This is a pre-copyedited, author-produced PDF of an article accepted for publication in Journal of Financial Econometrics following peer review. The version of record David Lando, Mamdouh Medhat, Mads Stenbo Nielsen, and Søren Feodor Nielsen Additive Intensity Regression Models in Corporate Default Analysis. Journal of Financial Econometrics (Summer 2013) 11 (3): 443-485 is available online at doi: http://dx.doi.org/10.1093/jjfinec/nbs018

Additive Intensity Regression Models in Corporate Default Analysis*

David Lando†

Mamdouh Medhat

Mads Stenbo Nielsen

Søren Feodor Nielsen

This version: August 31, 2012

*All authors are at the Department of Finance, Copenhagen Business School.

†Corresponding author: David Lando, Solbjerg Plads 3, A4, DK-2000 Frederiksberg, Denmark, +45 3815 3613, dl.fi@cbs.dk.

Abstract

We consider additive intensity (Aalen) models as an alternative to the multiplicative intensity (Cox) models for analyzing the default risk of a sample of rated, non-financial U.S. firms. The setting allows for estimating and testing the significance of time-varying effects. We use a variety of model checking techniques to identify misspecifications. In our final model we find evidence of time-variation in the effects of distance-to-default and short-to-long term debt, we identify interactions between distance-to-default and other covariates, and the quick ratio covariate is significant. None of our macroeconomic covariates are significant.

JEL Classification: G32, G33, C41, C52

Keywords: Default risk modeling, Aalen's additive regression model, martingale residual processes

1 Introduction

Intensity regression models provide flexible and powerful tools for studying one of the most basic questions of credit risk modeling: Which observable variables influence the default risk of corporations? Consequently, the models are useful for risk management of loan portfolios by providing a strong statistical basis for credit scoring. They also play an important role in academic studies investigating risk premia on corporate bonds or credit default swaps. There, the purpose of the hazard regressions is to provide estimates of the "physical" or "real-world" default probabilities. Combining these with "implied" or "risk-neutral" default probabilities obtained from prices of financial instruments, we can measure the risk premium required by investors for assuming default risk.

The statistical analysis of default data dates back at least to Beaver (1966) and Altman (1968), but the use of survival analysis techniques (parametric and semiparametric Cox regressions and non-parametric methods) is more recent. Examples of this literature includes Shumway (2001), Lando and Skødeberg (2002), Fledelius, Lando, and Nielsen (2004), Couderc and Renault (2004), Das, Duffie, Kapadia, and Saita (2007), Duffie, Saita, and Wang (2007), Chava, Stefanescu, and Turnbull (2011), Duffie, Eckner, Horel, and Saita (2009), Lando and Nielsen (2010), Figlewski, Frydman, and Liang (2012).

The studies that employ regression models typically look at Cox models. Some of these studies use a non-parametric baseline intensity – others used a constant baseline. But all other parameters remain fixed over time. Furthermore, for reasons discussed later in the paper, there is often very little model checking after the insignificant explanatory variables have been eliminated.

In this paper, we use additive Aalen models as an alternative to the Cox model. Using both non-parametric and semi-parametric versions we are able to study not only whether explanatory variables are significant, but also whether their effects vary with time. Both graphical techniques and formal tests are employed. To allow for a comparison with a Cox regression study by Lando and Nielsen (2010), we use the exact same data set.

We find that both the additive structure and the use of time varying coefficients change the conclusions of Lando and Nielsen (2010) somewhat. Model checking leads us to identify outliers from the data and helps us resolve problems with model misspecification. We find evidence of time-variation in the effects of distance-to-default and short-to-long term debt, and we identify the effect of interactions between distance-to-default and two other covariates: the quick ratio and (log) pledgable assets. Before removing outlier, the quick ratio covariate is insignificant – this, however, changes after the removal of outliers. None of our macroeconomic covariates are significant which may indicate that their effects are captured through their influence on firm-specific covariates.

The flow of the paper is as follows: We first recall the specification of non-parametric and semi-parametric Aalen models. We then summarize the estimation and testing procedures used, and after a data review we set up a model using a time-varying baseline intensity and firm-specific variables only. This leads us to conclude that the effects of certain firm-specific covariates are time-varying and that apparent model misspecifications may be resolved by including two interaction terms. We then replace the time varying baseline intensity by a constant baseline and global covariates (including a trailing monthly default rate), and replace regression functions by constant parameters for those firm-specific covariates which did not have significantly time-varying effects. After describing and implementing our model checking procedure, we discover that we need to remove several outliers

from the data. We finish up by testing the revised model and looking at the model check once more, before we conclude.

2 General model setup

Intensity models of default focus on describing the default time(s) of a debt-issuing firm through a stochastic intensity process. Fix a probability space $(\Omega, \mathcal{F}, P)$ and a finite time horizon $[0, T]$. For a cohort of n firms, the default-history of firm i is summarized by a piecewise-constant, right-continuous counting process $(N_{it})_{t\in[0,T]}$ with jumps of size 1 at the firm's default times. The counting processes are assumed adapted to a common filtration $(\mathcal{F}_t)_{t\in[0,T]}$, corresponding to the flow of information. In the absolute continuous case, the default intensity of firm i is the non-negative, integrable, $(\mathcal{F}_t)$-predictable process $(\lambda_{it})_{t\in[0,T]}$ such that

$$M_{it} = N_{it} - \int_0^t \lambda_{is}\, ds$$

is an $(\mathcal{F}_t)$-local martingale. Intuitively,

$$\lambda_{it} = \lim_{h\to 0} \frac{1}{h} E\left(N_{i(t+h)} - N_{it} \,\middle|\, \mathcal{F}_t\right) = \lim_{h\to 0} \frac{1}{h} P\left(N_{i(t+h)} - N_{it} = 1 \,\middle|\, \mathcal{F}_t\right), \tag{2.1}$$

so λ_{it} is the $\mathcal{F}_t$-conditional mean arrival rate of default: Given $\mathcal{F}_t$ and survival up to time t, firm i's probability of default within $[t, t+h)$ is $\lambda_{it}h + o(h)$.

Usually, $(N_{it})_{t\in[0,T]}$ is a one-jump process, corresponding to a single default. In this case, $N_{it} = 1_{(\tau_i \le t)}$, where τ_i is the single stochastic default-time for firm i. However, ambiguous definitions of real-world defaults and the possibility of restructuring will cause some firms to have several registered defaults over time. We therefore allow the counting processes in this paper to take any nonnegative integer value.

In the following, we will write the intensity as

$$\lambda_{it} = Y_{it}\alpha_i(t), \tag{2.2}$$

where $(Y_{it})_{t\in[0,T]}$ is a left-continuous, $(\mathcal{F}_t)$-predictable "at-risk indicator process," taking the value 1 if firm i is at risk of defaulting just before time t, and 0 otherwise. Here, $\alpha_i(t)$ is the $(\mathcal{F}_t)$-predictable "pre-default" intensity that may depend on covariates and past events. With a slight abuse of language we will also refer to $\alpha_i(t)$ as the intensity of default of firm i.

In regression models, the variation in the intensities across firms is solely due to covariates. This means that $\alpha_i(t) = \alpha(t \,|\, \mathbf{x}_{it})$ for a common function α, specifying the functional form of dependency on a p-dimensional, locally bounded vector $\mathbf{x}_{it} = (x_{i1,t}, \ldots, x_{ip,t})^T$ of covariate values at time t for firm i. The covariates might be constant (industry classification, for example), but will in this paper always be time-varying. Some covariates will be specific to firm i, and some will be macroeconomic variables shared by all firms. The predictability condition on the intensity then boils down to predictability of covariates. In practice, this means that covariate values entering the models at time t are required to be known just before time t.

The focus of this paper is the specification of $\alpha(t \,|\, \mathbf{x}_{it})$, i.e. determining which firm-specific and macroeconomic variables are significant explanatory variables, and how well α describes the data.

2.1 Relative and excess survival regression: The Cox and Aalen models

The Cox model was introduced by Cox (1972) in a survival data setting and extended to the general counting process framework by Andersen and Gill (1982). In this model, the intensity for firm i as

$$\alpha(t \,|\, \mathbf{x}_{it}) = \alpha_0(t) \, \exp(\boldsymbol{\beta}^T \mathbf{x}_{it}),$$

where $\alpha_0(t)$ is a locally integrable baseline intensity, which is left unspecified, while the vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ of regression coefficients gives the time-constant effects of the covariates. The baseline intensity $\alpha_0(t)$ corresponds to the default intensity at time t when all covariates are identically equal to zero. The model thus assumes that all firm-specific intensities are proportional to the same baseline intensity.

Consider two time-t covariate vectors $\mathbf{x}_{1t}$ and $\mathbf{x}_{2t}$, and assume that these are identical except for the jth coordinate, where $x_{2j,t} = x_{1j,t} + 1$. Forming the ratio of the intensities then gives

$$\frac{\alpha(t \,|\, \mathbf{x}_{2t})}{\alpha(t \,|\, \mathbf{x}_{1t})} = \exp\left(\boldsymbol{\beta}^T (\mathbf{x}_{2t} - \mathbf{x}_{1t})\right) = e^{\beta_j},$$

so the effect at time t of a one-unit increase in the jth covariate, when all other covariates are kept fixed, is to multiply the intensity by the "relative risk" e^{β_j}. Note that e^{β_j} is constant over time – the Cox model thus assumes that covariate effects are time-invariant and proportional to a baseline intensity.

In the models due to Aalen (1980, 1989), covariate effects act in an additive way on a baseline intensity. In the nonparametric case, the additive model specifies the intensity for firm i as

$$\alpha(t \,|\, \mathbf{x}_{it}) = \beta_0(t) + \boldsymbol{\beta}(t)^T \mathbf{x}_{it}, \tag{2.3}$$

where $\beta_0(t)$ is a locally integrable baseline intensity, left unspecified, and $\boldsymbol{\beta}(t) = (\beta_1(t), \ldots, \beta_p(t))^T$ is a vector of locally integrable regression coefficient functions, also left unspecified. The vector $\boldsymbol{\beta}(t)$ gives the time-varying effects of the covariates. The baseline $\beta_0(t)$ again corresponds to the intensity at time t when all covariates are identically equal to zero. We will also consider semiparametric versions of the additive model, where some covariate effects are time-constant parameters.

As for the Cox model, consider two time-t covariate vectors $\mathbf{x}_{1t}$ and $\mathbf{x}_{2t}$ that are identical except for the jth coordinate, where $x_{2j,t} = x_{1j,t} + 1$. Subtracting the intensities then gives

$$\alpha(t \,|\, \mathbf{x}_{2t}) - \alpha(t \,|\, \mathbf{x}_{1t}) = \boldsymbol{\beta}(t)^T \, (\mathbf{x}_{2t} - \mathbf{x}_{1t}) = \beta_j(t),$$

so the effect at time t of a one-unit increase in the jth covariate, when all other covariate are kept fixed, is to add the "excess-" or "absolute risk" $\beta_j(t)$ to the intensity.

In classical survival applications, the time-scale is usually duration-time, where t is measured as age or time from entry to exit for each subject at risk. In historical default studies, however, the natural time-scale is calender-time. The fact that this does not vary across firms makes it impossible to simultaneously identify a time-dependent baseline intensity and the effects of global time-dependent covariates. In the Cox model, this is handled by using a constant baseline, and similarly in the additive models, one sets $\beta_0(t) = \theta_0$ for all t and some real parameter θ_0, which leads to a semiparametric additive model.

2.2 Contrasting the Cox and the Aalen models

Using additive models for default intensities is unconventional, since one would a priori prefer models where intensities are forced to stay positive. In fact, an advantage of the Cox model is that intensities are born strictly positive. In additive models, we always run the risk of negative intensities, either as a result of estimation, or when extrapolating to more extreme covariate values.

Why, then, do we propose applying additive models in default studies? A main reason is the flexibility gained by relaxing the assumption of time-constant covariate effects. The additive models allow for simple estimation of time-varying effects using least-squares methods known from ordinary linear regression, and the resulting estimators are on a closed form that is easy to interpret and study. No smoothing is needed during estimation. There are extensions of the Cox model that incorporate time-varying effects, they require an iterative estimation procedure with smoothing in each iteration, which does not produce closed-form estimators and may blur the time-variation of effects due to repeated smoothing (see Zucker and Karr (1990) or Martinussen and Scheike (2006) and references therein).

The additive structure of the Aalen model is more robust towards model misspecification than the multiplicative structure of the Cox model. When we talk about model misspecification in this paper, it is in the sense of misspecification of the intensity. This may be a question of omitting relevant covariates or including them in an incorrect functional form (e.g. specifying a linear effect of a covariate when the true effect is nonlinear).

If a covariate is omitted, this corresponds to conditioning on a smaller filtration in the conditional expectation in (2.1). With an additive structure on this conditional mean and time varying parameters, the intensity will still be additive, though the remaining covariates may need to be transformed. In a Cox model, the proportionality is ruined when covariates are omitted, and the relative risk estimates will be biased, as shown by Struthers and Kalbfleisch (1986).

Including covariates in the wrong functional form in an additive model is a true misspecification, but the parameter estimates will still be interpretable as e.g. linear effects ("how much does the intensity change per change in the covariate overall?"). In a Cox model, the relative risk estimates are not interpretable as relative risks when the model is not correctly specified.

In the case of a wrong functional form for covariates, $(M_{it})_{t\in[0,T]}$ is no longer a martingale. This means that traditional variance estimates, which are based on martingale theory, will be biased. We will therefore use alternative variance estimators that are robust towards this type of misspecification. Having robust variance estimates and martingale based variance estimates allows us to detect model misspecification by comparing the two. Any large differences may suggest that the martingale based variance estimator is biased and this will be due to model misspecification. This idea is quite similar to the "Information Matrix Test" of White (1982).

The additive structure also permits methods from ordinary linear regression to carry over to the additive models. We will, for instance, in our data analysis in Section 4, include linear interaction terms as a supplement to the marginal effects of covariates and use a method known from ordinary linear regression to identify potential outliers that may be a source of model misspecification.

Studying excess- or absolute risks may give a more nuanced picture of risk-factor importance than solely relying on relative risks. Relative risks may be misleading and overstate the actual importance of risk-factors,

especially if the event is relatively rare, as has historically been the case with defaults. If a one unit increase in a covariate raises annual default probabilities from, say, 1 bps to 2 bps, the covariate's relative risk increase is 100% per year (a relative risk of 2), while its absolute risk increase is only 0.01% per year (an excess risk of 1bps). In this example, the relative risk may indicate an economically important default-predictor, while the excess risk may indicate that the same covariate is only moderately important or perhaps even economically insignificant.

As a final note, it is straightforward to obtain a non-negative estimate of the (past) intensity of a specific firm: Estimate the integrated intensity based on the parameter estimates obtained from the Aalen model and modify this estimate to be non-decreasing (by pooling adjacent violators – see Robertson, Wright, and Dykstra (1988)). Smoothing this modified integrated estimate will produce a non-negative estimate of the intensity.

2.3 Frailty and dynamic effects

Unobserved or latent effects are receiving increased attention in empirical default studies. Such effects may be due to omitted covariates or covariates subject to measurement error, but they may also correspond to effects which are actually unobservable.

There are two dominating approaches for correcting for unobserved effects in survival models. First, one may include "frailty" effects, where latent risk factors proxy unobserved effects – frailty is thus the survival analog of random effects and is often used to model dependence between event times. For instance, Duffie, Eckner, Horel, and Saita (2009) found evidence of a frailty process influencing historic U.S. corporate default probabilities by including a latent Ornstein-Uhlenbeck process alongside observable risk factors in a Cox model. A key purpose of this paper is to conduct a thorough check of time-varying effects, functional form, and interactions, hoping to detect possible sources of misspecification that might otherwise show up as frailty effects. We therefore do not include frailty effects in our paper but focus on means of teasing out more information on the effects of observable variables.

Second, one may use the internal history of the observed counting processes as a correction for missing effects. This is done by including time-dependent covariates directly linked to the observed history of the counting processes in the regression models. We follow Aalen, Fekjær, Borgan, and Husebye (2004), who included such covariates in an additive setting similar to ours, and call such covariates "dynamic." In our data analysis in Section 4, we will include a global dynamic covariate in the form of the trailing monthly default rate for the previous month as a correction for unobserved effects in an additive regression. This dynamic covariate may at a given time be viewed as reflecting the instantaneous default risk in the cohort. A significant effect may thus suggest that the additive model at hand is missing global effects that drive default intensities upwards.

