DETERMINANTS OF CORPORATE DEFAULT: A BMA APPROACH

2012

Carlos González-Aguado
and Enrique Moral-Benito

Documentos de Trabajo
N.º 1221

BANCO DE ESPAÑA
Eurosistema

DETERMINANTS OF CORPORATE DEFAULT: A BMA APPROACH (*)

Carlos González-Aguado

BLUECAP

Enrique Moral-Benito

BANCO DE ESPAÑA

(*) We would like to thank Cristian Bartolucci, Max Bruche and Joan Llull for helpful comments and suggestions. Research funding from the Spanish Ministry of Science and Innovation, Consolider Grant CSD2006-00016 is gratefully acknowledged. Corresponding author: enrique.moral@gmail.com.

Documentos de Trabajo. N.º 1221

2012

The Working Paper Series seeks to disseminate original research in economics and finance. All papers have been anonymously refereed. By publishing these papers, the Banco de España aims to contribute to economic analysis and, in particular, to knowledge of the Spanish economy and its international environment.

The opinions and analyses in the Working Paper Series are the responsibility of the authors and, therefore, do not necessarily coincide with those of the Banco de España or the Eurosystem.

The Banco de España disseminates its main reports and most of its publications via the INTERNET at the following website: http://www.bde.es.

ISSN: 1579-8666 (on line)

Abstract

Model uncertainty hampers consensus on the main determinants of corporate default. We employ Bayesian model averaging (BMA) techniques in order to shed light on this issue. Empirical findings suggest that the most robust determinants of corporate default are firm-specific variables such as the ratio of working capital to total assets, the ratio of retained earnings to total assets, the ratio of total liabilities to total assets and the standard deviation of the firm's stock return. In contrast, aggregate variables do not seem to play a relevant role once firm-specific characteristics (observable and unobservable) are taken into consideration.

Keywords: Default probabilities, Bayesian model averaging, Credit Risk.

JEL classification: G33, C1.

Resumen

Identificar los determinantes de la quiebra de empresas es una cuestión relevante en la literatura de riesgo de crédito. La ausencia de consenso en la especificación del modelo empírico adecuado hace que la metodología de promediado Bayesiano de modelos (BMA, por sus siglas en inglés) sea especialmente apropiada en este contexto. En el presente documento consideramos estas técnicas de promediado con el objetivo de explicar la variabilidad en la probabilidad de quiebra de un panel de empresas estadounidenses. Los resultados sugieren que los determinantes más importantes son características específicas de la empresa, como la ratio de capital sobre total de activos, la proporción de beneficios retenidos sobre los activos totales, la ratio de pasivo total sobre activos totales, y la varianza del retorno de su valor en bolsa. Sin embargo, variables macroeconómicas como los tipos de interés o el crecimiento del PIB no parecen desempeñar un papel relevante una vez que las características de la empresa (observables y no observables) se tienen en cuenta.

Palabras claves: quiebra de empresas, promediado de modelos, riesgo de crédito.

Códigos JEL: G33, C11.

1 Introduction

Understanding which variables are more relevant in predicting default risk at the firm level is one relevant question in the credit risk literature. Whether debt instruments are considered on a stand-alone basis, or within a portfolio context, default probabilities play a critical role in risk assessment and valuation. By (better) anticipating corporate default events, banks are able to better manage risk in their portfolios and investors can value more accurately credit products. For this reason, there exists an enormous literature aiming to understand which the main determinants of corporate defaults are.

Since the seminal paper by Altman (1968), several authors investigated this issue by estimating default prediction models based on binary response specifications (e.g. Ohlson, 1980; Zmijewski, 1984). These models were the basis for more recent developments on reduced-form analysis of default determinants, as for example Shumway (2001), Hillegeist et al. (2004) or Bonfim (2009).[1] The core of the empirical approach in this literature is based on selecting a single model after what amounts to a search in the space of all possible models. Then, researchers typically use the selected models under the implicit assumption that these particular models generated the data. This standard practice results in many different specifications considered in the literature. As a consequence, there is no consensus about the key determinants of firm defaults.

Bayesian model averaging (BMA) represents a promising alternative to address this challenge. By estimating all candidate models resulting from different combinations of regressors, BMA naturally avoids the model selection step. Furthermore, besides appropriate confidence bands,[2] BMA also provides a measure of robustness of each variable in predicting firm default, the Posterior Inclusion Probability (see the Appendix for more details on BMA). Another relevant issue in the literature of corporate default determinants is the existence of firm-specific unobservable heterogeneity correlated with the observable heterogeneity (i.e. firm characteristics proxied by the regressors). Among others, Bonfim (2009) considers Gaussian firm-specific effects in her empirical model under the assumption that they are uncorrelated with the other right-hand-side variables (i.e. random-effects specification). We go a step further in this dimension by considering a conditional fixed-effects logit model which avoids imposing distributional assumptions

[1] Doubly-stochastic intensity models have also been used to produce multiperiod default prediction (e.g. Duffie et al., 2007).

[2] BMA's standard error analogues incorporate both the estimated variances in individual models as well as the variance in estimates of the coefficients across different models.

about the unobservable heterogeneity and, on the other hand, it allows the firm-specific effects to be freely correlated with the regressors.

