Diseño y Construcción de Scorecards con Herramientas de Data Mining

Cristián Bravo R.
cbravo@dii.uchile.cl

Banco de Crédito e Inversiones
2 al 5 de Julio, 2011

Agenda del Módulo

- Preparación de datos para generación de scorecards.
  - Selección de Variables.
  - Transformaciones Notables.
  - Segmentación de Universos.
- Modelos avanzados:
  - Análisis de Supervivencia.
  - Redes neuronales en Credit Scoring.
  - Modelos de optimización.
  - SVMs.
  - Consideraciones de Basilea II y uso avanzado de modelos.

Agenda del Módulo

- Inferencia en los Rechazados.
- Construcción de un Scorecard.
  - Transformación a log-odds.
- Puntos de Corte.
- Stress Testing.
- Seguimiento.
- Loss Given Default

Implementación y Estrategias de punto de corte

Implementación de Modelos

- Dos cosas se deben considerar al implementar modelos.
  - Qué preguntar al cliente.
  - El punto de corte utilizado.
- Qué preguntar al cliente.
  - Es mala idea preguntar solamente lo que incluye el scoring.
    - Permite a los ejecutivos conocer el modelo detrás.
    - Impacta en las futuras actualizaciones.
- Opción: Aumentar las preguntas con otra información relevante.
  - Trabajo intra-área.

Punto de Corte

- Los modelos entregan salidas continuas (entre 0 y 1) interpretables como probabilidades.
- ¿Qué valor utilizar para decidir si un evento es de clase 0 o 1 (bueno o malo)?
- Estrategias:
  - Según máxima verosimilitud: corte en 0,5.
  - Según mejor resultado.
  - Según costo asociado.
  - Según decisión de comité.

Punto de Corte Transformación a Tabla.

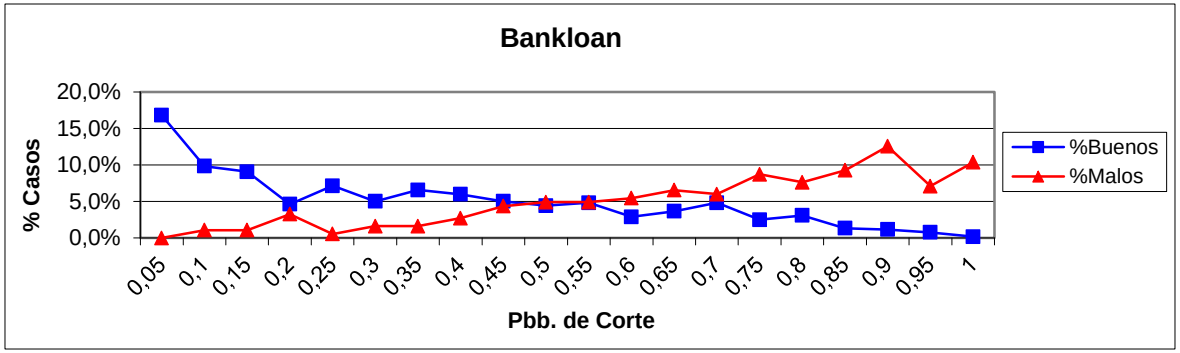
- Tomar salida continua y redondearla a intervalos: {0,05; 0,1; 0,15; …; 0,95; 1}
- Construir tabla de punto de corte:

Probabilidad Corte	%Buenos	%Malos	% Cobertura
0,05	16,8%	0,00%	8,41%
0,1	9,9%	1,09%	5,48%
0,15	9,1%	1,09%	5,09%
0,2	4,6%	3,28%	3,96%
0,25	7,2%	0,55%	3,85%
0,3	5,0%	1,64%	3,33%
0,35	6,6%	1,64%	4,11%
0,4	6,0%	2,73%	4,36%
0,45	5,0%	4,37%	4,70%
0,5	4,4%	4,92%	4,68%
0,55	4,8%	4,92%	4,88%
0,6	2,9%	5,46%	4,18%
0,65	3,7%	6,56%	5,12%
0,7	4,8%	6,01%	5,42%
0,75	2,5%	8,74%	5,63%
0,8	3,1%	7,65%	5,37%
0,85	1,4%	9,29%	5,32%
0,9	1,2%	12,57%	6,86%
0,95	0,8%	7,10%	3,94%
1	0,2%	10,38%	5,29%

Punto de Corte

Transformación a Tabla (2)

- La tabla anterior se puede graficar:



- Cortar en 0,5 corresponde a decir que ambos errores "cuestan" lo mismo.
  - Estadísticamente, a 0,5 se le llama "estimador de máxima verosimilitud".
  - Existen estrategias mejores, ad-hoc al uso que se le de al modelo y el entorno asociado.

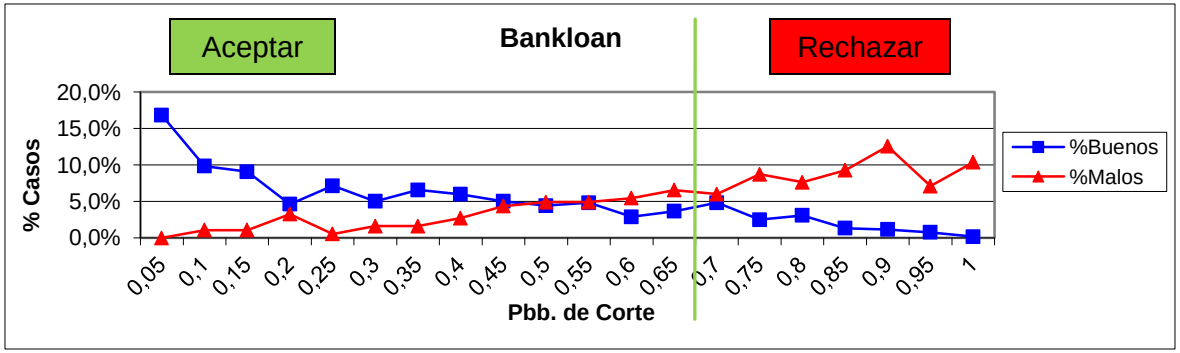
Punto de Corte
Tabla de Efectividad Acumulada

- Para estudiar este efecto se construye la siguiente tabla en base a la anterior:
- Muestra el efecto para cada punto de corte.
- Punto "óptimo" se encuentra en 0,7.

