

Weißbach, Rafael; Mollenhauer, Thomas

Conference Paper

Modelling Rating Transitions

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2011: Die Ordnung der Weltwirtschaft:
Lektionen aus der Krise - Session: Interest Rates, No. A16-V2

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Weißbach, Rafael; Mollenhauer, Thomas (2011) : Modelling Rating Transitions, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2011: Die Ordnung der Weltwirtschaft: Lektionen aus der Krise - Session: Interest Rates, No. A16-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft

This Version is available at:

<https://hdl.handle.net/10419/48698>

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.

Modelling Rating Transitions

Rafael Weißbach*

Chair for Statistics, University of Rostock, Germany

Thomas Mollenhauer

NORD/LB, Credit Risk Control, Hannover, Germany

Abstract

The time-continuous discrete-state Markov process is a model for rating transitions. One parameter, namely the intensity to migrate to an adjacent rating state, implies an ordinal rating to have an intuitive metric. State-specific intensities generalize the state-stationarity. Observing Markov processes from a multiplicative intensity model, the maximum likelihood parameter estimators for both models can be written as a martingale transform of the processes that count transitions between the rating states. A Taylor expansion reveals consistency and asymptotic normality of the parameter estimates, resulting in a χ^2 -distributed likelihood ratio of state-stationarity and the state-specific model. This extends to time-stationarity. Simulations contrast the asymptotic results with finite samples. An application to a sufficiently large set of credit rating histories shows that the one-parameter model can be a good starting point.

Keywords: Rating, Metricality, Multiple Markov process, Counting process, Likelihood ratio

2000 MSC: 60J28, 62M02, 62P05

*Corresponding author: Prof. Dr. Rafael Weißbach, Chair of Statistics, Faculty for Economic and Social Sciences, University of Rostock, 18051 Rostock, Germany, Phone: +49-381-4984428, Fax: +49-381-4984401.

Email address: rafael.weissbach@uni-rostock.de (Rafael Weißbach)

URL: <http://www.wiwi.uni-rostock.de/vwl/statistik/> (Rafael Weißbach)

1. Introduction

The homogenous Markov process, with stationary transition intensities, remains the starting point for rating-transition modelling (Bluhm et al., 2002, pg. 197ff). Several defects of the model, such as instationarity, non-Markovian behavior or intra-temporal dependence have been investigated (Altman and Kao 1992; Lando and Skødeberg 2002; Bangia et al. 2002; Frydman and Schuerman 2008; Kiefer and Larson 2007; Koopman and Lucas 2008; Weißbach and Walter 2010). Our general statistical objective is a parsimonious model, and Forest et al. (1998) formulate a one-parameter model for ratings, not, however, originating from the Markov process model. Here, we claim three properties that enable the formulation of a one-parameter Markov process model. First, the Merton model (Merton 1974) for an asset value suggests that a firm can only migrate from one rating state to an adjacent rating state, that is, up or down. All other transition intensities must be zero and observations of multiple class transitions are attributable to discontinuous observation and considered here as missing data. Second, a rating should be constructed so as to be metric, and not only ordinal. If changing rating classes does not depend on the specific state, i.e. is state-stationary, we will see that the rating can be equipped with a simple metric. Third transition intensities should be time-stationary.

We assume the first claim and take it into account in the data analysis. The second restriction, namely whether rating class changes are class-specific, is our primary research question. We study a likelihood ratio test on the null hypothesis of metricality. The formalization of the null hypothesis is a constant transition intensity for all rating classes and the alternative hypothesis, of only ordinal ratings, is formalized by letting each rating class have its two specific transition intensities in the directions of upgrade and downgrade. The third property will be studied briefly.

Maximum likelihood (ML) estimation for the generator of the homogeneous Markov process dates back to Albert (1962). We study the consistency and asymptotic normality of the estimator for the state-specific intensity and the state-stationary intensity. The results originate from the representation of the estimators as martingale transforms that arise for the transition count between rating states. The martingale limit theorem by Rebolledo (1980) suggests studying the predictable covariation process with inequalities by E. Lenglart and R. Gill. The time-stationarity can be studied by generalizing to a piecewise stationary. Through an argument relating to the profile likelihood, the likelihood ratio test statistics that compare one the one hand state-stationary versus state-specific and on the other time-stationary versus piecewise stationary are both asymptotically χ^2 -distributed.

Our application is credit risk, in particular, the rating transition intensities in an internal rating system, loosely related to the expert-rating discussed by Kiefer (2010). We show that our data may not be modeled significantly by one parameter, even if time-stationarity is imposed by transformation of the time. However, the model appears to be close to reality, simulation studies foster such an impression.

2. Models

Consider the time-continuous discrete-state Markov processes $\mathbf{X} = \{X_t, t \in [0, T]\}$ defined on a probability space $(\Omega, \mathfrak{F}, P)$. The ordered states $1, \dots, k$, e.g. rating classes, end in an absorbing state k (e.g. bankruptcy). We denote X_t as the state of an asset at time t , after a certain origin. Denote by $m_h(t) = P(X_t = h)$ the unconditional probability of state h at time t . The data are transition histories $\mathbf{X}_i = \{X_t^i, t \in [0, T]\}$ for each of the $i = 1, \dots, n$ assets within a sample.

2.1. State-Stationarity

The homogeneous, i.e. time-stationary, process is determined by the infinitesimal generator of the process $\mathbf{Q} = (q_{hj})_{h,j=1,\dots,k}$ with transition intensities

$$q_{hj} = \lim_{u \rightarrow 0^+} \frac{P(X_u = j | X_0 = h)}{u}.$$

Note that $q_{hh} = -\sum_{j=1, j \neq h}^k q_{hj}$ and $q_{kj} = 0$. If transition to any other than the adjacent class is impossible, \mathbf{Q} is determined by elements on the first off-diagonals. It is useful to collect the indices for all non-zero intensities in set $\mathcal{I}_1 = \{(h, j) : h = 1, \dots, k-1; j = 1, \dots, k; |h-j| = 1\}$ and to define set $\mathcal{I}_2 = \mathcal{I}_1 \setminus \{(1, 2)\}$.

Definition 1. *Let the intensities on $[0, T]$ be*

$$q_{hj} = \begin{cases} q & \text{if } (h, j) = (1, 2) \\ q + \gamma_{hj} & \text{if } (h, j) \in \mathcal{I}_2 \end{cases}$$

with $q > 0$ and $\gamma_{hj} \in (-q, \infty)$.

In the one-parameter case of $\gamma_{hj} = 0$, the mapping $(h, j) \mapsto |h-j| \cdot q$ is a metric on \mathcal{I}_1 , whereas the same mapping for $\gamma_{hj} \neq 0$ is not.

We have no intention to analyze on asset level so that, compared to the analysis of all transition histories $\mathbf{X}_1, \dots, \mathbf{X}_n$, there is no loss of information when using the vector of initial ratings X_0^1, \dots, X_0^n together with the processes

$$N_{hj}(t) = \#\{s \in [0, t], i = 1, \dots, n | X_{s-}^i = h, X_s^i = j\}, \quad t \in [0, T], (h, j) \in \mathcal{I}_1$$

counting the number of transitions from state h to j until time t in the entire sample. Additionally, let the processes $Y_h(t)$ denote the number of assets in state h at time t . For large samples, this constitutes a clear reduction in the number of random processes. We impose two additional assumptions.

(A1) For fixed t and $n \rightarrow \infty$ in probability (\xrightarrow{P})

$$\frac{Y_h(t)}{n} \xrightarrow{P} m_h(t).$$

(A2) The counting processes N_{hj} must follow a multiplicative intensity model, i.e. with collection of q and γ_{hj} in vectors $\boldsymbol{\gamma} := (\gamma_{21}, \gamma_{23} \dots, \gamma_{k-1,k})' \in \mathbb{R}^{2k-4}$ and $\boldsymbol{\theta} := (q, \boldsymbol{\gamma})' \in \mathbb{R}^{2k-3}$ they have the intensity process

$$\lambda_{hj}(t; \boldsymbol{\theta}) = Y_h(t)q_{hj}, \quad (h, j) \in \mathcal{I}_1.$$

Due to the law of large numbers, assumption (A1) is fulfilled if the Markov processes are independent. Independence is also a sufficient condition for (A2).

As usual in the analysis of durations, only a partial likelihood can be evaluated (see Andersen et al., 1997, equation 2.7.4')

$$\begin{aligned} \log(L) &= \int_0^T \log(Y_1(t)) + \log(q) dN_{12}(t) \\ &+ \int_0^T \sum_{(h,j) \in \mathcal{I}_2} \log(Y_h(t)) + \log(q + \gamma_{hj}) dN_{hj}(t) \\ &- \int_0^T q Y_1(t) dt - \int_0^T \sum_{(h,j) \in \mathcal{I}_2} (q + \gamma_{hj}) Y_h(t) dt. \end{aligned} \quad (1)$$

We now collect some useful properties of model (A2) with the generator of Definition 1 and the log-likelihood (1).

Lemma 1. *For all $(h, j) \in \mathcal{I}_1$ and all $t \in [0, T]$ there exist first, second and third partial derivatives of the intensity processes $\lambda_{hj}(t; \boldsymbol{\theta})$ defined in (A2), of the logarithmic intensity processes $\log \lambda_{hj}(t; \boldsymbol{\theta})$ and of the logarithmic likelihood $\log L(\boldsymbol{\theta})$ with respect to q and γ_{rs} , $(r, s) \in \mathcal{I}_2$. These partial derivatives are continuous in $\boldsymbol{\theta}$.*

The proof of this and the following lemmas are collected in the Appendix. Denote by $\boldsymbol{\theta}_0$ the true parameter.

Lemma 2. *For all $(h, j), (r, s) \in \mathcal{I}_2$ it holds that*

$$\begin{aligned} \frac{1}{n} \int_0^T \sum_{(\mu, \nu) \in \mathcal{I}_1} \left(\frac{\partial}{\partial \gamma_{hj}} \log \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) \right) \left(\frac{\partial}{\partial \gamma_{rs}} \log \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) \right) \\ \cdot \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) dt = 0, \end{aligned} \quad (2)$$

if $(h, j) \neq (r, s)$ then, under assumption (A1)

$$\frac{1}{n} \int_0^T \sum_{(\mu, \nu) \in \mathcal{I}_1} \left(\frac{\partial}{\partial \gamma_{hj}} \log \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) \right)^2 \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) dt \xrightarrow[n \rightarrow \infty]{P} \sigma_{hj} \quad (3)$$

and

$$\begin{aligned} & \frac{1}{n} \int_0^T \sum_{(\mu, \nu) \in \mathcal{I}_1} \left(\frac{\partial}{\partial q} \log \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) \right) \left(\frac{\partial}{\partial \gamma_{hj}} \log \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) \right) \\ & \cdot \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) dt \xrightarrow[n \rightarrow \infty]{P} \sigma_{hj} \end{aligned} \quad (4)$$

with $0 < \sigma_{hj} < \infty$.

Furthermore there exists a $\sigma_q (> 0)$, such that

$$\frac{1}{n} \int_0^T \sum_{(\mu, \nu) \in \mathcal{I}_1} \left(\frac{\partial}{\partial q} \log \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) \right)^2 \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) dt \xrightarrow[n \rightarrow \infty]{P} \sigma_q. \quad (5)$$

Lemma 3. For all $(h, j) \in \mathcal{I}_2$ and all $\varepsilon > 0$ it holds that

$$\begin{aligned} & \frac{1}{n} \int_0^T \sum_{(\mu, \nu) \in \mathcal{I}_1} \left(\frac{\partial}{\partial q} \log \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) \right)^2 \\ & \cdot \mathbf{1} \left(\left| \frac{1}{\sqrt{n}} \frac{\partial}{\partial q} \log \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) \right| > \varepsilon \right) \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) dt \end{aligned} \quad (6)$$

and

$$\begin{aligned} & \frac{1}{n} \int_0^T \sum_{(\mu, \nu) \in \mathcal{I}_1} \left(\frac{\partial}{\partial \gamma_{hj}} \log \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) \right)^2 \\ & \cdot \mathbf{1} \left(\left| \frac{1}{\sqrt{n}} \frac{\partial}{\partial \gamma_{hj}} \log \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) \right| > \varepsilon \right) \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) dt \end{aligned} \quad (7)$$

converge in probability to zero for $n \rightarrow \infty$.

