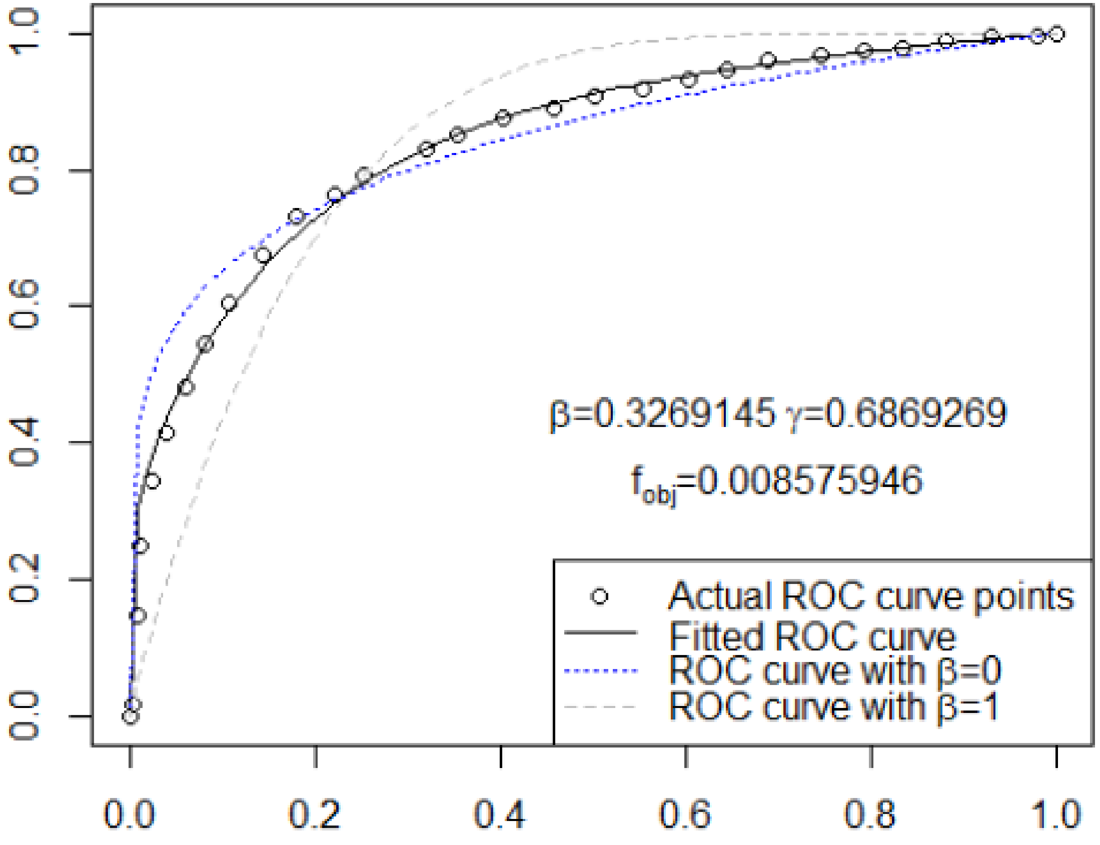


Figure 10: A curve fitted to the empirical data on a logit score from [4].

[4] B. Engelmann, D. Tasche (2003): Testing Rating Accuracy, "Risk Magazine", 1/2003.

Source	Shape param. (β)	Gini param. (γ)	Objective function (f_{obj})	Comments
[9]	0.555	0.705	0.011	Polish anti-fraud model
[4]	0.327	0.687	0.009	German SME logit model
[4]	0.257	0.456	0.010	Altman's Z on German SMEs
[7]	0.524	0.460	0.005	Unnamed European institution
[7]	0.520	0.461	0.005	As above, but with more data read from a Lorenz curve graph
[2]	0.268	0.570	0.018	A "CB (Credit Bureau?) score" of unknown origin
[2]	0.736	0.552	0.005	An "App (application?) score" of unknown origin
[3]	0.221	0.608	0.023	Credit cards - genetic algorithm
[3]	0.317	0.671	0.014	Credit cards - neural networks
[5]	0.993	0.687	0.029	A model based on logistic regression small bank - indiv. entrepreneurs
[5]	0.657	0.451	0.014	As above - small businesses

Table 1: Results of fitting a linear combination of two fractal ROC curves to empirical ROC curves.

