Do-It-Yourself Microfinance Scoring Trees

Mark Schreiner

Microfinance Risk Management
www.microfinance.com

Originally presented in Spanish as
"Arboles estadísticas en microfinanzas: Haz tu propio scorecard en tu tiempo libre"
Third Seminar on Banking and Microfinance in Latin America

November 12-13, 2001

The Central Bank of the Dominican Republic
Santo Domingo

Agenda

- **Introduction: What is scoring?**
- **Example: Statistical trees**
- **Comparison: Trees v. other methods**
- **Guidelines for using scoring**

Not just trees, but rather scoring!

Introduction: What is scoring?

Forecasts of risk that suppose that:

- **Quantified characteristics affect risk:**
  - **Borrower** (Age, type of business)
  - **Loan** (Amount disbursed, term)
  - **Lender** (Experience, loan officer)

- **The future will be like the past**

Scoring links characteristics with risk.

Why use scoring?

- **Microlenders already use scoring!**
  — Qualitative (implicit, sixth sense)
  — Quantitative (explicit, statistical)

- **Finance is risk management. Statistical scoring quantifies risk and makes its evaluation consistent and explicit**

Why use scoring? (cont.)

- **Scoring focuses evaluation and collections effort where it counts most**

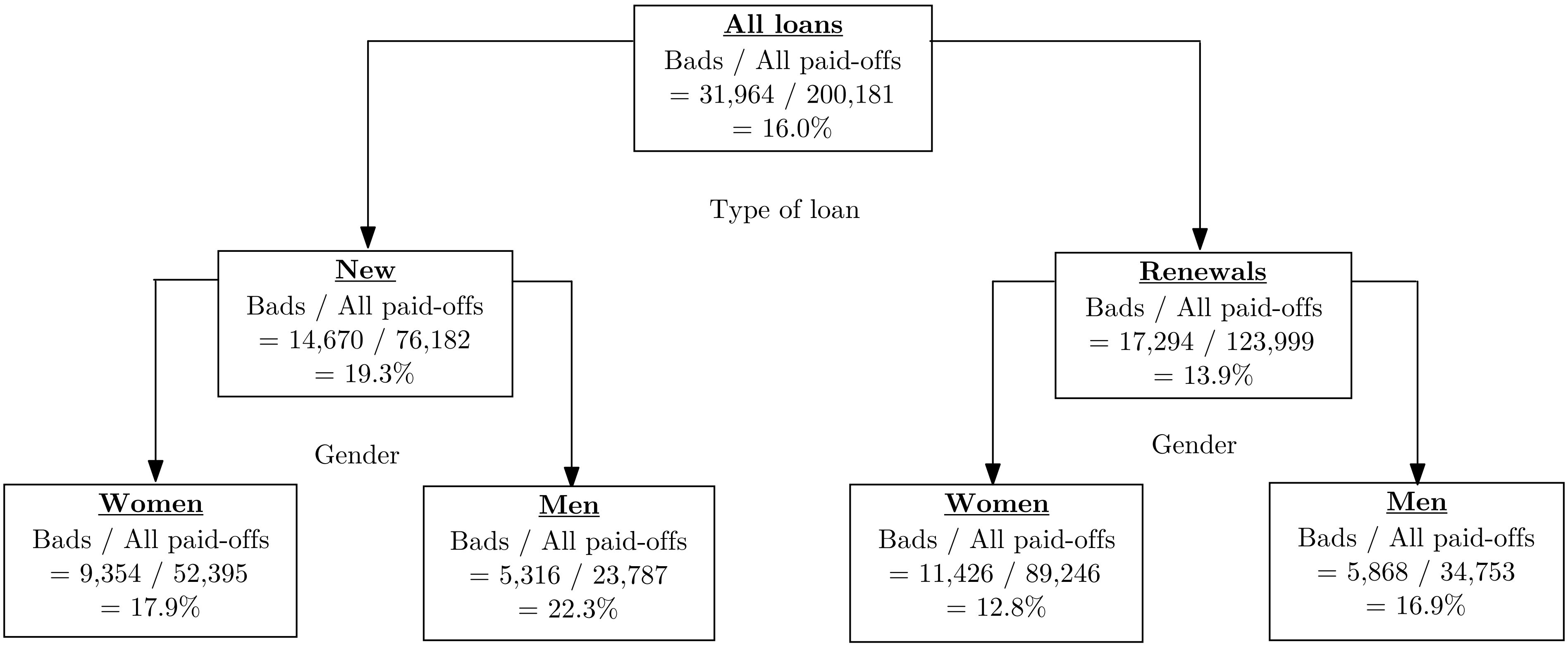
- **In microfinance, scoring does not replace loan officers; rather, it complements their subjective evaluation of the client's household and business**

