

# Localising Forward Intensities for Multiperiod Corporate Default \*

Dedy Dwi Prastyo<sup>1,2</sup> and Wolfgang Karl Härdle<sup>1,3</sup>

<sup>1</sup> Humboldt-Universität zu Berlin, Ladislaus von Bortkiewicz Chair of Statistics  
Center for Applied Statistics and Economics (C.A.S.E.)  
Unter den Linden 6, 10099 Berlin, Germany

<sup>2</sup> Department of Statistics, Institut Teknologi Sepuluh Nopember (ITS)  
Jl. Arief Rahman Hakim, Surabaya 60111, Indonesia

<sup>3</sup> Singapore Management University  
50 Stamford Road, Singapore 178899

## Abstract

Using a local adaptive Forward Intensities Approach (FIA) we investigate multiperiod corporate defaults and other delisting schemes. The proposed approach is fully data-driven and is based on local adaptive estimation and the selection of optimal estimation windows. Time-dependent model parameters are derived by a sequential testing procedure that yields adapted predictions at every time point. Applying the proposed method to monthly data on 2000 U.S. public firms over a sample period from 1991 to 2011, we estimate default probabilities over various prediction horizons. The prediction performance is evaluated against the global FIA that employs all past observations. For the six months prediction horizon, the local adaptive FIA performs with the same accuracy as the benchmark. The default prediction power is improved for the longer horizon (one to three years). Our local adaptive method can be applied to any other specifications of forward intensities.

*Key words:* Accuracy ratio, Forward default intensity, Local adaptive, Mutiperiod prediction

*JEL Classification:* C41, C53, C58, G33

---

\*This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 'Economic Risk', Humboldt-Universität zu Berlin. Dedy Dwi Prastyo was also supported by Directorate General for Higher Education, Indonesian Ministry of Education and Culture through Department of Statistics, Institut Teknologi Sepuluh Nopember (ITS), Indonesia. We would like to thank the Risk Management Institute (RMI), the National University of Singapore (NUS) for the data used in this study and for the partial financial support under Credit Research Initiative (CRI) project. We would also like to thank the International Research Training Group (IRTG) 1792. Email: prastyod@hu-berlin.de or dedy-dp@statistika.its.ac.id (Prastyo), haerdle@wiwi.hu-berlin.de (Härdle).

# 1 INTRODUCTION

Credit risk analysis plays an essential role in finance in order to measure default risk that can put stakeholders on financial complication. As a consequence of Basel's proposed capital requirement on credit asset, bank and financial institution have to develop their internal credit risk system. Two key elements for the internal credit rating are (Härdle and Prastyo, 2014): (i) compute probability of default (PD) and (ii) estimate the loss given default (LGD). The PD is the probability of failing to pay debt in full over a particular time horizon. The LGD is the percentage of loss over the total exposure upon default that can be estimated by identifying its distribution on defaulters with similar attributes. This paper employs a corporate PD prediction methodology.

Many stochastic models and statistical techniques have been developed to measure the likelihood that a debtor will fail to service its obligation in full. One of the techniques for default prediction is discriminant analysis. This and also classification settings such as support vector machines (Härdle et al. (2009); Chen et al. (2011); Härdle et al. (2014)) though fall short when the interest is in a time varying context. Logit and probit regression are designed to estimate PD directly but have rarely been employed in a time series setting. Recently, the implementation of hazard rate models has received much attention, see Shumway (2001); Chava and Jarrow (2004); Campbell et al. (2008); Bharath and Shumway (2008). A general fitting problem though remains: one disregards companies delist for reason other than that of defaults. This negligence results in censoring biases. One such issue is addressed in intensity based models e.g. Duffie et al. (2007) that treats both default and other exit events of the delisted companies.

An advanced default analysis views individual defaults along with their state predictors: the default (non-default) companies are tagged with common risk factors and firm-specific attributes. Duffie et al. (2009) accommodated unobservable common factors as frailty

factor. Duan et al. (2012) modeled the dynamic data panel by using a forward-intensity specification. It models the default term structure by a new reduced form approach that takes into account both default and other type of exits from companies that are delisted from the market. The PD for different horizon is then computed as a function of different input variables. The advantage of this specification is that one does not need to bother about modeling the high-dimensional state variable process. It contrasts the spot intensity model (Duffie et al., 2007) which requires specification and estimation of the time series of state variables. This may be quite challenging for a high dimensional set of variables.

The forward intensity approach specifies a parametric form that is per se constant although varies over time as depicted by Figure 1. In addition, Figure 2 and 3 exhibit parameters estimates of forward intensities specification that are sensitive to the length of estimation windows. This is where a time varying approach comes into play. More precisely, in this paper a time varying parameter model is approximated by a local constant parametric form. The aim is to implement and localise such parameters for forward intensities. The technique presented selects a data-driven estimation window that allows for flexible forecasts. The key idea is to employ a sequential testing procedure to identify this time interval of constant parameters. Corporate PDs are then computed based on this data interval. By controlling the risk of false alarm, i.e. the algorithm stops earlier than an oracle interval, the algorithm selects the longest possible window for which parameter constancy cannot be rejected.

The proposed framework builds on the local parametric approach (LPA) proposed by Spokoiny (1998). LPA involves processes that are stationary only locally: (i) consider only recent data, (ii) imply sub setting of data using some localisation scheme. Methods developed in this LPA framework are local change point (LCP) (Mercurio and Spokoiny, 2004), local model selection (Katkovnik and Spokoiny, 2008), and stagewise aggregation

(Belomestny and Spokoiny, 2007). The studies done by Chen et al. (2008); Giacomini et al. (2009); Chen et al. (2010); and Chen and Niu (2014) showed that LCP work well in practice.

The contribution of this paper is to introduce an adaptive calibration technique for forward intensities in a multiperiod corporate default setting motivated by our preliminary analysis (Figure 1, 2 and 3). We apply LCP as in Chen and Niu (2014) to the forward intensity model of Duan et al. (2012) to detect the largest interval of homogeneity, i.e. the interval where a local constant parametric form describes the data well, and to provide an adaptive estimate as the one associated with the interval found. The proposed method tests the null hypothesis of homogeneous interval with no change points against the alternative hypothesis of at least one change point being present.

Our empirical analysis uses 2000 U.S. public firms for the period from 1991 to 2011. There are two macroeconomic factors which act as common variables and ten firm-specific variables. The empirical implementation indicates that the parameter estimate of one-year simple return of S&P500 index is sensitive to the length of estimation interval. The parameter estimate of another macroeconomic factor, i.e. 3-month US Treasury bill interest rate, varies less for different calibration window length. The volatility-adjusted leverage (measured as distance-to-default), company size, market-to-book ratio, and idiosyncratic volatility are robust to the estimation interval length whereas the remaining firm-specific attributes that represent liquidity and profitability are not. We estimate default probabilities over various prediction horizons: one month, three months, six months, one year, two years, and three years. The prediction performance is evaluated against the global forward intensities approach that employs all past observations. For the six months prediction horizon, both the global and the local forward intensities approaches perform with the same accuracy. The default prediction power is improved for the longer horizon (one to three years).

The remainder of the paper is structured as follows. The forward intensity approach and the LPA are introduced in Section 2 and 3, respectively. The data and empirical findings on PD prediction are provided in Section 4. Section 5 concludes this study.

## 2 LOCALISING FORWARD INTENSITIES

### 2.1 Forward Intensities Approach (FIA)

Time of default for the  $i$ -th firm is denoted by  $\tau_{Di}$ . We first describe FIA in the pure default case then move on to combined exits for horizon  $\tau$ . In the following sections we drop index  $i$  for simplicity. For known default intensity  $\lambda_s$ , the survival probability in  $[t, t + \tau]$  is:

$$P(\tau_D > t + \tau) = \exp\left(-\int_t^{t+\tau} \lambda_s ds\right). \quad (1)$$

It is reasonable to assume that the intensity  $\lambda_s$  is driven by time dependent state variables  $X_s$ ,  $X_s \in \mathbb{R}^p$ , of which their future evolution is unknown. Hence, given a model for the dynamics of  $X_s$  one may obtain  $\lambda_s = \lambda(X_s)$  by forecasting the path of  $X_s$ . The state variables typically contain common factors  $W$  and firm-specific attributes  $U$ ,  $X_t = (W_t, U_t)$ . Consider conditioning on a filtration  $\{\mathcal{F}_t : t \geq 0\}$ , where  $\mathcal{F}_t$  is generated by:

$$\{(U_s, D_s) : s \leq \min(t, \tau_D)\} \cup \{W_s : s \leq t\},$$

with  $D_s$  a Poisson process for default with intensity  $\lambda(X_s)$ .

Given that from time  $t$  one would like to understand the default occurrence until  $t + \tau$  one needs therefore to simulate the future path of  $X_s$ . Following the idea of Duffie et al. (2007) technique, i.e.  $\lambda_s = \lambda(X_s; \theta_t)$ ,  $\theta_t$  is a vector of parameters obtained based on  $X_t$ ,

the survival probability (1) reads now as:

$$\mathbf{P}(\tau_D > t + s | \mathcal{F}_t) \stackrel{\text{def}}{=} \mathbf{E} \left[ \exp \left\{ - \int_t^{t+s} \lambda(X_u; \theta_t) du \right\} | X_t \right]. \quad (2)$$

Forecasting the time series of  $X_s$  is quite challenging though particularly when the dimension  $p$  is high. An alternative solution is to specify a forward default intensity:

$$\lambda_t(s) \stackrel{\text{def}}{=} \lim_{\Delta t \rightarrow 0} \frac{\mathbf{P}(t + s < \tau_D \leq t + s + \Delta t | \tau_D \geq t + s)}{\Delta t}, \quad (3)$$

as a function of  $X_t$  alone,  $\lambda_t(s) = \lambda(X_t, s)$ . Duan et al. (2012) proposed  $\lambda_t(s) = \lambda(\theta_s; X_t)$  that directly employs  $X_t$  instead of using an imposed dynamic of  $X_s$ . In this paper, we generalise this idea via time varying parameters,  $\lambda_t(s) = \lambda(\theta_{s,t}; X_t)$ . Table 1 summarizes the specification of intensity at time  $s$ .