Once model uncertainty and unobservable heterogeneity at the firm level are taken into account, in this paper we find that four firm-specific variables are the most robust predictors of firm default. In particular, these variables are the standard deviation of the firm stock return, the ratio of working capital to total assets (as a proxy of the firm's liquidity), the ratio of retained earnings to total assets (proxy of firm's profitability/financial leverage), and the ratio of total liabilities to total assets (proxy of firm's leverage). In contrast, aggregate variables such as GDP growth, interest rates or the industrial production index do not seem to directly affect firm default once observable and unobservable firm characteristics are accounted for.

2 Corporate Default and Model Uncertainty

In order to investigate the main determinants of firm default we consider a reduced-form binary choice model in the spirit of Ohlson (1980), Shumway (2001) and Bonfim (2009) among others. Let d_{it} denote a dummy variable indicating whether firm i in year t defaults ($d_{it} = 1$) or stays in business ($d_{it} = 0$). Moreover, x_{it} represents a vector of firm-specific characteristics (typically accounting ratios), and m_t a vector of macroeconomic variables. Given the ingredients above, we estimate the following model:

$$d_{it} = x'_{it-1}\beta + m'_{t-1}\delta + u_{it} \tag{1}$$

with:

$$d_{it} = 1 \quad \text{if} \quad x'_{it-1}\beta + m'_{t-1}\delta + u_{it} \geq 0 \tag{2}$$

$$d_{it} = 0 \quad \text{if} \quad x'_{it-1}\beta + m'_{t-1}\delta + u_{it} < 0 \tag{3}$$

where β and γ are vectors of parameters to be estimated.

Given the above, we can thus denote by $F(x'_{it-1}\beta + m'_{t-1}\delta)$ the probability of default for a given set of regressors and parameters being F some probability function determined by a distributional assumption on the error term. In the absence of a positive theory of default, we select the logistic function given its popularity in the credit risk literature and its convenience when including firm specific effects in the model (see below).

Two important remarks are in place at this stage. First, following earlier literature (e.g. Shumway, 2001; Carling et al., 2007; Bonfim, 2009) we include macroeconomic variables in the empirical model because aggregate shocks might trigger simultaneous defaults (see Hackbarth et al., 2006). Second, given the variation in default rates across industries documented in Li and Zhao (2006), we allow for specific industry effects in our estimation exercises.

Unobservable firm-specific characteristics such as the ability of the CEO or the company's goodwill might also be relevant determinants of firm default. Therefore, we go a step further in the literature and consider firm-specific effects freely correlated with the regressors with the aim of capturing firm-specific characteristics unobservable for the econometrician (i.e. we decompose u_{it} as $\eta_i + v_{it}$ without making any assumption about the relationship between η_i and x_{it-1}). In particular, we make use of a conditional fixed-effects logit specification which avoids the requirement of distributional assumptions on the firm effects (η_i) since $\sum_t d_{it}$ is a sufficient statistic for the effects in this particular case. This approach is the equivalent to fixed effects models in the linear framework in which the effects can be ruled out by demeaning (see for instance Wooldridge, 2001).[3]

The selection of appropriate predictors of default conforming the x and m vectors remains a challenge to credit risk researchers and practitioners. Model uncertainty hampers consensus on which regressors to include in the model. This concern is already present in Ohlson (1980, p. 112) *"To be sure, as is the case in any parametric analysis, a model (i.e. a set of regressors) must be specified, so there is always room for misspecification of the basic probability model."* In practice, this problem results in a number of different vectors of bankruptcy determinants considered in the literature. While virtually all empirical analyses focuses on accounting ratios (e.g. Zmijewski, 1984), the variety of ratios considered is enormous; even in the early eighties, Chen and Shimerda (1981) identified a total of 65 different ratios studied in the literature. Altman (1968) grouped the large set of potential default predictors given by accounting ratios in five different categories: liquidity, profitability, leverage, solvency, and activity. More recently, authors such as Shumway (2001), Carling et al. (2007), Jacobson et al. (2008), and Bonfim (2009) incorporated a variety of macro variables to test the impact of aggregate fluctuations on firm default.

In order to select the set of regressors to include in the empirical specification, researchers typically construct the models based on model selection heuristics, and use the

[3]Results based on the normal distribution (i.e. probit specification) are available upon request and very similar to those based on the logit specification. For the case with firm-specific effects, a probit fixed-effects specification is not available given the lack of a sufficient statistic for the firm-effects in the probit likelihood function.

final model chosen as if it is the true model that generated the data. This is the standard practice in credit risk modeling despite the existence of methods for taking model uncertainty into account. Bayesian model averaging represents an agnostic alternative to the usual approach based on selecting a single regression and deciding which variable is important depending on its associated t-ratio. The key idea of BMA is to consider and estimate all the possible regressions, and then report a weighted average as the estimate of interest. Therefore, model averaging is an agnostic approach in the sense that a researcher relying on this approach holds the view that the true single model is unknown and probably unknowable. Then, the best she can do is to consider all the possible alternatives instead of basing her conclusions on one single model (more details on BMA can be found in the Appendix).[4] In this paper we consider the BMA methodology in order to investigate the most robust determinants of corporate default.