Punto de Corte	% Buenos Aceptados	%Malos Rechazados	% Acierto Global	% Cobertura Total
0,05	16,8%	100,0%	38,6%	8,41%
0,1	26,7%	98,9%	45,6%	13,89%
0,15	35,8%	97,8%	52,0%	18,99%
0,2	40,4%	94,5%	54,6%	22,95%
0,25	47,6%	94,0%	59,7%	26,80%
0,3	52,6%	92,3%	63,0%	30,13%
0,35	59,2%	90,7%	67,4%	34,24%
0,4	65,2%	88,0%	71,1%	38,60%
0,45	70,2%	83,6%	73,7%	43,30%
0,5	74,7%	78,7%	75,7%	47,99%
0,55	79,5%	73,8%	78,0%	52,86%
0,6	82,4%	68,3%	78,7%	57,05%
0,65	86,1%	61,7%	79,7%	62,16%
0,7	90,9%	55,7%	81,7%	67,59%
0,75	93,4%	47,0%	81,3%	73,22%
0,8	96,5%	39,3%	81,6%	78,59%
0,85	97,9%	30,1%	80,1%	83,91%
0,9	99,0%	17,5%	77,7%	90,77%
0,95	99,8%	10,4%	76,4%	94,71%
1	100,0%	0,0%	73,9%	100,00%

Punto de Corte
Corte según punto óptimo.

- La estrategia anterior corresponde a elegir aquel punto con mejor efectividad total.
- Ventajas:
  - Es mejor que utilizar 0,5 pues mejora efectividad global.
  - Es sencillo de encontrar.
- Desventajas:
  - No considera costo del error.
  - Poca flexibilidad.



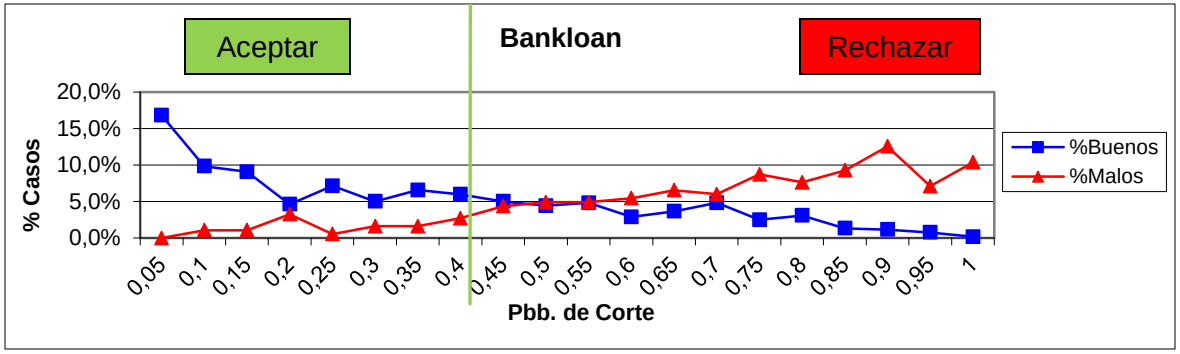
Punto de Corte
Corte según costo de errores.

- Siempre es posible definir un "costo" de errar en un caso particular.
  - Costo aceptar un cliente malo: Saldo promedio perdido por crédito.
  - Costo rechazar un cliente bueno: Utilidad promedio por crédito.
- Así, para cada punto de corte podemos definir:

$$C(corte) = C_{aceptar\ malo} * (\%malos\ acep.) + C_{rechazar\ buenos} * (\%buenos\ rech.)$$

Punto de Corte
Corte según costo de errores (2)

- Ejemplo:
  - Costo rechazar bueno = 1.
  - Costo aceptar malo = 2.
- Corte óptimo en 0,4 o 0,35.



Probabilidad Corte	Costo Buenos	Costo Malos	Costo Total
0,05	0,83	0,00	0,83
0,1	0,73	0,02	0,75
0,15	0,64	0,04	0,69
0,2	0,60	0,11	0,71
0,25	0,52	0,12	0,64
0,3	0,47	0,15	0,63
0,35	0,41	0,19	0,59
0,4	0,35	0,24	0,59
0,45	0,30	0,33	0,63
0,5	0,25	0,43	0,68
0,55	0,21	0,52	0,73
0,6	0,18	0,63	0,81
0,65	0,14	0,77	0,90
0,7	0,09	0,89	0,98
0,75	0,07	1,06	1,13
0,8	0,03	1,21	1,25
0,85	0,02	1,40	1,42
0,9	0,01	1,65	1,66
0,95	0,00	1,79	1,79
1	0,00	2,00	2,00

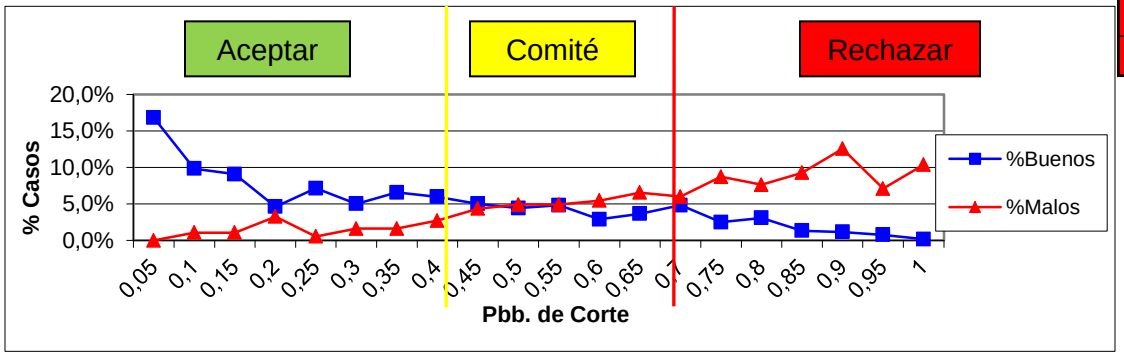
Punto de Corte
Estrategias Mixtas (Comité).

