

Neyman smooth goodness-of-fit tests for the marginal distribution of dependent data

Axel Munk · Jean-Pierre Stockis · Janis Valeinis ·
Götz Giese

Received: 2 July 2007 / Revised: 13 July 2009 / Published online: 17 November 2009
© The Institute of Statistical Mathematics, Tokyo 2009

Abstract We establish a data-driven version of Neyman’s smooth goodness-of-fit test for the marginal distribution of observations generated by an α -mixing discrete time stochastic process $(X_t)_{t \in \mathbb{Z}}$. This is a simple extension of the test for independent data introduced by Ledwina (J Am Stat Assoc 89:1000–1005, 1994). Our method only requires additional estimation of the cumulative autocovariance. Consistency of the test will be shown at essentially any alternative. A brief simulation study shows that the test performs reasonable especially for the case of positive dependence. Finally, we illustrate our approach by analyzing the validity of a forecasting method (“historical simulation”) for the implied volatilities of traded options.

Keywords Neyman’s smooth test · Goodness-of-fit · Strongly mixing process · Implied volatility

A. Munk (✉)
Institute for Mathematical Stochastics, Georg-August University of Göttingen,
Maschmühlenweg 8-10, 37073 Göttingen, Germany
e-mail: munk@math.uni-goettingen.de

J.-P. Stockis
Department of Mathematics, University of Kaiserslautern,
Postfach 3049, 67653 Kaiserslautern, Germany
e-mail: stockis@mathematik.uni-kl.de

J. Valeinis
Department of Mathematics, Faculty of Physics and Mathematics,
University of Latvia, Zellu 8, 1002 Riga, Latvia
e-mail: valeinis@lu.lv

G. Giese
Commerzbank AG, 60261 Frankfurt, Germany
e-mail: Goetz.Giese@commerzbank.com

1 Introduction

Let $(X_t)_{t \in \mathbb{Z}}$ be a real-valued strictly stationary process defined on a probability space (Ω, \mathcal{F}, P) . For any two σ -fields \mathcal{A} and $\mathcal{B} \subset \mathcal{F}$ define the following measure of dependence

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup |P(A \cap B) - P(A)P(B)|,$$

where this latter sup is taken over all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. For $-\infty \leq J \leq L \leq \infty$ define $F_J^L := \sigma(X_k, J \leq k \leq L)$. $(X_t)_{t \in \mathbb{Z}}$ is called strongly mixing or α -mixing (see, e.g. [Bradley 2002](#)) if $\alpha(n) := \alpha(F_{-\infty}^0, F_n^\infty) \rightarrow 0$ when $n \rightarrow \infty$.

Assume that an α -mixing strictly stationary process $(X_t)_{t \in \mathbb{Z}}$ has marginal distribution function F . We want to test the following simple hypothesis:

$$H_0 : F = U[0, 1] \quad \text{versus} \quad H_1 : F \neq U[0, 1],$$

where $U[0, 1]$ denotes the uniform distribution on the interval $[0, 1]$. Note that testing $H_0 : F = F_0$ for some general continuous distribution F_0 can be reduced to this situation by transforming the data to $F_0(X_t)$, $t \in \mathbb{Z}$.

The problem of testing the goodness-of-fit of a parametric family is a classical theme in statistics. During the last decades the asymptotical power and various concepts of efficiency for such well-known goodness-of-fit tests as the Kolmogorov–Smirnov or the Cramer–von Mises tests have been analyzed and it has been concluded that these tests do not behave like well-balanced omnibus tests for non-parametric alternatives. There are only very few directions of deviations from the null hypothesis for which these tests are of reasonable finite sample size power. In contrast, smooth goodness-of-fit tests obey much better overall performance against a broad range of alternatives (see [Rayner and Best 1989](#); [Inglot et al. 1997](#) among many others). The above results renewed interest in a class of tests already introduced by [Neyman \(1937\)](#).

The topic of goodness-of-fit tests for dependent data has been less intensively investigated in the past. Here follows a brief survey. [Gasser \(1975\)](#) explored the χ^2 test for correlated data by simulations. [Moore \(1982\)](#) and [Gleser and Moore \(1983\)](#) analyzed the χ^2 test of Pearson under the positive dependency. They concluded that the null hypothesis is rejected too often. The same conclusion holds for tests based on the empirical distribution function, that is, positive dependence is confounded with lack of fit. Later, [Chanda \(1981, 1999\)](#) analyzed for the Pearson test the effects of dependence for observations generated by linear, bilinear and Volterra processes.

For specific but important dependence structures such as ARMA processes there is much more literature on goodness-of-fit tests. For some survey of results, see [Ducharme and Micheaux \(2004\)](#) and the references therein. In this work, a smooth test of normality for the innovations of an ARMA process has been established using a data-driven procedure introduced by [Ledwina \(1994\)](#).

Some results have been recently obtained for a stationary α -mixing processes $(X_t)_{t \in \mathbb{Z}}$ by [Ignaccolo \(2004\)](#). This extends the tests associated with projection density estimators considered by [Bosq \(1989, 2002\)](#). In particular, the χ^2 test of Pearson and the Neyman smooth test for a fixed number of components follow from this general

framework. [Neumann and Paparoditis \(2000\)](#) consider tests based on the Bickel–Rosenblatt statistic under an exponential α -mixing rate.

Similarly as [Ledwina \(1994\)](#) and [Inglot and Ledwina \(1996\)](#) among others, we establish a simple data-driven version of the Neyman smooth goodness-of-fit test for testing uniformity of the marginals of α -mixing processes. In principle, we extend [Ignaccolo \(2004\)](#) result making it data driven in the following way.

For testing H_0 Neyman’s smooth test statistic has the following form

$$R_k = \sum_{j=1}^k \left\{ n^{-\frac{1}{2}} \sum_{i=1}^n \phi_j(X_i) \right\}^2, \quad k = 1, 2, \dots, \tag{1}$$

where ϕ_0, ϕ_1, \dots , is an orthonormal system in $L_2[0, 1]$ with $\phi_0(x) = 1$. Further on we restrict $\{\phi_j\}$ to be the orthonormal Legendre polynomials on $[0, 1]$ (for the general case see Remark 2), sometimes denoted as shifted Legendre polynomials (see [Abramowitz and Stegun 1972](#)). These polynomials can be defined recursively and are, for $j = 0, 1, 2, 3, 4$,

$$\begin{aligned} \phi_0(x) &= 1, \\ \phi_1(x) &= \sqrt{12}(x - 1/2), \\ \phi_2(x) &= \sqrt{5}(6(x - 1/2)^2 - 1/2), \\ \phi_3(x) &= \sqrt{7}(20(x - 1/2)^3 - 3(x - 1/2)), \\ \phi_4(x) &= 210(x - 1/2)^4 - 45(x - 1/2)^2 + 9/8. \end{aligned}$$

For a strictly stationary α -mixing process $(X_t)_{t \in \mathbb{Z}}$ under some natural conditions (for our purposes we will need only a subset of those conditions, see Sect. 3), [Ignaccolo \(2004\)](#) has shown in Theorem 3.1 that under H_0

$$R_k \rightarrow_d \lambda_1 Y_1 + \lambda_2 Y_2 + \dots + \lambda_k Y_k, \tag{2}$$

where $Y_j \sim \chi_1^2$ are independent r.v.s and the coefficients λ_j are the eigenvalues of the matrix with elements

$$\sigma_{jl} = E_0(\phi_j(X_1)\phi_l(X_1)) + \sum_{i=2}^{\infty} E_0(\phi_j(X_1)\phi_l(X_i)) + \sum_{i=2}^{\infty} E_0(\phi_j(X_i)\phi_l(X_1)) \tag{3}$$

for $1 \leq j, l \leq k$. Let E_0 and Var_0 denote further on the expectation and variance under H_0 , i.e. with respect to the Lebesgue measure on $[0, 1]$. [Ignaccolo \(2004\)](#) test is based on R_k and requires to estimate all eigenvalues in (2). This may cause, however, some practical difficulties because estimation of the eigenvalues can be numerically difficult and instable, in particular, when several eigenvalues are close to each other (see [Eaton and Tyler 1991](#)). Moreover, for the test proposed by [Neumann and Paparoditis \(2000\)](#) bandwidth selection is required, which is well known to be problematic already in the independent case.