Lemma 4. The matrix $\boldsymbol{\Sigma} = (\Sigma_{ij})_{i,j=1,\dots,2k-3} \in \mathbb{R}^{(2k-3) \times (2k-3)}$, defined by

$$\boldsymbol{\Sigma} := \begin{pmatrix} \sigma_q & \sigma_{21} & \sigma_{23} & \sigma_{32} & \dots & \sigma_{k-1,k} \\ \sigma_{21} & \sigma_{21} & 0 & 0 & \dots & 0 \\ \sigma_{23} & 0 & \sigma_{23} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \sigma_{k-1,k-2} & 0 & \dots & 0 & \sigma_{k-1,k-2} & 0 \\ \sigma_{k-1,k} & 0 & \dots & 0 & 0 & \sigma_{k-1,k} \end{pmatrix}$$

with values σ_{hj} , $(h, j) \in \mathcal{I}_2$, and σ_q from Lemma 2, is positive definite.

Lemma 5. *There exist neighborhoods Θ_0^q and Θ_0^{rs} , $(r, s) \in \mathcal{I}_2$, of q_0 and γ_{rs0} so that, with $\Theta_0 := \Theta_0^q \times \Theta_0^{21} \times \dots \times \Theta_0^{k-1,k}$, for all $(h, j) \in \mathcal{I}_1$ there is a constant H_{hj} , not depending on $\boldsymbol{\theta}$, such that for all $t \in [0, T]$ and all $m, r, s = 1, \dots, 2k - 3$*

$$\sup_{\boldsymbol{\theta} \in \Theta_0} \left| \frac{\partial^3}{\partial \theta_m \partial \theta_r \partial \theta_s} \log \lambda_{hj}(t; \boldsymbol{\theta}) \right| \leq H_{hj}. \quad (8)$$

Furthermore exists a positive constant $C < \infty$, such that

$$\frac{1}{n} \int_0^T \sum_{(\mu, \nu) \in \mathcal{I}_1} H_{\mu\nu} \cdot \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) dt \xrightarrow[n \rightarrow \infty]{P} C \quad (9)$$

and for all $\varepsilon > 0$ the following holds

$$\frac{1}{n} \int_0^T \sum_{(\mu, \nu) \in \mathcal{I}_1} H_{\mu\nu} \cdot \mathbf{1} \left(\frac{1}{\sqrt{n}} \cdot (H_{\mu\nu})^{\frac{1}{2}} > \varepsilon \right) \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) dt \xrightarrow[n \rightarrow \infty]{P} 0. \quad (10)$$

For the subsequent results, we can denote for $j = 1, \dots, 2k - 3$, the score statistic by $U_T^j(\boldsymbol{\theta})$, minus the second partial derivative of the log-likelihood (1) by $\mathcal{J}_T^{jl}(\boldsymbol{\theta})$ and the third derivatives by $R_T^{jlm}(\boldsymbol{\theta})$, for $j, l, m = 1, \dots, 2k - 3$.

Lemma 6. *For all $j = 1, \dots, 2k - 3$ the score statistic $U_T^j(\boldsymbol{\theta}_0)/n$ converges for $n \rightarrow \infty$ towards zero in probability.*

Lemma 7. *With the definition of Σ from Lemma 4, the score statistic vector $\mathbf{U}_T(\boldsymbol{\theta}_0)/\sqrt{n}$ converges in distribution (\xrightarrow{d}) to a Gaussian random vector*

$$\frac{1}{\sqrt{n}} \mathbf{U}_T(\boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(\mathbf{0}, \Sigma).$$

Lemma 8. *The matrix $\mathcal{J}_T(\boldsymbol{\theta}_0)/n$ converges, for $n \rightarrow \infty$, in a stochastic sense towards the Σ defined in Lemma 4.*

Lemma 9. *For $C < \infty$, defined in Lemma 5, $\forall j, l, m = 1, \dots, 2k - 3$ and uniformly $\forall \boldsymbol{\theta} \in \Theta_0$, it holds that $|R_T^{jlm}(\boldsymbol{\theta})/n|$ has asymptotically almost surely bound C .*

2.2. Time-Stationarity

When assessing the question whether transition intensities are time-stationary, there is no need to restrict the model to adjacent class transitions. A model that can distinguish a time-stationary from an in-stationary process proposed in Weißbach and Walter (2010) defines $q_{hj}(t)$ as a step function, i.e. imposes a piecewise stationary model.

Definition 2. Let the intensities on $[0, T]$ with the given change-points $0 = t_0 < t_1 < \dots < t_{b-1} < t_b = T$ be

$$q_{hj}(t) = \mathbb{1}_{[0, t_1)}(t)q_{hj} + \sum_{l=2}^b \mathbb{1}_{[t_{l-1}, t_l)}(t)(q_{hj} + \delta_{hjl})$$

with $q_{hj} > 0$ and $\delta_{hjl} \in (-q_{hj}, \infty)$, $l = 2, \dots, b$.

The counting processes of Section 2.1 now change to

$$N_{hj}(t) = \#\{s \in [0, t], i = 1, \dots, n | X_{s-}^i = h, X_s^i = j\}, t \in [0, T], j \neq h$$

with condition (A2) being extended to an index set $\mathcal{I} = \{j \neq h, h \neq k\}$. The likelihood alters to

$$\begin{aligned} \log(L) &= \int_0^{t_1} \sum_{j \neq h} \log(Y_h(t)) + \log(q_{hj}) dN_{hj}(t) \\ &+ \sum_{l=2}^b \int_{t_{l-1}}^{t_l} \sum_{j \neq h} \log(Y_h(t)) + \log(q_{hj} + \delta_{hjl}) dN_{hj}(t) \\ &- \sum_{j \neq h} \left[\int_0^{t_1} Y_h(t) q_{hj} dt + \sum_{l=2}^b \int_{t_{l-1}}^{t_l} Y_h(t) (q_{hj} + \delta_{hjl}) dt \right] \end{aligned}$$

Straight forward calculations yield that Lemma 1 is now equally true with \mathcal{I}_1 replaced by \mathcal{I} (Lemma 4.3 in Weißbach and Walter (2010)). Lemma 2 becomes the first part of Lemma 4.4, the adapted version of Lemma 3 is Lemma 4.5, and Lemma 4 results in the second part of Lemma 4.4 together with Lemma 4.1. Lemma 5 alters to Lemma 4.7. Lemma 6 is also true, again with the inequality of Lengart (1977) and the adapted version of Lemma 2. Lemma 7 follows from the Theorem of Rebolledo (1980) and the adapted versions of Lemmas 2 and 3. Lemma 8 and 9 are now Lemma 4.6. and Lemma 4.8, respectively.

3. Estimation and Testing

3.1. State-Stationarity

In order to test for equally spaced classes in Model 1, we formulate the null hypothesis

$$H_0 : \gamma_{hj} = 0 \quad \forall h, j \in \mathcal{I}_2, \quad (11)$$

with the alternative

$$H_1 : \exists \gamma_{hj} \neq 0. \quad (12)$$

Our aim is maximum likelihood estimation in the restricted and the unrestricted parameter space. Finally, we construct a test for the one-parameter, metric model with the likelihood ratio.

Theorem 1. *Given $\int_0^T Y_\mu(s)ds > 0$ for all $\mu = 1, \dots, k-1$, the maximum likelihood estimators in the unrestricted parameter space for q and γ_{hj} , $(h, j) \in \mathcal{I}_2$, are*

$$\hat{q} = \frac{N_{12}(T)}{\int_0^T Y_1(s)ds} \quad (13)$$

and

$$\hat{\gamma}_{hj} = \frac{N_{hj}(T)}{\int_0^T Y_h(s)ds} - \frac{N_{12}(T)}{\int_0^T Y_1(s)ds}. \quad (14)$$

If the numerators in (13) or (14) are zero, the quotient is set to zero. Define $N_{\bullet\bullet}(t) := \sum_{(h,j) \in \mathcal{I}_1} N_{hj}(t)$, $t \in [0, T]$. Then, in the restricted parameter space, i.e. under null hypothesis (11), the ML estimator of q is

$$\hat{q} = \frac{N_{\bullet\bullet}(T)}{\sum_{(h,j) \in \mathcal{I}_1} \int_0^T Y_h(s)ds}. \quad (15)$$

The resulting likelihood ratio is

$$\frac{L(\hat{q}, \mathbf{0})}{L(\hat{q}, \hat{\gamma})} = \prod_{\substack{t \in [0, T]: \exists (\mu, \nu) \in \mathcal{I}_1, \\ N_{\mu\nu}(t-) \neq N_{\mu\nu}(t)}} \left(\left(\frac{\hat{q}}{\hat{q}} \right)^{\Delta N_{12}(t)} \prod_{(h,j) \in \mathcal{I}_2} \left(\frac{\hat{q}}{\hat{q} + \hat{\gamma}_{hj}} \right)^{\Delta N_{hj}(t)} \right).$$

Proof. For the unrestricted parameter space, compare the partial derivatives of $\log L(\boldsymbol{\theta})$ with respect to q and γ_{hj} , $(h, j) \in \mathcal{I}_2$ with zero. For $\int_0^T Y_\mu(s) ds > 0$, $\mu = 1, \dots, k-1$, this is, due to Lemma 1 (especially (A.3)), equivalent to

$$\frac{N_{12}(T)}{q} + \sum_{(h,j) \in \mathcal{I}_2} \frac{N_{hj}(T)}{q + \gamma_{hj}} - \sum_{(h,j) \in \mathcal{I}_1} \int_0^T Y_h(s) ds = 0$$

and

$$\frac{N_{hj}(T)}{q + \gamma_{hj}} - \int_0^T Y_h(s) ds = 0.$$

Hence (13) and (14) hold. For the sufficient condition of a maximum, the Hessian of $\log L(\boldsymbol{\theta})$, evaluated in $\hat{\boldsymbol{\theta}}$, needs to be negative definite. The second derivatives have already been calculated for Lemma 1, see (A.4), and inserting $\hat{\boldsymbol{\theta}}$ results in

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{(\partial q)^2} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = - \sum_{(h,j) \in \mathcal{I}_1} \frac{\left(\int_0^T Y_h(s) ds \right)^2}{N_{hj}(T)}$$

and for all $(h, j) \in \mathcal{I}_2$, in

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{(\partial \gamma_{hj})^2} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial q \partial \gamma_{hj}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = - \frac{\left(\int_0^T Y_h(s) ds \right)^2}{N_{hj}(T)}.$$

Any other second derivative is zero. Hence, the Hessian has the same form as $\boldsymbol{\Sigma}$ (multiplied by -1) in Lemma 4. The latter was found to be positive definite, so that the Hessian is negative definite. \hat{q} and $\hat{\gamma}_{hj}$, $(h, j) \in \mathcal{I}_2$, are ML-estimators for q and γ_{hj} in the unrestricted space.

In order to calculate the ML-estimator in the restricted parameter space, the derivative of $\log L(q, \mathbf{0})$ needs to be set to zero, resulting in (15). The sufficient condition is fulfilled, because the second derivative, see again (A.4), evaluated in \hat{q} , is negative

$$-\frac{N_{\bullet\bullet}(T)}{q^2} \Big|_{q=\hat{q}} = - \frac{\left(\sum_{(h,j) \in \mathcal{I}_1} \int_0^T Y_h(s) ds \right)^2}{N_{\bullet\bullet}(T)} < 0. \quad (16)$$

Inserting \hat{q} and $\hat{\gamma}_{hj}$, $(h, j) \in \mathcal{I}_2$, as well as \hat{q} in the likelihood, see (1), yields

$$L(\hat{q}, \hat{\gamma}) = \left[\prod_t (\hat{q} \cdot Y_1(t))^{\Delta N_{12}(t)} \prod_{(h,j) \in \mathcal{I}_2} ((\hat{q} + \hat{\gamma}_{hj}) \cdot Y_h(t))^{\Delta N_{hj}(t)} \right] \cdot \exp \left(- \sum_{(h,j) \in \mathcal{I}_1} \frac{N_{hj}(T)}{\int_0^T Y_h(s) ds} \cdot \int_0^T Y_h(s) ds \right)$$

and

$$L(\hat{q}, 0) = \left[\prod_t \prod_{(h,j) \in \mathcal{I}_1} (\hat{q} \cdot Y_h(t))^{\Delta N_{hj}(t)} \right] \cdot \exp(-N_{\bullet\bullet}(T)),$$

where $\Delta N_{hj}(t) := N_{hj}(t) - N_{hj}(t-)$, resulting in the requested likelihood ratio. ■

It is evident that in the estimation of q_{hj} as $\hat{q} + \hat{\gamma}_{hj}$, in the numerator, only transitions from rating h to j are used, whereas for \hat{q} , transitions to any adjacent rating class enter. Therefore, in the one-parameter model, default probabilities are estimable, even if no default has occurred in the portfolio.

