A Likelihood Ratio Test for Stationarity of Rating Transitions

Rafael Weißbach* & Ronja Walter

Institut für Wirtschafts- und Sozialstatistik

Fakultät Statistik, Technische Universität Dortmund, Dortmund, Germany

December 2, 2008

Abstract

For a time-continuous discrete-state Markov process as model for rating transitions, we study the time-stationarity by means of a likelihood ratio test. For multiple Markov process data from a multiplicative intensity model, maximum likelihood parameter estimates can be represented as martingale transform of the processes counting transitions between the rating states. As a consequence, the profile partial likelihood ratio is asymptotically χ^2-distributed. An internal rating data set reveals highly significant instationarity.

---

*address for correspondence: Rafael Weißbach, Institut für Wirtschafts- und Sozialstatistik, Fakultät Statistik, Technische Universität Dortmund, 44221 Dortmund, Germany, email: Rafael.Weissbach@tu-dortmund.de, Fon: +49/231/7555419, Fax: +49/231/7555284.

JEL classifications. C33, C34, C41

Keywords. Stationarity, Multiple Markov process, Counting process, Likelihood ratio, Panel data.

1 Introduction

The homogenous Markov process with stationary transitions intensities remains to be the staring point for rating migration modelling (Bluhm et al., 2002, pg. 197ff). The mounting evidence for non-Markovian property - e.g. due to significant dependence on regressors - is rich, see Lando and Skødeberg (2002); Altman and Kao (1992); Bangia et al. (2002); Frydman and Schuerman (March 2007). For estimation of non-markovian transition intensities see e.g. Meira-Machado et al. (2006). More recently, evidence for the inhomogeneity, i.e. the instationarity of the transition intensities, has appeared (Kiefer and Larson (2007); Weißbach et al. (2008)). For estimation of instationary transition intensities see e.g. Weißbach (2006). Here we study a likelihood ratio test for stationarity on basis of multiple Markov processes, i.e. for panel data of debtors. In case of only one transitory state an approximation of the alternative parameter space can be found, for instance, with Laguerre polynomials in Kiefer (1985). Here, with several transitory rating states, the unknown hazard rates in the alternative are approximated by step-functions. Piecewise constant hazards occurs in Bayesian duration time Lancaster (2004). The goodness-of-fit aspect of the constant hazard rate requires a profile likelihood, being of current interest (Murphy and van der Vaart (2000)).

Time-dependence of the intensities can be interpreted as continuous-time generalization of time-variability in Markov dependence of the Markov chain. In this sense, the paper is an extension of test for stationary dependence in discrete time Markov chains by Anderson and Goodman (1957).

The partial profile likelihood ratio is asymptotically χ^2-distributed due to the asymptotic normality of the maximum likelihood (ML) estimates for the piece-wise constant hazard rates. For globally constant hazard rates Albert (1962) established the maximum likelihood generator for the time-continuous finite-state Markov process. The normality of our estimate results from its representation as a martingale transform. The main building block is the martingale that arises by compensating the processes that count transitions between the rating states. Finally, a martingale limit theorem by Rebolledo (1980) applies. Certain extent of the proof is to study the predictable covariation process with Lenglart's inequality.

Our application is credit risk, and in detail stationarity of rating transition intensities in an internal rating system. Further application is conceivable, for instance, in labor market dynamics.

2 Model

We consider Markov processes $X = \{X_t, t \in [0,T]\}$ defined on a probability space $(\Omega, \mathfrak{F}, P)$ with the set of states $K = \{1, \ldots, k\}$ (e.g. rating classes) where state k is an absorbing state (e.g. bankruptcy). We denote X_t as the state of an individual at time t after certain origin. The process is determined by the transition matrices

$$P(s,t) \quad = \quad (p_{hj}(s,t))_{h,j\in K} \in \mathbb{R}^{k\times k}; \quad s,t \in [0,T], s \leq t.$$

where the transition probabilities $p_{hj}(s,t) = P(X_t = j \mid X_s = h) \, \forall h,j \in K$ give the conditional probability for a transition from state h to j within the time period s till t. Denote by $m_h(t)$ the probability of state h at time t. The infinitesimal generator of the process is defined by the transition intensities

$$q_{hj}(t) \quad = \quad \lim_{u\to 0+} \frac{p_{hj}(t,t+u)}{u}.$$

Stationarity denotes the situation where those intensities are constant over time. In this case, the transition matrices can be represented as a matrix exponential of $Q = (q_{hj})_{h,j\in K}$. It holds that $p_{kj}(s,t) = q_{kj}(t) = 0$ with $j \neq k$.

Our model, encompassing stationarity, are piecewise constant intensities.

Definition 2.1 *Let the intensities on* $[0, T]$ *with the change-points* $t_l,\ l = 1, \ldots, b-1$ *and* $t_0 = 0,\ t_b = T$ *be*

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

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

The fragmentation of the parameter space may be chosen differently for different rating class combinations. For the ease of clarity, here only equal spacing is considered.

The data are transition histories $\mathbf{X}_i = \{X_t^i, t \in [0, T]\}$ for each of the $i = 1, \ldots, n$ individuals in a sample. We observe a panel continuously in time Compared to the analysis of all transition histories $\mathbf{X}_1, \ldots, \mathbf{X}_n$, there is no loss of information when using the vector of initial ratings $X_0^1, \ldots, X_0^n$ together with the processes

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

counting the number of transitions from state h to j until time t in the entire sample. Additionally, the processes $Y_h(t)$ denote the number of individuals in state h at time t. For large samples, this is a clear reduction in the number of random processes. The data situation is depicted in Figure 1.

