



FACULTAD
DE CIENCIAS
ECONÓMICAS



Universidad
Nacional
de Córdoba

REPOSITORIO DIGITAL UNIVERSITARIO (RDU-UNC)

Predicción de fracaso en empresas latinoamericanas utilizando el método del vecino más cercano para predecir efectos aleatorios en modelos mixtos

Norma Patricia Caro, Verónica Arias, Pablo Ortiz

Artículo publicado en Revista de Métodos Cuantitativos para la Economía y la Empresa
Volumen 24, Diciembre 2017 – ISSN: 1886-516X



Este obra está bajo una [Licencia de Creative Commons Reconocimiento-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-sa/4.0/).



Predicción de fracaso en empresas latinoamericanas utilizando el método del vecino más cercano para predecir efectos aleatorios en modelos mixtos

CARO, NORMA PATRICIA

Facultad de Ciencias Económicas

Universidad Nacional de Córdoba (Argentina)

Correo electrónico: pacaro@eco.unc.edu.ar

ARIAS, VERÓNICA

Facultad de Ciencias Económicas

Universidad Nacional de Córdoba (Argentina)

Correo electrónico: varias@eco.unc.edu.ar

ORTIZ, PABLO

Facultad de Ciencias Económicas

Universidad Nacional de Córdoba (Argentina)

Correo electrónico: pabort@eco.uncor.edu

RESUMEN

En la presente década, en economías emergentes como las latinoamericanas, se han comenzado a aplicar modelos logísticos mixtos para predecir el fracaso financiero de las empresas. No obstante, existen limitaciones subyacentes a la metodología, vinculadas a la factibilidad de predicción del estado de nuevas empresas que no han formado parte de la muestra de entrenamiento con la que se estimó el modelo.

En la literatura se han propuesto diversos métodos de predicción para los efectos aleatorios que forman parte de los modelos mixtos, entre ellos, el del vecino más cercano. Este método es aplicado en una segunda etapa, luego de la estimación de un modelo que explica la situación financiera (en crisis o sana) de las empresas mediante la consideración del comportamiento de sus ratios contables.

En el presente trabajo, se consideraron empresas de Argentina, Chile y Perú, estimando los efectos aleatorios que resultaron significativos en la estimación del modelo mixto.

De este modo, se concluye que la aplicación de este método permite identificar empresas con problemas financieros con una tasa de clasificación correcta superior a 80%, lo cual cobra relevancia en la modelación y predicción de este tipo de riesgo.

Palabras claves: fracaso empresarial; ratios contables; modelos mixtos; predicción; vecino más cercano.

Clasificación JEL: M41; C53; C40.

MSC2010: 62H10; 62P05; 91G70.

Artículo recibido el 19 de abril de 2016 y aceptado el 21 de abril de 2017.

Prediction of Failure in Latin-American Companies Using the Nearest-Neighbor Method to Predict Random Effects in Mixed Models

ABSTRACT

In the present decade, in emerging economies such as those in Latin-America, mixed logistic models have been started applying to predict the financial failure of companies. However, there are limitations for the methodology linked to the feasibility of predicting the state of new companies that have not been part of the training sample which was used to estimate the model.

In the literature, several methods have been proposed for predicting random effects in the mixed models such as, for example, the nearest neighbor. This method is applied in a second step, after estimating a model that explains the financial situation (in crisis or healthy) of companies by considering the behavior of its financial ratios.

In this study, companies from Argentina, Chile and Peru were considered, estimating the random effects that were significant in the estimation of the mixed model.

Thus, we conclude that the application of these methods allow for identifying companies with financial problems with a correct classification rate of over 80%, which becomes important in modeling and predicting this risk.

Keywords: business failure; accounting ratios; mixed model; prediction; nearest neighbors.

JEL classification: M41; C53; C40.

MSC2010: 62H10; 62P05; 91G70.



1. Introducción

El fracaso empresarial es un tema de interés en el mundo, particularmente cuando se llega a situaciones desfavorables críticas, como lo es el cierre de una empresa, debido a las importantes consecuencias negativas que puede conllevar tanto desde el punto de vista económico como social.

Esto ha suscitado el interés por anticiparse a tales situaciones a través del diagnóstico de estos eventos, cobrando relevancia la información subyacente de los estados contables.

Precisamente, la información financiera ha sido de fundamental importancia para identificar el estado de crisis de las empresas a partir de las cuales se han desarrollado las diferentes investigaciones. En el presente trabajo, para modelar la situación financiera (en crisis o sana) de las empresas mediante la consideración del comportamiento de sus ratios contables, se aplica un modelo logístico mixto. En estos modelos, los efectos aleatorios son estimados individualmente para cada empresa. Avanzando en este sentido, se plantea como objetivo principal de este trabajo la predicción de los efectos aleatorios, ya que constituyen un problema debido a la dificultad en la clasificación de nuevas empresas (que no forman parte de la muestra de estimación) del estado de crisis financiero de las mismas. A raíz de esto, Tamura y Giampaoli (2013) y Tamura, *et al.* (2013) desarrollaron y aplicaron métodos de estimación que permiten completar el procesamiento y predecir la crisis en nuevas empresas.

