Estimation in discontinuous Bernoulli mixture models applicable in credit rating systems with dependent data

von

Daniel Tillich und Christoph Lehmann
Estimation in discontinuous Bernoulli mixture models applicable in credit rating systems with dependent data

Daniel Tillich∗, Christoph Lehmann†

August 30, 2016

∗Daniel Tillich, Technische Universität Dresden, Faculty of Business and Economics, Chair of Quantitative Methods, esp. Statistics, 01062 Dresden, Germany. E-Mail: daniel.tillich@tu-dresden.de.
†Christoph Lehmann, Technische Universität Dresden, Faculty of Business and Economics, Chair of Quantitative Methods, esp. Statistics, 01062 Dresden, Germany. E-Mail: christoph.lehmann@tu-dresden.de.
Abstract:

OBJECTIVE: We consider the following problem from credit risk modeling: Our sample \((X_i, Y_i), 1 \leq i \leq n\), consists of pairs of variables. The first variable \(X_i\) measures the creditworthiness of individual \(i\). The second variable \(Y_i\) is the default indicator of individual \(i\). It has two states: \(Y_i = 1\) indicates a default, \(Y_i = 0\) a non-default. A default occurs, if individual \(i\) cannot meet its contractual credit obligations, i.e. it cannot pay back its outstandings regularly. In a first step, our objective is to estimate the threshold between good and bad creditworthiness in the sense of dividing the range of \(X_i\) into two rating classes: One class with good creditworthiness and a low probability of default and another class with bad creditworthiness and a high probability of default.

METHODS: Given observations of individual creditworthiness \(X_i\) and defaults \(Y_i\), the field of change point analysis provides a natural way to estimate the breakpoint between the rating classes. In order to account for dependency between the observations, the literature proposes a combination of three model classes: These are a breakpoint model, a linear one-factor model for the creditworthiness \(X_i\), and a Bernoulli mixture model for the defaults \(Y_i\). We generalize the dependency structure further and use a generalized link between systematic factor and idiosyncratic factor of creditworthiness. So the systematic factor cannot only change the location, but also the form of the distribution of creditworthiness.

RESULTS: For the case of two rating classes, we propose several estimators for the breakpoint and for the default probabilities within the rating classes. We prove the strong consistency of these estimators in the given non-i.i.d. framework. The theoretical results are illustrated by a simulation study. Finally, we give an overview of research opportunities.

Keywords: regression with jump, change point, split point, credit risk, rating classification, default probability, dependence, strong consistency
1 Introduction

Let \((X_i, Y_i), 1 \leq i \leq n\), be bivariate observations. The variable \(X_i \in \mathbb{R}\) measures the creditworthiness of debtor \(i\), which is a person, a firm or a country. In credit risk, this variable is also called a score. High values represent a high creditworthiness and vice versa. The variable \(Y_i\) is the so called default indicator variable of debtor \(i\) with two possible states. The value \(Y_i = 1\) is taken in the case of default of debtor \(i\) and \(Y_i = 0\) in the case of non-default. Thereby, a default event occurs if a debtor does not meet its contractual credit obligations.

The idea of the joint behavior of the variables is as following. With increasing creditworthiness the corresponding probability of default (PD) decreases stepwise. As a consequence, different rating classes can be constructed with class specific default probabilities (also called risk levels) \(\pi_k\). The rating class borders \(\theta_k\) can be interpreted as breakpoints. A breakpoint indicates a structural change within the default probabilities. Finally, the main target is an estimation of the breakpoints, and thereby the rating classes, as well as the corresponding risk levels. In the contribution at hand, we start with two classes, i.e. there exists exactly one breakpoint \(\theta\). In the credit risk framework one could simply think of good and bad creditworthiness.

Up to now, the issue of breakpoint and risk level estimation has been discussed mainly for the i.i.d. case, e.g. by Dempe and Stute [2002] and Ferger and Klotsche [2009]. Therein, different types of breakpoint estimators—namely the maximum likelihood (ML) estimator, the Dempe-Stute (DS) estimator and the plug-in (PI) estimator—were developed and investigated. These estimators turned out to be strongly consistent under quite general assumptions. Tillich [2013] reviewed these contributions and found connections to the credit risk literature.

An important question in credit risk is the dependence structure of debtors, because the assumption of independence is unrealistic. For modeling the dependence structure, Tillich and Ferger [2015] propose a linear factor model for the creditworthiness variable \(X_i\) (see Section 2 for more details). The factor model incorporates a systematic factor affecting all scores \(X_i\) and therefore causing dependence. In this non-i.i.d. framework, Tillich and Ferger [2015] investigate the Dempe-Stute estimator for the breakpoint and plug-in estimators for the risk levels and show the strong consistency of these estimators.

But there is still another problem: In application, measured creditworthiness can be bounded, see Tillich [2013]. Hence, the systematic factor does not only cause a shift of the creditworthiness distribution, but changes its shape as well. Thus, the next step is a
Table 1: Overview on different theoretical main results and assumed setting of contributions in the field of breakpoint (and risk level) estimation.

<table>
<thead>
<tr>
<th>contribution</th>
<th>model</th>
<th>investigation of estimators for breakpoint $\theta$</th>
<th>risk levels $\pi_1$, $\pi_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>DS</td>
<td>ML</td>
</tr>
<tr>
<td>Dempfle and Stute [2002]</td>
<td>i.i.d.</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ferger and Klotzche [2009]</td>
<td>i.i.d.</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tillich [2013]</td>
<td>linear link</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tillich and Ferger [2015]</td>
<td>non-i.i.d.</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hähle [2014]</td>
<td>non-linear link</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>contribution at hand</td>
<td>link</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

generalized, namely a non-linear, link of systematic and individual effect. Based on the working paper leading to [Tillich and Ferger [2015], Hähle [2014] developed the concept of proving that the already proposed estimators are strongly consistent in the non-linear case. However, some important questions referring to the requirements on the non-linear link function still remain unanswered. This is the starting point of the contribution at hand.

As a concluding overview on the different works in the field of breakpoint (and risk level) estimation, Table 1 provides an orientation on the theoretical main results of the several contributions.

This contribution is structured as follows. Section 2 describes the model background for the observations $(X_i, Y_i)$. Basically, we assume a non-i.i.d. setting with a generalized model in the sense of a non-linear link function. In this setting, we investigate the maximum-likelihood estimator (Section 3), the Dempfle-Stute estimator (Section 4), and some plug-in estimators (Sections 5 & 6). In Section 7 some technical assumptions are clarified and illustrated with examples. In addition, we support the theoretical work with a short simulation study (Section 8). Section 9 concludes the paper and provides an outlook.
2 One-factor, Breakpoint, and Bernoulli Mixture Model

As can be seen from Table 1, the former contributions are based on different assumptions on the dependence structure of the pairs \((X_i, Y_i), 1 \leq i \leq n\). The starting point was to assume the bivariate observations \((X_i, Y_i)\) to be i.i.d., cf. [Dempe and Stute 2002, Ferger and Klotsche 2009, Tillich 2013]. The next step was to take dependency into account as done in [Tillich and Ferger 2015]. Therein a linear one-factor model (in connection with a Bernoulli mixture model) was used to create dependence between the random vectors \((X_i, Y_i)\), i.e.

\[ X_i = Z + U_i, \quad 1 \leq i \leq n, \quad (2.1) \]

where the latent variables \(Z, U_1, \ldots, U_n\) are assumed to be mutually independent.

Within the present contribution we extend the idea from \((2.1)\) by assuming that \(X_i\) is a more general transformation of \(Z\) and \(U\), i.e.

\[ X_i = T(Z, U_i), \quad 1 \leq i \leq n, \quad (2.2) \]

where \(T : \mathbb{R}^2 \to \mathbb{R}\) is a real-valued function of two real variables. The latent variables \(Z, U_1, \ldots, U_n\) are assumed to be mutually independent. The random variables \(U_i\) are assumed to be identically distributed with distribution function \(F_U\). Hence, the score variables \(X_i\) are identically distributed with distribution function \(F_X\). The marginal distributions as well as the joint distributions of \(Z\) and the \(U_i\), and therewith the distribution of \(X_i\), are unknown. During the work, we will need additional assumptions about \(T, Z, U\) and \(X\). We will discuss them on the spot.

Analogously to the linear one-factor model, the random variable \(Z\) can be interpreted as a systematic risk factor, which influences all the scores \(X_i\). In the context of credit risk, it reflects the economic situation as a whole. The so called idiosyncratic factors \(U_i\) represent individual effects on the creditworthiness. By using a non-linear link function \(T\), the systematic factor \(Z\) can not only cause a shift of the distribution of \(U\), but a more complex change of its whole shape.

The next step is to connect the scores with the default event, i.e. linking \(X_i\) and \(Y_i\). This is done by the assumption of conditional independence and a two-state conditional
default probability,

\[ P(Y \in \cdot \mid X = x) = \bigotimes_{i=1}^{n} \text{Ber}(m(x_i)), \]  

(2.3)

where \( X = (X_1, \ldots, X_n) \) is the random vector of the scores with realization \( x = (x_1, \ldots, x_n) \in \mathbb{R}^n \) and \( Y = (Y_1, \ldots, Y_n) \) is the random vector of the corresponding Bernoulli variables. Finally, \( m : \mathbb{R} \to [0,1[ \) is a one-step function given by

\[ m(x) = \pi_1 \cdot I_{\{x \leq \theta\}} + \pi_2 \cdot I_{\{x > \theta\}} = \begin{cases} 
\pi_1, & \text{if } x \leq \theta, \\
\pi_2, & \text{if } x > \theta,
\end{cases} \]  

(2.4)

with risk levels \( 0 < \pi_1 < 1, 0 < \pi_2 < 1, \pi_1 \neq \pi_2 \), and breakpoint \( \theta \in \mathbb{R} \) (see Figure 1).

In the context of credit risk, Equation (2.3) can be interpreted as follows. Given the realizations \( x_i \) of all scores \( X_i \), the default variables \( Y_i \) are conditional independent and Bernoulli distributed with Bernoulli parameter \( m(x_i) \). Hence, the conditional default probability of debtor \( i \) depends only on its own creditworthiness \( x_i \) and it takes either the value \( \pi_1 \) or \( \pi_2 \).

Using factor models and the assumption of conditional independence is quite common in credit risk modeling. Often, the factor model is combined with a threshold model, see e. g. Bluhm et al. [2010], Höse and Huschens [2010], McNeil et al. [2005] or Schönbucher [2003]. Within this type of model, the default \( Y_i = 1 \) occurs if and only if \( X_i \leq c_i \), where \( c_i \) is a fixed threshold, i. e. \( Y_i = I_{\{X_i \leq c_i\}} \). In contrast, the model used here does not contain such a deterministic link between \( X_i \) and \( Y_i \).

Both model types have in common that they can be seen as a two-step random experiment. At first, the scores \( X_i \) realize. Then, determined via a model specific connection, the variables \( Y_i \) realize. In our case the connection is modeled by the step function \( m \).
Consequently, and in contrast to the threshold model mentioned above, for every \( x_i \in \mathbb{R} \), the outcomes \( Y_i = 0 \) and \( Y_i = 1 \) are possible. The variable \( m(X_i) \) serves as the random success (in our context: default) probability of the Bernoulli variable \( Y_i \). Thus, the given model is a Bernoulli mixture model and it follows, compare for instance Bluhm et al. \[2010\] pp. 55-56:

\[
\text{Cov}[Y_i, Y_j] = \text{Cov}[m(X_i), m(X_j)].
\]

The dependency between the scores \( X_i \) caused by Equation (2.2) leads to a dependency of the unconditional variables \( Y_i \). Compare Tillich \[2016a\] for a detailed discussion of the special case (2.1).

Now the aim of the statistician is to estimate the model parameters, namely the breakpoint \( \theta \) and the risk levels \( \pi_1 \) and \( \pi_2 \). For this, we make

**Assumption 2.1** The link function \( T: \mathbb{R}^2 \to \mathbb{R} \) as well as the distributions of \( Z \) and \( U_1, \ldots, U_n \) from (2.2) do not depend on the parameters \( \theta, \pi_1 \) and \( \pi_2 \).

In order to estimate the breakpoint \( \theta \) from the observations \((X_i, Y_i)\), a necessary assumption is \( 0 < F_X(\theta) < 1 \). If this assumption is not fulfilled, there are no score observations \( X_i \) to the left or to the right of \( \theta \). Thus, a detection of \( \theta \) would be impossible, see also Dempfle and Stute \[2002\] p. 235]. Are there observations on either side of the breakpoint \( \theta \), we try to exploit the structural change in the probabilities of default \( \pi_1 \) and \( \pi_2 \) in order to estimate \( \theta \). Which further assumptions are needed, we will investigate in the next sections.

### 3 Maximum likelihood estimator for the breakpoint

Let \( Q_n = P_X \) denote the distribution of \( X \) and let \( \mu \) denote the local counting measure on \( \{0, 1\} \), i.e.

\[
\mu = \delta_0 + \delta_1,
\]

where \( \delta_w \) is the Dirac measure (unit mass) at point \( w \), compare Schmidt \[2011\] pp. 49-50] or Schilling \[2011\] p. 26]. For every Borel-set \( A \subseteq \mathbb{R}^n \) and \( B \subseteq \{0, 1\}^n \), it holds

\[
P((X, Y) \in A \times B) = \int_A P(Y \in B \mid X = x) \, P_X(dx)
\]

\[
= \int_A \bigotimes_{i=1}^n \text{Ber}(m(x_i))(B) \, P_X(dx)
\]
\[ 
= \int_A \int_B \prod_{i=1}^n m(x_i)^{y_i} (1 - m(x_i))^{1-y_i} \, \mu^n(dy) \, P_X(dx).
\]

From Assumption 2.1 and (2.2), it follows that the distribution \( Q_n \) of \( X \) does not depend on the unknown parameters \( \theta, \pi_1 \) and \( \pi_2 \). Thus, the random vector \((X, Y)\) has the \((Q_n \otimes \mu^n)\)-density
\[
\prod_{i=1}^n m(x_i)^{y_i} (1 - m(x_i))^{1-y_i}.
\]

Hence, the likelihood function is the same as in the i.i.d. case, see Feige and Klotsche [2009]. Since \( m \) is of type (2.4), the log-likelihood function can be written as
\[
\ell_n(\theta) = \sum_{i=1}^n 1\{x_i \leq \theta\} \left( y_i \ln \pi_1 + (1 - y_i) \ln(1 - \pi_1) \right) 
+ \sum_{i=1}^n 1\{x_i > \theta\} \left( y_i \ln \pi_2 + (1 - y_i) \ln(1 - \pi_2) \right) 
= \sum_{i=1}^n 1\{x_i \leq \theta\} \left( y_i \ln \left( \frac{\pi_1}{\pi_2} \frac{1 - \pi_2}{1 - \pi_1} \right) + \ln \frac{1 - \pi_1}{1 - \pi_2} \right) 
+ \sum_{i=1}^n (y_i \ln \pi_2 + (1 - y_i) \ln(1 - \pi_2)).
\] (3.1)

In both steps, we use \( 1 = 1\{x_i \leq \theta\} + 1\{x_i > \theta\} \). See [Tillich 2013, pp. 58-59] for a more detailed calculation.