The following two theorems demonstrate good behavior of the ML-estimate.

Theorem 2. *For a sample of Markov processes with an intensity as in Definition 1, let assumptions (A1) and (A2) be fulfilled. Then, in the unrestricted parameter space, the estimator $\hat{\theta}$ and in the restricted space, the estimator \hat{q} of Theorem 1 are consistent.*

Proof. Due to Lemma 1 and $\theta \in \Theta_0$, defined in Lemma 5, the Taylor expansion of the score statistic can now be written as

$$U_T^j(\theta) = U_T^j(\theta_0) - \sum_{l=1}^{2k-3} (\theta_l - \theta_{l0}) \mathcal{J}_T^{jl}(\theta_0) + \frac{1}{2} \sum_{l=1}^{2k-3} \sum_{m=1}^{2k-3} (\theta_l - \theta_{l0}) (\theta_m - \theta_{m0}) R_T^{jlm}(\theta^{j*}), \quad (17)$$

with θ^{j*} on the connecting line between θ and θ_0 . Matrix Σ of Lemma 4 is positive definite, so that there exists a constant $\beta > 0$, such that, for all

$\mathbf{t} \in \mathbb{R}^{2k-3}$ with $|\mathbf{t}| = 1$ it holds that $\mathbf{t}'\Sigma\mathbf{t} \geq \beta$. Here and throughout $|\cdot|$ denotes the Euclidean norm. For any $\varepsilon > 0$, one can chose some $\delta = \delta(\varepsilon) > 0$ fulfilling $\delta < \varepsilon$, $\{\boldsymbol{\theta} : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \leq \delta\} \subseteq \Theta_0$ and $\delta < \beta/(3(2k-3)^3(C+1))$. As a consequence of Lemma 6 to Lemma 9, there exists a $n_0(\varepsilon) \in \mathbb{N}$, such that, with a probability larger than $1 - \varepsilon$, for all $n \geq n_0(\varepsilon)$ hold the inequalities for $j, l, m = 1, \dots, 2k-3$ (where applicable)

$$\left| \frac{1}{n} U_T^j(\boldsymbol{\theta}_0) \right| < \delta^2, \quad 0 \leq \left| \frac{1}{n} R_T^{jlm}(\boldsymbol{\theta}^{j*}) \right| < C+1 \quad (18)$$

and

$$\left| \frac{1}{n} \mathcal{J}_T^{jl}(\boldsymbol{\theta}_0) - \Sigma_{jl} \right| < \delta. \quad (19)$$

Dividing (17) by n results in

$$\begin{aligned} & \frac{1}{n} U_T^j(\boldsymbol{\theta}) + \sum_{l=1}^{2k-3} (\theta_l - \theta_{l0}) \frac{1}{n} \mathcal{J}_T^{jl}(\boldsymbol{\theta}_0) \\ &= \underbrace{\frac{1}{n} U_T^j(\boldsymbol{\theta}_0) + \frac{1}{2} \sum_{l=1}^{2k-3} \sum_{m=1}^{2k-3} (\theta_l - \theta_{l0}) (\theta_m - \theta_{m0}) \frac{1}{n} R_T^{jlm}(\boldsymbol{\theta}^{j*})}_{(*)} \\ \Leftrightarrow & \frac{1}{n} U_T^j(\boldsymbol{\theta}) + \sum_{l=1}^{2k-3} (\theta_l - \theta_{l0}) \left(\Sigma_{jl} + \frac{1}{n} \mathcal{J}_T^{jl}(\boldsymbol{\theta}_0) - \Sigma_{jl} \right) = (*) \\ \Leftrightarrow & \frac{1}{n} U_T^j(\boldsymbol{\theta}) + \sum_{l=1}^{2k-3} (\theta_l - \theta_{l0}) \Sigma_{jl} \\ &= (*) + \sum_{l=1}^{2k-3} (\theta_l - \theta_{l0}) \left(\Sigma_{jl} - \frac{1}{n} \mathcal{J}_T^{jl}(\boldsymbol{\theta}_0) \right). \end{aligned}$$

Due to $|\theta_l - \theta_{l0}| \leq |\boldsymbol{\theta} - \boldsymbol{\theta}_0|$, if $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| \leq \delta$ then (18) and (19) yield $|\frac{1}{n} U_T^j(\boldsymbol{\theta}) + \sum_{l=1}^{2k-3} (\theta_l - \theta_{l0}) \Sigma_{jl}| \leq |(*)| + \sum_{l=1}^{2k-3} |\theta_l - \theta_{l0}| |\mathcal{J}_T^{jl}(\boldsymbol{\theta}_0)/n - \Sigma_{jl}| \leq \delta^2 + \delta(2k-3)|\boldsymbol{\theta} - \boldsymbol{\theta}_0| + (2k-3)^2 |\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2 \frac{C+1}{2} \leq 3(2k-3)^2(C+1)\delta^2$. The inequality holds equally for any component of the vector. With the mapping $h(\mathbf{x}) = \delta \cdot \mathbf{x} + \boldsymbol{\theta}_0$

and because $-1 < (\theta_j - \theta_{j0})/\delta < 1$ it follows for $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| = \delta$ that

$$\begin{aligned}
& \sum_{j=1}^{2k-3} \frac{1}{n} U_T^j \left(h \left(\frac{\boldsymbol{\theta} - \boldsymbol{\theta}_0}{\delta} \right) \right) \frac{\theta_j - \theta_{j0}}{\delta} = \sum_{j=1}^{2k-3} \frac{1}{n} U_T^j(\boldsymbol{\theta}) \frac{\theta_j - \theta_{j0}}{\delta} \\
\leq & \underbrace{- \sum_{j=1}^{2k-3} \sum_{l=1}^{2k-3} \Sigma_{jl} \frac{\theta_j - \theta_{j0}}{\delta} \frac{\theta_l - \theta_{l0}}{\delta}}_{= \frac{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'}{|\boldsymbol{\theta} - \boldsymbol{\theta}_0|} \cdot \boldsymbol{\Sigma} \cdot \frac{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)}{|\boldsymbol{\theta} - \boldsymbol{\theta}_0|} \geq \beta} \delta + \sum_{j=1}^{2k-3} 3(2k-3)^2 (C+1) \delta^2 \\
\leq & -\beta \cdot \delta + 3(2k-3)^3 (C+1) \delta^2 =: (**).
\end{aligned}$$

and because $\delta < \beta/(3(2k-3)^3(C+1))$ follows

$$(**) = (-\beta + 3(2k-3)^3(C+1)\delta)\delta < (-\beta + \beta)\delta = 0$$

As a result of Aitchison and Silvey (1958), if, for a continuous mapping $g : \mathbb{R}^s \rightarrow \mathbb{R}^s$ it holds that $\mathbf{x}'g(\mathbf{x}) < 0$ for all \mathbf{x} where $|\mathbf{x}| = 1$, than there exists a $\tilde{\mathbf{x}}$ with $|\tilde{\mathbf{x}}| < 1$ and $g(\tilde{\mathbf{x}}) = 0$. Mapping $(\mathbf{U}_T \circ h)/n$ is continuous in $(\boldsymbol{\theta} - \boldsymbol{\theta}_0)/\delta$, so that, for $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| = \delta$ follows the existence of $\tilde{\boldsymbol{\theta}} \in \Theta_0$, such that, for all $n \geq n_0(\varepsilon)$ and with a probability larger than $1 - \varepsilon$,

$$\frac{1}{n} \mathbf{U}_T \left(h \left(\frac{\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0}{\delta} \right) \right) = \frac{1}{n} \mathbf{U}_T(\tilde{\boldsymbol{\theta}}) = 0,$$

where $|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| < \delta < \varepsilon$ is fulfilled. Moreover ε can be arbitrarily small, so that $\tilde{\boldsymbol{\theta}}$ is a consistent solution. On the other hand, ML-estimator $\hat{\boldsymbol{\theta}}$ from Theorem 1 is the unique solution of $\mathbf{U}_T(\boldsymbol{\theta}) = 0$ and, as a result, $\hat{\boldsymbol{\theta}} \equiv \tilde{\boldsymbol{\theta}}$. ■

Together with the consistency, $\hat{\boldsymbol{\theta}}$ and \hat{q} are asymptotically Gaussian.

Theorem 3. *For the unrestricted estimator $\hat{\boldsymbol{\theta}}$ of Theorem 1 and $\boldsymbol{\Sigma}$ of Lemma 4 holds*

$$\sqrt{n} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^{-1}).$$

Proof. One part of proving Theorem 3 is the martingale limit theorem of Rebolledo (1980), and Lemma 3 is the respective Lindeberg condition. Lemma 4 ensures that the covariance matrix $\boldsymbol{\Sigma}^{-1}$ is positive definite. Looking at this in more detail, analogous to (17) we may, for any $j = 1, \dots, 2k -$

3, expand the score statistic U_T^j , divided by \sqrt{n} , in $\boldsymbol{\theta}_0$ due to Lemma 1. Evaluation in the ML-estimator $\hat{\boldsymbol{\theta}}$ yields

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}} U_T^j(\hat{\boldsymbol{\theta}}) = \frac{1}{\sqrt{n}} U_T^j(\boldsymbol{\theta}_0) - \sum_{l=1}^{2k-3} \sqrt{n} (\hat{\theta}_l - \theta_{l0}) \frac{1}{n} \mathcal{J}_T^{jl}(\boldsymbol{\theta}_0) \\ &\quad + \frac{1}{2} \sum_{l=1}^{2k-3} \sum_{m=1}^{2k-3} \sqrt{n} (\hat{\theta}_l - \theta_{l0}) (\hat{\theta}_m - \theta_{m0}) \frac{1}{n} R_T^{jlm}(\boldsymbol{\theta}^{j*}). \end{aligned}$$

As $\hat{\boldsymbol{\theta}}$ is a consistent estimator and $\boldsymbol{\theta}^{j*}$ is on the line between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$, $\hat{\boldsymbol{\theta}}$ and finally $\boldsymbol{\theta}^{j*}$ will be in Θ_0 with a probability approaching one. Taking the absolute value, using the triangular inequality and Lemma 9, we have the stochastic upper limit

$$\begin{aligned} &\left| -\frac{1}{\sqrt{n}} U_T^j(\boldsymbol{\theta}_0) + \sum_{l=1}^{2k-3} \sqrt{n} (\hat{\theta}_l - \theta_{l0}) \frac{1}{n} \mathcal{J}_T^{jl}(\boldsymbol{\theta}_0) \right| \\ &\leq \frac{1}{2} C \sum_{l=1}^{2k-3} \sqrt{n} |\hat{\theta}_l - \theta_{l0}| \sum_{m=1}^{2k-3} |\hat{\theta}_m - \theta_{m0}|. \end{aligned}$$

In matrix notation

$$\begin{aligned} &\left| \sqrt{n} \cdot \frac{1}{n} \mathcal{J}_T(\boldsymbol{\theta}_0) \cdot (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \frac{1}{\sqrt{n}} \mathbf{U}_T(\boldsymbol{\theta}_0) \right| \\ &\leq \frac{C(2k-3)}{2} \sum_{l=1}^{2k-3} \sqrt{n} |\hat{\theta}_l - \theta_{l0}| \sum_{m=1}^{2k-3} |\hat{\theta}_m - \theta_{m0}| \\ &\leq \left(\frac{C(2k-3)^2}{2} \sum_{l=1}^{2k-3} |\hat{\theta}_l - \theta_{l0}| \right) \cdot \left| \sqrt{n} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right| \\ &= \left(\frac{C(2k-3)^2}{2} \sum_{l=1}^{2k-3} |\hat{\theta}_l - \theta_{l0}| \right) \cdot \left| \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma} \sqrt{n} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right| \\ &\leq \left(\frac{C(2k-3)^2}{2} \sum_{l=1}^{2k-3} |\hat{\theta}_l - \theta_{l0}| \right) \cdot \|\boldsymbol{\Sigma}^{-1}\|_2 \cdot \left| \boldsymbol{\Sigma} \sqrt{n} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right|. \end{aligned}$$