There are only two further assumptions:

(A1) We assume for fixed t

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

(A2) The counting processes must follow a multiplicative intensity model, i.e. have intensity process

$$\lambda_{hj}(t) = Y_h(t) q_{hj}(t),\ h, j \in K, j \neq h.$$

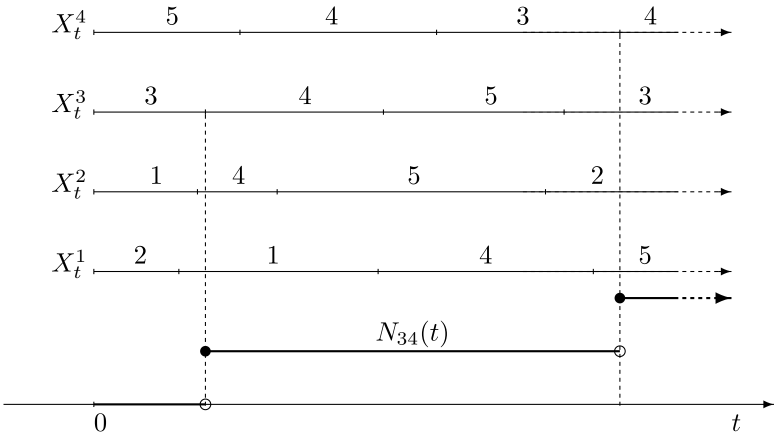

Figure 1: Four Markov processes and counting process $N_{34}(t)$

As usual in the analysis of durations, only a partial likelihood can be evaluated (see Andersen et al., 1993, equation 2.7.4")

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

where $\sum_{j\neq h} = \sum_{h=1}^{k-1} \sum_{\substack{j=1 \\ j\neq h}}^{k}$.

In order to test on stationarity of the intensities, the null hypothesis can be formulated as

$$H_0 : \delta_{hj2} = \ldots = \delta_{hjb} = 0 \; \forall j \neq h, h, j \in K, \qquad (2)$$

with the alternative

$$H_1 : \exists \; \delta_{hjl} \neq 0. \qquad (3)$$

3 Results

Our aim is to construct a likelihood ratio test on stationarity in a multiplicative intensity model. Statistics of the likelihood ratio are usually asymptotically χ^2 distributed under certain regularity conditions. In our case there are two obstacles. First there is certainly right censoring at time T, so only a partial likelihood can be used, additionally, transition histories may be lost to follow-up. Also, the q_{hj} are nuisance parameters, requiring a profile likelihood.

Denote the partial likelihood ratio by

$$\Delta \quad = \quad \frac{L((\hat{q}_{hj})_{h,j\in K, j\neq h})}{L((\tilde{\hat{q}}_{hj}, \hat{\delta}_{hjl})_{h,j\in K, j\neq h, l=2,\ldots,b})}, \tag{4}$$

where $\hat{q}_{hj}$ are the ML-estimates in the case of stationarity and $\tilde{\hat{q}}_{hj}$ resp. $\hat{\delta}_{hjl}$ are the ML-estimates in case of a piecewise stationary process with (b-1) change-points.

In the following theorems we are able to show, that the asymptotic distribution of the test statistic still remains χ^2.

Theorem 1 *For a sample of Markov processes with intensity as in Definition 2.1, let assumptions (A1) and (A2) be fulfilled. Then the partial ML-estimators of the parameters are asymptotically normal distributed*

$$\sqrt{n}\begin{pmatrix} \tilde{\hat{q}}_{hj} - q_{hj0} \\ \hat{\delta}_{hjl} - \delta_{hjl0} \end{pmatrix}_{j\neq h, h,j\in K, l=2,\ldots,b} \xrightarrow{d} N\left(0, \Sigma^{-1}\right),$$

where q_{hj0} and δ_{hjl0} denote the true parameters.

The representation and estimation of Σ is described later. Clearly, the asymptotic normality of the estimate vector maybe used to construct confidence ellipsoids for the parameter vector, resulting in confidence sets for the rating transition probabilities comparable to Christensen et al. (2004). For instance, confidence sets for the δ_i can be used for inclusion rules in order to answer not only the equality hypothesis (3) but also the equivalence hypothesis (see Munk and Weißbach (1999)). Additionally, Wald and score tests can be derived with the asymptotic normality. However, as the Wald test is not scale-invariant and the score test lacks power, we construct a likelihood ratio test.

Corollary 2 *Under the assumptions of Theorem 1 it holds*

$$-2\log(\Delta) \overset{n\to\infty}{\sim} \chi^2_{(b-1)(k-1)^2}.$$

As expected, the degrees of freedom depend on the number of change-points $(b-1)$, and additionally on the number of states k in the model.

After we know that the test statistic of the likelihood ratio is χ^2 distributed, we aim at its explicit form. With explicit expressions of the ML-estimates the test statistic is computable.

Theorem 3 *The ML-estimate in* (4) *under the null hypothesis* (2) *has the following representation*

$$\hat{q}_{hj} = \frac{N_{hj}(T)}{\int_0^T Y_h(t)dt}.$$

Under the alternative (3) *one obtains*

$$\tilde{\hat{q}}_{hj} = \frac{N_{hj}(t_1^-)}{\int_0^{t_1} Y_h(t)dt}.$$

With the definition $\hat{q}_{hjl} = \frac{N_{hj}(t_l^-)-N_{hj}(t_{l-1}^-)}{\int_{t_{l-1}}^{t_l} Y_h(t)dt}, \quad l = 2, \ldots, b$ *it holds*

$$\hat{\delta}_{hjl} = \hat{q}_{hjl} - \tilde{\hat{q}}_{hj}, \qquad l = 2, \ldots, b.$$

As a consequence, $-2\log(\Delta)$ *has the form*

$$-2\sum_{j\neq h}\left[N_{hj}(t_1^-)\log\left(\frac{\hat{q}_{hj}}{\tilde{\hat{q}}_{hj}}\right) + \sum_{l=2}^{b}(N_{hj}(t_l^-) - N_{hj}(t_{l-1}^-))\log\left(\frac{\hat{q}_{hj}}{\hat{q}_{hjl}}\right)\right]. \quad (5)$$

As one can see, $\tilde{\tilde{q}}_{hj}$ only depends on the number of transitions from h to j and the number of individuals in state h until time t_1. The similar behavior can be observed with the $\hat{q}_{hjl}$. They only depend on the transitions and number of individuals in state h between time t_{l-1} and t_l. The estimates are only derived by the transition counts and duration times one obtains if defining time t_{l-1} as starting point 0 and t_l as the end of a study.

