Received 13 August 2013
Revised 23 September 2013
11 November 2013
Accepted 12 November 2013

Models for predicting default: towards efficient forecasts

Fernando Castagnolo
Citigroup, London, UK, and
Gustavo Ferro
*Instituto de Economía UADE,
Universidad Argentina de la Empresa and CONICET,
Buenos Aires, Argentina*

Abstract

Purpose – The purpose of this paper is to assess and compare the forecast ability of existing credit risk models, answering three questions: Can these methods adequately predict default events? Are there dominant methods? Is it safer to rely on a mix of methodologies?

Design/methodology/approach – The authors examine four existing models: O-score, Z-score, Campbell, and Merton distance to default model (MDDM). The authors compare their ability to forecast defaults using three techniques: intra-cohort analysis, power curves and discrete hazard rate models.

Findings – The authors conclude that better predictions demand a mix of models containing accounting and market information. The authors found evidence of the O-score's outperformance relative to the other models. The MDDM alone in the sample is not a sufficient default predictor. But discrete hazard rate models suggest that combining both should enhance default prediction models.

Research limitations/implications – The analysed methods alone cannot adequately predict defaults. The authors found no dominant methods. Instead, it would be advisable to rely on a mix of methodologies, which use complementary information.

Practical implications – Better forecasts demand a mix of models containing both accounting and market information.

Originality/value – The findings suggest that more precise default prediction models can be built by combining information from different sources in reduced-form models and combining default prediction models that can analyze said information.

Keywords Financial crisis, Efficiency, Credit risk, Empirical analysis, Predicting default models

Paper type Research paper

1. Introduction

Keeping in mind the recent financial crisis, we address the usefulness of existing default prediction models and the relevance of the information they consider. Defaults are extremely rare events: their prediction models aim to identify defaulting firms and alert agents not to finance them.

During 2007, contagion of the US mortgage market crisis generated a global event with losses spilled over the financial and banking markets worldwide. Financial innovation such us mortgage-backed securities (MBS) and collateralized debt obligations (CDO), enabled institutions and investors around the world to invest in the American housing market. As American real estate prices went down, institutions heavily exposed to those instruments experienced important losses, and begun to unwind their positions, generating further price declines and losses given the interconnection of the financial institutions. The FED facilitated access to the discount window and offered a special credit

The Journal of Risk Finance
Vol. 15 No. 1, 2014
pp. 52-70

1526-5943
DOI 10.1108/JRF-08-2013-0057

line (term auction facility). In the UK deposit insurance was extended to the deposits of the failed Northern Rock institution. Problems in the USA aggravated with the rescue of Bear Stearns, the bankruptcy of Lehman Brothers – which was a turning point – the bail out of the private insurance company AIG, and the rescue of the mortgage originators Freddie Mac and Fannie Mae (government sponsored enterprises). As a consequence of the crisis, GDP and foreign trade fell and capital flows reversed their direction. Several countries announced bank recapitalization and other support facilities during 2008 and 2009 (Laeven and Valencia, 2010; Claessens *et al.*, 2010). What followed Lehman bankruptcy, was a \$700 billion bailout package (Troubled Asset Relief Program) prepared by the Secretary of the Treasury, and approved by Congress.

Bank diversification is associated to disintermediation and migration from the paradigm "originate to hold" to the "originate to sale" business. The logic of transforming loans into securities, and to trade them out of the balance sheet, has incorporated opacity and the need of extending business to substandard clients. The disintermediation incorporated fragility by lowering the quality of the assets (Claessens, 2002).

During 2010, the US Congress enacted Dodd-Frank Wall Street Reform and Consumer Protection Act. The reform tries to mitigate systemic risk, empowering a Financial Stability Oversight Council (FSOC) to act as the systemic risk regulator. The FSOC is integrated by the Treasury, the FED and federal regulators of different financial business (Bhatia, 2011). As a result of banks' massive incurred losses, the Basel Committee on Banking Supervision introduced Basel III banking regulatory framework, increasing capital requirements, putting caps on leverage, establishing limits to counter-party risk, and setting new liquidity requirements. Basel III expanded the scope of the incremental risk capital charge from default risk to both default risk and credit rating transition risk. This credit rating transition risk comprises the probability of losses resulting from an internal or external credit rating downgrade or upgrade (van der Ploeg, 2010). Basel III seeks to increase bank liquidity and decrease bank leverage. Sabato (2010) finds that with the precedent Basel Capital Accord, credit scoring models have been given unprecedented significance. Most financial institutions worldwide have either developed or modified existing internal credit risk models.

According to Hassan *et al.* (2004), there are particular corporate governance problems in the banking sector, because its stock holders are not the solely beneficiaries of the productivity of the institutions. Bad corporate governance in banks has the potential of externalities to other parts of the economy, given the interconnections between finance and the real sector. In the same line, Cocris and Ungureanu (2007) suggest that leverage, moral hazard and opacity of the bank assets collaborate to make the banks different. Nam (2006) emphasizes leverage and regulation, but suggests that the intertemporal nature and complexity of the transactions gives rise to opacity of banks assets. Good corporate governance practices allow banks to effectively monitor the quality of their assets. Arun and Turner (2004) highlight the opposite vision of regulators (looking at the safety of the whole banking system, and then trying to lower risks) and that of the bank managers, motivated by packages of incentives to take risk and also being able to hide risks in the books, in part thanks to the intertemporal character of the operations. Good corporate governance practices can lower the risk of the whole system, and predicting models are a key feature of the job to be done.

The sub-prime and the Euro zone crises have shown the dangerous consequences of not being able to correctly measure and anticipate credit risk. Defaults were severely underestimated by the major rating agencies (Blöchlinger, 2013). Our objective is to assess and compare the forecast ability of existing credit risk models and to propose possible alternatives in answer to three questions:

(1) Can those methods adequately predict defaults?

(2) Are there dominant methods?

(3) Given the results, is it safer to rely on a mix of methods?

We examine four existing models: O-score, Z-score, Campbell, and Merton distance to default model (MDDM), as well as test a Naive alternative of Merton model (NAMM). They differ in the information they provide – accounting or market based.