For convenience we now introduce a rescaled version of the Neyman statistic (1)

$$N_k = (12\sigma^2)^{-1} R_k = (12\sigma^2)^{-1} \sum_{j=1}^k \left\{ n^{-\frac{1}{2}} \sum_{i=1}^n \phi_j(X_i) \right\}^2 \tag{4}$$

with

$$\sigma^2 = \sum_{t=-\infty}^{+\infty} \text{Cov}(X_0, X_t). \tag{5}$$

Roughly speaking, the factor $(12\sigma^2)^{-1}$ adjusts R_k for the additional dependency, such that under H_0 the new statistic N_k has the same limiting distribution as R_k in the independent case. In fact, then

$$\sigma_{11} = \sum_{t \in \mathbb{Z}} E(\phi_1(X_0)\phi_1(X_t)) = 12 \sum_{t \in \mathbb{Z}} \text{Cov}(X_0, X_t) = 12\sigma^2. \tag{6}$$

The Neyman test R_k has regained its importance after [Ledwina \(1994\)](#) proposed a data driven selection rule for the choice of a proper k in (1) even if k has to be chosen among a (possibly) infinite number of models $1, \dots, d(n)$, where $d(n) \rightarrow \infty$ as $n \rightarrow \infty$. Based on ideas of Ledwina, Kallenberg and Inglot, in this paper, we propose the use of a modified selection rule based on N_k resulting in a very simple procedure, which adapts automatically to the dependent case. This is the main contribution of this paper.

In addition to the case of independent observations, for the implementation of our test only σ^2 in (5) has to be estimated additionally. Hence, the resulting test is surprisingly simple. The Neyman smooth test statistic for α -mixing data is only a multiple of the test statistic in the independent case with a proper estimate of σ^2 . As a consequence, in situations where the assumption of independence is questionable we simply suggest to adjust the test statistic by an estimate of σ^2 and subsequent to proceed as in the independent case.

Our results can also be regarded as a kind of robustness analysis towards dependent data of the data-driven Neyman smooth test for the i.i.d. case. First, note that $\text{Var}_0(X_0) = 1/12$. If $12\sigma^2 > 1$ (positive dependence) then the Neyman smooth test will reject too often, otherwise (negative dependence) it will be conservative.

To summarize, our main goal is to show consistency of the rescaled Neyman test using modified Schwarz’s selection rule in the α -mixing case, thus extending the result of [Ignaccolo \(2004\)](#). Simulations are performed for α -mixing processes generated from some ARMA processes, in particular AR(1) processes, which under some conditions have the α -mixing property with exponential mixing rate (see, e.g. [Bradley 1986](#), p. 179). We would like to stress that various other important classes of processes such as GARCH and bilinear processes also obey the α -mixing property (see [Mokkadem 1990](#)). Our major finding is that for positively correlated observations the proposed test works quite well. A Monte Carlo study reveals the finite sample properties even better than for the independent case. In contrast, for negatively correlated observations, estimation of the variance (5) causes severe problems and the test cannot be recommended.

The paper is organized as follows. In Sect. 2 the smooth goodness-of-fit test statistic for α -mixing processes and the corresponding model selection rule is discussed. Section 3 is devoted to technical assumptions and consistency. A brief simulation study on critical values, level and power is performed in Sect. 4. In Sect. 5 we apply our approach to the so-called historical simulation of implied option volatilities, a commonly method employed by banks to measure the market risk of their trading portfolio. Our analysis reveals the foreign exchange implied volatilities EUR-GBP and EUR-USD as nearly uncorrelated and the results of our test coincide with the test based on R_k for the independent case. For USD and EUR swaption volatilities we found serial dependency, and we rejected uniformity of the volatilities based on N_k (and also on R_k). Hence, this seems to be an indication for insufficient modeling of the implied volatilities by a non-parametric heteroscedastic dynamical system.

All proofs are deferred to the Appendix.

2 Test statistic

As mentioned in the Introduction for a strictly stationary α -mixing process $(X_t)_{t \in \mathbb{Z}}$ we suggest to use the test statistic N_k defined in (4) with σ^2 in (5), which is a rescaled version of Neyman’s smooth test for independent data, i.e. $N_k = (12\sigma^2)^{-1}R_k$.

We will show that under H_0 the Schwarz’s data-driven selection rule selects the dimension 1, thus the statistic N_k reduces to N_1 . Moreover, if $k = 1$ the eigenvalue λ_1 is equal to (6). Therefore, N_k under H_0 has the same limiting distribution as in the independent case, which is the χ^2_1 distribution.

For fixed k from (2) follows that under H_0 ,

$$N_k \rightarrow_d \left(\frac{\lambda_1}{\sigma_{11}}\right) Y_1 + \left(\frac{\lambda_2}{\sigma_{11}}\right) Y_2 + \dots + \left(\frac{\lambda_k}{\sigma_{11}}\right) Y_k. \tag{7}$$

Ignaccolo (2004) has derived the convergence rates of $R_k = 12\sigma^2 N_k$ to the limiting distribution for fixed k depending on rates of mixing coefficients $\alpha(n)$. Naturally, if σ^2 is finite we have the same asymptotic rates.

Remark 1 In the case of i.i.d. r.v.s, one has $12\sigma^2 = 1$ under H_0 , and we get the statistic introduced by Neyman (1937). It follows that if the model is misspecified and if $1/(12\sigma^2) < 1$, i.e. $\sigma^2 > 1/12$, this results in a higher rejection rate. If $\sigma^2 < 1/12$ we get a conservative test. Similarly as for Pearson’s test (see, e.g. Moore 1982; Gleser and Moore 1983), we find again that positive dependence of the data will lead to an increase of the actual level of the Neyman smooth test in the i.i.d. case. Therefore, in situations where independence might be a questionable assumption our proposal is simply to multiply the test statistic by $1/(12\hat{\sigma}^2)$, where $\hat{\sigma}^2$ denotes an appropriate estimator of σ^2 in (5), and then to proceed as in the independent case. We will show in the following that this will yield a consistent test, in particular, when the dimension k is automatically selected from the data properly.

Remark 2 It is easy to generalize our results to an arbitrary orthonormal system $\varphi_0, \varphi_1, \dots$ in $L_2[0, 1]$ with $\varphi_0 \equiv 1$. In order to present the smooth test in this more general framework write $\varphi_1(X) = (g(X) - E_0g(X))/\sqrt{\text{Var}_0g(X)}$ for some function g ,

where E_0 and Var_0 denote the expectation and the variance under the null hypothesis, respectively. The smooth test statistic is now

$$(\sigma^2/\text{Var}_0g(X))^{-1} \sum_{j=1}^k \left\{ n^{-\frac{1}{2}} \sum_{i=1}^n \varphi_j(X_i) \right\}^2,$$

where $\sigma^2 = \sum_{t=-\infty}^{+\infty} \text{Cov}(g(X_0), g(X_t))$. For the system of Legendre polynomials $g(x) = x$, and we get the test statistic N_k in (4). For the frequently used cosine system, given by $\varphi_j(x) = \sqrt{2} \cos(j\pi x)$, we have $g(x) = \sqrt{2} \cos(\pi x)$ and $\text{Var}_0g(X) = 1$. Note that the α -mixing property will not be affected by the choice of the orthogonal system because under measurable transformations the α -mixing property will be maintained with the dependence coefficients $\alpha(n)$ being no greater than the corresponding ones for $(X_t)_{t \in \mathbb{Z}}$ (see Bradley 1986, p. 170).

Finally, to construct an omnibus test statistic we have to choose a model among those corresponding to $k = 1, \dots, d(n)$ for some $d(n) \rightarrow \infty$ as $n \rightarrow \infty$. There is a vast literature on the proper choice of the dimension k in the independent case (see, e.g. Kallenberg 2002). Ledwina (1994) introduced a consistent data-driven procedure using Schwarz (1978) selection rule based on an expansion of the maximum likelihood function. For our purposes we will use a modification of Schwarz’s rule introduced by Kallenberg and Ledwina (1997a) and analyzed also by Kallenberg and Ledwina (1997b), Inglot et al. (1997) and Janic-Wroblewska and Ledwina (2000). First, the modified Schwarz’s selection rule is applied to select a suitable dimension k , i.e.

$$S_{\text{mod}} = \min\{k : 1 \leq k \leq d(n), R_k - k \log n \geq R_j - j \log n, j = 1, \dots, d(n)\}, \tag{8}$$

where $d(n)$ denotes an upper bound of the dimension k which may tend to infinity as $n \rightarrow \infty$. Then the test $N_{S_{\text{mod}}}$ is performed.

