*The Canadian Journal of Statistics*
Vol. xx, No. yy, 2012, Pages 1–25
*La revue canadienne de statistique*


# Testing for intercept-scale switch in linear autoregression

Florian Ketterer and Hajo Holzmann*

*Philipps-Universität Marburg,*

*Key words and phrases:* switching autoregression; likelihood ratio test; time series

*MSC 2010:* Primary 62F03; secondary 62M10

*Abstract:* Autoregressive models with switching regime are a frequently used class of non-linear time series models, which are popular in finance, engineering, and other fields. We consider linear switching autoregressions in which intercept and variance possibly switch simultaneously, while the autoregressive parameters are structural and hence the same in all states, and propose quasi-likelihood-based tests for a regime switch in this class of models. Our motivation is from financial time series, where one expects states with high volatility and low mean together with states with low volatility and higher mean. We investigate the performance of our tests in a simulation study, and give an application to a series of IBM monthly stock returns. *The Canadian Journal of Statistics* xx: 1–25; 2012 © 2012 Statistical Society of Canada

*Résumé:* Insérer votre résumé ici. We will supply a French abstract for those authors who can't prepare it themselves. *La revue canadienne de statistique* xx: 1–25; 2012 © 2012 Société statistique du Canada

## 1. INTRODUCTION

Autoregressive models with switching regime are a frequently used class of non-linear time series models, which are popular in finance, engineering, and other fields.
Often, only a single parameter is allowed to switch, all other parameters are taken as structural parameters, which are supposed to be the same for all states, see e.g. Hamilton (1989, 2008), or Piger (2009), who apply regime switching models for the identification of business cycle turning points. Compared to a model in which all parameters are allowed to switch, a single switching parameter has two major advantages: First, the states have easier and better interpretations, since they only affect a single parameter like variance or drift. Second, allowing all parameters to switch often leads to too many parameters, and the estimation results become pretty unstable. However, in particular in financial applications to series of stock returns, one might expect periods of high volatility to be accompanied by slightly lower means, whereas periods of low volatility have slightly higher means. Thus, in this context, two parameters should be allowed to switch simultaneously, while other model parameters can be modeled as structural parameters. Therefore, in this paper we suppose that the observed stationary time series $(X_t)_{t\in\mathbb{Z}}$ follows the model

$$X_t = \zeta_{S_t} + \sum_{j=1}^{p} \phi_j X_{t-j} + \sigma_{S_t}\epsilon_t, \tag{1}$$

where $\epsilon_t \overset{iid}{\sim} N(0,1)$, and where $(S_t)_{t\in\mathbb{Z}}$ is a stationary (unobserved) Markov chain on a finite state space $\mathcal{M} = \{1,\dots,m\}$. In case of two states $m = 2$, the parameters of interest are the entries $a_{21}, a_{12}$ of the transition matrix $P = (a_{ij})_{i,j=1,2}$ of the Markov chain, the state-dependent

* *Author to whom correspondence may be addressed.*
*E-mail: holzmann@mathematik.uni-marburg.de*

© 2012 Statistical Society of Canada / Société statistique du Canada

parameters $(\zeta_1, \sigma_1), (\zeta_2, \sigma_2) \in Z \times \Sigma \subset \mathbb{R}^2$ as well as the structural autoregressive parameters $\boldsymbol{\phi} \in \boldsymbol{\Phi} \subset \mathbb{R}^p$ which are the same for all states.

Our main purpose in this paper is to propose a test for the hypothesis $\mathcal{M} = \{1\}$ of no regime switch against the alternative of two states $\mathcal{M} = \{1, 2\}$ in model (1). The asymptotic behaviour of likelihood-based methods is however difficult due to the following reasons:
1. Testing for homogeneity in simple finite mixture models results in loss of identifiability under the null-hypothesis, see e.g. Dacunha-Castelle & Gassiat (1999),
2. For the normal model, additional technical difficulties arise due to loss of "strong identifiability", see Chen & Chen (2003),
3. For normal models, if the variance is allowed to switch as well and is not constrained, the likelihood function becomes unbounded and the Fisher information is infinite for certain parameter combinations (e.g. Chen & Li 2009),
4. Finally, additional difficulties arise if the Markov dependence structure of the regime is incorporated in the test statistic: Even for compact parameter spaces, Gassiat & Keribin (2000) show that the LRT for regime switching may not converge in distribution at all.

Despite these difficulties we can construct tests with a tractable asymptotic distribution by using a quasi likelihood which neglects the dependence structure in the regime under the alternative, and by using the EM-test technique as in Chen & Li (2009) and Li, Chen & Marriot (2009) for i.i.d. mixtures and in Holzmann & Ketterer (2011) for switching autoregressions with single switching parameter.

Switching regime models are used as an alternative to GARCH-type models to capture the conditional heteroscedasticity in financial time series, with subsequent applications of volatility estimation and estimation of risk measures such as the value at risk. Indeed, a Markov-switching of the standard deviation of the residuals (as in our model (1)) already results in conditional heteroscedasticity. Simple hidden Markov models with state-dependent normal distributions for daily log-returns of stocks or stock indices are used by Rydén, Teräsvirta, & Asbrink (1998) and Velucchi (2009). While Rydén, Teräsvirta, & Asbrink (1998) fix the mean at zero, Velucci (2009) allows both mean and standard deviations to switch, and concludes that there are two regimes: one with low returns and high volatility and a second with high returns and low volatility. A slightly different version of model (1) is used in Bhar & Hamori (2004) for modeling monthly returns of stock indices from the G7 countries. They suggest their Markov-switching stock return model as follows,

$$X_t - \mu_{S_t} = \phi(X_{t-1} - \mu_{S_{t-1}}) + \sigma_{S_t}\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0,1). \tag{2}$$

Similarly to Velucchi (2009) they conclude that there are two states, one with high $\sigma$ and small $\mu$, the other with smaller $\sigma$ and high $\mu$. When comparing models, in model (1) a level shift in the mean occurs immediately when changing the state of the underlying Markov chain while the mean level in model (2) approaches the new value gradually over several periods. Without regime switch both models reduce to simple linear autoregressions. Using our test methodology, we investigate whether a similar phenomenon of joint regime switch in mean and variance can also be observed in model (1) when applied to a series of monthly IBM stock returns from January 1926 to December 1999. Finally, we mention that some authors consider regime-switching ARCH and GARCH models, in which the parameters of a GARCH-type model are allowed to switch (see Lange & Rahbeck 2009). However, these models are quite involved, and the actual effect of a regime switch in the conditional heteroscedasticity structure is not particularly transparent.

This paper is organized as follows. In Section 2 we propose our tests for regime switch in model (1). Section 3 contains a simulation section in which we investigate the finite sample properties of the proposed tests. In Section 4, we apply the methodology to the series of IBM stock returns. Sketches of proofs are given in an appendix, complete arguments can be found in the supplementary material Ketterer and Holzmann (2012).