We apply three techniques to compare the models' ability to forecast defaults: intra-cohort analysis, power curves and discrete hazard rate models. We conclude that better predictions demand a mix of models containing both accounting and market information.

This paper continues the search for a superior default predictive model, being the first one, to the knowledge of the authors, to apply such a broad set of tools to compare different models on one dataset. Previous literature has focused either on one models' statistical power or its *ex post* predictive ability by the results shown by one of the measuring tools. This paper combines a broad set of measuring tools and concludes favourably for a model that combines accounting, market data, and the structural form of the MDDM, as all of them add default predictive power to each other. Future research would be required to propose a superior structural form that combines them. After this introduction, Section 2 reviews the literature on default predictive models. Section 3 presents the data and the methodology, Section 4 the results, and Section 5 concludes.

2. Literature review

This section examines the construction of several credit risk models that are available in the literature and their tests of predictive accuracy.

2.1 Credit risk modeling

Default rate intensities are a necessary input to credit derivative pricing models. These inputs can be estimated implicitly using debt prices or explicitly through actual bankruptcies, balance sheet and market data (Chava and Jarrow, 2004).

2.1.1 Z-score and O-score. Altman' (1968) Z-score is a multivariate discriminant analysis (MDA) based on five accounting ratios representing proxies of liquidity, profitability, leverage, solvency, and the activity level that is generally associated with a firm's probable bankruptcy[1]:

$$Z = 0.12(WC/TA) + 0.14(RE/TA) + 0.33(EBIT/TA) + 0.006(MVE/BVD) + 0.999(S/TA) \quad (1)$$

where:

WC/TA	working capital to total assets.
RE/TA	retained earnings to total assets.

EBIT/TA earnings before interest and tax to total assets.

MVE/BVD market value of equity to book value of debt.

S/TA sales to total assets.

MDA assumes that defaulted and non-defaulted firm predictors are normally distributed, with an equal variance-covariance matrix. Ohlson (1980) points out that it also imposes matching procedures that tend to be arbitrary and provides an output that has little intuitive interpretation.

Thus, Ohlson (1980) proposed a logit model of nine explanatory variables (O-score):

$$O = b1(Size/GNP) + b2(TL/TA) + b3(WC/TA) + b4(CL/CA) + b5\ OENEG + b6(NI/TA) + b7(FO/TL) + b8\ INTWO + b9CHIN \tag{2}$$

where:

bi are the corresponding logit coefficients of the explanatory variables.

Size/GNP log of total assets to gross national product.

TL/TA total liabilities to total assets.

WC/TA working capital to total assets.

CL/CA current liabilities to current assets.

NI/TA net income to total assets.

FO/TL funds provided by operations to total liabilities.

OENEG indicator that is 1 if (total liabilities $>$ total assets), or 0 otherwise.

INTWO indicator that is 1 if (net income $<$ 0) for the last two years, or 0 otherwise.

CHIN change in net income.

2.1.2 Other reduced forms. Campbell *et al.* (2008) combine both recent and lagged accounting and market information in a logit model of eight variables to predict default:

$$C = c1\ (TL/MTA) + c2\ OSIGMA + c3\ ORSIZE + c4\ CASHMTA + c5\ (MVE/BE) + c6\ Price + c7\ (NI/MTA - Avg) + c8\ (EXRET - Avg) \tag{3}$$

where:

ci are the corresponding logit coefficients of the explanatory variables.

TL/MTA total liabilities to market value of assets.

OSIGMA daily stock return standard deviation of the last three months.

ORSIZE log ratio of market capitalization to S&P-500.

CASHMTA current assets to market value of assets.

MVE/BE market to book ratio.

Price	log price per share.
NI/MTA-Avg	net income to average market value of assets.
EXRET-Avg	average log excess return relative to S&P-500.

2.1.3 Merton distance to default model. The MDDM (Merton, 1974) is based on market and accounting information and makes it possible to determine a firm's probabilities of default at any point in time but imposes two restrictive assumptions. First, default can only occur at maturity, as if the firm had issued a zero coupon security that matures at time T. Second, the total value of a firm follows a geometric (i.e. exponential) Brownian motion.

This model exploits the fact that equity and a call option on the underlying asset have identical payoffs, and that equity is a residual claim on the assets after all other obligations have been met. As such, shareholders will exercise their option and pay off all debts at time T if the value of the firm's assets is greater than the face value of its liabilities. Otherwise, the firm files for bankruptcy and ownership is assumed to be transferred without cost to the creditors, while the payoff to shareholders is zero (Hillegeist *et al.*, 2004).

Default in the Merton setting occurs when the ratio of the value of assets to debt is less than one. The distance to default at time t (DD_t) indicates how many standard deviations the log of this ratio should deviate from its mean for default to occur (Vassalou and Xing, 2004). DD_t can be defined as:

$$DD_t = [ln(V_{A,t}/X_t) + (u - 0.5\,s^2 A)T]/s_A t^{-0.5} \qquad (4)$$

where:

V_A	is the value of the firm.
u	is the expected continuously compounded return on $V_{A.}$
s_A	is the standard deviation of V_A.
X	is the debt face value.
T	its maturity time.

2.1.4 Naive alternative to the Merton model. Bharath and Shumway (2008) proposed a NAMM. This retains the structural form of the original model but estimates its inputs more simply. It approximates the debt face value with the sum of current liabilities plus half long-term liabilities and estimates the volatility of assets by:

$$Naïve\ \ s_A = [MVE/(MVE + BVD)]s_E + [BVD/(MVE + BVD)]Naïve\ \ s_D \qquad (5)$$

where:

MVE	is the market value of equity.
BVD	is the book value of debt.
s_E	is the stock volatility.
Naïve s_D	g + h s_E.

Factor g is included to represent the term structure of volatility, while h allows for volatility associated with default risk (Bharath and Shumway, 2008)[2]. Finally, this model defines:

$$Naïve_u = r_{it-1}$$

This simplified specification easily calculates and retains the structure of the MDDM. Therefore, evidence of the limitations of the MDDM can be found if its predictive power is similar to that of the NAMM.

