



UNIVERSIDAD
DE LA REPUBLICA
URUGUAY

dECON
Facultad de Ciencias Sociales

Un modelo para predecir solicitudes de concurso de acreedores en Uruguay

Guillermina Suárez

Tesis de Maestría presentada al Programa de Maestría en Economía, Facultad de Ciencias Sociales de la Universidad de la República, como parte de los requisitos necesarios para la obtención del título de Magister en Economía.

Director de tesis:

Jorge Ponce

Montevideo – Uruguay

Enero 2022

RESUMEN

El objetivo de este estudio es elaborar un modelo predictivo para la caída en concurso de acreedores de empresas uruguayas, utilizando principalmente la información proveniente de los estados financieros, en forma de ratios. El método de estimación es el análisis logístico y las variables incluidas en el modelo refieren a la rentabilidad de la empresa, su nivel de liquidez y su estructura de financiamiento. Las tres variables resultan significativas a la hora de predecir hasta un año antes la solicitud de concurso de acreedores: empresas con menor rentabilidad, menor liquidez y mayor endeudamiento tienen mayor probabilidad de participar en un concurso de acreedores. El modelo es capaz de clasificar correctamente en promedio un 84% de las firmas, y este porcentaje es válido tanto para las empresas concursadas como para las saludables. En líneas generales, los resultados se mantienen cuando las observaciones clasificadas no pertenecen a la muestra utilizada para estimar el modelo. Asimismo, independientemente de la proporción de empresas saludables y concursadas consideradas, el modelo es bueno para discriminar ambos tipos de empresas, y es robusto al método a través del cual son seleccionadas las empresas saludables. El incremento del horizonte temporal de predicción y la introducción de dinámica temporal aparecen como importantes líneas de investigación a futuro.

ABSTRACT

The purpose of this study is to develop a bankruptcy prediction model for Uruguayan companies, mainly using information from their financial statements, in the form of ratios. The estimation method is the logistic analysis and the variables included in the model refer to the company's profitability, its level of liquidity and financing structure. All three variables are significant when predicting up to a year before filing for bankruptcy: companies with lower profitability, lower liquidity and higher debt are more likely to fall in bankruptcy. The model is capable of correctly classify on average 84% of the firms, and this percentage is valid for both bankrupt and healthy companies. In general, the results remain valid when the predicted observations do not belong to the sample used to estimate the model. Further, regardless of the proportion of healthy and bankrupt companies considered, the model is fairly good at disentangling both types of companies, and it is robust to the healthy companies' selection methods. The increase in the prediction time horizon and the introduction of temporal dynamics emerge as important lines of future research.

Tabla de contenido

1.	Introducción	1
2.	Revisión Bibliográfica	3
3.	Datos y metodología	7
3.1	Datos	7
3.2	Modelo empírico	9
3.3	Selección de la muestra y grado de desbalance	11
3.4	Medidas de desempeño	14
4.	Resultados	15
4.1	Estimación inicial	15
4.2	Incremento del ratio de empresas saludables vs concursadas	18
5.	Pruebas de Robustez	22
5.1	Cambios en los métodos de selección de empresas saludables	22
5.2	Predicción fuera de la muestra	24
6.	Ampliación del horizonte temporal y dinámica	26
6.1	Predicción dos períodos antes	26
6.2	Dinámica temporal	27
7.	Conclusiones	29
8.	Bibliografía	30
9.	Anexo	33

1. Introducción

La predicción de dificultades financieras o quiebra de las empresas ha sido un tema de amplio estudio en las últimas décadas. Poder anticipar dichas dificultades resulta de interés para varias partes: la propia empresa, sus empleados, sus acreedores, las instituciones financieras, inversionistas, entre otros. Estas complicaciones pueden conducir al cierre o liquidación de la empresa, lo que en mayor o menor medida, implica un costo para cada uno de los involucrados e incluso puede tener un impacto negativo en la sociedad y la economía (Tian, et al., 2015).

El objetivo de la presente investigación es elaborar un modelo predictivo para la caída en concurso de acreedores de empresas uruguayas, utilizando principalmente la información proveniente de los estados financieros, en forma de ratios, y combinada con variables referidas a las características observables de la empresa, como su sector de actividad y tamaño. La contribución de este estudio es, por una parte, la elaboración de un modelo empírico con datos de empresas uruguayas sobre las solicitudes admitidas en concurso de acreedores, y por otra, la creación de una base de datos con información contable de dichas empresas.

La motivación para realizar la investigación es evaluar el rol de los ratios financieros a la hora de prevenir los procesos concursales en Uruguay, evitando así pérdidas importantes para las partes involucradas. Para ello, resulta de importancia, y no sólo en los contextos económicos negativos, contar con información contable ordenada, sin errores y que permita visualizar de forma clara la situación financiera de las empresas.

La utilización de una definición judicial del fracaso empresarial no sólo provee un criterio objetivo para separar a la población de empresas en dos subgrupos, sino que además permite determinar el momento del fracaso con exactitud (Balcaen & Ooghe, 2006). La definición de concurso de acreedores queda determinada por la Ley de Proceso Concursal¹, que establece que la declaración judicial de concurso procede respecto de cualquier deudor que se encuentre en estado de insolvencia. Se considera en estado de insolvencia, independientemente de la existencia de pluralidad de acreedores, al deudor que no puede cumplir con sus obligaciones.

Las firmas utilizadas para elaborar la muestra de este estudio han participado en un proceso concursal en el período 2011-2020, conformando el grupo de empresas concursadas. Como contraparte, el grupo de firmas saludables se compone de empresas que cuentan con al menos un estado contable disponible en igual período y que no han participado de un concurso de acreedores. El modelo estimado no resulta específico

¹ Ley N° 18387.

de una industria o de un tamaño de firma particular, dado que se incluyen en la muestra todos los sectores de actividad de la economía plausibles de solicitar concurso de acreedores, y empresas micro, pequeñas, medianas y grandes.

El método de estimación utilizado es el análisis logístico y las variables incluidas en el modelo refieren a la rentabilidad de la empresa, su nivel de liquidez y su estructura de financiamiento. Las tres variables resultan significativas a la hora predecir hasta un año antes la solicitud de concurso de acreedores. Empresas con menor rentabilidad, menor liquidez y mayor endeudamiento tienen mayor probabilidad de participar en un concurso de acreedores. El modelo, estimado en 100 muestras aleatorias diferentes, es capaz de clasificar correctamente en promedio un 84% de las firmas, y dicho porcentaje es válido tanto para las empresas concursadas (sensibilidad) como para las saludables (especificidad). Esto indica que un 16% de las firmas de cada grupo son clasificadas incorrectamente. El área bajo la curva ROC da señales de que el modelo es adecuado para distinguir entre ambos tipos de firmas, con un valor mínimo de 0,896 y un máximo de 0,935, en las 100 estimaciones.

En primera instancia, se consideran en el modelo los ratios en niveles y únicamente para el balance cerrado inmediatamente antes de la solicitud de concurso, para luego testear diferentes tamaños muestrales, diferentes métodos de selección de las empresas saludables, la ampliación del horizonte temporal de predicción y la inclusión de la evolución en el tiempo de las variables. Los resultados sugieren que, independientemente de la proporción de empresas saludables y concursadas consideradas, el modelo es bueno para identificar y separar ambos tipos de firmas, aunque es posible ver una reducción en el efecto de las variables sobre la probabilidad de solicitar concurso a medida que se incrementa la cantidad de empresas saludables en la muestra en relación con las concursadas. Por otra parte, los resultados indican que el modelo no es altamente sensible al método a través del cual son seleccionadas las empresas saludables. La estimación que utiliza la información de dos períodos previos a la solicitud no muestra un efecto de la rentabilidad sobre la probabilidad de solicitar concurso y reduce el desempeño respecto al modelo inicial, pero continúa siendo buena para distinguir ambos tipos de firmas. Por último, la inclusión de la dinámica temporal no brinda los mismos resultados, siendo una importante línea de investigación a futuro.

El trabajo se estructura de la siguiente forma. En la sección 2 se realiza una breve revisión de la literatura, en la sección 3 se presentan los datos y la metodología a utilizar, los resultados y las pruebas de robustez son presentados en las secciones 4 y 5 respectivamente. En la sección 6 se realiza la introducción de la dinámica temporal y, por último, en la sección 7 se incluyen las conclusiones.

2. Revisión Bibliográfica

El fracaso empresarial es un fenómeno complejo para el que las variables causales son muchas veces de difícil acceso (Bardos, 2007). Sin embargo, la disponibilidad de información financiera, en conjunto con su procesamiento y análisis, han hecho posible la predicción de dicho fenómeno. En particular, los ratios financieros han demostrado ser buenos síntomas de la salud de la empresa, apareciendo como una importante herramienta a la hora de predecir estas dificultades. Desde la década de los sesenta, numerosos estudios buscan aplicar diferentes métodos estadísticos con el fin de poder anticipar el fracaso empresarial (Sandin & Porporato, 2008). Uno de los pioneros en el análisis de ratios financieros con fines predictivos es el trabajo de Beaver (1966), quien desarrolla un modelo univariante para predecir el fracaso de las firmas, y provee los primeros resultados que indican la utilidad de la información contable para este objetivo, señalando además las diferentes capacidades predictivas o relevancia de los ratios analizados.

Pocos años después, Altman (1968) introduce un modelo estadístico basado en análisis discriminante múltiple (MDA por sus siglas en inglés), llamado Z-score, que logra clasificar correctamente un 95% del total de 66 firmas incluidas en la muestra. Este modelo constituye la base de un nuevo modelo ajustado (Altman, et al., 1977) y de sucesivos estudios que buscan aplicarlo en diferentes datos, períodos y países. Edminster (1972) logra superar la limitante asociada al tamaño de los activos, al confirmar los resultados obtenidos por Altman para empresas medianas y grandes en una muestra de pequeñas empresas. Taffler (1983) utiliza las mismas técnicas estadísticas para desarrollar un Z-model para firmas del Reino Unido, en tanto que Sandin y Porporato (2008) aplican el Z-score en firmas argentinas en la década de los noventa, obteniendo buenos resultados predictivos.