The second inequality follows from $|\hat{\theta}_m - \theta_{m0}| \leq |\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|$. The spectral norm, denoted by $\|\cdot\|_2$, is the consistent matrix norm of the Euclidean norm. Due to the consistency of $\hat{\boldsymbol{\theta}}$, the leading term converges to zero and, together with Lemma 8, we obtain

$$\left| \frac{1}{\sqrt{n}} \mathbf{U}_T(\boldsymbol{\theta}_0) - \boldsymbol{\Sigma} \sqrt{n} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right| \leq \varepsilon_n \cdot \left| \boldsymbol{\Sigma} \sqrt{n} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right|.$$

Theorem 10.1 from Billingsley (1961) ensures $\Sigma\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(\mathbf{0}, \Sigma)$. And with the Continuous Mapping Theorem, we finally obtain

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(\Sigma^{-1}\mathbf{0}, \Sigma^{-1}\Sigma(\Sigma^{-1})') = \mathcal{N}(\mathbf{0}, \Sigma^{-1}). \quad \blacksquare$$

Clearly, the asymptotic normality of the estimate vector may be used to construct confidence ellipsoids for the parameter vector, resulting in confidence sets for the rating transition probabilities comparable to those in Christensen et al. (2004). For instance, confidence sets for the γ_{hj} can be used for inclusion rules, in order to confirm or reject both the equality hypothesis (3) and the equivalence hypothesis (see Munk and Weißbach, 1999). For the unrestricted parameter space of Definition 1 the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\theta}}/\sqrt{n}$ as given in Theorem 3 is the inverse of Σ defined in Lemma 4. Let us extend the notation $\boldsymbol{\theta}' = (q, \boldsymbol{\gamma}')$ by

$$\Sigma^{-1} := \begin{pmatrix} \Sigma_q & \Sigma_{q\boldsymbol{\gamma}} \\ \Sigma_{\boldsymbol{\gamma}q} & \Sigma_{\boldsymbol{\gamma}} \end{pmatrix},$$

where $\Sigma_q \in \mathbb{R}$ and $\Sigma_{\boldsymbol{\gamma}} \in \mathbb{R}^{(2k-4) \times (2k-4)}$. By results on the inverse of a partitioned matrix it is $\Sigma_q = \sigma_q - \sum_{(h,j) \in \mathcal{I}_2} \sigma_{hj}$, $\Sigma_{q\boldsymbol{\gamma}} = -\Sigma_q \mathbf{1}_{1 \times (2k-4)}$ and $\Sigma_{\boldsymbol{\gamma}} = \text{diag}(\sigma_{21}^{-1}, \sigma_{23}^{-1}, \sigma_{32}^{-1}, \dots, \sigma_{k-1,k}^{-1}) + \Sigma_q \mathbf{1}_{(2k-4) \times (2k-4)}$. As usual, *diag* denotes a diagonal matrix and $\mathbf{1}$ a matrix which contains 1 in every element. By standard arguments, the elements of Σ are consistently estimated by the Fisher information, i.e. minus the second derivatives of the log-Likelihood given in (A.4), with inserting the estimates $\hat{\boldsymbol{\theta}}$ of Theorem 1 and dividing by n .

For the restricted parameter space, the second derivative of the likelihood (1) yields as variance of (15) (see (16))

$$\text{Var}(\hat{q}) = \frac{N_{\bullet\bullet}(T)}{n \left(\sum_{(h,j) \in \mathcal{I}_1} \int_0^T Y_h(s) ds \right)^2}. \quad (20)$$

Additionally, Wald and score tests can be derived from the asymptotic

normality. However, we restrict our analysis to the likelihood ratio test as an example.

Corollary 4. *Under hypothesis (11), the profile partial likelihood ratio has an asymptotic distribution*

$$-2 \log \frac{L(\hat{q}, \mathbf{0})}{L(\hat{q}, \hat{\gamma})} \stackrel{n \rightarrow \infty}{\sim} \chi_{2k-4}^2. \quad (21)$$

Proof. Note first that, evaluated in $\hat{\boldsymbol{\theta}}$ from Theorem 1, \mathcal{J}_T/n is a consistent estimator for $\boldsymbol{\Sigma}$. The reason is that for all $j, l = 1, \dots, 2k-3$, a Taylor expansion of \mathcal{J}_T^{jl}/n around the true $\boldsymbol{\theta}_0$ is

$$\frac{1}{n} \mathcal{J}_T^{jl}(\hat{\boldsymbol{\theta}}) = \frac{1}{n} \mathcal{J}_T^{jl}(\boldsymbol{\theta}_0) - \sum_{m=1}^{2k-3} \left(\hat{\theta}_m - \theta_{m0} \right) \frac{1}{n} R_T^{jlm}(\boldsymbol{\theta}^{j*}).$$

The second term on the right hand side converges, due to Theorem 2 and Lemma 9, for $n \rightarrow \infty$, to zero in a stochastic sense, so that Lemma 8 yields the consistency.

Noting $\mathbf{U}_T(\hat{\boldsymbol{\theta}}) = 0$, consider the Taylor expansion of $\log L$ at $\hat{\boldsymbol{\theta}}$, evaluated in $\boldsymbol{\theta}_0$,

$$\begin{aligned} \log L(\boldsymbol{\theta}_0) &= \log L(\hat{\boldsymbol{\theta}}) - \frac{1}{2} \sum_{j=1}^{2k-3} \sum_{l=1}^{2k-3} \left(\theta_{j0} - \hat{\theta}_j \right) \left(\theta_{l0} - \hat{\theta}_l \right) \mathcal{J}_T^{jl}(\hat{\boldsymbol{\theta}}) \\ &+ \frac{1}{6} \sum_{j=1}^{2k-3} \sum_{l=1}^{2k-3} \sum_{m=1}^{2k-3} \left(\theta_{j0} - \hat{\theta}_j \right) \left(\theta_{l0} - \hat{\theta}_l \right) \left(\theta_{m0} - \hat{\theta}_m \right) R_T^{jlm}(\boldsymbol{\theta}^\dagger). \end{aligned}$$

The last summand converges as a consequence of Theorems 2 and 3, together with Lemma 9 and Slutsky's lemma, for $n \rightarrow \infty$, towards zero in a stochastic sense. We can replace $\mathcal{J}_T(\hat{\boldsymbol{\theta}})/n$ by $\boldsymbol{\Sigma}$, because the resulting error can be absorbed in the last summand, due to Lemma 8 and Slutsky's lemma. For convergence in distribution, the last summand can be ignored, again due to Slutsky's lemma, so that

$$-2 \log \frac{L(\boldsymbol{\theta}_0)}{L(\hat{\boldsymbol{\theta}})} \approx \sqrt{n} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \right)' \boldsymbol{\Sigma} \sqrt{n} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \right).$$

Let diagonal matrix $\mathbf{\Lambda}$ contain the (positive) eigenvalues of $\mathbf{\Sigma}_\gamma$ and let $\mathbf{\Gamma}$ be the matrix of the respective eigenvectors, such that $\mathbf{\Sigma}_\gamma^{1/2} := \mathbf{\Gamma}\mathbf{\Lambda}^{1/2}\mathbf{\Gamma}'$. Standard arguments of the profile likelihood ratio yield, with notation \mathbf{I} as the identity matrix:

$$-2 \log \frac{L(\hat{\mathbf{q}}, \mathbf{0})}{L(\hat{\mathbf{q}}, \hat{\boldsymbol{\gamma}})} \approx (\sqrt{n}\hat{\boldsymbol{\gamma}})' \mathbf{\Sigma}_\gamma^{-1} (\sqrt{n}\hat{\boldsymbol{\gamma}}) = \left(\mathbf{\Sigma}_\gamma^{-1/2} \sqrt{n}\hat{\boldsymbol{\gamma}} \right)' \left(\mathbf{\Sigma}_\gamma^{-1/2} \sqrt{n}\hat{\boldsymbol{\gamma}} \right).$$

Theorem 3 and the continuous mapping theorem result in $\mathbf{\Sigma}_\gamma^{-1/2} \sqrt{n}\hat{\boldsymbol{\gamma}} \stackrel{H_0}{\mathcal{N}(\mathbf{0}, \mathbf{I})}$, for $n \rightarrow \infty$. ■

3.2. Time-Stationarity

The null hypothesis for time-stationarity in Definition 2 is

$$H_0 : \delta_{hj2} = \dots = \delta_{hjb} = 0 \quad \forall j \neq h, h \neq k.$$

Theorem 1 becomes Theorem 3 of Weißbach and Walter (2010). The restricted estimator obviously is $\hat{q}_{hj} = N_{hj}(T) / \int_0^T Y_h(t) dt$. The unrestricted is for $l = 2, \dots, b$

$$\tilde{q}_{hj} = \frac{N_{hj}(t_1^-)}{\int_0^{t_1} Y_h(t) dt} \quad \text{and} \quad \hat{\delta}_{hjl} = \frac{N_{hj}(t_l^-) - N_{hj}(t_{l-1}^-)}{\int_{t_{l-1}}^{t_l} Y_h(t) dt} - \tilde{q}_{hj}.$$

For Theorem 2 there is no analogue in Weißbach and Walter (2010), it is an assumption there. However, as in the analysis of state-stationarity the proof follows from Aitchison and Silvey (1958), the uniqueness of the estimator from the adapted version of Theorem 1, and adapted versions of Lemmas 1 and 5 to 9. Theorem 3 and Corollary 4 are the alteration of Theorem 1 in Weißbach and Walter (2010) and Corollary 2, respectively.

4. Simulations

The finite sample properties of the likelihood ratio test on state-stationarity developed in Corollary 4 are now studied in a Monte Carlo simulations. Size

and power are studied under conditions that will be realistic for the application Section 5. Furthermore, the approximation of the parameter estimate with the Gaussian distribution in Theorem 3 is assessed.

The path $t \mapsto X(t, \omega)$, $\omega \in \Omega$ of a Markov process from Definition 1 has transition times $J_0 := 0$, $J_{n+1} := \inf \{t \geq J_n : X_t \neq X_{J_n}\}$, $n \in \mathbb{N}_0$, with $\inf \emptyset = \infty$. Define according durations (see also Figure 1)

$$D_n := \begin{cases} J_n - J_{n-1}, & \text{falls } J_{n-1} < \infty \\ \infty, & \text{sonst} \end{cases}, \quad n \in \mathbb{N}. \quad (22)$$

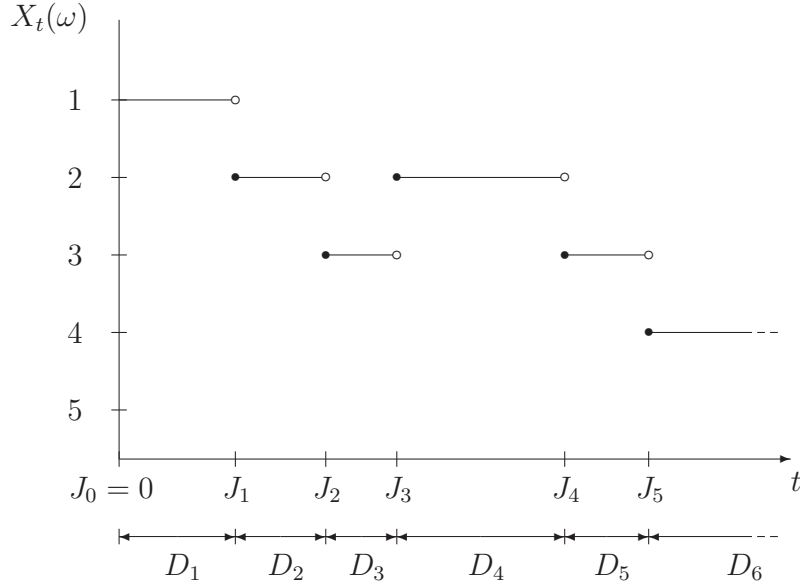


Figure 1: Transition times J_i and durations $D_{i'}$ of Markov process \mathbf{X} with $k = 5$ states and intensities of Definition 1

The embedded discrete-time Markov chain $\mathbf{Z} := (Z_i, i \in \mathbb{N}_0)$ with $Z_i := X_{J_i}$, $i \in \mathbb{N}_0$ allows for an alternative representation of the markov process \mathbf{X} . A matrix $\mathbf{Q} \in \mathbb{R}^{k \times k}$ is the generator of \mathbf{X} if and only if the durations D_1, \dots, D_i , conditional on Z_0, \dots, Z_{i-1} are independent exponentially distributed random variables with parameters $-q_{Y_0 Y_0}, \dots, -q_{Y_{i-1} Y_{i-1}}$ and the

embedded markov chain \mathbf{Z} has transition matrix $\Pi = (\pi_{hj})_{h,j=1,\dots,k}$ with entries

$$\pi_{hj} := \begin{cases} -\frac{q_{hj}}{q_{hh}}, & \text{if } h \neq j \text{ and } q_{hh} \neq 0 \\ 0, & \text{if } h \neq j \text{ and } q_{hh} = 0 \end{cases}$$

and π_{hh} is zero if $q_{hh} \neq 0$ or one otherwise. Hence for a Markov process starting in $h_0 \neq k$ a path can be simulated on $[0, T]$ as follows.