4 Proofs

The score statistic, evaluated at the true parameters, is a martingale transform. The vector of parameter estimates is asymptotically normal, see Theorem 1, almost immediately implying the test statistic $-2\log\Delta$ to follow a χ^2-distribution, see Theorem 2. Explicit formulae for parameter estimates and the likelihood ratio of Theorem 3 facilitate applications.

4.1 Proof of Theorem 1

The normality of the estimates results from the necessary condition for the ML property. The partial derivatives of the log-likelihood are equal to zero, hence, the leading term in a Taylor-expansion, the score statistic, equals (minus) the residual terms. The linear expansion of the classical case, is replaced by a quadratic. But at first we need some prerequisites,

Note that for all $h \in K$

$$\frac{1}{n}\int_{t_i}^{t_j} Y_h(t)dt \leq \frac{n(t_j - t_i)}{n} = t_j - t_i, \qquad i, j = 0, \ldots, b, i < j. \tag{6}$$

Lemma 4.1 *The matrix A with*

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & A_n \end{pmatrix}$$

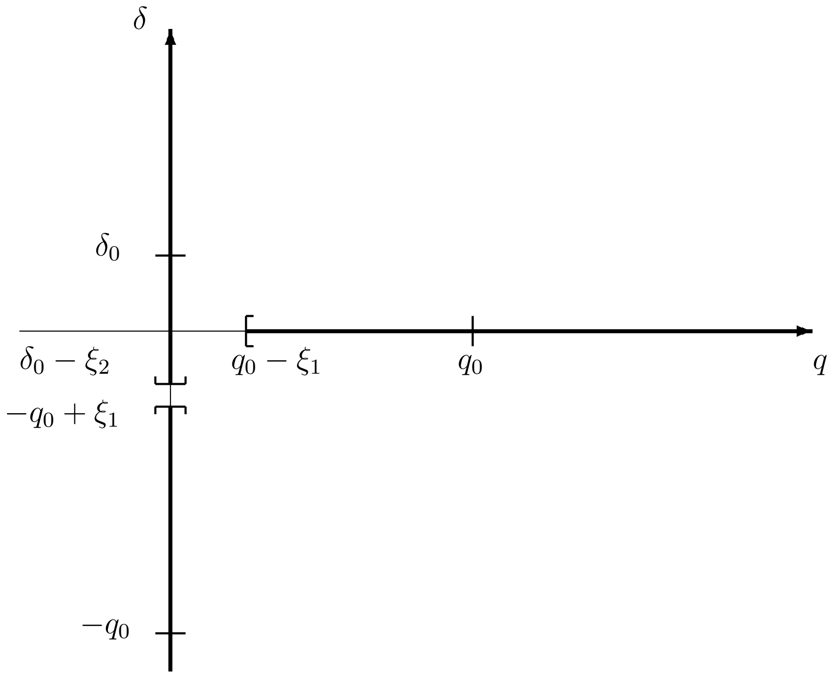

Figure 2: Environment of q_0 and δ_0

where

$$A_i = \begin{pmatrix} a_i + c_i & a_i \\ a_i & a_i \end{pmatrix}, \qquad a_i, c_i > 0,$$

is positive definite.

Proof: It is necessary that all eigenvalues e of A are positive. With some matrix algebra one can show $det(A - eI) = \prod_{i=1}^{n} det(A_i - eI)$. Therefore, it suffices to prove that the A_i have positive eigenvalues. Then $e_{ij} = (2a_i + c_i)/2 \pm \sqrt{(2a_i + c_i)^2/4 - a_i c_i} > 0, \quad j = 1, 2$ with $a_i, c_i > 0$. $\square$

Lemma 4.2 *For $q \in (0, \infty)$ and $\delta \in (-q, \infty)$, there exist, for all true parameters q_0 and δ_0, $\xi_1, \xi_2 > 0$ so that the neighborhood $\Theta_0^q = [q_0 - \xi_1, \infty) \subset (0, \infty)$ and $\Theta_0^\delta = [\delta_0 - \xi_2, \infty) \subset (-q_0 + \xi_1, \infty)$.*

Proof: This is based on the openness of the parameter space see Figure 2. $\square$

For simplification we now restrict, for the meantime, to the case of only one change-point, namely

$$\lambda_{hj}(t) \quad = \quad \mathbb{1}_{[0,t_1)}(t) q_{hj} Y_h(t) + \mathbb{1}_{[t_1,T]}(t)(q_{hj} + \delta_{hj}) Y_h(t), \;\; h, j \in K, j \neq h. \tag{7}$$

Lemma 4.3 *The first to third partial derivatives of the intensity process* (7) *and the log-intensity process with respect to the parameters* q_{hj} *and* δ_{hj} *exist and are continuous. Additionally, the first to third partial derivatives of the log-likelihood* (1) *exist.*

Proof: The first partial derivatives of the intensity process have the form

$$\frac{\partial \lambda_{hj}(t)}{\partial q_{hj}} = Y_h(t) \quad \text{and}$$
$$\frac{\partial \lambda_{hj}(t)}{\partial \delta_{hj}} = \mathbb{1}_{[t_1,T]}(t) Y_h(t).$$