En la misma línea que estos estudios y para el caso de Uruguay, un modelo para predecir problemas financieros en firmas manufactureras es elaborado por Pascale (1988). El modelo es estimado mediante MDA en una muestra de 44 firmas manufactureras con problemas financieros (liquidaciones, bancarrotas, acuerdos con acreedores o similares) y 41 firmas del mismo sector que no experimentaron ninguno de estos problemas. Incluyendo tres ratios vinculados al nivel de actividad de la firma, la rentabilidad y la estructura del endeudamiento, dicho modelo es capaz de clasificar correctamente un 92% de las empresas, y en particular cerca del 98% de las firmas con problemas.

En la década de los ochenta comienzan a aparecer nuevos modelos basados en la probabilidad condicional, considerados menos demandantes que los basados en MDA (Balcaen & Ooghe, 2006). Ohlson (1980) fue pionero en aplicar el análisis logístico (LA por sus siglas en inglés) en la predicción de bancarrota. Su modelo, que incluye 105 firmas en bancarrota y 2.058 saludables, determina cuatro factores básicos que afectan la probabilidad de fracaso: el tamaño de la firma, su estructura financiera, su performance y su

liquidez, y logra clasificar correctamente el 96% de las firmas. Zmijewski (1984), pionero en el análisis Probit, también utiliza una muestra desbalanceada para estimar su modelo, e incluye tres ratios (rentabilidad, apalancamiento financiero y liquidez) para analizar los posibles sesgos asociados a los modelos de predicción y su importancia. Posteriormente otros modelos basados en el análisis logístico o Probit, entre otros, son elaborados por Zavgren (1985) para firmas del sector industrial de Estados Unidos; Ho, et al. (2013) para la industria del papel en el mismo país; Becchetti y Sierra (2003) para firmas manufactureras en Italia; y Keasey y McGuinness (1990) o Charitou, et al. (2004) para empresas industriales de Reino Unido.

Aziz y Dar (2006) realizan un análisis de 89 estudios que desarrollan modelos de predicción de bancarrota, y encuentran que un 64% de los métodos aplicados en los mismos corresponde a los métodos estadísticos clásicos, en tanto que MDA y LA en conjunto constituyen un 77% de dichos métodos. Recientemente, y a la par de las mejoras computacionales, se han desarrollado modelos que aplican métodos alternativos en la predicción de dificultades financieras, como redes neuronales, árboles de decisión o *support vector machine*². Varios de estos métodos son analizados por Balcaen y Ooghe (2004) con relación a los métodos estadísticos clásicos, y llegan a la conclusión de que no es claro que los primeros puedan alcanzar una mejor performance en predecir el fracaso corporativo que los últimos, si bien son más sofisticados e intensivos computacionalmente. En resumen, aunque existen diferentes métodos aplicables en estos tipos de modelos, no parece haber consenso sobre cuál provee mejores estimaciones, y en muchos casos la elección del mismo se encuentra sujeta a la disponibilidad de información, el desarrollo computacional o la interpretación de los resultados. En esta investigación, por su facilidad de estimación e interpretación, el método de estimación seleccionado es el análisis logístico.

Como es mencionado por Zhou, et al. (2012), a la hora de construir los modelos de predicción de dificultades financieras, un tema importante refiere a la disponibilidad de información. La mayor parte de los estudios utilizan la información proveniente de los estados contables, y algunos añaden variables macroeconómicas, pero qué variables incluir en el modelo no es una pregunta fácil de responder. Muchos estudios optan por seleccionar aquellos ratios más frecuentemente nombrados en la literatura o en base a su experiencia previa (Beaver, 1966; Altman, 1968; Ohlson, 1980; Taffler, 1983; Pascale, 1988; Becchetti & Sierra, 2003). Sin embargo, Zhou, et al. (2012) identifican más de 500 variables diferentes utilizadas en 128 artículos, y señala que, dado que diferentes países pueden tener distintas políticas contables o tipos de estados contables por sectores de actividad, muchas veces es difícil replicar algunos de los ratios mencionados. Por otra parte, la selección de los ratios en base a su popularidad puede ser problemática por

² En Bae (2012) se puede encontrar una breve descripción de estos métodos y su aplicación para el caso de la industria manufacturera en Corea del Sur.

el hecho de que precisamente esos ratios tienen mayor probabilidad de ser manipulados y por tanto ser menos confiables (Balcaen & Ooghe, 2006). Por último, Ooghe, et al. (1995) advierten que la elección de los ratios financieros como variables independientes del modelo asume de forma subyacente que los estados contables presentan una imagen fiel de la situación de la firma, pero esto no siempre ocurre.

Kirkos (2015) plantea la inexistencia de un marco de referencia capaz de definir ciertas prácticas deseables en términos de la selección de los *inputs* del modelo. Diferentes métodos han sido utilizados, solos o combinados, en la literatura revisada por el autor: árboles de decisión, *forward and backward stepwise search, pruning*, análisis de componentes principales, matrices de correlaciones, entre otros. Pero destaca el hecho de que casi todos los estudios implementan aproximaciones diferentes, reafirmando la falta de una estrategia de selección de variables ampliamente aceptada. La ausencia de un fundamento teórico subyacente en la elección de los ratios relevantes, puede provocar que los resultados sean específicos de la muestra utilizada (Oz & Yelkenci, 2017). En esta investigación, teniendo presente las limitaciones comentadas, los ratios incluidos en el modelo son seleccionados en base a tres aspectos: parsimonia del modelo (reducido número de variables), simplicidad en el cálculo, y referencia a diferentes dimensiones de la situación de la empresa (independencia).

Por último, algunos estudios recientes señalan el carácter estático que generalmente tienen este tipo de modelos predictivos y destacan la importancia de incluir la historia de la firma (Shumway, 2001; du Jardin, 2014; du Jardin, 2017; Volkov, et al., 2017). Siempre que utilizamos los estados contables de una empresa, estamos mirando únicamente una foto de un momento determinado del tiempo, en tanto que la quiebra, o en este caso la solicitud de concurso, es el desenlace de un proceso que se desarrolla a lo largo de un período de tiempo (Kirkos, 2015). Por lo tanto, resulta razonable pensar que la inclusión de la dinámica temporal puede ser importante. Shumway (2001) elabora un modelo de riesgo o multi Logit que incorpora toda la información histórica disponible de las firmas, y permite la inclusión de covariables que cambian en el tiempo. En igual dirección, du Jardin (2014, 2017) busca métodos que permitan incluir la trayectoria en el tiempo, a fin de encontrar mejores predicciones en el mediano plazo. Estos tipos de modelos dinámicos resultan más exigentes, en cuanto a la cantidad de información disponible a lo largo del tiempo para cada firma incluida, en relación con los llamados modelos estáticos. En este estudio se realiza una aproximación inicial y básicamente exploratoria del tema, quedando pendiente su abordaje en mayor profundidad en estudios futuros.

Finalmente cabe señalar, que gran parte de las investigaciones relativas al análisis de ratios financieros son realizadas para países desarrollados, o para un sector en particular de una economía. En ese sentido, esta investigación no sólo contribuye con un nuevo modelo para un país en desarrollo, sino que además incluye todos los sectores de la economía plausibles de solicitar concurso de acreedores. Por otro lado, como fue

mencionado, una limitación frecuente de varios estudios destinados a construir modelos de predicción de bancarrota es la disponibilidad de información, no sólo por el hecho de que la bancarrota es un evento poco frecuente, pero también por la dificultad para acceder a estados contables completos y sin errores. Aziz y Dar (2006) encuentran que un 42% de los 89 artículos revisados utilizan muestras de menos de 100 firmas, y además muchas muestras se componen de firmas que cotizan públicamente en bolsa, dado que es más fácil acceder a su información. En esta línea, la presente investigación contribuye a la literatura con un modelo que incluye una cantidad considerable de firmas que fracasan, y de diferentes tamaños.

3. Datos y metodología

3.1 Datos

La base de datos es de elaboración propia con información proporcionada por la Liga de Defensa Comercial (LIDECO), sujeta al anonimato de las empresas incluidas. La misma cuenta con información de empresas concursadas, que son aquellas sobre las que se presentó³ y fue admitida una solicitud de concurso de acreedores en el período 2011-2020. Se incluyen en la base las firmas que tienen al menos un estado contable cerrado en los tres años inmediatamente anteriores a la solicitud de concurso⁴. En total la muestra incluye 291 empresas concursadas con 831 estados contables: 254 cuentan con tres estados contables cerrados inmediatamente anteriores a la solicitud de concurso, 32 con dos y 5 con un solo estado contable. La distribución de firmas por año de solicitud y por tipo de concurso (voluntario o necesario) es la siguiente:

Tabla 1: Cantidad de firmas concursadas por año de solicitud y tipo de concurso

Año de Presentación	Necesario	Voluntario	Total
2011	0	13	13
2012	1	10	11
2013	0	9	9
2014	1	13	14
2015	1	37	38
2016	2	45	47
2017	0	51	51
2018	0	43	43
2019	2	41	43
2020	1	21	22
Total	8	283	291

Fuente: elaboración propia.

Por otra parte, las empresas saludables son todas aquellas disponibles en la base de datos de LIDECO⁵ que no tuvieron una presentación concursal admitida en el período, y cuentan con al menos un estado contable cerrado entre 2008 y 2020⁶. Esta muestra se compone de 3.720 firmas, que suman un total de 13.284 estados contables. No se incluyen en la base las empresas de intermediación financiera ni vinculadas al Estado

³ De forma voluntaria, considerado un concurso voluntario, o por un tercero interesado, considerado un concurso necesario.

⁴ Para obtener la información proveniente de los estados contables, se consultaron los expedientes judiciales de las presentaciones a concurso admitidas en el período, disponibles en la base de LIDECO.

⁵ Dicha base de datos se nutre de estados contables que bien fueron presentados en la Auditoría Interna de la Nación (AIN) y adquiridos por LIDECO o proporcionados por las empresas asociadas a la institución.

⁶ Las solicitudes de concurso corresponden al período 2011-2020, los estados contables hasta tres años cerrados previos a la solicitud, por lo que el primer estado contable corresponde al año 2008 y el último a 2020.

(administración pública, entes autónomos, servicios descentralizados, etc.), por no encontrarse comprendidas en la Ley de Proceso Concursal.

Además de la información contable, se cuenta con el sector de actividad de cada una de las empresas. Para ambos tipos de firmas el sector mayoritario es Comercio:

Tabla 2: Porcentaje de firmas por sector de actividad

Sector de Actividad	Concursadas	Saludables
Primario	4,5%	11,9%
Industria	23,7%	16,0%
Servicios	22,3%	28,4%
Construcción	9,3%	4,6%
Comercio	40,2%	39,1%
Total	100%	100%

Fuente: elaboración propia.

