Analysis of Conditional Probability of Default by the Major Private Debtors in the Colombian Financial System

José Eduardo Gómez González
Inés Paola Orozco Hinojosa
Nancy Eugenia Zamudio Gómez*

I. Introduction

The private corporate sector is the primary debtor in the Colombian financial system (commercial loans account for 54.9% of the total gross portfolio). Consequently, it is extremely important to measure and monitor the risk this sector of the economy might pose to the financial system. Ever since the crisis in the late nineties, Colombian companies have not experienced a comparable situation. Today, the quality indicators for the commercial loan portfolio are at historic lows, and the portfolio has begun to grow, following the standstill in 2003-2005. The non-performing/total loan ratio for companies was 1.63% at June 2006, while real growth in the private commercial loan portfolio was 18.3%.

Coupled with a good economic situation and good corporate performance in recent years, the foregoing poses no imminent risk to financial stability. However, the mid-term risks are still out there, which means this type of risk must continue to be measured and monitored. For example, a hefty increase in commercial loans is good, as it helps to fund investment projects. Nonetheless, an unexpected shock to corporate creditworthiness might be a source of risk to the financial system, because of possible deterioration in the loan portfolio.

---

* Gómez González is a member of the Cornell University Department of Economics. Orozco Hinojosa and Zamudio Gómez are researchers with the Financial Stability Department of the Monetary and Reserves Division at Banco de la República. The opinions expressed herein are those of the authors and do not necessarily reflect the opinions of Banco de la República or its Board of Governors. The useful comments from Dairo Estrada, Carlos Amaya and Andrés Murcia are gratefully acknowledged.

The objective of this paper is to discover the primary determinants of the risk rate[1] or conditional probability of default on financial obligations by companies in the Colombian private sector.[2] Estimates of maximum partial verisimilitude were done with a duration model, using the Camel[3] model variables as input.

The results indicate the size of the debt is the main determinant of conditional probability of default on corporate obligations to the financial system: specifically, the larger the corporate debt the greater the probability of corporate default. Profitability, size and belonging to certain sectors of the economy are other variables that determine this probability. Finally, probability of default on financial obligations was found to be negatively dependent on duration; that is, the longer a company's time to default, the less it is likely to default.

This paper is divided into four parts, including this introduction. The second part contains a theoretical review of the duration model, with emphasis on the risk function proposed by Cox (1972), and a description of the estimation procedure. The figures and results of the estimate are presented in the third section and the conclusions, in the fourth.

II. The Duration Model

The duration model used to estimate the probability of major corporate borrowers defaulting on loans from the Colombian financial system is described in this section, as is the procedure for arriving at that estimate. A duration model was used to analyze the time it takes companies to default. The particular question to be answered with a model of this type is: what is the probability that a company will default on its financial obligations at moment *t*, given that it has not done so up to that point?

Duration models have been used widely in labor economics to determine how long agents remain unemployed and how this variable changes with the economic cycle. Recently, these models were applied in studies on financial economics, such as the one by Gómez and Kiefer (2006), where the authors used a duration model to estimate the amount of time before credit institutions in Colombia's financial system fail in the wake of a negative economic shock.

---

[1] In this paper, the term risk is equivalent to the concept of *hazard* in duration models.

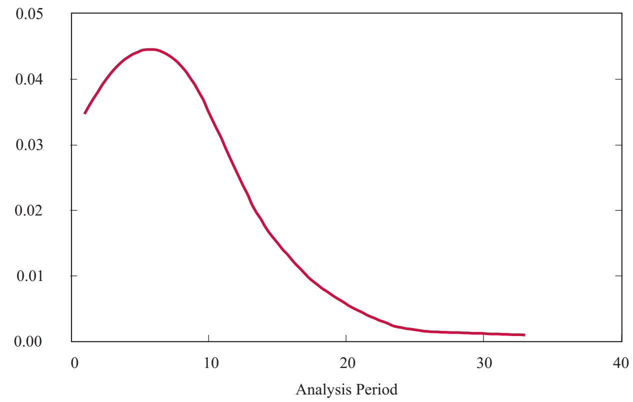
[2] The probability of default is conditioned by companies not having defaulted on obligations to the financial system up to moment *t*.

[3] Camel is the acronym for capital protection, asset quality, management efficiency, earning strength and liquidity risk.

