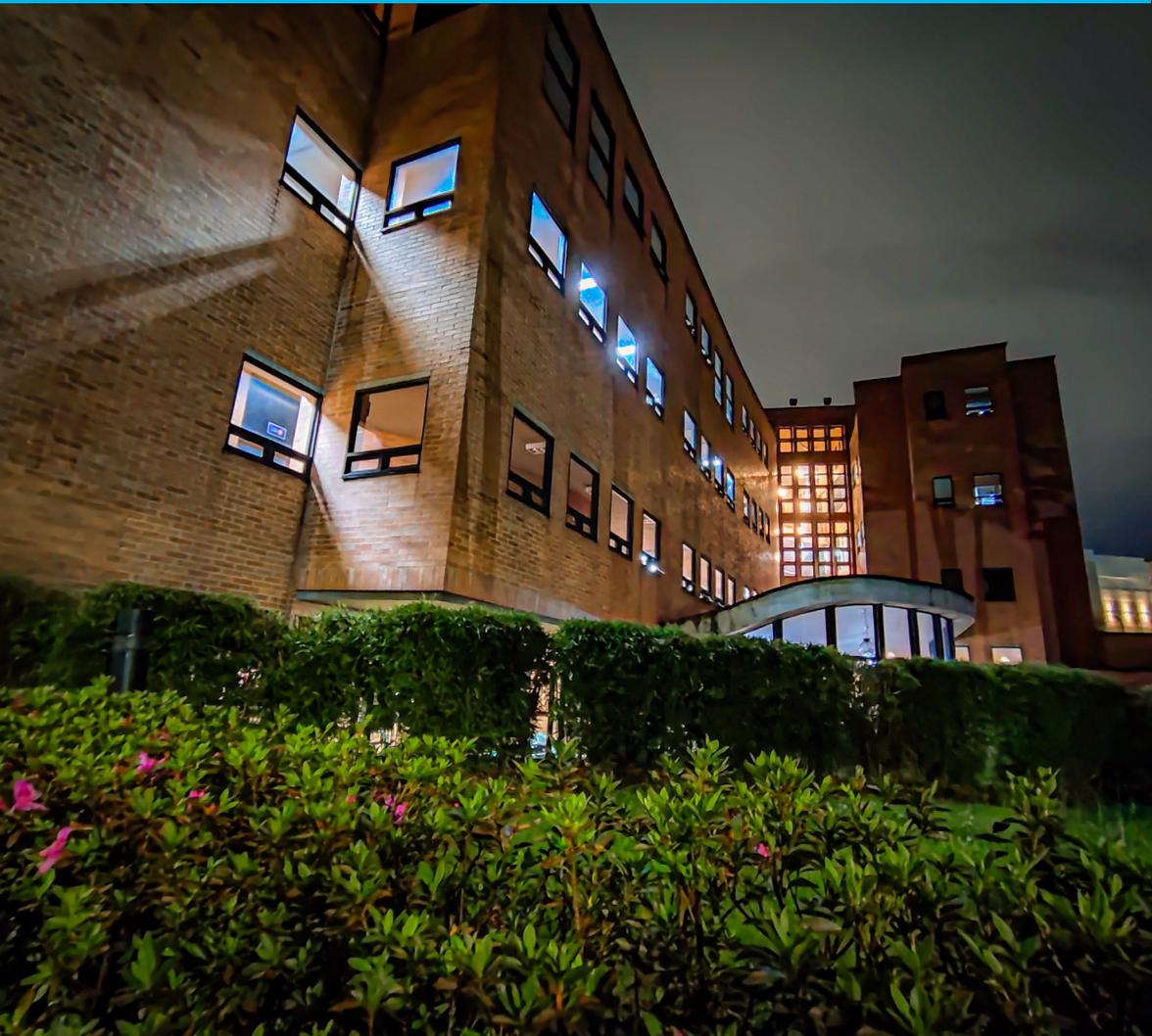


CUADERNOS DE ECONOMÍA

ISSN 0121-4772



Facultad de Ciencias Económicas
Escuela de Economía
Sede Bogotá



UNIVERSIDAD
NACIONAL
DE COLOMBIA

ASESORES EXTERNOS

COMITÉ CIENTÍFICO

Ernesto Cárdenas

Pontificia Universidad Javeriana-Cal

José Félix Cataño

Universidad de los Andes

Philippe De Lombaerde

NEOMA Business School y UNU-CRIS

Edith Klímovsky

Universidad Autónoma Metropolitana de México

José Manuel Menudo

Universidad Pablo de Olavide

Gabriel Mísas

Universidad Nacional de Colombia

Mauricio Pérez Salazar

Universidad Externado de Colombia

Fábio Waltenberg

Universidade Federal Fluminense de Rio de Janeiro

EQUIPO EDITORIAL

Daniela Cárdenas

Karen Tatiana Rodríguez

María Paula Moreno Mojca

Estudiante auxiliar

Proceditor Ltda.

Corrección de estilo, armada electrónica,
finalización de arte, impresión y acabados
Tel. 757 9200, Bogotá D. C.

Luis Tarapuez

Área de Comunicaciones - Facultad de Ciencias Económicas

Fotografía de la cubierta

Indexación, resúmenes o referencias en

SCOPUS

Thomson Reuters Web of Science

(antiguo ISI)-SciELO Citation Index

ESCI (Emerging Sources Citation Index) - Clarivate Analytics

EBSCO

Publindex - Categoría B - Colciencias

SciELO Social Sciences - Brasil

RePEc - Research Papers in Economics

SSRN - Social Sciences Research Network

EconLit - Journal of Economic Literature

IBSS - International Bibliography of the Social Sciences

PAIS International - CSA Public Affairs Information Service

CLASE - Citas Latinoamericanas en Ciencias Sociales y Humanidades

Latindex - Sistema regional de información en línea

HLAS - Handbook of Latin American Studies

DOAJ - Directory of Open Access Journals

CAPEs - Portal Brasileiro de Informação Científica

CIBERA - Biblioteca Virtual Iberoamericana España / Portugal

DIALNET - Hemeroteca Virtual

Ulrich's Directory

DOTEC - Documentos Técnicos en Economía - Colombia

LatAm-Studies - Estudios Latinoamericanos

Redalyc

Universidad Nacional de Colombia

Carrera 30 No. 45-03, Edificio 310, primer piso

Correo electrónico: revcuaeo_bog@unal.edu.co

Página web: www.ceconomia.unal.edu.co

Teléfono: (571)3165000 ext. 12308, AA. 055051, Bogotá D. C., Colombia

Cuadernos de Economía Vol. 43 No. 91 - 2024

El material de esta revista puede ser reproducido citando la fuente. El contenido de los artículos es responsabilidad de sus autores y no compromete de ninguna manera a la Escuela de Economía, ni a la Facultad de Ciencias Económicas, ni a la Universidad Nacional de Colombia.

UNIVERSIDAD NACIONAL DE COLOMBIA

Rectora

Dolly Montoya Castaño

Vicerrectora Sede Bogotá (E)

Lorena Chaparro Díaz

FACULTAD DE CIENCIAS ECONÓMICAS

Decana

Juanita Villaveces

ESCUELA DE ECONOMÍA

Directora

Nancy Milena Hoyos Gómez

CENTRO DE INVESTIGACIONES PARA EL DESARROLLO

- CID

Karoll Gómez

DOCTORADO Y MAESTRÍA EN CIENCIAS ECONÓMICAS Y PROGRAMA CURRICULAR DE ECONOMÍA

Coordinadora

Olga Lucía Manrique

CUADERNOS DE ECONOMÍA

EDITOR

Gonzalo Cómbita

Universidad Nacional de Colombia

CONSEJO EDITORIAL

Marta Juanita Villaveces

Universidad Nacional de Colombia

Liliana Chicaíza Becerra

Universidad Nacional de Colombia

Manuel Muñoz Conde

Universidad Nacional de Colombia

Mario García Molina

Universidad Nacional de Colombia

Iván Montoya

Universidad Nacional de Colombia

Iván D. Hernández

Universidad de Ibagué

Juan Miguel Gallego

Universidad del Rosario

Paula Herrera Idárraga

Pontificia Universidad Javeriana

Esteban Pérez Caldentey

Comisión Económica para América Latina y el Caribe

Noemi Levy

Universidad Nacional Autónoma de México

Juan Carlos Moreno Brid

Universidad Nacional Autónoma de México

Matías Vernengo

Bucknell University

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivadas 2.5 Colombia.

Usted es libre de:

Compartir - copiar, distribuir, ejecutar y comunicar públicamente la obra

Bajo las condiciones siguientes:

- **Atribución** — Debe reconocer los créditos de la obra de la manera especificada por el autor o el licenciante. Si utiliza parte o la totalidad de esta investigación tiene que especificar la fuente.
- **No Comercial** — No puede utilizar esta obra para fines comerciales.
- **Sin Obras Derivadas** — No se puede alterar, transformar o generar una obra derivada a partir de esta obra.

Los derechos derivados de usos legítimos u otras limitaciones reconocidas por la ley no se ven afectados por lo anterior.



El contenido de los artículos y reseñas publicadas es responsabilidad de los autores y no refleja el punto de vista u opinión de la Escuela de Economía de la Facultad de Ciencias Económicas o de la Universidad Nacional de Colombia.

