

## Artículos

# *El comportamiento predictivo del cliente financiero de La Plata luego de la crisis 2001. Un enfoque multivariado*

Simonato, Fernando R.\*

*El mercado financiero hace mas de 2 años sufrió la crisis más grande de la historia de la Argentina. La incautación de depósitos de grandes, medianos y pequeños ahorristas socavo la esencia del sistema que basa su órtaleza en 2 pilares fundamentales: La confianza y la seguridad. Para este trabajo se realizo un relevamiento a través de información primaria de las dijérentes tipologías de clientes y sus relaciones con las entidades financieras, para poder determinar las variables más importantes en la influencia de las /uturas decisiones que ellos tomaran. La metodología de análisis de la información primaria es la utilización de la técnica multivariada predictiva: Regresión Logística. Esta técnica permitirá determinar las variables que mas impacto tendrían en el comportamiento del cliente en el futuro mediato.*

La lealtad y la fidelización en las personas es fruto del ejercicio de la libertad. No podemos considerar fiel a quien mantiene al lado por imposibilidad de escoger o de irse. Solo puede considerarse fiel a quien, pudiendo dejar de serlo. Opta por la lealtad. El valor de la lealtad y de la fidelidad se fundamenta en el ejercicio de la voluntad del ser humano quien, en el uso de la libertad, decide tomar una opción excluyendo, al tiempo a las otras.

Para entender un poco mejor el estudio de la fidelidad, analicemos la infidelidad humana en un sentido amplio como la cara opuesta a la fidelidad, en la que encontramos que esta tiene 2 dimensiones (Ver 1)

\*La infidelidad carnal, que se produce cuando se comete el acto infiel.

\*La imaginaria, que se da cuando solo se imagina y se la desea, despierto o en sueños.

En nuestro sistema financiero antes, pero en especial luego del corralito, comenzaron a darse ambos tipos de infidelidades en el mercado. Una pregunta que sistemáticamente se hacen los consultores del sistema financiero es si el ser humano puede ser plenamente y en ambas dimensiones fiel de por vida.

Todo banco aspira a ser él, el objeto de la libre elección de sus clientes. Toda entidad financiera aspira a ser colmada con la lealtad de aquellos. No obstante, no son muchas las que lo consiguen, en especial, después de haber roto el lazo principal (entre la entidad financiera y el cliente) de

la existencia de las mismas, que es la confianza.

Es por ello se intento estudiar a través de una aproximación, el comportamiento del cliente financiero desde la siguiente óptica:

El grado de Confianza o Desconfianza en su entidad financiera para dejar o no el dinero que le tiene retenido el banco como consecuencia del corralito financiero a través de la regresión logística.

En principio el análisis general basado en las entidades bancarias pretende plantear mediante el Análisis Logit un modelo sobre los clientes que fueron afectados por el Corralito financiero de fines de 2001.

### **METODOLOGIA**

\*Recolección de información secundaria para analizar la situación del sistema financiero.

\*Utilización de información a través de una Investigación primaria descriptiva mediante cuestionario estructurado y personalizado.

\*La investigación se realizo del 5 de Mayo hasta el 1 de Agosto de 2003.

\*La muestra alcanzo las 820 personas de la Ciudad de La Plata, de las cuales el 86% era bancarizadas.

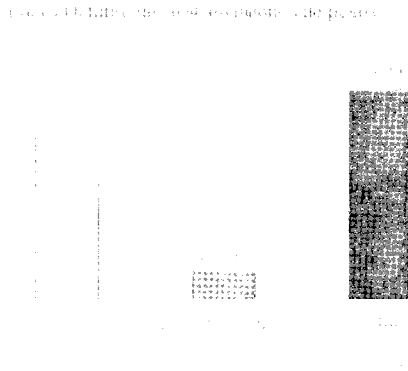
\*El nivel de confianza del estudio alcanzo el 95% con un margen de error de ' 345 %

\*Se trata de determinar la magnitud de cada variable clave (drivers) a tener en cuenta, para lo cual se tuvieron en cuenta variables demográficas, socioeconómicas, actitudinales y comportamentales.

