Journal of the Operational Research Society (2012) 63, 306–321

www.palgrave-journals.com/jors/

A time-dependent proportional hazards survival model for credit risk analysis

J-K Im[1], DW Apley[1*], C Qi[2] and X Shan[3]

[1]Northwestern University, Evanston, IL, USA; [2]Consultant, Lake Forest, IL, USA; and [3]Consultant, Deerfield, IL, USA

In the consumer credit industry, assessment of default risk is critically important for the financial health of both the lender and the borrower. Methods for predicting risk for an applicant using credit bureau and application data, typically based on logistic regression or survival analysis, are universally employed by credit card companies. Because of the manner in which the predictive models are fit using large historical sets of existing customer data that extend over many years, default trends, anomalies, and other temporal phenomena that result from dynamic economic conditions are not brought to light. We introduce a modification of the proportional hazards survival model that includes a time-dependency mechanism for capturing temporal phenomena, and we develop a maximum likelihood algorithm for fitting the model. Using a very large, real data set, we demonstrate that incorporating the time dependency can provide more accurate risk scoring, as well as important insight into dynamic market effects that can inform and enhance related decision making.

Journal of the Operational Research Society (2012) **63,** 306–321. doi:10.1057/jors.2011.34

Published online 11 May 2011

Keywords: risk: predicting consumer credit risk; failure models: survival analysis for credit risk modelling; data analysis: estimating survival models

1. Introduction

In the credit industry, profits realized on loan products or credit cards depend heavily on whether customers pay interest regularly, miss payments, default on their loans, etc. Because of this, the industry has invested substantial research into identifying high-risk credit applicants. For this purpose, companies calculate numerical scores of applicants' creditworthiness based on the applicants' credit bureau data, such as past credit activity, and their application data. This process is generally referred to as consumer credit risk assessment, in contrast to corporate and portfolio credit risk assessment, which involves very different considerations.

Regarding consumer credit risk, which is the subject of this paper, Rosenberg and Gleit (1994), Hand and Henley (1997), Thomas (2000), Thomas *et al* (2002), Baesens *et al* (2003), Thomas *et al* (2004), Thomas *et al* (2005), and Crook *et al* (2007) reviewed several different risk assessment objectives and approaches. The risk assessment schemes generally fall into three main areas: credit scoring, behaviour scoring, and profit scoring. The objective of credit scoring (also called application scoring) is to help lenders discriminate between high-risk applicant and low-risk applicant, in terms of likely default behaviour, and make decisions whether to accept an applicant based on the score. The objective of behaviour scoring is to help lenders make better decisions in managing existing customers by assessing likely future performance, based on past performance of the customer. The objective of profit scoring is to assess the likely profitability of an applicant, which is related to default likelihood, but involves a number of other considerations as well. In this work, we focus on the credit scoring objective of predicting default behaviour of applicants.

The most common credit scoring approach uses logistic regression (LR) (Stepanova and Thomas, 2002). One specifies a period of time (eg, 24 months), and then fits an LR model to historical customer data for predicting the probability an applicant will default within the specified time period, as a function of the applicant's credit bureau and application variables (henceforth referred to as predictor variables or predictors). As a lower predicted probability means better creditworthiness, one usually sets a cut-off threshold and approves credit to those having predicted probability less than the threshold. Numerous other statistical methods that attempt to fit more complex models with higher degrees of nonlinearity between the predictors and the response, such as support vector machines, neural networks, and Bayesian network classifiers, have also been

*Correspondence: *DW Apley, Department of Industrial Engineering & Management Sciences, Northwestern University, 2145 Sheridan Rd. (C210), Evanston, IL 60208-3119, USA.*
E-mail: apley@northwestern.edu

investigated for credit scoring. Baesens *et al* (2003) summarized and benchmarked many such methods but concluded that the more complex models generally perform quite similarly to LR, in terms of predicting default probability. Moreover, Hand (2006) argued that potential performance improvements attainable using more complex models are often offset by other sources of uncertainty that are exacerbated by the added complexity.

Survival analysis is an alternative to LR that is still reasonably simple, in that it does not involve an overly parameterized model. In survival analysis for credit risk scoring, the objective is to model the *distribution* of the time T to default (or of some other event associated with default), as opposed to the LR objective of predicting the probability of default within a single, specified period of time. The distribution is allowed to be a function of an applicant's predictor variables via a proportional hazards (PH) survival model [see Leemis (1995) for comprehensive treatments of PH and other survival models in the context of reliability analysis]. Banasik *et al* (1999) and Stepanova and Thomas (2002) have investigated PH survival models for credit risk scoring. For the sole purpose of predicting the probability of default within a single specified period, traditional PH survival modelling has little advantage over LR. Stepanova and Thomas (2002) have found their performances nearly indistinguishable. However, survival analysis, with the predicted distribution of T that it provides, offers a number of other advantages: First, it provides a consistent means of predicting probability of default within many different periods of time (eg, 12 month default rate, 24 month default rate, etc). Second, it possesses an inherent mechanism for taking into consideration the most recent data. In contrast, in LR, if one wishes to predict the probability of default within 24 months, customers joining within the past 24 months cannot be included when fitting the model. Third, it provides more complete information on the predicted behaviour of T via its predicted distribution. Relevant information includes a point estimate of T (eg, the mean of the predicted distribution) for each applicant, as well as a quantitative understanding of the uncertainty one may expect in T (eg, via the upper and lower 0.05 quantiles of the predicted distribution). Knowledge obtained from the predicted distribution of T can be useful in the broader context of profit modelling.

An additional characteristic of survival modelling, which we develop as the main contribution of this work, is that it can be conveniently modified to incorporate dynamic economic conditions into the model. Thomas (2000) has pointed out the important impact that dynamic conditions can have on credit risk. For example, the recent global economic downturn substantially increases an applicant's default likelihood, relative to someone with the same predictor variables applying in the past when the economy was healthier. In light of this, we develop an extension of the PH survival model that takes into account dynamic economic conditions. We refer to this as a time-dependent proportional hazards (TDPH) survival model. The TDPH survival model represents the effects of dynamic economic conditions in a direct manner, without the need to identify a set of underlying macroeconomic factors that best characterizes the current state of the economy (in terms of its impact on consumer credit risk) and include them as additional predictor variables, as was done in Tang *et al* (2007) and Bellotti and Crook (2009). In a sense that will become clear later, the TDPH survival modelling automatically identifies a single, scalar, time-varying factor that represents the net collective effect of all dynamic economic conditions on default likelihood. We demonstrate that in a dynamic economic environment, this offers modest but significant (statistically and practically) improvements in credit scoring accuracy *versus* the industry standard LR. Equally importantly, the approach provides a means of identifying and understanding the net effect of dynamic market conditions in a simple, quantitative manner that can be easily incorporated into other predictive decision-making strategies. We demonstrate these points using a very large, real data set from a consumer credit company.

Besides the aforementioned works that include time-varying macroeconomic factors in the PH model, Stepanova and Thomas (2002) and Tang *et al* (2007) incorporated a type of time dependency by including interactions between the predictors and time as additional terms in the model. This is equivalent to assuming that the effects (ie, the coefficients in the PH model) of the predictors change as linear functions of the time since a customer joined. This was intended to represent changes in the effects of certain predictor variables that become more or less pronounced the longer one has been a customer. It was not intended to capture the effects of dynamic market conditions.

In terms of identifying an external time-dependent factor, the TDPH model is similar to the dual-time dynamics (DtD) model of Breeden (2007), Breeden *et al* (2008), and Breeden and Thomas (2008). The DtD approach represents the collective effect of external time-dependent factors as an additional additive term in a generalized additive model (GAM). The commonality of the DtD and TDPH methods is that neither requires macroeconomic variables to be identified in advance and included in the model. Rather, they both represent the collective effect of such time-dependent factors using a single, scalar function of calendar time. The specifics of how this is accomplished are quite different, however. The GAM model is better suited for modelling continuous variables (eg, profit, loss, or exposure), whereas the TDPH model is better suited for modelling the hazard function of a time-related event (eg, credit default).

The format of the remainder of the paper is as follows. In Section 2 we provide background on the traditional PH survival model. In Section 3 we introduce our TDPH extension of the model. In Section 4 we derive the maximum likelihood estimates (MLEs) of the TDPH survival model parameters. In Section 5 we demonstrate its use in credit risk scoring with the real data set and discuss a number of implementation issues and performance characteristics. As we will show, the TDPH survival model appears to fit the data quite well, especially in terms of capturing market dynamics. We argue that it is a reasonable and useful approach for modelling the time-to-default.

2. Background on PH survival modelling for T

Let $f(t)$ and $F(t)$ denote the probability density function (pdf) and cumulative distribution function (cdf) of the time T to default ($T=0$ corresponds to the time of approval). The hazard function is defined as $h(t)=(1-F(t))^{-1}f(t)$ and is interpreted as the instantaneous likelihood of defaulting at time t, given that the customer has not defaulted prior to time t. From the definition of the hazard function, it can be shown that $F(t) = 1 - \exp\left(-\int_0^t h(u)du\right)$. We refer interested readers to Leemis (1995) or Meeker and Escobar (1998) for derivations of these and related results and, in general, for more comprehensive treatments of survival modelling.

Strictly speaking, time-to-default data are discrete, because all records are generated based on monthly payment behaviour. However, for tractability purpose, we represent the distribution of T as continuous and take the probability that a customer defaults in month number t (for $t=1, 2, \ldots$) to be $F(t)-F(t-1)$.