The content of all published articles and reviews does not reflect the official opinion of the Faculty of Economic Sciences at the School of Economics, or those of the Universidad Nacional de Colombia. Responsibility for the information and views expressed in the articles and reviews lies entirely with the author(s).

UN MODELO JERÁRQUICO PARA LA PREDICCIÓN DE INSOLVENCIA EMPRESARIAL. APLICACIÓN DE ANÁLISIS DISCRIMINANTE Y ÁRBOLES DE CLASIFICACIÓN

Dante Domingo Terreno
Jorge Orlando Pérez
Silvana Andrea Sattler

Terreno, D. D., Pérez, J. O., & Sattler, S. A. (2024). Un modelo jerárquico para la predicción de insolvencia empresarial. Aplicación de análisis discriminante y árboles de clasificación. *Cuadernos de Economía*, 43(91), 51-76.

El estudio plantea la utilización de un modelo jerárquico basado en la ratio de capacidad de pago, definido como la relación de los resultados operativos con las deudas, para la predicción de insolvencia en un contexto turbulento como la economía argentina. Las técnicas aplicadas para la construcción de los modelos son el análisis discriminante y árboles de clasificación. La evidencia indica que el

D. D. Terreno

Universidad Nacional de Córdoba, Córdoba, Argentina. Correo electrónico: dante.terreno@unc.edu.ar

J. O. Pérez

Universidad Católica de Córdoba, Córdoba, Argentina. Correo electrónico: jorgeoperez53@gmail.com

S. A. Sattler

Universidad Nacional de Córdoba, Córdoba, Argentina. Correo electrónico: silvana.sattler@unc.edu.ar

Sugerencia de citación: Terreno, D. D., Pérez, J. O., & Sattler, S. A. (2024). Un modelo jerárquico para la predicción de insolvencia empresarial. Aplicación de análisis discriminante y árboles de clasificación. *Cuadernos de Economía*, 43(91), 51-76. <https://10.15446/cuadernos.v43n91.105115>

Este artículo fue recibido el 4 de octubre 2022, ajustado el 7 de junio de 2023 y su publicación aprobada el 27 de julio de 2023.

modelo de capacidad de pago en dos momentos en el tiempo supera al modelo *Z-score* calibrado, aplicando el método de análisis discriminante lineal como el de árboles de clasificación, si bien este último mostró ser superior.

Palabras clave: insolvencia; economía argentina; ratio de capacidad de pago; aprendizaje automático.

JEL: G01, M41, C53, F34.

Terreno, D. D., Pérez, J. O., & Sattler, S. A. (2024). A hierarchical model for enterprise insolvency prediction. Application of discriminant analysis, and classification trees. *Cuadernos de Economía*, 42(91), 51-76.

The purpose of this study is the utilization of a hierarchical model based on the capability to pay ratio, definite for the relation operative earnings with liability, for insolvency prediction, in the turbulent environment how Argentine`s economy, in the decade last. The applicate techniques for building the models are discriminant analysis and classification trees. In general, the evidence shows that the capability to pay at two moments in the time model is higher than the calibrate *Z-score* model, so much applying discriminant analysis as classification trees, but the last method it showed to be better.

Keywords: Insolvency, Argentine economy, capability to pay ratio, machine learning.

JEL: G01, M41, C53, F34.

INTRODUCCIÓN

El crédito constituye una de las principales fuentes de recursos para el crecimiento de las empresas y a través de estos se canaliza parte de los excedentes de la economía. Este círculo virtuoso es afectado por el riesgo de incobrabilidad ocasionando la disminución de la fluidez e incrementando el costo del crédito (Bazarbash, 2019). En las últimas décadas se ha producido un importante interés en el desarrollo de modelos empíricos para la predicción de la insolvencia o crisis financiera. Beaver (1966) y Altman (1968) han sido los pioneros en la utilización de ratios contables para la predicción de la bancarrota y, a partir de allí, se produjo un gran número de investigaciones en esa línea. Los modelos empíricos están basados en la regularidad estadística, y las variables explicativas son seleccionadas de una especie de fondo común de variables que, en algunos casos, tienen un fundamento económico, y en otros el respaldo empírico de trabajos anteriores, pero sin explicitar un marco teórico (Tascón y Castaño, 2012).

La investigación, en general, se ha orientado en encontrar modelos con una mayor capacidad predictiva que los estudios precedentes, ya sea para un ámbito determinado o por algún método estadístico. Pero, los modelos utilizados se han apoyado en la aplicación de modelos lineales multivariantes con una estructura horizontal. A diferencia de dichos estudios, en este se plantea un modelo jerárquico que se soporta en la relación de los resultados operativos con las deudas para la predicción de la crisis financiera. El modelo jerárquico es coincidente con la metodología de análisis de estados financieros que va de lo general a lo particular (*top-down*) (Wahlen *et al.*, 2014). La relación entre los resultados operativos y las deudas es designada como ratio de la capacidad de pago y tiene por objetivo medir la capacidad de pago del acreedor. Este incluye en el numerador los resultados operativos como un indicador del potencial de generar efectivo para atender las obligaciones y, en el denominador, las deudas como referencia a los intereses y amortización de los pasivos que debe afrontar la empresa. Por otra parte, se aplica una técnica no paramétrica como árboles de clasificación utilizada por *machine learning*, además del método de análisis discriminante. Si bien Frydman *et al.* (1985) aplicaron esta misma metodología para la predicción de la insolvencia, el contexto de esta investigación es totalmente diferente.

El estudio es efectuado sobre empresas del Mercado de Valores de Buenos Aires entre los años 2014 y 2019. En la última década, la economía argentina enfrenta fuertes desequilibrios macroeconómicos, una recesión prolongada e inflación, lo que implica un contexto menos predecible. En tal situación se propone examinar el poder predictivo de la ratio de capacidad de pago con base en tres hipótesis: (i) aplicando el método de análisis discriminante lineal (ADL), la precisión en la predicción de la insolvencia de la ratio de capacidad de pago en un momento o en dos momentos en el tiempo es similar o superior al *Z-score* de Altman (1968) calibrado; (ii) la ratio de capacidad de pago calculado en dos momentos en el tiempo por el método de árboles de clasificación tiene una mayor precisión que por el ADL; y (iii) la ratio de capaci-

dad de pago en dos momentos en el tiempo desagregados en endeudamiento, margen y rotación, muestra una mejor precisión que el modelo sin desagregar.

La capacidad de pago intenta brindar un esquema que sirva de marco común para la predicción de la insolvencia por ratios contables. Adicionalmente, los analistas a partir del pronóstico por la capacidad de pago pueden profundizar en las causas subyacentes de los problemas financieros de la empresa.

REVISIÓN BIBLIOGRÁFICA

Modelos estructurales

Un enfoque alternativo a los modelos empíricos son los modelos estructurales para la predicción de la bancarrota. Merton (1974) en su trabajo sobre la estructura de riesgo de la tasa de interés define que la bancarrota ocurrirá si el valor de mercado de los activos cae por debajo del punto crítico determinado por el valor de libros de los pasivos. La aplicación más conocida de Merton es el método Moody's KMV. El método estima el valor futuro (esperado) de mercado de los activos de la empresa a partir del valor de mercado de sus acciones y su volatilidad; luego, se construye una medida con base en el número de desviaciones estándar entre el valor esperado de la empresa y el punto predeterminado de incumplimiento (DD, *distance-to-default*). Moody's KMV se sirve de su gran base de datos histórica para el cálculo de las probabilidades de incumplimiento¹ (Bharath y Shumway, 2008). Bharath y Shumway (2008) indican que, si bien el modelo de Merton no muestra buenos resultados estadísticos, su estructura funcional es útil para la predicción de la bancarrota.

Modelos empíricos

La primera etapa, de tipo descriptiva, consistió en comparar las ratios de las empresas solventes y en bancarrota. Posteriormente, en la etapa predictiva se trata de anticipar la bancarrota a través de la utilización de ratios contables, los primeros trabajos en esa línea son de Beaver (1966) y Altman (1968). El primero encuentra como el mejor predictor a la ratio de flujos de caja² sobre deudas. Altman (1968) desarrolla un modelo multivariante aplicando el método de ADL, el que denomina *Z-score*. El modelo fue elaborado con datos del periodo comprendido entre los años 1946 y 1965 y está compuesto por un conjunto de ratios: liquidez, autofinanciación, rentabilidad del activo, endeudamiento y rotación de activo. El estudio evidencia que el mejor predictor es la rotación de los activos y la probabilidad de predecir la bancarrota a un año antes de que ocurra es del 95% y dos años antes del 70%. En la revisión del *Z-score* Altman *et al.* (1977) elaboran el *Zeta-Score* para el periodo más reciente (desde 1969 hasta 1975),

¹ Basado en un análisis empírico de incumplimientos, KMV ha encontrado que el punto de incumplimiento más frecuente es un valor aproximadamente igual a los pasivos corrientes más el 50% de los pasivos a largo plazo (Altman, 2002).

² Los flujos de caja son medidos como resultados netos antes de depreciación y amortizaciones.

en este caso, la precisión de la clasificación alcanza el 96% un año antes y el 70% los cinco años anteriores. Las ratios más significativas para la predicción fueron la de rentabilidad acumulada y estabilidad de ingresos.