What is a statistical tree?

- **Like all other scoring methods, trees link past characteristics with past arrears and suppose that the future will be like the past**

- **Trees closely resemble expert systems:**
  - **Trees use quantitative experience as recorded in an electronic data base**
  - **Expert systems use qualitative experience as recorded in people's minds**

Simple example of a statistical tree



Risk forecasts by the statistical tree:

16.3%	**21.0%**	**13.1%**	**16.9%**

Output of an expert system:

Normal	**High risk**	**Low risk**	**Normal**

Simple example tree from Colombia

- **Microlender: Women's World Banking, with Hans Dellien**
- **"Bad": 7 days arrears/installment or 30 days in a row**

Data base	Year	Risk	# loans
Construction:	1992-98	15.5%	141,759
Test:	1999	17.6%	68,759

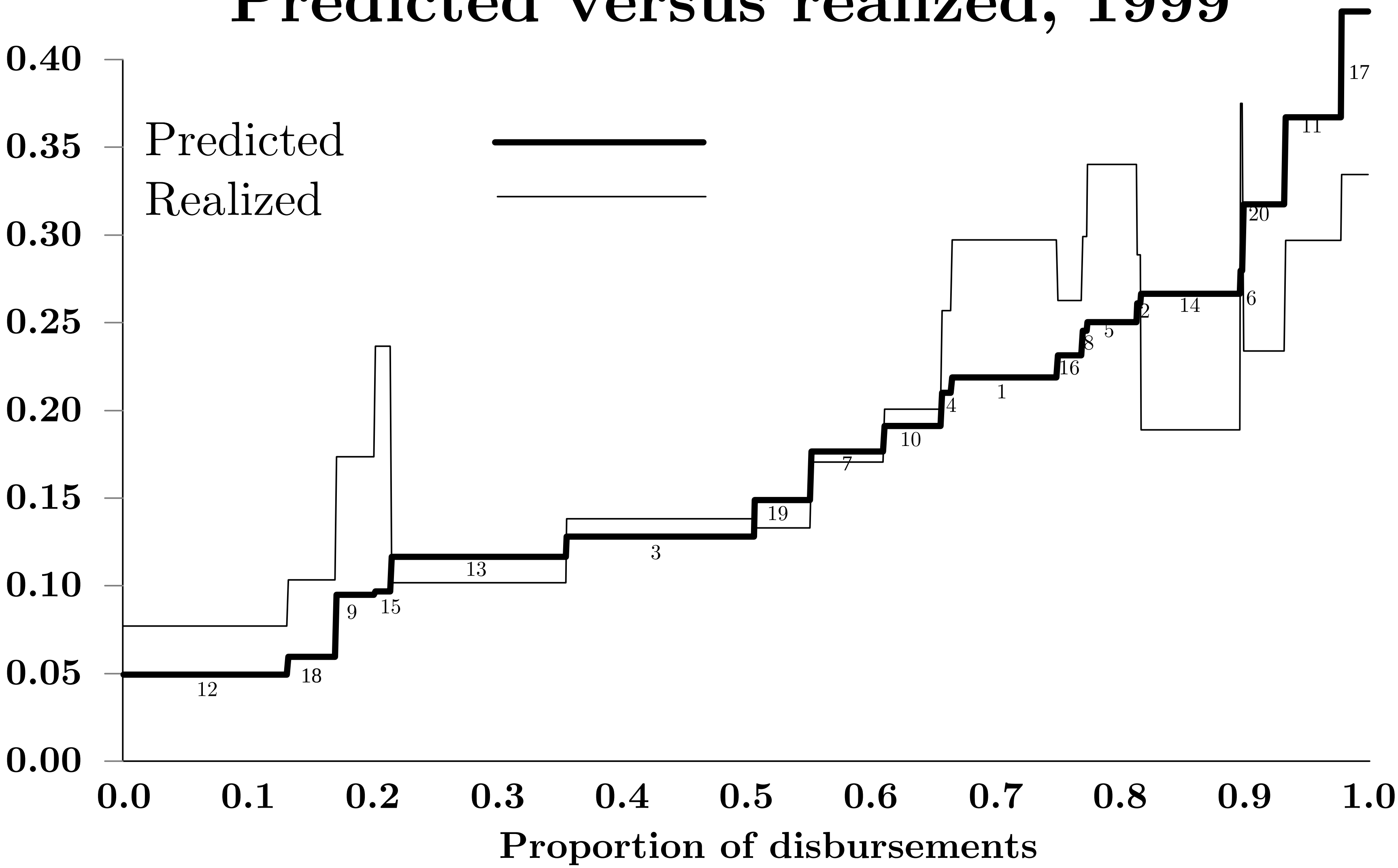
Test: Compare risk predicted by tree built w/data up to end of 1998 with risk realized in 1999

