

Rafael Weißbach, Rafael; Voß, Sebastian

Conference Paper

A score-test on measurement errors in rating transition times

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Econometric Theory, No. C20-V3

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Rafael Weißbach, Rafael; Voß, Sebastian (2014) : A score-test on measurement errors in rating transition times, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Econometric Theory, No. C20-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/100532>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Accepted Manuscript

A score-test on measurement errors in rating transition times

Sebastian Voß, Rafael Weißbach

PII: S0304-4076(14)00016-5

DOI: <http://dx.doi.org/10.1016/j.jeconom.2014.01.004>

Reference: ECONOM 3863

To appear in: *Journal of Econometrics*

Received date: 7 November 2012

Revised date: 16 August 2013

Accepted date: 7 January 2014



Please cite this article as: Voß, S., Weißbach, R., A score-test on measurement errors in rating transition times. *Journal of Econometrics* (2014), <http://dx.doi.org/10.1016/j.jeconom.2014.01.004>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A Score-Test on Measurement Errors in Rating Transition Times

Sebastian Voß*

Rafael Weißbach†

January 10, 2014

We model credit rating histories as continuous-time discrete-state Markov processes. Infrequent monitoring of the debtors' solvency will result in erroneous observations of the rating transition times, and consequently in biased parameter estimates. We develop a score test against such measurement errors in the transition data that is independent of the error distribution. We derive the asymptotic χ^2 -distribution for the test statistic under the null by stochastic limit theory. The test is applied to an international corporate portfolio, while accounting for economic and debtor-specific covariates. The test indicates that measurement errors in the transition times are a real problem in practice.

1 Introduction

Credit ratings are used to appraise the credit worthiness of a debtor on an ordinal scale. As the credit worthiness evolves over time, the rating has to be modeled as a stochastic process. The objective is to estimate rating transition probabilities and especially default probabilities, as they are a major input to credit risk management. A lot of research has addressed the topic recently. Nickell et al. (2000), Hu et al. (2002), and Stefanescu et al. (2009), for example, use a discrete-time approach to model the rating transitions. Discrete-time models however, do not utilize all the information which is available to the analyst, as it does not account for the exact times at which the transitions occur, i.e. the rating at the beginning and at the end of a period - typically a year - is considered, while everything which happens in between this time span is neglected. Hence, a continuous-time approach, i.e. an analysis of durations, is the more appropriate choice and the continuous-time, finite-state Markov process remains the starting point for modeling credit rating transition histories (see Jarrow et al., 1997). Therefore, Koopman et al. (2008), Duffie et al. (2009) and Kim et al. (2012) formulate models for the Markov transition intensity. We pursue a similar approach and ask for a measurement error in the transition times. In linear and non-linear models, measurement error in the covariate is known to cause biased parameter estimates (see e.g. Amemiya, 1985). In our non-linear

*Sebastian Voß, Institute for Business and Social Statistics, Department of Statistics, Technische Universität Dortmund, D-44221 Dortmund, Germany, email: svoss@statistik.tu-dortmund.de, Fon: +49-231-7553869.

†Corresponding author: Rafael Weißbach, Department of Economics, Faculty of Economic and Social Sciences, University of Rostock, Ulmenstraße 69, D-18057 Rostock, Germany, email: rafael.weissbach@uni-rostock.de, Fon: +49-381-4984428.

JEL: C41, C52, G33

Keywords: Measurement error, multiple spells, rating, score

duration analysis the time (or age) acts as a covariate to the categorical dependent variable and therefore a similar effect is to be expected. Correcting for measurement error in time-discrete Markov chains has also been considered in Magnac and Visser (1999) in the context of labor market research. An example for measurement error in continuous-time labor market research can be found in van den Berg and van der Klaauw (2001), where it is referred to as a recall error.

Rating agencies like Moody's, Standard & Poor's or Fitch distinguish between 20-22 rating states. Large banks usually follow such granularity in their internal rating for comparability reasons. In either case, as the rating evolves over time, hundreds of possible combinations of a transition from an origin state to a destination state can occur. Estimating their probabilities, even in simple intensity models, exceeds available sample sizes. The bulk of rating transitions is to adjacent rating states. Under the assumption that the asset value of a debtor is a continuous-state process (Merton, 1974), rating changes of more than one state should theoretically even be impossible. Hence, we model only the intensities of transitions into an adjacent rating state, implying that rating changes of more than one state can only materialize through a series of one-state transitions. According to Kadam and Lenk (2008), the assumption of homogeneity among debtors leads to a misjudgment of a banks' credit risk. Therefore, following Koopman et al. (2008, 2009) and Kim et al. (2012), we include covariates in the model to account for debtor-specific and economic differences. Weißbach and Mollenhauer (2011) find that the transition intensity does not depend on the rating state, so that we can simplify our model by assuming that all one-state upgrades and all one-state downgrades follow the same intensity process. This reduces the number of parameters to a manageable level.

A continuous-time model requires an inspection of a debtors' rating in small periodic intervals, in order to generate suitable data for estimating the parameters of the model. According to Mählmann (2006), this is not necessarily the case in practice. If the ratings are reviewed at irregular intervals, the transition times are registered with a measurement error. Neglecting these errors leads to biased parameter estimators. In order to account for measurement errors, the intensity model must be extended by a frailty term, a latent stochastic process. One possible approach is to assume that the frailties follow a suitable parametric distribution family and to estimate the parameters of this family together with those of the model. However, as pointed out by Heckman and Singer (1984a,b), the estimators for the model parameters depend heavily on the choice of the parametric family. Note, that in discrete-time models, selecting the parametric family does not need to be an issue (see Magnac and Visser, 1999).

We propose initially testing for measurement errors in the data and construct a score test for the null hypothesis that the transition times are registered without errors. The test extends a concept from Lancaster (1990) which does not depend on a specific distributional assumption for the measurement error terms. We apply this test to a 10-year dataset of an international corporate portfolio of a large German bank and obtain a positive result.

The paper is organized as follows. In Section 2.1, we introduce a simple intensity model, which serves as our model under the null hypothesis of no measurement errors in the data. In Section 2.2, we expand the intensity to account for measurement errors and construct the score test. The proof that the test statistic asymptotically follows a χ^2 -distribution is delivered in Section 2.3. By means of Monte Carlo simulations, in Section 3, we analyze the consequences of ignoring the measurement errors, study the Taylor approximation intrinsic to the concept of Lancaster (1990), and investigate finite sample properties of the score test. Finally, we apply the test to the abovementioned rating dataset in Section 4.

2 Derivation of the test

The rating $(R_i(t), t \in [0, \tau])$ of a debtor $i \in \{1, \dots, n\}$ in the observation period $[0, \tau]$ is perceived as a continuous-time, discrete-state Markov process with finite state space $\{1, \dots, D\}$, where the lowest rating D implies bankruptcy and is considered as absorbing. Assuming that the initial rating of a debtor i does not depend on the intensity parameters, the rating process $(R_i(t), t \in [0, \tau])$ can - without loss of information - be decomposed into the right-continuous counting processes $(N_{hi}(t), t \in [0, \tau])$ and the left-continuous risk processes $(Y_{hi}(t), t \in [0, \tau])$, where the index $h \in \{\uparrow, \downarrow\}$ distinguishes between one-state upgrades (\uparrow) and one-state downgrades (\downarrow). This simple distinction is based on former model building (see Weißbach and Walter, 2010; Weißbach and Mollenhauer, 2011), but can easily be generalized to allow for transitions between any two rating states by replacing the index h by two indices denoting origin and destination state without hindering the test derivation. The process $Y_{hi}(t) \in \{0, 1\}$ takes the value 0, if the loan is paid back, if a debtor is declared bankrupt or - in the case of $h = \uparrow$ - if the debtor is in the highest possible rating state. In order to account for different macroeconomic situations or debtor-specific attributes, we include covariates in our model. The d -dimensional covariate process of debtor i is denoted by $(\mathbf{X}_i(t), t \in [0, \tau])$ and is assumed to be *cadlag* and almost surely bounded. Moreover, the covariate processes of the n debtors are considered independent and identically distributed.

2.1 Estimating the model without measurement error

We propose the following model for the intensity process of $(N_{hi}(t), t \in [0, \tau])$:

$$\lambda_{hi}(t) = Y_{hi}(t) \lambda_{0h} \exp\left(\boldsymbol{\beta}_h^\top \mathbf{X}_i(t)\right), \quad h \in \{\downarrow, \uparrow\}, \quad (1)$$

where $\boldsymbol{\beta}_h \in \mathbb{R}^d$ is the vector of covariate effects and $\lambda_{0h} \in \mathbb{R}^+$ is a constant baseline intensity. Note that $t = 0$ marks the credit origination, so that the time-axis of the model is not the calendar-time, but the age of the debtor in the banks' credit portfolio. In this way, as argued in Weißbach and Walter (2010), the processes $(N_{hi}(t), t \in [0, \tau])$, $i = 1, \dots, n$, can be considered independent, as they are not observed under similar conditions for a given t .

Let $T_{hi1}, \dots, T_{hiN_{hi}(\tau)}$ be the $N_{hi}(\tau)$ ages at one-state transitions of type h , observed for debtor i . According to Andersen et al. (1993), p. 585, the likelihood of the model (1) is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n \prod_{h \in \{\downarrow, \uparrow\}} \prod_{j=1}^{N_{hi}(\tau)} \left(\lambda_{0h} \exp\left(\boldsymbol{\beta}_h^\top \mathbf{X}_i(T_{hij})\right) \right) \exp\left(-\int_0^\tau Y_{hi}(t) \lambda_{0h} \exp\left(\boldsymbol{\beta}_h^\top \mathbf{X}_i(t)\right) dt\right) \quad (2)$$

with $\boldsymbol{\theta} = (\lambda_{0(\uparrow)}, \boldsymbol{\beta}_\uparrow^\top, \lambda_{0(\downarrow)}, \boldsymbol{\beta}_\downarrow^\top)^\top$. As usual, the parameter vector $\boldsymbol{\theta}$ can be estimated by maximizing the logarithm of $L(\boldsymbol{\theta})$. A closed form for $\hat{\boldsymbol{\theta}}$ does not exist, so that the estimation requires numeric methods, which is quite cumbersome, due to the time-varying covariates. Therefore, we decided to estimate the model parameters in two steps. We first determine $\hat{\boldsymbol{\beta}}_h$ using the partial likelihood of Andersen and Gill (1982), which arises from a factorization of (2) and is given by

$$L_h^P(\boldsymbol{\beta}_h) = \prod_{i=1}^n \prod_{j=1}^{N_{hi}(\tau)} \frac{\exp(\boldsymbol{\beta}_h^\top \mathbf{X}_i(T_{hij}))}{\sum_{k=1}^n Y_{hk}(T_{hij}) \exp(\boldsymbol{\beta}_h^\top \mathbf{X}_k(T_{hij}))}. \quad (3)$$

In the second step, an estimator for λ_{0h} is obtained by inserting $\hat{\beta}_h$ - the maximum point of (3) - into the logarithm of (2). Maximization of $\log L$ then leads to

$$\hat{\lambda}_{0h}(\hat{\beta}_h) = \frac{\sum_{i=1}^n N_{hi}(\tau)}{\sum_{i=1}^n \int_0^\tau Y_{hi}(t) \exp\left(\hat{\beta}_h^\top \mathbf{X}_i(t)\right) dt}. \quad (4)$$

Let the true parameter values of the model be denoted by $\lambda_{0h}^{(0)}$ and $\beta_h^{(0)}$. It then follows from Andersen and Gill (1982), Theorem 4.1, that $\hat{\beta}_h$ is a consistent estimator for $\beta_h^{(0)}$, if the following assumptions are fulfilled:

- (A1) $P(N_{hi}(\tau) < \infty) = 1$.
- (A2) $P(Y_{hi}(t) = 1 \forall t \in [0, \tau]) > 0$ for all $h \in \{\uparrow, \downarrow\}$, $i = 1, \dots, n$.
- (A3) The Fisher-Information matrix \mathcal{I}^P of L_h^P from (3) is positive definite.
- (A4) $\lambda_{0h}^{(0)} \in (c_h^l, c_h^u)$, with $0 < c_h^l < c_h^u < \infty$.

If, in addition,

- (A5) There exists a compact neighborhood \mathcal{B}_h around $\beta_h^{(0)}$ and bounded, deterministic functions $\mu_h(t, \beta_h)$ on $[0, \tau]$, so that

$$\sup_{t \in [0, \tau]} \left| \frac{1}{n} \sum_{i=1}^n Y_{hi}(t) \exp\left(\beta_h^\top \mathbf{X}_i(t)\right) - \mu_h(t, \beta_h) \right| \xrightarrow{\mathcal{P}} 0$$

for all $\beta_h \in \mathcal{B}_h$, with $\mu_h(t, \beta_h)$ continuous in β_h for all $t \in [0, \tau]$ and

$$\int_0^\tau \mu(t, \beta_h) dt \neq 0 \quad \forall \beta_h \in \mathcal{B}_h.$$

holds, $\hat{\lambda}_{0h}(\hat{\beta}_h)$ is a consistent estimator for $\lambda_{0h}^{(0)}$. This can be shown by first proving that $\hat{\lambda}_{0h}(\beta_h^{(0)}) \xrightarrow{\mathcal{P}} \lambda_{0h}^{(0)}$, using Martingale limit theory and the inequality of Lenglar (1977). Consistency of $\hat{\lambda}_{0h}(\hat{\beta}_h)$ then follows from the fact that $\hat{\lambda}_{0h}(\beta_h)$ converges uniformly in \mathcal{B}_h against a continuous function $\lambda_{0h}(\beta_h)$. Details are omitted here for the sake of brevity.

2.2 Measurement error model and test statistic

In order to obtain data for estimating the parameters of this continuous-time model, it is essential to check the solvency of the debtors at small periodic intervals, which may not be possible in practice. If the ratings are reviewed at irregular intervals, transition times may be registered with a delay, which means that there are measurement errors in the data. In order to account for such measurement errors, the model (1) must be extended by a frailty term, an unobservable stochastic process. Let $\mathbf{V}_h = (V_{hij}, i \in \{1, \dots, n\}, j \in \{1, \dots, N_{hi}(\tau) + 1\})$, $h \in \{\uparrow, \downarrow\}$, be two independent families of positive i.i.d. random variables with expectation 1 and let

$$W_{hi}(t) := V_{hi(N_{hi}(t-) + 1)}. \quad (5)$$

The left-continuous process $W_{hi}(t)$ is piecewise constant with random starting value V_{hi1} and takes on a new random value immediately after a transition of type h . An intensity model to account for measurement error in the transition times is then given by

$$\lambda_{hi}^*(t, W_{hi}(t)) = W_{hi}(t) \lambda_{hi}(t) \quad (6)$$

with $\lambda_{hi}(t)$ given by (1). According to this model, the measurement errors result from a data generating intensity, which alternates randomly from transition to transition around the originally assumed model $\lambda_{hi}(t)$. Model (6) can also be interpreted as mixture of Markov processes moving with different speeds, see Frydman (2005) and Frydman and Schuermann (2008). Note that the measurement error process $W_{hi}(t)$ can take values greater than 1 which increases the intensity $\lambda_{hi}(t)$. At first glance, this implies that the transition time is registered too early, while in practice, only a delay should be possible. However, as the average delay is not identifiable, it is already included in $\lambda_{hi}(t)$ and a transition time of an increased intensity is only registered too early with respect to the non-identifiable average delay.