Let $x_1, x_2, \ldots, x_M$ denote a set of M predictor variables for an applicant, and define the predictor vector $\mathbf{x}=[x_1, x_2, \ldots, x_M]'$. In survival analysis, perhaps the most popular way to allow the distribution of T to depend on a set of predictor variables is through a PH survival model, defined as follows. Denoting the hazard function for a customer with predictors $\mathbf{x}$ by $h(t;\mathbf{x})$ to indicate its explicit dependence on $\mathbf{x}$, a PH survival model represents $h(t;\mathbf{x})=h_0(t)\psi(\mathbf{x})$, where $h_0(t)$ denotes the hazard function for some baseline distribution $f_0(t)$, and $\psi(\mathbf{x})$ is some appropriate function of the predictors. Two common choices for $\psi(\mathbf{x})$ are the exponential function $e^{\beta_0+\boldsymbol{\beta}'\mathbf{x}}$ or the logistic (s-shaped) function $(1+e^{\beta_0+\boldsymbol{\beta}'\mathbf{x}})^{-1}e^{\beta_0+\boldsymbol{\beta}'\mathbf{x}}$, where the $M\times 1$ vector $\boldsymbol{\beta}$ and the scalar β_0 are parameters to be estimated, along with the parameters of the baseline distribution $f_0(t)$ [equivalently, the parameters of $h_0(t)$]. Notice that for a specific applicant (ie, for a fixed value of $\mathbf{x}$), the hazard function $h(t,\mathbf{x})$ is proportional to the baseline hazard function; hence the terminology PH survival model.

The application of survival models to credit risk assessment has been studied by a number of researchers. Banasik *et al* (1999) and Stepanova and Thomas (2002) applied a PH survival model to credit scoring, while Stepanova and Thomas (2001) applied it to behavioural scoring. Baesens *et al* (2005) discussed the use of neural networks for survival analysis and Andreeva *et al* (2007) used a PH survival model to model profitability. Bellotti and Crook (2009) incorporated macroeconomic variables using a PH survival model.

The objective in survival analysis is to estimate the parameters of $h_0(t)$ and $\psi(\mathbf{x})$ from a set of historical data, after which the predicted distribution for a new applicant can be calculated via the relationship $F(t;\mathbf{x}) = 1 - \exp\left(-\int_0^t h(u;\mathbf{x})du\right)$ with $h(t;\mathbf{x})=\psi(\mathbf{x})h_0(t)$. Notice that we have added the argument $\mathbf{x}$ to the pdf and cdf to indicate that they depend on the predictor variables for an applicant. One quantity that can be extracted from the predicted distribution is the probability that an applicant will default within the specific time period, which is what LR produces. For example, the predicted probability of default within 24 months is simply $F(24;\mathbf{x})$, the predicted cdf evaluated at month $t=24$. Conceptually, this is the area under the predicted pdf $f(t;\mathbf{x})$ between $t=0$ and $t=24$, as depicted in Figure 1.

For predicting default rates within a specific period of time, Stepanova and Thomas (2002) found that there was no significant difference between the receiver operating characteristic (ROC) curves of PH survival modelling *versus* LR. This is not surprising considering that LR directly models the probability of default within a specific time period, whereas this is indirectly obtained in PH survival modelling. In the remainder of the paper, we demonstrate that if one modifies the PH survival model to incorporate dynamic market conditions (as described in Section 3), the conclusions are different. In particular, the TDPH model significantly outperforms LR for scoring applicants, at least for the data set that we consider.

3. The TDPH survival model

Over an extended time period, all markets are dynamic. To illustrate, Figure 2 shows the percentage of customers

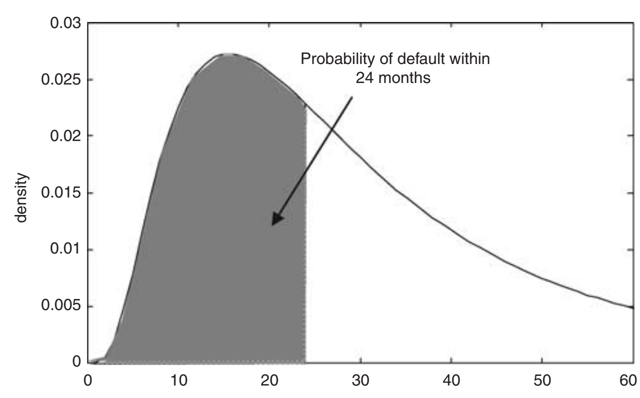

Figure 1 Predicted pdf of T and probability of default within 24 months (shaded area).

who defaulted within their first 9 months, for three different vintages (ie, customers joining in the three different quarters: Quarter 2 of 2004, Quarter 4 of 2005, and Quarter 4 of 2007). The customers in all three vintages fell into the same FICO scoring band (between 675 and 705). Hence, if one ignored market trends and attempted to predict default probability based only on the applicants' predictor variables, one would naively conclude that customers in the three vintages all have the same default probability. In reality, Figure 2 shows that the default probability is much higher for the Q4 2007 vintage, because of the severe economic downturn in 2008. Owing to confidentiality concerns, the numerical values for the default rate are omitted in Figure 2.

To account for such temporal effects, one potential approach is to incorporate macroeconomic variables into the PH survival model. Tang *et al* (2007) and Bellotti and Crook (2009) incorporated macroeconomic variables such as interest rate and unemployment rate as additional predictors to improve the model fit and the prediction of default probability. One drawback of this approach is that it may be quite difficult to identify the right economic factors to include. For example, there was a temporary spike in defaults in December, 2005 (which we will discuss later) that was triggered by changes in the bankruptcy laws. This could not have been realistically modelled via a dependency on common macroeconomic variables. The approach that we develop avoids this difficulty by more directly modelling the effects of dynamic market phenomena.

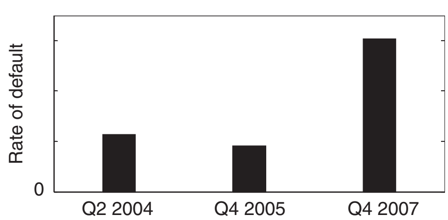

Figure 2 Actual rates of 9-month default for three vintages of customers, all with similar FICO scores.

For a customer that joins during month τ, denote his or her hazard function by $h(t; \mathbf{x}, \tau)$ to explicitly indicate its dependence not only on $\mathbf{x}$, but also on the time at which the customer joins. In our TDPH survival model, we represent $h(t; \mathbf{x}, \tau) = h_0(t)\psi(\mathbf{x})\gamma(\tau + t)$, where $\gamma(\tau + t)$ is a scalar function of time that can be viewed as a time-dependency factor. One might regard γ as a macroeconomic factor, but more precisely, it represents the aggregated effects of external time-dependent factors that were not accounted for by the predictor variables in the PH survival model. Notice that $\tau + t$ represents absolute time, and t represents the time relative to when the customer joined. Notice also that the first month in which a customer can default is $\tau + 1$, corresponding to $t = 1$. Throughout this paper, we use a piecewise constant function for γ that is constant over each quarter. Figure 3 shows the estimated γ function (denoted $\hat{\gamma}$) for our real data set. We discuss the estimation algorithm and other model details more fully in Sections 4 and 5. Here, we include it only to help illustrate the model by providing a concrete example of what the γ function might look like. Such a nonparametric γ function avoids the overly-restrictive assumptions that are implicit in a parametric choice for γ, such as linear, quadratic, or exponential. In general, nonparametric functions often involve the risk of overfitting, but credit scoring data sets are so large that this is hardly a problem for our application (our data set consisted of roughly 200 000 customers and even this number represented $< 10\%$ of the total number of available customer records).

Regarding the effect of the γ function and how one should interpret it, consider two customers with the exact same predictors $\mathbf{x}$, but with Customer 1 booked in January, 2006 and Customer 2 in August, 2007. At month $t = 12$ for each customer, the hazard function $h(12; \mathbf{x}, \tau)$ for Customer 2 is roughly double that of Customer 1, because γ for August, 2008 is roughly double γ for January, 2007. This means that if neither customer defaulted through their first 11 months, the likelihood that Customer 2 defaults in their month 12 is roughly double the corresponding value for Customer 1.

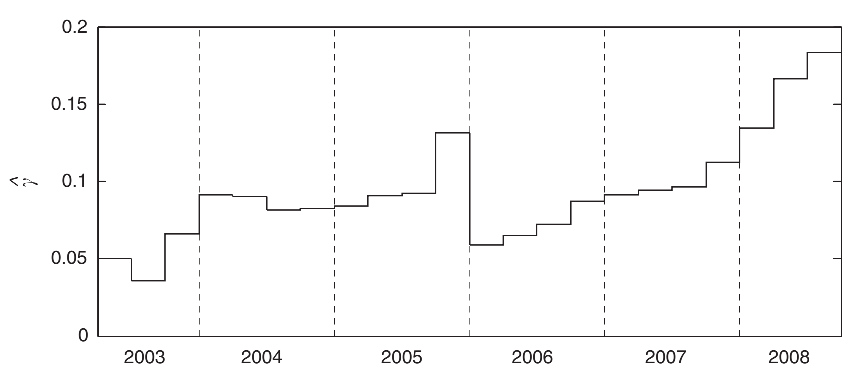

Figure 3 Estimated γ for the real data set. The dashed vertical lines represent the changes in calendar year.