Table 1: The specifications of the default intensity.

$\lambda_s$	=	$\lambda(X_s; \theta_t)$	, Duffie et al. (2007)
$\lambda_t(s)$	=	$\lambda(\theta_s; X_t)$	, Duan et al. (2012)
$\lambda_t(s)$	=	$\lambda(\theta_{s,t}; X_t)$	, Our approach

The idea behind the FIA in the pure default case is as follows. One observes firm  $i$ ,  $i = 1, \dots, N$ , over an entire sample period  $[0, T]$  and records its default time  $\tau_{Di}$  and state variables  $X_{it}$ . At any time  $t$ ,  $0 < t \leq t + s \leq T$ , the PD can be predicted for next  $s$  period using information  $X_t$  and evaluated against the true one. This predicted PD is derived by an estimate of the forward default intensity  $\widehat{\lambda}_t(s) = \lambda(\widehat{\theta}_s; X_t)$ . The  $\widehat{\theta}_s$  is calibrated by maximizing corresponding likelihood function over  $[0, t]$  that will be presented in (24). The forward default intensity specifies an explicit dependence of default intensity in the future,  $\lambda_s$ , to the values of state variables at the time of prediction  $t$ .

Denote the (differentiable) conditional cdf of  $\tau_D$  evaluated at  $s$ :

$$F_t(s) = 1 - \mathbf{P}(\tau_D > t + s | \mathcal{F}_t), \quad (4)$$

with the conditional survival probability:

$$\mathbf{P}(\tau_D > t + s | \mathcal{F}_t) = \mathbf{E} \left\{ \exp \left( - \int_t^{t+s} \lambda_u du \right) | X_t \right\}. \quad (5)$$

The hazard rate is the event rate at time  $t$  conditional on survival time  $t$  or later. The forward intensity is a hazard function where the survival time is evaluated at a fixed horizon of interest. Hence, the forward default intensity (3) can be rewritten as:

$$\lambda_t(s) = \frac{F'_t(s)}{1 - F_t(s)} = \psi_t(s) + \psi'_t(s)s, \quad (6)$$

with  $\psi_t(s)$  defined as:

$$\begin{aligned} \psi_t(s) &\stackrel{\text{def}}{=} - \frac{\log \{1 - F_t(s)\}}{s}, \\ &= - \frac{\log \mathbf{E} \left\{ \exp \left( - \int_t^{t+s} \lambda_u du \right) | X_t \right\}}{s}. \end{aligned} \quad (7)$$

Thus,  $\psi_t(s)s = \int_0^s \lambda_t(u) du$  is the cumulative forward default intensity and  $\exp\{-\psi_t(s)s\}$  is the survival probability. The proof is given in Appendix A. The forward default intensity  $\lambda_t(s)$  as defined in (3) is then formulated as:

$$\lambda_t(s) = \exp\{-\psi_t(s)s\} \lim_{\Delta t \rightarrow 0} \frac{\mathbf{E} \left\{ \int_{t+s}^{t+s+\Delta t} \exp \left( - \int_t^{t+u} \lambda_v dv \right) \lambda_u du | \tau_D \geq t + s \right\}}{\Delta t}. \quad (8)$$

The conditional probabilities to survive (9) and to default (10) now are, respectively:

$$\mathbf{P}(\tau_D > t + s | \mathcal{F}_t) = \exp \left\{ - \int_t^{t+s} \lambda_t(u) du \right\} = \exp \{ -\psi_t(s)s \} \quad (9)$$

$$\mathbf{P}(\tau_D \leq t + s | \mathcal{F}_t) = \int_t^{t+s} \exp \left\{ - \int_t^{t+u} \lambda_t(v) dv \right\} \lambda_t(u) du. \quad (10)$$

A company traded in a stock exchange can be delisted because of a default event or other reasons, such as merger or aquisition operations. Duffie et al. (2007) modelled these two events via a doubly stochastic process driven by two independent mechanisms with intensities  $\lambda_t$  and  $\phi_t$ . Denote  $\tau_O$  as the time of other exit. Recall  $\lambda_s = \lambda(X_s; \theta_t)$  and specify  $\phi_s = \phi(X_s; \theta_t)$ , by law of iterated expectation and conditional on filtration  $\mathcal{F}_t$ , the probability to survive (11) and to default (12) over  $[t, t + \tau]$  are:

$$\mathbf{P}(\tau_D, \tau_O > t + s, | \mathcal{F}_t) \stackrel{\text{def}}{=} \mathbf{E} \left[ \exp \left\{ - \int_t^{t+s} (\lambda_u + \phi_u) du \right\} | X_t \right], \quad (11)$$

$$\mathbf{P}(\tau_D, \tau_O \leq t + s, | \mathcal{F}_t) \stackrel{\text{def}}{=} \mathbf{E} \left[ \int_t^{t+s} \exp \left\{ - \int_t^{t+u} (\lambda_v + \phi_v) dv \right\} \lambda_u du | X_t \right]. \quad (12)$$

A default event cannot happen after a company exited from the market. Thus these two events are competing and not fully independent. The independency assumption of the two processes will blur the distinction between competing and independent risk. Duan et al. (2012) proposed a forward intensity approach that enables us to work in a more convenient way.

Time of default and other exits together, hereinafter called combined exit, is denoted by  $\tau_C$ , with  $\tau_C \leq \tau_D$ . Applying the same procedure as in estimating forward default intensity from the pure default process, denote the (differentiable)

$$G_t(s) = 1 - \mathbf{P}(\tau_C > t + s | \mathcal{F}_t) \quad (13)$$



as the conditional cdf of  $\tau_C$  evaluated at  $s$  with conditional survival probability:

$$\mathbf{P}(\tau_C > t + s | \mathcal{F}_t) = \mathbf{E} \left\{ \exp \left( - \int_t^{t+s} g_u du \right) | X_t \right\}. \quad (14)$$

The forward combined exit intensity  $g_t(s)$ :

$$g_t(s) \stackrel{\text{def}}{=} \lim_{\Delta t \rightarrow 0} \frac{\mathbf{P}(t + s < \tau_C \leq t + s + \Delta t | \tau_C \geq t + s)}{\Delta t} \quad (15)$$

can be rewritten as:

$$g_t(s) = \frac{G'_t(s)}{1 - G_t(s)} = \psi_t(s) + \psi'_t(s)s. \quad (16)$$

The  $\psi_t(s)$  in (7) is rewritten in term of  $g_u$ . Thus,  $\psi_t(s)s = \int_0^s g_t(u) du$  and the conditional survival probability (14) is given by:

$$\mathbf{P}(\tau_C > t + s | \mathcal{F}_t) = \exp \{ -\psi_t(s)s \}. \quad (17)$$

The instantaneous default intensity at horizon  $t+s$  (forward default intensity from doubly Poisson processes) is defined as:

$$f_t(s) \stackrel{\text{def}}{=} \exp \{ -\psi_t(s)s \} \lim_{\Delta t \rightarrow 0} \frac{\mathbf{P}(t + s < \tau_D = \tau_C \leq t + s + \Delta t | Q)}{\Delta t}, \quad (18)$$

$$= \exp \{ -\psi_t(s)s \} \lim_{\Delta t \rightarrow 0} \frac{\mathbf{E} \left\{ \int_{t+s}^{t+s+\Delta t} \exp \left( - \int_t^u g_v dv \right) \lambda_u du | Q \right\}}{\Delta t}, \quad (19)$$

with  $Q$  the event that  $\tau_D = \tau_C \geq t + s$ . The default probability over  $[t, t + s]$  is

$$\mathbf{P}(\tau_C \leq t + s | \mathcal{F}_t) = \int_t^{t+s} \exp \left\{ - \int_t^{t+u} g_t(v) dv \right\} f_t(u) du. \quad (20)$$

Duan et al. (2012) deal with  $f_{it}(s)$  and  $g_{it}(s)$  as functions of state variables  $X_{it}$  for firm

$i$ , with  $f_{it}(s) > 0$  and  $g_{it}(s) \geq f_{it}(s)$ . More precisely, with  $f_{it}(s) = f_{it}(\theta_s; X_{it})$  and  $g_{it}(s) = g_{it}(\theta_s; X_{it})$ :

$$f_{it}(s) = \exp \{ \alpha^\top(s) X_{it} \}, \quad (21)$$

$$g_{it}(s) = f_{it}(s) + \exp \{ \beta^\top(s) X_{it} \}, \quad (22)$$

with  $X_{it} = (1, x_{it,1}, x_{it,2}, \dots, x_{it,p})^\top$  that include macroeconomic factors ( $W_t$ ) as common factors and firm-specific attributes ( $U_{it}$ ). The survival and default probabilities are assumed to depend only upon  $W$  and  $U$  such that different firms are  $\mathcal{F}_t$ -conditionally independent among themselves. If it is not the case, the dependency must arise from their sharing of  $W$  and/or any correlation among  $U$ . This conditional independence assumption is in essence similar to the doubly stochastic assumption. Therefore, one firm's exit neither feedback to the state variables nor influence the exit probabilities of other firms. This approach is identical to the spot intensity formulation of Duffie et al. (2007) when  $s = 0$ .

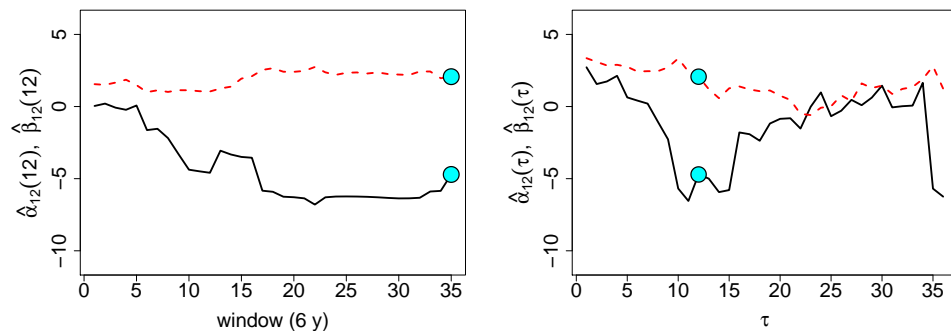


Figure 1: The  $\hat{\alpha}_j(\tau) = \hat{\alpha}_{12}(12)$  (solid) and  $\hat{\beta}_j(\tau) = \hat{\beta}_{12}(12)$  (dashed) for idiosyncratic volatility. Left: fixed  $\tau = 12$  over 35 rolling windows (length: 6 years). Right: The 35-th window with time end December 2011,  $\tau = 0, 1, \dots, 36$ . Solid circles represent the same estimates.