En este trabajo se propone avanzar hacia un análisis completo de la vulnerabilidad empresarial de empresas en países latinoamericanos, particularmente sudamericanos, incluyendo la predicción del efecto aleatorio para nuevas unidades. Se aplica el método del vecino más cercano para predecir el efecto aleatorio de nuevas empresas y se compara con los resultados obtenidos cuando se supone que el efecto aleatorio es nulo (método *naive*). El análisis se realiza sobre empresas de Argentina, Chile y Perú. Las economías latinoamericanas han experimentado un sólido crecimiento a lo largo de la década de 2000, impulsado por la sostenida demanda externa (especialmente de economías emergentes como China), en conjunción con un importante crecimiento de los precios de los productos primarios (alcanzando niveles históricos) y con demandas internas muy dinámicas. Estos factores propiciaron un crecimiento de la región a un ritmo anual promedio de casi el 5%

durante el período 2003-2008. Por otro lado, este contexto estuvo acompañado de una adecuada gestión macroeconómica: entre 2000 y 2007, los países de la región pudieron reducir en promedio sus deudas públicas en 15 puntos porcentuales y pasar de un déficit fiscal de 2.4% del PIB a un superávit de 0.4% (OECD, 2010; CEPAL, 2011; OECD/ECLAC, 2011, 2012). La relativa estabilidad macroeconómica alcanzada y la notable mejora de los términos de intercambio en Latinoamérica proveyeron de importantes recursos para que los gobiernos latinoamericanos lleven adelante políticas tendentes a mitigar la pobreza y facilitar el acceso a los servicios públicos. Esta expansión continua fue interrumpida tras la crisis financiera de 2008-2009, aunque la solidez alcanzada en los años anteriores permitió a los gobiernos de la región desplegar una intensa actividad contracíclica, propiciando que el impacto de la crisis fuese menos profundo y la recuperación más acelerada que en los países desarrollados. En el corto plazo, la crisis impactó a los países de Latinoamérica a través de la disminución de los flujos comerciales globales, los cuales pudieron recuperarse apuntaladas por las economías emergentes. No obstante, tras la recuperación económica global a partir de 2010, en los últimos años el ritmo de crecimiento nuevamente mostró debilitamiento, lo que se tradujo en el deterioro de la demanda global con una consecuente caída en los precios de las exportaciones de los países latinoamericanos y una desaceleración de sus economías. Particularmente, a la baja *performance* de las economías desarrolladas se le sumó la ralentización de China e India, que han sido un motor importante para el crecimiento global y de Latinoamérica en particular. Al contexto externo se le suma que la demanda interna no mantuvo su desempeño debido a que, tras las altas tasas de crecimiento de la década, la brecha del producto se ha cerrado en muchos países de la región. A modo de ejemplo, Chile, Colombia y Perú convergen a su producto potencial desde niveles ligeramente superiores; aunque economías como Argentina, Brasil y Venezuela tuvieron una desaceleración que ha llevado sus niveles de actividad incluso por debajo de su potencial.

El análisis se realiza sobre empresas de Argentina, Chile y Perú. La elección de estos países se debe a que son algunos de los más representativos de la región. Los países latinoamericanos no considerados en este estudio se deben a que no poseen Bolsa de Valores (como en el caso Uruguay) o bien no se cuenta con el dato de vulnerabilidad financiera para poder realizar el presente estudio (como es el caso de Brasil).

En la siguiente sección, se realiza una breve presentación de antecedentes. Luego se describen tanto la muestra seleccionada como las variables medidas, para continuar con la aplicación de técnicas de análisis y herramientas estadísticas que nos permitirán evaluar si la información contable es útil para predecir el fracaso empresarial. Finalmente, se exponen los resultados obtenidos, las conclusiones y unas consideraciones generales sobre el trabajo desarrollado.

2. Antecedentes

Los primeros estudios con rigor estadístico se realizaron en el marco del análisis univariado (Beaver, 1968). Altman (1968) incursiona en métodos multivariados capaces de captar la estructura multidimensional de las empresas, centrandó la atención en el análisis discriminante múltiple. En Caro (2014), se presenta una revisión de la literatura presentando modelos de predicción de crisis que se construyeron en economías desarrolladas y en economías emergentes, aplicando diferentes métodos estadísticos.

Es en la década de 1960 cuando surgen los primeros estudios ante la necesidad de contar con modelos que permitan pronosticar las situaciones de crisis financieras (Altman, 1968). En ellos, la metodología aplicada se basaba en un diseño apareado de empresas y se utilizaron métodos de discriminación lineal y cuadrática. En la década de 1980 aparecen los primeros cuestionamientos al diseño apareado, por ser no aleatorio (Ohlson, 1980; Zmijewski, 1984), y se avanza en la modelación con regresión logística o modelo probit.

Debido a la importancia de incorporar la historia de cada empresa, se comenzaron a aplicar modelos para datos longitudinales, como es el modelo logístico mixto, que incorpora en su análisis los balances de cada empresa en un horizonte temporal. En este sentido, Jones y Hensher (2004) demuestran que el modelo logístico mixto supera ampliamente la *performance* del modelo logístico estándar. Jones y Hensher (2007), luego de obtener conclusiones válidas importantes con la aplicación del modelo mixto, completan su primera investigación aplicando el modelo logit anidado multinomial y realizan una comparación con el logit estándar y el logit mixto. Por último, Hensher y Jones (2007) mencionan que los recientes estudios realizados (Jones y Hensher, 2004, 2007) han introducido un sustento teórico y econométrico utilizando modelos de probabilidad avanzados, tales como modelos logit anidados, logit mixtos, multinomial de clase latente y

logit de componente de error, avanzando en el tema aplicando métodos de optimización en la estimación del modelo.

Altman (1968) y Jones y Hensher (2004) son los principales antecedentes en cuanto a esta temática y a la metodología que se aplica. Esta línea de investigación se ha visto enriquecida por los numerosos aportes de diversos autores en torno a nuevos enfoques que han contribuido, por un lado, a mejorar los resultados de los modelos propuestos, reflejada en una importante evolución en el concepto de crisis financiera, las variables empleadas en los modelos y las metodologías (Tascón y Castaño, 2012); y, por otro lado, a investigar el comportamiento de las empresas en economías emergentes (Caro *et al.*, 2013; Caro, 2015).

