



Trabajo de Licenciatura en
Administración de Empresas

ANÁLISIS DE RIESGO CREDITICIO.
*MODELO PARA DETERMINAR EL RIESGO DE
INCOBRABILIDAD DE CRÉDITOS PARA
INDIVIDUOS.*

Autor: Agustina Precedo Legajo: 14137

Mentor: Víctor Álvarez

Victoria, 30 de mayo de 2007

A mi familia



Universidad de
San Andrés

Resumen

El objetivo de este estudio es estimar un modelo que ayude a los dadores de crédito a decidir la aceptación o el rechazo de una solicitud de préstamo de individuos en relación de dependencia. Para ello fue necesario hallar las variables significativas para predecir el *default* (entendido como el atraso en los pagos).

A tal efecto se estimaron modelos de tipos lineal, probit y logit sobre dos muestras distintas: una estratificada (mitad cumplidores y mitad incumplidores) y otra de selección aleatoria. Ambas muestras comprenden 1.000 individuos, hombres y mujeres, clientes del banco que proveyó la información, que a diciembre del 2003 se encontraban calificados en situación 1 (sin atrasos) y que a esa fecha estaban en relación de dependencia.

Una vez estimados los modelos, se calcularon indicadores de bondad de ajuste regresándolos sobre las muestras que les dieron origen, obteniendo valores satisfactorios. Luego se verificó la eficiencia predictiva de los modelos regresándolos sobre una tercera muestra de “control”, obteniendo también resultados satisfactorios. A su vez se utilizaron estos indicadores para seleccionar el tipo de modelo con mejor aptitud predictiva. Mediante el test de significatividad se concluyó que las variables explicativas para predecir el default son: sexo, edad, antigüedad en el empleo, ingreso, deuda final, deuda final/ingreso y cantidad de bancos final.

Los modelos obtenidos de la muestra aleatoria resultaron ineficaces porque ésta tenía muy pocos casos de incumplimiento. En cambio resultaron muy eficientes aquellos calculados sobre la muestra estratificada para la cual se forzó la selección a fin de incluir cumplidores e incumplidores por mitades. Con el modelo logit se obtuvo el mayor porcentaje de aciertos, aunque las diferencias con los modelos lineal y probit no fueron significativas.

Con el modelo elaborado en este trabajo, la gerencia de créditos tendrá una herramienta que le permitirá decidir el otorgamiento de créditos fijando los límites aceptables de cometer el error de otorgar crédito a incumplidores o el error de rechazar solicitudes de cumplidores. Utilizando el gráfico de “porcentaje acumulado de errores”, al fijar los porcentajes admisibles de cada tipo de error, automáticamente se determinan los cortes correspondientes en la probabilidad de *default* que definen las tres decisiones de los oficiales de crédito acerca de la solicitud de crédito: aprobarla, elevarla para una decisión en niveles jerárquicos superiores o rechazarla.

Índice Temático	Pág.
1. Introducción	1
2. Teoría de riesgo crediticio	3
2.1. Creciente interés por el riesgo crediticio	3
2.2. Prácticas tradicionales para medir el riesgo crediticio	3
2.3. Un estudio empírico del riesgo de default de los individuos	6
3. Metodología y Muestra	8
3.1. Metodología	8
3.1.1. Tipos de modelo	8
3.1.2. Tests de especificación	10
3.1.3. Coeficientes beta y efectos marginales	13
3.1.4. Bondad de ajuste	15
3.2. Muestra	16
3.2.1. Variables	16
3.2.2. Condiciones de las muestras	18
3.2.3. Datos	19
4. Regresiones y Cuestiones econométricas	21
4.1. Modelo Lineal	21
4.1.1. Estimación	21
4.1.2. Tests de especificación	22
4.1.3. Coeficientes beta	23
4.1.4. Bondad de ajuste	23
4.2. Modelos probit y logit	24
4.2.1. Estimación	25
4.2.2. Tests de especificación	25
4.2.3. Efectos marginales en las medias	26
4.2.4. Bondad de ajuste	27
4.3. Comparación de modelos	28
4.3.1. Modelos aplicados a las muestras originales	28
4.3.2. Modelos aplicados a la muestra aleatoria de control	29
5. Conclusiones	31
6. Bibliografía	36
7. Glosario	37
8. Anexo	38

Índice de Cuadros	Pág.
Cuadro 1: Estadísticas descriptivas	38
Cuadro 2: Matriz de correlación de la muestra estratificada	38
Cuadro 3: Modelo lineal completo	39
Cuadro 4: Modelo lineal restringido	39
Cuadro 5: Probit completo	39
Cuadro 6: Probit restringido	40
Cuadro 7: Efectos marginales probit	40
Cuadro 8: Logit completo	40
Cuadro 9: Logit restringido	41
Cuadro 10: Efectos marginales logit	41
Cuadro 11: Indicadores de bondad de ajuste	28
Cuadros 12-17: PRPC sobre sí, muestra estratificada	41
Cuadro 18-23: PRPC sobre sí, muestra aleatoria	41
Cuadro 24: Indicadores de bondad de ajuste sobre muestra aleatoria de control	29
Cuadros 25-30: PRPC sobre muestra aleatoria de control, muestra estratificada	42
Cuadros 31-36: PRPC sobre muestra aleatoria de control, muestra aleatoria	42
Cuadro 37: Predicciones del modelo logit completo para distintos cortes	43
Cuadro 38: Análisis de aciertos y errores por deciles	44

Índice de Gráficos

Gráfico 1: Predicciones del modelo logit completo para distintos cortes	44
Gráfico 2: Error tipo 1	45
Gráfico 3: Error tipo 2	45
Gráfico 4: Porcentaje acumulado de errores	45

1. Introducción

Los bancos, financieras, cajas de crédito, tarjetas de crédito y cooperativas de crédito necesitan evaluar el riesgo crediticio de sus clientes para mantener o modificar las líneas de crédito otorgadas. Esta evaluación también es necesaria para aceptar o rechazar solicitudes de crédito de nuevos clientes y establecer los límites de créditos a otorgar. El cálculo del riesgo también es necesario para determinar la tasa de interés a cobrar.

Se pueden clasificar los créditos en aquellos otorgados a individuos y los otorgados a empresas. A su vez los primeros pueden clasificarse en prendarios, hipotecarios, descubiertos en cuenta corriente, tarjetas de crédito, etc. En este trabajo se analizarán créditos para individuos menores a \$10.000 sin diferenciar las categorías mencionadas.

El objetivo de este estudio es, en primer lugar, determinar las variables que resultan significativas para predecir el *default* (entendido como el atraso en los pagos) de individuos tomadores de crédito en el sistema financiero. Luego se buscará estimar un algoritmo o modelo que ayude a los dadores de crédito a decidir la aceptación o el rechazo de una solicitud de crédito.

La utilización de modelos de tal tipo permite homogeneizar las decisiones de otorgamiento tomadas en toda la entidad. Esto se logra mediante la utilización de sistemas informáticos a disposición de los oficiales de crédito de las distintas sucursales que utilizan los algoritmos de decisión determinados por la gerencia.

En la siguiente sección se describen las razones del creciente interés por el riesgo crediticio, las prácticas tradicionales para medirlo y un estudio empírico realizado sobre cuentas de tarjetas de créditos para individuos. Posteriormente, en la tercera sección, se detallan los distintos tipos de modelos que se estimarán en el trabajo (lineal, probit y logit), junto con los tests de especificación, la interpretación que se dará a los coeficientes beta y a los efectos marginales y las medidas de bondad de ajuste que se utilizarán. A continuación se describen los datos de las muestras y las variables que se incluirán en las estimaciones por ser significativas para predecir el “*default*”. En la cuarta sección se exponen los resultados del modelo lineal de probabilidad, los problemas que enfrenta este tipo de estimación y dos posibles soluciones, el modelo probit y el modelo logit, con sus resultados. Luego se comparan los resultados de los distintos modelos para estimar qué

modelo predice mejor el *default* de los individuos. Finalmente, en la quinta sección, se presentan algunas conclusiones y sugerencias que surgen del análisis.



Universidad de
San Andrés

2. Teoría de riesgo crediticio

2.1 Creciente interés por el riesgo crediticio

La administración y la medición del riesgo crediticio se hallan en un período de cambios importantes. La nueva generación de profesionales financieros está aplicando ideas y tecnologías innovadoras al diseño de modelos de riesgo crediticio (Saunders, 1999: 1).

¿Por qué ahora? Las siguientes son algunas de las razones que, según Saunders, explican el creciente interés por este tema: a) el aumento permanente de bancarrotas a nivel mundial, posiblemente debido al incremento de la competencia global; b) la expansión de los mercados de capitales; c) el incremento de la competencia por tomadores de créditos de menor solvencia; d) las crisis bancarias en países desarrollados como Suiza y Japón que han demostrado que los valores de los activos reales son difíciles de predecir; e) las ventajas de aplicar la tecnología de la información (TI) para el diseño de modelos predictivos; f) la aplicación de las Normas de Basilea como guía de las políticas de regulación crediticia de los Bancos Centrales.

Según Gross y Souleles (2002: 319) la mayoría de los estudios en esta materia se han referido al riesgo de *default* de las empresas y no a nivel personal. Dado el auge de créditos para individuos, se realizará un análisis del riesgo de *default* personal que ayude a entender este fenómeno.

2.2 Prácticas tradicionales para medir el riesgo crediticio

Como se mencionó anteriormente es poca la literatura dedicada a este tipo de deuda. De todas formas la teoría referida al riesgo crediticio a nivel corporativo puede aplicarse a los individuos. A continuación se expone lo que los autores escriben acerca del riesgo corporativo.

Saunders (1999: 7) plantea tres clases de modelos tradicionales para abordar esta medición: los “*expert systems*”, los “*rating systems*” y los “*credit-scoring systems*”.

En un “*expert system*” una persona que pertenece a la compañía prestamista, basándose en su experiencia y juicio subjetivo, pondera ciertos factores claves para luego decidir acerca del otorgamiento del crédito. Los factores primordiales a analizar son cinco: el carácter del tomador del préstamo, es decir su reputación y su historial de repago; su capital en relación

con el endeudamiento (coeficiente de endeudamiento); la capacidad de repago, determinada por la generación de utilidades para atender el pago de la deuda y sus intereses; las garantías, tanto las constituidas sobre determinados activos a favor de algunos acreedores o garantías de terceros; y el ciclo económico, es decir la etapa en el ciclo del negocio en el que se encuentra la compañía, en especial para industrias sujetas a ciclos.

Uno de los “*rating systems*” más antiguos es el elaborado por la *U.S Office of the Comptroller of the Currency* (OCC). Este sistema es utilizado por los bancos y otros entes reguladores para evaluar la suficiencia de sus reservas para préstamos incobrables. El sistema en cuestión ubica un préstamo en una de las cinco categorías siguientes, a las que les corresponde un porcentaje de reserva específico: “*Other assets especially mentioned*” (0%), “*Substandard assets*” (20%), “*Doubtful assets*” (50%), “*Loss assets*” (100%), “*Pass/performing*” (0%).

La mayoría de los análisis de riesgo crediticio, tanto para consumidores como para préstamos comerciales, se basa en “*credit-scoring systems*”. La idea es identificar previamente algunos factores clave que determinan la probabilidad de *default* y ponderarlos según su importancia. En algunos casos esta ponderación puede interpretarse como la probabilidad de *default*, en otros, como una forma de clasificación: ubica a un potencial tomador de crédito en una clase según su puntaje y los puntos de corte de ésta.

Resulta útil plantear un ejemplo de este tipo de modelos ya que el objetivo primordial de este estudio es hallar un algoritmo que ayude a los dadores de crédito a aceptar o rechazar una solicitud de crédito, dependiendo de la probabilidad de *default* del tomador.

El “*Z-score model*” de Altman (1968) se utiliza para clasificar a los prestatarios. A partir de una muestra que contiene empresas solventes y otras quebradas y utilizando regresiones lineales, el modelo que plantea Altman para clasificar préstamos comerciales, es el siguiente:

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 1.0X_5$$

Donde: X_1 = capital de trabajo/total activo

X_2 = resultados no asignados/total activo

X_3 = resultados antes de intereses e impuestos/total activo

X_4 = valor de mercado de acciones/total pasivo

X_5 = ventas/total activo

Si al ingresar los ratios contables de un prestatario (las X_i), ponderados por los coeficientes de la ecuación, ésta resulta en un valor de Z menor a un valor crítico predeterminado (1.81), el oficial de crédito clasificará al tomador como “malo” y rechazará su solicitud de crédito.

Este modelo tiene dos tipos de limitaciones: limitaciones específicas, que surgen de contener únicamente variables contables, y limitaciones que son comunes a todo modelo lineal cuando se utiliza para predecir probabilidades.

Son varios los problemas que surgen de utilizar un modelo lineal de elección binaria para estimar la probabilidad de un evento.

Un problema es el de heteroscedasticidad. En el modelo lineal la varianza de los errores es una función del vector $(x_1; \dots; x_n)$, en adelante X , y por lo tanto no es constante. Un supuesto que debe cumplirse para la estimación eficiente de mínimos cuadrados ordinarios (MCO) es el de homoscedasticidad, es decir, de varianza constante de los errores. Al no cumplirse este supuesto, la técnica de MCO resulta en estimadores de β 's ineficientes y errores estándar sesgados, lo que afecta los estadísticos empleados en los tests de significatividad.