The first to third derivatives with respect to any other δ_{il} or $q_{il}, \;\; i, l = 1, .., k$ exist and equal zero. The first to third derivative of the log-intensity process also exists, because $q_{hj} > 0$ and $q_{hj} + \delta_{hj} > 0$ (see Definition 2.1). The third derivatives result to

$$\frac{\partial^3 \log(\lambda_{hj}(t))}{\partial q_{hj}^3} \quad = \quad \frac{2 \; \mathbb{1}_{[0,t_1)}(t)}{q_{hj}^3} + \frac{2 \; \mathbb{1}_{[t_1,T]}(t)}{(q_{hj} + \delta_{hj})^3} \tag{8}$$

and

$$\frac{\partial^3 \log(\lambda_{hj}(t))}{\partial \delta_{hj}^3} \quad = \quad \frac{2 \; \mathbb{1}_{[t_1,T]}(t)}{(q_{hj} + \delta_{hj})^3}. \tag{9}$$

They are obviously continuous in q_{hj} and δ_{hj}. The mixed second and third derivatives with respect to δ_{hj} and q_{hj} obtain the same form as the second and third derivatives with respect to δ_{hj}. It is also easy to show that the first three derivatives of the log-likelihood exist and are continuous in q_{hj} and δ_{hj} because the log-likelihood (1) is an additive composition of the intensity processes. □

Now we derive the asymptotic distribution of the ML-estimators. The Taylor series expansions of the score statistics $U_T^i(\hat{\theta}) = \frac{\partial \log L}{\partial \theta_i}\Big|_{\theta=\hat{\theta}}$ around the true parameters q_{hj0} and δ_{hj0} are:

$$\begin{aligned} 0 = \frac{1}{\sqrt{n}} U_T^i(\hat{\theta}) =& \frac{1}{\sqrt{n}} U_T^i(\theta_0) - \sum_{l=1}^{2(k-1)^2} \sqrt{n}(\hat{\theta}_l - \theta_{l0}) \frac{1}{n} \mathfrak{I}_T^{il}(\theta_0) \\ &+ \sum_{l=1}^{2(k-1)^2} \sqrt{n}(\hat{\theta}_l - \theta_{l0}) \frac{1}{2n} \sum_{m=1}^{2(k-1)^2} (\hat{\theta}_m - \theta_{m0}) R_T^{ilm}(\theta^*) \end{aligned} \tag{10}$$

where

$$\theta \quad = \quad \begin{pmatrix} q_{hj} \\ \delta_{hj} \end{pmatrix}_{j \neq h, h, j \in K} \in \mathbb{R}^{2(k-1)^2} \tag{11}$$

denotes the parameter vector, and $\hat{\theta}$ its ML-estimates. Here $\mathfrak{I}_T(\theta)$ denotes minus the Hesse matrix, and $R_T^{ilm}(\theta)$ the third partial derivatives of the log-likelihood, while θ^* is on the line segment between $\hat{\theta}$ and the true parameter θ_0. If we want to apply Billingsley (1961, Theorem 10.1), $\frac{1}{n}\mathfrak{I}_T^{il}(\theta_0)$, in the linear term, must converge to a covariance matrix. The quadratic term must be asymptotically negligible.

The constant term $\frac{1}{\sqrt{n}} U_T^i(\theta_0)$ is a local square integrable martingale, as a function of T, and normality can be studied with the martingale central limit theorem (Rebolledo, 1980; Andersen et al., 1993, Theorem II.5.1). To this end, two properties must be shown. First, its covariation processes must converge in probability to a covariance matrix. The covariation processes mainly depend on the partial derivatives of the intensity processes.

Lemma 4.4 *Let δ_{hj0} and q_{hj0} be the true parameters. For $\theta_{il} \in \{\{q_{il}\} \cup \{\delta_{il}\}, i, l \in K, i \neq k\}$ and $\theta_{xy} \in \{\{q_{xy}\} \cup \{\delta_{xy}\}, x, y \in K, x \neq y\}$, without the case where $i, x = h$ and $l, y = j$, it holds*

$$\begin{aligned} &\tfrac{1}{n} \int_0^T \tfrac{\partial \log(\lambda_{hj}(t))}{\partial \theta_{il}} |_{q_{hj0}, \delta_{hj0}} \tfrac{\partial \log(\lambda_{hj}(t))}{\partial \theta_{xy}} |_{q_{hj0}, \delta_{hj0}} \lambda_{hj}(t, q_{hj0}, \delta_{hj0}) dt \\ &= 0. \end{aligned} \tag{12}$$

The only covariances that do not vanish are

$$\frac{1}{n}\int_{t_1}^{T}\frac{\mathbb{1}_{[t_1,T]}(t)Y_h(t)}{(q_{hj0}+\delta_{hj0})}dt \quad \xrightarrow{P} \quad \int_{t_1}^{T}\frac{m_h(t)}{q_{hj0}+\delta_{hj0}}dt =: a_{hj} > 0 \tag{13}$$

and

$$\frac{1}{n}\int_{0}^{t_1}\frac{\mathbb{1}_{[0,t_1)}Y_h(t)}{q_{hj0}}dt \quad \xrightarrow{P} \quad \int_{0}^{t_1}\frac{m_h(t)}{q_{hj0}}dt =: c_{hj} > 0. \tag{14}$$

Hence, the covariance matrix Σ has on the diagonal matrices described by

$$\Sigma_{hj} = \begin{pmatrix} a_{hj}+c_{hj} & a_{hj} \\ a_{hj} & a_{hj} \end{pmatrix}, \qquad a_{hj}, c_{hj} > 0,$$

with $h \in K, j \in K,\ j \neq h$. All other entries equal zero, and the Σ is positive definite.

Proof: Equation (12) is clear. The convergence in (13) and (14) follow with (A1) and Helland (1983). Therefore, the covariation processes converge to a finite function. It also applies, with Lemma 4.1, that Σ is positive definite.$\square$

Second, we need to prove the Lindeberg condition.