Ignoring measurement errors in the data can have severe consequences, as we show in a Monte Carlo simulation in Section 3. Accounting for the errors however, requires detailed insight into the properties of $W_{hi}(t)$, which are difficult to determine, due to unobservability. Therefore, it makes sense to initially check for the existence of measurement errors in the data with a statistical test. We develop such a test based on Lancaster (1990), Chapter 11.2.2. It is of interest to determine whether the variance of the measurement error terms V_{hij} differs significantly from zero. If this is not the case, the process $W_{hi}(t)$ takes a constant value of 1 and $\lambda_{hi}^*(t, W_{hi}(t))$ equals the intensity $\lambda_{hi}(t)$ from (1), indicating that there are no measurement errors in the data. The likelihood of model (6) arises from (2) by replacing $\lambda_{hi}(t)$ with $\lambda_{hi}^*(t, W_{hi}(t))$. Due to the fact that it depends on the unobservable measurement error processes $W_{hi}(t)$, it is however not possible to use this likelihood for parameter inference. Hence, the test is based on the marginal likelihood of model (6), i.e. the expected value of the likelihood with respect to the terms V_{hij} . This marginal likelihood contains the variance of the V_{hij} as a parameter and is therefore suitable for testing the hypothesis of interest. Specifically, the likelihood of model (6) is replaced by an approximation, to avoid distributional assumptions on the measurement error terms. Lancaster (1990) constructs such a test for uncensored survival data, which is extended by Jaggia (1997) for censored observations. We extend their idea to counting processes and prove the asymptotic distribution of the test statistic under the null hypothesis.

As there are n debtors i and two transition types h , the likelihood of the model (6) consists of $2n$ similar factors L_{hi}^* with the full likelihood given by $\prod_{hi} L_{hi}^*$. Let $\mathbf{V}_{hi} := (V_{hij}, j \in \{1, \dots, N_{hi}(\tau) + 1\}) \subset \mathbf{V}_h$ be the underlying measurement error terms of the process $W_{hi}(t)$ and let $\boldsymbol{\theta}_h := (\lambda_{0h}, \boldsymbol{\beta}_h^\top)^\top$. Then, the hi^{th} factor of the likelihood is given by

$$\begin{aligned} L_{hi}^*(\boldsymbol{\theta}_h, \mathbf{V}_{hi}) &= \prod_{j=1}^{N_{hi}(\tau)} \lambda_{hi}^*(T_{hij}, W_{hi}(T_{hij})) \exp\left(-\int_0^\tau \lambda_{hi}^*(t, W_{hi}(t)) dt\right) \\ &= \prod_{j=1}^{N_{hi}(\tau)} \left[V_{hij} \lambda_{0h} \exp\left(\boldsymbol{\beta}_h^\top \mathbf{X}_i(T_{hij})\right) \exp\left(-V_{hij} \int_{T_{hi(j-1)}}^{T_{hij}} Y_{hi}(t) \lambda_{0h} \exp\left(\boldsymbol{\beta}_h^\top \mathbf{X}_i(t)\right) dt\right) \right] \\ &\quad \times \exp\left(-V_{hi(N_{hi}(\tau)+1)} \int_{T_{hiN_{hi}(\tau)}}^\tau Y_{hi}(t) \lambda_{0h} \exp\left(\boldsymbol{\beta}_h^\top \mathbf{X}_i(t)\right) dt\right) \end{aligned} \quad (7)$$

with $T_{hi0} := 0$. So L_{hi}^* decomposes into $N_{hi}(\tau) + 1$ factors $L_{hi1}^*, \dots, L_{hi(N_{hi}(\tau)+1)}^*$ - one for each transition of type h and one for the time between the last observed transition and τ .

In order to avoid a distributional assumption on the V_{hij} when determining the expected value of the likelihood, the $N_{hi}(\tau) + 1$ factors of L_{hi}^* are replaced by a Taylor approximation of order two around $E(V_{hij}) = 1$. This serves as a reasonable approximation, if the variance of the measurement error terms V_{hij}

is small, which implies that they usually take values close to 1. Let

$$z_{hij}(\boldsymbol{\theta}_h) := \int_{T_{hi(j-1)}}^{T_{hij}} Y_{hi}(t) \lambda_{0h} \exp\left(\boldsymbol{\beta}_h^\top \mathbf{X}_i(t)\right) dt \quad (8)$$

for $j = 1, \dots, N_{hi}(\tau) + 1$ with $T_{hi(N_{hi}(\tau)+1)} := \tau$. Then the approximation for L_{hij}^* , $j = 1, \dots, \dots, N_{hi}(\tau)$, is given by

$$\begin{aligned} \tilde{L}_{hij}^*(\boldsymbol{\theta}_h, V_{hij}) &:= \lambda_{0h} \exp\left(\boldsymbol{\beta}_h^\top \mathbf{X}_i(T_{hij})\right) \exp(-z_{hij}(\boldsymbol{\theta}_h)) \\ &\times \left(1 - (V_{hij} - 1)(z_{hij}(\boldsymbol{\theta}_h) - 1) + \frac{(V_{hij} - 1)^2}{2} (z_{hij}^2(\boldsymbol{\theta}_h) - 2z_{hij}(\boldsymbol{\theta}_h))\right), \end{aligned} \quad (9)$$

while the approximation for $L_{hi(N_{hi}(\tau)+1)}^*$ becomes

$$\begin{aligned} \tilde{L}_{hi(N_{hi}(\tau)+1)}^*(\boldsymbol{\theta}_h, V_{hi(N_{hi}(\tau)+1)}) &:= \exp(-z_{hi(N_{hi}(\tau)+1)}(\boldsymbol{\theta}_h)) \\ &\times \left(1 - (V_{hi(N_{hi}(\tau)+1)} - 1)z_{hi(N_{hi}(\tau)+1)}(\boldsymbol{\theta}_h)\right. \\ &\left.+ \frac{(V_{hi(N_{hi}(\tau)+1)} - 1)^2}{2} z_{hi(N_{hi}(\tau)+1)}^2(\boldsymbol{\theta}_h)\right). \end{aligned} \quad (10)$$

Due to the independence of the V_{hij} , $E(L^*) = \prod_{hij} E(L_{hij})$ holds and an approximation of the marginal likelihood is given by

$$\begin{aligned} \bar{L}(\boldsymbol{\theta}^*) &:= \prod_{i=1}^n \prod_{h \in \{\uparrow, \downarrow\}} \prod_{j=1}^{N_{hi}(\tau)+1} E(\tilde{L}_{hij}^*) \\ &= L(\boldsymbol{\theta}) \prod_{i=1}^n \prod_{h \in \{\uparrow, \downarrow\}} \left(\prod_{j=1}^{N_{hi}(\tau)} \left[1 + \eta_h (z_{hij}^2(\boldsymbol{\theta}_h) - 2z_{hij}(\boldsymbol{\theta}_h))\right] \right) \left[1 + \eta_h z_{hi(N_{hi}(\tau)+1)}^2(\boldsymbol{\theta}_h)\right] \end{aligned} \quad (11)$$

with $\eta_h := \text{Var}(V_{hij})/2$, $\boldsymbol{\theta}^* := (\boldsymbol{\theta}_\uparrow^\top, \eta_\uparrow, \boldsymbol{\theta}_\downarrow^\top, \eta_\downarrow)^\top$ and $L(\boldsymbol{\theta})$ given by (2). The likelihood \bar{L} depends on the original parameters $\boldsymbol{\theta}$ and the measurement error variance parameters η_\uparrow and η_\downarrow . At first glance, it seems that \bar{L} provides a universal and easy way to account for measurement errors. It is, however, not useful for estimating $\boldsymbol{\theta}$ together with $\boldsymbol{\eta} := (\eta_\uparrow, \eta_\downarrow)$, as the likelihood can take negative values for $\eta_h > 1$. It is, however, a suitable foundation for testing the null hypothesis ' $H_0 : \eta_\uparrow = \eta_\downarrow = 0$ ', because \bar{L} is a reasonable approximation of the real marginal likelihood of model (6) if the variance of the V_{hij} is small, i.e. close to H_0 . Under the null hypothesis, the value of $\boldsymbol{\eta}$ is on the boundary of the parameter space. As a consequence, the ML-estimator does not exhibit the usual properties (see Godfrey, 1990), such as asymptotic normality. Test procedures, like the likelihood ratio- or Wald-test should therefore be avoided, as the distribution of these tests depend on the asymptotic normality of the maximum likelihood estimate, derived by the asymptotic normality of the score vector and a Taylor expansion (see e.g. Weißbach and Walter, 2010, Formula 10). If the Taylor expansion fails, resorting to the score vector for test construction seems natural. This test procedure requires only the estimation of $\boldsymbol{\theta}$, while $\boldsymbol{\eta}$ is determined by H_0 . The basic idea of the score test is that the expected value of the score vector - i.e. the first derivative of the log likelihood - disappears at the true parameter values, when H_0 is valid, implying that H_0 can be rejected, if the difference of the score vector evaluated at $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\eta} = 0$ to zero is 'too big'. We show in Section 2.3, that this remains true in our model, even though $\boldsymbol{\eta}$ is on the boundary of the parameter space.

According to Rao (1948), the test statistic of the score test for ‘ $H_0 : \eta_{\uparrow} = \eta_{\downarrow} = 0$ ’ is based on the score vector of $\boldsymbol{\eta}$, evaluated at the estimator for $\boldsymbol{\theta}^*$ under H_0 , i.e.

$$\bar{U}_{\boldsymbol{\eta}}(\hat{\boldsymbol{\theta}}) := \frac{\partial}{\partial \boldsymbol{\eta}} \log \bar{L}(\boldsymbol{\theta}^*) \Big|_{\boldsymbol{\theta}^* = (\hat{\boldsymbol{\theta}}_{\uparrow}^{\top}, 0, \hat{\boldsymbol{\theta}}_{\downarrow}^{\top}, 0)^{\top}}, \quad (12)$$

with $\hat{\boldsymbol{\theta}} := (\hat{\boldsymbol{\theta}}_{\uparrow}^{\top}, \hat{\boldsymbol{\theta}}_{\downarrow}^{\top}) = (\hat{\lambda}_{0\uparrow}, \hat{\boldsymbol{\beta}}_{\uparrow}^{\top}, \hat{\lambda}_{0\downarrow}, \hat{\boldsymbol{\beta}}_{\downarrow}^{\top})^{\top}$ denoting the two-step-ML-estimators of model (1). The two components of (12) are given by

$$\bar{U}_{\eta_h}(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \left(\sum_{j=1}^{N_{hi}(\tau)} \left(z_{hij}^2(\hat{\boldsymbol{\theta}}_h) - 2z_{hij}(\hat{\boldsymbol{\theta}}_h) \right) + z_{hi(N_{hi}(\tau)+1)}^2(\hat{\boldsymbol{\theta}}_h) \right), \quad h \in \{\uparrow, \downarrow\}. \quad (13)$$

The test statistic also depends on the estimated Fisher information matrix of \bar{L}_i - the i^{th} factor of \bar{L} in (11) - which is given by

$$\hat{\boldsymbol{I}}(\hat{\boldsymbol{\theta}}) = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \boldsymbol{\theta}^* (\partial \boldsymbol{\theta}^*)^{\top}} \log \bar{L}_i(\boldsymbol{\theta}^*) \Big|_{(\hat{\boldsymbol{\theta}}_{\uparrow}^{\top}, 0, \hat{\boldsymbol{\theta}}_{\downarrow}^{\top}, 0)^{\top}}. \quad (14)$$

Here and onwards ‘ $\boldsymbol{\theta}^* =$ ’ is suppressed in the argument for which the derivative is to be evaluated. Due to the fact that \bar{L}_i consists of an upgrade- and a downgrade-factor, the second derivatives with respect to one parameter of each transition type disappear. The components of $\hat{\boldsymbol{I}}(\hat{\boldsymbol{\theta}})$ that differ from zero are given by

$$\begin{aligned} \frac{\partial^2}{(\partial \eta_h)^2} \log \bar{L}_i(\boldsymbol{\theta}^*) \Big|_{(\hat{\boldsymbol{\theta}}_{\uparrow}^{\top}, 0, \hat{\boldsymbol{\theta}}_{\downarrow}^{\top}, 0)^{\top}} &= - \sum_{j=1}^{N_{hi}(\tau)} \left(z_{hij}^2(\hat{\boldsymbol{\theta}}_h) - 2z_{hij}(\hat{\boldsymbol{\theta}}_h) \right)^2 - z_{hi(N_{hi}(\tau)+1)}^4(\hat{\boldsymbol{\theta}}_h) \\ \frac{\partial^2}{(\partial \lambda_{0h})^2} \log \bar{L}_i(\boldsymbol{\theta}^*) \Big|_{(\hat{\boldsymbol{\theta}}_{\uparrow}^{\top}, 0, \hat{\boldsymbol{\theta}}_{\downarrow}^{\top}, 0)^{\top}} &= - \frac{N_{hi}(\tau)}{\hat{\lambda}_{0h}^2} \\ \frac{\partial^2}{\partial \beta_{hi} \partial \beta_{hi'}} \log \bar{L}_i(\boldsymbol{\theta}^*) \Big|_{(\hat{\boldsymbol{\theta}}_{\uparrow}^{\top}, 0, \hat{\boldsymbol{\theta}}_{\downarrow}^{\top}, 0)^{\top}} &= - \hat{\lambda}_{0h} \int_0^{\tau} Y_{hi}(t) X_{il}(t) X_{il'}(t) \exp \left(\hat{\boldsymbol{\beta}}_h^{\top} \mathbf{X}_i(t) \right) dt \\ \frac{\partial^2}{\partial \eta_h \partial \lambda_{0h}} \log \bar{L}_i(\boldsymbol{\theta}^*) \Big|_{(\hat{\boldsymbol{\theta}}_{\uparrow}^{\top}, 0, \hat{\boldsymbol{\theta}}_{\downarrow}^{\top}, 0)^{\top}} &= \sum_{j=1}^{N_{hi}(\tau)} \left(\frac{2}{\hat{\lambda}_{0h}} \left(z_{hij}^2(\hat{\boldsymbol{\theta}}_h) - z_{hij}(\hat{\boldsymbol{\theta}}_h) \right) \right) + \frac{2}{\hat{\lambda}_{0h}} z_{hi(N_{hi}(\tau)+1)}^2(\hat{\boldsymbol{\theta}}_h) \\ \frac{\partial^2}{\partial \eta_h \partial \beta_{hl}} \log \bar{L}_i(\boldsymbol{\theta}^*) \Big|_{(\hat{\boldsymbol{\theta}}_{\uparrow}^{\top}, 0, \hat{\boldsymbol{\theta}}_{\downarrow}^{\top}, 0)^{\top}} &= \sum_{j=1}^{N_{hi}(\tau)} \left(2 \left(z_{hij}(\hat{\boldsymbol{\theta}}_h) - 1 \right) \frac{\partial}{\partial \beta_{hl}} z_{hij}(\boldsymbol{\theta}_h) \Big|_{\hat{\boldsymbol{\theta}}_h} \right) \\ &\quad + 2z_{hi(N_{hi}(\tau)+1)}(\hat{\boldsymbol{\theta}}_h) \frac{\partial}{\partial \beta_{hl}} z_{hi(N_{hi}(\tau)+1)}(\boldsymbol{\theta}_h) \Big|_{\hat{\boldsymbol{\theta}}_h} \\ \frac{\partial^2}{\partial \lambda_{0h} \partial \beta_{hl}} \log \bar{L}_i(\boldsymbol{\theta}^*) \Big|_{(\hat{\boldsymbol{\theta}}_{\uparrow}^{\top}, 0, \hat{\boldsymbol{\theta}}_{\downarrow}^{\top}, 0)^{\top}} &= - \int_0^{\tau} Y_{hi}(t) X_{il}(t) \exp \left(\hat{\boldsymbol{\beta}}_h^{\top} \mathbf{X}_i(t) \right) dt, \end{aligned} \quad (15)$$