## 2.2 Local Parametric Dynamics

As can be deduced from Figure 1, the parameters in (21) and (22) may vary over time, i.e.  $\alpha_{jt}(\tau)$  and  $\beta_{jt}(\tau)$ ,  $j = 1, \dots, p$ , are time local parameters. Time varying coefficients are typically assumed as: (i) smooth functions of time (Cai et al. (2000); Fan and Zhang (2008)) or (ii) piecewise constant functions (Bai and Perron, 1998). In contrast to these approaches that aim at establishing a time varying model for the whole sample period, our approach is local and data-driven. It is focused on an instantaneous calibration of (20).

The LPA (Local Parametric Approach) aims at finding a balance between parameter variability (precision) and modelling bias by taking into account the past information which is statistically identified as being relevant. In fact one determines time localized parameters: for any particular time point  $t$ , there exists a past suitable window over which the time varying parameters in (21) and (22) are approximately constant. This is in fact the basic idea of the LPA, that is to select a window that guarantees a localised stable model. This is realised by a sequential test based on comparing the increase of the log likelihood process relative to critical values (Spokoiny, 2009).

Denote an interval  $I = [t - m, t]$  as a right-end fixed interval of  $m$  observation at time  $t$ . Suppose that our sample has period  $[0, T]$  for each interval  $I$ . Then, the local likelihood (for the horizon  $\tau$ ) based on (21)(22) in interval  $I$ :

$$L_{I,\tau}(\alpha, \beta) = \prod_{i=1}^N \prod_{\substack{t=0 \\ t \in I}}^{T-1} L_{\tau,i,t}(\alpha, \beta), \quad (23)$$

where  $N$  is the number of companies at  $t$  and

$$\alpha = \begin{pmatrix} \alpha_0(0) & \alpha_0(1) & \cdots & \alpha_0(\tau - 1) \\ \alpha_1(0) & \alpha_1(1) & \cdots & \alpha_1(\tau - 1) \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_p(0) & \alpha_p(1) & \cdots & \alpha_p(\tau - 1) \end{pmatrix}; \beta = \begin{pmatrix} \beta_0(0) & \beta_0(1) & \cdots & \beta_0(\tau - 1) \\ \beta_1(0) & \beta_1(1) & \cdots & \beta_1(\tau - 1) \\ \vdots & \vdots & \ddots & \vdots \\ \beta_p(0) & \beta_p(1) & \cdots & \beta_p(\tau - 1) \end{pmatrix}.$$

Let  $t_{0i}$  be the first time that firm  $i$  appeared in the sample. If the firm does not appear in sample in  $t$  or is already delisted before  $t$ , i.e.  $t_{0i} > t$  or  $\tau_{Ci} \leq t$ , then the likelihood is set to 1 and is transformed to 0 in log-likelihood such that

$$\begin{aligned} L_{\tau,i,t}(\alpha, \beta) &= \mathbf{1}_{\{t_{0i} \leq t, \tau_{Ci} > t + \tau\}} \mathbf{P}_t(\tau_{Ci} > t + \tau) \\ &\quad + \mathbf{1}_{\{t_{0i} \leq t, \tau_{Di} = \tau_{Ci} \leq t + \tau\}} \mathbf{P}_t(\tau_{Ci}; \tau_{Di} = \tau_{Ci} \leq t + \tau) \\ &\quad + \mathbf{1}_{\{t_{0i} \leq t, \tau_{Di} \neq \tau_{Ci}, \tau_{Ci} \leq t + \tau\}} \mathbf{P}_t(\tau_{Ci}; \tau_{Di} \neq \tau_{Ci} \& \tau_{Ci} \leq t + \tau) \\ &\quad + \mathbf{1}_{\{t_{0i} > t\}} + \mathbf{1}_{\{\tau_{Ci} \leq t\}}, \end{aligned} \tag{24}$$

with  $\mathbf{P}_t(\tau_{Ci}) = \mathbf{P}(\tau_{Ci} | \mathcal{F}_t)$ .

For numerical application, (14),(17) are approximated by:

$$\mathbf{P}_t(\tau_{Ci} > t + \tau) = \exp \left\{ - \sum_{s=0}^{\tau-1} g_{it}(s) \Delta t \right\}, \tag{25}$$

with  $\Delta t = 1/12$  to represent that prediction horizon is measured in month. Therefore,

the probability of exit due to default and other reasons, respectively:

$$\begin{aligned}
& P_t(\tau_{Ci}; \tau_{Di} = \tau_{Ci} \leq t + \tau) \tag{26} \\
& = \begin{cases} 1 - \exp\{-f_{it}(0)\Delta t\} & \text{if } \tau_{Ci} = t + 1, \\ \exp\left\{-\sum_{s=0}^{\tau_{Ci}-t-2} g_{it}(s)\Delta t\right\} \times [1 - \exp\{-f_{it}(\tau_{Ci} - t - 1)\Delta t\}] \\ & \text{if } t + 1 < \tau_{Ci} \leq t + \tau, \end{cases}
\end{aligned}$$

$$\begin{aligned}
& P_t(\tau_{Ci}; \tau_{Di} \neq \tau_{Ci} \& \tau_{Ci} \leq t + \tau) \tag{27} \\
& = \begin{cases} \exp\{-f_{it}(0)\Delta t\} - \exp\{-g_{it}(0)\Delta t\} & \text{if } \tau_{Ci} = t + 1, \\ \exp\left\{-\sum_{s=0}^{\tau_{Ci}-t-2} g_{it}(s)\Delta t\right\} \times \\ & [\exp\{-f_{it}(\tau_{Ci} - t - 1)\Delta t\} - \exp\{-g_{it}(\tau_{Ci} - t - 1)\Delta t\}] \\ & \text{if } t + 1 < \tau_{Ci} \leq t + \tau, \end{cases}
\end{aligned}$$

The forward intensities in the discretized version, i.e.  $f_{it}(\tau)$  and  $g_{it}(\tau)$ , should be understood as at time  $t$  for the period  $[t + \tau, t + \tau + 1]$  because horizon index  $s$  in (25)-(27) starts from zero. Forward intensity is basically spot intensity for one month ahead. Duan et al. (2012) derived the large sample properties of the likelihood (23) constructed from overlapped periods. This likelihood can be numerically maximized to obtain  $\hat{\alpha}$  and  $\hat{\beta}$ .

The log likelihood of (23) separates into a sum of terms involving  $\alpha$  and  $\beta$ . We can maximize its two components individually to obtain  $\hat{\alpha}$  and  $\hat{\beta}$ . In addition, the likelihood for  $\alpha$  or  $\beta$  can be decomposed to terms involving  $\alpha(\tau)$  or  $\beta(\tau)$  only. This property enables us to estimate  $\alpha$  and  $\beta$  without performing estimation sequentially from shorter to longer

prediction horizon. Thus for horizon  $s = 0, 1, \dots, \tau - 1$ ,

$$\hat{\alpha}(s) = \max_{\alpha(s)} \log L\{\alpha(s)\} = \max_{\alpha(s)} \log \left[ \sum_{i=1}^n \sum_{t=0}^{T-s-1} L_{i,t}\{\alpha(s)\} \right], \quad (28)$$

$$\hat{\beta}(s) = \max_{\beta(s)} \log L\{\beta(s)\} = \max_{\beta(s)} \log \left[ \sum_{i=1}^n \sum_{t=0}^{T-s-1} L_{i,t}\{\beta(s)\} \right], \quad (29)$$

with

$$\begin{aligned} L_{i,t}\{\alpha(s)\} &= \mathbf{1}_{\{t_{0i} \leq t, \tau_{Ci} > t+s+1\}} \exp\{-f_{it}(s)\Delta t\} \\ &\quad + \mathbf{1}_{\{t_{0i} \leq t, \tau_{Di} = \tau_{Ci} \leq t+s+1\}} [1 - \exp\{-f_{it}(s)\Delta t\}] \\ &\quad + \mathbf{1}_{\{t_{0i} \leq t, \tau_{Di} \neq \tau_{Ci}, \tau_{Ci} \leq t+s+1\}} \exp\{-f_{it}(s)\Delta t\} \\ &\quad + \mathbf{1}_{\{t_{0i} > t\}} + \mathbf{1}_{\{\tau_{Ci} \leq t+s+1\}}, \end{aligned} \quad (30)$$

$$\begin{aligned} L_{i,t}\{\beta(s)\} &= \mathbf{1}_{\{t_{0i} \leq t, \tau_{Ci} > t+s+1\}} \exp\{-[g_{it}(s) - f_{it}(s)]\Delta t\} \\ &\quad + \mathbf{1}_{\{t_{0i} \leq t, \tau_{Di} = \tau_{Ci} \leq t+s+1\}} \\ &\quad + \mathbf{1}_{\{t_{0i} \leq t, \tau_{Di} \neq \tau_{Ci}, \tau_{Ci} \leq t+s+1\}} [1 - \exp\{-[g_{it}(s) - f_{it}(s)]\Delta t\}] \\ &\quad + \mathbf{1}_{\{t_{0i} > t\}} + \mathbf{1}_{\{\tau_{Ci} \leq t+s+1\}}, \end{aligned} \quad (31)$$

where  $g_{it}(s) - f_{it}(s) = \exp\{\beta_0(s) + \beta_1(s)x_{it,1} + \dots + \beta_p(s)x_{it,p}\}$ .