Por último, se construye una variable que busca medir el tamaño de las empresas en base a sus ingresos operativos, y según las franjas determinadas en la categorización de micro, pequeñas y medianas empresas, establecida en el decreto 504/2007⁷. Dichas franjas se determinan en unidades indexadas, por lo que para cada año se define el valor de las mismas en pesos uruguayos, utilizando el valor promedio anual de la unidad indexada. De esta forma, para cada empresa y estado contable, queda definida la variable tamaño. Más del 70% de las empresas concursadas son pequeñas y medianas, en tanto que más del 80% de las saludables son medianas y grandes.

Tabla 3: Porcentaje de firmas por tamaño

Tamaño	Concursadas	Saludables
Grande	10,8%	35,9%
Mediana	41,9%	46,9%
Pequeña	29,4%	9,6%
Micro	10,2%	2,9%
Sin clasificar*	7,7%	4,7%
Total	100%	100%

Fuente: elaboración propia.

Notas: *Para las empresas cuyos estados contables son presentados en dólares americanos no se calcula la variable “tamaño”, así como para aquellas cuyos ingresos operativos no son positivos.

⁷ Ley N° 16.201: Declaración de interés nacional. Promoción de las micro, pequeñas y medianas empresas y Decreto N° 504/007: Reformulación de la categorización de micro, pequeñas y medianas empresas.

3.2 Modelo empírico

En este contexto nos encontramos frente a un problema de clasificación: deseamos separar e identificar a las empresas como concursadas o saludables. Por su facilidad de estimación e interpretación de los resultados, el método seleccionado es el análisis logístico.

Por simplicidad, se parte de un modelo simple con tres variables o ratios, que refieren a tres importantes aspectos relativos a la salud de la empresa: la rentabilidad, la liquidez y el endeudamiento. Esperamos que empresas con mayor rentabilidad, mayor liquidez y menor endeudamiento, tengan menor probabilidad de participar en un concurso de acreedores. La construcción de las variables en forma de ratios hace posible que los efectos de la inflación no deban ser tomados en cuenta, dado que afectan de igual forma al numerador y denominador, lo mismo ocurre para los estados contabilizados en diferentes monedas.

Como es señalado por Edmister (1972), es posible obtener mejores resultados con un menor número de ratios seleccionados cuidadosamente. Asimismo, como los ratios tienden a ser muy similares en la información que contienen o desean medir, debemos tener cuidado de que los mismos den cuenta de diferentes dimensiones de la situación financiera de la empresa.

Por otra parte, Grice y Dugan (2001) alertan sobre la pérdida de precisión generada al aplicar modelos sobre períodos e industrias distintos a aquellos en los que fueron desarrollados. Es común que las variables seleccionadas para un modelo específico fueran las mejores discriminantes para esa muestra, pero que no lo sean bajo otras condiciones. Por lo general, la importancia relativa de cada variable suele variar en el tiempo. En particular concluyen que tanto el modelo de Ohlson (1980) como el de Zmijewski (1984) son sensibles a los períodos de tiempo en los que son utilizados, sin embargo, este último no resulta sensible a las clasificaciones por industria. Oz y Sigma-Mugan (2018) analizan la validez de 5 modelos diferentes, aplicados sobre datos provenientes de países en desarrollo, y encuentran que es importante reestimar el modelo cuando es aplicado en las economías emergentes, con el objetivo de reducir la pérdida de precisión.

Se incluyen por lo tanto en el modelo tres variables de control: el año de balance, el sector de actividad y el tamaño de la empresa. La inclusión de las *dummies* por año de balance pretende capturar efectos macroeconómicos, vinculados a la fluctuación de la actividad económica o las condiciones externas, asociados al año en cuestión. La inclusión del sector de actividad controla estas mismas fluctuaciones, pero afectando a un sector de la economía en particular, al mismo tiempo que las diferencias entre los ratios propias de la operativa de cada sector (por ejemplo, puede ser que un sector determinado necesite operar

con mayor liquidez que otro). Por último, la variable tamaño da cuenta de las diferencias en los ratios asociadas a esta característica de la empresa⁸.

El modelo incluye un intercepto y tres variables explicativas, más las tres variables de control. Buscamos estimar la probabilidad de que una empresa sea admitida en concurso de acreedores, $P(X_{ik}, \beta)$, donde $X_{ik} = (X_{i1}, X_{i2}, \dots, X_{ik})$ es el vector de k predictores para cada observación i y $\beta = (\beta_1, \beta_2, \dots, \beta_k)$ es el vector de k parámetros a estimar. P es una función de probabilidad, en este caso la función logística. La estimación se realiza maximizando el logaritmo de la verosimilitud: $l(\beta)$.

$$P = (1 + \exp\{-y_i\})^{-1} \quad (1)$$

$$y_i = \beta_0 + \beta_1 ROA_i + \beta_2 LIQ_i + \beta_3 FINL_i + \beta_4 AÑO_i + \beta_5 INDUSTRIA_i + \beta_6 TAMAÑO_i + \varepsilon_i \quad (2)$$

Donde:

y = Variable dependiente binaria que vale 1 si la firma participa de un proceso concursal en el período y cero en otro caso.

ROA = Resultado Operativo⁹ / Activos Totales. LIQ = Activo Corriente / Pasivo Corriente. FINL = Pasivo No Corriente¹⁰ / Activos Totales.

AÑO = Año del balance. INDUSTRIA = Sector de actividad de la firma. TAMAÑO = Tamaño de la firma.

En primera instancia, se consideran en el modelo las variables en niveles y únicamente para el balance cerrado inmediatamente antes de la solicitud de concurso, dejándonos un total de 282 empresas concursadas e igual cantidad de balances, y 3.185 empresas saludables con 10.118 balances¹¹. Esta especificación es tomada como base, para luego probar diferentes tamaños muestrales, métodos de selección de las empresas saludables, ampliar el horizonte temporal de predicción e incluir la dinámica en el tiempo de las variables.

⁸ Otra posible variable de control, que no fue posible capturar en esta instancia, es la antigüedad o edad de la firma, considerada como la diferencia entre el año de balance y el inicio de actividades.

⁹ Resultado Operativo Neto: comprende los ingresos operativos o ventas netas menos el costo de ventas, los gastos de administración y ventas y las amortizaciones.

¹⁰ Como variable alternativa para el endeudamiento se consideró el apalancamiento o endeudamiento total, definido como Pasivos Totales sobre Activos Totales. Sin embargo, se encontró una elevada correlación negativa entre dicho ratio y el de rentabilidad (Resultado Operativo/Activos) en las diferentes muestras aleatorias seleccionadas. Por lo tanto, se incluyó en el modelo el endeudamiento de largo plazo sobre activos totales como variable alternativa, proporcionando buenos resultados. Resultaría interesante, para futuras investigaciones, profundizar sobre la estructura de financiamiento de las firmas (relaciones pasivo activo de corto y largo plazo, diferenciación entre deudas financieras y comerciales), y cómo afecta la probabilidad de solicitud de concurso.

¹¹ Se excluyen los estados contables correspondientes a los años 2008 y 2009, dado que el primer año de solicitud de concurso es 2011 y por lo tanto el primer estado contable cerrado inmediatamente previo a dicha solicitud es 2010.

3.3 Selección de la muestra y grado de desbalance

Es común, al elaborar este tipo de modelos, encontrarnos con una muestra desbalanceada: existen disponibles muchas más observaciones de empresas saludables que de empresas concursadas. Este desbalance proviene de la distribución real en la economía: la proporción de empresas que fallan o en este caso solicitan concurso, es muy pequeña en el total de empresas, es decir que el evento que buscamos predecir es poco común. Este hecho nos lleva a tomar dos decisiones vinculadas al tamaño y selección de la muestra: por un lado, qué proporción de empresas saludables y concursadas considerar, y por otro, cómo seleccionar dichas empresas. Modelos como el de Altman (1968) o Beaver (1966) seleccionan las empresas saludables en igual proporción que las malas, emparejándolas por industria y/o tamaño de las empresas. Otros estudios como el de Ohlson (1980) o Shumway (2001) utilizan todas las observaciones disponibles.

Básicamente contamos con tres métodos de muestreo en este caso: submuestrear la clase mayoritaria (seleccionar una parte de las empresas saludables disponibles), utilizar todas las observaciones disponibles, y sobremuestrear la clase minoritaria (muestrear consecutivamente las empresas concursadas hasta alcanzar una muestra balanceada). Zhou (2013), en su estudio sobre los efectos de diferentes métodos de muestreo en la performance de los modelos de predicción de bancarrota, concluye que, cuando hay cientos de observaciones disponibles de empresas que fallan, el método de submuestreo es mejor que el de sobremuestreo, ya que no hay una diferencia significativa en la performance de los modelos, y el último método requiere de mayor tiempo de cálculo y computacional. Por tanto, nos concentraremos entonces en las diferencias de las dos primeras opciones.

Como es señalado por Veganzones y Séverin (2018), los modelos de clasificación, en este caso de predicción de bancarrota, se encuentran diseñados para optimizar la precisión general, lo cual puede resultar en el desarrollo de modelos subóptimos: al centrarse en la correcta clasificación de la clase mayoritaria e ignorar a la minoritaria, se pueden obtener igualmente buenos resultados de precisión general. Para ilustrar esta idea, pensemos en un modelo con una muestra muy desbalanceada: 10.000 empresas saludables y 100 empresas malas. Si este modelo clasifica a todas las empresas como buenas, clasifica correctamente un 99% de los casos, pese a que todas las malas se clasificaron incorrectamente. Por lo tanto, estos tipos de modelos tienden a clasificar correctamente a la clase mayoritaria y erróneamente a la minoritaria, es decir se encuentran sesgados hacia la clase mayoritaria.

Esto nos lleva a su vez a considerar las definiciones de error de tipo I y II: el primero ocurre cuando clasificamos como saludable una empresa concursada, y el segundo cuando una empresa saludable es clasificada como una concursada. Tian, et al. (2015), analizan diferentes métodos de muestreo en un modelo Logit, y señala que la elección del método no puede realizarse sin tener en cuenta el objetivo del modelo y los errores de clasificación que deseamos, y por lo tanto, la probabilidad de corte. En un caso donde

deseamos separar las empresas buenas de las malas, y el error de tipo I y II implican los mismos costos, es recomendable seleccionar las empresas saludables en lugar de utilizar todos los datos disponibles. Pero si los costos son asimétricos, es necesario además elegir un punto de corte apropiado para reducir los errores de clasificación. Si el interés es encontrar una probabilidad precisa de falla, el método que utiliza todas las observaciones disponibles logra aproximarse mejor al valor real de dicha probabilidad.