(\*) Simonato, Fernando. Contador Público (UNLP), MBA University of Dallas (USA), Magister en Marketing Internacional (UNCE), Profesor Ad de Comercialización (UNLP), Miembro de la American Marketing Association (USA), Profesor Visitante de Universidades Extranjeras Consultor Trainer de Empresas Nacionales e Internacionales.

## LA CONFIANZA DEL SISTEMA BANCARIO

En el grafico (Ver 2) debajo muestra la evolución de los depósitos en el sistema y en el mismo se muestra claramente que a lo largo de este año los depósitos del Sector Publico tuvieron un crecimiento muy importante y un claro estancamiento de los depósitos del sector privado.



El crecimiento de los depósitos públicos se asemeja al superávit fiscal primario del estado. Los depósitos del sector privado que son esencialmente ahorros voluntarios no crecieron o no fueron volcados al sector financiero. Que significaría este comportamiento de los clientes financieros? Que el componente de la tasa de interés que ofrece el mercado (riesgo institucional o CONFIANZA DEL SISTEMA) es insuficiente para convencer al cliente financiero para que preste sus ahorros.

Es por ello que el trabajo se propone mediante el análisis logit estudiar los diferentes "drivers" o variables comportamentales del cliente financiero que impactaran en su relación con la entidad bancaria para poder así analizar el efecto que tendría el retorno de la confianza al sistema a través de la posibilidad o no de dejar los depósitos privados incautados en las entidades.

## EL ANALISIS LOGIT

El modelo LOGIT ( también llamado modelo de regresión logística, es un método de análisis explicativo y debe ser utilizado cuando se necesita explicar un fenómeno o una conducta

## EL ANALISIS LOGIT

El modelo LOGIT (Ver 3), también llamado modelo de regresión logística, es un método de análisis explicativo y debe ser utilizado cuando se necesita explicar un fenómeno o una conducta humana que implica una dicotomía, como por ejemplo tener o no tener un teléfono celular, tener o no un seguro médico privado, ir o no de vacaciones... y cuando las variables explicativas disponibles están medidas en escalas

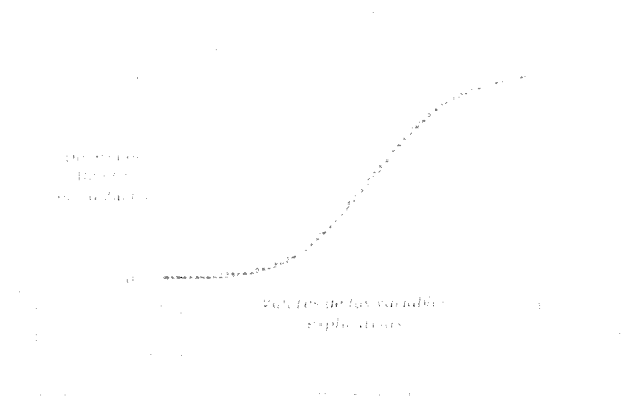
cualitativas (ordinales y nominales) y cuantitativas (contínuas).

El modelo LOGIT, al igual que la mayoría de los modelos explicativos tiene una doble finalidad:

*En primer lugar* determinar cuáles de las variables explicativas del modelo explican la presencia o ausencia del fenómeno estudiado, y en qué medida. En este caso que se quiere explicar la posibilidad de que los clientes financieros DEJARIAN EL DINERO EN EL BANCO O NO DEJARIAN EN DINERO EN EL BANCO a partir de un conjunto de variables sociodemográficas, comportamentales y actitudinales los resultados de la aplicación del modelo nos dirán si este conjunto de variables influyen de forma significativa sobre el hecho de dejar o no dejar el dinero en la institución financiera, y cuáles son las variables explicativas con mayor poder diferenciador entre los clientes que dejarían o no dejarían

*En segundo lugar* determinar con fin predictivo el grupo al que pertenecerá un individuo pendiente de clasificación en base a las respuestas o valores que toma dicho individuo en la serie de variables que más explican la presencia o ausencia del fenómeno estudiado. Siguiendo con el ejemplo anterior, el modelo Logit asignará a estos individuos una probabilidad de DEJAR O NO EL DINERO EN EL BANCO, basándose exclusivamente en el conjunto de variables explicativas utilizadas.