All the firm-month observations are classified into following categories  $X^0 = (x_1^0, \dots, x_{N_0}^0)^\top$ ,  $X^1 = (x_1^1, \dots, x_{N_1}^1)^\top$ , and  $X^2 = (x_1^2, \dots, x_{N_2}^2)^\top$ , where  $X^0$ ,  $X^1$ , and  $X^2$  contain all firm-month observations that survive, default, and exit due to other reasons, respectively. The  $N_0$ ,  $N_1$ , and  $N_2$  are number of observations in each category. Therefore, we can express

the horizon-specific log-likelihood:

$$\begin{aligned} \log L\{\alpha(s)\} &= - \sum_{i=1}^{N_0} \exp(x_i^0 \alpha) \Delta t + \sum_{i=1}^{N_1} \log [1 - \exp\{-\exp(x_i^1 \alpha) \Delta t\}] \\ &\quad - \sum_{i=1}^{N_2} \exp(x_i^2 \alpha) \Delta t, \end{aligned} \quad (32)$$

$$\log L\{\beta(s)\} = - \sum_{i=1}^{N_0} \exp(x_i^0 \beta) \Delta t + \sum_{i=1}^{N_1} \log [1 - \exp\{-\exp(x_i^1 \beta) \Delta t\}]. \quad (33)$$

In LPA, the maximum likelihood estimates (MLE) of  $\theta = \{\alpha, \beta\}$  over horizon  $\tau$  in the data interval  $I$  is maximizing (23):

$$\tilde{\theta}_I = \arg \max_{\theta \in \Theta} L_{I,\tau}(\alpha, \beta). \quad (34)$$

The interval  $I$  controls the estimation quality and addresses the tradeoff between estimation efficiency and local flexibility. The quality of  $\tilde{\theta}_I$  as the estimator of the true time varying parameter vector  $\theta_t^*$  is assessed by Kullback-Leibler (KL) divergence. Discarding the time subscript and keep an asterisk (\*) for notational convenience, the KL divergence of approximate distribution  $\mathcal{P}_{\tilde{\theta}_I}$  from the true distribution  $\mathcal{P}_{\theta^*}$  is  $\mathcal{K}_{I,\tau}\{\tilde{\theta}_I, \theta^*\} = \mathbb{E}_{\theta^*} \left\{ \log(\mathcal{P}_{\theta^*} / \mathcal{P}_{\tilde{\theta}_I}) \right\}$ . Let  $N_I$  is number of observations in interval  $I$ , the KL divergence of distributions that belong to exponential family can be represented in term of (local) likelihood:

$$\mathcal{K}_{I,\tau}\{\tilde{\theta}_I, \theta^*\} = N_I^{-1} \left\{ L_{I,\tau}(\tilde{\theta}_I) - L_{I,\tau}(\theta^*) \right\} \quad (35)$$

that measures the expectation (under  $\mathcal{P}_{\theta^*}$ ) of the information lost when  $\mathcal{P}_{\tilde{\theta}_I}$  is used to approximate  $\mathcal{P}_{\theta^*}$ . By introducing the  $r$ -th power of that likelihood difference, define a

loss function:

$$L_{I,\tau}(\tilde{\theta}_I, \theta^*) \stackrel{\text{def}}{=} \left| L_{I,\tau}(\tilde{\theta}_I) - L_{I,\tau}(\theta^*) \right| \quad (36)$$

that obeys a parametric risk bound:

$$\mathbb{E}_{\theta^*} \left| L_{I,\tau}(\tilde{\theta}_I, \theta^*) \right|^r \leq \mathcal{R}_r(\theta^*), \quad (37)$$

where  $\mathcal{R}_r(\theta^*)$  denotes a constant depending on  $r > 0$  and  $\theta^*$ , see Spokoiny (2009). In the exponential family set up, the parametric risk bound is parameter invariant, i.e.  $\mathcal{R}_r(\theta^*) = \mathcal{R}_r$ . Different values of  $r$  lead to different risk bounds (37), critical values and adaptive estimates. Higher values of  $r$  lead to selection of longer intervals of homogeneity. We follow the recommendation of Čížek et al. (2009) and consider  $r = 0.5$  and  $r = 1$ .

### 3 LOCAL PARAMETRIC FRAMEWORK

In practice, the interval of homogeneity is unknown and needs to be selected among a finite set of  $K$  candidates. The aim is to well approximate the time varying parameter  $\theta_t$  model by a locally constant parametric model. The approximation quality is measured by the KL divergence (35). Denote  $\Delta_{I_k,\tau}(\theta) = \sum_{t \in I_k} \mathcal{K}_{I_k,\tau} \{ \mu_t, \mu_t(\theta) \}$  as a measure of discrepancy between the true (unknown) data generating process  $\mu_t$  and the parametric model  $\mu_t(\theta)$  with intensities (21) and (22) for intervals  $I_k$ . Let for some  $\theta \in \Theta$ ,

$$\mathbb{E} \{ \Delta_{I_k}(\theta) \} \leq \Delta, \quad (38)$$

where  $\Delta \geq 0$  denotes a small modelling bias (SMB) for interval  $I_k$ . Consider  $(K + 1)$  nested intervals (with fixed right-end point  $t$ )  $I_k = [t - m_k, t]$  of length  $|I_k| = m_k$  for



any particular time  $t$ ,  $I_K \supset \cdots \supset I_k \supset \cdots \supset I_1 \supset I_0$ . The oracle, i.e. theoretically optimal, choice  $I_{k^*}$  of the interval sequence is defined as the largest interval for which the SMB condition (38) holds. In practice of course  $\Delta_{I_k}$  is unknown and therefore the oracle choice of  $k^*$  cannot be implemented directly. One therefore mimics the oracle choice via sequential testing for  $k$  ideal situation, where  $k = 1, \dots, K$  (Spokoiny, 2009).

### 3.1 Homogeneity Interval Test for Fixed $\tau$

The interval selection algorithm chooses the (optimal) length of interval where at each  $I_k$ , it tests the null hypothesis on parameter homogeneity against the alternative of a change point within  $I_k$ . We write  $\tilde{\theta}_k$  instead of  $\tilde{\theta}_{I_k}$  as estimates obtained at interval  $I_k$ . The adaptive estimates  $\hat{\theta}_k$  is the MLE at the interval of homogeneity, i.e.  $\hat{\theta}_k = \tilde{\theta}_k$ . For  $I_0$ , one puts  $\hat{\theta}_0 = \tilde{\theta}_0$ . One iteratively extends the subsets and sequentially tests for possible change points in the next longer interval. For a fixed horizon, a likelihood ratio test (LRT) is employed at each  $I_k$  with test statistic (Chen and Niu, 2014):

$$T_{k,\tau} = \left| L_{I_k}(\tilde{\theta}_k) - L_{I_k}(\hat{\theta}_{k-1}) \right|^r, \quad k = 1, \dots, K. \quad (39)$$

Assume parameter homogeneity in  $I_{k-1}$  has been established at a given time point  $t$ , the hypothetical homogeneity in interval  $I_k$  is tested by measuring the difference between their corresponding estimates.

Once a set of critical values  $\mathfrak{z}_{1,\tau}, \dots, \mathfrak{z}_{K,\tau}$  is generated via a Monte Carlo simulation, the sequential testing procedure is accomplished. If  $T_{k,\tau} > \mathfrak{z}_{k,\tau}$ , then the procedure terminates and selects interval  $I_{k-1}$  such that  $\hat{\theta}_k = \hat{\theta}_{k-1} = \tilde{\theta}_{k-1}$ . Otherwise, the interval  $I_k$  is accepted as homogeneous and one continues to update the estimate  $\hat{\theta}_k = \tilde{\theta}_k$ . The adaptive estimation is done through comparing the test statistic (39) at every step  $k$  with the corresponding critical value  $\mathfrak{z}_{k,\tau}$ . One then searches for the longest interval of

homogeneity  $I_{\widehat{k}}$  for which the null hypothesis is not rejected:

$$\widehat{\theta} = \widetilde{\theta}_{\widehat{k}}, \quad \widehat{k} = \max_{k \leq K} \{k : T_{\ell, \tau} \leq \mathfrak{z}_{\ell, \tau}, \quad \ell \leq k\}. \quad (40)$$

The smallest interval is always considered to be homogeneous. If the null is rejected at the first step, then  $\widehat{\theta} = \widetilde{\theta}_0$ . Otherwise, we sequentially repeat this test until we find a change point at  $k$ , accordingly  $\widehat{\theta} = \widetilde{\theta}_{\widehat{k}}$ , or exhaust all interval such that  $\widehat{\theta} = \widetilde{\theta}_K$ .

### 3.2 Critical Values for Fixed $\tau$

Under the hypothesis of parameter homogeneity, the correct choice of interval is the largest one,  $I_K$ . The critical values are chosen in a way such that the probability of selecting  $k < K$ , "false alarm", is minimized. For a fixed  $\tau$ , in case  $k$  is selected (instead of  $K$ ) and thus  $\widehat{\theta} = \widetilde{\theta}_k$  instead of  $\widetilde{\theta}_K$ , the loss as defined in (36) is  $L_{I_K}(\widetilde{\theta}_K, \widehat{\theta}) = L_{I_K}(\widetilde{\theta}_K) - L_{I_K}(\widehat{\theta})$  and stochastically bounded by

$$\mathbb{E}_{\theta^*} \left| L_{I_K}(\widetilde{\theta}_K) - L_{I_K}(\widehat{\theta}) \right|^r \leq \rho \mathcal{R}_r(\theta^*), \quad (41)$$

with parametric risk bound generated from (true) simulated parameter  $\theta^*$ :

$$\mathcal{R}_r(\theta^*) = \mathbb{E}_{\theta^*} \left| L_{I_K}(\widetilde{\theta}_K) - L_{I_K}(\theta^*) \right|^r. \quad (42)$$

The  $\mathcal{R}_r(\theta^*)$  is finite (see Appendix B) and can be numerically computed with the knowledge of  $\theta^*$ .