Otro problema del modelo lineal es que predice valores de z (\hat{z}) negativos o mayores a la unidad. Esto no tiene sentido si se tiene en cuenta que la probabilidad de un evento toma valores en un rango entre cero y uno.

Por último, el hecho de tratarse de una función lineal dificulta la interpretación de los coeficientes beta. En este modelo, un incremento de una unidad en x_j tiene un efecto constante de β_j en la probabilidad de éxito, manteniendo las demás variables constantes. Es decir, el efecto de un cambio en x_j en la probabilidad es el mismo sin importar para qué valor de x_j se mide. En general, en los modelos de probabilidad, los efectos de las variables independientes son decrecientes a medida que la probabilidad predicha se acerca a los extremos, cero o uno. Según Long éste es el problema más grave (Long, 1997: 40).

Una manera de superar estos problemas es utilizar un modelo de tipo probit o logit en lugar del modelo lineal.

Para subsanar ambos tipos de limitaciones, en este estudio se proponen modelos de tipo probit y logit además del modelo de regresión lineal, con variables de diversa índole, y no únicamente contable.

2.3 Un estudio empírico del riesgo de default de los individuos

Gross y Souleles (2002: 319) estiman que entre 1994 y 1997 la cantidad de *defaults* personales en Estados Unidos aumentó un 75%. Esto produjo pérdidas a los prestamistas que equivalían a una importante proporción de los pagos de intereses que obtenían, lo que elevaba el costo promedio del crédito.

Los autores plantean dos explicaciones para estas tendencias. La primera, a la que denominan “*risk effect*”, es el aumento de la participación de prestatarios de menor calidad crediticia en el conjunto total. Los prestatarios de menor calidad crediticia son los principales responsables del incremento en el nivel de *defaults* debido a que han obtenido créditos adicionales en los últimos años en mayor proporción de lo que lo han hecho los buenos pagadores. Muchos analistas hacen alusión al aumento de oferentes de tarjetas de crédito y de los límites de crédito de estas.

La segunda explicación se basa en los costos de *default*, que incluyen los costos sociales, legales y de información. Generalmente el *default* afecta el status social, por sus consecuencias tanto pecuniarias como no-pecuniarias. Muchos analistas argumentan que en la actualidad estos costos sociales han disminuido dado que el *default* se ha vuelto más común. En cuanto a los otros costos, legales y de información, cabe suponer que también han disminuido por el incremento en la cantidad de abogados dedicados a las quiebras y su popularidad. Este “*demand effect*” explica por qué la gente se ha vuelto más indiferente al *default*.

Este último efecto implica un cambio en la relación entre el *default* y las variables que los prestamistas utilizan para predecir el *default*. Más allá de que los prestamistas puedan reducir las pérdidas ocasionadas por el efecto riesgo y el efecto demanda, mejorando la composición del riesgo de sus carteras; para hacer frente a la caída en los costos sociales de *default* precisan cambios más sustanciales en las reglas que utilizan para otorgar préstamos.

Estos autores utilizan datos de cuentas de tarjetas de crédito de un gran número de individuos, correspondientes a distintos oferentes, para analizar el *default* de los individuos.

Un estudio similar es el que se realizará para detectar qué variables son las que mejor predicen el *default* de los individuos que son asistidos crediticiamente por un conjunto de entidades financieras.



Universidad de
San Andrés

3. Metodología y Muestra

3.1 Metodología

3.1.1 Tipos de modelo

En este estudio se estimarán tres tipos de modelo de respuesta binaria: el lineal, el probit y el logit.

Modelo lineal

Este modelo es lineal de regresión múltiple con variable dependiente binaria. La especificación de este tipo de modelo está dada por la siguiente ecuación:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

Donde x_{ji} representa cada una de las variables independientes de la observación i , β_j es el parámetro de la variable j y ε_i es el término de error de la observación i . La variable dependiente y_i es igual a 1 cuando el evento en estudio ocurre e igual a 0 cuando el evento no ocurre.

Cuando y_i es una variable aleatoria binaria, si se asume que la esperanza condicional del término de error en X_i es cero, $E(\varepsilon_i | X_i) = 0$, la esperanza condicional de y_i en X_i es igual a la probabilidad de que el evento ocurra:

$$E(y_i | X_i) = [1 \times \Pr(y_i = 1 | X_i)] + [0 \times \Pr(y_i = 0 | X_i)] = \Pr(y_i = 1 | X_i)$$

De esta forma se puede escribir al modelo lineal como:

$$\Pr(y_i = 1 | X_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_k x_{ki} = \beta X_i$$

Esta fórmula, que se lee “probabilidad de y sub i condicional en x sub i ”, expresa la probabilidad de que la variable dependiente (y) sea igual a 1. Las x_1, x_2, \dots, x_k representan las variables independientes y las $\beta_1, \beta_2, \dots, \beta_k$ son los parámetros que deben estimarse. El modelo se llama lineal porque tanto las variables independientes como los parámetros están elevados a la 1. La expresión βX_i se utiliza para reemplazar la fórmula de la probabilidad. Los valores de $x_{1i}, x_{2i}, \dots, x_{ki}$ son los que corresponden a la observación i de la muestra en estudio.

Modelo probit

Tanto el modelo probit como el logit requieren supuestos acerca de la distribución de los errores.

En el modelo probit se supone que los errores se distribuyen normalmente y se asume que la $E(\varepsilon_i | X_i) = 0$, al igual que en el modelo lineal, y que la $\text{Var}(\varepsilon_i | X_i) = 1$. En este tipo de modelo, la probabilidad de que ocurra el evento, dado el vector X_i , es igual a la función de distribución acumulada de una variable aleatoria normal tipificada $\Phi(z)$, evaluada en βX_i .

Así, se puede representar el modelo probit con la siguiente ecuación:

$$\Pr(y_i = 1 | X_i) = \int_{-\infty}^{\beta X_i} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{t^2}{2}\right] dt = \Phi(\beta X_i)$$

Donde la probabilidad de que ocurra el evento para la observación i , surge de integrar la anterior función (Φ) respecto del valor dado por la suma del producto de cada variable independiente y su respectivo parámetro más el intercepto (β_0). En el modelo probit, $\Phi(\beta X_i)$ toma valores entre cero y uno para cualquier valor de los parámetros y de las variables independientes (Wooldridge, 2006: 622).

Modelo logit

En el modelo logit los errores se distribuyen logísticamente y se asume que la $E(\varepsilon_i | X_i) = 0$, igual que en los modelos anteriores, y que la $\text{Var}(\varepsilon_i | X_i) = \pi^2 / 3$. Por lo tanto la probabilidad de que ocurra el evento, condicional a X_i , es igual a la función de distribución acumulada de una variable aleatoria logística estandarizada $\lambda(t)$, evaluada en βX_i :

$$\Pr(y_i = 1 | X_i) = \frac{\exp(\beta X_i)}{1 + \exp(\beta X_i)} = \lambda(\beta X_i)$$

En este modelo, la probabilidad de que se de el evento en la observación i resulta de evaluar la anterior función (λ) en el valor que surge de sumar las variables independientes multiplicadas por sus parámetros más el intercepto (β_0). En el modelo logit, $\lambda(\beta X_i)$ toma valores entre cero y uno para todos los números reales βX_i (Wooldridge, 2006: 621).

Para estimar el modelo lineal, el modelo probit y el modelo logit se empleará el programa aplicativo Stata 7.

3.1.2 Tests de especificación

Con el objeto de hallar la especificación más precisa para estimar la probabilidad de *default* de los individuos, se realizarán tests de significatividad individual y global sobre los coeficientes de las variables independientes incluidas en cada tipo de modelo (lineal, probit y logit).

Test de significatividad individual

Para evaluar la significatividad individual de los coeficientes se contrasta la hipótesis nula $H_0: \beta_j = 0$ frente a una alternativa de dos colas $H_A: \beta_j \neq 0$, es decir, se prueba si determinada variable tiene o no efecto parcial sobre la variable dependiente. Esto puede hacerse de dos formas.

La primera es ver si el *estadístico t* del coeficiente estimado para la muestra (t_{β_j}) cae o no en la región crítica. Se rechaza H_0 si el valor absoluto del estadístico t es mayor a un valor crítico ($c_{p, gl}$). Este valor crítico surge una vez definido el percentil (p) y los grados de libertad (gl) de la distribución t . Debido a que se trata de un test de dos colas, el percentil será igual a 1 menos el nivel de significatividad (ns) dividido 2. El nivel de significatividad es la probabilidad de error tipo I, es decir, la probabilidad de rechazar H_0 cuando ésta es verdadera; este nivel es elegido por el investigador. Por otra parte, los grados de libertad se calculan restando del número de observaciones en la muestra (n), la cantidad de variables explicativas en el modelo (k) más 1.

Por lo tanto se rechaza H_0 si:

$$|t_{\beta_j}| > c_{p, gl}$$

$$\text{Donde } t_{\beta_j} = \frac{\hat{\beta}_j}{\hat{se}(\beta_j)} \quad ; \quad p = 1 - (ns/2) \quad ; \quad gl = n - k - 1$$

Siendo $\hat{\beta}_j$ el coeficiente estimado para la variable independiente j y $\hat{se}(\beta_j)$ la raíz cuadrada del estimador de la varianza de β_j (Sosa Escudero, 1999) o lo que es lo mismo, el error estándar del coeficiente estimado.

Stata 7 calcula para cada coeficiente en la regresión su error estándar y, en los casos de regresiones lineales, el *estadístico t*, y para regresiones de tipo probit y logit, el *estadístico z*, que surge de la distribución normal estándar. De todas formas el estadístico *z* se comporta asintóticamente, es decir para niveles altos de *n*, como una distribución *t*, por lo que el resultado del test no varía si se utiliza una distribución *t* en lugar de una distribución normal. La ecuación del estadístico z_{β_k} es la misma que la del estadístico t_{β_k} para muestras grandes, que es el caso en estudio.

Otra forma de realizar un test de significatividad individual de los coeficientes es comparar el *p-valor* del estadístico *t*, calculado para el coeficiente, con el nivel de significatividad elegido. En lugar de realizar el método de la región crítica para distintos niveles de significatividad, “es más informativo contestar a la siguiente pregunta: dado el valor observado del estadístico *t*, ¿cuál es el nivel de significatividad más pequeño al que se rechazaría la hipótesis nula?” (Wooldridge, 2006: 142). A este nivel se lo conoce como el *p-valor*. Se obtiene calculando la probabilidad de que una variable aleatoria con distribución *t* (*T*) sea mayor que el valor absoluto del estadístico *t* estimado para el parámetro.

Con este método, se rechaza H_0 si:

$$p\text{-valor}_{t_{\beta_k}} < ns$$

$$\text{Donde } p\text{-valor}_{t_{\beta_k}} = \Pr(|T| > |t_{\beta_j}|)$$

El programa Stata 7 provee el estadístico *t* estimado para cada coeficiente y su *p-valor*. En este análisis se utilizará el método del *p-valor* en lugar del método de la región crítica ya que de esta manera la comparación es automática, dado que el nivel de significatividad (*ns*) lo define uno mismo, mientras que el valor crítico debe buscarse en la tabla de la distribución *t* de student o en la tabla de la distribución normal estándar, según el caso.

Test de significatividad global

Para evaluar la significatividad global de los coeficientes se contrasta la hipótesis nula $H_0: \beta_1 = \beta_2 = \beta_3 \dots = 0$ frente a una alternativa en la que al menos uno de los β_j es distinto de cero, de esta manera se prueba si es cierto que ninguna de las variables explicativas incluidas en el modelo tiene efecto sobre la variable dependiente.

Debido a que el programa Stata 7 calcula el estadístico F con su p-valor para las regresiones lineales y el estadístico del cociente de verosimilitud con su p-valor para las regresiones de tipo probit y logit, se emplearán distintos estadísticos para evaluar la significatividad global de los coeficientes en los distintos tipos de modelos.

El *estadístico F*, que se usa para contrastar la hipótesis nula en las regresiones lineales, se define como:

$$F \equiv \frac{R^2 / k}{(1 - R^2) / (n - k - 1)}$$

Donde R^2 es el coeficiente de determinación de la regresión de y_i sobre X_i , k y $(n - k - 1)$ son los grados de libertad.

Igual que para el test de significatividad individual de los coeficientes, existen dos formas de llevar a cabo un test de significatividad global: el método de la región crítica y el método del p-valor.

Para poder realizar el primer método es necesario obtener un valor crítico ($c_{p, k, n-k-1}$) contra el cual comparar el estadístico F estimado. Este valor crítico surge de una distribución $F_{k, n-k-1}$ y depende del percentil, que ahora se define como 1 menos el nivel de significatividad, y de los grados de libertad en el numerador y denominador de la ecuación del estadístico F.