2.2 Evidence of default prediction performance

The first studies on default prediction are due to Beaver (1966) and Altman (1968). Merton (1974), with his MDDM introduced a model based on both market and accounting information. Until the end of the 1970s, discriminant analysis introduced by the former authors remained the dominant method in the prediction of failure (van der Ploeg, 2010). Martin (1977) introduced the logit prediction model, without restrictive assumptions on the distributional properties of the variables. Ohlson (1980) proposed a logit model of nine explanatory variables. Lim (1999) started a series of studies comparing the performance of the prediction models. Kealhofer and Kurbat (2001), Kealhofer (2003), Hillegeist *et al.* (2004), Stein (2005), Bharath and Shumway (2008) and Campbell *et al.* (2008) continue that line of work, trying to conclude the superiority of one model with respect to the others, and proposing alternatives. Bharath and Shumway (2008) conclude that the MDDM model does not appear to produce a sufficient statistics for default. Campbell *et al.* (2008) compare and contrast some models in accessing default risks. van der Ploeg (2010) goes beyond the comparisons of model performance and try to predict credit rating transitions. Hayden (2003) evaluates rating models for three different default definitions with a unique data set on credit risk analysis for the Austrian market. Kocenda and Vojtek (2009) estimate models which are compared in terms of efficiency and power to discriminate between low and high risk clients, by employing data from a new European Union economy (Czech Republic). John *et al.* (2007) evaluate the efficiency of Altman's Z-score model for credit risk evaluation through empirical data for Indian banks, and propose a new revised model.

The evidence of different default prediction models' performance is inconclusive and findings can be divided into two groups. Hillegeist *et al.* (2004), Kealhofer and Kurbat (2001), Kealhofer (2003) and Lim (1999) indicate that the MDDM outperforms and contains more default-related information than any other accounting-based and rating-based model. Conversely, Campbell *et al.* (2008), Stein (2005) and Bharath and Shumway (2008) highlight the MDDM's incompleteness and find reduced forms that outperform it.

Among the first group, Hillegeist *et al.* (2004) proposes a discrete hazard model that incorporates the economy-wide percentage rate of defaults by listed companies over the past year as the baseline hazard rate. They conclude that O-score outperforms Z-score, but that MDDM outperforms both. Kealhofer and Kurbat (2002) conclude that all the default predictive information of Moody's ratings and accounting variables were already present in a KMV setting[3] of the MDDM. They also find that the latter had fewer incorrect identifications of default than the alternatives. Kealhofer (2003) and Lim (1999) present analogous results.

On the other hand, Campbell *et al.* (2008) offer evidence of the MDDM's incompleteness. They propose a multivariate logit model that incorporates a wide range of explanatory variables to predict default. Stein (2005) presents additional evidence of the MDDM's failure.

Bharath and Shumway (2008) test the structural form of the MDDM and the algorithms used to derive its inputs and provide three main findings: first, the MDDM does not suffice to predict default since it can be improved by adding explanatory variables. Second, the algorithm specified by the MDDM to calculate total firm value and volatility does not add default prediction power. Third, the MDDM's usefulness is based solely on its structural form. They conclude that the MDDM probability "is a marginally useful default forecaster, but it is not a sufficient statistic for default" (Bharath and Shumway, 2008).

Evidence about the completeness of the each model is contradictory. Even when comparing credit risk models with the same methodologies, the literature does not agree on their suitability, nor in the existence of a superior one.

3. Data and methodology

We use accounting, market and macroeconomic information to compare the models, analyzing all non-financial firms in the intersection of the CRSP and compustat databases (accounting and market data) between January 1990 and December 2010. The sample of quarterly observations contains 328 actual defaults, bankruptcies or liquidations from 10,439 firms.

The data were divided into two sub-samples according to the bankruptcy indicator (i.e. defaulted and non-defaulted observations). Table I presents summary statistics of input variables of credit risk and variables notation. Defaulted observations present a negative mean WC/TA, indicating that their current assets cannot, on average, cover their liabilities. They also show a less negative RE/TA and a lower NI/TA than non-defaulted observations, a lower MVE and a higher ratio of MVE/BVD. Finally, they exhibit a much higher rate of CL/CA, suggesting that defaults are likely to be generated by liquidity problems and lower stock return (and higher stock volatility) rather than non-defaulted firms.

The implementation of the models faces us with several difficulties. For example:

- MDDM relies on non-observable inputs;
- there are missing data in the databases; and
- statistical O-score and Z-score present limitations.

The s_E and V_E can be derived from market price information and the amount of shares outstanding. Instead, the s_A, $V_{A,}$ and u are not directly observable and must be estimated to empirically test the Merton model[4].

Data to construct the inputs of the models were unavailable for all companies, making it necessary to make some adjustments:

- The ratio of sales to total assets (over 50 percent of missing observations) was replaced by NI/TA.
- Funds provided by operation (FO) showed a similar pattern of missing observations so it was replaced with net income to total liabilities (NI/TL). The quarters with missing stock price, total value of assets, or total value of liabilities were eliminated.

