Internet Appendix: Discrete-time hazard model in "Social Capital and the Viability of Stakeholder-Oriented Firms: Evidence from Savings Banks"

Charlotte Ostergaard
BI Norwegian Business School

Ibolya Schindele
BI Norwegian Business School

Bent Vale
Norges Bank

March 2015

Discrete-time hazard model of savings bank exit

To record event occurrence, we divide the time from branching deregulation into equal-sized intervals of length one year, with interval j defined as $(\, j-1, j\,]$. Interval $j=1$ is thus the first year following the date of branching deregulation, 1 January 1984.[1]

Let $\textsc{t}$ denote the time (years) elapsed from branching deregulation to the observed exit of savings bank i, i.e. we have observations on n independent and identically distributed random variables, where n is the number of banks observed at the beginning of interval 1. The failure function, $P(j) = prob(\textsc{t} \leq j)$, is the cumulative distribution function of $\textsc{t}$ with probability mass function $p(j)$. It defines, in turn, the survival function $S(j) = 1 - P(j) = prob(\textsc{t} > j)$ which is the probability that the duration of the lifetime of a randomly chosen bank exceeds j periods. Since each bank does not survive for the same number of periods after deregulation, we denote the last period of the lifetime of bank i, j_i.

The modeling of the economic relationship between the probability of survival and the explanatory variables focuses on the "hazard rate" rather than the survival function. The hazard rate is defined as the probability of the event of exit during interval j, conditional on survival up to that point in time. Below, we outline our estimation approach which follows Allison (1982) and Jenkins (2005).[2]

Let the hazard rate for bank i in year j be defined as

$$h_{ij} = prob(\textsc{t}_i = j |\, \textsc{t}_i \geq j, x_{ij})\,, \tag{1}$$

where x_{ij} is a $(k \times 1)$ vector of bank-specific (constant or time-varying) explanatory variables. We explain how we construct the explanatory variables, $\mathbf{x_i}$, in detail below but the general point is that $\mathbf{x_i}$ measures the characteristics of bank i and the markets in which it operates, among others, the level of social capital.

[1] Although it is possible to uncover the exact day of a bank's exit, we prefer to model the process in discrete rather than continuous time to match the frequency of the explanatory variables, most of which are available only annually.

[2] Jenkins (2005) is a valuable exposition of duration analysis and its implementation. For discrete-time methods, see also Singer and Willett (1993).

We specify a proportional odds logistic model for the hazard rate:

$$\log\left[\frac{h_{ij}}{1-h_{ij}}\right] = \log\left[\frac{h_{0j}}{1-h_{0j}}\right] + \beta' x_{ij} \tag{2}$$

$$\Leftrightarrow h_{ij} = \frac{1}{1+e^{-[\theta_{0j}+\beta' x_{ij}]}} \, . \tag{3}$$

In (2), the log-odds of the hazard rate for each bank depends linearly on x_{ij} and a "baseline" hazard of risk over time, $logit(h_{0j}) = \theta_{0j}$. Since the hazard rate is a (conditional) probability, it lies between zero and one, while the log of the odds ratio accordingly lies between minus and plus infinity. The baseline hazard is common to all banks and a function of observation time only. It is the underlying process driving the event of exit when the individual bank characteristics equal zero. In our setting, the baseline hazard captures the underlying process of consolidation in the Norwegian banking sector following deregulation.

We specify a functional form for θ_{0j},

$$\theta_{0j} = \alpha_0 + \alpha_1 \log(j) + \alpha_2 [\log(j)]^2 \, . \tag{4}$$

Ignoring first the quadratic term in (4), the sign of α_1 controls the pattern of duration dependence for the population of savings banks. When α_1 is negative the hazard rate is monotonically decreasing over time for all banks, and vice versa for positive α_1. When α_1 is zero, the baseline probability of exit is constant for all observation intervals. We include a quadratic term to capture the fact that the hazard rate cannot continuously decrease or increase forever, given that the population of banks at the beginning of the sample is fixed. In practice, the form in (4) was chosen based on a preliminary non-parametric estimation of the baseline hazard, with the aim of capturing the "shape" of the process of consolidation in a parsimonious manner, preserving degrees of freedom. As a robustness check, we estimate our main survival regression using time dummy variables in place of (4).