- Mejor idea: 2 puntos de corte!
  - 1º según costo.
  - 2º según efectividad.
- En Zona Intermedia manejar el riesgo por parte del usuario (comité de decisión, más antecedentes, etc.)
  - Maximiza rentabilidad, si decisión de comité es correcta.
  - Entrega máxima flexibilidad, al ser intervalos modificables.

Punto de Corte Estrategias Mixtas (Comité).

- Se acepta con rapidez en un intervalo y se mejora efectividad en todo el resto.
- Si se cambia estrategia, se ajustan cortes.

Probabilidad Corte	% Buenos Aceptados	%Malos Rechazados	% Acierto Global	% Cobertura Total
0,05	16,8%	100,0%	38,6%	8,41%
0,1	26,7%	98,9%	45,6%	13,89%
0,15	35,8%	97,8%	52,0%	18,99%
0,2	40,4%	94,5%	54,6%	22,95%
0,25	47,6%	94,0%	59,7%	26,80%
0,3	52,6%	92,3%	63,0%	30,13%
0,35	59,2%	90,7%	67,4%	34,24%
0,4	65,2%	88,0%	71,1%	38,60%
0,45	70,2%	83,6%	73,7%	43,30%
0,5	74,7%	78,7%	75,7%	47,99%
0,55	79,5%	73,8%	78,0%	52,86%
0,6	82,4%	68,3%	78,7%	57,05%
0,65	86,1%	61,7%	79,7%	62,16%
0,7	90,9%	55,7%	81,7%	67,59%
0,75	93,4%	47,0%	81,3%	73,22%
0,8	96,5%	39,3%	81,6%	78,59%
0,85	97,9%	30,1%	80,1%	83,91%
0,9	99,0%	17,5%	77,7%	90,77%
0,95	99,8%	10,4%	76,4%	94,71%
1	100,0%	0,0%	73,9%	100,00%



Stress Testing

Validación de Modelos

- Una vez implementado, es necesario validar que el funcionamiento del score sea el **deseado**.
- Esto es un requerimiento legal.
  - Se deben entregar resultados periódicos a la SBIF.
- Tres tipos:
  - **Stress Testing**: Probar el modelo en situaciones extremas, pero plausibles.
  - **Seguimiento**: Análisis inicial para revisar si se tienen cambios en la población.
  - **Backtesting**: Realizar seguimiento al modelo para ver su funcionamiento.

Stress Testing

- Corresponde a una medición del **funcionamiento del modelo en casos extremos**.
- Se basa en estimar posibles escenarios de cambio severo y aplicar el modelo.
  - Requerimiento de Basilea II a la PD (NO al scoring).
- Dos tipos:
  - Análisis de Sensibilidad: Apunta a cambios en los modelos por cambios en las variables económicas (como la tasa de interés interbancaria).
  - **Análisis de Escenarios**: Cambios en el modelo por cambios en las variables. Este es el que mayor impacto tiene en credit scoring.

Características del Stress Testing

- Miu y Ozdemir (2009) identifican las siguientes características del Stress Testing bajo Basilea II:
  - Su propósito es **identificar el riesgo que se corre y mitigarlo** frente a cambios en el sistema, o en la composición de créditos.
  - Su cobertura es a nivel de portafolio de créditos (se testea el modelo a nivel global dentro del segmento).
  - La metodología usada es la de **juicio experto**, pues cada escenario a probar depende del scorecard.
  - La periodicidad es tanta como se necesite. Siempre es necesario al **finalizar la construcción** de un sistema.

Construcción de Escenarios

- Se pueden identificar los siguientes pasos.
  1. Identificar los riesgos más importantes.
     - Se deben identificar aquellas variables que marcan el riesgo de la cartera. Por ejemplo, en el caso de modelos a consumidor, considerar edad, mora anterior, etc.
  2. Definir la cobertura.
     - Identificar los productos a los que se aplicará el test. Si existen más de un universo, se deben generar portafolios de créditos que sean representativos.

Construcción de Escenarios (II)

3. Definir el tipo de shock y escenarios.
   - Existen dos formas de medir un shock:
     - Dada la posibilidad que un escenario ocurra, medir el peor resultado posible (Worst Case Scenario).
     - Dado un impacto en la cartera, qué combinación de posibilidades deben darse para llegar a este valor.
   - En modelos de crédito, se argumenta que es necesario realizar el stress por segmentos de riesgo (score) y medir los cambios en ellos.

Realizando Stress Testing

- Una vez listos los escenarios se realiza el stress test.
  - Se considera la cartera de créditos y se modifica la(s) variable(s) acorde al tipo de escenario diseñado.
  - Se mide efecto en el riesgo de crédito global.
- Preguntas a responder:
  - ¿Cómo se mueve la probabilidad de default del portafolio dado el cambio en los parámetros?
  - ¿Afecta el escenario a los índices de riesgo estimados?
  - ¿Cómo cambia la tasa de aceptación de clientes?
  - ¿Cuánto cambia la utilidad dado el escenario?