Altman fue uno de los primeros en adaptar sus modelos a fin de poder aplicarlos en otros países tanto desarrollados como en economías emergentes (Altman *et al.*, 1977). En Latinoamérica surgieron así nuevos trabajos de investigación (Altman *et al.*, 1977; Altman *et al.*, 1979; Pascale, 1988; Swanson y Tybout, 1988; Chocce *et al.*, 2002; Sandin y Porporato, 2008; Zurita, 2008; Montalván *et al.*, 2011; Caro *et al.*, 2013; Caro, 2015) con el objetivo de estudiar el comportamiento de los indicadores contables para predecir el riesgo de insolvencia.

Por un lado, existe una vasta literatura que aplica, en diferentes investigaciones, metodología de corte transversal, lo que no se considera adecuado si se incorporan observaciones de varios períodos de tiempo (estados contables) de cada una de las empresas, donde el supuesto de independencia no se cumple. El interés por abordar esta temática se extiende a países no desarrollados. En un primer lugar, se aplican los modelos de Altman (1968) a economías emergentes sin adaptación alguna, siendo que se trata de ambientes económicos y legales con comportamientos empresariales diferentes. Por tal motivo, Altman (2005) desarrolla el modelo *Emerging Market Scoring (EMS)* como una herramienta que permite determinar una calificación para empresas en economías emergentes a partir de una serie de ajustes a los modelos aplicados en Estados Unidos. De esta forma, un inversor puede evaluar el valor relativo de los créditos en estas economías. En una primera instancia, el modelo Z Score (Altman *et al.*, 1998) se aplicó a empresas mexicanas y luego se practicaron ajustes a efectos de considerar la vulnerabilidad de la empresa a devaluaciones corrientes, al sector al que pertenece y a su posición competitiva en la industria.

En cuanto a los países considerados en el presente trabajo, las principales investigaciones realizadas con aplicación de modelos mixtos han indicado que las ratios financieras de rentabilidad, el flujo de fondos, el endeudamiento, el volumen de ventas y el capital de trabajo han sido significativos para estimar el estado de las empresas (Caro *et al.*, 2013; Caro y Díaz, 2015; Giampaoli *et al.*, 2016; Caro, 2016). No obstante, cabe destacar que ninguno de los modelos construidos, aplicando la metodología que se considera adecuada para este fin, ha podido completar la predicción del estado de fracaso con la finalidad de validar los resultados en nuevas empresas. Esto se debió a que no se había desarrollado un método para ello, siendo Tamura *et al.* (2013) y Tamura y Giampaoli (2013)¹ quienes comienzan con el desarrollo de estos métodos.

3. Muestra y variables

Las unidades de análisis son empresas que cotizan en las Bolsas de Valores respectivas de cada uno de los países que constituyen la presente investigación: Argentina, Chile y Perú, durante el período comprendido entre 2000 y 2011². Del total de empresas cotizantes se seleccionó una muestra representativa de la población para constituir una muestra de entrenamiento, utilizada para ajustar el modelo logístico mixto, y una muestra de validación para la predicción sobre nuevas unidades.

Otro criterio tomado en cuenta en la selección de la muestra, es la inclusión de empresas en crisis y empresas sanas. Por ello, la muestra cuenta con empresas que informan el estado de crisis financiera, ya sea por suspender la cotización de sus acciones o bien por cotizar en ruedas reducidas (sanción que la Bolsa le impone cuando poseen signos de vulnerabilidad financiera, como altas pérdidas o cesación de pagos, entre otras) o encontrarse en procesos judiciales de concurso preventivo o de quiebra. Dado que la cantidad de empresas en crisis no es muy numerosa, se ha incorporado a la mayoría de ellas en la muestra, mientras que las empresas sanas constituyeron una muestra aleatoria del total de empresas sanas de la población.

¹Presentado en el *28th International Workshop on Statistical Modeling*.

² En el caso de Argentina, el horizonte temporal fue desde 2003 a 2011, a fin de no incluir la crisis del 2001 y unificar el análisis en un período de estabilidad económica.

Para cada una de las empresas en crisis se tomaron como mínimo dos años previos a la manifestación del estado de crisis, contemplando hasta seis períodos, según la disponibilidad de la información.

A los fines de la estimación del estado de las empresas, las mismas se consideraron según presentasen crisis financiera o no. Este estado (en crisis o sanas) constituye la variable dependiente (véase Tabla 1).

Tabla 1. Cantidad de empresas en cada muestra por país y estado

| Países | Entrenamiento | | Validación | |
|-----------|---------------|-----------|------------|-----------|
| | Sanas | En crisis | Sanas | En crisis |
| Argentina | 31 | 13 | 8 | 4 |
| Chile | 89 | 22 | 37 | 11 |
| Perú | 37 | 13 | 39 | 1 |

Si bien la cantidad de empresas cotizantes en países latinoamericanos no es numerosa, el hecho de trabajar con datos anuales de cada empresa, incrementa la cantidad de observaciones para estimar los modelos.

Las variables independientes utilizadas, siguiendo a Caro *et al.* (2013), son los ratios financieras que se calculan con los estados contables de publicación al cierre de ejercicio (véase Tabla 2).

Tabla 2. Descripción de las variables utilizadas en el modelo

| Variable | Forma de cálculo | Elementos del Ratio | Interpretación |
|---------------------------------------------|-------------------------|---------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ratio de flujo de fondos (FF_AT) | $FF_AT = (FFO/AT)*100$ | FFO: flujo de fondos generado por las operaciones AT: activo total | Mide la posición de efectivo originado por las operaciones respecto al total del activo. |
| Ratio de liquidez (E_AT) | $E_AT = (E/AT)*100$ | E: disponibilidades (efectivo e inversiones que se convierten en efectivo dentro de las 24 hs.) AT: activo total | Representa la relación entre los recursos más líquidos de la empresa con respecto al total del activo. |
| Ratio de rotación sobre activo total (V_AT) | $V_AT = (V/AT)*100$ | V: total de ventas AT: activo total | Expresa la eficiencia con que la empresa puede usar sus activos para generar ventas. |
| Ratio de endeudamiento (D_PN) | $D_PN = (D/PN)*100$ | D: total de deudas PN: patrimonio neto | Mide la relación de las deudas (capital de terceros) por cada peso de capital aportado por los dueños de la empresa. |
| Ratio de capital de trabajo (CT_AT) | $CT_AT = (CT/AT)*100$ | CT: capital de trabajo (activo corriente menos pasivo corriente) AT: activo total | Expresa el grado de fluidez del activo total e indica en qué medida la inmovilización del activo total resulta neutralizada por el capital de trabajo. |
| Ratio de rentabilidad económica (GE_AT) | $GE_AT = (Uai/AT)*100$ | Uai: utilidad antes de intereses e impuestos AT: activo total | Mide la rentabilidad en función de las ganancias de la explotación. |