Details of tree from Colombia

#				Telephone o	1992-1998				1999			
Class	**Type**	**Gende**	**Collateral**	**past arrears**	Bads	Cases	% port.	% bad	Bads	Cases	% port.	% bad
1	New	Woman	Personal	Telephone	2,519	11,520	8.1	21.9	1,748	5,880	8.6	29.7
2				No telephone	96	368	0.3	26.1	67	232	0.3	28.9
3			Other	Telephone	2,647	20,673	14.6	12.8	1,429	10,332	15.0	13.8
4				No telephone	252	1,200	0.8	21.0	131	510	0.7	25.7
5		Man	Personal	Telephone	1,592	6,362	4.5	25.0	940	2,763	4.0	34.0
6				No telephone	104	372	0.3	28.0	54	144	0.2	37.5
7			Other	Telephone	1,585	8,978	6.3	17.7	683	4,003	5.8	17.1
8				No telephone	159	648	0.5	24.5	73	244	0.4	29.9
9	Renew	Woman	Personal	0 days	517	5,448	3.8	9.5	374	2,156	3.1	17.3
10				up to 2 days	1,453	7,600	5.4	19.1	645	3,213	4.7	20.1
11				>2 days	1,329	3,621	2.6	36.7	914	3,078	4.5	29.7
12			Other	0 days	1,076	21,789	15.4	4.9	695	9,021	13.1	7.7
13				up to 2 days	2,279	19,542	13.8	11.7	989	9,721	14.1	10.2
14				>2 days	1,906	7,153	5.0	26.6	1,031	5,458	7.9	18.9
15		Man	Personal	0 days	247	2,552	1.8	9.7	206	871	1.3	23.7
16				up to 2 days	862	3,724	2.6	23.1	356	1,356	2.0	26.3
17				>2 days	813	1,902	1.3	42.7	525	1,570	2.3	33.4
18			Other	0 days	452	7,614	5.4	5.9	279	2,702	3.9	10.3
19				up to 2 days	1,150	7,729	5.5	14.9	417	3,135	4.6	13.3
20				>2 days	941	2,964	2.1	31.7	551	2,356	3.4	23.4