3 Additive regression models

In this section, we describe how the nonparametric and semiparametric additive models are estimated, and how we test the relevant hypotheses of significance and possible time-variation of regression coefficients.

The focus is on two main model structures. First, the nonparametric additive model (2.3) with all covariate effects as unspecified functions of time, restated here for convenience

$$\alpha(t \,|\, \mathbf{x}_{it}) = \beta_0(t) + \beta_1(t)x_{i1,t} + \cdots + \beta_p(t)x_{ip,t}. \tag{3.1}$$

Second, the semiparametric sub-model first introduced by MacKeague and Sasieni (1994), where more structure is put on some of the regression coefficients. This is of interest when some effects are believed to be time-invariant, but also when including global time-dependent covariates of both the macroeconomic or dynamic type. Covariates with time-varying effects are collected in the p-dimensional covariate vector $\mathbf{x}_{it}$, whereas the q-dimensional vector $\mathbf{z}_{it} = (z_{i1,t}, \ldots, z_{iq,t})^T$ captures the time-invariant effects. Both are assumed to be predictable and locally bounded. The semiparametric additive model then specifies the intensity as

$$\alpha(t \,|\, \mathbf{x}_{it}, \mathbf{z}_{it}) = \beta_0(t) + \beta_1(t)x_{i1,t} + \cdots + \beta_p(t)x_{ip,t} + \theta_1 z_{i1,t} + \cdots + \theta_q z_{iq,t}, \tag{3.2}$$

where $\beta_0(t), \ldots, \beta_p(t)$ are as before while $\theta_1, \ldots, \theta_q$ are real-valued parameters giving the time-invariant effects of a one-unit increase in each component of $\mathbf{z}_{it}$. Note that, as discussed in Section 2.1, $\beta_0(t) = \theta_0$ when we include global time-dependent covariates in the regressions, in order to make all parameters identifiable. Martinussen and Scheike (2006) propose a resampling-based inference procedure that allows the time-invariance of effects to be tested, so that an initial nonparametric additive model may be reduced to a semiparametric. This is implemented in the `aalen`-function as a part of their `timereg` package in `R` (R Development Core Team (2011)) which will be used in the data analyses of Sections 4 and 6.

3.1 The nonparametric additive regression model

In the general model (3.1), the regression functions $\beta_j(t)$ for $j = 0, 1, \ldots, p$ are unrestricted and consequently difficult to estimate nonparametrically. However, similar to estimating a cumulative distribution function rather than a density or a cumulative hazard instead of the hazard itself, it turns out that the cumulative regression functions,

$$B_j(t) = \int_0^t \beta_j(s)\, ds, \qquad j = 1, \ldots, p,$$

are easier to estimate than the regression functions themselves. As in the case with the density or hazard, estimators of the regression functions may be obtained by smoothing the cumulative estimates.

Estimation

The basic idea is to estimate the cumulative regression coefficients by step functions. We can write the increment of $(N_{it})_{t\in[0,T]}$ over the small time interval $[t, t + dt)$ as

$$dN_{it} = Y_{it} dB_0(t) + \sum_{j=1}^{p} Y_{it} x_{ij,t}\, dB_j(t) + dM_{it}, \tag{3.3}$$

where

$$M_{it} = N_{it} - \int_0^t \left(Y_{is}\beta_0(s) + \sum_{j=1}^{p} Y_{is} x_{ij,s} \beta_j(s) \right) ds$$

defines a local martingale $(M_{it})_{t\in[0,T]}$ due to the assumed local boundedness of $x_{ij,t}$ and local integrability of $\beta_j(t)$ for all i and j. At each time t, the model (3.3) has the form of an ordinary linear regression with dN_{it} as the response, $Y_{it}x_{ij,t}$ as the predictors, $dB_j(t) = \beta_j(t)\, dt$ as the parameters of interest, and dM_{it} as the noise. For the cohort of n firms, the full model may thus be written as

$$d\mathbf{N}_t = \mathbf{X}_t\, d\mathbf{B}(t) + d\mathbf{M}_t, \tag{3.4}$$

where $\mathbf{N}_t = (N_{1t}, \ldots, N_{nt})^T$, $\mathbf{B}(t) = (B_0(t), \ldots, B_p(t))^T$, and $\mathbf{M}_t = (M_{1t}, \ldots, M_{nt})^T$, while $\mathbf{X}_t$ is the $n \times (1+p)$-dimensional, locally bounded matrix with ith row $(Y_{it}, Y_{it}x_{i1,t}, \ldots, Y_{it}x_{ip,t})$. When $\mathbf{X}_t$ has full rank, the ordinary least squares estimator of the increment of $\mathbf{B}(t)$ is given by

$$d\widehat{\mathbf{B}}(t) = \mathbf{X}_t^- \, d\mathbf{N}_t,$$

where $\mathbf{X}_t^- = (\mathbf{X}_t^T\mathbf{X}_t)^{-1}\mathbf{X}_t^T$ is the usual least squares generalized inverse of $\mathbf{X}_t$. When $\mathbf{X}_t$ has less than full rank, $d\mathbf{B}(t)$ is not identifiable from the data, and we put $d\widehat{\mathbf{B}}(t) = 0$. Also, note that $d\widehat{\mathbf{B}}(t) = 0$ when $d\mathbf{N}_t = 0$, so that all the increments are at default times.

To obtain an estimator for the vector $\mathbf{B}(t)$ of cumulative regression functions, we let $J(t)$ be the indicator of $\mathbf{X}(t)$ having full rank and aggregate $d\widehat{\mathbf{B}}(t)$ over the ordered default-times $\tau_1 < \tau_2 < \cdots$ to obtain

$$\widehat{\mathbf{B}}(t) = \int_0^t J(s)\, \mathbf{X}_s^- \, d\mathbf{N}_s = \sum_{\tau_k \le t} J(\tau_k)\, \mathbf{X}_{\tau_k}^- \, \triangle\mathbf{N}_{\tau_k}, \tag{3.5}$$

where $\triangle\mathbf{N}_{\tau_k}$ is a vector of zeros except for a one at the component corresponding to the firm with a default at τ_k. Note that when there are no covariates in the model (i.e. $p = 0$), the estimator (3.5) is just the usual Nelson-Aalen estimator of the cumulative hazard – in this sense, the nonparametric additive model is the natural generalization of nonparametric hazard estimation to the situation with covariates.

Using (3.4), we obtain

$$\begin{aligned} \widehat{\mathbf{B}}(t) - \mathbf{B}(t) &= \int_0^t J(s)\, \mathbf{X}_s^- \, d\mathbf{M}_s + \int_0^t (J(s) - 1)\, d\mathbf{B}(s) \\ &= \int_0^t J(s)\, \mathbf{X}_s^- \, d\mathbf{M}_s + o_P\left(\tfrac{1}{\sqrt{n}}\right), \end{aligned}$$

where the last equality holds under reasonable regularity assumptions (Martinussen and Scheike, 2006). Hence, the deviation $\widehat{\mathbf{B}}(t) - \mathbf{B}(t)$ is a vector-valued local martingale except for a negligible remainder term. The asymptotic properties of the estimator (3.5) may be obtained by the martingale central limit theorem. In case of model misspecifications, the process $(\mathbf{M}_t)_{t\in[0,T]}$ is no longer a local martingale, and care must be taken when deriving the asymptotic properties of the estimator $\widehat{\mathbf{B}}(t)$. Both cases will be treated below.

When the model is well-specified, we can estimate the covariance function of $\widehat{\mathbf{B}}(t)$ by the optional variation process of the martingale part of the deviation $\widehat{\mathbf{B}}(t) - \mathbf{B}(t)$:

$$\widehat{\mathbf{\Sigma}}_{\mathrm{mar}}(t) = \sum_{\tau_k \le t} J(\tau_k)\, \mathbf{X}_{\tau_k}^- \, \mathrm{diag}(\triangle\mathbf{N}_{\tau_k})\, {\mathbf{X}_{\tau_k}^-}^T. \tag{3.6}$$

If the model is misspecified, $\widehat{\mathbf{\Sigma}}_{\mathrm{mar}}(t)$ will be biased. In this case Martinussen and Scheike (2006) show that

$$\sqrt{n}(\widehat{\mathbf{B}}(t) - \mathbf{B}(t)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{Q}_i(t) + o_P(1), \tag{3.7}$$

with

$$\mathbf{Q}_i(t) = \int_0^t J(s) \left(n^{-1}\mathbf{X}_s^T\mathbf{X}_s\right)^{-1} X_{is}^T \left(dN_{is} - X_{is}\, d\mathbf{B}(s)\right),$$

and where X_{it} is the ith row of $\mathbf{X}_t$. When $n \to \infty$ $J(t)\left(n^{-1}\mathbf{X}_t^T\mathbf{X}_t\right)^{-1}$ converges in probability, and (3.7) is a normalized sum of independent and identically distributed processes. The covariance function of $\widehat{\mathbf{B}}(t)$ is

consistently estimated by

$$\widehat{\mathbf{\Sigma}}_{\text{rob}}(t) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\mathbf{Q}}_i(t)\widehat{\mathbf{Q}}_i(t)^T, \tag{3.8}$$

where $\widehat{\mathbf{Q}}_i(t)$ is obtained by replacing $\mathbf{B}(t)$ with its estimator in $\mathbf{Q}_i(t)$. Since $\widehat{\mathbf{\Sigma}}_{\text{rob}}(t)$ is derived without use of the local martingale property of $(\mathbf{M}_t)_{t\in[0,T]}$, it is robust to model misspecifications of the intensity, and is therefore a robust covariance function estimator.

Large sample properties

When the model is correctly specified and regularity conditions are fulfilled, it follows from the martingale central limit theorem that the normalized deviation $\sqrt{n}(\widehat{\mathbf{B}}(t) - \mathbf{B}(t))$ converges (for $n \to \infty$ and fixed T) in distribution to a mean-zero multivariate Gaussian martingale with a covariance function that may be estimated consistently by the martingale-based estimator $\widehat{\mathbf{\Sigma}}_{\text{mar}}(t)$. In case of misspecifications, when $(M_{it})_{t\in[0,T]}$ is not a martingale, the main asymptotic result is instead that $\sqrt{n}(\widehat{\mathbf{B}}(t) - \mathbf{B}(t))$, using (3.7), converges in distribution to a mean-zero Gaussian process (but not a martingale) with a covariance function which may be estimated consistently by the robust estimator $\widehat{\mathbf{\Sigma}}_{\text{rob}}(t)$. Martinussen and Scheike (2006) give the details and proofs.

The asymptotic results imply that an approximate $100(1-\alpha)\%$ martingale-based pointwise confidence band for the jth cumulative regression coefficient is given by

$$\widehat{B}_j(t) \pm z_{1-\alpha/2}\sqrt{\widehat{\sigma}^2_{\text{mar},jj}(t)},$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution and $\widehat{\sigma}^2_{\text{mar},jj}(t)$ is the jth diagonal element of $\widehat{\mathbf{\Sigma}}_{\text{mar}}(t)$. Alternatively, an approximate $100(1-\alpha)\%$ robust confidence band for the jth cumulative regression coefficient may be obtained as

$$\widehat{B}_j(t) \pm z_{1-\alpha/2}\sqrt{\widehat{\sigma}^2_{\text{rob},jj}(t)},$$

where $\widehat{\sigma}^2_{\text{rob},jj}(t)$ is the jth diagonal element of $\widehat{\mathbf{\Sigma}}_{\text{rob}}(t)$.

In large samples, if the two types of variance estimates (i.e. the martingale-based and the robust) are markedly different, it is an indication of model misspecification.

Kernel smoothing

$\widehat{B}_j(t)$ estimates the cumulated regression function $\int_0^t \beta_j(s)ds$. However, we are really interested in its slope – i.e. the regression function $\beta_j(t)$ itself. As assessing the slope of a step function graphically may be difficult, we smooth the estimators of the cumulative regression functions to obtain estimators of the regression functions themselves.

We will use kernel function smoothing, as first proposed in a survival model setting by Ramlau-Hansen (1983). Then $\boldsymbol{\beta}(t) = (\beta_0(t), \ldots, \beta_p(t))$ is estimated at time t as a weighted sum of the increments of $\widehat{\mathbf{B}}(t)$ over the interval $[t-b, t+b]$,

$$\widehat{\boldsymbol{\beta}}(t) = \frac{1}{b}\sum_{\tau_k} K\left(\frac{t-\tau_k}{b}\right) \triangle\widehat{\mathbf{B}}(\tau_k), \tag{3.9}$$

where $b > 0$ is a bandwidth, determining the size of the interval $[t-b, t+b]$, while $K(x)$ is a bounded kernel-function vanishing outside $[-1, 1]$ and integrating to 1, determining the weights. A typical choice is the Epanechnikov kernel, where $K(x) = \frac{3}{4}(1-x^2)$ for $|x| \leq 1$, and zero otherwise, but Aalen, Borgan, and Gjessing (2008) discuss other kernels. Like other kernel estimators, (3.9) suffers from boundary effects: For small values of t (i.e. when $t - b < 0$) the estimator is severely biased towards zero. We handle this problem by using a boundary kernel, as also discussed by Aalen et al. (2008).

When the model is correctly specified and $(\mathbf{M}_t)_{t\in[0,T]}$ is a vector-valued local martingale, an estimator of the covariance function of $\widehat{\boldsymbol{\beta}}(t)$ is simply obtained as

$$\widehat{\mathrm{Cov}}\,\widehat{\boldsymbol{\beta}}(t) = \frac{1}{b^2}\sum_{\tau_k} K\left(\frac{t-\tau_k}{b}\right)^2 \triangle\widehat{\boldsymbol{\Sigma}}_{\mathrm{mar}}(\tau_k),$$

where $\triangle\widehat{\boldsymbol{\Sigma}}_{\mathrm{mar}}(\tau_k) = J(\tau_k)\,\mathbf{X}^-_{\tau_k}\,\mathrm{diag}(\triangle\mathbf{N}_{\tau_k})\,\mathbf{X}^{-\,T}_{\tau_k}$ is the increment of the martingale-based covariance function estimator (3.6). If the model is misspecified, the covariance function of the function $\widehat{\boldsymbol{\beta}}(t)$ has the same form as above, but with $\triangle\widehat{\boldsymbol{\Sigma}}_{\mathrm{mar}}(\tau_k)$ replaced by $\triangle\widehat{\boldsymbol{\Sigma}}_{\mathrm{rob}}(\tau_k)$, i.e, the increment of the robust covariance function estimator (3.8). These covariance function estimators may be combined with the large sample properties of $\widehat{\mathbf{B}}(t)$ to construct approximate confidence bands for $\widehat{\boldsymbol{\beta}}(t)$ in the usual way.

There exist techniques (such as cross-validation) for objectively choosing the bandwidth with an optimal (in some sense) trade-off between bias and variances, but these will not be considered in this paper. We simply rely on a subjective assessment of what is a reasonable degree of smoothing.

Resampling-based inference

The tests we use will be based on a resampling procedure presented and implemented by Martinussen and Scheike (2006) – however, Aalen et al. (2008) give a more traditional martingale-based approach.

We are primarily interested in testing two hypotheses: The hypothesis of no effect of the jth covariate,

$$H_0^{\mathrm{effect}}:\ \beta_j(t) = 0 \ \text{ for all } \ t \in [0,T].$$

and the hypothesis of a time-invariant effect of the jth covariate,

$$H_0^{\mathrm{time}}:\ \beta_j(t) = \theta_j \ \text{ for all } \ t \in [0,T].$$

The null of H_0^{time} is the semi-parametric additive model with a parametric coefficient for the effect of the jth covariate. Both hypotheses may be of interest over a shorter time-interval than the entire study time, but this is usually not the case, and will not be considered here. Since the estimated cumulative regression coefficients have nicer distributional properties than their smoothed counterparts, the hypotheses are usually formulated in the equivalent forms

$$H_0^{\mathrm{effect}}:\ B_j(t) = 0 \quad \text{and} \quad H_0^{\mathrm{time}}:\ B_j(t) = \theta_j t,$$

both for all $t \in [0,T]$.