Note that the second sum in (3.2) does not contain the parameter \( \theta \). Thus, as in the i.i.d. case studied by Feige and Klotsche [2009] p. 98], the maximum likelihood (ML) estimator for \( \theta \) is given by
\[
\theta_n^* := \arg\max_{x \in \mathbb{R}} S_n^*(x),
\] (3.3)

where the criterion function to be maximized is given by
\[
S_n^*(x) := \frac{1}{n} \sum_{i=1}^n 1\{X_i \leq x\} \left( \alpha Y_i + \beta \right)
\] (3.4)

with
\[
\alpha := \ln \left( \frac{\pi_1}{\pi_2} \frac{1 - \pi_2}{1 - \pi_1} \right) \quad \text{and} \quad \beta := \ln \frac{1 - \pi_1}{1 - \pi_2}.
\] (3.5)

For the following remark also see [Tillich 2013, pp. 59-60].
Remark 3.1 (On the criterion function $S_n^*$)

(i) The realization of the random function $S_n^*$ differs from the first part of the log-likelihood function by the factor $\frac{1}{n}$. This has no influence on the maximum point of the considered functions. Thus, it is justified to call $\theta_n^*$ ML estimator.

(ii) Using the factor $\frac{1}{n}$, we get an alternative representation of the empirical process $S_n^*$:

$$S_n^*(x) = \alpha H_n(x) + \beta F_n(x), \quad x \in \mathbb{R},$$

where $F_n$ denotes the random empirical distribution function of the scores

$$F_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}$$

and $H_n$ is the corresponding marked empirical distribution function

$$H_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} Y_i.$$ 

(iii) If $\pi_1 > \pi_2$, then $\beta < 0 < \alpha$ and $\alpha + \beta = \ln \frac{\pi_1}{\pi_2} > 0$ and vice versa.

Let $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ denote the order statistics of this sample $X_1, X_2, \ldots, X_n$. In particular, $X_{1:n}$ is the minimum and $X_{n:n}$ is the maximum of the sample. From Remark 3.1(iii) we deduce the following properties of the empirical process $S_n^*$ (see also Ferger and Klotsche [2009, p. 98] and Tillich [2013, p. 60]):

(i) $S_n^*: \mathbb{R} \to \mathbb{R}$ is a random step function. The jumps can only occur at the observations $X_i$. The trajectories of $S_n^*$ are right continuous with left limits (rcdl/lc), i. e. for all $x_0 \in \mathbb{R}$ the left limit $S_n^*(x_0-):= \lim_{x \uparrow x_0} S_n^*(x)$ and the right limit $S_n^*(x_0+):= \lim_{x \downarrow x_0} S_n^*(x)$ exist and $S_n^*(x_0+)=S_n^*(x_0)$.

(ii) It holds $S_n^*(x)=0$ for all $x < X_{1:n}$ and $S_n^*(x)=\alpha \bar{Y}_n + \beta$ for all $x \geq X_{n:n}$, where

$$\bar{Y}_n := \frac{1}{n} \sum_{i=1}^{n} Y_i = \lim_{x \to \infty} H_n(x)$$

is the overall success (default) rate. The exact functional form in between is random, but due to (2.4) and Remark 3.1(iii) we expect an increase followed by a decrease.
The function $S_n^*$ does not have a unique maximum point. But there exists a unique smallest maximum point, whenever the value of $S_n^*$ is greater than zero on at least one interval $[X_{i:n}, X_{(i+1):n}]$, $i = 1, \ldots, n-1$, or $[X_{n:n}, +\infty]$. In this case $\theta_n^*$ is taken to be that smallest maximizer. Otherwise, we use the smallest observation, i.e. $\theta_n^* = X_{1:n}$.

Just like Ferger and Klotsc he [2009], we start with the temporary assumption that the risk levels $\pi_1$ and $\pi_2$ are known. Hence, also $\alpha$ and $\beta$ are known. In order to prove strong consistency of the ML estimator $\theta_n^*$ we want to apply the Argmax-Theorem of Ferger [2009, p. 25, Thm.4.6], which can also be found in Ferger and Klotsc he [2009, p. 124, Thm. A.1]. A multivariate extension of this can be found in Ferger [2015, p. 28, Thm. 3.3]. To this we additionally have to prove the following:

(P1) $S_n^*$ converges uniformly to its limit process $S^*$ almost surely (a.s.).

(P2) The limit process $S^*$ of $S_n^*$ is rcll (càdlàg).

(P3) The breakpoint $\theta$ is a.s. the well-separated maximum point of $S^*$, i.e.

$$S^*(\theta) > \sup \{ S^*(x) : |x - \theta| > \epsilon \} \text{ for all } \epsilon > 0.$$  

If these properties are fulfilled, then the convergence of the processes is transferred to their maximum points. During the proof of (P1)-(P3) we will collect requirements to formulate a corresponding theorem.

We begin with studying the two random parts of (3.6). Note, that the following structure is very similar to the case of linear links studied by Tillich and Ferger [2015, Section 3.2]. Furthermore, by $\| \cdot \|$ we denote the supremum norm, i.e. $\|f\| := \sup_{x \in \mathbb{R}} |f(x)|$. By $\| \cdot \|_A$ we denote its restriction to the set $A$, i.e. $\|f\|_A := \sup_{x \in A} |f(x)|$.

**Assumption 3.1** Let $I_X$ be an open interval with $P(X \in I_X) = 1$. Furthermore assume that there exists a function $T_z^{-1} = T^{-1}(z, \cdot) : I_X \to \mathbb{R}$, such that with probability 1 it holds

$$X_i \leq x \Leftrightarrow U_i \leq T^{-1}(Z, x) \text{ for all } x \in I_X.$$  

If Assumption 3.1 is fulfilled, then for all $x \in I_X$ it holds

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq x\}} = \frac{1}{n} \sum_{i=1}^{n} I_{\{T(Z, U_i) \leq x\}} = \frac{1}{n} \sum_{i=1}^{n} I_{\{U_i \leq T^{-1}(Z, x)\}} = G_n(T^{-1}(Z, x)).$$
where

\[ G_n(u) = \frac{1}{n} \sum_{i=1}^{n} I_{\{U_i \leq u\}} \]

is the empirical distribution function of \( U \). In doing so we use the idea that the random variables \( U_i, 1 \leq i \leq n \), are a random sample from the distribution of \( U \). Thus,

\[
0 \leq \| F_n - F_U(T^{-1}(Z, \cdot)) \|_{I_X} = \sup_{x \in I_X} | F_n(x) - F_U(T^{-1}(Z, x)) | \\
= \sup_{x \in I_X} | G_n(T^{-1}(Z, x)) - F_U(T^{-1}(Z, x)) | \\
\leq \sup_{u \in \mathbb{R}} | G_n(u) - F_U(u) | \\
= \| G_n - F_U \|.
\]

If \( I_X \neq \mathbb{R} \), then we can extend the composition \( F_U(T^{-1}(Z, \cdot)) \) as follows: Let

\[ x_{\inf} := \inf I_X \quad \text{and} \quad x_{\sup} := \sup I_X. \]

- If \( x_{\inf} > -\infty \), then define \( F_U(T^{-1}(Z, \cdot)) := 0 \) for all \( x \leq x_{\inf} \).
- If \( x_{\sup} < +\infty \), then define \( F_U(T^{-1}(Z, \cdot)) := 1 \) for all \( x \geq x_{\sup} \).

Note that \( F_n(x) = 0 \) for all \( x \leq x_{\inf} \) and \( F_n(x) = 1 \) for all \( x \geq x_{\sup} \). Thus,

\[ F_n(x) - F_U(T^{-1}(Z, \cdot)) = 0 \quad \text{for all} \quad x \notin I_X \]

and consequently

\[ 0 \leq \| F_n - F_U(T^{-1}(Z, \cdot)) \| = \| F_n - F_U(T^{-1}(Z, \cdot)) \|_{I_X} \leq \| G_n - F_U \|. \quad (3.10) \]

By the theorem of Glivenko-Cantelli (see e.g. [Billingsley 1995, p. 269] or [Shorack 2000, p. 223]), it holds \( \| G_n - F_U \| \xrightarrow{a.s.} 0 \) as \( n \to \infty \). From this and (3.10), it results

\[ \| F_n - F_U(T^{-1}(Z, \cdot)) \| \xrightarrow{a.s.} 0 \quad (n \to \infty). \quad (3.11) \]

Next, we decompose the marked empirical distribution function \( H_n \) into two random
functions $R_n$ and $M_n$ as follows:

\[
H_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} Y_i \\
= \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} (Y_i - m(X_i)) + \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} m(X_i) \\
=: R_n(x) + M_n(x).
\]

The proof that $R_n$ converges uniformly to zero with probability 1, i.e.

\[
\|R_n\| \overset{a.s.}{\to} 0 \quad (n \to \infty),
\]

is completely the same as in Tillich and Ferger [2015, pp. 767-769].

Now we have a look at the second part $M_n$. Because $m$ is a one-step function, see (2.4), it holds

\[
M_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} m(X_i) \\
= \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} \left( \pi_1 \mathbb{1}_{\{X_i \leq \theta\}} + \pi_2 \mathbb{1}_{\{X_i > \theta\}} \right) \\
= \pi_1 \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\} \cap \{X_i \leq \theta\}} + \pi_2 \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\} \cap \{X_i > \theta\}} \\
= \begin{cases} 
\pi_1 F_n(x), & \text{if } x \leq \theta, \\
\pi_1 F_n(\theta) + \pi_2 (F_n(x) - F_n(\theta)), & \text{if } x > \theta,
\end{cases}
\]

This in connection with (3.11) motivates the following definition:

\[
H(x) := \begin{cases} 
\pi_1 F_U(T^{-1}(Z, x)), & \text{if } x \leq \theta, \\
\pi_2 F_U(T^{-1}(Z, x)) + (\pi_1 - \pi_2) F_U(T^{-1}(Z, \theta)), & \text{if } x > \theta.
\end{cases}
\]

If $I_X \neq \mathbb{R}$, the composition $F_U(T^{-1}(Z, \cdot))$ is extended as above. Since

\[
0 \leq \|M_n - H\| = \sup_{x \in \mathbb{R}} |M_n(x) - H(x)|
\]
\[
\begin{align*}
\max \left\{ \sup_{x \in (-\infty, \theta]} |M_n(x) - H(x)|, \sup_{x \in [\theta, \infty[} |M_n(x) - H(x)| \right\} \\
= \max \left\{ \sup_{x \in (-\infty, \theta]} \left| \pi_1 F_n(x) - \pi_1 F_U(T^{-1}(Z, x)) \right|, \right. \\
\left. \sup_{x \in [\theta, \infty[} \left| \pi_2 F_n(x) + (\pi_1 - \pi_2) F_n(\theta) \\
-(\pi_2 F_U(T^{-1}(Z, x)) + (\pi_1 - \pi_2) F_U(T^{-1}(Z, \theta))) \right| \right\} \\
\leq \max \left\{ \pi_1 \sup_{x \in (-\infty, \theta]} \left| F_n(x) - F_U(T^{-1}(Z, x)) \right|, \\
\pi_2 \sup_{x \in [\theta, \infty[} \left| F_n(x) - F_U(T^{-1}(Z, x)) \right| \\
+ |\pi_1 - \pi_2| \left| F_n(\theta) - F_U(T^{-1}(Z, \theta)) \right| \right\} \\
\leq \max \left\{ \pi_1 \sup_{x \in \mathbb{R}} \left| F_n(x) - F_U(T^{-1}(Z, x)) \right|, \\
\pi_2 \sup_{x \in \mathbb{R}} \left| F_n(x) - F_U(T^{-1}(Z, x)) \right| \\
+ (\pi_1 + \pi_2) \sup_{x \in \mathbb{R}} \left| F_n(x) - F_U(T^{-1}(Z, x)) \right| \right\} \\
= (\pi_1 + 2\pi_2) \| F_n - F_U(T^{-1}(Z, \cdot)) \|,
\end{align*}
\]

By (3.11) we get
\[
\|M_n - H\| \xrightarrow{a.s.} 0 \quad (n \to \infty). \tag{3.16}
\]

Due to the decomposition (3.12) we have
\[
0 \leq \|H_n - H\| = \|R_n + M_n - H\| \leq \|R_n\| + \|M_n - H\|.
\]

From this as well as (3.13) and (3.16) it follows
\[
\|H_n - H\| \xrightarrow{a.s.} 0 \quad (n \to \infty). \tag{3.17}
\]
Together with \((3.11)\), this yields the almost sure uniform convergence of \(S^*_n\) to the limit process (extended where necessary)

\[ S^* := \alpha H + \beta F_U(T^{-1}(Z, \cdot)), \tag{3.18} \]

i. e.

\[ \|S^*_n - S^*\| \xrightarrow{a.s.} 0 \quad (n \to \infty), \]

since

\[ 0 \leq \|S^*_n - S^*\| = \|\alpha H_n + \beta F_n - (\alpha H + \beta F_U(T^{-1}(Z, \cdot)))\| \]

\[ \leq |\alpha|\|H_n - H\| + |\beta|\|F_n - F_U(T^{-1}(Z, \cdot))\|. \]

Thus, we have proven \((P1)\).

To show \((P2)\) and \((P3)\), we have a more detailed look at the limit process \(S^*\). It holds

\[ S^*(x) = \alpha H(x) + \beta F_U(T^{-1}(Z, x)) \]

\[ = \begin{cases} 
(\alpha \pi_1 + \beta) F_U(T^{-1}(Z, x)), & \text{if } x \leq \theta, \\
(\alpha \pi_2 + \beta) F_U(T^{-1}(Z, x)) + \alpha (\pi_1 - \pi_2) F_U(T^{-1}(Z, \theta)), & \text{if } x > \theta.
\end{cases} \tag{3.19} \]

By demanding that the composition \(F_U(T^{-1}(Z, \cdot))\) is rcll, i. e. by demanding that for all \(x_0 \in \mathbb{R}\) there exist

\[ F_U(T^{-1}(Z, x_0-)) = \lim\limits_{x \downarrow x_0} F_U(T^{-1}(Z, x)) \quad \text{and} \]

\[ F_U(T^{-1}(Z, x_0+)) = \lim\limits_{x \uparrow x_0} F_U(T^{-1}(Z, x)) = F_U(T^{-1}(Z, x_0)), \]

\((3.19)\) guarantees that the left limits of \(S^*\) exist and that \(S^*\) is continuous from the right on the open intervals \([-\infty, \theta[ \text{ und } ]\theta, \infty[\). Now we only have to check continuity from the right in point \(\theta:\)

\[ \lim_{x \downarrow \theta} S^*(x) = \lim_{x \downarrow \theta} ((\alpha \pi_2 + \beta) F_U(T^{-1}(Z, x)) + \alpha (\pi_1 - \pi_2) F_U(T^{-1}(Z, \theta))) \]

\[ = (\alpha \pi_2 + \beta) \lim_{x \downarrow \theta} F_U(T^{-1}(Z, x)) + (\alpha \pi_1 - \alpha \pi_2) F_U(T^{-1}(Z, \theta)) \]

\[ = (\alpha \pi_1 + \beta) F_U(T^{-1}(Z, \theta)) \]
\[ S^* = S^*(\theta). \]

Hence, \( S^* \) is rcll and (P2) is proven.