Lemma 4.5 *For any $\varepsilon > 0$ and $j \neq h \in K$ it holds*

$$\frac{1}{n}\int_{0}^{t_1}\frac{Y_h(t)}{q_{hj0}}dt\mathbb{1}_{(\varepsilon,\infty)}\left(\left|\frac{1}{\sqrt{n}q_{hj0}}\right|\right) \xrightarrow{P} 0$$

and

$$\frac{1}{n}\int_{t_1}^{T}\frac{Y_h(t)}{(q_{hj0}+\delta_{hj0})}dt\mathbb{1}_{(\varepsilon,\infty)}\left(\left|\frac{1}{\sqrt{n}(q_{hj0}+\delta_{hj0})}\right|\right) \xrightarrow{P} 0,$$

as n converges to ∞.

Proof: This follows with (6) and

$$\lim_{n\to\infty}\mathbb{1}_{(\varepsilon,\infty)}\left(\left|\frac{1}{\sqrt{n}q_{hj0}}\right|\right) = \lim_{n\to\infty}\mathbb{1}_{(\varepsilon,\infty)}\left(\left|\frac{1}{\sqrt{n}(q_{hj0}+\delta_{hj0})}\right|\right) = 0.$$

□

Lemma 4.4 and 4.5 now imply that $\frac{1}{\sqrt{n}} U_T^i(\theta_0)$ is normal distributed with mean 0 and covariance matrix Σ.

After the constant term we now come to the linear term of the Taylor expansion (10).

Lemma 4.6 $\frac{1}{n}\mathfrak{I}_T^{il}(\theta_0)$ *converges to* Σ, *as* $n \to \infty$.

Proof: One is able to write the entries of $\frac{1}{n}\mathfrak{I}_T(\theta_0)$ as a sum of the term of the left side of (12) and

$$-\frac{1}{n}\int_0^T \sum_{j\neq h} \frac{\partial^2}{\partial\theta_i\theta_l} \log \lambda_{hj}(s,\theta_0)dM_{hj}(s), \tag{15}$$

where $M_{hj}(t) = N_{hj}(t) - \int_0^t \lambda_{hj}(s)ds$. The first term converges to the entries of Σ because of Lemma 4.4. The second term, depending on the true parameters, represents a local square integrable martingale and converges in probability to zero. We can show this with its variation process

$$\begin{aligned} &\frac{1}{n}\int_0^{t_1} \sum_{j\neq h} \frac{q_{hj0}Y_h(t)}{q_{hj0}^4}dt + \frac{1}{n}\int_{t_1}^T \sum_{j\neq h} \frac{(q_{hj0}+\delta_{hj0})Y_h(t)}{(q_{hj0}+\delta_{hj0})^4}dt \\ \leq\ & \sum_{j\neq h}\frac{t_1}{q_{hj0}^3} + \sum_{j\neq h}\frac{T-t_1}{(q_{hj0}+\delta_{hj0})^3} < \infty, \end{aligned}$$

converging to a finite quantity and Lenglart's inequality (see Lenglart, 1977). □

With the following, we can prove that $\frac{1}{n}R_T^{ilm}(\theta^*)$ is bounded in probability by a constant M, hence, the quadratic term in the Taylor expansion vanishes as n converges to ∞.

The third partial derivatives of the log likelihood with respect to q_{hj}

(divided by n) have the form

$$\frac{1}{n}\int_0^{t_1}\frac{2}{q_{hj}^3}dN_{hj}(t)+\frac{1}{n}\int_{t_1}^{T}\frac{2}{(q_{hj}+\delta_{hj})^3}dN_{hj}(t). \tag{16}$$

The third partial derivatives with respect to δ_{hj} or mixed partial derivatives of both are represented by only the second term.

Lemma 4.7 *There exist neighborhoods Θ^q_{hj0} and Θ^δ_{hj0} around the true parameters and a predictable process $H_{hjn}(t)$ not depending on q_{hj} and δ_{hj} with*

$$\begin{aligned}\sup_{q_{hj}\in\Theta^q_{hj0}}\left|\frac{\partial^3\log(\lambda_{hj}(t))}{\partial q_{hj}^3}\right| &\leq H_{hjn}(t),\\ \sup_{\delta_{hj}\in\Theta^\delta_{hj0}}\left|\frac{\partial^3\log(\lambda_{hj}(t))}{\partial \delta_{hj}^3}\right| &\leq H_{hjn}(t).\end{aligned} \tag{17}$$

And it holds that

$$\frac{1}{n}\int_0^T\sum_{j\neq h}H_{hjn}(t)\lambda_{hj}(t,q_{hj0},\delta_{hj0})dt<\infty. \tag{18}$$

Proof: It exists with Lemma 4.2 for all q_{hj0} and δ_{hj0} a $(\xi^q_{hj},\xi^\delta_{hj})>0$ with $\Theta^q_{hj0}=[q_{hj0}-\xi^q_{hj},\infty)\subset(0,\infty)$ and $\Theta^\delta_{hj0}=[\delta_{hj0}-\xi^\delta_{hj},\infty)\subset(-q_{hj0}+\xi^q_{hj},\infty)$ $\forall j\neq h, h,j\in K$. Define

$$H_{hjn}(t)=\frac{2\ \mathbb{1}_{[0,t_1)}(t)}{(q_{hj0}-\xi^q_{hj})^3}+\frac{2\ \mathbb{1}_{[t_1,T]}(t)}{(q_{hj0}-\xi^q_{hj}+\delta_{hj0}-\xi^\delta_{hj})^3}.$$

For all $q_{hj}\in\Theta^q_{hj0}$ and $\delta_{hj}\in\Theta^\delta_{hj0}$, with (8) and (9) one obtains (17). As all mixed derivatives equal the third derivative with respect to δ_{hj} or zero, their supremum is less or equal $H_{hjn}(t)$ as well. Now it holds with (6)

$$\begin{aligned}&\frac{1}{n}\int_0^T\sum_{j\neq h}H_{hjn}(t)\lambda_{hj}(t,q_{hj0},\delta_{hj0})dt \\ \leq\ &\sum_{j\neq h}\left(\frac{2t_1q_{hj0}}{(q_{hj0}-\xi^q_{hj})^3}+\frac{2(T-t_1)(q_{hj0}+\delta_{hj0})}{(q_{hj0}-\xi^q_{hj}+\delta_{hj0}-\xi^\delta_{hj})^3}\right)<\infty.\end{aligned} \tag{19}$$

$\square$

Lemma 4.8 *With Lemma 4.7,* (16) *also converges to a deterministic* $M < \infty$.