Por lo tanto se rechaza H_0 si:

$$F > c_{p, k, n-k-1}$$

Donde $p = 1 - \alpha$

El *estadístico del cociente de verosimilitud*, que se utiliza para el test de significatividad global en las regresiones de tipo probit y logit, se define como el doble de la diferencia de la función logarítmica de verosimilitud del modelo no restringido (ℓ_{nr}) y la función logarítmica de verosimilitud del modelo restringido (ℓ_r). El modelo restringido es el modelo que incluye únicamente el intercepto, ya que según la H_0 , todos los parámetros de las variables independientes son iguales a cero. Entonces se puede representar este estadístico de la siguiente manera:

$$LR = 2(\ell_{nr} - \ell_r)$$

Donde el modelo no restringido:

$$\Pr(y_i = 1 | X_i) = \Phi (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_k X_{ik}) \text{ para el probit y}$$

$$\Pr(y_i = 1 | X_i) = \lambda (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_k X_{ik}) \text{ para el logit}$$

Y el modelo restringido:

$$\Pr(y_i = 1 | X_i) = \Phi (\beta_0) \text{ para el probit y}$$

$$\Pr(y_i = 1 | X_i) = \lambda (\beta_0) \text{ para el logit}$$

Es menester encontrar un valor crítico contra el cual comparar este estadístico para poder llevar a cabo el método de la región crítica. Bajo H_0 este estadístico tiene una distribución chi-cuadrado. El valor crítico depende del percentil $(1 - \alpha)$ y del número de restricciones que se imponen al modelo original. Tratándose de un test de significatividad global, el número de restricciones es igual al número de variables independientes en el modelo estimado (k). Así se rechaza H_0 si:

$$LR > c_{p, k}$$

Por último, se puede realizar el test de significatividad global sobre los coeficientes comparando el *p-valor* del estadístico F , en el caso de regresiones lineales, o del estadístico del cociente de verosimilitud, en los casos de regresiones probit y logit, con el nivel de significatividad elegido.

Con este método, se rechaza H_0 si:

$$p\text{-valor}_F < \alpha \text{ para las regresiones lineales}$$

$$p\text{-valor}_{LR} < \alpha \text{ para las regresiones probit y logit}$$

3.1.3 Coeficientes beta y efectos marginales

Modelo lineal

En el modelo lineal la probabilidad de éxito, es decir, de que el evento ocurra, es lineal en los parámetros. Como se explicó anteriormente (ver página 8), la ecuación del modelo lineal es la siguiente:

$$\Pr(y_i = 1 | X_i) = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_k X_{ik} = \beta X_i$$

Por consiguiente, β_j mide el cambio en la probabilidad de éxito cuando x_j cambia en una unidad, ceteris paribus:

$$\Delta \Pr(y_i = 1 | X_i) = \beta_j \Delta x_j$$

De esta manera, al estimar la ecuación del modelo lineal con la mecánica de mínimos cuadrados ordinarios (MCO) se obtiene la siguiente función:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki}$$

Donde \hat{y}_i es la probabilidad predicha de éxito, $\hat{\beta}_0$ es la probabilidad predicha de éxito cuando cada una de las variables independientes es igual a cero y $\hat{\beta}_j$ mide el cambio predicho en la probabilidad de éxito frente a un incremento unitario de x_j .

Modelo probit y logit

Para el modelo probit, al igual que para el logit, se utiliza la técnica de máxima verosimilitud, en lugar de MCO, para estimar la siguiente ecuación:

$$\Pr(y_i = 1 | X_i) = F(\beta X_i)$$

Donde F es la función de distribución acumulada para la distribución normal estándar (ϕ) en el modelo probit, y la función de distribución acumulada para la distribución logística (λ) en el modelo logit.

El *efecto marginal* de la variable x_j es la pendiente de la curva de probabilidad en un gráfico de $\Pr(y=1 | X)$ contra x_j , manteniendo todas las demás variables constantes. Surge de calcular la derivada parcial de la ecuación anterior con respecto a x_j :

$$\frac{\partial \Pr(y = 1 | X)}{\partial x_j} = \frac{\partial F(\beta X)}{\partial x_j} = \frac{dF(\beta X)}{d\beta X} \frac{\partial \beta X}{\partial x_j} = f(\beta X) \beta_j$$

El signo del efecto marginal viene dado por el signo de β_j , puesto que $f(\beta X)$, la derivada de $F(\beta X)$ respecto de βX , es siempre positiva. La magnitud del efecto marginal depende del valor de β_j y de βX , por lo tanto depende de los valores de las variables independientes que se usen para calcularlo.

El programa Stata 7 calcula los *efectos marginales en las medias*, es decir, cómo cambia la probabilidad predicha cuando cambia en una unidad la variable x_j , fijando todas las demás variables independientes en sus valores medios.

Si x_j es binaria, como el sexo en los modelos estimados, entonces el efecto marginal es la diferencia entre la probabilidad predicha cuando x_j toma el valor de 1 y la probabilidad predicha cuando x_j vale 0, fijando las demás variables en sus medias.

3.1.4 Bondad de ajuste

R^2 y *pseudo- R^2*

En el modelo lineal, como en todos los modelos de regresiones lineales, la medida más conocida de bondad de ajuste es el R^2 , el coeficiente de determinación, que determina la proporción de la variación de y que puede ser explicada por las variables independientes incluidas en el modelo. El R^2 aumenta cuando aumenta la cantidad de variables incluidas en el modelo, por lo que resulta más preciso utilizar el R^2 -ajustado que ajusta por la cantidad de variables incluidas.

Para los modelos probit y logit existen varias medidas de bondad de ajuste a las que se las denomina *pseudo- R^2* por su similitud al R^2 de las regresiones lineales. La que se utilizará en esta investigación es la medida de McFadden (1974):

$$\text{pseudo-}R^2_{MF} = 1 - (\ell_{nr} / \ell_r), \quad \text{donde } |\ell_{nr}| \leq |\ell_r|$$

Como en la fórmula del estadístico del cociente de verosimilitud, ℓ_{nr} es la función logarítmica de verosimilitud del modelo no restringido y ℓ_r es la función logarítmica de verosimilitud del modelo restringido que incluye únicamente el intercepto (α).

Si las variables independientes no tienen poder explicativo entonces $\ell_{nr} = \ell_r$, $\ell_{nr}/\ell_r = 1$, y el *pseudo- R^2_{MF}* es cero. Lo que significa que la ganancia por incorporar variables explicativas en el modelo, en términos de la bondad de ajuste, es baja.

Por otro lado, cuando $|\ell_{nr}| < |\ell_r|$ el *pseudo- R^2_{MF}* es positivo y el modelo que incorpora variables explicativas resulta superior al modelo con sólo una constante (α).

Porcentaje de respuestas predichas correctamente

Otra medida de bondad de ajuste es el porcentaje de respuestas predichas correctamente (PRPC), que puede utilizarse en cualquier tipo de estimación. Lo que se hace es calcular para cada observación i la probabilidad estimada de que y_i sea igual a 1 $P \hat{f}(y_i=1 | X_i)$. Luego se fija un valor de corte (γ_i), generalmente 0.5, de tal manera que si $P \hat{f}(y_i=1 | X_i)$ es superior a este valor de corte, la predicción de y_i (\hat{y}_i) es 1 y si es inferior, \hat{y}_i es 0. Finalmente se calcula el porcentaje de observaciones en las que la predicción de y_i , \hat{y}_i , coincide con el valor real de y_i . Representado en ecuaciones:

$$\hat{y}_i = \begin{cases} 0 & \text{si } P \hat{f}(y_i=1 | X_i) \leq \gamma_i \\ 1 & \text{si } P \hat{f}(y_i=1 | X_i) > \gamma_i \end{cases}$$

$$\text{PRPC} = \frac{\text{RPC}}{n}$$

Donde RPC es el número de observaciones en las que $\hat{y}_i=1$ cuando $y_i=1$ y en las que $\hat{y}_i=0$ cuando $y_i=0$.

3.2 Muestra

3.2.1 Variables

Las variables que se incluirán en las estimaciones y sus fuentes son las siguientes:

Variable dependiente:

“Default”: El Banco Central de la República Argentina (BCRA) obliga a los bancos a informar, todos los días 20 de cada mes respecto del cierre del mes anterior, el monto de las deudas de cada uno de sus deudores y su calificación. Las calificaciones tienen un rango del 1 al 6. La situación 1, llamada “normal”, comprende los créditos sin atrasos o con atrasos menores a 31 días. La situación 2, “con riesgo potencial”, comprende aquellos créditos con atrasos entre 32 y 90 días. La situación 3, “con problemas”, comprende los que tienen atrasos entre 91 y 180 días. La situación 4, “con alto riesgo de insolvencia”, comprende los que tienen atrasos entre 181 y 365 días. La situación 5, “irrecuperable”, comprende los que tienen atrasos mayores a 365 días. La situación 6, “irrecuperable por disposición técnica”, comprende a los deudores irrecuperables de entidades financieras liquidadas.

A los fines de este estudio, se define que está en situación de “**default**” aquel deudor que recibe una calificación de 3, 4, 5 o 6. Se considera que el momento del default es cuando dicho deudor recibe por primera vez alguna de las mencionadas calificaciones. Aunque lo habitual es que a medida que empeora el cumplimiento de un deudor, éste vaya pasando gradualmente de la situación 1 a la 5, ocurre a veces que alguien pase de la 1 directamente a la 4 o 5 debido al atraso de los bancos en reflejar los cambios de situación.

Esta variable, “default”, es binaria, tomando valor 1 si el deudor cayó en default (situaciones 3, 4, 5 o 6) y 0 si se mantuvo en situaciones 1 o 2. Este análisis consiste en hallar un modelo para predecir el default, por lo tanto ésta es la variable dependiente.

Variables independientes:

a) “**Sexo**”: ésta es una variable binaria que adquiere valor 1 cuando se trata de un hombre y 0 en caso de una mujer. En ambas muestras, la selección de los individuos es aleatoria respecto de esta variable. Las fuentes de esta información son las solicitudes de crédito presentadas al banco por los individuos al inicio de su relación crediticia.

b) “**Edad**”: ésta es una variable discreta expresada en años cuyas fuentes son las mismas solicitudes de crédito.

c) “**Antigüedad en el empleo**”: ésta es una variable discreta que expresa la antigüedad en el empleo al momento de ocurrir el default (primera situación 3, 4, 5 o 6) para aquellos que sufrieron default en el período analizado, y para aquellos que no incurrieron en *default*, su antigüedad en el empleo a diciembre de 2006. Esta variable toma los siguientes valores: 0 si el individuo se encuentra desempleado, 1 si tiene una antigüedad de 1 a 12 meses, 2 si tiene una antigüedad de 13 a 24 meses, y así sucesivamente. La fuente es la ANSES (Administración Nacional de la Seguridad Social).

d) “**Ingreso**”: es el salario bruto del prestatario expresado en miles de pesos. Esta información proviene de las solicitudes de crédito presentadas al inicio de la relación crediticia de cada individuo con el banco que proveyó la información. Si bien el ingreso pudo haber variado entre dicho momento y el del incumplimiento, se lo empleará de todos modos, ya que normalmente el salario de una persona no varía significativamente.

e) “**Deuda final**”: es el monto total del endeudamiento de cada individuo en el conjunto de bancos que lo asisten crediticiamente expresado en miles de pesos. Si la persona ha

incumplido, se toma la deuda al momento del incumplimiento, de lo contrario, se toma la deuda a diciembre de 2006. La fuente de esta información es el BCRA.

f) **“Deuda final/Ingreso”**: esta variable es el resultado del cociente de dos variables, deuda e ingreso.

g) **“Deuda final/Deuda inicial”**: es el cociente entre **Deuda Final** y **Deuda Inicial**. Se entiende como Deuda Final la definida en e). Deuda Inicial es la que tenían en el conjunto de bancos en diciembre de 2003. Su fuente es el BCRA.

h) **“Deuda final - Deuda inicial”**: es la resta de los conceptos anteriores.

i) **“(Deuda final -Deuda inicial)/Ingreso”**: es la definida en h) dividido el ingreso.

j) **“Cantidad de Bancos final”**: es el total de entidades bancarias que lo asistían al momento del incumplimiento o a diciembre de 2006. Su fuente es el BCRA.

k) **“Cantidad de Bancos final/Cantidad de Bancos inicial”**: es el cociente entre la cantidad de bancos final (concepto anterior) y la cantidad de bancos inicial. La cantidad inicial de bancos es la cantidad de bancos que lo asistían en diciembre de 2003. Su fuente es el BCRA.

l) **“Cantidad Bancos final - Cantidad Bancos inicial”**: es la resta de los conceptos anteriores.

3.2.2 Condiciones de las muestras

La estimación de máxima verosimilitud (en inglés: *maximum likelihood estimation*) consiste en encontrar los parámetros que maximizan la función logarítmica de verosimilitud (Long, 1997: 53):

$$\ln L(\beta|y,X) = \sum_{y=1} \ln F(\beta X) + \sum_{y=0} \ln [1 - F(\beta X)]$$

Esto se hace probando de forma iterativa distintos parámetros hasta encontrar los parámetros que maximicen esta función; es decir que igualen a cero la pendiente de esta función. Cuando se encuentran los parámetros buscados se dice que hay convergencia. Para obtener los estimadores de máxima verosimilitud, técnica que se utiliza en las regresiones no lineales, es necesario que haya convergencia.

La convergencia ocurre más rápidamente cuanto mayor sea el número de observaciones en la muestra y cuanto mayor sea la razón del número de observaciones por el número de variables independientes. Además, según Long (1997: 60), la convergencia puede dificultarse si una de las categorías de una variable categórica tiene pocos casos.