with

$$\frac{\partial}{\partial \beta_{hl}} z_{hij}(\boldsymbol{\theta}_h) \Big|_{\hat{\boldsymbol{\theta}}_h} = \hat{\lambda}_{0h} \int_{T_{hi(j-1)}}^{T_{hi j}} Y_{hi}(t) X_{il}(t) \exp \left(\hat{\boldsymbol{\beta}}_h^{\top} \mathbf{X}_i(t) \right) dt \quad (16)$$

for $j \in \{1, \dots, N_{hi}(\tau) + 1\}$. Let $\boldsymbol{\lambda}_0 := (\lambda_{0\uparrow}, \lambda_{0\downarrow})^{\top}$ and

$$\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}}) := \hat{\boldsymbol{I}}_{-\boldsymbol{\lambda}_0 - \boldsymbol{\lambda}_0}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{I}}_{-\boldsymbol{\lambda}_0 \boldsymbol{\lambda}_0}(\hat{\boldsymbol{\theta}}) \hat{\boldsymbol{I}}_{\boldsymbol{\lambda}_0 \boldsymbol{\lambda}_0}^{-1}(\hat{\boldsymbol{\theta}}) \hat{\boldsymbol{I}}_{\boldsymbol{\lambda}_0 - \boldsymbol{\lambda}_0}(\hat{\boldsymbol{\theta}}), \quad (17)$$

with $\hat{\mathcal{I}}_{\pm\lambda_0 \pm\lambda_0}$ denoting that the rows respectively columns of $\hat{\mathcal{I}} \in \mathbb{R}^{2(d+2) \times 2(d+2)}$, which belong to the parameter λ_0 are retained (+) or deleted (-), i.e. $\hat{\Sigma}(\hat{\theta})$ is the Schur complement of the partitioned matrix $\hat{\mathcal{I}}(\hat{\theta})$. Furthermore, let $\hat{\Sigma}_{\eta\eta} \in \mathbb{R}^{2 \times 2}$ be the elements of $\hat{\Sigma} \in \mathbb{R}^{2(d+1) \times 2(d+1)}$, which belong to η . We show in the next section that under H_0 and some mild regularity conditions

$$\frac{1}{n} \bar{U}_{\eta}^{\top}(\hat{\theta}) \hat{\Sigma}_{\eta\eta}^{-1}(\hat{\theta}) \bar{U}_{\eta}(\hat{\theta}) \xrightarrow{D} \chi_2^2. \quad (18)$$

2.3 Asymptotic distribution of the test statistic

In order to prove that (18) holds, we need to show that under the null hypothesis ' $\eta_{\uparrow} = \eta_{\downarrow} = 0$ ', the score vector converges uniformly to a normally distributed random variable with an expected value of 0 and the Fisher-Information as the covariance matrix. When dealing with counting processes, this is usually done by using the fact that the score vector of the likelihood is a Martingale. Due to the fact the \bar{L} emerges from an approximation of the counting process likelihood, it is however no longer clear whether this property holds. We show that \bar{L} is essentially a probability density function of the durations between two transitions, which enables us to use density properties to calculate the expected value and covariance matrix. The uniform convergence to the normal distribution then follows according to a theorem of Donsker.

In addition to assumptions **(A1)**-**(A5)**, which ensure the consistency of the two-step-ML-estimator, some further regularity assumptions are required:

(A6) The durations over which the debtors are observed are realizations of the i.i.d. random variables C_1, \dots, C_n . The distribution- and density function of the censoring time C_i , $i = 1, \dots, n$, are denoted by G and g respectively and do not depend on the parameters of model (6). Furthermore, the censoring times are independent of the transition times.

(A7) Let

$$\bar{U}_{hi}(\theta_h) := \left. \frac{\partial}{\partial \theta^*} \log \bar{L}_{hi}(\theta^*) \right|_{(\theta_h^{\top}, 0)}, \quad (19)$$

with \bar{L}_{hi} denoting the factor of \bar{L} that belongs to debtor i and transition type h , i.e.

$$\begin{aligned} \bar{L}_{hi}(\theta_h^*) &= \left(\prod_{j=1}^{N_{hi}(\tau)} \left[\lambda_{0h} \exp(\beta_h^{\top} \mathbf{X}_i(T_{hij})) \right] \exp \left(- \int_0^{\tau} Y_{hi}(t) \lambda_{0h} \exp(\beta_h^{\top} \mathbf{X}_i(t)) \right) \right) \\ &\quad \times \left[1 + \eta_h z_{hi}^2(N_{hi}(\tau)+1)(\theta_h) \right] \left[1 + \eta_h (z_{hij}^2(\theta_h) - 2z_{hij}(\theta_h)) \right]. \end{aligned} \quad (20)$$

Then, $\bar{U}_{\uparrow i}(\theta_{\uparrow}^{(0)})$ and $\bar{U}_{\downarrow i}(\theta_{\downarrow}^{(0)})$ are uncorrelated for all $i = 1, \dots, n$.

(A8) There exists a compact neighborhood Θ around $\theta^{(0)} := (\theta_{\uparrow}^{(0)\top}, \theta_{\downarrow}^{(0)\top})^{\top}$, so that $\mathcal{I}(\theta) := \text{Cov}(\bar{U}_i(\theta))$ with $\bar{U}_i(\theta) = (\bar{U}_{\uparrow i}^{\top}(\theta_{\uparrow}), \bar{U}_{\downarrow i}^{\top}(\theta_{\downarrow}))^{\top}$ is positive definite for all $\theta \in \Theta$.

(A9) $\hat{\mathcal{I}}(\hat{\theta})$ from equation (14) is positive definite.

Because the starting points of the rating histories are distributed randomly across the observation period of length τ , the time horizon over which a certain debtor is observed is a random variable. Assumption **(A6)** ensures that the censoring times of the debtors do not depend on the parameters of λ_{0h} and β_h . Assumption

(A7) allows us to examine upgrades and downgrades separately in the following proofs. As the upgrade- and downgrade factor of the likelihood \bar{L} do not contain any common parameters, the upgrade factor does not contribute anything to the estimation of the downgrade parameters and vice versa. Hence, the upgrade- and downgrade components of the score vector are uncorrelated. However, the formal proof of this property is very technical. Therefore, we decided to allocate it to the regularity conditions. Finally, assumptions (A8) and (A9) ensure the invertibility of the Fisher-Information matrix and its estimator.

In order to show that (18) holds, we need the following Lemmas. The proofs are presented in the Appendix.

Lemma 1 *Assume that assumption (A6) holds and ‘ $H_0 : \eta_{\uparrow} = \eta_{\downarrow} = 0$ ’ is valid. Then, $E[\bar{U}_{hi}(\boldsymbol{\theta}_h^{(o)})] = 0$ for $i = 1, \dots, n$ and $h \in \{\uparrow, \downarrow\}$ with $\boldsymbol{\theta}_h^{(o)} = (\lambda_{0h}^{(o)}, \boldsymbol{\beta}_h^{(o)\top})^\top$.*

This result is essential for the score test to work, even though under H_0 , the parameter of interest is on the boundary of the parameter space. In the appendix, we show that \bar{L} is essentially a probability density function, which allows us to prove the result, using standard density techniques. An alternative approach would be to use the fact that the z_{hij} (see (8)) conditional on $\mathbf{X}_i(t)$ and $Y_{hi}(t)$ are standard exponentially distributed (see e.g. Cook and Lawless, 2007, p. 92). Knowing this, it can easily be seen that the summands of $\bar{U}_{hi}(\boldsymbol{\theta}_h^{(o)})$ have an expected value of zero. This approach, however, does not help with the proof of Lemma 2, which is the reason we decided against it.

Lemma 2 *Under the conditions of Lemma 1*

$$\text{Cov}[\bar{U}_{hi}(\boldsymbol{\theta}_h^{(o)})] = -E \left[\left(\frac{\partial^2}{\partial \boldsymbol{\theta}_h^* (\partial \boldsymbol{\theta}_h^*)^\top} \log \bar{L}_{hi}(\boldsymbol{\theta}_h^*) \right)_{(\boldsymbol{\theta}_h^{(o)\top}, 0)^\top} \right] =: \mathcal{I}_h(\boldsymbol{\theta}_h^{(o)})$$

for $i = 1, \dots, n$ and $h \in \{\uparrow, \downarrow\}$.

Lemma 3 *Suppose assumptions (A6) and (A7) are satisfied. Let Θ be a compact neighborhood of $\boldsymbol{\theta}^{(o)} := (\boldsymbol{\theta}_{\uparrow}^{(o)\top}, \boldsymbol{\theta}_{\downarrow}^{(o)\top})^\top$ and $\bar{\mathbf{u}}(\boldsymbol{\theta}) := E[\bar{U}_i(\boldsymbol{\theta})]$. Then,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\bar{U}_i(\boldsymbol{\theta}) - \bar{\mathbf{u}}(\boldsymbol{\theta})) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}))$$

with $\xrightarrow{\mathcal{D}}$ denoting uniform convergence in distribution for all $\boldsymbol{\theta} \in \Theta$ (see Borovkov (1998), p. 220). In particular, this implies that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{U}_i(\boldsymbol{\theta}^{(o)}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}^{(o)})), \quad (21)$$

with

$$\mathcal{I}(\boldsymbol{\theta}^{(o)}) = \begin{pmatrix} \mathcal{I}_{\uparrow}(\boldsymbol{\theta}_{\uparrow}^{(o)}) & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_{\downarrow}(\boldsymbol{\theta}_{\downarrow}^{(o)}) \end{pmatrix}.$$

Lemma 4 *Suppose assumptions (A1)-(A6) hold and ‘ $H_0 : \eta_{\uparrow} = \eta_{\downarrow} = 0$ ’ is valid. Then*

$$\hat{\mathcal{I}}(\hat{\boldsymbol{\theta}}) = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \boldsymbol{\theta}^* (\partial \boldsymbol{\theta}^*)^\top} \log \bar{L}_i(\boldsymbol{\theta}^*) \Big|_{(\hat{\boldsymbol{\theta}}_{\uparrow}^\top, 0, \hat{\boldsymbol{\theta}}_{\downarrow}^\top, 0)} \xrightarrow{\mathcal{P}} \mathcal{I}(\boldsymbol{\theta}^{(o)}),$$

i.e. $\hat{\mathcal{I}}(\hat{\boldsymbol{\theta}})$ is a consistent estimator of $\mathcal{I}(\boldsymbol{\theta}^{(o)})$.

Obviously, if $\boldsymbol{\eta} = 0$, the derivatives of the log-likelihood of models (1) and (6) with respect to $\boldsymbol{\theta}$ are equal. As $\hat{\lambda}_{0h}$ is the ML estimator of model (1) and hence, $\hat{\boldsymbol{\theta}}_h$ is the root of $\partial/\partial\lambda_{0h} \log L$, $\bar{\mathbf{U}}_{\lambda_0} := \partial/\partial\lambda_0 \log \bar{L}|_{\boldsymbol{\eta}=0}$ disappears, when replacing $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$, which means that there is a functional relationship between $\hat{\boldsymbol{\theta}}$ and $\bar{\mathbf{U}}_{\lambda_0}$. Although the test statistic (18) does not depend directly on $\bar{\mathbf{U}}_{\lambda_0}$, there is a connection, due to the fact that $\bar{\mathbf{U}}_{\lambda_0}$ and the other components of $\bar{\mathbf{U}}(\boldsymbol{\theta}) := \partial/\partial\boldsymbol{\theta}^* \log \bar{L}(\boldsymbol{\theta}^*)|_{\boldsymbol{\eta}=0}$, denoted by $\bar{\mathbf{U}}_{-\lambda_0}$, are not uncorrelated. Hence, the test statistic is not based on $\bar{\mathbf{U}}_{-\lambda_0}$, but on a modification, which does not depend on $\bar{\mathbf{U}}_{\lambda_0}$.

Lemma 5 *Suppose, assumptions (A1)-(A6) are satisfied. Let $\boldsymbol{\theta}$ be an element of Θ from Lemma 3. Then, $\bar{\mathbf{U}}_{\lambda_0}(\boldsymbol{\theta})$ and*

$$\bar{\mathbf{U}}_{-\lambda_0}^*(\boldsymbol{\theta}) = \bar{\mathbf{U}}_{-\lambda_0}(\boldsymbol{\theta}) - \mathcal{I}_{-\lambda_0\lambda_0}(\boldsymbol{\theta})\mathcal{I}_{\lambda_0\lambda_0}^{-1}(\boldsymbol{\theta})\bar{\mathbf{U}}_{\lambda_0}(\boldsymbol{\theta})$$

are uncorrelated.

By means of the properties presented in Lemmas 1-5, we are now able to prove the asymptotic distribution of the test statistic.