Proof: First, (16) is less or equal to the integral over H_{hjn} with respect to $dN_{hj}(t)$. This integral is the optional variation process and (19) the predictable variation process of the same martingale. The asymptotic equality (and hence the boundedness of (16)) follows by the martingale central limit theorem, if we can show that

$$\sum_{j\neq h} \frac{2q_{hj0}}{(q_{hj0}-\xi^q_{hj})^3}\frac{1}{n}\int_0^{t_1} Y_h(t)dt\mathbb{1}_{(\varepsilon,\infty)}\left(\sqrt{\frac{2}{n(q_{hj0}-\xi^q_{hj})^3}}\right)$$
$$+ \sum_{j\neq h} \frac{2(q_{hj0}+\delta_{hj0})}{(q_{hj0}-\xi^q_{hj}+\delta_{hj0}-\xi^\delta_{hj})^3}\frac{1}{n}\int_{t_1}^{T} Y_h(t)dt$$
$$\mathbb{1}_{(\varepsilon,\infty)}\left(\sqrt{\frac{2}{n(q_{hj0}-\xi^q_{hj}+\delta_{hj0}-\xi^\delta_{hj})^3}}\right)$$

converges for $n\to\infty$ to 0. This holds because of the same argument as in the proof of Lemma 4.5. $\square$

Because $\frac{1}{n}U^i_T(\theta_0) \xrightarrow{P} 0$ and Lemmata 4.6 and 4.8 the ML-estimate $\hat\theta$ exists and is consistent.

With (10) and Lemma 4.8 it holds:

$$\sum_{l=1}^{2(k-1)^2} \sqrt{n}(\hat\theta_l-\theta_{l0})\frac{1}{n}\mathfrak{I}^{il}_T(\theta_0) - \frac{1}{\sqrt{n}}U^i_T(\theta_0)$$
$$\leq \frac{1}{2}M\sum_{m=1}^{2(k-1)^2}(\hat\theta_m-\theta_{m0})\sum_{l=1}^{2(k-1)^2}\sqrt{n}(\hat\theta_l-\theta_{l0}).$$

Now it follows with Lemma 4.6:

$$\left|\frac{1}{\sqrt{n}}U_T(\theta_0)-\Sigma\sqrt{n}(\hat\theta-\theta_0)\right| \leq \varepsilon_n|\sqrt{n}(\hat\theta-\theta_0)|$$

where

$$\varepsilon_n = \frac{2(k-1)^2}{2}M\sum_{m=1}^{2(k-1)^2}|\hat\theta_m-\theta_{m0}| \overset{n\to\infty}{\to} 0$$

because of the consistency of $\hat{\theta}$. Here $|.|$ denotes the absolute norm.

This has the form

$$|u_n - v_n| \leq \varepsilon_n |\Sigma^{-1} v_n|.$$

With a similar proof as to Billingsley (1961, Theorem 10.1), the normality of the score statistic implies now the normality of the ML-estimates.

As $\hat{\theta}$ converges to θ_0, Lemma 4.6 ensures that $\frac{1}{n}\mathfrak{I}_T(\hat{\theta})$ is a consistent estimate of Σ. The proof for $(b-1) > 1$ is analogous to that for only one change-point and is omitted here for the sake of brevity.

4.2 Proof of Corollary 2

For the proof of Theorem 1 the order of δ_{hj} and q_{hj} in parameter θ (see (11)) was necessary for Lemma 4.4. Here another order will be convenient. Let $\hat{\vartheta} = (\hat{\delta}, \tilde{\hat{q}})$ be the unrestricted ML-estimator, where the vector $\hat{\delta}$ includes all $\hat{\delta}_{hj}$ and $\tilde{\hat{q}}$ all $\tilde{\hat{q}}_{hj}$ (in case of $b-1=1$), and $\hat{\vartheta}_0 = (0, \hat{q})$, where $\hat{q}$ includes all $\hat{q}_{hj}$. We want to show that

$$-2 \log \frac{L(\hat{\vartheta}_0)}{L(\hat{\vartheta})} \overset{n\to\infty}{\sim} \chi^2_{(b-1)(k-1)^2}.$$

With Theorem 1 we have that

$$\begin{pmatrix} \hat{\delta} - \delta \\ \tilde{\hat{q}} - q \end{pmatrix} \xrightarrow{d} N\left(0, \Gamma^{-1} = \begin{pmatrix} \Gamma^{\delta} & \Gamma^{\delta,q} \\ \Gamma^{q,\delta} & \Gamma^{q} \end{pmatrix}\right) \tag{20}$$

where Γ is a rearrangement of Σ. Now under $H_0 : \vartheta = (0, q)$ with standard arguments of the profile likelihood ratio

$$-2 \log \frac{L(\hat{\vartheta}_0)}{L(\hat{\vartheta})} \doteq (\hat{\delta} - \delta)(\Gamma^{\delta})^{-1}(\hat{\delta} - \delta).$$

Together with equation (20) we find that $-2\log\Delta$ is χ^2 distributed. We obtain $(k-1)^2$ degrees of freedom for $(b-1) = 1$ change-point since $\dim(\delta) = (k-1)^2$ because of the defaulting class k. With $(b-1) > 1$ we achieve the same result with $(b-1)(k-1)^2$ degrees of freedom.