Por lo tanto se trata de un modelo probabilístico que permite predecir si un fenómeno va a ocurrir es decir, que Dejaría el Dinero (probabilidad cercana o igual a 1) o no va a ocurrir, No Dejaría el Dinero (probabilidad cercana igual a 0). Esto significa que los valores predichos de la variable a explicar van a fluctuar entre 0 y 1, siendo, en la mayoría de los casos la relación entre la variable a explicar y las variables explicativas del tipo expuesto en la figura 1.1 (**Relación entre la variable a explicar y las variables explicativas**).



Este tipo de relación expuesto en la figura anterior se obtiene a partir de una distribución logística o función LOGIT, cuya fórmula en el caso de trabajar en P variables explicativas es:

Siendo  $\Pi(x)$  la probabilidad de que el fenómeno ocurra y  $1 - \Pi(x)$  la probabilidad de que el fenómeno no ocurra. Se considera que este es el tipo de distribución más adecuado para explicar una variable dicotómica, puesto que se ha comprobado que muchas facetas del comportamiento humano se ajustan a esta modelización.

En muchos aspectos el modelo LOGIT y el análisis de regresión lineal se parecen. Por ejemplo, el proceso de cálculo de ambos modelos permite estimar los parámetros  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ . Pero la medición dicotómica de la variable a explicar provoca un nuevo planteamiento de los conceptos ya utilizados en el análisis de regresión lineal ( Ver 4)

En el análisis de regresión lineal el método más comúnmente utilizado en la estimación de los parámetros  $\beta$  es el de mínimos cuadrados. Bajo los supuestos habituales de una regresión lineal el método de mínimos cuadrados proporciona unas estimaciones de los parámetros con un número apreciable de propiedades estadísticas.

Por desgracia, cuando el método de mínimos cuadrados se aplica a un modelo cuya variable a explicar es dicotómica los estimadores pierden estas propiedades.

La estimación de los parámetros  $\beta$  en un modelo LOGIT se efectuará aplicando el método de máxima verosimilitud. De forma general este método proporciona valores, para los parámetros desconocidos, que maximizan la probabilidad de obtener el conjunto de valores observados de la variable a explicar.

En la variable uno  $V_1$  se clasificaron los individuos que Dejarían el Dinero en el Banco (Valor 1) frente a los que No Dejarían el Dinero en el Banco (valor 0). El resto de las variables y sus respectivas codificaciones del estudio fueron las siguientes:

La premisa de la regresión logística es determinar que tipología de cliente y que perfil tendrían los que Dejarían el Dinero en su Banco respecto de aquellos que No lo Dejarían tomándose variables socioeconómicas y actitudinales de los bancos para con los clientes.

Un primer aspecto a tener en cuenta es que, antes de iniciar la estimación del modelo, todas las variables explicativas que entrarán en el mismo deberán ser dicotómicas, de no ser así, se procederá a transformarlas. Es imprescindible efectuar esta transformación puesto que en el caso de variables explicativas nominales, los valores tomados por estas variables son meramente identificadores, y no tienen ninguna significación numérica. En general, si una variable categórica toma  $k$  posibles valores, se necesitarán  $k - 1$  variables dicotómicas.

### INTERPRETACIÓN DE LOS PARAMETROS $\beta$

En el caso de una regresión lineal, el valor obtenido en la estimación de un parámetro  $\beta_i$ , nos dice cuál va a ser la variación en la variable a explicar cuando la variable  $X$  asociada experimenta un incremento en una unidad, siempre y cuando las demás variables explicativas se mantengan constantes. En este contexto podemos definir el parámetro  $\beta$  como sigue:

$$\beta_p = f(x_p + 1) - f(x_p)$$

Siendo:

$$f(x_p + 1) = \beta_0 + \beta_1 x_1 + \dots + \beta_p (x_p + 1) + \dots + \beta_k x_k$$

$$f(x_p) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \dots + \beta_k x_k$$