Ohlson (1980) emplea el método de regresión logística, no obstante la popularidad del análisis discriminante. El estudio comprende a empresas manufactureras, 105 en bancarrota y 2058 solventes en el periodo 1970-1976. Las variables clasificadoras son el tamaño de la empresa, medidas de la estructura financiera y medidas de desempeño económico y de liquidez. Desarrolla tres modelos diferentes para predecir la bancarrota, a un año, a dos años, y en uno o dos años. La precisión de la clasificación a un año alcanza el 96,12%; en dos años el 95,55%; y en el modelo a uno y dos años el 92,84%. Otro estudio que aplica dicho método es el de Zmijewski (1984), que utiliza tres ratios: rentabilidad, endeudamiento y liquidez. Jones y Hensher (2004) usan el modelo regresión logística mixto que, a diferencia del anterior, incorpora la heterogeneidad de las unidades.

A partir de la década de los noventa, ha tenido un gran desarrollo la aplicación de técnicas de *machine learning* (árboles de clasificación, *random forests*, redes bayesianas, etc.) para la predicción de la bancarrota. En uno de los primeros trabajos, Frydman *et al.* (1985) aplican el método de particionamiento recursivo (árboles de clasificación) y encuentran que dicha técnica produce una mejora de aproximadamente el 10% en relación con los modelos tradicionales. Entre otros estudios de este tipo, se destacan los de Nanni y Lumini (2009), Barboza *et al.* (2017), Mullainathan y Spiess (2017) y Athey (2018), estos dos últimos analizan la técnica de *machine learning* en el contexto del riesgo crediticio.

De las investigaciones en economías latinoamericanas sobresale la de Altman *et al.* (1979) y Altman *et al.* (1998) que aplican el *Z-score* a empresas de economías emergentes, Brasil y México, respectivamente. Altman *et al.* (1998) desarrollan un *scoring* (puntuación) para bonos corporativos en economías emergentes, específicamente para México. Otro importante trabajo, el de Montalván *et al.* (2011), constata para Perú que los factores macroeconómicos afectan la probabilidad de insolvencia empresarial, como así también la rentabilidad de los activos y el índice de solvencia. En Argentina, Caro *et al.* (2013) y Sandin y Porporato (2007) aplican el método de regresión logística mixta y ADL, respectivamente. Caro *et al.* (2018) utilizan el método de árboles de decisión para los países de Argentina, Chile y Perú. Los estudios de Argentina muestran buenos resultados estadísticos, pero sin mayor justificación económica.

DESCRIPCIÓN DE LA ECONOMÍA ARGENTINA

La década que va desde el 2011 al 2020 fue considerada la década “perdida” en cuanto al crecimiento del producto bruto interno (PBI); a partir del año 2012 la economía ha mostrado un total estancamiento. Dicho periodo se puede dividir en dos etapas, aquella comprendida entre los años 2011 y 2015 corresponde al man-

dato presidencial de Cristina Fernández de Kirchner; mientras que desde el 2015 hasta el 2019 a la presidencia de Mauricio Macri. La primera etapa estuvo marcada por la caída de los precios de los productos exportados, la disminución de los volúmenes exportados y el déficit comercial en hidrocarburos que lleva a una caída de las reservas internacionales (Frugoli y Pascual, 2020). En el 2012 aparece el déficit fiscal primario que sin acceso a los mercados internacionales la financiación se produce vía emisión monetaria, acelerando el proceso inflacionario y la consecuente devaluación del tipo de cambio.

Por otra parte, el gobierno de Macri libera el mercado de cambios provocando la primera devaluación en el año 2015. El fuerte desequilibrio de las cuentas públicas, no corregido en los primeros años, estuvo financiado por una creciente participación del endeudamiento externo, aprovechando la reapertura del acceso a los mercados financieros internacionales. En abril del 2018, la percepción de que los mercados internacionales no estaban dispuestos a continuar el flujo de financiamiento llevó al gobierno a solicitar un préstamo *stand-by* al Fondo Monetario Internacional. En esta época comenzó la recuperación de las exportaciones que registraron cuatro años consecutivos de crecimiento (2016-2019) (Frugoli y Pascual, 2020). En el 2019 la contracción económica se acompañó de un proceso de aceleración de la inflación, impulsada por la depreciación de la moneda, alcanzando el 52,90 % anual. Recién al final del mandato el gobierno logra una importante reducción del déficit primario, alcanzando un 0,44 % del PBI, pero con dos años consecutivos de caída del PBI. En la tabla 1 se detalla la evolución de los principales indicadores económicos.

Tabla 1.

Principales indicadores de la economía argentina

Año	% IPC ^{ab}	% PBI ^c	TC cierre ^d
2010	27,03	10,350	3,976
2011	23,28	6,150	4,304
2012	23,01	- 1,050	4,917
2013	31,94	2,000	6,518
2014	39,01	- 2,600	8,552
2015	31,57	2,700	13,005
2016	40,90	- 2,100	15,875
2017	25,00	2,800	18,774
2018	47,10	- 2,600	37,808
2019	52,90	- 2,100	59,895

Fuente: elaboración propia con base en: ^a Para los años 2016-2019: Instituto Nacional de Estadística y Censos (INDEC) (2020). ^b Para el periodo 2010-2015: Índice de precios al consumidor de Dirección Provincial de Estadística y Censos (2020). ^c INDEC (2020).

^d Banca Personal Nación (2020).

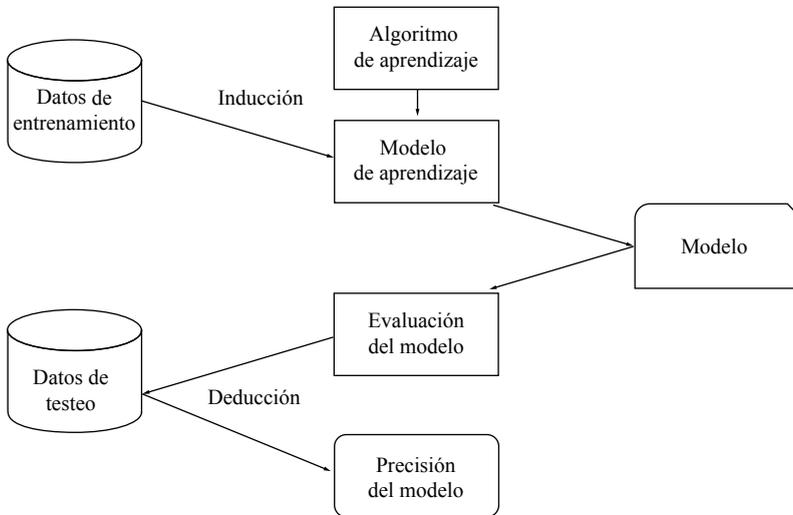
METODOLOGÍA

Método general para resolver problemas de clasificación

Un modelo de clasificación constituye una representación abstracta de la relación entre un conjunto de variables predictoras y determinadas categorías (por ejemplo, bancarota y no bancarota). Las técnicas de clasificación son un método sistemático para construir modelos de clasificación desde un conjunto de datos de entrada (árboles de decisión, redes neuronales, redes bayesianas, etc.). Cada técnica emplea un algoritmo de aprendizaje (*learning algorithm*) que identifica el modelo que mejor relaciona las variables predictoras y las categorías. El modelo generado por el algoritmo debería ajustarse bien a los datos y predecir correctamente la categoría (Tan *et al.*, 2016).

Figura 1.

Método general para la construcción de modelos de clasificación



Fuente: elaboración propia con base en Tan *et al.* (2016).

La figura 1 muestra el proceso general de clasificación. Los datos de entrenamiento consisten en un conjunto de datos categorizados que son empleados para construir el modelo de clasificación a través de la utilización de un algún algoritmo de aprendizaje. Una vez elaborado el modelo, este se aplica a un conjunto de observaciones en el cual es desconocida la categoría, datos de testeo, a fin de determinar el desempeño. La evaluación del modelo de clasificación se basa en la cantidad de observaciones correctas e incorrectamente clasificadas (Tan *et al.*, 2016). Los errores posibles, si se toma el caso de predicción de la bancarota, pueden ser: (i) tipo I es la falla en identificar una empresa como solvente cuando en realidad entra en crisis; y (ii) tipo II es identificar una empresa como solvente cuando

en realidad cae en crisis. Entre los métodos de validación más conocidos se registran el método de validación aparente (*apparent validation*) y la validación cruzada de k iteraciones (*k-fold cross-validation*). En el primero, se usan los mismos datos de entrenamiento para la validación, si bien este método sería válido para una muestra suficientemente grande y es útil como una primera aproximación. En el segundo, los datos se dividen en k subconjuntos de manera aleatoria, de los cuales uno de los subconjuntos se utiliza como datos de testeo y el resto ($k-1$) como datos de entrenamiento. El proceso de validación cruzada es repetido durante k iteraciones, con cada uno de los posibles subconjuntos de datos de testeo.

Entre las técnicas de clasificación se encuentran las técnicas paramétricas y no paramétricas. Los métodos paramétricos suponen la forma funcional de $f(x)$ conocida, por ejemplo, lineal, reduciendo el problema a determinar los parámetros de los que depende el modelo y que permiten un mejor ajuste de los datos. En lugar de eso, los modelos no paramétricos no buscan encontrar los parámetros de una función conocida, sino que emplean formas funcionales flexibles que se ajustan a los datos y aproximan la función objetivo. Estos métodos al ser más flexibles pueden adaptarse a patrones más complejos, como una estructura jerárquica. Entre los métodos paramétricos se aplica el ADL, entre los no paramétricos árboles de clasificación.