4. Maximum likelihood estimation of the TDPH survival model parameters

To estimate the parameters of the TDPH survival model, we develop an MLE method that is a modification of the standard MLE approach commonly used in PH survival modelling (Leemis, 1995). Using the relationship between F and h for survival models in general, the cdf of T for a customer with predictors $\mathbf{x}$ joining in month τ can be expressed as $F(t;\mathbf{x},\tau;\psi,h_0,\gamma) = 1-\exp\big(-\int_0^t h_0(u)\psi(\mathbf{x})\gamma(\tau+u)du\big)$ for the TDPH survival model. For clarity, we have added arguments to F to explicitly indicate its dependence on the parameters. Because γ is modelled as piecewise constant with quarterly pieces, and $\psi(\mathbf{x})$ does not depend on the variable u in the integrand, it is convenient to break the integral up into components that correspond to the pieces over which γ is constant. Doing this, we show in Appendix B that the cdf is given by

$$\begin{aligned}
&F(t;\,\mathbf{x},\,\tau;\,h_0,\,\psi,\,\gamma)\\
&\quad= 1-\exp\left(-\psi(\mathbf{x})\sum_{k=1}^{w(\tau,t)}\gamma_{\lceil\tau/3\rceil+k-1}\cdot\right.\\
&\qquad\times\big(-\log\big(1-F_0(s_{\tau,k})\big)\\
&\qquad\left.+\log\big(1-F_0(s_{\tau,k-1})\big)\big)\right).
\end{aligned}$$

Precise definitions of the notation, as well as further details, are given in Appendix B. Conceptually, $w(\tau,t)$ denotes the number of different intervals over which the piecewise constant γ is constant (ie, the number of pieces) between the time of approval (τ) and the time in question ($\tau+t$); and $\gamma_{\lceil\tau/3\rceil+k-1}$ and $s_{\tau,k}$ represent the value of γ over the kth interval, and the upper boundary of the interval, respectively.

Now consider the data $\{t_i,\mathbf{x}_i,\tau_i,\delta_i: i=1,2,\ldots,N\}$ from which the model parameters are to be estimated, where N denotes the total number of customers in the (training) data set. We have added a subscript i to t, $\mathbf{x}$ and τ to denote the customer index. The binary variable δ_i indicates whether the ith customer defaulted at any point in time (1 for default; 0 otherwise). If $\delta_i=1$, then t_i denotes the time at which the ith customer defaulted. Otherwise, t_i denotes the time at which the ith customer was censored (eg, $t_i=15$ if the ith customer's account was closed in good standing after 15 months as an active customer; or $t_i=18$ if the ith customer joined 18 months prior to the last month of the data set and remained a customer in good standing for the duration).

The log-likelihood function for the ith customer is

$$\begin{aligned}
&l(h_0,\,\psi,\,\gamma|t_i,\,\mathbf{x}_i,\,\tau_i,\,\delta_i)\\
&\quad=\delta_i\,\log(F(t_i;\,\mathbf{x}_i,\,\tau_i;\,h_0,\,\psi,\,\gamma)\\
&\qquad-F(t_i-1;\,\mathbf{x}_i,\,\tau_i;\,h_0,\,\psi,\,\gamma))\\
&\qquad+(1-\delta_i)\log(1-F(t_i;\,\mathbf{x}_i,\,\tau_i;\,h_0,\,\psi,\,\gamma)).
\end{aligned}$$

Notice that if $\delta_i=1$, the log-likelihood for the ith customer reduces to $\log(F(t_i;\mathbf{x}_i,\tau_i;h_0,\psi,\gamma)-F(t_i-1;\mathbf{x}_i,\tau_i;h_0,\psi,\gamma))$, which is the log of the probability that the customer defaulted in month t_i. If $\delta_i=0$, the log-likelihood reduces to $\log(1-F(t_i;\mathbf{x}_i,\tau_i;h_0,\psi,\gamma))$, the log of the probability that the customer remains in good standing until their censored time. Also notice that, given the time τ_i at which the customer joined (and hence the knowledge of whether t_i represents a default or a censor), the binary random variable δ_i is completely determined by t_i. Hence, when deriving the likelihood, it is sufficient to consider only the distribution of t_i. In essence, δ_i is used simply as a convenient notation. Assuming the customers are independent (conditioned on τ), the log-likelihood function for the entire data is

$$\begin{aligned}
&l(h_0,\,\psi,\,\gamma|\mathbf{t},\,\mathbf{X},\,\boldsymbol{\tau},\,\boldsymbol{\delta})\\
&\quad=\sum_{i=1}^{N}(\delta_i\log(F(t_i;\,\mathbf{x}_i,\,\tau_i;\,h_0,\,\psi,\,\gamma)\\
&\qquad-F(t_i-1;\,\mathbf{x}_i,\,\tau_i;\,h_0,\,\psi,\,\gamma))\\
&\qquad+(1-\delta_i)\log(1-F(t_i;\,\mathbf{x}_i,\,\tau_i;\,h_0,\,\psi,\,\gamma))),
\end{aligned}$$

where $\mathbf{t}$, $\boldsymbol{\tau}$, and $\boldsymbol{\delta}$ are $N\times 1$ column vectors whose ith elements are t_i, τ_i, and δ_i, respectively, and $\mathbf{X}$ is the $N\times M$ matrix with ith row $\mathbf{x}_i'$.

The log-likelihood $l(h_0,\psi,\gamma\ |\mathbf{t},\mathbf{X},\boldsymbol{\tau},\boldsymbol{\delta})$ is maximized to estimate the parameters of h_0, ψ, and γ. Notice that the parameters of the piecewise constant γ are simply the values of the pieces, which we represent as a $Q\times 1$ vector of parameters $\boldsymbol{\gamma}=[\gamma_1,\gamma_2,\ldots,\gamma_Q]'$, where Q denotes the total number of quarters over which the data are available. To implement the MLE, one can use a gradient-based optimization algorithm. In Appendix C, we derive expressions for the log-likelihood and its gradient for the special case of a lognormal $h_0(t)$, an exponential function for $\psi(\mathbf{x})$, and a piecewise constant γ. Matlab code for this special case is available upon request from the authors.

Regarding confidence intervals on the parameters, for the large sample sizes commonly found in consumer credit risk modelling, asymptotic results for MLEs may provide reasonable approximate confidence intervals. Alternatively, bootstrap methods could be used. Although bootstrapped confidence intervals would be conceptually straightforward, the computational expense may be prohibitive, considering that it takes a substantial amount of time to fit a single TDPH model.

5. Results and discussion

We applied our approach using a set of credit card customer performance data from a major US financial institution. The data consist of (1) all customers who were approved between January 2003 and July 2008 and who defaulted at some point within this period, and (2) a 1% random sample of all customers who were approved between January 2003 and July 2008 and who did not default. That is, our data set is obtained by random under-sampling of the majority class (those who did not default) and by using the entire minority class (those who did default). Other sampling methods that may have resulted in better performance are certainly possible. See Batista *et al* (2004) and the recent discussion in Chawla *et al* (2004) on balancing data sets to improve classification performance. In practice, if one under-samples the majority class when constructing the training set, it is advisable to use the entire data as a test set. However, we do not do this in this paper because of data accessibility issues.

In the combined set, there were a total of $N = 212\,742$ customers. A 2/3 random sample was used for training and the rest 1/3 for testing. Of the thousands of potential predictor variables, 75 were selected by domain experts as good candidates to consider for this study. Of these, for illustrative purposes, we chose to include only 10 rather significant ones (selected by simply using a forward sequential feature selection method for predicting the 9 month default probability using LR with all 75 predictors). For all analyses, the predictors were standardized before fitting the model. Some comments on the data treatment are provided in Appendix A.

5.1. The fitted TDPH model and comparison with standard PH and LR models

Using the MLE algorithm described in Section 4, we simultaneously estimated the values of the time-dependency function γ (which are plotted in Figure 3), as well as the parameters of the exponential function ψ and the lognormal base hazard function h_0. We focus on the case of an exponential $\psi(\mathbf{x})$, because we believe it provides better ability to discriminate between moderately high-risk customer and very high-risk customer. The s-shape of the logistic $\psi(\mathbf{x})$ may better suit for discriminating between moderately low risk and moderately high-risk customers. Overall, based on the performance metrics that we discuss later, the exponential $\psi(\mathbf{x})$ resulted in the most effective scoring. The estimated parameters of the base lognormal distribution were $\hat{\mu} = 2.83$ and $\hat{\sigma} = 0.557$ (see Appendix C for their definitions). The estimated TDPH $\hat{\boldsymbol{\beta}}$ for the 10 (standardized) predictors are plotted in Figure 4, along with their counterparts obtained by fitting a PH model, and LR models for predicting 9 month default and 24 month default, using the same 10 predictors. Although $\hat{\boldsymbol{\beta}}$ has slightly different meaning in these models, one would expect to see the same general trends. Figure 4 indicates that this is the case. The elements of $\hat{\boldsymbol{\beta}}$ are quite similar for all four models.

From Figure 3, which plots the $\hat{\gamma}$ values, we see two interesting phenomena. First, the pronounced upward spike in Q4 2005 is clearly an anomaly. The reason for this anomaly was changes in the bankruptcy laws that went into effect in January, 2006, making it more difficult to declare bankruptcy. As a result, there was a marked increase in the number of defaults shortly before the law went into effect, explaining the upward spike in $\hat{\gamma}$ in Q4 2005. For the same reason, there was drop in the number of defaults after the law went into effect, which resulted in a somewhat lower $\hat{\gamma}$ over the first half of 2006. The second interesting phenomenon is the gradual increase in $\hat{\gamma}$ over the last 2 years, which began to accelerate in 2008. This is clearly due to the economic downturn and credit crisis of 2008. We also note the oddly low values of $\hat{\gamma}$ over the first year of the data set. We believe this does not truly reflect economic conditions at that time. Rather, it is probably due to the base distribution providing a poor fit of the actual default probabilities in the very early months (eg, the first four months) of an account. We discuss this further in Section 5.3.