Theorem 1 *Suppose, assumptions (A1)-(A9) hold and ' $H_0 : \eta_{\uparrow} = \eta_{\downarrow} = 0$ ' is valid. Let*

$$\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}}) := \hat{\mathcal{I}}_{-\lambda_0-\lambda_0}(\hat{\boldsymbol{\theta}}) - \hat{\mathcal{I}}_{-\lambda_0\lambda_0}(\hat{\boldsymbol{\theta}})\hat{\mathcal{I}}_{\lambda_0\lambda_0}^{-1}(\hat{\boldsymbol{\theta}})\hat{\mathcal{I}}_{\lambda_0-\lambda_0}(\hat{\boldsymbol{\theta}})$$

and let the rows and columns of $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}}) \in \mathbb{R}^{2(d+1) \times 2(d+1)}$ that belong to the second derivatives of $\boldsymbol{\eta}$ in $\hat{\mathcal{I}}_{-\lambda_0-\lambda_0}(\hat{\boldsymbol{\theta}})$ be denoted by $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\hat{\boldsymbol{\theta}}) \in \mathbb{R}^{2 \times 2}$. Then

$$\frac{1}{n}\bar{\mathbf{U}}_{\boldsymbol{\eta}}^{\top}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(\hat{\boldsymbol{\theta}})\bar{\mathbf{U}}_{\boldsymbol{\eta}}(\hat{\boldsymbol{\theta}}) \xrightarrow{\mathcal{D}} \chi_2^2.$$

Proof: Let $\boldsymbol{\theta} \in \Theta$ and $\bar{\mathbf{u}}_{-\lambda_0}^*(\boldsymbol{\theta}) := \mathbb{E}[\bar{\mathbf{U}}_{-\lambda_0}^*(\boldsymbol{\theta})]$. Because $\bar{\mathbf{U}}_{-\lambda_0}^*$ is a linear transformation of $\bar{\mathbf{U}}$, it follows with Lemma 3 that

$$\frac{1}{\sqrt{n}}(\bar{\mathbf{U}}_{-\lambda_0}^*(\boldsymbol{\theta}) - \bar{\mathbf{u}}_{-\lambda_0}^*(\boldsymbol{\theta})) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$$

with

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathcal{I}_{-\lambda_0-\lambda_0}(\boldsymbol{\theta}) - \mathcal{I}_{-\lambda_0\lambda_0}(\boldsymbol{\theta})\mathcal{I}_{\lambda_0\lambda_0}^{-1}(\boldsymbol{\theta})\mathcal{I}_{\lambda_0-\lambda_0}(\boldsymbol{\theta}).$$

Let the components of $\bar{\mathbf{U}}_{-\lambda_0}^*$ that belong to $\boldsymbol{\eta}$ be denoted by $\bar{\mathbf{U}}_{\boldsymbol{\eta}}^*$ with expected value $\bar{\mathbf{u}}_{\boldsymbol{\eta}}^*$. Then

$$\frac{1}{\sqrt{n}}(\bar{\mathbf{U}}_{\boldsymbol{\eta}}^*(\boldsymbol{\theta}) - \bar{\mathbf{u}}_{\boldsymbol{\eta}}^*(\boldsymbol{\theta})) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\theta})),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\theta}) \in \mathbb{R}^{2 \times 2}$ are the rows and columns of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, which belong to $\bar{\mathbf{U}}_{\boldsymbol{\eta}}^*$. With (A8), it follows that

$$\frac{1}{\sqrt{n}}\boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}(\boldsymbol{\theta})(\bar{\mathbf{U}}_{\boldsymbol{\eta}}^*(\boldsymbol{\theta}) - \bar{\mathbf{u}}_{\boldsymbol{\eta}}^*(\boldsymbol{\theta})) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbb{I}_2). \quad (22)$$

If H_0 is true, Lemma 1 yields

$$\bar{\mathbf{u}}_{-\lambda_0}^*(\boldsymbol{\theta}^{(0)}) = \mathbb{E}[\bar{\mathbf{U}}_{-\lambda_0}(\boldsymbol{\theta}^{(0)})] - \mathcal{I}_{-\lambda_0\lambda_0}(\boldsymbol{\theta}^{(0)})\mathcal{I}_{\lambda_0\lambda_0}^{-1}(\boldsymbol{\theta}^{(0)})\mathbb{E}[\bar{\mathbf{U}}_{\lambda_0}(\boldsymbol{\theta}^{(0)})] = 0$$

and hence, it follows with (22), that

$$\frac{1}{\sqrt{n}}\boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1/2}(\boldsymbol{\theta}^{(0)})\bar{\mathbf{U}}_{\boldsymbol{\eta}}^*(\boldsymbol{\theta}^{(0)}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbb{I}_2). \quad (23)$$

Equation (23) remains valid, if $\theta^{(0)}$ in $\bar{U}_\eta^*(\theta^{(0)})$ is replaced by $\hat{\theta}$ - a consistent estimator under H_0 - because of the uniform convergence stated in equation (22) (see Davidson, 1994, Theorem 21.6) and the fact that $\bar{U}_\eta^*(\theta^{(0)})$ and $\bar{U}_{\lambda_0}(\theta^{(0)})$ are uncorrelated (Lemma 5). As argued in the paragraph above Lemma 5, there is a functional relationship between $\hat{\theta}$ and \bar{U}_{λ_0} which also has an influence on \bar{U}_η , as \bar{U}_{λ_0} and \bar{U}_η are correlated. Hence, it is essential to use the uncorrelated modification \bar{U}_η^* in the test statistic. Because of $\bar{U}_{\lambda_0}(\hat{\theta}) = 0$, it follows directly from the definition, that $\bar{U}_{-\lambda_0}^*(\hat{\theta}) = \bar{U}_{-\lambda_0}(\hat{\theta})$ and therefore, that $\bar{U}_\eta^*(\hat{\theta}) = \bar{U}_\eta(\hat{\theta})$. Consequently,

$$\frac{1}{\sqrt{n}} \Sigma_{\eta\eta}^{-1/2}(\theta^{(0)}) \bar{U}_\eta(\hat{\theta}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbb{I}_2). \quad (24)$$

Lemma 4, together with **(A9)** and the Continuous Mapping Theorem yield $\hat{\Sigma}_{\eta\eta}^{-1/2}(\hat{\theta}) \xrightarrow{\mathcal{P}} \Sigma_{\eta\eta}^{-1/2}(\theta^{(0)})$. Finally, it results from Slutsky's Theorem, that

$$\frac{1}{\sqrt{n}} \hat{\Sigma}_{\eta\eta}^{-1/2}(\hat{\theta}) \bar{U}_\eta(\hat{\theta}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbb{I}_2). \quad (25)$$

Consequently, the scalar product of the statistic in (25) follows a χ_2^2 -distribution. \square

3 Monte Carlo simulations

Three open topics are feasible for investigating in Monte Carlo simulations. (i) Is it at all necessary to estimate θ differently in the presence of measurement errors. (ii) Is the approximation (11) of the marginal likelihood useful for testing $\eta \neq 0$. (iii) Which sample size is necessary to meet the nominal level of the test and achieve an adequate power. In order to investigate these topics, we simulate rating histories based on a rating system with 21 states using model (6) with gamma-distributed measurement error terms V_{hij} and three covariates of different type, while the origin rating at age onset is assigned randomly. The model parameters are given by $\theta_\downarrow = (0.3, 1, 1.5, -1)^\top$ and $\theta_\uparrow = (0.3, -1, 1.5, 1)^\top$, while the observation lengths are determined by Exp(0.4)-distributed random variables with a maximum length of $\tau = 10$. Each of the simulations is based on 5000 iterations.

i) Neglecting measurement errors: In order to demonstrate the need to account for measurement errors in the intensity model, we depict the consequences of neglecting these errors and using model (1), when in fact model (6) is the true one. Therefore, we simulate $n = 250$ rating histories with measurement errors of varying magnitude and use the two step ML estimator to determine the model parameters. The results for the parameters of the downgrade intensity are shown in Figure 1. It becomes evident that ignoring measurement errors in the data leads to a bias of the parameter estimator towards zero. Moreover, the covariate effects are estimated with a higher variance. Both effects increase with the variance of the measurement error terms. The effect on the estimators of the upgrade intensity is obviously the same, as the upgrade and downgrade model have the same structure.

Ridder and Verbakel (1983) analyzed the consequences of omitted covariates in a single spell model and observed a similar effect on the parameter estimators. Therefore, we investigate the impact of this form of latent heterogeneity in our model as well. Omitted covariates $\tilde{\mathbf{X}}_i$, which are constant over time, are included in our model by replacing the time-dependent measurement error processes $W_{hi}(t)$ in (6), with the random variables $V_{hi} := \exp(\tilde{\beta}_h^\top \tilde{\mathbf{X}}_i)$. We once again generated $n = 250$ rating histories using this model with the V_{hi} following a gamma distribution, while the parameters are estimated under the assumption of

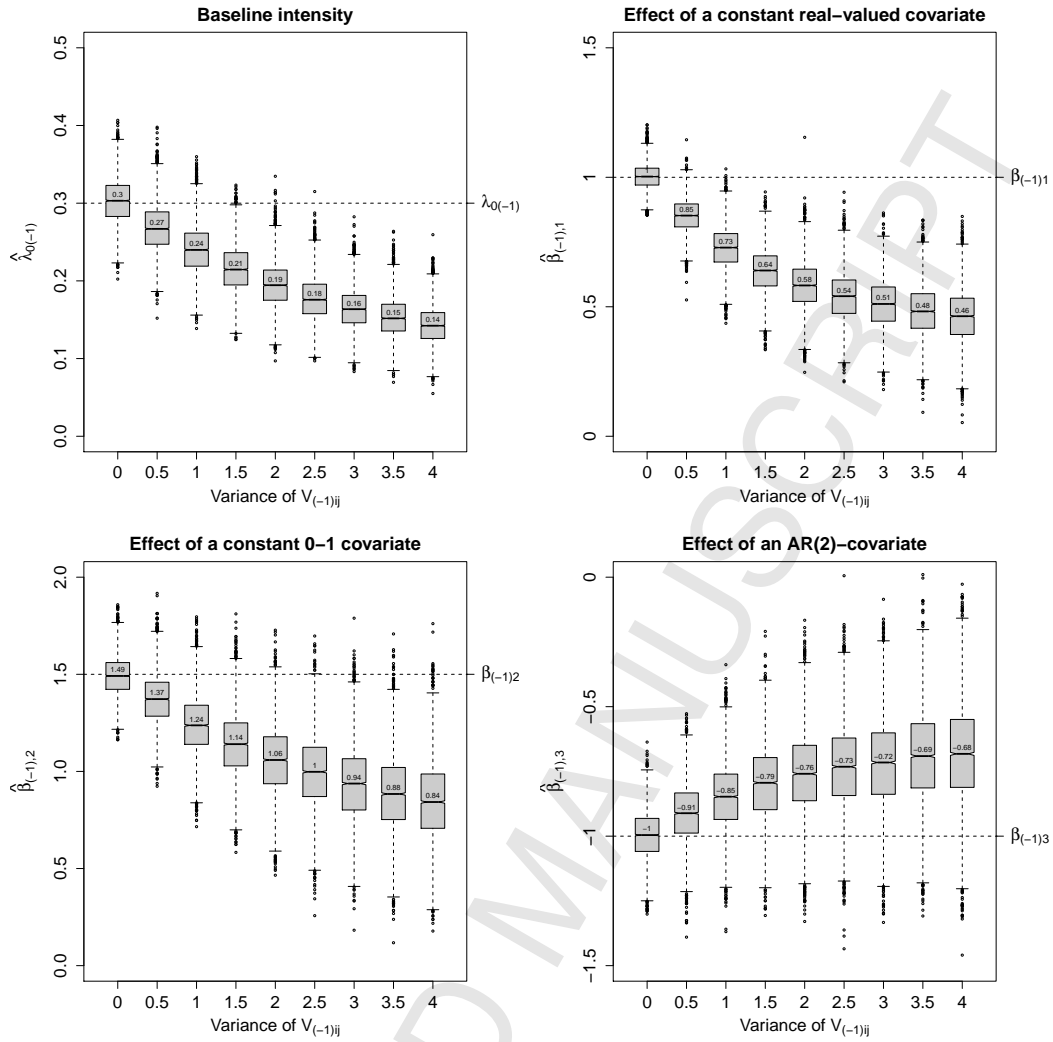


Figure 1: Estimation of the downgrade intensity parameters ignoring measurement errors

model (1). The results are shown in Figure 2 and are to some extent similar to the results with ignored measurement errors as the estimated covariate effects are also biased towards zero. However, the bias of the estimated baseline intensity is much smaller than in the situation of ignored measurement errors and for omitted covariates with small variances, λ_{0j} is even overestimated. On the other hand, the variance of the parameter estimators seems to be more sensitive to omitted covariates than to ignored measurement errors.

Interestingly, in their empirical study, Duffie et al. (2009) find a considerable difference in the regression constant with frailty, compared to without frailty. In contrast, for the other regression coefficients, the differences are not so pronounced. Here, find that the estimate of the baseline hazard is affected more by the measurement error than by a neglected covariate. Still models are different, because their endpoint is default, and fortunately, the timing can be determined quite precisely for legal aspects. Rating acts as a regressor, and their frailty models the unfortunately obvious cross-sectional dependence. Our endpoint is rating transition (be it to the default state or any other) and the frailty models the unfortunately possible measurement error. Fortunately, cross-sectional dependence is very unlikely or small in our study, due to the measurement of portfolio age.