Remark 3 In Sect. 7.1, Inglot and Ledwina (1996) have shown the asymptotic equivalence of Schwarz’s rule based on the penalized maximum likelihood function and S_{mod} . In the case of weak dependence such as a mixing property, it can be expected that the asymptotic behavior of S_{mod} will be the same as in the independent case. The behavior of sums of random variables from mixing sequences, in general, is similar to the independent case (central limit type theorems, large deviation results can be proved using the “blocking technique”, which basically allows to transfer the respective results from the independent case (see, e.g. Bosq 1989)).

Consider now the following selection rule replacing the Neyman statistic R_k with N_k in (8)

$$S_{\text{mod}2} = \min\{k : 1 \leq k \leq d(n), N_k - k \log n \geq N_j - j \log n, j = 1, \dots, d(n)\}. \tag{9}$$

Evidently in the case of positive dependence ($\sigma^2 > 1/12$), $S_{\text{mod}2}$ will have a bigger concentration on the dimension 1 under H_0 (see Theorem 1). For negative

dependence the converse is true. Therefore, the use of $S_{\text{mod}2}$, taking into account dependence, is recommended if prior knowledge of positive correlation is available (see Tables 1, 2).

Table 1 5% Empirical critical values of $T_{S_{\text{mod}}}$, $T_{S_{\text{mod}2}}$ and T_k (with fixed $k = d(n)$) of K_1 and K_3 models with $q = 3$ in (13), and empirical counts of S_{mod} and $S_{\text{mod}2}$ under uniformity and $d(n) = 10, n = 50$, based on 10,000 samples in each case

Critical values	$d(n)$									
	1	2	3	4	5	6	7	8	9	10
$T_{S_{\text{mod}}}, K_1$	3.88	4.60	4.99	5.06	5.14	5.20	5.20	5.20	5.20	5.21
$T_{S_{\text{mod}2}}, K_1$	3.88	4.22	4.26	4.26	4.26	4.26	4.26	4.26	4.26	4.26
T_k, K_1	3.88	4.96	5.96	6.81	7.63	8.50	9.25	10.05	10.81	11.69
$T_{S_{\text{mod}}}, K_3$	3.77	4.32	4.62	4.92	5.12	5.23	5.31	5.39	5.39	5.40
$T_{S_{\text{mod}2}}, K_3$	3.77	3.88	3.89	3.89	3.89	3.89	3.89	3.89	3.89	3.89
T_k, K_3	3.77	4.51	5.14	5.61	6.07	6.60	7.12	7.55	7.93	8.33
Counts	s									
	1	2	3	4	5	6	7	8	9	10
$S_{\text{mod}} = s, K_1$	9260	528	127	31	28	12	6	2	4	2
$S_{\text{mod}2} = s, K_1$	9862	125	11	1	0	0	1	0	0	0
$S_{\text{mod}} = s, K_3$	8144	1244	325	105	61	31	30	18	11	31
$S_{\text{mod}2} = s, K_3$	9953	41	5	1	0	0	0	0	0	0

Table 2 5% Empirical critical values of $T_{S_{\text{mod}}}$, $T_{S_{\text{mod}2}}$ and T_k (with fixed $k = d(n)$) of K_2 and K_4 models with $q = 3$ in (13), and empirical counts of S_{mod} and $S_{\text{mod}2}$ under uniformity and $d(n) = 10, n = 50$, based on 10,000 samples in each case

Critical values	$d(n)$									
	1	2	3	4	5	6	7	8	9	10
$T_{S_{\text{mod}}}, K_2$	3.92	8.98	9.69	9.86	9.92	9.93	9.93	9.93	9.94	9.94
$T_{S_{\text{mod}2}}, K_2$	3.92	9.42	12.13	15.01	17.60	19.14	20.50	21.59	22.35	22.80
T_k, K_2	3.92	9.55	12.75	16.21	19.47	21.96	24.71	27.51	30.26	32.87
$T_{S_{\text{mod}}}, K_4$	3.94	28.13	29.95	31.53	31.78	32.07	32.10	32.29	32.29	32.31
$T_{S_{\text{mod}2}}, K_4$	3.94	28.13	32.14	38.66	42.69	49.99	54.36	59.83	65.12	70.56
T_k, K_4	3.94	28.13	32.66	39.61	44.72	52.06	56.70	61.92	67.34	73.27
Counts	s									
	1	2	3	4	5	6	7	8	9	10
$S_{\text{mod}} = s, K_2$	9269	582	89	38	14	4	0	1	3	0
$S_{\text{mod}2} = s, K_2$	7252	1319	504	323	206	127	86	83	56	44
$S_{\text{mod}} = s, K_4$	8415	1374	79	76	7	20	1	13	3	12
$S_{\text{mod}2} = s, K_4$	2568	1711	615	813	578	636	561	705	746	1067

3 Consistency

For a strictly stationary α -mixing process $(X_t)_{t \in \mathbb{Z}}$ assume the following

- (A) $\alpha(n) \leq a\rho^n$, for some $a > 0$, $0 < \rho < 1$.
- (B) $E|X_t|^\gamma < +\infty$ for some $\gamma > 2$.
- (C) $\sigma^2 = \sum_{t=-\infty}^{+\infty} \text{Cov}(X_0, X_t) > 0$.
- (D) $d(n) = o(\log n / \log \log n)$.

We will show that the test statistics $N_{S_{\text{mod}}}$ and $N_{S_{\text{mod}2}}$ are consistent against essentially all alternatives of interest. The idea is that under the null hypothesis $N_{S_{\text{mod}}}$ has a limiting χ_1^2 law, while under alternatives $N_{S_{\text{mod}}} \rightarrow \infty$. The proofs are deferred to the appendix.

Let P_0 denote that the marginals of $(X_t)_{t \in \mathbb{Z}}$ are uniformly distributed on $[0, 1]$.

Theorem 1 *For a strictly stationary α -mixing process $(X_t)_{t \in \mathbb{Z}}$ assume (A) and (D).*

a) *Then*

$$\lim_{n \rightarrow \infty} P_0(S_{\text{mod}} = 1) = 1.$$

b) *If we assume furthermore that (B) and (C) hold, then under H_0 , as $n \rightarrow \infty$,*

$$N_{S_{\text{mod}}} \rightarrow_d \chi_1^2.$$

Corollary 2 *Assume $\hat{\sigma}^2$ is a consistent estimator of $\sigma^2 = \sum_{t=-\infty}^{+\infty} \text{Cov}(X_0, X_t)$ and further assume (A)–(D). Then*

$$(12\hat{\sigma}^2)^{-1} R_{S_{\text{mod}}} \rightarrow_d \chi_1^2.$$

Corollary 3 *Under the assumptions (A)–(D) Theorem 1 and Corollary 2 hold also for the selection rule $S_{\text{mod}2}$ and the test statistic $N_{S_{\text{mod}2}}$.*

Now let us investigate the asymptotical behavior of $N_{S_{\text{mod}}}$ under alternatives. Let X_t have a marginal distribution $P \neq P_0$ on $[0, 1]$. Suppose that

$$E_P \phi_1(X) = \dots = E_P \phi_{K-1}(X) = 0, \quad E_P \phi_K(X) \neq 0 \tag{10}$$

for some $K = K(P) \geq 2$. Consistency of $N_{S_{\text{mod}}}$ will be proved for any alternative of the form (10). It will be assumed that $\lim_{n \rightarrow \infty} \inf d(n) \geq K$, which is certainly the case if $\lim_{n \rightarrow \infty} d(n) = \infty$, since K is fixed (see Assumption D).

Theorem 4 *Let $(X_t)_{t \in \mathbb{Z}}$ be a strictly stationary α -mixing process. Assume that (A) holds. Then under any alternative $P \neq P_0$ with K defined in (10), as $n \rightarrow \infty$*

$$\lim_{n \rightarrow \infty} P(S_{\text{mod}} \geq K) = 1 \quad \text{and} \quad N_{S_{\text{mod}}} \xrightarrow{P} \infty.$$

The following result follows immediately from Theorem 4.

Corollary 5 *Under the assumption (A) Theorem 4 holds for the selection rule $S_{\text{mod}2}$ and the test statistic $N_{S_{\text{mod}2}}$.*

From Theorems 1 and 4 the consistency of $N_{S_{\text{mod}}}$ follows. Note that similar result for fixed k has been shown by Ignaccolo (2004).