[2] P. Beling, Z. Covaliu and R.M. Oliver (2005): Optimal scoring cutoff policies and efficient frontiers, Journal of the Operational Research Society (2005) 56, 1016–1029.

[3] Dario Bernardo, Hani Hagras, Edward Tsang (2013): A Genetic Type-2 fuzzy logic based system for financial applications modelling and prediction, Conference Paper in IEEE International Conference on Fuzzy Systems.

[4] B. Engelmann, D. Tasche (2003): Testing Rating Accuracy, "Risk Magazine", 1/2003.

[5] Michiko Miyamoto (2014): Credit Risk Assessment for a Small Bank by Using a Multinomial Logistic Regression Model, International Journal of Finance and Accounting 2014, 3(5): 327-334.

[7] M. Rezac; F. Rezac (2011): How to measure the quality of credit scoring models, Finance a Uver, Volume 61(5), p. 486, Charles University, Faculty of Social Sciences.

[9] Bartosz Wójcicki, Grzegorz Migut (2010): Wykorzystanie skoringu do przewidywwania wyłudzeń w Invest Banku. w: Scoring w Zarzadzaniu Ryzykiem, Statsoft, p. 47.

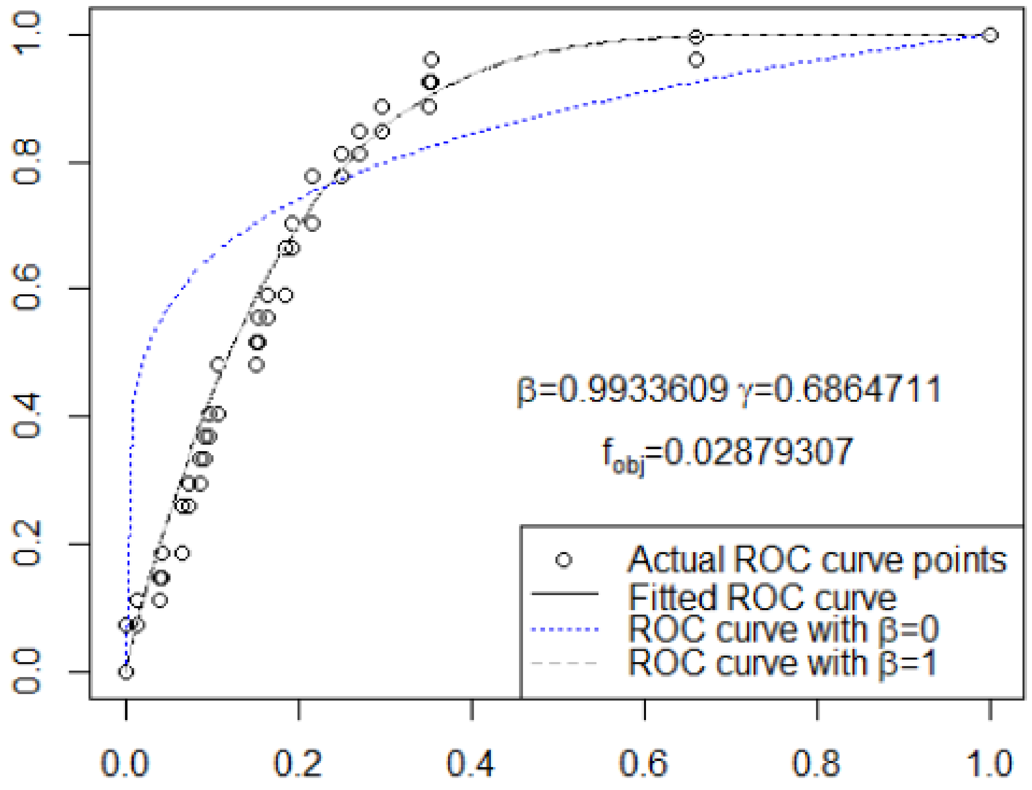


Figure 11: A curve fitted to the empirical data on a model for individual entrepreneurs from [5].