## 2. TESTING FOR SWITCHING INTERCEPT AND VARIANCE

In this section, we develop methods for testing for homogeneity in model (1) using (penalized) likelihood based tests. In our theoretical derivations, we shall assume for the standard deviations that $\Sigma = [\delta, \infty)$ for some $\delta > 0$, which guarantees that the likelihood functions are bounded. Other possibilities proposed in the literature include bounding the ratio of the variances from below (Hathaway 1985), or penalizing deviations of $\sigma$ from homogeneity, in particular smaller values of $\sigma$ (Chen, Tan & Zhang 2008, Chen & Li 2009). We shall explore the latter possibility in the simulation study. To ensure identifiability of the parameters and for the uniqueness of the order $p$, we suppose that under the null model, i.e. no regime switch, $(X_t)_{t\in\mathbb{Z}}$ is a causal AR($p$) process.

### 2.1. Quasi Likelihood

Let $\varphi(\cdot)$ be the density of a standard normal variate and denote by $\varphi(\cdot;\mu,\sigma) = (1/\sigma)\varphi\big((\cdot - \mu)/\sigma\big)$ the corresponding location scale family and $X_k^p = (X_k, \ldots, X_{k-p+1})$. Conditional on $X_{k-1}^p = x_{k-1}^p$ and $S_k = i$, $X_k$ has density

$$\varphi\big(x_k; \mu(\zeta_i, \boldsymbol{\phi}; x_{k-1}^p), \sigma_i\big),$$

where

$$\mu(\zeta_i, \boldsymbol{\phi}; x_{k-1}^p) = E[X_k | X_{k-1}^p = x_{k-1}^p, S_k = i] = \zeta_i + \sum_{j=1}^p \phi_j x_{k-j}.$$

The conditional log-likelihood given the initial observations $X_0^p = (X_0, \ldots, X_{-p+1})$ (we start indexing from $-p+1, -p+2, \ldots$) and the initial unobserved state $S_0 = i_0$ is given by

$$\tilde{l}_n(\boldsymbol{\omega}) = \log\Big(\sum_{i_1=1}^2 \cdots \sum_{i_n=1}^2 \prod_{k=1}^n a_{i_{k-1},i_k}\, \varphi(X_k; \mu(\zeta_{i_k}, \boldsymbol{\phi}; X_{k-1}^p), \sigma_{i_k})\Big) \tag{3}$$

where $\boldsymbol{\omega} = (a_{12}, a_{21}, (\zeta_1, \sigma_1), (\zeta_2, \sigma_2), \boldsymbol{\phi}^T)^T$. We shall use the full model likelihood in Equation (3) for computing the model selection criteria AIC and BIC in our application in Section 4. However, for the reasons discussed in the introduction, we shall base our test not on the full model likelihood in Equation (3), but rather on a quasi likelihood which neglects the dependence of the regime, and which is defined as follows

$$\begin{aligned} &l_n(\alpha, \zeta_1, \zeta_2, \sigma_1, \sigma_2, \boldsymbol{\phi}) \\ =&\sum_{k=1}^n \log\big((1-\alpha)\varphi(X_k; \mu(\zeta_1, \boldsymbol{\phi}; X_{k-1}^p), \sigma_1) + \alpha\varphi(X_k; \mu(\zeta_2, \boldsymbol{\phi}; X_{k-1}^p), \sigma_2)\big), \end{aligned} \tag{4}$$

where $\boldsymbol{\psi} = (\alpha, (\zeta_1, \sigma_1), (\zeta_2, \sigma_2), \boldsymbol{\phi}^T)^T$ and $(1-\alpha, \alpha)^T$ corresponds to the stationary distribution of the hidden Markov chain $(S_t)_{t\in\mathbb{Z}}$. In terms of the parameters of the quasi likelihood, the hypothesis of no regime switch is formulated as

$$H : (\zeta_1, \sigma_1) = (\zeta_2, \sigma_2) \qquad \text{or} \qquad \alpha(1-\alpha) = 0.$$

A methodologically simple approach would be to test $H$ via a likelihood ratio statistic based on the quasi likelihood,

$$QLR_n = 2\big(\max_{\alpha,\zeta_1,\zeta_2,\sigma_1,\sigma_2,\boldsymbol{\phi}} l_n(\alpha,\zeta_1,\zeta_2,\sigma_1,\sigma_2,\boldsymbol{\phi}) - l_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\sigma}_0,\widehat{\sigma}_0,\widehat{\boldsymbol{\phi}}_0)\big), \tag{5}$$

where $\widehat{\zeta}_0$, $\widehat{\sigma}_0$ and $\widehat{\boldsymbol{\phi}}_0$ are the maximizers of $l_n$ under the null. Cho & White (2007) use this approach in the simpler case of a single switching parameter. However, in case of simultaneously switching intercept and scale, the asymptotic distribution of the QLR test is unknown and seems to be very difficult to derive. The tests which will be introduced are designed for alternatives $(\zeta_1,\sigma_1)\neq(\zeta_2,\sigma_2)$ and (due to essentially fixing the weights and penalizing small weights) not against small deviations of the mixture weight from 0 or 1.

## 2.2. Tests for homogeneity

A simple possibility is to consider a finite set of fixed values $\mathcal{J}=\{\alpha_1,\ldots,\alpha_J\}$ for $\alpha$. Define

$$R_n(\alpha_j) = 2\big(l_n(\alpha_j,\widehat{\zeta}_{1,\alpha_j},\widehat{\zeta}_{2,\alpha_j},\widehat{\sigma}_{1,\alpha_j},\widehat{\sigma}_{2,\alpha_j},\widehat{\boldsymbol{\phi}}_{\alpha_j}) - l_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\sigma}_0,\widehat{\sigma}_0,\widehat{\boldsymbol{\phi}}_0)\big), \qquad \alpha_j\in\mathcal{J},$$

where $(\widehat{\zeta}_{1,\alpha_j},\widehat{\zeta}_{2,\alpha_j},\widehat{\sigma}_{1,\alpha_j},\widehat{\sigma}_{2,\alpha_j},\widehat{\boldsymbol{\phi}}_{\alpha_j})$ is the maximizer of $l_n(\alpha,\zeta_1,\zeta_2,\sigma_1,\sigma_2,\boldsymbol{\phi})$ subject to $\alpha=\alpha_j$, and set

$$R_n(\mathcal{J}) = \max_{\alpha_j\in\mathcal{J}} R_n(\alpha_j).$$

**Theorem 1.** *Under the null model, we have for finite $\mathcal{J}\subset(0,1)$, whenever $1/2\in\mathcal{J}$,*

$$R_n(\mathcal{J}) \xrightarrow{d} \chi^2_2, \qquad n\to\infty,$$

*where $\chi^2_\nu$ denotes the $\chi^2$-distribution with $\nu>0$ degrees of freedom and $\xrightarrow{d}$ denotes convergence in distribution.*

We provide an outline of the proof of the theorem in the appendix, full details are provided in Ketterer & Holzmann (2012). While the simulations in Section 3 show that the test $R_n(1/2)$ keeps the nominal level quite well (already for small sample sizes), additional simulations (not displayed) indicate that if the set $\mathcal{J}$ contains further elements, as e.g. $\mathcal{J}=\{0.1,0.3,0.5\}$, the test is highly anticonservative even for moderately large sample sizes, and therefore the asymptotic approximation should not be used.