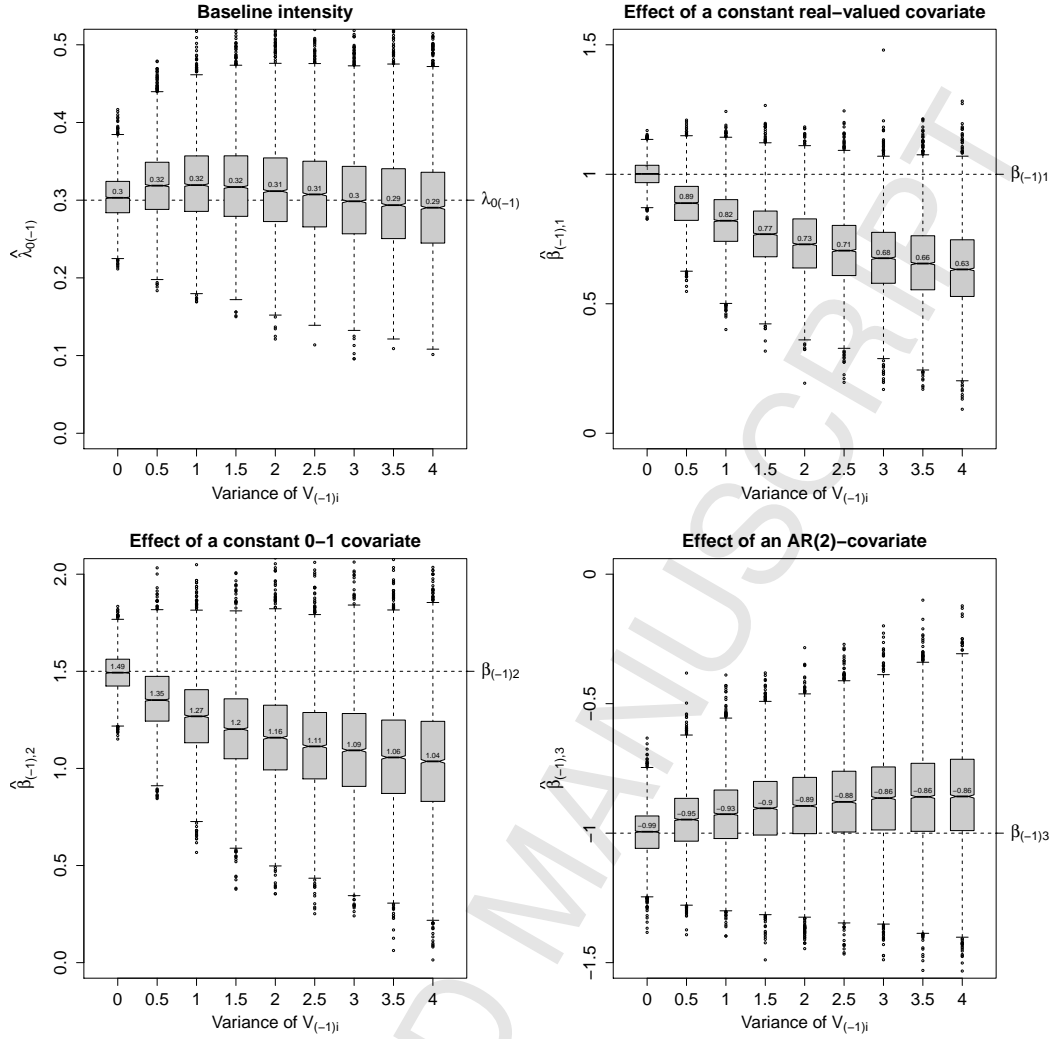


Figure 2: Estimation of the downgrade intensity parameters with omitted covariates

ii) Approximation of the marginal likelihood: The test is based on the Taylor approximation \tilde{L}^* , instead of the unobservable likelihood L^* of model (6). If the null hypothesis is valid, that is no measurement errors in the data, \tilde{L}^* and L^* are equal by construction. In order to investigate the quality of this approximation near the null hypothesis, we once again simulate $n = 250$ rating histories in each iteration - this time with rather small measurement errors - and estimate $E[\log \tilde{L}^*]$ and $E[\log L^*]$. The results are shown in Figure 3. It is evident that the difference between the likelihood and approximation disappears quite fast and uniformly, indicating that \tilde{L}^* is a reasonable foundation for the score test. As rejection in a score test results from a slope of the (marginal) likelihood that is too far from zero, the results suggests that a test based on the approximation of the likelihood might even have a higher power than a test based on the true likelihood, due to the higher slope of the approximation, if the variance of the errors differs from zero.

iii) Power of the score test: In order to investigate the properties of the score test in finite samples, we simulated rating histories of $n \in \{25, 50, 100, 250, 500, 1000\}$. The estimated distribution of the test statistic under $H_0 : \eta_{\uparrow} = \eta_{\downarrow} = 0$ is shown in Figure 4. The test statistic seems to converge rather fast. A reasonable level of approximation is already reached for $n = 250$. Table 1 shows the power of the test using a level of

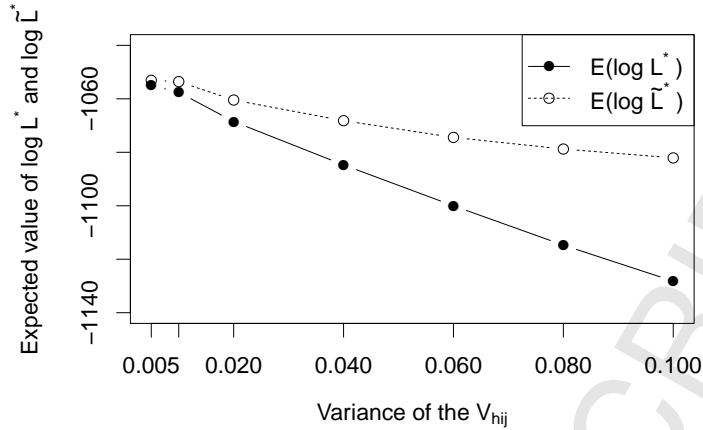


Figure 3: Difference between the marginal log likelihood $\log L^*$ and the approximation $\log \tilde{L}^*$

significance of 5%. The test exceeds the specified probability of a Type I error for all investigated sample sizes. However, with $n = 1000$, the desired level of 5% approximately holds and datasets in practice are usually of that sample size. Overall, the score test seems to be quite powerful near H_0 , as it detects even measurement errors with a small variance. As pointed out in the Monte Carlo study i), omitted covariates have a similar effect on the parameter estimators to ignored measurement errors. Therefore, the test is of course also sensitive to this kind of latent heterogeneity, and a positive test result does not automatically imply measurement errors in the transition times. The rejection of H_0 does, however, always imply that the estimators are biased and if a sufficient number of covariates is included in the model, measurement errors are the more likely reason.

4 Application

We apply the test to an internal rating history of an international corporate portfolio of a large German bank. The bank classifies its debtors into 21 rating states. The data contains 760 international corporate debtors and covers the period from 1997 until 2007. During this time, 233 one-state upgrades and 255 one-state downgrades were observed. The simulations in Section 3 suggest that the test is also sensitive to omitted

Table 1: Power of the score test ($\alpha = 5\%$)

Variance of V_{hij}	Sample size					
	25	50	100	250	500	1000
0.0	0.128	0.123	0.100	0.079	0.065	0.055
0.5	0.096	0.324	0.715	0.997	1.000	1.000
1.0	0.211	0.669	0.953	1.000	1.000	1.000
1.5	0.292	0.792	0.978	1.000	1.000	1.000
2.0	0.339	0.828	0.989	1.000	1.000	1.000
2.5	0.359	0.849	0.991	1.000	1.000	1.000

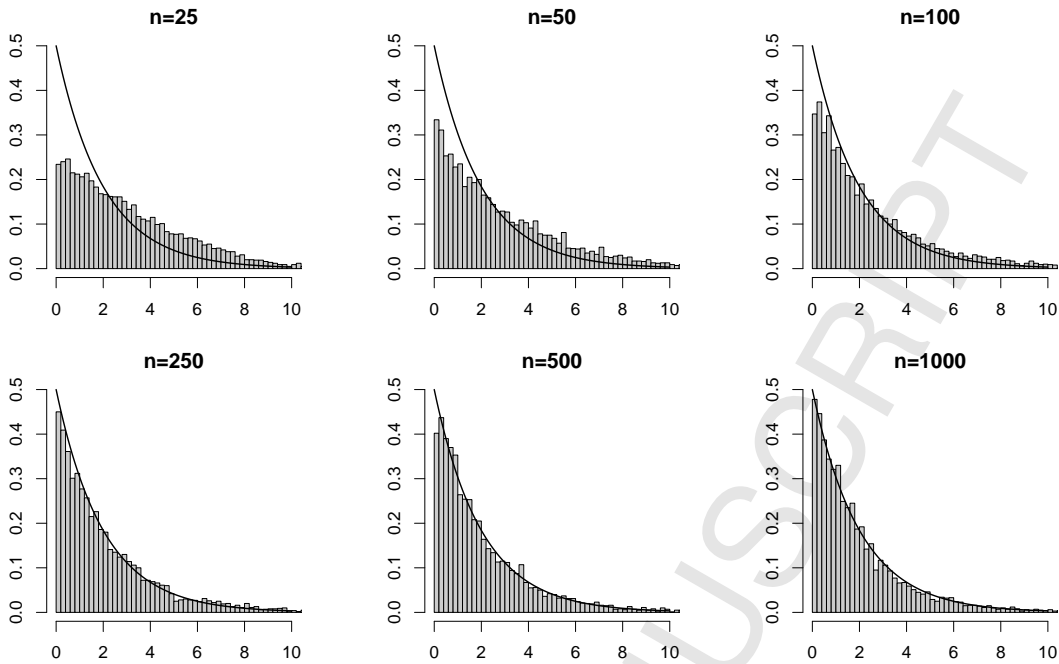


Figure 4: Empirical distribution of the test statistic under H_0 in finite samples and χ^2_2 -approach (solid line)

covariates. Hence, we strive to account for as many covariates that are established for credit risk as possible. Koopman et al. (2011) account for 100 covariates, which does not seem sensible here as our data base is considerably smaller. In considering their categories of bank lending conditions, macroeconomic and business cycle indicators, micro-level business and stock market returns and volatilities we use, after pre-screening, five economic industry indicators and industry dummies. For instance, GDP growth is also considered in Duffie et al. (2007). Specifically, we use GDP change, GDP trend, change in industrial production, 3-month interbank interest rates and stock market volatility. For each debtor, we choose the time series according to its location. For example, we use the GDP and industrial production of the USA, the USD-LIBOR and the volatility of the SP500 for an US-corporate and the GDP and industrial production of Germany, the EURIBOR and the volatility of the STOXX50 for a German debtor. The macroeconomic indicators are standardized, so that zero always indicates an average value of the respective covariate. In contrast to the latter two data sets being for the US, for our international data, we include country dummies, because, according to Kadam and Lenk (2008) and Nickell et al. (2000), the debtors' location has an impact on the transition probabilities. Regional dummies are also used by Hu et al. (2002). Kadam and Lenk (2008) also find evidence, that the transition behavior depends on the industrial sector of a debtor, which is why we include the industry as a covariate as well.

The dataset contains debtors from 9 regions and 18 industries, which have to be encoded as dummy variates with one reference category each. Together with the effects of the economic indicators and the baseline intensities, each of the two intensity models (1) contains 31 parameters that are estimated with the two-step ML-procedure described in Section 2.1. The results are summarized in Table 2. The fact that many regressors are not significant does not cause us to eliminate them from the model, because model selection here is not empirically driven.

The effects of the GDP change (ΔGDP) and the GDP trend ($\Delta^2\text{GDP}$) are intuitive. An above-average

Table 2: Estimated model parameters and standard errors

Macroeconomic indicators	$\hat{\beta}_{\uparrow l}$	$\hat{\sigma}(\hat{\beta}_{\uparrow l})$	$\hat{\beta}_{\downarrow l}$	$\hat{\sigma}(\hat{\beta}_{\downarrow l})$
Δ GDP	0.051	(0.087)	-0.056	(0.081)
Δ^2 GDP	0.010	(0.108)	-0.064	(0.090)
Δ Industrial production	0.259	(0.137)	0.232	(0.122)
3-month interbank interest rates	-0.135	(0.097)	-0.132	(0.101)
Volatility of stock returns	-0.201	(0.122)	-0.422	(0.101)
Regional effects (reference: Germany)	$\hat{\beta}_{\uparrow l}$	$\hat{\sigma}(\hat{\beta}_{\uparrow l})$	$\hat{\beta}_{\downarrow l}$	$\hat{\sigma}(\hat{\beta}_{\downarrow l})$
USA	0.165	(0.255)	-0.127	(0.316)
UK	0.147	(0.371)	0.485	(0.222)
China	-0.039	(0.740)	0.783	(0.555)
France	0.542	(0.397)	-0.750	(0.587)
Spain	0.272	(0.413)	0.623	(0.336)
Italy	-0.674	(0.725)	0.522	(0.324)
Japan	-1.396	(1.481)	-0.658	(0.911)
Other	0.556	(0.330)	-0.270	(0.426)
Industrial effects (reference: Miscellaneous)	$\hat{\beta}_{\uparrow l}$	$\hat{\sigma}(\hat{\beta}_{\uparrow l})$	$\hat{\beta}_{\downarrow l}$	$\hat{\sigma}(\hat{\beta}_{\downarrow l})$
Utilities, Electric & Waste Mgt.	0.001	(0.232)	0.030	(0.225)
Oil & Mining	-0.185	(0.384)	-0.053	(0.337)
Steel & Metal Products	0.359	(0.266)	0.154	(0.302)
Transportation	-0.004	(0.273)	0.513	(0.197)
IT, Optical & Precision Instruments	0.305	(0.274)	-0.026	(0.363)
Chemicals, Plastic & Rubber	-0.013	(0.343)	0.326	(0.362)
Automotive	0.680	(0.290)	0.469	(0.289)
Machinery & Equipment	0.021	(0.376)	-0.173	(0.416)
Business Products	-0.053	(0.526)	0.162	(0.359)
Business Services	0.123	(0.300)	-0.208	(0.446)
Construction	-0.028	(0.573)	-0.173	(0.316)
Real Estate	0.232	(0.306)	0.145	(0.322)
Medical Services, Instruments & Pharmaceuticals	-0.268	(0.530)	0.590	(0.385)
Consumer Goods (Production)	-0.391	(0.499)	-0.017	(0.359)
Food, Beverage & Tobacco	-0.291	(0.439)	0.352	(0.306)
Consumer Goods (Retail)	-0.191	(0.275)	0.477	(0.230)
Telephone, Broadcast Media & Cable TV	-0.566	(0.471)	0.724	(0.304)
Baseline intensities	$\hat{\lambda}_{0\uparrow}$	$\hat{\sigma}(\hat{\lambda}_{0\uparrow})$	$\hat{\lambda}_{0\downarrow}$	$\hat{\sigma}(\hat{\lambda}_{0\downarrow})$
	$3.0 \cdot 10^{-4}$	$(5.8 \cdot 10^{-4})$	$2.8 \cdot 10^{-4}$	$(4.8 \cdot 10^{-4})$

GDP-change yields an increase in the upgrade hazard and a decrease in the downgrade hazard, while negative values of both covariates have the opposite effect. In contrast to the other macroeconomic indicators, the effect of GDP is admittedly rather low. The change in industrial production (Δ Industrial production) in a debtors' home country has about the same positive effect on the upgrade- and the downgrade intensity. This implies that growth in industrial production leads to an increase in rating movements in both directions. The positive effect on the upgrade intensity is rather intuitive. The positive effect on the downgrade intensity

can be explained by problems in selling the produced goods, due to a slow reaction to a changing economic situation. Hence, phases of high industrial production yield a higher variance of the rating. The 3-month interbank interest rates have a negative effect on both intensities, implying that high interest rates stabilize the ratings, while low interest rates increase their volatility. This seems plausible, as 3-month interbank interest rates like the LIBOR and the EURIBOR serve as an indicator of general conditions in the financial market. Favorable market conditions entice businesses to invest or expand, which may have positive as well as negative consequences for their solvency in the medium term. High stock market volatility has a stabilizing effect on the ratings, whereas the estimated decrease in the downgrade intensity is much higher than in the upgrade intensity. This leads to the conclusion that the bank tends to keep the ratings constant in economically unstable times and is even more conservative with rating downgrades. This also has been reported in Kim et al. (2012).