Por otra parte, si el método elegido consiste en seleccionar las empresas saludables, otro aspecto importante a definir es cómo seleccionarlas: podemos realizar un match por una o más características de las empresas concursadas, o podemos seleccionarlas aleatoriamente. Un posible problema asociado al emparejamiento de las firmas saludables en base a un criterio, como la industria o el tamaño, señalado por Ooghe, et al. (1995), es que las variables asociadas a dicho criterio pueden ser explicativas del fracaso de la firma, generando por tanto un sesgo en los resultados de la clasificación. Además, al seleccionar las firmas saludables en base a características de las empresas malas, es posible que la muestra de empresas saludables no sea representativa de la población real.

Por último, es necesario mencionar que, independientemente del método seleccionado, existen dos tipos de sesgo inevitablemente asociados a estos modelos, que son analizados por Zmijewski (1984). Por un lado, el sesgo denominado *choice-based*, se encuentra asociado al hecho de que primero observamos la variable dependiente, es decir, identificamos a las empresas como fallidas o saludables, y luego las incluimos en el modelo, muchas veces sobre representando a las empresas malas. Este sesgo tiende a reducirse a medida que la muestra se aproxima a la proporción real de firmas fallidas en el total. Por otro lado, el sesgo denominado de selección o *complete data*, ocurre cuando incluimos en la muestra empresas que tienen información completa, y las observaciones con información incompleta no son aleatorias en la población. Si las firmas con mayor probabilidad de fallar suelen tener datos incompletos (no han cerrado estados contables, no llevan contabilidad, etc.), la probabilidad de fallar de una firma dado que tiene datos completos, es menor que la probabilidad de fallar de una observación seleccionada aleatoriamente de la población.

Si bien el análisis realizado por Zmijewski (1984) concluye que ambos sesgos no implican cambios significativos en la clasificación general y en los ratios de predicción, o no indican diferentes resultados cualitativos en los modelos testeados, Balcaen y Ooghe (2006) remarcan la importancia de tener en cuenta los problemas de selectividad de muestreo, que pueden traer aparejados sesgos en los parámetros, afectar los niveles de precisión y generar restricciones en el uso predictivo de los modelos.

A su vez, podríamos pensar que existe un sesgo en esta muestra en particular: las observaciones son tomadas de una base que podría no ser representativa de la economía, precisamente por cómo fueron recolectados

los datos, es decir, puede existir una tendencia a incluir ciertos tipos de empresas saludables. Para intentar reducir este posible sesgo, se decide utilizar una submuestra de las empresas saludables disponibles, seleccionándolas aleatoriamente de forma tal que la distribución final por sectores de actividad sea similar a la distribución sectorial de la economía. Para ello, en base a información disponible del Instituto Nacional de Estadística (INE)¹², se determina una distribución promedio por sector de actividad¹³ y se aplica la misma para seleccionar aleatoriamente a las empresas saludables de la base inicial, resultando en la siguiente distribución:

Tabla 4: Distribución promedio por sector de actividad, similar a la economía

Sector de Actividad	Saludables
Primario	1,5%
Industria	9,5%
Servicios	50%
Construcción	5%
Comercio	34%
Total	100%

Fuente: elaboración propia.

Con este procedimiento la base de empresas saludables queda compuesta por 1.608 empresas y un total de 4.761 balances. A partir de esta muestra, las empresas son seleccionadas en forma aleatoria, a fin de reducir el posible sesgo de emparejamiento. En cuanto a la proporción de empresas saludables y concursadas, se partirá de una muestra con 50% de empresas de cada clase. Nos referiremos a esta muestra como ratio 1:1, una empresa saludable por cada empresa concursada. Posteriormente se incrementará este ratio para intentar acercarnos un poco más a la proporción real de empresas que fallan en la economía: por cada concursada se seleccionarán 1, 2, 4, 8 y 12 empresas saludables. Llamaremos ratio 2:1 al que selecciona dos buenas por cada mala, es decir dos terceras partes de buenas y una tercera parte de malas, y así sucesivamente. A medida que se incrementan dichos ratios se ajustará el punto de corte para la clasificación.

¹² Directorio de Empresas y Establecimientos, Entidades Jurídicas con actividad económica del Sector Privado, por tramos de personal ocupado, según división de actividad.

¹³ Promedio de las distribuciones sectoriales para los años bajo estudio, excluyendo los sectores considerados especiales (actividades de la administración pública (O), de los hogares (T) y los organismos extraterritoriales (U)) y el sector de actividades financieras y seguros (K), dado que no se incluyen en esta base de datos.

3.4 Medidas de desempeño

Con el objetivo de evitar que el modelo se sobreajuste a los datos, se decide realizar cada estimación 100 veces, con una muestra aleatoria distinta cada vez. Para cada conjunto de estimaciones se muestra la distribución de los coeficientes, así como el promedio de sus valores, errores estándar¹⁴ y porcentajes de significación de las variables¹⁵. Para poder evaluar la performance del modelo en las diferentes muestras se utilizan cuatro medidas de desempeño comúnmente utilizadas en los modelos de clasificación, y se presentan iguales medidas resumen que para los coeficientes:

- *Porcentaje Correctamente Clasificado o Precisión General (PCORR)*: se calcula como la cantidad de empresas clasificadas correctamente, independientemente de si son malas o buenas, sobre el total de empresas clasificadas.
- *Sensibilidad (SENS)*: es el porcentaje de empresas malas que han sido correctamente clasificadas, es decir las empresas concursadas que efectivamente lo son, sobre el total de concursadas. Los falsos negativos o errores de tipo I (empresas malas clasificadas como buenas), pueden calcularse como 1 menos el porcentaje de sensibilidad.
- *Especificidad (SPEC)*: es el porcentaje de empresas buenas que han sido correctamente clasificadas, es decir las empresas saludables que efectivamente lo son, sobre el total de saludables. Los falsos positivos o errores de tipo II (empresas buenas clasificadas como malas), pueden calcularse como 1 menos el porcentaje de especificidad.
- *Área bajo la curva ROC (AUROC)*: es una medida que resume la capacidad del modelo para distinguir entre los casos malos y buenos o positivos y negativos, tomando en cuenta el *trade-off* existente entre sensibilidad y especificidad. Para graficar la curva ROC se utilizan los verdaderos positivos o sensibilidad en el eje vertical, y los falsos positivos o 1 menos especificidad en el eje horizontal. El mejor modelo de predicción sería aquel que se encuentre en la coordenada (0,1), es decir un 100% de sensibilidad y especificidad, en tanto que una clasificación aleatoria daría como resultado una línea de 45°. Así, los resultados por encima de dicha curva se consideran buenos, y por el contrario malos los que están por debajo (peores que adivinar al azar el resultado). Por lo tanto, cuanto más alejada de la línea de 45° esté la curva ROC, mejor será el modelo, esto es, cuanto mayor sea el área debajo de la curva. Por lo general, un modelo con área mayor a 0,75 es considerado bueno, y mayor a 0,90 muy bueno.

¹⁴ El error estándar refiere a la dispersión entre los 100 coeficientes estimados.

¹⁵ Contraste de significación individual, estadístico t, nivel de significación $\alpha=5\%$.

4. Resultados

4.1 Estimación inicial

La primera muestra utilizada para estimar el modelo, como fue indicado, se compone de todas las empresas concursadas con el primer estado contable disponible previo a la solicitud de concurso, y las empresas saludables seleccionadas aleatoriamente de una base con similar distribución sectorial que la economía, con un ratio 1:1, una empresa saludable por cada concursada. En total se cuenta con 282 observaciones de empresas concursadas, que se mantienen en cada una de las 100 muestras, y 282 observaciones de empresas saludables que varían aleatoriamente.

En la Tabla 5 se pueden observar los valores promedio de los 100 coeficientes y medidas de desempeño estimadas, sus errores estándar y mínimos/máximos, así como la cantidad de veces que es significativa la variable de interés en las 100 estimaciones:

Tabla 5: Coeficientes y desempeños, estimación inicial

*Selección aleatoria en base de empresas saludables con similar distribución sectorial que la economía
Cutoff: 0,5, Ratio 1:1, Promedios 100 estimaciones*

	Media	EE	MIN	MAX	% Sig
ROA	-2,218	1,018	-3,774	0,041	85
LIQ	-0,363	0,051	-0,513	-0,235	100
FINL	2,898	0,968	0,852	5,557	90
PCORR	84,08	1,34	79,89	86,55	
SENS	83,88	1,48	79,39	87,40	
SPEC	84,27	1,64	80,00	87,82	
AUROC	0,917	0,009	0,896	0,935	

Notas: Ratio 1:1: una empresa saludable por cada concursada. Cut off: se clasifica como positivo (concurada) si la probabilidad predicha es mayor al valor del cut off. % Sig: porcentaje de coeficientes significativos en las 100 estimaciones. ROA: Resultado Operativo/Activo. LIQ: Activo corriente/Pasivo Corriente. FINL: Pasivo No Corriente/Activo. Variables de control: Año, Industria, Tamaño. PCORR: Porcentaje Correctamente Clasificado (Clasificadas Correctamente/Total). SENS: Sensibilidad (Casos positivos correctamente clasificados/Total positivos). SPEC: Especificidad (Casos negativos correctamente clasificados/Total negativos). AUROC: Área bajo la curva ROC (capacidad del modelo para distinguir entre positivos y negativos, relación entre SENS y 1-SPEC).