Critical values must ensure that the loss associated with false alarm is at most a  $\rho$ -fraction of the parametric risk bound,  $\mathcal{R}_r(\theta^*)$ , of the oracle estimate  $\widetilde{\theta}_K$ . The  $\rho$  can be interpreted

as the false alarm rate probability for  $r \rightarrow 0$ . Accordingly, an estimate  $\hat{\theta}_k$  should satisfy

$$\mathbb{E}_{\theta^*} \left| L_{I_k}(\tilde{\theta}_k) - L_{I_k}(\hat{\theta}_k) \right|^r \leq \frac{\rho^k}{K} \mathcal{R}_r(\theta^*). \quad (43)$$

Comparing test statistics (39) with critical values, if  $T_{k,\tau} > \mathfrak{z}_{k,\tau}$  one accepts  $\hat{\theta}_k = \hat{\theta}_{k-1} = \tilde{\theta}_{k-1}$ , otherwise  $\hat{\theta}_k = \tilde{\theta}_k$ . In general,  $\tilde{\theta}_k$  differs from  $\hat{\theta}_k$  only if a change point is detected at the first  $k$  steps.

The critical values  $\mathfrak{z}_k$  that satisfy (43) are found numerically by Monte Carlo simulation because the sampling distribution of the test statistic is unknown even asymptotically. Relatively large critical values lead to a higher probability of selecting subintervals everywhere, resulting in selecting longer intervals of homogeneity. On the other hand, small critical values lead to favor shorter intervals, discarding useful observations in the past and increasing modelling bias. The optimal critical values are the minimum values to just accept the parametric risk bound at each interval.

## 4 DATA AND EMPIRICAL FINDINGS

Following the notation that was introduced in section 2.1, the state variable for firm  $i$  at time  $t$  is  $X_i = (W, U_i)$ , where a vector  $W$  is common to all firms in the same economy and a firm-specific vector  $U_i$  is observable from the date of firm's financial statement is firstly released until the month before the firm exits (if it does). Our data is a subset of dataset used by Duan et al. (2012) consisting of 2000 U.S. public firms over the period from 1991 to 2011 obtained from CRI database. There are two common variables and ten firm-specific variables.

Common variables ( $W_t$ ) are: (i) trailing one-year simple return on S&P500 index (ii) 3-month U.S. Treasury bill rate, hence  $W \in \mathbb{R}^2$ . Firm-specific variables ( $U_{it}$ ) are: (i)

Table 2: State variable.

$X_j$	
$X_1$	: Index return
$X_2$	: Interest rate
$X_3, X_4$	: $DTD_{level}, DTD_{trend}$
$X_5, X_6$	: $(CASH/TA)_{level}, (CASH/TA)_{trend}$
$X_7, X_8$	: $(NI/TA)_{level}, (NI/TA)_{trend}$
$X_9, X_{10}$	: $SIZE_{level}, SIZE_{trend}$
$X_{11}$	: M/B
$X_{12}$	: IdV

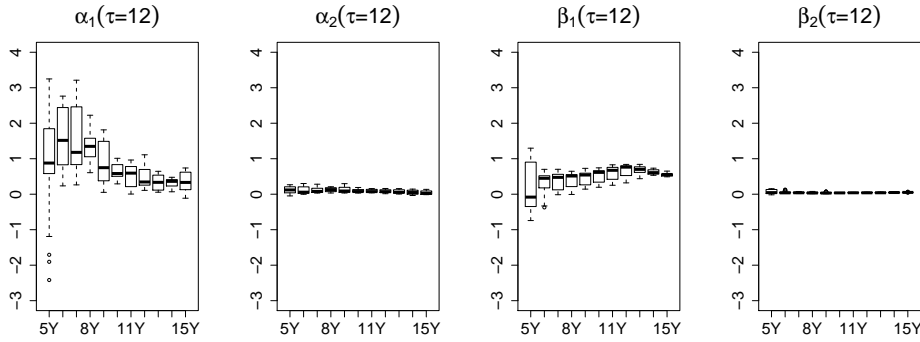


Figure 2: Box-plots of parameters estimates corresponding to macroeconomic factors of 12 months forward default (two left) and other exit (two right) intensities over 35 windows. Each box-plot represents estimation interval length 5, 6, ..., 15 years.

volatility-adjusted leverage; measured as distance-to-default (DTD) in a Merton-type model which are adjusted as in Duan et al. (2012), (ii) liquidity; measured as a ratio of cash and short term investment to total assets, (iii) profitability; measured as a ratio of net income to total assets, (iv) relative size; measured as the logarithm of the ratio of market capitalization to the economy's average market capitalization, (v) market-to-book asset ratio, and (vi) idiosyncratic volatility. The first four characteristics are transformed into level and trend. The level is computed as the one-year average of the measure. The trend is computed as the current value of the measure minus the one-year average of the measure. By doing this transformation, two firms with the same current value for all

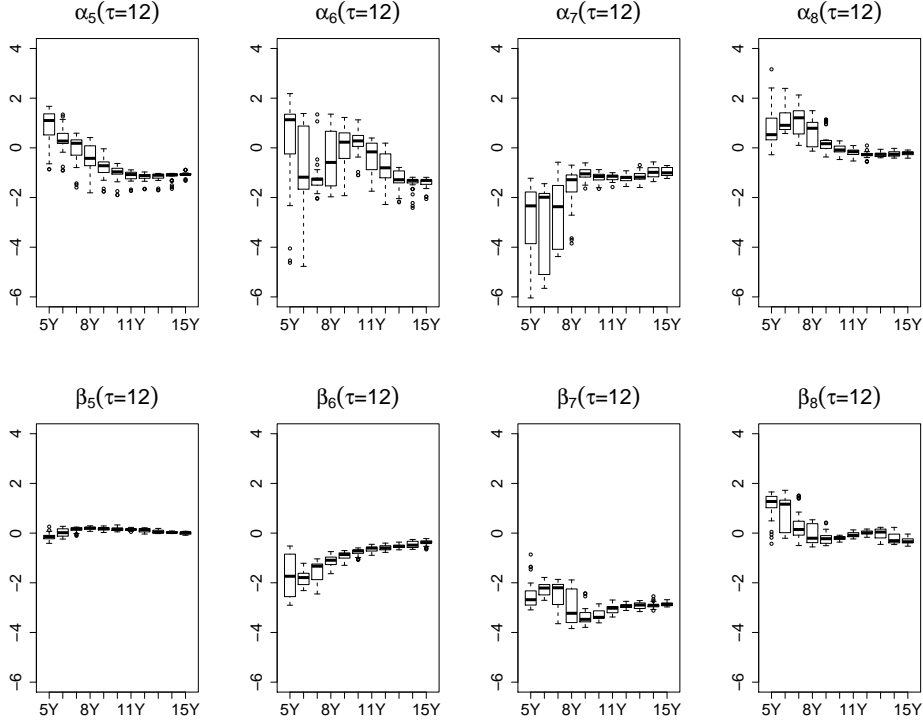


Figure 3: Box-plots of parameters estimates corresponding to  $X_j, j = 5, 6, 7, 8$ , of 12 months forward default (upper) and other exit (bottom) intensities over 35 windows. Each box-plot represents estimation interval length 5, 6,  $\dots$ , 15 years.

measures may have different PD. The state variables are summarized in Table 2.

#### 4.1 Robustness on Estimation Window Length

The estimates of parameters that determine the forward intensities,  $\theta = \{\alpha, \beta\}$ , depend on the length of estimation intervals. Figure 2 displays box-plots for the estimates of  $\alpha_j(\tau)$  and  $\beta_j(\tau)$ , where horizon  $\tau = 12$  months, that correspond to macroeconomic factors: (i) index return and (ii) interest rate. We explore the estimates for the different horizons, but we do not document them here, whereas the information extracted are similar to one year outlook. The estimate of one-year simple return of S&P500 index is sensitive to the length of estimation interval, particularly for default scheme. It contrasts to

parameter estimate of 3-months U.S. Treasury bill rate that exhibits more robustness to the calibration windows length. This finding is in line with Duan et al. (2012) that showed the standard error of  $\alpha_1(\tau)$  is much larger than  $\alpha_2(\tau)$ 's as well as that for  $\beta_1(\tau)$  and  $\beta_2(\tau)$ .

Figure 3 exhibits the estimates corresponding to liquidity and profitability, both for level and trend measures, that are sensitive to the length of estimation windows. The sensitivity increases for default events rather than other exits. The idiosyncratic volatility estimate has similar behavior as showed in Figure (1). This is also true for the intercept estimates, that empirically have negative values. This confirms the Duan et al. (2012)'s empirical result that showed the standard errors of aforesaid covariates are much larger than those of the remaining covariates. Thus, DTD, company size, and market-to-book ratio are robust to the estimation interval length.

## 4.2 Set-up for Homogeneity Intervals Test

The complete localising analysis includes two main steps: (a) design the set-up of the procedure and (b) data-driven search for the longest interval of homogeneity, the interval where a local constant parametric form describes the data well. At the first step, the delisting process is approximated by forward intensity approach with forward intensities as in (21) and (22). Next, select the interval candidates  $\{I_0, \dots, I_5\} = \{60, 72, 96, 120, 144, 180\}$  months, equivalently  $\{5, 6, 8, 10, 12, 15\}$  years, and compute test statistic  $T_{k,\tau}$  in (39). The likelihood ratio  $L_{I_k,\tau}(\tilde{\theta}_{I_k}, \hat{\theta}_{I_k})$  in (43) should be divided by an effective sample size of the corresponding interval because different firms have different life time observation. Therefore, the corresponding term  $(k/K)$  can be discarded. The critical values (40) are computed using Monte Carlo simulation. The parametric risk bound (42) is computed based on the true parameter generated from the average of es-

estimates over 35 moving windows of the longest interval (15 years). Our set up considers generating true parameter based on the grand average of estimates over interval  $I_k$  and over rolling windows at each interval. Though the risk bound is fulfilled under the simulated interval of homogeneity, it is parameter-invariant. The critical values depend on hyperparameters  $r$  and  $\rho$  that are counterparts of the usual significant level. We set two model risk levels  $r = \{0.5, 1\}$  to represent modest and conservative, respectively, and two significance levels  $\rho = \{0.5, 0.75\}$ . The conservative risk level lead to, theoretically, a selection of longer intervals of homogeneity that yield more precise estimates, but increase the modelling bias. Fortunately, this effect can be controlled according to (37).