4 Simulation study

First, we stress that only prior knowledge of the α -mixing property is needed to perform the test $N_{S_{\text{mod}}}$ and $N_{S_{\text{mod}2}}$ introduced earlier. In this section we will deal with data generated from some ARMA processes. In order to construct a sequence of α -mixing r.v.s having marginal $U[0, 1]$ distribution under H_0 we proceed as follows:

1. First simulate X_1, \dots, X_n from a stationary AR(1) process $\{X_t\}_{t \in \mathbb{Z}}$ defined as

$$X_t - \theta X_{t-1} = Z_t, \quad (11)$$

where $\{Z_t\}_{t \in \mathbb{Z}}$ is an innovation process which is weakly stationary with mean 0 and autocovariance $E(Z_t Z_{t+h}) = \sigma_Z^2 < \infty$ if $h = 0$ and 0 otherwise, and $|\theta| \leq 1$ is the coefficient of the process. Note that for our simulation study dealing with a simple hypothesis we do not estimate the coefficient θ , instead fixing it in advance.

2. Then generate data from (11) with innovations $Z_t \sim N(0, 1 - \theta^2)$. The generated process $\{X_t\}_{t \in \mathbb{Z}}$ will have marginal $N(0, 1)$ distribution.
3. Finally, transform obtained data with Φ , where Φ denotes the $N(0, 1)$ cumulative distribution function. We obtain a sample $X'_1, \dots, X'_n = \Phi(X_1), \dots, \Phi(X_n)$ which is α -mixing and has $U[0, 1]$ marginal distributions.

All these steps have been implemented with statistical software package **R**. In general, it is not easy to simulate a stationary sequence with a specified marginal distribution and autocorrelation structure (for some methods see, e.g. [Song and Hsiao \(1993\)](#) and references therein). Note that for this method the specific AR(1) covariance structure is not relevant. For our simulation study rather it is important that the α -mixing property is maintained (see also Remark 2).

To be precise we will test the following null hypothesis $H_0 : X_1, \dots, X_n$ have the marginal $N(0, 1)$ distribution. This testing problem has another equivalent interpretation, i.e. we test $H_0 : X'_1, \dots, X'_n$ have the marginal $U[0, 1]$ distribution.

4.1 Critical values

The null distribution of the selection rules S_{mod} and $S_{\text{mod}2}$ has been simulated for testing the hypothesis described above. In the following, we summarize briefly the results from a more extensive simulation study (for samples sizes $n = 50, 100$ and $n = 200$) performed by the authors. Here, we analyze the finite-sample behavior with $n = 50$ and $d(n) = 10$. Moreover, ϕ_j s have been chosen as the orthonormal Legendre polynomials on $[0, 1]$. To compare our results with the case of independent data and to highlight the additional difficulties encountered with dependent data, simulations have been performed similarly as described by [Ledwina \(1994\)](#), [Kallenberg and Ledwina \(1995a\)](#) in the independent case. In the following Monte Carlo simulation study, we performed 10,000 replications in each simulation scenario.

We obtain a test for uniformity using the following test statistic

$$T_k = (12\hat{\sigma}^2)^{-1} \sum_{j=1}^k \left\{ n^{-1/2} \sum_{i=1}^n \phi_j[\Phi(X_i)] \right\}^2,$$

where $\hat{\sigma}^2$ is an estimator of $\sigma^2 = \sum_{t=-\infty}^{+\infty} \text{Cov}(\Phi(X_0), \Phi(X_t))$.

An estimator for σ^2 in case of an α -mixing process can be found in [Ignaccolo \(2004\)](#). Basically it involves the estimation of autocovariance structure normalized by some weight function. In case of ARMA processes we require estimation of the autocovariance function $\gamma(h) := \text{Cov}(X_{t+h}, X_t)$ for all $t, h \in \mathbb{Z}$. For stationary processes and, in particular, for ARMA processes a common estimate for $\gamma(h)$ is given by

$$\hat{\gamma}(h) = (n - h)^{-1} \sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X}) \quad \text{for } 0 \leq h \leq n - 1 \quad (12)$$

(see [Brockwell and Davis 1991](#), Chapter 7) and this can be used for an estimator of σ^2 in the following way. We estimate

$$\hat{\sigma}^2 = \hat{\gamma}(0) + 2 \sum_{j=1}^q \hat{\gamma}(j), \quad (13)$$

where q denotes the lag of the last autocovariance $\gamma(q)$, which has to be estimated. For simulated data X'_1, \dots, X'_n autocovariances decay exponentially to zero. Thus, we suggest to truncate simply the autocovariances rounding them to three decimal places and estimate σ^2 as in (13) (for similar recommendation see also [Hannan 1970](#)). In practice, this estimator can also be used for general α -mixing stationary processes with rapidly (exponentially) decreasing autocovariance structures.

In the following, we distinguish the case of negative and positive dependence in order to highlight the difference between these two cases. We analyze α -mixing processes generated from AR(1) processes (defined in (11)) with coefficients $\theta = 0.3, -0.3, 0.6$ and $\theta = -0.6$ (further on we denote these models as K_1, K_2, K_3 and K_4) as described above. Here, coefficients ± 0.3 and ± 0.6 have been chosen to represent moderate and rather strong positive (negative) dependence, respectively. More general ARMA processes can be treated similarly. However, it is easier to classify the positive or negative dependence in the case of AR(1) processes.

Positive dependence. Table 1 shows the simulated critical values for the models K_1, K_3 , respectively. The nominal level has been always chosen to be $\alpha = 0.05$ and σ^2 has been estimated as described in (13). For comparison, we also show the simulated critical values of T_k for fixed $k = 1, \dots, d(n)$.

The critical values are expected to be close to the upper χ^2_1 -quantile, being 3.84 for $\alpha = 0.05$. However, similarly as for the independent case (see, e.g. [Kallenberg and Ledwina 1995a,b](#)) for finite sample sizes, higher dimensions or more components

have also been chosen under H_0 . Moreover, we have an additional effect due to the ratio λ_i/σ_{11} in (7), which will be discussed in Sect. 4.3.

Negative dependence. Here, we analyze the models K_2 and K_4 for sample sizes $n = 50$, $d(n) = 10$ and nominal level $\alpha = 0.05$ (see Table 2).

For negatively correlated data the critical values are much higher than for the positive dependence case. σ_{11} gets small causing the ratios λ_i/σ_{11} to be much bigger than one in (7) (see Sect. 4.3 for more details). Clearly, S_{mod} is preferable as we already noted in Remark 3. We conclude that for moderate sample sizes in the case of strong negative dependence the test does not work well, still it can be used in the case of slight negative dependence.

4.2 Simulated power

Ledwina (1994) and Kallenberg and Ledwina (1995a) have investigated a broad range of alternatives with different patterns of the density. Following these authors we consider alternatives of the form

$$g_1(x) = c(\theta) \exp \left\{ \sum_{j=1}^k \theta_j \phi_j(x) \right\}, \tag{14}$$

where $\theta = (\theta_1, \dots, \theta_k)^T \in \mathbb{R}^k$ is a parameter vector and θ^T denotes the transpose of a vector θ . Here, $c(\theta)$ is a normalizing constant, such that g_1 integrates to one. We also analyze alternatives of the form

$$g_2(x) = 1 + \rho \cos(j\pi x). \tag{15}$$

As we have seen in Sect. 4.1 the test does not work well in the negative dependence case. Thus, analyzing power we consider only the case of positive dependence. We generate data from alternatives (14) and (15) with the dependence structure of model K_1 . In the first Monte Carlo study, we simulate an AR(1) process with $N(0, 1)$ marginals. Transforming the data from K_1 via Φ we obtain an α -mixing process with $U[0, 1]$ marginals. Finally, we apply the transformation by the inverse cumulative distribution function obtaining the data from the densities (14) and (15). We consider the case $n = 50$ and $d(n) = 10$ only.

For power analysis we use different simulated critical values shown in Table 3. The following notation will be used for critical values (when $n = 50$): Cr_1 , independent case with $R_{S_{\text{mod}}}$; Cr_2 , K_1 model with S_{mod} ; Cr_3 , K_1 model with $S_{\text{mod}2}$.