The model applied in the present study is the one most widely used in literature: Cox's semi-parametric proportional risks model (1972). The justification for selecting it and not others, such as the exponential model or Weibull's model, is the non-monotonicity of the risk function. As Graph 1 illustrates, in the early periods, this function increases to a maximum, then declines monotonically.[4]

GRAPH 1

SMOOTHED RISK FUNCTION



Studies based on the assumption that the passage of time will have a particular effect on conditional risk suppose, for example, that the impact of changes in macroeconomic conditions that affect all institutions equally generates a monotonic change in conditional risk that continues over time. One of the advantages of developing non-parametric estimates of the risk rate, such as the ones in this paper, is that they do not imply assumptions of this type. This allows for a more adequate and reliable estimate of the coefficients in the conditional model.

A. Risk Functions and Survival[5]

The probability distribution of the durations is defined as:

(1) $F(t) = Prob(T < t)$

It is, however, common to define the "survival" function in models of this type:

(2) $S(t) = 1 - F(t)$
$S(t) = Prob\ (T \geq t)$

The equation (2) is defined as the probability that random variable T is equal to or greater than a certain value t. Working with a survival function is equivalent to working with a probability function, whatever it may be.

The most useful function in a duration model analysis is the risk function that determines the conditional probability of a company defaulting on its obligations, given that it has not defaulted so far. It is defined as:

4 Exponential distribution and Weibull's distribution impose a certain parameterization of the risk function. The former assumes it should be constant over time; the latter assumes it should grow continuously, decline or remain constant.

5 See Kiefer (1988) for a more detailed explanation of duration models.

$$(3) \quad h(t) = f(t) / S(t)$$

Where $f(t)$ is the probability density function. In the case of the Cox model (1972), the specific risk function is provided by:

$$(4) \quad h(t) = h_0(t)\ \psi(x.\ \beta)$$

Where $h_0(t)$ is the baseline risk function (namely, an unknown parameter that has to be estimated), and $\psi(x.\ \beta) = \exp\ (x'\beta)$ is a vector of explicative variables and unknown coefficients. It is convenient to assume that the form of function $\psi(x.\ \beta)$ is exponential, as this ensures the risk function is not negative, without imposing sign constrains on the interest parameters.

B. Maximum Likelihood Estimation

This method, developed by Cox (1972), allows us to estimate the β parameters without having to specify a particular baseline risk function form $h_0(t)$. The crucial point of this estimate is that the contribution to the partial verosimilitude function of duration *i* is provided by:

$$(5) \quad \frac{h(t_i .\ x_i .\ \beta)}{\sum_{j=1}^{n} h(t_i .\ x_j .\ \beta)}$$

This implies that:

$$(6) \quad \frac{h(t_i .\ x_i .\ \beta)}{\sum_{j=1}^{n} h(t_i .\ x_j .\ \beta)} = \frac{h_0(t)\ \psi(x_i .\ \beta)}{h_0(t) \sum_{j=1}^{n} \psi(x_j .\ \beta)} = \frac{\psi(x_i .\ \beta)}{\sum_{j=1}^{n} \psi(x_j .\ \beta)}$$

And, therefore, this does not depend on the duration.

The verisimilitude function is constructed as the product of the individual contributions given in equation (6). The logarithm of this function is provided by:

$$(7) \quad l(\beta) = \sum_{i=1}^{n} \{\ln \psi(x_i .\ \beta) - \ln[\sum_{j=1}^{n} \psi(x_j .\ \beta)]\}$$

As equation (7) shows, given the absence of the baseline risk function, the order of the durations contains information on the unknown coefficients, which are obtained by maximizing that function.

III. EMPIRICAL EXERCISE

A. Data and Variables

Figures on the two thousand primary debtors in the Colombian financial system were used for this exercise. They contain the history of each firm's loan portfolio classifications, are quarterly and extend from 1997-IV to 2006-I.[6] After some weeding, the total number of companies comes to 989.[7]

A Camel-type model[8] was chosen as the base model for the estimate. Although generally used in bank assessment and ranking exercises, some of its variables can be regarded as possible determinants of the probability of company default; others can be eliminated or substituted with better indicators.

Capitalization, asset quality, management or efficiency, profits and liquidity are the variables that represent the Camel model. According to the *Financial Stability Report,* particularly its regular review of stylized events in Colombia's private corporate sector, two variables in this model are irrelevant to explaining the financial difficulties of Colombian firms, or are not equivalent for the case of banks, which is precisely where the applications of this model are concentrated. For example, asset quality is not a determinant variable of corporate difficulties; in the case of banks, the loan portfolio quality index is. Moreover, the variable generally used to measure efficiency is the ratio of administrative and labor costs to assets. In the case of companies, this is more a size variable, than one of efficiency or management.