Martinussen and Scheike (2006) propose testing the hypothesis of no influence, H_0^{effect}, by the supremum test statistic

$$T_{\mathrm{sup}} = \sup_{t\in[0,T]}\left|\frac{\sqrt{n}\,\widehat{B}_j(t)}{\sqrt{\widehat{\sigma}^2_{\mathrm{rob},jj}(t)}}\right|, \tag{3.10}$$

where, again, $\widehat{\sigma}^2_{\text{rob},jj}(t)$ is the jth diagonal element of $\widehat{\mathbf{\Sigma}}_{\text{rob}}(t)$ given in (3.8). This evaluates the maximal deviation of the estimated cumulative regression coefficient $\widehat{B}_j(t)$ from the zero function, relative to its variation.

With regards to testing the hypothesis of time-invariance, H_0^{time}, Martinussen and Scheike (2006) propose the process

$$\sqrt{n}\left(\widehat{B}_j(t) - \frac{\widehat{B}_j(T)}{T}t\right) \quad \text{for} \quad t \in [0,T] \tag{3.11}$$

as a basic starting point for evaluating the time-invariance of the jth regression coefficient – the idea is that $\widehat{B}_j(T)/T$ is an estimator of the time-constant coefficient θ_j under the null. This process may then be turned into the Kolmogorov-Smirnov test statistic

$$T_{\text{KS}} = \sup_{t\in[0,T]} \sqrt{n}\left|\widehat{B}_j(t) - \frac{\widehat{B}_j(T)}{T}t\right|, \tag{3.12}$$

or the Cramer-von Mises test statistic

$$T_{\text{CvM}} = n\int_0^T \left(\widehat{B}_j(t) - \frac{\widehat{B}_j(T)}{T}t\right)^2 dt. \tag{3.13}$$

The former is a maximal deviations test statistic, sensitive to single large deviations from the null, while the latter is a sum of squared deviations type statistic, sensitive to small but persistent deviations from the null.

To test the hypotheses, Martinussen and Scheike (2006) propose evaluating the variability of test statistics through a resampling-scheme that approximates the distribution of the estimated vector of cumulative regression coefficients $\widehat{\mathbf{B}}(t)$. Based on the iid. representation (3.7), their main result is that, conditional on the data $(N_{it}, Y_{it}, \mathbf{x}_{it})$ for $i = 1,\ldots,n$, the normalized deviation $\sqrt{n}(\widehat{\mathbf{B}}(t) - \mathbf{B}(t))$ has the same limiting distribution as

$$\mathbf{R}(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n u_i \widehat{\mathbf{Q}}_i(t),$$

where $u_1,\ldots,u_n$ are iid. standard normally distributed stochastic variables, and $\widehat{\mathbf{Q}}_i(t)$ is as in (3.8). The result utilizes the "conditional multiplier central limit theorem" (van der Vaart and Wellner, 1996). Note that $\mathbf{R}(t)$ a weighted sum of the observable $\widehat{\mathbf{Q}}_i(t)/\sqrt{n}$ with standard normally distributed weights. In fact, any normalized distribution could have been used for the weights, but the choice of a normal distribution fits well with the limiting distribution.

Obtaining P-values for the tests may be done through replication of $\mathbf{R}(t)$. The idea is to hold the observed data fixed whilst repeatedly generating series of iid. standard normal variables $u_1^{(r)},\ldots,u_n^{(r)}$, and approximating the distribution of T_{sup} by the empirical distribution of the processes

$$\sup_{t\in[0,T]}\left|\frac{\sqrt{n}R_j^{(r)}(t)}{\sqrt{\widehat{\sigma}^2_{\text{rob},jj}(t)}}\right|; \quad r = 1,2,\ldots,$$

where $R_j^{(r)}(t)$ denotes the rth resample of the jth element of $\mathbf{R}(t)$. Similarly, approximations of the distributions of T_{KS} and T_{CvM} are obtained by approximating the distribution of the process (3.11) by the empirical distribution of the processes

$$\sqrt{n}\left(R_j^{(r)}(t) - \frac{R_j^{(r)}(T)}{T}t\right); \quad r = 1,2,\ldots.$$

Finally, the deviation of the estimated regression coefficient from the null of time-invariance may be assessed by plotting the observed process (3.11) as a function of study time along with a number of the resampled processes under the null. A similar behavioral pattern in the observed and resampled processes would suggest consistency with the null of time-invariance. The advantage of this graphical method is that it pinpoints where in time deviations from the null might occur.

3.2 The semiparametric additive regression model

The methods used to estimate and conduct inference in the semiparametric Aalen model (3.2) are to a large extent similar to the ones presented in the previous sections for the nonparametric model. We will therefore only briefly outline the procedure of finding estimators and the resampling-based method of inference. Martinussen and Scheike (2006) give a detailed account.

Estimation

For the semiparametric Aalen model, the increments of the counting processes are given as

$$d\mathbf{N}_t = \mathbf{X}_t\, d\mathbf{B}(t) + \mathbf{Z}_t\, \boldsymbol{\theta} dt + d\mathbf{M}_t, \tag{3.14}$$

where $\mathbf{N}_t$, $\mathbf{X}_t$, $\mathbf{B}(t)$, and $\mathbf{M}_t$ are as in (3.4), while $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^T$ and $\mathbf{Z}_t$ is the $n \times q$-dimensional, locally bounded matrix with ith row $(Y_{it} z_{i1,t}, \ldots, Y_{it} z_{iq,t})$.

We estimate the unknown regression functions and parameters by minimizing the integrated sum of squares,

$$\int_0^T J(t) \big(d\mathbf{N}_t - \mathbf{X}_t\, d\mathbf{B}(t) - \mathbf{Z}_t\, \boldsymbol{\theta} dt\big)^T \big(d\mathbf{N}_t - \mathbf{X}_t\, d\mathbf{B}(t) - \mathbf{Z}_t\, \boldsymbol{\theta} dt\big), \tag{3.15}$$

where, again, $J(t)$ is the indicator of $\mathbf{X}_t$ having full rank. Re-writing $d\mathbf{N}_s - \mathbf{X}_s\, d\mathbf{B}(s) - \mathbf{Z}_s\, \boldsymbol{\theta} dt$ as the sum of its projections on span $\mathbf{X}_t$ and its orthogonal complement, we obtain, when $J(t) = 1$,

$$\mathbf{X}_t \big(\mathbf{X}_t^-\, d\mathbf{N}_t - d\mathbf{B}(t) - \mathbf{X}_t^-\, \mathbf{Z}_t\, \boldsymbol{\theta}\, dt\big) + \big(\mathbf{I} - \mathbf{X}_t\, \mathbf{X}_t^-\big) \big(d\mathbf{N}_t - \mathbf{Z}_t\, \boldsymbol{\theta}\, dt\big),$$

where $\mathbf{I}$ is the $n \times n$ identity matrix, while, as usual, $\mathbf{X}_t^-$ is the least squares generalized inverse of $\mathbf{X}_t$. Hence, (3.15) splits into a sum of two terms,

$$\int_0^T J(t) \big(\mathbf{X}_t^- d\mathbf{N}_t - d\mathbf{B}(t) - \mathbf{X}_t^-\, \mathbf{Z}_t\, \boldsymbol{\theta}\, dt\big)^T \mathbf{X}_t^T\, \mathbf{X}_t \big(\mathbf{X}_t^- d\mathbf{N}_t - d\mathbf{B}(t) - \mathbf{X}_t^-\, \mathbf{Z}_t\, \boldsymbol{\theta}\, dt\big)$$
$$+ \int_0^T J(t) \big(d\mathbf{N}_t - \mathbf{Z}_t\, \boldsymbol{\theta}\, dt\big)^T \big(\mathbf{I} - \mathbf{X}_t\, \mathbf{X}_t^-\big) \big(d\mathbf{N}_t - \mathbf{Z}_t\, \boldsymbol{\theta}\, dt\big),$$

as $\big(\mathbf{I} - \mathbf{X}_t\, \mathbf{X}_t^-\big)^T \big(\mathbf{I} - \mathbf{X}_t\, \mathbf{X}_t^-\big) = \big(\mathbf{I} - \mathbf{X}_t\, \mathbf{X}_t^-\big)$ and $\big(\mathbf{I} - \mathbf{X}_t\, \mathbf{X}_t^-\big)^T \mathbf{X}_t = \mathbf{0}$. Minimizing the latter term first and the former second yields the estimators

$$\widehat{\boldsymbol{\theta}} = \left(\int_0^T J(t)\, \mathbf{Z}_t^T \big(\mathbf{I} - \mathbf{X}_t \mathbf{X}_t^-\big)\, \mathbf{Z}_t\, dt\right)^{-1} \int_0^T J(t)\, \mathbf{Z}_t^T \big(\mathbf{I} - \mathbf{X}_t \mathbf{X}_t^-\big)\, d\mathbf{N}_t, \tag{3.16}$$

$$\widehat{\mathbf{B}}(t) = \int_0^t J(s)\, \mathbf{X}_s^- \big(d\mathbf{N}_s - \mathbf{Z}_s\, \widehat{\boldsymbol{\theta}}\, ds\big). \tag{3.17}$$

In the non-parametric case, the minimizer of the integrated sum of squares and the pointwise sums of squares is the same. Thus, the present estimator (3.17) generalizes the one given for the nonparametric case in (3.5).

Combining (3.16) and (3.17) with (3.14) gives the normalized deviations

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{U}^{-1}\,\mathbf{M}_T^{(1)}, \tag{3.18}$$

$$\sqrt{n}(\widehat{\mathbf{B}}(t) - \mathbf{B}(t)) = \mathbf{M}_t^{(2)} - \mathbf{V}(t)\,\mathbf{U}^{-1}\,\mathbf{M}_T^{(1)} + o_P(1), \tag{3.19}$$

where

$$\mathbf{U} = \frac{1}{n}\int_0^T J(t)\,\mathbf{Z}_t^T\left(\mathbf{I} - \mathbf{X}_t\mathbf{X}_t^-\right)\mathbf{Z}_t\,dt, \qquad \mathbf{V}(t) = \frac{1}{n}\int_0^t J(s)\,\mathbf{X}_s^-\,\mathbf{Z}_s\,ds,$$

$$\mathbf{M}_t^{(1)} = \frac{1}{\sqrt{n}}\int_0^t J(s)\,\mathbf{Z}_s\left(\mathbf{I} - \mathbf{X}_s\mathbf{X}_s^-\right)d\mathbf{M}_s, \qquad \mathbf{M}_t^{(2)} = \frac{1}{\sqrt{n}}\int_0^t J(s)\,\mathbf{X}_s^-\,d\mathbf{M}_s.$$

When the model is correctly specified, $(\mathbf{M}_t^{(1)})_{t\in[0,T]}$ and $(\mathbf{M}_t^{(2)})_{t\in[0,T]}$ are vector-valued, local martingales (due to the assumed local boundedness of $\mathbf{X}_t$ and $\mathbf{Z}_t$), thus showing that $\widehat{\boldsymbol{\theta}}$ is an unbiased estimator of $\boldsymbol{\theta}$, while $\widehat{\mathbf{B}}(t)$ is an approximately unbiased estimator of $\mathbf{B}(t)$. In case of a misspecified model, the process $(\mathbf{M}_t)_{t\in[0,T]}$ is no longer a local martingale. In the following, we handle the two cases separately when deriving variance-estimates and asymptotic properties for the estimators (3.16) and (3.17).

When the model fit is reasonable, the deviations (3.18) and (3.19) imply that a martingale-based estimator of the covariance matrix of $\widehat{\boldsymbol{\theta}}$ is given by

$$\widehat{\boldsymbol{\Psi}}_{\mathrm{mar}} = \mathbf{U}^{-1}\,[\mathbf{M}^{(1)}](T)\,\mathbf{U}^{-1},$$

while a martingale-based estimator of the covariance function of $\widehat{\mathbf{B}}(t)$ is given by

$$\begin{aligned}\widehat{\boldsymbol{\Upsilon}}_{\mathrm{mar}}(t) &= [\mathbf{M}^{(2)}](t) + \mathbf{V}(t)\,\widehat{\boldsymbol{\Psi}}_{\mathrm{mar}}\,\mathbf{V}(t)^T \\ &\quad -[\mathbf{M}^{(1)},\mathbf{M}^{(2)}](t)\,\mathbf{U}^{-1}\,\mathbf{V}(t)^T - \mathbf{V}(t)\,\mathbf{U}^{-1}\,[\mathbf{M}^{(1)},\mathbf{M}^{(2)}](t),\end{aligned}$$

where the optional variation and covariation processes of $(\mathbf{M}_t^{(1)})_{t\in[0,T]}$ and $(\mathbf{M}_t^{(2)})_{t\in[0,T]}$ are

$$[\mathbf{M}^{(1)}](t) = \frac{1}{n}\int_0^t J(s)\,\mathbf{Z}_s^T\left(\mathbf{I} - \mathbf{X}_s\mathbf{X}_s^-\right)\mathrm{diag}(d\mathbf{N}_s)\left(\mathbf{I} - \mathbf{X}_s\mathbf{X}_s^-\right)\mathbf{Z}_s,$$

$$[\mathbf{M}^{(2)}](t) = n\int_0^t J(s)\,\mathbf{X}_s^-\,\mathrm{diag}(d\mathbf{N}_s)\,\mathbf{X}_s^-,$$

$$[\mathbf{M}^{(1)},\mathbf{M}^{(2)}](t) = \int_0^t J(s)\,\mathbf{Z}_s^T\left(\mathbf{I} - \mathbf{X}_s\mathbf{X}_s^-\right)\mathrm{diag}(d\mathbf{N}_s)\,\mathbf{X}_s^{-T}.$$

If the semiparametric additive model is misspecified, $\widehat{\boldsymbol{\Psi}}_{\mathrm{mar}}$ and $\widehat{\boldsymbol{\Upsilon}}_{\mathrm{mar}}(t)$ will be biased. To address this, Martinussen and Scheike (2006) show that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \frac{1}{\sqrt{n}}\,\mathbf{U}^{-1}\sum_{i=1}^n \mathbf{K}_i,$$

where

$$\mathbf{K}_i = \int_0^T J(t)\left(Z_{it}^T - \mathbf{Z}_t^T\,\mathbf{X}_t\left(\mathbf{X}_t^T\mathbf{X}_t\right)^{-1}X_{it}^T\right)\left(dN_{it} - X_{it}\,d\mathbf{B}(t) - Z_{it}\,\boldsymbol{\theta}\,dt\right),$$

with X_{it} as the ith row of $\mathbf{X}_t$ and Z_{it} as the ith row of $\mathbf{Z}_t$. When n is large, this represents the deviation $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ as a normalized sum of iid. terms, showing that the covariance function of $\widehat{\boldsymbol{\theta}}$ may be estimated by the sandwich-type estimator

$$\widehat{\boldsymbol{\Psi}}_{\text{rob}} = \mathbf{U}^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \widehat{\mathbf{K}}_i \, \widehat{\mathbf{K}}_i^T \right) \mathbf{U}^{-1},$$

where $\widehat{\mathbf{K}}_i$ is obtained by replacing $\boldsymbol{\theta}$ and $\mathbf{B}(t)$ with their estimators in $\mathbf{K}_i$. As $\widehat{\boldsymbol{\Psi}}_{\text{rob}}$ was derived without relying on the martingale property, it is robust to model misspecifications. Similarly,

$$\sqrt{n}(\widehat{\mathbf{B}}(t) - \mathbf{B}(t)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{O}_i(t) + o_p(1),$$

where

$$\mathbf{O}_i(t) = \int_0^t J(s) \left(n^{-1} \mathbf{X}_s^T \mathbf{X}_s \right)^{-1} X_{is}^T \left(dN_{is} - X_{is} \, d\mathbf{B}(s) - Z_{is} \, \boldsymbol{\theta} \, ds \right) - \mathbf{V}(t) \, \mathbf{U}^{-1} \, \mathbf{K}_i.$$

Hence, copying the above argument, this suggests that the covariance function of $\widehat{\mathbf{B}}(t)$ may be estimated by the robust estimator

$$\widehat{\boldsymbol{\Upsilon}}_{\text{rob}}(t) = \frac{1}{n} \sum_{i=1}^{n} \widehat{\mathbf{O}}_i(t) \, \widehat{\mathbf{O}}_i(t)^T. \tag{3.20}$$

Large sample properties

The asymptotic properties of $\widehat{\boldsymbol{\theta}}$ are in case of a well-specified model obtained through a standard application of the martingale central limit theorem, which under suitable regularity assumptions implies that $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ converges (for $n \to \infty$ and fixed T) in distribution to a mean-zero, multivariate normal variable with a covariance which is consistently estimated by $\widehat{\boldsymbol{\Psi}}_{\text{mar}}$. If the model is not well specified, the same convergence result holds, but $\widehat{\boldsymbol{\Psi}}_{\text{mar}}$ has to be replaced by $\widehat{\boldsymbol{\Psi}}_{\text{rob}}$, which may also be shown to be a consistent estimator of the asymptotic covariance. The details are given by Martinussen and Scheike (2006).