4. Método

Con el objetivo de modelar la situación financiera (en crisis o sana) de las empresas mediante la consideración del comportamiento de sus ratios contables, se utilizó un modelo logístico mixto.

Los problemas con respuesta dicotómica son modelados usualmente a través de regresión logística, donde la probabilidad de que la variable respuesta asuma el valor 1 (empresa en crisis) es:

$$E(y_i / x_i) = \Pr(y_i = 1 / x_i) \quad i = (1, 2, \dots, n) \quad (1)$$

$$\Pr(y_i = 1 / x_i) = h(\beta_1 + \beta_2 x_i) = h(z_i) \quad i = (1, 2, \dots, n) \quad (2)$$

donde β_i son los coeficientes de la regresión.

La transformación logit es:

$$\text{logit} \{ \Pr(y_i = 1 / \mathbf{x}_i) \} = \ln \left\{ \frac{\Pr(y_i = 1 / \mathbf{x}_i)}{1 - \Pr(y_i = 1 / \mathbf{x}_i)} \right\} = z_i. \quad (3)$$

donde:

y_i : variable que representa el estado (1: en crisis; 0: sana);

\mathbf{x}_i : covariables.

En el presente trabajo, los datos utilizados son las observaciones y mediciones presentadas por cada empresa a través del tiempo, constituyendo un panel de datos (datos longitudinales). Estos cuentan con una estructura jerárquica (empresas medidas repetidamente en el tiempo), sin cumplirse el supuesto de independencia necesario en la regresión logística tradicional. Esta situación se resuelve con la utilización del modelo logístico mixto, en el que se distinguen efectos fijos y aleatorios. Los fijos tienen como finalidad comparar los resultados de la variable dependiente para los distintos valores de las variables independientes; mientras que los efectos aleatorios analizan la variabilidad no explicada por la respuesta.

La estructura del modelo es:

$$Y_{ij} = \mathbf{X}'_{ij} \boldsymbol{\beta} + \mathbf{Z}'_{ij} \mathbf{b}_i + e_{ij} \quad (4)$$

donde el primer término hace referencia a los efectos fijos, cuyo vector $\boldsymbol{\beta}$ está formado por los parámetros de las variables independientes (\mathbf{X}_i) que se estiman; el segundo término

corresponde a los efectos aleatorios (\mathbf{b}), con su matriz de diseño (\mathbf{Z}_i), que siguen una distribución normal donde se explicita una estructura de varianza (G):

$$\mathbf{b}_j \sim N(\mathbf{0}, G) \quad (5)$$

y, por último, el término de error asociado a todo modelo.

En cada uno de los mercados analizados, se aplica un modelo logístico mixto con dos efectos aleatorios. En estos modelos, los efectos aleatorios son estimados individualmente para cada empresa, por lo que no es posible hacer predicción directamente para el caso de nuevas unidades, pues no se conocen esos valores.

El modelo aplicado en este trabajo es:

$$Y_{ij} = \beta_0 + (\beta_1 + b_{1j})GE_{AT_{ij}} + (\beta_2 + b_{2j})FF_{AT_{ij}} + \beta_3E_{AT_{ij}} + \beta_4V_{AT_{ij}} + \beta_5D_{PN_{ij}} + \beta_6CT_{AT_{ij}} \quad (6)$$

donde:

β_k es el coeficiente que representa los k efectos fijos del modelo, siendo k el número de variables independientes X_k ;

b_{kj} es el k -ésimo efecto aleatorio de la empresa j ;

X_{kij} es la k -ésima variable independiente correspondiente al período i de la empresa j . Por ejemplo, $X_{1ij} = GE_{AT_{ij}}$ es el índice de rentabilidad en el período i de la empresa j , y así con cada uno de los indicadores mencionados en la Tabla 2;

ε_{ij} es el término de error.

Los efectos aleatorios siguen una distribución normal:

$$\mathbf{b}_j = \begin{bmatrix} b_{1j} \\ b_{2j} \end{bmatrix} / \mathbf{x}_{ij} \sim N_2(0, \Psi) ; \Psi = \begin{pmatrix} d_{11} & 0 \\ 0 & d_{22} \end{pmatrix} \quad (7)$$

Nuevos métodos fueron propuestos por Tamura *et al.* (2013) y Tamura y Giampaoli (2013) para predecir los efectos aleatorios de nuevas observaciones. Uno de ellos es el método de predicción por regresión lineal (LRPM) desarrollado por Tamura *et al.* (2013) para un modelo logístico mixto con k efectos aleatorios. Esta metodología considera el ajuste de un modelo de regresión para predecir los efectos aleatorios con base al conjunto de datos de la muestra de entrenamiento.

Este modelo considera que la variable respuesta son los efectos aleatorios que se pretenden predecir ($\hat{\alpha}_i$), de cada i -ésima empresa de la muestra de entrenamiento. En lo sucesivo, los parámetros estimados de esos modelos de regresión son usados para predecir el efecto aleatorio para una nueva empresa (l), que no pertenece al conjunto de entrenamiento.

Luego de obtener las estimaciones, a través del modelo mixto, es necesario que las covariables disponibles a nivel de observación (período anual) sean agregadas a nivel de grupo (empresa). Para ello, se puede calcular el promedio, la mediana o bien tomar el valor mínimo o el máximo.