Throughout for power study of $N_{S_{\text{mod}}}$ and $N_{S_{\text{mod}2}}$ we use critical values Cr_2 and Cr_3 , respectively. Power results for g_1 are shown in Table 4. In order to be comparable to the i.i.d. case, additionally we have simulated power for g_1 alternative in the independent case. For $R_{S_{\text{mod}}}$ we use the critical values Cr_1 . Clearly in all cases, $S_{\text{mod}2}$ concentrates more on the dimension 1 resulting in better test performance. In view of Theorem 4 we could expect the power of $N_{S_{\text{mod}2}}$ to be worse in this case. However, when θ is one- or two-dimensional parameter, $S_{\text{mod}2}$ is good enough to detect the alternatives under

Table 3 5% Empirical critical values when $n = 50$ based on 10, 000 samples in each case

Critical values	$d(n)$									
	1	2	3	4	5	6	7	8	9	10
Cr_1	3.79	5.41	5.84	6.07	6.14	6.14	6.14	6.14	6.14	6.14
Cr_2	3.88	4.60	4.99	5.06	5.14	5.20	5.20	5.20	5.20	5.21
Cr_3	3.88	4.22	4.26	4.26	4.26	4.26	4.26	4.26	4.26	4.26

Table 4 Estimated power (%) for g_1 alternative defined in (14) of $N_{S_{mod}}, N_{S_{mod2}}$ and N_k (with fixed $k = d(n)$) based on K_1 model and empirical counts of S_{mod} and S_{mod2} when $d(n) = 10, n = 50$, based on 10, 000 samples in each case; $\theta_1 = 0.3, \theta_2 = (0.25, -0.35)$ and $\theta_3 = (0, 0, 0.4)$ in (15)

Power	$d(n)$									
	1	2	3	4	5	6	7	8	9	10
$N_{S_{mod2}}, \theta_1$	36.31	34.20	33.94	33.95	33.95	33.95	33.95	33.95	33.95	33.95
$N_{S_{mod}}, \theta_1$	36.31	32.44	29.99	29.77	29.39	29.02	29.03	29.05	29.08	28.99
$R_{S_{mod}}, \theta_1$	55.88	42.61	39.90	38.38	38.01	38.02	38.01	38.01	38.02	38.02
$N_{S_{mod2}}, \theta_2$	20.35	56.09	56.14	56.16	56.16	56.17	56.18	56.18	56.18	56.18
$N_{S_{mod}}, \theta_2$	20.35	57.81	54.63	54.18	53.42	52.82	52.85	52.85	52.85	52.73
$R_{S_{mod}}, \theta_2$	21.16	55.21	52.89	51.29	50.68	50.67	50.66	50.66	50.66	50.66
$N_{S_{mod2}}, \theta_3$	4.90	5.61	23.72	23.91	23.95	23.95	23.95	23.95	23.95	23.95
$N_{S_{mod}}, \theta_3$	4.90	6.55	53.53	53.70	53.22	52.72	52.74	52.75	52.76	52.62
$R_{S_{mod}}, \theta_3$	5.42	6.55	57.09	57.36	57.50	57.59	57.61	57.62	57.62	57.62

For estimation of $\hat{\sigma}^2$ we use $q = 3$ in (13)

consideration. Therefore, $N_{S_{mod2}}$ gives better power performance than $R_{S_{mod}}$ for low dimensional alternatives. For higher dimensional alternatives S_{mod} gives much better power performance.

Finally, we have analyzed power for the alternatives g_2 defined in (15) (see Table 5). We conclude that the selection rule S_{mod} and the test behave similarly as in the independent case. However, power is much worse, if we use S_{mod2} and $N_{S_{mod2}}$.

Certainly, from this Monte Carlo study the performance of Neyman test for the general α -mixing case cannot be fully concluded. However, from this study we conclude that in the case of positive dependence the test works well, but cannot be recommended in the negative dependence case.

4.3 Comments on the simulation study

According to [Kallenberg and Ledwina \(1995b\)](#) in the independent case, the first-order result $R_{S_{mod}} \rightarrow_d \chi_1^2$ is often not sharp enough to obtain accurate approximations of the null distribution for small to moderate sample sizes. For example, when $n = 50, d(n) = 10$ and $\alpha = 0.05$ the simulated critical value is equal to 6.14 (see Table 3)

Table 5 Estimated power (%) for g_2 alternative defined in (15) of $N_{S_{\text{mod}}}$, $N_{S_{\text{mod}2}}$ and N_k (with fixed $k = d(n)$) based on K_1 model when $n = 50$ and $d(n) = 10$ based on 10, 000 samples in each case

$\{\rho, j\}$		$d(n)$									
		1	2	3	4	5	6	7	8	9	10
{0.4;1}	$N_{S_{\text{mod}2}}$	34.19	32.02	31.82	31.82	31.82	31.82	31.82	31.82	31.82	31.82
	$N_{S_{\text{mod}}}$	34.19	30.40	28.26	28.02	27.48	27.09	27.11	27.13	27.13	27.03
	$R_{S_{\text{mod}}}$	52.19	39.16	36.08	34.78	34.26	34.23	34.22	34.23	34.23	34.23
{0.5;2}	$N_{S_{\text{mod}2}}$	4.93	30.41	30.34	30.35	30.35	30.35	30.35	30.35	30.35	30.35
	$N_{S_{\text{mod}}}$	4.93	33.80	32.04	33.84	33.57	33.14	33.17	33.18	33.19	33.17
	$R_{S_{\text{mod}}}$	8.41	60.44	58.65	58.47	58.10	58.13	58.15	58.16	58.16	58.16
{0.7;4}	$N_{S_{\text{mod}2}}$	4.86	9.62	9.63	14.83	14.89	14.92	14.93	14.93	14.94	14.94
	$N_{S_{\text{mod}}}$	4.86	11.04	11.45	45.07	45.68	52.20	52.47	52.53	52.57	52.56
	$R_{S_{\text{mod}}}$	6.28	15.53	15.26	45.71	46.32	52.40	52.60	52.72	52.73	52.73

For estimation of $\hat{\sigma}^2$ we use $q = 3$ in (13)

instead of the asymptotically correct value 3.84, i.e. the 0.95-quantile of the χ_1^2 distribution. In the independent case a second-order approximation is suggested to solve this problem (see [Kallenberg and Ledwina 1995b](#)); however, in the dependent case this becomes more difficult, because the underlying dependence structure will be pertinent.

Further, a valid analysis of data with serial dependency often requires larger data sets as for the independent case, and hence small sample corrections are often not of major interest. Second, the simulated critical values of $N_{S_{\text{mod}}}$ and $N_{S_{\text{mod}2}}$ will depend on the particular dependency structure of the α -mixing process under consideration. Hence, a second-order correction is not as easy applicable as in the independent case, because it requires knowledge of the specific dependency. We believe that a bootstrap modification (using, e.g. a block bootstrap) could yield improvements; however, we did not pursue this issue further. Moreover, we found that for the case of positive dependence the critical values are much closer to 3.84 (when $n = 50$, $d(n) = 10$ and $\alpha = 0.05$ the simulated critical value is equal to 4.26 for the K_1 model and 3.89 for the K_3 model, see Table 1). Note, that this approximation is more precise as in the independent case.

To our knowledge negatively correlated data in the context of goodness-of-fit tests have not been analyzed in the literature so far. As mentioned above in the case of negative dependency, the test $N_{S_{\text{mod}}}$ works bad (see Table 2). To find a good finite sample approximation for the case of negatively correlated observations is very challenging and we cannot recommend to apply our method in this situation.

Although under the null hypothesis, S_{mod} and $S_{\text{mod}2}$ select the dimension 1 asymptotically with probability one for moderate sample sizes; the dimension 2 is also selected quite frequently (sometimes even more than 5% in our simulation study, see Tables 1, 2). For an explanation of this fact in the independent case see Theorem 3.1 and Remark 3.1 in [Kallenberg and Ledwina \(1995b\)](#).

Let us therefore investigate more detailed the situation when the selection rules S_{mod} or $S_{\text{mod}2}$ select the dimension 2 under H_0 . In this case we have $k = 2$ and under H_0

$$N_2 \rightarrow_d \left(\frac{\lambda_1}{\sigma_{11}}\right) Y_1 + \left(\frac{\lambda_2}{\sigma_{11}}\right) Y_2, \tag{16}$$

where $Y_1, Y_2 \sim \chi_1^2$ are independent r.v.s, λ_1 and λ_2 are the eigenvalues of the matrix with elements from (3) and $\sigma_{11} = 12\sigma^2 = 12 \sum_{t=-\infty}^{+\infty} \text{Cov}(X_0, X_t)$. Now the difference from the independent case is that we have additional multipliers λ_1/σ_{11} and λ_2/σ_{11} (which are equal to one in the independent case) which behavior we will analyze now for Legendre polynomials.