As in Chen & Li (2009) and Holzmann & Ketterer (2011), one can also construct a so-called EM-test. This is based on a penalized quasi likelihood function of the form

$$pl_n(\alpha,\zeta_1,\zeta_2,\sigma_1,\sigma_2,\boldsymbol{\phi}) = l_n(\alpha,\zeta_1,\zeta_2,\sigma_1,\sigma_2,\boldsymbol{\phi}) + p(\alpha), \quad p(\alpha) = C\log(1-|1-2\alpha|), \tag{6}$$

for some fixed $C>0$. For each $\alpha_j\in\mathcal{J}$, we compute

$$M_n^{(0)}(\alpha_j) = R_n(\alpha_j) + 2\{p(\alpha_j)-p(0.5)\}.$$

Then for each $\alpha_j$ we perform a fixed finite number $K$ of EM-steps using $pl_n$ to obtain updated EM-estimates $(\alpha_j^{(k)},\zeta_{1j}^{(k)},\zeta_{2j}^{(k)},\sigma_{1j}^{(k)},\sigma_{2j}^{(k)},\boldsymbol{\phi}_j^{(k)})$, $k=1,\ldots,K$, and test statistics

$$M_n^{(k)}(\alpha_j) = 2\left\{pl_n(\alpha_j^{(k)},\zeta_{1j}^{(k)},\zeta_{2j}^{(k)},\sigma_{1j}^{(k)},\sigma_{2j}^{(k)},\boldsymbol{\phi}_j^{(k)}) - pl_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\sigma}_0,\widehat{\sigma}_0,\widehat{\boldsymbol{\phi}}_0)\right\}.$$

The EM-statistic $EM_n^{(K)}(\mathcal{J})$ is then set to the maximal value

$$EM_n^{(K)}(\mathcal{J}) = \max\left\{M_n^{(K)}(\alpha_j), j = 1,\dots,J\right\}.$$

The precise formulation of the test statistic $EM_n^{(K)}(\mathcal{J})$ is given in the appendix. On the one hand, by performing the EM-iteration, one hopes to improve the power properties of the test, on the other hand, by penalizing values $\alpha_j \in \mathcal{J}$ with $\alpha_j \neq 1/2$, the test is less anticonservative than the corresponding version of $R_n(\mathcal{J})$. The asymptotic distribution of $EM_n^{(K)}(\mathcal{J})$ under the null is the same as that of $R_n(\mathcal{J})$,

$$EM_n^{(K)}(\mathcal{J}) \xrightarrow{d} \chi_2^2, \qquad n \to \infty. \tag{7}$$

Here, it is necessary to include $1/2 \in \mathcal{J}$ in order to obtain the supposed asymptotic distribution. If not so, the asymptotic distribution of $EM_n^{(K)}(\mathcal{J})$ would be a location shifted $\chi_2^2$-distribution where the shift $\max_{\alpha_j \in \mathcal{J}} 2\{p(\alpha_j) - p(0.5)\}$ is due to the penalty on $\alpha$. A proof of (7) can be found in Ketterer (2011).

## 3. SIMULATIONS

Here we present some of the results of an extensive simulation study of the proposed tests. For the EM-test and $R_n(\mathcal{J})$, we always choose $\mathcal{J} = \{0.1, 0.3, 0.5\}$, and further consider $R_n(1/2)$. For the lower bound $\delta$ on the $\sigma_i$'s one may choose a fixed fraction (e.g. $1/5$) of the value $\hat{\sigma}_0$ estimated under the homogeneous model. Alternatively, following Chen, Tan, & Zhang (2008) and Chen & Li (2009) we add the additional penalty

$$\tilde{p}_n(\sigma_1) + \tilde{p}_n(\sigma_2), \qquad \text{where} \quad \tilde{p}_n(\sigma) = -0.5 \cdot \left(\hat{\sigma}_0^2/\sigma^2 + \log(\sigma^2/\hat{\sigma}_0^2)\right)$$

to the penalized quasi likelihood function (6). For the penalty on $\alpha$ in Equation (6), we choose $C = 3$ in general and $C = 1$ if the penalty on $\sigma$ given above is added. Indeed, if $C = \infty$, the EM-test reduces to the test against the fixed proportion of $1/2$, and if $C = 0$ to the simple quasi likelihood ratio test. In a further paper, Chen & Li (2011) give recommendations (in a slightly different setting of the EM-test) how to determine suitable values of $C$ by simulations, mainly to keep the level under the hypothesis. A more simple approach would be to first estimate the parameters under the hypothesis, and then to try different values for $C$ in simulations when generating data from the estimated model, and then choose $C$ minimal such that the test still keeps its nominal level. We simply investigated several fixed values (between $0.5$ and 3) for some models and parameter values, and found that 1 (in case of penality on $\sigma$) and 3 (without further penalty) worked reasonably well in the scenarios under investigation in terms of both level and power.

### 3.1. Simulated sizes

First, we simulate the sizes of the two tests for the following data generating processes (DGP).
DGP 1: $X_t = 0.5X_{t-1} + \epsilon_t$ where $\epsilon_t \overset{iid}{\sim} N(0,1)$,
Model 1: $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$.
The results for various sample sizes are displayed in Table 4.2 (without penalty on $\sigma$) and Table 4.2 (with penalty on $\sigma$).
DGP 2: $X_t = 0.6X_{t-1} - 0.3X_{t-2} + \epsilon_t$ where $\epsilon_t \overset{iid}{\sim} N(0,1)$,
Model 2: $X_t = \zeta_{S_t} + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \sigma_{S_t}\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$.

The results for various sample sizes can be found in Table 4.2 (without penalty on $\sigma$) and Table 4.2 (with penalty on $\sigma$).

Without penalty on $\sigma$, the tests are slightly anticonservative for the small sample size $n = 200$ in both scenarios, but keep the level reasonably well for larger sample sizes. Adding the penalty on $\sigma$ improves the behaviour of the EM-test, but makes $R_n(1/2)$ somewhat conservative.