Source: Computed by Schreiner with data from WWB/Colombia

Predicted versus realized, 1999



Lessons from the simple tree

- It predicted well: Loans with low predicted risk turned out to have low realized risk

- It highlights attributes linked with lower risk:
  - Being a woman
  - Owning a telephone
  - Not having required a personal guarantee
  - Having had low previous arrears

- In some segments, it did not predict well

Effects on profitability

- Management chooses how to use scoring weighing trade-offs among breadth, depth, and length of outreach

- Effects on profitability and on number of disbursements can be estimated beforehand

- Assumptions:
  - Net gain of avoiding a bad: $250
  - Net cost of losing a good: $100

<u>Effects on profitability</u> (cont.)

- **With tree in 1999, if WWB had rejected all renewal applications with > 2 days of previous arrears:**

— Avoid	1,439 bads (31% of rejects)	
	32% of all bads	
— Disburse	16.9% fewer loans	
— Lose	3,209 goods (69% rejects)	
— Save	\$250·1,439	=\$359,750
— Lose	\$100·3,209	=\$320,900

- **Net effect: save \$38,850, plus time and effort**

Example of simple regression model

(Specialized statistical work)

$$
\begin{aligned}
\text{Risk} = {} & 0.100 \\
& + 0.010 \cdot \text{Loan sequence} \\
& - 0.001 \cdot \text{Age} \\
& - 0.015 \cdot \text{Woman} \\
& + 0.001 \cdot \text{Amount disbursed} \\
& + 0.020 \cdot \text{Longest previous arrears} \\
& - 0.045 \cdot \text{Ownership of telephone}
\end{aligned}
$$

Use of regression scorecard

Renewal for 100 to 30-year-old man with a telephone and 4 days previous arrears:

Risk	$= 0.100$	
	$+ 0.010 \cdot 2$	**(Sequence)**
	$- 0.001 \cdot 30$	**(Age)**
	$- 0.015 \cdot 0$	**(Man)**
	$+ 0.001 \cdot 100$	**(Disbursement)**
	$+ 0.020 \cdot 4$	**(Previous arrears)**
	$- 0.045 \cdot 1$	**(Owns telephone)**
	$= 0.225 = 22.5\%$	

