# Variable selection in proportional hazards cure model with time-varying covariates, application to US bank failures

Alessandro Beretta[a] and Cédric Heuchenne[a,b]

[a]HEC-Liège, Management School of the University of Liège, Belgium; [b]ISBA, Université catholique de Louvain, Belgium

**ARTICLE HISTORY**
Compiled October 1, 2018

**ABSTRACT**
From a survival analysis perspective, bank failure data are often characterized by small default rates and heavy censoring. This empirical evidence can be explained by the existence of a subpopulation of banks likely immune from bankruptcy. In this regard, we use a mixture cure model to separate the factors with an influence on the susceptibility to default from the ones affecting the survival time of susceptible banks. In this paper, we extend a semi-parametric proportional hazards cure model to time-varying covariates and we propose a variable selection technique based on its penalized likelihood. By means of a simulation study, we show how this technique performs reasonably well. Finally, we illustrate an application to commercial bank failures in the United States over the period 2006-2016.

**KEYWORDS**
Bank failures; Survival Analysis; Mixture Cure Model; Time-varying covariates; Penalized likelihood; SCAD

## 1. Introduction

As a consequence of failures, acquisitions and mergers, the number of commercial banks in the United States has shrunk by two third in the last three decades (see Figure 1). Most of the defaults took place during the 1980s, with the savings and loans crisis, and, more recently, with the global financial crisis originated in the subprime mortgage market at the end of 2007 (see Figure 2).

The literature concerning the analysis and prediction of bank failures started growing during the 1980s, just after the savings and loans crisis (for a review, see [10, 40]). In the former studies, single period classification models (e.g. discriminant analysis or probit/logit regressions) have been used to investigate the probability of default or to discriminate between healthy and troubled institutions, conditionally on banks-specific information (for a survey, see [9]). Several authors criticized such techniques for their parametric nature and because they do not directly include the time-to-failures in the modeling efforts. For these reasons, Lane *et al.* [27] and Whalen [43] were the first who introduced the Cox's proportional-hazards (PH) model [7] into the study of bank

Corresponding Author: Alessandro Beretta. Email: a.beretta@ulg.ac.be. Note: An oral presentation of this paper was given at the 10th International Conference of the ERCIM WG on Computational and Methodological Statistics, Senate House, University of London, UK, 16-18 December 2017

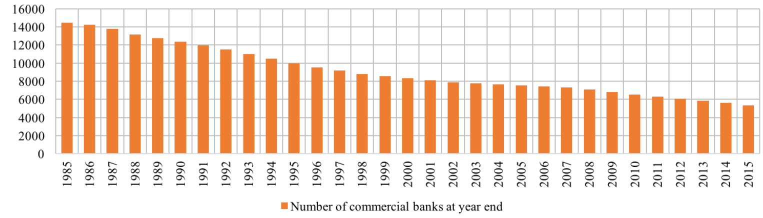


**Figure 1.** Evolution of the number of FDIC-insured commercial banks during the period 1985-2015 (source: FDIC, Table CB01 available at https://www5.fdic.gov/hsob/)

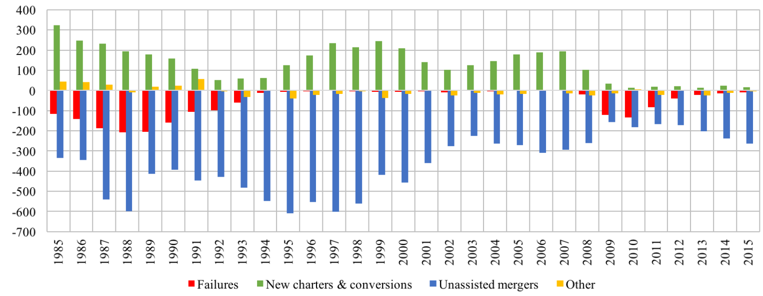


**Figure 2.** Changes in Number of Institutions FDIC-Insured Commercial Banks (source: FDIC, Table CB02 available at https://www5.fdic.gov/hsob/)

failures. Compared to single period classification models, its main advantage lies in the ability to directly model the survival time, rather than the mere default probability over a predetermined time horizon. In addition, as a semi-parametric model, it does not require any *a priori* distributional assumption for the baseline hazard function. In a similar way, Wheelock and Wilson [44] used a competing-risks PH model to identify the characteristics that make a bank more likely to fail or be acquired, with emphasis on management qualities.

Despite its advantages, the Cox's PH model relies on the assumption that the entire population will eventually experience the event of interest. However, bank failures data are characterized by a high censoring proportion and a quite low default rate (as we will see in Section 5). Hence, a more appropriate model choice would be the adoption of a *split-population survival-time model* (also known as *mixture cure model*). Originally developed in biostatistics to study long-term survivors of cancer in clinical trials, it allows to consider a portion of the population as cured (i.e. immune, non-susceptible to the event of interest) and to separate the factors influencing the probability to be susceptible (known in the literature as *incidence*) from the ones affecting the time-to-failure distribution of susceptible individuals (known as *latency*). In recent years, mixture cure models have also been investigated in the context of credit scoring [29, 42]. In general, they contain a binary classifier (e.g. logistic regression) and a survival component to model the incidence and latency distributions, respectively. Several parametric mixture cure models have been proposed in the literature [2, 5, 11, 18–20, 24]. But in order to avoid the specification of a distribution for the survival times, Kuk and Chen [26] introduced a semi-parametric alternative based on a proportional-hazards assumption. Later on, its estimation procedure has been improved by Sy and Taylor [39] and Peng and Dear [33], which proposed an EM algorithm for the joint estimation of the regression parameters and the baseline hazard function. The mixture cure model have been introduced into the literature on bank failures by Cole and Gunther [6]. They consider a fully parametric form of the survival time distribution (log-logistic) and time-independent covariates observed at the end of 1985 to predict survival times during the period 1986-1992.

The first contribution of our paper is the extension of the semi-parametric PH cure model of Sy and Taylor [39] to time-dependent covariates. With this model, we investigate commercial bank failures in the United States during the period 2006-2016 using quarterly bank-specific explanatory variables[1], as proxies for capital adequacy, asset quality, earnings, management efficiency and liquidity. As far as we know at the time of writing this article, in the banking-related literature, only two attempts have been made to incorporate time-varying covariates in a mixture cure model. DeLeonardis and Rocci [8] replaced the survival (latency) component with a discrete-time equivalent of the PH model to study Italian SMEs' credit risk. Dirick *et al.* [12] were the first to introduce macroeconomic time-dependent variables in the context of credit risk for portfolios of consumer loans. These are considered as "external covariates", according to the definition of Kalbfleisch and Prentice [25], and incorporated in a classical semi-parametric PH cure model as if they were time-independent, using their values at failure or censoring times. Among other research fields, the most recent attempt we are aware of to accommodate time-varying covariates in a cure model has been made in the medical literature by Shi and Yin [37]. Using a methodological paradigm for dynamic prediction known as landmark approach, they predict the survival probabilities at a given landmark time point given the information available at that time point only, discarding all the subjects who are no longer at risk and the rest of the full history of the covariates.

Our second contribution is the proposal of a variable selection technique for the semi-parametric PH cure model with time-varying covariates, which is based on a penalized version of its likelihood. In general, the simplest and most direct solution for variable selection would be the traditional best subset selection procedures. Given a set of $m$ covariates, they allow to search among the $2^m$ possible variable combinations the one that give the best value of a certain criterion, such as AIC [1], BIC [36] or Mallows' $C_p$ [31]. In particular, Dirick *et al.* [13] developed an Akaike information criterion for variable selection in the mixture cure model with multiple competing risks. However, these procedures suffer from significant drawbacks [14]: the lack of stability [3] and the high computational effort required (which increase exponentially with the number of variables). To overcome these issues, different variable selection techniques involving a penalization of the likelihood have been proposed in the statistical literature. They allow to simultaneously estimate the model's coefficients and select the most significant variables by shrinking the coefficient values towards zero. This kind of techniques varies for the estimation method and the type of penalty used. Among the most popular ones, the ridge [22], LASSO [41], adaptive LASSO [46] and SCAD [14, 15] penalties. In our study, we make use of the SCAD penalty, which is intuitively more appealing [14]. Nevertheless, due to the non-concavity and non-differentiability of this function, the maximization of a SCAD penalized log-likelihood is more challenging. In this regard, we use the MM algorithm proposed by Hunter and Li [23]. Unlike the Local Quadratic Approximation (LQA) algorithm of Fan and Li [14], it allows to have a penalized log-likelihood which becomes differentiable (adding a slight perturbation) and it never forces to eliminate permanently a covariate at any iteration of the selection procedure.

[1] Data are available through the FDIC Research Information System database, in the Statistics on Depository Institution (SDI) reports, which are based on the regulatory Call Reports filled by banks. Downloadable from the website https://www5.fdic.gov/sdi.

The rest of the article is organized as follows. In Section 2 and 3, we present the PH cure model with time varying covariates and the penalized likelihood technique for variable selection, respectively. A simulation study is provided in Section 4 to evaluate the performance of the proposed methodology. In Section 5, we describe the United States commercial banks dataset and present the results of our analysis. Finally, a conclusion is given in Section 6.