3 Data

Our dataset consists of quarterly data from 1980:Q1 to 2005:Q4 for a maximum of $4,367$ US firms comprising 593 defaults. The dependent variable of our analysis is an indicator variable that marks firm-quarters in which a firm defaulted. We obtain the information on the defaulting firms from the Altman-NYU Salomon Center Corporate Bond Default Master Database. Each entry in the database lists the name of the issuer of the bond and the date of default.

The set of candidate determinants of firm default we consider comprises eight firm-specific variables and four macroeconomic variables collected from CompuStat and the US Federal Reserve Board respectively. In particular, the eight firm-specific variables are five accounting ratios (net income to total assets, total liabilities to total assets, working capital to total assets, retained earnings to total assets, and current assets to current liabilities), the term structure, the recovery rate, and the standard deviation of the firm's stock return. As for the macroeconomic/aggregate variables we consider, these are the rate of growth of GDP, the 1-year return on the S&P500 index, the interest rate, and the US industrial production index. Following earlier literature, we lag all regressors to ensure that they are observable in the beginning of the period in which default is observed, and we winsorize all variables at the 1%-level.

[4] Note however that the BMA approach considers all models given by different combinations of right hand side variables but it is not able to deal with the uncertainty surrounding the distributional assumption in discrete choice models.

The set of candidate default determinants considered in this paper is only a subset of that identified in earlier literature. This is so because, as recently found by Ciccone and Jarocinski (2010) and Moral-Benito (2012), the fewer the potential candidate determinants considered the smaller the sensitivity of the results within the BMA framework.

4 Empirical Results

Table 1 presents the results when estimating by means of BMA the $4,096$ candidate models resulting from all possible combinations of the 12 candidate determinants of corporate default.[5] In particular, it reports the posterior mean, which can be interpreted as a weighted average of all the model-specific coefficient estimates, and the posterior inclusion probability, which indicates the relevance of each regressor in explaining variation in firm defaults. For the implementation of the BMA methodology we employ the unit information prior on the parameter space and uniform priors on the model space following the recommendation in Eicher et al. (2011) (see the Appendix for more details).

In columns (1) and (2) of Table 1 we see the baseline logit specification with neither industry effects nor firm effects. Columns (3) and (4) present the results when including a set of industry dummies. Finally, in columns (5) and (6) we include firm-specific effects in the equation to account for unobserved heterogeneity at the firm level. Taking into consideration all the three specifications, four variables emerge as robust determinants of firm default since they have posterior inclusion probabilities of 1. According to Raftery (1995), evidence for a regressor with a posterior inclusion probability from $50-75\%$ is called weak, from $75-95\%$ positive, from $95-99\%$ strong, and $>99\%$ very strong.[6] Moreover, their ratios of posterior means to posterior standard deviations are larger than two; while these ratios are not distributed according to the usual t-distribution, Sala-i-Martin et al. (2004) note that in most cases, having a ratio around two in absolute value indicates an approximate 95-percent Bayesian coverage region that excludes zero.

The first variable labeled as robust is the standard deviation of the firm stock return, which positively affect the probability of default. This result confirms the finding in Shumway (2001) that less volatile firms are safer than volatile firms. The standard deviation of the firm stock return is a market driven and firm-specific variable which might

[5] Results based on the probit specification for the model-specific step are in line with those presented in Table 1 and are available upon request.

[6] Note also that Masanjala and Papageorgiou (2008) indicate that a PIP of 0.50 corresponds approximately to an absolute t-ratio of one.

Table 1: BMA Determinants of Firm Default

	PIP (1)	P. Mean (2)	PIP (3)	P. Mean (4)	PIP (5)	P. Mean (6)
GDP Growth$_{t-1}$	0.02	-0.16	0.02	-0.18^*	0.62	-0.23^*
Ind. Production Index$_{t-1}$	0.98	-0.11^*	0.98	-0.12^*	0.35	-0.11^*
S&P 500$_{t-1}$	0.83	3.46^*	0.83	3.47^*	0.25	2.88^*
Interest Rate$_{t-1}$	0.02	0.10	0.02	0.10	0.58	0.27
Term Structure$_{t-1}$	0.01	-0.13	0.01	-0.14	0.01	-0.01
σ(Stock Return)$_{t-1}$	1.00	2.41^*	1.00	2.41^*	1.00	1.14^*
Recovery Rate$_{t-1}$	0.05	-0.02	0.06	-0.02^*	0.01	0.00
Net Income/T.Assets $_{t-1}$	0.95	-1.46^*	0.84	-1.48^*	0.35	-1.91
T.Liabilities/T.Assets$_{t-1}$	1.00	1.54^*	1.00	1.62^*	1.00	3.75^*
Work. Capital/T.Assets$_{t-1}$	1.00	-1.87^*	1.00	-2.03^*	1.00	-2.19^*
Ret. Earnings/T.Assets$_{t-1}$	0.99	0.74^*	1.00	0.74^*	1.00	1.27^*
C.Assets/C.Liabilities$_{t-1}$	0.00	0.01	0.00	0.02	0.02	-0.08
Specification	Logit	Logit	Logit	Logit	Logit	Logit
Industry Effects	NO	NO	YES	YES	NO	NO
Firm Effects	NO	NO	NO	NO	YES	YES