To help gauge how well the TDPH model fits the data, Figure 5 shows empirical distributions of T over a 50 month time period for customers in the data set that joined in Q2 2004, for three different segments of customers, segmented based on their predictor values. The figure also shows the fitted pdfs of T, obtained by plugging each segment's median predictor values and τ value into the fitted TDPH model. The same fitted TDPH model was used to obtain all three fitted pdfs in Figure 5. Notice that

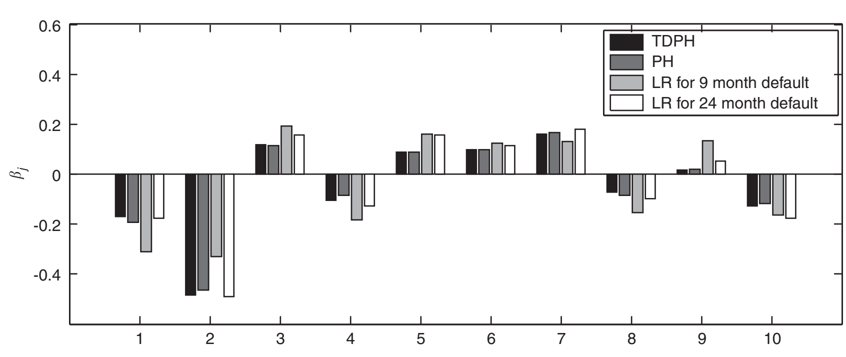

Figure 4 The components of β for TDPH, PH, and LR.

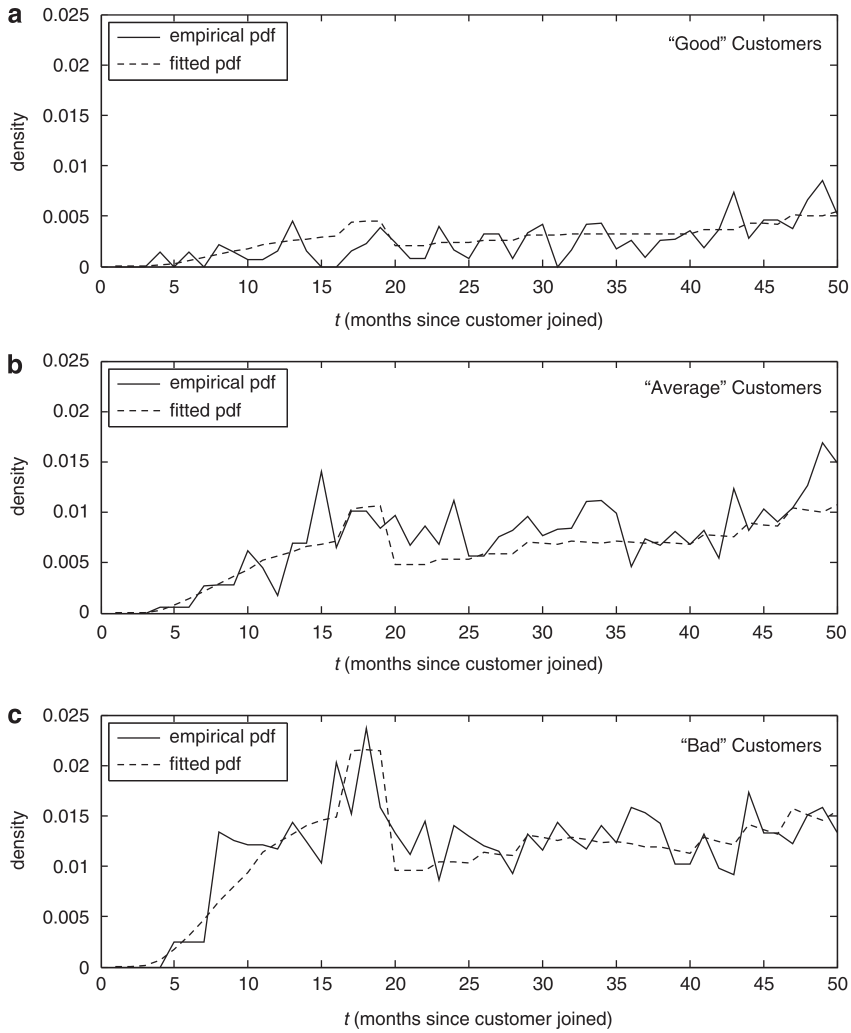

Figure 5 Comparison of empirical distributions of T *versus* pdfs from the fitted TDPH model for the Q2 2004 vintage and three different segments based on $\psi(\mathbf{x})$ values.

the fitted pdfs are quite rough. This is entirely because of the $\gamma(\tau+t)$ term in the hazard function $h(t, \mathbf{x}, \tau) = \psi(\mathbf{x})h_0(t)\gamma(\tau+t)$ for the TDPH model. If $\gamma(\tau+t)$ were constant for $t = 1, 2, \ldots$, then the corresponding fitted pdf would indeed appear much smoother (see Figure 6). The Q2 2004 vintage was chosen for Figure 5 because this is the earliest period that avoids the uncharacteristically low $\hat{\gamma}$ values due to initial conditions discussed above. The three segments of customers were defined based on $\psi(\mathbf{x})$ values for the customers: 'good' customers had $\psi(\mathbf{x})$ values between the 0.05 and 0.25 quantiles of all $\psi(\mathbf{x})$ values, 'average' customers had $\psi(\mathbf{x})$ values between the 0.40 and 0.60 quantiles, and 'bad' customers had $\psi(\mathbf{x})$ values between the 0.75 and 0.95 quantiles. For each empirical distribution, we excluded all customers who were censored before the end of the 50 month time period. All three empirical distributions in Figure 5 exhibit a spike between months 15 and 20, which corresponds to the Q4 2005 spike due to the bankruptcy law changes. The spike is clearly much more pronounced in the 'bad' customers, which is intuitively reasonable ('good' customers had much less inclination to rush to default before the new, stricter bankruptcy laws went into effect). Based on Figure 5 and similar figures shown later (as well as results for other vintages and segments that are not shown here for brevity), the fitted pdfs from the TDPH survival model appear reasonably close to the empirical distributions, indicating that the TDPH survival model provides a reasonably good fit to the data.

For the purpose of assessing the potential improvement in fit obtained by adding the γ function in the TDPH survival model, Figure 6 plots the same empirical distribution that is plotted in Figure 5(c) (for bad customers), and compares it with the fitted pdf from the same TDPH survival model but with γ held at a fixed value over the 50 month period. The fixed value of γ was taken to be the

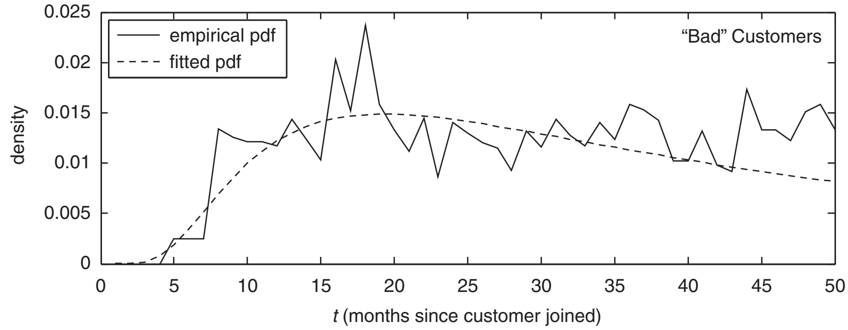

Figure 6 Comparison of the empirical distribution of T *versus* the pdf from the fitted TDPH model, but with γ held fixed. Compare with Figure 5(c).

average γ over the same 50 month period. Without the time-varying γ term, the model clearly cannot capture temporal phenomena. By comparing Figures 5(c) and 6, we see that including the time-varying γ substantially improves the ability of the TDPH model to represent the data.

To further assess how well the TDPH model fits the data, Figure 7 shows a different comparison of the empirical distributions *versus* the fitted pdfs. Figure 7 is similar to Figure 5, except that it shows three different vintages of customers (customers approved in Q1 2004, Q1 2005, and Q1 2006). All customers in all three plots in Figure 7 were 'average' customers, in the sense that they had similar $\psi(\mathbf{x})$ values within a narrow band around the 50th percentile. In the t axis, the plots extend until the final month of the data set (August, 2008). In Figures 7(a) and (b), the bankruptcy spike can be seen around months 23 and 11, respectively. The worsening economic conditions of 2008 can be noticed near the end of all three plots. Although the fitted pdf in Figure 7(b) appears to overestimate the probability of default, the overall fit from Figures 5 and 7 appears quite reasonable for this data set. Notice that all fitted pdfs shown in Figures 5 and 7 were obtained from the *same* TDPH survival model, a single model fitted to the entire data set.

5.2. Comparison in terms of credit risk scoring

In this section, we compare the ROC curves and related performance measures for four different methods of scoring customers in the test set based on 9-month default rates. As mentioned at the beginning of Section 5, this test set is independent of the training set that was used to fit the model. ROC curves are industry standard methods for comparing two or more scoring algorithms (Thomas *et al*, 2004). To assess the statistical significance of the results, for each pair of methods, we calculate bootstrapped 95% confidence intervals for the difference in their areas under ROC curve (AUC) and the difference in their Kolmogorov–Smirnov (KS) statistics. The AUC for a particular method is simply the area under the ROC curve for the method. In the context of credit scoring, the KS statistic for a particular method is the KS distance between two distributions, the first being the distribution of scores for all good customers, and the second being the distribution of scores for all bad customers (Thomas *et al*, 2002). In general, the larger the AUC and the KS statistic for a scoring method, the better the method.

The four models are the TDPH survival model, the standard PH survival model, a standard LR model, and an LR model with the TDPH γ. Regarding the latter, we fit a standard LR model, but included as an additional predictor variable (denoted by x_{11}) the value of $\hat{\gamma}$ estimated when fitting the TDPH model. More specifically, for a customer joining in month τ, the value assigned to x_{11} for that customer is the average $\hat{\gamma}$ value for the 9 following months from Figure 3. In practice, scoring a new customer using the TDPH model or the LR model with the TDPH γ would involve the forecast of the near-future γ values, whereas here we use the actual estimated $\hat{\gamma}$ values. As will be discussed in Section 5.4, because γ changes relatively smoothly for the most part (aside from unusual events like the spike due to bankruptcy law changes), reasonably accurate extrapolation into the near-future is not infeasible.