Se considera, para cada m -ésimo efecto aleatorio estimado, un modelo capaz de explicar la relación entre las covariables y los efectos aleatorios de la forma:

$$\hat{\alpha}_{mi} = w_{mi}^t \lambda_m + \varepsilon_{mi} \quad \text{con } \varepsilon_{mi} \sim N(0, \sigma_m^2) \text{ independientes}; \quad (8)$$

donde:

λ_m es el vector de coeficientes de regresión desconocidos;

w_{mi} es el vector de covariables de cada empresa y cada efecto aleatorio, agregado a nivel de empresa.

Así, es posible predecir la probabilidad de respuesta de un j -ésima observación dentro de una m -ésima nueva empresa, usando la función logística del modelo mixto.

Este método no se presenta en la aplicación empírica debido a la falta de cumplimiento de los supuestos de la regresión lineal múltiple (independencia de residuos, normalidad y homoscedasticidad).

Otro método propuesto por Tamura *et al.* (2013), que es el que se aplica en este trabajo, es el método de predicción del vecino más cercano considerando el modelo logístico mixto con k efectos aleatorios. La ventaja de esta técnica es que no requiere distribución alguna para los efectos aleatorios.

Este método es una técnica no paramétrica, comúnmente utilizada en clasificación supervisada. Para la predicción de los efectos aleatorios de las nuevas empresas, se consideran los valores conocidos de estos efectos de sus vecinos más cercanos.

Al tratarse de una variable continua, es necesario considerar alguna medida de centralidad (media, mediana u otra) para realizar la imputación del valor correspondiente.

Se selecciona la cantidad de vecinos a considerar de manera que se maximice la *performance* de predicción del modelo mixto.

Siguiendo a Tamura *et al.* (2013) se calculan las distancias (euclídea, de Mahalanobis, *City Block* u otra) entre las nuevas empresas (base de validación) y aquellas con efectos aleatorios conocidos (base de entrenamiento), considerando su vector de características \mathbf{g} .

Para $i'=1, \dots, q'$, donde $i' \in Q'$ (siendo Q' el conjunto de empresas de la base de validación) e $i=1, \dots, q$, donde $i \in Q$ (siendo Q el conjunto de empresas de la base de entrenamiento), se computan las distancias $d(i', i)$ entre $g_{i'}$ y g_i , y luego se ordenan las distancias de manera creciente:

$$d(i', .) = (d(i', 1.), d(i', 2.), \dots, d(i', q)) \quad (9)$$

Posteriormente se computan medidas de centralidad para los efectos aleatorios conocidos correspondientes a los l primeros elementos de las distancias ordenadas:

$$\hat{\alpha}_{mi'}^* = (\hat{\alpha}_{m1}, \hat{\alpha}_{m2}, \dots, \hat{\alpha}_{ml}) \quad (10)$$

Finalmente, se introducen los efectos aleatorios en el predictor lineal de la regresión logística mixta, haciendo posible la predicción de la probabilidad de crisis para la nueva empresa.

5. Resultados

En una primera etapa, se realiza un análisis descriptivo de las empresas que constituyen la muestra, lo que se refleja en la Tabla 3. Las empresas en crisis presentan menores niveles de rentabilidad, flujo de fondos, efectivo, volumen de negocio, capital de trabajo, en promedio, que las empresas sanas. Por otro lado, el nivel de endeudamiento promedio es mayor en las empresas en crisis. Estos indicadores poseen valores extremos, lo que hace que su variabilidad sea elevada; por ello, como medida representativa se analizaron las medianas obteniendo iguales conclusiones entre empresas con problemas y sin ellos.

Tabla 3. Análisis descriptivo de las variables utilizadas en el modelo logístico mixto según país y estado de las empresas (muestra de entrenamiento)

| Variable | | Argentina | | Chile | | Perú | |
|----------|--------------|-----------|-----------|---------|-----------|--------|-----------|
| | | Sanas | En crisis | Sanas | En crisis | Sanas | En crisis |
| GE_AT | Media | 1,24 | -10,71 | 6,41 | -10,11 | 17,66 | 0,74 |
| | Desv. típica | 6,59 | 23,57 | 10,74 | 31,31 | 17,18 | 6,41 |
| | Mediana | 0,80 | -8,30 | 5,94 | -1,36 | 13,05 | 0,78 |
| | Min. | -17,53 | -128,54 | -147,35 | -271,55 | -27,76 | -21,95 |
| | Max. | 38,25 | 36,71 | 85,99 | 21,80 | 90,57 | 20,41 |
| FF_AT | Media | 7,25 | -1,22 | 9,36 | -9,28 | 14,39 | 2,75 |
| | Desv. típica | 9,60 | 23,05 | 9,82 | 36,92 | 13,23 | 4,87 |
| | Mediana | 7,54 | 3,55 | 8,27 | -0,72 | 11,95 | 1,87 |
| | Min. | -64,51 | -126,99 | -29,89 | -266,70 | -21,41 | -10,05 |
| | Max. | 34,58 | 25,12 | 126,01 | 83,33 | 64,74 | 17,65 |
| E_AT | Media | 2,46 | 1,95 | 1,07 | 1,01 | 9,30 | 1,19 |
| | Desv. típica | 3,94 | 4,49 | 2,72 | 2,61 | 12,51 | 2,23 |
| | Mediana | 1,39 | 0,55 | 0,43 | 0,36 | 3,98 | 0,48 |
| | Min. | 0,01 | 0,00 | 0,00 | 0,00 | 0,11 | 0,00 |
| | Max. | 51,86 | 31,38 | 34,07 | 25,73 | 64,12 | 13,72 |
| V_AT | Media | 86,24 | 35,83 | 57,64 | 33,04 | 78,43 | 44,91 |
| | Desv. típica | 68,54 | 39,66 | 53,89 | 38,12 | 47,79 | 56,96 |
| | Mediana | 73,68 | 23,57 | 50,71 | 23,39 | 68,56 | 23,98 |
| | Min. | 0,06 | -12,46 | 0,00 | 0,00 | 2,81 | 1,80 |
| | Max. | 374,85 | 117,35 | 85,99 | 21,80 | 285,67 | 286,41 |
| D_PN | Media | 82,11 | 241,31 | 70,36 | 371,66 | 67,88 | 311,31 |
| | Desv. típica | 68,49 | 373,87 | 58,43 | 2018,65 | 45,68 | 908,67 |
| | Mediana | 64,62 | 133,52 | 56,97 | 64,87 | 60,22 | 87,01 |
| | Min. | 0,02 | 0,25 | 0,06 | 0,04 | 5,17 | 26,46 |
| | Max. | 451,88 | 2020,31 | 419,94 | 21167,50 | 219,23 | 6491,88 |
| CT_AT | Media | 10,72 | -24,25 | 11,98 | 11,69 | 18,05 | -3,93 |
| | Desv. típica | 20,66 | 71,09 | 14,39 | 36,63 | 18,39 | 19,28 |
| | Mediana | 7,12 | -0,31 | 7,84 | 4,46 | 14,00 | -0,25 |
| | Min. | -65,27 | -263,85 | -18,77 | -90,85 | -32,98 | -66,37 |
| | Max. | 66,35 | 44,49 | 93,09 | 99,95 | 68,51 | 36,25 |