## 2. Proportional Hazards Cure model with time-varying covariates

Let $Y$ be the indicator that a bank is susceptible ($Y = 1$) or immune ($Y = 0$) to default, $T$ the time to failure (defined only when $Y = 1$), $S(t|Y = 1)$ the survival function for susceptible banks, and $p = P(Y = 1)$ the probability of being susceptible. Assuming that a fraction of the banks population is immune to bankruptcy, the marginal survival function of $T$ is defined as

$$S(t) = (1 - p) + pS(t|Y = 1). \tag{1}$$

The probability of being susceptible (i.e. incidence) is modeled by a logistic regression model

$$p(\mathbf{x}) = \mathbb{P}(Y = 1|\mathbf{x}) = \exp(\mathbf{x'b})/(1 + \exp(\mathbf{x'b})), \tag{2}$$

where $\mathbf{x} = (1, x_1, ..., x_q)'$ is a vector of time-fixed covariates (including the intercept) and $\mathbf{b}$ a vector of unknown coefficients. Whereas, the conditional survival function for susceptible banks (i.e. latency) is modeled by a [7] Proportional Hazards model with time-dependent covariates

$$h(t|Y = 1, \mathbf{z}(t)) = h_0(t|Y = 1)e^{\mathbf{z}'(t)\boldsymbol{\beta}},$$

where $\mathbf{z}(t) = (z_1(t), ..., z_p(t))'$ is a vector of time-varying covariates, $h_0(t|Y = 1)$ is an arbitrary conditional baseline hazard function and $\boldsymbol{\beta}$ is a vector of unknown coefficients. Then, model (1) can be rewritten as

$$S(t|\bar{\mathbf{z}}(t), \mathbf{x}) = 1 - p(\mathbf{x}) + p(\mathbf{x})e^{-\int_0^t h_0(u|Y=1)e^{\mathbf{z}'(u)\boldsymbol{\beta}}du}.$$

Denote the observed data by $(t_i, \delta_i, \bar{\mathbf{z}}_i(t_i), \mathbf{x}_i)$, for $i = 1, ..., n$, where $t_i$ is the observed failure or censoring time, $\delta_i$ is the censoring indicator taking value 1 if $t_i$ is uncensored and 0 otherwise, $\mathbf{x}_i$ is a vector of time-fixed covariates and $\bar{\mathbf{z}}_i(t_i)$ denotes a matrix of time varying covariates observed up to $t_i$, which are assumed to be constant within a predefined partition of the time scale forming $J$ intervals $(0, s_1], (s_1, s_2], ..., (s_{J-1}, s_J]$. The observed (incomplete) data likelihood can be written as

$$L(\mathbf{b}, \boldsymbol{\beta}, \mathbf{h}_0) = \prod_{i=1}^{n} \left\{ p_i h_0(t_i|Y = 1)e^{\mathbf{z}_i'(t_i)\boldsymbol{\beta}} e^{-\int_0^{t_i} h_0(u|Y=1)e^{\mathbf{z}_i'(u)\boldsymbol{\beta}}du} \right\}^{\delta_i} \times \left\{ (1 - p_i) + p_i e^{-\int_0^{t_i} h_0(u|Y=1)e^{\mathbf{z}_i'(u)\boldsymbol{\beta}}du} \right\}^{(1-\delta_i)},$$

where $p_i = \mathbb{P}(Y_i = 1|\mathbf{x}_i)$ and $\mathbf{h}_0$ is a vector containing the values of the baseline hazard function, which is assumed to be piecewise constant between failure times, $h_0(t|Y=1)=h_{0j}$, for $t \in (t_{(j-1)}, t_{(j)}]$, where $t_{(1)} \leq ... \leq t_{(k)}$ denotes the $k$ ordered failure times.

In order to obtain estimates for $\mathbf{b}$ and $\boldsymbol{\beta}$, as in the case of the PH cure model with fixed covariates [39], it is not possible to directly maximize $L(\mathbf{b}, \boldsymbol{\beta}, \mathbf{h}_0)$, nor to derive a partial likelihood (as in the standard Cox's PH model). But, as we know that $y_i = 1$ only when $\delta_i = 1$, the vector $\mathbf{y} = \{y_i : i = 1, ..., n\}$ is partially unobserved and we can treat the estimation as a missing data problem, using an Expectation-Maximization (EM) algorithm. The complete (unobserved) data likelihood can be defined as

$$L_C(\mathbf{b}, \boldsymbol{\beta}, \mathbf{h}_0; \mathbf{y}) = \underbrace{\prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{(1-y_i)}}_{L_1(\mathbf{b};\mathbf{y})} \times \underbrace{\prod_{i=1}^{n} \left[h_0(t_i|Y=1)e^{\mathbf{z}_i'(t_i)\boldsymbol{\beta}}\right]^{\delta_i y_i} \left[e^{-\int_0^{t_i} h_0(u|Y=1)e^{\mathbf{z}_i'(u)\boldsymbol{\beta}}\, du}\right]^{y_i}}_{L_2(\boldsymbol{\beta},\mathbf{h}_0;\mathbf{y})}. \quad (3)$$

In the rest of the article, we denote the complete data log-likelihood by $\ell_C(\mathbf{b}, \boldsymbol{\beta}, \mathbf{h}_0; \mathbf{y}) = \ell_1(\mathbf{b}; \mathbf{y}) + \ell_2(\boldsymbol{\beta}, \mathbf{h}_0; \mathbf{y})$, the sum of its logistic and PH components, respectively.

Given the observed data $\mathbf{O} = \{(t_i, \delta_i, \bar{\mathbf{z}}_i(t_i), \mathbf{x}_i); i = 1, ..., n\}$ and some starting values for the parameters $\boldsymbol{\theta} = (\mathbf{b}, \boldsymbol{\beta}, \mathbf{h}_0)$, the algorithm iterates between the E-step and the M-step until convergence. The E-step, takes the expectation of $\ell_C(\mathbf{b}, \boldsymbol{\beta}, \mathbf{h}_0; \mathbf{y})$ with respect to the conditional distribution of the unobserved elements in $\mathbf{y}$, given the current parameter values $\boldsymbol{\theta}^{(m)}$ and the observed data $\mathbf{O}$. Since $y_i$ is a linear term in $\ell_C(\mathbf{b}, \boldsymbol{\beta}, \mathbf{h}_0; \mathbf{y})$, for all censored $i$, we only need to compute

$$\pi_i^{(m)} = E[Y_i = 1|\boldsymbol{\theta}^{(m)}, \mathbf{O}] = \left.\frac{p_i e^{-\int_0^{t_i} h_0(u|Y=1)e^{\mathbf{z}_i'(u)\boldsymbol{\beta}} du}}{(1-p_i) + p_i e^{-\int_0^{t_i} h_0(u|Y=1)e^{\mathbf{z}_i'(u)\boldsymbol{\beta}} du}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(m)}}, \quad (4)$$

which represents the probability to be allocated in the susceptible group. Therefore, at the $m$-th iteration, we replace $\mathbf{y}$ in (3) with $\mathbf{y}^{*(m)} = \{\delta_i + (1-\delta_i)\pi_i^{(m)} : i = 1, ..., n\}$.

The M-step involves the maximization of the expected complete data log-likelihood $\ell_C\left(\mathbf{b}, \boldsymbol{\beta}, \mathbf{h}_0; \mathbf{y}^{*(m)}\right)$ with respect to $\mathbf{b}$, $\boldsymbol{\beta}$ and $\mathbf{h}_0$, which can be achieved by the simultaneous maximization of its logistic and PH components. In particular, regarding the maximization of $\ell_2\left(\boldsymbol{\beta}, \mathbf{h}_0; \mathbf{y}^{*(m)}\right)$, we adopt a profile likelihood technique similar to the one used by [4] in the case of a standard Cox's PH model. It can be shown that the MLEs of the baseline hazard function at failure times given $\boldsymbol{\beta}$ can be derived as

$$h_{0_j} = \frac{d_j}{\left(t_{(j)} - t_{(j-1)}\right) \sum_{i \in R_j} y_i e^{\mathbf{z}_i'(t_{(j)})\boldsymbol{\beta}}}, \quad (5)$$

where $d_j$ is the number of failures occuring at $t_{(j)}$ and $R_j$ is the risk set at $t_{(j)}^-$. Then, substituting (5) into $L_2(\boldsymbol{\beta}, \mathbf{h}_0; \mathbf{y}^{*(m)})$ gives the partial likelihood:

$$Ł_3(\boldsymbol{\beta}; \mathbf{y}^{*(m)}) = \prod_{j=1}^{k} \frac{e^{\mathbf{s}'(t_{(j)})\boldsymbol{\beta}}}{\left\{\sum_{i \in R_j} y_i^{*(m)} e^{\mathbf{z}_i'(t_{(j)})\boldsymbol{\beta}}\right\}^{d_i}}, \quad (6)$$

where $\mathbf{s}(t_{(j)}) = \sum_{i \in D_j} \mathbf{z}_i(t_{(j)})$ is the sum of the covariate vectors of all banks belonging to $D_j$, the set of tied failures occuring at $t_{(j)}$. In (5) and (6), we used the Peto-Breslow approximation for tied failure times [25].

Summarizing, at the $m$-th iteration, in the E-step, we first compute $\mathbf{y}^{*(m)}$ given the current parameter estimates $\boldsymbol{\theta}^{(m)}$. Afterwards, in the M-step, we simultaneously maximize $L_1\left(\mathbf{b}; \mathbf{y}^{*(m)}\right)$ and $L_3\left(\boldsymbol{\beta}; \mathbf{y}^{*(m)}\right)$ to obtain $\hat{\mathbf{b}}^{(m+1)}$ and $\hat{\boldsymbol{\beta}}^{(m+1)}$. The estimates of the baseline hazard function $\hat{\mathbf{h}}_0^{(m+1)}$ are then computed replacing $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}^{(m+1)}$ and $\mathbf{y}$ by $\mathbf{y}^{*(m)}$ into (5). Finally, the parameter estimates for the next iteration are given by $\boldsymbol{\theta}^{(m+1)} = \left(\hat{\mathbf{b}}^{(m+1)}, \hat{\boldsymbol{\beta}}^{(m+1)}, \hat{\mathbf{h}}_0^{(m+1)}\right)$ and the algorithm terminates when the norms of $\left(\hat{\mathbf{b}}^{(m+1)} - \hat{\mathbf{b}}^{(m)}\right)$ and $\left(\hat{\boldsymbol{\beta}}^{(m+1)} - \hat{\boldsymbol{\beta}}^{(m)}\right)$ are lower than a given tolerance threshold.

In order to ensure model identifiability and avoid numerical instabilities in the estimation of $\mathbf{b}$ and $\boldsymbol{\beta}$, it is important for the baseline survival function estimates $\hat{S}_0(t|Y = 1)$ to approach zero as $t$ approaches $t_{(k)}$, the last failure time. However, using the Breslow-type estimator in (5), this condition may not always be satisfied. As in Sy and Taylor [39], in order to meet this constraint, in the E-step, we impose $\pi_i^{(m)} = 0$, for all $t_i > t_{(k)}$ and all $m$. Thus, in our application, all the banks with a censoring time greater than the last failure time are treated as non-susceptible to default. This can be justified by the substantial proportion of banks surviving until the end of the study (2017) and by the sufficient follow-up period after the global financial crisis (2008-2012), where most of the defaults occur.

### *2.1. Inference*

In general, the iterative procedure of the EM algorithm itself does not directly provide an estimate of the covariance matrix of the MLEs. For this purpose, several methods based on the observed information matrix have been proposed [32].