La distribución de los valores de los 100 coeficientes estimados con sus intervalos de confianza se muestra en el Gráfico 1, y la distribución de las medidas de desempeño en las 100 estimaciones en el Gráfico 2:

Gráfico 1: Coeficientes en 100 muestras aleatorias, estimación inicial

Selección aleatoria en base de empresas saludables con similar distribución sectorial que la economía

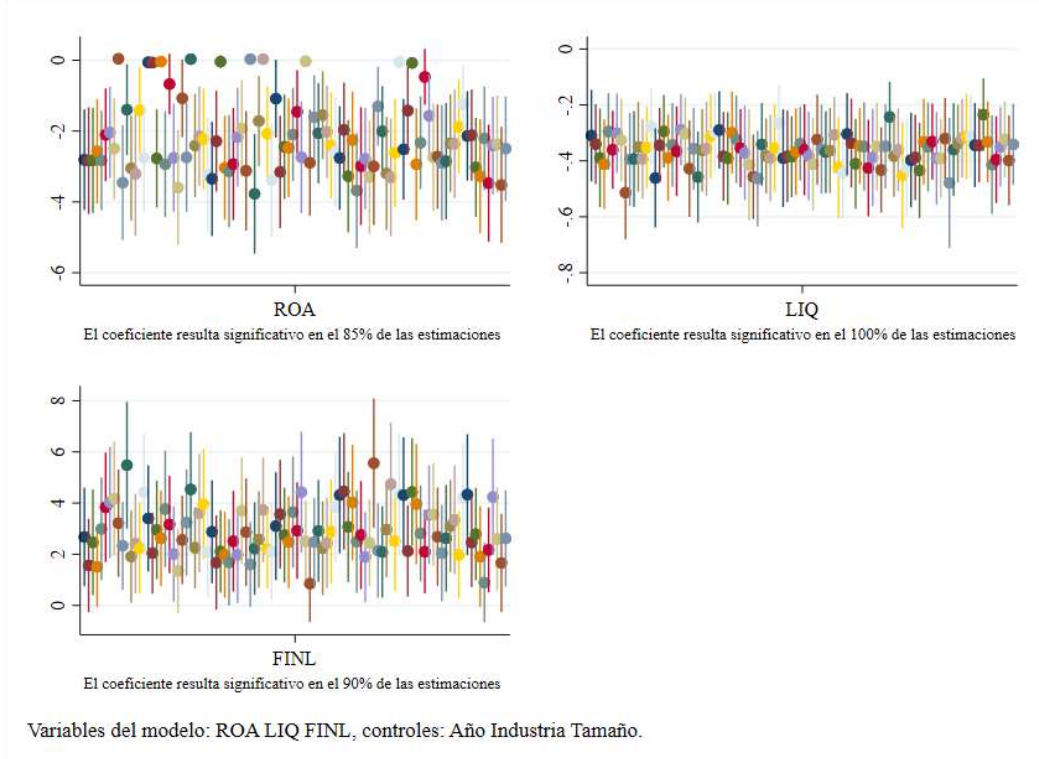
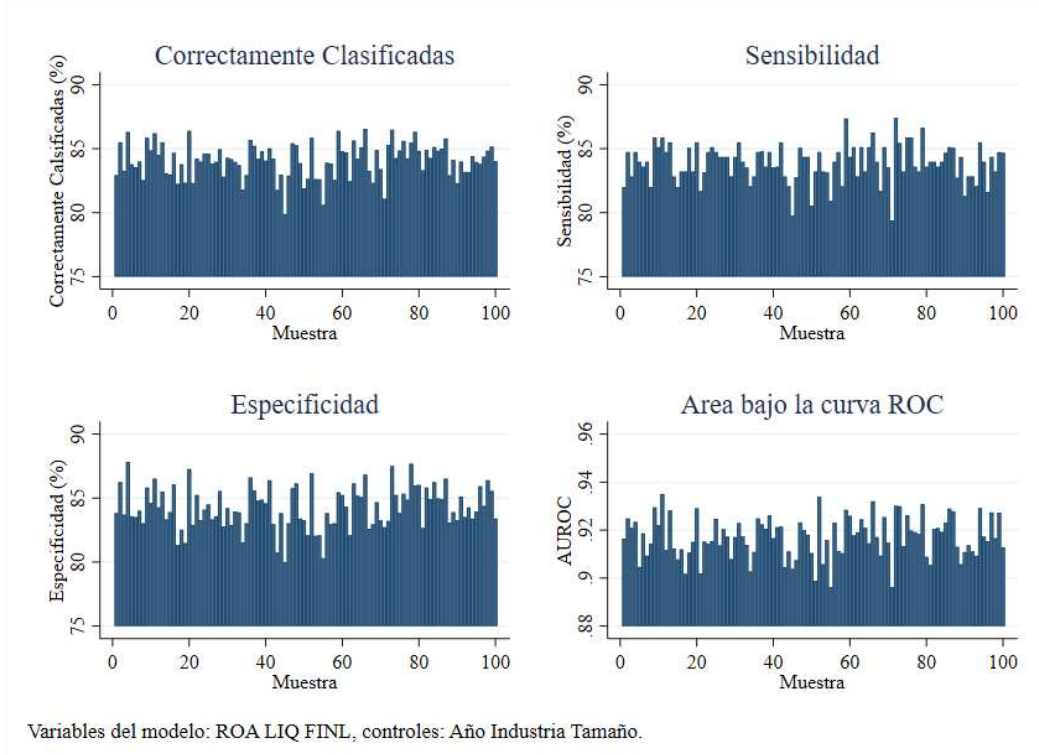


Gráfico 2: Desempeños en 100 muestras aleatorias, estimación inicial

Selección aleatoria en base de empresas saludables con similar distribución sectorial que la economía



Los ratios de rentabilidad y de liquidez, como es esperado, se relacionan negativamente con la probabilidad de entrar en concurso: a mayor rentabilidad y liquidez, menor dicha probabilidad. El ratio de endeudamiento, por su parte, se vincula de forma positiva: a mayor endeudamiento, mayor la probabilidad de solicitar concurso de acreedores. Las variables ROA y LIQ resultan significativas en 85% y 100% de las 100 estimaciones respectivamente, en tanto que FINL en el 90%.

En cuanto a las medidas de desempeño, el modelo, en promedio, es capaz de clasificar correctamente 84% de las firmas, y este porcentaje es válido tanto para las empresas concursadas (sensibilidad) como para las saludables (especificidad). Esto indica que un 16% de las firmas de cada grupo son clasificadas incorrectamente. El área bajo la curva ROC da señales de que el modelo es muy bueno para distinguir entre ambos grupos, con un valor mínimo de 0,896 y un máximo de 0,935.

Por último, los efectos marginales presentados en la Tabla 6 confirman la importancia del endeudamiento en la probabilidad de solicitar concurso de acreedores: ante un incremento de 1% en el ratio de endeudamiento, esto es, ante un incremento de 1% de la relación pasivo no corriente a activo, dicha probabilidad aumentaría en promedio en un 0,342%. Este resultado no resulta extraño al tener en cuenta que el endeudamiento constituye una pieza clave en la declaración del proceso concursal: se considera insolvente a aquella firma que no puede cumplir con sus obligaciones. Por su parte la liquidez, que resulta significativa en el 100% de las estimaciones y que, por referir a la capacidad de cumplir con las obligaciones de corto plazo, podría ser también una variable clave, aparece con el menor efecto marginal promedio de los tres ratios incluidos: los resultados indican una reducción de 0,043% en la probabilidad de solicitar concurso por cada 1% de incremento en el ratio. La rentabilidad tiene un efecto marginal promedio de -0,261%, confirmando su importancia en la probabilidad de solicitar concurso, aunque con la menor proporción de significación.

Tabla 6: Efectos marginales, estimación inicial

*Selección aleatoria en base de empresas saludables con similar distribución sectorial que la economía
Promedios 100 estimaciones*

	Media	EE	MIN	MAX
ROA	-0,261	0,119	-0,444	0,005
LIQ	-0,043	0,006	-0,057	-0,029
FINL	0,342	0,111	0,101	0,628

Notas: ROA: Resultado Operativo/Activo. LIQ: Activo corriente/Pasivo Corriente. FINL: Pasivo No Corriente/Activo.

Para cada una de las variables de control incluidas el modelo se realiza un test de Wald de significación conjunta ($\alpha=1\%$), resultando las tres variables significativas en el 100% de las estimaciones. En la Tabla A.1 (ver Anexo) se presentan las correlaciones (promedio, mínimo y máximo) entre las variables incluidas en el modelo para las 100 estimaciones.

4.2 Incremento del ratio de empresas saludables vs concursadas

Como fue mencionado, se procede en esta etapa a incrementar la cantidad de empresas saludables por cada empresa concursada, en orden de ir aproximándonos a la proporción real de empresas concursadas en el total. Para ello se continúan seleccionando de forma aleatoria a las empresas saludables, de la misma base que la estimación inicial, aumentando la cantidad seleccionada a 2, 4, 8 y 12 firmas por cada concursada.

Al mismo tiempo, y a fin de reflejar correctamente el objetivo del modelo, que es discriminar lo mejor posible a las empresas concursadas de las saludables, se ajusta el punto de corte de probabilidad para clasificar a una empresa como mala. La forma elegida para ajustarlo se basa en el método de Liu¹⁶, que busca maximizar la discriminación entre empresas malas y buenas mediante la maximización del producto de sensibilidad y especificidad. En cada una de las 100 estimaciones se calcula el punto de corte óptimo acorde a este método, para luego utilizar el valor medio de dichos puntos para evaluar el desempeño. A medida que se incrementa el ratio este punto de corte promedio tiende a disminuir, lo que implica que para que una empresa sea clasificada como mala tiene que superar un umbral cada vez más bajo.

El resultado de las estimaciones para cada uno de los ratios se presenta en la Tabla 7. Los signos de las variables de interés se mantienen a medida que se incrementa el ratio de empresas saludables y concursadas, al igual que la importancia y significación de la variable asociada a la liquidez, esto último con excepción del ratio 12:1, donde la significación se reduce del 100% al 77%. La variable de endeudamiento comienza a adquirir mayor significación, es decir, mayor cantidad de coeficientes significativos en las 100 estimaciones con respecto a la estimación inicial, llegando a ser significativa en todas las estimaciones, en tanto que la variable de rentabilidad pierde significación, para aumentar luego a 82% en el último ratio. En cuanto a las medidas de desempeño, se puede observar una pequeña reducción en la performance del modelo a medida que se incrementa el ratio de saludables vs concursadas, aunque continúa siendo muy bueno para discriminar los dos tipos de empresas. Los errores de clasificación de las empresas concursadas tienden a aumentar, pero nuevamente en una magnitud muy reducida. En resumen, el modelo mantiene su nivel de performance pese al incremento del ratio saludables vs concursadas.