For each of the four methods, the score assigned to a customer is the predicted probability of 9-month default for each method. In the LR models, 9-month default probability is modelled directly. In the TDPH and PH methods, the default distribution is modelled, and the 9-month default probability for a customer is taken as the integral of the fitted pdf through month $t = 9$. A higher score translates to higher risk.

Figure 8 compares the ROC curves for the four different scoring methods. Each curve is for all customers within the entire window of data, excluding customers who were censored before reaching 9 months. Each point on an ROC curve is associated with a particular score: The horizontal axis value is the fraction of good customers (ie, those who did not default within 9 months) who scored below that score, and the vertical axis is the fraction of bad customers (ie, those who did default within 9

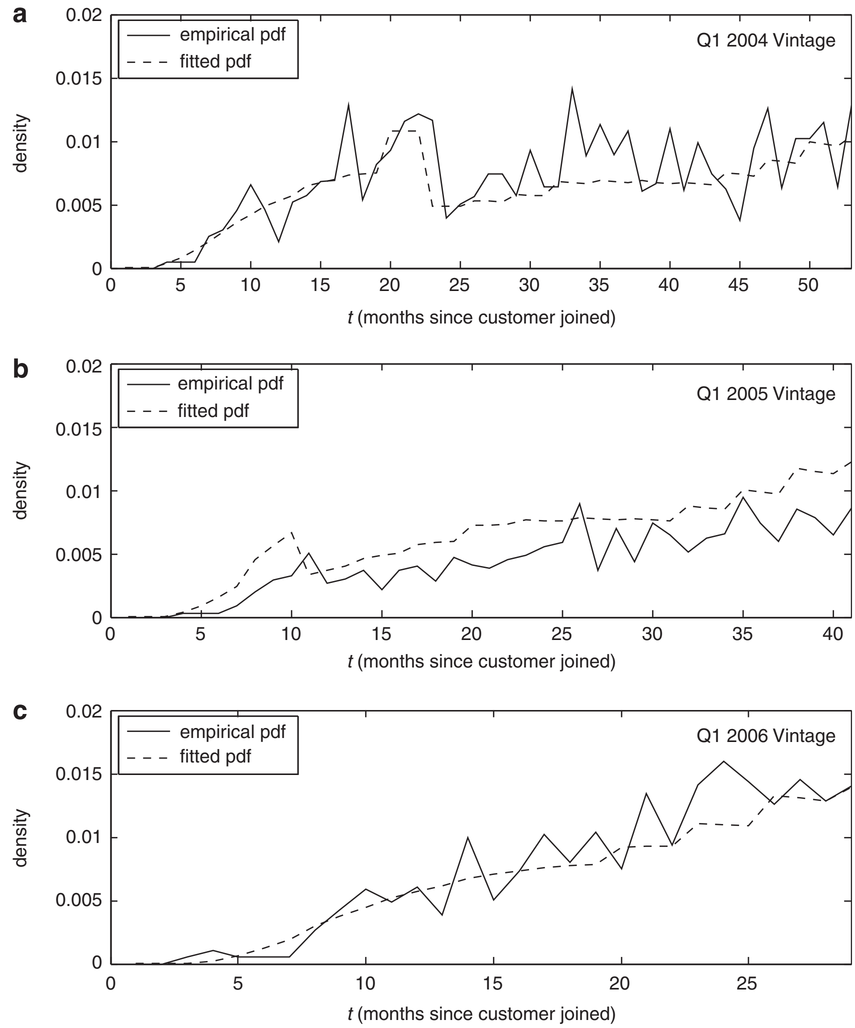

Figure 7 Comparison of empirical distributions of T *versus* pdfs from the fitted TDPH model for three different vintages of customers, all with similar (average) $\psi(\mathbf{x})$ values.

months) who scored below that score. The ROC curve for the randomized classification rule in Figure 8, which is a straight line with ordinate values equal to abscissa values, is included as a reference. Tables 1 and 2 show bootstrapped 95% confidence intervals for the difference in the AUC and the KS statistic for each pair of methods.

While all four methods have somewhat similar performance in terms of their ROC curves, the TDPH model does perform better than the LR model in terms of the KS statistics. The similar performance of LR *versus* the standard PH method is consistent with what Stepanova and Thomas (2002) observed. While LR with TDPH γ is significantly better than all other methods in terms of AUC differences (ie, their confidence intervals do not include 0), in terms of their KS statistics, the only pair with a statistically significant difference is TDPH *versus* LR.

It is interesting to note that the most effective method in Figure 8 was LR with the TDPH γ as an additional predictor. Enhancing LR with the TDPH γ clearly improves upon the performance of LR. Considering that LR and LR with TDPH γ were designed specifically for 9-month default, which gives them an inherent advantage over PH and TDPH under these performance measures, it is quite significant that TDPH performs as good as LR. Thus, inclusion of a time-dependency factor via TDPH modelling appears to have potential benefit for the objective of scoring. Moreover, there are other benefits of TDPH modelling, relative to LR modelling, that relate to the dynamic characteristic either inherent to PH (Banasik *et al*, 1999) or captured by the additional TDPH γ term. This benefit is most significant when economic conditions change dramatically, for example, during credit crises. We can see this more clearly by repeating the ROC, AUC, and KS statistic comparisons, but including only two short vintage windows with very different γ levels when constructing the ROC curves (the models were

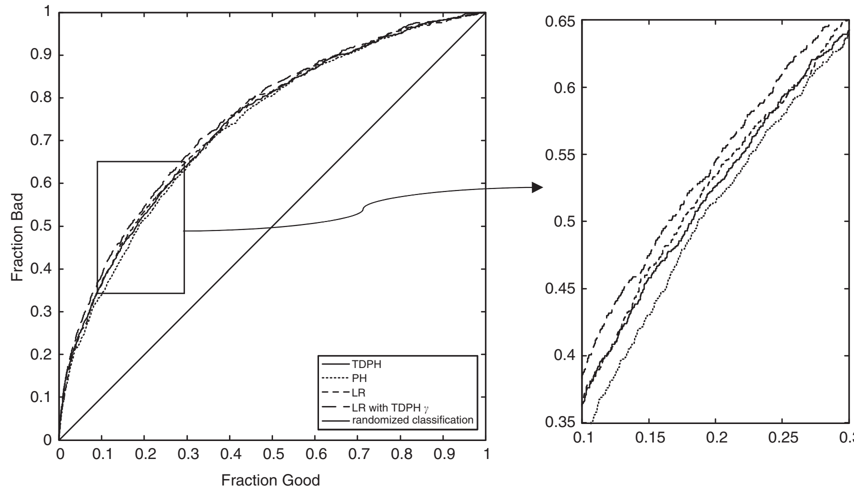

Figure 8 ROC curves over a time window that includes all customers in the test data set.

Table 1 95% confidence intervals for the AUC difference for each pair of methods (row method minus column method) over the entire data window, corresponding to Figure 8

	LR with TDPH γ	*PH*	*LR*
TDPH	$[-0.0153, -0.0108, -0.0065]$	$[0.0027, 0.0074, 0.0120]$	$[-0.0069, -0.0010, 0.0046]$
LR with TDPH γ		$[0.0121, 0.0183, 0.0246]$	$[0.0051, 0.0098, 0.0143]$
PH			$[-0.0123, -0.0085, -0.0048]$

The three quantities within the brackets are the confidence interval lower bound, the median bootstrapped AUC difference, and the confidence interval upper bound.

Table 2 95% confidence intervals for the KS difference for each pair of methods (row method minus column method) over the entire data window, corresponding to Figure 8

	LR with TDPH γ	*PH*	*LR*
TDPH	$[-0.0102, 0.0046, 0.0191]$	$[-0.0040, 0.0091, 0.0224]$	$[0.0011, 0.0173, 0.0336]$
LR with TDPH γ		$[-0.0131, 0.0047, 0.0220]$	$[-0.0011, 0.0127, 0.0273]$
PH			$[-0.0060, 0.0081, 0.0220]$

The three quantities within the brackets are the confidence interval lower bound, the median bootstrapped KS difference, and the confidence interval upper bound.

still fit using the entire set of customers, though). The two vintages were customers who joined in Q4 2004 and in Q3 2007. We chose these two quarters because their respective γ levels over the 9 months subsequent to their joining are quite different (see Figure 3). Hence, there is more potential for a time-dependency factor to improve the relative scoring of customers. Figure 9 and Tables 3 and 4 compare ROC curves, AUC and KS statistics over the two short vintage windows. In this case, both methods that incorporate the time-dependency factor (TDPH and LR with TDPH γ) have significantly better performance than their counterparts that do not take into account time dependency. Although many of the KS differences in Table 2 are not statistically significant (ie, their confidence interval includes 0), many in Table 4 are. For example, LR with TDPH γ shows statistically significant improvements over standard LR. The median KS improvement is approximately 3.2%, which is considered in the industry to be quite practically significant.

