Duration analysis of new firms in the banking industry

Jesus Orbe and Vicente Núñez-Antón

Abstract—This paper studies the duration or survival time of commercial banks active in the Moscovian three month Rouble deposits market, during the 1994-1997 period. The privatization process of the Russian commercial banking industry, after the 1988 banking reform, caused a massive entry of new banks followed by a period of high rates of exit. As a consequence, many firms went bankrupt without refunding their deposits. Therefore, both for the banks and for the banks' depositors, it is of interest to analyze which are the significant characteristics that motivate the exit or the closing of the bank. We propose a different methodology based on penalized weighted least squares which represents a very general, flexible and innovative approach for this type of analysis. The more relevant results are that smaller banks exit sooner, banks that enter the market in the last part of the study have shorter durations. As expected, the more experienced banks have a longer duration in the market. In addition, the mean survival time is lower for banks which offer extreme interest rates.

Keywords—Banking, Censored, Duration, Kaplan-Meier.

I. INTRODUCTION

THE accelerated development of the Russian commercial banking industry has its origin in 1988, right after the banking reform. This rapid evolution of the market was the result of the null entry barriers that caused high rates of entry and that were consequently followed by a period of high rates of exit. As a consequence, many banks have to exit the market without refunding their deposits. Therefore, both for the banks and for the banks' depositors, it is of interest to analyze which are the significant characteristics that motivate the exit or the closing of the bank. The survival of new firms has been previously studied, specially in the manufacturing industry, [1] and [2] for firms in United States, [3] for the German industry, [4] for Portugal, [5] for Italy, [6] for the Netherlands and [7] and [8] for Spain. In the context of the banking industry, [9,10,11,12] used hazard models to analyze the bank failure process in the United States, and [13] used the same methodology for the Russian market. This methodology studies the effect of the covariates on the hazard function. This function specifies the instantaneous rate of leaving the market at time t, given that the bank has not left the market up till time t. The most applied model is the semiparametric proportional hazard model proposed by Cox [14].

This paper proposes an alternative and flexible approach to study the direct effects of relevant factors on the survival time instead of on the hazard rate. In addition, and taking as a reference the analysis in [13], we propose the use of a different methodology that allows us to estimate the unknown

Authors are with Dpt. Econometría y Estadística, University of the Basque Country, Avda. Lehendakari Agirre 83, 48015 Bilbao, Spain (email: jesus.orbe@ehu.es).

functional form of a covariate's effect and that does not require to have any assumption on the probability distribution of the survival time. Section II introduces the model and its estimation process. Section III describes the data. The results of the analysis are included in Section IV. Finally, Section V includes some general conclusions.

II. METHODOLOGY

The methodology used in this study is based on a regression model to evaluate the effect of k covariates, $X_i = (x_{i1}, \ldots, x_{ik})'$, on the duration of the i-th bank, t_i. Thus the relation between the covariates and the logarithm of the duration, $Z_i = \ln t_i$, is modelled as:

$$Z_i = X_i'\beta + \epsilon_i, \qquad i = 1, \ldots, n \qquad (1)$$

One of the main reasons to consider this alternative model is the clear and easy interpretation of the results, because here we measure the direct effect of the covariates on the duration instead of on the conditional probability of exit of the market at time t, modelling the so called hazard function or hazard rate $\lambda(t)$

$$\lambda(t) = \lim_{\Delta t \to 0^+} \frac{P(t \le T \le t + \Delta t | T \ge t)}{\Delta t}$$

In duration or survival data analysis, and because of censorship, we do not observe completely the duration variable T. Instead of it we have the observed duration variable Y. That is, if the observation is not censored, we observe the duration of the bank in the market, $Y = T$. On the other hand, if the observation is censored, C, we only know that the duration is larger than this value, i.e. $T \ge C$.