### 3.2. Power comparison of several tests

Next we conduct a power comparison between several tests. For proper estimation of the power we shall use simulated critical values. Precisely, for given alternative, we generate a single large sample ($n = 10000$) from this alternative and fit a corresponding null model to this sample by maximum likelihood. Note that this will approximate the null model with minimal Kullback-Leibler distance to the given alternative. From this null model, we generate 10000 samples of size 200, and in each case compute the test statistics and then compute the finite sample critical values for the test statistics as empirical quantiles. Although the asymptotic properties of the quasi likelihood ratio test (abbreviated QLRT) given in Equation (5) are unknown, we use it, with simlulated critical values as described above, as a benchmark test.

Furthermore, we evaluate the power of the EM-test designed for linear switching autoregressive models with a single switching parameter, namely a switching intercept, see Holzmann and Ketterer (2011). We denote the corresponding test statistic by $\widetilde{EM}_n^{(K)}(\mathcal{J})$, here.

DGP 1: $X_t = (-1)^{S_t} \cdot 0.1 + 0.5X_{t-1} + (\mathbb{1}_{\{S_t=1\}} + 2 \cdot \mathbb{1}_{\{S_t=2\}})\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0, 1)$ and various combinations $a_{12}$ and $a_{21}$ leading to different values of $\alpha$.

Model 1(a): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0, 1)$, for $EM_n^{(K)}(\mathcal{J})$, $QLR_n$ and $R_n(1/2)$

Model 1(b): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0, 1)$, for $\widetilde{EM}_n^{(K)}(\mathcal{J})$.

The results (without penalty on $\sigma$) are presented in Table 4.2 (there, the transition probabilities of the hidden Markov chain are in bold print if and only if the regime reduces to an i.i.d. sample, i.e. $a_{12} = 1 - a_{21}$ holds). The power of the compared tests mainly depends on the stationary distribution of $(S_t)_{t\in\mathbb{Z}}$. The particular transition probabilities of the hidden Markov chain do not significantly influence the power of the corresponding tests. In all scenarios, the test based on fixed $\alpha = 1/2$ under the alternative slightly outperforms the EM-test: Due to the large value $C = 3$ in Equation (6), the EM-test takes its maximum for the starting value $\alpha_j = 1/2$. The EM-test designed for linear switching autoregressive models with possibly switching intercept under the alternative shows the lowest power of the tests under consideration. Therefore, we should only use this test if there is a priori knowledge that the scale parameter in both regimes is almost identical. We also included the penalty on $\sigma$ in all test statistics. The simulated critical values for the QLRT, $R_n(\mathcal{J})$ and the EM-test are much smaller in this case, which also leads to higher power of these tests (see Table 4.2). For small values of $\alpha$ under the alternative, $R_n(\mathcal{J})$, the QLRT and also the EM - test have higher power than $R_n(1/2)$, as could be expected.

DGP 2: $X_t = (-1)^{S_t} + 0.3X_{t-1} + (\mathbb{1}_{\{S_t=1\}} + 1.1 \cdot \mathbb{1}_{\{S_t=2\}})\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0, 1)$ and various combinations $a_{12}$ and $a_{21}$ leading to different values of $\alpha$.

Model 2(a): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0, 1)$, for $EM_n^{(K)}(\mathcal{J})$, $QLR_n$ and $R_n(1/2)$

Model 2(b): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0, 1)$, for $\widetilde{EM}_n^{(K)}(\mathcal{J})$.

The results (without penalty on $\sigma$) are presented in Table 4.2. Here, the power of our tests highly depends on the stationary distribution as well as on the transition probabilities of the hidden Markov chain. Since the scale parameter of the innovation process is quite close in both regimes, the EM-test designed for linear switching autoregressive models with possibly switching intercept under the alternative outperforms the tests which allow for a switch in both intercept and scale. We also included the penalty on $\sigma$ in all test statistics, the effect of which is similar as in the above scenario, see Table 4.2).

Finally, we compare the power of the tests under consideration with the power of the MQLRT designed for linear switching autoregressive models with possibly switching scale parameter of the innovations as introduced in Holzmann and Ketterer (2011), denoted by $M_n$.

DGP 3: $X_t = (-1)^{S_t} \cdot 0.7 + 0.5X_{t-1} + (1.8 \cdot \mathbb{1}_{\{S_t=1\}} + \cdot\mathbb{1}_{\{S_t=2\}})\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$ and various combinations $a_{12}$ and $a_{21}$ leading to different values of $\alpha$.

Model 3(a): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $EM_n^{(K)}(\mathcal{J})$, $QLR_n$ and $R_n(1/2)$,

Model 3(b): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $\widetilde{EM}_n^{(K)}(\mathcal{J})$

Model 3(c): $X_t = \zeta + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $M_n$.

The results (without penalty on $\sigma$) can be found in Table 4.2. Here, the tests designed for switching intercept and variance under the alternative outperform $\widetilde{EM}_n^{(K)}(\mathcal{J})$ as well as $M_n$. In all scenarios, the test based on fixed $\alpha = 1/2$ has slightly higher power than the corresponding EM-test. Since $M_n$ and $\widetilde{EM}_n^{(K)}(\mathcal{J})$ have lower power than the tests designed for two switching parameters, $EM_n^{(K)}(\mathcal{J})$ or $R_n(1/2)$ should be preferred if there is no a priori knowledge that the scale parameter (resp. the intercept) in both regimes is almost identical. We also included the penalty on $\sigma$ in all test statistics, the effect of which is similar as in the above scenarios, see Table 4.2).

Summarizing, regarding level together with power we recommend the use of either $R_n(1/2)$ without penalties, or of $EM_n^{(1)}(\mathcal{J})$ with the additional penalty on $\sigma$, using an appropriate value for the tuning constant $C > 0$.

## 4. APPLICATION

In this section, we apply our methods to the series of monthly log returns $(X_t)_{t=-3,\ldots,884}$ of IBM stock from January 1926 to December 1999. The returns are in percentage and include dividends. The data can be obtained from `http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts/m-ibmspln.dat`

### 4.1. Marginal distribution

A kernel density estimate (see Figure 1, right) indicates slight asymmetry and fat tails. The empirical skewness coefficient and kurtosis are given by $-0.2369$ and $4.9278$, respectively. To deal with skewness and kurtosis in the unconditional distribution of stock returns finite normal mixtures have been applied quite often, see e.g. Kon (1984). In a first step, we test one against two components in a normal mixture model using penalized likelihood based tests. The hypothesis of a single component is rejected (with $p$-value $< 0.001$) by every test under consideration. Testing against an alternative with possibly distinct means and variances using the EM-test introduced in Chen and Li (2009), we see that the alternative two-component model has almost identical means ($\overline{\mu}_1 = 1.37$ and $\overline{\mu}_2 = 0.94$), quite different standard deviations ($\overline{\sigma}_1 = 4.80$ and $\overline{\sigma}_2 = 9.89$) and the relative size of component 2 is $\overline{\alpha} = 0.30$.

