

# Kolmogorov-Smirnov-type testing for the partial homogeneity of Markov processes - with application to credit risk

Rafael Weißbach<sup>†\*</sup> & Holger Dette<sup>‡</sup>

<sup>†</sup>Institut für Wirtschafts- und Sozialstatistik, Universität Dortmund, Dortmund, Germany

<sup>‡</sup>Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum, Germany

6. Januar 2006

---

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

JEL classifications. C52, G18. AMS subject classification 62M02

Keywords. Rating migrations, Credit risk, Markov processes, hypothesis testing.

## Abstract

In banking the default behavior of the counterpart is of interest not only for the pricing of transactions under credit risk but also for the assessment of portfolio credit risk. We develop a test against the hypothesis that default intensities are constant over time within a homogeneous group of counterparts under investigation, e.g. a rating class. The Kolmogorov-Smirnov-type test builds on the asymptotic normality of counting processes in event history analysis. Right-censoring accommodates for Markov process with more than one no-absorbing state. A simulation study and an example of rating migrations support the usefulness of the test.

## 1 Introduction

Financial credit risk is currently of interest in connection with the recent release of new regulatory capital requirements for financial intermediaries (Basel Committee on Banking Supervision (June 2004)). Owners of credit portfolios have to hold capital corresponding to a high quantile of the credit-portfolio loss distribution. Banks are allowed - after approval of the national regulatory forces - to estimate the default probabilities (PD) of their counterparts based on internal rating data. These probabilities enter the capital charges in the basic and advanced version of the internal ratings based approach and are hence of great interest. Additionally, several internal models are currently used to calculate the loss distribution. The economic capital for the specific portfolio (Gordy (2000)) might well be different from the lump-sum regulatory percentages. The rating transition matrix is a parameter in some of them, e.g. CreditMetrics (see Gupton et al. (1997); Finger (1998)). Moreover, the rating migration intensities are used in banking to calculate expected losses which arise not only from defaulted counterparts but also from counterparts which are down rated and thus owning their debt has less value. Rational prices valued by the “fundamental pricing formula” (Duffie

and Singleton (1999)) depend on the intensities. It is important to note that markov process models are also applied to fields beside finance, take Hoem (1977) as an (early) example in labour market econometrics.

Typically, estimation of the rating transition matrix bases on a homogeneous Markov process, i.e. the assumption that the migrations have constant default intensities. The maximum likelihood estimate of the generator is on the one hand theoretically well established (Albert (1962)) and on the other hand popular in credit risk modelling (Bluhm et al. (2002)). The estimate is only unbiased if the model holds, and Basel II requests estimating the PD - which is the last column of the migration matrix - free of any (known) bias. Recently, Lando and Skødeberg (2002) analysed *continuous* rating migrations and raised doubts about the Markovian property. Applying a Cox regression, they reject for many transitions the hypothesis of a deterministic intensity and suspect a “rating drift”. We restrict ourselves to the markovian assumption and consider the problem of homogeneity. Practically speaking: “Do rating transitions depend on time?” A test for the homogeneity of the *discrete* Markov chain modelling of credit rating migrations is found in Kiefer and Larson (2004). They are concerned with answering the question on basis of given transitions matrices for several periods and find that the homogeneous markov model is sufficient. However, they raise doubts about the adequacy of their data for a rigorous proof. In contrast to Weißbach et al. (2005) - who construct a global likelihood ratio test against the specific alternative of structural breaks - we consider the singles migrations (as in Lando and Skødeberg (2002)) and derive a globally consistent test. The Kolmogorov-Smirnov idea modifies the well-known one-sample log-rank statistic (see e.g. Andersen et al. (1993)), using a general asymptotic result for counting processes by Hjort (1990). Although, for non-homogeneous Markov processes, nonparametric alternatives for the estimation of the migration matrix exist, e.g. the Aalen-Johansen estimator (see Aalen (1978); Aalen and Johansen (1978); Fleming (1978)), we propose to model the generator based on the

assessment of the single transitions, to enable “partial homogeneity”. We see no need in nonparametric estimation for the entire matrix. For those transitions where the homogeneity need not be rejected, constant intensities reflect the data. For the transitions where the homogeneity is rejected, we propose to use kernel estimates as described e.g. in Weißbach (2006). It is important to note that we take into account censoring of observation because defaults (and even rating migration) are rare events. Rating histories without movement belong to the population under risk and demand for censored models in continuous modelling.