Simulación de Montecarlo

- Corresponde al método de simulación estática más conocido y utilizado.
  - Usos van desde el riesgo de un proyecto hasta física de partículas.
- Parte de los siguientes supuestos:
  - Se tiene un fenómeno que resulta de una función de variables aleatorias independientes.
  - Si n es grande, se cumple que:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \frac{3\sigma}{\sqrt{n}}\right) \approx 99.8\%$$

  - Lo anterior sigue del teorema central del límite.

Simulación de Montecarlo (II)

- Podemos ahora definir un procedimiento para estimar el valor más probable de esta variable.
- Si $Y = f(X_1, X_2, \dots, X_n)$, con X_i variables aleatorias distribuidas según una ley F_{X_i}, en el rango $[l_i, u_i]$ el valor del fenómeno Y puede ser estimado como:
  1. Generar n vectores $(X_{1i}, X_{2i}, \dots, X_{ni})$ siguiendo la ley de probabilidad F_{X_i}, o en caso de no conocerla, generando números uniformes en cada intervalo $[l_i, u_i]$.
     - Lo anterior tiene sentido si n es grande, pues los números serán finalmente normales.
  2. Calcular $Y_i = f(X_{1i}, X_{2i}, \dots, X_{ni})$.
  3. El estimador más probable para Y será $\overline{Y(n)} = \sum_{i=1}^{n} \frac{Y_i}{n}$

Eligiendo el Número de Experimentos

- ¿Cuán grande debe ser n?
- Se puede determinar a partir de la expresión del error. De los supuestos:

$$\varepsilon = \frac{3\sigma}{\sqrt{n}} \Rightarrow n = \left(\frac{3\sigma}{\varepsilon}\right)^2$$

- Entonces, es posible estimar un límite para este error siguiendo estos pasos:
  1. Calcular $Y_{min} = \min\{f(X_1, \ldots, X_n)\}$ y $Y_{max} = \max\{f(X_1, \ldots, X_n)\}$. Esto se realiza a partir de los rangos de las variables.
  2. Calcular la desviación de estos dos valores y utilizar como sigma.
  3. Usar el promedio de estos dos valores y dividir por la tolerancia deseada.
  4. Usar fórmula para obtener un estimador de n.
  5. Una vez hechos los experimentos, calcular el error real a partir de la fórmula.

Backtesting

Backtesting

- Backtesting es realizar seguimiento, una vez que ya se dispone de los resultados del crédito.
- Es necesario que existan unos cuantos **ciclos de crédito** (apertura/cierre) para realizar el estudio.
  - Necesito tener la salida real de los créditos.
- Mediciones:
  - Acierto del Scorecard
  - Reporte de no pagadores (delinquency report).
  - Reporte de cohortes (vintage).
  - Reporte de migración de defaulters.

Acierto del Scorecard

- Corresponde a medir la cantidad de clientes por cada **rango de score** que **resultaron defaulters**.

Score	Accounts	Active	%	Bad	%	Exp Bad %
0–169	200	198	99%	35	18%	23.0%
170–179	348	300	86%	37	12%	18.0%
180–189	435	367	84%	40	11%	14.0%
190–199	466	387	83%	37	10%	10.0%
200–209	2,456	1,876	76%	126	7%	8.0%
210–219	4,563	3,600	79%	213	6%	5.0%
220–229	5,678	4,325	76%	166	4%	4.0%
230–239	7,658	4,598	60%	106	2%	2.0%
240–249	5,786	3,546	61%	35	1%	0.8%
250+	4,987	2,176	44%	26	1%	0.5%
Total	32,577	21,373	66%	821	4%	

- Esta medición puede estar dividida por el tiempo en libros.

Reporte de No Pagadores

- Se espera medir la cantidad de personas en cada rango de score que efectivamente fueron no pagadores.
- Se divide el proceso por rangos de riesgo.
  - Normal – Mora 123 – Mora 30 – Mora 60 – Vencida – Castigo.

Score	Accounts	Active	%	Current	%	1–29 days	%	30–59	%	60–89	%	90+	%	Writeoff	%	Bankrupt	%
0–169	200	198	99%	127	64%	22	10.9%	15	7.50%	12	6.10%	9	4.50%	6	3.00%	8	4.00%
170–179	348	300	86%	207	69%	36	12.0%	20	6.80%	9	3.00%	9	3.10%	9	3.10%	9	3.00%
180–189	435	367	84%	264	72%	48	13.0%	15	4.00%	12	3.20%	10	2.80%	8	2.20%	10	2.80%
190–199	466	387	83%	286	74%	43	11.0%	21	5.50%	11	2.80%	8	1.94%	10	2.56%	9	2.20%
200–209	2,456	1,876	76%	1,482	79%	225	12.0%	43	2.30%	47	2.48%	18	0.96%	39	2.10%	22	1.16%
210–219	4,563	3,600	79%	2,952	82%	342	9.5%	93	2.58%	83	2.30%	23	0.65%	67	1.87%	40	1.10%
220–229	5,678	4,325	76%	3,676	85%	389	9.0%	93	2.16%	67	1.54%	14	0.32%	51	1.18%	35	0.80%
230–239	7,658	4,598	60%	4,046	88%	359	7.8%	87	1.90%	41	0.90%	18	0.40%	28	0.60%	18	0.40%
240–249	5,786	3,546	61%	3,333	94%	142	4.0%	35	1.00%	18	0.50%	—	0.00%	7	0.20%	11	0.30%
250+	4,987	2,176	44%	2,089	96%	44	2.0%	17	0.80%	9	0.40%	9	0.40%	4	0.20%	4	0.20%
Total	32,577	21,373	66%	18,463	86%	1,648	7.71%	441	2.06%	307	1.44%	118	0.55%	230	1.08%	165	0.77%

Reporte de Cohortes (Vintage)

- Se utiliza para:
  - Identificar **evolución de los no pagadores** a lo largo del tiempo.
  - Identificar **cohortes riesgosas** (periodos del tiempo con alta tasa de defaulters).
- Construcción:
  - Se secciona la cartera por periodo de apertura del crédito.
  - Se calcula la tasa de defaulters para cada periodo.