But modeling the series of log returns for IBM stock by finite mixtures would only be appropriate if the data did not exhibit autocorrelation. The empirical partial autocorrelation indicates an AR(1) process, and the autocorrelation function of the squares gives evidence for conditional heteroscedasticity. Therefore we use time series models for the analysis.

### 4.2. Autoregressive model

To start, we fit an AR(1) model

$$X_t = \zeta + \phi X_{t-1} + \sigma\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0,1),$$

yielding the estimate $(\widehat{\zeta}, \widehat{\phi}, \widehat{\sigma}) = (1.157, 0.077, 6.698)$, to capture autocorrelation in our time series. Computing the residuals $(\widehat{\epsilon}_t)$ of the fitted model and testing for normality using the Anderson-Darling test ($A_n = 2.95$) we strongly reject $H : \epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$ by a $p$-value $< 0.001$ (for reasons why to use asymptotic critical values of the Anderson-Darling test for independent and identically distributed observations, cf. Pierce 1985) and which indicates lack-of-fit of the supposed AR(1) model. While Tsay (2002) fits an AR(1)-GARCH(1,1) to the monthly log returns of IBM stock, Kim, Nelson, & Startz (1998) and Bhar & Hamori (2004) suggest modeling monthly stock returns by Markov-switching autoregressive models, in general. We follow this approach and fit several linear switching autoregressive models to the data.

Testing for homogeneity in the model

$$X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0,1), \tag{8}$$

using the tests introduced in this paper, we clearly reject the hypothesis of no regime switch using the test against fixed $\alpha = 1/2$ under the alternative ($R_n(1/2) = 47.64$ without penalty on sigma, and $R_n(1/2) = 46.44$ with penalty on sigma) and the EM-test ($EM_n^{(2)}(\mathcal{J}) = 52.05$ without penalty on sigma and $C = 3$, and $EM_n^{(2)}(\mathcal{J}) = 53.03$ with penalty on sigma and $C = 1$) with a $p$-value $< 0.001$. The full model MLE for model (8) yields $(\widehat{a}_{12}, \widehat{a}_{21}) = (0.015, 0.052)$, $(\widehat{\zeta}_1, \widehat{\sigma}_1) = (1.266, 5.183)$, $(\widehat{\zeta}_2, \widehat{\sigma}_2) = (0.752, 10.383)$ and $\widehat{\phi} = 0.081$. Our analysis indicates that there are two regimes: Regime 1 with higher mean level in the (log) returns and lower variance and regime 2 with slightly lower mean level in the (log) returns and higher variance.

Our next analysis concerns testing if the intercept as well as the scale parameter of the innovations switch according to the hidden Markov-chain or if one of the parameters is equal in both scenarios. Fitting model (8) with two states using maximum likelihood for the full model, we have to find parameters $\boldsymbol{\omega} = (a_{12}, a_{21}, (\zeta_1, \sigma_1), (\zeta_2, \sigma_2), \phi)^T$ with $(\zeta_1, \sigma_1) \neq (\zeta_2, \sigma_2)$ and $(a_{12}, a_{21}) \in (0, 1]^2$ maximizing $\tilde{l}_n(\boldsymbol{\omega})$.

Assuming that $\sigma_1 \neq \sigma_2$, testing $H : \zeta_1 = \zeta_2$ in model (8) with two states is a regular problem and therefore the likelihood ratio statistic

$$T_n = 2\big(\max_{\boldsymbol{\omega}} \tilde{l}_n(\boldsymbol{\omega}) - \max_{\boldsymbol{\omega}:\zeta_1=\zeta_2} \tilde{l}_n(\boldsymbol{\omega})\big)$$

asymptotically follows a $\chi_1^2$ distribution. In our case, we have $T_n = 0.322$ ($p$-value=0.57). Therefore, we cannot reject the hypothesis $H : \zeta_1 = \zeta_2$.

Assuming that $\zeta_1 \neq \zeta_2$, testing the hypothesis $H : \sigma_1 = \sigma_2$ in model (8) is also a regular problem and therefore the likelihood ratio statistic

$$T_n = 2\big(\max_{\boldsymbol{\omega}} \tilde{l}_n(\boldsymbol{\omega}) - \max_{\boldsymbol{\omega}:\sigma_1=\sigma_2} \tilde{l}_n(\boldsymbol{\omega})\big)$$

asymptotically follows a $\chi_1^2$ distribution. In our case, we have $T_n = 112.8643$ ($p$-value $< 0.001$). Therefore, we clearly reject the hypothesis $H : \sigma_1 = \sigma_2$.

## Tables and Figures

TABLE 1: DGP: $X_t = 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, Model: $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$; number of replications: 20000, no penalty on $\sigma$

| Sample Size | Levels (%) | $EM_n^{(0)}(\mathcal{J})$ | $EM_n^{(1)}(\mathcal{J})$ | $EM_n^{(2)}(\mathcal{J})$ | $R_n(1/2)$ |
|---|---|---|---|---|---|
| $n = 200$ | 10% | 11.8 | 12.0 | 12.0 | 11.1 |
| | 5% | 6.4 | 6.4 | 6.4 | 5.8 |
| | 1% | 1.4 | 1.4 | 1.4 | 1.2 |
| $n = 500$ | 10% | 10.9 | 10.9 | 10.9 | 10.7 |
| | 5% | 5.6 | 5.6 | 5.6 | 5.4 |
| | 1% | 1.4 | 1.4 | 1.4 | 1.3 |
| $n = 1000$ | 10% | 10.4 | 10.4 | 10.4 | 10.4 |
| | 5% | 5.3 | 5.3 | 5.3 | 5.2 |
| | 1% | 1.1 | 1.1 | 1.1 | 1.1 |

TABLE 2: DGP: $X_t = 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, Model: $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$; number of replications: 20000, penalty on $\sigma$ included.

| Sample Size | Levels (%) | $EM_n^{(1)}(\mathcal{J})$ | $EM_n^{(2)}(\mathcal{J})$ | $R_n(\mathcal{J})$ | $R_n(1/2)$ |
|---|---|---|---|---|---|
| $n = 200$ | 10% | 9.4 | 9.4 | 13.7 | 8.7 |
| | 5% | 4.6 | 4.7 | 7.2 | 4.2 |
| | 1% | 1.0 | 1.0 | 1.4 | 0.8 |
| $n = 500$ | 10% | 9.7 | 9.7 | 12.8 | 9.3 |
| | 5% | 5.0 | 5.0 | 6.5 | 4.6 |
| | 1% | 0.9 | 0.9 | 1.4 | 0.8 |

## Acknowledgements

The authors would like to thank the editor Professor Chen, the associate editor and the two reviewers for their detailed and helpful comments. Financial support from the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

## BIBLIOGRAPHY

Bhar, R., & Hamori, S. (2004). *Hidden Markov Models*. Advanced Studies in Theoretical and Applied Econometrics. Kluwer Academic Publishers, Dortrecht.