En la muestra inicial, tomada al azar, los individuos con incumplimiento representaron el 10.5% del total. Esto enmarca dicha muestra en el problema planteado por Long (1997: 59): pocos casos con valor 1 para la variable dependiente, que en este estudio es categórica. Para subsanar este problema y lograr una rápida y eficiente convergencia se utilizará para este estudio una muestra de 1.000 individuos en la que la mitad de los individuos son casos de default ($y_i=1$) y la otra mitad son casos de buen cumplimiento ($y_i=0$). A este tipo de muestra se la denomina “estratificada”.

El muestreo estratificado puede ocasionar sesgos en los estimadores. Para comparar la eficiencia de las dos metodologías (análisis con muestra estratificada y con muestra sin estratificar), se construirá un modelo a partir de una muestra de 1.000 individuos de selección aleatoria y otro a partir de una muestra de la misma cantidad con selección estratificada (50% cumplidores - 50% incumplidores). Luego, se aplicarán ambos modelos sobre una tercera muestra aleatoria también de 1.000 individuos. Finalmente, se compararán los distintos indicadores de bondad de ajuste de ambos modelos para determinar cuál resulta más eficiente.

3.2.3 Datos

Las variables descritas anteriormente caracterizan a individuos que pertenecen a una muestra de clientes facilitada por un banco. Dicho banco pidió que su nombre se mantenga en confidencialidad y suministró la base con los nombres y números de documento reemplazados por un número de serie.

El banco confeccionó la muestra estratificada tomando del total de sus deudores 500 individuos al azar del subconjunto de incumplidores y 500 también al azar del subconjunto de cumplidores. La forma de lograr que la selección sea al azar fue elegir a los que tenían el mismo valor en el ante penúltimo dígito del número de su documento de identidad.

Para las otras dos muestras aleatorias aplicó la misma metodología pero sin tomar en cuenta si eran cumplidores o incumplidores.

Como se expuso anteriormente, los modelos se estimarán sobre dos muestras distintas: una estratificada y otra aleatoria. Ambas muestras comprenden 1.000 individuos, hombres y mujeres, clientes del banco que proveyó la información, que a diciembre del 2003 se encontraban calificados en situación 1, con deudas inferiores a \$10.000 y que a esa fecha estaban en relación de dependencia.

La muestra estratificada, está compuesta en un 50% de individuos que habían incurrido en incumplimiento en el período 12/2003 – 12/2006 y el restante 50% que se mantuvo en situación 1 o 2 en dicho lapso. La muestra aleatoria resultó tener un 10.5% de incumplidores y un 89.5% de buenos pagadores. En el **Cuadro 1** en el Anexo se muestran las estadísticas descriptivas de cada una de estas dos muestras. Por último, la tercera muestra aleatoria, que será utilizada como control, resultó tener un 8.8% de incumplidores.



Universidad de
San Andrés

4. Regresiones y Cuestiones econométricas

En esta sección se analizarán, en primer lugar, los resultados obtenidos en las regresiones lineales de probabilidad y a continuación los obtenidos en las estimaciones de los modelos probit y logit (para la muestra estratificada y la muestra aleatoria). Por último se compararán los distintos indicadores de precisión de los diversos modelos con el fin de concluir cuál es el que tiene más aptitud para utilizarse para decidir la aceptación o el rechazo de una solicitud de crédito. La comparación de los índices de bondad de ajuste se hará sobre los resultados que provienen de regresar los modelos sobre las muestras originales y luego sobre los resultados que surgen de regresarlos sobre una nueva muestra aleatoria.

4.1 Modelo Lineal

Como puede observarse en el **Cuadro 2**, existe una alta correlación entre algunas variables independientes, lo que puede originar problemas en las estimaciones. Para superar estos problemas, en este estudio se trabajará eliminando variables que resulten altamente correlacionadas con otras (correlación mayor a 0.5), a saber: deuda final dividido deuda inicial, deuda final menos deuda inicial, deuda final menos deuda inicial dividido ingreso, cantidad de bancos final dividido cantidad de bancos inicial y cantidad de bancos final menos cantidad de bancos inicial. Aunque eran esperables las mencionadas correlaciones porque algunas de ellas son funciones de las mismas variables, se las incluyó en el modelo inicial a fin de elegir de entre los pares con correlación mayor a 0.5 las de p-valor menor a 0.10 (como se analizará en la sección 4.1.2).

4.1.1 *Estimación*

En una primera estimación (**Cuadro 3**) se regresó la variable dependiente, “default”, contra todas las variables independientes que resultaron no altamente correlacionadas: sexo, edad, antigüedad en el empleo, ingreso, deuda final, deuda final dividido ingreso y cantidad de bancos final. La ecuación resultante para la *muestra estratificada* es la siguiente:

$$Pr(Default_i | X_i) = 0.7042 + 0.1825Sexo_i - 0.0073Edad_i - 0.0044Antemp_i - 0.1004Ingreso_i - 0.0021DF_i + 0.000007DF/Ing_i + 0.1019CBF_i$$

Para la *muestra aleatoria* la ecuación es:

$$Pr(Default_i | X_i) = 0.1895 + 0.0129Sexo_i - 0.0021Edad_i - 0.0024Antemp_i - 0.0258Ingreso_i - 0.0018DF_i - 0.000088DF/Ing_i + 0.0524CBF_i$$

4.1.2 Tests de especificación

Como se dijo al explicar la metodología, resulta más simple la técnica del p-valor que la técnica de la región crítica, por lo que se hará referencia a los p-valores y se ignorarán los estadísticos t y F .

Con respecto al *test de significatividad global*, puede observarse que las variables incluidas en los modelos anteriores son conjuntamente significativas. Esto se debe a que el p-valor del estadístico F ($Prob>F=0.000$ al pie del **Cuadro 3**) es menor que cualquier nivel de significatividad que se elija. Por lo tanto puede rechazarse la hipótesis nula, de que ninguna de las variables explicativas tiene efecto sobre la variable dependiente, con cualquier nivel de significatividad, es decir sin incurrir en error tipo I.

A pesar de ser conjuntamente significativas, algunas variables independientes no son individualmente significativas. Cuando se realiza el *test de significatividad individual* sobre cada variable, no siempre puede rechazarse la hipótesis nula $H_0: \beta_j = 0$ con un nivel de significatividad del 10%. Aquellas variables para las que no puede rechazarse la hipótesis nula, y por lo tanto resultan no significativas a nivel individual, son las que tienen un p-valor ($P>|t|$ en la quinta y novena columna del **Cuadro 3**) mayor al nivel de significatividad elegido, en este estudio: 0.10.

De esta manera, en la *muestra estratificada* el cociente entre la deuda final y el ingreso no es significativo. Al regresar la variable dependiente contra las variables que sí resultan individualmente significativas (**Cuadro 4**) se obtiene el siguiente modelo “restringido” (en contraposición al denominado “completo”, en el que se incluyen todas las variables independientes):

$$Pr(Default_i | X_i) = 0.7101 + 0.1841Sexo_i - 0.0073Edad_i - 0.0045Antemp_i - 0.1020Ingreso_i - 0.0021DF_i + 0.1016CBF_i$$

Por otro lado, en la *muestra aleatoria* no superan el test de significatividad individual el sexo y el cociente entre la deuda final y el ingreso. Teniendo en cuenta únicamente a las variables individualmente significativas, el modelo restringido para la muestra aleatoria es:

$$Pr(\text{Default}_i | X_i) = 0.1955 - 0.0021 \text{ Edad}_i - 0.0024 \text{ Antemp}_i - 0.0248 \text{ Ingreso}_i - 0.0018 \text{ DF}_i + 0.0523 \text{ CBF}_i$$

4.1.3 Coeficientes beta

En 3.1.3 se dijo que en el modelo lineal el coeficiente β_j mide el cambio en la probabilidad predicha de default cuando x_j varía en una unidad. Dado que este modelo es lineal en los parámetros, la interpretación de los coeficientes beta es directa y vale para cualquier nivel de las variables. A continuación se analizará el impacto que tienen las variables individualmente significativas al 10% sobre la probabilidad de incumplir la cancelación de un crédito, cuando se mantienen fijas las demás variables.

En el modelo estimado a partir de la *muestra estratificada*, la probabilidad de incumplir de los hombres, manteniendo fijos todos los demás factores, es 0.1841 más alta que la de las mujeres. Un aumento de un año en la edad del individuo disminuye su probabilidad de incumplir en 0.0073. A medida que aumenta la antigüedad en el empleo del prestatario su probabilidad de incumplir disminuye en 0.0045 puntos por año. Por cada \$1.000 que aumenta el ingreso del individuo su probabilidad de default disminuye en 0.1020; y el mismo aumento en la deuda final con los bancos disminuye dicha probabilidad en 0.0021. Esto es así porque tanto el ingreso como la deuda del individuo están expresados en miles de pesos. Un aumento unitario en la cantidad de bancos que lo asisten crediticiamente aumenta la probabilidad de incumplir del individuo en 0.1016 puntos.

Con la excepción del sexo, que resulta no significativo, en el modelo lineal restringido para la muestra aleatoria los coeficientes beta tienen el mismo signo que los coeficientes del mismo modelo para la muestra estratificada. Sin embargo, el impacto que tiene un cambio en las variables sobre la probabilidad de default es aproximadamente la mitad del impacto que tienen en el modelo de la muestra estratificada. Para conocer las magnitudes de los cambios en la probabilidad de incumplir producto de cambios en las variables independientes ver los coeficientes en la sexta columna del **Cuadro 4**.

4.1.4 Bondad de ajuste

Para determinar la bondad de ajuste en el modelo lineal se utilizan el R^2 y el R^2 -ajustado. Estos indicadores determinan la proporción de la variación de y que puede ser explicada por las variables independientes incluidas en el modelo. Como el R^2 aumenta cuando

aumenta la cantidad de variables incluidas en el modelo, resulta más preciso utilizar el R^2 -ajustado que está modificado en función de la cantidad de variables incluidas. Esto se comprueba comparando los resultados de ambos indicadores contra el PRPC.

El R^2 -ajustado, aplicado a la *muestra estratificada*, nos indica que el modelo lineal restringido a las variables significativas (con R^2 -ajustado igual a 0.3131) es mejor que el modelo completo con todas las variables (cuyo R^2 -ajustado es igual a 0.3129). Esto es así a pesar de que el R^2 indica lo contrario (0.3172 para el restringido y 0.3177 para el completo). El cálculo de los respectivos PRPC (76.3% para el restringido y 75.9% para el completo) confirman los mejores resultados que arroja el modelo calculado sobre el conjunto restringido de variables (ver **Cuadro 11**).

En la *muestra aleatoria* sucede lo mismo que en la estratificada: el R^2 es mayor para el modelo lineal completo (0.0767 contra 0.0756 para el restringido) y el R^2 -ajustado es mayor para el modelo lineal restringido (0.0710 contra 0.0702 para el completo). Confirmando la afirmación de Long (citada en la sección 3.2.2) el modelo puede invalidarse si una de las categorías de una variable categórica tiene pocos casos. En la muestra aleatoria bajo análisis, justamente es la variable dependiente la que tiene pocos casos (10.5%). Al observar las tablas de PRPC (**Cuadros 18-19**) se detecta que ambos modelos, completo y restringido, aplicados a la muestra aleatoria que tiene el inconveniente de tener pocos casos de incumplidores, predicen que todos los individuos en la muestra cumplirán con su deuda. Por tal motivo, ambos modelos predicen que el 100% cumplirá, por lo tanto, acertando en el 89.5% de los casos. Aunque este resultado parece alentador por el alto índice de aciertos, lo cierto es que resulta incapaz de detectar los incumplidores.

4.2 Modelos probit y logit

Igual que para los modelos lineales de probabilidad, en una primera etapa se estimaron los modelos completos, que incluyen todas las variables independientes no altamente correlacionadas, para las regresiones probit y logit en la muestra estratificada y aleatoria. En una segunda etapa se volvieron a estimar estos modelos pero esta vez sin incluir a las variables que resultaron no significativas individualmente. A estos modelos se los denominó probit restringido y logit restringido.

4.2.1 Estimación

La ecuación del modelo *probit* completo para la *muestra estratificada* es la siguiente (**Cuadro 5**):

$$Pr(Default_i | X_i) = \Phi (0.5580 + 0.6254Sexo_i - 0.0246Edad_i - 0.0154Antemp_i - 0.3178Ingreso_i - 0.0138DF_i + 0.000048DF/Ing_i + 0.4163CBF_i)$$

El modelo *probit* completo de la *muestra aleatoria* tiene la siguiente función:

$$Pr(Default_i | X_i) = \Phi (-0.7064 + 0.1100Sexo_i - 0.0112Edad_i - 0.0144Antemp_i - 0.2380Ingreso_i - 0.0344DF_i - 0.0024DF/Ing_i + 0.3157CBF_i)$$

Por otro lado, los modelos *logit* que incluyen a todas las variables independientes (**Cuadro 8**) pueden representarse con las siguientes ecuaciones:

$$Pr(Default_i | X_i) = \lambda (0.7795 + 1.0695Sexo_i - 0.0408Edad_i - 0.0246Antemp_i - 0.5298Ingreso_i - 0.0442DF_i + 0.0001DF/Ing_i + 0.8054CBF_i) \text{ para la muestra estratificada}$$

$$Pr(Default_i | X_i) = \lambda (-0.9324 + 0.1911Sexo_i - 0.0219Edad_i - 0.0260Antemp_i - 0.5174Ingreso_i - 0.0746DF_i - 0.0033DF/Ing_i + 0.5756CBF_i) \text{ para la muestra aleatoria.}$$

4.2.2 Tests de especificación

Las variables incluidas en los modelos anteriores son *conjuntamente* significativas puesto que los p-valores del estadístico del cociente de verosimilitud ($Prob > \chi^2$ en los **Cuadros 5 y 8**) son nulos. Esto indica que puede rechazarse la hipótesis de que ninguna variable tiene efecto sobre la probabilidad de default a cualquier nivel de significatividad que se elija.