[5] Michiko Miyamoto (2014): Credit Risk Assessment for a Small Bank by Using a Multinomial Logistic Regression Model, International Journal of Finance and Accounting 2014, 3(5): 327-334.

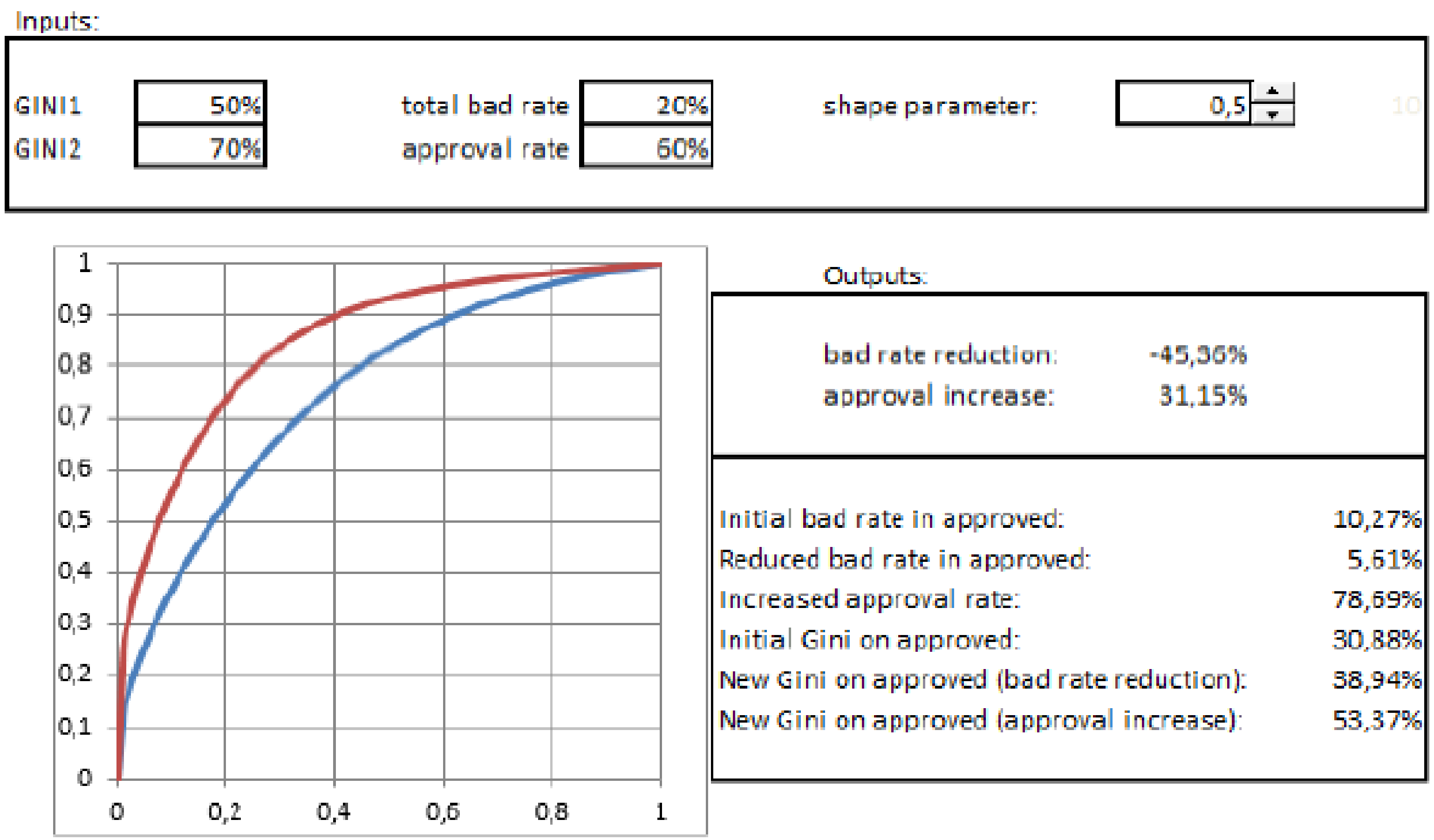


Figure 12: Interface of MS Excel calculation engine enabling modelling impact of Gini change.