1. Set $i := 1$.
2. Draw the duration in state h_{i-1} as exponentially distributed D_i with parameter

$$-q_{h_{i-1}h_{i-1}} = \begin{cases} q_{12}, & \text{if } h_{i-1} = 1 \\ q_{h_{i-1},h_{i-1}-1} + q_{h_{i-1},h_{i-1}+1}, & \text{if } h_{i-1} \in \{1, \dots, k-1\} \end{cases}.$$

3. If $\sum_{i'=1}^i D_{i'} \leq T$, draw Bernoulli-distributed random variable that assumes value 1 with probability $-q_{h_{i-1},h_{i-1}+1}/q_{h_{i-1}h_{i-1}}$. If the value becomes 1, the process migrates in time $\sum_{i'=1}^n D_{i'}$ to state $h_i := h_{i-1} + 1$, and otherwise to state $h_i := h_{i-1} - 1$. If $\sum_{i'=1}^i D_{i'} > T$ stop.
4. If $h_i \neq k$, set $i := i + 1$ and go to 2. If $h_i = k$ stop, state k is absorbing.

Repeating the algorithm results in n sample paths $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$, where sample of $n = 100$, $n = 1,000$ and $n = 10,000$ observations are studied. The number of processes starting in the the $k - 1$ states are equal. The follow-up period is $T = 5$ and $T = 10$ years. Size and power are calculated as percentage of likelihood ratio tests rejecting among $n_{sim} = 10,000$ simulated samples. Consider now a simulation for a one-parameter generator \mathbf{Q}^0 with five state and $q_{hj}^0 \equiv 0.076$ for all $(h, j) \in \mathcal{I}_1$. The magnitude of the generator entries are taken from an empirical study on rating transitions of Weißbach et al. (2009). The processes follow Definition 1 and have $\gamma_{hj}^0 = 0$. In view of

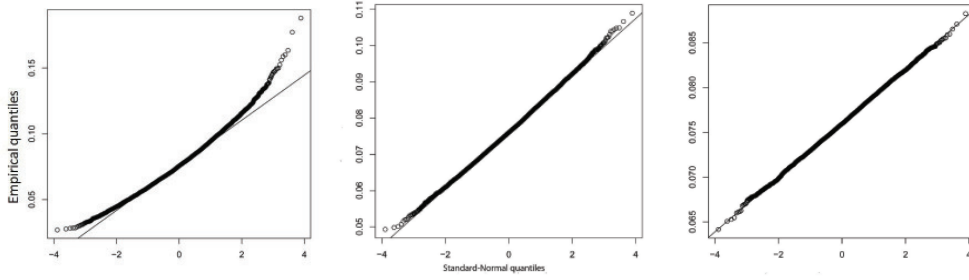


Figure 2: Quantile-Quantile plots for unrestricted estimator \hat{q}^0 of q^0 for sample sizes of $n = 100$ observed over $T = 10$ years (left) and $n = 1,000$ observed over $T = 5$ years (middle) and quantile-quantile plot for restricted estimator $\hat{\tilde{q}}^0$ for sample size $n = 1,000$ observed over $T = 5$ years (right)

Theorems 2 and 3, consistency and normality of the estimator need finite-sample assessment. When estimating in the unrestricted parameter space the components of $\sqrt{n}(\hat{\theta} - \theta_0)$ are compared with the Gaussian quantiles in a quantile-quantile plot. Figure 2 displays a small sample and a medium sample size scenario for $\hat{\theta}_1 = \hat{q}^0$. The sample size of $n = 1,000$ observed over $T = 5$ years will resemble our application later on. Especially the fit of the normal approximation for $n = 1,000$ is already good. The consistency of the estimator follows from the decrease in deviation visible from the y-axis label. The increase in efficiency for information that the γ_{hj}^0 are zero is also documented. The left panel of Figure 2 the quantile-quantile plot for the restricted estimation $\hat{\tilde{q}}^0$ for $n = 1,000$ observations is displayed. Again the normal approximation is good, and compared the unrestricted estimation for 1,000 observations (middle panel of Figure 2) the decrease in deviation is visible. The comparison is extended to Table 1 where the mean of the estimates is derived from the simulation. The restricted results for $\hat{\tilde{q}}^0$ are compared to \hat{q}^0 and $\hat{q}^0 + \hat{\gamma}_{21}^0$, results for $\hat{q}^0 + \hat{\gamma}_{23}^0$, $\hat{q}^0 + \hat{\gamma}_{32}^0$, $\hat{q}^0 + \hat{\gamma}_{34}^0$, $\hat{q}^0 + \hat{\gamma}_{43}^0$ and $\hat{q}^0 + \hat{\gamma}_{45}^0$ are very similar to those for $\hat{q}^0 + \hat{\gamma}_{21}^0$. The decrease in standard error is perceptible, because, of course more transitions contribute to the estimation of $\hat{\tilde{q}}^0$ as compared to \hat{q}^0 or $\hat{q}^0 + \hat{\gamma}_{hj}^0$.

Table 1: Simulation average of estimates for generator $\hat{\mathbf{Q}}^0$ in restricted and unrestricted parameter space (with simulated standard errors)

$\hat{\mathbf{Q}}^0$	Period	Sample size		
		$n = 100$	$n = 1,000$	$n = 10,000$
restricted parameter space				
\hat{q}^0	$T = 5$	$0.0759^{\pm 0.0096}$	$0.0760^{\pm 0.0030}$	$0.0760^{\pm 0.0010}$
	$T = 10$	$0.0760^{\pm 0.0069}$	$0.0760^{\pm 0.0022}$	$0.0760^{\pm 0.0007}$
unrestricted parameter space				
\hat{q}^0	$T = 5$	$0.0769^{\pm 0.0248}$	$0.0762^{\pm 0.0078}$	$0.0760^{\pm 0.0024}$
	$T = 10$	$0.0767^{\pm 0.0178}$	$0.0760^{\pm 0.0056}$	$0.0760^{\pm 0.0017}$
$\hat{q}^0 + \hat{\gamma}_{21}^0$	$T = 5$	$0.0772^{\pm 0.0255}$	$0.0761^{\pm 0.0078}$	$0.0760^{\pm 0.0025}$
	$T = 10$	$0.0769^{\pm 0.0180}$	$0.0762^{\pm 0.0055}$	$0.0760^{\pm 0.0018}$

The generator maybe also be used to assess the size of the likelihood ratio test of Corollary 4. However, to assess also power and other aspect like what happens when the parameter in the alternative becomes close to the hypothesis we consider twelve generators. For instance the generator with five states

$$\mathbf{Q}^1 := \begin{pmatrix} & 0.019 & 0 & 0 & 0 \\ 0.01 & & 0.072 & 0 & 0 \\ 0 & 0.015 & & 0.11 & 0 \\ 0 & 0 & 0.106 & & 0.2 \\ 0 & 0 & 0 & 0 & \end{pmatrix}$$

follows Definition 1 and has $\gamma_{hj}^1 \neq 0$, hence allows to assess the power of the test. The first six generators describe processes with $k = 5$ states, the second six generators belong to a three-states model. The intensities of generators \mathbf{Q}^{0*} and \mathbf{Q}^6 up to \mathbf{Q}^{10} derive from the generators \mathbf{Q}^0 to \mathbf{Q}^5 each by coarsen-

ing, states 1 and 2 as well as 3 and 4 are combined. Generators $\mathbf{Q}^0, \mathbf{Q}^5, \mathbf{Q}^{0/2}$ and \mathbf{Q}^{10} determine a one-parameter model, in the generators \mathbf{Q}^1 to \mathbf{Q}^4 and \mathbf{Q}^6 to \mathbf{Q}^9 at least two entries are different. The intensities in the generators $\mathbf{Q}^3, \mathbf{Q}^4, \mathbf{Q}^5, \mathbf{Q}^8, \mathbf{Q}^9$ and \mathbf{Q}^{10} result from dividing the entries of the generators $\mathbf{Q}^1, \mathbf{Q}^2, \mathbf{Q}^0, \mathbf{Q}^6, \mathbf{Q}^7$ and $\mathbf{Q}^{0/2}$ by two. The entries of the generators $\mathbf{Q}^2, \mathbf{Q}^4, \mathbf{Q}^7$ and \mathbf{Q}^9 are chosen such that all transitions intensities $q_{hj}, (h, j) \in \mathcal{I}_1$, but one coincide. These processes are ‘closer’ to the one-parameter model than the others. For completeness it suffices to report the generators \mathbf{Q}^3 and \mathbf{Q}^0 . There is $q_{hj}^3 = 0.076, (h, j) \in \mathcal{I}_1 \setminus \{(3, 2)\}$ and $q_{32}^3 = 0.038$ whereas $q_{hj}^0 = 0.076$ for all $(h, j) \in \mathcal{I}_1$. The nominal level for all tests is $\alpha = 5\%$. For

Table 2: Size of likelihood ratio test for one-parameter generators, sample sizes $n = 100$, $n = 1.000$ and $n = 10.000$ and observation periods $T = 5$ and $T = 10$ years

Generator	Period	Sample size		
		$n = 100$	$n = 1,000$	$n = 10,000$
$k = 5$				
\mathbf{Q}^0	$T = 5$	0.054	0.054	0.050
	$T = 10$	0.053	0.053	0.050
\mathbf{Q}^5	$T = 5$	0.069	0.049	0.050
	$T = 10$	0.058	0.049	0.047
$k = 3$				
$\mathbf{Q}^{0/2}$	$T = 5$	0.053	0.052	0.053
	$T = 10$	0.053	0.054	0.049
\mathbf{Q}^{10}	$T = 5$	0.066	0.048	0.046
	$T = 10$	0.055	0.048	0.048

virtually all situations depicted in Table 2 the actual size is close to the nominal level of 5%. The exception for generator \mathbf{Q}^5 where 100 entities observed

over 5 years result in a size of 6.9% can be explained by the fact that in this scenario only 4.6 transitions per sample occur on average over the 10,000 simulations. Table 3 shows that the test is consistent, with increasing sam-

Table 3: Power of LR test for different alternatives, sample sizes $n = 100$, $n = 1,000$ and $n = 10,000$ and observation periods $T = 5$ and $T = 10$ years

Generator	Period	Sample size		
		$n = 100$	$n = 1,000$	$n = 10,000$
$k = 5$				
\mathbf{Q}^1	$T = 5$	1.000	1.000	1.000
	$T = 10$	1.000	1.000	1.000
\mathbf{Q}^2	$T = 5$	0.191	0.990	1.000
	$T = 10$	0.361	1.000	1.000
\mathbf{Q}^3	$T = 5$	0.979	1.000	1.000
	$T = 10$	1.000	1.000	1.000
\mathbf{Q}^4	$T = 5$	0.134	0.808	1.000
	$T = 10$	0.190	0.992	1.000
$k = 3$				
\mathbf{Q}^6	$T = 5$	0.995	1.000	1.000
	$T = 10$	1.000	1.000	1.000
\mathbf{Q}^7	$T = 5$	0.231	0.988	1.000
	$T = 10$	0.395	1.000	1.000
\mathbf{Q}^8	$T = 5$	0.903	1.000	1.000
	$T = 10$	0.996	1.000	1.000
\mathbf{Q}^9	$T = 5$	0.153	0.836	1.000
	$T = 10$	0.227	0.983	1.000

ple size, longer observation period and increasing intensities urge the power towards one. Furthermore, generators \mathbf{Q}^2 , \mathbf{Q}^4 , \mathbf{Q}^7 and \mathbf{Q}^9 as compared \mathbf{Q}^1 ,

Q^3 , Q^6 and Q^8 show that power decreases when the alternative approaches the hypothesis. The comparison for the state numbers $k = 3$ and $k = 5$ suggests that number of states has no impact on the test. Combining states 1 and 2 as well as 3 and 4 increases the number of entities at risk in the new two states, which should, *ceteris paribus*, increase the power because the number of transitions should increase. However, the coarser states are associated with intensities which are only half as large. This compensates the increase in power completely. On average the results from the simulation are satisfactory with respect to power.