Dependent variable is the default dummy. PIP refers to the Posterior Inclusion Probability, our measure of robustness. P. Mean is the Posterior Mean described in the Appendix. * denotes that the ratio of posterior mean to posterior standard deviation is larger than two; while the ratio of posterior mean to posterior standard deviation is not distributed according to the usual t-distribution, Sala-i-Martin et al. (2004) note that in most cases, having a ratio around two in absolute value indicates an approximate 95-percent Bayesian coverage region that excludes zero.

be interpreted as a proxy of the operating leverage. Second, the ratio of total liabilities to total assets can also be labeled as robust predictor of corporate default. This ratio measures the firm's leverage and was also found to be significant by Zmijewski (1984). The posterior mean indicates that the higher the leverage the higher the probability of default. Third, the ratio of working capital to total assets is also a crucial determinant of default as found in Ohlson (1980). This ratio represents a proxy of the firm's liquidity, and, as expected, the higher the liquidity the safer the firm. Lastly, the fourth robust determinant of default emerging from our approach is the ratio of retained earnings to total assets. Altman (1968, 1993) also finds that this variable is a statistically significant predictor of default. While Chen and Shimerda (1981) considered this ratio as representative of the firm's financial leverage, Altman (1968) argues that it measures the cumulative profitability over time.[7]

[7] As already pointed out in Altman (1968), this ratio implicitly considers the age of a firm since it might take some time to build up the cumulative profits.

Finally, the macroeconomic variables we include in the analysis cannot be considered robust determinants of corporate default according to our findings. This is so because in our preferred specification with firm fixed effects their PIP is lower than 0.75 providing only weak evidence for them (Raftery, 1995). In contrast to previous literature (e.g. Bonfim, 2009), this finding indicates that after accounting for firm-specific effects and model specification uncertainty, aggregate variables do not seem to trigger simultaneous defaults as suggested in Hackbarth et al. (2006). Since our study formally considers model uncertainty in the computation of standard errors, we argue that (as shown in Table A1 in the Appendix) ignoring this extra uncertainty might explain previous findings in the literature based on standard errors conditional on a single model.

5 Concluding Remarks

Based on model selection heuristics and significance of estimates, many authors investigate the most relevant factors in determining firm default. The problem with this approach is that confidence bands are not wide enough because, in the absence of clear theoretical guidance, the uncertainty surrounding the model selection step is ignored. In this paper we address this caveat taking into account model uncertainty by means of Bayesian model averaging techniques.

Our empirical findings suggest that four variables are key predictors of corporate default: the standard deviation of the firm stock return, the ratio of working capital to total assets, the ratio of retained earnings to total assets, and the ratio of total liabilities to total assets. In contrast, none of the aggregate variables we consider seems to directly affect corporate default once model uncertainty and firm-specific (observable and unobservable) characteristics are accounted for.

References

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

[2] Altman, E. (1993) "Corporate financial distress and bankruptcy: A complete guide to predicting and avoiding distress and profiting from bankruptcy" New York: John Wiley and Sons.

[3] Bonfim, D. (2009) "Credit risk drivers: Evaluating the contribution of firm level information and of macroeconomic dynamics" *Journal of Banking and Finance*, vol. 33, pp. 281-299.

[4] Carling, K., T. Jacobson, J. Linde, and K. Roszbach (2007) "Corporate credit risk modelling and the macroeconomy" *Journal of Banking and Finance*, vol. 31, pp. 845-868.

[5] Chen, K. and T. Shimerda (1981) "An empirical analysis of useful financial ratios" *Financial Management*, vol. 10, pp. 51-61.

[6] Ciccone, A. and M. Jarocinski (2010) "Determinants of economic growth: Will data tell?" *American Economic Journal: Macroeconomics*, vol. 2(4), pp. 223-247.

[7] Duffie, D., L. Saita, and K. Wang (2007) "Multi-period corporate default prediction with stochastic covariates" *Journal of Financial Economics*, vol. 83, pp. 635-665.

[8] Eicher, T., C. Papageorgiou, and A. Raftery (2011) "Default Priors and Predictive Performance in Bayesian Model Averaging with Application to Growth Determinants" *Journal of Applied Econometrics*, vol. 26, pp. 30-55.

[9] Hackbarth, D., J. Miao, and E. Morellec (2006) "Capital structure, credit risk and macroeconomic conditions" *Journal of Financial Economics*, vol. 82, pp. 519-550.

[10] Hillegeist, S., E. Keating, D. Cram, and K. Lundstedt (2004) "Assessing the probability of bankruptcy" *Review of Accounting Studies*, vol. 9, pp. 5-34.

[11] Jacobson, T., R. Kindell, J. Linde, and K. Roszbach (2008) "Firm default and aggregate fluctuations" *CEPR Discussion Paper*, No. 7083.

[12] Kass, R. and L. Wasserman (1995): "A Reference Bayesian Test for Nested Hypothesis with Large Samples" *Journal of the American Statistical Association*, vol. 90, pp. 928-934.

[13] Leamer, E. (1978) "Specification Searches" New York: John Wiley and Sons.

[14] Li, X. and X. Zhao (2006) "Macroeconomic effects in corporate default" mimeo, University of York.

[15] Masanjala, W. and C. Papageorgiou (2008) "Rough and Lonely Road to Prosperity: a Reexamination of the Sources of Growth in Africa using Bayesian Model Averaging" *Journal of Applied Econometrics*, vol. 23, pp. 671-682.