The variables included in the model and several statistics descriptive of these variables are presented in Table 1. The *time to failure* variable is equal to the number quarters before a company's loan portfolio rating changes from A/B to C/D/E, or what is considered herein as failure or default. Two important aspects with respect to this variable are shown in Table 1. First, the companies in this sample take 15 quarters, on average, to default on their obligations to the financial system. Secondly, the sample contains companies that defaulted and companies that never defaulted.

---

6 Data as of 1997 were used to cover the period prior to the crisis in the late nineties.

7 The simple was trimmed several times before the estimate was made. The initial quarter is 1997-IV, which is considered the base period. With this assumption, the companies that defaulted on loans during the base period were the first to be eliminated, followed by those with no available information for the next quarter (1998-I). The final criterion for remaining in the sample was having balance sheet and earning statement data for the base period.

8 See Gilbert, Meyer and Vaughn (2000) for a more detailed explanation of this model.

TABLE 1

DESCRIPTIVE STATISTICS OF THE VARIABLES INCLUDED IN THE MODEL

	Average	Deviation	Minimum	Maximum
Time to failure	15.341	12.681	1.000	33.000
Debt	0.334	0.182	0.000	1.314
Liquidity	2.015	7.021	0.058	204.356
Size	16.602	1.480	7.631	20.876
Capitalization	0.437	0.223	-0.898	0.989
Dummy profitability	0.497	0.500	0.000	1.000
Dummy industry	0.434	0.496	0.000	1.000
Dummy construction	0.131	0.338	0.000	1.000

Source: Office of the National Superintendent of Financial Institutions, National Superintendent of Corporate Affairs, and the authors' calculations.

The *debt* is the debt over assets ratio. It was 33% on average. The liquidity indicator is the ratio of liquid assets to liquid liabilities. On average, it shows the companies' short-term assets covered more than twice the liabilities nearest to maturity. The size measure was constructed as the sales logarithm, and capitalization is equal to equity over assets.

Three dichotomic variables were included in the estimate; profitability was constructed as profit before taxes over assets, and the respective dummy variable is equal to 1 when the company has negative profitability. On the basis of Table 1, we can infer that approximately half the companies in the sample showed negative profitability in 1997. Two sector variables for industry and construction were developed the same way. They are equal to 1, if the company belongs to these sectors and to 0 if it does not.[9]

B. Estimate and Results

The results of the estimate are presented in Table 1. To facilitate interpretation, it shows the coefficients and not the risk rates.[10] The combined significance test indicates the included variables are relevant to explaining duration. All the variables show the expected sign, except the liquidity variable, but it is not significant. Therefore, one can assume that its effect on the risk rate is 0.

9 The intention of these dichotomic variables is to control sectoral effects. The industrial sector was chosen because it is the most representative of the sample, and the construction sector, because it is one of the most fragile throughout the period in question.

10 The estimate shows the hazard ratios rather than the coefficients. The hazard ratios logarithm is calculated to obtain the coefficients.

Estimate by Maximum Partial Verosimilitude

Variable	Coefficient	Standard Error
Dummy Profitability	0.375242 ***	0.0993396
Debt	1.314651 ***	0.3511115
Liquidity	-0.000951	0.0052542
Size	-0.076329 **	0.0347549
Capitalization	-0.246420	0.3022769
Dummy Industry	-0.277751 **	0.1104563
Dummy Construction	0.513085 ***	0.1334809
Number of Observations	989	
Likelihood Log	-3049.3886	
LR chi2(7)	151.2	
Prob > chi2	0.0000	

** 95% significant.
*** 99% significant.

One of the most important results is the effect of the debt. It has the largest coefficient and indicates that, all else being constant, an increase in the companies' debt spells greater conditional probability of default during the period analyzed. With the profitability variable coefficient, the indication is that a company's loss increases the risk rate. The size variable indicates the largest companies are less likely to default, since they are regarded as firms in a higher category, where default on debts can be more costly.