Análisis discriminante lineal

El análisis discriminante es una técnica que tiene como finalidad construir un modelo predictivo para pronosticar el grupo al que pertenece una observación (o individuo) a partir de determinadas características observadas que delimitan su perfil. La expresión funcional es la que se muestra en la ecuación (1).

$$y = f(x_1, x_2, x_3 \dots x_n) \quad (1)$$

Donde:

y : variables categóricas.

x_j : variables predictoras numéricas.

j : 1, 2, 3... n.

El objetivo del análisis discriminante es obtener una función discriminante que, de acuerdo con el *score*, posibilita clasificar los individuos, correspondiendo la expresión (2).

$$D = a_1 X_1 + a_2 X_2 + a_3 X_3 + \dots + a_n X_n \quad (2)$$

Donde:

D : *score* discriminante.

a_j : coeficientes (o pesos) discriminantes.

X_j : variables predictoras.

El algoritmo para la solución consiste en encontrar los valores de los coeficientes a_1, a_2, \dots, a_n que maximicen la variación entre grupos y minimicen la variación dentro de cada grupo. Las variaciones intergrupos e intragrupos se calculan a partir de las correspondientes sumas de los cuadrados de las desviaciones de las puntuaciones discriminantes con respecto a los valores de las variables independientes, iguales a las medias de cada grupo (*centroides*), esto es:

$$\begin{aligned}\bar{D}_1 &= a_1 \overline{X_1^{(1)}} + a_2 \overline{X_2^{(1)}} + \dots + a_n \overline{X_n^{(1)}} \\ \bar{D}_2 &= a_1 \overline{X_1^{(2)}} + a_2 \overline{X_2^{(2)}} + \dots + a_n \overline{X_n^{(2)}}\end{aligned}\quad (3)$$

El punto de corte discriminante (D') no es más que la media de las puntuaciones discriminantes medias de cada grupo:

$$D' = \frac{\bar{D}^{(1)} + \bar{D}^{(2)}}{2}\quad (4)$$

Finalmente, se debe aplicar el siguiente criterio para clasificar un individuo determinado:

$D_i < D'$ se clasifica el individuo en el grupo 1.

$D_i > D'$ se clasifica el individuo en el grupo 2.

La aplicación de ADL se apoya en una serie de supuestos básicos: normalidad multivariante, homogeneidad de matrices de varianza-covarianza, linealidad y ausencia de multicolinealidad. En ocasiones, sin embargo, puede que algunos de ellos no se cumplan, esto no supone necesariamente la invalidación de los resultados (Hair *et al.*, 1999).

Árboles de clasificación

Un árbol de clasificación es un conjunto de reglas organizadas en una estructura jerárquica, mientras que la clasificación final de una observación es el resultado del cumplimiento de las reglas desde el nodo raíz hasta sus hojas. La estructura de árbol está compuesta por tres niveles de nodos, un único nodo en la cima llamado nodo raíz, nodos intermedio y nodos hoja, donde cada rama representa una regla de decisión.

Breiman *et al.* (1984) introdujeron el algoritmo conocido como CART (*classification and regression trees*) para la construcción de árboles de decisión. CART es un método no paramétrico de clasificación binaria, en el que cada partición genera una regla de decisión y esto se sucede en forma recursiva hasta finalizar la construcción del árbol, por eso es conocido como particionamiento recursivo. El proceso paso a paso para construir un árbol de decisión consiste en:

a) Construcción del árbol máximo (T_{\max})

Sea un vector $x = (x_1, x_2, x_3, \dots, x_n)$ de variables predictoras, el proceso consiste en buscar la mayor homogeneidad en cada nodo observando las siguientes características: (i) cada división de los nodos depende del valor de una sola variable predictora, y (ii) la separación del árbol es evaluada y determinada con base en la medida de impureza. Primero se busca la variable predictora con la mayor impureza y luego se establece una regla y se separan los datos. Esto es efectuado de manera recursiva hasta que es imposible disminuir la impureza del árbol. Una de las posibles medidas de impureza es el valor de la entropía:

$$\text{Entropía}(S) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 \quad (5)$$

Sea, p_1 y p_2 la proporción clasificada en cada grupo. Si p_1 y p_2 son 0,50, entonces la entropía está en su máximo 1. Si la p_1 y p_2 es 1, entonces el grupo es completamente puro y la entropía está en su mínimo 0.

b) Poda de árbol

Cuando el árbol es demasiado grande y complejo, no necesariamente es el árbol óptimo. Porque como el árbol es elaborado a medida de los datos de entrenamiento, pierde generalidad y precisión al aplicarse a una muestra distinta a la de entrenamiento. Este problema se conoce como sobreidentificación. Por ello, es menester generar un proceso de reducción o poda. La poda por definición es la eliminación de ramas y el reemplazo de estas por nodos de hoja hasta minimizar el error de clasificación errónea. Existen distintos criterios para la poda del árbol, el utilizado por el algoritmo CART es el costo de complejidad que se puede describir a partir de:

$$R(T) = \text{Resub}(T) + \alpha |T| \quad (6)$$

$R(T)$ tasa de error obtenida por *k-fold cross-validation*, $\text{Resub}(T)$ la tasa de clasificación errónea estimada por resustitución³, α una constante no negativa interpretada como el costo por la complejidad del árbol y $|T|$ es el número de nodos terminales del árbol T . Para $\alpha = 0$ el árbol óptimo es igual a T_{max} , para cualquier valor positivo de α el óptimo es un subárbol de T_{max} , o sea a medida que α se incrementa se poda más el árbol y para un valor suficientemente alto de α el árbol no muestra divisiones. A partir de (6) se genera un anidamiento de secuencias de árboles menor complejidad y aumento de $\text{Resub}(T)$, siendo seleccionado un α que minimice el valor de $R(T)$. Según Breiman *et al.* (1984) el valor adecuado es $k = 10$.

Datos

Este estudio es ejecutado sobre una muestra de empresas no financieras listadas en el Mercado de Valores de Buenos Aires, se excluyen las empresas de capital extranjero y las empresas de mayor tamaño (Arcor y YPF) para una mayor homogeneidad de la muestra. La muestra está constituida por 14 empresas que entra-

³ La resustitución estima el porcentaje de error cometido con los datos de entrenamiento.

ron en crisis financiera en los años 2018 y 2019, y por 28 empresas consideradas solventes en el mismo periodo. El hecho que se tuvo en cuenta para clasificar una empresa en crisis fue la individualización efectuada por la Bolsa de Comercio de Buenos Aires por causas económicas (incumplimiento en el pago de obligaciones negociables, pérdida de parte de su patrimonio, patrimonio neto negativo o la presentación en concurso preventivo) e información obtenida de publicaciones económicas (*Infobae* y *El Cronista Comercial*). Las observaciones están constituidas por los datos de los estados financieros anuales consolidados, elaborados bajo las Normas Internacionales de Información Financiera, por un periodo de cinco años comprendido entre los años 2013-2017 y 2014-2018.

Ratio principal: capacidad de pago de la deuda

La crisis financiera o fracaso empresarial se define como la incapacidad para atender obligaciones financieras a su vencimiento. Operacionalmente, se dice que una empresa ha fracasado cuando se produce algunos de los siguientes eventos: bancarrota, incumplimiento del pago de un bono o sus intereses, sobregiro en la cuenta bancaria y no pago de los dividendos preferidos (Beaver, 1966). La relación entre resultados y el total de los pasivos o ratio de la capacidad de pago se corresponde con el concepto de crisis financiera mencionado. Cuando dicha ratio cae por debajo de un determinado nivel es muy probable que la empresa entre en insolvencia, debido a que no genera suficiente efectivo para atender sus obligaciones. La ratio propuesta tiene su antecedente en Beaver (1966) y Frydman *et al.* (1985). Pero, a diferencia de la ratio propuesta se utilizan en el numerador los flujos de efectivo en vez de resultados contables.

La ratio de capacidad de pago de la deuda tiene un enfoque similar al del modelo de Moody's KMV; este determina la probabilidad de bancarrota por la distancia entre el valor de mercado de la empresa y los pasivos mapeados por la experiencia histórica. En la ratio de capacidad de pago propuesto, el punto de corte se establece con base en datos contables históricos, y en lugar del valor de mercado son utilizados los resultados contables.