A simulation study confirms the consistency as well as the sufficient power of the test in practice. An application to internal rating migration data of a cooperating bank finds inhomogeneities for few migrations to neighboring rating classes (up- and down).

## 2 A consistent test for homogeneity

For the derivation, we consider the case of only one performing and one (absorbing) non-performing rating class. The intensity of the Markov process expresses equivalently as hazard rate of the default time. The application to a rating system with several performing states describes Section 4. Especially for the general model, but as well for the two-stage model, right-censoring of transition times needs consideration. We assume to observe right-censored default times  $X_i = \min(T_i, C_i)$  with censoring indicator  $\delta_i = 1_{\{X_i=T_i\}}$ . The default times  $T_1, \dots, T_n$  are assumed to be independent identically distributed with distribution function  $F$ , the censoring time  $C_1, \dots, C_n$  are assumed to be independent identically distributed with distribution function  $G$ . The  $T_i$ 's are assumed to be independent of the (censoring times)  $C_i$ 's. We are interested in the hazard rate

$$\lambda(t) := \frac{f(t)}{1 - F(t)}$$

of the actual default times  $T_i$ .

As test for a known hazard rate, the “one-sample log-rank” test, was proposed by Breslow (1975). In the context of counting processes the test is discussed in Andersen et al. (1993). Questions whether statistical characteristics are independent of time are typical in financial statistics. The idea is to compare the increments of the unrestricted estimator of the cumulative hazard rate with those of the estimator parametrically arising from the hazard rate model under the hypothesis  $H_0 : \lambda(t) = \lambda_0(t)$ .

We want to test for a constant hazard rate, i.e. for an exponential model. The time-invariance is a typical question in finance as it constitutes a reduction in dimension of the parameter space. Simplified estimation is the result. E.g. Dette and von Lieres und Wilkau (2003) investigate the volatility in the Black-Scholes stochastic differential equation used for the description of stock, exchange rate or interest rate development.

As a prerequisite we need to recall the estimate of the hazard rate under the null hypothesis, or simply the parameter of an exponential distribution (with density  $f(x) = \lambda e^{-\lambda x}$ ). Note that  $\lambda^{-1}$  is the expectation of the exponential distribution. For the random right-censoring the ratio of “number of uncensored events” to “exposure to risk”,

$$\lambda_n := \frac{n_u}{\sum_{i=1}^n X_i}, \tag{1}$$

is the long-known maximum likelihood estimate. The number of uncensored observations is defined as  $n_u := \sum_{i=1}^n \delta_i$ .

We construct the test statistic for the hypothesis

$$H_0 : \quad \text{“Is } \lambda(t) \text{ constant?”} \quad \Leftrightarrow \quad \lambda(t) \equiv \lambda_0 \quad \text{unknown}$$

in the same manner as above. The differences of the increments of the unrestricted estimator of the cumulative hazard and those of the estimator arising for an exponential distribution are added up.

The non-parametrical Nelson-Aalen estimator of the cumulative hazard

rate  $\Lambda(t) := \int_0^t \lambda(s) ds$  is

$$\hat{\Lambda}(t) = \int_{0 < s \leq t} H(s) dN(s) = \sum_{i: X_{(i)} \leq t} \frac{\delta_{(i)}}{n - i + 1},$$

where  $N(t) := \#\{i : X_i \leq t, \delta_i = 1\}$  is the process counting the defaults,  $J(t) := I\{Y(t) > 0\}$  and  $H(t) := \frac{J(t)}{Y(t)}$  with  $\frac{0}{0} := 0$ . The population under risk at time  $t$  is denoted by  $Y(t) := \#\{i : X_i \geq t\}$  (see Nelson (1972); Aalen (1978)). The definition of  $H(s)$  “stops” the estimation of the cumulative hazard rate at the largest observation  $X_{(n)}$  (irrespective of censoring).