Our proposed estimation method is based on the weighted least squares (WLS) methodology. At the end of the study some banks do not leave the market. Therefore, their durations are censored and, as is well known, in the presence of censorship, the ordinary least squares methodology produces biased and inconsistent estimators. In order to solve this problem, we use the WLS approach where the weights take into account the effect of censorship. Thus, the estimator of coefficients β can be obtained by weighted least squares, minimizing the next expression

$$\sum_{i=1}^{n} W_i (\ln y_i - X_i'\beta)^2 \qquad (2)$$

where $\ln y_i$ denotes the logarithm of the observed duration variable and W_i is the Kaplan-Meier weight, that is calculated as the contribution of the y_i observation on the Kaplan-Meier

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:4, No:6, 2010

estimator of the distribution function of the duration variable T (see [15]),

$$\hat{F}(y) = 1 - \prod_{i=1}^{n} \left[\frac{n - \text{Rank}(y_i)}{n - \text{Rank}(y_i) + 1} \right]^{1_{\{\delta_i = 1, y_i \le y\}}}$$

where $\delta_i = 0$ when the correspondent observed duration is a censored observation and $\delta_i = 1$ when the observed duration is not a censored observation. An easy form to compute the weights W_i can be described as:

1) Order the sample of n observed durations, $y_1 < y_2 < \ldots < y_n$, and put the same weight, $\frac{1}{n}$, on all observations.
2) Start with the smallest value of the sample Y, y_1. If it is not censored, keep its weight; otherwise, if it is censored, put zero weight on it and redistribute its weight between the observed durations that are larger than this observation.
3) Repeat previous step until the largest value of the sample, y_n.

The solution of minimizing problem in Equation (2) can be written as:

$$\hat{\beta} = (X'WX)^{-1}X'W \ln Y \qquad (3)$$

where W is a diagonal matrix formed with the Kaplan-Meier weights W_i, $i = 1, 2, \ldots, n$.

This estimation method has been studied in [16], the consistency of the β estimator, and [17], its asymptotic normality, for the case of the linear regression models with censored data. In our case, we need a method that considers both linear and nonlinear effects or unknown effects. Thus, we propose using a generalization of model (1), where we will introduce a nonparametric component, where we do not specify any functional form for the effect of one covariate on the duration variable. The motivation for this lies on the specification of the effect of the covariate representing the relative interest rate. Reference [13] points out the possible quadratic relationship for this covariate on the hazard, indicating that the higher the deposit interest rate offered by the bank, the higher its risk. In addition, banks which are offering low interest rates may not be interested in attracting customers and, thus, are trying to exit the market. Reference [13] specified a quadratic relationship for this covariate but the estimation of the effect was not statistically significant. This paper allows for a general specification of the functional form of the effect of this covariate, without assuming any parametric form for it. Besides, it does not need to assume any distribution for the duration variable. Reference [13] uses a parametric proportional hazard regression model assuming a parametric probability distribution, Gompertz distribution, for the basic hazard function. Our proposal will avoid any incorrect specification of the model and its consequences (see, e.g., the study presented in [18]). We use the technique of penalized weighted least squares (PWLS) and, thus, the estimators are obtained by minimizing (in matrix form)

$$(\ln Y - X\beta - Nf)' W(\ln Y - X\beta - Nf) + \alpha \int f''(z_2)^2 dz_2, \qquad (4)$$

with respect to β and $f(\cdot)$. In Equation (4), z_2 takes the value of the covariate introduced in a nonparametric form (i.e., the nonlinear covariate), vector f contains the values $f_j = f(z_{2j})$ for the d distinct values $z_{21} < z_{22} < \ldots < z_{2d}$, of the covariate Z_2. The connection between $z_{21} < z_{22} < \ldots < z_{2d}$ and $z_{21} < z_{22} < \ldots z_{2n}$ is given through the $n \times d$ incidence matrix N which assigns the respective value of the covariate Z_2 to the ith individual; that is, $N_{ij} = 1$ if $Z_{2i} = z_{2j}$. There is no functional assumption on $f(\cdot)$ in Equation (4) and this contains a penalized term that considers the "smoothness" of $f(\cdot)$, by using the integrated squared second derivatives. The parameter α controls for the relevance of the goodness-of-fit and for the smoothness of $f(\cdot)$. The purpose is to have a reasonable goodness-of-fit together with a smooth $f(\cdot)$. The effect of the censorship is handled with the previously defined W_i weights. Reference [19] studied this estimation process and used a simulation study to check on the properties of the proposed methodology.