First note that for positive definite, symmetric matrix defined through its elements in (3) λ_1 and λ_2 are positive numbers equal to

$$\lambda_{1,2} = \frac{(\sigma_{11} + \sigma_{22}) \pm \sqrt{(\sigma_{11} + \sigma_{22})^2 - 4(\sigma_{11}\sigma_{22} - \sigma_{12}^2)}}{2}, \tag{17}$$

where $\sigma_{12} = \sigma_{21}$. Now, let us analyze the behavior of $\lambda_{\text{max}}/\sigma_{11}$, where λ_{max} denotes the biggest eigenvalue of (17). Obviously an upper bound is given by

$$\frac{\lambda_{\text{max}}}{\sigma_{11}} \leq 1 + \frac{\sigma_{22}}{\sigma_{11}}. \tag{18}$$

Moreover, if $\sigma_{12} \ll \sigma_{11}$ then

$$\frac{\lambda_{\text{max}}}{\sigma_{11}} \approx \frac{\sigma_{11} + \sigma_{22}}{2\sigma_{11}} + \sqrt{\frac{(\sigma_{11} - \sigma_{22})^2}{4\sigma_{11}^2} + \left(\frac{\sigma_{12}}{\sigma_{11}}\right)^2} \approx \frac{\sigma_{11} + \sigma_{22} + |\sigma_{11} - \sigma_{22}|}{2\sigma_{11}}. \tag{19}$$

Now it follows that $\begin{cases} \frac{\lambda_{\text{max}}}{\sigma_{11}} \approx 1, & \text{if } \sigma_{11} \geq \sigma_{22} \\ \frac{\lambda_{\text{max}}}{\sigma_{11}} \approx \frac{\sigma_{22}}{\sigma_{11}} & \text{if } \sigma_{11} < \sigma_{22} \end{cases}$.

Either we make a rough upper bound as in (18) or we have $\sigma_{12} \ll \sigma_{11}$ and thus (19), it is clear that the ratio σ_{22}/σ_{11} should be analyzed.

Consider now

$$\sigma_{11} = E_0(\phi_1(X_1)^2) + \sum_{i=2}^{\infty} E_0(\phi_1(X_1)\phi_1(X_i)) + \sum_{i=2}^{\infty} E_0(\phi_1(X_i)\phi_1(X_1)), \tag{20}$$

$$\sigma_{22} = E_0(\phi_2(X_1)^2) + \sum_{i=2}^{\infty} E_0(\phi_2(X_1)\phi_2(X_i)) + \sum_{i=2}^{\infty} E_0(\phi_2(X_i)\phi_2(X_1)), \tag{21}$$

where $\phi_1(y) = \sqrt{12}(y - 0.5)$ and $\phi_2(y) = \sqrt{5}(6(y - 0.5)^2 - 0.5)$ are the first two orthonormal Legendre polynomials. Here, E_0 is taken with respect to $U[0, 1]$ distribution. First note that $E_0(\phi_1(X_1)^2) = E_0(\phi_2(X_1)^2) = 1$.

Denote each term in (20) and (21) with $A1_{1i} = E_0(\phi_1(X_1)\phi_1(X_i))$ and $A2_{1i} = E_0(\phi_2(X_1)\phi_2(X_i))$ for $i = 1, \dots, n$. Thus, $A1_{1i}$ denotes the covariance between $\phi_1(X_1)$ and $\phi_1(X_i)$.

Table 6 Empirical characteristics of K_3, K_4 models, $n = 500, 000$

	h						
	1	2	3	4	5	6	7
Cov_h, K_3	0.0485	0.0289	0.0173	0.0105	0.0063	0.0038	0.0022
$Cond1, K_3$	-0.0016	-0.0013	-0.0009	-0.0006	-0.0004	-0.0002	-0.0001
Cov_h, K_4	-0.0486	0.0289	-0.0172	0.0104	-0.0064	0.0038	-0.0022
$Cond1, K_4$	0.0048	-0.0014	0.0013	-0.0008	0.0004	-0.0003	0.0001

Under “*Cond1*” we mean the left hand side of (22), i.e. $E_0(X_1^2 X_i^2) + 14/15 E_0(X_1 X_i) - 2E_0(X_1^2 X_i) - 1/90$. Cov_h denotes the covariance between X_i and X_{i+h} for the lag h

Lemma 6 Let X_1, \dots, X_n each have $U[0, 1]$ marginal distribution. If and only if

$$E_0(X_1^2 X_i^2) + \frac{14}{15} E_0(X_1 X_i) - 2E_0(X_1^2 X_i) - \frac{1}{90} \leq 0 \tag{22}$$

then $A2_{1i} \leq A1_{1i}$.

In Table 6, we show some covariances between X_i and X_{i+h} for the lag $h = 1, \dots, 7$ and also the value of the left side of the inequality (22) for models K_3 and K_4 considered earlier in our simulations (see Sect. 4.1).

For the model K_3 , which represents the case of positive dependence, the condition (22) is satisfied for all lags h considered in Table 6, therefore $A2_{1i} \leq A1_{1i}$ for every i and also $\sigma_{22} \leq \sigma_{11}$. Thus, both $\lambda_1/\sigma_{11} \leq 1$ and $\lambda_2/\sigma_{11} \leq 1$ in (16) and the critical values can only be smaller than those obtained in the independent case by [Kallenberg and Ledwina \(1995b\)](#).

In the case of negative dependence (the model K_4 in Table 6) the situation is more complicated because the negative covariances and the positive ones alternate subsequently for different lags h . However, we observe that in this case $\lambda_i/\sigma_{11} \geq 1$ in (16).

To summarize: the ratios λ_i/σ_{11} in (16) play an important role in the case of dependent observations (in the independent case they are equal to one). In the case of positive dependency they improve the accuracy of the first-order approximation of the null distribution. In the case of negative dependence we have an opposite effect. When $\sigma_{11} = 12\sigma^2$ becomes very small (in the case of very strong negative dependence), the estimation of σ^2 is unstable and therefore problematic. We again stress that we do not recommend to use the test in that case.

5 Data example

Finally, we will discuss an example drawn from the management of market risk in investment banks, which motivated our work to a large extent. One of the central tasks of a risk manager is to generate a forecast of the market parameters (interest rates, stock prices, etc.) that determine the market value of a trading portfolio. Once the parameter distribution for the next business day has been estimated, the resulting distribution of the portfolio value can be calculated, which is the basis for quantile-

based risk measures such as “value-at-risk”. Artzner et al. (1999) discuss the concept of coherent risk measures, in general.

One of the most popular approaches for forecasting market parameters is the so-called historical simulation method (see Duffie and Pan 1997 for a review of early literature). The key idea of this approach is to sample potential parameter values (“scenarios”) directly from the historically observed realizations. (Banking supervisors require a time series length of one year, i.e. 254 business days, for this purpose). While this approach makes no assumptions regarding the parameter distribution, it requires identifying i.i.d. r.v. as the stochastic drivers of the time series. For interest rates and stock prices it is common to consider the observed relative or absolute parameter changes as the realizations of i.i.d. variables. The observed changes are then applied to the current parameter value in order to generate a vector of forecasts for the next business day. Pritsker (2001) provides a detailed discussion of this and related historical simulation approaches.

The market parameters analyzed in the following are time series of implied volatilities, which have been derived from exchange traded options (see Hull 2002 for an excellent introduction to the topic). Our data sample consists of four different time series of implied volatilities Y_t over a period of several years: EUR and USD swaption volatilities with option maturity one year as well as EUR/USD and EUR/GBP at-the-money foreign exchange volatilities for maturities of one year as quoted by Reuters. For these parameters, the nature of the stochastic drivers is less evident and we employ the following non-parametric autoregressive model (see, e.g. Neumann and Kreiss 1998) to describe the relative change in implied volatility, $X_t = Y_t/Y_{t-1} - 1$, at time t

$$X_t = g(X_{t-1}) + \sigma(X_{t-1})\epsilon_t. \tag{23}$$

Here, the ϵ_t represent a sequence of i.i.d. r.v.s such that $E[\epsilon_t] = 0$ and $E[\epsilon_t^2] = 1$. The functions g and σ^2 are unknown (smooth) functions modeling the expectation and variance of the non-parametric time series and have each been estimated by a local linear estimator (see Fan and Gijbels 1996; Fan and Yao 1998). The respective bandwidths have been chosen by cross validation. Obviously, model (23) is a generalization of the simple case where the X_t alone are treated as i.i.d. variables (see Barone-Adesi et al. 1999 for a similar approach).