Table I. Descriptive statistics

Variable	Notation	Non-defaulted sub-sample Mean	Median	SD	Defaulted sub-sample Mean	Median	SD
WC/TA	Working capital to total assets	0.15	0.23	15.47	−0.20	0.00	1.22
RE/TA	Retained earnings to total assets	−2.29	0.05	315.37	−3.09	−0.77	9.89
MVE	Market value of equity (million)	10.615	1.599	75.589	5.813	2.991	9.740
MVE/BVD	Market value of equity to book value of debt	30.90	0.37	347.33	168	1.03	1,825
NI/TA	Net income to total assets	−0.06	0.01	9.91	−0.18	−0.09	0.46
TL/TA	Total liabilities to total assets	0.63	0.47	16.75	1.28	0.95	2.04
CL/CA	Current liabilities to current assets	1.20	0.50	55.72	4.47	1.05	44.52
Price	Log share price	2.43	2.71	1.47	3.07	3.10	1.00
Size/GNP	Log total assets to gross national product	5.10	4.98	2.21	4.23	4.31	1.69
CHNI	Change in net income	0.00	0.02	0.62	−0.09	−0.07	0.65
Ret Asset	Market stock return	−0.07	−0.01	1.87	−1.49	−0.59	3.63
EXRET	Log excess return relative to S&P-500	−0.08	−0.02	1.86	−1.45	−0.50	3.63
SIGMA	Yearly standard deviation of stock return	0.81	0.24	1.79	2.53	0.94	2.53
MVE/BVE	Market to book value of equity	4.65	0.07	1,676.35	1,121.90	0.05	17,821.47
FVD	Face value of debt according to KMV (million)	7.849	0.432	114.458	2.571	0.434	21.145
MTA	Market value of assets (million)	23.140	3.788	109.834	8.774	3.443	29.987
CASHMTA	Current assets to market value of assets	0.69	0.28	1.03	0.32	0.14	0.70
TL/MTA	Total liabilities to market value of assets	0.97	0.60	1.33	0.36	0.09	1.02
EBIT/TA	Earnings before interest and tax to total assets	0.003	0.00002	0.75	−0.03	0.0002	0.37
NIQ/LTQ	Net income to total liabilities	−0.07	0.01	1.66	−0.07	−0.08	1.37
Rsize	Log ratio of firm’s market cap to that of CRSP index	−8.95	−8.78	2.15	−8.28	−8.14	1.60
NI/MTA	Net income to market value of assets	0.00	0.00	0.18	0.01	−0.02	1.01
NI/MTA-AVG	Net income to average market value of assets	−0.32	0.00	55.44	−0.05	−0.01	0.50
EXRET-Avg	Average stock excess return	−0.07	−0.03	0.97	−1.22	−0.49	1.93
Naive_Sigma_D	Naive volatility of debt	0.25	0.11	0.45	0.61	0.18	0.66
Naive_Sigma_A	Naive volatility of assets	0.68	0.17	1.60	1.82	0.42	2.27

Source: Own elaboration on compustat/CRSP

- Furthermore, when considering data with 50 percent missing observations (or more), and given the significant differences between the sub-samples in Table II, the non-defaulted sub-sample was further divided into four groups of firms according to their GVKEY codes.
- Finally, missing observations were replaced with their corresponding sub-sample's mean.

We estimated both Z-scores and O-scores using a multiperiod logit model with white adjusted error terms, as implemented by Campbell *et al.* (2008). This specification is equivalent to a discrete-time hazard model with a determined hazard function (Shumway, 2001).

Taking X as a vector of explanatory variables, a as a constant, and b as the coefficient vector, the multiperiod logit model defines the probability of bankruptcy (P) of firm i at time t as:

$$P_{i,t} = e^{\alpha+Xi,tbt}/(1+e^{\alpha+Xi,tbt}) \tag{6}$$

As the predictors should be available prior to the event of failure (Ohlson, 1980), these two scores were estimated using a ten-year rolling window approach. For the first window, data from 1990 to 1999 were defined as "in-sample" and used to estimate the logit coefficients that would predict default in 2000. For the second window, the "in-sample" period was 1991-2000 and the coefficients were used to predict default in 2001, and so on. This approach produced 11 sets of ten-year "in-sample" and one-year "out-of-sample" periods.

Table II presents the estimated logit coefficients and the level of significance. The default explanatory power of the predictors is not constant in time, but most predictors are significant at reasonable levels throughout most of the sample. Surprisingly, Size/GNP and CL/CA are only statistically significant in a few subperiods. TL/TA, INTWO, CHNI and MVE/BVD coefficients show the expected sign, while NIQ/LTQ shows the opposite-than-expected sign.

4. Methodology

This subsection explains the three methods implemented to compare the default prediction models: discrete hazard-rate models, intra-cohort analysis and power curve tests[5].

4.1 Hazard rate models

Although a multiperiod logit model overcomes the limitations of a single period logit model, it fails to model time varying changes in the underlying risk of bankruptcy that induces cross-sectional dependence in the data (Hillegeist *et al.*, 2004). Furthermore, the multi period logit model estimates biased and inconsistent parameters (Shumway, 2001). To overcome these limitations, a discrete hazard-rate model, defined by equation (6), considers a time varying hazard rate (a_t) that recognizes the existence of variables that are not firm specific. The time varying hazard-rate affects all firms equally and modifies the underlying probability of default over time. Fluctuations in the baseline hazard-rate will cause observations to be cross-sectionally correlated across time (Hillegeist *et al.*, 2004).

4.2 Intra-cohort analysis

This compares two credit risk models and determines whether one has predictive information that is not contained in the other. Observations are sorted by one default