In order to show that \( \theta \) is a well-separated maximum point of \( S^* \) (P3), note that

\[ \alpha \pi_2 + \beta < 0 < \alpha \pi_1 + \beta \]

for all \( \pi_1 \neq \pi_2 \in [0, 1] \), see Ferger and Klotsche [2009, p. 99, (2.7)]. Consequently, for all \( x < \theta \) we have

\[
S^*(\theta) - S^*(x) = (\alpha \pi_1 + \beta)F_U(T^{-1}(Z, \theta)) - (\alpha \pi_1 + \beta)F_U(T^{-1}(Z, x))
\]

\[
\geq 0, \text{ if } F_U(T^{-1}(Z, \cdot)) \text{ was increasing}
\]

\[
> 0, \text{ if } F_U(T^{-1}(Z, \cdot)) \text{ was strictly increasing}
\]

and for all \( x > \theta \) we have

\[
S^*(\theta) - S^*(x) = (\alpha \pi_1 + \beta)F_U(T^{-1}(Z, \theta))
\]

\[
- [(\alpha \pi_2 + \beta)F_U(T^{-1}(Z, x)) + \alpha(\pi_1 - \pi_2)F_U(T^{-1}(Z, \theta))] \\
= (\alpha \pi_1 + \beta - \alpha \pi_1 + \alpha \pi_2)F_U(T^{-1}(Z, \theta)) - (\alpha \pi_2 + \beta)F_U(T^{-1}(Z, x))
\]

\[
= (\alpha \pi_2 + \beta)(F_U(T^{-1}(Z, \theta)) - F_U(T^{-1}(Z, x))).
\]

Putting things together, we get

\[ S^*(\theta) - S^*(x) \geq 0 \quad \text{for all } x \in \mathbb{R}, \]

i. e. the unknown parameter \( \theta \) is a maximum point of \( S^* \), if the composite function \( F_U(T^{-1}(Z, \cdot)) \) is increasing. To ensure that \( \theta \) is a well-separated maximizer of \( S^* \), i. e.

\[ S^*(\theta) > \sup \{ S^*(x) : |x - \theta| > \varepsilon \} \quad \text{for all } \varepsilon > 0 \]

we could demand that the composite function \( F_U(T^{-1}(Z, \cdot)) \) is strictly increasing. But it would also suffice to demand that the composition is increasing and that for all possible realizations \( z \) of \( Z \), there does not exist an interval containing \( \theta \) on which \( F_U(T^{-1}(z, \cdot)) \) is constant. Compare Ferger and Klotsche [2009, p. 99, Lemma 2.1 (1)] and Tillich and...
In this way, (P3) is proven. This enables us to recap the findings in the following

**Theorem 1** Let the model assumptions (2.2), (2.3) and (2.4) hold. If there exists an open interval $I_X$ with $P(X \in I_X) = 1$ and a function $T_z^{-1} = T_z^{-1}(z, \cdot) : I_X \to \mathbb{R}$, which for all $u \in \mathbb{R}$, $x \in I_X$ and all possible realizations $z$ of $Z$ fulfills the following three properties:

(i) $X_i \leq x \Leftrightarrow U_i \leq T_z^{-1}(Z, x)$ for all $1 \leq i \leq n$,
(ii) the composition $F_U \circ T_z^{-1}$ is right-continuous (cadlag) and increasing, and
(iii) there does not exist an interval containing $\theta$ on which $F_U \circ T_z^{-1}$ is constant,

then the ML estimator $\theta_n^*$ is strongly consistent for $\theta$, i.e.

$$\theta_n^* \xrightarrow{a.s.} \theta \text{ as } n \to \infty.$$

**Remark 3.2 (On the assumptions of Theorem 1)**

- The properties (i)-(iii) are quite complex. In Section 7 one can find sufficient conditions for them and some examples.
- Properties (i) and (iii) imply that the composition is strictly increasing in $\theta$ and that

$$0 < P(X_i \leq \theta) < 1.$$

Thus it is ensured that with probability 1 there are observations $X_i$ to the left and to the right of $\theta$ for eventually all $n \in \mathbb{N}$.

## 4 Dempfle-Stute estimators for the breakpoint

If the risk levels $\pi_1$ and $\pi_2$ are unknown, the following estimators for the breakpoint $\theta$ have been proposed:

If $\pi_1 > \pi_2$:

$$\hat{\theta}_n := \arg\max_{x \in \mathbb{R}} S_n(x),$$

If $\pi_1 < \pi_2$:

$$\tilde{\theta}_n := \arg\min_{x \in \mathbb{R}} S_n(x) = \arg\max_{x \in \mathbb{R}} -S_n(x).$$
if $\pi_1 \neq \pi_2$:

$$\hat{\theta}_n := \operatorname{argmax}_{x \in \mathbb{R}} |S_n(x)|. \quad (4.3)$$

The corresponding criterion function is

$$S_n(x) := H_n(x) - \bar{Y}_n F_n(x), \quad (4.4)$$

where

- $F_n(x)$ is the empirical distribution function from (3.7),
- $H_n(x)$ is the marked empirical distribution function from (3.8), and
- $\bar{Y}_n$ is the overall success (in our context: default) rate from (3.9).

There are different alternative representations of the criterion function, compare Tillich and Ferger [2015, pp. 764-765] and Tillich [2013, Section 3.2.2]. All of them result in the same estimators.

All of the estimators (4.1)-(4.3) are subsumed under the term Dempe-Stute estimators because of the first occurrence of the one-sided versions (4.1) and (4.2) in Dempe and Stute [2002]. The two-sided version (4.3) can be found in Ferger and Klotsche [2009]. The one- and two-sided estimators are connected as follows (Tillich [2013, p. 53] or Tillich and Ferger [2015]):

$$\bar{\theta}_n = \begin{cases} 
\hat{\theta}_n, & \text{if } \max_{x \in \mathbb{R}} S_n(x) > \max_{x \in \mathbb{R}} -S_n(x), \\
\check{\theta}_n, & \text{if } \max_{x \in \mathbb{R}} S_n(x) < \max_{x \in \mathbb{R}} -S_n(x), \\
\min\{\hat{\theta}_n, \check{\theta}_n\}, & \text{if } \max_{x \in \mathbb{R}} S_n(x) = \max_{x \in \mathbb{R}} -S_n(x).
\end{cases}$$

**Remark 4.1 (Connections to credit risk)**

Note that there are some interesting connections between the Dempe-Stute estimators, the Kolmogorov-Smirnov statistic, and the curve of the Receiver Operating Characteristic (ROC). The latter two can be used to measure the quality of a credit scoring system. Compare Tillich and Ferger [2015, Remark 2] or Tillich [2013, pp. 45-46] for details.

For the properties of $S_n$, we repeat the corresponding statements of Tillich and Ferger [2015] one to one. Note that there is a close analogy to the properties of the empirical process $S^*_n$ of Section 3. Again, $X_{1:m} \leq X_{2:m} \leq \ldots \leq X_{n:m}$ denote the order statistics of the sample $X_1, X_2, \ldots, X_n$. D. Tillich, C. Lehmann
• The empirical process $S_n$ is a step function which jumps exactly at the observations $X_i$. It is right continuous with left limits (rcll).

• The process $S_n$ vanishes outside the range of the observations, i.e. $S_n(x) = 0$, if $x < X_{1:n}$ or $x \geq X_{n:n}$. Within the range of the observations, we expect $S_n$ to have a triangular form. Because of (2.4) we expect an increase followed by a decrease, if $\pi_1 > \pi_2$, and the other way around, if $\pi_1 < \pi_2$.

• The random function $S_n$ does not have a unique maximum (minimum) point, but a smallest maximum (minimum) point exists as long as the path of $S_n$ runs above (below) the abscissa at least for one interval $[X_i:n, X_{i+1:n}]$, $1 \leq i \leq n - 1$.

In this case, the one-sided estimator $\hat{\theta}_n$ ($\tilde{\theta}_n$) is taken to be that smallest maximum (minimum) point if $\pi_1 > \pi_2$ ($\pi_1 < \pi_2$). Otherwise we use the smallest observation, i.e. $\hat{\theta}_n = X_{1:n}$ or $\tilde{\theta}_n = X_{1:n}$. The formation of the two-sided version is carried out completely analogously to that of $\hat{\theta}_n$. Here, a smallest maximum point of $|S_n|$ does not exist if and only if $S_n(x) = 0$ for all $x \in \mathbb{R}$.

As in Section 3, our goal is to prove strong consistency of the breakpoint estimators. And again, we make use of the Argmax-Theorem of Ferger [2009, Theorem 4.6] (also in Ferger and Klotsche [2009, Theorem A.1]). We obtain the following result.

**Theorem 2** Let the model assumptions (2.2), (2.3) and (2.4) hold. If there exists an open interval $I_X$ with $P(X \in I_X) = 1$ and a function $T_{z^{-1}}^{-1} = T_{z^{-1}}^{-1}(z, \cdot) : I_X \to \mathbb{R}$, which for all $u \in \mathbb{R}$, $x \in I_X$ and all possible realizations $z$ of $Z$ fulfills the following three properties:

(i) $X_i \leq x \Leftrightarrow U_i \leq T_{z^{-1}}^{-1}(Z, x)$ for all $1 \leq i \leq n$,

(ii) the composition $F_U \circ T_{z^{-1}}^{-1}$ is rcll (càdlàg) and increasing, and

(iii) there does not exist an interval containing $\theta$ on which $F_U \circ T_{z^{-1}}^{-1}$ is constant,

then the Dempe-Stute estimators are strongly consistent for $\theta$, i.e.

$$\hat{\theta}_n \xrightarrow{a.s.} \theta \text{ as } n \to \infty, \text{ if } \pi_1 > \pi_2,$$

$$\tilde{\theta}_n \xrightarrow{a.s.} \theta \text{ as } n \to \infty, \text{ if } \pi_1 < \pi_2,$$

$$\bar{\theta}_n \xrightarrow{a.s.} \theta \text{ as } n \to \infty.$$
Proof of Theorem 2: We first deal with the case \( \pi_1 > \pi_2 \) and the one-sided estimator \( \hat{\theta}_n \). The two other cases are treated analogously. See the end of the proof for details.

We have already seen that the conditions of the Argmax-Theorem regarding \( S_n \) are met: The criterion function \( S_n \) is rcll and the estimator \( \hat{\theta}_n \) belongs to the criterion function’s maximum points. Thus, we additionally have to prove the following properties:

\((P1')\) \( S_n \) converges uniformly to its limit process \( S \) almost surely (a. s.).

\((P2')\) The limit process \( S \) of \( S_n \) is rcll (càdlàg).

\((P3')\) The breakpoint \( \theta \) is a. s. the well-separated maximum point of \( S \), i. e.

\[ S(\theta) > \sup \{ S(x) : |x - \theta| > \epsilon \} \text{ for all } \epsilon > 0. \]

If these properties are fulfilled, then the convergence of the processes is transferred to their maximum points.

Based on (4.4) and motivated by (3.9), (3.11), and (3.17), define (where necessary extended as in Section 3)

\[ S(x) := H(x) - \tilde{H}F_U(T^{-1}(Z,x)), \quad x \in \mathbb{R}, \]

with \( H \) from (3.15) and

\[ \tilde{H} := \lim_{x \to \infty} H(x) = \pi_2 + (\pi_1 - \pi_2)F_U(T^{-1}(Z,\theta)). \]

In order to prove \((P1')\), note that

\[
0 \leq \| S_n - S \| \\
= \| H_n - \tilde{Y}_n F_n - (H - \tilde{H}F_U(T^{-1}(Z, \cdot))) \| \\
= \| H_n - H - \tilde{Y}_n F_n + \tilde{Y}_n F_U(T^{-1}(Z, \cdot)) - \tilde{Y}_n F_U(T^{-1}(Z, \cdot)) + \tilde{H}F_U(T^{-1}(Z, \cdot)) \| \\
\leq \| H_n - H \| + |\tilde{Y}_n| \cdot \| F_n - F_U(T^{-1}(Z, \cdot)) \| + |\tilde{Y}_n - \tilde{H}| \cdot \| F_U(T^{-1}(Z, \cdot)) \| \\
\leq \| H_n - H \| + \| F_n - F_U(T^{-1}(Z, \cdot)) \| + |\tilde{Y}_n - \tilde{H}|. \tag{4.7}
\]

The last relation holds because \( 0 \leq \tilde{Y}_n \leq 1 \), since \( Y_i \in \{0,1\} \), and because \( F_U \) is a cumulative distribution function, which has a supremum norm equal to one.