Los resultados operativos o de explotación, conocidos en inglés como *core earnings*, son los que reflejan mejor la capacidad de generar flujos de efectivo (Messim y Penman, 2001). La expresión matemática de la capacidad de pago es la que se muestra en la ecuación (7).

$$CP = \frac{Rop}{PT} \quad (7)$$

Rop son los resultados operativos y surgen de la diferencia entre ventas menos costo de ventas y los gastos operativos y PT es el pasivo total e incluye deudas operativas y financieras. Si la expresión (7) se la divide numerador y denominador por los activos totales (AT), se la puede presentar en función de la ratio de endeudamiento y rentabilidad del activo, resultando las fórmulas (8) y (9).

$$End = \frac{PT}{AT} \quad (8)$$

$$RA = \frac{Rop}{AT} \quad (9)$$

Reemplazando, se obtiene:

$$CP = \frac{RA}{End} \quad (10)$$

Por otra parte, si se divide y multiplica el numerador la expresión (10) por las ventas totales (VT) se puede desagregar la RA en margen de ganancias (Mg) y en la rotación de los activos ($Rota$), resultando las fórmulas (11), (12) y (13).

$$Mg = \frac{Rop}{VT} \quad (11)$$

$$Rota = \frac{VT}{AT} \quad (12)$$

$$RA = Mg \times Rota \quad (13)$$

Reemplazando en (10) por esta última fórmula, se tiene:

$$CP = \frac{(Mg \times Rota)}{End} \quad (14)$$

Con base en las ratios desarrolladas se plantean los siguientes modelos. El primer modelo, univariante, incluye la capacidad de pago en un momento en el tiempo, siendo:

$$CP1 = f (CP_t) \quad (15)$$

Donde CP_t es la capacidad de pago del año actual. El periodo de análisis es de cuatro años, esto presupone que es posible anticipar la insolvencia en dicho plazo. La insolvencia, en la mayoría de los casos, es un proceso que sucede a lo largo del tiempo. De ahí, es interesante considerar la capacidad de pago en dos momentos en el tiempo, al cierre del año actual y al cierre del año anterior, como se muestra en la ecuación (16).

$$CP2 = f (CP_t, CP_{t-1}) \quad (16)$$

Donde CP_{t-1} es la capacidad de pago del año anterior.

Si la expresión (16) se la desagrega de acuerdo con (14), es posible plantear el modelo desagregado:

$$CPD = f(End_t, Mg_t, Rota_t, End_{t-1}, Mg_{t-1}, Rota_{t-1}) \quad (17)$$

De acuerdo con Esplin *et al.* (2014) la desagregación mejora la predicción del pronóstico de la rentabilidad, lo cual se supone que se puede trasladar a la predicción de la insolvencia.

Z-score

El modelo *Z-score* es utilizado como un punto de referencia para la evaluación de los modelos de capacidad de pago, la función es la que se muestra en la ecuación (18).

$$Z = a_1 X1 + a_2 X2 + a_3 X3 + a_4 X4 + a_5 X5 \quad (18)$$

Donde:

$$X1 = \frac{\text{Capital corriente}}{\text{Activos totales}}$$

$$X2 = \frac{\text{Resultados retenidos}}{\text{Activos totales}}$$

$$X3 = \frac{\text{Resultados antes de intereses e impuestos}}{\text{Activos totales}}$$

$$X4 = \frac{\text{Patrimonio}}{\text{Activos totales}}$$

$$X5 = \frac{\text{Ventas}}{\text{Activos totales}}$$

El *Z-score* fue desarrollado para empresas manufactureras en el contexto económico de Estados Unidos entre los años 1946 y 1965, lo cual difiere completamente del contexto actual de Argentina, por lo tanto, el modelo es reestimado con los datos de este estudio. En el análisis previo se evaluó la aplicación del *Z-score* modificado para empresas no manufactureras, pero los resultados no mostraron diferencias.

Análisis previo

En esta etapa previa, se analizan las variables individualmente. La tabla 2 informa la media, mediana, desviación estándar y el estadístico-F para el análisis de varianzas (ANOVA), agrupados por las empresas solventes y en crisis. Los valores asumidos por los ratios de capacidad de pago, rentabilidad del activo, margen de ganancias y rotación de los activos de las empresas solventes son superiores a las empresas en crisis y, lógicamente, la ratio de endeudamiento es superior en las empresas en crisis.

Tabla 2.
Estadísticos descriptivos

	Solventes			Crisis				
			Desv.			Desv.	ANOVA	
Ratios	Media	Mediana	Estándar	Media	Mediana	Estándar	F-value	
A-Ratio de capacidad de pago								
CP_t	0,3120	0,2724	0,2048	0,1311	0,1148	0,1680	32,620	***
End_t	0,5699	0,5920	0,1599	0,7298	0,7424	0,1498	38,930	***
RA_t	0,1555	0,1489	0,0842	0,0852	0,0900	0,1038	22,220	***
Mg_t	0,1923	0,1499	0,1697	0,0617	0,0862	0,1050	27,770	***
$Rota_t$	1,1971	1,0399	0,8520	1,0758	1,0287	0,5228	0,953	
CP_{t-1}	0,2941	0,2641	0,1879	0,1576	0,1313	0,1792	20,520	***
End_{t-1}	0,5646	0,5843	0,1574	0,6901	0,6985	0,1685	22,620	***
RA_{t-1}	0,1481	0,1495	0,0807	0,0973	0,0977	0,1105	11,480	***
Mg_{t-1}	0,1319	0,1414	0,4429	0,0531	0,0862	0,1851	1,627	
$Rota_{t-1}$	1,1774	1,0152	0,8339	1,0784	0,9857	0,5228	0,656	
B-Z-score								
X1	0,1895	0,1899	0,1867	0,1481	0,1481	0,1501	2,079	
X2	0,3028	0,2757	0,1505	0,1280	0,1523	0,1720	44,630	***
X3	0,1619	0,1567	0,0844	0,0875	0,0940	0,1062	24,330	***
X4	0,9418	0,6893	0,7133	0,4369	0,3472	0,3454	25,050	***
X5	1,1971	1,0399	0,8520	1,0758	1,0287	0,5228	0,953	

Nota: ANOVA: análisis de varianza, *** nivel de significación menor a 0,001.

Fuente: elaboración propia con base en datos de estados financieros publicados por la Comisión Nacional de Valores de Argentina.

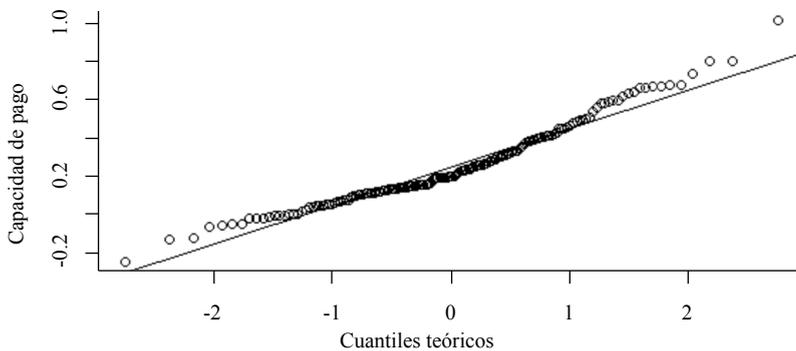
Además, las ratios de la capacidad de pago, el endeudamiento y la rentabilidad de los activos mostraron que las diferencias son significativas para el periodo actual y anterior, de empresas solventes y en crisis, en cambio el margen de ganancias lo es únicamente para el periodo actual. Contrario a lo supuesto, la rotación de los activos no exhibe diferencias significativas, posiblemente, debido a que la rotación depende en gran medida del tipo de actividad y del contexto recesivo de la economía. En el *Z-score* las diferencias de media resultan significativas en la autofinanciación (x2), rentabilidad del activo (x3) y endeudamiento (x4).

El ADL debe cumplir determinados supuestos: normalidad e igualdad de varianzas. La distribución de los datos de la ratio de capacidad de pago merece una atención especial. La figura 2 es un método gráfico que permite observar cuán cerca

está la distribución de un conjunto de datos de la distribución normal teórica. La figura 2 indica una distribución alejada de la normal de los datos de capacidad de pago actual. La curva muestra una trayectoria cóncava debido a que la relación entre las ratios de endeudamiento y rentabilidad no son independientes. Según las teorías de las jerarquías financieras, un aumento de endeudamiento indica la incapacidad de la empresa para generar fondos internamente (Myers, 1984). Para solucionar el problema de la falta de normalidad son utilizados logaritmos. La tabla 2 muestra que el Log (CP_t) y el Log (CP_{t-1}) de acuerdo con Shapiro-Wilk tiene una distribución cercana a la normal.

Figura 2.

Cuantil-cuantil (Q-Q plots): CP_t



Fuente: elaboración propia con Stata 16 y datos de los estados financieros.

La prueba de normalidad multivariante, según se describe en la tabla 3, muestra que las variables del modelo CP2, de acuerdo con el test de Nardia, tienen una distribución normal multivariante. En cambio, se rechaza la distribución normal multivariante para las variables del modelo *Z-score*, existiendo normalidad univariante para las variables x_1 (liquidez) y x_3 (rentabilidad).

Tabla 3.

Test de normalidad univariante y multivariante

Tests	Estadístico	
a-Modelo CP2		
Nardia test (multivariante)		
Asimetría	5,5361	***
Curtosis	1,3999	***
Shapiro-Wilk (univariante)		
Log (CP_t)	0,9859	**

(Continúa)

Tests	Estadístico	
Log (CP _{t-1})	0,9842	**
b-Modelo Z-score		
Asimetría	417,7700	
Curtosis	15,7500	
Shapiro-Wilk (univariante)		
x1	0,9877	***
x2	0,9534	
x3	0,9902	***
x4	0,8213	
x5	0,8667	

Nota: *** nivel de significación superior a 0,10; ** nivel de significación superior a 0,05.
Fuente: elaboración propia con Stata16 y datos de los estados financieros.

La tabla 4 muestra que de acuerdo con el test de Box las varianzas son homogéneas para el modelo CP2, no así para el Z-score. Cuando las varianzas no son homogéneas, no es apropiada la estructura lineal y parece ser más eficiente la estructura cuadrática. Los modelos lineales y cuadráticos revelaron una precisión similar con la muestra original, pero con la muestra de prueba indican una clara superioridad de la estructura lineal (Altman, 2013).