III. DATA

The dataset, provided by Martin Carree, shows information about commercial banks active[1] in the Moscovian three months Rouble deposits market during the 1994-1997 period[2]. Reference [13] analyzed the effect of the covariates on the hazard of the license duration, from its license date until its exit. We study the effect of the covariates on time from the entry of the bank in the market until its exit[3]. The previous time, from the license date until the entry in the market, is used to built the covariate experience (*experience*). It is expected that more experienced banks suffer less from a liability of a newness and, therefore, have higher durations. Usually, other relevant factor is the size. Small banks have more risk and, thus, have lower durations (i.e., the liability of smallness). The size is measured using the share of the bank in the Moscovian three months Rouble deposits market. We use the reciprocal of this share (*sizeinv*). This higher risk of small and young firms is a common characteristic in different industries (see, e.g., [2], [6], [7], [10] and [20]), and it is well described in [21]. Another measure of the risk is related to the interest rate offered to depositors. The risk increases and the survival time decreases with the interest rate. This risk is measured by using the ratio between the interest rate offered by the bank and the mean interest rate offered at the market (*interestrate*). As a consequence of this financial crisis, the government hardened its monetary policy, increasing the probability of exit. We use a covariate measuring the period in which the bank enters the market (*period*).

---

[1] See [13] to understand when a bank is considered active in the Moscovian three months Rouble deposits market.

[2] For a detailed description of the dataset see [13].

[3] Some observations are censored because some banks remain on the market at the end of the study.

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:4, No:6, 2010

IV. RESULTS

We propose the estimation of a partial regression model for censored data, previously described in Section II. The covariates: *experience*, *period* and *sizeinv* are introduced parametrically (X matrix), and the estimation of each β coefficient measures the effect of the covariate on the duration of the bank. The covariate *interestrate* (Z_2 vector) is introduced nonparametrically and its effect on the duration is measured by the estimation of function $f(\cdot)$. Thus, we do not assume any functional form for the effect of the *interestrate* covariate on the duration and therefore we can check if the effect is quadratic, as is supposed in [13], and, also if this effect is statistically significant (in [13] results nonsignificant). That is, this is the proposed model:

$$
\begin{aligned}
\ln t_i = \ & \beta_1 + \beta_2 sizeinv_i + \beta_3 experience_i + \\
& + \ \beta_4 period_i + f(interestrate_i) + \eta_i
\end{aligned}
\tag{5}
$$

TABLE I
ESTIMATION OF THE EFFECT OF THE COVARIATES INTRODUCED IN THE PARAMETRIC COMPONENT OF THE MODEL

Variable	$\hat{\beta}$	95% **Bootstrap Confidence Interval**
period	-0.1053	(-0.1588, -0.0740)
experience	0.0707	(0.0531, 0.1167)
sizeinv	-0.0003	(-0.0005, -0.0002)

Table 1 shows that the estimation of the effects of the covariates, introduced linearly through the parametric term, are statistically significant. The signs of the estimated coefficients are contrary to the ones obtained in [13]. This makes sense because we estimate the effect on the duration or survival time and not on the hazard function. Thus, when we have a positive effect of some covariate on the duration, the effect of this covariate on the hazard function is negative. The results are similar, the negative sign of *sizeinv* indicates that smaller banks exit sooner. The effect of *period* suggests that banks that enter in the market in the last part of the study have lower durations. As expected, the sign of *experience* shows that more experienced banks have a longer duration in the market. These results are consistent with the ideas and conclusions in others studies (see e.g. [2], [6], [7], [10] and [20]). As for the estimation of the *interestrate* effect, captured in the nonparametric term, Figure 1 presents a U inverted functional form effect (on the duration). This agrees with the U quadratic effect (on the hazard function) obtained in [13], but in his case it was not significant. Therefore, Figure 1 shows that the mean survival time is lower for banks which offer extreme interest rates. It seems that banks which offer the highest interest rates have more risk and exit sooner, and banks which offer the lowest rates are reflecting that they are not interested in attracting customers and are trying to exit the market. The confidence intervals reported in Table 1 and in Figures 1 have been calculated by using bootstrap resampling techniques (see, e.g., [22]).
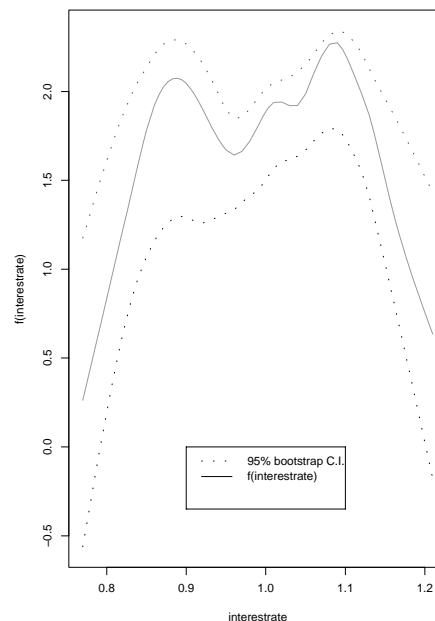