The situation is slightly more complicated for $\widehat{\mathbf{B}}(t)$ due to the more involved nature of the corresponding deviation (3.19), which depends on $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$. When the model is well-specified and the necessary regularity conditions are fulfilled, Martinussen and Scheike (2006) show that combining the martingale central limit theorem and the continuous mapping theorem gives that $\sqrt{n}(\widehat{\mathbf{B}}(t) - \mathbf{B}(t))$ converges in distribution to a mean-zero, multivariate Gaussian process (i.e. not a martingale). Its asymptotic covariance function is estimated consistently by $\widehat{\boldsymbol{\Upsilon}}_{\text{mar}}(t)$. If the model is misspecified, the same result holds, but $\widehat{\boldsymbol{\Upsilon}}_{\text{mar}}(t)$ has to be replaced by $\widehat{\boldsymbol{\Upsilon}}_{\text{rob}}(t)$. Note that the asymptotic distribution of $\sqrt{n}(\widehat{\mathbf{B}}(t) - \mathbf{B}(t))$ is not a martingale in either case. These results may in large samples be used to construct approximate pointwise confidence bands for $\widehat{\mathbf{B}}(t)$ in the usual way.

Resampling-based inference

Three hypotheses arise naturally when conducting inference in the semiparametric additive model. The first two are the null hypotheses of no influence and time-invariance of the time-varying regression coefficients that we also considered for the nonparametric model. The third is the simple null hypothesis $\theta_l = \tilde{\theta}$, where the lth time-constant regression coefficient is tested against the null of some known, real-valued parameter $\tilde{\theta}$. Usually,

$\tilde{\theta} = 0$, in which case we are testing the influence of the lth time-constant covariate effect. This hypothesis is tested in the usual manner using e.g. the Wald test statistic, which in case of a well-specified model is given by

$$T_{\text{Wald}} = \frac{\widehat{\theta}_l - \tilde{\theta}}{\sqrt{\psi^2_{\text{mar},ll}}},$$

where $\psi^2_{\text{mar},ll}$ is the lth diagonal element of $\widehat{\boldsymbol{\Psi}}_{\text{mar}}$. Alternatively, in case of a misspecified model, we instead use $\psi^2_{\text{rob},ll}$, the lth diagonal element of $\widehat{\boldsymbol{\Psi}}_{\text{rob}}$. In either case, the asymptotic properties of $\widehat{\boldsymbol{\theta}}$ imply that T_{Wald} will under the null be approximately standard normally distributed in large samples.

With regards to the hypotheses of no influence and time-invariance for the a priori time-varying regression coefficients, Martinussen and Scheike (2006) propose a resampling-based method analogous to what was done for the nonparametric additive model. As for the nonparametric additive model, we can assess the deviation of an estimated, time-varying regression coefficient from the null of time-invariance by plotting the observed process (3.11), as a function of study time, along with a number of the resampled processes under the null.

4 Data and model specification

Our empirical results are based on an analysis of U.S. industrial corporate defaults occurring between 1982 and 2006. In this period, the U.S. economy suffered three major economic recessions between '81-'82, '91-'92, and '01-'02. We use the exact same data set as Lando and Nielsen (2010), but, for convenience, we repeat the description of the data below. Other default studies have used the same data supplemented with additional defaults from other sources – e.g. Li and Zhao (2006), Das et al. (2007), Davydenko (2007), and Le (2007). Our main purpose here is to investigate the role of time-varying coefficients and the importance of model checking in an additive intensity regression setting.

4.1 Data

The sample includes all U.S. industrial firms with a debt issue registered in Moody's Default Risk Service Database, or DRSD (Moody's Investor's Service©, 2010), which essentially covers the period since 1970, including universal identifiers facilitating the combination with other data sources. However, due to sparseness of data, the study time was chosen as the period between January 1st 1982 and January 1st 2006, giving a total of 289 observation months. The sample was restricted to firms for which accompanying stock market data from CRSP (Center for Research in Security Prices©, 2010) and accounting data from CompuStat (Standard & Poor's©, 2010) could be obtained, and for which there were at least 6 months of available data. All consecutive default events occurring within a 1-month horizon of any previously registered default ascribed to the same parent company were excluded from the data in order to correct for observations of multiple defaults caused by parent-subsidiary relations within the same corporate family. The final result is a study cohort of 2,557 firms comprising a total of 370 defaults.

The database classifies any of the following 9 events as constituting a default: *Chapter 7, chapter 11, distressed exchange, grace period default, missed interest payment, missed principle payment, missed principal and interest payment, prepackaged chapter 11*, and *suspension of payments*. In particular, we do not correct the timing of a "Distressed exchange," which in the DRSD is registered as the time of completion of the exchange,

although as suggested by Davydenko (2007), it would probably be more appropriate to instead collect separate information on the announcement date of the exchange.

Of the 370 realized defaults in the cohort, 25 defaults happened to firms which had already defaulted earlier. To be precise, 22 firms experienced 2 defaults and a single firm defaulted 4 times.

Covariate specification

The idiosyncratic covariate specification to be employed in the regressions consists of the following 5 balance sheet variables obtained from CRSP and CompuStat:

- **Quick ratio** for the previous month, calculated as the book value of cash and short-term investments added the book value of total receivables, all relative to the book value of total current liabilities. This measures the firm's ability to use its near cash assets to immediately extinguish current liabilities.
- **Pledgeable assets** for the previous month, calculated as the book value of total current assets plus the book value of net property, plant, and equipment. This measures the firm's ability to convert liquid assets to collateral that may be transferred to a lender to secure debt.
- **Trailing 1-year equity return** for the previous month. This measures the firm's efficiency at generating profits from shareholders' equity (or assets less liabilities) and is thus a reflection of how well the firm uses investment funds to generate growth on earnings.
- **Trailing 1-year distance-to-default** for the previous month and estimated over a one-year rolling window. Our distance measure is the option-implied measure as used for example in Duffie et al. (2007) (see page 660 of that paper). This measures roughly how far - measured in standard deviations of log returns - assets are from hitting a default triggering boundary. The default triggering level is the sum of the notional of short-term debt and half of the notional of long-term debt.
- **Percentage short-term debt** for the previous month calculated as the book value of debt in current liabilities divided by the sum of the book value of debt in current liabilities and the book value of total long-term debt. This is a measure of the firm's vulnerability to a sudden funding shock.

All the above are time-dependent covariates which are clearly predictable due to the 1-month lagging. In case of missing monthly values, the latest quarterly observation was substituted as a proxy, and if also this was missing, the latest yearly observation was used. In addition, we examined the influence of the firm's book asset value, its equity value, and the value of its fixed assets. The book assets and the pledgeable assets were highly correlated, and including both in the analysis would give problems with collinearity. A preliminary analysis showed that our estimates are almost identical regardless of which one we choose. The equity return was preferred over the equity value since the latter seemed to cause instabilities during estimation and unintuitive results due to excessively large values and high correlation with other covariates. The fixed assets also caused instabilities during estimation, and were dropped altogether from the covariate specification.

The macroeconomic covariate specification consisted of the following 7 variables obtained from CRSP and the U.S. Federal Reserve Board:

- **Trailing 1-year return on the S&P 500-index** for the previous month. The change on this index is considered one of the major predictors for the future state of the U.S. economy.

- **Spread between yields on Moody's Baa- and AAA-rated corporate bonds** for the previous month. This is a measure of the credit risk that the market is factoring on lower grade bonds. A widening usually suggests that the market is forecasting greater credit risk due to a slowing economy.
- **Trailing 1-year percentage change in U.S. Consumer Price Index** (CPI) for the previous month. The index measures the average price of consumer goods and services purchased by households and its percentage change thus measures the level of inflation.
- **Trailing 1-year percentage change in average weekly earnings** for the previous month. This is an indicator of short-term earnings growth.
- **Trailing 1-year percentage change in U.S. domestic crude oil First Purchase Price** (FPP) for the previous month. High oil prices might have a negative impact on U.S. economic growth.
- **Spread between the 10- and 1-year U.S. Treasury yields** for the previous month. This measures the extra cost of holding long-term debt compared to the cost of holding short-term debt.
- **Trailing 1-year percentage change in U.S. unemployment rate** (UR) as percentage of civilian labor force (seasonally adjusted) for the previous month. The rate itself is a major macroeconomic indicator, and its change measures the difference between labour relationships newly broken and labour relationships newly initiated.

As for the idiosyncratic covariates, predictability is ensured through the 1-month lagging. Alternative versions of the above macroeconomic variables were also considered along with the U.S. industrial production and several versions of the U.S. gross domestic product (GDP), but only the above showed some signs of reasonable significance or did not cause instability during estimation.

Lastly, the following global dynamic covariate will be used:

- **Trailing monthly default rate** for the previous month. The trailing monthly default rate is calculated at each month as the realized number of defaults during that month relative to the number of firms at risk at the beginning of the month.

  Including the lagged version, and not the monthly default rate itself, ensures predictability. This dynamic covariate may at a given time be viewed as a reflection of the instantaneous default risk in the cohort – a significant effect may thus suggest missing global effects driving default intensities upwards.

[Figure 1 about here.]

The at-risk and empirical default patterns of the study cohort are illustrated in Figure 1. The upper left panel shows that the final dataset contains a minimum of 1,005, an average of 1,142, and a maximum of 1,363 firms at-risk at any given point in the study time. The upper and lower right panels show the effect on the cohort of the recessions in the U.S. economy. The monthly default rate reaches its maximum of 0.611% during the '01-'02 recession. Note that the lower right panel also depicts the path of the trailing monthly default rate used in the analysis in a 1-month lagged version as a global dynamic covariate.

[Figure 2 about here.]

Figure 2 shows nonparametric Nelson-Aalen estimates of the default intensity in the cohort due to the passing of time itself – i.e. without correcting for covariate effects. The left panel shows the cumulative estimate,

From this graph, the slope seems fairly constant except for during the recessions of '91-'92 and '01-'02. The smoothed plot in the right panel, obtained through kernel smoothing of the increments of the Nelson-Aalen estimator, is better at showing the time-variation. It captures the same tendencies as the empirical default patterns in Figure 1 and clearly shows that the default intensity is much larger during the recessions compared to periods of economic upturns. During the '01-'02 recession the smoothed monthly default intensity peaks at just over 3%.

Table 1 shows summary statistics for the covariates included in our analysis and we indicate the expected sign of the effects of each covariate on the default intensities of firms. The firm-specific covariates are grouped with respect to defaults and non-defaults to give a rough idea of whether the values of covariates of defaulting firms differ from those of non-defaulted firms in a way consistent with the expected sign of the effect of the covariate. This seems to be the case, although the ratio of short-to-long term debt is seen to be of almost the same magnitude (or maybe even slightly greater) for non-defaulting firms compared to defaulting firms across all statistics.

Note that the table shows extreme maximum values for both the quick ratio, the pledgeable assets, and the 1-year equity return covariates when compared to their 75% quantiles. The influence of these potential outliers will be examined during model check in Section 6.

[Table 1 here in separate landscape page.]

[Figure 3 about here.]

4.2 Nonparametric Aalen analysis

Initially, a nonparametric additive model including time-varying effects for all five idiosyncratic covariates was fitted using the `aalen` function from the `timereg` package in `R` (R Development Core Team (2011)) with robust standard errors. Apart from the additive specification replacing the Cox specification used by e.g. Lando and Nielsen (2010), our main departure in this initial fit is the time-varying coefficients on the idiosyncratic covariates and a general time-varying baseline intensity.

As mentioned in Section 2.1, it is impossible to simultaneously identify a time-varying baseline and the time-constant effects of global (both macroeconomic and dynamic) covariates when the time-scale is calendar-time. We therefore initially fit the model with a time-varying baseline, which can be viewed as a time-varying proxy for all global tendencies. Afterwards, in Section 4.3, we consider the semiparametric submodel where the time-varying baseline is replaced by a constant term and where the 7 macroeconomic covariates and the global dynamic covariate all have time-constant effects.

Figure 3 illustrates the first interesting observation of our study: The 1-year distance-to-default is, in the terminology of Kalbfleisch and Prentice (2002, p. 199), "responsive," in the sense that it alone weakened the effect and increased the misspecification of both the quick ratio and the pledgeable assets. Such behavior is a sign of missing interactions in the model.

Estimated cumulative regression coefficients $\widehat{B}_j(t) = \int_0^t \widehat{\beta}_j(s)\, ds$ for the quick ratio and the pledgeable assets from the reference model without interactions are shown in Figure 3 with confidence bands along with the corresponding estimates from an extended model including interactions between the two and the 1-year

distance-to-default. The estimates from the model not including interactions are considerably smaller than the estimates from the extended model and have robust confidence bands which are much wider than the bands based on martingale theory. Including the interaction terms corrects this behavior. In fact, the estimates from the extended model are almost identical to estimates from a model not including the 1-year distance-to-default (not shown). Note that the interactions are economically plausible, as it is reasonable to expect that the effect of a firm's distance-to-default on default risk depends on asset level and short-term liquidity. However, even in the model including interactions, the robust confidence bands for the quick ratio coefficient are still somewhat larger near the end of the study time than the ones based on martingale-theory, suggesting that there is still some misspecification with regards to this covariate. Inspired by this finding, we tried all other possible combinations of two-way interactions, but none seemed to improve the overall fit.

[Figure 4 here on separate page.]

Estimated cumulative regression coefficients from the extended model including the two interactions are shown with confidence bands in Figure 4 for all but the quick ratio and the pledgeable assets, which were given in the right panels of Figure 3. The effect of each of the firm specific covariates is as expected. The percentage short-term debt is seen to increase default intensities, but the effect seems to wear off near the end of the study time, from around 2004 and onwards. On the other hand, default intensities fall as we increase the 1-year equity return, the 1-year distance-to-default, the quick ratio, or the book asset value. The pledgeable assets are unimportant until about 1987, or for nearly the first 5 years of the study time. The cumulative estimates for the interactions are both increasing, but have much lower magnitude than the effects of the 1-year distance-to-default, the quick ratio, and the (log) pledgeable assets, so that they only dampen, or correct, but do not completely remove the marginal effects of the covariates. The baseline is increasing suggesting that, given the covariate specification used here, it proxies for default increasing effects.

[Table 2 about here.]

Judging from the confidence bands of the cumulative coefficients, all effects are significant over most of the study time. The robust bands for all but the quick ratio and its interaction with the 1-year distance-to-default are almost identical to the martingale-based bands, suggesting limited model misspecification with respect to the rest of the covariates.

To determine if the changes in the slopes of the cumulative estimates are significant, the observed test processes (3.11) for the extended model including interactions are plotted in Figure 5 along with 50 resampled test processes under the null of time-invariance. Only the observed test processes corresponding to the short-to-long term debt ratio, the 1-year distance-to-default, and the interaction between the latter and the pledgeable assets truly exhibit extreme behavior over longer periods of time compared to the resampled test processes. The rest are at best borderline extreme, and only over limited periods of time. Hence, we only expect the three mentioned effects to be significantly time-varying.

The graphical considerations are supported by test statistics and associated P-values based on 1,000 resamples of each cumulative coefficient in Table 2. All the estimated cumulative coefficients are highly significant. The Kolmogorov-Smirnov and Cramer-von Mises tests of time-invariance agree fairly well for all cumulative coefficients, and the hypothesis of a time-constant effect is only clearly rejected for the effects of

the short-to-long term debt ratio, the 1-year distance-to-default, and the interaction between the latter and the (log) pledgeable assets. The test results correspond well with the impressions from the plots.