Para la *muestra estratificada*, las variables que resultan *individualmente* significativas al 10% son las mismas en el modelo *probit* y *logit*. La única variable para la cual no puede rechazarse la hipótesis de insignificancia es, en ambos modelos, el cociente entre deuda e ingreso. Al estimar nuevamente estos modelos sin incluir este cociente, surgen las especificaciones “restringidas” de los **Cuadros 6 y 9**:

$$Pr(Default_i | X_i) = \Phi (0.5890 + 0.6360Sexo_i - 0.0246Edad_i - 0.0163Antemp_i - 0.3277Ingreso_i - 0.0131DF_i + 0.4135CBF_i) \text{ para el modelo probit y}$$

$Pr(Default_i | X_i) = \lambda (0.8576 + 1.0919Sexo_i - 0.0409Edad_i - 0.0269Antemp_i - 0.5557Ingreso_i - 0.0404DF_i + 0.7932CBF_i)$ para el modelo logit.

A diferencia de lo que ocurre con la muestra estratificada, en los modelos probit y logit de la *muestra aleatoria* las variables significativas (al 10%) a nivel *individual* no son exactamente las mismas. Así para el modelo probit no resulta individualmente significativo ni el sexo ni el cociente entre deuda e ingreso. Para el modelo logit además de estas variables tampoco es significativa la antigüedad en el empleo. Al excluir estas variables los resultados de las regresiones son:

$Pr(Default_i | X_i) = \phi (-0.6619 - 0.0112Edad_i - 0.0152Antemp_i - 0.2192Ingreso_i - 0.0364DF_i + 0.3157CBF_i)$ para el modelo probit y

$Pr(Default_i | X_i) = \lambda (-0.7848 - 0.0286Edad_i - 0.5092Ingreso_i - 0.0774DF_i + 0.5710CBF_i)$ para el modelo logit.

4.2.3 Efectos marginales en las medias

Mientras que en el modelo lineal el efecto de las variables independientes en la variable dependiente depende únicamente de β , por lo que la interpretación es directa, en los modelos probit y logit este efecto varía según el valor que asume cada una de las variables independientes. Como no es factible calcular los efectos para todos los niveles posibles de las distintas variables, en este estudio se optó por calcularlos para los niveles medios de éstas. Así, lo que se expone a continuación son los *efectos marginales en las medias* de las variables que resultaron individualmente significativas al 10%.

Probit

Con respecto al sexo, una variable binaria, el efecto marginal es la diferencia entre la probabilidad predicha de incumplimiento de los hombres y la probabilidad predicha de incumplimiento de las mujeres, fijando las demás variables en sus medias. Para el modelo probit de la *muestra estratificada* éste es de 0.2448, como surge de aplicar el Stata 7, y resulta no significativo en la muestra aleatoria. Para un individuo con características promedio, un año de edad adicional disminuye la probabilidad de incumplimiento en 0.0098. La probabilidad de incumplir, para un individuo con las mismas características, baja en 0.0061 cuando aumenta en un año su permanencia en el empleo y cae 0.1267 puntos al incrementarse en \$1.000 su salario. Por otro lado, un aumento de \$1.000 en la

deuda disminuye esta probabilidad en 0.0055 puntos y un banco prestamista adicional la incrementa en 0.1660 puntos.

La misma estimación en la *muestra aleatoria* arroja efectos marginales en las medias muy inferiores a los enunciados para la muestra estratificada aunque con el mismo signo (**Cuadro 7**).

Logit

Los efectos marginales en las medias del modelo logit para la *muestra estratificada* son sustancialmente mayores que los del modelo probit para la misma muestra. Sin embargo ocurre lo opuesto con la *muestra aleatoria*: son mayores los efectos marginales en las medias del modelo probit respecto del logit (**Cuadros 7 y 10**). Esto se debe a las distintas curvas de probabilidad, y por lo tanto las distintas pendientes, que trazan los modelos probit y logit cuando varía la cantidad de incumplidores en la muestra.

4.2.4 Bondad de ajuste

Probit

En las estimaciones probit de la *muestra estratificada*, para el modelo completo y el restringido, tanto los valores del pseudo- R^2 (0.2850 y 0.2839 respectivamente) como del PRPC (76.9% y 76.8% respectivamente) indican que el modelo completo se ajusta mejor a la realidad que el restringido (**Cuadro 11**).

En la *muestra aleatoria* el pseudo- R^2 indica lo mismo, es decir, que el modelo completo ajusta mejor (0.1338 contra 0.1309 para el restringido). Al igual que en el modelo lineal, se produce el problema planteado por Long, por lo tanto, el PRPC resulta indistinto para el modelo completo y el restringido (89.2%), ya que en ambos casos el modelo predice casi el 100% como cumplidores (**Cuadros 20 y 21**).

Logit

En la *muestra estratificada* el pseudo- R^2 y el PRPC revelan informaciones contradictorias. El primero indica que el modelo logit restringido (0.3279) ajusta mejor que el completo (0.2943). El segundo muestra que el modelo logit completo predice correctamente (77%) más casos que el logit restringido (76.9%). Considerando que el PRPC es un indicador de

bondad de ajuste más preciso que el pseudo- R^2 , por ser afectado este último por la cantidad de variables incluidas en el modelo, el modelo logit completo resulta mejor que el restringido.

Contrariamente, en la *muestra aleatoria* ambos indicadores revelan la misma información: el modelo logit completo ajusta mejor que el restringido (el pseudo- R^2 es 0.1348 y el PRPC es 89.3% para el completo y el pseudo- R^2 es 0.1279 y el PRPC es 89.2% para el restringido).

4.3 Comparación de modelos

4.3.1 Modelos aplicados a las muestras originales

Cuadro 11: Indicadores de Bondad de ajuste

Modelos aplicados sobre la muestra de origen	Muestra estratificada				Muestra aleatoria			
	R2	R2-ajustado	Pseudo R2	PRPC	R2	R2-ajustado	Pseudo R2	PRPC
Lineal completo	0.3177	0.3129	-	75.90%	0.0767	0.0702	-	89.50%
√ Lineal restringido	0.3172	0.3131	-	76.30%	0.0756	0.0710	-	89.50%
√ Probit completo	-	-	0.2850	76.90%	-	-	0.1338	89.20%
Probit restringido	-	-	0.2839	76.80%	-	-	0.1309	89.20%
√ Logit completo	-	-	0.2943	77.00%	-	-	0.1348	89.30%
Logit restringido	-	-	0.3279	76.90%	-	-	0.1279	89.20%

En el cuadro anterior se ve claramente que la aplicación de cualquiera de los tres modelos a la *muestra aleatoria* (con pocos casos de *default*) arroja resultados insatisfactorios: los tres tipos de R^2 bajos y el porcentaje de respuestas predichas correctamente casi igual al porcentaje de casos de cumplidores ya que los modelos predicen casi un 100% de cumplimiento. Por tal motivo el PRPC resulta tan alto aunque el modelo es totalmente ineficiente para predecir los casos de *default*.

Luego de analizar la divergencia entre los indicadores de bondad de ajuste que surgen de la aplicación de los distintos modelos a la *muestra estratificada*, se selecciona el modelo que mejor PRPC tiene. Se elige este indicador porque en definitiva representa en forma más directa y práctica lo que se está buscando: la mayor proporción de respuestas correctas. En otras palabras, se elige el modelo que mejor resultados arroja. Por otro lado, el PRPC es el único indicador común a todos los modelos.

4.3.2 Modelos aplicados a la muestra aleatoria de control

Hasta aquí se han utilizado dos muestras, una estratificada y otra aleatoria, para obtener los parámetros de los tres modelos en estudio: lineal, probit y logit. Luego se calcularon los indicadores de bondad de ajuste (R^2 's y PRPC) aplicando los modelos con los parámetros obtenidos sobre las mismas muestras que les dieron origen. Ahora es necesario validar la calidad de predicción de estos modelos aplicándolos a una tercera muestra que se denomina de "control". Para esto se calcularon los PRPC de cada modelo aplicado a la muestra de control.

La muestra de control fue seleccionada aleatoriamente. En ésta, el 8.8% de los individuos habían incurrido en incumplimiento en el período 12/2003 – 12/2006 y el restante 91.2% se mantuvo en situación 1 o 2 en dicho lapso.

Cuadro 24: Indicadores de Bondad de ajuste cuando los modelos se regresan sobre una muestra aleatoria de control

Modelos aplicados sobre la muestra aleatoria de control	Muestra estratificada	Muestra aleatoria
	PRPC	PRPC
√ Lineal completo	67.00%	91.20%
Lineal restringido	66.90%	91.20%
√ Probit completo	67.60%	91.50%
Probit restringido	66.90%	91.50%
√ Logit completo	68.70%	91.50%
Logit restringido	68.00%	91.50%

El cuadro anterior muestra la ineficiencia de los modelos obtenidos de la muestra aleatoria, ya que si bien predicen correctamente el total de los cumplidores erra en la totalidad de los incumplidores. Por lo tanto se continuará con el análisis de los modelos obtenidos de la muestra estratificada, aplicándolos sobre la de control.

Para los tres tipos de modelos las especificaciones completas tienen un mayor PRPC que las especificaciones restringidas. El tipo de modelo que mejor predice el *default* es el logit porque tiene el mayor PRPC, aunque las diferencias no son significativas.

Una vez seleccionado el modelo logit completo de la muestra estratificada se analiza qué corte (γ) resulta óptimo. Para esto se armó una tabla de frecuencia de predicción contra observado, al igual que para calcular el PRPC cuando se fijó el corte en 0.5, esta vez para

cada corte entre 0.1 y 1. En el **Cuadro 37** se muestra, a modo de resumen, el porcentaje de respuestas predichas correctamente y de errores tipo 1 y 2, por separado y conjuntamente, para cada corte. En el Gráfico 1 se ilustran estos resultados. Puede observarse que el corte con el mínimo porcentaje total de errores y por lo tanto el mayor PRPC es 0.5.



Universidad de
San Andrés

5. Conclusiones

Análisis de las variables independientes

Las variables significativas al 10% (esto significa con el 10% de probabilidad de incurrir en error tipo I) que quedaron incluidas en el modelo completo son: sexo, edad, antigüedad en el empleo, ingreso, deuda final, deuda final/ingreso y cantidad de bancos final. Fueron descartadas: deuda final/deuda inicial, deuda final-deuda inicial, (deuda final-deuda inicial)/ingreso, cantidad de bancos final/cantidad de bancos inicial, cantidad de bancos final-cantidad de bancos inicial. Estas fueron eliminadas como resultado de haber tomado los pares de variables con alta correlación, elegido la variable más significativa (con menor p-valor) y descartado la otra.

Las variables elegidas y sus correspondientes parámetros quedan plasmadas en el modelo logit, de especificación completa, obtenido de la muestra estratificada que se expone a continuación. Fue éste el modelo seleccionado porque es el que mejor predice el *default* ya que tiene el mayor PRPC, aunque las diferencias con los otros modelos no son significativas. La fórmula del modelo logit es la siguiente:

$$Pr(\text{Default}_i | X_i) = \lambda (0.7795 + 1.0695\text{Sexo}_i - 0.0408\text{Edad}_i - 0.0246\text{Antemp}_i - 0.5298\text{Ingreso}_i - 0.0442\text{DF}_i + 0.0001\text{DF/Ing}_i + 0.8054\text{CBF}_i) \text{ para la muestra estratificada}$$

-De la muestra en estudio se verifica que las mujeres son más cumplidoras que los hombres. En este modelo el sexo es una variable binaria donde a las mujeres se les asigna el valor 0 y a los hombres el 1. Como el valor de la variable es ascendente de mujer a hombre (de 0 a 1) y el parámetro de esta variable es positivo (+1.0695), esto indica que los hombres tienen mayor probabilidad de default que las mujeres.

-A medida que aumenta la edad de la persona, mejora su perfil de cumplimiento. Esto resulta lógico porque con los años las personas adquieren mayor experiencia en la administración de sus recursos y en la atención de sus compromisos; también con los años normalmente aumentan sus ingresos. Esto se refleja en el modelo con el signo negativo del beta correspondiente a la edad (-0.0408).

-Con mayor antigüedad en el empleo disminuye la probabilidad de incumplimiento, esto se debe a la mayor estabilidad en el origen de los recursos para atender las obligaciones y al

incremento del ingreso debido al progreso jerárquico. Su beta es negativo reflejando esta relación inversa (-0.0246).

-Con el aumento del ingreso baja la probabilidad de default en la medida que los mayores salarios exceden con mayor holgura el monto para cubrir las necesidades mínimas. Así lo refleja el beta negativo (-0.5298).