Our sample is right-censored as we do not observe the life duration of banks that survive from the time of deregulation until the end of our sample. We only know that these banks did not exit prior to 2005, the end of our sample period, as, by nature, banks can only exit once.[3]

Define an indicator variable, δ_i equal to one if bank i exits during the sample and zero otherwise (censoring). The general form of the likelihood function corresponding to the observations of τ_i is

$$\begin{aligned} L &= \prod_{i,\text{uncensored}} p(j_i) \prod_{i,\text{censored}} [1 - P(j_i)] \\ &= \prod_{i=1}^{n} p(j_i)^{\delta_i} [1 - P(j_i)]^{(1-\delta_i)} \end{aligned} \quad (5)$$

There is a one-to-one relationship between the survival function and the hazard rate and (5) can therefore be rewritten in terms of the latter, $S(j) = \Pi_{k=1}^{j}(1 - h_k)$. In our setting, the probability functions must be further modified for left-truncation—the relevant starting date for our "experiment" is the year of deregulation, 1984, but we observe the population of banks only three years later, from 1987.

Let j_τ denote the point of truncation (the year of 1987, common to all banks). The truncated conditional probability functions can be written in terms of the hazard rate as

$$p(j_i | j_i > j_\tau) = \frac{h_{ij_i} \prod_{k=1}^{j_i - 1}(1 - h_{ik})}{\prod_{k=1}^{j_\tau}(1 - h_{ik})} = h_{ij_i} \prod_{k=j_\tau}^{j_i - 1}(1 - h_{ik}) \quad (6)$$

for censored observations and

$$1 - P(j_i | j_i > j_\tau) = \frac{\prod_{k=1}^{j_i}(1 - h_{ik})}{\prod_{k=1}^{j_\tau}(1 - h_{ik})} = \prod_{k=j_\tau}^{j_i}(1 - h_{ik}) \quad (7)$$

for uncensored observations respectively. The corresponding unconditional expressions are respectively

$$prob(\tau_i > j_i) = S(j_i) = (1 - h_{i1})(1 - h_{i2})...(1 - h_{ij_i}) = \prod_{k=1}^{j_i}(1 - h_{ik}) \quad (8)$$

[3]Censoring is indeed one reason why an OLS regression of life duration on bank and municipality-characteristics would be an inappropriate estimation approach for the issue at hand. The alternative approach of defining a binary dependent variable that equals one if a bank exits during the sample period ignores important information regarding the timing of exit, see Allison (1982) for a discussion of such issues.

and

$$prob(\mathrm{T}_i = j_i) = h_{ij_i} S(j_i - 1) = h_{ij_i} \prod_{k=1}^{j_i-1} (1 - h_{ik})\,. \tag{9}$$

Substituting into the likelihood function we obtain

$$L = \prod_{i=1}^{n} \Big[h_{ij_i} \prod_{k=j_\tau}^{j_i-1} (1 - h_{ik}) \Big]^{\delta_i} \Big[\prod_{k=j_\tau}^{j_i} (1 - h_{ik}) \Big]^{1-\delta_i}\,. \tag{10}$$

Brown (1975) and Allison (1982) demonstrate that (10) can be reformulated as the likelihood function for a binary dependent variable, y_{ij}, where

$$y_{ij} = \begin{cases} 1, & \text{if bank } i \text{ exits during interval } j \\ 0, & \text{if bank } i \text{ does not exit during interval } j \end{cases}\,. \tag{11}$$

Hence, if the event of exit occurs for bank i during, say, the fifth year of observation, y_{ij} equals zero in years one to four, and one in year five. For banks that are not observed to exit during our sample, y_{ij} equals zero in all periods. Essentially, this formulation converts the problem into a panel with a binary bank-specific dependent variable where the time dimension refers to the number of observation periods for each bank. The panel is unbalanced because not all banks survive for the same number of years. The reformulated likelihood function becomes

$$L = \prod_{i=1}^{n} \Big[\prod_{k=j_\tau}^{j_i} h_{ik}^{y_{ik}} (1 - h_{ik})^{(1-y_{ik})} \Big]\,. \tag{12}$$

The likelihood in (12) has the standard form for a logistic binary dependent variable, y_{ik}, with probabilities h_{ik} and $(1 - h_{ik})$ respectively (given that h_{ik} is logistic by assumption). Hence, (2) may be estimated as a logit regression with y_{it} as the dependent variable and α_0, $\log(j)$, $(\log(j))^2$, and x_{ij} as explanatory variables. The total number of observations equals $\sum_{i=1}^{n} (j_i - j_\tau)$ and bank i is observed for j_i periods.

References

Allison, P. D. (1982), 'Discrete-time methods for the analysis of event data', *Sociological Methodology* **13**(1), 61–98.

Brown, C. C. (1975), 'On the use of indicator variables for studying the time dependence of parameters in a responce-time model', *Biometrics* **31**, 863–872.

Jenkins, S. P. (2005), *Survival Analysis*, Unpublished Manuscript, Institute for Social and Economic Research, University of Essex.

Singer, J. D. and Willett, J. B. (1993), 'It's about time: Using discrete-time survival analysis to study duration and the timig of events', *Journal of Educational Statistics* **18**(2), 155–195.