¹⁶ Phil Clayton, 2013. "CUTPT: Stata module for empirical estimation of cutpoint for a diagnostic test," Statistical Software Components S457719, Boston College Department of Economics. <<https://ideas.repec.org/c/boc/bocode/s457719.html>>

Tabla 7: Coeficientes y desempeños, incremento del ratio saludables vs concursadas
Selección aleatoria en base de empresas saludables con similar distribución sectorial que la economía
Promedios 100 estimaciones

RATIO 2:1, N=846					
Cutoff: 0,3107781	Media	EE	MIN	MAX	% Sig
ROA	-1,830	1,085	-3,350	0,029	78
LIQ	-0,380	0,056	-0,585	-0,268	100
FINL	2,805	0,639	1,117	5,341	98
PCORR	82,97	1,028	79,63	85,21	
SENS	85,06	1,146	82,82	87,79	
SPEC	81,96	1,200	78,07	84,76	
AUROC	0,914	0,006	0,896	0,929	
RATIO 4:1, N=1.410					
Cutoff: 0,1965069	Media	EE	MIN	MAX	% Sig
ROA	-1,002	1,107	-3,357	0,022	47
LIQ	-0,453	0,072	-0,643	-0,318	100
FINL	2,514	0,566	1,197	3,935	100
PCORR	82,74	0,770	80,75	84,76	
SENS	83,64	0,939	81,30	85,88	
SPEC	82,52	0,876	80,34	85,05	
AUROC	0,910	0,004	0,898	0,918	
RATIO 8:1, N=2.538					
Cutoff: 0,1106012	Media	EE	MIN	MAX	% Sig
ROA	-0,631	0,990	-3,086	0,012	50
LIQ	-0,510	0,062	-0,606	-0,367	100
FINL	2,329	0,327	1,583	3,158	100
PCORR	82,58	0,576	81,13	83,84	
SENS	83,24	0,791	81,68	85,50	
SPEC	82,50	0,621	80,83	83,91	
AUROC	0,908	0,003	0,901	0,916	
RATIO 12:1, N=3.666					
Cutoff: 0,082311	Media	EE	MIN	MAX	% Sig
ROA	-1,097	0,957	-3,012	0,413	82
LIQ	-0,286	0,219	-0,574	-0,001	77
FINL	2,116	0,260	1,525	2,761	100
PCORR	82,26	1,109	78,98	84,16	
SENS	81,82	1,959	77,48	86,26	
SPEC	82,29	1,218	78,70	84,30	
AUROC	0,900	0,011	0,879	0,913	

Notas: Ratio "x":1: "x" empresas saludables por cada concursada. Cut off: se clasifica como positivo (concurada) si la probabilidad predicha es mayor al valor del cut off. N = número de observaciones aproximado (varía en cada estimación según las variables disponibles). % Sig: porcentaje de coeficientes significativos en las 100 estimaciones. ROA: Resultado Operativo/Activo. LIQ: Activo corriente/Pasivo Corriente. FINL: Pasivo No Corriente /Activo. Variables de control: Año, Industria, Tamaño. PCORR: Porcentaje Correctamente Clasificado (Clasificadas Correctamente/Total). SENS: Sensibilidad (Casos positivos correctamente clasificados: Concursadas Clasificadas Correctamente/Total Concursadas). SPEC: Especificidad (Casos negativos correctamente clasificados: Saludables Clasificadas Correctamente/Total Saludables). AUROC: Área bajo la curva ROC (capacidad del modelo para distinguir entre positivos y negativos, relación entre SENS y 1-SPEC).

Donde podemos observar un mayor impacto del incremento de este ratio es en los efectos marginales, presentados en la Tabla 8. En las tres variables se puede observar una reducción de su efecto parcial. En el ratio 1:1 vemos que un incremento de 1% en la rentabilidad conlleva una reducción promedio en la probabilidad de solicitar concurso de 0,261, en tanto que dicha magnitud pasa a 0,085 en el ratio 4:1, y a 0,053 en el ratio 12:1. La liquidez reduce su efecto parcial promedio más lentamente, pasando de -0,043 en el ratio 1:1 a -0,040 en el ratio 4:1, y finalmente -0,014 en el ratio 12:1. Por su parte el endeudamiento reduce su efecto marginal promedio de 0,342 en el ratio 1:1 a 0,219 en el 4:1, y finalmente 0,102 en el ratio 12:1, manteniéndose de todas formas como la variable con mayor influencia en la probabilidad de solicitar concurso en los diferentes ratios de empresas saludables y concursadas.

Tabla 8: Efectos marginales, incremento del ratio saludables vs concursadas
Selección aleatoria en base de empresas saludables con similar distribución sectorial que la economía
Promedios 100 estimaciones

RATIO 2:1, N=846				
Cutoff: 0,3107781	Media	EE	MIN	MAX
ROA	-0,197	0,116	-0,358	0,003
LIQ	-0,042	0,007	-0,061	-0,030
FINL	0,306	0,069	0,121	0,580
RATIO 4:1, N=1.410				
Cutoff: 0,1965069	Media	EE	MIN	MAX
ROA	-0,085	0,094	-0,277	0,002
LIQ	-0,040	0,007	-0,056	-0,026
FINL	0,219	0,048	0,102	0,352
RATIO 8:1, N=2.538				
Cutoff: 0,1106012	Media	EE	MIN	MAX
ROA	-0,037	0,057	-0,175	0,001
LIQ	-0,031	0,004	-0,036	-0,022
FINL	0,140	0,019	0,096	0,188
RATIO 12:1, N=3.666				
Cutoff: 0,082311	Media	EE	MIN	MAX
ROA	-0,053	0,047	-0,161	0,022
LIQ	-0,014	0,010	-0,027	0,000
FINL	0,102	0,012	0,075	0,130

Notas: Ratio "x":1: "x" empresas saludables por cada concursada. Cut off: se clasifica como positivo (concurada) si la probabilidad predicha es mayor al valor del cut off. ROA: Resultado Operativo/Activo. LIQ: Activo corriente/Pasivo Corriente. FINL: Pasivo No Corriente/Activo. Variables de control: Año, Industria, Tamaño.

Los resultados sugieren por lo tanto que independientemente de la proporción de empresas saludables y concursadas consideradas, el modelo es bueno para identificar y separar ambos tipos de empresas. Además, las tres variables incluidas resultan de importancia para realizar dicha clasificación, en particular destacan

los ratios de liquidez y endeudamiento. Sin embargo, es posible ver una reducción en el efecto de las variables sobre la probabilidad de solicitar concurso a medida que se incrementa la cantidad de empresas saludables en la muestra en relación con las concursadas.

5. Pruebas de Robustez

5.1 Cambios en los métodos de selección de empresas saludables

Podríamos preguntarnos si estos resultados se mantienen si cambia la forma en que fueron seleccionadas las empresas saludables. Por un lado, podrían seleccionarse también de forma aleatoria, pero sin considerar como base una muestra con similar distribución sectorial que la economía, sino simplemente seleccionarlas entre todas las firmas disponibles. Por otro lado, se podría considerar emparejar a las firmas por alguna característica de las empresas concursadas, como su sector de actividad, o realizar un *match* según los años en los que presentan los balances.

Por lo tanto, se vuelven a realizar 100 estimaciones para el ratio 1:1, seleccionando a las firmas saludables según estos métodos alternativos, y se presentan los resultados en la Tabla 9. Los valores promedio de los coeficientes cambian ligeramente, pero se mantiene su signo en todos los casos. El ratio de liquidez continúa siendo el más importante, en el sentido de que es significativo en el 100% de las estimaciones, mientras que los de rentabilidad y endeudamiento, si bien incrementan el porcentaje de significación, mantienen magnitudes similares a las de la estimación inicial. En cuanto a la performance del modelo, en los tres casos es levemente menor en las cuatro medidas de desempeño, en particular en las muestras con emparejamiento, sin embargo, estas diferencias no son elevadas.

Tabla 9: Coeficientes y desempeños, cambios en la selección de empresas saludables
Promedios 100 estimaciones

Selección aleatoria en toda la base					
Ratio 1:1, Cutoff: 0,5	Media	EE	MIN	MAX	% Sig
ROA	-2,384	0,847	-3,369	0,043	90
LIQ	-0,376	0,050	-0,531	-0,255	100
FINL	3,088	1,188	1,100	6,478	91
PCORR	83,61	1,312	80,34	86,25	
SENS	83,71	1,351	79,77	87,02	
SPEC	83,50	1,750	79,42	87,23	
AUROC	0,912	0,009	0,892	0,936	
Selección aleatoria por sector de actividad					
Ratio 1:1, Cutoff: 0,5	Media	EE	MIN	MAX	% Sig
ROA	-2,453	0,788	-3,665	0,034	93
LIQ	-0,385	0,054	-0,517	-0,242	100
FINL	3,168	1,060	0,544	6,553	95
PCORR	83,01	1,348	79,29	85,69	
SENS	82,82	1,436	79,01	86,64	
SPEC	83,19	1,882	77,53	87,82	
AUROC	0,907	0,009	0,879	0,924	
Selección aleatoria por año de balance					
Ratio 1:1, Cutoff: 0,5	Media	EE	MIN	MAX	% Sig
ROA	-2,681	0,903	-4,525	0,039	92
LIQ	-0,379	0,074	-0,616	-0,257	100
FINL	3,631	0,953	1,512	6,676	99
PCORR	81,25	1,238	77,63	83,86	
SENS	79,38	2,035	74,81	83,59	
SPEC	83,04	1,593	78,81	86,62	
AUROC	0,896	0,010	0,871	0,920	

Notas: Ratio 1:1: una empresa saludable por cada concursada. Cut off: se clasifica como positivo (concurada) si la probabilidad predicha es mayor al valor del cut off. % Sig: porcentaje de coeficientes significativos en las 100 estimaciones. ROA: Resultado Operativo/Activo. LIQ: Activo corriente/Pasivo Corriente. FINL: Pasivo No Corriente/Activo. Variables de control: Año, Industria, Tamaño. PCORR: Porcentaje Correctamente Clasificado (Clasificadas Correctamente/Total). SENS: Sensibilidad (Casos positivos correctamente clasificados: Concursadas Clasificadas Correctamente/Total Concursadas). SPEC: Especificidad (Casos negativos correctamente clasificados: Saludables Clasificadas Correctamente/Total Saludables). AUROC: Área bajo la curva ROC (capacidad del modelo para distinguir entre positivos y negativos, relación entre SENS y 1-SPEC).

Por su parte, los efectos marginales mantienen el mismo signo y similar magnitud, aunque en las tres estimaciones y para todas las variables, los efectos marginales son mayores que en la estimación inicial (Tabla 10).