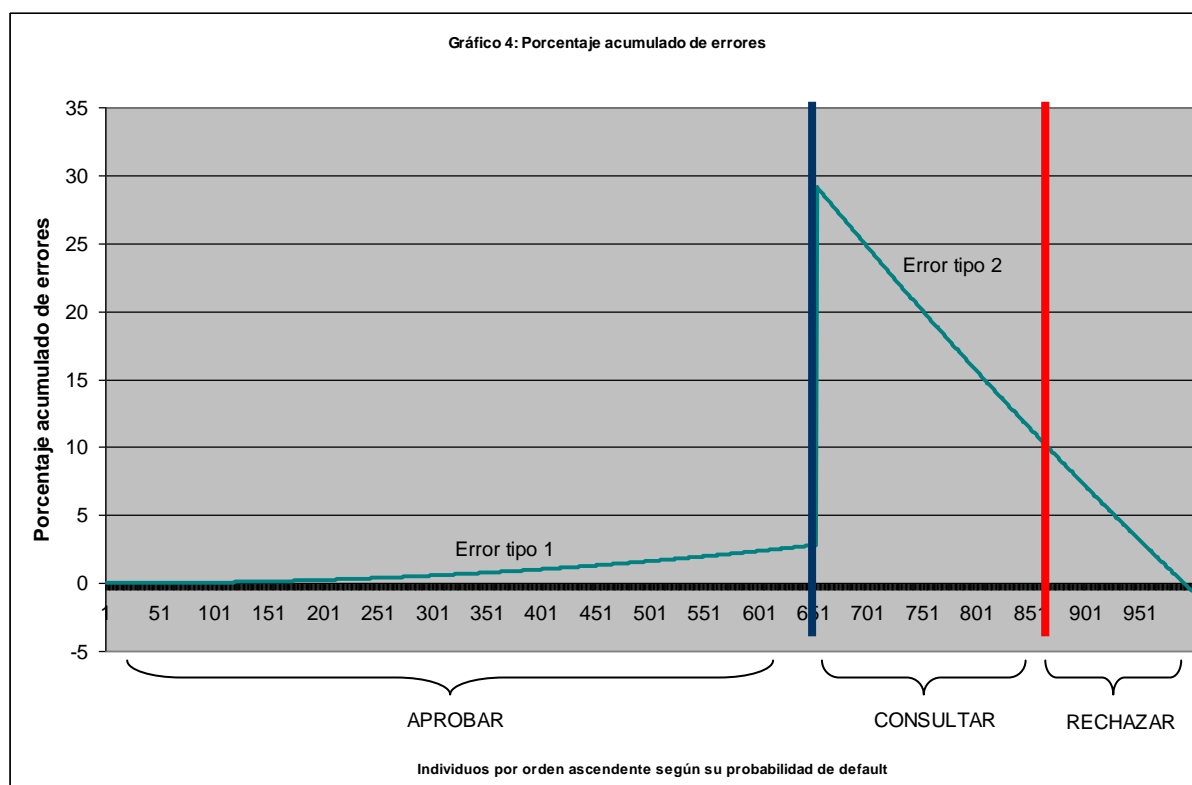
-A mayor deuda menor probabilidad de incumplimiento, aunque esto parezca contradictorio, podría deberse a que los bancos aumentan su asistencia crediticia a los clientes con mejor historial de cumplimiento. Otro motivo podría ser que las deudas de mayor monto por lo general corresponden a hipotecas o prendas, cuyo largo plazo de cancelación se traduce en cuotas relativamente pequeñas respecto al capital adeudado y por lo tanto no afectan significativamente el flujo de fondos mensual de los asalariados. En contraposición, las deudas por tarjetas de créditos, créditos personales y descubiertos en cuentas corrientes, tienen un corto plazo de cancelación, y aunque son de montos menores que los créditos hipotecarios o prendarios, pueden comprometer la capacidad de pago de los individuos, provocando el incumplimiento. Esta relación inversa corresponde a un beta negativo de -0.0442.

-A medida que aumenta la relación entre el total de la deuda sobre los ingresos del individuo será menor su capacidad de repago y por lo tanto mayor la probabilidad de default. Esta es una relación directa con un beta positivo de +0.0001.

-A medida que aumenta la cantidad de bancos que asisten a un individuo aumenta su probabilidad de default. El mantenimiento de una cuenta corriente implica un significativo costo y esfuerzo administrativo para cada prestatario. Por lo tanto se puede deducir que, quien se administre adecuadamente, debería relacionarse con la mínima cantidad de bancos necesaria para obtener el financiamiento requerido. Si recurre a una excesiva cantidad de bancos estaría demostrando que en los bancos que ya lo asisten no obtiene el crédito que necesita. Tal vez, según su perfil, los bancos a los que recurre le otorgan créditos con límites muy bajos que lo obligan a recurrir a más entidades, logrando un endeudamiento que supera el aceptable por cada banco individualmente. Esto se refleja en un beta positivo de +0.8054.

Aplicación del modelo de decisión propuesto

Elegido el modelo logit completo que surge de la muestra estratificada, se calcularon las probabilidades de default y se ordenaron los individuos según las mismas. Luego se calculó y graficó la distribución de los errores de aprobar incumplidores (error tipo 1) y los errores de rechazar cumplidores (error tipo 2) con 0.5 como corte (ver **Cuadro 38 y Gráficos 2 y 3** para entender de dónde surge el gráfico).



Decisión	Porcentaje de error aceptable	P(BX)	Porcentaje de individuos
Aprobar	<3% individuos que incumplen	< 0.5	65.30%
Consultar		> 0.5 y < 0.75	21.30%
Rechazar	<10% individuos que cumplen	> 0.75	13.40%

La gerencia de créditos de un banco deberá determinar hasta qué porcentaje de individuos incumplidores se permite aprobar (en el ejemplo 3%), y por otro lado, qué cantidad de cumplidores consiente rechazar (en el ejemplo 10%). Para facilitar esta toma de decisiones se presentan, en el gráfico de referencia, las opciones posibles: “aprobar”, “consultar” y “rechazar”. Fijando los porcentajes admisibles de cada tipo de error, automáticamente se determinan los cortes correspondientes en la probabilidad de default (0.5 y 0.75; probabilidades de default correspondientes a los individuos 653 y 866 respectivamente

según **Gráfico 4**). Esta herramienta, además de facilitar la comprensión del riesgo implícito en cada tramo de decisión, permite apreciar los porcentajes de solicitudes que los oficiales de crédito decidirán aprobar o rechazar y cuántas son las que deberán elevar para una decisión en niveles superiores. Las solicitudes que no se decidan por aprobación o rechazo en el primer nivel decisorio, tienen un mayor costo al requerir más personal y más tiempo en su evaluación. La intervención de niveles jerárquicos superiores podría evitarse si, por ejemplo, quienes deben decidir en primera instancia tuvieran la instrucción de requerir garantías de terceros a los solicitantes que queden clasificados en el tramo de “consultar”.

El sistema informático por el cual se pondría en práctica esta herramienta operaría de la siguiente manera. El oficial de crédito será el encargado de ingresar los datos del solicitante y el sistema automáticamente le informará: “aprobar”, “consultar” y “rechazar”. Para esto el sistema, utilizando el modelo logit, calcula la probabilidad de incumplimiento en base a los datos de cada individuo. Luego el sistema asigna una de las alternativas de decisión según dicha probabilidad.

Esta herramienta permite homogeneizar las decisiones ajustadas a los criterios establecidos por la gerencia. Además facilita la aplicación inmediata por los oficiales de crédito, en todas las sucursales, de cualquier cambio que posteriormente se decida puesto en ejecución a través del cambio en los parámetros del sistema decisorio.

Limitaciones

Este modelo podría ofrecer resultados más precisos de haberse contado con más variables, tales como: nivel educacional de los individuos, estado civil, carga de familia, propiedad inmueble, propiedad automotor, domicilio, etc.

Para lograr un mejor ajuste del modelo se debería haber hecho sobre una muestra de por lo menos 20.000 individuos, que es una cantidad que puede manejarse con el Stata 7, herramienta utilizada para realizar este trabajo. Al ser sólo 1.000 individuos los comprendidos en la muestra estudiada, para algunas variables, como la edad, en los valores extremos no hay individuos suficientes para hacer estadísticas válidas.

Estos modelos deberían recalcularse periódicamente para reflejar el cambio en el peso relativo de las variables a lo largo del tiempo debido a cambios en las condiciones económicas, financieras y demográficas.

Posibles extensiones

Sería muy útil hacer un análisis del costo de equivocarse de dar un crédito a un incumplidor y por otro lado el de rechazar una solicitud de crédito de un cumplidor.

En el primer caso, se produce la pérdida del capital prestado. Pero dicha pérdida depende del tiempo que transcurre entre el otorgamiento y el incumplimiento, ya que a lo largo de este período el banco fue cobrando comisiones e intereses, que compensan en parte la pérdida de capital. Además, en el momento del default, el deudor podría estar utilizando sólo una porción de la línea de crédito otorgada. Otro cálculo que debería hacerse es la posibilidad de recupero a través de la ejecución de garantías prendarias, hipotecarias o de terceros, o la demanda judicial del deudor.

En el segundo caso, se trata de la pérdida de una ganancia potencial. Esta pérdida podría ser más gravosa que la anterior si se mide la utilidad que hubiera generado el cliente rechazado en términos de comisiones e intereses a devengarse a lo largo de una relación de muchos años.

En función de lo dicho anteriormente, se debería calcular el plazo estimado desde el otorgamiento del crédito hasta el default, determinar la utilidad esperada en dicho período, estimar el monto expuesto a riesgo por debajo de la línea de crédito otorgada y la duración esperada de la relación con el cliente. Para ello es menester desarrollar un modelo adecuado.

6. Bibliografía

- Altman, E.I. (1968): “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy”, *The Journal of Finance*, Septiembre 1968.
- Gross, D y Souleles, N. (2002): “An empirical analysis of personal bankruptcy and delinquency”, *The Society for Financial Studies*.
- Long, J. (1997): *Regression Models for Categorical and Limited Dependent Variables. Advanced Quantitative Techniques in the Social Sciences, Volume 7*, SAGE Publications, Thousand Oaks.
- McFadden, D.L. (1974): *Conditional Logit Analysis of Qualitative Choice Behavior*, Academic Press, New York.
- Saunders, A. (1999): *Credit Risk Measurement. New Approaches to Value at Risk and Other Paradigms*, John Wiley & Sons Inc., New York.
- Sosa Escudero, W. (1999): “Tópicos de Econometría Aplicada. Notas de Clase”, Universidad Nacional de La Plata, La Plata.
- Wooldridge, J. (2006): *Introducción a la econometría. Un enfoque moderno*, Thomson Editores Spain, Madrid.

San Andrés

7. Glosario	Página
Coefficiente beta	12
Coefficiente de determinación	14
Convergencia	18
<i>Credit-scoring system</i>	4
<i>Default</i>	16
<i>Demand effect</i>	6
Efecto marginal	13
Efecto marginal en la media	13
Esperanza condicional	7
Estadístico del cociente de verosimilitud	11
Estadístico F	11
Estimación de máxima verosimilitud	18
<i>Expert system</i>	3
Función de distribución acumulada logística estandarizada	8
Función de distribución acumulada normal tipificada	8
Homoscedasticidad	5
Modelo lineal	7
Modelo logit	8
Modelo probit	8
Muestra de control	28
Muestra estratificada	18
Nivel de significatividad	9
P-valor	10
Porcentaje de respuestas predichas correctamente	14
Pseudo- R^2	14
<i>Rating system</i>	4
<i>Risk effect</i>	6
Test de significatividad global	10
Test de significatividad individual	9
<i>Z-score model</i> de Altman	4

8. Anexo

Cuadro 1: Estadísticas descriptivas

Variable	Muestra estratificada					Muestra aleatoria				
	Num. Obs. (n)	Media	Desv. Estándar	Mínimo	Máximo	(n)	Media	Desv. Estándar	Mínimo	Máximo
Default	1000	0.50	0.50	0.00	1.00	1000	0.11	0.31	0.00	1.00
Sexo	1000	0.67	0.47	0.00	1.00	1000	0.63	0.48	0.00	1.00
Edad	1000	43.20	10.81	24.00	87.00	1000	44.02	10.46	22.00	77.00
AntEmp	1000	10.52	7.62	0.00	37.00	1000	10.63	8.12	0.00	37.00
Ingreso	1000	2.03	1.35	0.00	7.46	1000	2.09	1.53	0.00	29.39
DF	1000	9.39	23.82	0.00	302.67	1000	6.67	14.31	0.00	195.85
DF/Ing	1000	191.77	1571.25	0.00	38500.00	1000	7.88	89.62	0.00	2314.00
DF/DI	1000	5.91	15.36	0.00	278.45	1000	9.30	25.82	0.00	367.00
DF-DI	1000	6.94	23.57	-7.99	294.07	1000	5.13	13.99	-8.10	188.55
(DF-DI)/Ing	1000	106.08	1406.14	-2994.00	37000.00	1000	1.20	150.71	-4200.00	2114.00
CBF	1000	2.51	1.73	0.00	11.00	1000	1.74	1.26	0.00	8.00
CBF/CBI	1000	1.65	1.21	0.00	11.00	1000	1.35	0.95	0.00	6.00
CBF-CBI	1000	0.78	1.63	-4.00	10.00	1000	0.39	1.16	-3.00	5.00