[16] Moral-Benito, E. (2012) "Determinants of economic growth: A Bayesian panel data approach" *The Review of Economics and Statistics*, vol. 94.

[17] Ohlson, J. (1980) "Financial ratios and the probabilistic prediction of bankruptcy" *Journal of Accounting Research*, vol. 18, pp. 109-131.

[18] Raftery, A. (1995) "Bayesian Model Selection in Social Research" *Sociological Methodology*, vol. 25, pp. 111-163.

[19] Sala-i-Martin X., G. Doppelhofer, and R. Miller (2004) "Determinants of long-term growth: A Bayesian Averaging of Classical Estimates (BACE) Approach" *American Economic Review*, vol. 94(4), pp. 813-835.

[20] Shumway, T. (2001) "Forecasting bankruptcy more efficiently: A simple hazard model" *Journal of Business*, vol. 74, pp. 101-124.

[21] Wooldridge, J. (2001) "Econometric Analysis of Cross Section and Panel Data" Cambridge, Massachusetts: The MIT Press.

[22] Zmijewski, M. (1984) "Methodological issues related to the estimation of financial distress prediction models" *Journal of Accounting Research*, vol. 22, pp. 59-82.

A Appendix

A.1 Single Model Estimation Results

In this appendix we estimate the corporate default specification based on a single model instead of BMA. In particular we choose the full model with all the twelve candidate determinants. Despite any other combination of regressors could be considered, the full model is the most appropriate to evaluate the robustness of all candidate determinants if we restrict ourselves to consider one single combination of regressors. Analogously to Table 1 in the main text, Table A1 presents the estimates from the full model based on three different logit specifications: pooled (columns 1 and 2), with industry effects (columns 3 and 4) and with firm fixed effects (columns 5 and 6).

Table A1: Single Model Determinants of Firm Default

	Coeff. (1)	S.E. (2)	Coeff. (3)	S.E. (4)	Coeff. (5)	S.E. (6)
GDP Growth$_{t-1}$	−0.01	0.07	−0.03	0.07	−0.13	0.09
Ind. Production Index$_{t-1}$	−0.11*	0.04	−0.11*	0.04	−0.07	0.04
S&P 500$_{t-1}$	3.90*	0.91	3.88*	0.91	2.92*	1.12
Interest Rate$_{t-1}$	0.10*	0.05	0.10*	0.05	0.29*	0.09
Term Structure$_{t-1}$	−0.11	0.09	−0.11	0.09	−0.03	0.11
σ(Stock Return)$_{t-1}$	2.37*	0.12	2.37*	0.13	1.17*	0.18
Recovery Rate$_{t-1}$	−0.01	0.01	−0.01*	0.00	0.00	0.01
Net Income/T.Assets $_{t-1}$	−1.40*	0.31	−1.42*	0.33	−1.84*	0.77
T.Liabilities/T.Assets$_{t-1}$	1.55*	0.26	1.66*	0.29	3.58*	0.66
Work. Capital/T.Assets$_{t-1}$	−1.81*	0.26	−1.92*	0.26	−2.46*	0.52
Ret. Earnings/T.Assets$_{t-1}$	0.74*	0.16	0.78*	0.18	1.35*	0.35
C.Assets/C.Liabilities$_{t-1}$	0.01	0.03	0.02	0.03	−0.08	0.08
Specification	Logit	Logit	Logit	Logit	Logit	Logit
Industry Effects	NO	NO	YES	YES	NO	NO
Firm Effects	NO	NO	NO	NO	YES	YES

Dependent variable is the default dummy. * denotes significant at 5%.

Given the t-ratios (or z-scores) resulting from single model estimations, it is common in the literature to label as robust those regressors with coefficient estimates significantly different from zero. According to this practice and the estimates in Table A1, we would conclude that five firm-specific variables (the standard deviation of the firm stock return, the ratio of working capital to total assets, the ratio of retained earnings to total assets, the ratio of net income to total assets, and the ratio of total liabilities to total assets) and two macroeconomic variables (the 1-year return on the S&P500 index and the interest rate) appear as significant/robust determinants of corporate default.[8] However, as we have shown in Table 1, once BMA is considered and model uncertainty is taken into consideration (note that BMA allows incorporating across-model uncertainty into the standard errors, and also obtaining the posterior inclusion probability as an additional robustness check) only four firm-specific variables remain robustly correlated with firm default while no macroeconomic variable remains significant/robust.

A.2 Bayesian Model Averaging

Model uncertainty arises because the lack of clear theoretical guidance on the choice of default probability covariates results in a wide set of possible specifications. Therefore, researcher's uncertainty about the value of the parameter of interest in a regression exists at distinct two levels. The first one is the uncertainty associated with the parameter conditional on a given empirical model. This level of uncertainty is of course assessed in virtually every empirical study. What is not fully assessed is the uncertainty associated with the specification of the empirical model. It is typical for a given paper that the specification of the regression is taken as essentially known; while some variations of a baseline model are often reported, via different choices of control variables, standard empirical practice does not systematically account for the sensivity of claims about the parameter of interest to model choice.