From Section 3, we know that the first two summands of (4.7) converge to zero almost surely, if the conditions of Theorem 2 are met. Thus, we only have to show that \( |\tilde{Y}_n - \tilde{H}| \)
converges to zero almost surely as \(n\) tends to infinity. To that, decompose \(\hat{Y}_n\) as follows:

\[
\hat{Y}_n = R^*_n + M^*_n
\]

with

\[
R^*_n := \frac{1}{n} \sum_{i=1}^{n} (Y_i - m(X_i)) \quad \text{and} \quad M^*_n := \frac{1}{n} \sum_{i=1}^{n} m(X_i).
\]

The first part \(R^*_n\) converges to zero almost surely as \(n \to \infty\), i.e.

\[
|R^*_n| \xrightarrow{a.s.} 0 \quad \text{as} \quad n \to \infty.
\] (4.8)

As in the proof of (3.13), the reasoning is exactly the same as in [Tillich and Ferger, 2015, p. 770]. Compare also [Hähle, 2014, pp. 44-45] for some more details. The second part \(M^*_n\) is the limit of \(M_n\) from (3.12) for \(x \to \infty\). With (3.14) we get

\[
M^*_n = \lim_{x \to \infty} M_n(x) = \lim_{x \to \infty} \pi_2 F_n(x) + (\pi_1 - \pi_2) F_n(\theta) = \pi_2 + (\pi_1 - \pi_2) F_n(\theta).
\]

Hence, from

\[
0 \leq |\hat{Y}_n - \bar{H}| \leq |R^*_n| + |M^*_n - \bar{H}|
\]

\[
= |R^*_n| + |\pi_1 - \pi_2| \cdot |F_n(\theta) - F_U(T^{-1}(Z, \theta))|
\]

\[
\leq |R^*_n| + |\pi_1 - \pi_2| \cdot \|F_n - F_U(T^{-1}(Z, \cdot))\|,
\]

and (4.8) and (3.11), we conclude

\[
|\hat{Y}_n - \bar{H}| \xrightarrow{a.s.} 0 \quad \text{as} \quad n \to \infty.
\] (4.9)

This in connection with (3.11), (3.17), and (4.7) shows (P1'), i.e.

\[
\|S_n - S\| \xrightarrow{a.s.} 0 \quad \text{as} \quad n \to \infty.
\]

Now, we have a closer look at the properties of the limit process \(S\). From (4.5) in connection with (3.15) and (4.6), it results

\[
S(x) = \begin{cases} 
(\pi_1 - \pi_2)(1 - F_U(T^{-1}(Z, \theta))) F_U(T^{-1}(Z, x)) & \text{if } x \leq \theta, \\
(\pi_1 - \pi_2) F_U(T^{-1}(Z, \theta))(1 - F_U(T^{-1}(Z, x))) & \text{if } x > \theta.
\end{cases}
\]
Note that in both cases the first two factors are positive and do not depend on \( x \). Thus, Condition (i) of Theorem 2, i.e. the composition \( F_U \circ T_z^{-1} \) is rcll and increasing, yields that the limit process \( S \) is rcll (F2), especially
\[
\lim_{x \downarrow \theta} S(x) = S(\theta),
\]
and that \( S \) is increasing on \([-\infty, \theta]\) and decreasing on \([\theta, \infty[\). Condition (iii) of Theorem 2 ensures that the breakpoint \( \theta \) is the well-separated maximum point of \( S \). Thus, also (F3) is fulfilled and the Argmax-Theorem yields strong consistency of the one-sided Dempe-Stute estimator \( \hat{\theta}_n \).

If \( \pi_1 < \pi_2 \), then all the arguments above remain valid for \(-S_n \) instead of \( S_n \) and \(-S\) instead of \( S \). If only \( \pi_1 \neq \pi_2 \) is known, replace \( S_n \) by \(|S_n|\) and \( S \) by \(|S|\) and note that \( ||S_n| - |S|| \leq ||S_n - S|| \). This completes the proof. \( \Box \)

## 5 Estimators for the risk levels

As Ferger and Klotsche [2009] and Tillich and Ferger [2015], we use relative success (in our context: default) frequencies for estimating the risk levels \( \pi_1 \) and \( \pi_2 \). We start with the random functions
\[
\pi_{1,n}(x) := \frac{\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} Y_i}{\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}} \quad x \geq X_{1:n},
\]
and
\[
\pi_{2,n}(x) := \frac{\sum_{i=1}^{n} \mathbb{1}_{\{X_i > x\}} Y_i}{\sum_{i=1}^{n} \mathbb{1}_{\{X_i > x\}}} \quad x < X_{n:n}.
\]
Again, \( X_{1:n} \) and \( X_{n:n} \) denote the minimum and the maximum of the sample \( X_1, \ldots, X_n \). Note that the denominator is the number of observations whose scores \( X_i \) are below/above the threshold \( x \). In the numerator, we find the number of successes (defaults) in the respective two groups. If there are no observations in one of the two groups separated by the threshold \( x \), then set the corresponding value equal to zero. By expanding the fractions by \( \frac{1}{n} \), using \( \mathbb{1}_{\{X_i > x\}} = 1 - \mathbb{1}_{\{X_i \leq x\}} \) in the second equation, and employing the definitions (3.7), (3.8) and (3.9), we get
\[
\pi_{1,n}(x) = \frac{H_n(x)}{F_n(x)} \quad x \geq X_{1:n}, \quad \text{and} \quad \pi_{2,n}(x) = \frac{Y_n - H_n(x)}{1 - F_n(x)} \quad x < X_{n:n}.
\]  

(5.1)
If Assumption 3.1, p. 8, is fulfilled, then (3.11) and (3.17) yield
\[\pi_{1,n}(x) \xrightarrow{a.s.} \tilde{\pi}_1(x) := \frac{H(x)}{F_U(T^{-1}(Z,x))} \quad \text{as } n \to \infty \] (5.2)
and together with (4.9) also
\[\pi_{2,n}(x) \xrightarrow{a.s.} \tilde{\pi}_2(x) := \frac{\tilde{H} - H(x)}{1 - F_U(T^{-1}(Z,x))} \quad \text{as } n \to \infty \] (5.3)
for all \(x\) with \(P(0 < F_U(T^{-1}(Z,x)) < 1) = 1\) by application of the continuous mapping theorem (CMT), see e. g. Serfling [1980, p. 24] or Davidson [1994, p. 286, Thm. 18.8].

The objects \(\tilde{\pi}_1\) and \(\tilde{\pi}_2\) are random functions, since they depend on the random variable \(Z\).

By (3.15) and (4.6) we obtain
\[\tilde{\pi}_1(x) = \begin{cases} \pi_1, & \text{if } x \leq \theta, \\ \pi_1 F_U(T^{-1}(Z,\theta)) + \pi_2 (F_U(T^{-1}(Z,x)) - F_U(T^{-1}(Z,\theta))) F_U(T^{-1}(Z,x)) , & \text{if } x > \theta, \end{cases} \] (5.4)
and
\[\tilde{\pi}_2(x) = \begin{cases} \pi_1 (F_U(T^{-1}(Z,\theta)) - F_U(T^{-1}(Z,x))) + \pi_2 (1 - F_U(T^{-1}(Z,\theta))) , & \text{if } x < \theta, \\ \pi_2 , & \text{if } x \geq \theta. \end{cases} \] (5.5)

In (5.5), the case \(x = \theta\) originally belongs to the upper part of the case analysis. Since \(\tilde{\pi}_2(\theta) = \pi_2\), it is assigned below. This facilitates the subsequent work. Furthermore it should be noted that we are not interested in estimating the whole functions \(\tilde{\pi}_1\) and \(\tilde{\pi}_2\), but only the risk levels \(\pi_1\) and \(\pi_2\). So we merely have to choose a reasonable argument for \(\pi_{1,n}\) and \(\pi_{2,n}\).

Assume for a moment that the breakpoint \(\theta\) is known. Then
- calculating and zeroizing the first derivatives of the loglikelihood function (3.1) with respect to \(\pi_1\) and \(\pi_2\) and
- checking the Hessian matrix of second derivatives for negative-definiteness
shows that the relative frequencies
\[\pi^*_1(n) := \pi_{1,n}(\theta) = \frac{H_n(\theta)}{F_n(\theta)} \quad \text{and} \quad \pi^*_2(n) := \pi_{2,n}(\theta) = \frac{Y_n - H_n(\theta)}{1 - F_n(\theta)}\]
are the maximum likelihood (ML) estimators for the risk levels $\pi_1$ and $\pi_2$ given $0 < F_n(\theta) < 1$, which by (3.11) is fulfilled almost surely for eventually all $n \in \mathbb{N}$ if $P(0 < F_U(T^{-1}(Z, \theta)) < 1) = 1$. Under this condition, we also obtain the strong consistency of the ML estimators $\hat{\pi}^*_1,n$ and $\hat{\pi}^*_2,n$ for the parameters $\pi_1$ and $\pi_2$ from (5.2) and (5.4) or (5.3) and (5.5), respectively.

Now reconsider the case where the breakpoint $\theta$ is unknown. From the above, one can derive the idea of substituting the unknown parameter by an appropriate estimator. Using the Dempe-Stute estimators for $\theta$ from Section 4, we obtain

$$\hat{\pi}_1,n := \pi_1,n(\hat{\theta}_n) \quad \text{and} \quad \hat{\pi}_2,n := \pi_2,n(\hat{\theta}_n), \quad \text{if } \pi_1 > \pi_2,$$

$$\bar{\pi}_1,n := \pi_1,n(\bar{\theta}_n) \quad \text{and} \quad \bar{\pi}_2,n := \pi_2,n(\bar{\theta}_n), \quad \text{if } \pi_1 < \pi_2,$$

(5.6)

$$\check{\pi}_1,n := \pi_1,n(\check{\theta}_n) \quad \text{and} \quad \check{\pi}_2,n := \pi_2,n(\check{\theta}_n), \quad \text{if } \pi_1 \neq \pi_2.$$

**Theorem 3** Let the model assumptions (2.2), (2.3) and (2.4) hold. If there exists an open interval $I_X$ with $P(X \in I_X) = 1$ and a function $T^{-1}(z, \cdot) : I_X \rightarrow \mathbb{R}$, which for all $u \in \mathbb{R}$, $x \in I_X$ and all possible realizations $z$ of $Z$ fulfills the following four properties:

(i) $X_i \leq x \Leftrightarrow U_i \leq T^{-1}(Z, x)$ for all $1 \leq i \leq n$,

(ii) the composition $F_U \circ T^{-1}$ is right (càdlàg) and increasing,

(iii) there does not exist an interval containing $\theta$ on which $F_U \circ T^{-1}$ is constant, and

(iv) the composition $F_U \circ T^{-1}$ is continuous at $\theta$,

then the risk level estimators from (5.6) are strongly consistent, i.e. as $n \rightarrow \infty$ it holds

$$\hat{\pi}_{1,n} \xrightarrow{a.s.} \pi_1 \quad \text{and} \quad \hat{\pi}_{2,n} \xrightarrow{a.s.} \pi_2, \quad \text{if } \pi_1 > \pi_2,$$

$$\bar{\pi}_{1,n} \xrightarrow{a.s.} \pi_1 \quad \text{and} \quad \bar{\pi}_{2,n} \xrightarrow{a.s.} \pi_2, \quad \text{if } \pi_1 < \pi_2,$$

$$\check{\pi}_{1,n} \xrightarrow{a.s.} \pi_1 \quad \text{and} \quad \check{\pi}_{2,n} \xrightarrow{a.s.} \pi_2, \quad \text{if } \pi_1 \neq \pi_2.$$

**Remark 5.1 (On the assumptions of Theorem 3)**

a) Note the difference in the assumptions of the Theorems 2 and 3 versus Theorem 3.
In (iv), we additionally demand $F_U \circ T^{-1}$ to be continuous at $\theta$. This implies that the functions $\tilde{\pi}_1$ and $\tilde{\pi}_2$ are continuous in $\theta$. See (5.4) and (5.5).
b) Condition (iii) implies $P(0 < F_U(T^{-1}(Z, \theta)) < 1) = 1$, which itself is a prerequisite for (5.2) and (5.3).

**Proof of Theorem 3.** We only consider the case $\pi_1 > \pi_2$. The other cases are treated analogously. Start with (5.6) and (5.4) to get

$$0 \leq |\hat{\pi}_{1,n} - \pi_1| = |\pi_{1,n}(\hat{\theta}_n) - \hat{\pi}_1(\theta)| = |\pi_{1,n}(\hat{\theta}_n) - \hat{\pi}_1(\hat{\theta}_n) + \hat{\pi}_1(\hat{\theta}_n) - \hat{\pi}_1(\theta)| \leq |\pi_{1,n}(\hat{\theta}_n) - \hat{\pi}_1(\hat{\theta}_n)| + |\hat{\pi}_1(\hat{\theta}_n) - \hat{\pi}_1(\theta)| =: D_{1,1} + D_{1,2}. \quad (5.7)$$

Analogously, (5.6) and (5.5) yield

$$0 \leq |\hat{\pi}_{2,n} - \pi_2| \leq |\pi_{2,n}(\hat{\theta}_n) - \hat{\pi}_2(\hat{\theta}_n)| + |\hat{\pi}_2(\hat{\theta}_n) - \hat{\pi}_2(\theta)| =: D_{2,1} + D_{2,2}. \quad (5.8)$$

As stated in Remark 5.1a), the functions $\hat{\pi}_1$ and $\hat{\pi}_2$ are continuous in $\theta$. Moreover, the conditions of Theorem 3 also ensure that the Dempe-Stute estimators from (4.1)-(4.3) are strongly consistent (see Theorem 2). Thus,

$$D_{1,2} \xrightarrow{a.s.} 0 \quad \text{and} \quad D_{2,2} \xrightarrow{a.s.} 0 \quad \text{as} \quad n \to \infty \quad (5.9)$$

follows from the Continuous Mapping Theorem (CMT, see e.g. Serfling [1980, p. 24] or Davidson [1994, p. 286, Thm. 18.10]).

In order to show that $D_{1,1}$ and $D_{2,1}$ also converge to zero almost surely, we study the behaviour of $\pi_{1,n}$ and $\pi_{2,n}$ from (5.1) in a small neighborhood of $\theta$. From (iv) and (iii), i.e. because the composition $F_U \circ T_z^{-1}$ is continuous and strictly increasing in $\theta$, it follows that there is an $r > 0$ such that $P(0 < F_U(T^{-1}(Z, \theta - r)) < F_U(T^{-1}(Z, \theta + r)) < 1) = 1$. Hence $0 < F_n(\theta - r) \leq F_n(\theta + r) < 1$ for eventually all $n$ almost surely because of (3.11). Define

$$B_r(\theta) := [\theta - r, \theta + r[$$

and note that for all $x \in B_r(\theta)$ it holds

$$F_U(T^{-1}(Z, \theta - r)) \leq F_U(T^{-1}(Z, x)) \leq F_U(T^{-1}(Z, \theta + r)) \quad (5.10)$$
and

\[ F_n(\theta - r) \leq F_n(x) \leq F_n(\theta + r). \] (5.11)

Furthermore, recall that \( \hat{\theta}_n \in B_r(\theta) \) a.s. for eventually all \( n \in \mathbb{N} \) according to Theorem 2. Therefore

\[
0 \leq D_{1,1} = \left| \pi_{1,n}(\hat{\theta}_n) - \tilde{\pi}_1(\hat{\theta}_n) \right| \\
\leq \sup_{x \in B_r(\theta)} \left| \pi_{1,n}(x) - \tilde{\pi}_1(x) \right| \\
= \sup_{x \in B_r(\theta)} \left| \frac{H_n(x)}{F_n(x)} - \frac{H(x)}{F_u(T^{-1}(Z,x))} \right| \\
= \sup_{x \in B_r(\theta)} \left| \frac{H_n(x)}{F_n(x)} - \frac{H(x)}{F_n(x)} + \frac{H(x)}{F_n(x)} - \frac{H(x)}{F_u(T^{-1}(Z,x))} \right| \\
= \sup_{x \in B_r(\theta)} \frac{1}{F_n(x)} \left( H_n(x) - H(x) \right) + H(x) \left( \frac{1}{F_n(x)} - \frac{1}{F_u(T^{-1}(Z,x))} \right) \\
\leq \sup_{x \in B_r(\theta)} \frac{|H_n(x) - H(x)|}{F_n(\theta - r)} + \sup_{x \in B_r(\theta)} H(x) \frac{|F_u(T^{-1}(Z,x)) - F_n(x)|}{F_n(x)F_u(T^{-1}(Z,x))} \\
\leq \frac{\|H_n - H\|}{F_n(\theta - r)} + \frac{\|F_n - F_u(T^{-1}(Z, \cdot))\|}{F_n(\theta - r)F_u(T^{-1}(Z, \theta - r))}. \] (5.12)