At the second step, the smallest tested interval  $I_0$  is initially assumed to be homogeneous. If  $I_{k-1}$  is negatively tested on the presence of a change point, one continues with  $I_k$  by employing (39) to detect a potential change point in  $I_k$ . If no change point is found, then  $I_k$  is accepted as time-homogeneous. We sequentially repeat these tests until we find a change point or exhaust all interval. The latest (longest) interval accepted as time-homogeneous is used for estimation. The whole search and estimation in second step is repeated at different end time point  $T$  without reiterating the first step because the critical values  $\mathfrak{z}_{k,\tau}$ , for fixed  $\tau$ , depend only on the approximating parametric model and interval length  $m_k = |I_k|$ , not on the time point  $T$ .

Figure 4 and 5 depict the estimated length of interval of homogeneity over recent 35 windows at each horizon  $\tau = \{1, 3, 6, 12, 24, 36\}$  months. The 35-th window has time-end  $T$  at December 2011. The preceding windows are one month back-shifting. Therefore the time-end  $T$  of the  $\{1\text{-st}, 2\text{-nd}, \dots, 35\text{-th}\}$  windows are at  $\{28.02.2009, 31.03.2009, \dots, 31.12.2011\}$ . As expected, the selected intervals of homogeneity are shorter in the modest risk level ( $r = 0.5$ ) than in the conservative risk level ( $r = 1$ ). Particularly for conservative case, the test procedure apparently selects longer estimation period at recent windows. This result may happens because we employ the most recent state variable data in Monte Carlo

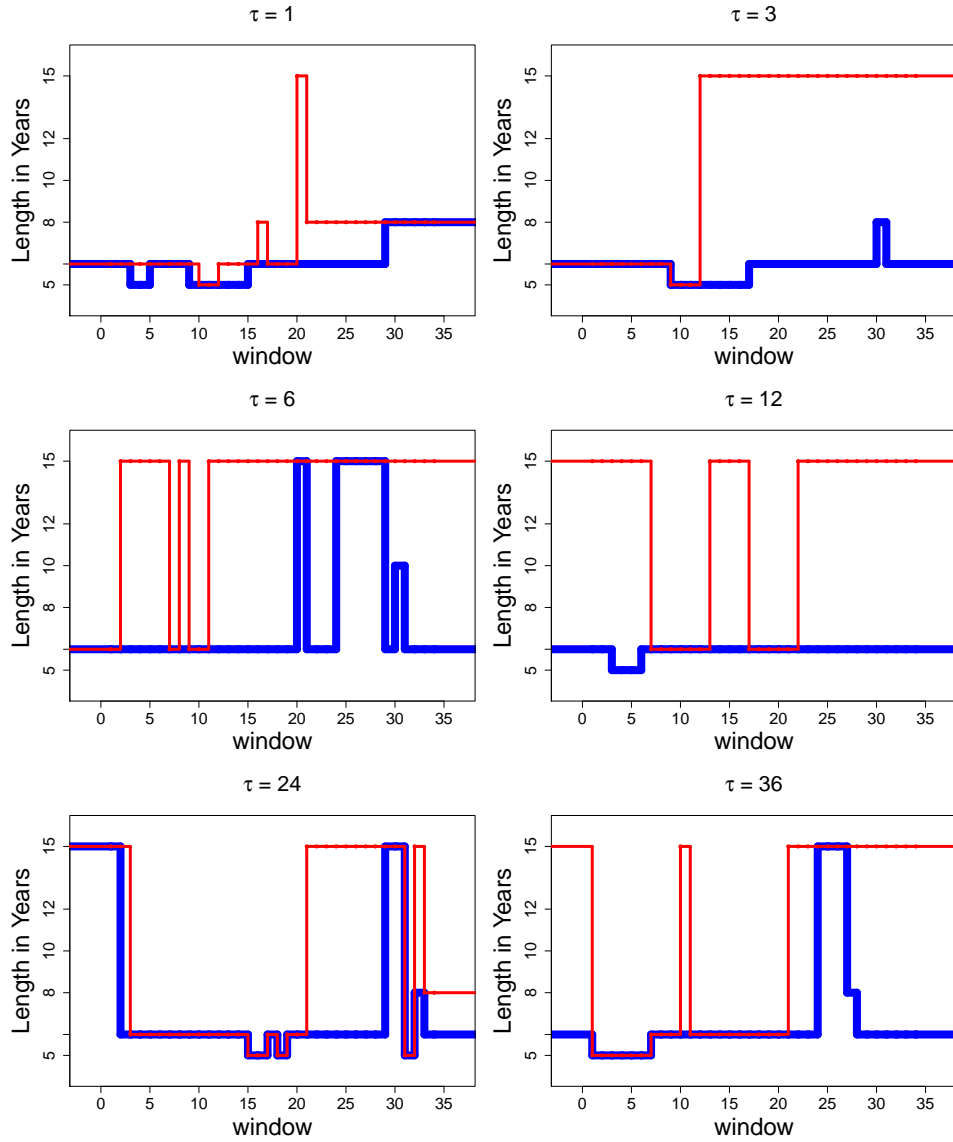


Figure 4: Estimated length of *interval of homogeneity* (in years) for 35 last windows in case of a modest ( $r = 0.5$ , blue) and conservative ( $r = 1$ , red) modelling risk level, with  $K = 5$  and  $\rho = 0.50$ .

simulation to obtain the critical values. This leads to select longer interval such that we can employ more observations to obtain the consistent estimators from the overlapped likelihood.



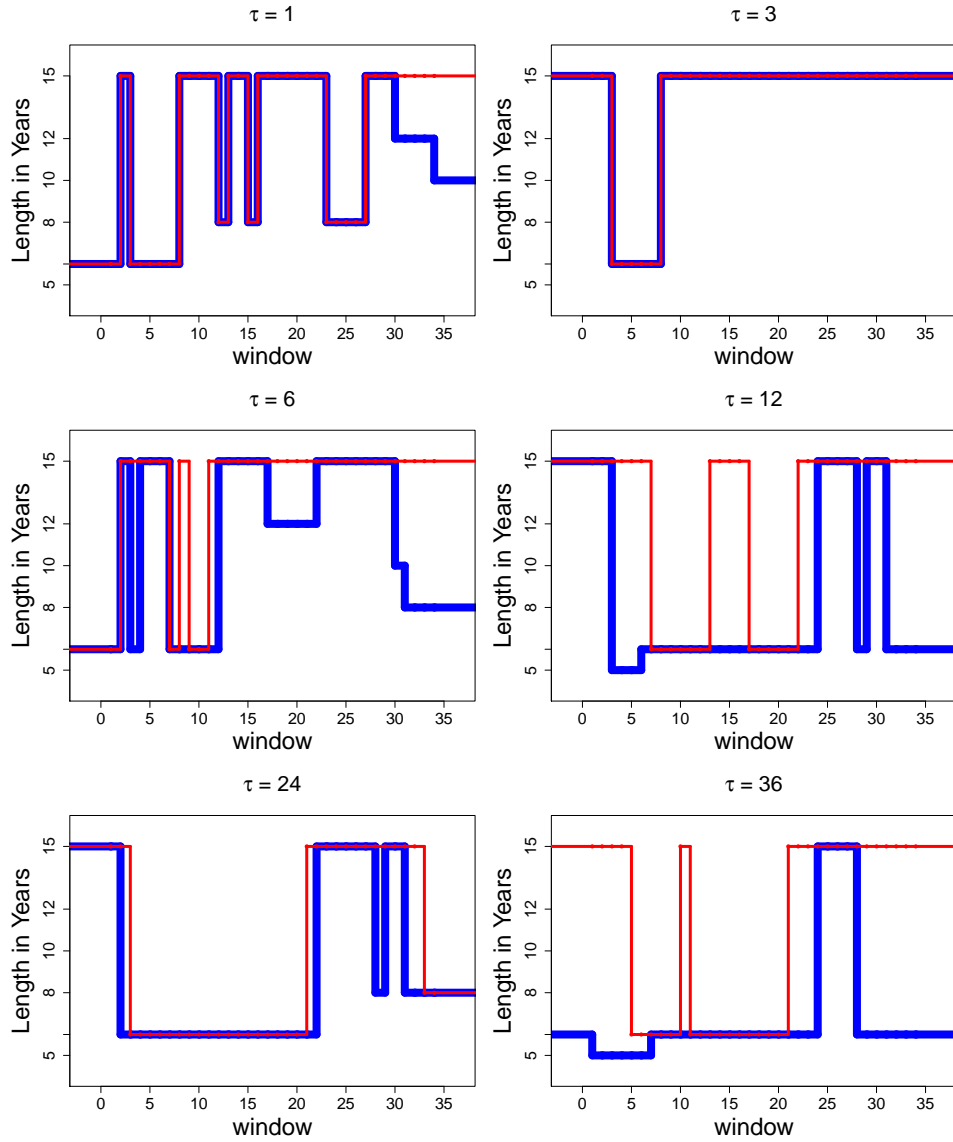


Figure 5: Estimated length of *interval of homogeneity* (in years) for 35 last windows in case of a modest ( $r = 0.5$ , blue) and conservative ( $r = 1$ , red) modelling risk level, with  $K = 5$  and  $\rho = 0.75$ .