La interpretación de un parámetro  $\beta_i$  en un modelo LOGIT no es tan evidente. Teniendo en cuenta la relación que existe entre la variable a explicar y las variables explicativas, un parámetro  $\beta_i$  quedaría definido

$$\beta_p = f(x_p + 1) - f(x_p)$$

Siendo:

$$f(x_p + 1) = \ln \left( \frac{\prod (x_p + 1)}{1 - \prod (x_p + 1)} \right)$$

$$f(x_p) = \ln \left( \frac{\prod (x_p)}{1 - \prod (x_p)} \right)$$

El Coeficiente  $\beta_i$  representa, pues el cambio que experimentaría la relación entre la probabilidad de que un fenómeno ocurra y la probabilidad de que no ocurra, cuando la variable X, aumente en una unidad y las demás variables explicativas se mantengan constantes.

La interpretación de un parámetro  $\beta_3$  no se efectuará directamente sino

que se hará a partir de un ratio de probabilidad derivado (Ratio de ventaja u odds ratio). Dicha interpretación también dependerá del tipo de variable explicativa de que se trate: dicotómica o de más de dos categorías.

**Variables en la Ecuación** Nel 3.

Cuadro 1.6								
Var.	VARIABLES	$\beta_p$	S.E.	Wald	df	Sig.	R	Exp(B)
V2	HAPODIDO(1)	0.264	0.825	0.1024	1	0.749	-0.31	1.303
V3	LOHALLAM(1)	0.996	0.724	1.892	1	0.169	-0.33	2.707
V4	VIVIENDA(1)	0.094	0.819	0.013	1	0.048	-0.173	1.098
V5	HACECUA	0.035	0.050	0.50	1	0.048	-0.15	1.036
V6	CALISTRI			3.155	2	0.206	0.25	
V6	CALISTRI(1)	0.093	0.873	0.011	1	0.042	0.00	1.097
V6	CALISTRI(2)	-1.241	0.805	2.376	1	0.123	0.22	0.289
V7	SUPRIMIO(1)	2.150	0.900	5.706	1	0.017	0.088	8.589
V8	CESÓLAS(1)	1.548	0.916	2.856	1	0.091	0.211	4.704
V9	NOCAMBIO(1)	4.494	1.208	13.839	1	0.000	-0.43	89.51
V10	CELULAR(1)	1.203	0.790	2.320	1	0.013	-0.243	3.331
V11	INGRESME	-0.001	0.001	1.000	1	0.017	-0.147	0.999
	Constante	-3.827	1.560	6.017	1	0.014		0.022

Los coeficientes Beta que aparecen en la primer columna cuantifican la tasa de incremento o decremento de la curva mostrada anteriormente, es decir la incidencia que cada variable independiente tiene en la ocurrencia del fenómeno (dejaría o no dejaría el dinero).

El signo de estos coeficientes (positivo o negativo) señala si la curva asciende o desciende, es decir, el incremento (o disminución) en la probabilidad de que ocurra el fenómeno analizado (dejar el dinero silo llaman para acordar una devolución consensuada o no) ante un aumento (o disminución) unitario en la variable independiente.

La tercer columna del cuadro 1.6, como explicamos arriba, muestra los resultados del test de Wald para término. Bajo el titulo df se encuentran los grados de libertad (números de categorías menos 1) a la derecha la significación del test de Wald.

La anteuúltima columna muestra la correlación parcial de cada variable con el término dependiente, un instrumento muy útil para detectar la presencia de Multicolinealidad en la regresión con mas de un término independiente.

La última columna - Exp (1 - es el logaritmo neperiano inverso del coeficiente Beta y se interpreta como el aumento en la razón (odds) de que suceda el fenómeno analizado cuando los coeficientes de ' ecuación cambian de 0 a 1.

Si analizamos la Variable (V3) vemos que si al cliente el Banco hubiera llamado para acordar algún tipo de devolución del dinero d corralito es 2.707 veces más probable que hubiera dejado el dinero en banco que aquellos bancos que no lo hubieran hecho.

Por lo tanto, la información útil, en un modelo LOGIT, para interpretar

la relación entre cada variable explicativa y la variable a explicar será el ratio de probabilidad (odds ratio) asociado a cada 1J representado por Exp (B).