Note that while the overall conclusions are comparable to findings by e.g. Lando and Nielsen (2010), we obtain a description of how the effects of covariates vary with time and how this may cause model misspecification.

[Figure 5 here in seperate landscape page.]

[Table 3 about here.]

4.3 Semiparametric Aalen analysis

The next step is to introduce the macroeconomic covariates and the global dynamic covariate instead of the general time-varying baseline intensity in a semiparametric Aalen model. As mentioned above, to identify the parameters, this means that we have to use a constant baseline intensity. Based on the results from the nonparametric Aalen model, the semiparametric model is fitted allowing for time-varying effects of the short-to-long term debt ratio, the 1-year distance-to-default, and the interaction between the 1-year distance-to-default and the (log) pledgeable assets.

Table 3 shows the time-constant parameter estimates from this model, while estimated regression coefficients $\widehat{\beta}_j(t)$ for the short-to-long term debt ratio, the 1-year-distance-to-default and the interaction between the latter and the (log) pledgeable assets are shown as smoothed functions with confidence bands in Figure 6.

Consider first the time-constant effects. The most pronounced time-constant default decreasing effect is the 1-year equity return, which when increased by 1% lowers intensities by 1.52 percentage points. Note, however, that using robust standard errors, the effects of the quick ratio and the interaction between the quick ratio and the 1-year distance-to-default are insignificant. The robust standard errors for these two effects are also seen to be somewhat larger than the martingale-based standard errors. This shows that the misspecifications first observed in Figures 3 and 4 are also present for the corresponding time-constant effects. In fact, had the Wald test statistics for these two effects been calculated using the martingale-based standard errors, we would have mistakingly judged them as being clearly significant. The robust standard errors for the rest of the time-constant coefficients, both idiosyncratic, macroeconomic, and dynamic, are, however, almost identical to the standard errors based on martingale theory.

[Figure 6 about here.]

The table also shows that the global dynamic effect is of large magnitude and highly significant, suggesting a high level of unobserved variation in the cohort: A 1% increase in the monthly default rate implies an excess risk of 6.97 percentage points, which is a large shift, but plausible given that the observed monthly default rate in the data has a maximum of 0.611%. On the other hand, none of the macroeconomic effects are significant at conventional levels. A slight indication of a default decreasing effect is found for the 1-year CPI change, but only at the 10% level. In contrast to what is expected from univariate economic reasoning, the Baa-AAA yield spread decreases intensities, but the effect is very small, and its sign is therefore not a big concern. Finally, note that the 1-year return on the S&P500 index has, as expected, a default decreasing effect in the additive model,

although not significant at conventional levels. This particular covariate has been reported in other studies to have the "wrong" sign in Cox models, see for example Duffie et al. (2009) and Lando and Nielsen (2010).

Turning to the time-varying coefficients $\widehat{\beta}_j(t)$ in Figure 6, the variational pattern over time is in all three cases closely connected to the recessions in the U.S. economy, with all three effects being most pronounced during the recessions as compared to in-between periods. The magnitude of the short-to-long term debt ratio effect varies greatly depending on the recession at hand: A 1% increase corresponds to an excess risk of around 7.00 percentage points during the '91-'92 recession, while the same increase implies an excess risk of almost 14.00 percentage points, or twice as much, during the '01-'02 recession. The effect of the 1-year distance-to-default is, on the other hand, fairly much the same during both mentioned recessions, where an increase of one standard deviation lowers intensities by about 3 percentage points. Comparing this with the time-constant estimates in Table 3, the 1-year distance-to-default has the most influential default decreasing effect of all included covariates. The effect of the interaction is a time-varying correction to the marginal effects of the 1-year distance-to-default and the pledgeable assets.

5 Model check

Despite the existence of several goodness-of-fit methods for survival models, actual model checking is still somewhat overlooked in practice. The problem is that classical model check procedures looking at residuals all originate from similar methods used in ordinary statistics, but the way in which they are to be interpreted in the survival data setup with censoring is unclear. For instance, the usual martingale residuals first considered by Barlow and Prentice (1988) are not normally distributed – hence, the interpretation of such residuals is not as straightforward as in classical regression analysis, and it may therefore be difficult to judge the quality of a model-fit from a plot of the residuals.

Nevertheless, model checking may actually lead us to alter the model specification. In our case, it leads us to conclude that potential problems with the fit of the additive model is due to over-influence of firms with extreme covariate values. Before presenting these findings, we describe graphical model checking methods for the nonparametric Aalen model.

An advantage of the nonparametric Aalen model is that its additive structure fits well with martingale theory, producing residual processes which are exact local martingales when the model is correct – this is neither the case for the semiparametric additive model nor for the Cox model, for which the residual processes are only approximate (asymptotic) local martingales. As the semiparametric Aalen model is a submodel of its nonparametric counterpart, we will focus on model-check for the nonparametric model.

Aalen et al. (2008) discuss a graphical method based on so-called "martingale residual processes" – a method that has also been applied to the Cox model. Using the nonparametric additive model form (3.4), define the martingale residual process $(\mathbf{M}_{\text{res},t})_{t\in[0,T]}$ as the accumulated difference between the vector of counting processes and the vector of estimated cumulative intensity processes at the time points where the model is estimable,

$$\mathbf{M}_{\text{res},t} = \int_0^t J(s)\, d\mathbf{N}_s - \int_0^t J(s)\, \mathbf{X}_s\, d\widehat{\mathbf{B}}(s). \tag{5.1}$$

Inserting the expression (3.5) for the estimator $\widehat{\mathbf{B}}(t)$, applying the model form (3.4), and using the definition of the least squares generalized inverse $\mathbf{X}_t^-$, we get

$$\begin{aligned}\mathbf{M}_{\text{res},t} &= \int_0^t J(s)\left(\mathbf{I}-\mathbf{X}_s\mathbf{X}_s^-\right)\, d\mathbf{N}_s = \int_0^t J(s)\left(\mathbf{I}-\mathbf{X}_s\mathbf{X}_s^-\right)\left(\mathbf{X}_s\, d\mathbf{B}(s)+d\mathbf{M}_s\right)\\ &= \int_0^t J(s)\left(\mathbf{I}-\mathbf{X}_s\mathbf{X}_s^-\right)\, d\mathbf{M}_s.\end{aligned}$$

Under the assumption that $\mathbf{X}_t$ is locally bounded, this proves that the process $(\mathbf{M}_{\text{res},t})_{t\in[0,T]}$ is a vector-valued local martingale when the nonparametric additive model is true. As mentioned, this exact result does not carry over to semi-parametric models – whether of the Aalen or the Cox type.

5.1 Covariate misspecifications: Grouped martingale residual processes

The firm specific martingale residual processes are typically not of much use on their own due to the relatively few number of recurrent default events. It is more useful to aggregate over groups of firms with respect to covariate values and conduct graphical model-check based on the grouped residual processes. Specifically, assume that a grouping of the firms is given, and let $G(t)$ denote the set of all firms belonging to group G at time t. The grouping is allowed to depend on time, such that firms may move from one group to another during the study time, as would naturally be the case when grouping is based on time-dependent covariates. It is, however, essential that the grouping is predictable, such that information needed to group firms at time t is available just before time t. The martingale residual process of group G at time t then takes the form

$$M_{\text{res},t}^{(G)} = \int_0^t \sum_{i\in G(s)} dM_{\text{res},it}, \tag{5.2}$$

where $M_{\text{res},it}$ is the ith element of the time-t vector $\mathbf{M}_{\text{res},t}$ in (5.1). These grouped residual processes may then be plotted as functions of study time in order to assess the model fit with respect to covariates. If the model fits well, the resulting plots will fluctuate around zero and show no overdispersion or particular trends.

5.2 The model fit as a whole: Model-based covariance of residual processes

The martingale residual processes may also be of use in a different manner, aimed at judging the model fit as a whole. The method is explained by Aalen et al. (2008), and involves estimating a model-based covariance function for the martingale residual processes. The idea is that if we standardize martingale residual processes by dividing them with their estimated standard deviation at each time t, then the time-varying mean and standard deviation of the standardized processes should stay close to 0 and 1, respectively, if the model fits well.

To elaborate, assume that the nonparametric additive model is true, so that the process $(\mathbf{M}_{\text{res},t})_{t\in[0,T]}$ is a vector-valued local martingale. Its covariance function may then be estimated by its optional variation process

$$\boldsymbol{\Sigma}_{\text{res}}(t) = \sum_{\tau_k\le t} J(\tau_k)\left(\mathbf{I}-\mathbf{X}_{\tau_k}\mathbf{X}_{\tau_k}^-\right)\,\text{diag}(\mathbf{X}_{\tau_k}\mathbf{X}_{\tau_k}^-\triangle\mathbf{N}_{\tau_k})\,\left(\mathbf{I}-\mathbf{X}_{\tau_k}\mathbf{X}_{\tau_k}^-\right)^T. \tag{5.3}$$

Standardized residual processes are then obtained at each time t by calculating

$$\frac{M_{\text{res},it}}{\sqrt{\sigma_{\text{res},ii}^2(t)}};\quad i=1,\ldots,n,$$

where $\sigma_{\text{res},ii}^2(t)$ is the ith diagonal element of $\boldsymbol{\Sigma}_{\text{res}}(t)$. A plot of the mean and standard deviation of the standardized residual processes as a function of time will then give a graphical assessment of the model fit as a whole: If the model gives a reasonable fit, the mean will stay close to 0, and the standard deviation to 1.

5.3 Outliers and over-influence: Diagonal cumulative hat processes

The leverages are a useful diagnostic for finding observations that may be overly influential in linear regression models. The leverages are the diagonal elements of the so-called hat matrix, which is given by the product of the design matrix and its least squares generalized inverse. For the additive model, we define a cumulative hat matrix as

$$\mathbf{H}_{\text{cum}}(t) = \sum_{\tau_k \leq t} \mathbf{X}_{\tau_k} \mathbf{X}^{-}_{\tau_k}. \tag{5.4}$$

The idea is to look for observed firm processes with particularly high influence by plotting the diagonal elements $h_{\text{cum},ii}(t)$ of $\mathbf{H}_{\text{cum}}(t)$, called the diagonal cumulative hat processes, as functions of the successive default times in the cohort.

In ordinary linear regression, the average value of the diagonal elements of the hat matrix is r/n, where r is the rank of the hat matrix, which is the number of independent parameters of the model, and n is the number observations. Diagonal elements well above r/n are said to have high leverage, and Aalen et al. (2008), amongst others, recommend investigation of observations with diagonal hat matrix element above $2r/n$.

With respect to the nonparametric additive model, this theory may be applied at each default time in the cohort. Hence, the outlying process criterion is

$$h_{\text{cum},ii}(t) > 2r \sum_{\tau_k \leq t} \frac{1}{Y_{\bullet\,\tau_k}}, \tag{5.5}$$

where $Y_{\bullet\,\tau_k}$ is the number of firms at risk at default time τ_k.

6 Revising the model

The initial analyses of Section 4 using the Aalen models gave intuitive results about the influence of the included risk factors that are largely consistent with results from Cox models found by e.g. Duffie et al. (2009) and Lando and Nielsen (2010). There were, however, also signs of model misspecifications. First, robust standard errors were considerably larger than regular standard errors for several covariates, even after correcting for interactions. Second, several significantly time-varying effects where identified, indicating that the time-invariant proportionality assumption of the Cox model applied elsewhere is a misspecification for these covariates. Finally, the significance of the global dynamic effect suggests that the covariate specification applied here is missing essential elements. Hence, there is good reason for a deeper analysis of the misspecifications and the data. All results presented in this section are based on own implementations in `R` (R Development Core Team (2011)).

6.1 Martingale residual processes

Graphical assessment of martingale residual processes, suitably grouped with respect to the values of covariates, may give a good indication of functional form misspecifications and outliers due to extreme covariate values. We study the martingale residual processes of the nonparametric additive model including all five idiosyncratic covariates and the two interactions.

Figure 7 shows grouped martingale residual processes (5.2) for each of the five idiosyncratic covariates. Grouping is done at each default time according to the quantiles of each covariate, with 50 equidistant groups

in each case: Group 1 corresponds to values between the 0% and 2% quantiles, group 2 corresponds to values between the 2% and 4% quantiles, and so on. This allowed for the isolation of the very extreme covariate values while maintaining a reasonable amount of firms in each group at each default time.

If the model fits well, all residual processes would stay close to zero and not show signs of over-dispersion or trends. The impression from the figure is therefore that deviations from the model mostly occur in the extreme (both low and high) quantiles of the covariates, with especially the high quantiles causing misspecifications. This is expected, but the plots allow us to give a detailed account of the misspecifications. More defaults than expected occur in the highest quantile groups of the quick ratio, the 1-year equity return, and the 1-year distance-to-default. There are slightly more defaults than expected in the highest quantile group of the short-to-long term debt ratio, but the deviation is limited to around the year 2001, which had a historically high default rate. On the other hand, fewer defaults than expected occur in the lowest quantile groups of the 1-year-equity return and the 1-year distance-to-default, while it is in the highest quantile groups of the (log) pledgeable assets that fewer defaults than expected occur. In the latter case, the deviation is again limited to around the year 2001. Most of the non-extreme quantile groups show no truly alarming signs.

The deviations in the high quantile groups of the quick ratio, the 1-year equity return, and the 1-year distance-to-default are either due to outliers, or due to certain firms having excessively high values of these covariates, yet still defaulting. Analogously, the deviations in the lowest quantile groups of the 1-year equity return and the 1-year distance-to-default are due to certain firms having negative values for these covariates, yet still not defaulting. The deviation in the highest quantile groups of the (log) pledgeable assets is somewhat unintuitive, but may be due to a considerable amount of defaults amongst firms with relatively high values of pledgeable assets around the year 2001.

The deviations seen in the figure suggest that simple transformations or inclusion of, say, quadratic terms, would not correct the misspecifications, since they are limited to the extreme quantile groups. Transformations yielding reasonable residual processes would have to "squeeze" together the distribution of the covariates, so that outlying values are neutralized, and such would tend to blur effects. Transformations of the form $x \mapsto -e^{-ax}$ for $a > 0$ were used in a preliminary version of this paper. Hence, a detailed analysis of influential observations, particularly ones due to extreme covariate values, is necessary.

[Figure 7 here on separate page.]

[Figure 8 about here.]

6.2 Model-based covariance of residual processes

The grouped martingale residual processes revealed a considerable amount of model misspecification at the extreme quantiles of the covariates. However, looking back at the results from Section 4.2, robust standard errors were quite close to martingale-based standard errors for most covariates, suggesting that the misspecifications might not be of crucial overall importance for the final results.

To get an idea if the model fits as a whole, we standardize the martingale residual processes using the model-based covariance function (5.3) and compare in Figure 8 their time-varying mean and standard deviation with what would be the true values if the model fit well: The mean should stay close to 0 while the standard

deviation should stay close to 1 if the model gives an adequate description of the data. The figure thus shows a misspecification with respect to the standard deviation of the standardized residual processes, which increases with time to a maximum value of just over 2. Hence, the standard deviation of the calculated residual processes are greater than is expected under the model. The plot therefore shows that while the additive model on average captures the patterns present in the data throughout the study time, the magnitude of variation in the data which the model fails to explain is considerable. It thus seems that while the model is not entirely wrong in its description of the defaults in the data, it apparently lacks essential explanatory risk factors.

[Figure 9 about here.]

6.3 Diagonal cumulative hat processes

To check whether the deviations from martingale-behavior seen in Figures 7 and 8 are due to over-influence of some firms in the cohort, we calculated the cumulative hat matrix (5.4) for the nonparametric additive model as a function of the successive default times in the cohort.

The left panel of Figure 9 shows the 2,557 diagonal cumulative hat processes corresponding to each firm in the cohort along with the expected hat matrix and the outlying process criterion (5.5). Several processes exceed the outlying process criterion, meaning that these firms have the highest influence on the results of the analysis. In total, 177 processes were identified as exceeding the outlying process criterion at some default time, and these are emphasized in the right panel of the figure. These firms were the ones with the most extreme (both high and low) values of the idiosyncratic covariates. Among those firms there were 19 defaults distributed across 16 firms, with 3 firms defaulting twice.

When examining the 177 firms more closely, it was clear that they all have some covariate values that are outliers which, for example, were many magnitudes higher than the maximum in the rest of the cohort. This identification of over-influential firms confirms the observations from the study of the grouped martingale residual processes, where it was noted that deviations from the additive model were primarily limited to the extreme quantile groups.