In our study, we use a methodology similar to the one used to obtain approximate standard errors for the Cox's PH Cure model with fixed covariates [38]. First, we replace $\mathbf{h}_0$ in the incomplete-data log-likelihood by its estimator:

$$\tilde{\mathbf{h}}_0(\boldsymbol{\beta}; \mathbf{y}^*) = \left\{ \frac{d_j}{\left(t_{(j)} - t_{(j-1)}\right) \sum_{i \in R_j} y_i^* e^{\mathbf{z}_i'(t_{(j)})\boldsymbol{\beta}}} : j = 1, ..., k \right\}.$$

This is a profile likelihood technique used to reduce the dimension of the observed information matrix, which may become too large to be inverted when there are too many nuisance parameters. Second, in order to facilitate the calculation of the second-order derivatives, we use the method of Louis [30] according to which the observed information matrix can be computed in terms of the Hessian matrix of the complete-data log-likelihood. Hence, recalling that $\mathbf{y}^*$, at its turn, depends on $(\mathbf{b}, \boldsymbol{\beta}, \mathbf{h}_0)$, the first-order derivatives can be obtained as

$$\frac{\mathrm{d}\ell\left[\mathbf{b}, \boldsymbol{\beta}, \tilde{\mathbf{h}}_0\left(\boldsymbol{\beta}, \mathbf{y}^*(\mathbf{b}, \boldsymbol{\beta}, \hat{\mathbf{h}}_0)\right)\right]}{\mathrm{d}(\mathbf{b}, \boldsymbol{\beta})} = \frac{\mathrm{d}\ell_C\left[\mathbf{b}, \boldsymbol{\beta}, \tilde{\mathbf{h}}_0\left(\boldsymbol{\beta}, \mathbf{y}^*(\mathbf{b}, \boldsymbol{\beta}, \hat{\mathbf{h}}_0)\right); \mathbf{y}^*(\mathbf{b}, \boldsymbol{\beta}, \hat{\mathbf{h}}_0)\right]}{\mathrm{d}(\mathbf{b}, \boldsymbol{\beta})}$$

Note that here we made an additional approximation replacing $\mathbf{h}_0$ in $\mathbf{y}^*$ by its final estimated values $\hat{\mathbf{h}}_0$ because otherwise $\tilde{\mathbf{h}}_0$ would still depend on $(\mathbf{b}, \boldsymbol{\beta})$ through $\mathbf{y}^*$. The derivation of all the components of the Hessian matrix is then straightforward.

In table 1 and 2, as a result of a small simulation study, we provide the coverage probabilities for the approximate 95% confidence intervals of $\hat{\mathbf{b}}$ and $\hat{\boldsymbol{\beta}}$, respectively, under different levels of censoring and cure rate. As expected, the best coverage probabilities are obtained for the lowest censoring proportions. But, in general, they tend to the optimal value as the sample size increase. For a fixed level of censoring and an increase (resp. decrease) in the cure proportion, we observe an an increase (resp. decrease) in the quality of the coverage probabilities in the incidence component only. The latency component is not affected, which seems to be intuitive as the the amount of uncensored information remain the same.

| | | | | | **Logistic regression coefficients** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Cens** | **%** | **Cure** | **%** | **SS** | $b_0$ | 1.5 | 0 | -0.75 | 0 | -1.5 | 0 | 0.75 | 0 |
| L | 40% | H | 30% | 100 | 95.4% | 94.2% | 93.8% | 95.5% | 95.4% | 93.7% | 94.7% | 95.3% | 95.3% |
| | | | | 250 | 94.1% | 94.8% | 95.7% | 93.8% | 93.9% | 94.3% | 95.3% | 96.2% | 94.9% |
| | | | | 500 | 93.0% | 92.5% | 94.2% | 95.1% | 94.8% | 94.3% | 95.0% | 94.3% | 94.7% |
| | | | | 1000 | 93.7% | 95.2% | 95.1% | 94.8% | 94.1% | 94.4% | 95.3% | 94.4% | 95.2% |
| L | 40% | M | 20% | 100 | 96.4% | 96.8% | 94.9% | 94.9% | 95.3% | 96.3% | 95.6% | 95.8% | 94.9% |
| | | | | 250 | 93.0% | 92.4% | 93.1% | 93.2% | 93.6% | 94.2% | 93.8% | 94.6% | 94.5% |
| | | | | 500 | 93.3% | 94.8% | 93.9% | 93.5% | 94.7% | 95.0% | 93.4% | 93.8% | 94.3% |
| | | | | 1000 | 93.8% | 94.6% | 93.9% | 95.1% | 95.1% | 94.0% | 95.4% | 96.2% | 95.8% |
| M | 75% | H | 56% | 100 | 88.8% | 94.9% | 92.9% | 94.7% | 94.6% | 93.9% | 93.9% | 93.7% | 94.6% |
| | | | | 250 | 90.0% | 93.6% | 95.1% | 94.9% | 94.6% | 93.7% | 94.3% | 95.0% | 94.1% |
| | | | | 500 | 92.9% | 94.1% | 95.3% | 94.1% | 94.3% | 95.0% | 94.8% | 93.7% | 95.3% |
| | | | | 1000 | 93.7% | 95.2% | 94.1% | 94.2% | 94.8% | 94.9% | 95.4% | 95.0% | 94.6% |
| M | 75% | M | 38% | 100 | 74.5% | 88.8% | 93.5% | 90.9% | 93.8% | 88.9% | 94.1% | 91.8% | 94.3% |
| | | | | 250 | 81.0% | 90.1% | 94.1% | 93.5% | 94.6% | 91.3% | 93.2% | 92.5% | 93.8% |
| | | | | 500 | 83.5% | 90.4% | 94.6% | 93.1% | 94.4% | 92.5% | 94.4% | 92.7% | 94.2% |
| | | | | 1000 | 87.3% | 91.8% | 93.8% | 94.3% | 95.0% | 92.1% | 94.5% | 92.3% | 93.3% |

**Table 1.** Coverage probabilities for the approximate 95% confidence intervals for the logistic regression coefficients (different simulations settings Low/Moderate censoring and Low/Moderate cure rate).

| | | | | | **Cox's PH regression coefficients** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Cens** | **%** | **Cure** | **%** | **SS** | -0.7 | 0 | 1 | 0 | -0.5 | 0.75 | 0 | 0 |
| L | 40% | H | 30% | 100 | 86.7% | 92.4% | 84.9% | 92.5% | 90.8% | 88.5% | 91.3% | 90.8% |
| | | | | 250 | 89.5% | 93.9% | 90.7% | 94.3% | 94.5% | 90.4% | 93.8% | 93.4% |
| | | | | 500 | 90.0% | 96.6% | 90.4% | 93.2% | 93.7% | 92.2% | 94.7% | 94.8% |
| | | | | 1000 | 91.5% | 94.7% | 89.5% | 94.5% | 94.6% | 92.3% | 94.1% | 96.3% |
| L | 40% | M | 20% | 100 | 85.5% | 92.1% | 85.5% | 90.7% | 88.4% | 88.2% | 90.3% | 90.4% |
| | | | | 250 | 89.1% | 94.1% | 90.1% | 93.0% | 92.1% | 91.8% | 93.9% | 94.1% |
| | | | | 500 | 90.2% | 95.8% | 90.5% | 93.7% | 93.4% | 91.4% | 94.0% | 94.3% |
| | | | | 1000 | 89.3% | 95.1% | 90.7% | 95.1% | 93.5% | 92.3% | 93.9% | 95.2% |
| M | 75% | H | 56% | 100 | 74.8% | 85.6% | 77.8% | 85.0% | 83.7% | 80.9% | 86.3% | 84.2% |
| | | | | 250 | 86.9% | 91.2% | 87.9% | 92.4% | 92.3% | 89.1% | 91.3% | 92.2% |
| | | | | 500 | 91.7% | 92.5% | 88.9% | 93.4% | 94.1% | 92.0% | 93.9% | 93.8% |
| | | | | 1000 | 91.1% | 94.0% | 90.8% | 93.1% | 92.8% | 93.5% | 94.7% | 93.1% |
| M | 75% | M | 38% | 100 | 76.9% | 85.9% | 77.5% | 84.6% | 82.6% | 83.5% | 83.7% | 84.5% |
| | | | | 250 | 88.4% | 93.6% | 87.4% | 93.2% | 91.3% | 91.5% | 92.3% | 92.1% |
| | | | | 500 | 93.0% | 95.0% | 92.6% | 94.6% | 93.7% | 92.7% | 92.6% | 94.0% |
| | | | | 1000 | 92.4% | 95.2% | 91.6% | 95.3% | 94.1% | 93.5% | 93.9% | 93.3% |

**Table 2.** Coverage probabilities for the approximate 95% confidence intervals for the Cox's PH regression coefficients (different simulations settings Low/Moderate censoring and Low/Moderate cure rate).

## 3. Penalized PH Cure model with time-varying covariates

In this section, we propose a penalized maximum likelihood technique for variable selection in the Cox's PH Cure model with time varying covariates. It is based on the Smoothly Clipped Absolute Deviation (SCAD) penalty introduced by Fan and Li [14], which is defined by

$$p_\lambda(|\beta_j|) = \begin{cases} \lambda|\beta_j|, & \text{if } |\beta_j| \leq \lambda \\ \frac{(a^2-1)\lambda^2-(|\beta_j|-a\lambda)^2}{2(a-1)}, & \text{if } \lambda < |\beta_j| \leq a\lambda \\ \frac{(a+1)\lambda^2}{2}, & \text{if } |\beta_j| > a\lambda \end{cases},$$

for some $a > 2$ and $\lambda > 0$. These unknown parameters $(a, \lambda)$ are usually selected by means of a data-driven method, such as cross-validation or generalized cross-validation over a two-dimensional grid. However, in the context of linear regression and generalized linear models, Fan and Li [14] suggest that setting $a = 3.7$ appears to be a reasonable choice in order to reduce the potentially high computational cost. This solution is supported by a simulation study which indicates that $a$ selected by cross-validation do not improve the performance of the variable selection method. In the following section, we will show how this still holds true for our methodology and in the reminder of this article we will consider $a = 3.7$ (except where otherwise stated).

The properties of SCAD penalized estimators have been extensively studied by Fan and Li [14, 15] under different frameworks: linear regression, robust linear regression, generalized linear model, Cox's PH model and frailty model. They suggest that compared to other penalties (e.g. LASSO) the SCAD estimators simultaneously possess three properties: unbiasedness for large true coefficients, sparsity and continuity. In addition, with a proper choice of the tuning parameter, they show that SCAD possess the oracle property (i.e. true zero coefficients are automatically estimated as zero).