Table II. Logit coefficients of O-score and Z-score

In-sample period	1990-1999	1991-2000	1992-2001	1993-2002	1994-2003	1995-2004	1996-2005	1997-2006	1998-2007	1999-2008	2000-2009
Out-of-sample	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
n	132,329	144,019	150,345	154,966	157,865	159,421	159,803	158,419	155,088	151,355	147,408
Defaults	238	234	209	209	201	189	181	167	137	117	90
O-score coefficients [a]											
Size/GNP	0.033	0.030	0.008	0.031	0.045	0.047	0.066	0.070**	0.058	0.052	0.088**
TL/TA	−0.090***	−0.065***	−0.005	−0.012***	−0.012***	−0.008***	−0.008***	−0.007***	−0.008***	−0.007***	−0.007**
WC/TA	−0.083***	−0.061**	−0.003	−0.009***	−0.012***	−0.008***	−0.009***	−0.007**	−0.008***	−0.007***	−0.007**
CL/CA	−0.0024***	−0.0008***	0.0001	−0.0005	−0.0005	−0.0005	−0.0005	−0.0002	−0.0002	−0.0001	−0.0001*
OENEG	−1.875***	−1.770***	−1.872***	−1.768***	−1.734***	−1.701***	−1.703***	−1.681***	−1.769***	−1.767***	−1.611***
NI/TA	−0.065***	−0.036	−0.008	−0.007*	−0.003	−0.008***	−0.008***	−0.007***	−0.008***	−0.007***	−0.007***
NIQ/LTQ	−0.006***	−0.006***	−0.007***	−0.007***	−0.007***	−0.008***	−0.007***	−0.007***	−0.007***	−0.007***	−0.015**
INTWO	−1.996***	−1.920***	−1.754***	−1.807***	−1.805***	−1.780***	−1.688***	−1.666***	−1.648***	−1.481***	−1.292***
CHNI	0.492***	0.524***	0.470***	0.494***	0.555***	0.507***	0.430***	0.407**	0.455**	0.495**	0.356*
Constant	7.291***	7.351***	7.562***	7.536***	7.551***	7.599***	7.525***	7.577***	7.855***	7.924***	7.823***
Z-score coefficients											
WC/TA	−0.025*	−0.013**	−0.007	−0.014***	−0.005***	−0.005***	−0.006***	−0.006***	−0.004**	−0.004**	−0.001
RE/TA	0.003632*	0.001954***	0.001174	0.002054***	0.000654*	0.000629*	0.000682*	0.000867**	0.000532	0.000440	−0.000002
EBIT/TA	−0.147***	−0.149***	−0.143***	−0.150***	−0.142***	−0.142***	−0.152***	−0.142***	−0.119***	−0.125***	−0.082***
MVE/BVD	0.00003***	0.00003***	0.00013***	0.00014***	0.00014***	0.00014***	0.00017***	0.00017***	0.00020***	0.00021***	0.00023***
NI/TA	0.0030	−0.0157	−0.0240	−0.0025*	−0.0003	−0.0002	−0.0002	−0.0017	−0.0001	0.0008	0.0006
Constant	−6.319***	−6.425***	−6.589***	−6.616***	−6.677***	−6.749***	−6.801***	−6.874***	−7.056***	−7.193***	−7.429***

Notes: Statistically significant at: *10, **5 and ***1 percent; [a]Original coefficients' signs were inverted since the original specification provides scores increasing in default probability measure into percentiles, and each percentile is sorted according to a second default measure, forming cohorts. Each cohort has the same default risk according to the first measure. So, if the second measure adds explanatory power to the first one, there should be a relatively higher default rate for the low quality firms within each cohort according to the second measure (Kealhofer *et al.*, 2002). To compare default measures, Kealhofer and Kurbat (2001) propose converting each measure into its percentile rank and combining the defaults across cohorts by their percentile scores.

4.3 Power curves

Power curves indicate the percentage of defaults forecasted correctly, given the percentage of non-defaults forecasted incorrectly (Kealhofer *et al.*, 2002), or the trade-off between the defaulting firms to which the model avoids lending, and the proportion of firms the model excludes (Crosbie and Bohn, 2003).

Agents determine a cut-off value, v, and decide not to finance companies whose values fall below v. Defining a cut-off value of v implies two errors: a type I error consists of identifying a company that actually defaulted as a non-default firm. A type II error implies identifying a company that subsequently does not default as a default firm. There is a trade-off between these errors since a high v minimizes type I error but maximizes type II error.

Defining $t_1(v)$, $t_2(v)$ as the type I and II errors (for cut-off v), respectively, the power curve for index i is defined as:

$$P_i(x) = 1 - t_{1i}(t_{2i}^{-1}(x)), \tag{7}$$

One measure is more powerful than another if it produces fewer type I errors than the other, when both produce type II errors equal to x:

$$p_i(x) > p_j(x) \tag{8}$$

When equation (8) is satisfied for all admissible levels of a type II error, the power curve for index i is uniformly more powerful than that for index j.

5. Results

5.1 Information content

We first compare the information content of the four proposed default prediction models defining 1990-1999 as the in-sample period and 2000-2010 as the out-of-sample period.

Table III presents an intra-cohort analysis for the MDDM as the second sort measure. It indicates that the MDDM does not add significant default-related information to either NAMM or Z-score, but it does add certain information to the O-score.

Conversely, as shown in Table IV, the O-score adds default-related information to all the other measures. Such relationship is significant when either Z-scores or Naive DD cohorts are considered since defaults are concentrated in the lowest quality deciles of O-score.

MDDM and O-score are based on different information (mostly market-based in the former, only accounting in the latter), thus an adequate default prediction model should not be based solely on one source of information.

As with the intra-cohort analysis findings a power curve indicates that the O-score presents the best trade-off between type I and II errors, followed by the MDDM,

	Cohort 1-5				Cohort 6-10		
Decile	MDDM deciles within O-score	MDDM deciles within Naïve DD	MDDM deciles within Z-score	Decile	MDD deciles within O-score	MDD deciles within Naïve DD	MDD deciles within Z-score
1	13	12	4	1	5	6	14
2	17	10	7	2	0	7	10
3	5	3	2	3	0	2	3
4	8	3	0	4	0	5	8
5	4	3	0	5	1	2	5
6	5	3	2	6	0	2	3
7	1	0	0	7	0	1	1
8	4	2	2	8	1	3	3
9	1	2	2	9	1	0	0
10	10	8	16	10	15	16	8

Notes: Cohort indicates decile of the first sorting default measure; intra-cohort analysis determines whether the second default measure adds default related information to the first one; MDDM as second sort

Table III.
Intra-cohort analysis

	Cohort 1-5				Cohort 6-10		
Decile	O-score deciles within MDDM	O-score deciles within Naïve DD	O-score deciles within Z-score	Decile	O-score deciles within MDDM	O-score deciles within Naïve DD	O-score deciles within Z-score
1	32	23	10	1	8	18	31
2	5	5	2	2	7	8	11
3	5	1	4	3	3	8	5
4	2	2	0	4	1	1	3
5	0	1	1	5	1	0	0
6	2	5	5	6	5	2	2
7	2	3	3	7	3	2	2
8	2	3	3	8	2	1	1
9	0	2	4	9	5	2	0
10	0	1	3	10	4	2	0

Notes: Cohort indicates decile of the first sorting default measure; intra-cohort analysis determines whether the second default measure adds default related information to the first one; O-score as second sort; number of defaults

Table IV.
Intra-cohort analysis

as Figure 1 shows. The O-score power curve lies above the power curve of the alternative models, indicating that the O-score has the lowest type I error for any given level of type II error, or conversely, the lowest type II error for any given level of type I error. So, O-scores are uniformly more powerful than any other considered measure.