- Solving integral/differential equations
- ROC curve model (linear combination of two fractal ROC curves)
- BOBYQA-based code in R for fitting the model curve to real data point
- Fitting model ROC curve to real data
- C++/Excel calculation engine to model impact of Gini change
- Results from the calculation engine

Approval rate	10%	20%	30%	40%	50%	60%	70%	80%
Bad rate	Bad rate reduction							
5%	-5.9%	-6.9%	-7.4%	-7.3%	-6.7%	-5.7%	-4.6%	-3.3%
10%	-5.9%	-6.9%	-7.3%	-7.1%	-6.3%	-5.3%	-4.1%	-2.9%
15%	-5.9%	-6.8%	-7.2%	-6.8%	-5.9%	-4.8%	-3.6%	-2.5%
20%	-5.9%	-6.8%	-7.0%	-6.4%	-5.4%	-4.3%	-3.2%	-2.1%
25%	-5.8%	-6.7%	-6.7%	-6.0%	-5.0%	-3.8%	-2.7%	-1.8%
30%	-5.8%	-6.6%	-6.4%	-5.6%	-4.5%	-3.3%	-2.3%	-1.5%
35%	-5.7%	-6.4%	-6.1%	-5.1%	-4.0%	-2.9%	-2.0%	-1.2%
40%	-5.7%	-6.2%	-5.7%	-4.6%	-3.5%	-2.5%	-1.7%	-1.0%
45%	-5.6%	-6.0%	-5.2%	-4.1%	-3.0%	-2.1%	-1.4%	-0.8%
50%	-5.5%	-5.6%	-4.7%	-3.6%	-2.6%	-1.8%	-1.1%	-0.7%

Table 2: Bad rate reduction resulting from Gini coefficient's increase from 0.5 to 0.52 for various combination of total bad rate and approval rate (β assumed to be 0.5) – simulated with C++ calculation engine.

Approval rate	10%	20%	30%	40%	50%	60%	70%	80%
β	Bad rate reduction							
0,05	-4.7%	-4.6%	-4.5%	-4.2%	-3.9%	-3.5%	-3.0%	-2.3%
0,1	-4.8%	-4.8%	-4.6%	-4.4%	-4.1%	-3.6%	-3.0%	-2.3%
0,15	-4.9%	-4.9%	-4.8%	-4.6%	-4.2%	-3.6%	-2.9%	-2.2%
0,2	-4.9%	-5.1%	-5.0%	-4.8%	-4.3%	-3.6%	-2.9%	-2.1%
0,25	-5.1%	-5.3%	-5.3%	-5.0%	-4.4%	-3.7%	-2.9%	-2.1%
0,3	-5.2%	-5.5%	-5.5%	-5.1%	-4.5%	-3.7%	-2.8%	-2.0%
0,35	-5.3%	-5.7%	-5.8%	-5.3%	-4.6%	-3.7%	-2.8%	-1.9%
0,4	-5.4%	-6.0%	-6.1%	-5.6%	-4.7%	-3.8%	-2.8%	-1.9%
0,45	-5.6%	-6.3%	-6.4%	-5.8%	-4.8%	-3.8%	-2.8%	-1.8%
0,5	-5.8%	-6.7%	-6.7%	-6.0%	-5.0%	-3.8%	-2.7%	-1.8%
0,55	-6.1%	-7.1%	-7.1%	-6.3%	-5.1%	-3.8%	-2.7%	-1.7%
0,6	-6.4%	-7.6%	-7.5%	-6.5%	-5.2%	-3.9%	-2.7%	-1.7%
0,65	-6.7%	-8.2%	-8.0%	-6.8%	-5.3%	-3.9%	-2.7%	-1.7%
0,7	-7.2%	-8.8%	-8.5%	-7.1%	-5.4%	-3.9%	-2.6%	-1.6%
0,75	-7.8%	-9.6%	-9.0%	-7.3%	-5.5%	-3.9%	-2.6%	-1.6%
0,8	-8.7%	-10.7%	-9.6%	-7.7%	-5.7%	-4.0%	-2.6%	-1.6%
0,85	-10.0%	-11.9%	-10.4%	-8.0%	-5.8%	-4.0%	-2.6%	-1.5%
0,9	-12.1%	-13.6%	-11.2%	-8.3%	-5.9%	-4.0%	-2.6%	-1.5%
0,95	-16.3%	-15.8%	-12.1%	-8.7%	-6.0%	-4.0%	-2.6%	-1.5%