In our context, the penalized complete data log-likelihood is obtained adding two penalty terms to (3):

$$\begin{aligned} \ell_C^P(\mathbf{b}, \boldsymbol{\beta}, \mathbf{h}_0; \mathbf{y}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = & \underbrace{\ell_1(\mathbf{b}; \mathbf{y}) - n\sum_{j=2}^{q+1} p_{\lambda_{1j}}(|b_j|)}_{\ell_1^P(\mathbf{b};\mathbf{y},\boldsymbol{\lambda}_1)} \\ & + \underbrace{\ell_3(\boldsymbol{\beta}; \mathbf{y}) - n\sum_{l=1}^{p} p_{\lambda_{2l}}(|\beta_l|)}_{\ell_3^P(\boldsymbol{\beta};\mathbf{y},\boldsymbol{\lambda}_2)}. \end{aligned}$$

Note that in the logistic regression model we do not penalize the intercept and the tuning parameters (one for each regression coefficient) can be chosen according to the procedure presented afterwards in Section 3.1.

For fixed values of the tuning parameters in $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$, the penalized MLEs of $\mathbf{b}$, $\boldsymbol{\beta}$ and $\mathbf{h}_0$ are obtained by means of an EM algorithm similar to the one in the unpenalized model. The E-step does not change, because the unobserved elements of $\mathbf{y}$ are not included in the penalty terms. The only difference lies in the M-step, which requires, at the $m$-th iteration, the maximisation of $\ell_C^P\left(\mathbf{b}, \boldsymbol{\beta}, \mathbf{h}_0; \mathbf{y}^{*(m)}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2\right)$, which is achieved by separately maximizing $\ell_1^P\left(\mathbf{b}; \mathbf{y}^{*(m)}, \boldsymbol{\lambda}_1\right)$ and $\ell_3^P\left(\boldsymbol{\beta}; \mathbf{y}^{*(m)}, \boldsymbol{\lambda}_2\right)$.

Due to the introduction of the SCAD penalties, the maximization of these likelihood functions is not a trivial task, because they are non-concave and non-differentiable at the origin. Fan and Li [14] proposed a new algorithm based on the Newton-Raphson method, but with the gradient computed using a local quadratic approximation (LQA) of the penalty functions. However, this algorithm has a major limitation: at each iteration, whenever a coefficient is sufficiently close to zero, the associated covariate is removed permanently from the model. In order to address this issue, caused by the non-differentiability at the origin of the LQA, Hunter and Li [23] introduced a perturbation. Then, they use a minorize-maximize (MM) algorithm to maximize this new perturbed likelihood function, which is now differentiable. This is the technique that we use in this paper to obtain estimates for $\mathbf{b}$ and $\boldsymbol{\beta}$ at each iteration of the EM algorithm.

**Remark**. Despite different authors [14, 17, 23] provide a sandwich formula to compute the standard errors of the coefficient estimates, they do not present any method to conduct inference (e.g. construction of the confidence intervals). According to Leeb and Pötscher [28], the finite sample distribution of the SCAD penalized estimator is typically highly non-normal and its large sample behavior depends on the choice of the tuning parameters. For this reason, confidence intervals constructed on the basis of standard asymptotic theory may not have good coverage. That is why, in this article, we rely on inference from the unpenalized model, which is refitted with the selected covariates from the penalized technique, as in Pfeiffer *et al.* [34].

### *3.1. Selection of the tuning parameters*

The selection of the tuning parameters ($\boldsymbol{\lambda}_1$,$\boldsymbol{\lambda}_2$) is based on the Generalized Information Criterion (GIC) for non-concave penalized likelihood functions, proposed by Zhang *et al.* [45]. In the PH cure model context, we can rewrite it as

$$\mathrm{GIC}\left(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2\right) = \frac{1}{n}\left[-2\ell\left(\hat{\boldsymbol{\theta}}\right) + \kappa_n e(\hat{\boldsymbol{\theta}}; \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)\right],$$

where $\ell\left(\hat{\boldsymbol{\theta}}\right)$ is the observed (incomplete) data log-likelihood evaluated at the penalized MLEs $\hat{\boldsymbol{\theta}} = (\hat{\mathbf{b}}', \hat{\boldsymbol{\beta}}')'$, $\kappa_n$ a positive number controlling the properties of variable selection and $e(\hat{\boldsymbol{\theta}}; \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ the effective number of parameters. It is straightforward to see how GIC becomes AIC, resp. BIC, when $\kappa_n = 2$, resp. $\kappa_n = \ln(n)$. In the framework of generalized linear models, Zhang *et al.* [45] analysed the theoretical properties of GIC. They found that the BIC-type selector is consistent when a set of candidate model contains the true model and the AIC-type selector is asymptotically loss efficient, if the true model is approximated by a family of candidate models. In practice, the choice between these two criteria depends on how much the investigator wants to penalize model complexity. A larger value of $\kappa_n$ increases the chances of selecting a more parsimonious model.

The effective number of parameters is computed as

$$e(\hat{\boldsymbol{\theta}}; \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \mathrm{tr}\left\{\left[\nabla^2\ell\left(\hat{\boldsymbol{\theta}}\right) - n\boldsymbol{\Sigma}_{\lambda,\varepsilon}\left(\hat{\boldsymbol{\theta}}\right)\right]^{-1}\nabla^2\ell\left(\hat{\boldsymbol{\theta}}\right)\right\},$$

where $\nabla^2\ell\left(\hat{\boldsymbol{\theta}}\right)$ is the Hessian matrix of the observed (incomplete) data log-likelihood with respect to $\mathbf{b}$ and $\boldsymbol{\beta}$, evaluated at the penalized MLEs, and

$$\boldsymbol{\Sigma}_{\lambda,\varepsilon}\left(\hat{\boldsymbol{\theta}}\right) = \text{diag}\left[0, \frac{p'_{\lambda_1}(|b_2|)}{\varepsilon + |b_2|}, ..., \frac{p'_{\lambda_1}(|b_{q+1}|)}{\varepsilon + |b_{q+1}|}, \frac{p'_{\lambda_2}(|\beta_1|)}{\varepsilon + |\beta_1|}, ..., \frac{p'_{\lambda_2}(|\beta_p|)}{\varepsilon + |\beta_p|}\right].$$

As the number of parameters $(b + q)$ increases, the computational cost to find the value of $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ which minimise the GIC increase exponentially. For this reason, as in Fan and Li [16], considering that the magnitude of the tuning parameters should be proportional to the variability of the coefficient estimates in the standard model with no penalisation, we set $\lambda_{1j} = \lambda_1 \text{SE}(\hat{b}_j^U)$, for $j = 2, .., q + 1$, and $\lambda_{2k} = \lambda_2 \text{SE}(\hat{\beta}_k^U)$, for $k = 1, .., p$, where $\text{SE}(\hat{b}_j^U)$ and $\text{SE}(\hat{\beta}_k^U)$ corresponds to the standard errors of the MLEs obtained from the standard model. This allows to significantly reduce the dimensions of the tuning parameter space to a 2-dimensional space and, hence, to considerably reduce the computing time.

## 4. Simulation study

In this section, a comprehensive simulation study[2] is presented to assess the finite sample performance of the PH Cure model with time-varying covariates and its variable selection procedure. Performance is measured either in terms of model errors and model's ability to select the variables with true non-zero coefficients. Different scenarios are considered according to different censoring levels, cure rates and sample sizes. Furthermore, we investigate the sparsity of the regression coefficients vectors $\mathbf{b}$ and $\boldsymbol{\beta}$, as well as the magnitude of their elements.

### *4.1. Simulation design*

In order to simulate data for this study, we generated: (i) *failure times* from a Cox's PH model with a constant unit baseline hazard function and time varying covariates, using the method of Hendry [21], (ii) *censoring times* from an exponential distribution with parameter $\lambda_C$ and truncated above 5, and (iii) *cure indicators* from a logistic regression model.

The survival component depend on 8 time-varying covariates, which are piecewise constant over $J = 50$ equally spaced intervals, between 0 and 5, and follow a VAR model $\mathbf{z}_{i,j} = \boldsymbol{\Phi}\mathbf{z}_{i,j-1} + \mathbf{e}_j$, where $\mathbf{e}_j \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and $\Phi_{p,q} = \frac{2^{-q}}{p}$, for $p, q = 1, ..., 8$. Whereas, the logistic regression component depends on 8 time-fixed covariates following a multivariate normal distribution $\mathbf{x}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ with zero means and covariance matrix $\Sigma_{p,q} = 0.5^{|p-q|}$, for $p, q = 1, ..., 8$.

At first, we evaluate the effects of varying sample size, censoring proportion and cure rate. The regression coefficients vectors are set equal to $\boldsymbol{\beta}_0 = (-0.7, 0, 1, 0, -0.5, 0.75, 0, 0)'$ and $\mathbf{b}_0 = (b_0, 1.5, 0, -0.75, 0, -1.5, 0, 0.75, 0)'$. Then, we use different parametrizations of $b_0$ and $\lambda_C$ to obtain the different scenarios depicted in Table 3.

In a second set of simulations, we study the impact of sparsity and magnitude of the coefficients on the performance of the variable selection technique.

[2]Conducted with a code developed by the authors, using C++ and the Armadillo library of Sanderson and Curtin [35]. Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11.

| Censoring | | Cure | | $\lambda_C$ | $b_0$ |
|---|---|---|---|---|---|
| Low | 40% | High | 30% | 0.13 | 1.45 |
| Low | 40% | Medium | 20% | 0.55 | 2.35 |
| Medium | 75% | High | 56.25% | 1.33 | -0.44 |
| Medium | 75% | Medium | 37.5% | 2.7 | 0.89 |
| High | 95% | High | 71.25% | 8.9 | -1.56 |

**Table 3.** Simulation set I. Parametrizations of the different scenarios. The percentages refer to the total number of individuals in the sample.