The penultimate inequality results from the monotonicity statements in (5.10) and (5.11). The numerators in (5.12) converge a.s. to zero due to (3.17) and (3.11). Equation (3.11) also implies a.s. pointwise convergence of the term \( F_n(\theta - r) \) in the denominators. Its limit \( F_u(T^{-1}(Z, \theta - r)) \) is a.s. greater than zero by construction of \( B_r(\theta) \). Thus, the CMT yields

\[ D_{1,1} \xrightarrow{a.s.} 0 \quad \text{as} \quad n \to \infty. \] (5.13)

In the same manner one shows

\[
0 \leq D_{2,1} = \left| \pi_{2,n}(\hat{\theta}_n) - \tilde{\pi}_2(\hat{\theta}_n) \right| \\
\leq \sup_{x \in B_r(\theta)} \left| \pi_{2,n}(x) - \tilde{\pi}_2(x) \right| \\
\leq \frac{\|H_n - H\|}{F_n(\theta - r)} + \frac{\|F_n - F_u(T^{-1}(Z, \cdot))\|}{F_n(\theta - r)F_u(T^{-1}(Z, \theta - r))}. \]
\[
\begin{align*}
    &= \sup_{x \in B_r(\theta)} \left| \frac{\tilde{Y}_n - H_n(x)}{1 - F_n(x)} - \frac{\bar{H} - H(x)}{1 - F_U(T^{-1}(Z, x))} \right| \\
    &= \sup_{x \in B_r(\theta)} \left| \frac{\tilde{Y}_n - H_n(x)}{1 - F_n(x)} - \frac{\tilde{H} - H(x)}{1 - F_n(x)} + \frac{\bar{H} - H(x)}{1 - F_n(x)} - \frac{\bar{H} - H(x)}{1 - F_U(T^{-1}(Z, x))} \right| \\
    &= \sup_{x \in B_r(\theta)} \left| \frac{1}{1 - F_n(x)} \left( \tilde{Y}_n - \bar{H} - (H_n(x) - H(x)) \right) \\
    &\quad + (\bar{H} - H(x)) \left( \frac{1}{1 - F_n(x)} - \frac{1}{1 - F_U(T^{-1}(Z, x))} \right) \right| \\
    &\leq \sup_{x \in B_r(\theta)} \left| \frac{1}{1 - F_n(x)} \left( \tilde{Y}_n - \bar{H} - (H_n(x) - H(x)) \right) \right| \\
    &\quad + \sup_{x \in B_r(\theta)} \left| H_n(x) - H(x) \right| \\
    &\quad + \sup_{x \in B_r(\theta)} \left| \bar{H} - H(x) \right| \left| \frac{1 - F_U(T^{-1}(Z, x)) - (1 - F_n(x))}{(1 - F_n(x))(1 - F_U(T^{-1}(Z, x)))} \right| \\
    &\leq \frac{|\tilde{Y}_n - \bar{H}|}{1 - F_n(\theta + r)} + \sup_{x \in B_r(\theta)} \frac{|H_n(x) - H(x)|}{1 - F_n(\theta + r)} \\
    &\quad + \sup_{x \in B_r(\theta)} \frac{1}{(1 - F_n(\theta + r))(1 - F_U(T^{-1}(Z, \theta + r)))} \frac{|F_n(x) - F_U(T^{-1}(Z, x))|}{1 - F_n(\theta + r)} \\
    &\leq \frac{|\tilde{Y}_n - \bar{H}|}{1 - F_n(\theta + r)} + \frac{\|H_n - H\|}{1 - F_n(\theta + r)} + \frac{\|F_n - F_U(T^{-1}(Z, \cdot))\|}{(1 - F_n(\theta + r))(1 - F_U(T^{-1}(Z, \theta + r)))}
\end{align*}
\]

and consequently
\[
D_{2,1} \xrightarrow{a.s.} 0 \quad \text{as} \quad n \to \infty \tag{5.14}
\]

by (4.9), (3.17), and (3.11). Connecting all the intermediate results (5.7)–(5.14), it follows
\[
|\hat{\pi}_{1,n} - \pi_1| \xrightarrow{a.s.} 0 \quad \text{and} \quad |\hat{\pi}_{2,n} - \pi_2| \xrightarrow{a.s.} 0 \quad \text{as} \quad n \to \infty,
\]
which finishes the proof. \qed
6 Plug-in estimators for the breakpoint

Recall the definition of the maximum likelihood (ML) estimator from Section 3, see (3.3)–(3.5). It involves the knowledge of the risk levels $\pi_1$ and $\pi_2$. Thus, the ML estimator should be written more precisely as

$$\theta^*_n = \theta^*_n(\pi_1, \pi_2)$$

and the criterion function from (3.4) should be written more precisely as

$$S^*_n(x) = S^*_n(x; \pi_1, \pi_2).$$

If the risk levels $\pi_1$ and $\pi_2$ are unknown, we can replace them by some estimators from Section 5. Take for instance the case $\pi_1 > \pi_2$. Here we could define

$$\hat{\theta}^*_n := \theta^*_n(\hat{\pi}_{1,n}, \hat{\pi}_{2,n}) = \theta^*_n(\hat{\theta}_n, \pi_{2,n}(\hat{\theta}_n)),$$

i. e. we insert the estimators $\hat{\pi}_{1,n}$ and $\hat{\pi}_{2,n}$, which themselves are based on the Dempe-Stute estimator $\hat{\theta}_n$, into the ML estimator for the breakpoint. Thus, it results a two-step plug-in estimator for the breakpoint.

In more detail, let

$$\hat{\alpha}_n := \ln \frac{\hat{\pi}_{1,n}(1 - \hat{\pi}_{2,n})}{\hat{\pi}_{2,n}(1 - \hat{\pi}_{1,n})}, \quad \tilde{\alpha}_n := \ln \frac{\tilde{\pi}_{1,n}(1 - \tilde{\pi}_{2,n})}{\tilde{\pi}_{2,n}(1 - \tilde{\pi}_{1,n})}, \quad \bar{\alpha}_n := \ln \frac{\bar{\pi}_{1,n}(1 - \bar{\pi}_{2,n})}{\bar{\pi}_{2,n}(1 - \bar{\pi}_{1,n})},$$

$$\hat{\beta}_n := \ln \frac{1 - \hat{\pi}_{1,n}}{1 - \hat{\pi}_{2,n}}, \quad \tilde{\beta}_n := \ln \frac{1 - \tilde{\pi}_{1,n}}{1 - \tilde{\pi}_{2,n}}, \quad \bar{\beta}_n := \ln \frac{1 - \bar{\pi}_{1,n}}{1 - \bar{\pi}_{2,n}},$$

$$\hat{S}^*_n := \hat{\alpha}_n H_n + \hat{\beta}_n F_n, \quad \tilde{S}^*_n := \tilde{\alpha}_n H_n + \tilde{\beta}_n F_n, \quad \bar{S}^*_n := \bar{\alpha}_n H_n + \bar{\beta}_n F_n, \quad (6.1)$$

with risk level estimators from (5.6), $H_n$ from (3.8) and $F_n$ from (3.7). Define the breakpoint estimators

$$\hat{\theta}^*_n := \arg\max_{x \in \mathbb{R}} \hat{S}^*_n(x), \quad \tilde{\theta}^*_n := \arg\max_{x \in \mathbb{R}} \tilde{S}^*_n(x), \quad \bar{\theta}^*_n := \arg\max_{x \in \mathbb{R}} \bar{S}^*_n(x), \quad (6.2)$$

in the cases

$$\pi_1 > \pi_2, \quad \pi_1 < \pi_2, \quad \pi_1 \neq \pi_2,$$
Remark 6.1 (Existence of plug-in estimators for the breakpoint)

(i) The estimators in (6.2) are not defined if the corresponding \( \alpha \) and/or \( \beta \) value is not defined. This is the case if a corresponding plug-in estimator for the risk levels is equal to zero or one. Thus, for the existence of plug-in estimators for the breakpoint, it is necessary

- that there are at least two observations \( X_i \) to the left and to the right of the corresponding Dempe-Stute estimator from the first step, and
- that under the corresponding observations \( Y_i \) there is at least one success (default) and one failure (non-default).

Written as formulas this means in the case of \( \pi_1 > \pi_2 \) and the one-sided estimator \( \hat{\theta}_n \):

\[
0 < \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq \hat{\theta}_n\}} Y_i < \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq \hat{\theta}_n\}} \quad \text{and} \quad 0 < \sum_{i=1}^{n} \mathbb{1}_{\{X_i > \hat{\theta}_n\}} Y_i < \sum_{i=1}^{n} \mathbb{1}_{\{X_i > \hat{\theta}_n\}}.
\]

(ii) Under the assumptions of Theorem 3, the plug-in estimators for the breakpoint exist almost surely for eventually all \( n \in \mathbb{N} \), since the plug-in estimators for the risk levels converge a.s. to \( \pi_1 \) or \( \pi_2 \), which are not equal to zero or one.

Theorem 4 Let the model assumptions (2.2), (2.3), and (2.4) hold. If there exists an open interval \( I_X \) with \( P(X \in I_X) = 1 \) and a function \( T^{-1}_z = T^{-1}(z, \cdot) : I_X \to \mathbb{R} \), which for all \( u \in \mathbb{R}, x \in I_X \) and all possible realizations \( z \) of \( Z \) fulfills the following four properties:

(i) \( X_i \leq x \iff U_i \leq T^{-1}(Z, x) \) for all \( 1 \leq i \leq n \),

(ii) the composition \( F_{U} \circ T^{-1}_z \) is rcll (càdlàg) and increasing,

(iii) there does not exist an interval containing \( \theta \) on which \( F_{U} \circ T^{-1}_z \) is constant, and

(iv) the composition \( F_{U} \circ T^{-1}_z \) is continuous at \( \theta \),

then the breakpoint estimators from (6.2) are strongly consistent, i.e. as \( n \to \infty \) it holds

\[
\hat{\theta}_n^* \xrightarrow{a.s.} \theta, \quad \text{if } \pi_1 > \pi_2,
\]

\[
\hat{\theta}_n^* \xrightarrow{a.s.} \theta, \quad \text{if } \pi_1 < \pi_2,
\]
Proof of Theorem 4. Again, we make use of the Argmax-Theorem of Ferger [2009].

(1) Note that the criterion functions from (6.1) are rcll and that the corresponding estimators from (6.2) belong to the respective criterion functions’ maximum points as requested.

(2) Recall the definitions of $\alpha$ and $\beta$ from (3.5). From Theorem 3 and the continuous mapping theorem, it follows

$$
\hat{\alpha}_n^* \xrightarrow{a.s.} \alpha, \quad \text{if } \pi_1 > \pi_2,
$$

$$
\hat{\beta}_n^* \xrightarrow{a.s.} \beta, \quad \text{if } \pi_1 < \pi_2,
$$

$$
\bar{\alpha}_n^* \xrightarrow{a.s.} \alpha, \quad \text{if } \pi_1 \neq \pi_2.
$$

Furthermore, note that $0 \leq H < 1$, since from (3.15) we get

$$
0 \leq H(x) = \begin{cases} 
\pi_1 F_U(T^{-1}(Z, x)) & \text{if } x \leq \theta, \\
\pi_1 F_U(T^{-1}(Z, \theta)) + \pi_2 \left( F_U(T^{-1}(Z, x)) - F_U(T^{-1}(Z, \theta)) \right) & \text{if } x > \theta.
\end{cases}
$$

Now we can prove the almost sure uniform convergence of the criterion functions towards the limit process $S^*$ from (3.18). Exemplarily we consider the case $\pi_1 > \pi_2$, the other cases are treated analogously:

$$
0 \leq \| \hat{S}_n^* - S^* \| = \| \hat{\alpha}_n^* H_n + \hat{\beta}_n^* F_n - \alpha H - \beta F_U(T^{-1}(Z, \cdot)) \|
$$

$$
\leq \| \hat{\alpha}_n^* H_n - \hat{\alpha}_n H + \hat{\alpha}_n H - \alpha H \|
$$

$$
+ \| \hat{\beta}_n^* F_n - \beta F_n + \beta F_n - \beta F_U(T^{-1}(Z, \cdot)) \|
$$

$$
\leq \left( \frac{\| H_n - H \|}{\alpha} \right)^{\alpha} \frac{\| \hat{\alpha}_n - \alpha \|}{\alpha} \frac{\| H \|}{\alpha}^{\alpha} + \left( \frac{\| F_n - F \|}{\beta} \right)^{\beta} \left( \| F_n - F_U(T^{-1}(Z, \cdot)) \| \right)^{\beta}
$$

D. Tillich, C. Lehmann
By (3.11), (3.17), and (6.3), we get
\[ \| \hat{S}^*_n - S^* \| \xrightarrow{a.s.} 0 \quad (n \to \infty). \]

(3) As seen in Section 3, the limit process $S^*$ is rcll and the unknown breakpoint $\theta$ is its well-separated maximum point.

Thus, the Argmax-Theorem yields the desired result. \[ \square \]

7 On the transformation $T$ and the random variables $X$, $U$, and $Z$

Recall the model equation (2.2):
\[ X_i = T(Z, U_i), \quad 1 \leq i \leq n. \]

The latent random variables $Z, U_1, \ldots, U_n$ are assumed to be mutually independent. The random variables $U_i$ are identically distributed with distribution function $F_U$. The marginal distributions as well as the joint distributions of $Z$ and the $U_i$, and therefore the distribution of $X_i$, are unknown. The Function $T : \mathbb{R}^2 \to \mathbb{R}$ is a real-valued function of two real variables.

As demonstrated above, further assumptions on $T$, $X$, $U$, and $Z$ are required, namely (i)--(iv) of the following lemma. Because there is a complex interaction between these conditions, we will state a situation where all the properties (i)--(iv) are fulfilled. For this purpose, let
\[ T_z := T(z, \cdot) : \mathbb{R} \to \mathbb{R} \]
be the restriction of $T$ to $\{z\} \times \mathbb{R}$ for all $z \in \mathbb{R}$.

**Lemma 7.1** Let $X$, $U$, and $Z$ real random variables and let $T : \mathbb{R}^2 \to \mathbb{R}$ with $X = T(Z, U)$.

**A** Let $I_U$ be an open interval (finite or infinite) with $P(U \in I_U) = 1$.

**B** Let $Z$ a random variable such that with probability 1

1. the restriction $T_z$ is strictly increasing on $I_U$,
2. the restriction $T_z$ is continuous on $I_U$,

3. the image $T_z(I_U)$ does not depend on $z$. Denote this image set by $I_X$.