Many researchers consider that one promising approach to account for model uncertainty is to employ Bayesian model averaging (BMA) techniques to construct parameter estimates that formally address the dependence of model-specific estimates on a given model. The basic idea behind BMA is to estimate the distribution of unknown parameters of interest across different models. The fundamental principle of BMA is to treat models and related parameters as unobservable, and to estimate their distributions based on the observable data. In contrast to classical estimation, BMA copes with model uncertainty by allowing for all possible models to be considered, which consequently reduces the biases of parameters.

[8] As in the case of the BMA analysis in the main text, results based on the probit specification are virtually the same and available upon request.

Formally, consider a generic representation of an empirical model of the form:

$$\Psi = \theta X + \epsilon \tag{4}$$

where Ψ is the dependent variable of interest (i.e. the corporate default indicator), and X represents a set of predictors or covariates. Imagine that there exist potentially very many empirical models, each given by a different combination of explanatory variables (*i.e.* different vectors X), and each with some probability of being the 'true' model. This is the starting idea of the BMA methodology.

Using the Bayesian jargon, a model is formally defined by a likelihood function and a prior density. Suppose we have K possible explanatory variables. We will have 2^K possible combinations of regressors, that is to say, 2^K different models - indexed by M_j for $j = 1, ..., 2^K$- which all seek to explain y -the data-. M_j depends upon parameters θ^j. In cases where many models are being entertained, it is important to be explicit about which model is under consideration. Hence, the posterior for the parameters calculated using M_j is written as:

$$g\left(\theta^j|y, M_j\right) = \frac{f\left(y|\theta^j, M_j\right) g\left(\theta^j|M_j\right)}{f\left(y|M_j\right)} \tag{5}$$

and the notation makes clear that we now have a posterior, a likelihood, and a prior for each model. The logic of Bayesian inference suggests that we use Bayes' rule to derive a probability statement about what we do not know (*i.e.* whether a model is correct or not) conditional on what we do know (*i.e.* the data). This means the posterior model probability can be used to assess the degree of support for M_j. Given the prior model probability $P\left(M_j\right)$ we can calculate the posterior model probability using Bayes Rule as:

$$P\left(M_j|y\right) = \frac{f\left(y|M_j\right) P\left(M_j\right)}{f\left(y\right)} \tag{6}$$

Since $P\left(M_j\right)$ does not involve the data, it measures how likely we believe M_j to be the correct model before seeing the data. $f\left(y|M_j\right)$ is often called the marginal (or integrated) likelihood, and is calculated using (5) and a few simple manipulations. In particular, if we integrate both sides of (5) with respect to θ^j, use the fact that $\int g\left(\theta^j|y, M_j\right) d\theta^j = 1$ (since probability density functions integrate to one), and rearrange, we obtain:

$$f\left(y|M_j\right) = \int f\left(y|\theta^j, M_j\right) g\left(\theta^j|M_j\right) d\theta^j \tag{7}$$

The quantity $f(y|M_j)$ given by equation (7) is the marginal probability of the data, because it is obtained by integrating the joint density of (y, θ^j) given y over θ^j. The ratio of integrated likelihoods of two different models is the Bayes Factor and it is closely related to the likelihood ratio statistic, in which the parameters θ^j are eliminated by maximization rather than by integration.

Moreover, considering θ a function of θ^j for each $j = 1, ..., 2^K$, we can also calculate the posterior density of the parameters for all the models under consideration:

$$g(\theta|y) = \sum_{j=1}^{2^K} P(M_j|y)\, g(\theta|y, M_j) \tag{8}$$

If one is interested in point estimates of the parameters, one common procedure is to take expectations across (8):

$$E(\theta|y) = \sum_{j=1}^{2^K} P(M_j|y)\, E(\theta|y, M_j) \tag{9}$$

Following Leamer (1978), we calculate the posterior variance as:

$$\begin{aligned} V(\theta|y) &= \sum_{j=1}^{2^K} P(M_j|y)\, V(\theta|y, M_j) \\ &+ \sum_{j=1}^{2^K} P(M_j|y)\left(E(\theta|y, M_j) - E(\theta|y)\right)^2 \end{aligned} \tag{10}$$

Inspection of (10) shows that the posterior variance incorporates both the estimated variances of the individual models as well as the variance in estimates of the θ's across different models. Hence, the uncertainty at the two different levels mentioned above is taken into account.

Moreover, the BMA methodology allows constructing a ranking of variables ordered by their *robustness*. In this particular case, *robustness* as default intensity determinants. In order to construct our measure of *robustness*, we estimate the posterior probability that a particular variable h is included in the regression, and we interpret it as the probability of that the variable belongs in the true empirical model. In other words, variables with high posterior probabilities of being included are considered as *robust* determinants of default probabilities. This is called the *posterior inclusion probability* for variable h, and it is calculated as the sum of the posterior model probabilities for all of the models including that variable:

$$\text{posterior inclusion probability} = P(\theta_h \neq 0|y) = \sum_{\theta_h \neq 0} P(M_j|y) \tag{11}$$