Tabla 10: Efectos marginales, cambios en la selección de empresas saludables
Promedios 100 estimaciones

Selección aleatoria en toda la base	Media	EE	MIN	MAX
ROA	-0,293	0,102	-0,416	0,005
LIQ	-0,046	0,006	-0,067	-0,034
FINL	0,378	0,136	0,136	0,767
Selección aleatoria por sector de actividad	Media	EE	MIN	MAX
ROA	-0,309	0,098	-0,453	0,004
LIQ	-0,049	0,006	-0,063	-0,032
FINL	0,401	0,133	0,077	0,851
Selección aleatoria por año de balance	Media	EE	MIN	MAX
ROA	-0,358	0,117	-0,529	0,006
LIQ	-0,051	0,010	-0,081	-0,034
FINL	0,486	0,120	0,227	0,781

Notas: ROA: Resultado Operativo/Activo. LIQ: Activo corriente/Pasivo Corriente. FINL: Pasivo No Corriente/Activo.

De esta forma, podemos concluir que los resultados encontrados en la estimación inicial son robustos, no siendo sustancialmente modificados por el método a través del cual son seleccionadas las empresas saludables.

5.2 Predicción fuera de la muestra

Con el objetivo de evaluar la robustez del modelo, se analiza además su performance fuera de la muestra, es decir, en un subconjunto de los datos distinto al utilizado para estimarlo. Dado que no se dispone de más observaciones de empresas concursadas, la aproximación se realiza dividiendo a dicha muestra en dos: un 75% de las observaciones son utilizadas para estimar el modelo, mientras que el restante 25% se utiliza para testarlo. Este procedimiento, igual que en la estimación inicial, es realizado 100 veces en distintas muestras. Contamos con 282 empresas concursadas y se seleccionan 282 saludables de forma aleatoria de una muestra con similar distribución sectorial que la economía. Finalmente, del total de firmas (564) de la muestra, tres cuartas partes sirven para estimar el modelo, en tanto que la predicción se realiza para la cuarta parte restante, y se evalúa con las medidas de desempeño presentadas anteriormente.

En la Tabla 11 se presenta el promedio de las medidas de desempeño en las 100 estimaciones, sus errores estándar, mínimos y máximos. Los resultados sugieren que el modelo continúa siendo bueno para distinguir ambos tipos de firmas, aun cuando las observaciones que son clasificadas no pertenecen a la muestra utilizada para estimarlo. Sin embargo, se observa una reducción en la cantidad de empresas clasificadas correctamente, tanto en el total como en las concursadas (sensibilidad) y saludables (especificidad), a pesar del leve incremento en el promedio del valor bajo la curva ROC.

Tabla 11: Desempeños fuera de la muestra, estimación inicial
Selección aleatoria en base de empresas saludables con similar distribución sectorial que la economía
Promedios 100 estimaciones

	Media	EE	MIN	MAX
PCORR	81,32	3,39	71,85	90,30
SENS	81,36	4,57	68,57	91,94
SPEC	81,36	4,98	69,57	92,86
AUROC	0,919	0,012	0,888	0,949

Notas: PCORR: Porcentaje Correctamente Clasificado (Clasificadas Correctamente/Total). SENS: Sensibilidad (Casos positivos correctamente clasificados: Concursadas Clasificadas Correctamente/Total Concursadas). SPEC: Especificidad (Casos negativos correctamente clasificados: Saludables Clasificadas Correctamente/Total Saludables). AUROC: Área bajo la curva ROC (capacidad del modelo para distinguir entre positivos y negativos, relación entre SENS y 1-SPEC).

6. Ampliación del horizonte temporal y dinámica

6.1 Predicción dos períodos antes

El modelo inicial puede predecir correctamente la solicitud de concurso de una empresa utilizando la información contenida en el estado contable inmediatamente anterior a dicha solicitud, pero podríamos preguntarnos si es posible prever este hecho con mayor anticipación. Para comenzar a ampliar el horizonte temporal de la predicción, se estima el modelo utilizando la información proveniente del segundo estado contable previo a la solicitud, es decir, dos períodos antes de que ocurra.

La muestra se compone de todas las empresas concursadas con el segundo estado contable disponible previo a la solicitud de concurso, un total de 284 firmas, y de las empresas saludables, seleccionadas aleatoriamente de una base con similar distribución sectorial que la economía, totalizando 5.970 balances, para el período 2009-2019. En la Tablas 12 y 13 se pueden observar los resultados.

Tabla 12: Coeficientes y desempeños, dos períodos previo a la solicitud de concurso
Selección aleatoria en base de empresas saludables con similar distribución sectorial que la economía
Cutoff: 0,5, Ratio 1:1, Promedios 100 estimaciones

	Media	EE	MIN	MAX	% Sig
ROA	-0,781	0,488	-1,808	0,015	18
LIQ	-0,267	0,034	-0,374	-0,195	100
FINL	3,045	0,886	1,309	5,918	97
PCORR	81,26	1,40	77,97	85,39	
SENS	81,45	1,58	76,34	84,73	
SPEC	81,08	1,93	77,32	86,03	
AUROC	0,895	0,010	0,871	0,920	

Notas: Ratio 1:1: una empresa saludable por cada concursada. Cut off: se clasifica como positivo (concurada) si la probabilidad predicha es mayor al valor del cut off. % Sig: porcentaje de coeficientes significativos en las 100 estimaciones. ROA: Resultado Operativo/Activo. LIQ: Activo corriente/Pasivo Corriente. FINL: Pasivo No Corriente/Activo. Variables de control: Año, Industria, Tamaño. PCORR: Porcentaje Correctamente Clasificado (Clasificadas Correctamente/Total). SENS: Sensibilidad (Casos positivos correctamente clasificados/Total positivos). SPEC: Especificidad (Casos negativos correctamente clasificados/Total negativos). AUROC: Área bajo la curva ROC (capacidad del modelo para distinguir entre positivos y negativos, relación entre SENS y 1-SPEC).

Tabla 13: Efectos marginales, dos períodos previo a la solicitud de concurso
Selección aleatoria en base de empresas saludables con similar distribución sectorial que la economía
Promedios 100 estimaciones

	Media	EE	MIN	MAX
ROA	-0,102	0,064	-0,244	0,002
LIQ	-0,035	0,004	-0,045	-0,026
FINL	0,399	0,114	0,173	0,798

Notas: ROA: Resultado Operativo/Activo. LIQ: Activo corriente/Pasivo Corriente. FINL: Pasivo No Corriente/Activo.

Los ratios asociados a la liquidez y el endeudamiento mantienen su importancia para predecir la solicitud de concurso de acreedores incluso dos años antes de que ocurra, sin embargo, esto no es así para el ratio asociado a la rentabilidad, que si bien mantiene el signo esperado sólo es significativo en el 18% de las estimaciones. Pese a ello el modelo continúa siendo bueno para distinguir entre las firmas saludables y concursadas, pudiendo clasificar correctamente un poco más del 80% de las firmas. En cuanto a los efectos marginales, el endeudamiento confirma su importancia, afectando la probabilidad de solicitar concurso en 0,399% al incrementarse en un 1%, en tanto que la liquidez reduce dicha probabilidad en 0,035% al aumentar en un 1%.

Por último, se realiza el mismo procedimiento que en la estimación inicial para evaluar los desempeños fuera de la muestra, obteniéndose un porcentaje promedio de firmas clasificadas correctamente cercano al 79%. El modelo continúa siendo bueno con un valor promedio de área debajo de la curva ROC de 0,897 (ver Tabla A.2, Anexo).

En resumen, la estimación que utiliza la información de dos períodos previo a la solicitud no muestra un efecto de la rentabilidad sobre la probabilidad de solicitar concurso, y reduce la performance respecto al modelo inicial, tanto dentro como fuera de la muestra. Sin embargo, continúa siendo bueno para distinguir ambos tipos de firmas con un horizonte temporal mayor, constituyendo una aproximación inicial hacia la predicción en el mediano plazo.

6.2 Dinámica temporal

Como fue mencionado, estudios más recientes destacan la importancia de incluir dinámica temporal en las variables del modelo. Para intentar realizar una primera aproximación a este punto, se prueba estimar el modelo inicial sustituyendo las variables en niveles por los cambios en los ratios de un año a otro.

Para ello se considera una muestra que solo incluye a aquellas empresas que cuentan con al menos dos estados contables consecutivos¹⁷: el inmediatamente anterior a la solicitud de concurso, y el que lo precede. Esto genera una muestra de 275 empresas concursadas, para las que se calculan las nuevas variables como el valor de cada ratio en el último año, menos el valor correspondiente al año anterior. Para las empresas saludables, que tienen al menos dos balances consecutivos en el período, se elabora nuevamente una muestra con similar distribución sectorial que la economía, que se compone de 1.010 empresas con 2.944 observaciones, a partir de la cual se seleccionan las firmas aleatoriamente.

¹⁷ Se excluyen las empresas que cambian la moneda de presentación de sus estados contables en años consecutivos.

En la Tabla A.3 (ver Anexo) se presentan los resultados promedio de las 100 estimaciones. Como puede observarse, sólo el cambio en la rentabilidad parece tener importancia en la predicción de la solicitud de concurso. Esto podría indicar que la rentabilidad es el ratio que se deteriora más tarde o cerca del final del proceso concursal, o que la caída de la rentabilidad es uno de los desencadenantes finales de dicho proceso. Un año antes de la solicitud de concurso, los resultados muestran que los tres ratios son importantes en la predicción, en tanto que dos años antes la rentabilidad no es significativa en la mayoría de las estimaciones. Sin embargo, los resultados que incluyen la dinámica temporal indican que el cambio en la rentabilidad sí es significativo, sugiriendo que este ratio es el que sufre un mayor deterioro en el último año. Es decir, con malos ratios de liquidez y endeudamiento, tanto en el primer como el segundo año previo a la solicitud, la caída de la rentabilidad en el último año, podría ser el indicador final que confirma las dificultades de la empresa que posteriormente llevan a la solicitud de concurso.

Sobre este punto será necesario investigar en el futuro otras alternativas, para dar cuenta del carácter temporal del proceso de solicitud de concurso de acreedores, en conjunto con la predicción en mediano plazo.

7. Conclusiones

La predicción de dificultades financieras o quiebra de las empresas ha sido un tema de amplio estudio en las últimas décadas. Muchas de estas dificultades pueden determinar el cierre o liquidación de la empresa, lo que en mayor o menor medida, implica un costo para cada uno de los involucrados e incluso puede tener un impacto negativo en la sociedad y la economía. Los ratios financieros han demostrado ser buenos síntomas de la salud de la empresa, apareciendo como una importante herramienta a la hora de predecir este fenómeno.