Tabla 4.

Box M-test: homogeneidad de varianzas

Detalle	Chi-Sq (aprox.)		gl
a-Modelo CP2	6,424	***	3
b-Modelo Z-score	84,374		15

Nota: *** nivel de significación superior a 0,10.

Fuente: elaboración propia con Stata16 y datos de los estados financieros.

DESARROLLO DE LOS MODELOS Y RESULTADOS EMPÍRICOS

Hipótesis 1

Para responder a la primera hipótesis, las ratios de capacidad de pago en un momento y en dos momentos en el tiempo son evaluadas con el Z-score, aplicando el ADL. Los modelos elaborados se exponen en la tabla 5.

Tabla 5.

Funciones discriminantes

Modelo	Función discriminante	Criterio de decisión
CP1	$CP' = 15,872 \log(CP_i)$	$CP' > 1,2907$ Solvente $CP' < 1,2907$ Crisis
CP2	$CP'' = 12,983 \log(CP_i) + 4,140 \log(CP_{i-1})$	$CP'' > 1,4004$ Solvente $CP'' < 1,4004$ Crisis
Z-score	$Z = -0,692 X1 + 3,891 X2 + 4,454 X3 +$ $0,444 X4 + 0,09 X5$	$Z > 1,6818$ Solvente $Z < 1,6818$ Crisis

Nota: estimado por el método de *maximum likelihood* (máxima verosimilitud), probabilidades *a priori*: 0,50.

Fuente: elaboración propia con Stata16 y datos de los estados financieros.

La tabla 6 entrega los resultados de la clasificación utilizando los datos de la muestra de entrenamiento (muestra original). El modelo CP2 presenta el mayor porcentaje de empresas correctamente clasificadas (75,45%), supera levemente al *Z-score* (74,11%) y al modelo CP1 (73,22%). En el análisis de los errores, el modelo CP1 evidencia el menor error tipo I (17,86%), si bien este es más costoso, tiene el mayor error tipo II (35,71%). El error tipo I del modelo CP2 (19,64%) es inferior al *Z-score* (23,21%), mientras que los errores tipo II son bastantes similares. Entre los modelos CP1 y CP2, este último exhibe un menor error tipo II – 6,25% (29,46% – 35,71%).

Tabla 6.

Test de validación aparente (muestra original), análisis discriminante

Modelo CP1					
Tipo de error	Correctas	Incorrectas	Total	% correctas	% incorrectas
Tipo I	46	10	56	82,14	17,86
Tipo II	72	40	112	64,29	35,71
Total*			168	73,22	26,79
Modelo CP2					
Tipo de error	Correctas	Incorrectas	Total	% correctas	% incorrectas
Tipo I	45	11	56	80,36	19,64
Tipo II	79	33	112	70,54	29,46
Total*			168	75,45	24,55
Z-score					
Tipo de error	Correctas	Incorrectas	Total	% correctas	% incorrectas
Tipo I	43	13	56	76,79	23,21
Tipo II	80	32	112	71,43	28,57
Total*			168	74,11	25,89

Nota: * balanceado en un 50%; no existen costos diferenciales de clasificación errónea.

Fuente: elaboración propia con Stata16 y datos de los estados financieros.

La tabla 7 señala los resultados de la validación cruzada. El modelo CP1 tiene un porcentaje levemente superior de empresas correctamente clasificadas por la validación cruzada en relación con la validación aparente del 0,36 % (73,58 % – 73,22 %). En cambio, para el modelo CP2 y *Z-score* la disminución del porcentaje correctamente clasificado es de – 0,82 % (74,63 % – 75,45 %) y – 1,51 % (72,60 % – 74,11 %), respectivamente. El modelo CP2 muestra el mayor porcentaje de empresas correctamente clasificadas (74,63 %), supera al modelo CP1 (73,58 %) y al *Z-score* (72,60 %). En el análisis de los errores, el modelo CP1 reporta el menor error tipo I (17,50 %), pero tiene el mayor error tipo II (35,35 %). El error tipo I del modelo CP2 (20,20 %) es inferior al *Z-score* (25,30 %), mientras que en el error tipo II el *Z-score* (29,50 %) es levemente inferior al modelo CP2 (30,55 %). En síntesis, los resultados de la validación aparente y cruzada revelan un desempeño del modelo de capacidad de pago en dos momentos superiores al *Z-score*, en particular un error tipo I bastante inferior (– 5,10 %).

Tabla 7.

Test de validación cruzada, análisis discriminante

Modelo CP1					
Tipo de error	% correctas	% incorrectas	% total	% correctas	% incorrectas
Tipo I	27,50	5,83	33,33	82,50	17,50
Tipo II	43,10	23,57	66,67	64,65	35,35
Total*			100,00	73,58	26,43
Modelo CP2					
Tipo de error	% correctas	% incorrectas	% total	% correctas	% incorrectas
Tipo I	26,60	6,73	33,33	79,80	20,20
Tipo II	46,30	20,37	66,67	69,45	30,55
Total*			100,00	74,63	25,38
Z-score					
Tipo de error	% correctas	% incorrectas	% total	% correctas	% incorrectas
Tipo I	24,90	8,43	33,33	74,70	25,30
Tipo II	47,00	19,67	66,67	70,50	29,50
Total*			100,00	72,60	27,40

Nota: * balanceado en un 50 %; método de validación: *cross-validation* con $k = 10$ y 5 repeticiones; no existen costos diferenciales de clasificación errónea.

Fuente: elaboración propia con Stata16 y datos de los estados financieros.

Hipótesis 2

Por esta hipótesis, corresponde evaluar el modelo CP2 por el método de árboles de clasificación en relación con el mismo modelo por ADL, medido en la hipótesis anterior.

En la figura 3 se observa el árbol de clasificación del modelo CP2, del cual surgen las siguientes reglas:

- a. Si $CP_t \geq 0,22$ la empresa es clasificada como solvente;
- b. Si $CP_t < 0,22$ y $CP_{t-1} < 0,22$ la empresa es clasificada como en crisis;
- c. Si $(CP_t < 0,22$ y $CP_{t-1} \geq 0,22)$ la empresa es clasificada como solvente.

El nodo raíz indica un total de 168 observaciones (112 solventes y 56 en crisis), cuando CP_t es mayor o igual al 22 % (equivalente a 4,55 años de pasivo sobre resultados), clasifica una importante proporción (72/80) de observaciones como empresas solventes. Cuando CP_t es menor al 22 % y CP_{t-1} es mayor o igual al 22 % clasifican una parte importante de las observaciones (15/18) como empresas sanas. Cuando CP_t y CP_{t-1} son menores al 22 % clasifica una alta proporción (45/70) de observaciones en empresas en crisis. La secuencia demuestra, claramente, que la mayor posibilidad de que la empresa entre en crisis es cuando la capacidad de pago en ambos momentos es inferior al valor crítico.

Los resultados del test de validación aparente del modelo CP2 por árboles de clasificación, expuestos en la tabla 8, muestran un porcentaje de clasificaciones correctas superior al construido por ADL en 3,57 % (79,02 % – 75,45 %). En la validación cruzada, expuesta en la tabla 9, el porcentaje de clasificaciones correctas por el método de árboles de clasificación (77,57 %) es superior al ADL (74,63 %), en especial exhibe una significativa reducción del error tipo II del – 8,21 % (22,34 % – 30,55 %). Si se lo compara con el *Z-score* por ADL, hay una mejora en el porcentaje de clasificaciones correctas del 4,97 % (77,57 % – 72,60 %). Por lo tanto, respondiendo a la segunda hipótesis, el modelo de capacidad de pago con dos momentos por el método de árboles de clasificación alcanza una mayor precisión que por el método de ADL. Esto demuestra que los patrones de clasificación no lineales y de tipo jerárquico tienen una mayor capacidad predictiva que una función lineal.

Tabla 8.

Test de validación aparente, árboles de clasificación: modelo CP2

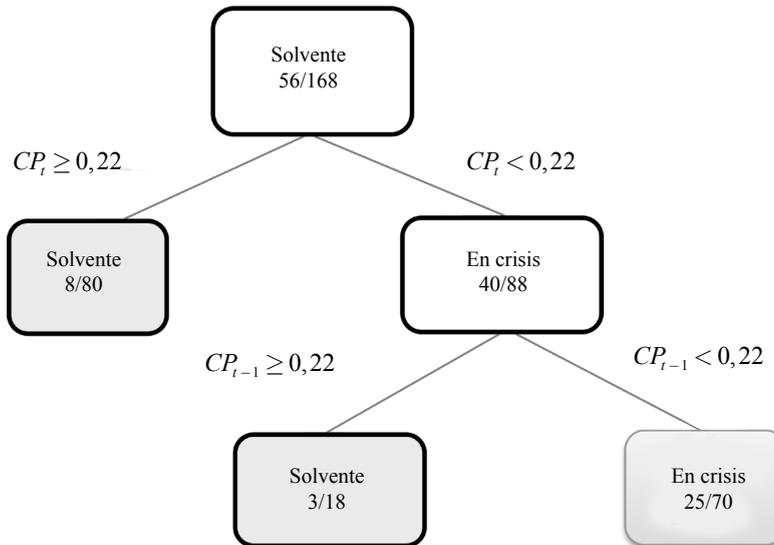
	Correctas	Incorrectas	Total	% correctas	% incorrectas
Tipo I	45	11	56	80,36	19,64
Tipo II	87	25	112	77,68	22,32
Total*			168	79,02	20,98

Nota: * balanceado al 50 %; no existen costos diferenciales de clasificación errónea.