Chen, H., & Chen, J. (2003). Tests for homogeneity in normal mixtures in the presence of a structural parameter. *Statistica Sinica*, 13, 351-365.

Chen, J., & Li, P. (2009). Hypothesis test for Normal Mixture Models: the EM Approach. *The Annals of Statistics*, 37, 2523–2542.

TABLE 3: DGP: $X_t = 0.6X_{t-1} - 0.3X_{t-2} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, Model: $X_t = \zeta_{S_t} + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \sigma_{S_t}\epsilon_t$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$; number of replications: 20000, no penalty on $\sigma$

| Sample Size | Levels (%) | $EM_n^{(0)}(\mathcal{J})$ | $EM_n^{(1)}(\mathcal{J})$ | $EM_n^{(2)}(\mathcal{J})$ | $R_n(1/2)$ |
|---|---|---|---|---|---|
| $n = 200$ | 10% | 12.5 | 12.8 | 13.0 | 11.6 |
| | 5% | 6.7 | 6.9 | 7.0 | 6.0 |
| | 1% | 1.7 | 1.8 | 1.8 | 1.5 |
| $n = 500$ | 10% | 11.1 | 11.1 | 11.1 | 10.8 |
| | 5% | 5.9 | 5.9 | 5.9 | 5.7 |
| | 1% | 1.2 | 1.2 | 1.2 | 1.1 |
| $n = 1000$ | 10% | 10.8 | 10.8 | 10.8 | 10.7 |
| | 5% | 5.4 | 5.4 | 5.4 | 5.3 |
| | 1% | 1.0 | 1.0 | 1.0 | 1.0 |

TABLE 4: DGP: $X_t = 0.6X_{t-1} - 0.3X_{t-2} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, Model: $X_t = \zeta_{S_t} + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \sigma_{S_t}\epsilon_t$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$; number of replications: 20000, penalty on $\sigma$ included.

| Sample Size | Levels (%) | $EM_n^{(1)}(\mathcal{J})$ | $EM_n^{(2)}(\mathcal{J})$ | $R_n(\mathcal{J})$ | $R_n(1/2)$ |
|---|---|---|---|---|---|
| $n = 200$ | 10% | 9.9 | 9.9 | 13.9 | 9.2 |
| | 5% | 5.1 | 5.1 | 7.2 | 4.7 |
| | 1% | 1.1 | 1.1 | 1.6 | 1.0 |
| $n = 500$ | 10% | 9.4 | 9.4 | 13.1 | 9.3 |
| | 5% | 4.6 | 4.6 | 6.9 | 4.5 |
| | 1% | 0.9 | 0.9 | 1.5 | 0.9 |

Chen, J., & Li, P. (2011). Tuning the EM-test for the order of finite mixture models. *The Canadian Journal of Statistics*, 39, 389-404.

Chen, J., Tan, X., & Zhang, R. (2008). Consistency of penalized MLE for normal mixtures in mean and variance. *Statistica Sinica*, 18, 443–465.

Cho, J.S., & White,H. (2007). Testing for regime switching. *Econometrica*, 75, 1671–1720.

Dacunha-Castelle, D., & Gassiat, E. (1999). Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes. *The Annals of Statistics*, 27, 1178–1209.

Gassiat E., & Keribin C. (2000). The likelihood ratio test for the number of components in a mixture with Markov regime. *ESAIM P&S*, 4, 25–52.

Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57, 357–384.

Hamilton, J.D. (2008). Regime-Switching Models. In S.N.Durlauf and L.E.Blume (Eds.)*The New Palgrave Dictionary of Economics, Second Edition*. Palgrave McMillan Ltd.

Hathaway, R.J. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *The Annals of Statistics*, 13, 795–800.

TABLE 5: Nominal level: 5% DGP: $X_t = (-1)^{S_t} \cdot 0.1 + 0.5X_{t-1} + (\mathbb{1}_{\{S_t=1\}} + 2 \cdot \mathbb{1}_{\{S_t=2\}})\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$, sample size: 200, number of replications: $5,000$. Model (a): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $EM_n^{(K)}(\mathcal{J})$, $QLR_n$ and $R_n(1/2)$ and Model (b): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $\widetilde{EM}_n^{(K)}(\mathcal{J})$. Let $\alpha = a_{12}/(a_{12}+a_{21})$ be the stationary distribution of the hidden Markov chain $(S_t)_{t\in\mathbb{Z}}$, no penalty on $\sigma$

| $a_{12}$ | $a_{21}$ | $\alpha$ | $EM_n^{(0)}(\mathcal{J})$ | $EM_n^{(1)}(\mathcal{J})$ | $\widetilde{EM}_n^{(0)}(\mathcal{J})$ | $\widetilde{EM}_n^{(1)}(\mathcal{J})$ | $QLR_n$ | $R_n(1/2)$ |
|---|---|---|---|---|---|---|---|---|
| 0.1 | 0.1 | 0.5 | 60.6 | 60.0 | 15.1 | 15.8 | 42.7 | 60.9 |
| 0.3 | 0.3 | | 62.7 | 62.6 | 15.9 | 16.5 | 41.9 | 63.5 |
| **0.5** | **0.5** | | 62.6 | 62.4 | 14.4 | 15.1 | 46.7 | 63.1 |
| 0.9 | 0.9 | | 63.6 | 63.5 | 14.6 | 15.4 | 40.9 | 64.4 |
| **0.4** | **0.6** | 0.4 | 71.9 | 71.5 | 19.2 | 20.5 | 56.3 | 72.5 |
| 0.3 | 0.45 | | 71.5 | 71.0 | 18.3 | 19.5 | 54.7 | 72.6 |
| 0.2 | 0.3 | | 70.4 | 70.3 | 18.5 | 19.7 | 51.9 | 71.2 |
| 0.1 | 0.15 | | 68.0 | 67.7 | 18.8 | 20.0 | 54.1 | 69.1 |
| **0.2** | **0.8** | 0.2 | 65.4 | 64.9 | 18.6 | 20.9 | 52.4 | 66.2 |
| 0.1 | 0.4 | | 63.4 | 63.1 | 18.1 | 19.6 | 51.9 | 64.3 |
| **0.1** | **0.9** | 0.1 | 42.5 | 42.1 | 10.4 | 11.7 | 36.1 | 43.5 |
| 0.05 | 0.45 | | 39.9 | 39.4 | 9.0 | 10.2 | 36.2 | 40.9 |

TABLE 6: As in Table 4.2 but including the penalty on $\sigma$ for all test statistics.

| $a_{12}$ | $a_{21}$ | $\alpha$ | $EM_n^{(1)}(\mathcal{J})$ | $QLR_n$ | $R_n(\mathcal{J})$ | $R_n(1/2)$ |
|---|---|---|---|---|---|---|
| 0.1 | 0.1 | 0.5 | 60.6 | 58.5 | 60.0 | 61.0 |
| **0.5** | **0.5** | | 63.3 | 61.3 | 64.3 | 64.0 |
| **0.4** | **0.6** | 0.4 | 73.8 | 71.2 | 72.3 | 73.3 |
| **0.2** | **0.8** | 0.2 | 67.4 | 68.6 | 69.4 | 67.1 |
| **0.1** | **0.9** | 0.1 | 45.0 | 50.7 | 50.1 | 43.6 |
| 0.05 | 0.45 | | 43.3 | 48.1 | 47.7 | 41.9 |

Holzmann, H., & Ketterer, F. (2011). Feasible tests for regime switching in autoregressive models. Working paper, Marburg University.