Analizando la conformación de la ecuación, observamos que los coeficientes del test de Wald son significativos, excepto la variable V2 (HAPODIDO).

La ecuación de regresión es la siguiente:

$$VI = -3.827 + 0.264 (V2) + 0.996 (V3 + 0.094 (V4) + 0.035 (V5) + 0.093 (V6) (1) - 1.241 "V6, (2), + 1.334 (V6) (3) H- 2.150 "V7 + 1.548 (V8 H- 4.494 (V9 + 1.203 (VIQ) 0.001 (Vii)$$

Cabe también la posibilidad de evaluar la bondad del ajuste del modelo mediante la creación de una tabla de clasificación donde comparamos los valores observados de la variable a explicar con los valores de la variable a explicar resultado de la predicción. Así podemos ver fácilmente cuántos objetos se asignan correctamente, esto es, vuelven a tomar el valor observado al aplicarse el modelo LOGIT hallado. (Ver 5)

GRUPO DE SEGMENTOS	Predicted Group Membership	Predicted Group Membership	Porcentaje
	No dejaría	Dejaría	
Grupo que No Dejaría el Dinero	54	6	90 %
Grupo que Dejaría el Dinero	12	30	71,4 %
Overall			82,4 %

En nuestro caso, un 82.4% de los encuestados obtienen un valor

predicho de la variable a explicar igual al valor observado. Aunque este porcentaje sea elevado, la función obtenida ajusta mejor en la gente que No dejaría el dinero en su banco actual si hubiera recibido una llamada o comunicación de la institución financiera, que las personas que si dejarían el dinero. Un 90% de los entrevistados que no dejarían el dinero vuelven a su grupo de origen aplicando la función estimada, cuando este porcentaje es solo del 71.4% de las personas que si dejarían el dinero.

Por ejemplo tomemos ipersona con las siguientes características:

**Perfil de la Persona**

- \*V02 = Que no pudo retirar el dinero (0).
- \*V03 = Que el banco no lo llamo (0).
- \*V04 = Que posee vivienda (1).
- \*V05 = Que tiene 5 años de antigüedad (5).
- \*V06 = Que dio tina calificación alta a su Banco (1).
- \*V07 = Que No suprimió las principales operaciones (0).
- \*V08 = Que No ceso las operaciones con su Banco (0).
- \*V09 = Que no cambio (0).
- \*V10 = Que posee celular (1).
- \*V11 = Que tiene ingresos por 2300 pesos mensuales (2300).

(DD = Dejar el Dinero)

$$P(DD) = \frac{e^{-3.827 + 0.264v + 0.996v + 0.094v + 0.035 Vs \dots}}{1 + e^{-3.827 + 0.264v - 0.996y + 0.094v + 0.035}}$$

(NDD = No Dejar el Dinero)

$$P(NDD) = \frac{1}{1 + e^{-3.827 + 0.264v + 0.996y + 0.094v + 0.035}}$$

La Probabilidades de dejar el dinero serian las siguientes:

$$\text{Persona} = -3.827 - 0.264(0) + 0.996(0) + 0.094(1) + 0.035(5) + \\ - 0.093(0) - 1.241(0) + 1.334(1) + 2.150(0) - \\ + 1.548(0) + 4.494(1) + 1.203(1) - 0.001(2300)$$

Persona = 1.173

$$P(x) = \frac{e^{1.173}}{1 + e^{1.173}}$$

$P(x) = 0.7637$ . Por lo tanto la Probabilidad de dejar el dinero que tiene la Persona. De manera que la Probabilidad de No dejar el dinero que tiene el perfil de la Persona es de 0.2363.

$$\text{Odds} = P/1-P = 0.7637/0.2363 = 3.2319$$

En este caso el perfil de la Persona tiene una preferencia por dejar el dinero en el Banco si su entidad lo hubiera llamado para acordar una devolución consensuada de 3.23 veces mayor a no dejarlo.