Due to computational reasons, we focus only on an intermediate case (medium censoring and high cure rate) and a sample size of 500. We use the same data generating process with 8 covariates in both model components, but we consider two levels of sparsity: (i) *high*, when $\boldsymbol{\beta}_0 = (\beta_1, -\beta_1, 0, 0, 0, 0, 0, 0)'$ and/or $\mathbf{b}_0 = (b_0, b_1, -b_1, 0, 0, 0, 0, 0, 0)'$, and (ii) *low*, when $\boldsymbol{\beta}_0 = (\beta_1, -\beta_1, \beta_1, -\beta_1, \beta_1, -\beta_1, 0, 0)'$ and/or $\mathbf{b}_0 = (b_0, b_1, -b_1, b_1, -b_1, b_1, -b_1, 0, 0)'$. In addition, for each of these combinations, we consider two different levels of coefficients' magnitude: (i) *high*, when $b_1 = 1.5$ and/or $\beta_1 = 1.5$, and (ii) *low*, when $b_1 = 0.5$ and/or $\beta_1 = 0.5$. In Table 4, we depict the different parametrizations of $b_0$ and $\lambda_C$ to obtain the different scenarios.

| Sparsity $\mathbf{b}_0$ | Sparsity $\boldsymbol{\beta}_0$ | Magnitude | $b_0$ | $\lambda_C$ |
|---|---|---|---|---|
| High | High | Low | -0.27 | 0.83 |
| High | High | High | -0.36 | 1.59 |
| Low | Low | Low | -0.29 | 0.97 |
| Low | Low | High | -0.46 | 2.35 |
| High | Low | Low | -0.27 | 0.98 |
| High | Low | High | -0.36 | 2.33 |
| Low | High | Low | -0.29 | 0.82 |
| Low | High | High | -0.46 | 1.60 |

**Table 4.** Simulation set II. Parametrizations of the different scenarios.

For each scenario, after generating 1000 datasets, we computed the model errors of the standard cure PH model (Section 2) with all the covariates (i.e. Full model) and with the covariates having true non-zero coefficients only (i.e. Oracle model). These errors are compared with the ones obtained from the penalized method with SCAD penalties (Section 3). In this latter case, the tuning parameters $\lambda_1$ and $\lambda_2$ are optimized according to a BIC criterion, as we know in advance that the true model is among the tested ones, and a grid-search over the values $\{0.2, 0.4, ..., 2\}$. Additionally, we investigate the choice of $a = 3.7$ in the SCAD penalty function, making a comparison with the case where the tuning parameters $a_1$ and $a_2$ are selected from the grid $\{2.1, 3.0, 3.7, 4.5, 5.5\}$.

### *4.2. Simulation results*

The results of the two sets of experiments are summarized in Table 5 and 6, where we provide the Median of the Model Errors (MME), the Median of the Relative Model Errors (MRME), as well as the average number of correct and incorrect zeros identified by the penalized approach. In this regard, the model errors for the logistic regression and Cox's PH components are respectively defined as

$$E\left[\left((1+e^{-\mathbf{x}'\hat{\mathbf{b}}})^{-1} - (1+e^{-\mathbf{x}'\mathbf{b}_0})^{-1}\right)^2\right]$$

and

$$E\left[\left(S(T|\bar{\mathbf{z}}(T),\hat{\boldsymbol{\beta}}) - S(T|\bar{\mathbf{z}}(T),\boldsymbol{\beta}_0)\right)^2\right].$$

In Table 5, for all considered models, we can observe that the model errors decrease with larger sample sizes or lower levels of censoring. Although this is true for both model components, for a given level of censoring, a decrease in the cure proportion has a negative impact on the errors of the logistic regression component only. If we look at the average number of correct and incorrect zeros, we can see that the penalized method with SCAD penalties perform quite well in identifying the variables with true non-zero coefficients and removing the incorrect ones. As we noticed for the model errors, an increase in the sample size or a decrease in the censoring percentage has a positive effect, increasing (resp. decreasing) the number of correct (resp. incorrect) zeros. Whereas, for a given level of censoring, the lower the cure proportion, the lower (resp. higher) the number of correct (resp. incorrect) zeros in the logistic regression component. Again, we do not observe any effect on the Cox's PH coefficients. Looking at the Median of the Relative Model Errors (MRME), as one would expect, among the three estimated models (Full, SCAD and Oracle), the errors of the Oracle model are always the lowest. The fact that the penalized model errors tend towards them as the sample size increase is a further evidence of the good performance of the proposed methodology.

These conclusions are also supported by the outcome from the second set of simulations (see Table 6). The scenarios with a higher coefficients' magnitude and higher levels of sparsity show better performance. The worst results are obtained in situations of low sparsity and low magnitude, especially in the logistic regression component. However, after increasing the sample size, we observed much better results.

Finally, comparing the results obtained from the penalized method with the parameter $a$ selected by cross-validation with the ones with $a = 3.7$ (SCAD$^a$ and SCAD$^b$, respectively), it is possible to see how the selection of the parameter $a$ has little impact both on model errors and number of correct/incorrect zeros. Therefore, we can conclude that $a = 3.7$ seems to be a reasonable choice.

## 5. Application to US bank failures

### *5.1. Data and explanatory variables*

The data used in this study cover the period 2006-2016 and the 7492 commercial banks insured by the Federal Deposit Insurance Corporation (FDIC) during the first quarter of 2006. New insured banks after this date are not taken into account due to model design. We define the event of interest (i.e. bankruptcy) as a bank's closure by the FDIC. According to this definition, 419 (5.6%) institutions failed during the period under consideration and 7073 (94.4%) are right-censored, meaning that for these banks we do not know when bankruptcy will occur, given that it can occur. Among these non-failed institutions, 2314 (30.9%) are lost to follow-up due to mergers/acquisitions, while the remaining 4759 (63.5%) are still at risk at the end of the study period. This substantial portion of long-term survivors, is a first evidence of the existence of possibly non-susceptible banks and a justification for the adoption of a PH Cure model. As we shown in the simulation study (Section 4), the high level of censoring (94.4%) is less of a concern in this application thanks to the quite large sample size.

| | | | | Logistic regression | | | | Cox's model | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cens. | Cure | Size | Method | $MME$ | $MRME$ | C-0's | IC-0's | $MME$ | $MRME$ | C-0's | IC-0's |
| Low | High | 100 | Full | 0.0221 | 1 | - | - | 0.0102 | 1 | - | - |
| | | | SCAD$^a$ | 0.0234 | 1.06 | 3.488 | 1.098 | 0.0093 | 0.92 | 3.631 | 0.198 |
| | | | SCAD$^b$ | 0.0239 | 1.08 | 3.360 | 1.067 | 0.0094 | 0.92 | 3.656 | 0.200 |
| | | | Oracle | 0.0104 | 0.47 | 4 | 0 | 0.0083 | 0.81 | 4 | 0 |
| | | 250 | Full | 0.0077 | 1 | - | - | 0.0038 | 1 | - | - |
| | | | SCAD$^a$ | 0.0057 | 0.74 | 3.804 | 0.354 | 0.0034 | 0.87 | 3.769 | 0.004 |
| | | | SCAD$^b$ | 0.0061 | 0.79 | 3.713 | 0.331 | 0.0033 | 0.87 | 3.756 | 0.004 |
| | | | Oracle | 0.0040 | 0.51 | 4 | 0 | 0.0033 | 0.85 | 4 | 0 |
| | | 500 | Full | 0.0040 | 1 | - | - | 0.0020 | 1 | - | - |
| | | | SCAD$^a$ | 0.0021 | 0.53 | 3.927 | 0.033 | 0.0017 | 0.84 | 3.796 | 0.000 |
| | | | SCAD$^b$ | 0.0021 | 0.53 | 3.910 | 0.044 | 0.0017 | 0.83 | 3.771 | 0.000 |
| | | | Oracle | 0.0020 | 0.50 | 4 | 0 | 0.0016 | 0.82 | 4 | 0 |
| | | 1000 | Full | 0.0019 | 1 | 0 | 0 | 0.0010 | 1 | - | - |
| | | | SCAD$^b$ | 0.0010 | 0.52 | 3.954 | 0 | 0.0008 | 0.83 | 3.715 | 0 |
| | | | Oracle | 0.0010 | 0.50 | 4 | 0 | 0.0008 | 0.82 | 4 | 0 |
| Low | Medium | 100 | Full | 0.0310 | 1 | - | - | 0.0107 | 1 | - | - |
| | | | SCAD$^a$ | 0.0374 | 1.20 | 3.183 | 1.933 | 0.0099 | 0.93 | 3.644 | 0.219 |
| | | | SCAD$^b$ | 0.0422 | 1.36 | 3.146 | 2.085 | 0.0100 | 0.94 | 3.625 | 0.217 |
| | | | Oracle | 0.0127 | 0.41 | 4 | 0 | 0.0086 | 0.80 | 4 | 0 |
| | | 250 | Full | 0.0100 | 1 | - | - | 0.0043 | 1 | - | - |
| | | | SCAD$^a$ | 0.0119 | 1.18 | 3.740 | 0.833 | 0.0036 | 0.83 | 3.735 | 0.002 |
| | | | SCAD$^b$ | 0.0122 | 1.21 | 3.633 | 0.823 | 0.0036 | 0.83 | 3.727 | 0.002 |
| | | | Oracle | 0.0047 | 0.47 | 4 | 0 | 0.0034 | 0.80 | 4 | 0 |
| | | 500 | Full | 0.0049 | 1 | - | - | 0.0022 | 1 | - | - |
| | | | SCAD$^a$ | 0.0030 | 0.62 | 3.835 | 0.267 | 0.0018 | 0.83 | 3.785 | 0.000 |
| | | | SCAD$^b$ | 0.0032 | 0.65 | 3.808 | 0.279 | 0.0018 | 0.82 | 3.779 | 0.000 |
| | | | Oracle | 0.0024 | 0.48 | 4 | 0 | 0.0017 | 0.80 | 4 | 0 |
| | | 1000 | Full | 0.0024 | 1 | - | - | 0.0011 | 1 | - | - |
| | | | SCAD$^b$ | 0.0013 | 0.53 | 3.94 | 0.012 | 0.0009 | 0.82 | 3.766 | 0 |
| | | | Oracle | 0.0012 | 0.51 | 4 | 0 | 0.0009 | 0.81 | 4 | 0 |
| Medium | High | 100 | Full | 0.0480 | 1 | - | - | 0.0282 | 1 | - | - |
| | | | SCAD$^a$ | 0.0581 | 1.21 | 3.190 | 1.660 | 0.0279 | 0.99 | 3.700 | 1.344 |
| | | | SCAD$^b$ | 0.0622 | 1.29 | 3.054 | 1.802 | 0.0279 | 0.99 | 3.713 | 1.373 |
| | | | Oracle | 0.0221 | 0.46 | 4 | 0 | 0.0185 | 0.66 | 4 | 0 |
| | | 250 | Full | 0.0149 | 1 | - | - | 0.0104 | 1 | - | - |
| | | | SCAD$^a$ | 0.0170 | 1.14 | 3.748 | 0.740 | 0.0098 | 0.94 | 3.871 | 0.392 |
| | | | SCAD$^b$ | 0.0169 | 1.13 | 3.642 | 0.694 | 0.0098 | 0.94 | 3.867 | 0.413 |
| | | | Oracle | 0.0074 | 0.49 | 4 | 0 | 0.0080 | 0.77 | 4 | 0 |
| | | 500 | Full | 0.0073 | 1 | - | - | 0.0050 | 1 | - | - |
| | | | SCAD$^a$ | 0.0044 | 0.60 | 3.873 | 0.192 | 0.0040 | 0.82 | 3.908 | 0.050 |
| | | | SCAD$^b$ | 0.0046 | 0.64 | 3.829 | 0.210 | 0.0040 | 0.82 | 3.940 | 0.050 |
| | | | Oracle | 0.0036 | 0.50 | 4 | 0 | 0.0039 | 0.78 | 4 | 0 |
| | | 1000 | Full | 0.0033 | 1 | - | - | 0.0024 | 1 | - | - |
| | | | SCAD$^b$ | 0.0018 | 0.53 | 3.953 | 0.007 | 0.0020 | 0.82 | 3.96 | 0 |
| | | | Oracle | 0.0017 | 0.52 | 4 | 0 | 0.0019 | 0.80 | 4 | 0 |
| Medium | Medium | 100 | Full | 0.0961 | 1 | - | - | 0.0253 | 1 | - | - |
| | | | SCAD$^a$ | 0.1008 | 1.05 | 3.431 | 2.858 | 0.0245 | 0.97 | 3.738 | 1.313 |
| | | | SCAD$^b$ | 0.1058 | 1.10 | 3.529 | 3.129 | 0.0241 | 0.95 | 3.688 | 1.298 |
| | | | Oracle | 0.0472 | 0.49 | 4 | 0 | 0.0158 | 0.63 | 4 | 0 |
| | | 250 | Full | 0.0303 | 1 | - | - | 0.0094 | 1 | - | - |
| | | | SCAD$^a$ | 0.0348 | 1.15 | 3.517 | 1.475 | 0.0089 | 0.94 | 3.850 | 0.350 |
| | | | SCAD$^b$ | 0.0354 | 1.17 | 3.392 | 1.440 | 0.0089 | 0.94 | 3.850 | 0.375 |
| | | | Oracle | 0.0145 | 0.48 | 4 | 0 | 0.0069 | 0.74 | 4 | 0 |
| | | 500 | Full | 0.0138 | 1 | - | - | 0.0046 | 1 | - | - |
| | | | SCAD$^a$ | 0.0160 | 1.17 | 3.723 | 0.685 | 0.0038 | 0.82 | 3.902 | 0.035 |
| | | | SCAD$^b$ | 0.0163 | 1.18 | 3.663 | 0.715 | 0.0038 | 0.82 | 3.910 | 0.042 |
| | | | Oracle | 0.0070 | 0.51 | 4 | 0 | 0.0035 | 0.76 | 4 | 0 |
| | | 1000 | Full | 0.0068 | 1 | - | - | 0.0023 | 1 | - | - |
| | | | SCAD$^b$ | 0.0048 | 0.69 | 3.837 | 0.204 | 0.0018 | 0.80 | 3.955 | 0 |
| | | | Oracle | 0.0037 | 0.54 | 4 | 0 | 0.0018 | 0.78 | 4 | 0 |
| High | High | 1000 | Full | 0.0193 | 1 | - | - | 0.0124 | 1 | - | - |
| | | | SCAD$^b$ | 0.0245 | 1.27 | 3.741 | 1.097 | 0.0130 | 1.05 | 3.883 | 0.872 |
| | | | Oracle | 0.0127 | 0.66 | 4 | 0 | 0.0093 | 0.76 | 4 | 0 |
| | | 2000 | Full | 0.0097 | 1 | - | - | 0.0057 | 1 | - | - |
| | | | SCAD$^b$ | 0.0116 | 1.20 | 3.887 | 0.404 | 0.0052 | 0.91 | 3.957 | 0.203 |
| | | | Oracle | 0.0065 | 0.68 | 4 | 0 | 0.0042 | 0.74 | 4 | 0 |
| | | 4000 | Full | 0.0050 | 1 | - | - | 0.0027 | 1 | - | - |
| | | | SCAD$^b$ | 0.0034 | 0.69 | 3.964 | 0.037 | 0.0022 | 0.84 | 3.978 | 0.011 |
| | | | Oracle | 0.0033 | 0.66 | 4 | 0 | 0.0022 | 0.82 | 4 | 0 |