Simulations for testing time-stationarity in the model of Section 3.2 are presented in Weißbach and Walter (2010).

5. Analyzing Credit Ratings

In banks, transition probabilities, especially for the transition to default, should be estimated with internal default data (see Basel Committee on Banking Supervision, 2004, paragraph 461ff). WestLB AG granted access to an internal system of credit ratings with 20 non-default rating classes and one default class, so that $k = 21$. The rating histories of 3,699 counterparts have been observed over ten years from 1997 until 2006. Internal bank ratings are particularly necessary for counterparts without an external rating (see Kiefer, 2010). An early analysis revealed that the internal credit ratings are inhomogeneous, mainly attributable to the first-year behavior (see Weißbach and Walter, 2010). Hence, shifting the time origin to the beginning of the second year yields a homogeneous process. Transitions in the data set that cross several rating classes are classified as missing information, any rating history with such a defect is deleted from the data. In the end, 1,220 rating histories remain, migrating 957 times in total to an adjacent rating class. The average observation time per obligo is 1.6 years. By (20) the standard error is given as 0,016. The internal rating process starts at the credit origination, so

Table 4: LR test for the one-parameter model (11) from Corollary 4.

k	$2k - 4$	$-2 \log LR$	$\chi_{2k-4,0.95}^2$	p -value
21	38	56.26	53.38	0.028

that the intra-temporal dependence for rating changes, counted in quarters, found for instance in Koopman and Lucas (2008), is avoided. The random entry to the portfolio samples histories, independently of the credit cycle. Assumptions (A1) and (A2) are fulfilled.

The LR test of Corollary 4 is performed in Table 4. The p -value of 0.028 reveals that the rating process should not be considered as one-parametric. However, given the large data set and the well-known finding that any econometric model must be rejected if there is a sufficiently large data set, we find the goodness of the one-parameter model remarkable. There is one additional aspect. The LR-test statistic $-2 \log LR$ is only asymptotical. In a Monte Carlo simulation, we assess its finite sample properties, subject to the conditions of the data. We study the Type I error, using the generator estimated with \tilde{q} . We simulate 1,220 rating histories in each of 10,000 simulation loops. Each data set has the same average observation time of 1.6 years and is distributed over the rating classes at credit origination with the empirical distribution of the portfolio. The bootstrapped distribution of $-2 \log LR$, as both a histogram and as a kernel estimate, is given in Figure 3 and compared to the χ_{38}^2 -density.

The comparison with the χ^2 -distribution reveals a good fit in general. However, the simulated 95%-quantile of 55.6 is slightly larger than the χ_{2k-4}^2 quantile of 53.4 and is now even closer to the test statistic of 56.26. Further simulations of the power for various generators did not reveal much new. As expected for a one-parameter model, the power of 1 is reached quickly, once

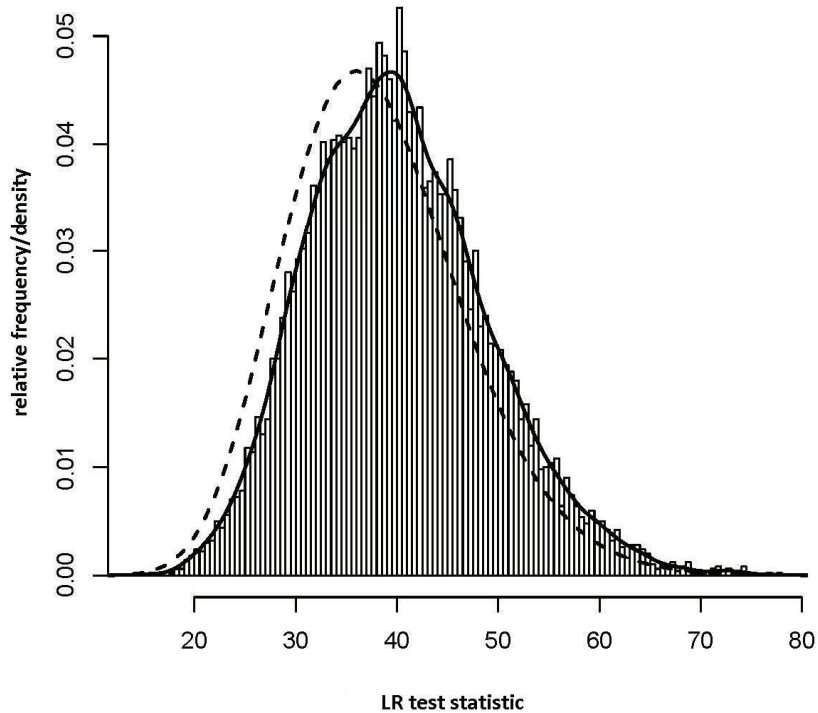


Figure 3: The asymptotic (dashed line) and the bootstrapped distribution of LR test statistic (histogram with solid line as kernel smoothing)

the sample size n exceeds 100. Changing the length of the study T and the number of rating classes k , does not result in a new assessment.

6. Discussion

For the description of rating transitions by means of a Markov process, we assume that rating state have similar widths. With the economic understanding that ratings should change state by one (and only one) grade at a time, we propose a one-parametric structure for the process generator. This implies a metric on the set of rating classes. A likelihood ratio test for the one-parametric distribution relies on the theory of counting processes. The

analysis of credit ratings reveals infrequent rating activity to be an obstacle to the continuous-time model. Rating transitions not to the neighboring rating class, for instance defaults, yield a large share of missing data. Even though we rule them out, it must be admitted that they are not strictly uninformative; rating transitions to remote classes imply that the transition to the adjacent class have already occurred. However, the magnitude of the bias must still be quantified.

Of course, many generalizations to multi-parameter models are conceivable. The case of a piecewise time-stationary model is one example studied here shortly. Additionally, it is possible that downgrades and upgrades do not behave equivalently so that their intensities should not be forced to be equal. However, considerable care must be exercised, because any new parameters will decrease the precision of established parameters. Additionally, it is not clear whether a better model would need to be Markovian, given the empirical evidence in favor of non-Markovian behavior. Dependence and covariates are additional sources for new parameters. Even if one wants to restrict the analysis to metric rating-state models, the proposed one-parameter model is not unique, and other metric models are easily constructed. The search for the next parameter is hence not an obvious process.

Acknowledgement: We would like to thank Walter Krämer for initiating the project, Ronja Walter and two anonymous referees for their valuable suggestions. Furthermore, we like to thank Claudia Gregor-Lawrenz and Patrick Tschiersch for advise in the application. The financial support from the Deutsche Forschungsgemeinschaft is gratefully acknowledged (SFB 823), The views expressed here are those of the authors and do not necessarily reflect the opinion of NORD/LB. All computations have been processed in GNU Octave, MATLAB and R.

References

- Aitchison, J., Silvey, S., 1958. Maximum-likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics* 29, 813–828.
- Albert, A., 1962. Estimating the infinitesimal generator of a continuous time, finite state Markov process. *Annals of Mathematical Statistics* 38, 727–753.
- Altman, E., Kao, D., 1992. Rating drift of high yield bonds. *Journal of Fixed Income* 1, 15–20.
- Andersen, P., Borgan, O., Gill, R., Keiding, N., 1997. *Statistical models based on counting processes*. Springer, New York.
- Bangia, A., Diebold, F., Kronimus, A., Schlagen, C., Schuermann, T., 2002. Rating migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking and Finance* 26, 235–264.
- Basel Committee on Banking Supervision, 2004. *International convergence of capital measurement and capital standards*. Tech. rep., Bank for International Settlements.
- Billingsley, P., 1961. *Statistical Inference for Markov Processes*. The University of Chicago Press, Chicago.
- Bluhm, C., Overbeck, L., Wagner, C., 2002. *An Introduction to Credit Risk Modeling*. Chapman & Hall.
- Christensen, J., Hansen, E., Lando, D., 2004. Confidence sets for continuous-time rating transition probabilities. *Journal of Banking and Finance* 28, 2575–2602.
- Forest, L., Belki, B., Suchower, S., Q3 1998. A one-parameter representation of credit risk and transition matrices. Tech. rep., CreditMetrics.

- Frydman, H., Schuerman, T., 2008. Credit rating dynamics and Markov mixture models. *Journal of Banking and Finance* 32, 1062–1075.
- Gill, R., 1983. Discussion of the papers by Helland and Kurtz. *Bulletin of the International Statistical Institute* 50, 239–243.
- Kiefer, N., 2010. Default estimation and expert information. *Journal of Business and Economic Statistics* 28, 320–328.
- Kiefer, N., Larson, C., 2007. A simulation estimator for testing the time homogeneity of credit rating transitions. *Journal of Empirical Finance* 14, 818–835.
- Koopman, S., Lucas, A., 2008. A non-gaussian panel time series model for estimating and decomposing default risk. *Journal of Business and Economic Statistics* 26, 510–525.
- Lando, D., Skødeberg, T., 2002. Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking and Finance* 26, 423–444.
- Lenglart, E., 1977. Relation de domination entre deux processus. *Annales de l'Institut Henri Poincaré* 13, 171–179.
- Merton, R., 1974. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* 29, 449–470.
- Munk, A., Weißbach, R., 1999. $1 - \alpha$ equivariant confidence rules for convex alternatives are $\alpha/2$ -level tests. *Journal of the American Statistical Association* 94, 1311–1320.
- Rebolledo, R., 1980. Central limit theorem for local martingales. *Probability Theory and Related Fields* 51, 269–286.

Weißbach, R., Tschiersch, P., Lawrenz, C., 2009. Testing time-homogeneity of rating transitions after origination of debt. *Empirical Economics* 36, 575–596.

Weißbach, R., Walter, R., 2010. A likelihood ratio test for stationarity of rating transitions. *Journal of Econometrics* 155, 188–194.