The regional effects describe the intensity-difference of the debtors' business location with respect to the reference category Germany in an average economic situation. It is conspicuous, that Japanese debtors have a very stable rating, as both intensities decrease with respect to Germany. The opposite effect can be observed with respect to debtors from Spain, which yields a very volatile rating during the observation period.

The estimated industry effects of the automotive sector, the steel and metal product sector and the real estate sector reveal increased transition intensities in both directions. This is plausible, as these industries are strongly cyclical from an economic point of view and therefore, the respective debtors frequently undergo rating movements in both directions. Another notable industry is the telephone, broadcast media and cable TV sector, which exhibits an increased downgrade intensity and a decreased upgrade intensity. This may be explained by the fact that the majority of debtors in the data is from Germany. In 1998 the telephone monopoly of German Telekom was eradicated, which led to an overflow of the market with telephone businesses in the following years, and many of these businesses did not prove equal to the task.

A critical assumption in our model is the proportionality of the intensities. In order to assess a possible misspecification, we checked this assumption using the test of Grambsch and Therneau (1994). They argue that departure from proportionality can often be expressed by introducing time-dependent covariate effects. Hence, we tested $H_0 : \beta_h(t) = \beta_h$ versus $H_1 : \beta_h(t) = \beta_h + \gamma_h \mathbf{g}(t)$ for several functions of the observation time $\mathbf{g}(t)$, but were in no case able to reject H_0 . This is supported by the results of Duffie et al. (2009), who also test negatively for the proportional intensity assumption. Another possible cause of misspecification is the constant baseline intensity that we assume. In our model, fluctuations of the intensity over time are driven solely by the macroeconomic covariates. Weißbach and Walter (2010) however, find that the intensity of a rating change is significantly lower in the first year after origination of debt. In our dataset, this does not seem to be the case, though, as about a third of the rating changes occur during that time span, while the first year also accounts for about a third of the total observation time.

The estimated parameters are used to calculate the test statistic of the score test, which results in a value of 7.61, indicating a p-value of 0.022. Hence, the null hypothesis ' $H_0 : \eta_{\uparrow} = \eta_{\downarrow} = 0$ ' is rejected with respect to a significance level of 5%, which implies that there is evidence of measurement errors in the data.

5 Discussion

Given that find empirical evidence of measurement errors in the transition times, one possible direction for future work could be to explore reliable methods to account for these errors in rating models. Measurement errors are a form of latent heterogeneity in the data, which is usually modeled by expanding the intensity model by a frailty-process and using the marginal likelihood of the transition times for the parameter inference. However, the estimation results depend on the the properties of this frailty-process. Heckman and Singer (1984a,b) point out, that the estimators for the parameters of interest are sensitive to the choice of the parametric family for the frailties and, due to the fact that they are unobservable, it is unclear how to obtain an indication of their distribution. Heckman and Singer (1984a) propose a non-parametric procedure as an alternative. This procedure, however, turns out to be numerically cumbersome and dependent on a starting value, see Trussel and Richards (1985) for a detailed discussion.

For the homogeneous Markov process model, a presumably simpler approach is to formulate the likelihood in terms of the transition matrices (Bladt and Sørensen (2005, 2009)). The computational difficulty that transition matrices cannot be written in closed form as a function of the intensities can be overcome by numerically stable procedures for calculating derivatives and eigenvalues, or by an EM algorithm. However, the consistency proof for the parameter relies on the ergodicity of the Markov processes, which is not the case for rating histories when we assume the default to be an absorbing state, but this can be based on an argmax theorem (see Kremer and Weißbach, 2013). Even so, it must be conceded that if dates of rating affirmations are available, which is not the case in our data set, such techniques may well increase the efficiency. Bladt and Sørensen (2005, 2009), however, do not include covariates in their model. When determining the intensity parameters with the likelihood of the transition times, neglected covariates bias the estimators (Ridder and Verbakel, 1983). Whether this has a similar effect on the estimation method of Bladt and Sørensen (2005, 2009) is an interesting topic for future investigation.

Another direction for future work could be to formulate more complex models. Extending the model for non-proportional hazards (as in Kim et al., 2012), to include transitions to beyond the adjacent class (as in Weißbach and Walter, 2010), the non-homogeneous Markov process (as in Kim et al., 2012; Weißbach et al., 2009), to cross-sectional dependence (as in Koopman et al., 2011) or to non-Markovian hazards (as in Lando and Skødeberg, 2002; Koopman et al., 2008) can be done, as long as after marginalizing out the measurement error in a quadratic Taylor expansion, the likelihood factors have closed-form solution for derivatives with respect to the error. But one must always keep in mind that further parameters reduce the quality of the asymptotic approximation. For instance, Weißbach and Walter (2010) developed a test on stationarity of rating transitions, in a model, which allows for (instantaneous) transition between any states (apart from the absorbing). Even for a rating system with only eight rating states and thousands of rating histories, the nominal level of the test is not met.

Note however, that it is not feasible to interpret the measurement error as a random parameter and to apply Bayesian algorithms. In that case, the error variance, which is the parameter of interest in our case, would be a gearing parameter and not a model parameter. Of course, one could impose a prior distribution on the error variance and decide for the measurement error, if the credibility region does not contain the zero. However, in this case, a non-parametric distribution of the error must be assumed in the model, in order to achieve robustness.

Acknowledgement

We are indebted to the editor and two anonymous referees for detailed remarks. Additionally, we would like to thank Claudia Gregor-Lawrenz (WestLB), Patrick Tschiersch (WestLB), Walter Krämer and Dominik Wied for data and valuable suggestions. The financial support from the Deutsche Forschungsgemeinschaft is also gratefully acknowledged (Grant WE3573 “Cross-sectional dependence in durations - with application to finance and demography” and Graduate School GK1032 “Statistical Modelling”). All computations have been processed in R.

References

- Y. Amemiya. Instrumental variable estimator for the nonlinear errors-in-variables model. *Journal of Econometrics*, 28:273–289, 1985.
- P.K. Andersen and R.D. Gill. Cox’s regression model for counting processes: A large sample study. *The Annals of Statistics*, 10:1100–1120, 1982.
- P.K. Andersen, O. Borgan, R.D. Gill, and N. Keiding. *Statistical Models based on counting processes*. Springer, New York, 1993.
- M. Bladt and M. Sørensen. Statistical inference for discretely observed Markov jump processes. *Journal of the Royal Statistical Society, Series B*, 67:395–410, 2005.
- M. Bladt and M. Sørensen. Efficient estimation of transition rates between credit ratings from observations at discrete time points. *Quantitative Finance*, 9:147–160, 2009.
- A.A. Borovkov. *Mathematical Statistics*. Gordon & Breach, 1998.
- R.J. Cook and J.F. Lawless. *The statistical analysis of recurrent events*. Springer, New York, 2007.
- J. Davidson. *Stochastic Limit Theory*. Oxford, 1994.
- D. Duffie, L. Saita, and K. Wang. Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83:635–665, 2007.
- D. Duffie, A. Eckner, G. Horel, and L. Saita. Frailty correlated default. *Journal of Finance*, 64:2089–2123, 2009.
- M. Fisz. *Wahrscheinlichkeitsrechnung und mathematische Statistik*. VEB Deutscher Verlag der Wissenschaften, Berlin, 1989.
- H. Frydman. Estimation in the mixture of Markov chains moving with different speeds. *Journal of the American Statistical Association*, 100:1046–1053, 2005.
- H. Frydman and T. Schuermann. Credit rating dynamics and Markov mixture model. *Journal of Banking & Finance*, 32:1062–1075, 2008.
- L.G. Godfrey. *Misspecification tests in econometrics*. Cambridge University Press, Cambridge, 1990.

- P.M. Grambsch and T.M. Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81:515–526, 1994.
- J.J. Heckman and B. Singer. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52:271–320, 1984a.
- J.J. Heckman and B. Singer. Econometric duration analysis. *Journal of Econometrics*, 24:63–132, 1984b.
- Y. Hu, R. Kiesel, and W. Perraudin. The estimation of transition matrices for sovereign credit ratings. *Journal of Banking & Finance*, 26:1383–1406, 2002.
- S. Jaggia. Alternative forms of the score test for heterogeneity in a censored exponential model. *The Review of Economics and Statistics*, 79:340–343, 1997.
- R. Jarrow, D. Lando, and S. Turnbull. A Markov model for the term structure of credit risk spreads. *Review of Financial Studies*, 10:481–523, 1997.
- A. Kadam and P. Lenk. Bayesian inference for issuer heterogeneity in credit ratings migration. *Journal of Banking & Finance*, 32:2267–2274, 2008.
- Y.-D. Kim, L. James, and R. Weißbach. Bayesian analysis of multi-state event history data: Beta-dirichlet process prior. *Biometrika*, 99:127–140, 2012.
- S.J. Koopman, A. Lucas, and A.B. Monteiro. The multi-state latent factor intensity model for credit rating transitions. *Journal of Econometrics*, 142:399–424, 2008.
- S.J. Koopman, R. Kräussl, A. Lucas, and A.B. Monteiro. Credit cycles and macro fundamentals. *Journal of Empirical Finance*, 16:42–54, 2009.
- S.J. Koopman, A. Lucas, and B. Schwaab. Modeling frailty-correlated defaults using many macroeconomic covariates. *Journal of Econometrics*, 162:312–325, 2011.
- A. Kremer and R. Weißbach. Consistent estimation for discretely observed Markov jump processes with an absorbing state. *Statistical Papers*, 54:993–1007, 2013.
- T. Lancaster. *The Econometric Analysis of Transition Data*. Cambridge University Press, Cambridge, 1990.
- D. Lando and T.M. Skødeberg. Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking & Finance*, 26:423–444, 2002.
- E. Lenglart. Relation de domination entre deux processus. *Annales de l'Institut Henri Poincaré*, 13:171–179, 1977.
- T. Magnac and M. Visser. Transition models with measurement errors. *The Review of Economics and Statistics*, 81:466–474, 1999.
- T. Mählmann. Estimation of rating class transition probabilities with incomplete data. *Journal of Banking & Finance*, 30:3235–3256, 2006.
- R. Merton. On pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29:449–470, 1974.

- P. Nickell, W. Perraudin, and S. Varotto. Stability of rating transitions. *Journal of Banking & Finance*, 24: 203–227, 2000.
- C.R. Rao. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, 44:50–57, 1948.
- G. Ridder and W. Verbakel. On the estimation of the proportional hazard model in the presence of heterogeneity. Technical Report 22, University of Amsterdam, Faculty of Actuarial Science and Econometrics, 1983.
- C. Stefanescu, R. Tunaru, and S. Turnbull. The credit rating process and estimation of transition probabilities: A Bayesian approach. *Journal of Empirical Finance*, 16:216–234, 2009.
- J. Trussel and T. Richards. Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer strategy. *Sociological Methodology*, 15:242–276, 1985.
- G.J. van den Berg and B. van der Klaauw. Combining micro and macro unemployment duration data. *Journal of Econometrics*, 102:271–309, 2001.
- A.W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.
- R. Weißbach and T. Mollenhauer. Modelling rating transitions. *Journal of the Korean Statistical Society*, 4:469–485, 2011.
- R. Weißbach, P. Tschiersch, and C. Lawrenz. Testing time-homogeneity of rating transitions after origination of debt. *Empirical Economics*, 36:575–596, 2009.
- R. Weißbach and R. Walter. A likelihood ratio test for stationarity of rating transitions. *Journal of Econometrics*, 155:188–194, 2010.

Appendix

Proof of Lemma 1: The proof is not exactly the same for upgrades and downgrades. We only present it in detail for the downgrades, i.e. for $h = \downarrow$, and point out the necessary modifications for $h = \uparrow$ at the end of the proof. We start by deriving the conditional joint density of the durations between the downgrades and the duration between the last downgrade and the censoring time for a debtor i under the assumption of the intensity $\lambda_{hi}^*(t)$ from (6). Let the duration between the $(j-1)$ th and j th downgrade of debtor i be denoted by $D_{\downarrow ij}$. Using the Markov property of the rating, it can be shown that the distribution function of $D_{\downarrow ij}$ conditional on $(D_{\downarrow i(j-1)}, \dots, D_{\downarrow i1})^\top = (d_{\downarrow i(j-1)}, \dots, d_{\downarrow i1})^\top =: \mathbf{d}^{(j-1)}$, $\mathbf{X}_i(t) = \mathbf{x}_i(t)$ and $W_{\downarrow i}(t) = w_{\downarrow i}(t)$ is for $j \in \mathbb{N}$ given by

$$F_{D_{\downarrow ij} | \mathbf{d}^{(j-1)}, \mathbf{x}_i(t), w_{\downarrow i}(t)}(d_{\downarrow ij}) = 1 - \exp \left(-v_{\downarrow ij} \int_{\sum_{k=1}^{j-1} d_{\downarrow ik}}^{\sum_{k=1}^j d_{\downarrow ik}} \lambda_{0\downarrow i}^{(0)} \exp \left(\boldsymbol{\beta}_{\downarrow}^{(0)\top} \mathbf{x}_i(t) \right) dt \right), \quad (26)$$

where $v_{\downarrow ij}$ is the value of $w_{\downarrow i}(t)$ at the j th downgrade. The associated conditional density function is given by

$$f_{D_{\downarrow ij}|\mathbf{d}^{(j-1)}, \mathbf{x}_i(t), w_{\downarrow i}(t)}(d_{\downarrow ij}) = v_{\downarrow ij} \lambda_{0\downarrow i}^{(0)} \exp\left(\boldsymbol{\beta}_{\downarrow}^{(0)\top} \mathbf{x}_i \left(\sum_{k=1}^j d_{\downarrow ik}\right)\right) \times \exp\left(-v_{\downarrow ij} \int_{\sum_{k=1}^{j-1} d_{\downarrow ik}}^{\sum_{k=1}^j d_{\downarrow ik}} \lambda_{0\downarrow i}^{(0)} \exp\left(\boldsymbol{\beta}_{\downarrow}^{(0)\top} \mathbf{x}_i(t)\right) dt\right). \quad (27)$$

Let $\Delta_i \in \{0, 1\}$ be an indicator that specifies whether the last downgrade of debtor i is from state $D - 1$ to state D , which implies that the observation of the debtor ends because of insolvency. In this case, only $D_{\downarrow i1}, \dots, D_{\downarrow iN_{\downarrow i}(\tau)}$ are observed. Debtor insolvency is a rare event, though. For the majority of debtors, Δ_i will take the value 0, which implies that there is an additional observation that carries information about the downgrade process, namely the duration between the last downgrade and the censoring time. We only discuss the more common case of $\Delta_i = 0$ in detail.