La Tabla 4 muestra los resultados obtenidos con el modelo logístico mixto para cada uno de los países (Argentina, Chile y Perú). La rentabilidad económica (GE_AT), con signo negativo, fue el único ratio que resultó estadísticamente significativo en los tres países. El flujo de fondos (FF_AT) resultó significativo en Argentina y Chile, mientras que la liquidez (E_AT) lo fue en Perú. Tanto la rotación del activo total (V_AT) como el endeudamiento (D_PN) fueron significativos para el caso Argentina y Chile, siendo este último ratio significativo también para Perú. Finalmente, el capital de trabajo (CT_AT) resulta un ratio estadísticamente significativo en Chile y Perú, pero no así en Argentina.

Sólo los índices de rentabilidad GE_AT y de flujo de fondos operativo FF_AT resultaron tener una amplia variabilidad y ser significativos para representar la

heterogeneidad no observada (véase Tabla 5). Los datos fueron procesados con el procedimiento nlmixed del *software* SAS³.

Tabla 4. Resultados del modelo mixto para cada país (realizado a partir de la información de las empresas que conforman las bases de entrenamiento)

| Variable | Argentina | | Chile | | Perú | |
|-----------|----------------------------|----------|----------------------------|----------|----------------------------|---------|
| | Estimador (error estándar) | p-valor | Estimador (error estándar) | p-valor | Estimador (error estándar) | p-valor |
| Constante | 1,227 (0,901) | 0,181 | -1,405 (0,449) | 0,002*** | 1,519 (1,178) | 0,203 |
| GE_AT | -0,162 (0,092) | 0,086* | -1,225 (0,314) | 0,000*** | -0,591 (0,235) | 0,016** |
| FF_AT | -0,269 (0,121) | 0,032** | -0,425 (0,123) | 0,001*** | -0,132 (0,141) | 0,356 |
| E_AT | 0,364 (0,233) | 0,126 | -0,092 (0,075) | 0,224 | -0,536 (0,255) | 0,041** |
| V_AT | -0,076 (0,022) | 0,001*** | -0,010 (0,005) | 0,079* | -0,012 (0,016) | 0,478 |
| D_PN | 0,015 (0,007) | 0,031** | 0,011 (0,004) | 0,005*** | 0,022 (0,012) | 0,080* |
| CT_AT | 0,008 (0,011) | 0,481 | 0,043 (0,016) | 0,009*** | -0,169 (0,072) | 0,022** |

Efectos aleatorios para Argentina, Perú y Chile (GE_AT, FF_AT).

Nota: error estándar entre paréntesis.

*** Significativo al 0,01; ** Significativo al 0,05; * Significativo al 0,10.

Tabla 5. Pruebas de razón de verosimilitud para determinar significatividad de los efectos aleatorios para cada país (realizado a partir de la información de las empresas que conforman las bases de entrenamiento)

| Modelo propuesto | Argentina | | | Chile | | | Perú | | |
|----------------------------|----------------------|------|---------|----------------------|-------|---------|----------------------|------|---------|
| | -2log(verosimilitud) | Dif* | p-valor | -2log(verosimilitud) | Dif* | p-valor | -2log(verosimilitud) | Dif* | p-valor |
| (A) Un efecto aleatorio | 112,7 | | | 276,6 | | | 85,2 | | |
| (B) Dos efectos aleatorios | 105,3 | 7,4 | 0,0033 | 230,9 | 45,52 | | 76,5 | 8,7 | 0,0032 |

*Dif: diferencia entre $-2\log(\text{verosimilitud})$ del modelo propuesto (B) con dos efectos aleatorios, GE_AT y FF_AT, y del modelo de referencia (A) con un efecto aleatorio, GE_AT.

Para la predicción del efecto aleatorio para nuevas empresas se probó el método de regresión lineal, pero no resultó apropiado para el problema considerado en el presente trabajo, ya que las covariables disponibles no resultaron significativas y no se cumplían los supuestos del modelo de regresión lineal (normalidad de los errores, independencia y homoscedasticidad).

³ Statistical Analysis System (sistema integrado realizado por SAS Institute Inc.)

Se empleó el método de predicción del vecino más cercano basado en la distancia euclídea de las nuevas empresas a los vectores de covariables (ratios) de las que contribuyeron a la construcción del modelo mixto, agregando los datos de distintos períodos a nivel de empresas a través de la mediana.