Comparison: Trees versus other methods

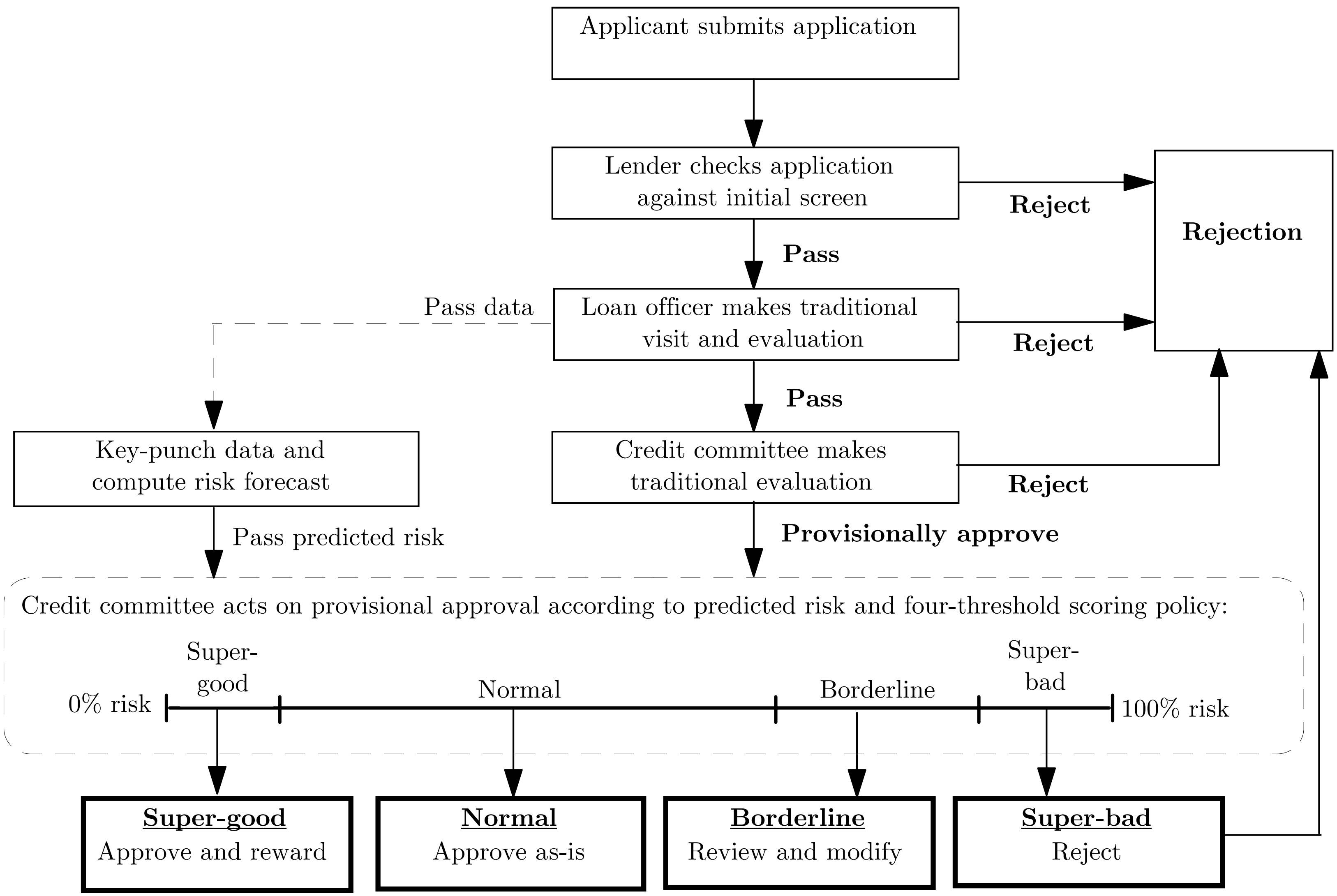
Criterion	Trees	Regression	Expert system
Multivariate	Yes	Yes	Yes
Statistical	Yes	Yes	No
Reveals links with risk	No	Yes	No
Tolerates dirty data	Yes	No	Yes
Data requirements	Medium	High	None
Ease of construction	Easy	Difficult	Easy
Ease systems integration	Equal		
Ease of understanding	Simple	Complex	Simplest
Predictive power	Better	Best	Good

Stages in scoring

1. Accumulation of data
2. Statistical work
3. Integration in systems
4. Daily use in branches

$\therefore$ The hard part is daily use in branches; whether the other stages work can be established before they are completed

Where scoring fits in traditional underwriting



The “Scoring Follow-up Report”