Under these modeling assumptions the historical simulation method works as follows: For a fixed business day $t^* \geq 254$, we estimate model (23) over the time period $\{t^* - 254, \dots, t^*\}$ and calculate the time series of residuals via

$$\hat{\epsilon}_t = \frac{X_t - \hat{g}_{t^*}(X_{t-1})}{\hat{\sigma}_{t^*}(X_{t-1})}, \quad t = t^* - 254, \dots, t^*$$

(we introduce the index t^* to emphasize that for each business day we perform a new estimation of g and σ). From the residuals we generate a vector of forecasts for the volatility at time $t^* + 1$

$$\hat{Y}_{t^*+1}^{(j)} = (1 + \hat{X}_{t^*+1}^{(j)})Y_{t^*}^{(j)}, \tag{24}$$

where

$$\hat{X}_{t^*+1}^{(j)} = \hat{g}_{t^*}(X_t^*) + \hat{\sigma}_{t^*}(X_t^*)\hat{\epsilon}_{t^*+1-j}, \quad j = 1, \dots, 254.$$

We proceed in this manner for every $t^* \in \{254, \dots, N\}$, where the time series length N varies for the four different data sets investigated.

The standard way of testing whether model (24) generates reliable forecasts is to consider the transformed variables $U_{t^*+1} = \hat{F}_{t^*+1}(Y_{t^*+1})$. Here $\hat{F}_{t^*+1}(y)$ denotes the empirical c.d.f. constituted from the $\hat{Y}_{t^*+1}^{(j)}$ in (24). Under the null hypothesis that the model always generates the correct (conditional) parameter distribution, the (U_{t^*+1}) are independent and follow a uniform distribution (see Rosenblatt 1952). If the forecasts have a systematic tendency to over or underestimate the historical volatilities this will result in deviations from uniformity.

Table 7 p Values of Neyman tests

p Values	Swaption-EUR			Swaption-USD			EUR-GBP		
	q								
	0	1	2	0	1	2	0	1	2
$N_{S_{\text{mod}}}$	0.72	0.70	0.71	0.53	0.59	0.55	0.75	0.74	0.74
$R_{S_{\text{mod}}}$	0.70	0.70	0.70	0.50	0.50	0.50	0.76	0.76	0.76

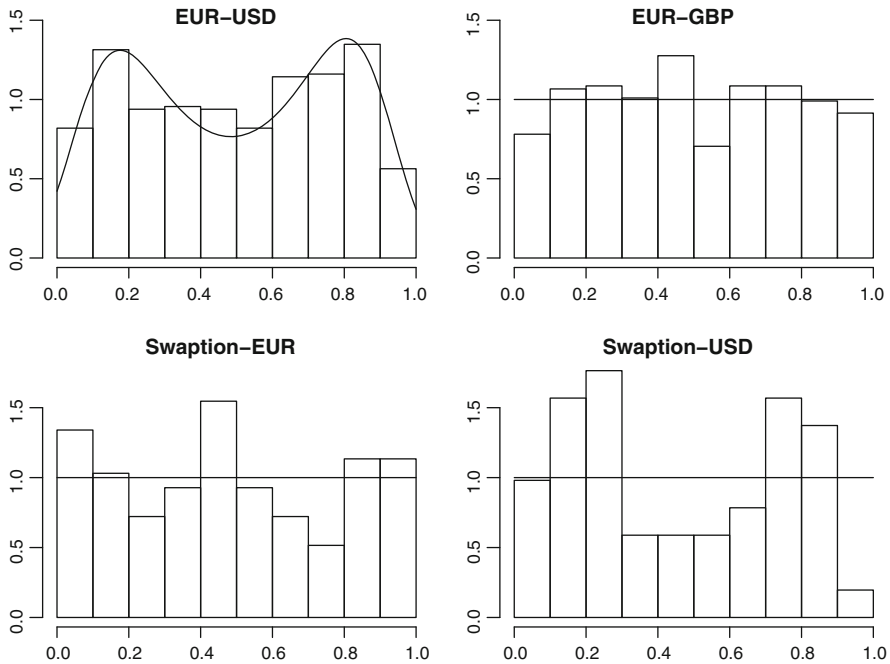


Fig. 1 Histograms and the corresponding projection density estimators for the implied volatility data

Table 8 Values of Neyman tests

Values	EUR-USD		
	q		
	0	1	2
$N_{S_{\text{mod}}}$	52.13	53.83	54.32
$R_{S_{\text{mod}}}$	48.73	48.73	48.73

Therefore, to investigate the validity of the method we use the test statistic $N_{S_{\text{mod}}}$. This will be compared with results for the original Neyman statistic $R_{S_{\text{mod}}}$. Let $n = N - 254$ be the number of transformed observations available. After estimating the auto-covariances for $q = 0, 1, 2$ in (13) we present the corresponding p -values of the test statistic $N_{S_{\text{mod}}}$ in Table 7 for three data sets—Swaption-EUR, Swaption-USD and EUR-GBP with the respective sample sizes $n = 97, 51, 525$. In Fig. 1 we have displayed the histograms of the implied volatilities together with their Legendre polynomial projection density estimators according to the result of the model selector S_{mod} . Note, that for EUR-GBP, Swaption-EUR and Swaption-USD S_{mod} selects the dimension 1 and uniformity is not rejected at a 5% level. In these cases the resulting density estimator is simply the uniform density (the first Legendre polynomial). Table 7 also provides p -values for the Neyman smooth test statistic $R_{S_{\text{mod}}}$ in the independent case. We conclude that these values are very close to those for $N_{S_{\text{mod}}}$. We mention that the time series of the (U_{t^*+1}) was found to be almost uncorrelated for the foreign exchange rates, for the swaptions we observed autocorrelations up to 0.2. As can also be seen from Table 7, the p -values of statistic $N_{S_{\text{mod}}}$ still do not change much for larger values of q in (13).

The fourth data set EUR-USD is different. Table 8 shows some values of $N_{S_{\text{mod}}}$ with $n = 586$ (see Fig. 1). In this case we reject the uniformity at 5% level (all p -values were found to be less than 10^{-4}). In accordance with this, S_{mod} chooses the dimension $d(n) = 4$, which increases drastically the value of $N_{S_{\text{mod}}}$. The corresponding Legendre polynomial projection estimator of order 4 is displayed in Fig. 1.

Finally, it is interesting to stress that the classical Kolmogorov–Smirnov test does not reject uniformity in this case (the p -value is approximately 0.085).

Acknowledgments A. Munk acknowledges support of the DFG, grant FOR 916 and J. Valeinis acknowledges support of the Georg Lichtenberg program “Applied Statistics and Empirical Methods”. We thank R. Ignaccolo and T. Mikosch for pointing out some references and T. Ledwina for helpful comments. The comments of two referees are gratefully acknowledged which led to a much better presentation of the paper.

Appendix

(i) *Proof of Theorem 1* First we introduce some notation. Let

$$R_j = \sum_{k=1}^j \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_k(X_i) \right)^2,$$

$$R_k^* := \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_k(X_i) \right)^2, \text{ thus } R_j = \sum_{k=1}^j R_k^*.$$

In order to prove part (a) of Theorem 1 it is enough to prove that

$$P_0(S_{\text{mod}} \neq 1) = P_0 \left(\max_{2 \leq j \leq d(n)} (R_j - j \log n) > R_1 - \log n \right) \rightarrow 0$$

as $n \rightarrow \infty$. First note that

$$\begin{aligned} & P_0 \left(\max_{2 \leq j \leq d(n)} (R_j - j \log n) > R_1 - \log n \right) \\ &= P_0(\exists j : 2 \leq j \leq d(n) : R_j - R_1 > (j - 1) \log n) \\ &\leq P_0(\exists k : 2 \leq k \leq d(n) : R_k^* > \log n) \\ &\leq \sum_{k=2}^{d(n)} P_0(R_k^* > \log n) \leq \sum_{k=2}^{d(n)} E_0(R_k^*) / (\log n), \end{aligned}$$

where the last inequality follows from the Markov’s inequality. Due to the strict stationarity of $(X_t)_{t \in \mathbb{Z}}$ we have

$$E_0(R_k^*) = \frac{1}{n} \sum_{i,j} E_0(\phi_k(X_i)\phi_k(X_j)) \leq 2 \sum_{i=0}^n |\text{cov}_0(\phi_k(X_0), \phi_k(X_i))|.$$

Now, using Corollary 1.2 of Rio (1993) and his comment on p. 593 on geometric rates of mixing (see also Theorem 1 of Doukhan et al. (1994), in particular, the point 4 of the discussion in p. 68) we get that for each k and for some $C < \infty$

$$\sum_{i=0}^n |\text{cov}_0(\phi_k(X_0), \phi_k(X_i))| \leq C \log k.$$

To this end, note that $E_0(\phi_k^2(X_1)\log^+|\phi_k(X_1)|) \leq C' \log k^{1/2} = 1/2C' \log k$, $C' < \infty$, since for orthonormal Legendre polynomial system $\max_{1 \leq j \leq k} \sup_x |\phi_j(x)| = O(k^{1/2})$ and $E_0(\phi_k^2(X_1)) = 1$.