6.4 Analysis based on clean data

Based on the results from the model check, we conduct an analysis of the data, omitting the 177 firms from the cohort who were identified as over-influential. Omitting these firms resulted in a "clean" data set consisting of 2,380 firms, comprising a total of 351 defaults. The 19 omitted defaults were fairly evenly distributed across study time, and an analysis of the at-risk and default patterns in the clean data set showed no particular changes compared to the full data set.

Model check for the cleaned data

Grouped martingale residual processes for a nonparametric additive model fitted to the clean data, including the two interactions, showed much less dispersion in the extreme quantile groups and generally a better fit across all covariates. As shown in Table 4, and detailed below, martingale-based and robust standard errors are also in close correspondence. The standard deviation of the standardized residual processes did not change much

compared to the analysis on the full data set. There may therefore be room for additional covariates which we have not considered. Overall, however, the model fit was considerably improved by omitting outliers in the data. The trade-off is, of course, that the model is then not fitted to the full data set. On the other hand, we avoid ad-hoc transformations of the functional form that are sensitive to a few extreme outliers.

Model fits for the cleaned data

[Table 4 about here.]

A nonparametric Aalen model was fitted to the clean data including all five idiosyncratic covariates and the two interactions. Only the effects of the 1-year distance-to-default and the short-to-long term debt ratio had significantly time-varying effects – the hypothesis of time-invariance was clearly not rejected for the rest of the effects. Hence, cleaning the data removed the time-variation of the effect of the interaction between the 1-year distance-to-default and the (log) pledgeable assets compared to the analysis on the full data set.

Based on these findings, a semiparametric additive model was fitted to the clean data with time-varying effects for the 1-year distance-to-default and the short-to-long term debt ratio, while the rest of the idiosyncratic covariates had time-constant effects. The time-varying baseline was replaced by a constant term, the macroeconomic covariates, and the global dynamic covariate. Table 4 gives the estimates for the time-constant effects. The smoothed coefficients for the time-varying effects of the 1-year distance-to-default and the short-to-long term debt ratio did not change compared to what is given in Figure 6, and are not repeated here.

Consider first the time-constant idiosyncratic effects in Table 4. Comparing with the results in Table 3 for the semiparametric model fitted to the full data set, the effects of the quick ratio and its interaction with the 1-year distance-to-default are now of considerably larger magnitude, and their robust standard errors are in good correspondence with the martingale-based standard errors. Both effects are therefore significant regardless of which standard errors we use to construct test statistics. Thus, cleaning the data has corrected the misspecification first observed for these effects. In fact, the quick ratio now has the most pronounced default decreasing time-constant effect, lowering default intensities by 2.18 percentage points per unit increase.

Finally, with respect to the global effects, the borderline significant effects of the 1-year CPI change found in Table 3 is now clearly insignificant at conventional levels. On the other hand, the global dynamic effect is still highly significant, and its magnitude has not changed much compared to the model fitted to the full data set. In conclusion, the significance of the global dynamic effect was not an artifact of over-influence of certain firms. It is conceivable, that the dynamic covariate captures an effect related to the business cycle which our macro variables do not capture and which is not transmitted through the firm specific covariates. Note that this fits well with the analysis of the standardized martingale residual processes which indicated that the model is missing essential risk factors, even after cleaning the data. This is a topic of future research.

7 Conclusion

This paper has studied additive models for stochastic default intensities. We model the mean arrival rate of default events using nonparametric and semiparametric versions of Aalen's additive regression model. Using both firm-specific and global covariates, we fit the models and conduct model check on a sample of rated, non-financial U.S. corporates, covering the period 1982 to 2005.

In additive models, covariates act in an additive way on a baseline intensity. The nonparametric model has time-varying effects for all covariates, while the semiparametric model allows for some parametric, time-constant effects. This makes the additive models more flexible than the often applied Cox model in describing covariate effects and how they may change over time.

An advantage of the Cox models is that their log-linear structure ensures that estimated intensities are positive. The way we estimate the Aalen model does not rule out negative values of estimated intensities. Negative intensities are inevitable for extrapolations using extreme values of covariates, but even for data with firms that have very small default intensities, estimated intensities for observed covariates may be negative. The possibility of negative intensities and the non-parametric nature of the intensity function imply that we should not think of Aalen models as a tool for prediction. Its purpose is to analyze how macro-economic and firm-specific covariates have affected the default intensity in the past. The strength of the Aalen model is its ability to model time-varying effects of covariates in a flexible, non-parametric manner.

We present the theory behind estimation and inference for both nonparametric and semiparametric additive models, including how to test whether a covariate effect is time-varying. We use model checking techniques based on visual inspection of martingale residual processes that help us identify model misspecifications, as well as a method for identifying outliers that is very similar to what is known from ordinary linear regression. Our model checking strongly influences the final model specification.

In our final model we find evidence of time-variation in the effects of distance-to-default and short-to-long term debt, and we identify the effect of interactions between distance-to-default and two other covariates: the quick ratio and (log) pledgable assets. In our final specification, which excludes outliers, the effect of the interaction terms is not time-dependent. Furthermore, the quick ratio covariate is significant. None of our macroeconomic covariates are significant which may indicate that their effects are captured through their influence on firm-specific covariates.

Figure legends

Figure 1. At-risk and empirical default patterns of study cohort. *Upper left panel: Number of firms at-risk at each time point in the study time, with the time-average of 1,142 firms at-risk indicated by the horizontal line. Upper Right panel: Yearly number of defaults. Lower right panel: Monthly default rate in % at each time point in the study time.*

Figure 2. Nelson-Aalen estimates of cohort default intensity. *Left panel: Estimated cumulative default intensity due to the passing of time itself with approximate 95% martingale-based pointwise confidence bands. Right panel: Smoothed default intensity due to the passing of time itself with approximate 95% martingale-based pointwise confidence bands. Smoothing done using the Epanechnikov kernel with bandwidth 1.1 years and boundary correction at the left endpoint of the study time.*

Figure 3. Missing interactions. *Estimated cumulative regression coefficients for the quick ratio and (log) pledgeable assets covariates with martingale-based* 95% *confidence bands (dotted lines) and robust* 95% *confidence bands (dashed lines). Left panels: Estimates from the model without interactions. Right panels: Estimates from extended model including interactions. Note that the value axis' are not the same for the two sets of plots.*

Figure 4. Cumulative regression coefficients. *Estimated cumulative regression coefficients for the baseline, the 1-year equity return, the 1-year distance-to-default, the short-to-long term debt ratio, and the interactions of the extended nonparametric additive model including interactions with martingale-based* 95% *confidence bands (dotted lines) and robust* 95% *confidence bands (dashed lines). The estimates form this model for the quick ratio and the (log) pledgeable assets are shown in the right panels of Figure 3.*

Figure 5. Testing for time-invariance. *Observed test processes (3.11) for the cumulative regression coefficients from the extended nonparametric additive model including interactions (thick black lines), along with 50 resampled test processes under the null of time-invariance (thin grey lines).*

Figure 6. Smoothed regression coefficients. *Smoothed estimates (3.9) of regression coefficients for the short-to-long term debt ratio, the 1-year-distance-to-default, and the interaction between the latter and the (log) pledgeable assets from the semiparametric model with martingale-based* 95% *confidence bands (dotted lines) and robust* 95% *confidence bands (dashed lines). Smoothing done using the Epanechnikov kernel with bandwidth 1.3 years and boundary correction at the left endpoint of the study time.*

Figure 7. Covariate misspecifications. *Grouped martingale residual processes (5.2) for the nonparametric Aalen model including all five idiosyncratic covariates and the two interactions. Grouping done at each default time according to the quantiles of each idiosyncratic covariate, with 50 equidistant groups in each case: Group 1 corresponds to values between the 0% and 2% quantiles, group 2 corresponds to values between the 2% and 4% quantiles, and so on. Groups with outlying residual processes are marked by group number.*

Figure 8. The model fit as a whole. *Mean (lower curve) and standard deviation (upper curve) of standardized martingale residual processes as functions of study time from the nonparametric Aalen model including all five idiosyncratic covariates and the two interactions. The true values under the model, 0 and 1, respectively, are indicated by the dotted lines.*

Figure 9. Outliers and over-influence. *Analysis of over-influence through the cumulative hat matrix (5.4) of the nonparametric additive model including all five idiosyncratic covariates and the two interactions. The plots show the 2,557 firm-specific hat processes (thin grey lines) as functions of the default times in the cohort, along with the expected hat matrix (lower thick solid line) and the outlying process criterion (upper thick solid line). In the right plot, the 177 processes which at some default time exceed the outlying process criterion are emphasised (thin dotted lines).*

Figure 10. Model-based mean default intensity. *Estimated mean monthly default intensity from the semiparametric additive model fitted to the clean data. Estimate obtained by averaging the idiosyncratic covariate values over all firms at risk at each monthly observation time (and also including the global covariates) and applying the semiparametric additive form (3.2). Estimates of time-constant regression coefficient were given in Table 4. Estimated cumulative regression coefficients for the 1-year distance-to-default and the short-to-long term debt ratio were smoothed using the Epanechnikov kernel with bandwidth 1.4 years and boundary correction at the left endpoint of the study time.*

References

O. O. Aalen. A model for non-parametric regression analysis of life times. In J. Rosinski W. Klonecki, A. Kozek, editor, *Mathematical Statistics and Probability Theory*, volume 2 of *Lecture Notes in Statistics*, pages 1–25, New York, USA, 1980. Springer-Verlag.

O. O. Aalen. A linear regression model for the analysis of life times. *Statistics in medicine*, 8:907–925, 1989.

O. O. Aalen, H. W. Fekjær, Ø. Borgan, and E. Husebye. Dynamic analysis of multivariate failure time data. *Biometrics*, 60(3):764–773, September 2004.

O. O. Aalen, Ø. Borgan, and H. K. Gjessing. *Survival and Event History Analysis: A Process Point Of View*. Springer-Verlag, New York, USA, 1st edition, 2008.

E. I. Altman. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23:589–609, 1968.

P. K. Andersen and R. D. Gill. Cox's regression model for counting processes: A large sample study. *Annals of Statistics*, 10:1100–1120, 1982.

W. E. Barlow and R. L. Prentice. Residuals for relative risk regression. *Biometrika*, No. 75:65–74, 1988.

W. Beaver. Financial ratios and the prediction of failure. *Journal of Accounting Research. Supplement : Empirical research in accounting : Selected studies 1966*, 4:77–111, 1966.

S. Chava, C. Stefanescu, and S. Turnbull. Modeling the loss distribution. *Management Science*, 57(7):1267–1287, 2011.

F. Couderc and O. Renault. Time-to-default : Life-cycle, global and industry cycle impacts. Unpublished manuscript, University of Geneva and Warwick Business School, 2004.

D. R. Cox. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 34:187–220, 1972.

S. Das, D. Duffie, N. Kapadia, and L. Saita. Common failings: How corporate defaults are correlated. *Journal of Finance*, 62:93–117, 2007.

S. A. Davydenko. When do firms default? A study of the default boundary. Unpublished manuscript, Rotman School of Management, University of Toronto, 2007.

D. Duffie, L. Saita, and K. Wang. Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83:635–665, 2007.

D. Duffie, A. Eckner, G. Horel, and L. Saita. Frailty correlated default. *The Journal of Finance*, LXIV(5): 2089–2123, October 2009.

S. Figlewski, H. Frydman, and W. Liang. Modeling the effect of macroeconomic factors on corporate default and credit rating transitions. *International Review of Economics & Finance*, 21(1):87–105, 2012.

P. Fledelius, D. Lando, and J. Perch Nielsen. Non-parametric analysis of rating transition and default data. *Journal of Investment Management*, 2(2):71–85, 2004.

J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. Wiley, Hoboken, N.J., 2nd edition, 2002.

D. Lando and M. S. Nielsen. Correlation in Corporate Defaults: Contagion or Conditional Independence? *Journal of Financial Intermediation*, 19(3):335–372, 2010. http://ssrn.com/paper=1338381.

D. Lando and T. Skødeberg. Analyzing Rating Transitions and rating Drift with Continuous Observations. *The Journal of Banking and Finance*, 26:423–444, 2002.

A. Le. Separating the components of default risk : A derivative-based approach. Unpublished manuscript, Kenan-Flagler Business School, University of North Carolina, 2007.

X. Li and X. Zhao. Macroeconomic effect in corporate default. Unpublished manuscript, York University, 2006.

I. W. MacKeague and P. D. Sasieni. A partly paramteric additive risk model. *Biometrika*, No. 81:501–514, 1994.

T. Martinussen and T. H. Scheike. *Dynamic Regression Models for Survival Data*. Springer-Verlag, New York, USA, 1. edition, 2006.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

H. Ramlau-Hansen. Smoothing counting processes by means of kernel functions. *Annals of Statistics*, 11: 453–466, 1983.

T. Robertson, F.T. Wright, and R. Dykstra. *Order restricted statistical inference*. Wiley series in probability and mathematical statistics: Probability and mathematical statistics. Wiley, 1988.

T. Shumway. Forecasting bankruptcy more efficiently : A simple hazard model. *Journal of Bus.*, 74:101–124, 2001.

C. A. Struthers and J. D. Kalbfleisch. Misspecified proportional hazard models. *Biometrika*, 73:363–369, 1986.

A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag New York, 1996.

H. White. Maximum Likelihood estimation of Misspecified Models. *Econometrica*, 50(1-25), 1982.

D. M. Zucker and A. F. Karr. Nonparametric survival analysis of with time-dependent covariate effects: A penalized partial likelihood approach. *Annals of Statistics*, 18:329–353, 1990.

Acknowledgments

The authors would like to thank Torben Martinussen, Thomas Scheike, Eric Renault, two anonymous referees, and participants at the SoFie Conference on liquidity, credit risk, and extreme events in Chicago, and at the annual D-CAF membership meeting, for helpful remarks and discussions.

Tables

Table 1. Covariate summary statistics.

Summary statistics for the five idiosyncratic, the seven macroeconomic, and the global dynamic covariate included in the analysis. The idiosyncratic covariates are grouped according to defaults and non-defaults, while the macroeconomic covariates and the global dynamic covariate are for the full cohort. Expected default intensity increasing effects are indicated with "(+)," while expected default intensity decreasing effects are indicated with "(÷)." Effects where the sign is not clear, even when based on univariate economic reasoning, are indicated with "(?)."

	Min	25% quant.	Median	Mean	75% quant.	Max	SE
Idiosyncratic covariates							
(÷) **Quick ratio**							
Defaults	0.005	0.540	0.898	1.677	1.392	900.900	11.750
Non-defaults	0.003	0.666	0.980	1.460	1.484	828.500	4.611
(÷) **Pledgeable assets** ($mio.)							
Defaults	1.244	138.900	304.800	1,168.000	911.800	42,120.000	3,109.598
Non-defaults	0.050	213.800	667.300	2,953.000	2,070.000	228,400.000	8,590.747
(÷) **1-year equity return** ($mio.)							
Defaults	−0.996	−0.408	−0.092	0.089	0.310	9.948	0.875
Non-defaults	−0.999	−0.161	0.092	0.228	0.410	9.980	0.748
(÷) **1-year distance-to-default**							
Defaults	−3.659	0.580	1.595	1.862	2.896	11.580	1.787
Non-defaults	−5.119	2.440	4.009	4.337	5.910	19.430	2.695
(+) **Short-to-long term debt**							
Defaults	0.000	0.026	0.086	0.185	0.254	1.000	0.233
Non-defaults	0.000	0.031	0.109	0.199	0.290	1.000	0.228
Macroeconomic covariates							
(÷) 1-year S&P 500 return ($mio.)	−0.275	−0.027	0.106	0.090	0.218	0.534	0.171
(+) Baa-AAA yield spread	0.550	0.740	0.900	1.010	1.210	2.690	0.385
(÷) 1- year CPI change (%)	1.100	2.300	3.000	3.052	3.700	6.400	1.090
(?) 1-year earnings change (%)	0.700	2.400	3.000	3.041	3.500	6.500	0.884
(?) 1-year oil price change (%)	−0.615	−0.152	−0.008	0.079	0.324	1.963	0.397
(+) Treasury yield spread	−0.370	0.490	1.260	1.337	2.140	3.290	0.976
(+) 1-year UR change (%)	−28.700	−8.900	−4.300	0.296	7.400	46.200	13.898
global dynamic covariate							
(+) Monthly default rate (%)	0.000	0.000	0.093	0.138	0.189	0.611	0.138

Table 2. Initial nonparametric Aalen analysis.