5.3. Some implementation issues

Of the many baseline distributions that one might consider, we have investigated both a Weibull and a lognormal distribution. We have focused on results for a lognormal distribution, because it appears to fit our data better (for reasons discussed later in this section). In addition to how well the distributions fit the data, there are other considerations in choosing a baseline distribution. One advantage of the Weibull baseline distribution is

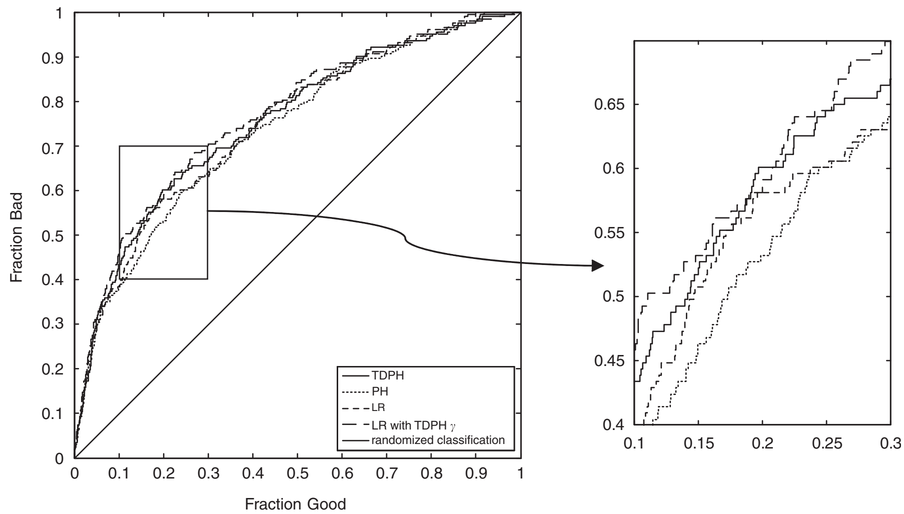

Figure 9 ROC curves over two short vintage windows. Incorporating the time-dependency factor improves scoring performance.

Table 3 95% confidence intervals for the AUC difference for each pair of methods (row method minus column method) over two short vintage windows, corresponding to Figure 9

	LR with TDPH γ	*PH*	*LR*
TDPH	$[-0.0205, -0.0118, -0.0025]$	$[0.0079, 0.0185, 0.0289]$	$[-0.0062, 0.0068, 0.0199]$
LR with TDPH γ		$[0.0149, 0.0304, 0.0442]$	$[0.0065, 0.0185, 0.0302]$
PH			$[-0.0211, -0.0118, -0.0021]$

The three quantities within the brackets are the confidence interval lower bound, the median bootstrapped AUC difference, and the confidence interval upper bound.

Table 4 95% confidence intervals for the KS difference for each pair of methods (row method minus column method) over two short vintage windows, corresponding to Figure 9

	LR with TDPH γ	*PH*	*LR*
TDPH	$[-0.0438, -0.0147, 0.0144]$	$[0.0093, 0.0443, 0.0776]$	$[-0.0170, 0.0185, 0.0578]$
LR with TDPH γ		$[0.0207, 0.0582, 0.0967]$	$[0.0023, 0.0323, 0.0701]$
PH			$[-0.0529, -0.0250, 0.0047]$

The three quantities within the brackets are the confidence interval lower bound, the median bootstrapped KS difference, and the confidence interval upper bound.

computational expense: We have found that the computational expense required to evaluate the gradient of the log-likelihood (within the gradient-based MLE algorithm) is much lower when we use a Weibull distribution instead of a lognormal distribution.

Another advantage of the Weibull distribution is that the likelihood function is less sensitive to certain changes in the distribution parameters, making the optimization algorithm more computationally stable. In contrast, for a lognormal baseline distribution, the likelihood function is quite sensitive to certain changes in the standard deviation parameter σ. Figure 10 illustrates why by showing two different pdfs for T with a lognormal baseline distribution.

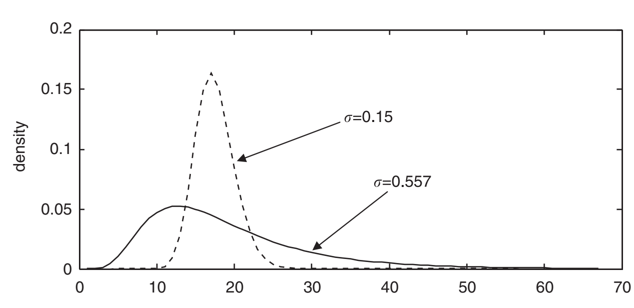

Figure 10 Two pdfs of T for a lognormal baseline distribution with all parameters common except for σ.

Both pdfs have the same predictor (**x**) values and use a constant γ function. The only difference is that two different values of σ are considered ($\sigma = 0.557$, which was the fitted value using our entire data set, and $\sigma = 0.15$). For the $\sigma = 0.15$ case, the probability of default within the first 3 months is virtually zero. However, albeit quite rare, there do exist a few customers with similar predictor values who defaulted within the first three months. The likelihood and its gradient become essentially zero for these customers if a value such as $\sigma = 0.15$ is tried during the optimization search. Similarly, if working with the log-likelihood, very large negative numbers that exceed the internal limits of the software may occur. This can be avoided if one writes their own optimization code and includes checks to avoid calculating the complete likelihood for choice of parameters for which it is virtually zero. But if one uses commercial optimization software, one may have little control over the parameter values that are considered within a gradient-based line search. Perhaps the simplest way to avoid such problems is impose a lower bound on σ as a constraint. In our algorithms we have represented σ via $\sigma = \exp(s) + 0.2$ and optimized over s, which results in a lower bound of 0.2 for σ.

In spite of involving numerical issues that require more care, a lognormal baseline distribution appears to fit the data much better than a Weibull distribution, in particular over the early months. Figure 11 shows a typical fitted pdf of T using a Weibull baseline distribution. Notice that the shape of the Weibull distribution assigns much higher probability to the early months (eg, $t<4$) than would a lognormal distribution with similar first two moments. By inspection of the empirical distributions in Figures 5–7, it is evident that in reality there are very few defaults in the early months, which is much more consistent with the shape of a lognormal distribution than with a Weibull distribution. Hence, the lognormal provided a better fit to the data. An additional consequence of the lack of fit in the early months for a Weibull baseline distribution relates to the underestimation of γ over the first few months of the data (see Figure 3). Figure 3 used a lognormal baseline distribution, but the underestimation of γ in the early months was even more extreme when a Weibull baseline distribution was used.

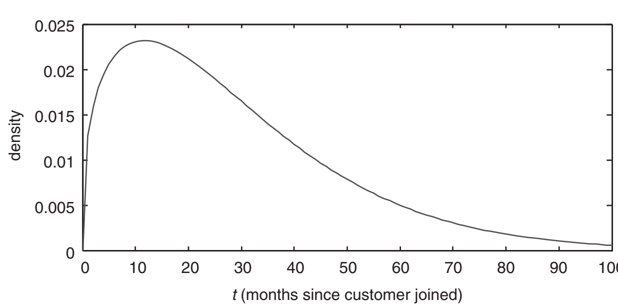

Figure 11 pdf of T for a Weibull baseline distribution.

5.4. Scoring new customers

An additional benefit of having a time-dependency factor in the model is that it provides an inherent mechanism for adjusting the customer acceptance threshold in a manner that attempts to control the collective default rate for new customers under changing market conditions. For example, suppose one uses a TDPH model to score applicants and that an acceptance threshold was set in Q1 2007 that gave a desired collective default rate at that time. If we use the same acceptance threshold in Q4 2008 (refer to Figure 3), then we achieve the same collective default rate in spite of the worsened economic conditions. This is because the higher γ value in Q4 2008 is already incorporated into the applicants' scoring. If one were using standard LR, one would have to tighten the acceptance criterion to keep the overall bad rate the same as it was in 2007, and it may not be at all clear how much to tighten it. The TDPH model provides an inherent, consistent means of accomplishing this.

The rationale for this strategy involves the assumption that future γ values after the customer joins remain close to the γ value when the customer applied or can be reasonably extrapolated into the near future. There is no guarantee that this will be the case, of course, especially in the present context of dynamic market conditions. However, it is reasonable to expect that future γ values will be closer to the most recent value than to some average value over the past years. For example, consider a customer who applies at the end of Q3 2008, at which time $\gamma = 0.18$ roughly (refer to Figure 3). At that point in time, it would have been reasonable to expect that the γ values for Q4 2008 and beyond would be closer to 0.18 than to the historical average of roughly 0.1.

One could also take this a step further and attempt to extrapolate γ into the near future. Again considering a customer who applies at the end of Q3 2008, it appears from Figure 3 that γ is increasing at that time. Hence, one might extrapolate values for γ that continue to increase over 2009, perhaps somewhere between 0.2 and 0.25. Figure 12 illustrates possible extrapolations that could be used for the purpose of scoring new applicants or predicting future performance of existing customers.

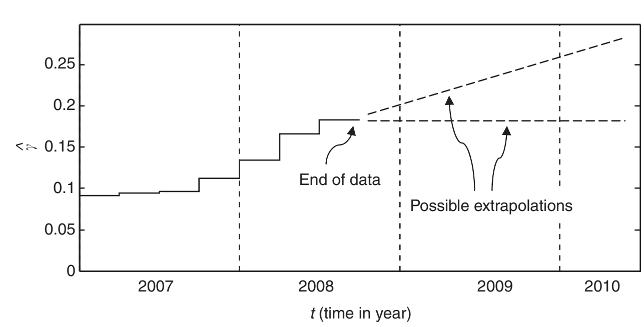

Figure 12 Possible extrapolations of γ.

5.5. Other uses

One desirable consequence of fitting a TDPH model is that it produces an estimate of the time-dependency function γ, which can be viewed as the net effects of overall market conditions, as they impact default behaviour. This can be useful for explanatory analyses, for helping to identify and quantify unusual events and trends that affect default, such as the bankruptcy law spike in Q4 2005 or the economic meltdown of 2008.