The duration observed after $N_{\downarrow i}(\tau)$ downgrades is the minimum of $D_{\downarrow i(N_{\downarrow i}(\tau)+1)}$ and the duration between the $N_{\downarrow i}(\tau)$ th downgrade and the censoring time conditional on $N_{\downarrow i}(\tau) = n_{\downarrow i}$, $\mathbf{X}_i(t) = \mathbf{x}_i(t)$, $W_{\downarrow i}(t) = w_{\downarrow i}(t)$, $(D_{hi n_{\downarrow i}}, \dots, D_{hi 1})^\top = \mathbf{d}^{(n_{\downarrow i})}$ and $\Delta_i = 0$. The duration between the $n_{\downarrow i}$ th downgrade and the censoring time is given by the random variable $\tilde{C}_i = C_i - \sum_{j=1}^{n_{\downarrow i}} d_{\downarrow ij}$ with the condition that $C_i > \sum_{j=1}^{n_{\downarrow i}} d_{\downarrow ij}$. The density of $C_i > \sum_{j=1}^{n_{\downarrow i}} d_{\downarrow ij}$ is given by

$$\tilde{g}(c) := \frac{g(c)}{1 - G\left(\sum_{j=1}^{n_{\downarrow i}} d_{\downarrow ij}\right)} \mathbb{1}_{[\sum_{j=1}^{n_{\downarrow i}} d_{\downarrow ij}, \infty)}(c).$$

Determining the conditional density of $\min(\tilde{C}_i, D_{\downarrow i(N_{\downarrow i}(\tau)+1)})$ yields

$$\begin{aligned} & f_{\tilde{D}_{\downarrow i}|\mathbf{d}^{(j)}, \mathbf{x}_i(t), w_{\downarrow i}(t), n_{\downarrow i}, \Delta_i=0}(\tilde{d}_{\downarrow i}) \\ &= \frac{\partial}{\partial \tilde{d}_{\downarrow i}} P\left(\tilde{D}_{\downarrow i} \leq \tilde{d}_{\downarrow i} \mid N_{\downarrow i}(\tau) = n_{\downarrow i}, \Delta_i = 0, \mathbf{d}^{(j-1)}, \mathbf{x}_i(t), w_{\downarrow i}(t)\right) \\ &= \frac{\partial}{\partial \tilde{d}_{\downarrow i}} P\left(\tilde{C}_i \leq \tilde{d}_{\downarrow i}, \tilde{C}_i \leq D_{\downarrow i(n_{\downarrow i}+1)} \mid \mathbf{d}^{(j-1)}, \mathbf{x}_i(t), w_{\downarrow i}(t)\right) \\ &= \frac{\partial}{\partial \tilde{d}_{\downarrow i}} \int_0^{\tilde{d}_{\downarrow i}} \int_v^\infty f_{D_{\downarrow i(n_{\downarrow i}+1)}|\mathbf{d}^{(n_{\downarrow i})}, \mathbf{x}_i(t), w_{\downarrow i}(t)}(u) \tilde{g}\left(v + \sum_{j=1}^{n_{\downarrow i}} d_{\downarrow ij}\right) dudv \\ &= \frac{\partial}{\partial \tilde{d}_{\downarrow i}} \int_0^{\tilde{d}_{\downarrow i}} \left(1 - F_{D_{\downarrow i(n_{\downarrow i}+1)}|\mathbf{d}^{(n_{\downarrow i})}, \mathbf{x}_i(t), w_{\downarrow i}(t)}(v)\right) \tilde{g}\left(v + \sum_{j=1}^{n_{\downarrow i}} d_{\downarrow ij}\right) dv \\ &= \left(1 - F_{\tilde{D}_{\downarrow i}|\mathbf{d}^{(n_{hi})}, \mathbf{x}_i(t), w_{\downarrow i}(t)}(\tilde{d}_{\downarrow i})\right) \tilde{g}\left(\tilde{d}_{\downarrow i} + \sum_{j=1}^{n_{\downarrow i}} d_{\downarrow ij}\right). \end{aligned} \quad (28)$$

In order to show that the likelihood factor $\bar{L}_{\downarrow i}$ is essentially the joint density function of $T_{\downarrow i1}, \dots, T_{\downarrow in_{\downarrow i}}, C_i$, the distribution and density functions (26) and (27) are now approximated using the same principle as for

the likelihood in Section 2.1. This yields

$$\begin{aligned} \bar{F}_{D_{\downarrow ij}|\mathbf{d}^{(j-1)}, \mathbf{x}_i(t)}(d_{\downarrow ij}) &:= 1 - \exp \left[- \int_{\sum_{k=1}^{j-1} d_{\downarrow ik}}^{\sum_{k=1}^j d_{\downarrow ik}} \lambda_{0\downarrow i}^{(0)} \exp \left(\boldsymbol{\beta}_{\downarrow}^{(0)\top} \mathbf{x}_i(t) \right) dt \right] \\ &\times \left[1 + \eta_{\downarrow} \left(\int_{\sum_{k=1}^{j-1} d_{\downarrow ik}}^{\sum_{k=1}^j d_{\downarrow ik}} \lambda_{0\downarrow i}^{(0)} \exp \left(\boldsymbol{\beta}_{\downarrow}^{(0)\top} \mathbf{x}_i(t) \right) dt \right)^2 \right] \end{aligned} \quad (29)$$

and

$$\begin{aligned} \bar{f}_{D_{\downarrow ij}|\mathbf{d}^{(j-1)}, \mathbf{x}_i(t)}(d_{\downarrow ij}) &:= \lambda_{0\downarrow i}^{(0)} \exp \left(\boldsymbol{\beta}_{\downarrow}^{(0)\top} \mathbf{x}_i \left(\sum_{k=1}^j d_{\downarrow ik} \right) \right) \exp \left[- \int_{\sum_{k=1}^{j-1} d_{\downarrow ik}}^{\sum_{k=1}^j d_{\downarrow ik}} \lambda_{0\downarrow i}^{(0)} \exp \left(\boldsymbol{\beta}_{\downarrow}^{(0)\top} \mathbf{x}_i(t) \right) dt \right] \\ &\times \left[1 + \eta_{\downarrow} \left\{ \left(\int_{\sum_{k=1}^{j-1} d_{\downarrow ik}}^{\sum_{k=1}^j d_{\downarrow ik}} \lambda_{0\downarrow i}^{(0)} \exp \left(\boldsymbol{\beta}_{\downarrow}^{(0)\top} \mathbf{x}_i(t) \right) dt \right)^2 \right. \right. \\ &\left. \left. - 2 \int_{\sum_{k=1}^{j-1} d_{\downarrow ik}}^{\sum_{k=1}^j d_{\downarrow ik}} \lambda_{0\downarrow i}^{(0)} \exp \left(\boldsymbol{\beta}_{\downarrow}^{(0)\top} \mathbf{x}_i(t) \right) dt \right\} \right]. \end{aligned} \quad (30)$$

For $\eta_{\downarrow} \in [0, 1]$, \bar{F} is left-continuous with left-hand limit 0 and right-hand limit 1 and hence, a proper distribution function, according to the definition of Kolmogoroff (see Fisz (1989), Definition 2.2.1). As \bar{f} is the derivation of \bar{F} , it is a regular density function with $\int_0^{\infty} \bar{f} = 1$.

The joint distribution of $\mathbf{D}_{\downarrow i} := (D_{\downarrow i1}, \dots, D_{\downarrow in_{hi}})$ and $\tilde{D}_{\downarrow i}$ is the product of the conditional densities (27) and (28). Replacing f and F by \bar{f} and \bar{F} then yields an approximation of the joint distribution:

$$\begin{aligned} &f_{\mathbf{D}_{\downarrow i}, \tilde{D}_{\downarrow i}|\mathbf{x}_i(t), w_{\downarrow i}(t), n_{\downarrow i}, \Delta_i=0}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}) \\ &= \left(\prod_{j=1}^{n_{hi}} f_{D_{\downarrow ij}|\mathbf{d}^{(j-1)}, \mathbf{x}_i(t), w_{\downarrow i}(t)}(d_{\downarrow ij}) \right) \\ &\quad \times \left(1 - F_{\tilde{D}_{\downarrow i}|\mathbf{d}^{(n_{hi})}, \mathbf{x}_i(t), w_{\downarrow i}(t)}(\tilde{d}_{\downarrow i}) \right) \tilde{g} \left(\tilde{d}_{\downarrow i} + \sum_{j=1}^{n_{\downarrow i}} d_{\downarrow ij} \right) \end{aligned} \quad (31)$$

$$\begin{aligned} &\approx \left(\prod_{j=1}^{n_{hi}} \bar{f}_{D_{\downarrow ij}|\mathbf{d}^{(j-1)}, \mathbf{x}_i(t)}(d_{\downarrow ij}) \right) \\ &\quad \times \left(1 - \bar{F}_{\tilde{D}_{\downarrow i}|\mathbf{d}^{(n_{hi})}, \mathbf{x}_i(t)}(\tilde{d}_{\downarrow i}) \right) \tilde{g} \left(\tilde{d}_{\downarrow i} + \sum_{j=1}^{n_{\downarrow i}} d_{\downarrow ij} \right) \end{aligned} \quad (32)$$

$$=: \bar{f}_{\downarrow i}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}; \eta_{\downarrow})$$

Note that under the null hypothesis, $\eta_{\downarrow} = 0$ (31) equals (32), as the measurement error terms $v_{\downarrow ij}$ take the value 1 for all j . Hence, $\bar{f}_{\downarrow i}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}; 0)$ denotes the joint distribution of $\mathbf{D}_{\downarrow i}$ and $\tilde{D}_{\downarrow i}$ under H_0 .

With $T_{\downarrow ij} = \sum_{k=1}^j D_{\downarrow ik}$, $C_i = \tilde{D}_{\downarrow i} + \sum_{k=1}^j D_{\downarrow ik}$ and the fact, that $Y_{\downarrow i}(t) = 1$ for $t < C_i$, it follows that with given transition times, covariate path and risk process

$$\bar{L}_{\downarrow i}(\boldsymbol{\theta}_{\downarrow}^{(0)}, \eta_{\downarrow}) = \frac{\bar{f}_{\downarrow i}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}; \eta_{\downarrow})}{\tilde{g} \left(\tilde{d}_{\downarrow i} + \sum_{j=1}^{n_{\downarrow i}} d_{\downarrow ij} \right)}. \quad (33)$$

Hence, apart from the factor \tilde{g} , the likelihood $\bar{L}_{\downarrow i}$ is a conditional density of duration times. Moreover

$$\frac{\partial}{\partial \boldsymbol{\theta}_{\downarrow}^*} \log \bar{L}_{\downarrow i}(\boldsymbol{\theta}_{\downarrow}^*) \Big|_{(\boldsymbol{\theta}_{\downarrow}^{(0)\top}, \eta_{\downarrow})^\top} = \frac{\partial}{\partial (\boldsymbol{\theta}_{\downarrow}^{(0)\top}, \eta_{\downarrow})^\top} \log [\bar{f}_{\downarrow i}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}; \eta_{\downarrow})], \quad (34)$$

i.e. the log likelihood and the logarithm of the density have the same derivative, as \tilde{g} does not depend on $(\boldsymbol{\theta}_{\downarrow}^{(0)\top}, \eta_{\downarrow})^\top$ due to **(A6)**. From the fact that $\bar{f}_{\downarrow i}$ is a density function, it follows that

$$\int_0^\infty \dots \int_0^\infty \bar{f}_{\downarrow i}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}; \eta_{\downarrow}) d\tilde{d}_{\downarrow i} \dots dd_{\downarrow i1} = 1. \quad (35)$$

As the integration limits of (35) do not depend on the parameters, integrating and differentiating with respect to $(\boldsymbol{\theta}_{\downarrow}^{(0)\top}, \eta_{\downarrow})^\top$ is interchangeable. Hence, differentiation of both sides of (35) and using (34) yields

$$\begin{aligned} & \int_0^\infty \dots \int_0^\infty \frac{\partial}{\partial (\boldsymbol{\theta}_{\downarrow}^{(0)\top}, \eta_{\downarrow})^\top} \bar{f}_{\downarrow i}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}; \eta_{\downarrow}) d\tilde{d}_{\downarrow i} \dots dd_{\downarrow i1} \\ &= \int_0^\infty \dots \int_0^\infty \frac{\partial}{\partial (\boldsymbol{\theta}_{\downarrow}^{(0)\top}, \eta_{\downarrow})^\top} \log \bar{f}_{\downarrow i}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}; \eta_{\downarrow}) \bar{f}_{\downarrow i}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}; \eta_{\downarrow}) d\tilde{d}_{\downarrow i} \dots dd_{\downarrow i1} \\ &= \int_0^\infty \dots \int_0^\infty \left(\frac{\partial}{\partial \boldsymbol{\theta}_{\downarrow}^*} \log \bar{L}_{\downarrow i}(\boldsymbol{\theta}_{\downarrow}^*) \Big|_{(\boldsymbol{\theta}_{\downarrow}^{(0)\top}, \eta_{\downarrow})^\top} \right) \bar{f}_{\downarrow i}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}; \eta_{\downarrow}) d\tilde{d}_{\downarrow i} \dots dd_{\downarrow i1} \\ &= 0. \end{aligned} \quad (36)$$

As already mentioned, $\bar{f}_{\downarrow i}$ is the true conditional density of $\mathbf{D}_{\downarrow i}$ and $\tilde{D}_{\downarrow i}$ if $\eta_{\downarrow} = 0$. Therefore, under H_0 , the third row of (36) is the conditional expectation of the the log likelihood and (36) implies that

$$\mathbb{E} \left[\left(\frac{\partial}{\partial \boldsymbol{\theta}_{\downarrow}^*} \log \bar{L}_{\downarrow i}(\boldsymbol{\theta}_{\downarrow}^*) \right) \Big|_{(\boldsymbol{\theta}_{\downarrow}^{(0)\top}, 0)^\top} \Big| N_{\downarrow i}(\tau), \mathbf{X}_i(t), \Delta_i = 0 \right] = 0. \quad (37)$$

This also holds for $\Delta = 1$, that is for the (rare) situation that the observation of a debtor ends with a downgrade from the second to last rating state into insolvency. The only difference in the proof is that in this case, the factor of $\bar{f}_{\downarrow i}$, which belongs to the duration between the last downgrade and the censoring time, is omitted. Hence, $\mathbb{E}[\bar{U}_{\downarrow i}(\boldsymbol{\theta}_{\downarrow}^{(0)}) | N_{\downarrow i}(\tau), \mathbf{X}_i(t), \Delta_i] = 0$ and, according to the law of iterated expectation, $\mathbb{E}[\bar{U}_{\downarrow i}(\boldsymbol{\theta}_{\downarrow}^{(0)})] = 0$.