Reporte de Cohortes (II)

Open Date	*1 Qtr*	*2 Qtr*	*3 Qtr*	*4 Qtr*	*5 Qtr*
Jan–03	0.00%	0.44%	0.87%	1.40%	2.40%
Feb–03	0.00%	0.37%	0.88%	1.70%	2.30%
Mar–03	0.00%	0.42%	0.92%	**1.86%**	**2.80%**
Apr–03	0.00%	**0.65%**	**1.20%**	**1.90%**	
May–03	0.00%	0.10%	0.80%	1.20%	
Jun–03	0.00%	0.14%	0.79%	1.50%	
Jul–03	0.00%	0.23%	0.88%		
Aug–03	0.00%	0.16%	0.73%		
Sep–03	0.00%	0.13%	0.64%		
Oct–03	0.20%	**0.54%**			
Nov–03	0.00%	**0.46%**			
Dec–03	0.00%	0.38%			
Jan–04	0.00%				
Feb–04	0.00%				
Mar–04	0.00%				

Reporte de Migración de No Pagadores

- Los créditos que caen en mora pueden **cambiar de estado**.
  - Ej: Un crédito que en un mes tiene 30 días de mora puede pasar a 0 (cliente pagó) o caer en mora 60.
  - Esto debería estar **indexado por el score**.

			This Month															
	Previous Month		Current		1–29 days		30–59 days		60–89 days		90–119 days		120–179 days		180+		Bankrupt	
Previous Month	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
Current	54,782	72%	52,591	96.0%	2,082	3.8%											210	0.2%
1–29 days	12,640	17%	10,112	80.0%	632	5.0%	1,871	14.8%									85	0.2%
30–59 days	3,254	4%	1,302	40.0%	325	10.0%	472	14.5%	1,139	35.0%							59	0.5%
60–89 days	2,271	3%	182	8.0%	204	9.0%	227	10.0%	227	10.0%	1,413	62.2%					39	0.8%
90–119 days	1,449	2%	55	3.8%	65	4.5%	80	5.5%	87	6.0%	72	5.0%	1,065	73.1%			42	2.1%
120–179	887	1.2%	16	1.8%	12	1.4%	19	2.1%	20	2.2%	27	3.0%	64	4.0%	550	78.0%	150	7.5%
180+	632	0.8%	1	0.2%	8	1.2%	11	1.8%	16	2.5%	13	2.0%	3	3.0%	371	60.3%	210	29.0%
Total	75,915		64,258	85%	3,329	4%	2,679	3.5%	1,488	2.0%	1,524	2.0%	1,132	1.5%	921	1.2%	585	0.8%

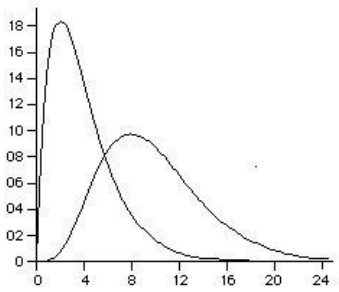
Seguimiento y Concept Drift

Seguimiento

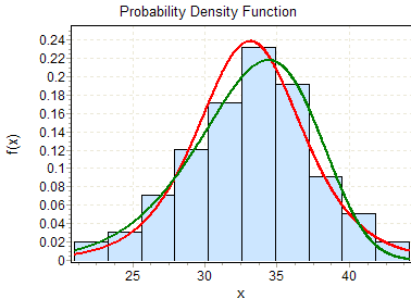
- Apunta a detectar **cambios drásticos en los parámetros o la población** antes que estos ocurran.
- NO son requerimiento legal (aún!), pero permiten adelantarse a los cambios.
- Existe interés actual por ellos, pero **no hay mucho desarrollo**. Dos ejemplos:
  - Test de Estabilidad de Parámetros
  - Test de Estabilidad del Sistema

Test de Estabilidad de Parámetros

- Se considera un dataset futuro.
- ¿Está correcta la distribución de la variable?



Not OK!



OK? Not OK?

- Pequeños cambios son aún buenos para el modelo, no existe un límite definido para que el modelo sea dañado.

Test Empírico de Estabilidad

- Una opción es usar los coeficientes del modelo (Betas) como proxies.
- Procedimiento:
  - Obtener intervalos de confianza del modelo:

$$\hat{\beta} - 2\sigma_{\beta} \leq \hat{\beta} \leq \hat{\beta} + 2\sigma_{\beta} \Rightarrow \widehat{\beta_{min}} \leq \hat{\beta} \leq \widehat{\beta_{max}}$$

  - Calcular desviación máxima: $C \in \left[\frac{\widehat{\beta_{min}}}{\hat{\beta}}, \frac{\widehat{\beta_{max}}}{\hat{\beta}}\right]$
  - Se estiman las medias del dataset original y las del dataset nuevo:

$$\frac{\bar{x}'}{\bar{x}} \in \left[\frac{\widehat{\beta_{min}}}{\hat{\beta}}, \frac{\widehat{\beta_{max}}}{\hat{\beta}}\right]$$

- Test sencillo que permite revisar rápidamente si las poblaciones han cambiado mucho.
  - Problema: No sirve para variables categóricas.