4.3 Proof of Theorem 3

In order to obtain the partial ML-estimators and the explicit test statistic, we need the first derivatives with respect to q_{hj} and δ_{hjl}. They result to

$$\begin{aligned}\frac{\partial \log(L)}{\partial q_{hj}} &= \frac{N_{hj}(t_1^-)}{q_{hj}} + \sum_{l=2}^{b} \frac{N_{hj}(t_l) - N_{hj}(t_{l-1}^-)}{q_{hj}+\delta_{hjl}} - \int_0^T Y_h(t)dt, \\ \frac{\partial \log(L)}{\partial \delta_{hjl}} &= \frac{N_{hj}(t_l) - N_{hj}(t_{l-1}^-)}{q_{hj}+\delta_{hjl}} - \int_{t_1}^T Y_h(t)dt.\end{aligned}$$

In the case of stationary intensities where $\delta_{hjl} = 0 \ \forall j \neq h \ h, j \in K, l = 2, \ldots, b$ you obtain, by equating with zero and solving the resulting equation, the partial ML-estimators of Albert (1962)

$$\hat{q}_{hj} = \frac{N_{hj}(T)}{\int_0^T Y_h(t)dt}.$$

With piecewise constant intensities the partial ML-estimators are

$$\begin{aligned}\tilde{\hat{q}}_{hj} &= \frac{N_{hj}(t_1^-)}{\int_0^{t_1} Y_h(t)dt} \\ \hat{q}_{hjl} &= \frac{N_{hj}(t_l^-) - N_{hj}(t_{l-1}^-)}{\int_{t_{l-1}}^{t_l} Y_h(t)dt} \qquad l = 2, \ldots, b \\ \hat{\delta}_{hjl} &= \hat{q}_{hjl} - \tilde{\hat{q}}_{hj} \qquad l = 2, \ldots, b.\end{aligned}$$

Now we obtain the partial likelihood ratio

$$\begin{aligned}\Delta &= \frac{L((\hat{q}_{hj})_{h,j\in K, j\neq h})}{L((\tilde{\hat{q}}_{hj}, \hat{\delta}_{hjl})_{h,j\in K, j\neq h, l=2,\ldots,b})} \\ &= \prod_{t\in[0,t_1)} \prod_{j\neq h} \left(\frac{\hat{q}_{hj}}{\tilde{\hat{q}}_{hj}}\right)^{\Delta N_{hj}(t)} \prod_{l=2}^{b} \prod_{t\in[t_{l-1},t_l]} \prod_{j\neq h} \left(\frac{\hat{q}_{hj}}{\tilde{\hat{q}}_{hj} + \hat{\delta}_{hjl}}\right)^{\Delta N_{hj}(t)}\end{aligned}$$

and the test statistic $-2\log(\Delta)$ equals

$$-2\sum_{j\neq h} \left[N_{hj}(t_1^-) \log\left(\frac{\hat{q}_{hj}}{\tilde{\hat{q}}_{hj}}\right) + \sum_{l=2}^{b} (N_{hj}(t_l^-) - N_{hj}(t_{l-1}^-)) \log\left(\frac{\hat{q}_{hj}}{\tilde{\hat{q}}_{hj} + \hat{\delta}_{hjl}}\right)\right].$$

5 Application

Their capital ratio is important for banks. It is dependent on the rating transitions of the portfolio counterparts in two ways. Economically, it is sensitive to changes in portfolio risk Kleff and Weber (2008). Legally, the capital is a function of the transition probabilities, especially for the transition to default, and may 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 8 non-default rating classes and one default class. Rating histories of $3,699$ counterparts were observed over seven years from 1.1.1997 until 31.12.2003. Internal rating starts at credit origination, dampening the expected impact of calendar time - via the business cycle - (see Bangia et al., 2002). The transition histories may assumed to be independent or at least to fulfill assumptions (A1) and (A2).

The nonparametric Johansen-Aalen estimates of the transition matrix $\hat{P}(s,t)$ for different off-sets present an indication for the instationary behavior of rating transitions, e.g. $\hat{P}(0,t)$ and $\hat{P}(1,t)$ being both theoretically equal for a stationary process. Figure 3 shows the dissimilarity for the rating combinations $\hat{p}_{34}(0,t)$ and $\hat{p}_{34}(1,t)$.

Simultaneous inference for all rating combinations corrects for spurious effect. The simultaneous test for stationarity of rating transitions, based on the test statistic $-2\log(\Delta)$, however, is only asymptotical due to Corrollary 2. A Monte Carlo simulation can serve to assess its finite sample properties under the conditions of the data at hand. We studied the type I error using the generator estimated with $\hat{q}_{hj}$ of Theorem 3 (as in Casjens et al., 2007). At a nominal significance level of 5% the actual size for a sample size of 7000 rating histories was found to be 0.75%. This means, the test is considerably conservative, causing interpretation problems, when the test does not reject.

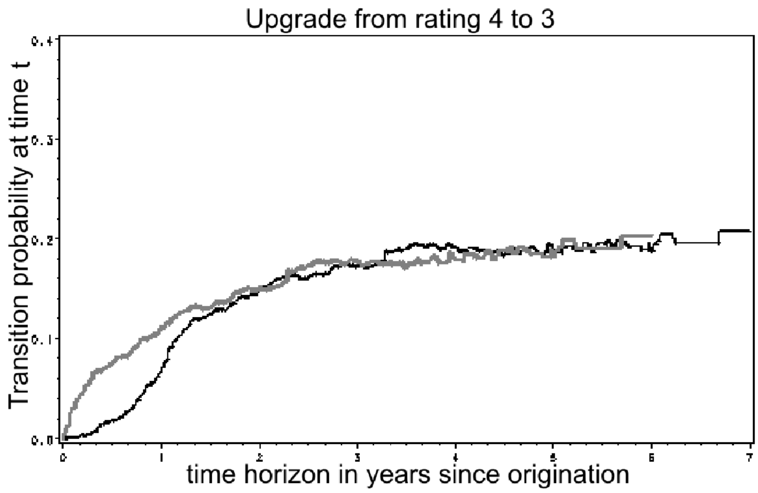

Figure 3: Nonparametric estimates: t-years transition probability at origin (black line) and after one year (grey line)