Fuente: elaboración con RStudio y datos de los estados financieros.

Figura 3.

Árbol de clasificación: modelo CP2



Nota: No: solvente, Sí: en crisis; algoritmo de árboles de clasificación: CART; algoritmo de división: *information* (ver expresión 1); método de poda: costo de complejidad (α) (ver expresión 2). En cada nodo el número a la derecha señala el total de observaciones y el número a la izquierda el número de clasificaciones incorrectas. Para seleccionar el árbol óptimo fue utilizado el valor de mayor precisión y el valor para el modelo final fue de $\alpha = 0,02381$. Los valores de α fueron determinados por el proceso de *cross-validation* $k = 10$ y repeticiones = 5.

Fuente: elaboración propia con RStudio y datos de los estados financieros.

Tabla 9.

Test de validación cruzada, árboles de clasificación: modelo CP2

Tipo de error	% correctas	% incorrectas	% total	% correctas	% incorrectas
Tipo I	25,80	7,50	33,30	77,48	22,52
Tipo II	51,80	14,90	66,70	77,66	22,34
Total*				77,57	22,43

Nota: * balanceado al 50%; método de validación: *cross-validation* con $k = 10$ y 5 repeticiones; no existen costos diferenciales de clasificación errónea.

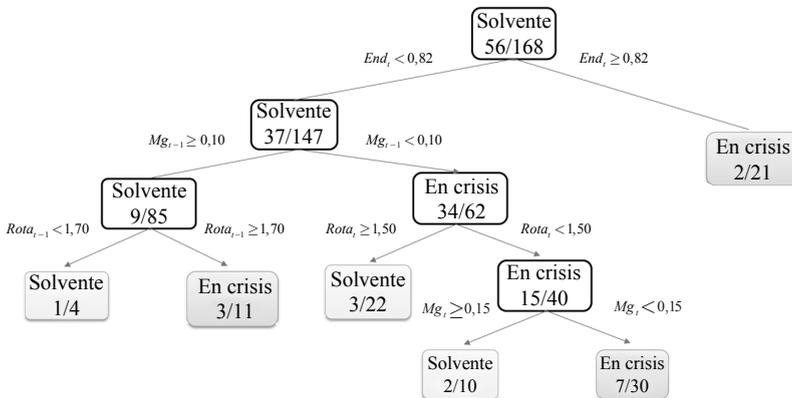
Fuente: elaboración con RStudio y datos de los estados financieros.

Hipótesis 3

Para contrastar la hipótesis 3, se debe evaluar el modelo de capacidad de pago en dos momentos desagregado (CPD) elaborado con árboles de clasificación para ser comparado con el mismo modelo sin desagregar (CP2). En la figura 4 se ilustra el árbol de clasificación CPD, del cual surgen las siguientes reglas:

- a. Si $End_{t-1} \geq 0,82$ la empresa es clasificada como en crisis;
- b. Si $(End_t < 0,82 \text{ y } Mg_{t-1} \geq 0,10 \text{ y } Rota_{t-1} < 1,7)$ la empresa es clasificada como solvente;
- c. Si $(End_t < 0,82 \text{ y } Mg_{t-1} < 0,10 \text{ y } Rota_{t-1} \geq 1,7)$ la empresa es clasificada como en crisis;
- d. Si $(End_t < 0,82 \text{ y } Mg_{t-1} < 0,10 \text{ y } Rota_t \geq 1,5)$ la empresa es clasificada como solvente;
- e. Si $(End_t < 0,82 \text{ y } Mg_{t-1} < 0,10 \text{ y } Rota_t \geq 1,5 \text{ y } Mg_t \geq 0,15)$ la empresa es clasificada como solvente;
- f. Si $(End_t < 0,82 \text{ y } Mg_{t-1} < 0,10 \text{ y } Rota_t < 1,5 \text{ y } Mg_t < 0,15)$ la empresa es clasificada como en crisis.

Figura 4.
Árbol de clasificación: modelo CP3



Nota: No: solvente, Sí: en crisis; algoritmo de árboles de clasificación: CART; algoritmo de división: *information* (ver expresión 1); método de poda: costo de complejidad (α) (ver expresión 2). En cada nodo el número a la derecha indica el total de observaciones y el número a la izquierda el número de clasificaciones incorrectas. Para seleccionar el árbol óptimo fue utilizado el valor de mayor precisión y el valor del modelo final fue de $\alpha = 0,00$. Los valores de α fueron determinados por el proceso de *cross-validation* $k = 10$ y repeticiones = 5.
Fuente: elaboración con RStudio y datos de los estados financieros.

Del análisis de la secuencia del árbol surgen cuestiones interesantes. La primera variable es el nivel de endeudamiento (del año actual), cuando la empresa tiene un pasivo que representa un 82% del activo o más, clasifica directamente a la observación como empresa en crisis (19/21). Cuando la empresa tiene un margen del año anterior superior al 10%, depende del nivel de rotación del año anterior. Si tiene una rotación menor a 1,70 las observaciones son clasificadas como empresas solventes (73/74), pero si tiene una rotación mayor al 1,70 las observaciones son consideradas

empresas en crisis (8/11). Esto último se explica porque en determinadas situaciones una alta rotación puede ser consecuencia de la falta de los activos.

Tabla 10.

Test de validación aparente, árbol de clasificación: modelo CPD

Tipo de error	Correctas	Incorrectas	Total	% correctas	% incorrectas
Tipo I	50	6	56	89,29	10,71
Tipo II	100	12	112	89,29	10,71
Total*			168	89,29	10,71

Nota: * balanceado en un 50 %; método de validación: *cross-validation* con $k = 10$ y 5 repeticiones; no existen costos diferenciales de clasificación errónea.

Fuente: elaboración con RStudio y datos de los estados financieros.

Por otra parte, de acuerdo con el test ANOVA la rotación de los activos entre empresas solventes y en crisis no es significativa, pero sí es relevante en la clasificación no paramétrica. Los resultados del test de validación aparente del modelo CPD por árboles de clasificación, expuestos en la tabla 10, muestran un porcentaje de categorizaciones correctas sumamente alto (89,29 %) y un error tipo I y II del 10,71 %. Si se lo compara con los resultados del modelo CP2, la precisión del modelo CPD resulta muy superior, siendo la diferencia de 10,27 % (89,29 % – 79,02 %).

Tabla 11.

Test de validación cruzada, árbol de clasificación: modelo CPD

Tipo de error	% correctas	% incorrectas	% total	% correctas	% incorrectas
Tipo I	22,70	10,60	33,30	68,17	31,83
Tipo II	54,40	11,80	66,20	82,18	17,82
Total*				75,18	24,83

Nota: * balanceado en un 50 %; método de validación: *cross-validation* con $k = 10$ y 5 repeticiones; no existen costos diferenciales de clasificación errónea.

Fuente: elaboración con RStudio y datos de los estados financieros.

Los resultados del test de validación cruzada, expuestos en la tabla 11, muestran que el modelo CP2 exhibe un porcentaje de aciertos levemente superior al modelo CPD del 2,39 % (77,57 % – 75,18 %). Por lo que, dando respuesta a la tercera hipótesis, el modelo CPD no ostenta una mayor precisión que el modelo sin desagregar (CP2). No obstante, el modelo desagregado revela la relación entre las variables en la explicación de la insolvencia.

CONCLUSIONES

Este estudio plantea la utilización de la ratio de capacidad de pago, definida por la relación de los resultados operativos con las deudas, para la predicción de la insol-

vencia en el contexto turbulento de la economía argentina en la última década. Las conclusiones obtenidas de la validación cruzada demuestran que: primero, aplicando el ADL el modelo de capacidad de pago en un momento en el tiempo exhibe una precisión algo inferior al *Z-score*, no obstante que el error tipo I es bajo (17,50%) es muy importante el error tipo II. En cambio, el modelo en dos momentos muestra una precisión superior al *Z-score*, el error tipo I es menor en un 5,10%. Segundo, el modelo de capacidad de pago en dos momentos por el método de árboles de clasificación evidencia una mejora en la precisión en relación con el método de ADL, por lo cual disminuye el error tipo I un 2,78%. Esto demuestra que los patrones de clasificación no lineales y de tipo jerárquico tienen una mayor capacidad predictiva que una función lineal. El punto de insolvencia de la ratio de capacidad de pago actual de acuerdo con el árbol de clasificación es del 22%, equivalente a 4,55 años de pasivo sobre resultados operativos.

Tercero, el modelo de capacidad de pago en dos momentos, desagregado en las ratios de endeudamiento, margen y rotación utilizando el método de árboles de clasificación no reporta una mayor precisión que el modelo sin desagregar. No obstante, el modelo desagregado revela la relación de las ratios en la explicación de la insolvencia y la importancia, principalmente, del nivel de endeudamiento actual y el margen de ganancias del año anterior. La rotación de los activos que no había resultado significativa en el análisis por ANOVA, sí mostró ser relevante en un método no paramétrico.

En general, los resultados evidencian la capacidad predictiva de los modelos basados en ratio de capacidad de pago, en particular referida a dos momentos en el tiempo. Este estudio constituye un aporte para la toma de decisiones de los inversores y entidades financieras en el otorgamiento de créditos y resulta en un punto de partida para profundizar en las causas subyacentes de los problemas financieros de la empresa o en las fortalezas de esta.