For $h = \uparrow$, it has to be taken into account that the risk process $Y_{\uparrow i}(t)$ can take the value 0 on an interim basis, if a debtor is in the highest possible rating state. Moreover, it is not possible for the observation time to end with an upgrade, so there is always a positive duration between the last upgrade and the censoring time. The only challenge in accounting for these two differences lies in the notation. The principle of the proof is exactly the same as for $h = \downarrow$. \square

Proof of Lemma 2: As in the previous proof, we only consider the case $h = \downarrow$ and $\Delta_i = 0$. Differentiating both sides of (35) twice with respect to $(\boldsymbol{\theta}_{\downarrow}^{(0)\top}, \eta_{\downarrow})^\top$ yields

$$0 = \int_0^\infty \cdots \int_0^\infty \frac{\partial^2}{\partial (\boldsymbol{\theta}_{\downarrow}^{(0)\top}, \eta_{\downarrow})^\top \partial (\boldsymbol{\theta}_{\downarrow}^{(0)\top}, \eta_{\downarrow})} \log \bar{f}_{\downarrow i}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}; \eta_{\downarrow}) \bar{f}_{\downarrow i}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}; \eta_{\downarrow}) d\tilde{d}_{\downarrow i} \cdots dd_{\downarrow i1} \\ + \int_0^\infty \cdots \int_0^\infty \frac{\partial}{\partial (\boldsymbol{\theta}_{\downarrow}^{(0)\top}, \eta_{\downarrow})^\top} \log \bar{f}_{\downarrow i}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}; \eta_{\downarrow}) \frac{\partial}{\partial (\boldsymbol{\theta}_{\downarrow}^{(0)\top}, \eta_{\downarrow})} \bar{f}_{\downarrow i}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}; \eta_{\downarrow}) d\tilde{d}_{\downarrow i} \cdots dd_{\downarrow i1}$$

and hence,

$$\int_0^\infty \cdots \int_0^\infty \frac{\partial}{\partial \boldsymbol{\theta}_{\downarrow}^*} \log \bar{L}_{\downarrow i}(\boldsymbol{\theta}_{\downarrow}^*) \Big|_{(\boldsymbol{\theta}_{\downarrow}^{(0)\top}, \eta_{\downarrow})^\top} \left(\frac{\partial}{\partial \boldsymbol{\theta}_{\downarrow}^{*\top}} \log \bar{L}_{\downarrow i}(\boldsymbol{\theta}_{\downarrow}^*) \Big|_{(\boldsymbol{\theta}_{\downarrow}^{(0)\top}, \eta_{\downarrow})^\top} \right)^\top \\ \times \bar{f}_{\downarrow i}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}; \eta_{\downarrow}) d\tilde{d}_{\downarrow i} \cdots dd_{\downarrow i1} \\ = - \int_0^\infty \cdots \int_0^\infty \frac{\partial^2}{\partial \boldsymbol{\theta}_{\downarrow}^* \partial \boldsymbol{\theta}_{\downarrow}^{*\top}} \log \bar{L}_{\downarrow i}(\boldsymbol{\theta}_{\downarrow}^*) \Big|_{(\boldsymbol{\theta}_{\downarrow}^{(0)\top}, \eta_{\downarrow})^\top} \bar{f}_{\downarrow i}(\mathbf{d}_{\downarrow i}, \tilde{d}_{\downarrow i}; \eta_{\downarrow}) d\tilde{d}_{\downarrow i} \cdots dd_{\downarrow i1}. \quad (38)$$

If H_0 is valid, $\bar{f}_{\downarrow i}$ is the true density and (38) says that

$$\mathbb{E} \left[\left(\left(\frac{\partial}{\partial \boldsymbol{\theta}_{\downarrow}^*} \log \bar{L}_{\downarrow i}(\boldsymbol{\theta}_{\downarrow}^*) \right) \left(\frac{\partial}{\partial \boldsymbol{\theta}_{\downarrow}^{*\top}} \log \bar{L}_{\downarrow i}(\boldsymbol{\theta}_{\downarrow}^*) \right)^\top \right) \Big|_{(\boldsymbol{\theta}_{\downarrow}^{(0)\top}, 0)} N_{\downarrow i}(\tau), \mathbf{X}_i(t), \Delta_i = 0 \right] \\ = - \mathbb{E} \left[\left(\frac{\partial^2}{\partial \boldsymbol{\theta}_{\downarrow}^* \partial \boldsymbol{\theta}_{\downarrow}^{*\top}} \log \bar{L}_{\downarrow i}(\boldsymbol{\theta}_{\downarrow}^*) \right) \Big|_{(\boldsymbol{\theta}_{\downarrow}^{(0)\top}, 0)} N_{\downarrow i}(\tau), \mathbf{X}_i(t), \Delta_i = 0 \right]. \quad (39)$$

This result also holds for $\Delta_i = 1$. Lemma 1, the theorem of iterated expectation and equation (39) then yield

$$\text{Cov} \left[\bar{U}_{\downarrow i}(\boldsymbol{\theta}_{\downarrow}^{(0)}) \right] \\ = \mathbb{E} \left[\left(\left(\frac{\partial}{\partial \boldsymbol{\theta}_{\downarrow}^*} \log \bar{L}_{\downarrow i} \right) \left(\frac{\partial}{\partial \boldsymbol{\theta}_{\downarrow}^{*\top}} \log \bar{L}_{\downarrow i} \right)^\top \right) \Big|_{(\boldsymbol{\theta}_{\downarrow}^{(0)\top}, 0)} \right] \\ = \mathbb{E} \left[\mathbb{E} \left[\left(\left(\frac{\partial}{\partial \boldsymbol{\theta}_{\downarrow}^*} \log \bar{L}_{\downarrow i} \right) \left(\frac{\partial}{\partial \boldsymbol{\theta}_{\downarrow}^{*\top}} \log \bar{L}_{\downarrow i} \right)^\top \right) \Big|_{(\boldsymbol{\theta}_{\downarrow}^{(0)\top}, 0)} N_{\downarrow i}(\tau), \mathbf{X}_i(t), \Delta_i \right] \right] \\ = - \mathbb{E} \left[\mathbb{E} \left[\left(\frac{\partial^2}{\partial \boldsymbol{\theta}_{\downarrow}^* \partial \boldsymbol{\theta}_{\downarrow}^{*\top}} \log \bar{L}_{\downarrow i} \right) \Big|_{(\boldsymbol{\theta}_{\downarrow}^{(0)\top}, 0)} N_{\downarrow i}(\tau), \mathbf{X}_i(t), \Delta_i \right] \right] \\ = - \mathbb{E} \left[\left(\frac{\partial^2}{\partial \boldsymbol{\theta}_{\downarrow}^* \partial \boldsymbol{\theta}_{\downarrow}^{*\top}} \log \bar{L}_{\downarrow i} \right) \Big|_{(\boldsymbol{\theta}_{\downarrow}^{(0)\top}, 0)} \right].$$

With the modifications addressed at the end of the previous proof, the argumentation also applies for $h = \uparrow$. \square

Proof of Lemma 3: Let $i \in \{1, \dots, n\}$ and $\boldsymbol{\theta} \in \Theta$ be arbitrary. The components of $\bar{U}_i(\boldsymbol{\theta})$ are given by

$$\begin{aligned} \frac{\partial}{\partial \lambda_{0h}} \log \bar{L}_i(\boldsymbol{\theta}^*) \Big|_{(\boldsymbol{\theta}_\uparrow^\top, 0, \boldsymbol{\theta}_\downarrow^\top, 0)^\top} &= \frac{N_{hi}(\tau)}{\lambda_{0h}} - \int_0^\tau Y_{hi}(t) \exp(\boldsymbol{\beta}_h^\top \mathbf{X}_i(t)) dt, \\ \frac{\partial}{\partial \beta_l} \log \bar{L}_i(\boldsymbol{\theta}^*) \Big|_{(\boldsymbol{\theta}_\uparrow^\top, 0, \boldsymbol{\theta}_\downarrow^\top, 0)^\top} &= \sum_{j=1}^{N_{hi}(\tau)} X_{il}(T_{hij}) - \int_0^\tau Y_{hi}(t) \lambda_{0h} X_{il}(t) \exp(\boldsymbol{\beta}_h^\top \mathbf{X}_i(t)) dt, \end{aligned}$$

$l = 1, \dots, d$, and

$$\frac{\partial}{\partial \eta_h} \log \bar{L}_i(\boldsymbol{\theta}^*) \Big|_{(\boldsymbol{\theta}_\uparrow^\top, 0, \boldsymbol{\theta}_\downarrow^\top, 0)^\top} = \sum_{i=1}^n \left(\sum_{j=1}^{N_{hi}(\tau)} (z_{hij}^2(\boldsymbol{\theta}_h) - 2z_{hij}(\boldsymbol{\theta}_h)) + z_{hi(N_{hi}(\tau)+1)}^2(\boldsymbol{\theta}_h) \right)$$

with

$$z_{hij}(\boldsymbol{\theta}_h) = \int_{T_{hi(j-1)}}^{T_{hij}} Y_{hi}(t) \lambda_{0h} \exp(\boldsymbol{\beta}_h^\top \mathbf{X}_i(t)) dt$$

(see (8)). Thus, $\bar{U}_i(\boldsymbol{\theta})$ is clearly infinitely differentiable and due to the boundedness of the underlying random variables, it is bounded as well. Hence, if we regard $\bar{U}_i(\boldsymbol{\theta})$ as a function of the random variables, i.e. $\bar{U}_i(\boldsymbol{\theta}) = f_{\boldsymbol{\theta}}(N_{hi}(t), Y_{hi}(t), \mathbf{X}_i(t), h \in \{\uparrow, \downarrow\}, t \in [0, \tau])$, $\{f_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$ is a Donsker class, according to van der Vaart (1998), example 19.9. As $\bar{U}_i(\boldsymbol{\theta})$ are independent for $i = 1, \dots, n$, it follows from van der Vaart (1998), p. 269, that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\bar{U}_i(\boldsymbol{\theta}) - \bar{\mathbf{u}}(\boldsymbol{\theta})) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta})).$$

As uniform convergence implies pointwise convergence, this means in particular that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\bar{U}_i(\boldsymbol{\theta}^{(0)}) - \bar{\mathbf{u}}_i(\boldsymbol{\theta}^{(0)})) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}^{(0)})).$$

According to Lemma 1 $\bar{\mathbf{u}}_i(\boldsymbol{\theta}^{(0)}) = 0$, while Lemma 2 and (A7) yield

$$\mathcal{I}(\boldsymbol{\theta}^{(0)}) = \begin{pmatrix} \mathcal{I}_\uparrow(\boldsymbol{\theta}_\uparrow^{(0)}) & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_\downarrow(\boldsymbol{\theta}_\downarrow^{(0)}) \end{pmatrix}.$$

□

Proof of Lemma 4: According to the law of large numbers

$$\hat{\mathcal{I}}(\boldsymbol{\theta}) = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \boldsymbol{\theta}^* (\partial \boldsymbol{\theta}^*)^\top} \log \bar{L}_i(\boldsymbol{\theta}^*) \Big|_{(\boldsymbol{\theta}_\uparrow^\top, 0, \boldsymbol{\theta}_\downarrow^\top, 0)} \xrightarrow{\mathcal{P}} \mathcal{I}(\boldsymbol{\theta})$$

for $\boldsymbol{\theta} = (\boldsymbol{\theta}_\uparrow^\top, \boldsymbol{\theta}_\downarrow^\top)^\top \in \Theta$ with

$$\mathcal{I}(\boldsymbol{\theta}) := \mathbb{E} \left[\left(\frac{\partial^2}{\partial \boldsymbol{\theta}^* (\partial \boldsymbol{\theta}^*)^\top} \log \bar{L}_i(\boldsymbol{\theta}^*) \right)_{(\boldsymbol{\theta}_\uparrow^\top, 0, \boldsymbol{\theta}_\downarrow^\top, 0)} \right].$$

It is obvious from (15) that all components of $\hat{\mathcal{I}}(\boldsymbol{\theta})$ are differentiable with respect to $\boldsymbol{\theta}$. According to Davidson (1994), p. 340, this is a sufficient condition for $\hat{\mathcal{I}}(\boldsymbol{\theta})$ to be equicontinuous. As a consequence, the components of $\hat{\mathcal{I}}(\boldsymbol{\theta})$ converge not only pointwise, but uniformly to the components of $\mathcal{I}(\boldsymbol{\theta})$. As $\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}$, the proof follows with Davidson (1994), Theorem 21.6. \square

Proof of Lemma 5: The proof of this Lemma is straightforward, using the properties of covariance matrices:

$$\begin{aligned}
\text{Cov} [\bar{U}_{\lambda_0}(\boldsymbol{\theta}), \bar{U}_{-\lambda_0}^*(\boldsymbol{\theta})] &= \text{Cov} [\bar{U}_{\lambda_0}(\boldsymbol{\theta}), \bar{U}_{-\lambda_0}(\boldsymbol{\theta}) - \mathcal{I}_{-\lambda_0\lambda_0}(\boldsymbol{\theta})\mathcal{I}_{\lambda_0\lambda_0}^{-1}(\boldsymbol{\theta})\bar{U}_{\lambda_0}(\boldsymbol{\theta})] \\
&= \text{Cov} [\bar{U}_{\lambda_0}(\boldsymbol{\theta}), \bar{U}_{-\lambda_0}(\boldsymbol{\theta})] - \text{Cov} [\bar{U}_{\lambda_0}(\boldsymbol{\theta}), \mathcal{I}_{-\lambda_0\lambda_0}(\boldsymbol{\theta})\mathcal{I}_{\lambda_0\lambda_0}^{-1}(\boldsymbol{\theta})\bar{U}_{\lambda_0}(\boldsymbol{\theta})] \\
&= \mathcal{I}_{\lambda_0-\lambda_0}(\boldsymbol{\theta}) - \text{Cov} [\bar{U}_{\lambda_0}(\boldsymbol{\theta}), \bar{U}_{\lambda_0}(\boldsymbol{\theta})] \mathcal{I}_{\lambda_0\lambda_0}^{-1}(\boldsymbol{\theta})\mathcal{I}_{-\lambda_0\lambda_0}^\top(\boldsymbol{\theta}) \\
&= \mathcal{I}_{\lambda_0-\lambda_0}(\boldsymbol{\theta}) - \mathcal{I}_{\lambda_0\lambda_0}(\boldsymbol{\theta})\mathcal{I}_{\lambda_0\lambda_0}^{-1}(\boldsymbol{\theta})\mathcal{I}_{\lambda_0-\lambda_0}(\boldsymbol{\theta}) \\
&= 0
\end{aligned}$$

\square