Finally, we get

$$\sum_{k=2}^{d(n)} E_0(R_k^*) / (\log n) \leq \frac{1}{2} \sum_{k=1}^{d(n)} C \log k / (\log n) = O(d(n) \log d(n) / (\log n))$$

which proves the statement (a) of Theorem 1. Under additional conditions B and C part (b) follows from Theorem 3.1 in Ignaccolo (2004). □

(ii) *Proof of Corollary 3* The constant $12\sigma^2$ in $S_{\text{mod}2}$ does not affect the asymptotic behavior of S_{mod} . The result follows from Theorem 1 and Slutsky’s lemma. □

(iii) *Proof of Theorem 4* Similarly as in [Inglot et al. \(1997\)](#), p. 1233 using ergodic theorem (see, e.g. [Hannan 1970](#), p. 203) for $k \in 1, \dots, K - 1$

$$\frac{R_k}{n} \xrightarrow{P} 0 \quad \text{and} \quad \frac{R_K}{n} \xrightarrow{P} \{\mathbb{E}_P \phi_K(X)\}^2 > 0.$$

It follows that

$$P(S_{\text{mod}} \geq K) \rightarrow 1.$$

To prove that $N_{S_{\text{mod}}} \xrightarrow{P} \infty$ first note that the condition on the mixing coefficients (Assumption (A)) assure that $\sigma^2 < \infty$ by Theorem 1.5 in [Bosq \(1998\)](#), p. 34. Now according to Theorem 3.3 in [Ignaccolo \(2004\)](#) the result follows. \square

Under H_0 we have

$$A1_{1i} = 12E_0(X_1 X_i) - 3 \in [-1; 1],$$

$$A2_{1i} = 180E_0(X_1^2 X_i^2 - X_1^2 X_i - X_1 X_i^2 + X_1 X_i) - 5 \in [-1; 1].$$

Now using elementary calculations the result follows. \square

References

- Abramowitz, M., Stegun I. A. (1972). Legendre functions and orthogonal polynomials. Ch. 22 in Chs. 8 and 22. In *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, 9th printing (pp. 331–339 and 771–802). New York: Dover.
- Artzner, P., Delbaen, F., Eber, J. M., Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9, 203–228.
- Barone-Adesi, G., Giannopoulos, K., Vosper, L. (1999). VaR without correlations for non-linear portfolios. *Journal of Futures Markets*, 19, 583–602.
- Bosq, D. (1989). Test du χ^2 généralisés. Comparaison avec le test du χ^2 classique. *Revue de Statistique Appliquée*, XXXVII, 43–52.
- Bosq, D. (1998). *Nonparametric statistics for stochastic processes* (Vol. 110). Lecture Notes in Statistics. Heidelberg: Springer.
- Bosq, D. (2002). Functional tests of fit. Goodness-of-fit tests and model validity. In *Statistics for industry and technology* (pp. 341–356). Boston: Birkhäuser.
- Bradley, R. C. (1986). Basic properties of strong mixing conditions. In E. Eberlein, M. S. Taqqu (Eds.), *Dependence in probability and statistics*, Progress in probability and statistics (Vol. 11, pp. 165–192). Boston: Birkhäuser.
- Bradley, R. C. (2002). *Introduction to strong mixing conditions*. Technical Report, 1, Bloomington: Indiana University.
- Brockwell, J. P., Davis, A. R. (1991). *Time series: Theory and methods* (2nd ed.). New York: Springer.
- Chanda, K. C. (1981). Chi-square goodness of fit tests based on dependent observations. In C. Taillie, G. P. Patil, B. A. Baldessari (Eds.), *Statistical distributions in scientific work* (Vol. 5, pp. 35–49). Dordrecht: D. Reidel Publishing Company.
- Chanda, K. C. (1999). Chi-squared tests of goodness-of-fit for dependent observations. *Asymptotics, non-parametrics and time series*. Statistics: Textbooks and monographs, 158 (pp. 743–756). New York: Dekker.
- Doukhan, P., Massart, P., Rio, E. (1994). The functional central limit theorem for strongly mixing processes. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 30, 63–82.

- Ducharme, G. R., Micheaux, P. L. (2004). Goodness-of-fit tests of normality for the innovations in ARMA models. *Journal of Time Series Analysis*, 25, 373–395.
- Duffie, D., Pan, J. (1997). An overview of value at risk. *Journal of Derivatives*, 7, 7–49.
- Eaton, M. L., Tyler, D. E. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *The Annals of Statistics*, 19, 260–271.
- Fan, J., Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman and Hall.
- Fan, J., Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85, 645–660.
- Gasser, T. (1975). Goodness-of-fit tests for correlated data. *Biometrika*, 62, 563–570.
- Gleser, L. J., Moore, D. S. (1983). The effect of dependence on chi-squared and empiric distribution tests of fit. *The Annals of Statistics*, 11, 1100–1108.
- Hannan, E. J. (1970). *Multiple time series*. New York: Wiley.
- Hull, J. C. (2002). *Options, futures, and other derivatives* (5th ed.). New York: Pearson Education.
- Ignaccolo, R. (2004). Goodness-of-fit tests for dependent data. *Journal of Nonparametric Statistics*, 16, 19–38.
- Inglot, T., Ledwina, T. (1996) Asymptotic optimality of data driven Neyman's tests for uniformity. *The Annals of Statistics*, 24, 1982–2019.
- Inglot, T., Kallenberg, W. C. M., Ledwina, T. (1997). Data driven smooth tests for composite hypotheses. *The Annals of Statistics*, 25, 1222–1250.
- Janic-Wroblewska, A., Ledwina, T. (2000). Data driven test for two-sample problem. *Scandinavian Journal of Statistics*, 27, 281–298.
- Kallenberg, W. C. M. (2002). The penalty in data driven Neyman's tests. *Mathematical Methods of Statistics*, 11, 323–340.
- Kallenberg, W. C. M., Ledwina, T. (1995a). Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *The Annals of Statistics*, 23, 1594–1608.
- Kallenberg, W. C. M., Ledwina, T. (1995b). On data driven Neyman's tests. *Probability and Mathematical Statistics*, 15, 409–426.
- Kallenberg, W. C. M., Ledwina, T. (1997a). Data driven smooth tests when the hypothesis is composite. *Journal of The American Statistical Association*, 92, 1095–1105.
- Kallenberg, W. C. M., Ledwina, T. (1997b). Data driven smooth tests for composite hypotheses: Comparison of powers. *Journal of Statistical Computation and Simulation*, 59, 101–121.
- Ledwina, T. (1994). Data driven version of Neyman's smooth test of fit. *Journal of the American Statistical Association*, 89, 1000–1005.
- Mokkadem, A. (1990). Propriétés de mélange des processus autorégressifs polynomiaux. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 26, 219–260.
- Moore, D. S. (1982). The effect of dependence on chi-squared tests of fit. *The Annals of Statistics*, 10, 1163–1171.
- Neumann, M. H., Kreiss, J. P. (1998). Regression type inference in nonparametric autoregression. *The Annals of Statistics*, 26, 1570–1613.
- Neumann, M. H., Paparoditis, E. (2000). On bootstrapping L_2 -type statistics in density testing. *Statistics and Probability Letters*, 50, 137–147.
- Neyman, J. (1937). Smooth test for goodness of fit. *Skandinavian Aktuarietidskrift*, 20, 150–199.
- Pritsker, M. (2001). The hidden risks of historical simulation. Federal Reserve Board, Washington. Available at <http://www.gloriamundi.org>.
- Rayner, J., Best, D. (1989). *Smooth tests of goodness-of-fit*. Oxford: Oxford University Press.
- Rio, E. (1993). Covariance inequalities for strongly mixing processes. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 29, 587–597.
- Rosenblatt, D. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23, 470–472.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464.
- Song, W.-M. T., Hsiao, L. C. (1993). Generation of autocorrelated random variables with a specified marginal distribution. In G. W. Evans, M. Mollaghasemi, E. C. Russell, W. E. Biles (Eds.), *Proceedings of the 1993 winter simulation conference* (pp. 374–377).