REFERENCIAS

1. Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609. <https://doi.org/10.2307/2978933>
2. Altman, E. I. (2002). Corporate distress prediction models in a turbulent economic and Basel II environment. *NYU Working Paper No. FIN-02-052*. Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1295810
3. Altman, E. I. (2013). Predicting financial distress of companies: revisiting the Z-score and ZETA[®] models. En A. R. Bell, C. Brooks & M. Prokopczuk (Eds.), *Handbook of Research Methods and Applications in Empirical Finance* (pp. 428-456). Edward Elgar Publishing. <https://doi.org/10.4337/9780857936097>

4. Altman, E. I., Baidya, T. K., & Dias, L. M. R. (1979). Previsão de problemas financeiros em empresas. *Revista de Administração de Empresas*, 19(1), 17-28. <https://doi.org/10.1590/S0034-75901979000100002>
5. Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). ZETATM analysis: a new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, 1(1), 29-54. [https://doi.org/10.1016/0378-4266\(77\)90017-6](https://doi.org/10.1016/0378-4266(77)90017-6)
6. Altman, E. I., Hartzell, J., & Peck, M. (1998). Emerging market corporate bonds — a scoring system. En R. Levich (Ed.), *Emerging Market Capital Flows* (pp. 391-400). Springer.
7. Athey, S. (2018). The impact of machine learning on economics. En A. Agrawal, J. Gans & A. Goldfarb (Eds.), *The Economics of Artificial Intelligence: An Agenda* (pp. 507-552). University of Chicago Press.
8. Banca Personal Nación (2020). Cotizaciones históricas. <https://www.bna.com.ar/Personas>
9. Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405-417. <https://doi.org/10.1016/j.eswa.2017.04.006>
10. Bazarbash, M. (2019). Fintech in financial inclusion: machine learning applications in assessing credit risk. *IMF Working Paper No. 19/109*. Social Science Research Network. <https://ssrn.com/abstract=3404066>
11. Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71-111. <https://doi.org/10.2307/2490171>
12. Bharath, S. T., & Shumway, T. (2008). Forecasting default with the Merton distance to default model. *The Review of Financial Studies*, 21(3), 1339-1369. <https://doi.org/10.1093/rfs/hhn044>
13. Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees*. CRC Press.
14. Caro, N. P., Díaz, M., & 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.
15. Caro, N. P., Guardiola, M., & Ortiz, P. (2018). Árboles de clasificación como herramienta para predecir dificultades financieras en empresas latinoamericanas a través de sus razones contables. *Contaduría y Administración*, 63(1), 1-14. <http://dx.doi.org/10.22201/fca.24488410e.2018.1148>
16. Dirección Provincial de Estadística y Censos (2020). Índices de Precios al Consumidor <http://www.estadistica.sanluis.gov.ar/indice-de-precios-al-consumidor-san-luis/>
17. Esplin, A., Hewitt, M., Plumlee, M., & Yohn, T. L. (2014). Disaggregating operating and financial activities: implications for forecasts of profitability. *Review of Accounting Studies*, 19(1), 328-362. <https://doi.org/10.1007/s11142-013-9256-5>

18. Frugoli, E., & Pascual, J. P. (2020). La situación fiscal y financiera de las dos últimas administraciones en Argentina. <http://fcece.org.ar/wp-content/uploads/informes/situacion-fiscal-financiera.pdf>
19. Frydman, H., Altman, E. I., & Kao, D. L. (1985). Introducing recursive partitioning for financial classification: the case of financial distress. *The Journal of Finance*, 40(1), 269-291. <https://doi.org/10.1111/j.1540-6261.1985.tb04949.x>
20. Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1999). *Análisis multivariante*. Prentice Hall.
21. Instituto Nacional de Estadística y Censos (INDEC). (2020). Estadísticas económicas: índice de precios al consumidor. <https://www.indec.gov.ar/indec/web/Nivel3-Tema-3-5>
22. Jones, S., & Hensher, D. A. (2004). Predicting firm financial distress: a mixed logit model. *The Accounting Review*, 79(4), 1011-1038. <https://doi.org/10.2308/accr.2004.79.4.1011>
23. Merton, R. C. (1974). On the pricing of corporate debt: the risk structure of interest rates. *The Journal of Finance*, 29(2), 449-470. <https://doi.org/10.2307/2978814>
24. Missim, D., & Penman, S. H. (2001). Ratio analysis and equity valuation: from research to practice. *Review of Accounting Studies*, 6, 109-154. <https://doi.org/10.1023/A:1011338221623>
25. Montalván, S. M., Delgado, F. I. A., O'Shee, D. F., & Yamashiro, M. A. (2011). Determinantes de la insolvencia empresarial en el Perú. *Academia. Revista Latinoamericana de Administración*, 47, 126-139.
26. Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87-106. <https://doi.org/10.1257/jep.31.2.87>
27. Myers, S. C. (1984). The capital structure puzzle. *The Journal of Finance*, 39, 575-592.
28. Nanni, L., & Lumini, A. (2009). An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 36(2), 3028-3033. <https://doi.org/10.1016/j.eswa.2008.01.018>
29. Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109-131. <https://doi.org/10.2307/2490395>
30. Sandin, A. R., & Porporato, M. (2007). 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. <https://doi.org/10.1108/10569210710844372>

31. Tan, P. N., Steinbach, M., & Kumar, V. (2016). *Introduction to Data Mining*. Pearson Education India.
32. Tascón, M., & Castaño, F. (2012). Variables y modelos para la identificación y predicción del fracaso empresarial: revisión de la investigación empírica reciente. *Revista de Contabilidad*, 15(1), 7-58. [https://doi.org/10.1016/S1138-4891\(12\)70037-7](https://doi.org/10.1016/S1138-4891(12)70037-7)
33. Wahlen, J. M., Baginski, S. P., & Bradshaw, M. (2014). *Financial Reporting, Financial Statement Analysis and Valuation*. Cengage Learning.
34. Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59-82. <https://doi.org/10.2307/2490859>



CUADERNOS DE ECONOMÍA

ISSN 0121-4772

ARTÍCULOS

- JOHN GARCÍA RENDÓN, MANUEL CORREA GIRALDO Y ALEJANDRO GUTIÉRREZ GÓMEZ
Efecto de la entrada en operación de la central hidroeléctrica más grande y de las energías renovables no convencionales en Colombia sobre el precio de bolsa 1
- ROBERTO ARPI, LUIS ARPI, RENE PAZ PAREDES Y ANTONIO SÁNCHEZ-BAYÓN
Desigualdad del ingreso laboral por grupo étnico en el Perú durante la pandemia de COVID-19 25
- DANTE DOMINGO TERRENO, JORGE ORLANDO PÉREZ Y SILVANA ANDREA SATTLER
Un modelo jerárquico para la predicción de insolvencia empresarial. Aplicación de análisis discriminante y árboles de clasificación 51
- SARA FLORES Y PAUL CARRILLO-MALDONADO
¿Mejora el comercio internacional con un tratado de libre comercio? El caso de Alianza del Pacífico 77
- MANUELA MAHECHA ALZATE
A theoretical framework to study accumulation regimes and crises in Colombia 99
- CAROLINA ROMÁN Y HENRY WILLEBALD
Transferencias de ingresos entre actividades productivas en Uruguay (1955-2022). Estabilidad, cambio y creciente dispersión 127
- JENNY LISSETH AVENDAÑO LÓPEZ, ÓSCAR HERNÁN CERQUERA LOSADA Y CRISTIAN JOSÉ ARIAS BARRERA
Modelo de probabilidad según condiciones socioeconómicas para el trabajo infantil rural y urbano en Colombia 175
- INMACULADA CEBRIÁN Y GLORIA MORENO
The path to labour stability for young spanish workers during the great recession 195
- MARÍA CRISTINA BOLÍVAR RESTREPO, LAURA CARLA MOISÁ ELICABIDE Y NICOLÁS ALBERTO MORENO REYES
Informalidad laboral femenina en Colombia: composición y determinantes socioeconómicos 231
- CÉSAR AUGUSTO GIRALDO PRIETO, JESÚS SANTIAGO SAAVEDRA SANTA Y LÍA CECILIA VALENCIA ÁLVAREZ
La educación financiera como mediadora entre la planeación financiera y el desempeño financiero en microemprendedores del sector solidario 265
- ERIKA SIERRA PÉREZ Y ALEXANDER VILLARRAGA ORJUELA
Efectos del desajuste educativo sobre los salarios de los jóvenes de 18 a 28 años: análisis en países de la Comunidad Andina 297
- ALEXANDER SANTOS NIÑO, WILDER ARLEHT ANGARITA OSORIO Y JOSÉ LUIS ALVARADO MARTÍNEZ
Estudio de la dinámica de préstamos y depósitos en un sistema económico cerrado a partir de modelos cinéticos de distribución 327
- JESÚS BOTERO, CRISTIAN CASTRILLÓN, ÁLVARO HURTADO, HUMBERTO FRANCO Y CHRISTIAN VARGAS
Formality and informality in an emerging economy: The case of Colombia 345

RESEÑA

- JUAN CARLOS VILLAMIZAR
The World that Latin America Created. The United Nations Economic Commission for Latin America in the Development Era de Margarita Fajardo, 2021

375