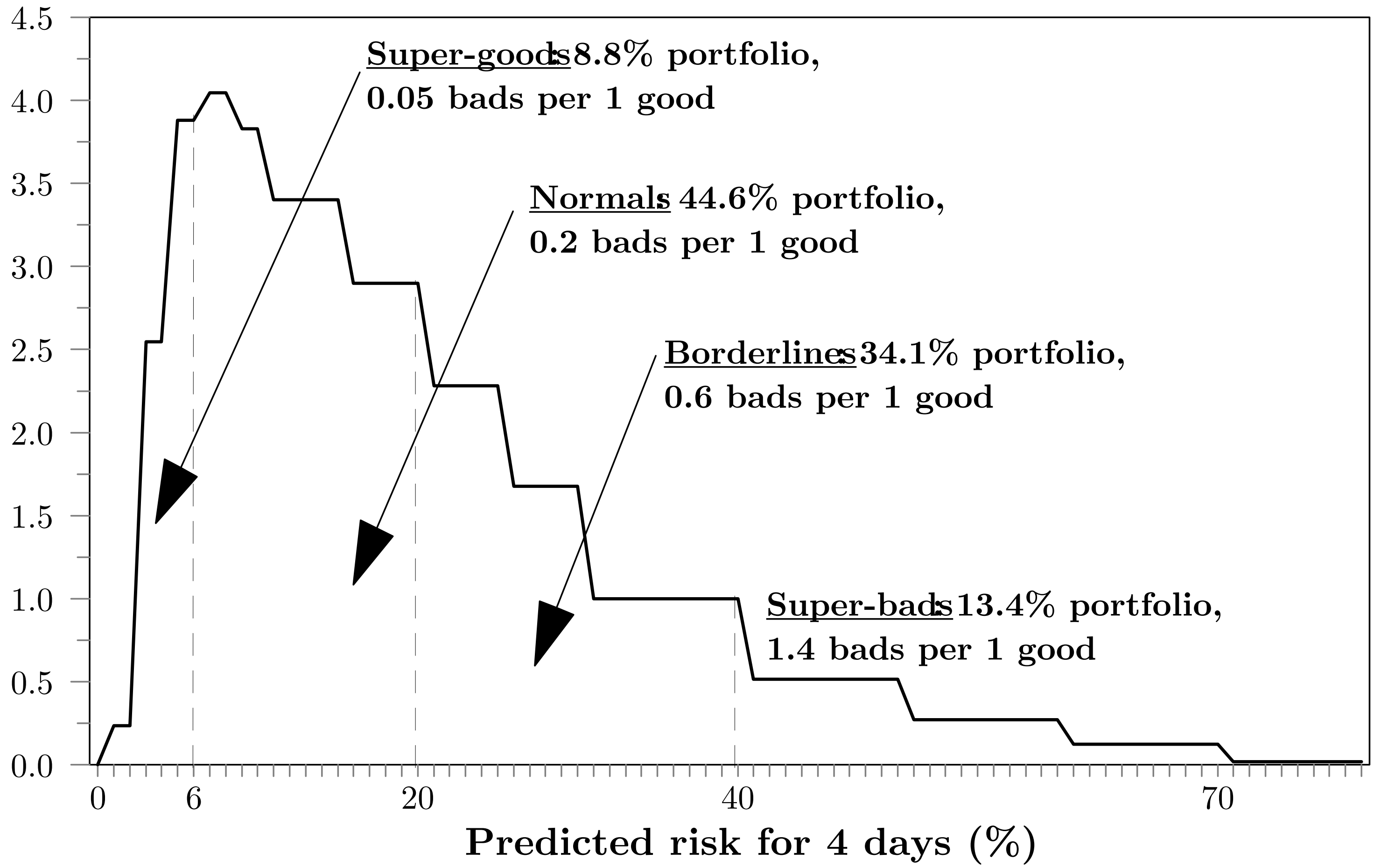
- **Central report in scoring**
- **Checks predictive power**
- **Shows distribution of portfolio risk for setting policy thresholds**
- **Indicates whether overrides made sense**

Scoring Follow-up Report, Risk of 4 days, Colombia

Date tested: 01/12/01 | **Date scorecard constructed**: 31/07/01 | **Branch**: All

Forecast risk (%)	# Loans out. (%)	Realized risk (%) by days since disbursement			
		0-90	91-180	181-270	271+
0-2	0.5	1.4	2.0	0.0	4.0
2-4	5.1	2.8	2.8	2.1	3.5
4-6	7.8	3.0	4.0	4.0	5.1
6-8	8.1	3.9	4.8	5.5	8.1
8-10	7.7	5.1	6.7	6.4	11.5
10-15	17.0	5.5	8.1	11.6	18.1
15-20	14.5	6.8	12.1	17.9	27.6
20-25	11.4	9.0	16.9	23.8	33.1
25-30	8.4	11.4	19.4	30.4	37.8
30-40	10.0	14.6	25.0	37.3	45.8
40-50	5.1	18.4	30.4	50.9	53.6
50-60	2.7	23.0	42.3	57.2	60.4
60-70	1.2	32.4	42.6	65.2	70.5
70+	0.5	34.3	62.9	65.5	77.9