Using estimate (1) for the hazard rate under the null hypothesis yields

$$\begin{aligned} \Lambda_0^*(t) &= \int_{0 < s \leq t} J(s) \lambda_n ds \\ &= \min(X_{(n)}, t) \lambda_n = \frac{n_u \min(X_{(n)}, t)}{\sum_{i=1}^n X_i} \end{aligned}$$

as the parametric estimate for the cumulative hazard rate, again capped at  $X_{(n)}$ .

The log-rank statistic weights the incremental differences with the number of observations under risk and adds them (Andersen et al. (1993)):

$$\begin{aligned} Z(t) &:= \int_0^t K(s) d\{\hat{\Lambda}(s) - \Lambda_0^*(s)\} \\ &= \int_0^t dN(s) - Y(s) \lambda_n ds \quad \text{for } K(s) = Y(s) \\ &= N(t) - E^*(t), \end{aligned}$$

where

$$E^*(t) := \int_0^t Y(s) \lambda_n ds = \frac{n_u \sum_{i=1}^n \min(X_i, t)}{\sum_{i=1}^n X_i}$$

is the expected number of defaults in  $[0, t]$  under hypothesis. The explicit test statistic is finally obtained as

$$Z(t) = \sum_{i=1}^n I\{X_i \leq t, \delta_i = 1\} - \frac{(\sum_{i=1}^n \delta_i)(\sum_{i=1}^n \min(X_i, t))}{\sum_{i=1}^n X_i}.$$

For fixed  $t$  the asymptotic distribution of the standardized test statistic is given by a normal distribution, i.e.

$$\frac{Z(t)}{\sqrt{\langle Z \rangle(t)}} \xrightarrow{d} N(0, 1)$$

if  $n \rightarrow \infty$ . Here the predictable variation process  $\langle Z \rangle(t)$  is again  $E^*(t)$  and the asymptotics is clear because of the central limit theorem and Slutsky's lemma (see for example Andersen et al. (1993).335).

However, the test, which rejects the null hypothesis of a constant hazard rate for large values of the statistic  $|Z(t)/\sqrt{\langle Z \rangle(t)}|$  is not uniformly consistent. E.g. if the hazard rates is of the form  $\lambda(s) = c\lambda + 2\lambda(1 - c)s/t$  for any  $c \in [0, 1)$ , the increments sum up to 0 in the limit, the power converges to the size of the test, not to 1. To avoid this defect, we propose to use a Kolmogorov-Smirnov- or Cramer-von-Mises-type test statistic of the process  $Z(t)$ .

For this purpose we define

$$\hat{\sigma}^2 := \int_0^T \frac{dN(s)}{n\lambda_n^2} = \frac{n_u}{n\lambda_n^2} = \frac{(\sum_{i=1}^n X_i)^2}{nn_u}.$$

For hazard rates bounded away from 0 on  $[0, T]$  it follows from Hjort (1990) that, under the null hypothesis of a constant hazard rate, the process

$$(Z(t)/(\sqrt{n}\lambda_n\hat{\sigma}))_{t \in [0, T]}$$

converges weakly in  $D[0, T]$  to a Brownian bridge, say  $W^0$ , in time  $p(t) = t/T$ , i.e.

$$\frac{Z}{\sqrt{n}\lambda_n\hat{\sigma}} \xrightarrow{d} W^0 \circ p.$$

By the continuous mapping theorem (see Shorack and Wellner (1986)) the Kolmogorov-Smirnov- and Cramer-von-Mises-type statistic show a similar

behaviour, that is

$$K_n = \max_{0 \leq t \leq T} \left| \frac{Z(t)}{\sqrt{n\lambda_n\hat{\sigma}}} \right| \xrightarrow{d} \max_{0 \leq t \leq T} |W^0(t/T)| \stackrel{d}{=} \max_{0 \leq t \leq 1} |W^0(t)|$$

$$C_n = \frac{1}{T} \int_0^T \left| \frac{Z(t)}{\sqrt{n\lambda_n\hat{\sigma}}} \right|^2 dt \xrightarrow{d} \frac{1}{T} \int_0^T |W^0(t/T)|^2 dt \stackrel{d}{=} \int_0^1 |W^0(t)|^2 dt$$