Finally, we use discrete hazard rate models to compare the information content of the default prediction models. Following Hillegeist *et al.* (2004), the ratio of defaults to companies under analysis in the previous year (lagged mean default rate) was defined

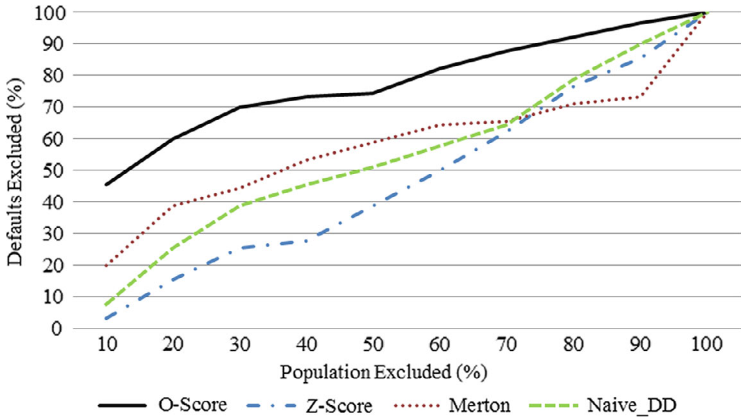

Figure 1.
Power curves: MDDD, Naive MDDM, O-score and Z-score

as the discrete hazard rate. To avoid the influence of outliers, the MDDM and NAMM measures were Winsorized to their fifth and 95th percentile. Values greater than the 95th percentile were set equal to that percentile, and values lower than the fifth percentile were set equal to the fifth percentile.

Hazard rate models of our sample, presented in Table V, indicate that only the MDDM and the O-scores are statistically significant predictors of default, while the NAMM and the Z-score model do not have the statistical power to predict defaults. Table V shows that models MDDM and O-score have the best trade-off between type I and II errors, supporting the findings indicated by the power curves. The reported log likelihood ratios indicate that the O-score is superior. Furthermore, model 1 in Table V combines both MDDM and O-score in a single model and demonstrates that neither statistic suffices to predict default on its own since both variables continue to be statistically significant when taken together (in a bivariate model).

5.2 Proposed reduced form: combining accounting and market information

We showed that the selected models, based on either accounting or market information, are not sufficient statistics to predict default on their own. We present a modified Campbell model (C', incorporating both accounting and market information) and analyze whether such specification outperforms the two significant default prediction measures. Defaults are predicted via a ten-year rolling window approach. Table VI details the coefficients of the proposed model.

	MDDM	NAMM	Z-score	O-score	Model 1
MDDM	0.0137*				0.0154*
NAMM		−0.0006			
Z-score			0.0603		
O-score				0.2062*	0.2113*
Likelihood ratio test	11.68*	0.01	0.62	34.01*	48.45*

Notes: Statistically significant at: *1 percent: discrete hazard rate model coefficients, considering the mean default rate of the previous year as the discrete hazard rate; all models consider only one predictor, except model 1 that combines Merton DD and O-score

Table V.
Relative information with discrete hazard rate models

Table VI. Modified Campbell logit coefficient

In-sample period	1990-1999	1991-2000	1992-2001	1993-2002	1994-2003	1995-2004	1996-2005	1997-2006	1998-2007	1999-2008	2000-2009
Out-of-sample	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
n	132,329	144,019	150,345	154,966	157,865	159,421	159,803	158,419	155,088	151,355	147,408
Defaults	238	234	209	209	201	189	181	167	137	117	90
C' coefficients											
NI/NTA-Avg	−0.0002	−0.0004***	−0.0004	−0.0007***	−0.0007***	−0.0006***	−0.0006***	−0.0006***	−0.0006**	−0.0006**	−0.0004**
EXRET-Avg	0.3775***	0.3402***	0.3403***	0.3487***	0.3472***	0.3562***	0.3586***	0.3369***	0.3457***	0.3334***	0.2957***
TL/MTA	1.2636***	0.6881**	0.6631***	0.6009*	0.5862*	0.5137	0.5805	0.5150	0.5780	0.4208	0.1545
SIGMA	−0.1741***	−0.1713***	−0.1717***	−0.1719***	−0.1873***	−0.1807***	−0.1768***	−0.1973***	−0.1856***	−0.1437***	−0.1088***
RSIZE	0.0020	0.0003	0.0031	0.0131	0.0222	0.0019	0.0176	0.0191	0.0725	0.0554	−0.0091
CASHMTA	−0.2270	−0.0431	0.1166	0.0688	0.0913	0.1616	0.1318	0.1293	0.3563	0.3409	0.3130
MVE/BE	0.00002	0.00001	0.00002	0.00001*	0.00001**	0.00001**	0.00001**	0.00001***	0.00001***	0.00001***	0.00001**
PRICE	0.1893	0.0999	0.1435	0.1403	0.1662	0.1890**	0.2144**	0.2189**	0.2448***	0.1250	−0.0548
Constant	5.7307***	6.1966***	6.2380***	6.4360***	6.5279***	6.3488***	6.4282***	6.5481***	6.9860***	7.2598***	7.5005***

Note: Statistically significant at: *10, **5 and ***1 percent

The comparative performance analysis yields inconclusive evidence. The results of the intra-cohort analysis indicate that the MDDM contains more default predictive information than the other models, particularly in the cohorts with a high default likelihood compared to the alternative models (Table VII).

The power curve test illustrates that no default measure is uniformly more powerful than the others and suggests that different models or even variables may be key to predicting default at different credit risk levels (Figure 2).