Se definió como medida de centralidad (predicción) de los efectos aleatorios de las nuevas empresas correspondientes a la base de validación, la media de los efectos aleatorios conocidos de las empresas de la base de entrenamiento elegidas como vecinos más cercanos de las empresas nuevas.

Para medir la *performance* del modelo, en la clasificación de nuevas empresas *en sanas* o *en crisis*, se comparó la clasificación binaria observada versus la clasificación estimada por el modelo en el punto de corte. El punto de corte fue definido como aquel donde la sensibilidad iguala a la especificidad en la base de entrenamiento de cada país. En base a esto se calcularon medidas de sensibilidad (verdaderos positivos entre la suma de falsos negativos y verdaderos positivos), especificidad (verdaderos negativos entre la suma de verdaderos negativos y falsos positivos) y tasa de clasificación correcta (la división de la suma de verdaderos positivos y verdaderos negativos entre el total de observaciones).

Estas medidas fueron calculadas según la clasificación de empresas de la base de validación que surgió del modelo con los efectos aleatorios estimados por el método del vecino más cercano y para el modelo que considera solo efectos fijos (sin estimación de los efectos aleatorios) o método *naive* (véase Tabla 6).

En todos los casos, la tasa de clasificación correcta fue muy buena, superando el 80% en el caso de Argentina y Perú, y el 90% en Chile. En el caso de Perú, el hecho de que la muestra de validación solo posea una única empresa en crisis hace que sea difícil detectarla. Para el caso de Argentina, la clasificación de las empresas nuevas utilizando el modelo con el efecto aleatorio estimado mediante el método del vecino más cercano, mejoró la especificidad y la tasa de clasificación correcta. En el caso de Chile, las medidas se mantienen entre uno y otro método. No es conveniente interpretar estas tasas en Perú, debido a que se trata de una única empresa en crisis en la muestra de validación.

De este modo, el método del vecino más cercano resulta útil para clasificar una nueva observación (empresa). Adicionalmente, cabe remarcar la pertinencia de trabajar de manera separada con cada país, dada las diferencias obtenidas en las estimaciones que

resultaron significativos en cada uno de ellos, aunque sus efectos aleatorios fueron los mismos en todos los países.

Tabla 6. *Medidas de performance del modelo por país (bases de validación)*

| Método | País | Argentina | Chile | Perú |
|--------------------|--------------------------------|-------------|-------------|-------------------|
| Vecino más cercano | Especificidad | 0,75 | 0,89 | 0,85 |
| | Sensibilidad | 1 | 0,91 | 0 |
| | Tasa de clasificación correcta | 0,83 | 0,90 | 0,85 ⁴ |
| Naive | Especificidad | 0,75 | 0,89 | 0,92 |
| | Sensibilidad | 0,75 | 0,91 | 0 |
| | Tasa de clasificación correcta | 0,75 | 0,90 | 0,90 |

6. Conclusiones

La necesidad de comparabilidad de los ratios financieros de diferentes empresas conduce a utilizar información de aquellas que cotizan en bolsa. Estas empresas tienen especiales exigencias en cuanto a la presentación y disponibilidad de información, lo que no es así en otro tipo de empresas.

Con respecto a los ratios utilizados y los resultados obtenidos, todos fueron significativos en al menos alguno de los países que constituyeron este estudio.

El ratio de rentabilidad económica (GE_AT) es uno de los indicadores que, a través de la literatura, ha resultado significativo para predecir el estado de crisis financiera de las empresas, lo que se confirma en esta investigación también para todos los países. Esto significa que la utilidades obtenidas respecto al total del activo manifiestan que a menores utilidades (o pérdidas), más posibilidad tiene la empresa de estar en crisis.

Respecto al índice de flujo de fondos (FF_AT), que constituye un indicador relativamente nuevo en esta área, indica que a menor nivel de flujo de fondos generados por

⁴ Valor aceptable, aunque no comparable con lo obtenido en otros métodos debido a que se trata de una única empresa en crisis en la muestra de validación.

las operaciones respecto del activo total existen más posibilidades de que las empresas estén en crisis. No obstante, este indicador solo en Argentina y Chile resultó significativo.

El índice de liquidez, si bien no siempre refleja signos de vulnerabilidad, sí lo hizo en Perú, donde a menor liquidez, mayor posibilidad de crisis financiera.

Otros de los indicadores que han reflejado un comportamiento determinado en las empresas son la rotación del activo (nivel de ventas en relación al total de activos) y el endeudamiento (deudas respecto al patrimonio neto), los cuales fueron significativos en la mayoría de los países.

Se confirma que los ratios seleccionados, en consistencia con la literatura mencionada, reflejan en mayor o menor medida el comportamiento de las empresas y permiten un diagnóstico eficaz de la situación de crisis financiera.

Es de destacar que los índices de rentabilidad y de flujo de fondos operativos resultaron ser efectos aleatorios significativos, que explican la heterogeneidad no observada. Si bien no hay una teoría de ratios que indique cuáles son los que se deberían utilizar para este objetivo de predecir el estado de vulnerabilidad financiera, los ratios utilizados, propuestos por la literatura, son los adecuados.

A efectos de predecir el estado de una nueva empresa, que no ha formado parte de la base de entrenamiento con la que se estimó el modelo mixto, se empleó para la estimación de los efectos aleatorios el método del vecino más cercano. La literatura avanza hacia métodos no paramétricos como éste, ya que los métodos paramétricos, como el de regresión múltiple, no siempre es adecuado por la falta de cumplimiento de los supuestos estadísticos del modelo, permitiendo así la clasificación de una nueva empresa. De esta manera se levanta una limitación que tenían los modelos mixtos en cuanto a la estimación y predicción de una nueva observación.