Finally, belonging to certain sectors of the economy can influence the risk rate. For example, being part of the industrial sector is tantamount to being part of a less volatile sector in terms of income. This implies a lower risk rate.[11] However, all things being constant, being part of the construction sector involves a higher probability of default. This result has been a constant in other exercise used to estimate corporate probability of failure (be it based on bankruptcy or default).[12]

Proportional risks are the primary assumption in Cox's model (1972); hence, the importance of validating it. The results of the proportional risks test are

11 Approximately 50% of the sample belongs to the industrial sector.

12 See the work by Arango, Zamudio and Orozco (2005) in the case of bankruptcy. See Chapter IV of this report in the case of default. The reason for this result is that the exercises consider a company's entire history. Therefore, although the construction sector has recovered and is in better situation, it faced adverse circumstances during the crisis in the nineties. The exercise includes those circumstances.

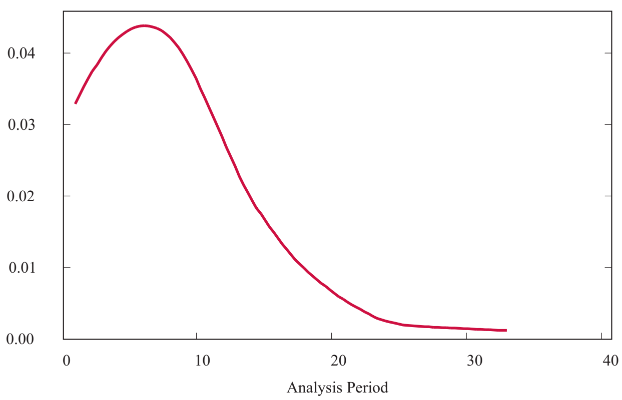
TABLE 3

PROPORTIONAL-HAZARDS ASSUMPTION TEST

	Rho	χ^2	Degrees of Freedom	Prob > χ^2
Dummy Profitability	0.029	0.430	1	0.514
Debt	0.007	0.020	1	0.891
Liquidity	0.029	0.480	1	0.490
Size	0.047	1.090	1	0.297
Capitalization	0.041	0.680	1	0.408
Dummy Industry	-0.061	1.810	1	0.178
Dummy Construction	-0.007	0.030	1	0.871
Global Test		**3.5**	**7**	**0.835**

GRAPH 2

COX PROPORTIONAL-HAZARDS REGRESSION



shown in Table 3, where the null hypothesis is that the slope of the coefficients is equal to 0. In other words, the coefficients would not vary over time. The test shows the individual results for each coefficient and for the global test. In each case, we cannot rule out the null hypothesis, which maintains the coefficients do not vary over time. Therefore, it is possible to conclude that the Cox proportional-hazards assumption is adequate in this case.

The estimated risk function of the model can be obtained once the estimate and the proportional-hazards test have been done. This function is presented in Graph 2 for the average values of the variables. Their pattern is similar to the risk function shown in Graph 1.[13] Conditional probability increases to a maximum point, then declines and is now at its lowest level, indicating a negative correlation between probability of default and duration. In other words, the longer it takes a company to default, the less its probability of default.

Graph 3 shows the risk function estimated for three types of situations. In the upper panel (A), the function is divided between companies with negative profitability and those with above-0 profitability. Both groups follow the same tendency; however, there is a major difference in level;

13 Graph 1 is the non-parametrically estimated risk function and pertains to the instantaneous conditional probability of default (in other words, it does not depend on the model's exogenous variables). Graph 2 shows the estimated risk function, where the risk function is expected to be similar to the one obtained non-parametrically, as is the case. This indicates the estimated model adjusts appropriately to the non-parametric model, which is closest to the empirical distribution of the duration.

the estimated conditional probability is greater for the group with losses in 1997, although the gap has been closing recently.

The estimated risk function for companies in the industrial sector is shown in the middle panel (B), compared to those in the other sectors. The lower panel (C) shows the conditional probability for companies in the construction sector compared to companies in the other sectors of the economy. The graphs show the tendency for all the groups is the same, but there are some differences in level. In particular, compared to the other sectors, being part of the industrial sector implies less conditional probability of default. On the contrary, being in the construction sector leads to a higher risk rate. As with profitability, these differences are becoming less and the gap is closing steadily.[14]