Ketterer, F. (2011). *Penalized likelihood based tests for regime switching in autoregressive models*. Ph.D.thesis, Marburg University, Germany.

Ketterer, F., & Holzmann, H. (2012). Testing for intercept-scale switch in linear autoregression: Technical details. Supplementary material.

Kim, C.J., Nelson, C.R., & Startz, R. (1998). Testing for me an reversion in heteroscedastic data based on Gibbs-sampling augmented randomization. *Journal of Empirical Finance*, 5, 131–154.

Kon, S.J. (1984). Models of stock returns – A comparison. *The Journal of Finance*, 39, 147–165.

Kreiss, J.P., & Neuhaus, G. (2006). *Einführung in die Zeitreihenanalyse*. Springer, Berlin. (in German)

Krengel, U. (1985). *Ergodic Theorems*. De Gruyter Studies in Mathematics, Berlin.

Lange, T., & Rahbek A. (2009). An Introduction to Regime Switching Time Series Models. In *Handbook of Financial Time Series* (R.A.Davis, J.P.Kreiss and T.Mikosch,eds.). Springer, New York.

TABLE 7: Nominal level: 5% DGP: $X_t = (-1)^{S_t} + 0.3X_{t-1} + (\mathbb{1}_{\{S_t=1\}} + 1.1 \cdot \mathbb{1}_{\{S_t=2\}})\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$, sample size: 200, number of replications: $5,000$. Model (a): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $EM_n^{(K)}(\mathcal{J})$, $QLR_n$ and $R_n(1/2)$ and Model (b): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $\widetilde{EM}_n^{(K)}(\mathcal{J})$, $\alpha = a_{12}/(a_{12}+a_{21})$, no penalty on $\sigma$.

| $a_{12}$ | $a_{21}$ | $\alpha$ | $EM_n^{(0)}(\mathcal{J})$ | $EM_n^{(1)}(\mathcal{J})$ | $\widetilde{EM}_n^{(0)}(\mathcal{J})$ | $\widetilde{EM}_n^{(1)}(\mathcal{J})$ | $QLR_n$ | $R_n(1/2)$ |
|---|---|---|---|---|---|---|---|---|
| 0.1 | 0.1 | 0.5 | 8.4 | 8.3 | 9.1 | 9.1 | 6.0 | 8.7 |
| 0.3 | 0.3 | | 26.8 | 26.7 | 38.6 | 38.2 | 17.9 | 27.8 |
| **0.5** | **0.5** | | 37.7 | 37.6 | 53.5 | 53.0 | 22.0 | 38.7 |
| 0.7 | 0.7 | | 24.9 | 25.0 | 38.8 | 38.3 | 15.6 | 25.0 |
| 0.9 | 0.9 | | 8.6 | 8.5 | 12.1 | 12.4 | 6.2 | 8.8 |
| **0.4** | **0.6** | 0.4 | 49.2 | 49.3 | 54.7 | 55.4 | 29.5 | 50.0 |
| 0.2 | 0.3 | | 27.4 | 27.3 | 35.5 | 35.5 | 17.6 | 28.3 |
| 0.1 | 0.15 | | 14.0 | 14.0 | 13.1 | 13.4 | 6.2 | 14.0 |
| **0.2** | **0.8** | 0.2 | 65.1 | 65.0 | 62.1 | 62.5 | 47.7 | 66.0 |
| 0.1 | 0.4 | | 46.7 | 46.3 | 45.7 | 45.5 | 29.6 | 47.4 |
| **0.1** | **0.9** | 0.1 | 50.0 | 49.6 | 44.7 | 44.2 | 36.4 | 50.3 |
| 0.05 | 0.45 | | 36.9 | 36.8 | 32.8 | 32.5 | 26.4 | 37.4 |

TABLE 8: As in Table 4.2 but including the penalty on $\sigma$ for all test statistics.

| $a_{12}$ | $a_{21}$ | $\alpha$ | $EM_n^{(1)}(\mathcal{J})$ | $QLR_n$ | $R_n(\mathcal{J})$ | $R_n(1/2)$ |
|---|---|---|---|---|---|---|
| 0.1 | 0.1 | 0.5 | 9.0 | 8.4 | 8.6 | 8.8 |
| **0.5** | **0.5** | | 38.8 | 34.9 | 35.8 | 38.2 |
| **0.4** | **0.6** | 0.4 | 50.2 | 46.9 | 48.8 | 50.5 |
| **0.2** | **0.8** | 0.2 | 66.6 | 66.1 | 67.3 | 67.9 |
| **0.1** | **0.9** | 0.1 | 51.6 | 54.1 | 54.3 | 52.6 |
| 0.05 | 0.45 | | 40.4 | 41.4 | 42.5 | 40.3 |


Li, P., Chen, J., & Marriot, P. (2009). Non-finite Fisher information and homogeneity: the EM approach. *Biometrika*, 96, 411–426.

Pierce, D.A. (1985). Testing normality in autoregressive models. *Biometrika*, 72, 293–297.

Piger, J. (2009). Econometrics: models of regime changes. In *Encyclopedia of Complexity and Systems Science* (R.A.Meyers, ed.). Springer, New York.

Rydén, T., Teräsvirta, T., & Asbrink, S. (1998). Stylized facts of daily return series and the hidden Markov model of absolute returns. *Journal of Applied Econometrics*, 13, 217-244.

Tsay, R.S. (2002). *Analysis of financial time series*. John Wiley and Sons, New York.

Velucchi, M. (2009). Regime Switching: Italian financial markets over a century. *Statistical Methods and Applications*, 18, 67–86.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics* 20, 595–601.