As a consequence rejecting the null hypothesis of a constant hazard rate for large values of the statistic  $K_n$  or  $C_n$  will yield a consistent test. In particular, if  $k_\alpha$  denotes the  $1 - \alpha$  quantile of the the distribution of the random variable  $\max_{0 \leq t \leq 1} |W^0(t)|$  the rule

$$\phi((X_1, \delta_1), \dots, (X_n, \delta_n)) := I_{\{\max_{0 \leq t \leq T} |\int_0^t dN(s) - Y(s)\lambda_n ds| > k_\alpha \hat{\sigma} \sqrt{n\lambda_n}\}} \quad (2)$$

corresponds to the Kolmogorov-Smirnov statistic and constitutes a test with asymptotic level  $\alpha$ . The inequality in the indicator function is most explicitly expressed as

$$\max_{0 \leq t \leq T} \left| \sum_{i=1}^n I\{X_i \leq t, \delta_i = 1\} - \frac{(\sum_{i=1}^n \delta_i)(\sum_{i=1}^n \min(X_i, t))}{\sum_{i=1}^n X_i} \right| > k_\alpha \sqrt{n\lambda_n},$$

while the quantiles  $k_\alpha$  can be found in Shorack and Wellner (1986), p. 143. For example, the 95% and 90% quantile are given by  $k_{0.05} \approx 1.36$  and  $k_{0.1} \approx 1.23$ , respectively. In the following section we will discuss properties of this test by means of a simulation study.

### 3 Simulation study

A Monte Carlo simulation assesses the validity of the asymptotic approximation. The hazard rate under hypothesis,  $\lambda_0$ , calibrates the mediocre one-year probability of default of 1% corresponding approximately to a rating of ‘Ba’ in Moody’s system (see Nickell et al. (2000)) or ‘BB’ in Standard & Poor’s system (see Gordy (2000)). Hence, the constant hazard of the survival (as



well as for the censoring) is  $-\log(0.99)$ . The censoring distribution with hazard rate of  $-\log(0.99)$  models a 50%-censoring. The percentage resembles the probability to change to another than the investigated rating class, given a migration happens. We specify  $T$  to be 100 or 200 years. The maximization of the statistics  $Z(t)$  over  $[0, T]$  is performed on a discrete grid. Clearly, the maximum increases with the fineness of the grid. However, monthly intervals are found to be sufficiently fine. 2000 simulation loops were used. Table 1 demonstrates for 50 observations a slightly conservative approximation. For 100 observation the approximation is equally good for both 100 years and 200 years support. For 500 observations the approximation is even slightly anti-conservative for 200 years of observation.

Clearly, the support of the exponential distribution is  $\mathbb{R}_0^+$ . However, Table 1 demonstrates that the restriction to 85%-quantile of the survival distribution, i.e. approximately  $[0, 200]$ , is sufficient for the hazard rate  $-\log 0.99$ . Less than 2% of the observations range larger than that. The 65%-quantile  $[0, 100]$  of the survival distribution does not appear to be sufficient, almost 15% of the observations range larger than that.

– insert Table 1 around here –

If the end of the support  $T$  is taken as additional censoring for the observations, the censoring distribution is not the same as the survival distribution anymore. However, the default distribution is not altered and the test can still decide whether its hazard rate is constant up to  $T$ . Now, we may decrease  $T$  in order to model realistic designs. Usually, only moderately recent default information for a set of companies is considered homogeneous and economically relevant. We believe that at least 5 years of observations should be assembled. Better are clearly 10, 20 or 50 years.

We need to explain a numerical effect first before interpreting the results. As the discrete evaluation points are less for monthly spacing and restricted support, the fineness has to be increased. Especially for a large number of

observations, the path of  $Z(t)$  gets finer, the number of in-discontinuities increases linearly and maxima are more likely to slip through the grid. The effect was found to be negligible up to 1000 observations (on a monthly grid). The maximization of the statistics  $Z(t)$  over  $[0, T]$  is performed on a weekly grid for 2500 observations.

– insert Table 2 around here —

Table 2 lists simulated sizes for time horizons found in the practice of credit risk. The censoring in a simulation sample may well be up to almost 100%. However, the size for 500 observations of 2.75% advocates the usefulness of the test even for a short history. It is interesting to see that short history can be traded-off against a larger sample and vice versa, an effect known since Albert (1962). Combinations of long history and many observations with obviously valid sizes are not displayed for computational ease.