Then the restriction $T_z : I_U \to \mathbb{R}$ has an inverse function $T_z^{-1} = T^{-1}(z, \cdot)$ defined on the open interval $I_X$ and the inverse is strictly increasing and continuous. Furthermore, with probability 1

(i) for all $x \in I_X$ it holds

$$X_i \leq x \iff T(Z, U_i) \leq x \iff T_Z(U_i) \leq x \iff U_i \leq T^{-1}(Z, x),$$

(7.1)

(ii) the composition $F_U \circ T_z^{-1} : I_X \to [0, 1]$ is real and increasing on $I_X$.

Even if $I_X \neq \mathbb{R}$, it holds

$$P(X \in I_X) = 1.$$ 

Thus, the (reasonably extended) composition $F_U \circ T_z^{-1}$ is a cumulative distribution function with

$$\lim_{x \to -\infty} F_U(T_z^{-1}(x)) = 0 \quad \text{and} \quad \lim_{x \to \infty} F_U(T_z^{-1}(x)) = 1.$$ 

Moreover, if $F_U$ is strictly increasing on $I_U$ and $\theta \in I_X$, then

(iii) there does not exist an interval containing $\theta$ on which $F_U \circ T_z^{-1}$ is constant.

If $F_U$ is continuous on $I_U$ and $\theta \in I_X$, then

(iv) the composition $F_U \circ T_z^{-1}$ is continuous at $\theta$.

**Proof of Lemma 7.1** By B1, function $T_z$ has an inverse function $T_z^{-1}$ on $T_z(I_U)$ and the inverse is strictly increasing and continuous. By B3, the inverse is defined on $I_X$ for almost all $z \in \mathbb{R}$. By B2, the set $I_X$ is an open interval with boundary points

$$x_{\inf} := \inf I_X \quad \text{and} \quad x_{\sup} := \sup I_X,$$

compare e.g. [Heuser 2000, p. 231, 37.1] or [Schröder 2008, p. 67, Thm. 3.36 and 3.38].

Let $u \in I_U$ and $x \in I_X$. Since $T_z$ is strictly increasing, see B1, it holds

$$u \leq T_z^{-1}(x) \iff T_z(u) \leq T_z(T_z^{-1}(x)) \iff T_z(u) \leq x.$$
Together with A this yields (7.1). Before we go on with proving (ii)–(iv), note that (7.1) is also fulfilled for weaker assumptions.

Remark 7.1 (Weaker assumptions and generalized inverse) Let \( T_z \) increasing and left continuous. For all \( x \in I_X \) define the generalized inverse

\[
T_z^\rightarrow(x) := \inf\{u \in I_U \mid T_z(u) > x\}.
\]

The function \( T_z^\rightarrow \) is increasing and right continuous. This is sufficient for (7.1), if \( T_z^{-1} \) is replaced with \( T_z^\rightarrow \).

If \( T_z \) is actually continuous (not only left continuous), then the generalized inverse \( T_z^\rightarrow \) is strictly increasing. If \( T_z \) is actually strictly increasing, then the generalized inverse \( T_z^\rightarrow \) is continuous.

For the remaining statements, we investigate the composition \( F_U \circ T_z^{-1} : I_X \to [0,1] \). It is increasing on \( I_X \). To see this, let \( x_1 < x_2 \in I_X \). Since \( T_z^{-1} \) is strictly increasing, it follows \( T_z^{-1}(x_1) < T_z^{-1}(x_2) \in I_U \). The assertion \( F_U(T_z^{-1}(x_1)) \leq F_U(T_z^{-1}(x_2)) \in [0,1] \) holds because \( F_U \) is increasing on \( I_U \). This is the second part of (ii). If \( F_U \) is actually strictly increasing on \( I_U \), then \( \leq \) becomes \( < \) and the composition is strictly increasing. If additionally \( \theta \in I_X \), then (iii) results.

Now we check the first part of (ii), i.e. the rcll-property of \( F_U \circ T_z^{-1} \). First note that this composition is bounded. And since it is increasing, all left and right limits exist, see e.g. Schröder [2008, p. 41, Thm. 2.37] or Heuser [2000, p. 155, 23.1]. Let \( x \in I_X \). There is a sequence \( (x_n)_{n \in \mathbb{N}} \subset I_X \) with \( x_n \downarrow x \) as \( n \to \infty \). Since \( T_z^{-1} \) is increasing and especially right continuous it holds \( T_z^{-1}(x_n) \downarrow T_z^{-1}(x) \) as \( n \to \infty \). Since \( F_U \) is a cumulative distribution function, which in particular is right continuous, we get right continuity of the composition:

\[
F_U \circ T_z^{-1}(x_n) = F_U(T_z^{-1}(x_n)) \to F_U(T_z^{-1}(x)) = F_U \circ T_z^{-1}(x).
\]

Thus, (ii) is proven. If \( F_U \) is actually continuous on \( I_U \), then the continuity of \( T_z^{-1} \) on \( I_X \) implies the continuity of the composition \( F_U \circ T_z^{-1} \) on \( I_X \), compare Heuser [2000, p. 215, 34.4]. If additionally \( \theta \in I_X \), then (iv) follows.

Let \( u_{\inf} := \inf I_U \) and \( u_{\sup} := \sup I_U \). From A we conclude that

\[
\lim_{u \downarrow u_{\inf}} F_U(u) = 0 \quad \text{and} \quad \lim_{u \uparrow u_{\sup}} F_U(u) = 1.
\]

The reasoning for this is the same as in Witting [1985, p. 19, 1.15 b].
Let \((x_n)_{n \in \mathbb{N}} \in I_X\) a sequence with \(x_n \downarrow x_{\text{inf}}\) as \(n \to \infty\). Then it follows \(u_n := T^{-1}_z(x_n) \downarrow u_{\text{inf}}\) and
\[
\lim_{x \downarrow x_{\text{inf}}} F_U(T^{-1}_z(x)) = \lim_{n \to \infty} F_U(T^{-1}_z(x_n)) = \lim_{n \to \infty} F_U(u_n) = 0.
\]

In the same manner one shows
\[
\lim_{x \uparrow x_{\text{sup}}} F_U(T^{-1}_z(x)) = 1.
\]

By (7.1) it results
\[
P(X \leq x_{\text{inf}}) = \lim_{x \downarrow x_{\text{inf}}} P(X \leq x) = \lim_{x \downarrow x_{\text{inf}}} P(U \leq T^{-1}_z(x)) = \lim_{x \downarrow x_{\text{inf}}} F_U(T^{-1}_z(x)) = 0
\]
and
\[
P(X < x_{\text{sup}}) = \lim_{x \uparrow x_{\text{sup}}} P(X \leq x) = \lim_{x \uparrow x_{\text{sup}}} P(U \leq T^{-1}_z(x)) = \lim_{x \uparrow x_{\text{sup}}} F_U(T^{-1}_z(x)) = 1.
\]

Thus, \(P(X \in I_X) = P(x_{\text{inf}} < X < x_{\text{sup}}) = 1\).

In summary, if \(I_X = \mathbb{R}\) then the composite function \(F_U \circ T^{-1}_z\) is a cumulative distribution function. If \(I_X \neq \mathbb{R}\), we can extend the composition to achieve a cumulative distribution function, namely as follows:

- If \(x_{\text{inf}} > -\infty\), then \(F_U \circ T^{-1}_z(x) := 0\) for all \(x \leq x_{\text{inf}}\).
- If \(x_{\text{sup}} < +\infty\), then \(F_U \circ T^{-1}_z(x) := 1\) for all \(x \geq x_{\text{sup}}\).

Thus, we get the following limits in all cases:
\[
\lim_{x \to -\infty} F_U(T^{-1}_z(x)) = 0 \quad \text{and} \quad \lim_{x \to \infty} F_U(T^{-1}_z(x)) = 1
\]
and the proof is complete.

\[\square\]

Example 7.1 The i.i.d. case, i.e.

\[X_i = T(Z, U_i) = U_i \quad \text{i.i.d.}\]

is covered by Lemma 7.1. Let \(I_U = \mathbb{R}\), hence \(P(U \in I_U) = 1\). The restriction of \(T\) is

\[T_z(u) = u \quad \text{for all } u, z \in \mathbb{R}.\]
It is strictly increasing and continuous on \( I_U = \mathbb{R} \). Its image \( T_z(I_U) = \mathbb{R} \) is independent of \( z \). Thus, there is an inverse on \( I_X = \mathbb{R} \), namely
\[
T_z^{-1}(x) = x \quad \text{for all } x, z \in \mathbb{R},
\]
and the equivalence in Lemma \([7.1](i)\) is fulfilled:
\[
T_z(u) \leq x \iff u \leq x \iff u \leq T_z^{-1}(x).
\]
For the composition \( F_U \circ T_z^{-1} \), it holds
\[
F_U(T_z^{-1}(x)) = F_U(x) \quad \text{for all } x, z \in \mathbb{R},
\]
\[
\lim_{x \to -\infty} F_U(T_z^{-1}(x)) = 0 \quad \text{for all } z \in \mathbb{R},
\]
\[
\lim_{x \to \infty} F_U(T_z^{-1}(x)) = 1 \quad \text{for all } z \in \mathbb{R}.
\]
Obviously, the composition is rcll and increasing (Lemma \([7.1](ii)\)) and the limits are 0 and 1, resp., since these properties hold for \( F_U = F_U \circ T_z^{-1} \). The remaining conditions (iii) and (iv) of our theorems are equal to the conditions in the i.i.d. case, compare e. g. [Ferger and Klotsche, 2009, Prop. 2.3, Prop. 5.5]. They can be achieved by requiring that \( F_U \) is strictly increasing and continuous on \( I_U = \mathbb{R} \), as stated in the lemma.

Example 7.2 The model used in [Tillich and Ferger, 2015], i. e.
\[
X_i = T(Z, U_i) = Z + U_i
\]
is also covered by Lemma \([7.1]\). Let \( I_U = \mathbb{R} \), hence \( P(U \in I_U) = 1 \). The restriction \( T_z \) of \( T \) is given as
\[
T_z(u) = z + u \quad \text{for all } u, z \in \mathbb{R}.
\]
It is strictly increasing and continuous on \( I_U = \mathbb{R} \) for all \( z \in \mathbb{R} \). Its image \( T_z(I_U) = \mathbb{R} \) is independent of \( z \). Thus, there is an inverse on \( I_X = \mathbb{R} \), namely
\[
T_z^{-1}(x) = x - z \quad \text{for all } x, z \in \mathbb{R}
\]
and the equivalence in Lemma \([7.1](i)\) is fulfilled:
\[
T_z(u) \leq x \iff z + u \leq x \iff u \leq x - z \iff u \leq T_z^{-1}(x).
\]
For the composition \( F_U \circ T_{z}^{-1} \), it holds
\[
F_U(T_{z}^{-1}(x)) = F_U(x - z) \quad \text{for all } x, z \in \mathbb{R},
\]
\[
\lim_{x \to -\infty} F_U(T_{z}^{-1}(x)) = 0 \quad \text{for all } z \in \mathbb{R},
\]
\[
\lim_{x \to \infty} F_U(T_{z}^{-1}(x)) = 1 \quad \text{for all } z \in \mathbb{R}.
\]

Hence, the composition is increasing and \( I_X \) and Lemma 7.1(ii) is fulfilled.

If \( F_U \) is strictly increasing on \( \mathbb{R} \), as demanded in the Theorems 1–3 of Tillich and Ferger [2015], then Lemma 7.1 yields that Condition (iii) of our Theorems 1–4 is fulfilled for all \( \theta \in \mathbb{R} \). The assumption on \( F_U \) can be weakened, as already mentioned in Remark 3(iii) of Tillich and Ferger [2015]. If \( F_U \) additionally is continuous on \( \mathbb{R} \), as demanded in the Theorems 2 and 3 of Tillich and Ferger [2015], then Condition (iv) of our Theorems 2 and 4 is fulfilled for all \( \theta \in \mathbb{R} \). Thus, our theorems have slightly weaker assumptions than the Theorems 1–3 of Tillich and Ferger [2015].

Note that the existence of an inverse function due to a strict decrease of \( T_z \) does not suffice for the equivalence in (7.1) as the following example shows. Compared to the model used in Tillich and Ferger [2015], we merely replaced the sum by the difference of systematic and idiosyncratic factor:

\[
X_i = T(Z, U_i) = Z - U_i,
\]
\[
T_{z}(u) = z - u \quad \text{for all } z, u \in \mathbb{R},
\]
\[
T_{z}^{-1}(x) = z - x \quad \text{for all } z, x \in \mathbb{R}.
\]

But the equivalence in (i) is not fulfilled:

\[
T(z, u) \leq x \iff z - u \leq x \iff u \geq z - x \iff u \geq T^{-1}(z, x).
\]

The relation is exactly the other way round as desired.

As yet, \( T_z(I_U) = \mathbb{R} \). Next we look at two examples with \( T_z(I_U) \subsetneq \mathbb{R} \). They are used in the simulation study in Section 8. In Example 7.3 we have \( I_U = \mathbb{R} \), while in Example 7.4 we use \( I_U = [0, 1[ \).  

**Example 7.3** Let

\[
X_i = T(Z, U_i) = \Phi(Z + U_i),
\]
7 ON THE TRANSFORMATION $T$ AND THE RANDOM VARIABLES $X$, $U$, AND $Z$

where $\Phi$ is the cumulative distribution function of a standard normal distribution $N(0, 1)$. Let $I_U = \mathbb{R}$, hence $P(U \in I_U) = 1$. The restriction $T_z$ of $T$ is given as

$$T_z(u) = \Phi(z + u) \quad \text{for all } u, z \in \mathbb{R}.$$  

It is strictly increasing and continuous on $I_U = \mathbb{R}$ for all $z \in \mathbb{R}$. Its image $T_z(I_U) = ]0, 1]$ is independent of $z$. Thus, there is an inverse on $I_X = ]0, 1]$, namely

$$T^{-1}_z(x) = \Phi^{-1}(x) - z \quad \text{for all } x \in ]0, 1], z \in \mathbb{R},$$

where $\Phi^{-1}$ denotes the inverse of $\Phi$. Furthermore, the equivalence in Lemma 7.1(i) is fulfilled:

$$T_z(u) \leq x \iff u \leq T^{-1}_z(x).$$

For the composition $F_U \circ T^{-1}_z$, it holds

$$F_U(T^{-1}_z(x)) = F_U(\Phi^{-1}(x) - z) \quad \text{for all } x \in ]0, 1], z \in \mathbb{R}.$$  

It is increasing and rcll on $I_X = ]0, 1]$ and hence Lemma 7.1(ii) is fulfilled. Furthermore,

$$\lim_{x \downarrow 0} F_U(T^{-1}_z(x)) = 0 \quad \text{and} \quad \lim_{x \uparrow 1} F_U(T^{-1}_z(x)) = 1 \quad \text{for all } z \in \mathbb{R}. \quad (7.2)$$

To show this, we exemplarily consider the second statement. Let $(x_n)_{n \in \mathbb{N}} \subset I_X$ be a sequence with $x_n \uparrow 1$ as $n \to \infty$. Then $T^{-1}_z(x_n) = \Phi^{-1}(x_n) - z \to \infty$ as $n \to \infty$. Since $F_U$ is a cumulative distribution function with $\lim_{u \to \infty} F_U(u) = 1$, it follows that $F_U(T^{-1}_z(x_n)) \to 1$ as $n \to \infty$.