Test Formal de Estabilidad

- Basado en reconstruir los resultados originales.
  1. Estimar salidas desde el score original.
  2. Re-calibrar modelo utilizando las salidas estimadas en el nuevo dataset.
  3. Testear si los parámetros han cambiado mucho. Asintóticamente, los beta son normales.
- Pruebas a realizar:

$$H_0: \beta'_j = \beta_j^{inf} \qquad H_0: \beta'_j = \beta_j^{sup}$$
$$H_a: \beta'_j < \beta_j^{inf} \quad y \quad H_a: \beta'_j > \beta_j^{sup}$$

Test de Estabilidad del Sistema

- Es otra medida empírica para calcular si el score se ha desviado mucho de la media.
- Basada en el cálculo del valor de información de la diferencia.
- Construcción:
  - Considerar rangos de score significativos.
  - Comparar, para la base de datos de origen y la nueva, los porcentajes de casos ingresados.
  - Calcular la medida de información:

$$MI = \sum(\%Actual - \%Esperado) * \ln\left(\%\frac{Actual}{\%Esperado}\right)$$

Test de Estabilidad del Sistema (II)

EXHIBIT 9.1 SYSTEM STABILITY REPORT

Score Range	Actual %	Expected %	(A-E)	A/E	ln(A/E)	Index
0–169	7%	10%	–3%	0.7000	–0.3567	0.0107
170–179	6%	10%	–4%	0.6000	–0.5108	0.0204
180–189	6%	10%	–4%	0.6000	–0.5108	0.0204
190–199	7%	10%	–3%	0.7000	–0.3567	0.0107
200–209	9%	10%	–1%	0.9000	–0.1054	0.0011
210–219	13%	10%	3%	1.3000	0.2624	0.0079
220–229	13%	10%	3%	1.3000	0.2624	0.0079
230–239	11%	10%	1%	1.1000	0.0953	0.0010
240–249	13%	10%	3%	1.3000	0.2624	0.0079
250+	15%	10%	5%	1.5000	0.4055	0.0203
Index						**0.1081**

- Menor a 0.1: No hay problema.
- Entre 0.1 y 0.25: Cambio menor.
- Mayor a 0.25: Cambio peligroso.

Loss Given Default

El Problema de la Pérdida Dado el Default

- Las normas contables nos obligan a provisionar un valor dado por:

$$Prov = PD * LGD * EAD$$

- Donde:
  - PD: Probabilidad de default del segmento de score en el que se encuentra el cliente en el largo plazo.
  - EAD: Exposición (cantidad en deuda) al momento de caer en default.
  - LGD: Porcentaje de pérdida que se da al minuto del default. **Pérdida dado el default**.

Filosofías de LGD y PD

- Basilea II permite interpretar PD y LGD de dos maneras distintas:
  - Asumiendo que el default de un portafolio es una característica **intrínseca de la población**, podemos estimar la PD de largo plazo del portafolio y asumir que las **pérdidas son cíclicas**.
  - La otra opción es inversa, asumir pérdidas constantes según segmento y probabilidades que se mueven según el riesgo.
- Recomendado (Miu y Ozdemir): Primer enfoque.
  - Es más robusto frente a cambios de política y fluctuaciones macroeconómicas.

Consecuencias de la Filosofía de Largo Plazo en PD

- Dada la construcción de scores normal, el utilizar filosofía de largo plazo tiene dos consecuencias en la modelación:
  - Es necesario obtener serie de la PD analizando la **evolución de los segmentos** de riesgo del score.
    - Segunda modelación, encima de la primera.
  - Es necesario modelar la LGD utilizando **técnicas avanzadas**, pues sigue tanto el ciclo como las características del cliente.

Variables para LGD

- Necesitamos **dataset supervisado**: variables descriptoras y porcentaje de pérdida para un grupo de defaulters.

- Además de las variables para score y de comportamiento, es necesario incorporar **variables macroeconómicas**.

Canada General Model

- Real GDP
- Industrial production
- Unemployment
- Corporate profits
- Slope of the yield curve
- High-yield spreads
- Equity indices
- Composite index of leading indicators for Canada
- Consumer credit
- Delinquency rates (DR)
- Short-term interest rate

Real Estate Model

General health of the economy:

- GDP
- Unemployment
- Interest rates
- Retail sales
- Composite Index of Leading Indicators
- Consumer Sentiment Index
- Durable goods
- S&P 500, TSE Composite Index

Specific to the sector:

- S&P/TSX Capped Real Estate Index
- New building permits
- New housing starts (Leading Indicator - Stats Canada)
- Residential real estate (U.S.)
- House price appreciation
- Housing affordability index
- Commercial real estate (U.S.)
    - Vacancy rates
    - Rents per square foot

Fuente: Miu y Odezimir, 2009.