The power assessment needs a specification of the alternative. Apart from generalizations of the Exponential distribution, linear increase of the hazard rate is another, with a special advantage: The distance to the hypothesis, the constant hazard, is easily expressed in terms of the slope of the hazard. The hazard rates circle around the point  $(T/2, \lambda_0)$ . The intercept in our simulation is least 20% of  $\lambda_0$ , corresponding to a maximal slope of 0.002 in our example where the end of the support is 10 years. Figure 1 of the simulated power reveal for 10 years of study period the difference in convergence for 100 and 500 observations. All other conditions of the simulation are as in the size simulations earlier. Whereas for 500 observations a slope of 0.002 is almost sufficient for a power of 100%, only 25% power result for 100 observations.

– insert Figure 1 around here —

## 4 Example

WestLB provided us with a rating transitions for an internal rating system with 8 non-default classes observed over the seven years 1997 through 2003. The time origin is the event of entering the credit portfolio (of the cooperating bank), hence counterparts with business relations before 1997 had to be ruled out. To protect the interest of the portfolio owner, only around 600 transitions for randomly chosen 10% of the counterparts (=360) are analysed. Only first transitions can be analysed, second (and third, etc.) transitions can not be incorporated into the analyses because the methodology does not allow for left truncation of observation. Clearly, we see room for methodological improvement to integrate residual transitions. No migration during the business relation (capped with the seventh year) as well as migration to another but the target class constitute censoring events. The preliminary analysis assumes the trivial rating system of only one rating class, i.e. combines all non-default classes. The test statistics (2) has the value 334.68 and does not exceed the critical value at level 5% of 353.57. However, that does not prove that no migration on the dis-aggregate level is inhomogeneous, as we will see. For the migrations to neighboring classes the results of the test are displayed in Table 3. For the upgrade from class 4 to 3 and for the downgrade from 3 to 4 the intensity proves not to be constant. Further modelling and estimation effort is necessary. All other migration are - based our the restricted data pool - unobtrusive. Constant intensity reflects the experience expressed in the data.

– insert Table 3 around here —

*Acknowledgement.* We would like to thank Frederik Kramer for engaged computational assistance and are grateful to WestLB AG, Düsseldorf for access to its database. Furthermore we acknowledge the financial support of DFG, SFB 475 “Reduction of Complexity in Multivariate Structures”,

project B1.

## Literatur

- O.O. Aalen. Nonparametric estimation of partial transition probabilities in multiple decrement models. *Annals of Statistics*, 6:534–545, 1978.
- O.O. Aalen and S. Johansen. An empirical transition matrix for non-homogeneous markov chains based on censored observation. *Scandinavian Journal of Statistics*, 5:141–150, 1978.
- A. Albert. Estimating the infinitesimal generator of a continuous time, finite state markov process. *Annals of Mathematical Statistics*, 38:727–753, 1962.
- P.K. Andersen, Ø. Borgan, R.D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer-Verlag New York, 1993.
- Basel Committee on Banking Supervision. International convergence of capital measurement and capital standards. Technical report, Bank for International Settlements, June 2004.
- C. Bluhm, L. Overbeck, and C. Wagner. *An Introduction to Credit Risk Modeling*. Chapman & Hall, 2002.
- N.E. Breslow. Analysis of survival data under the proportional hazards model. *International Statistical Review*, 43:45–58, 1975.
- H. Dette and C. von Lieres und Wilkau. On a test for a parameteric form of volatility in continuous time financial models. *Finance and Stochastic*, 7: 363–384, 2003.
- D. Duffie and K. Singleton. Modeling term structures of defaultable bonds. *Review of Financial Studies*, 12:687–720, 1999.