Cuadro 2: Matriz de correlación de la muestra estratificada

	Default	Sexo	Edad	AntEmp	Ingreso	DF	DF/Ing	DF/DI	DF-DI	(DF-DI)/Ing	CBF	CBF/CBI	CBF-CBI
Default	1	0.1379	-0.2382	-0.2062	-0.3604	-0.1183	0.0991	-0.0347	-0.1010	0.0571	0.3717	0.3409	0.3878
Sexo	0.1379	1	0.0607	-0.0760	0.1263	0.0347	0.0521	0.0144	0.0307	0.0288	0.0199	-0.0252	-0.0038
Edad	-0.2382	0.0607	1	0.4139	0.1869	-0.0394	-0.0882	-0.0804	-0.0580	-0.0599	-0.0449	-0.1333	-0.1077
AntEmp	-0.2062	-0.0760	0.4139	1	0.2141	-0.0140	-0.1636	-0.0457	-0.0211	-0.1014	-0.0038	-0.0865	-0.0780
Ingreso	-0.3604	0.1263	0.1869	0.2141	1	0.3068	-0.1788	0.1532	0.2882	-0.1096	-0.0909	-0.1244	-0.1079
DF	-0.1183	0.0347	-0.0394	-0.0140	0.3068	1	0.0232	0.5926	0.9969	0.0372	0.1524	0.0828	0.1265
DF/Ing	0.0991	0.0521	-0.0882	-0.1636	-0.1788	0.0232	1	0.0629	0.0246	0.9362	-0.0088	-0.0003	-0.0126
DF/DI	-0.0347	0.0144	-0.0804	-0.0457	0.1532	0.5926	0.0629	1	0.6129	0.0849	0.0844	0.1734	0.1490
DF-DI	-0.1010	0.0307	-0.0580	-0.0211	0.2882	0.9969	0.0246	0.6129	1	0.0422	0.1468	0.0966	0.1364
(DF-DI)/Ing	0.0571	0.0288	-0.0599	-0.1014	-0.1096	0.0372	0.9362	0.0849	0.0422	1	0.0085	0.0207	0.0109
CBF	0.3717	0.0199	-0.0449	-0.0038	-0.0909	0.1524	-0.0088	0.0844	0.1468	0.0085	1	0.6768	0.8437
CBF/CBI	0.3409	-0.0252	-0.1333	-0.0865	-0.1244	0.0828	-0.0003	0.1734	0.0966	0.0207	0.6768	1	0.8929
CBF-CBI	0.3878	-0.0038	-0.1077	-0.0780	-0.1079	0.1265	-0.0126	0.1490	0.1364	0.0109	0.8437	0.8929	1

Cuadro 3: Modelo lineal completo

Variable	Muestra estratificada				Muestra aleatoria			
	Coficiente	Error estándar	t	P> t	Coficiente	Error estándar	t	P> t
Sexo	0.1825011	0.0284006	6.43	0.000	0.0129494	0.0197194	0.66	0.512
Edad	-0.007252	0.0013510	-5.37	0.000	-0.0020939	0.0009759	-2.15	0.032
Ant Empleo	-0.004370	0.0019443	-2.25	0.025	-0.0023531	0.0012614	-1.87	0.062
Ingreso	-0.100431	0.0109556	-9.17	0.000	-0.0258279	0.0063227	-4.08	0.000
Deuda Final	-0.002149	0.0005954	-3.61	0.000	-0.0017806	0.0006870	-2.59	0.010
Deud F div Ing	0.0000072	0.0000086	0.83	0.407	-0.0000884	0.0001052	-0.84	0.401
CBF	0.1019152	0.0077835	13.09	0.000	0.0523684	0.0076945	6.81	0.000
Constante	0.7041831	0.0614605	11.46	0.000	0.1895137	0.0460622	4.11	0.000

n	1000	n	1000
F(7, 992)	65.99	F(7, 992)	11.78
Prob > F	0.000	Prob > F	0.000
R2	0.3177	R2	0.0767
R2-ajustado	0.3129	R2-ajustado	0.0702

Cuadro 4: Modelo lineal restringido

Variable	Muestra estratificada				Muestra aleatoria			
	Coficiente	Error estándar	t	P> t	Coficiente	Error estándar	t	P> t
Sexo	0.1840733	0.0283330	6.50	0.000	no significativa individualmente			
Edad	-0.0072604	0.0013507	-5.38	0.000	-0.0020939	0.0009746	-2.15	0.032
AntEmp	-0.0045367	0.0019336	-2.35	0.019	-0.0023953	0.0012582	-1.90	0.057
Ingreso	-0.1020180	0.0107859	-9.46	0.000	-0.0247756	0.0062310	-3.98	0.000
DF	-0.0021098	0.0005934	-3.56	0.000	-0.0018095	0.0006844	-2.64	0.008
CBF	0.1016485	0.0077757	13.07	0.000	0.0523169	0.0076905	6.80	0.000
Constante	0.7101390	0.0610308	11.64	0.000	0.1955259	0.0441269	4.43	0.000

n	1000	n	1000
F(6, 993)	76.9	F(5, 994)	16.26
Prob > F	0.000	Prob > F	0.000
R2	0.3172	R2	0.0756
R2-ajustado	0.3131	R2-ajustado	0.0710

Cuadro 5: Probit completo

Variable	Muestra estratificada				Muestra aleatoria			
	Coficiente	Error estándar	z	P> z	Coficiente	Error estándar	z	P> z
Sexo	0.6254027	0.1013082	6.17	0.000	0.1099605	0.1229858	0.89	0.371
Edad	-0.0246182	0.0048311	-5.10	0.000	-0.0112002	0.0061253	-1.83	0.067
AntEmp	-0.0154094	0.0067702	-2.28	0.023	-0.0143583	0.0086307	-1.66	0.096
Ingreso	-0.3177663	0.0409176	-7.77	0.000	-0.2380199	0.0691700	-3.44	0.001
DF	-0.0137715	0.0040601	-3.39	0.001	-0.0343965	0.0195376	-1.76	0.078
DFdivIng	0.0000478	0.0000461	1.04	0.300	-0.0024345	0.0201214	-0.12	0.904
CBF	0.4162626	0.0361780	11.51	0.000	0.3156562	0.0490010	6.44	0.000
Constante	0.5580462	0.2164718	2.58	0.010	-0.7064036	0.2916476	-2.42	0.015

n	1000	n	1000
LR chi2(7)	395.14	LR chi2(7)	89.87
Prob > chi2	0.000	Prob > chi2	0.000
Pseudo R2	0.2850	Pseudo R2	0.1338
Log verosimilitud	-495.57788	Log verosimilitud	-290.9952

Cuadro 6: Probit restringido

Variable	Muestra estratificada				Muestra aleatoria			
	Coefficiente	Error estándar	z	P> z	Coefficiente	Error estándar	z	P> z
Sexo	0.6359508	0.1009652	6.30	0.000	no significativa individualmente			
Edad	-0.0246496	0.0048316	-5.10	0.000	-0.0112300	0.0060962	-1.84	0.065
AntEmp	-0.0162727	0.0067317	-2.42	0.016	-0.0151623	0.0085973	-1.76	0.078
Ingreso	-0.3277234	0.0400829	-8.18	0.000	-0.2191981	0.0574542	-3.82	0.000
DF	-0.0131108	0.003941	-3.33	0.001	-0.0363963	0.0131154	-2.78	0.006
CBF	0.4135180	0.036059	11.47	0.000	0.3157149	0.0490338	6.44	0.000
Constante	0.5890013	0.2148992	2.74	0.006	-0.6619332	0.2742111	-2.41	0.016

n	1000.000	n	1000
LR chi2(6)	393.59	LR chi2(5)	87.94
Prob > chi2	0.000	Prob > chi2	0.000
Pseudo R2	0.2839	Pseudo R2	0.1309
Log verosimilitud	-496.35126	Log verosimilitud	-291.96389

Cuadro 7: Efectos Marginales probit

Variable	Muestra estratificada	Muestra aleatoria
	dF/dx	dF/dx
Sexo	0.2448289	0.0143811
Edad	-0.0098169	-0.0014956
Ant Emp	-0.0061448	-0.0019173
Ingreso	-0.1267147	-0.0317833
DF	-0.0054916	-0.0045931
DF div Ing	0.0000190	-0.0003251
CBF	0.1659918	0.0421503

Cuadro 8: Logit completo

Variable	Muestra estratificada				Muestra aleatoria			
	Coefficiente	Error estándar	Z	P> z	Coefficiente	Error estándar	z	P> z
Sexo	1.06946	0.1751064	6.11	0.000	0.1910860	0.2296453	0.83	0.405
Edad	-0.0408338	0.0083765	-4.87	0.000	-0.0219441	0.0114169	-1.92	0.055
AntEmp	-0.0246443	0.0117131	-2.10	0.035	-0.0259964	0.0168959	-1.54	0.124
Ingreso	-0.5297873	0.0732314	-7.23	0.000	-0.5173524	0.1300431	-3.98	0.000
DF	-0.0441914	0.0116838	-3.78	0.000	-0.0745665	0.0285460	-2.61	0.009
DFdivIng	0.0001292	0.0001175	1.10	0.271	-0.0033026	0.0118564	-0.28	0.781
CBF	0.8053548	0.0744571	10.82	0.000	0.5755771	0.0907270	6.34	0.000
Constante	0.7794522	0.3749958	2.08	0.038	-0.9324074	0.5380956	-1.73	0.083

n	1000	n	1000
LR chi2(7)	408.04	LR chi2(7)	90.55
Prob > chi2	0.000	Prob > chi2	0.000
Pseudo R2	0.2943	Pseudo R2	0.1348
Log verosimilitud	-489.12803	Log verosimilitud	-290.65797

Cuadro 9: Logit restringido

Variable	Muestra estratificada				Muestra aleatoria			
	Coficiente	Error estándar	z	P> z	Coficiente	Error estándar	z	P> z
Sexo	1.0918580	0.1745418	6.26	0.000	no significativa individualmente			
Edad	-0.0408733	0.0083728	-4.88	0.000	-0.0285501	0.0108110	-2.64	0.008
AntEmp	-0.0268634	0.0116293	-2.31	0.021	no significativa individualmente			
Ingreso	-0.5556627	0.0715850	-7.76	0.000	-0.5091614	0.1250617	-4.07	0.000
DF	-0.0404040	0.0110882	-3.64	0.000	-0.0773786	0.0267340	-2.89	0.004
CBF	0.7932421	0.0735091	10.79	0.000	0.5710224	0.0902514	6.33	0.000
Constante	0.8575530	0.3715065	2.31	0.021	-0.7847707	0.5122090	-1.53	0.125

n	1000	n	1000
LR chi2(10)	454.54	LR chi2(4)	85.96
Prob > chi2	0.000	Prob > chi2	0.000
Pseudo R2	0.3279	Pseudo R2	0.1279
Log verosimilitud	-465.87572	Log likelihood	-292.95393

Cuadro 10: Efectos Marginales logit

Variable	Muestra estratificada	Muestra aleatoria
	dF/dx	dF/dx
Sexo	0.2597555	0.0114653
Edad	-0.0102067	-0.0013444
AntEmp	-0.0061600	-0.0015926
Ingreso	-0.1324248	-0.0316943
DF	-0.0110460	-0.0045681
DFdivIng	0.0000323	-0.0002023
CBF	0.2013052	0.0352613

PRPC sobre sí mismas

Muestra estratificada

Muestra aleatoria

Cuadro 12: Predicciones Lineal completo

Predicción	Observado		Total
	0	1	
0	391	132	523
1	109	368	477
Total	500	500	1000
Errores	21.80%	26.40%	24.10%
PRPC	75.90%		

Cuadro 18: Predicciones Lineal completo

Predicción	Observado		Total
	0	1	
0	895	105	1000
1	0	0	0
Total	895	105	1000
Errores	0.00%	100.00%	10.50%
PRPC	89.50%		

Cuadro 13: Predicciones Lineal restringido

Predicción	Observado		Total
	0	1	
0	390	127	517
1	110	373	483
Total	500	500	1000
Errores	22.00%	25.40%	23.70%
PRPC	76.30%		

Cuadro 19: Predicciones Lineal restringido

Predicción	Observado		Total
	0	1	
0	895	105	1000
1	0	0	0
Total	895	105	1000
Errores	0.00%	100.00%	10.50%
PRPC	89.50%		

Cuadro 14: Predicciones probit completo

Predicción	Observado		Total
	0	1	
0	396	127	523
1	104	373	477
Total	500	500	1000
Errores	20.80%	25.40%	23.10%
PRPC	76.90%		

Cuadro 20: Predicciones probit completo

Predicción	Observado		Total
	0	1	
0	889	102	991
1	6	3	9
Total	895	105	1000
Errores	0.67%	97.14%	10.80%
PRPC	89.20%		

Cuadro 15: Predicciones probit restringido

Predicción	Observado		Total
	0	1	
0	394	126	520
1	106	374	480
Total	500	500	1000
Errores	21.20%	25.20%	23.20%
PRPC	76.80%		

Cuadro 21: Predicciones probit restringido

Predicción	Observado		Total
	0	1	
0	889	102	991
1	6	3	9
Total	895	105	1000
Errores	0.67%	97.14%	10.80%
PRPC	89.20%		

Cuadro 16: Predicciones logit completo

Predicción	Observado		Total
	0	1	
0	397	127	524
1	103	373	476
Total	500	500	1000
Errores	20.60%	25.40%	23.00%
PRPC	77.00%		

Cuadro 22: Predicciones logit completo

Predicción	Observado		Total
	0	1	
0	890	102	992
1	5	3	8
Total	895	105	1000
Errores	0.56%	97.14%	10.70%
PRPC	89.30%		

