A likelihood ratio test for stationarity of rating transitions

Weißbach, Rafael; Walter, Ronja

Postprint / Postprint
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:
Weißbach, R., & Walter, R. (2009). A likelihood ratio test for stationarity of rating transitions. *Journal of Econometrics*, *155*(2), 188-194. https://doi.org/10.1016/j.jeconom.2009.10.016

A likelihood ratio test for stationarity of rating transitions

Rafael Weißbach, Ronja Walter

PII: S0304-4076(09)00269-3
DOI: 10.1016/j.jeconom.2009.10.016
Reference: ECONOM 3269

To appear in: *Journal of Econometrics*

Received date: 2 December 2008
Revised date: 3 June 2009
Accepted date: 10 October 2009

Please cite this article as: Weißbach, R., Walter, R., A likelihood ratio test for stationarity of rating transitions. *Journal of Econometrics* (2009), doi:10.1016/j.jeconom.2009.10.016

A Likelihood Ratio Test for Stationarity of Rating Transitions

Rafael Weißbach* & Ronja Walter†

September 29, 2009

Abstract

We study the time-stationarity of rating transitions, modelled by a time-continuous discrete-state Markov process and derive a likelihood ratio test. For multiple Markov processes from a multiplicative intensity model, maximum likelihood parameter estimates can be written as martingale transform of the processes, counting transitions between the rating states, so that the profile partial likelihood ratio is asymptotically χ^2-distributed. An application to an internal rating data set reveals highly significant instationarity.

JEL classifications. C33, C34, C41

Keywords. Stationarity, Multiple Markov process, Counting process, Likelihood ratio, Multiple spells

*address for correspondence: Rafael Weißbach, Department of Economics, Faculty of Economic and Social Sciences, University of Rostock, Ulmenstraße 69, D-18057 Rostock, Germany, email: rafael.weissbach@uni-rostock.de, Fon: +49/381/4984429.

†Institute for Business and Social Statistics, Department of Statistics, Technische Universität Dortmund, D-44221 Dortmund, Germany, email: ronjawalter@statistik.tu-dortmund.de.

1 Introduction

The homogenous Markov process, with stationary transition intensities, remains the staring point for rating-transition modelling (Bluhm et al., 2002, pg. 197ff). Evidence of non-Markovian property - due, for example, to significant dependence on regressors - is mounting, see Lando and Skødeberg (2002), Altman and Kao (1992), Bangia et al. (2002) and Frydman and Schuerman (2007). For the estimation of non-Markovian transition intensities, see e.g. Meira-Machado et al. (2006). More recently, evidence of inhomogeneity, i.e. the instationarity of the transition intensities, has appeared (Kiefer and Larson, 2007; Weißbach et al., 2009). For the estimation of instationary transition intensities, see Weißbach (2006). In the present paper, we perform a likelihood ratio test for stationarity based on a multiple Markov process, i.e. for panel data of debtors. In the case of only one transitory state, an approximation of the alternative parameter space can be found, for instance, with Laguerre polynomials in Kiefer (1985). In our model, with several transitory rating states, the unknown hazard rates in the alternative are approximated by step-functions. Piecewise constant hazards occur in Bayesian duration analysis (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-dependent intensities can be interpreted as a continuous-time generalization of time-variability in dependence of the Markov chain. In this sense, the paper is an extension of a test for stationary dependence in discrete-time Markov chains from 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 piecewise constant hazard rates. For globally constant hazard rates, Albert (1962) established a 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 are the martingales that arise by counting transitions between the rating states. Finally, a martingale limit theorem by Rebolledo (1980) applies. A large part of the proof is to study the predictable covariation process, using Lenglart's inequality.

Our application is credit risk, in particular, the stationarity of rating transition intensities in an internal rating system. Additional applications might arise, for instance, in labor market dynamics.

2 The Model

We consider time-continuous discrete-state Markov processes $X = \{X_t, t \in [0,T]\}$ defined on a probability space $(\Omega, \mathfrak{F}, P)$. The set of states $K = \{1, \ldots, k\}$ includes states 1 to k (e.g. rating classes), where k is an absorbing state (e.g. bankruptcy). We denote X_t as the state of an asset at time t, after a certain origin, which means that we observe multiple spell data. 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 $p_{hj}(s,t) = P(X_t = j \mid X_s = h)$ give the conditional probability for a transition from state h to j, within the time period from s to t. Denote by $m_h(t)$ the unconditional 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 occurs whenever intensities are constant over time. In such cases, the transition matrices can be represented as a matrix exponential of

$Q = (q_{hj})_{h,j \in K}$, where $p_{kj}(s,t) = q_{kj} = 0$ for $j \neq k$.