Table 1: Likelihood ratio test for stationarity of internal rating transitions. The number of b ranges between 2 and 7

b	2	3	4	5	6	7
$-2\log(\Delta)$	93.9	125.9	289.3	345.8	447.3	626.2
p-value	0.009	0.535	< 0.001	< 0.001	< 0.001	< 0.001

In simulations for the type II error we found that, for doubling the hazard over the seven years, the power achieves virtually 100% for around n=1000 processes. For a linear - exponentiated Weibull - hazard function the results were similar.

Ultimately, we are interested in testing the null of stationarity (2), at the significance level $\alpha = 0.05$, against the alternative of transition intensities with structural breaks (3). We consider different equidistant partitions $0 = t_0 \leq t_1 \leq t_2 \leq \ldots \leq t_b = 7$ of the time interval $[0, 7]$. The maximum number of breaks is six, yielding seven one-year intervals.

The striking small p-values (see Table 1), except for $b = 3$, prove that rating transition intensities in this rating system are not stationary. Time since origination does influence rating transition probabilities significantly.

A argumentation of the result for $b = 3$ is local inconsistency of likelihood ratio tests. The construction of the test (5) implies that local instationarity within an interval of the alternative cannot be discovered by the test. A possible reason is the non-monotony of some of the intensities. In a simplified situation, Weißbach and Dette (2007) proposed a globally consistent test that will detect any alternative. From a practical point of view, this deficiency is accounted for here by processing our test on different partitions.

References

A. Albert. Estimating the infinitesimal generator of a continuous time, finite state Markov process. *Annals of Mathematical Statistics*, 38:727 –753, 1962.

E.I. Altman and D.L. Kao. Rating drift of high yield bonds. *Journal of Fixed Income*, 1992.

P.K. Andersen, Ø. Borgan, R.D. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer, New York, 1993.

T.W. Anderson and L.A. Goodman. Statistical inference about Markov chains. *Annals of Mathematical Statistics*, 28:89–110, 1957.

A. Bangia, F.X. Diebold, A. Kronimus, C. Schlagen, and T. Schuermann. Rating migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking and Finance*, 26:235–264, 2002.

Basel Committee on Banking Supervision. International convergence of capital measurement and capital standards. Technical report, Bank for International Settlements, 2004.

P. Billingsley. *Statistical Inference for Markov Processes.* The University of Chicago Press, Chicago, 1961.

C. Bluhm, L. Overbeck, and C. Wagner. *An Introduction to Credit Risk Modeling.* Chapman & Hall, 2002.

S. Casjens, F. Kramer, T. Mollenhauer, and R. Walter. Die optimale Kombination interner und externer Ratings. In W. von Schimmelmann and G. Franke, editors, *Interne und externe Ratings*, pages 123–162. Frankfurter Allgemeine Buch, Frankfurt, 2007.

J. Christensen, E. Hansen, and D. Lando. Confidence sets for continuous-time rating transition probabilities. *Journal of Banking and Finance*, 28: 2575–2602, 2004.

T.R. Fleming and D.P. Harrington. *Counting Processes and Survival Analysis.* Wiley, New York, 1991.

H. Frydman and T. Schuerman. Credit rating dynamics and markov mixture models. Technical report, Stern School of Business, New York University, March 2007.

I.S. Helland. Applications of central limit theorems for martingales with continuous time. *Bulletin of the International Statistical Institute*, 5:346–360, 1983.

N.M. Kiefer. Specification diagnostics based on Laguerre alternatives for econometric models of duration. *Journal of Econometrics*, 28:135–154, 1985.

N.M. Kiefer and C.E. Larson. A simulation estimator for testing the time homogeneity of credit rating transitions. *Journal of Empirical Finance*, 14:818–835, 2007.

V. Kleff and M. Weber. How do banks determine capital? Evidence from Germany. *German Economic Review*, 9:354–372, 2008.

T. Lancaster. *An Introduction to Modern Bayesian Econometrics*. Blackwell Publishing, 2004.

D. Lando and T. Skødeberg. Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking and Finance*, 26:423–444, 2002.

E. Lenglart. Relation de domination entre deux processus. *Annales de l´Institute Henri Poincaré*, 13:171–179, 1977.

L. Meira-Machado, J. Uña-Álvarez, and C. Cadarso-Suárez. Nonparametric estimation of transition probabilities in a non-markov illness-death model. *Lifetime Data Analysis*, 12:325–344, 2006.

A. Munk and R. Weißbach. $1-\alpha$ equivariant confidence rules for convex alternatives are $\alpha/2$-level tests. *Journal of the American Statistical Association*, 94:1311–1320, 1999.

S.A. Murphy and A.W. van der Vaart. On profile likelihood. *Journal of the American Statistical Association*, 95:449–465, 2000.

R. Rebolledo. Central limit theorem for local martingales. *Probability Theory and Related Fields*, 51:269–286, 1980.

R. Weißbach and H. Dette. Kolmogorov-Smirnov-type testing for the partial homogeneity of Markov processes. *Applied Stochastic Models in Business and Industry*, 23:223–234, 2007.

R. Weißbach, P. Tschiersch, and C. Lawrenz. Testing time-homogeneity of rating transitions after origination of debt. *Empirical Economics*, in press, 2008. doi: 10.1007/s00181-008-0212-3.

R. Weißbach. A general kernel functional estimator with general bandwidth. *Journal of Nonparametric Statistics*, 18:1–12, 2006.