**Table 5.** Results of the simulations under different designs. *Note*: in SCAD$^a$ the value of $a$ is selected by cross-validation, while in SCAD$^b$ is set equal to 3.7.

| | | | | | Logistic regression | | | | Cox's model | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sparsity $\mathbf{b}_0$ | Sparsity $\boldsymbol{\beta}_0$ | Magnitude | Size | Method | $MME$ | $MRME$ | C-0's | IC-0's | $MME$ | $MRME$ | C-0's | IC-0's |
| High | High | High | 500 | Full | 0.0073 | 1 | 6 | 2 | 0.0058 | 1 | 6 | 2 |
| | | | | SCAD$^a$ | 0.0023 | 0.31 | 5.925 | 0.000 | 0.0041 | 0.71 | 5.892 | 0.000 |
| | | | | SCAD$^b$ | 0.0023 | 0.31 | 5.934 | 0.000 | 0.0041 | 0.71 | 5.937 | 0.000 |
| | | | | Oracle | 0.0021 | 0.29 | 6 | 0 | 0.0040 | 0.69 | 6 | 0 |
| High | High | Low | 500 | Full | 0.0093 | 1 | 6 | 2 | 0.0033 | 1 | 6 | 2 |
| | | | | SCAD$^a$ | 0.0095 | 1.02 | 5.920 | 0.770 | 0.0026 | 0.78 | 5.849 | 0.027 |
| | | | | SCAD$^b$ | 0.0093 | 1.00 | 5.925 | 0.787 | 0.0026 | 0.78 | 5.940 | 0.031 |
| | | | | Oracle | 0.0029 | 0.32 | 6 | 0 | 0.0025 | 0.76 | 6 | 0 |
| | | | 1000 | Full | 0.0044 | 1 | 6 | 2 | 0.0018 | 1 | 6 | 2 |
| | | | | SCAD$^b$ | 0.0015 | 0.35 | 5.955 | 0.129 | 0.0014 | 0.78 | 5.955 | 0.000 |
| | | | | Oracle | 0.0014 | 0.32 | 6 | 0 | 0.0013 | 0.74 | 6 | 0 |
| High | Low | High | 500 | Full | 0.0066 | 1 | 6 | 2 | 0.0085 | 1 | 2 | 6 |
| | | | | SCAD$^a$ | 0.0018 | 0.28 | 5.923 | 0.000 | 0.0079 | 0.93 | 1.957 | 0.000 |
| | | | | SCAD$^b$ | 0.0018 | 0.28 | 5.931 | 0.000 | 0.0079 | 0.93 | 1.968 | 0.000 |
| | | | | Oracle | 0.0017 | 0.26 | 6 | 0 | 0.0078 | 0.91 | 2 | 0 |
| High | Low | Low | 500 | Full | 0.0090 | 1 | 6 | 2 | 0.0041 | 1 | 2 | 6 |
| | | | | SCAD$^a$ | 0.0104 | 1.16 | 5.927 | 0.793 | 0.0046 | 1.12 | 1.966 | 0.549 |
| | | | | SCAD$^b$ | 0.0102 | 1.13 | 5.930 | 0.810 | 0.0043 | 1.05 | 1.973 | 0.458 |
| | | | | Oracle | 0.0029 | 0.32 | 6 | 0 | 0.0038 | 0.91 | 2 | 0 |
| Low | High | High | 500 | Full | 0.0067 | 1 | 2 | 6 | 0.0057 | 1 | 6 | 2 |
| | | | | SCAD$^a$ | 0.0051 | 0.75 | 1.962 | 0.000 | 0.0041 | 0.72 | 5.842 | 0.000 |
| | | | | SCAD$^b$ | 0.0051 | 0.75 | 1.962 | 0.000 | 0.0041 | 0.72 | 5.910 | 0.000 |
| | | | | Oracle | 0.0049 | 0.73 | 2 | 0 | 0.0039 | 0.69 | 6 | 0 |
| Low | High | Low | 500 | Full | 0.0089 | 1 | 2 | 6 | 0.0034 | 1 | 6 | 2 |
| | | | | SCAD$^a$ | 0.0240 | 2.69 | 1.960 | 3.535 | 0.0027 | 0.80 | 5.826 | 0.039 |
| | | | | SCAD$^b$ | 0.0247 | 2.76 | 1.971 | 3.558 | 0.0027 | 0.81 | 5.920 | 0.036 |
| | | | | Oracle | 0.0070 | 0.78 | 2 | 0 | 0.0025 | 0.74 | 6 | 0 |
| Low | Low | High | 500 | Full | 0.0063 | 1 | 2 | 6 | 0.0084 | 1 | 2 | 6 |
| | | | | SCAD$^a$ | 0.0048 | 0.76 | 1.969 | 0.000 | 0.0078 | 0.94 | 1.960 | 0.000 |
| | | | | SCAD$^b$ | 0.0048 | 0.76 | 1.970 | 0.000 | 0.0078 | 0.94 | 1.972 | 0.000 |
| | | | | Oracle | 0.0046 | 0.73 | 2 | 0 | 0.0078 | 0.93 | 2 | 0 |
| Low | Low | Low | 500 | Full | 0.0085 | 1 | 2 | 6 | 0.0040 | 1 | 2 | 6 |
| | | | | SCAD$^a$ | 0.0239 | 2.82 | 1.975 | 3.505 | 0.0044 | 1.10 | 1.948 | 0.464 |
| | | | | SCAD$^b$ | 0.0248 | 2.92 | 1.984 | 3.532 | 0.0043 | 1.07 | 1.974 | 0.402 |
| | | | | Oracle | 0.0064 | 0.75 | 2 | 0 | 0.0037 | 0.92 | 2 | 0 |
| | | | 1000 | Full | 0.0044 | 1 | 2 | 6 | 0.0020 | 1 | 2 | 6 |
| | | | | SCAD$^b$ | 0.0044 | 0.99 | 1.981 | 0.811 | 0.0019 | 0.92 | 1.971 | 0.002 |
| | | | | Oracle | 0.0035 | 0.78 | 2 | 0 | 0.0018 | 0.91 | 2 | 0 |
| | | | 2000 | Full | 0.0021 | 1 | 2 | 6 | 0.0010 | 1 | 2 | 6 |
| | | | | SCAD$^b$ | 0.0016 | 0.77 | 1.992 | 0.020 | 0.0010 | 0.93 | 1.987 | 0.000 |
| | | | | Oracle | 0.0016 | 0.77 | 2 | 0 | 0.0009 | 0.92 | 2 | 0 |

**Table 6.** Results of the simulations under different levels of sparsity and magnitude of the coefficient vectors. *Note*: in SCAD$^a$ the value of $a$ is selected by cross-validation, while in SCAD$^b$ is set equal to 3.7.