Supremum test of significance (3.10), Kolmogorov-Smirnov test of time-invariance (3.12) and Cramer-von Mises test of time-invariance (3.13) for the cumulative coefficients from the extended nonparametric Aalen model including interactions. Associated P-values are based on 1,000 resampled test processes.

Effect	Supremum statistic	*P*-value	Kolmogorov-Smirnov statistic	*P*-value	Cramer-von Mises statistic	*P*-value
Baseline	12.10	0.00	0.42	0.01	1.69	0.00
Quick ratio	7.50	0.00	0.04	0.03	0.01	0.05
Pledgeable Assets (log)	8.54	0.00	0.05	0.01	0.02	0.01
1-year equity return	13.50	0.00	0.03	0.14	0.00	0.27
1-year distance-to-default	12.70	0.00	0.11	0.00	0.11	0.00
Short-to-long term debt	8.00	0.00	0.23	0.00	0.52	0.00
Distance-to-default × Quick	7.74	0.00	0.00	0.39	0.00	0.32
Distance-to-default × Assets	9.95	0.00	0.01	0.00	0.00	0.00

Table 3. Initial semiparametric Aalen analysis.

Parameter estimates, regular (martingale-based) standard errors, robust standard errors, Wald test statistics, and associated P-values for the time-constant effects from the semiparametric Aalen model. Test statistics and P-values are based on robust standard errors. Effects are grouped according to covariate type: Idiosyncratic, macroeconomic, and dynamic.

Effect	Estimate	Standard error	Robust SE	Wald statistic	*P*-value
Quick ratio	−0.000590	0.000095	0.000468	−1.26	0.21
Pledgeable Assets (log)	−0.009850	0.001250	0.001330	−7.41	1.30×10^{-13}
1-year equity return	−0.015200	0.000916	0.001000	−15.20	0.00
Distance-to-default × Quick	0.000235	0.000024	0.000114	2.02	0.04
Baseline	0.115000	0.018200	0.018500	6.21	5.09×10^{-10}
1-year S&P 500 return	−0.021200	0.019200	0.019100	−1.10	0.27
Baa-AAA yield spread	−0.009660	0.006360	0.006220	-1.55	0.12
1- year CPI change	−0.003790	0.002280	0.002220	-1.71	0.09
1-year earnings change	0.001540	0.003050	0.003080	0.5	0.61
1-year oil price change	−0.004990	0.005910	0.005830	−0.86	0.39
Treasury yield spread	0.002290	0.003250	0.003170	0.72	0.47
1- year UR change	0.000343	0.000272	0.000274	1.25	0.21
Monthly default rate	0.069700	0.023800	0.023600	2.95	3.31×10^{-3}

Table 4. Semiparametric Aalen analysis of cleaned data.

Parameter estimates, regular (martingale-based) standard errors, robust standard errors, Wald test statistics, and associated P-values for the time-constant effects from a semiparametric Aalen model fitted to the clean data. Test statistics are based on robust standard errors. Effects are grouped with respect to covariate type: Idiosyncratic, macroeconomic, and dynamic.

Effect	Estimate	Standard error	Robust SE	Wald statistic	*P*-value
Quick ratio	−0.021800	0.002150	0.002580	−8.45	0.00
Pledgeable Assets (log)	−0.012500	0.001810	0.001840	−6.79	1.10×10^{-11}
1-year equity return	−0.015600	0.000993	0.001060	−14.72	0.00
Distance-to-default × Quick	0.004860	0.000421	0.000520	9.34	0.00
Distance-to-default × Assets	0.003200	0.000371	0.000379	8.44	0.00
Baseline	0.157000	0.023200	0.023200	6.77	1.31×10^{-11}
1-year S&P 500 return	−0.020700	0.022000	0.021900	−0.95	0.34
Baa-AAA yield spread	−0.009920	0.007620	0.007430	−1.34	0.18
1- year CPI change	−0.003120	0.002650	0.002570	−1.21	0.22
1-year earnings change	0.001610	0.003580	0.003610	0.45	0.65
1-year oil price change	−0.006590	0.006360	0.006260	-1.05	0.29
Treasury yield spread	0.003450	0.003670	0.003590	0.96	0.34
1- year UR change	0.000353	0.000304	0.000305	1.16	0.24
Monthly default rate	0.077200	0.026700	0.026600	2.90	3.70×10^{-3}

Figures

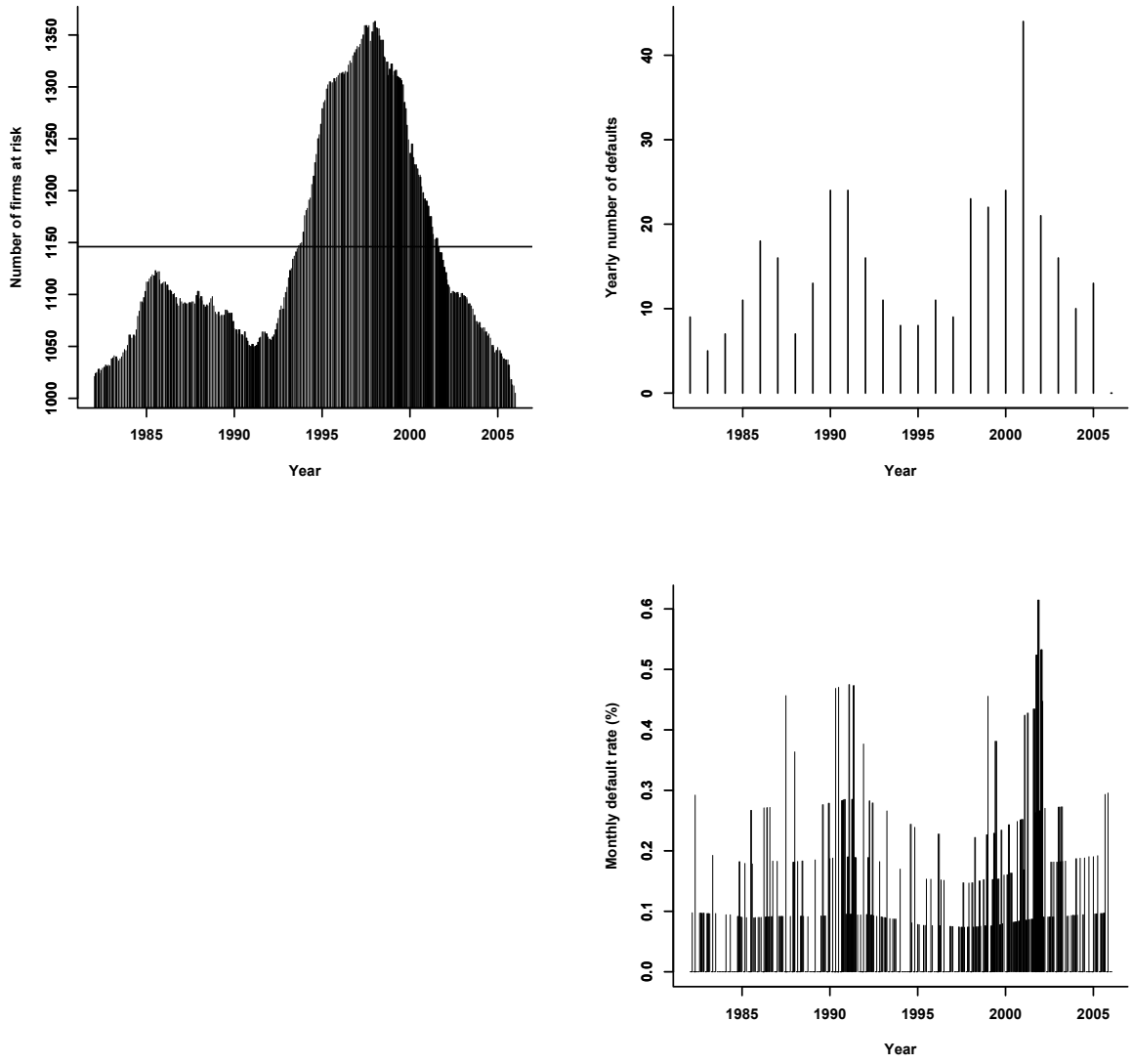

Figure 1. At-risk and empirical default patterns of the study cohort.

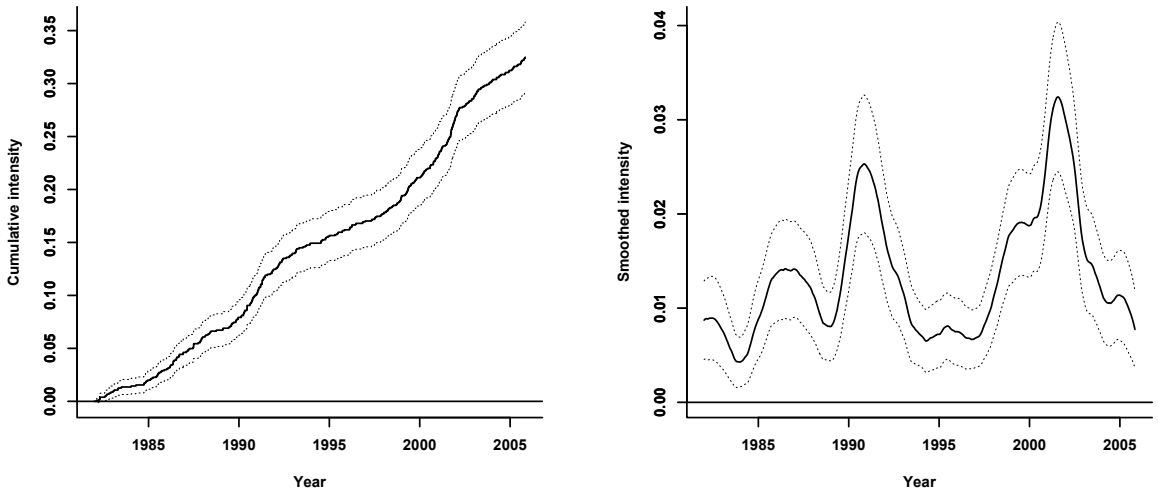

Figure 2. Nelson-Aalen estimates of the default intensity in the cohort.

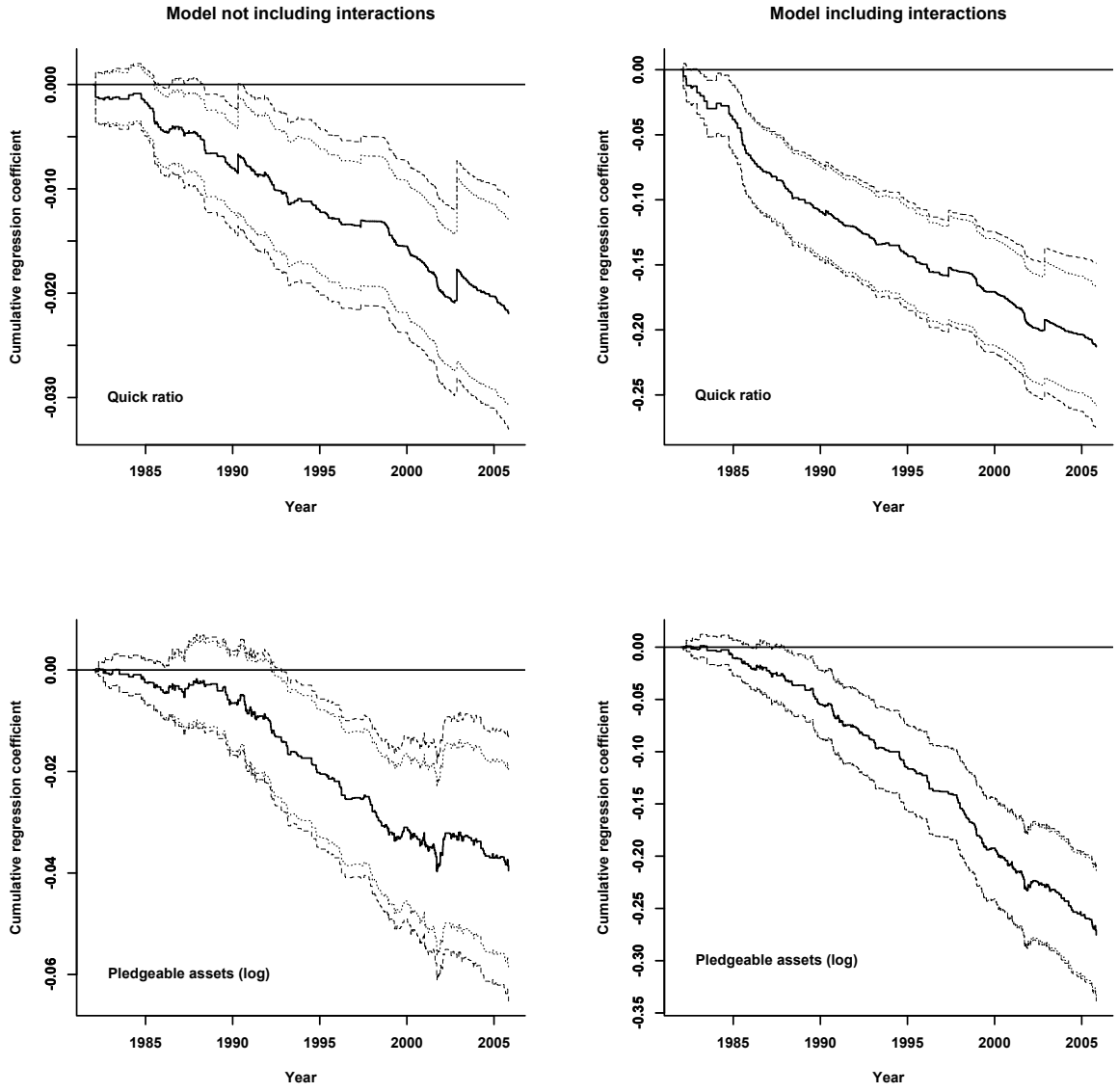

Figure 3. Missing interactions.

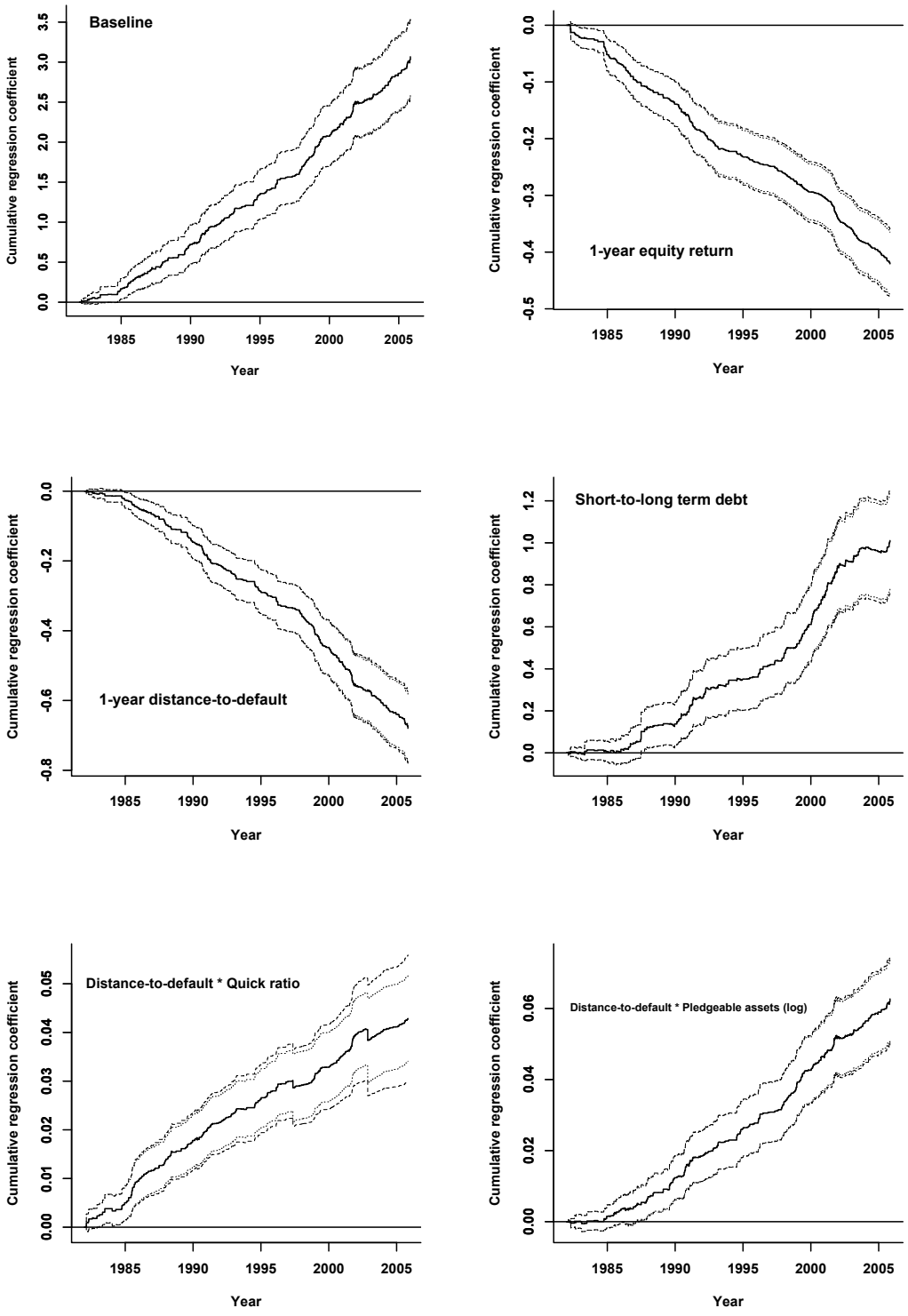

Figure 4. Cumulative coefficients from initial nonparametric Aalen analysis.