Management sets policy thresholds



Predicted versus realized risk, worst 30

Report date : 31/07/2001		Branch : All		Risk : 1 spell >=30 days				Loans : out. >270 days			List : 30 worst
						Realized risk					
Loan Code	Client name	Days out.	$ out.	Monthly payment	Next due	Current arrears	# spells	Days late/ installment	Longest spell	Bad?	Predicted risk (%)
79922	Javela, María	308	2,106	83	03-Aug	23	2	42.5	77	Bad	90
50973	Posada, María	334	1,860	71	29-Aug	0	3	21.1	36	Bad	81
71596	Arboleda, Nivelly	336	1,323	132	29-Aug	2	3	14.8	25	Good	80
80816	Beltrán, Dioselina	304	1,032	48	29-Aug	0	3	14.8	42	Bad	80
62037	Nuñez, Dolly	337	5,683	316	02-Aug	0	1	22.7	28	Good	72
45638	Cruz, Leonor	304	377	22	29-Aug	0	3	45.5	101	Bad	71
64823	Rivera, Antonia	304	603	39	29-Aug	23	2	22.2	39	Bad	68
61653	Marín, Graciela	337	5,763	283	02-Aug	0	4	14.5	25	Good	62
78800	Muñoz, Marco	304	2,003	111	29-Aug	0	3	25.7	67	Bad	60
24893	Silva, Oswaldo	304	388	29	29-Aug	86	2	36.0	86	Bad	59
65323	Ruíz, Asia	308	56	12	03-Aug	58	4	24.7	58	Bad	59
59506	Cardona, Graciela	334	188	51	29-Aug	0	2	11.9	18	Good	59
54093	Tejada, María	285	14,638	790	11-Aug	0	1	0.3	2	Good	58
71243	Castillo, Rosa	293	630	70	18-Aug	0	2	6.1	15	Good	58
22692	Tavárez, María	348	143	39	13-Aug	0	1	0.4	2	Good	58
99155	Marroquín, Libia	334	77	41	29-Aug	0	1	11.1	22	Good	58
18634	Rivera, Melida	334	470	50	29-Aug	191	2	82.7	191	Bad	57
74810	Marulanda, Pablo	304	331	27	29-Aug	23	3	25.8	54	Bad	56
20410	Valencia, Claudia	356	323	53	21-Aug	0	4	5.5	14	Good	55
60737	Suárez, Yolanda	335	275	40	03-Aug	0	1	0.5	2	Good	55
85854	Marín, Jorge	308	1,275	106	03-Aug	0	4	7.7	20	Good	55
42074	Lozano, Nevalia	292	251	19	18-Aug	86	2	52.0	93	Bad	54
30986	Berrios, Fanny	318	2,449	136	13-Aug	0	2	4.4	15	Good	54
31208	Gomez, Diofanor	306	6,049	291	01-Aug	0	3	4.5	12	Good	54
89020	Calderón, Editha	319	259	38	14-Aug	0	1	7.0	14	Good	54
8408	Marulanda, María	306	332	42	01-Aug	0	2	61.6	131	Bad	53
36244	Castillo, Brunilda	279	383	46	05-Aug	0	1	0.9	3	Good	52
5699	Ortíz, Nubia	334	570	46	29-Aug	0	2	15.5	39	Bad	52
7719	Montoya, Javier	281	100	17	07-Aug	36	3	12.6	36	Bad	52
40373	Moreno, Peregrino	304	381	50	29-Aug	177	4	68.9	177	Bad	51
Source: Regression scorecard and data base of Latin American microlender.								Ave. risk:		50	61