Once BMA has been described, it is important to note that we will employ a slight variation of this BMA methodology. This is so because the empirical default intensity model considered in the paper departs from the normal linear regression model. Therefore, elicitation of priors on the parameter space may be difficult and controversial since we face a situation where billions of non-regular models must be estimated. Moreover, Bayesian calculations are extremely cumbersome in this case because Bayes Factors may not have closed form and MCMC methods might be required for all the billions of models under consideration, which would make the problem computationally intractable. In particular, our approach uses the Schwarz asymptotic approximation to the Bayes Factor, and therefore substitutes equation (6) by:

$$P\left(M_j|y\right) = \frac{P\left(M_j\right)\left(NT\right)^{\frac{-k_j}{2}} f(y|\widehat{\theta}_j, M_j)}{\sum_{i=1}^{2^K} P\left(M_i\right)\left(NT\right)^{\frac{-k_i}{2}} f(y|\widehat{\theta}_i, M_i)} \tag{12}$$

where $f(y|\widehat{\theta}_j, M_j)$ is the maximized likelihood function for model j. Kass and Wasserman (1995) show that the Schwarz asymptotic approximation formula in (12) could also be obtained with a reasonable prior on the parameter space that is known as Unit Information Prior (UIP).[9]

Moreover, instead of equation (9) we use:

$$E\left(\theta|y\right) = \sum_{j=1}^{2^K} P\left(M_j|y\right) E\left(\theta|y, M_j\right) = \sum_{j=1}^{2^K} P\left(M_j|y\right) \widehat{\theta}_{ML}^{j} \tag{13}$$

where $\widehat{\theta}_{ML}^{j}$ is the ML estimate for model j (i.e. logit or probit in our case). Equation (13) was first considered by Raftery (1995) and it is true if we either assume diffuse priors on the parameter space for any given sample size, or have a large sample for any given prior on the parameter space.

For the BMA estimations in the paper we combine equations (12) and (13) (i.e. Unit Information Prior on the parameter space) together with a Uniform prior on the model space, which assumes that all models are equally probable a priori, i.e. $P(M_j) = \frac{1}{2^K} \forall j$. We consider this particular combination of priors because Eicher et al. (2011) conclude that this choice outperforms any other possible combination of priors previously considered in the BMA literature in terms of cross-validated predictive performance. This combination of priors should also identify the largest set of corporate default determinants.

[9] This prior on the parameter space is a multivariate normal with mean the MLE of the parameters and variance the inverse of the expected Fisher information matrix for one observation.

BANCO DE ESPAÑA PUBLICATIONS

WORKING PAPERS[1]

1101 GIACOMO MASIER AND ERNESTO VILLANUEVA: Consumption and initial mortgage conditions: evidence from survey data.
1102 PABLO HERNÁNDEZ DE COS AND ENRIQUE MORAL-BENITO: Endogenous fiscal consolidations.
1103 CÉSAR CALDERÓN, ENRIQUE MORAL-BENITO AND LUIS SERVÉN: Is infrastructure capital productive? A dynamic heterogeneous approach.
1104 MICHAEL DANQUAH, ENRIQUE MORAL-BENITO AND BAZOUMANA OUATTARA: TFP growth and its determinants: nonparametrics and model averaging.
1105 JUAN CARLOS BERGANZA AND CARMEN BROTO: Flexible inflation targets, forex interventions and exchange rate volatility in emerging countries.
1106 FRANCISCO DE CASTRO, JAVIER J. PÉREZ AND MARTA RODRÍGUEZ VIVES: Fiscal data revisions in Europe.
1107 ANGEL GAVILÁN, PABLO HERNÁNDEZ DE COS, JUAN F. JIMENO AND JUAN A. ROJAS: Fiscal policy, structural reforms and external imbalances: a quantitative evaluation for Spain.
1108 EVA ORTEGA, MARGARITA RUBIO AND CARLOS THOMAS: House purchase versus rental in Spain.
1109 ENRIQUE MORAL-BENITO: Dynamic panels with predetermined regressors: likelihood-based estimation and Bayesian averaging with an application to cross-country growth.
1110 NIKOLAI STÄHLER AND CARLOS THOMAS: FiMod – a DSGE model for fiscal policy simulations.
1111 ÁLVARO CARTEA AND JOSÉ PENALVA: Where is the value in high frequency trading?
1112 FILIPA SÁ AND FRANCESCA VIANI: Shifts in portfolio preferences of international investors: an application to sovereign wealth funds.
1113 REBECA ANGUREN MARTÍN: Credit cycles: Evidence based on a non-linear model for developed countries.
1114 LAURA HOSPIDO: Estimating non-linear models with multiple fixed effects: A computational note.
1115 ENRIQUE MORAL-BENITO AND CRISTIAN BARTOLUCCI: Income and democracy: Revisiting the evidence.
1116 AGUSTÍN MARAVALL HERRERO AND DOMINGO PÉREZ CAÑETE: Applying and interpreting model-based seasonal adjustment. The euro-area industrial production series.
1117 JULIO CÁCERES-DELPIANO: Is there a cost associated with an increase in family size beyond child investment? Evidence from developing countries.
1118 DANIEL PÉREZ, VICENTE SALAS-FUMÁS AND JESÚS SAURINA: Do dynamic provisions reduce income smoothing using loan loss provisions?
1119 GALO NUÑO, PEDRO TEDDE AND ALESSIO MORO: Money dynamics with multiple banks of issue: evidence from Spain 1856-1874.
1120 RAQUEL CARRASCO, JUAN F. JIMENO AND A. CAROLINA ORTEGA: Accounting for changes in the Spanish wage distribution: the role of employment composition effects.
1121 FRANCISCO DE CASTRO AND LAURA FERNÁNDEZ-CABALLERO: The effects of fiscal shocks on the exchange rate in Spain.
1122 JAMES COSTAIN AND ANTON NAKOV: Precautionary price stickiness.
1123 ENRIQUE MORAL-BENITO: Model averaging in economics.
1124 GABRIEL JIMÉNEZ, ATIF MIAN, JOSÉ-LUIS PEYDRÓ AND JESÚS SAURINA: Local versus aggregate lending channels: the effects of securitization on corporate credit supply.
1125 ANTON NAKOV AND GALO NUÑO: A general equilibrium model of the oil market.
1126 DANIEL C. HARDY AND MARÍA J. NIETO: Cross-border coordination of prudential supervision and deposit guarantees.
1127 LAURA FERNÁNDEZ-CABALLERO, DIEGO J. PEDREGAL AND JAVIER J. PÉREZ: Monitoring sub-central government spending in Spain.
1128 CARLOS PÉREZ MONTES: Optimal capital structure and regulatory control.
1129 JAVIER ANDRÉS, JOSÉ E. BOSCÁ AND JAVIER FERRI: Household debt and labour market fluctuations.
1130 ANTON NAKOV AND CARLOS THOMAS: Optimal monetary policy with state-dependent pricing.