The discrete hazard rate models in Table VIII show that the MDDM specification is statistically significant at explaining default, but, according to its log likelihood ratio, it does not explain the data better than the O-score. Model 2 shows that the three significant default measures are still significant at explaining default when considered together, thus providing evidence favoring models that combine different default prediction measures.

	Cohort 1-5			Cohort 6-10	
Decile	C' deciles within O-score	C' deciles within MDDM	Decile	C' deciles within O-score	C' deciles within MDDM
1	38	39	1	4	3
2	7	4	2	0	3
3	3	1	3	0	2
4	5	3	4	1	3
5	1	0	5	0	1
6	3	2	6	1	2
7	4	0	7	10	14
8	2	0	8	5	7
9	1	1	9	0	0
10	3	3	10	2	2

Notes: Cohort indicates decile of the first sorting default measure; intra-cohort analysis determines whether the second default measure adds default related information to the first one; modified Campbell as second sort; number of defaults

Table VII.
Intra-cohort analysis

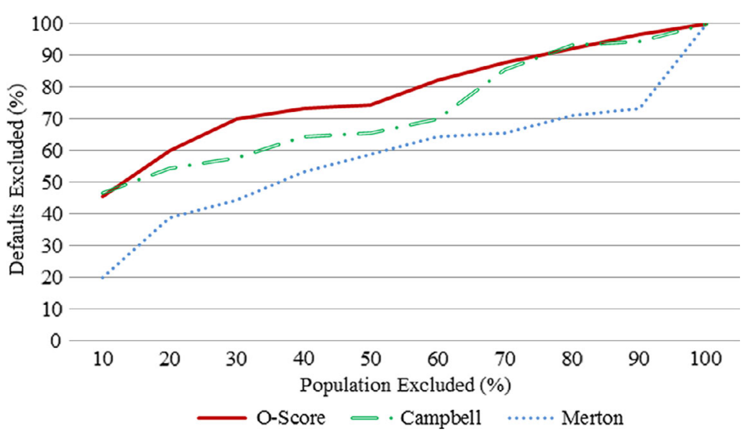

Figure 2.
Power curves: MDDM, O-score and Campbell score

6. Conclusions

The recent financial crisis started in the mortgage market of the USA and expanded globally, causing the failure or intervention of the largest financial institutions. Packages of fiscal and monetary aid were implemented, and major changes in legislation were introduced. The US Congress enacted Dodd-Frank Wall Street Reform and Consumer Protection Act, which tries to mitigate systemic risk, empowering a Financial Stability Oversight Council (FSOC) to act as the systemic risk regulator. Basel III expanded the scope of the incremental risk capital charge from default risk to both default risk and credit rating transition risk. An extended suspicion on the predictive power of default forecasting models was spread.

We assess and compare the forecast ability of credit risk models to answer three questions:

(1) Can these methods adequately predict defaults?

(2) Are there dominant methods?

(3) Given the results, is it safer to rely on a mix of methods?

We examined four existing models, each based mostly on one source of information: O-score, Z-score, MDDM, and NAMM. Evidence in the literature about the completeness of the each model is contradictory. Even when comparing credit risk models with the same methodologies, the literature does not agree on the suitability of the each model, nor in the existence of a superior one.

We tested the information content of the default prediction models with different methods (intra-cohort analysis, power curves and hazard rate models) and found evidence of the outperformance of the O-score relative to all the other models. O-score adds default related information to all the other measures and presents the best trade-off between type I and II errors. The MDDM in our sample has the second best trade-off between both errors, but it is not a sufficient default predictor on its own. Discrete hazard rate models suggest that combining both O-score and MDDM should improve the reliability of default prediction models.

These findings suggest that more precise default prediction models can be obtained by combining information from different sources in two ways:

(1) Using reduced-form models and incorporating variables that can identify the default risk present in the higher deciles of the modified Campbell model, for example, variables that measure the quality of the accounting information introduced in the model and macro variables that affect all firms equally according to market conditions.

	C’	Model 2
C’	0.0137*	−0.47985*
MDDM		0.01818
O-score		0.19698
Likelihood ratio test	26.39*	71.25*

Notes: Statistically significant at: *1 percent; discrete hazard rate model coefficients, with the mean default rate of the previous year as the discrete hazard rate

Table VIII. Discrete hazard rate model for significant measures

(2) Combining default prediction models that analyze the information from different sources. Given the statistical significance of the Merton model, the importance of the algorithms used to calculate its inputs, its defined structural form and the fact that it is not a statistical model, it is not clear how to incorporate information other than the one it already considers. Consequently, predicting default with solely one measure may not be adequate, and better default predictions are likely to be achieved by considering several reduced form models and a Merton distance to default at the same time.

The default predictive ability of different models and testing tools is likely to vary according to the dataset used to measure them. This paper is the first one to our knowledge, to combine such a broad set of models and testing tools to the same dataset. We also show that different sources of information add default predictive power to the other models. We find that the default prediction ability of scoring models (widely used in the banking industry) can be improved by incorporating both market and accounting data and by the MDDM structural form. While previous literature focused on accepting or rejecting either type of model or source of information. After this lengthy discussion the overall conclusion favors use of all models and sources of information, we expect further research to focus on the most appropriate way of combining them.

Notes

1. We use the terms "bankruptcy" and "default" interchangeably.
2. Following Bharath and Shumway (2008), we assume $g = 0.05$ and $h = 0.25$.
3. Kealhofer (2003) presents a detailed summary of the differences between the KMV setting and "standard" MDDM.
4. We follow the procedure Vassalou and Xing (2004) suggest and compute u as the average stock return for each company, replacing it with the average risk free interest rate when the estimated u is negative.
5. Further discussion of hazard-rate models is presented in Shumway (2001), while a deeper analysis of both intra-cohort analysis and power curve tests is provided by Kealhofer and Kurbat (2001) and Kealhofer (2003).

References

Altman, E. (1968), "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy", *Journal of Finance*, Vol. 23, pp. 589-609.