The estimated $\hat{\gamma}$ function can also be incorporated as a single additional predictor variable in many other methods for modelling credit risk. For example, this was done in the LR with TDPH γ model considered in Section 5.2. In fact, in terms of ROC performance as a credit risk scoring tool, the LR with TDPH γ model performed the best of all the models we considered. Another potential way to utilize the $\hat{\gamma}$ function relates to the approaches of Stepanova and Thomas (2002) and Tang *et al* (2007), who included interactions between the predictors and time (of the form tx_j) in the PH model. Their intent was to represent changes in the effects of certain predictor variables that become more or less pronounced the longer one has been a customer. Similarly, one might consider including interaction terms of the form $(t+\tau)x_j$ to represent changes in the effects of predictors due to dynamic market conditions, which would be equivalent to assuming the coefficients $\boldsymbol{\beta}$ experience linear trends over time. Alternatively, one might include interactions of the form $\gamma(t+\tau)x_j$. This would allow greater flexibility, in the sense of capturing transient and cyclical phenomena that are not linear trends, without increasing the number of parameters that must be estimated.

6. Conclusions

We have developed a TDPH survival model for modelling default behaviour of credit card customers in a manner that explicitly accounts for trends and short-lived phenomena in dynamic market conditions. We derived the MLEs and a gradient-based algorithm for estimating all parameters of the model. Using a very large set of real data from a major credit card company, we implemented the TDPH modelling approach to assess its effectiveness in representing the time-to-default distribution under dynamic market conditions and in scoring customers for credit risk. Our results indicate that the TDPH model fits the data well. As a scoring method that takes into account dynamic market conditions, the LR model using the TDPH γ as an additional predictor achieved roughly a 3.2% improvement in KS statistic over regular LR when we considered two different vintage windows of data under quite different market conditions. This result is consistent with the results of Tang *et al* (2007) and Bellotti and Crook (2009), who found modest but statistically significant improvements in predictive performance using macroeconomic variables with substantial dynamic variability. Overall, incorporating the TDPH γ into either the LR or the PH approaches improves the performance of these methods.

Additional benefits of the TDPH approach relate to the dynamic characteristics that the model captures: It provides an inherent mechanism for adjusting the customer acceptance threshold to keep constant the collective default rate of accepted customers in the face of dynamic market conditions. Given an acceptance threshold that was used in the past and the default rate that resulted, we can achieve the same collective default rate on average regardless of the changes in the economic conditions. Moreover, the TDPH model produces an estimate (the γ function) of the net effect of dynamic market conditions on default behaviour. The estimated γ function can facilitate explanatory data analyses or be incorporated into other modelling methods as an additional predictor variable.

Acknowledgements—The authors gratefully acknowledge numerous helpful comments from two anonymous referees and the editor.

References

Andreeva G, Ansell J and Crook JN (2007). Modeling profitability using survival combination scores. *Eur J Opl Res* **183**: 1537–1549.

Baesens B, Van Gestel T, Stepanova M, Van den Poel D and Vanthienen J (2005). Neural network survival analysis for personal loan data. *J Opl Res Soc* **56**: 1089–1098.

Baesens B, Van Gestel T, Viaene S, Stepanova M, Suykens J and Vanthienen J (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *J Opl Res Soc* **54**: 627–635.

Banasik J, Crook JN and Thomas LC (1999). Not if but when will borrowers default. *J Opl Res Soc* **50**: 1185–1190.

Batista GE, Prati RC and Monard MC (2004). A study of the behavior of several methods for balancing machine learning training data. *Sigkdd Explor* **6**: 20–29.

Bellotti T and Crook JN (2009). Credit scoring with macroeconomic variables using survival analysis. *J Opl Res Soc* **60**: 1699–1707.

Breeden JL (2007). Modeling data with multiple time dimensions. *Comp Stat Data Anal* **51**: 4761–4785.

Breeden JL and Thomas LC (2008). The relationship between default and economic cycle across countries for retail portfolios. *J Risk Model Validation* **2**(3): 11–44.

Breeden JL, Thomas LC and McDonald JW (2008). Stress-testing retail loan portfolios with dual-time dynamics. *J Risk Model Validation* **2**(2): 43–62.

Chawla NV, Japkowicz N and Kolcz A (2004). Editorial: Special issue on learning from imbalanced data sets. *Sigkdd Explor* **6**: 1–6.

Crook JN, Edelman DB and Thomas LC (2007). Recent developments in consumer credit risk management. *Eur J Opl Res* **183**: 1447–1465.

Hand DJ (2006). Classifier technology and the illusion of progress. *Stat Sci* **21**: 1–14.

Hand DJ and Henley WE (1997). Statistical classification methods in consumer credit scoring: A review. *J R Stat Soc Ser A* **160**: 523–541.

Leemis LM (1995). *Reliability: Probabilistic Models and Statistical Methods*. Prentice-Hall: Englewood Cliffs, NJ.
Meeker WQ and Escobar LA (1998). *Statistical Methods for Reliability Data*. Wiley-Interscience: New York.
Rosenberg E and Gleit A (1994). Quantitative methods in credit management: A survey. *Opns Res* **42**: 589–613.
Stepanova M and Thomas L C (2001). PHAB scores: Proportional hazards analysis behavioural scores. *J Opl Res Soc* **52**: 1007–1016.
Stepanova M and Thomas LC (2002). Survival analysis methods for personal loan data. *Opns Res* **50**: 277–289.
Tang L, Thomas LC, Thomas S and Bozzeto J-F (2007). It's the economy stupid: Modelling financial product purchases. *Int J Bank Market* **25**: 22–38.
Thomas LC (2000). A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers. *Int J Forecast* **16**: 149–172.
Thomas LC, Edelman DB and Crook JN (2002). *Credit Scoring and its Applications*. SIAM: Philadelphia, PA.
Thomas LC, Edelman DB and Crook JN (2004). *Readings in Credit Scoring: Recent Developments, Advances, and Aims*. Oxford University Press: Oxford, UK.
Thomas LC, Oliver RW and Hand DJ (2005). A survey of the issues in consumer credit modelling research. *J Opl Res Soc* **56**: 1006–1015.

Appendix A

Some data treatment issues

The data are described in Section 5. Ten predictor variables were chosen to include in the model, but due to confidentiality concerns, we do not describe them in this paper. We recommend standardizing all variables before implementing the MLE algorithm, for numerical reasons. Scaling of the variables will affect the direction in which a gradient-based algorithm searches for solutions to the parameter estimates. Regarding missing data, which are inevitable in large data sets, we used the following strategy. Predictor variable values that were missing for a customer were substituted with a linear regression prediction of the missing value, based on a covariance matrix calculated from all data that were not missing.

Appendix B

Derivation of the TDPH cdf

As mentioned in Section 4, the TDPH cdf of T can be written as $F(t;\mathbf{x},\tau;h_0,\psi,\gamma)=1-\exp\left(-\int_0^t h_0(u)\psi(\mathbf{x})\gamma(\tau+u)du\right)$. We break the integral up into components that correspond to the pieces over which γ is constant. Let $w(\tau,t)=\lceil(\tau+t)/3\rceil-\lceil\tau/3\rceil+1$ denote the number of different γ pieces between the time of approval (τ) and the time in question ($\tau+t$). The '3' in the denominator arises from the fact that t is in units of months, and the γ function is modelled as constant over each quarter. For a customer joining in month τ, the intervals (in terms of the relative month t) over which γ is constant can be written as

$$I_{\tau,1}=[s_{\tau,0},\,s_{\tau,1}]=\left[0,\,3\left\lceil\frac{\tau}{3}\right\rceil-\tau\right],$$
$$I_{\tau,2}=[s_{\tau,1},\,s_{\tau,2}]=\left[3\left\lceil\frac{\tau}{3}\right\rceil-\tau,\,3\left\lceil\frac{\tau}{3}\right\rceil-\tau+3\right]$$
$$\vdots$$
$$\begin{aligned}I_{\tau,w(\tau,t)-1}&=\left[s_{\tau,w(\tau,t)-2},\,s_{\tau,w(\tau,t)-1}\right]\\&=\left[3\left\lceil\frac{\tau+t}{3}\right\rceil-\tau-6,\,3\left\lceil\frac{\tau+t}{3}\right\rceil-\tau-3\right],\end{aligned}$$
$$\begin{aligned}I_{\tau,w(\tau,t)}&=\left[s_{\tau,w(\tau,t)-1},\,s_{\tau,w(\tau,t)}\right]\\&=\left[3\left\lceil\frac{\tau+t}{3}\right\rceil-\tau-3,\,t\right].\end{aligned}$$

Here $\lceil\cdot\rceil$ denotes the ceiling function returning the smallest integer not less than the argument. In the preceding expressions for the intervals, it is understood that the lower boundary of each is truncated at 0 and the upper boundary at t. The constant value of $\gamma(\cdot)$ over the interval $I_{\tau,k}$ is $\gamma_{\lceil\tau/3\rceil+k-1}$. Note that $s_{\tau,0}=0$ and $s_{\tau,w(\tau,t)}=t$ by definition. The cdf $F(t;\mathbf{x},\tau;h_0,\psi,\gamma)$ of T becomes

$$\begin{aligned}&F(t;\,\mathbf{x},\,\tau;\,h_0,\,\psi,\,\gamma)\\&=1-\exp\left(-\sum_{k=1}^{w(\tau,t)}\int_{s_{\tau,k-1}}^{s_{\tau,k}}h_0(u)\psi(\mathbf{x})\gamma_{\lceil\tau/3\rceil+k-1}du\right)\\&=1-\exp\left(-\psi(\mathbf{x})\sum_{k=1}^{w(\tau,t)}\int_{s_{\tau,k-1}}^{s_{\tau,k}}h_0(u)\gamma_{\lceil\tau/3\rceil+k-1}du\right)\\&=1-\exp\left(-\psi(\mathbf{x})\sum_{k=1}^{w(\tau,t)}\gamma_{\lceil\tau/3\rceil+k-1}\left(H_0(s_{\tau,k})-H_0(s_{\tau,k-1})\right)\right)\\&=1-\exp\left(-\psi(\mathbf{x})\sum_{k=1}^{w(\tau,t)}\gamma_{\lceil\tau/3\rceil+k-1}\left(-\log\left(1-F_0(s_{\tau,k})\right)\right.\right.\\&\qquad\left.\left.+\log\left(1-F_0(s_{\tau,k-1})\right)\right)\right),\end{aligned}$$

where $H_0(t)=\int_0^t h_0(u)du=-\log(1-F_0(t))$ is the cumulative hazard function of the baseline distribution.