Esta investigación elabora un modelo predictivo para la caída en concurso de acreedores de empresas uruguayas, utilizando principalmente la información proveniente de los estados financieros, en forma de ratios, y combinada con variables referidas a las características observables de la empresa, como su sector de actividad y tamaño. Los ratios utilizados refieren a la rentabilidad de la empresa, su nivel de liquidez y su estructura de financiamiento. Todos resultan importantes a la hora de predecir hasta un año antes la solicitud de concurso de acreedores: empresas con menor rentabilidad, menor liquidez y mayor endeudamiento tienen mayor probabilidad de participar en un concurso de acreedores.

El modelo es capaz de clasificar correctamente en promedio un 84% de las firmas, y este porcentaje es válido tanto para las empresas concursadas (sensibilidad) como para las saludables (especificidad). Esto indica que un 16% de las firmas de cada grupo son clasificadas incorrectamente. El área bajo la curva ROC da señales de que el modelo es muy bueno para distinguir entre ambos tipos de firmas. En líneas generales, estos resultados se mantienen cuando las observaciones que son clasificadas no pertenecen a la muestra utilizada para estimarlo.

Por otra parte, los resultados sugieren que independientemente de la proporción de empresas saludables y concursadas consideradas en el análisis, el modelo es bueno para identificar y separar ambos tipos de empresas, y que el mismo es robusto al método a través del cual son seleccionadas las empresas saludables. Al incrementar el horizonte temporal a dos períodos previos a la solicitud, la rentabilidad pierde efecto sobre la probabilidad de solicitar concurso, y se reduce la performance general del modelo, aunque continúa siendo bueno para distinguir ambos tipos de firmas. Por último, la inclusión de la dinámica temporal no brinda los mismos resultados, siendo una importante línea de investigación a futuro.

Este estudio constituye un pequeño paso en la exploración de los determinantes del fracaso de las firmas y su identificación temprana en Uruguay, con el fin de anticipar las dificultades que pueden conducir al concurso de acreedores, evitando las pérdidas asociadas al proceso. Es importante señalar, para el uso futuro del modelo, que dada la crisis actual debida al COVID-19 y las medidas tomadas por el gobierno para enfrentar la misma, este modelo puede no ser acertado, siendo necesaria su evaluación en dicho contexto.

8. Bibliografia

- Altman, E. I., 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), pp. 589-609.
- Altman, E. I., Haldeman, R. G. & Narayanan, P., 1977. ZETA analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, 1(1), pp. 29-54.
- Aziz, M. A. & Dar, H. A., 2006. Predicting corporate bankruptcy: where we stand?. *Corporate Governance*, 6(1), pp. 18-33.
- Bae, J. K., 2012. Predicting financial distress of the South Korean manufacturing industries. *Expert Systems with Applications*, 39(10), pp. 9159-9165.
- Balcaen, S. & Ooghe, H., 2004. Alternative methodologies in studies on business failure: do they produce better results than the classical statistical methods?. *Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium*, p. Paper Number 04/249.
- Balcaen, S. & Ooghe, H., 2006. 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 38(1), pp. 63-93.
- Bardos, M., 2007. What is at stake in the construction and use of credit scores?. *Computational Economics*, Volumen 29, pp. 159-172.
- Beaver, W., 1966. Financial Ratios As Predictors of Failure. *Journal of Accounting Research*, Volumen 4, pp. 71-711.
- Becchetti, L. & Sierra, J., 2003. Bankruptcy risk and productive efficiency in manufacturing firms. *Journal of Banking & Finance*, 27(11), pp. 2099-2120.
- Charitou, A., Neophytou, E. & Charalambous, C., 2004. Predicting corporate failure: empirical evidence for the UK. *European Accounting Review*, 13(3), pp. 465-497.
- du Jardin, P., 2014. Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research*, 242(1), pp. 286-303.
- du Jardin, P., 2017. Dynamics of firm financial evolution and bankruptcy prediction. *Expert Systems with Applications*, Volumen 75, pp. 25-43.
- Edminster, R. O., 1972. An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction. *Journal of Financial and Quantitative Analysis*, 7(2), pp. 1477-1493.

- Grice, J. S. & Dugan, M. T., 2001. The Limitations of Bankruptcy Prediction Models: Some Cautions for the Researcher. *Review of Quantitative Finance and Accounting*, Volumen 17, pp. 151-166.
- Ho, C. Y., McCarthy, P., Yang, Y. & Ye, X., 2013. Bankruptcy in the pulp and paper industry: market's reaction and prediction. *Empirical Economics*, Volumen 45, pp. 1205-1232.
- Keasey, K. & McGuinness, P., 1990. The Failure of UK industrial firms for the period 1976-1984: logistic analysis and entropy measures. *Journal of Business Finance & Accounting*, Volumen 13, pp. 119-135.
- Kirkos, E., 2015. Assessing methodologies for intelligent bankruptcy prediction. *Artificial Intelligence Review*, Volumen 43, pp. 83-123.
- Ohlson, J. A., 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), pp. 109-131.
- Ooghe, H., Joos, P. & De Bourdeaudhuij, C., 1995. Financial distress models in Belgium: The results of a decade of empirical research. *The International Journal of Accounting*, 30(3), pp. 245-274.
- Oz, I. O. & Sigma-Mugan, C., 2018. Bankruptcy prediction models' generalizability: Evidence from emerging market economies. *Advances in Accounting*, Volumen 41, pp. 114-125.
- Oz, I. O. & Yelkenci, T., 2017. A theoretical approach to financial distress prediction modeling. *Managerial Finance*, 43(2), pp. 212-213.
- Pascale, R., 1988. A multivariate model to predict firm financial problems: the case of Uruguay. *Studies in Banking and Finance*, Volumen 7, pp. 171-182.
- Sandin, A. & 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), pp. 295-311.
- Shumway, T., 2001. Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business*, 74(1), pp. 101-124.
- Taffler, R. J., 1983. The Assessment of Company Solvency and Performance Using a Statistical Model. *Accounting and Business Research*, 13(52), pp. 295-307.
- Tian, S., Yu, Y. & Zhou, M., 2015. Data Sample Selection Issues for Bankruptcy Prediction. *Risk, Hazards & Crisis in Public Policy*, 6(1), pp. 91-116.
- Veganzones, D. & Séverin, E., 2018. An investigation of bankruptcy prediction in imbalanced datasets. *Decision Support Systems*, Volumen 112, pp. 111-124.

Volkov, A., Benoit, D. F. & Van den Poel, D., 2017. Incorporating sequential information in bankruptcy prediction with predictors based on Markov for discrimination. *Decision Support Systems*, Volumen 98, pp. 59-68.

Zavgren, C. V., 1985. Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Analysis. *Journal of Banking & Finance*, 12(1), pp. 19-45.

Zhou, L., 2013. Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Systems*, Volumen 41, pp. 16-25.

Zhou, L., Lai, K. K. & Yen, J., 2012. Empirical models based on features ranking techniques for corporate financial distress prediction. *Computers & Mathematics with Applications*, 64(8), pp. 2484-2496.

Zmijewski, M. E., 1984. Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, Volumen 22, pp. 59-82.

9. Anexo

Tabla A.1: Correlaciones entre variables, estimación inicial

*Selección aleatoria en base de empresas saludables con similar distribución sectorial que la economía
Promedios 100 estimaciones*

	Media	EE	MIN	MAX
ROA_LIQ	0,010	0,008	0,002	0,033
LIQ_FINL	-0,032	0,022	-0,087	0,033
FINL_ROA	0,020	0,002	0,016	0,030

Notas: ROA: Resultado Operativo/Activo. LIQ: Activo corriente/Pasivo Corriente. FINL: Pasivo No Corriente /Activo.

Tabla A.2: Desempeños fuera de la muestra, dos períodos previo a la solicitud de concurso

*Selección aleatoria en base de empresas saludables con similar distribución sectorial que la economía
Promedios 100 estimaciones*

	Media	EE	MIN	MAX
PCORR	78,91	3,39	70,16	89,31
SENS	79,23	5,18	65,15	90,48
SPEC	78,73	5,11	66,67	91,53
AUROC	0,897	0,014	0,860	0,932

Notas: PCORR: Porcentaje Correctamente Clasificado (Clasificadas Correctamente/Total). SENS: Sensibilidad (Casos positivos correctamente clasificados: Concursadas Clasificadas Correctamente/Total Concursadas). SPEC: Especificidad (Casos negativos correctamente clasificados: Saludables Clasificadas Correctamente/Total Saludables). AUROC: Área bajo la curva ROC (capacidad del modelo para distinguir entre positivos y negativos, relación entre SENS y 1-SPEC).

Tabla A.3: Variables como cambio de un año al otro

*Selección aleatoria en base de empresas saludables con similar distribución sectorial que la economía
Promedios 100 estimaciones*

Cambio respecto al año anterior					
Ratio 1:1, Cutoff: 0,5	Media	EE	MIN	MAX	% Sig
chROA1	-1,613	0,454	-2,606	-0,368	88
chLIQ1	-0,007	0,025	-0,125	0,043	3
chFINL1	-0,369	0,251	-1,279	0,140	0
PCORR	80,89	1,388	78,03	84,54	
SENS	79,78	1,607	75,00	84,77	
SPEC	81,95	2,001	77,90	86,89	
AUROC	0,891	0,010	0,868	0,921	

Notas: Ratio 1:1: una empresa saludable por cada concursada. Cut off: se clasifica como positivo (concurada) si la probabilidad predicha es mayor al valor del cut off. % Sig: porcentaje de coeficientes significativos en las 100 estimaciones.

chROA1: ROA(t) - ROA(t-1). chLIQ1: LIQ(t) - LIQ(t-1). chFINL1: FINL(t) - FINL(t-1). Variables de control: Año, Industria, Tamaño. PCORR: Porcentaje Correctamente Clasificado (Clasificadas Correctamente/Total). SENS: Sensibilidad (Casos positivos correctamente clasificados: Concursadas Clasificadas Correctamente/Total Concursadas). SPEC: Especificidad (Casos negativos correctamente clasificados: Saludables Clasificadas Correctamente/Total Saludables). AUROC: Área bajo la curva ROC (capacidad del modelo para distinguir entre positivos y negativos, relación entre SENS y 1-SPEC).