Estos métodos contribuyen a la estimación del estado de vulnerabilidad en la que puede encontrarse una nueva empresa, en el mercado a través de la estimación de los efectos fijos y la predicción de los efectos aleatorios. La metodología utilizada hasta el momento en esta área de interés, como lo es la predicción de crisis financiera en empresas, limitaba este segundo objetivo, que a partir de las investigaciones realizadas ha podido ser superado. Ampliando el horizonte de análisis y disponiendo de una mayor cantidad de empresas vulnerables se podrán obtener mejores tasas de clasificación correcta.

Referencias

- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.
- Altman, E. I. (2005). An emerging market credit scoring system for corporate bonds. *Emerging Markets Review*, 6(4), 311-323.
- Altman, E. I.; Baidya, T. K. y Dias, L. M. R. (1979). Assessing potential financial problems for firms in Brazil. *Journal of International Business Studies*, 10(2), 9-24.
- Altman, E. I.; Haldeman, R. G. y Narayanan, P. (1977). ZETATM analysis A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, 1(1), 29-54.
- Altman, E. I.; Hartzell, J. y Peck, M. (1998). Emerging Markets Corporate Bonds: A Scoring System. En Levich R.M. (ed.): *Emerging Market Capital Flows*. The New York University Salomon Center Series on Financial Markets and Institutions, Vol. 2, Boston: Springer, pp. 391-400.
- Beaver, W. H. (1968). Alternative accounting measures as predictors of failure. *The Accounting Review*, 43(1), 113-122.
- Caro, N. P. (2014). Modelos de predicción de crisis financiera en empresas: una revisión de la literatura. *Revista Internacional Legis de Contabilidad y Auditoría*, 58, 135-183.
- Caro, N. P. (2015). Descripción de empresas en crisis financiera: el caso de Argentina en las décadas del 1990 y 2000. *Revista de Dirección y Administración de Empresas*, 22, 106-130.
- Caro, N. P. (2016) El modelo logístico mixto para predecir crisis financiera en empresas argentinas y chilenas. *Revista de Matemática: Teoría y Aplicaciones*, 23(1), 255-276.
- Caro, N. P.; Díaz, M. y Porporato, M. (2013). Predicción de quiebras empresariales en economías emergentes: uso de un modelo logístico mixto. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 16, 200-215.
- Caro, N. P. y Diaz, M. (2015). Factores determinantes para predecir la crisis financiera en empresas argentinas. *Cuadernos de Estudios Empresariales*, 25, 29-47.

- CEPAL. (2011). *Balance preliminar de las economías de América Latina y el Caribe 2010*. Santiago de Chile: Naciones Unidas CEPAL.
- Chocce, G. A. R.; González, P. A.; Aguirre, N. A.; Vega, P. L. y Carrazana, J. M. (2002). Modelos de clasificación y predicción de quiebra de empresas: Una aplicación a empresas chilenas. *Revista Fórum Empresarial*, 7(1), 33-52.
- Giampaoli, V.; Tamura, K. A.; Caro, N. P. y de Araujo, L.J.S. (2016). Financial crisis prediction on Latin American companies using mixed logistic regression model. *Chilean Journal of Statistics*, 7(1), 31-41
- Hensher, D. A. y Jones, S. (2007). Forecasting corporate bankruptcy: Optimizing the performance of the mixed logit model. *Abacus*, 43(3), 241-264.
- Jones, S. y Hensher, D. A. (2004). Predicting firm financial distress: A mixed logit model. *The Accounting Review*, 79(4), 1011-1038.
- Jones, S. y Hensher, D. A. (2007). Modelling corporate failure: A multinomial nested logit analysis for unordered outcomes. *The British Accounting Review*, 39(1), 89-107.
- Montalván, S. M.; O'Shee, D. F.; Delgado, F. A. y Yamashiro, M. A. (2011). Determinantes de la insolvencia empresarial en el Perú. *Academia, Revista Latinoamericana de Administración*, 47, 126-139.
- OECD (2010). *Perspectivas Económicas de América Latina 2011: En qué medida es clase media América Latina*. París: OECD Publishing, doi: <http://dx.doi.org/10.1787/leo-2011-es>.
- OECD/ECLAC (2011). *Perspectivas Económicas de América Latina 2012: Transformación del Estado para el Desarrollo*. París: OECD Publishing, doi: <http://dx.doi.org/10.1787/leo-2012-es>.
- OECD/ECLAC (2012). *Perspectivas económicas de América Latina 2013: Políticas de pymes para el cambio estructural*. París: OECD Publishing, doi: <http://dx.doi.org/10.1787/leo-2013-es>.
- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109-131.
- Pascale, R. (1988). A multivariate model to predict firm financial problems: the case of Uruguay. *Studies in Banking and Finance*, 7, 171-182.

- Sandin, A. R. y Porporato, M. (2008). Corporate bankruptcy prediction models applied to emerging economies: Evidence from Argentina in the years 1991-1998. *International Journal of Commerce and Management*, 17(4), 295-311.
- Swanson, E. y Tybout, J. (1988). Industrial bankruptcy determinants in Argentina. *Studies in Banking and Finance*, 7, 1-25.
- Tamura, K. A. y Giampaoli, V. (2013). New prediction method for the mixed logistic model applied in a marketing problem. *Computational Statistics & Data Analysis*, 66, 202-216.
- Tamura, K. A.; Giampaoli, V. y Noma, A. (2013). Nearest Neighbors Prediction Method for mixed logistic regression. En Muggeo, M. R.; Capursi, V.; Boscaino, G. y Lovison, G. (eds.): *Proceedings of the 28th International Workshop on Statistical Modeling*, Palermo: Università di Palermo, pp. 799-802.
- Tascón, M. y Castaño, F. (2012). Variables y modelos para identificación y predicción del fracaso empresarial: Revisión de la investigación reciente. *Revista de Contabilidad*, 15(1), 7-58.
- Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59-82.
- Zurita, F. (2008). La predicción de la insolvencia de empresas chilenas. *Economía Chilena*, 11(1), 93-116.