Appendix C

The log-likelihood and its gradient for lognormal $h_0(t)$ and exponential $\psi(x)$

In this section, we derive the log-likelihood function and its gradient (for use in a gradient-based algorithm for maximizing the likelihood) for the special case of a lognormal $h_0(t)$ and an exponential $\psi(\mathbf{x})$. For other baseline distributions or $\psi(\mathbf{x})$ functions, a similar approach can be used. For our special case, the baseline cdf and pdf are $F_0(t) = \Phi((\log(t)-\mu)/\sigma)$ and $f_0(t) = \phi((\log(t)-\mu)/\sigma)/(t\sigma)$ with mean parameter μ and standard deviation parameter σ, where Φ and ϕ denote the standard normal cdf and pdf. In order to avoid identifiability problems due to the confounding between e^{β_0} and γ, we define the exponential function as $\psi(\mathbf{x}) = e^{\boldsymbol{\beta}'\mathbf{x}}$, instead of $e^{\beta_0+\boldsymbol{\beta}'\mathbf{x}}$. For notational simplicity, we denote $F(t; \mathbf{x}_i, \tau_i; \mu, \sigma, \boldsymbol{\beta}, \boldsymbol{\gamma})$ by $F_i(t)$. From Section 4, the log-likelihood function for the entire data set is

$$\begin{aligned} l(\mu, \sigma, \boldsymbol{\beta}, \boldsymbol{\gamma} | \mathbf{t}, \mathbf{X}, \boldsymbol{\tau}, \boldsymbol{\delta}) &= \sum_{i=1}^{N} (\delta_i \log(F_i(t_i) - F_i(t_i - 1)) \\ &\quad + (1-\delta_i)\log(1 - F_i(t_i))), \end{aligned} \tag{C.1}$$

where expressions for the $F_i(t)$ are given in Appendix B.

To find the partial derivatives of l with respect to μ, σ, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$, notice that l depends on μ, σ, and $\boldsymbol{\gamma}$ via the terms $J_i(t) = \sum_{k=1}^{w(\tau_i,t)} \gamma_{\lceil \tau_i/3 \rceil + k - 1}(-\log(1 - F_0(s_{\tau_i,k})) + \log(1 - F_0(s_{\tau_i,k-1})))$ $(i = 1, 2, \ldots, N)$, and l depends on $\boldsymbol{\beta}$ via the terms $\psi(\mathbf{x}_i)$ $(i = 1, 2, \ldots, N)$. The relevant partial derivatives of $J_i(t)$ and $\psi(\mathbf{x}_i)$ are obtained from

$$\begin{aligned} J_i(t) &= \sum_{k=1}^{w(\tau_i,t)} \gamma_{\lceil \tau_i/3 \rceil + k - 1}(-\log(1 - F_0(s_{\tau_i,k})) \\ &\quad + \log(1 - F_0(s_{\tau_i,k-1}))) \\ &= \sum_{k=1}^{w(\tau_i,t)} \gamma_{\lceil \tau_i/3 \rceil + k - 1}\left(-\log\left(1 - \Phi\left(\frac{\log(s_{\tau_i,k}) - \mu}{\sigma}\right)\right)\right. \\ &\quad \left. + \log\left(1 - \Phi\left(\frac{\log(s_{\tau_i,k-1}) - \mu}{\sigma}\right)\right)\right), \end{aligned}$$

$$\begin{aligned} \frac{\partial J_i(t)}{\partial \mu} &= \sum_{j=1}^{w(\tau_i,t)} \frac{\gamma_{\lceil \tau_i/3 \rceil + k - 1}}{\sigma}\left(-\frac{\phi\left(\frac{\log(s_{\tau_i,k}) - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{\log(s_{\tau_i,k}) - \mu}{\sigma}\right)}\right. \\ &\quad \left. + \frac{\phi\left(\frac{\log(s_{\tau_i,k-1}) - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{\log(s_{\tau_i,k-1}) - \mu}{\sigma}\right)}\right), \end{aligned}$$

$$\begin{aligned} \frac{\partial J_i(t)}{\partial \sigma} &= \sum_{j=1}^{w(\tau_i,t)} \frac{\gamma_{\lceil \tau_i/3 \rceil + k - 1}}{\sigma^2} \\ &\quad \times \left(-\frac{(\log(s_{\tau_i,k}) - \mu)\phi\left(\frac{\log(s_{\tau_i,k}) - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{\log(s_{\tau_i,k}) - \mu}{\sigma}\right)}\right. \\ &\quad \left. + \frac{(\log(s_{\tau_i,k-1}) - \mu)\phi\left(\frac{\log(s_{\tau_i,k-1}) - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{\log(s_{\tau_i,k-1}) - \mu}{\sigma}\right)}\right), \end{aligned}$$

$$\frac{\partial J_i(t)}{\partial \gamma_{\lceil \tau_i/3 \rceil + k - 1}} = \begin{cases} -\log\left(1 - \Phi\left(\frac{\log(s_{\tau_i,k}) - \mu}{\sigma}\right)\right) \\ \quad + \log\left(1 - \Phi\left(\frac{\log(s_{\tau_i,k-1}) - \mu}{\sigma}\right)\right): \\ \quad k = 1, 2, \ldots, w(\tau_i, t), \\ 0 \text{ elsewhere} \end{cases}$$

,

$$\psi(\mathbf{x}_i) = \exp(\boldsymbol{\beta}'\mathbf{x}_i), \quad \text{and}$$

$$\frac{\partial \psi(\mathbf{x}_i)}{\partial \beta_j} = x_{i,j}\exp(\boldsymbol{\beta}'\mathbf{x}_i)$$

From the expressions in Appendix B, the partial derivatives of $F_i(t)$ are obtained from

$$F_i(t) = 1 - \exp(-\psi(\mathbf{x}_i)J_i(t)),$$

$$\frac{\partial F_i(t)}{\partial \mu} = \exp(-\psi(\mathbf{x}_i)J_i(t))\psi(\mathbf{x}_i)\frac{\partial J_i(t)}{\partial \mu},$$

$$\frac{\partial F_i(t)}{\partial \sigma} = \exp(-\psi(\mathbf{x}_i)J_i(t))\psi(\mathbf{x}_i)\frac{\partial J_i(t)}{\partial \sigma},$$

$$\frac{\partial F_i(t)}{\partial \beta_j} = \exp(-\psi(\mathbf{x}_i)J_i(t))\frac{\partial \psi(\mathbf{x}_i)}{\partial \beta_j}J_i(t) : j = 1, 2, \ldots, M, \text{ and}$$

$$\frac{\partial F_i(t)}{\partial \gamma_q} = \exp(-\psi(\mathbf{x}_i)J_i(t))\psi(\mathbf{x}_i)\frac{\partial J_i(t)}{\partial \gamma_q} : q = 1, 2, \ldots, Q.$$

From Equation (C.1), we also have

$$\begin{aligned} \frac{\partial l}{\partial \mu} &= \sum_{i=1}^{N}\left[\delta_i(F_i(t_i) - F_i(t_i - 1))^{-1}\right. \\ &\quad \times \left(\frac{\partial F_i(t_i)}{\partial \mu} - \frac{\partial F_i(t_i - 1)}{\partial \mu}\right) \\ &\quad \left. -(1-\delta_i)(1 - F_i(t_i))^{-1}\frac{\partial F_i(t_i)}{\partial \mu}\right], \end{aligned}$$

$$\begin{aligned} \frac{\partial l}{\partial \sigma} &= \sum_{i=1}^{N}\left[\delta_i(F_i(t_i) - F_i(t_i - 1))^{-1}\right. \\ &\quad \times \left(\frac{\partial F_i(t_i)}{\partial \sigma} - \frac{\partial F_i(t_i - 1)}{\partial \sigma}\right) \\ &\quad \left. -(1-\delta_i)(1 - F_i(t_i))^{-1}\frac{\partial F_i(t_i)}{\partial \sigma}\right], \end{aligned}$$

$$\begin{aligned} \frac{\partial l}{\partial \beta_j} &= \sum_{i=1}^{N}\left[\delta_i(F_i(t_i) - F_i(t_i - 1))^{-1}\right. \\ &\quad \times \left(\frac{\partial F_i(t_i)}{\partial \beta_j} - \frac{\partial F_i(t_i - 1)}{\partial \beta_j}\right) \\ &\quad \left. -(1-\delta_i)(1 - F_i(t_i))^{-1}\frac{\partial F_i(t_i)}{\partial \beta_j}\right] \end{aligned}$$

for $j=0,1,2,\ldots,M$, and

$$\frac{\partial l}{\partial \gamma_q}=\sum_{i=1}^{N}\left[\delta_i(F_i(t_i)-F_i(t_i-1))^{-1}\times\left(\frac{\partial F_i(t_i)}{\partial \gamma_q}-\frac{\partial F_i(t_i-1)}{\partial \gamma_q}\right)-(1-\delta_i)(1-F_i(t_i))^{-1}\frac{\partial F_i(t_i)}{\partial \gamma_q}\right]$$

for $q=1,2,\ldots,Q$.

Combining all of these gives the partial derivatives that constitute the gradient of l with respect to the parameters. The gradient can be used in an optimization algorithm for calculating the MLEs of the parameters. Matlab code is available upon request from the authors for the special case considered in this appendix.

Received March 2010;
accepted February 2011 after one revision