Defining $q_{hj}(t)$ as a step function can approximate any arbitrary function.

Definition 2.1 *Let the intensities on* $[0,T]$ *with the given change-points* $0 = t_0 < t_1 < \ldots < t_{b-1} < t_b = T$ *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 selected differently for different rating class combinations. Step functions are commonly used to approximate smooth functions, even though other approximations, for instance by wavelets, are conceivable.

The data are transition histories $\mathbf{X}_i = \{X_t^i, t \in [0,T]\}$ for each of the $i = 1, \ldots, n$ assets in a sample. We observe the panel continuously over 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, let the processes $Y_h(t)$ denote the number of assets 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.

We impose two additional assumptions:

(A1) For fixed t

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

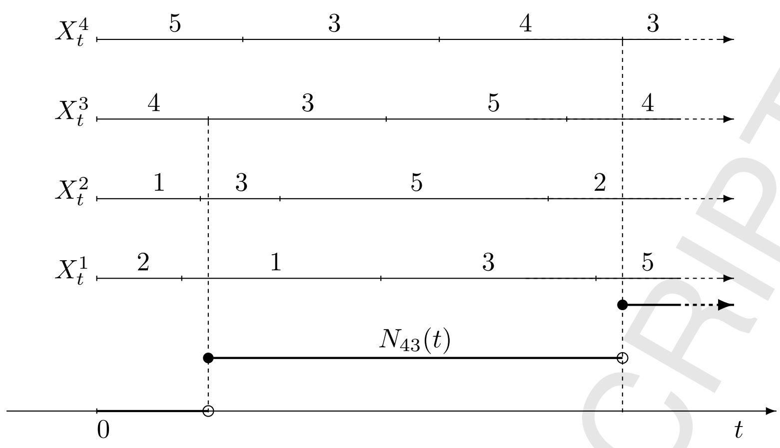

Figure 1: Four Markov processes and the counting process $N_{43}(t)$

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

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

Due to the law of large numbers, assumption (A1) is fulfilled if the Markov processes are independent. If independence cannot be expected, (A1) is a weaker assumption which, however, suffices for our results.

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

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

where $\sum_{j \neq h}$ is short for $\sum_{h=1}^{k-1} \sum_{\substack{j=1 \\ j \neq h}}^{k}$ and adds up all possible state combinations.

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

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

with the alternative

$$H_1 : \exists \; \delta_{hjl} \neq 0. \tag{3}$$

3 Results

Our aim is to construct a likelihood ratio test of stationarity in a multiplicative intensity model. Likelihood ratios are usually asymptotically χ^2 under certain regularity conditions. In our case, there are two obstacles. Firstly, there is right censoring, at time T or because of a loss to follow-up. Secondly, the q_{hj} are nuisance parameters, requiring a profile likelihood.

Denote the partial likelihood ratio by

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

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

The following theorems demonstrate that the asymptotic distribution of the test statistic remains χ^2, their proofs follow thereafter.

Theorem 1 *For a sample of Markov processes with an intensity as in Definition 2.1, let assumptions (A1) and (A2) be fulfilled. The partial ML-estimators of the parameters then converge in distribution ($\xrightarrow{d}$) to a Gaussian random vector*

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

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 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 δ_{hjl} 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). Additionally, Wald and score (Lagrange Multiplier) tests can be derived from the asymptotic normality. However, we restrict to the likelihood ratio test as an example.

Corollary 2 *Under the assumptions of Theorem 1 we have*

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

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.

With explicit expressions of the ML-estimates, the test statistic becomes computable.

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

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

Under the alternative (3), *one obtains*

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

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

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

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

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

It is evident, that $\tilde{\hat{q}}_{hj}$ depends on the number of transitions from h to j, as well as on the number of assets in state h until time t_1. Similar behavior can be observed with the $\hat{q}_{hjl}$. The latter only depend on the transitions and number of assets in state h between time t_{l-1} and t_l. The estimates are derived from the transition counts and duration times of a trimmed data set, defining time t_{l-1} as starting point 0 and t_l as the end of a particular 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, implying that the test statistic $-2\log\Delta$ follows a χ^2-distribution, see Theorem 2. Explicit formulae for parameter estimates and the likelihood ratio of Theorem 3 facilitate various 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, so that 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. However, we first 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*