Due to (7.2), the composition can be extended by 0 for all $x \leq 0$ and by 1 for all $x \geq 1$ to get a function, which is defined (as well as rcll and increasing) on $\mathbb{R}$. Conditions (iii) and (iv) of Lemma 7.1 can be achieved for instance by requiring that $F_U$ is strictly increasing and continuous on $\mathbb{R}$.

Example 7.4 Let $X_i = T(Z, U_i)$ with link function

$$T(z, u) = \begin{cases} u^2 & \text{if } z > 0 \text{ and } u \geq 0 \\ 0 & \text{else.} \end{cases}$$

Let $U$ be a random variable with $P(U \in ]0, 1[) = 1$ and let $Z$ be a random variable with
\[ P(Z > 0) = 1. \text{ For instance, let } U \sim \text{Beta}(c,d), \text{ i.e. } U \text{ has a beta distribution with parameters } c > 0 \text{ and } d > 0, \text{ and let } Z \sim \text{LN}(\mu, \sigma^2), \text{ i.e. } Z \text{ has a lognormal distribution with parameters } \mu \text{ and } \sigma^2, \text{ see e.g. Casella and Berger [2002, pp. 106 and 109] or Rinne [2008] pp. 340 and 306}. \]

Then with probability 1 the restriction \( T_z \) is strictly increasing and continuous on \( I_U = [0,1] \), and its image \( T_z(I_U) = ]0,1[ \) is independent of \( z \). Thus, there is an inverse \( T_z^{-1} \) on \( I_X = ]0,1[ \),

\[ T_z^{-1}(x) = x^{1/z} \text{ for all } x \in ]0,1[ , z > 0 \]

and the equivalence in Lemma 7.1(i) is fulfilled:

\[ T_z(u) \leq x \Leftrightarrow u \leq T_z^{-1}(x). \]

The composition \( F_U \circ T_z^{-1} \) can be extended to a cumulative distribution function as follows

\[
F_U(T^{-1}(z, x)) =
\begin{cases}
0, & \text{if } x \leq 0 \\
F_U(x^{1/z}), & \text{if } 0 < x < 1 \\
1, & \text{if } x \geq 1.
\end{cases}
\]

Consequently, Lemma 7.1(ii) is fulfilled, i.e. the (extended) composition is well and increasing.

Moreover, let \( \theta \in I_X = ]0,1[ \). Then the conditions (iii) and (iv) of Lemma 7.1 can be ensured by requiring that \( F_U \) is strictly increasing and continuous on \( I_U \), resp. If for instance \( U \sim \text{Beta}(c,d) \), then \( F_U \) is strictly increasing and continuous on \( I_U \).

8 Simulation study

In this section some simulations on breakpoint estimation were performed to illustrate the foregoing theoretical outcomes. We investigate the maximum likelihood (ML) estimator \( \theta_n^* \) from Section 3, the two-sided Dempsie-Stute (DS) estimator \( \bar{\theta}_n \) from Section 4, and the plug-in (PI) estimator \( \bar{\theta}_n^* \) from Section 6 as well as the risk level estimators \( \bar{\pi}_{1,n} \) and \( \bar{\pi}_{2,n} \) from Section 5. Thereby, the simulations use the link functions \( T \) from Examples 7.3 and 7.4.
8.1 Simulation of Example 7.3

The first simulation is based on Example 7.3 and the following is assumed:

\[ X_i = T(Z, U_i) = \Phi(Z + U_i), \]  

where \( \Phi \) denotes the cumulative distribution function of a standard normal distribution \( N(0, 1) \). Additionally to the declarations in Example 7.3, let

\[ Z \sim N(0, 1) \text{ independent of } U_i \sim N(0, 1), \quad 1 \leq i \leq n. \]  

Hence, \( X_i \) follows a so-called probit normal distribution, because its probit \( \Phi^{-1}(X_i) \) is normally distributed, \( \Phi^{-1}(X_i) \sim N(0, 2) \). Since \( F_U = \Phi \) is strictly increasing and continuous on \( I_U = \mathbb{R} \), Example 7.3 and Lemma 7.1 yield that the assumptions on \( T \) of Theorems 14 are fulfilled if \( \theta \in I_X = ]0, 1[ \).

To run the simulation, the random numbers are generated in a two-step procedure. In the first step, the realizations \( x_i \) of the score variables \( X_i \) are simulated according to (8.1) and (8.2). In the second step, given all \( X_i = x_i \), the corresponding default variables \( Y_i \) are generated as independent Bernoulli variables with default probability

\[ m(x_i) = \pi_1 \mathbb{1}_{\{x_i \leq \theta\}} + \pi_2 \mathbb{1}_{\{x_i > \theta\}} = \begin{cases} \pi_1, & \text{if } x_i \leq \theta, \\ \pi_2, & \text{if } x_i > \theta, \end{cases} \]

in agreement with (2.3) and (2.4).

Next, we have to fix the model parameters, i.e. the breakpoint \( \theta \) and the two risk levels \( \pi_1 \) and \( \pi_2 \). We investigate the scenarios stated in Table 2. The names and the risk levels of the scenarios are motivated by Tillich and Ferger [2015, Section 5]. The scenario ‘initial’ has very well separated risk levels. For the ‘risk level’ scenario the distance between the risk levels is smaller and \( \pi_1 < \pi_2 \). Finally, the ‘credit risk’

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
Name of the scenario & breakpoint & risk level \( \pi_1 \) & risk level \( \pi_2 \) \\
\hline
initial & \( \theta = 0.17 \) & 0.8 & 0.1 \\
\hline
risk level & and & 0.1 & 0.3 \\
\hline
credit risk & \( \theta = 0.5 \) & 0.008 & 0.002 \\
\hline
\end{tabular}
\end{table}
scenario is motivated by Tillich [2013, p. 114], who analyzed quarterly data of a retail portfolio and found average default rates of about 0.008 and 0.002. The breakpoint \( \theta = 0.17 \) corresponds to the first quartile of a probit normal distribution with parameters 0 and 2, while \( \theta = 0.5 \) corresponds to the median of this probit normal distribution. For the simulation, every combination of breakpoint and risk levels is regarded. In addition, the sample size \( n \) is varied from 1000 to 100,000.

Table 3 contains the estimated bias and the estimated root mean squared error (RMSE) of the breakpoint estimators \( \hat{\theta}_n^*, \hat{\theta}_n \) and \( \hat{\theta}_n^{**} \) for the scenarios from Table 2. There are performed 10,000 Monte Carlo replications for each scenario. The last column of Table 3 contains the number of valid replications, i.e., the number of replications where the plug-in estimate \( \hat{\theta}_n^* \) exists, compare the requirements for the existence of the estimates in Remark 6.1(i). The quite small number of valid replications in the credit risk scenarios can be explained by the risk levels near zero in connection with a small sample size, particularly in the case \( n = 1000 \). But in accordance with Remark 6.1(ii), the number of valid replications increases with increasing sample size.

In agreement with the consistency property of the estimators \( \hat{\theta}_n^*, \hat{\theta}_n \), and \( \hat{\theta}_n^{**} \) as stated in Theorems 1, 2, and 4 resp., absolute bias and RMSE are decreasing while the sample size is increasing. Thereby, the ML estimates show the smallest bias and RMSE in absolute values, which is caused by the known risk levels. In the case of unknown risk levels, the plug-in estimates are slightly better than the Dempe-Stute estimates. Moreover, the model parameters influence the results as expected:

- With only a few exceptions in the ‘credit risk’ scenario for small sample sizes, the estimates are better if the breakpoint equals the median of the unconditional distribution of the score instead of its quartile.

- The estimates are dependent on the difference and the location of the risk levels, i.e., the estimates in scenario ‘initial’ are better than those in scenario ‘risk level’, which themselves are better than the estimates in the ‘credit risk’ setting.

Additionally, Table 4 contains the estimated bias and estimated RMSE of the risk level estimators \( \hat{\pi}_{1,n} \) and \( \hat{\pi}_{2,n} \) based on the two-sided Dempe-Stute estimate \( \hat{\theta}_n \) for the breakpoint. The estimates \( \hat{\pi}_{1,n} \) and \( \hat{\pi}_{2,n} \) are used as input for the plug-in-estimator of the breakpoint. In agreement to the consistency property of the estimators \( \hat{\pi}_{1,n} \) and \( \hat{\pi}_{2,n} \) as stated in Theorem 3, absolute bias and RMSE are decreasing while the sample size is increasing. Note that small bias and RMSE have to be seen against the background of the absolute risk levels, especially in the ‘credit risk’ scenario.
Table 3: Simulation results based on Example 7.3. Bias and root mean squared error (RMSE) of breakpoint estimators $\hat{\theta}_n^*$, $\bar{\theta}_n$, and $\hat{\theta}_n^*$ for scenarios from Table 2 with varying sample sizes.

<table>
<thead>
<tr>
<th>Scenario name</th>
<th>Sample size</th>
<th>$\hat{\theta}_n^*$ (ML)</th>
<th>$\bar{\theta}_n$ (DS)</th>
<th>$\hat{\theta}_n^*$ (PI)</th>
<th>number of valid replications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breakpoint</td>
<td>$n$</td>
<td>bias</td>
<td>RMSE</td>
<td>bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>initial</td>
<td>1000</td>
<td>-0.0009</td>
<td>0.0231</td>
<td>0.0944</td>
<td>0.2426</td>
</tr>
<tr>
<td>$\theta = 0.17$</td>
<td>10000</td>
<td>-0.0002</td>
<td>0.0095</td>
<td>0.0551</td>
<td>0.1824</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>-0.0000</td>
<td>0.0037</td>
<td>0.0285</td>
<td>0.1297</td>
</tr>
<tr>
<td>risk level</td>
<td>1000</td>
<td>0.0127</td>
<td>0.0660</td>
<td>0.2030</td>
<td>0.3550</td>
</tr>
<tr>
<td>$\theta = 0.17$</td>
<td>10000</td>
<td>0.0019</td>
<td>0.0222</td>
<td>0.1220</td>
<td>0.2736</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>0.0004</td>
<td>0.0087</td>
<td>0.0774</td>
<td>0.2181</td>
</tr>
<tr>
<td>credit risk</td>
<td>1000</td>
<td>0.0416</td>
<td>0.2559</td>
<td>0.1989</td>
<td>0.3934</td>
</tr>
<tr>
<td>$\theta = 0.17$</td>
<td>10000</td>
<td>-0.0043</td>
<td>0.1003</td>
<td>0.1900</td>
<td>0.3570</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>-0.0035</td>
<td>0.0327</td>
<td>0.1189</td>
<td>0.2749</td>
</tr>
<tr>
<td>initial</td>
<td>1000</td>
<td>-0.0021</td>
<td>0.0141</td>
<td>-0.0064</td>
<td>0.1020</td>
</tr>
<tr>
<td>$\theta = 0.5$</td>
<td>10000</td>
<td>-0.0002</td>
<td>0.0040</td>
<td>-0.0024</td>
<td>0.0675</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>-0.0000</td>
<td>0.0015</td>
<td>-0.0002</td>
<td>0.0411</td>
</tr>
<tr>
<td>risk level</td>
<td>1000</td>
<td>0.0044</td>
<td>0.0481</td>
<td>0.0160</td>
<td>0.1684</td>
</tr>
<tr>
<td>$\theta = 0.5$</td>
<td>10000</td>
<td>0.0003</td>
<td>0.0119</td>
<td>0.0056</td>
<td>0.1065</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>0.0001</td>
<td>0.0028</td>
<td>0.0024</td>
<td>0.0737</td>
</tr>
<tr>
<td>credit risk</td>
<td>1000</td>
<td>-0.1104</td>
<td>0.2794</td>
<td>-0.0957</td>
<td>0.2995</td>
</tr>
<tr>
<td>$\theta = 0.5$</td>
<td>10000</td>
<td>-0.0259</td>
<td>0.1072</td>
<td>-0.0377</td>
<td>0.2051</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>-0.0035</td>
<td>0.0260</td>
<td>-0.0163</td>
<td>0.1353</td>
</tr>
</tbody>
</table>
Table 4: Simulation results based on Example 7.3 Bias and root mean squared error (RMSE) of risk level estimators $\bar{\pi}_{1,n}$ and $\bar{\pi}_{2,n}$ for scenarios from Table 2 with varying sample sizes.

<table>
<thead>
<tr>
<th>Scenario name</th>
<th>Sample size (n)</th>
<th>estimator $\bar{\pi}_{1,n}$</th>
<th>estimator $\bar{\pi}_{2,n}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breakpoint</td>
<td>bias</td>
<td>RMSE</td>
<td>bias</td>
</tr>
<tr>
<td>initial (\theta = 0.17)</td>
<td>1000</td>
<td>-0.1195</td>
<td>0.2601</td>
</tr>
<tr>
<td></td>
<td>10 000</td>
<td>-0.0787</td>
<td>0.2072</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>-0.0446</td>
<td>0.1544</td>
</tr>
<tr>
<td>risk level (\theta = 0.17)</td>
<td>1000</td>
<td>0.0581</td>
<td>0.0999</td>
</tr>
<tr>
<td></td>
<td>10 000</td>
<td>0.0422</td>
<td>0.0814</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>0.0295</td>
<td>0.0681</td>
</tr>
<tr>
<td>credit risk (\theta = 0.17)</td>
<td>1000</td>
<td>0.0057</td>
<td>0.0426</td>
</tr>
<tr>
<td></td>
<td>10 000</td>
<td>-0.0011</td>
<td>0.0038</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>-0.0011</td>
<td>0.0025</td>
</tr>
<tr>
<td>initial (\theta = 0.5)</td>
<td>1000</td>
<td>-0.0217</td>
<td>0.1059</td>
</tr>
<tr>
<td></td>
<td>10 000</td>
<td>-0.0115</td>
<td>0.0745</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>-0.0062</td>
<td>0.0527</td>
</tr>
<tr>
<td>risk level (\theta = 0.5)</td>
<td>1000</td>
<td>0.0150</td>
<td>0.0511</td>
</tr>
<tr>
<td></td>
<td>10 000</td>
<td>0.0090</td>
<td>0.0350</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>0.0050</td>
<td>0.0261</td>
</tr>
<tr>
<td>credit risk (\theta = 0.5)</td>
<td>1000</td>
<td>0.0046</td>
<td>0.0308</td>
</tr>
<tr>
<td></td>
<td>10 000</td>
<td>0.0001</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>-0.0002</td>
<td>0.0012</td>
</tr>
</tbody>
</table>
8.2 Simulation of Example 7.4

The second simulation is based on Example 7.4 and the following is assumed: Let

$$X_i = T(Z, U_i)$$

with link function

$$T(z, u) = \begin{cases} 
  u, & \text{if } z > 0 \text{ and } u \geq 0, \\
  0 & \text{else.}
\end{cases}$$

Moreover, let $U_i \overset{i.i.d.}{\sim} \text{Beta}(c, d)$ with $c = d = 1.5$ for all $i = 1, \ldots, n$. Then

$$P(U_i \in ]0, 1[) = 1, \ E[U_i] = 1/2, \text{ and } V[U_i] = 0.0625.$$  

Furthermore, let $Z \sim LN(\mu, \sigma^2)$ with $\mu = -0.0.025$ and $\sigma^2 = 0.05$. Then

$$P(Z > 0) = 1, \ E[Z] = 1, \text{ and } V[Z] = \exp(0.05) - 1 \approx 0.0513.$$  

All the random variables $Z$ and $U_i$, $i = 1, \ldots, n$, are mutually independent. As already stated in Example 7.4, the assumptions of Lemma 7.1 are fulfilled and therefore also the assumptions on the link function $T$ of Theorems 7.1 and 7.2.