Following previous studies on bank failures, we selected a set of bank-specific explanatory variables covering the 5 components of the well known CAMEL rating system: capital adequacy, asset quality, earnings, management efficiency and liquidity. Quarterly financial data from regulatory Call Reports filed by banks are retrieved from the Statistics on Depository Institution (SDI) reports in the FDIC Research Information System database. With few exceptions (EFF, SIZE and AGE), all covariates are expressed as ratios to Total Assets. See Tables 7-8 for a summary and descriptive statistics. Hereafter, we provide a brief description and the expected influences on incidence and latency distributions. In this regard, whenever we mention the positive (resp. negative) impact of a covariate, we refer to the fact that higher (resp. lower) values of one variable are associated with a decrease (resp. increase) in the probability to be immune from default or a decrease (resp. increase) in the survival time of susceptible banks.

| **Variable** | **Description** |
|---|---|
| **EQ** | Total Equity / Total Assets |
| **LOANS** | Total Loans / Total Assets |
| **LOANS_CI** | Commercial and Industrial Loans / Total Assets |
| **LOANS_RE** | Real Estate Loans / Total Assets |
| **OREA** | Other Real Estate Owned / Total Assets |
| **NPA** | Non-Performing Assets / Total Assets |
| **LLA** | Loan Loss Allowance / Total Assets |
| **PLLL** | Provision for loan and lease losses / Total Assets |
| **EFF** | Efficiency Ratio |
| **ROA** | Return on Assets |
| **CBD** | Cash & Balances due from depository institutions / Total Assets |
| **TD100** | Time Deposits of $100,000 or More / Total Assets |
| **COREDEP** | Retail (core) Deposits / Total Assets |
| **SIZE** | log( Total Assets ) |
| **AGE** | Bank's age (years since Established Date) |

**Table 7.** Definition of variables used in our analysis.

| Name | Min | Max | Mean | Median | Std. Dev. | Q-0.001 | Q-0.999 |
|---|---|---|---|---|---|---|---|
| EQ | -2.1495 | 1.0000 | 0.1121 | 0.1017 | 0.0602 | 0.0038 | 0.8996 |
| LOANS | 0.0000 | 1.0610 | 0.6278 | 0.6528 | 0.1669 | 0.0000 | 0.9658 |
| LOANS_CI | 0.0000 | 0.9684 | 0.0909 | 0.0755 | 0.0702 | 0.0000 | 0.5863 |
| LOANS_RE | 0.0000 | 1.0093 | 0.4356 | 0.4469 | 0.1771 | 0.0000 | 0.8844 |
| OREA | 0.0000 | 0.3061 | 0.0056 | 0.0011 | 0.0132 | 0.0000 | 0.1458 |
| NPA | 0.0000 | 8.6016 | 0.0208 | 0.0133 | 0.0372 | 0.0000 | 0.2528 |
| LLA | 0.0000 | 0.2227 | 0.0101 | 0.0088 | 0.0068 | 0.0000 | 0.0692 |
| PLLL | -1.6206 | 0.3111 | 0.0022 | 0.0006 | 0.0069 | -0.0088 | 0.0686 |
| EFF | -136.6150 | 253.7500 | 0.7225 | 0.6870 | 1.0693 | 0.0106 | 4.6668 |
| ROA | -7.6191 | 2.7421 | 0.0081 | 0.0089 | 0.0279 | -0.1302 | 0.1713 |
| CBD | 0.0000 | 1.0000 | 0.0779 | 0.0484 | 0.0852 | 0.0003 | 0.8810 |
| TD100 | 0.0000 | 0.9308 | 0.1534 | 0.1381 | 0.0897 | 0.0000 | 0.6735 |
| COREDEP | -0.1613 | 1.1519 | 0.7255 | 0.7495 | 0.1323 | 0.0000 | 0.9451 |
| SIZE | 4.1897 | 21.4740 | 12.0564 | 11.9123 | 1.3604 | 8.4152 | 19.4153 |
| AGE | 0.0082 | 224.5328 | 72.9650 | 83.4014 | 42.4148 | 0.7534 | 183.6178 |

**Table 8.** Summary statistics.

**Capital adequacy**. As a measure of stability and financial strength, we consider the *Total Equity Capital* (EQ). Because capital serves as a buffer to absorb losses during periods of financial and economic stress, better capitalized banks are expected to be less susceptible and to have longer survival time (if classified as susceptible). In just few words, EQ is expected to have a negative impact on both incidence and latency distribution.

**Asset quality**. Different variables are considered in this category. The amount of

*Total Loans* (LOANS), usually the least liquid and most risky asset. *Non-Performing Assets* (NPA), measured as the sum of assets past due 30 days or more (but still accruing interest) and assets in non-accrual status. *Other Real Estate Owned* (OREA), which usually include properties acquired through foreclosure after the borrower's default, are used as a signal of possibly deteriorating loans. *Allowance for Loan and Lease Losses* (LLA), which is a reserve that a bank calculates on the basis of its credit risk and that is used to cover future charge-offs. While LOANS, OREA and NPA are expected to have positive impacts, the effect of LLA is not certain. If it is true that a higher value of LLA may reflect a potentially higher credit risk, it is also true that banks with not enough reserves may be more susceptible to default, especially in case of adverse economic conditions.

**Earnings**. The capability of a bank to generate earnings is measured by the *Return on Assets* (ROA), which is the ratio of net income after taxes and extraordinary items to Total Assets (annualized). We expect higher values of this ratio to belong to stronger and safer banks. Hence, a negative impact on both incidence and latency components.

**Management efficiency**. The *efficiency ratio* (EFF), computed as the the proportion of net operating revenues absorbed by overhead expenses, is used to measure the ability of the management to turn non-interest costs into revenues. Given that lower values of this ratio indicates greater efficiency, we expect both incidence and latency distributions to be positively affected by this variable.

**Liquidity**. It is measured by *Cash & Balances due from depository institutions* (CBD), which by definition are the most liquid assets, and *Retail Deposits* (COREDEP), which is the most stable source of funding for lending activities. We expect both CBD and COREDEP to have negative impacts on both incidence and latency components.

**Miscellaneous**. Finally, the logarithm of Total Assets and the time since establishment are used to control for bank SIZE and AGE, respectively.

Due to a non-identifiability issue, it is important to remark that only time-invariant covariates can be introduced in the logistic regression component of the model. Instead of using covariate values observed at failure or censoring time, we decided to include the averages of the covariates over the entire history. From both an economic and a statistical perspective, we find this solution more attractive than comparing more heterogeneous data from different points in time.

### *5.2. Estimation results*

The coefficient estimates, standard errors and Wald test p-values obtained from the estimation of the unpenalized PH cure model on the aforementioned US banks datasets are listed in Table 9. Considering a significance level of 5%, the degree of capitalization (EQ), the concentration of real-estate loans (LOANS_RE), the amount of non-performing assets (NPA), the indicator of profitability (ROA) and a bank's size are all statistically significant covariates in both incidence and latency components. Measures of Loss Allowance for Loans and Leases (LLA), management efficiency (EFF) and liquidity (COREDEP,CBD) are statistically significant in explaining the probability to be immune from bankruptcy, while measures of loans concentrations (LOANS,LOANS_CI) are significant in explaining the survival time of susceptible banks. Although the surprising negative sign of efficiency (EFF) in the incidence part and the negative ones of LOANS_CI and LOANS_RE (as well as the positive one of ROA) on latency, which could be spurious results due to any kind of dependency among variables, most statistically significant variables have coefficients which are in line with our expectations. The different impact of SIZE on incidence and latency distribution could be explained by the fact that bigger banks are less likely to default, but they are more risky once classified as susceptible.

| | **Cure** | | | **Survival** | | |
|---|---|---|---|---|---|---|
| | **b** | s.e. | pval | $\boldsymbol{\beta}$ | s.e. | pval |
| Intercept | 11.8315 | 2.1581 | 0.0000 | - | - | - |
| EQ | -60.2630 | 5.5264 | 0.0000 | -30.8835 | 1.6448 | 0.0000 |
| LOANS | -0.4659 | 2.1614 | 0.8294 | 3.1498 | 1.1521 | 0.0063 |
| LOANS_CI | 3.9017 | 2.3934 | 0.1031 | -3.1047 | 1.2226 | 0.0111 |
| LOANS_RE | 4.0146 | 1.7745 | 0.0237 | -2.6145 | 0.9591 | 0.0064 |
| OREA | -12.5303 | 6.5440 | 0.0555 | 2.2342 | 1.2018 | 0.0630 |
| NPA | 40.8715 | 5.0453 | 0.0000 | 0.9071 | 0.2223 | 0.0000 |
| LLA | 81.0929 | 25.3546 | 0.0014 | -2.9360 | 2.8249 | 0.2986 |
| EFF | -0.1983 | 0.0920 | 0.0313 | 0.0016 | 0.0041 | 0.6924 |
| ROA | -140.8351 | 12.2354 | 0.0000 | 2.1318 | 0.8802 | 0.0154 |
| CBD | -21.6297 | 3.6814 | 0.0000 | 0.1558 | 0.8159 | 0.8486 |
| COREDEP | -12.3113 | 1.2560 | 0.0000 | -0.3645 | 0.3263 | 0.2638 |
| SIZE | -0.2946 | 0.1045 | 0.0048 | 0.1625 | 0.0304 | 0.0000 |
| AGE | 0.0006 | 0.0027 | 0.8256 | -0.0019 | 0.0015 | 0.1998 |

**Table 9.** Results from the unpenalized PH Cure model. The failure times are computed from the dates of closure by the FDIC and the covariates in the logistic part are the average of all past history.

After a first analysis using the unpenalized model, variable selection have been performed using the penalized technique to retain only the most relevant covariates. In this regard, as explained in Section 3.1, the minimization of the Generalized Information Criterion (GIC) has been performed with a grid-search over the values $\{0.1, 0.2, ..., 3\}$ and the magnitude of penalization has been kept proportional to the standard errors of the estimated coefficient in the unpenalized model. Concerning the GIC parameter controlling the properties of variable selection, we tried both $\kappa_n = 2$ and $\kappa_n = \ln(n)$ obtaining the same selected variables. BIC has been minimized ($= 0.1015$) at $\lambda_1 = 2.1$ and $\lambda_2 = 1$, while AIC ($= 0.0926$) at $\lambda_1 = 2.1$ and $\lambda_2 = 0.5$. Despite different authors [14, 17, 23] provide a sandwich formula for the covariance matrix of the coefficient estimates, the derivation of the p-values from the Wald test remains unclear. For this reason, we rely on inference from the unpenalized model with the selected covariates only.