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

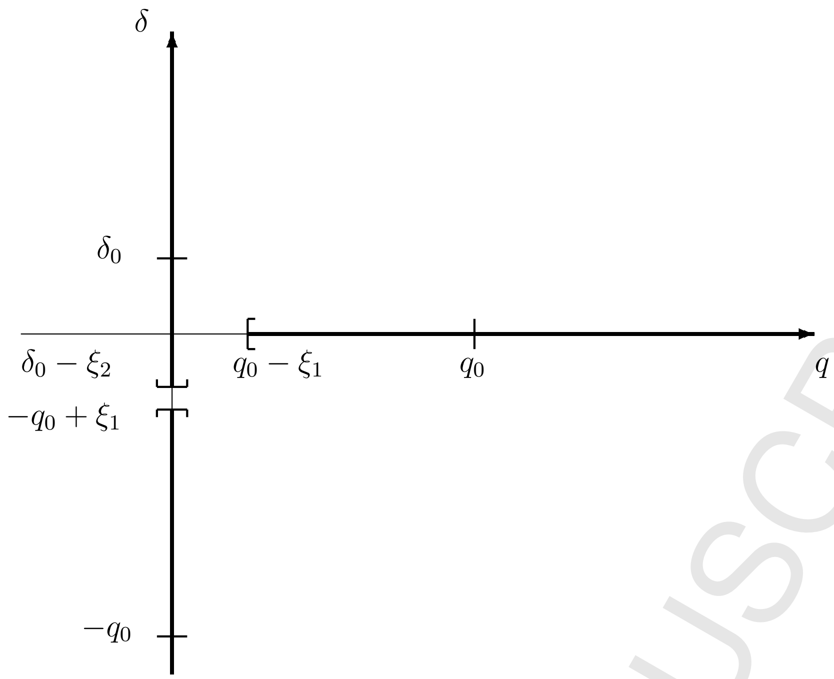

Figure 2: Environment of q_0 and δ_0

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: All eigenvalues e of A should be positive. Using 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 $a_i, c_i > 0$. □

Lemma 4.2 *For $q \in (0, \infty)$ and $\delta \in (-q, \infty)$ and for all true parameters q_0 and δ_0, there exist $\xi_1, \xi_2 > 0$, 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. □

In the interest of simplicity, we now restrict our analysis, for the time being, to the case of only one change-point, namely

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

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 in

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

and

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

They are obviously continuous in q_{hj} and δ_{hj}. The mixed second and third derivatives with respect to δ_{hj} and q_{hj} yield 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. □

We now 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 wish 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 by the means of the martingale central limit theorem (Rebolledo, 1980; Andersen et al., 1993, Theorem II.5.1). To this end, two properties must be demonstrated. First, its covariation processes must converge in probability to a covariance matrix. The covariation processes depend mainly 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*

$$\frac{1}{n} \int_0^T \frac{\partial \log(\lambda_{hj}(t))}{\partial \theta_{il}}\Big|_{\theta_0} \frac{\partial \log(\lambda_{hj}(t))}{\partial \theta_{xy}}\Big|_{\theta_0} \lambda_{hj}(t, \theta_0) dt = 0. \tag{12}$$

The only covariances that do not equal zero are

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

and

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

Hence, the covariance matrix Σ yields, 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.□

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

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

□

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

We now consider the linear term of the Taylor expansion (10).

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

Proof: It is possible to formulate the entries of $\frac{1}{n}\Im_T(\theta_0)$ as the sum of the terms 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).
□

In the following, we can show that $\frac{1}{n}R_T^{ilm}(\theta^*)$ is bounded in probability by a constant M, hence the quadratic term in the Taylor expansion disappears 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 $H_{hjn}(t)$ independent of q_{hj} and δ_{hj}, with*

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

Furthermore, 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 also less than or equal to $H_{hjn}(t)$. It now holds with (6)

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

□

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

Proof: First, (16) is less than 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 from 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 $\frac{1}{n}U^i_T(\theta_0)\stackrel{P}{\to}0$ and Lemmata 4.6 and 4.8, the ML-estimate $\hat\theta$ exists and is consistent.