Arun, T. and Turner, J. (2004), "Corporate governance of banks in developing economies: concepts and issues", *Corporate Governance: A International Review*, Vol. 12 No. 3.

Beaver, W. (1966), "Financial ratios as predictors of failure", *Journal of Accounting Research*, Vol. 4, pp. 71-102.

Bharath, S. and Shumway, T. (2008), "Forecasting default with the merton distance to default", *Review of Financial Studies*, Vol. 21, pp. 1339-1369.

Bhatia, A. (2011), "Consolidated regulation and supervisión in the United States", IMF Working Paper WP/11/23.

Blöchlinger, A. (2013), "The next generation of default prediction models: incorporating signal strength and dependency", SSRN, available at: http://ssrn.com/abstract=1080231 or http://dx.doi.org/10.2139/ssrn.1080231

Campbell, J., Hilscher, J. and Szilagyi, J. (2008), "In search of distress risk", *Journal of Finance*, Vol. 63, pp. 2899-2939.

Chava, S. and Jarrow, R. (2004), "Bankruptcy prediction with industry effects", SSRN, available at: http://dx.doi.org/10.2139/ssrn.287474

Claessens, S. (2002), "Benefits and costs of integrated financial services provision in developing countries", paper presented at Joint Netherland-US Roundtable on Financial Services Conglomerates, Washington, DC.

Claessens, S., Dell'Ariccia, G., Igan, D. and Laeven, L. (2010), "Lessons and policy implications from the global financial crisis", IMF Working Paper WP/10/44.

Cocris, V. and Ungureanu, M. (2007), "Why are bank special? Approach from the corporate governance perspective", available at: www.ssrn.com/abstract=1090291

Crosbie, P. and Bohn, J. (2003), *Modelling Default Risk: Modelling Methodology*, Moody's KMV Company, San Francisco, CA.

Hassan, M.K., Wolfe, D.R. and Maroney, N. (2004), "Corporate control and governance in banking", *Corporate Ownership & Control*, Vol. 1 No. 4, pp. 94-107.

Hayden, E. (2003), "Are credit scoring models sensitive with respect to default definitions? Evidence from the Austrian market", paper presented at EFMA 2003 Helsinki Meetings, SSRN, available at: http://ssrn.com/abstract=407709 or http://dx.doi.org/10.2139/ssrn.407709

Hillegeist, S., Keating, E. and Lundstedt, K. (2004), "Assessing the probability of bankruptcy", *Review of Accounting Studies*, Vol. 9 No. 1, pp. 5-34.

John, A., Banerjee, P. and Francis, V. (2007), "Modeling and empirical validation of revised Altman's credit risk model for Indian banks", SSRN, available at: http://dx.doi.org/10.2139/ssrn.960213

Kealhofer, S. (2003), "Quantifying credit risk I: default prediction", *Financial Analysis Journal*, Vol. 59, pp. 30-44.

Kealhofer, S. and Kurbat, M. (2002), *The Default Prediction Power of the Merton Approach, Relative to Debt Ratings and Accounting Variables*, KMV LLC, San Francisco, CA.

Kocenda, E. and Vojtek, M. (2009), "Default predictors and credit scoring models for retail banking", CESifo Working Paper Series No. 2862.

Laeven, L. and Valencia, F. (2010), "Resolution of banking crises: the good, the bad, and the ugly", IMF Working Paper WP/10/146.

Lim, F. (1999), "Comparative default predictive power of EDF's and agency debt ratings", unpublished paper, KMV Corporation, San Francisco, CA.

Martin, D. (1977), "Early warning of bank failure: a logit regression approach", *Journal of Banking & Finance*, Vol. 1.

Merton, R.C. (1974), "On the pricing of corporate debt: the risk structure of interest rates", *Journal of Finance*, Vol. 29, pp. 449-470.

Nam, S.-W. (2006), *Corporate Governance of Banks*, Asian Development Bank Institute, Tokyo.

Ohlson, J. (1980), "Financial ratios and the probabilistic prediction of bankruptcy", *Journal of Accounting Reasearch*, Vol. 18, pp. 109-131.

Sabato, G. (2010), "Credit risk scoring models", SSRN, available at: http://dx.doi.org/10.2139/ssrn.1546347

Shumway, T. (2001), "Forecasting bankruptcy more accurately: a simple hazard model", *Journal of Business*, Vol. 74, pp. 101-124.

Stein, R. (2005), "Evidence on the incompleteness of Merton-type structural models for default prediction", Technical Paper 1-2-1-2000, Moody's KMV Company, San Francisco, CA.

van der Ploeg, S. (2010), "Bank default prediction models: a comparison and an application to credit rating transitions", Masters' thesis, Department of Finance, Erasmus School of Economics, Erasmus University Rotterdam.

Vassalou, M. and Xing, Y. (2004), "Default risk in equity returns", *Journal of Finance*, Vol. 90, pp. 831-868.

Further reading

Blundell-Wignall, A. and Atkinson, P. (2010), "Thinking beyond Basel III: necessary solutions for capital and liquidity", *OECD Journal: Financial Market Trends*, Vol. 2010 No. 1.

Chow, J. and Surti, J. (2011), "Making banks safer: can Volcker and Vickers do it?", IMF Working Paper WP/11/236.

Georg, C. (2011), "Basel III and systemic risk regulation – what way forward?", Working Papers on Global Financial Markets No. 17, Universities Jena and Halle, January.

Rossi, A. and Timmermann, A. (2010), "What is the shape of the risk-return relationship?", paper presented at AFA 2010 Atlanta Meetings Paper, SSRN, available at: http://ssrn.com/abstract=1364750

About the authors

Fernando Castagnolo is a Financial Economist, and he holds degrees both in economics and finance, and a Master degree in finance by Warwick University (UK). He is currently working at Citigroup, London, UK.

Gustavo Ferro is a Research Professor at Instituto de Economía, Universidad Argentina de la Empresa, and CONICET. He holds a doctorate in economics at Universidad del CEMA, Buenos Aires, Argentina. Gustavo Ferro is the corresponding author and can be contacted at: gferro@uade.edu.ar