**CONCLUSION**

Las variables claves o "drivers" utilizados por los segmentos claramente identificados que deberían analizar los Bancos pensando cuales son las variables claves que podrían afectar el comportamiento a futuro de sus clientes y por ende a las Entidades Financieras en sus estrategias comerciales a implementar serian las siguientes:

- \* Aquellos clientes que han sido llamados por el banco para acordar algún tipo de devolución consensuada y posible con los clientes (v3),
- \* Que no han cambiado su relación con el banco (v9),
- \* Que calificaron muy alto a su banco (v6),
- \* Que tiene vivienda propia (v4), que tiene una antigüedad de 5 aflos (v5)
- \* Ingresos medios altos y altos (v11)

Finalmente como conclusión y aporte final podemos observar que las Entidades Financieras deberían utilizar mas operativamente los sistemas de Minería de Datos" (que se define como el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos) con que cuentan, lo que podría ayudar a modelizar el comportamiento predictivo de los actuales, pero fundamentalmente de los potenciales clientes que capte a futuro la banca agrupándolos entre los afectados y no afectados por el corralito financiero, midiendo el grado de confianza sobre el sistema cruzándolo con otras variables cualitativas que le permitan detectar que tipo de perfil tienen para así implementar una eficaz y eficiente técnica de segmentación que permita potenciar el éxito de una ágil estrategia de marketing.

La propuesta seria crear un modelo dentro de un Sistema (fundamentalmente comercial) Inteligente de Datos en una matriz como la que aparece debajo en donde el gerente o personal de atención (con distintos niveles de acceso) puedan tomar decisiones de alto impacto comercial en cada momento de contacto con los clientes (a través de cualquiera de los canales de marketing), en donde el decidor pueda mezclar variables internas y externas por cliente (o segmentos o mercados) para poder efectivizar y desarrollar la relación con los clientes.

MATRIZ DE VARIABLES o DRIVERS	Segmentos				
	Muy Bajo	Bajo	Moderado	Alto	Muy Alto
Gasto Medio Anual (En pesos)	\$ 780	\$ 1.800	\$ 3.600	\$ 7.200	\$ 15.000
Margen por cada Servicio	5%	5%	10%	20%	40%
Numero de visitas al año	3	12	24	48	144
Contribución Bruta por año	19,5	225	1.872	10.224	56.250
Costo de desarrollo y retencion por Cliente/año	156	156	180	220	220
Contribución Neta por año	(132,6)	(115,2)	1.321	9.144	54.600
Duración Media (años)	5	7	12	15	20
Tipo de descuento	0,9091	0,9091	0,9091	0,9091	0,9091
Costo de Captacion	148,2	148,2	148,2	200	200
Valor neto de la Relación	(562)	(370,2)	4.176	32.688	180.000
Nivel de Confianza (de 1 a 7)	1	3	5	6	7
Calidad de la Gerencia (de 1 a 7)	2	2	4	7	7
Precios de los Servicios (de 1 a 7)	6	6	5	4	4
Agilidad para Operar (de 1 a 7)	4	4	3	7	7
Diversidad de los Servicios (de 1 a 7)	4	4	5	7	7
información Disponible (de 1 a 7)	4	4	5	6	6
Ingreso Anual Familiar (En pesos)	\$ 8.400	\$ 12.000	\$ 18.500	\$ 80.000	\$ 200.000

(Muy bajo, bajo, moderado, alto, muy alto, categoría)

## BIBLIOGRAFIA

- \*FRYDMAN, ANDRES *Gestión comercial efectiva en tiempos de cambio. Capitulo 9, paginas 115 y 117. Editorial Ateneo.(2004)*
- \*CACHANOSKY, ROBERTO *Los argentinos, exiliados económicos. Suplemento Económico. Diario La Nación. 14 de noviembre del presente año.(2004)*
- \*MALHOTRA, NARESH *The Use of Linear Logia Modeis in Marketing Research". Journal of Marketing Research. Volumen 21. Febrero. Paginas 20-31. (1984)*
- \*SHARMA, S. *Applied Multivariate Techniques". Chapter 10, pages 317 a 329. John Wiley & Sons. New York.(1996)*
- \*PEREZ, CESAR *Técnicas Estadísticas con SPSS. Capitulo 11, paginas 409 a 415. Prentice Hall, España.(2001)*