With (10) and Lemma 4.8, it holds that:

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

Now, it follows with Lemma 4.6 that:

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

where

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

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. In this section, another order will be convenient. Let $(\hat{\delta}, \tilde{\hat{q}})$ 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 $(0, \hat{q})$ the restricted estimator, where $\hat{q}$ includes all $\hat{q}_{hj}$. We wish to show that

$$-2\log\frac{L(0,\hat{q})}{L(\hat{\delta},\tilde{\hat{q}})} \overset{n\to\infty}{\sim} \chi^2_{(b-1)(k-1)^2}.$$

With Theorem 1, we have:

$$\begin{pmatrix} \hat{\delta}-\delta_0 \\ \tilde{\hat{q}}-q_0 \end{pmatrix} \xrightarrow{d} N\left(0, \Gamma^{-1} = \begin{pmatrix} \Gamma^{\delta} & \Gamma^{\delta,q} \\ \Gamma^{q,\delta} & \Gamma^{q} \end{pmatrix}\right) \tag{20}$$

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

$$-2\log\frac{L(0,\hat{q})}{L(\hat{\delta},\tilde{\hat{q}})} \doteq (\hat{\delta}-\delta_0)\Gamma^{\delta}(\hat{\delta}-\delta_0).$$

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 in

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

In the case of stationary intensities where $\delta_{hjl} = 0 \ \forall j \neq h \ h, j \in K, l = 2, \dots, b$ we 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

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

We now obtain the partial likelihood ratio

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

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

Capital ratios are important for banks, and depend on the rating transitions of the portfolio counterparts in two ways. On average, the ratios are 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. The 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 over the business cycle (see Bangia et al., 2002). The transition histories may be 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 s may indicate instationary behavior of rating transitions, e.g. $\hat{P}(0,t)$ and $\hat{P}(1,t)$ must be n-asymptotically equal for a stationary process. Figure 3 shows the dissimilarity for the rating combinations $\hat{p}_{43}(0,t)$ and $\hat{p}_{43}(1,t)$.

Simultaneous inference for all rating combinations corrects for spurious effects. The simultaneous test for stationarity of rating transitions, based on the test statistic $-2\log(\Delta)$, however, is only asymptotical due to Corollary 2. A Monte Carlo simulation can serve to assess its finite sample properties, under the conditions of the data. We studied the type I error, using the generator estimated with $\hat{q}_{hj}$ of Theorem 3 for the data at hand (as in Casjens et al., 2007). At a nominal significance level of 5%, the actual size for a sample size of $7,000$ independent rating histories was found to be 0.75%. This means that the test is very conservative, causing interpretation problems, when the

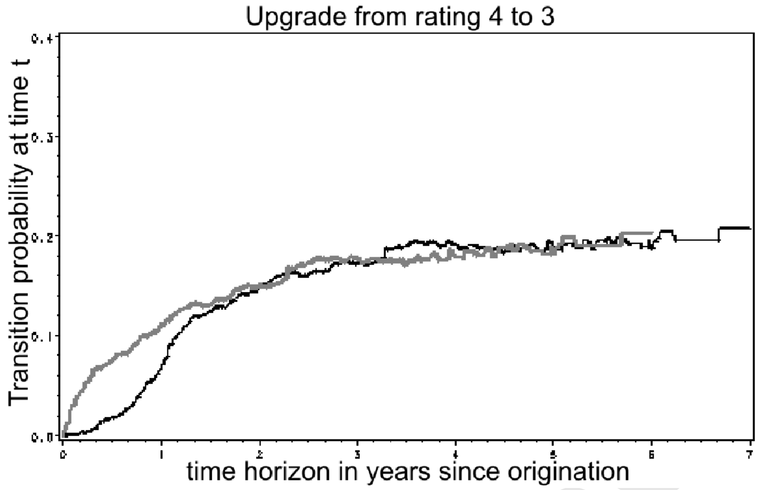

Figure 3: Nonparametric estimates for the t-years transition probability at credit origination (black line) and after one year (grey line)

test does not reject. In simulations for type II error, we found that, for a doubling of the hazard after $T/2$, the power achieves virtually 100% for around $n = 1,000$ processes. We also considered monotone exponentiated Weibull hazard functions for the simplified case of two rating classes with intensity

$$q_{12}(t) = \frac{\alpha\theta q_{12}(1 - exp(-(q_{12}t)^{\alpha}))^{\theta-1} exp(-(q_{12})^{\alpha})}{1 - (1 - exp(-(q_{12}t)^{\alpha}))^{\theta}}, \tag{21}$$

where $q_{12} = 0.1,\ \alpha = 0.9$ and $\theta = 1$ for a monotone decreasing and $q_{12} = 0.1,\ \alpha = 2$ and $\theta = 1$ for a monotone increasing shape (see Figure 4). The results for type II error were similar to the piecewise constant alternative.