Cuadro 17: Predicciones logit restringido

Predicción	Observado		Total
	0	1	
0	394	125	519
1	106	375	481
Total	500	500	1000
Errores	21.20%	25.00%	23.10%
PRPC	76.90%		

Cuadro 23: Predicciones logit restringido

Predicción	Observado		Total
	0	1	
0	888	101	989
1	7	4	11
Total	895	105	1000
Errores	0.78%	96.19%	10.80%
PRPC	89.20%		

PRPC sobre muestra aleatoria de control

Muestra estratificada

Muestra aleatoria

Cuadro 25: Predicciones Lineal completo

Predicción	Observado		Total
	0	1	
0	612	30	642
1	300	58	358
Total	912	88	1000
Errores	32.89%	34.09%	33.00%
PRPC	67.00%		

Cuadro 31: Predicciones Lineal completo

Predicción	Observado		Total
	0	1	
0	912	88	1000
1	0	0	0
Total	912	88	1000
Errores	0.00%	100.00%	8.80%
PRPC	91.20%		

Cuadro 26: Predicciones Lineal restringido

Predicción	Observado		Total
	0	1	
0	611	30	641
1	301	58	359
Total	912	88	1000
Errores	33.00%	34.09%	33.10%
PRPC	66.90%		

Cuadro 32: Predicciones Lineal restringido

Predicción	Observado		Total
	0	1	
0	912	88	1000
1	0	0	0
Total	912	88	1000
Errores	0.00%	100.00%	8.80%
PRPC	91.20%		

Cuadro 27: Predicciones probit completo

Predicción	Observado		Total
	0	1	
0	616	28	644
1	296	60	356
Total	912	88	1000
Errores	32.46%	31.82%	32.40%
PRPC	67.60%		

Cuadro 33: Predicciones probit completo

Predicción	Observado		Total
	0	1	
0	910	83	993
1	2	5	7
Total	912	88	1000
Errores	0.22%	94.32%	8.50%
PRPC	91.50%		

Cuadro 28: Predicciones probit restringido

Predicción	Observado		Total
	0	1	
0	608	27	635
1	304	61	365
Total	912	88	1000
Errores	33.33%	30.68%	33.10%
PRPC	66.90%		

Cuadro 34: Predicciones probit restringido

Predicción	Observado		Total
	0	1	
0	910	83	993
1	2	5	7
Total	912	88	1000
Errores	0.22%	94.32%	8.50%
PRPC	91.50%		

Cuadro 29: Predicciones logit completo

Predicción	Observado		Total
	0	1	
0	626	27	653
1	286	61	347
Total	912	88	1000
Errores	31.36%	30.68%	31.30%
PRPC	68.70%		

Cuadro 35: Predicciones logit completo

Predicción	Observado		Total
	0	1	
0	909	82	991
1	3	6	9
Total	912	88	1000
Errores	0.33%	93.18%	8.50%
PRPC	91.50%		

Cuadro 30: Predicciones logit restringido

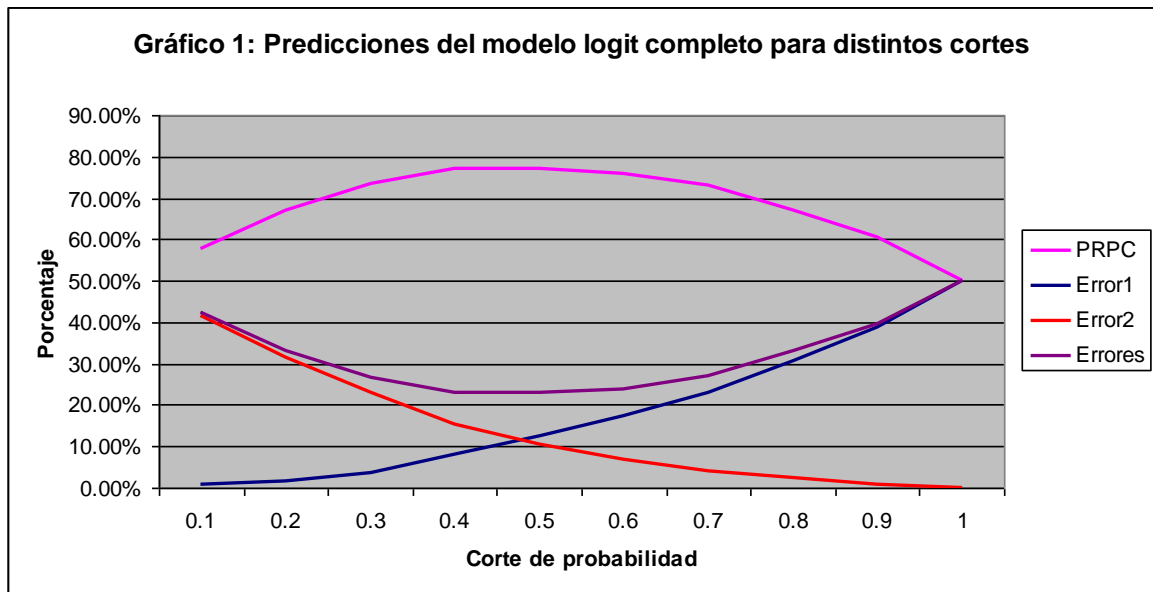
Predicción	Observado		Total
	0	1	
0	619	27	646
1	293	61	354
Total	912	88	1000
Errores	32.13%	30.68%	32.00%
PRPC	68.00%		

Cuadro 36: Predicciones logit restringido

Predicción	Observado		Total
	0	1	
0	909	82	991
1	3	6	9
Total	912	88	1000
Errores	0.33%	93.18%	8.50%
PRPC	91.50%		

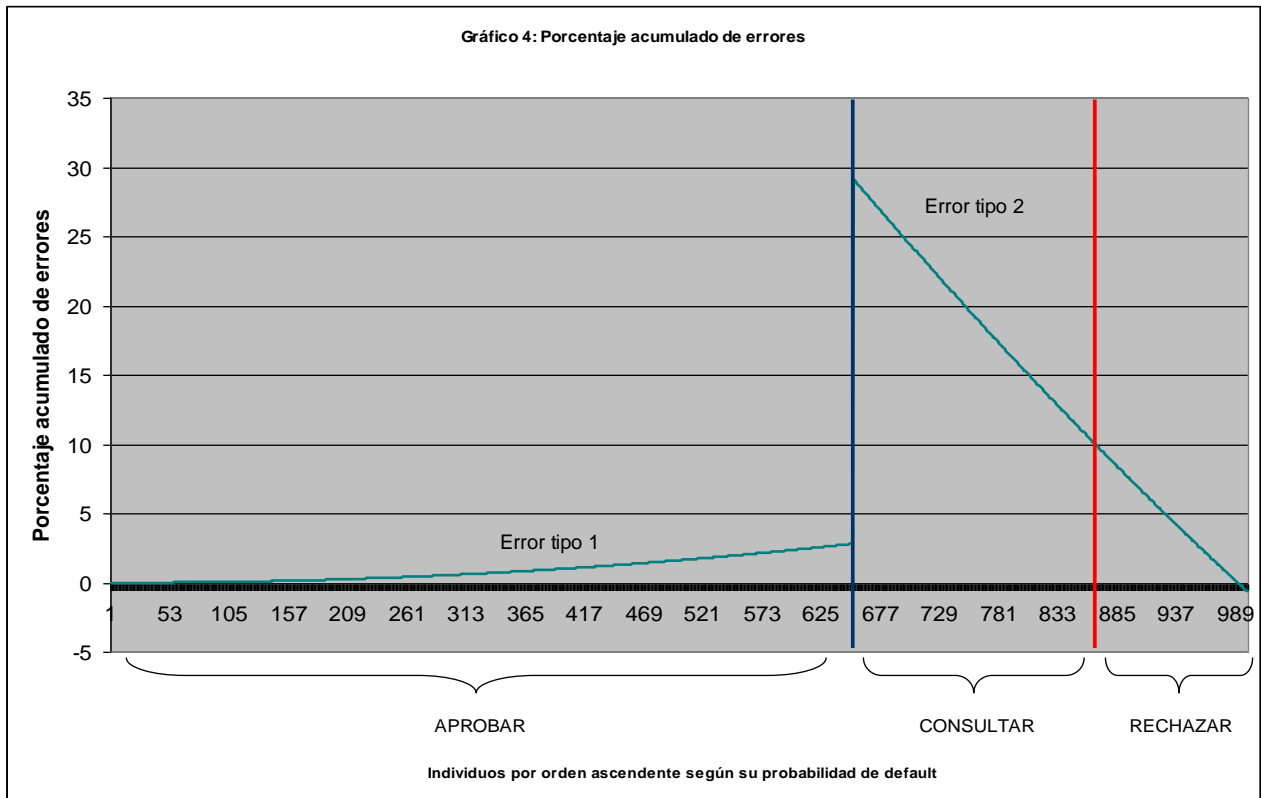
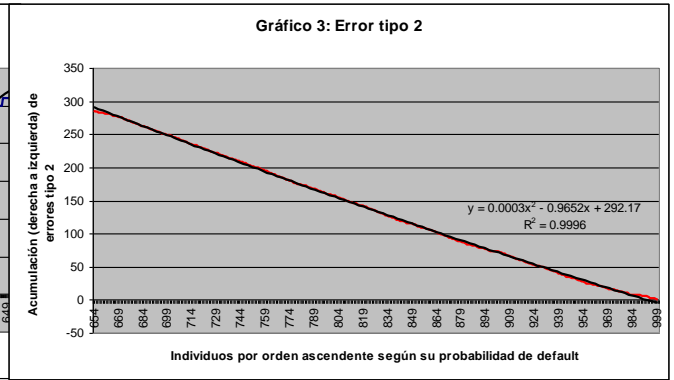
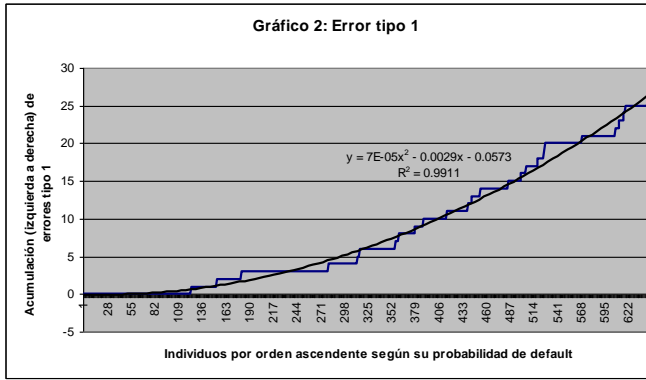
Cuadro 37: Predicciones del modelo logit completo para distintos cortes

Corte	PRPC	Error1	Error2	Errores
0.1	57.60%	0.80%	41.60%	42.40%
0.2	66.80%	1.80%	31.40%	33.20%
0.3	73.40%	3.60%	23.00%	26.60%
0.4	76.90%	7.90%	15.20%	23.10%
0.5	77.00%	12.70%	10.30%	23.00%
0.6	76.00%	17.20%	6.80%	24.00%
0.7	73.00%	22.90%	4.10%	27.00%
0.8	67.10%	30.60%	2.30%	32.90%
0.9	60.40%	38.70%	0.90%	39.60%
1	50.00%	50.00%	0.00%	50.00%



Cuadro 38: Análisis de aciertos y errores por deciles

Decil	P(XB)	Decil	Errores (cantidad)		Aciertos (cantidad)		Total de individuos por decil	Errores (%)		Aciertos (%)		Errores de ambos tipos	Aciertos de ambos tipos
			Aprobar incum_plidor	Rechazar cum_plidor	Rechazar incum_plidor	Aprobar cum_plidor		Aprobar incum_plidor	Rechazar cum_plidor	Rechazar incum_plidor	Aprobar cum_plidor		
1	0.0847	1	0	0	0	100	100	0.00%	0.00%	0.00%	100.00%	0.00%	100.00%
2	0.1494	2	3	0	0	97	100	3.00%	0.00%	0.00%	97.00%	3.00%	97.00%
3	0.2234	3	1	0	0	99	100	1.00%	0.00%	0.00%	99.00%	1.00%	99.00%
4	0.2937	4	6	0	0	94	100	6.00%	0.00%	0.00%	94.00%	6.00%	94.00%
5	0.3661	5	6	0	0	94	100	6.00%	0.00%	0.00%	94.00%	6.00%	94.00%
6	0.4561	6	5	0	0	95	100	5.00%	0.00%	0.00%	95.00%	5.00%	95.00%
		7a	6			47	53	11.32%	0.00%	0.00%	88.68%	11.32%	88.68%
7	0.5450	7b		38	9		47	0.00%	80.85%	19.15%	0.00%	80.85%	19.15%
8	0.6585	8	0	90	10	0	100	0.00%	90.00%	10.00%	0.00%	90.00%	10.00%
9	0.7802	9	0	85	15	0	100	0.00%	85.00%	15.00%	0.00%	85.00%	15.00%
10	0.9956	10	0	73	27	0	100	0.00%	73.00%	27.00%	0.00%	73.00%	27.00%



Decisión	Porcentaje de error aceptable	P(BX)	Porcentaje de individuos
Aprobar	<3% individuos que incumplen	< 0.5	65.30%
Consultar		> 0.5 y < 0.75	21.30%
Rechazar	<10% individuos que cumplen	> 0.75	13.40%