### 4.3 Measures of Accuracy

Figure 6 shows accuracy ratio (AR) computed from cumulative accuracy profile (CAP) (Sobehart et al., 2001), also known as power curve, over windows for a fixed horizon. The CAP evaluates the performance of a model based on default risk ranking. The higher

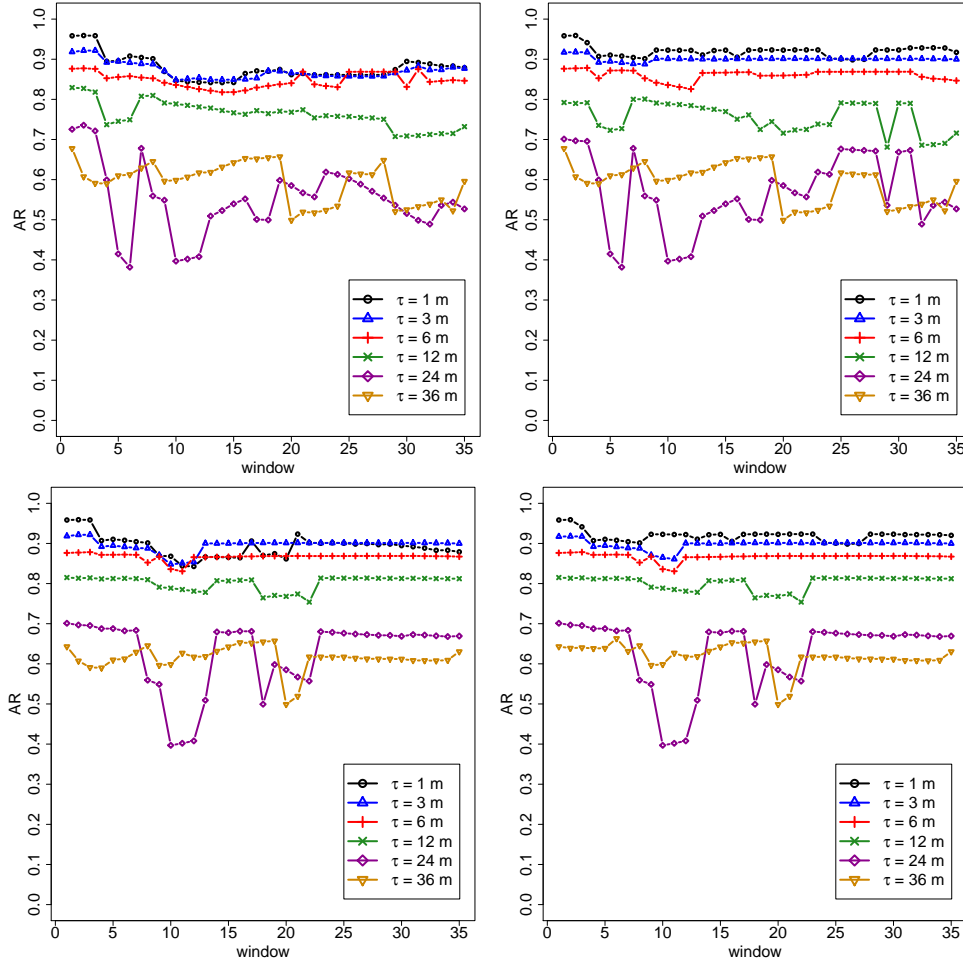


Figure 6: Accuracy ratios over windows. The first row:  $r = 0.5, \rho = 0.5$  (left) and  $r = 0.5, \rho = 0.75$  (right). The second row:  $r = 1, \rho = 0.5$  (left) and  $r = 1, \rho = 0.75$  (right).

PD implies the higher risk. The model discriminates well between healthy and distressed firms if the defaulting firms are assigned among the highest PD of all firms before they default. This leads to higher values of AR. The PD are taken to be non-overlapping. The one-year AR is based on PDs computed on, for example, 31.12.2001, 31.12.2002, ... and firms that default within one year of those dates whereas the three-years AR is based on PDs computed on 31.12.2001, 31.12.2004, ... and firms that default within three years of those dates.

Table 3: Accuracy-ratio-based performance comparison for horizon 1, 3, and 6 months.

window	$\tau = 1$				$\tau = 3$				$\tau = 6$				
	global	local			global	local			global	local			
		$r = 0.5, \rho = 0.5$	$r = 0.5, \rho = 0.75$	$r = 1, \rho = 0.5$		$r = 1, \rho = 0.75$	$r = 0.5, \rho = 0.5$	$r = 0.5, \rho = 0.75$		$r = 1, \rho = 0.5$	$r = 1, \rho = 0.75$	$r = 0.5, \rho = 0.5$	$r = 0.5, \rho = 0.75$
1		<	<	<		<	<	<		<	<	<	<
2		<	<	<		<	<	<		<	<	<	<
3		<	<	<		<	<	<		<	<	<	<
4		<	<	<		<	<	<		<	<	<	<
5	<	<	<	<	<	<	<	<	<	<	<	<	<
6	<	<	<	<	<	<	<	<	<	<	<	<	<
7	<	<	<	<	<	<	<	<	<	<	<	<	<
8	<	<	<	<	<	<	<	<	<	<	<	<	<
9	<	<	<	<	<	<	<	<	<	<	<	<	<
10	<	<	<	<	<	<	<	<	<	<	<	<	<
11	*	<	<	<	<	<	<	<	<	<	<	<	<
12	*	<	<	<	<	<	<	<	<	<	<	<	<
13	<	<	<	<	<	<	<	<	<	<	<	<	<
14	*	<	<	<	<	<	<	<	<	<	<	<	<
15	*	<	<	<	<	<	<	<	<	<	<	<	<
16	<	<	<	<	<	<	<	<	<	<	<	<	<
17	*	<	<	<	<	<	<	<	<	<	<	<	<
18	*	<	<	<	<	<	<	<	*	<	<	<	<
19	*	<	<	<	<	<	<	<	*	<	<	<	<
20	*	<	<	<	<	<	<	<	*	<	<	<	<
21	*	<	<	<	<	<	<	<	*	<	<	<	<
22	*	<	<	<	<	<	<	<	*	<	<	<	<
23	*	<	<	<	<	<	<	<	*	<	<	<	<
24	<	<	<	<	<	<	<	<	<	<	<	<	<
25	<	<	<	<	<	<	<	<	<	<	<	<	<
26	<	<	<	<	<	<	<	<	<	<	<	<	<
27	<	<	<	<	<	<	<	<	<	<	<	<	<
28	*	<	<	<	<	<	<	<	<	<	<	<	<
29	*	<	<	<	<	<	<	<	<	<	<	<	<
30	*	<	<	<	<	<	<	<	<	<	<	<	<
31		<	<	<	<	<	<	<	<	<	<	<	<
32		<	<	<	<	<	<	<	*	<	<	<	<
33		<	<	<	<	<	<	<	*	<	<	<	<
34		<	<	<	<	<	<	<	*	<	<	<	<
35	<	<	<	<	<	<	<	<	*	<	<	<	<

NOTE: The check mark ( $\checkmark$ ) denotes the corresponding approach results in higher AR whereas the star ( $*$ ) implies both the global and local FIA perform with the same accuracy. For one month horizon, the global FIA does better than the local FIA in 14 out of 35 windows, whereas our the local approach yields higher AR than those of the global approach for 7 windows. In the rest windows these two methods perform equally well. For three months horizon, the global FIA shows superiority over our local adaptive method. Both approaches perform with the same accuracy for six months horizon prediction.

The local adaptive approach results in very high AR at the first, second, and third windows particularly for one month horizon (about 95%). The ARs drop significantly

Table 4: Accuracy-ratio-based performance comparison for horizon 12, 24, and 36 months.

window	$\tau = 12$				$\tau = 24$				$\tau = 36$				
	global	local			global	local			global	local			
		$r = 0.5, \rho = 0.5$	$r = 0.5, \rho = 0.75$	$r = 1, \rho = 0.5$		$r = 1, \rho = 0.75$	$r = 0.5, \rho = 0.5$	$r = 0.5, \rho = 0.75$		$r = 1, \rho = 0.5$	$r = 1, \rho = 0.75$	$r = 0.5, \rho = 0.5$	$r = 0.5, \rho = 0.75$
1		✓	✓	✓			✓	✓		✓	✓	✓	✓
2		✓	✓	✓			✓	✓		✓	✓	✓	✓
3		✓	✓	✓			✓	✓					✓
4		✓	✓	✓			✓	✓					✓
5	✓						✓	✓					✓
6							✓	✓		✓	✓	✓	✓
7		✓	✓	✓			✓	✓					✓
8		✓	✓	✓			✓	✓		✓	✓	✓	✓
9		✓	✓	✓			✓	✓		✓	✓	✓	✓
10		✓	✓	✓			✓	✓		✓	✓	✓	✓
11		✓	✓	✓			✓	✓		✓	✓	✓	✓
12		✓	✓	✓			✓	✓		✓	✓	✓	✓
13		✓	✓	✓			✓	✓		✓	✓	✓	✓
14		✓	✓	✓			✓	✓		✓	✓	✓	✓
15		✓	✓	✓			✓	✓		✓	✓	✓	✓
16		✓	✓	✓			✓	✓		✓	✓	✓	✓
17		✓	✓	✓			✓	✓		✓	✓	✓	✓
18		✓	✓	✓			✓	✓		✓	✓	✓	✓
19		✓	✓	✓			✓	✓		✓	✓	✓	✓
20		✓	✓	✓			✓	✓		✓	✓	✓	✓
21		✓	✓	✓			✓	✓		✓	✓	✓	✓
22		✓	✓	✓			✓	✓		✓	✓	✓	✓
23		✓	✓	✓			✓	✓		✓	✓	✓	✓
24		✓	✓	✓			✓	✓		✓	✓	✓	✓
25		✓	✓	✓			✓	✓		✓	✓	✓	✓
26		✓	✓	✓			✓	✓		✓	✓	✓	✓
27		✓	✓	✓			✓	✓		✓	✓	✓	✓
28		✓	✓	✓			✓	✓		✓	✓	✓	✓
29		✓	✓	✓			✓	✓		✓	✓	✓	✓
30		✓	✓	✓			✓	✓		✓	✓	✓	✓
31		✓	✓	✓			✓	✓		✓	✓	✓	✓
32		✓	✓	✓			✓	✓		✓	✓	✓	✓
33		✓	✓	✓			✓	✓		✓	✓	✓	✓
34		✓	✓	✓			✓	✓		✓	✓	✓	✓
35		✓	✓	✓			✓	✓		✓	✓	✓	✓

NOTE: The check mark (✓) denotes the corresponding approach results in higher AR whereas the star (\*) implies both the global and local FIA perform with the same accuracy. For 12 months horizon, the local FIA outperforms the global FIA in 18 out of 35 windows, whereas the local FIA yields higher AR in the rest windows. The local adaptive method shows superiority over the global FIA for 24 months horizon (in 24 out of 35 windows). For 36 months horizon prediction, the local FIA method performs much better than the benchmark in 26 out of 35 windows.

at certain windows as exhibited in Figure 6. The proposed approach is able to generate accurate predictions, about 90%, for one month horizon. When the prediction horizon is extended to three and six months, the AR is still above 86% and 83%, respectively.