Appendix A. Proof of Lemmas

Appendix A.1. Lemma 1

For the first partial derivatives of $\lambda_{hj}(t; \boldsymbol{\theta})$, we have

$$\begin{aligned} \frac{\partial \lambda_{hj}(t; \boldsymbol{\theta})}{\partial q} &= Y_h(t), \quad (h, j) \in \mathcal{I}_1, \quad \frac{\partial \lambda_{hj}(t; \boldsymbol{\theta})}{\partial \gamma_{hj}} = Y_h(t), \quad (h, j) \in \mathcal{I}_2, \\ \frac{\partial \lambda_{hj}(t; \boldsymbol{\theta})}{\partial \gamma_{rs}} &= 0, \quad (h, j) \in \mathcal{I}_1, \quad (r, s) \in \mathcal{I}_2, \quad (h, j) \neq (r, s). \end{aligned}$$

All second and third derivatives of $\lambda_{hj}(t; \boldsymbol{\theta})$ with respect to q and γ_{rs} , $(r, s) \in \mathcal{I}_2$, exist and are zero. The partial derivatives of

$$\log \lambda_{hj}(t; \boldsymbol{\theta}) = \begin{cases} \log q + \log Y_1(t), & \text{if } (h, j) = (1, 2) \\ \log(q + \gamma_{hj}) + \log Y_h(t), & \text{if } (h, j) \in \mathcal{I}_2 \end{cases}$$

are

$$\begin{aligned} \frac{\partial \log \lambda_{hj}(t; \boldsymbol{\theta})}{\partial q} &= \begin{cases} \frac{1}{q}, & \text{if } (h, j) = (1, 2) \\ \frac{1}{q + \gamma_{hj}}, & \text{if } (h, j) \in \mathcal{I}_2 \end{cases}, \\ \frac{\partial \log \lambda_{hj}(t; \boldsymbol{\theta})}{\partial \gamma_{hj}} &= \frac{1}{q + \gamma_{hj}}, \quad (h, j) \in \mathcal{I}_2, \\ \frac{\partial \log \lambda_{hj}(t; \boldsymbol{\theta})}{\partial \gamma_{rs}} &= 0, \quad (h, j) \in \mathcal{I}_1, \quad (r, s) \in \mathcal{I}_2, \quad (h, j) \neq (r, s), \\ \frac{\partial^2 \log \lambda_{hj}(t; \boldsymbol{\theta})}{(\partial q)^2} &= \begin{cases} -\frac{1}{q^2}, & \text{if } (h, j) = (1, 2) \\ -\frac{1}{(q + \gamma_{hj})^2}, & \text{if } (h, j) \in \mathcal{I}_2 \end{cases}, \\ \frac{\partial^2 \log \lambda_{hj}(t; \boldsymbol{\theta})}{(\partial \gamma_{hj})^2} &= \frac{\partial^2 \log \lambda_{hj}(t; \boldsymbol{\theta})}{\partial q \partial \gamma_{hj}} = -\frac{1}{(q + \gamma_{hj})^2}, \quad (h, j) \in \mathcal{I}_2, \end{aligned} \tag{A.1}$$

$$\begin{aligned}
\frac{\partial^3 \log \lambda_{hj}(t; \boldsymbol{\theta})}{(\partial q)^3} &= \begin{cases} \frac{2}{q^3}, & \text{if } (h, j) = (1, 2) \\ \frac{2}{(q + \gamma_{hj})^3}, & \text{if } (h, j) \in \mathcal{I}_2 \end{cases}, \\
\frac{\partial^3 \log \lambda_{hj}(t; \boldsymbol{\theta})}{(\partial \gamma_{hj})^3} &= \frac{\partial^3 \log \lambda_{hj}(t; \boldsymbol{\theta})}{(\partial q)^2 \partial \gamma_{hj}} = \frac{\partial^3 \log \lambda_{hj}(t; \boldsymbol{\theta})}{\partial q (\partial \gamma_{hj})^2} \\
&= \frac{2}{(q + \gamma_{hj})^3}, \quad (h, j) \in \mathcal{I}_2.
\end{aligned} \tag{A.2}$$

All other second and third derivatives of $\log \lambda_{hj}(t; \boldsymbol{\theta})$, with respect to q and γ_{rs} , exist and are zero. Finally, for the partial derivatives of $\log L(\boldsymbol{\theta})$

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\theta})}{\partial q} &= \frac{N_{12}(T)}{q} + \sum_{(h,j) \in \mathcal{I}_2} \frac{N_{hj}(T)}{q + \gamma_{hj}} - \sum_{(h,j) \in \mathcal{I}_1} \int_0^T Y_h(s) ds, \\
\frac{\partial \log L(\boldsymbol{\theta})}{\partial \gamma_{hj}} &= \frac{N_{hj}(T)}{q + \gamma_{hj}} - \int_0^T Y_h(s) ds,
\end{aligned} \tag{A.3}$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{(\partial q)^2} = - \left(\frac{N_{12}(T)}{q^2} + \sum_{(h,j) \in \mathcal{I}_2} \frac{N_{hj}(T)}{(q + \gamma_{hj})^2} \right), \tag{A.4}$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{(\partial \gamma_{hj})^2} = \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial q \partial \gamma_{hj}} = - \frac{N_{hj}(T)}{(q + \gamma_{hj})^2},$$

$$\frac{\partial^3 \log L(\boldsymbol{\theta})}{(\partial q)^3} = \frac{2N_{12}(T)}{q^3} + \sum_{(h,j) \in \mathcal{I}_2} \frac{2N_{hj}(T)}{(q + \gamma_{hj})^3} \quad \text{and}$$

$$\frac{\partial^3 \log L(\boldsymbol{\theta})}{(\partial \gamma_{hj})^3} = \frac{\partial^3 \log L(\boldsymbol{\theta})}{\partial q (\partial \gamma_{hj})^2} = \frac{\partial^3 \log L(\boldsymbol{\theta})}{(\partial q)^2 \partial \gamma_{hj}} = \frac{2N_{hj}(T)}{(q + \gamma_{hj})^3}.$$

All other second and third partial derivatives of $\log L(\boldsymbol{\theta})$ with respect to q and γ_{rs} , exist and are zero. ■

Appendix A.2. Lemma 2

Together with (A.1) it holds for all $(h, j), (r, s) \in \mathcal{I}_2$

$$\begin{aligned}
&\sum_{(\mu, \nu) \in \mathcal{I}_1} \left(\frac{\partial}{\partial \gamma_{hj}} \log \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) \right) \left(\frac{\partial}{\partial \gamma_{rs}} \log \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) \right) \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) \\
&= \begin{cases} 0, & \text{if } (h, j) \neq (r, s) \\ \frac{Y_h(t)}{q_0 + \gamma_{hj_0}}, & \text{if } (h, j) = (r, s) \end{cases},
\end{aligned}$$

proving (2). In order to prove (3)-(5), we apply the theorem of Gill (1983). The four conditions that are required for the theorem to hold require that (A1) holds and that $Y_n(t) \leq n$. The remaining details for the proof of Lemma 2 are omitted here for the reason of space. \blacksquare

Appendix A.3. Lemma 3

Inserting (A.1) for all $\varepsilon > 0$, (6) results in

$$\begin{aligned}
& \frac{1}{n} \int_0^T \frac{1}{q_0^2} \cdot \mathbf{1} \left(\frac{1}{\sqrt{n} \cdot q_0} > \varepsilon \right) \cdot q_0 \cdot Y_1(t) dt \\
& + \frac{1}{n} \int_0^T \sum_{(\mu, \nu) \in \mathcal{I}_2} \frac{1}{(q_0 + \gamma_{\mu\nu 0})^2} \cdot \mathbf{1} \left(\frac{1}{\sqrt{n} \cdot (q_0 + \gamma_{\mu\nu 0})} > \varepsilon \right) \\
& \cdot (q_0 + \gamma_{\mu\nu 0}) \cdot Y_\mu(t) dt \\
= & \int_0^T \frac{Y_1(t)}{n \cdot q_0} \cdot \mathbf{1} \left(\frac{1}{\sqrt{n} \cdot q_0} > \varepsilon \right) dt \\
& + \sum_{(\mu, \nu) \in \mathcal{I}_2} \int_0^T \frac{Y_\mu(t)}{n \cdot (q_0 + \gamma_{\mu\nu 0})} \cdot \mathbf{1} \left(\frac{1}{\sqrt{n} \cdot (q_0 + \gamma_{\mu\nu 0})} > \varepsilon \right) dt =: (*)
\end{aligned}$$

and for all $(h, j) \in \mathcal{I}_2$, (7) becomes

$$\begin{aligned}
& \frac{1}{n} \int_0^T \frac{1}{(q_0 + \gamma_{hj0})^2} \cdot \mathbf{1} \left(\frac{1}{\sqrt{n} \cdot (q_0 + \gamma_{hj0})} > \varepsilon \right) \cdot (q_0 + \gamma_{hj0}) \cdot Y_h(t) dt \\
= & \int_0^T \frac{Y_h(t)}{n \cdot (q_0 + \gamma_{hj0})} \cdot \mathbf{1} \left(\frac{1}{\sqrt{n} \cdot (q_0 + \gamma_{hj0})} > \varepsilon \right) dt =: (**).
\end{aligned}$$

Because of $Y_h \leq n$

$$\begin{aligned}
0 \leq (*) & \leq \frac{T \cdot \mathbf{1} \left(\frac{1}{\sqrt{n} \cdot q_0} > \varepsilon \right)}{q_0} + \sum_{(\mu, \nu) \in \mathcal{I}_2} \frac{T \cdot \mathbf{1} \left(\frac{1}{\sqrt{n} \cdot (q_0 + \gamma_{\mu\nu 0})} > \varepsilon \right)}{q_0 + \gamma_{\mu\nu 0}} \\
& \xrightarrow{n \rightarrow \infty} 0
\end{aligned}$$

and $0 \leq (**.) \leq T / (q_0 + \gamma_{hj0}) \cdot \mathbf{1} (1 / (\sqrt{n} \cdot (q_0 + \gamma_{hj0}) > \varepsilon) \xrightarrow{n \rightarrow \infty} 0$. Especially,

$$\begin{aligned}
& \int_0^T \frac{Y_1(t)}{n \cdot q_0} \cdot \mathbf{1} \left(\frac{1}{\sqrt{n} \cdot q_0} > \varepsilon \right) dt \\
& + \sum_{(\mu, \nu) \in \mathcal{I}_2} \int_0^T \frac{Y_\mu(t)}{n \cdot (q_0 + \gamma_{\mu\nu 0})} \cdot \mathbf{1} \left(\frac{1}{\sqrt{n} \cdot (q_0 + \gamma_{\mu\nu 0})} > \varepsilon \right) dt \xrightarrow[n \rightarrow \infty]{P} 0
\end{aligned}$$

and

$$\int_0^T \frac{Y_h(t)}{n \cdot (q_0 + \gamma_{hj0})} \cdot \mathbb{1} \left(\frac{1}{\sqrt{n} \cdot (q_0 + \gamma_{hj0})} > \varepsilon \right) dt \xrightarrow[n \rightarrow \infty]{P} 0.$$

■

Appendix A.4. Lemma 4

With $n = 2k - 4$, define positive, real-valued constants $a_1 := \sigma_{21}, \dots, a_n = \sigma_{k-1,k}$ and $b := \sigma_q - \sum_{i=1}^n a_i$. By induction, we can show that for all dimensions $l \in \mathbb{N}$ with $l \leq n$ that $|\Sigma_l| = b \cdot \prod_{j=1}^l a_j + \sum_{i=l+1}^n \left(\prod_{j=1}^l a_j \right) a_i > 0$, and that Σ is hence positive definite because the subdeterminants are positive. The induction step uses the fact that $|\Sigma_l|$ is expanded in the last row. The two resulting summands contain a determinant of a somewhat peculiar matrix that can easily be expanded by its last column. ■

Appendix A.5. Lemma 5

The proof is similar to that of Lemma 4.7 in Weißbach and Walter (2010).