- C.C. Finger. Sticks and stones. Technical report, RiskMetrics Group publication, New York, 1998.
- T.R. Fleming. Nonparametric estimation for non-homogeneous markov processes in the problem of competing risks. *Annals of Statistics*, 6:1057–1070, 1978.
- M.B. Gordy. A comparative anatomy of credit risk models. *Journal of Banking and Finance*, 24:119–149, 2000.
- G.M Gupton, C.C. Finger, and M. Bhatia. CreditMetrics. Technical report, J.P.Morgan, 1997.
- N.L. Hjort. Goodness of fit tests in models for life history data based on cumulative hazard rates. *Annals of Statistics*, 18:1221–1258, 1990.
- J.M. Hoem. A markov chain model of working life tables. *Scandinavian Actuarial Journal*, 1:1–20, 1977.
- N.M. Kiefer and C.E. Larson. Testing simple markov structures for credit rating transitions. Technical report, Cornell University, 2004.
- D. Lando and T. Skødeberg. Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking and Finance*, 26:423–444, 2002.
- W. Nelson. Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14:945–966, 1972.
- P. Nickell, W. Perraudin, and S. Varotto. Stability of ratings transitions. *Journal of Banking and Finance*, 24:203, 2000.
- G.R. Shorack and J.A. Wellner. *Empirical processes with applications to statistics*. John Wiley & Sons, New York, 1986.

- R. Weißbach. A general kernel functional estimator with general bandwidth – strong consistency and applications. *Journal of Nonparametric Statistics*, to appear, 2006.
- R. Weißbach, P. Tschiersch, and C. Lawrenz. Testing homogeneity of time-continuous rating transitions. Technical report, University Dortmund, Technical Report 34, 2005.

Tabelle 1: Size of test  $\phi$  (2) for various number of observations and support 100 and 200 years. The default intensity is  $\lambda = -\log 0.99$ .

no. obs	100 years	200 years
50	0.0415	0.0305
100	0.0405	0.041
500	0.045	0.0525

Tabelle 2: Size of test for various number of observations and support 100 and 200 years. Parameters: Hazard rate underhypothesis  $\lambda_0 = -\log 0.99$ , 2000 simulations, monthly maximization.

no. obs	5 years	10 years	20 years	50 years	100 years	200 years
50	0.009	0.0215	0.03	0.0385	0.0495	0.048
100	0.021	0.025	0.032	0.0335	0.039	0.045
500	0.0275	0.041	0.0375	0.038	0.0415	0.048
1000	0.0395	0.038	0.036	NA	NA	NA
2500	0.045	0.0435	NA	NA	NA	NA



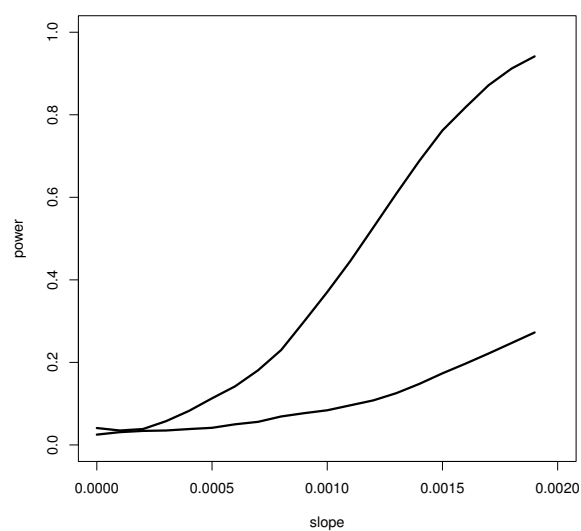


Abbildung 1: Power of the test for 10 years of  $n = 100$  (lower curve) and  $n = 500$  (upper curve) observations. The alternatives are linear hazard rates with slope as parameter of distance to the hypothesis.

Tabelle 3: Test statistics  $\phi$  (2) and critical value (at level 5%) for hypothesis that the downgrade (left) and upgrade (right) from a rating class into the neighboring class does not depend on the time since entry into the portfolio. \* marks the significant changes. Some transition (NA) have too few observation for the assessment.

Rating	$\phi$	crit. value	Rating	$\phi$	crit. value
1→2	NA	NA	2→1	NA	NA
2→3	17.02	25.81	3→2	95.59	150.32
3→4*	72.38	63.77	4→3*	58.54	54.46
4→5	57.81	67.93	5→4	24.82	32.31
5→6	34.89	49.36	6→5	3.91	10.44
6→7	NA	NA	7→6	NA	NA
7→8	NA	NA	8→7	NA	NA
8→def.	12.23	16.63			