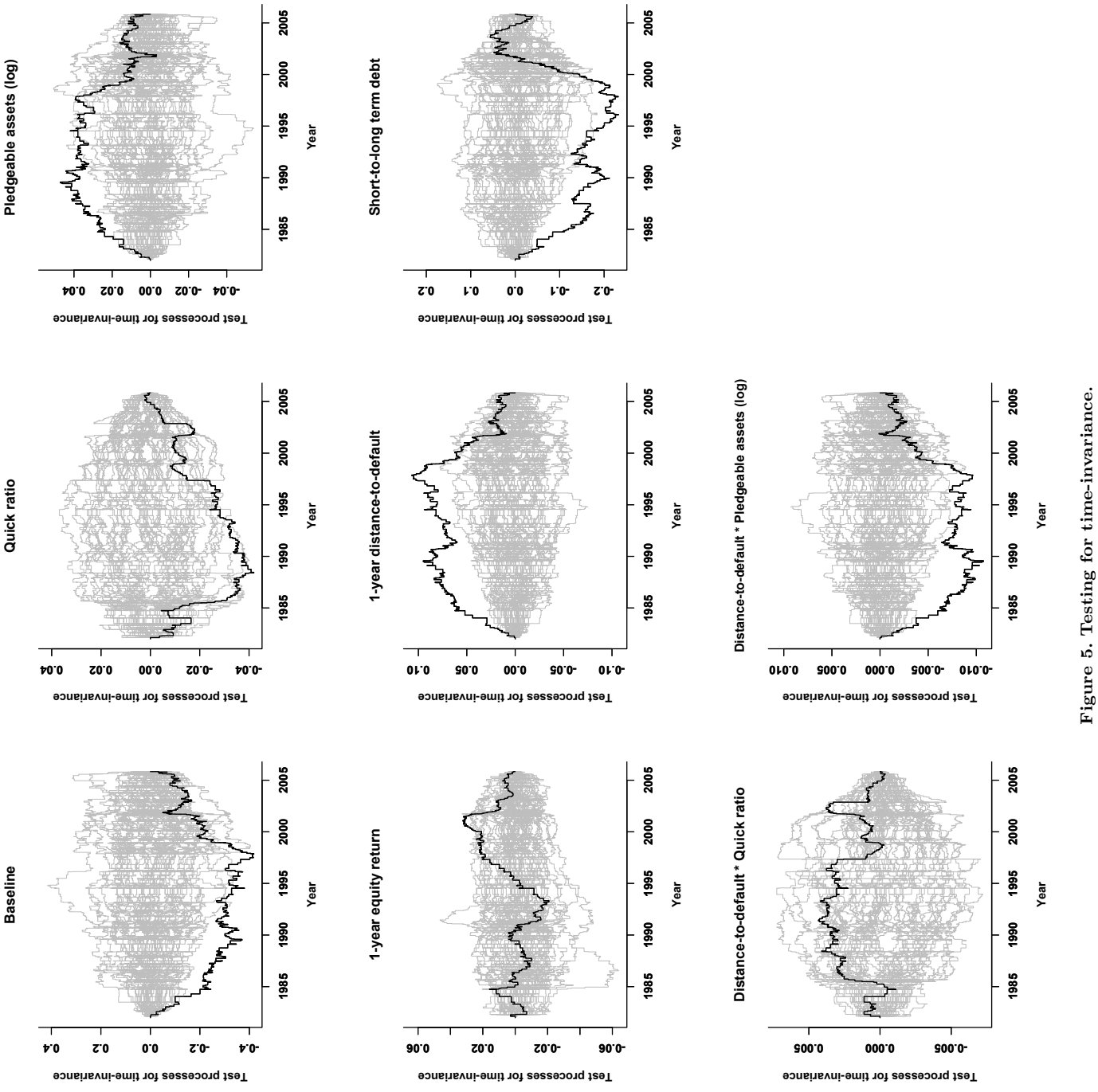

Figure 5. Testing for time-invariance.

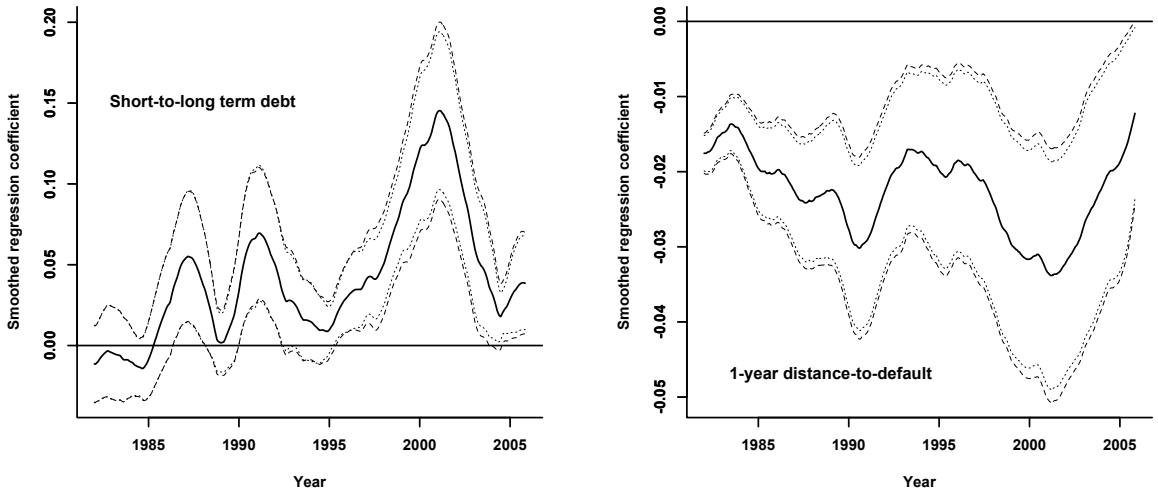

Figure 6. Smoothed regression coefficients.

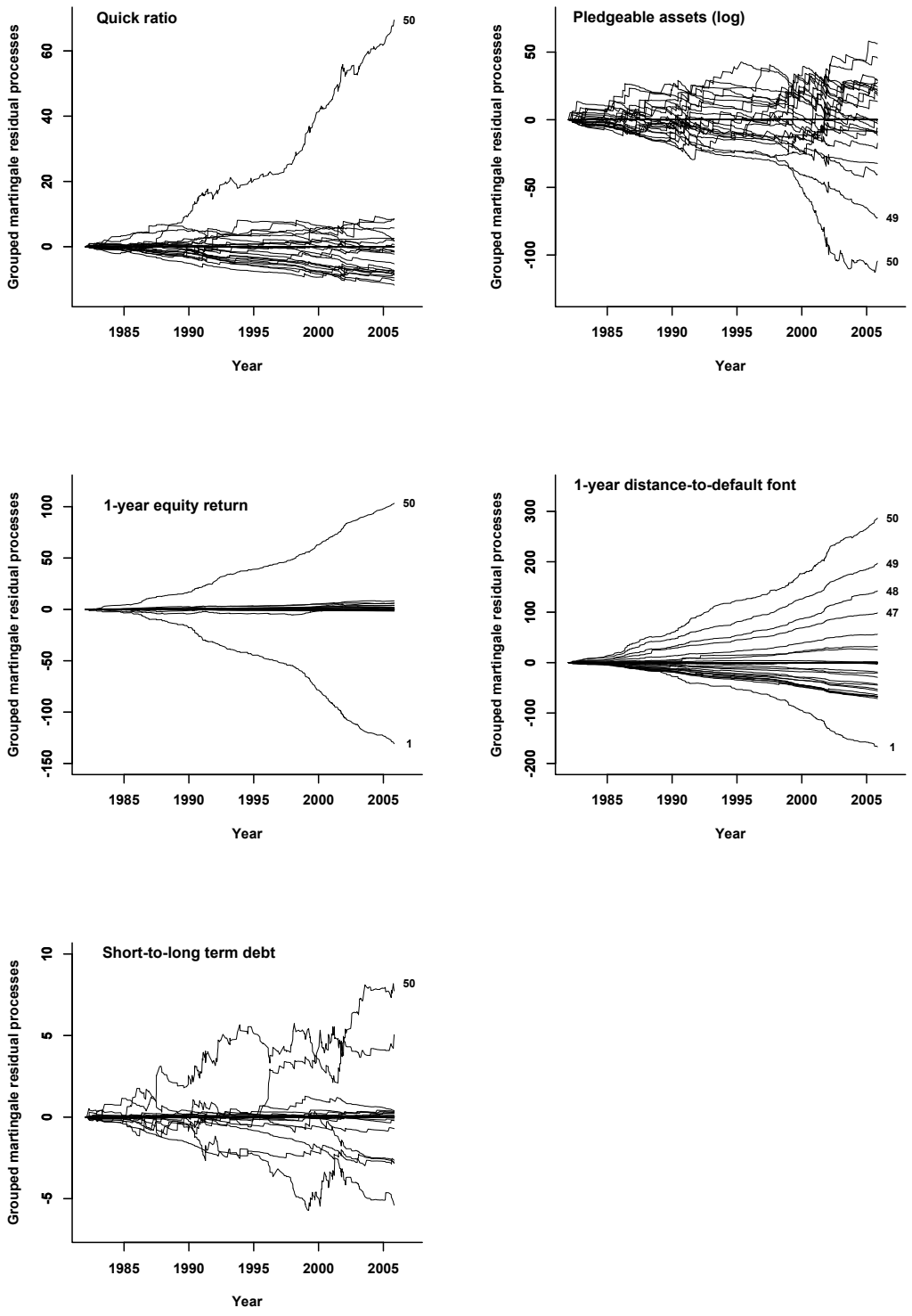

Figure 7. Covariate misspecifications.

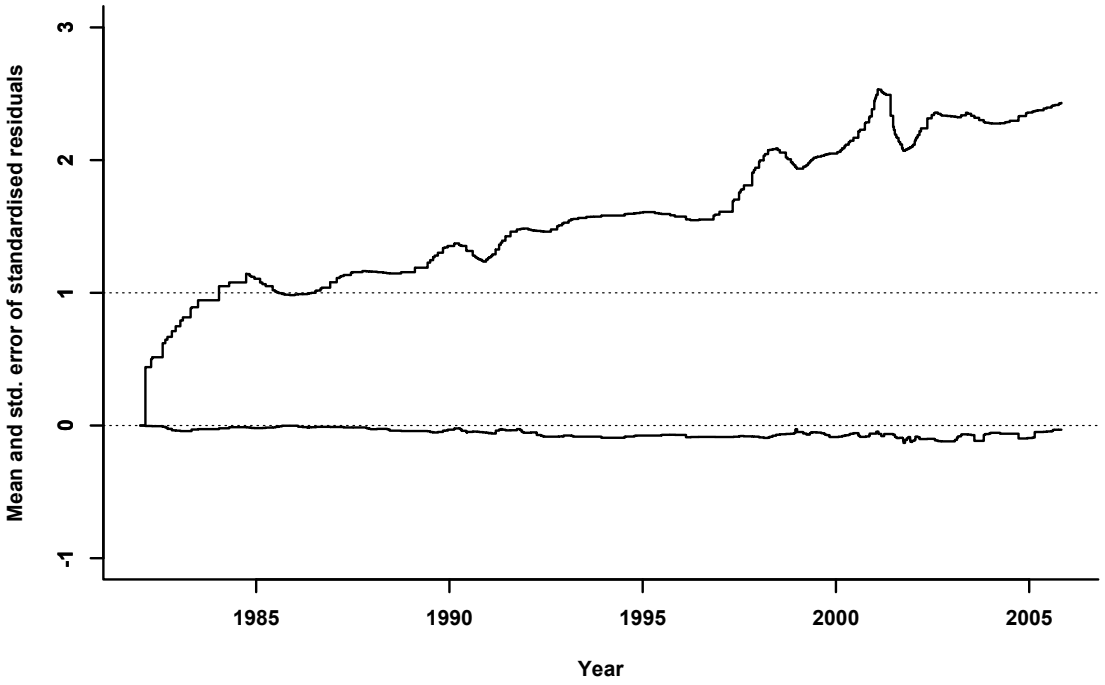

Figure 8. The model fit as a whole.

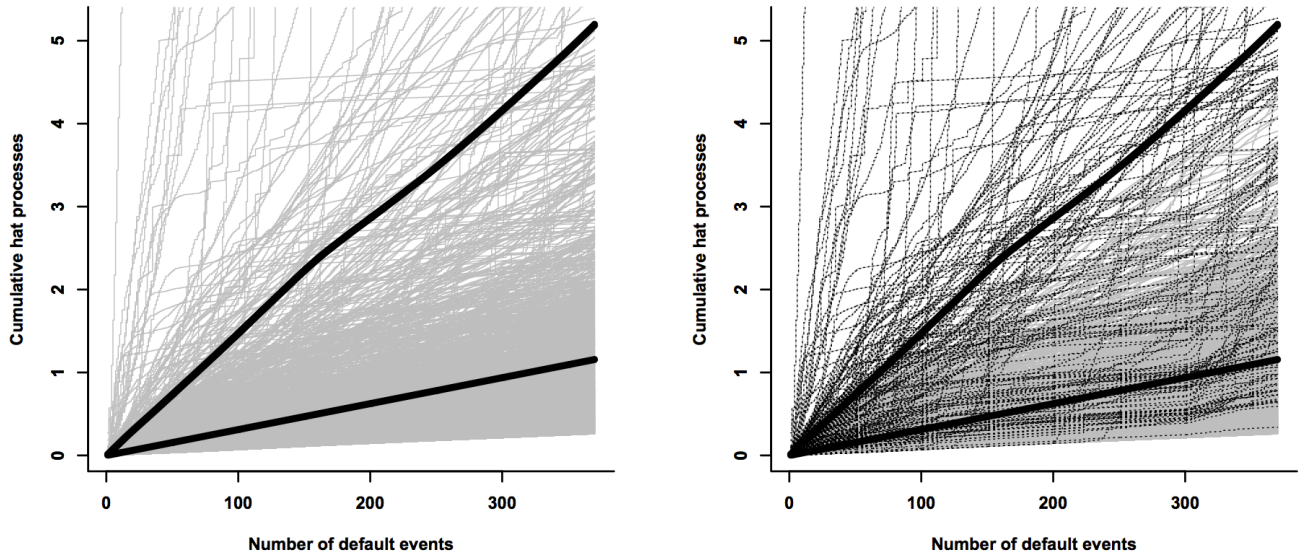

Figure 9. Outliers and over-influence.