For the simulation, we define the following scenarios for the breakpoint and the risk levels, see Table 5. The generation of the random numbers is performed analogously to Subsection 8.1.

Table 5 contains the estimated bias and the estimated root mean squared error (RMSE) of the breakpoint estimators $\hat{\theta}_n$ and $\bar{\theta}_n$ for the scenarios from Table 5 with different sample sizes. As in the simulation for Example 7.3, there were performed 10 000 Monte Carlo replications for each scenario. Additionally, Table 7 contains the

<table>
<thead>
<tr>
<th>Name of the scenario</th>
<th>breakpoint</th>
<th>risk level $\pi_1$</th>
<th>risk level $\pi_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>$\theta = 0.25$</td>
<td>0.8</td>
<td>0.1</td>
</tr>
<tr>
<td>risk level</td>
<td>and</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>credit risk</td>
<td>$\theta = 0.5$</td>
<td>0.008</td>
<td>0.002</td>
</tr>
</tbody>
</table>
Table 6: Simulation results based on Example 7.4. Bias and root mean squared error (RMSE) of breakpoint estimators $\hat{\theta}_n^*$, $\bar{\theta}_n$ and $\bar{\theta}_n^*$ for scenarios from Table 5 with varying sample sizes.

<table>
<thead>
<tr>
<th>Scenario name</th>
<th>Sample size</th>
<th>$\theta_n^*$ (ML)</th>
<th>$\bar{\theta}_n$ (DS)</th>
<th>$\bar{\theta}_n^*$ (PI)</th>
<th>number of valid replications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breakpoint</td>
<td>$n$</td>
<td>bias RMSE</td>
<td>bias RMSE</td>
<td>bias RMSE</td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>1000</td>
<td>-0.0009 0.0026</td>
<td>0.0046 0.0200</td>
<td>-0.0006 0.0068</td>
<td>10 000</td>
</tr>
<tr>
<td></td>
<td>10 000</td>
<td>-0.0001 0.0003</td>
<td>0.0007 0.0038</td>
<td>-0.0001 0.0003</td>
<td>10 000</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>-0.0000 0.0000</td>
<td>0.0001 0.0004</td>
<td>-0.0000 0.0000</td>
<td>10 000</td>
</tr>
<tr>
<td>risk level</td>
<td>1000</td>
<td>0.0017 0.0199</td>
<td>0.0627 0.1169</td>
<td>0.0338 0.0987</td>
<td>10 000</td>
</tr>
<tr>
<td></td>
<td>10 000</td>
<td>0.0002 0.0021</td>
<td>0.0127 0.0355</td>
<td>0.0012 0.0164</td>
<td>10 000</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>0.0000 0.0002</td>
<td>0.0020 0.0092</td>
<td>0.0000 0.0007</td>
<td>10 000</td>
</tr>
<tr>
<td>credit risk</td>
<td>1000</td>
<td>-0.0144 0.2048</td>
<td>0.1102 0.2616</td>
<td>0.0892 0.2595</td>
<td>4 490</td>
</tr>
<tr>
<td></td>
<td>10 000</td>
<td>-0.0141 0.0589</td>
<td>0.0540 0.1318</td>
<td>0.0406 0.1299</td>
<td>9 982</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>-0.0016 0.0072</td>
<td>0.0115 0.0375</td>
<td>0.0007 0.0217</td>
<td>10 000</td>
</tr>
<tr>
<td>initial</td>
<td>1000</td>
<td>-0.0007 0.0020</td>
<td>-0.0009 0.0025</td>
<td>-0.0007 0.0021</td>
<td>10 000</td>
</tr>
<tr>
<td></td>
<td>10 000</td>
<td>-0.0001 0.0002</td>
<td>-0.0001 0.0002</td>
<td>-0.0001 0.0002</td>
<td>10 000</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>-0.0000 0.0000</td>
<td>-0.0000 0.0000</td>
<td>-0.0000 0.0000</td>
<td>10 000</td>
</tr>
<tr>
<td>risk level</td>
<td>1000</td>
<td>0.0014 0.0158</td>
<td>0.0050 0.0195</td>
<td>0.0024 0.0177</td>
<td>10 000</td>
</tr>
<tr>
<td></td>
<td>10 000</td>
<td>0.0001 0.0015</td>
<td>0.0006 0.0023</td>
<td>0.0001 0.0015</td>
<td>10 000</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>0.0000 0.0002</td>
<td>0.0001 0.0002</td>
<td>0.0000 0.0002</td>
<td>10 000</td>
</tr>
<tr>
<td>credit risk</td>
<td>1000</td>
<td>-0.0763 0.2217</td>
<td>-0.0728 0.1767</td>
<td>-0.0913 0.2057</td>
<td>4 901</td>
</tr>
<tr>
<td></td>
<td>10 000</td>
<td>-0.0112 0.0526</td>
<td>-0.0212 0.0530</td>
<td>-0.0165 0.0540</td>
<td>9 999</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>-0.0011 0.0051</td>
<td>-0.0029 0.0082</td>
<td>-0.0012 0.0053</td>
<td>10 000</td>
</tr>
</tbody>
</table>
estimated bias and estimated RMSE of the risk level estimators $\bar{\pi}_{1,n}$ and $\bar{\pi}_{2,n}$ based on the two-sided Dempe-Stute estimates for the breakpoint. The estimates $\bar{\pi}_{1,n}$ and $\bar{\pi}_{2,n}$ are used as input for the plug-in-estimator of the breakpoint.

The two tables 6 and 7 show that the bias and the RMSE of the corresponding estimators are decreasing with an increasing sample size. This is in line with the property of consistency of Theorems 14. Also the influence of the model parameters, i. e. the breakpoint $\theta$ and the risk levels $\pi_1$ and $\pi_2$, is as described above in Subsection 8.1. And again, the ML estimates are better than the plug-in estimates, which themselves are better than the Dempe-Stute estimates with some exceptions in cases with small sample sizes. Noteworthy, the absolute bias and RMSE of the ML estimator and the plug-in estimator differ not that much here. This could be an indication of asymptotic equivalence of the ML and the plug-in estimators as in the i.i.d. case, see Ferger and Klotsche [2009, p. 115].

Comparing our simulation results for Examples 7.3 and 7.4 and those of Tillich and Ferger [2015] (who considered the one-sided Dempe-Stute estimator only), we see that the quality of the estimates is strongly influenced by the interplay of the (non-linear) link function $T$ and the joint distribution of the random variables $U_i$ and $Z$.

9 Conclusions and Outlook

The framework of this article is a binary regression model for observations $(X_i, Y_i) \in \mathbb{R} \times \{0, 1\}$ with a one-step regression function $m$. Function $m$ has three parameters of interest, namely the breakpoint $\theta$ and the risk levels $\pi_1$ and $\pi_2$. We investigated different estimators for these parameters under the assumption of non-i.i.d. observations. The dependence of the regressors $X_i$ is modeled by a possibly non-linear link function, i. e. by $X_i = T(Z, U_i)$. Based on this, a Bernoulli mixture model is used for the response variables $Y_i$.

We proved that all considered estimators, i. e. the maximum likelihood estimator (Section 3), the Dempe-Stute estimator (Section 4) and the plug-in estimator (Section 6) are strongly consistent for the unknown breakpoint. Furthermore, we proved strong consistency for the risk level estimators (Section 5) as well. In doing so, the assumptions on the link function $T$ turned out to be a crucial issue. In the primary representation, these assumptions are not that easily accessible. Thus, we clarified them more detailed and found that we can build on quite simple assumptions, namely on continuity and monotonicity (Section 7). This eases application and understanding.
Table 7: Simulation results based on Example 7.4. Bias and root mean squared error (RMSE) of risk level estimators $\hat{\pi}_{1,n}$ and $\hat{\pi}_{2,n}$ for scenarios from Table 5 with varying sample sizes.

<table>
<thead>
<tr>
<th>Scenario name</th>
<th>Sample size</th>
<th>estimator $\hat{\pi}_{1,n}$</th>
<th>estimator $\hat{\pi}_{2,n}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breakpoint</td>
<td>$n$</td>
<td>bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>initial</td>
<td>1 000</td>
<td>-0.0159</td>
<td>0.0652</td>
</tr>
<tr>
<td>$\theta = 0.25$</td>
<td>10 000</td>
<td>-0.0031</td>
<td>0.0209</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>-0.0003</td>
<td>0.0044</td>
</tr>
<tr>
<td>risk level</td>
<td>1 000</td>
<td>0.0258</td>
<td>0.0534</td>
</tr>
<tr>
<td>$\theta = 0.25$</td>
<td>10 000</td>
<td>0.0094</td>
<td>0.0253</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>0.0022</td>
<td>0.0098</td>
</tr>
<tr>
<td>credit risk</td>
<td>1 000</td>
<td>0.0074</td>
<td>0.0457</td>
</tr>
<tr>
<td>$\theta = 0.25$</td>
<td>10 000</td>
<td>-0.0001</td>
<td>0.0028</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>-0.0002</td>
<td>0.0010</td>
</tr>
<tr>
<td>initial</td>
<td>1 000</td>
<td>0.0005</td>
<td>0.0186</td>
</tr>
<tr>
<td>$\theta = 0.5$</td>
<td>10 000</td>
<td>0.0001</td>
<td>0.0059</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>-0.0000</td>
<td>0.0018</td>
</tr>
<tr>
<td>risk level</td>
<td>1 000</td>
<td>-0.0003</td>
<td>0.0148</td>
</tr>
<tr>
<td>$\theta = 0.5$</td>
<td>10 000</td>
<td>0.0001</td>
<td>0.0045</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>-0.0000</td>
<td>0.0014</td>
</tr>
<tr>
<td>credit risk</td>
<td>1 000</td>
<td>0.0048</td>
<td>0.0258</td>
</tr>
<tr>
<td>$\theta = 0.5$</td>
<td>10 000</td>
<td>0.0004</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td>100 000</td>
<td>0.0000</td>
<td>0.0004</td>
</tr>
</tbody>
</table>
Finally, we performed some simulations to illustrate the theoretical outcomes (Section 8). The results support the consistency property of the breakpoint estimators and the risk level estimators similar to the i.i.d. case and the non-i.i.d. case with a linear link function as in Tillich and Ferger [2015]. Comparing the estimators, we found that the maximum likelihood estimator leads to the best breakpoint estimates referring to the absolute bias and RMSE. This is not surprising because the ML estimator assumes the risk levels to be known. In the case of unknown risk levels, the plug-in estimator seems to be preferable in contrast to the Dempple-Stute estimator.

To check this conjecture, the theoretical investigation of convergence speed would be helpful. Thereby, the influence of a non-linear link function could be clarified. Another interesting point is the asymptotic distribution of the breakpoint and risk level estimators. It could form a basis for constructing confidence intervals and statistical tests for the model parameters. In order to gain more insights on the estimation procedure, a more comprehensive simulation study would be valuable. Its focus could also be aimed on some practical applications, e.g., credit risk.

From a practical point of view, also the following model modifications and extensions are of interest:

1. There is a need of estimation procedures in cases with more than two classes, i.e. more than one breakpoint. Obviously, an iterative application of the proposed estimators seems to be possible. However, a simultaneous treatment seems to be preferable. In both cases, the question of the correct number of classes and breakpoints raises.

2. Sometimes we have data from different (subsequent) periods. The usage of the proposed estimators seems to be possible also in this case. For a serious investigation by a simulation study and/or statistical theory, an appropriate model is needed. It has to take into account time dynamics. This refers especially to the modeling of the temporal dependence structure. To this, the established model structures can be adapted. A starting point for this field could be Tillich [2016b].

Finally, it should be noted that an application of the proposed methods is conceivable as well besides credit risk, for instance in medicine, environmental sciences or other fields of economics.

**Acknowledgment:** We thank Dietmar Ferger, Nicolas Frölich, Fabian Joost, Tony Klein, and René Mauer for advice, remarks, and hints.
Bibliography


D. Tillich, C. Lehmann


Dresdner Beiträge zu Quantitativen Verfahren (ISSN 0945-4802)

Ältere Ausgaben (1/94 – 47/08): http://wwqvs.file3.wcms.tu-dresden.de/f-db.htm


52/10  D. Tillich: Risikomaßzahlen für Kreditportfoliotranchen. 

53/10  S. Huschens: Kann es Rückzahlungswahrscheinlichkeiten von 100% geben? 


DOI: 10.13140/RG.2.1.2656.7768

57/12  S. Höse, S. Huschens: Credit Portfolio Correlations and Uncertainty. 

58/12  S. Fischer: Ratio calculandi periculi - Ein analytischer Ansatz zur Bestimmung der Verlustverteilung eines Kreditportfolios


60/14  C. Lehmann, D. Tillich: Consensus Information and Consensus Rating – A Note on Methodological Problems of Rating Aggregation. 

Erscheint in: Journal of Risk Model Validation.


63/16  S. Huschens: Chance (odd) versus Wahrscheinlichkeit (probability).

64/16  S. Huschens: Stetigkeit in der Statistik.

65/16  S. Huschens: Literaturlauswahl zur Statistik.

66/16  D. Tillich, C. Lehmann: Estimation in discontinuous Bernoulli mixture models applicable in credit rating systems with dependent data.

67/16  D. Tillich: Generalized Modeling and Estimation of Rating Classes and Default Probabilities Considering Dependencies in Cross and Longitudinal Section.