Técnicas de Modelamiento

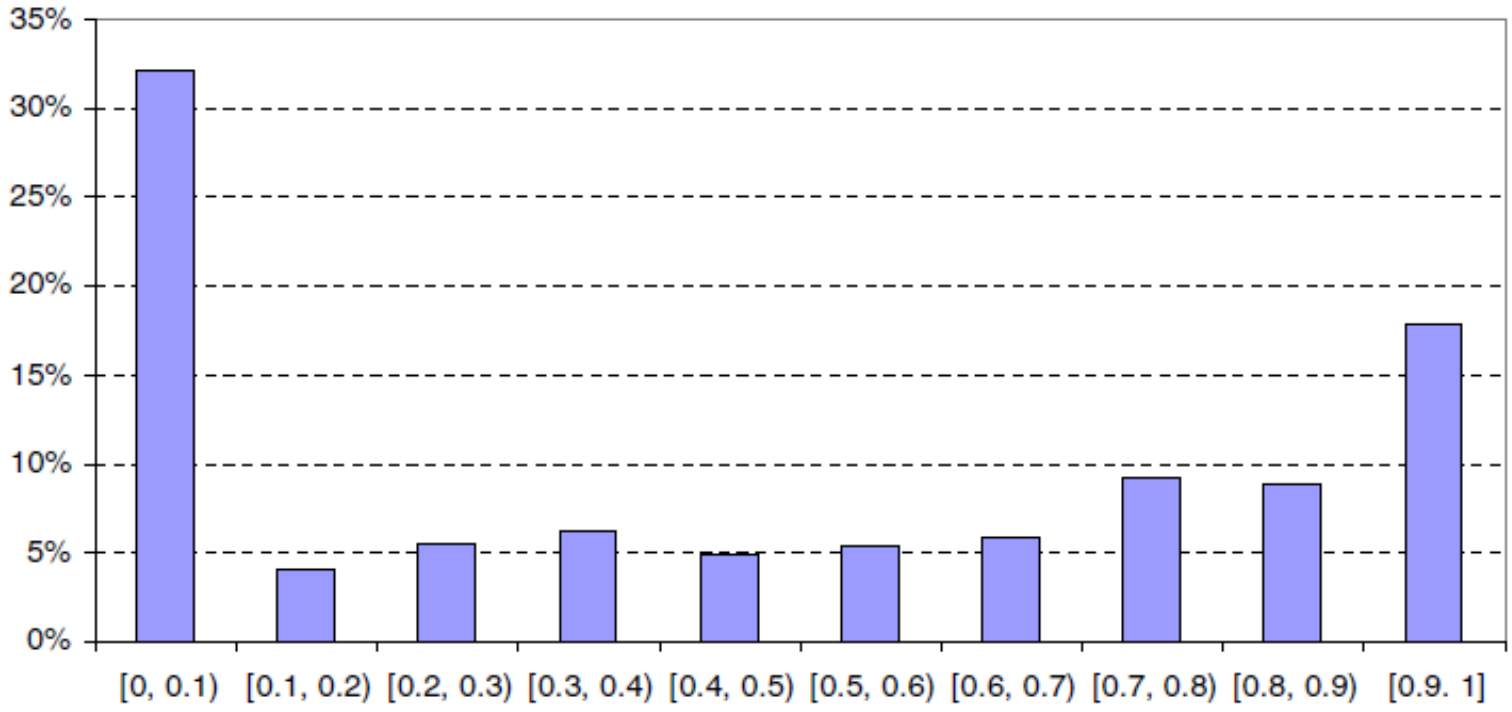
- El modelamiento de LGD está en pleno crecimiento.
  - 40% de las presentaciones en la *credit scoring conference* 2011 tratan de este tema.
- Existen dos tipos de técnicas:
  - No paramétricas: Regresión Logística, Árboles de Regresión y Redes Neuronales.
  - Paramétricos: Regresión de mínimos cuadrados, Gaussiana inversa, Gaussiana inversa con transformación beta.

Resultados en Dataset Real (Qi & Zhao 2011)

- Los autores probaron los distintos métodos llegando a las siguientes conclusiones:
  - Los métodos que mejores resultados obtienen son los no paramétricos
  - El método paramétrico que mejores resultados obtiene es la gaussiana inversa con transformación beta.
- Veremos algunas propiedades y características de los métodos.

Características de la LGD

- La LGD suele tener forma de "U".



Fuente: Qi & Zhao, 2011

Método de Respuesta Fraccionaria: Regresión Logística

- Modelo propuesto para LGD por Papke & Wooldridge en 1996.

$$E[LGD|x] = G(\beta x) = \frac{1}{1 + \exp(-\beta x)}$$

- Otra opción es utilizar modelo log-log

$$E[LGD|x] = \exp(-\exp(-\beta x))$$

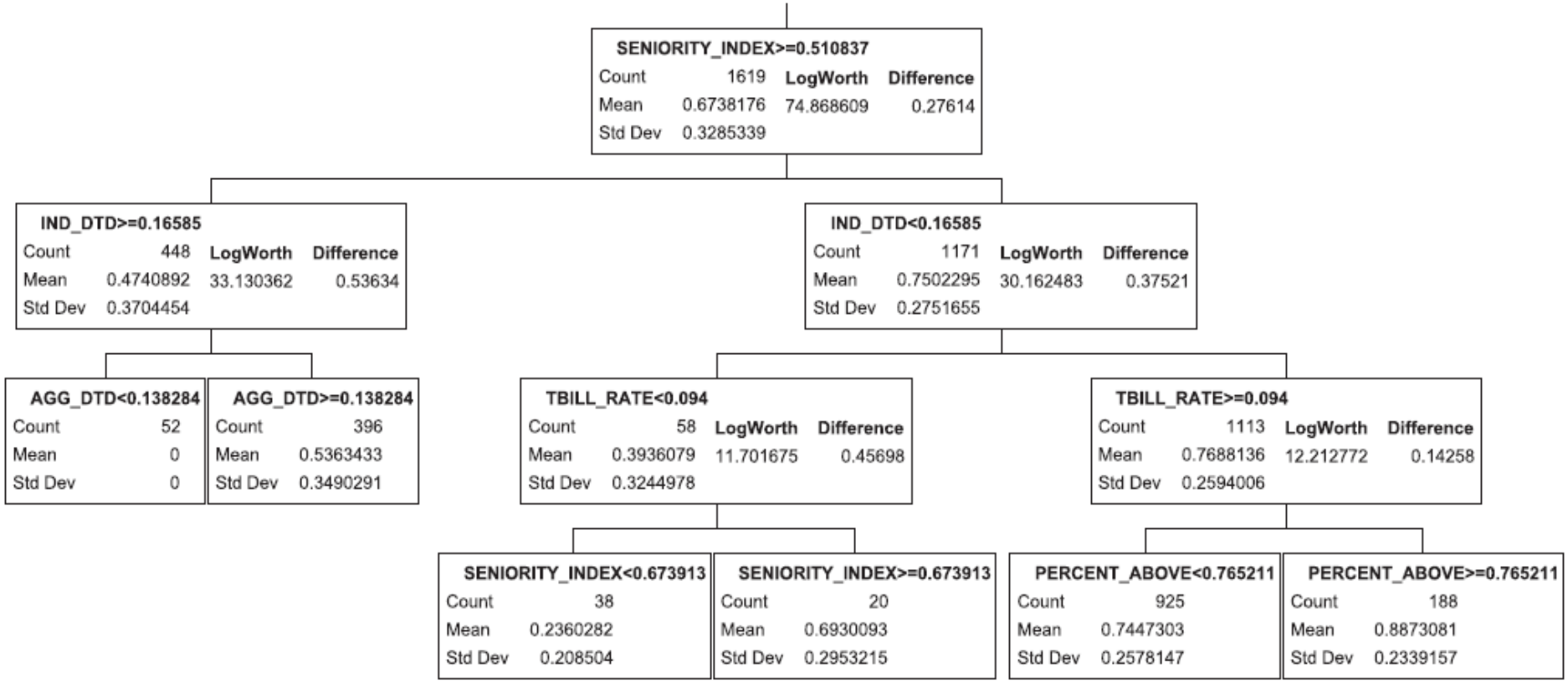
- Se utiliza máxima verosimilitud, por lo que el resultado tiene las mismas propiedades.