Appendix A.6. Lemma 6

Proof. In order to prove that the score statistic, divided by n , approaches zero, let us consider the gradient of the log-likelihood (1) for information only up to time in $0 \leq t \leq T$. By using the explicit representations of $\partial \lambda_{\mu\nu}(t; \boldsymbol{\theta}) / \partial \theta_j$ and $\partial \log \lambda_{\mu\nu}(t; \boldsymbol{\theta}) / \partial \theta_j$, given in the proof of Lemma 1, and $\partial \log \lambda_{\mu\nu}(s; \boldsymbol{\theta}) / \partial \theta_j = \frac{1}{\lambda_{\mu\nu}(s; \boldsymbol{\theta})} \cdot \frac{\partial}{\partial \theta_j} \lambda_{\mu\nu}(s; \boldsymbol{\theta})$ one can exchange the order of differentiation and integration, so that

$$U_t^j(\boldsymbol{\theta}_0) = \int_0^t \sum_{(\mu, \nu) \in \mathcal{I}_1} \frac{\partial}{\partial \theta_j} \log \lambda_{\mu\nu}(s; \boldsymbol{\theta}_0) (dN_{\mu\nu}(s) - \lambda_{\mu\nu}(s; \boldsymbol{\theta}_0) ds).$$

Due to a generalization of the Doob-Meyer decomposition (see Andersen et al., 1997, pg. 67), $N_{hj}(t)$ can be formulated as the sum of a compensator $\Lambda_{hj}(t) := \int_0^t \lambda_{hj}(s) ds$ and a local martingale $M_{hj}(t)$ and we can write

$U_t^j(\boldsymbol{\theta}_0)/n$ for any $0 \leq t \leq T$ as

$$\begin{cases} \int_0^t \frac{1}{n \cdot q_0} dM_{12}(s) + \sum_{(\mu, \nu) \in \mathcal{I}_2} \int_0^t \frac{1}{n \cdot (q_0 + \gamma_{\mu\nu 0})} dM_{\mu\nu}(s), & \text{if } \theta_j = q, \\ \int_0^t \frac{1}{n \cdot (q_0 + \gamma_{uv 0})} dM_{uv}(s), & \text{if } \theta_j = \gamma_{uv}. \end{cases} \quad (\text{A.5})$$

The integrands are, for a fixed n , deterministic and constant in t , thus constituting particularly predictable processes. As a result of multivariate counting process theory, the martingale transform $U_t^j(\boldsymbol{\theta}_0)/n$ is, for each $0 \leq t \leq T$, a local martingale with the predictable variation process $\langle U^j(\boldsymbol{\theta}_0)/n \rangle(t)$ that is

$$\begin{cases} \int_0^t \frac{1}{(n \cdot q_0)^2} \lambda_{12}(s; \boldsymbol{\theta}_0) ds + \sum_{(\mu, \nu) \in \mathcal{I}_2} \int_0^t \frac{1}{(n \cdot (q_0 + \gamma_{\mu\nu 0}))^2} \lambda_{\mu\nu}(s; \boldsymbol{\theta}_0) ds, & \text{if } \theta_j = q \\ \int_0^t \frac{1}{(n \cdot (q_0 + \gamma_{uv 0}))^2} \lambda_{uv}(s; \boldsymbol{\theta}_0) ds, & \text{if } \theta_j = \gamma_{uv}. \end{cases}$$

This is $(\int_0^t \sum_{(\mu, \nu) \in \mathcal{I}_1} \left(\frac{\partial}{\partial \theta_j} \log \lambda_{\mu\nu}(s; \boldsymbol{\theta}_0) \right)^2 \lambda_{\mu\nu}(s; \boldsymbol{\theta}_0) ds) / n^2$, which, multiplied by n , is known from Lemma 2 to converge. Hence, the predictable variation process itself converges to zero for large n . Applied to the inequality of Lengart (1977), namely that for all $\eta > 0$ and all $\varepsilon > 0$, it holds that

$$\mathbb{P} \left(\sup_{t \in [0, T]} \left| \frac{1}{n} U_t^j(\boldsymbol{\theta}_0) \right| > \eta \right) \leq \frac{\varepsilon}{\eta^2} + \mathbb{P} \left(\left\langle \frac{1}{n} U^j(\boldsymbol{\theta}_0) \right\rangle (T) > \varepsilon \right),$$

the last summand is asymptotically negligible. The result follows, because ε/η^2 may be arbitrarily small. \blacksquare

Appendix A.7. Lemma 7

Multiplying (A.5) by \sqrt{n} yields, for all $j = 1, \dots, 2k-3$, that $U_t^j(\boldsymbol{\theta}_0)/\sqrt{n}$ is $\int_0^t dM_{12}(s)/(\sqrt{n} \cdot q_0) + \sum_{(\mu, \nu) \in \mathcal{I}_2} \int_0^t dM_{\mu\nu}(s)/(\sqrt{n} \cdot (q_0 + \gamma_{\mu\nu 0}))$ if $\theta_j = q$ and equals

$\int_0^t dM_{uv}(s)/(\sqrt{n} \cdot (q_0 + \gamma_{uv 0}))$ if $\theta_j = \gamma_{uv}$. As above, $U_t^j(\boldsymbol{\theta}_0)/\sqrt{n}$ is a local

martingale, and has, because of (A.1), the predictable covariation process

$$\begin{aligned}
& \left\langle \frac{1}{\sqrt{n}}U^j(\boldsymbol{\theta}_0), \frac{1}{\sqrt{n}}U^l(\boldsymbol{\theta}_0) \right\rangle(t) \\
&= \frac{1}{n} \left\langle \sum_{(\mu,\nu) \in \mathcal{I}_1} \int_0^\cdot \frac{\partial}{\partial \theta_j} \log \lambda_{\mu\nu}(u; \boldsymbol{\theta}_0) dM_{\mu\nu}(u), \right. \\
&\quad \left. \sum_{(\eta,\rho) \in \mathcal{I}_1} \int_0^\cdot \frac{\partial}{\partial \theta_l} \log \lambda_{\eta\rho}(u; \boldsymbol{\theta}_0) dM_{\eta\rho}(u) \right\rangle(t) \\
&= \frac{1}{n} \sum_{(\mu,\nu) \in \mathcal{I}_1} \int_0^t \left(\frac{\partial}{\partial \theta_j} \log \lambda_{\mu\nu}(u; \boldsymbol{\theta}_0) \right) \cdot \left(\frac{\partial}{\partial \theta_l} \log \lambda_{\mu\nu}(u; \boldsymbol{\theta}_0) \right) \\
&\quad \cdot \lambda_{\mu\nu}(u; \boldsymbol{\theta}_0) du.
\end{aligned}$$

The rule $\langle \int_0^\cdot \mathbf{H} d\mathbf{M} \rangle(t) = \int_0^t \mathbf{H}(s) \text{diag}(\boldsymbol{\lambda}(s)) \mathbf{H}'(s) ds$ yields the last equality.

For $t = 0$, this is 0 and for $t = T$, one has, due to Lemma 2

$$\begin{cases} = 0, & \text{if } \theta_j = \gamma_{mr}, \theta_l = \gamma_{uv}, (m, r) \neq (u, v) \\ \xrightarrow[n \rightarrow \infty]{P} \sigma_{uv}, & \text{if } \theta_j = \theta_l = \gamma_{uv} \text{ or } \theta_j = q, \theta_l = \gamma_{uv} \\ \xrightarrow[n \rightarrow \infty]{P} \sigma_q, & \text{if } \theta_j = \theta_l = q \end{cases} .$$

Hence

$$\left\langle \frac{1}{\sqrt{n}}\mathbf{U}(\boldsymbol{\theta}_0) \right\rangle(0) \xrightarrow[n \rightarrow \infty]{P} 0 \quad \text{and} \quad \left\langle \frac{1}{\sqrt{n}}\mathbf{U}(\boldsymbol{\theta}_0) \right\rangle(T) \xrightarrow[n \rightarrow \infty]{P} \boldsymbol{\Sigma},$$

in a stochastic sense and component-wise. Together with Lemma 3, (A.1) and Rebolledo's theorem with $t \in \{0, T\}$ Lemma 7 follows. \blacksquare

Appendix A.8. Lemma 8

As in the proof of Lemma 6, integration and second differentiation can be interchanged, so that $\partial^2 \log L(\boldsymbol{\theta}_0) / (\partial \theta_j \partial \theta_l)$, divided by $-n$, can be formulated as

$$\begin{aligned}
& \frac{1}{n} \int_0^T \sum_{(\mu,\nu) \in \mathcal{I}_1} \frac{\partial^2}{\partial \theta_j \partial \theta_l} \lambda_{\mu\nu}(s; \boldsymbol{\theta}_0) ds - \frac{1}{n} \int_0^T \sum_{(\mu,\nu) \in \mathcal{I}_1} \frac{\partial^2}{\partial \theta_j \partial \theta_l} \log \lambda_{\mu\nu}(s; \boldsymbol{\theta}_0) dM_{\mu\nu}(s) \\
& - \frac{1}{n} \int_0^T \sum_{(\mu,\nu) \in \mathcal{I}_1} \left(\frac{\partial^2}{\partial \theta_j \partial \theta_l} \log \lambda_{\mu\nu}(s; \boldsymbol{\theta}_0) \right) \cdot \lambda_{\mu\nu}(s; \boldsymbol{\theta}_0) ds.
\end{aligned}$$

Omitting the second summand results in

$$\begin{cases} 0, & \text{if } \theta_j = \gamma_{mr}, \theta_l = \gamma_{uv}, (m, r) \neq (u, v), \\ \int_0^T \frac{Y_u(s)}{n \cdot (q_0 + \gamma_{uv})} ds, & \text{if } \theta_j = \theta_l = \gamma_{uv} \text{ or } \theta_j = q, \theta_l = \gamma_{uv}, \\ \int_0^T \frac{Y_1(s)}{n \cdot q_0} + \sum_{(\mu, \nu) \in \mathcal{I}_2} \frac{Y_\mu(s)}{n \cdot (q_0 + \gamma_{\mu\nu})} ds, & \text{if } \theta_j = \theta_l = q. \end{cases}$$

Here, as in Lemma 2, the second case $\xrightarrow{P} \sigma_{uv}$ and the third case $\xrightarrow{P} \sigma_q$. For the second summand, we have

$$\begin{cases} 0, & \text{if } \theta_j = \gamma_{mr}, \theta_l = \gamma_{uv}, (m, r) \neq (u, v), \\ \int_0^T \frac{-1}{n \cdot (q_0 + \gamma_{uv})^2} dM_{uv}(s), & \text{if } \theta_j = \theta_l = \gamma_{uv} \text{ oder } \theta_j = q, \theta_l = \gamma_{uv}, \\ \int_0^T \frac{-1}{n \cdot q_0^2} dM_{12}(s) \\ + \sum_{(\mu, \nu) \in \mathcal{I}_2} \int_0^T \frac{-1}{n \cdot (q_0 + \gamma_{\mu\nu})^2} dM_{\mu\nu}(s), & \text{if } \theta_j = \theta_l = q. \end{cases}$$

The same arguments as in the proof of Lemma 6 complete the proof

$$\frac{1}{n} \int_0^T \sum_{(\mu, \nu) \in \mathcal{I}_1} \frac{\partial^2}{\partial \theta_j \partial \theta_l} \log \lambda_{\mu\nu}(s; \boldsymbol{\theta}_0) dM_{\mu\nu}(s) \xrightarrow[n \rightarrow \infty]{P} 0.$$

■

Appendix A.9. Lemma 9

As in the last-but-one proof, integration and third differentiation can be interchanged by using Lemma 1. For all $j, l, m = 1, \dots, 2k - 3$ holds

$$\begin{aligned} \frac{\partial^3 \log L(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_l \partial \theta_m} &= \int_0^T \sum_{(\mu, \nu) \in \mathcal{I}_1} \frac{\partial^3}{\partial \theta_j \partial \theta_l \partial \theta_m} \log \lambda_{\mu\nu}(t; \boldsymbol{\theta}) dN_{\mu\nu}(t) \\ &- \int_0^T \sum_{(\mu, \nu) \in \mathcal{I}_1} \frac{\partial^3}{\partial \theta_j \partial \theta_l \partial \theta_m} \lambda_{\mu\nu}(t; \boldsymbol{\theta}) dt. \end{aligned}$$

Together with (8) and $\partial^3 \lambda_{\mu\nu}(t; \boldsymbol{\theta}) / (\partial \theta_j \partial \theta_l \partial \theta_m) \equiv 0$ from the proof of Lemma 1, one has, for all $n \in \mathbb{N}$, all $j, l, m = 1, \dots, 2k - 3$, and $\boldsymbol{\theta} \in \Theta_0$ by means of Lemma 5

$$\left| \frac{1}{n} R_T^{jlm}(\boldsymbol{\theta}) \right| = \left| \frac{1}{n} \cdot \frac{\partial \log L(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_l \partial \theta_m} \right| \leq \int_0^T \sum_{(\mu, \nu) \in \mathcal{I}_1} \frac{1}{n} H_{\mu\nu} dN_{\mu\nu}(t).$$

Along the arguments of Lemma 6, $\int_0^T \sum_{(\mu,\nu) \in \mathcal{I}_1} H_{\mu\nu} / nd N_{\mu\nu}(t)$ is the optional variation process of the local martingale $\int_0^\cdot \sum_{(\mu,\nu) \in \mathcal{I}_1} \sqrt{H_{\mu\nu} / nd} dM_{\mu\nu}(t)$, evaluated in T . The martingale has a predictable variation process $\int_0^\cdot \sum_{(\mu,\nu) \in \mathcal{I}_1} H_{\mu\nu} \lambda_{\mu\nu}(t; \boldsymbol{\theta}_0) / ndt$. Evaluated in T , the predictable variation converges by (9) in Lemma 5 towards C and, see (10), the assumptions of Rebolledo's theorem are thus fulfilled. The optional variation converges towards C . ■