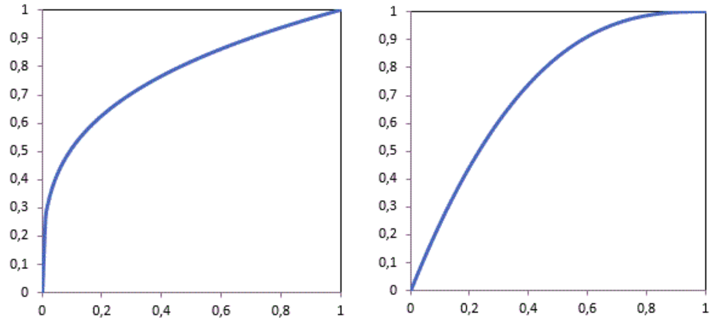
Table 3: Bad rate reduction resulting from Gini coefficient's increase from 0.5 to 0.52 for various combination of total bad rate and ROC curve shape parameter β (population bad rate assumed to be 0.25) – simulated with C++ calculation engine.

Gini change	Bad rate reduction
$0,45 \rightarrow 0,46$	-2.7%
$0,45 \rightarrow 0,47$	-5.3%
$0,45 \rightarrow 0,48$	-8.0%
$0,45 \rightarrow 0,49$	-10.6%
$0,45 \rightarrow 0,50$	-13.3%
$0,45 \rightarrow 0,51$	-15.9%
$0,45 \rightarrow 0,52$	-18.5%
$0,45 \rightarrow 0,53$	-21.1%
$0,45 \rightarrow 0,54$	-23.6%
$0,45 \rightarrow 0,55$	-26.2%

Table 4: Bad rate reduction resulting from Gini coefficient's increase from 0.45 (population bad rate assumed at 0.25, approval rate at 0.40, β at 0.5) – simulated with C++ calculation engine.

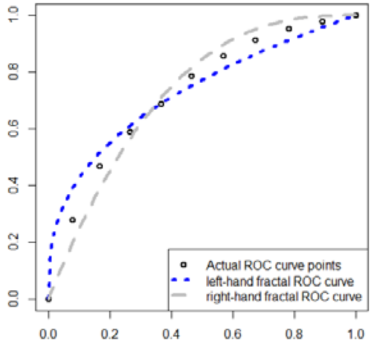
Summary

1



Fractal ROC curves do exist...

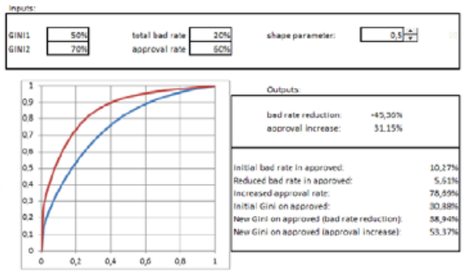
2



and are useful for modelling real-life ROC curves,

...which helps us assess impact of Gini improvement on credit losses

3



Example

GINI 0.50
↗ 0.52

Scored population bad rate = 25%
Approval rate = 40%
Beta = 0.5

Portfolio bad rate
↘ by 6%