To continue our empirical analysis of the internal ratings, 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.

[Table 1 about here]

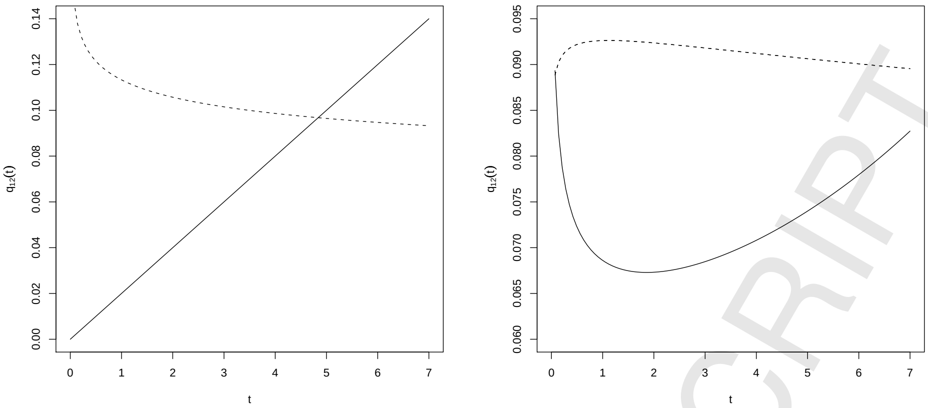

Figure 4: Monotone (left), convex and concave intensities (right) from an exponentiated Weibull family

The strikingly small p-values are listed in Table 1, and, 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 possible explanation of the result for $b = 3$ is potential 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 means of the test. A possible reason is the non-monotony of some of the intensities. For illustration purposes, the previous examples of the exponentiated Weibull family (21) allow for both a convex and a concave intensity shape with parameter values $q_{12} = 0.05,\ \alpha = 5$ and $\theta = 0.175$ for the convex shape and $q_{12} = 0.1,\ \alpha = 0.91$ and $\theta = 1.13$ for the concave shape, depicted in Figure 4. In a simulation study, again for $T = 7$, we tested against one change-point at $T/2$. For a sample size as large as $n = 10,000$, the convex-shaped intensity was associated with a type II error of 0.487, for the concave intensity the error was even 0.918.

In a simplified situation, Weißbach and Dette (2007) propose a globally consistent test that detects any alternative. From a practical point of view, this deficiency is accounted for here, by processing our test on different partitions.

6 Concluding Remarks

The question is whether a potential instationarity of rating transitions, detected by the proposed test, may not be due to ignored covariates. Systematic economic activity constitutes a documented covariate for rating transitions (Koopman et al., 2008). Systematic risk could lead to higher downgrade intensities during a recession, compared to upswings. This intuition is true for migrations measured in calendar time. However, in this case, rating histories are no longer independent of one another. Counting in portfolio time warrants, at least approximately, that assumption (A1), our proxy for independence, is valid.

Another aspect is that covariates, even though known to be influential, may not be available, so that the model may be under-specified. Heckman and Singer (1984) show, for single spell data, that under-specification causes negative duration dependence.

If stationarity is rejected, there may be microeconomic covariates, which influence the intensities. These may be time-dependent variables, such as return on investment of the obligor, clearly implying instationary intensities, as well as variables that are constant over time, such as trade, thus causing confounding problems. Modelling these variables and testing for time-stationarity of the baseline intensity may be possible, but with a model that is yet to be validated. Our aim was to show that, free of any model apart from the Markov assumption, portfolio age is a covariate that must be accounted for in further research on rating transitions.

Acknowledgement: We would like to thank two anonymous referees for their helpful comments, P. Tschiersch for pointing out some useful references and Y. Kim for the productive discussions. The financial support from the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures" and SFB 823 "Statistical modelling of nonlinear dynamic processes") is gratefully acknowledged.

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, 1:15–20, 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.

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

J.J. Heckman and B. Singer. Econometric duration analysis. Journal of Econometrics, 24:63–132, 1984.

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.

S.J. Koopman, A. Lucas, and A. Monteiro. The multi-state latent factor intensity model for credit rating transitions. Journal of Econometrics, 142: 399–424, 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 – with applications to the multivariate assessment of bioequivalence. 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, 36:575–596, 2009.

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

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