IV. CONCLUSIONS

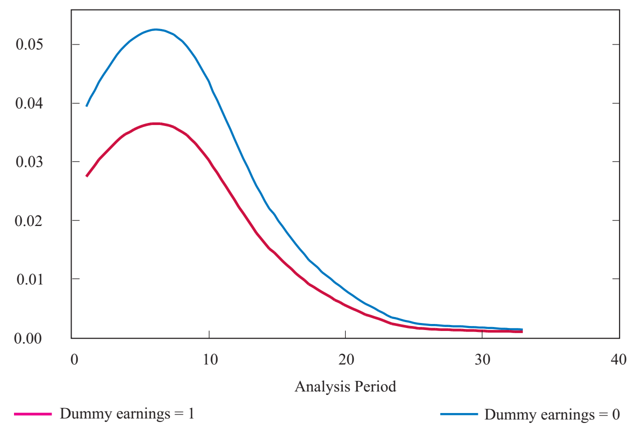
Using a duration model, this work estimates the conditional probability of loan default by firms in the private corporate sector. Specifically, it uses the Cox proportional-hazards model (1972) and develops an estimate of maximum partial verisimilitude, where the variables used originate initially with a Camel model adapted for the case of Colombian companies.

The results show the extent of corporate debt is the primary determinant of conditional probability of default. Other less important variables are company size and profitability. The impact belonging to certain sectors of the economy has on conditional probability of default is an interesting result. In particular, being part of

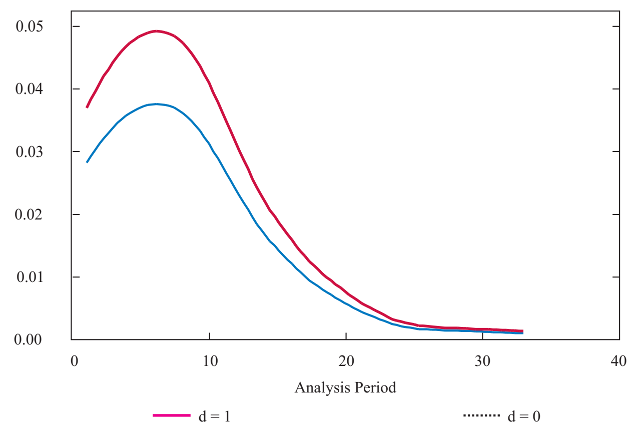
[14] The reduction in the gap between company groups also might be due to the convergence of non-conditional probability of default towards 0.

Graph 3

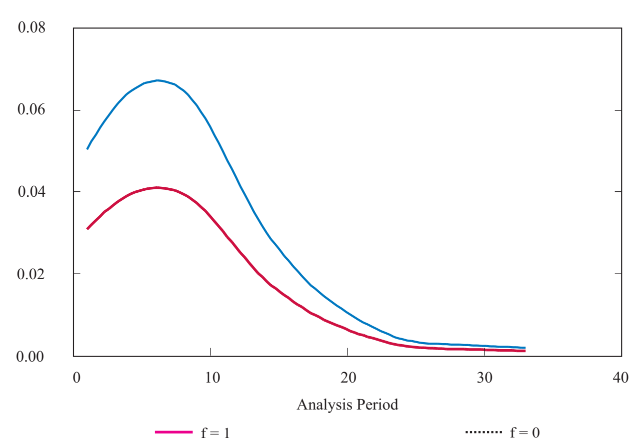
A. Cox Proportional-hazards Regression for Profitability



B. Cox Proportional-hazards Regression for the Industrial Sector



C. Cox Proportional-hazards Regression for the Construction Sector



industry generates less probability, while being part of the construction sector translates into higher probability.

One implication of the results is the negative correlation between probability of default and duration. In other words, the longer a company takes to default the less its probability of default. Finally, considering the excellent economic situation and good business performance in recent years, the private corporate sector clearly implies no imminent risk to financial stability at this time. Nevertheless, the mid-term risks continue, which means efforts to measure and monitor them must continue as well.

BIBLIOGRAPHY

Arango, J. P.; Zamudio, N. E.; Orozco, I. P. (2005), "Riesgo de crédito: un análisis desde las firmas" in *Reporte de Estabilidad Financiera*, December, Banco de la República, pp. 79-87.

Cox, R. (1972), "Regression Models and Life-Tables" in *Journal of the Royal Statistical Society*, No. B 34, pp. 187-220.

Gilbert, R.; Meyer, A.; Vaughan, M. (2000), "The Role of a CAMEL Downgrade in Bank Surveillance," Working Paper No. 2000-021A, St. Louis Federal Reserve Bank, U.S.A.

Gómez, J. E.; Kiefer, N. M. (2006), "Explaining Time to Bank Failure in Colombia during the Financial Crisis of the Late 1990," *Borradores de Economía*, No. 400, Banco de la República.

Kiefer, N. M. (1988), "Economic Duration Data and Hazard Functions," *Journal of Economic Literature*, No. 2, Vol. 26, pp. 646-679.