The one year prediction horizon drops the AR to the 75% – 80% range. For conservative modelling risk level ( $r = 1$ ) the accuracies are still above 60% for both two and three years horizon. We evaluate the performance of local adaptive approach against the global FIA that employs all past observations. This comparison is summarized in Table 3 and 4. The global FIA performs better than the localising algorithm for short outlook: one and three months horizon. Both two methods perform equally well for six months horizon prediction. Our local adaptive technique outperforms the benchmark for one year or longer horizon. This finding shows the accuracy prediction for long horizon can be increased by localising the time varying forward intensities and safely approximating them with constant.

## 5 CONCLUSION

In this paper we extend the idea of adaptive pointwise estimation to forward intensities calibration for multiperiod corporate default prediction. The FIA itself has simplicity substantially from the fact that no state variable forecasting model is required. Our approach addresses the inhomogeneity of parameters over time by optimally selecting the sample period over which parameters are approximately constant. The sequential LPA procedure provides an interval of homogeneity, the interval where a local constant parametric form describes the data well, that is used for modelling and prediction.

Applying the proposed method to monthly data on 2000 U.S. public firms over a sample period from 1991 to 2011, we estimate default probabilities over various prediction horizons. The default prediction performance is evaluated against the global FIA that employs all past observations. We utilize accuracy ratio from CAP curve to evaluate the performance of models based on default risk ranking. For the six months prediction horizon, the local adaptive approach performs with the same accuracy as the benchmark.

We show empirical evidence of increase in default prediction power for the longer horizon (one to three years).

The general framework of the FIA allows adjustments on forward intensities specification for future research. Our local adaptive method is data-driven and can be applied to those other specifications with different covariates, either macroeconomic or firm-specific drivers.

## APPENDIX A: CUMULATIVE FORWARD DEFAULT INTENSITY

This part shows the relationship between forward intensities and its cumulative. Denote a differentiable  $F_t(s)$ , the conditional cdf of  $\tau_D$  evaluated at  $t + s$ .

$$\begin{aligned} F_t(s) &= 1 - \exp\{-\psi_t(s)s\} \\ F'_t(s) &= -\exp\{-\psi_t(s)s\}\{-\psi'_t(s)s - \psi_t(s)\} \\ &= \exp\{-\psi_t(s)s\}\psi'_t(s)s + \exp\{-\psi_t(s)s\}\psi_t(s). \end{aligned}$$

Therefore

$$\begin{aligned} \lambda_t(s) &= \frac{F'_t(s)}{1 - F_t(s)} \\ &= \frac{\exp\{-\psi_t(s)s\}\psi_t(s) + \exp\{-\psi_t(s)s\}\psi'_t(s)s}{\exp\{-\psi_t(s)s\}} \\ &= \psi_t(s) + \psi'_t(s)s. \end{aligned}$$

This shows (6) and consequently:

$$\begin{aligned}
\int_0^s \lambda_t(u) du &= \int_0^s \psi_t(u) du + \int_0^s \psi_t'(u) u du \\
&= \int_0^s \psi_t(u) du + \psi_t(s)s - \int_0^s \psi_t(u) du \\
&= \psi_t(s)s.
\end{aligned}$$

## APPENDIX B: PARAMETRIC RISK BOUND

This part proves the parametric risk bound is finite.

Define  $\mathcal{E}(\mathfrak{z}) \stackrel{\text{def}}{=} \left\{ \theta^* : L_K(\tilde{\theta}_K) - L_K(\theta^*) \leq \mathfrak{z} \right\}$ , the parametric risk bound:

$$\begin{aligned}
\mathcal{R}_r(\theta^*) &= \mathbf{E}_{\theta^*} \left| L_K(\tilde{\theta}_K, \theta^*) \right|^r \\
&= - \int_{\mathfrak{z} \geq 0} \mathfrak{z}^r d\mathbf{P}_{\theta^*} \left\{ \left| L_K(\tilde{\theta}_K, \theta^*) \right| > \mathfrak{z} \right\} \\
&= r \int_0^\infty \mathfrak{z}^{r-1} \mathbf{P}_{\theta^*} \left\{ \left| L_K(\tilde{\theta}_K, \theta^*) \right| > \mathfrak{z} \right\} d\mathfrak{z} \\
&= r \int_0^\infty \mathfrak{z}^{r-1} \mathbf{P}_{\theta^*} \left\{ \left| L_K(\tilde{\theta}_K, \theta^*) \right| > \mathfrak{z}, \tilde{\theta}_K \in \mathcal{E}(\mathfrak{z}) \right\} d\mathfrak{z} \\
&\quad + r \int_0^\infty \mathfrak{z}^{r-1} \mathbf{P}_{\theta^*} \left\{ \left| L_K(\tilde{\theta}_K, \theta^*) \right| > \mathfrak{z}, \tilde{\theta}_K \notin \mathcal{E}(\mathfrak{z}) \right\} d\mathfrak{z} \\
&\leq 2r \int_0^\infty \mathfrak{z}^{r-1} e^{-\mathfrak{z}} d\mathfrak{z} \\
&< \infty
\end{aligned}$$

## REFERENCES

Bai, J. and Perron, P. (1998), “Estimating and Testing Linear Models with Multiple Structural Changes,” *Econometrica*, 66 (1), 47–78.

Belomestny, D. and Spokoiny, V. (2007), “Spatial Aggregation of Local Likelihood Esti-

- mates with Applications to Classification,” *The Annals of Statistics*, 35 (5), 2287–2311.
- Bharath, S. T. and Shumway, T. (2008), “Forecasting Default with the Merton Distance to Default Model,” *The Review of Financial Studies*, 21 (3), 1339–1369.
- Cai, Z., Fan, J., and Yao, Q. (2000), “Functional-Coefficient Regression Models for Non-linear Time Series,” *J. Am. Stat. Assoc.*, 95 (451), 941–956.
- Campbell, J. Y., Hilscher, J., and Szilagyi, J. (2008), “In Search of Distress Risk,” *Journal of Finance*, 63 (6), 2899–2939.
- Chava, S. and Jarrow, R. A. (2004), “Bankruptcy Prediction with Industry Effects,” *Review of Finance*, 8 (4), 537–569.
- Chen, S., Härdle, W., and Moro, R. (2011), “Modeling Default Risk with Support Vector Machines,” *Quantitative Finance*, 11 (1), 135–154.
- Chen, Y., Härdle, W., and Jeong, S.-O. (2008), “Nonparametric Risk Management with Generalized Hyperbolic Distribution,” *J. Am. Stat. Assoc.*, 103 (483), 910–923.
- Chen, Y., Härdle, W., and Pigorsch, U. (2010), “Localized Realized Volatility Modeling,” *J. Am. Stat. Assoc.*, 105 (492), 1376–1393.
- Chen, Y. and Niu, L. (2014), “Adaptive Dynamic Nelson-Siegel Term Structure Model with Applications,” *Journal of Econometrics*, 180 (1), 98–115.
- Čížek, P., Härdle, W., and Spokoiny, V. (2009), “Adaptive Pointwise Estimation in Time-Inhomogeneous Conditional Heterocedasticity Models,” *The Econometrics Journal*, 12 (2), 248–271.
- Duan, J.-C., Sun, J., and Wang, T. (2012), “Multiperiod Corporate Default Prediction - A Forward Intensity Approach,” *Journal of Econometrics*, 170 (1), 191–209.



- Duffie, D., Eckner, A., Horel, G., and Saita, L. (2009), “Frailty Correlated Default,” *Journal of Finance*, 64 (5), 2089–2123.
- Duffie, D., Saita, L., and Wang, K. (2007), “Multi-period Corporate Default Prediction with Stochastic Covariates,” *Journal of Financial Economics*, 83 (3), 635–665.
- Fan, J. and Zhang, W. (2008), “Statistical Methods with Varying Coefficient Models,” *Stat Interface*, 1 (1), 179–195.
- Giacomini, E., Härdle, W., and Spokoiny, V. (2009), “Inhomogeneous Dependence Modeling with Time-Varying Copulae,” *Journal of Business & Economic Statistics*, 27 (2), 224–234.
- Härdle, W., Lee, Y.-J., Schäfer, D., and Yeh, Y.-R. (2009), “Variable Selection and Oversampling in the Use of Smooth Support Vector Machines for Predicting the Default Risk of Companies,” *Journal of Forecasting*, 28 (6), 512–534.
- Härdle, W. and Prastyo, D. D. (2014), “Embedded Predictor Selection for Default Risk Calculation: A Southeast Asian Industry Study,” in *Handbook of Asian Finance: Financial Markets and Sovereign Wealth Funds*, eds. Chuen, D. L. K. and Gregoriou, G. N., Academic Press, vol. 1, pp. 131–148.
- Härdle, W. K., Prastyo, D. D., and Hafner, C. M. (2014), “Support Vector Machines with Evolutionary Model Selection for Default Prediction,” in *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, eds. Racine, J. S., Su, L., and Ullah, A., Oxford University Press, pp. 346–373.
- Katkovnik, V. and Spokoiny, V. (2008), “Spatially Adaptive Estimation via Fitted Local Likelihood Techniques,” *IEEE Transactions on Signal Processing*, 56 (3), 873–886.
- Mercurio, D. and Spokoiny, V. (2004), “Statistical Inference for Time-Inhomogeneous Volatility Models,” *The Annals of Statistics*, 32 (2), 577–602.

- Shumway, T. (2001), “Forecasting Bankruptcy More Accurately: A Simple Hazard Model,” *The Journal of Business*, 74 (1), 101–124.
- Sobehart, J., Keenan, S., and Stein, R. (2001), “Benchmarking Quantitative Default Risk Models: A Validation Methodology,” *Algo Research Quarterly*, 4 (1), 57–72.
- Spokoiny, V. (1998), “Estimation of A Function with Discontinuities via Local Polynomial Fit with An Adaptive Window Choice,” *The Annals of Statistics*, 26 (4), 1356–1378.
- (2009), “Multiscale Local Change Point Detection with Application to Value-at-Risk,” *The Annals of Statistics*, 37 (3), 1405–1436.