$$\sum_i l_i(\hat{\beta}) = \sum_i \left\{ \mathrm{LGD}_i \times \log\left[G(x_i\hat{\beta})\right] + (1 - \mathrm{LGD}_i) \times \log\left[1 - G(x_i\hat{\beta})\right] \right\}$$

Árboles de Regressión

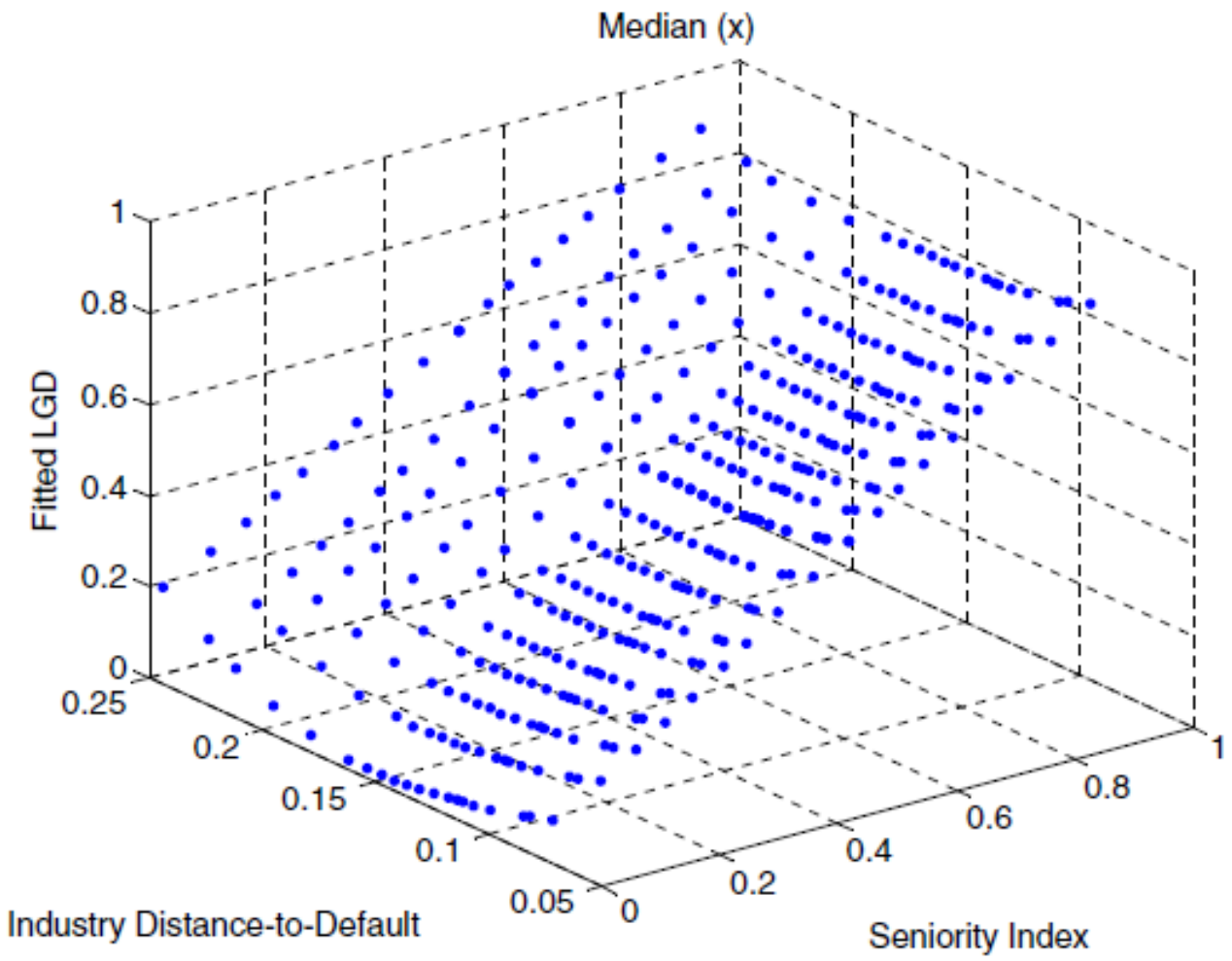
- Un árbol de regresión (Breiman *et al.* 1984) es idéntico a un árbol de clasificación salvo por las siguientes características:
  - La función objetivo es ahora la LGD, una característica continua.
  - En cada paso, el método divide el modelo según la **menor varianza** en la LGD que produce el dividir según una variable particular.
    - Utiliza χ^2 para obtener máximo corte por variable.
  - La LGD que se utiliza al predecir es finalmente el **promedio de las LGD** en la hoja.

Ejemplo



Fuente: Qi & Zhao, 2011

Redes Neuronales

- El método de redes neuronales corre una regresión en los valores de la LGD observados.
- Se regulariza el valor utilizando una **salida sigmoidal**.
- Dado que se está modelando una serie, que posee valores de 0 y 1 específicos, el error más apropiado es el **error cuadrático medio**.

Respuesta a Variables