At a first glance, in Table 10, it is interesting to see how the coefficients of problematic variables, which were statistically significant in the unpenalized model, have been shrunk toward zero. These results show that less profitable (ROA), less liquid (COREDEP) and younger (AGE) banks are significantly associated with lower probabilities to be immune from bankruptcy. While the susceptible banks with higher amounts of non-performing assets (NPA) or a bigger SIZE have significantly shorter survival times. Finally, capital adequacy (EQ) remains an important variable to explain both incidence and latency distributions.

## 6. Conclusions

Empirical evidence in the study of bank failures suggests that a significant proportion of institutions may never default. Mixture cure models were specifically conceived for this kind of problems, where a portion of the population is clearly not susceptible to the event of interest. Originally developed in biostatistics to study long-term survivors of cancer in clinical trials, it allows to separate the factors influencing the probability of default from the ones affecting the survival time of susceptible banks.

| | Cure | | | Survival | | |
|---|---|---|---|---|---|---|
| | **b** | s.e. | pval | $\boldsymbol{\beta}$ | s.e. | pval |
| Intercept | 10.7344 | 0.7440 | 0.0000 | - | - | - |
| EQ | -63.2297 | 4.3469 | 0.0000 | -26.3568 | 0.9488 | 0.0000 |
| NPA | - | - | - | 1.0204 | 0.2018 | 0.0000 |
| EFF | - | - | - | 0.0018 | 0.0044 | 0.6757 |
| ROA | -208.0610 | 10.4460 | 0.0000 | - | - | - |
| COREDEP | -10.4035 | 0.8085 | 0.0000 | - | - | - |
| SIZE | - | - | - | 0.2292 | 0.0067 | 0.0000 |
| AGE | -0.0046 | 0.0022 | 0.0366 | -0.0008 | 0.0013 | 0.5519 |

**Table 10.** Results from the penalized PH Cure model. The p-values are computed on the basis of the unpenalized model and the variables selected by the penalized technique. The failure times are computed from the dates of closure by the FDIC and the covariates in the logistic part are the average of all past history. After tuning, $BIC = 0.1015$ is minimized with $\lambda_1 = 2.1$ and $\lambda_2 = 1$. Identical results are obtained using an AIC criterion: $AIC = 0.0926$, which is minimized with $\lambda_1 = 2.1$ and $\lambda_2 = 0.5$.

In this paper, we extend the semi-parametric proportional hazards cure model of Sy and Taylor [39] to time-varying covariates and we propose a SCAD penalized likelihood method for the selection of the most relevant variables. Finite sample performance has been evaluated by means of a simulation study and we show how the penalty-based variable selection procedure performs reasonably well in identifying the most relevant variables. This methodology is applied to a large dataset of 7492 United States commercial banks insured by the FDIC, which includes bank-specific explanatory variables observed on a quarterly basis during the period 2006-2016. These covariates serve as proxies for capital adequacy, asset quality, earnings, management efficiency and liquidity. Our findings suggest that banks with a higher probability to be immune from default are characterized by a higher degree of capitalization, profitability, liquidity and a longer history. In addition, we find that among the banks classified as susceptible (i.e. not immune from default) the ones with shorter survival exhibit lower levels of capitalization, higher amounts of non-performing loans and a bigger size.

## References


[1] H. Akaike, *Information theory and an extension of the maximum likelihood principle*, 2nd International Symposium on Information Theory, Academiai Kiado, 1973.

[2] J.W. Boag, *Maximum likelihood estimates of the proportion of patients cured by cancer therapy*, J. R. Stat. Soc. Ser. B. Stat. Methodol. 11 (1949), 15–53.

[3] L. Breiman, *Heuristics of instability and stabilization in model selection*, Ann. Statist. 24 (1996), 2350–2383.

[4] N.E. Breslow, *Contribution to the discussion of the paper by D.R. Cox*, J. R. Stat. Soc. Ser. B. Stat. Methodol. 34 (1972), 216–217.

[5] A.B. Cantor and J.J. Shuster, *Parametric versus non-parametric methods for estimating cure rates based on censored survival data*, Stat. Med. 11 (1992), 931–937.

[6] R.A. Cole and J.W. Gunther, *Separating the likelihood and timing of bank failure*, J. Bank. Financ. 19 (1995), 1073–1089.

[7] D.R. Cox, *Regression models and life-tables*, J. R. Stat. Soc. Ser. B. Stat. Methodol. (1972), 187–220.

[8] D. De Leonardis and R. Rocci, *Default risk analysis via a discrete-time cure rate model*, Appl. Stoch. Models Bus. Ind. 30 (2014), 529–543.

[9] A. Demirgüç-Kunt, *Deposit-institution failures: a review of empirical literature*, Fed. Reserve Bank Cleveland Econ. Rev. 25 (1989), 2–19.

[10] Y. Demyanyk and I. Hasan, *Financial crises and bank failures: a review of prediction methods*, Omega 38 (2010), 315–324.

[11] J.W. Denham, E. Denham, K.B. Dear and G.V. Hudson, *The follicular non-hodgkin's lymphomas—I. the possibility of cure*, Eur. J. Cancer 32 (1996), 470–479.

[12] L. Dirick, T. Bellotti, G. Claeskens and B. Baesens, *Macro-economic factors in credit risk calculations: including time-varying covariates in mixture cure models*, J. Bus. Econom. Statist. (2016).

[13] L. Dirick, G. Claeskens and B. Baesens, *An akaike information criterion for multiple event mixture cure models*, European J. Oper. Res. 241 (2015), 449–457.

[14] J. Fan and R. Li, *Variable selection via nonconcave penalized likelihood and its oracle properties*, J. Amer. Statist. Assoc. 96 (2001), 1348–1360.

[15] J. Fan and R. Li, *Variable selection for Cox's proportional hazards model and frailty model*, Ann. Statist. 30 (2002), 74–99.

[16] J. Fan and R. Li, *New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis*, J. Amer. Statist. Assoc. 99 (2004), 710–723.

[17] J. Fan and H. Peng, *Nonconcave penalized likelihood with a diverging number of parameters*, Ann. Statist. 32 (2004), 928–961.

[18] V.T. Farewell, *The use of mixture models for the analysis of survival data with long-term survivors*, Biometrics 38 (1982), 1041–1046.

[19] V.T. Farewell, *Mixture models in survival analysis: Are they worth the risk?*, Canad. J. Statist. 14 (1986), 257–262.

[20] M. Ghitany, R.A. Maller and S. Zhou, *Exponential mixture models with long-term survivors and covariates*, J. Multivariate Anal. 49 (1994), 218–241.

[21] D.J. Hendry, *Data generation for the cox proportional hazards model with time-dependent covariates: a method for medical researchers*, Stat. Med. 33 (2014), 436–454.

[22] A.E. Hoerl and R.W. Kennard, *Ridge regression: applications to nonorthogonal problems*, Technometrics 12 (1970), 69–82.

[23] Hunter, D.R. and Li, R., *Variable selection using MM algorithms*, Ann. Statist. 33 (2005), 1617–1642.

[24] D. Jones, R. Powles, D. Machin and R. Sylvester, *On estimating the proportion of cured patients in clinical studies*, Biometrie-Praximetrie 21 (1981), 1–11.

[25] J.D. Kalbfleisch and R.L. Prentice, *The statistical analysis of failure time data*, John Wiley & Sons, 2011.

[26] A.Y. Kuk and C.H. Chen, *A mixture model combining logistic regression with proportional hazards regression*, Biometrika 79 (1992), 531–541.

[27] W.R. Lane, S.W. Looney and J.W. Wansley, *An application of the cox proportional hazards model to bank failure*, J. Bank. Financ. 10 (1986), 511–531.

[28] H. Leeb and B.M. Pötscher, *On the distribution of penalized maximum likelihood estimators: The LASSO, SCAD, and thresholding*, J. Multivariate Anal. 100 (2009), 2065–2082.

[29] F. Liu, Z. Hua and A. Lim, *Identifying future defaulters: A hierarchical bayesian method*, European J. Oper. Res. 241 (2015), 202–211.

[30] T.A. Louis, *Finding the observed information matrix when using the EM algorithm*, J. R. Stat. Soc. Ser. B. Stat. Methodol. 44 (1982), 226–233.

[31] C.L. Mallows, *Some comments on Cp*, Technometrics 15 (1973), 661–675.

[32] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*, John Wiley & Sons, 2007.

[33] Y. Peng and K.B. Dear, *A nonparametric mixture model for cure rate estimation*, Biometrics 56 (2000), 237–243.

[34] R.M. Pfeiffer, A. Redd and R.J. Carroll, *On the impact of model selection on predictor identification and parameter inference*, Computation. Stat. 32 (2017), 667–690.

[35] C. Sanderson and R. Curtin, *Armadillo: a template-based c++ library for linear algebra*, Journal of Open Source Software 1 (2016), 26–32.

[36] G. Schwarz, *Estimating the dimension of a model*, Ann. Statist. 6 (1978), 461–464.

[37] H. Shi and G. Yin, *Landmark cure rate models with time-dependent covariates*, Stat. Methods Med. Res. 26 (2017), 2042–2054.

[38] J.P. Sy and J.M. Taylor, *Standard errors for the cox proportional hazards cure model*,

Math. Comput. Model. 33 (2001), 1237–1251.
[39] J.P. Sy and J.M. Taylor, *Estimation in a cox proportional hazards cure model*, Biometrics 56 (2000), 227–236.
[40] J.B. Thomson, *Predicting bank failures in the 1980s*, Fed. Reserve Bank Cleveland Econ. Rev. 27 (1991), 9–20.
[41] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. R. Stat. Soc. Ser. B. Stat. Methodol. 58 (1996), 267–288.
[42] E.N. Tong, C. Mues and L.C. Thomas, *Mixture cure models in credit scoring: If and when borrowers default*, European J. Oper. Res. 218 (2012), 132–139.
[43] G. Whalen, *A proportional hazards model of bank failure: an examination of its usefulness as an early warning tool*, Fed. Reserve Bank Cleveland Econ. Rev. 27 (1991), 21–30.
[44] D.C. Wheelock and P.W. Wilson, *Why do banks disappear? The determinants of US bank failures and acquisitions*, Rev. Econ. Stat. 82 (2000), 127–138.
[45] Y. Zhang, R. Li, C.L. Tsai, *Regularization parameter selections via generalized information criterion*, J. Amer. Statist. Assoc. 105 (2010), 312–323.
[46] H. Zou, *The adaptive lasso and its oracle properties*, J. Amer. Statist. Assoc. 101 (2006), 1418–1429.