Predicted versus realized risk, best 30

Report date : 01/12/01		Branch : All		Risk : 1 spell >=30 days				Loans : out. >270 days			List : 30 best
						Realized risk					
Loan Code	Client name	Days out.	$ out.	Monthly payment	Next due	Current arrears	# spells	Days late/ installment	Longest spell	Bad?	Predicted risk (%)
62225	Valencia, Lucero	292	59	60	18-Aug	0	0	0.0	0	Good	0.5
38388	Betancourt, José	305	73	26	01-Aug	0	1	0.1	1	Good	0.5
88687	Valencia, Juan	279	35	36	05-Aug	0	0	0.0	0	Good	0.5
94799	Fernández, Zorrilla	281	289	38	07-Aug	0	0	0.0	0	Good	0.5
8154	Sánchez, Hernán	290	102	36	16-Aug	0	0	0.0	0	Good	0.5
38563	Escobar, Patricia	316	117	32	11-Aug	0	1	7.0	13	Good	0.5
27819	Echandia, Henry	322	102	36	17-Aug	0	0	0.0	0	Good	0.6
21502	Jaramillo, Ema	285	289	103	11-Aug	0	1	0.1	1	Good	0.6
71907	Guevara, César	295	87	31	20-Aug	0	0	0.0	0	Good	0.6
49562	Paz, María	336	768	167	01-Aug	0	1	0.8	5	Good	0.6
93142	Escobar, Mónica	284	35	36	10-Aug	0	0	0.0	0	Good	0.6
11221	Palomino, Fe	287	73	26	13-Aug	0	0	0.0	0	Good	0.7
88301	Garcia, Alberto	308	289	38	03-Aug	0	0	0.0	0	Good	0.7
77258	Arce, Eduardo	305	116	41	02-Aug	0	1	1.0	5	Good	0.7
1582	Contreras, Elena	318	147	77	13-Aug	0	1	0.1	1	Good	0.7
79476	Sánchez, Gonzalo	323	293	65	18-Aug	0	1	1.4	5	Good	0.7
985	Lopez, Flor	295	35	36	20-Aug	0	0	0.0	0	Good	0.7
85657	Torres, María	280	347	46	06-Aug	0	0	0.0	0	Good	0.7
16697	Chacón, Emilsa	293	73	26	18-Aug	0	1	4.0	20	Good	0.7
53165	Gutierrez, Lucila	356	153	55	21-Aug	0	0	0.0	0	Good	0.7
80399	López, Alejandro	291	460	86	17-Aug	0	1	0.1	1	Good	0.7
32949	Castaño, Alvaro	323	68	36	18-Aug	0	0	0.0	0	Good	0.7
94131	Duque, Lucia	287	219	78	13-Aug	0	0	0.0	0	Good	0.7
28050	Polanco, Gerardo	294	76	79	19-Aug	0	1	0.1	1	Good	0.7
30709	Fajardo, Carmen	349	101	103	14-Aug	0	0	0.0	0	Good	0.7
54730	Aristiza, Morena	287	73	26	13-Aug	0	0	0.0	0	Good	0.7
18377	Ceballos, Luís	314	168	45	09-Aug	0	0	0.0	0	Good	0.7
28881	Escobar, José	323	78	41	18-Aug	0	0	0.0	0	Good	0.8
34129	Muñoz, Edisón	283	461	86	09-Aug	0	0	0.0	0	Good	0.8
74078	Tabarez, Jesús	341	50	51	06-Aug	0	1	0.2	1	Good	0.8
								Ave. risk:		0.0	0.6

Source: Regression scorecard and data base of Latin American microlender.

Review of key points

- **Make your own tree: It is simple, quick, easy, and it can be powerfully predictive**

- **Test your scorecard with historical data**

- **Use scoring only for loans already provisionally approved under traditional underwriting norms**

- **Track scoring's performance in practice**