1. Previously published Working Papers are listed in the Banco de España publications catalogue.

1131 JUAN F. JIMENO AND CARLOS THOMAS: Collective bargaining, firm heterogeneity and unemployment.

1132 ANTON NAKOV AND GALO NUÑO: Learning from experience in the stock market.

1133 ALESSIO MORO AND GALO NUÑO: Does TFP drive housing prices? A growth accounting exercise for four countries.

1201 CARLOS PÉREZ MONTES: Regulatory bias in the price structure of local telephone services.

1202 MAXIMO CAMACHO, GABRIEL PEREZ-QUIROS AND PILAR PONCELA: Extracting non-linear signals from several economic indicators.

1203 MARCOS DAL BIANCO, MAXIMO CAMACHO AND GABRIEL PEREZ-QUIROS: Short-run forecasting of the euro-dollar exchange rate with economic fundamentals.

1204 ROCIO ALVAREZ, MAXIMO CAMACHO AND GABRIEL PEREZ-QUIROS: Finite sample performance of small versus large scale dynamic factor models.

1205 MAXIMO CAMACHO, GABRIEL PEREZ-QUIROS AND PILAR PONCELA: Markov-switching dynamic factor models in real time.

1206 IGNACIO HERNANDO AND ERNESTO VILLANUEVA: The recent slowdown of bank lending in Spain: are supply-side factors relevant?

1207 JAMES COSTAIN AND BEATRIZ DE BLAS: Smoothing shocks and balancing budgets in a currency union.

1208 AITOR LACUESTA, SERGIO PUENTE AND ERNESTO VILLANUEVA: The schooling response to a sustained Increase in low-skill wages: evidence from Spain 1989-2009.

1209 GABOR PULA AND DANIEL SANTABÁRBARA: Is China climbing up the quality ladder?

1210 ROBERTO BLANCO AND RICARDO GIMENO: Determinants of default ratios in the segment of loans to households in Spain.

1211 ENRIQUE ALBEROLA, AITOR ERCE AND JOSÉ MARÍA SERENA: International reserves and gross capital flows. Dynamics during financial stress.

1212 GIANCARLO CORSETTI, LUCA DEDOLA AND FRANCESCA VIANI: The international risk-sharing puzzle is at business-cycle and lower frequency.

1213 FRANCISCO ALVAREZ-CUADRADO, JOSE MARIA CASADO, JOSE MARIA LABEAGA AND DHANOOS SUTTHIPHISAL: Envy and habits: panel data estimates of interdependent preferences.

1214 JOSE MARIA CASADO: Consumption partial insurance of Spanish households.

1215 J. ANDRÉS, J. E. BOSCÁ AND J. FERRI: Household leverage and fiscal multipliers.

1216 JAMES COSTAIN AND BEATRIZ DE BLAS: The role of fiscal delegation in a monetary union: a survey of the political economy issues.

1217 ARTURO MACÍAS AND MARIANO MATILLA-GARCÍA: Net energy analysis in a Ramsey-Hotelling growth model.

1218 ALFREDO MARTÍN-OLIVER, SONIA RUANO AND VICENTE SALAS-FUMÁS: Effects of equity capital on the interest rate and the demand for credit. Empirical evidence from Spanish banks.

1219 PALOMA LÓPEZ-GARCÍA, JOSÉ MANUEL MONTERO AND ENRIQUE MORAL-BENITO: Business cycles and investment in intangibles: evidence from Spanish firms.

1220 ENRIQUE ALBEROLA, LUIS MOLINA AND PEDRO DEL RÍO: Boom-bust cycles, imbalances and discipline in Europe.

1221 CARLOS GONZÁLEZ-AGUADO AND ENRIQUE MORAL-BENITO: Determinants of corporate default: a BMA approach.

BANCO DE ESPAÑA
Eurosistema

Unidad de Servicios Auxiliares
Alcalá, 48 - 28014 Madrid
Telephone +34 91 338 6360. Fax +34 91 338 6488
E-mail: publicaciones@bde.es
www.bde.es