Fig. 1. Estimation of the nonparametric component. This figure shows the estimate of $f(z_2)$, that represents the effect of the *interestrate* covariate on the duration, together with the 95% bootstrap confidence bands

V. CONCLUSIONS

The results obtained strengthen the ideas put forward in [13] and add new insights to them. However, what is really important is that these results have been derived by using an alternative approach and applying a very general model. On the one hand, instead of assuming a quadratic form for the functional form of the effect of *interestrate*, we introduce it in a nonparametric form. On the other hand, we do not assume any probability distribution for the survival time. Therefore, this methodology is robust under incorrect specification of the functional form effects and under incorrect specification of the distribution probability and its consequences.

REFERENCES

[1] Audretsch DB, Mahmood, T. New firm survival: new results using a hazard function. *Review of Economics and Statistics* 1995; **77**: 97-103.
[2] Dunne T, Roberts MJ, Samuelson L. The growth and failure of US manufacturing plants. *Quarterly Journal of Economics* 1989; **104**: 671-698.
[3] Wagner J. The post-entry performance of new small firms in German manufacturing industries. *Journal of Industrial Economics* 1994; **42**: 141-154.
[4] Mata J. Firm growth during infancy. *Small Business Economics* 1994; **6**: 27-39.
[5] Arrighetti A. Entry, growth and survival of manufacturing firms. *Small Business Economics* 1994; **6**: 127-137.

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:4, No:6, 2010

[6] Audretsch DB, Mahmood T. Firm survival in the Netherlands. *Review of Industrial Organization* 2000; **16**: 1-11.

[7] Segarra A, Callejn M. New firms' survival and market turbulence: new evidence from Spain. *Review of Industrial Organization* 2002; **20**: 1-14.

[8] Esteve S, Sanchs A, Sanchs JA. The determinants of survival of Spanish manufacturing firms. *Review of Industrial Organization* 2004: **25**, 251-273.

[9] Shumway T. Forecasting bankruptcy more accurately: a simple hazard model. *Journal of Business* 2001; **74**: 101-124.

[10] Wheelock DC, Wilson PW. Why do banks disappear? The determinants of US bank failures and acquisitions. *Review of Economics and Statistics* 2000; **82**: 127-138.

[11] Whalen G. A proportional hazards model of bank failure: an examination of its usefulness as an early warning tool. *Economic Review (Federal Reserve Bank of Cleveland)* 1991; **27**: 21-31.

[12] Lane WS, Looney S, Wansley J. An application of the Cox proportional hazards model to bank failure. *Journal of Banking and Finance* 1986; **10**: 511-531.

[13] Carree MA. A hazard rate analysis of Russian commercial banks in the period 1994-1997. *Economic Systems* 2003; **27**: 255-269.

[14] Cox DR. Regression models and life-tables. *Journal of Royal Statistical Society-Series B* 1972; **34**: 187-220.

[15] Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 1958; **53**: 457-481.

[16] Stute W. Consistent estimation under random censorship when covariables are present *Journal of Multivariate Analysis* 1993; **45**: 89-103.

[17] Stute W. Distributional convergence under random censorship when covariables are present *Scandinavian Journal of Statistics* 1996; **23**: 461-471.

[18] Holländer N, Schumacher M. Estimating the functional form of a continuous covariate's effect on survival time. *Computational Statistics and Data Analysis* 2006; **50**: 1131-1151.

[19] Orbe J. Un Modelo de Regresión Parcial Censurado para Análisis de Supervivencia. Unpublished Ph.D. Thesis, Departamento de Econometría y Estadística, Facultad de Ciencias Económicas y Empresariales, Universidad del País Vasco-Euskal Herriko Univertsitatea, Bilbao, Spain, 2000.

[20] Evans DS. Test of alternative theories of firm growth. *Journal of Political Economy* 1987; **104**: 671-698.

[21] Jovanovic B. Selection and the evolution of industry. *Econometrica* 1982; **50**: 649-670.

[22] Davison AC, Hinkley DV. Bootstrap Methods and Their Application. Cambridge University Press: Cambridge, 1997.