TABLE 9: Nominal level: 5% DGP: $X_t = (-1)^{S_t} \cdot 0.7 + 0.5X_{t-1} + (1.8 \cdot \mathbb{1}_{\{S_t=1\}} + \cdot\mathbb{1}_{\{S_t=2\}})\epsilon_t$ with $\epsilon_t \stackrel{iid}{\sim} N(0,1)$, sample size: 200, number of replications: $5,000$. Model (a): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, $\epsilon_t \stackrel{iid}{\sim} N(0,1)$, for $EM_n^{(K)}(\mathcal{J})$, $QLR_n$ and $R_n(1/2)$, Model (b): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, $\epsilon_t \stackrel{iid}{\sim} N(0,1)$, for $\widetilde{EM}_n^{(K)}(\mathcal{J})$, and Model (c): $X_t = \zeta + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, $\epsilon_t \stackrel{iid}{\sim} N(0,1)$, for $M_n$. Let $\alpha = a_{12}/(a_{12}+a_{21})$ be the stationary distribution of the hidden Markov chain $(S_t)_{t\in\mathbb{Z}}$.

| $a_{12}$ | $a_{21}$ | $\alpha$ | $EM_n^{(0)}(\mathcal{J})$ | $EM_n^{(1)}(\mathcal{J})$ | $\widetilde{EM}_n^{(0)}(\mathcal{J})$ | $\widetilde{EM}_n^{(1)}(\mathcal{J})$ | $M_n$ | $QLR_n$ | $R_n(1/2)$ |
|---|---|---|---|---|---|---|---|---|---|
| 0.1 | 0.1 | 0.5 | 74.5 | 74.6 | 35.6 | 41.3 | 48.8 | 57.8 | 75.3 |
| 0.3 | 0.3 | | 84.5 | 84.5 | 54.2 | 59.6 | 50.2 | 67.0 | 85.4 |
| **0.5** | **0.5** | | 87.5 | 87.5 | 47.4 | 51.3 | 36.6 | 70.5 | 87.7 |
| 0.7 | 0.7 | | 85.5 | 85.3 | 51.4 | 55.4 | 39.7 | 69.6 | 86.3 |
| 0.9 | 0.9 | | 83.4 | 83.3 | 47.1 | 51.9 | 38.5 | 65.1 | 84.3 |
| **0.4** | **0.6** | 0.4 | 74.6 | 74.4 | 49.7 | 52.2 | 18.9 | 52.9 | 75.2 |
| 0.2 | 0.3 | | 70.8 | 70.6 | 42.4 | 45.4 | 22.8 | 50.6 | 71.7 |
| **0.2** | **0.8** | 0.2 | 26.3 | 26.2 | 23.3 | 23.6 | 4.3 | 15.3 | 26.7 |
| 0.1 | 0.4 | | 26.4 | 26.4 | 21.0 | 22.0 | 5.0 | 16.9 | 26.6 |

TABLE 10: As in Table 4.2 but including the penalty on $\sigma$ for all test statistics.

| $a_{12}$ | $a_{21}$ | $\alpha$ | $EM_n^{(1)}(\mathcal{J})$ | $QLR_n$ | $R_n(\mathcal{J})$ | $R_n(1/2)$ |
|---|---|---|---|---|---|---|
| 0.1 | 0.1 | 0.5 | 76.7 | 75.2 | 76.5 | 76.8 |
| **0.5** | **0.5** | | 87.0 | 85.3 | 86.3 | 87.6 |
| **0.4** | **0.6** | 0.4 | 75.1 | 72.9 | 74.3 | 74.9 |
| **0.2** | **0.8** | 0.2 | 28.6 | 27.4 | 27.5 | 28.5 |
| 0.1 | 0.4 | | 28.5 | 27.0 | 28.0 | 28.4 |

# Appendix

## The EM-test

In the following we describe the (quasi) EM-test for testing for homogeneity in model (1). Note that in the following algorithm we proceed in some steps via the ECM algorithm instead of the EM algorithm. If we use the EM algorithm, we have to derive the updated estimators $(\zeta_{1j}^{(k+1)}, \zeta_{2j}^{(k+1)}, \sigma_{1j}^{(k+1)}\sigma_{2j}^{(k+1)}, \boldsymbol{\phi}_j^{(k+1)})$ in Step 3 by maximizing

$$\sum_{t=1}^{n}(1-w_{tj}^{(k)})\log\varphi(X_t;\mu(\zeta_1,\boldsymbol{\phi};X_{t-1}^p),\sigma_1) + \sum_{t=1}^{n} w_{tj}^{(k)}\log\varphi(X_t;\mu(\zeta_2,\boldsymbol{\phi};X_{t-1}^p),\sigma_2),$$

where $w_{tj}^{(k)}$ will be described presently, simultaneously over $Z^2 \times [\delta,\infty)^2 \times \boldsymbol{\Phi}$.

**Step 0.** Choose $0 < \alpha_1 < \alpha_2 < \ldots < \alpha_J = 0.5$. Compute

$$(\widehat{\zeta}_0, \widehat{\sigma}_0, \widehat{\boldsymbol{\phi}}_0) = \arg\max_{\zeta,\sigma,\boldsymbol{\phi}} pl_n(0.5,\zeta,\zeta,\sigma,\sigma,\boldsymbol{\phi}).$$

Put $j = 1$ and $k = 0$.

**Step 1.** Put $\alpha_j^{(k)} = \alpha_j$.

**Step 2.** Compute

$$(\zeta_{1j}^{(k)}, \zeta_{2j}^{(k)}, \sigma_{1j}^{(k)}, \sigma_{2j}^{(k)}, \boldsymbol{\phi}_j^{(k)}) = \arg\max_{\zeta_1,\zeta_2,\sigma_1,\sigma_2,\boldsymbol{\phi}} pl_n(\alpha_j^{(k)},\zeta_1,\zeta_2,\sigma_1,\sigma_2,\boldsymbol{\phi})$$

TABLE 11: BIC and AIC for the corresponding models for monthly returns of IBM stock

| **BIC** | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ |
|---|---|---|---|---|---|
| $p = 1$ | 5891.56 | 5812.57 | **5799.38** | 5911.92 | 5805.84 |
| $p = 2$ | 5898.25 | 5825.86 | 5805.96 | 5918.60 | 5812.42 |
| $p = 3$ | 5904.65 | 5838.91 | 5812.71 | 5925.00 | 5819.17 |
| $p = 4$ | 5911.08 | 5849.97 | 5817.83 | 5931.43 | 5824.29 |

| **AIC** | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ |
|---|---|---|---|---|---|
| $p = 1$ | 5877.21 | 5774.30 | **5770.67** | 5883.21 | 5772.35 |
| $p = 2$ | 5879.11 | 5778.01 | 5772.47 | 5885.11 | 5774.15 |
| $p = 3$ | 5880.73 | 5781.50 | 5774.43 | 5886.73 | 5776.11 |
| $p = 4$ | 5882.37 | 5782.99 | 5774.77 | 5888.37 | 5776.45 |

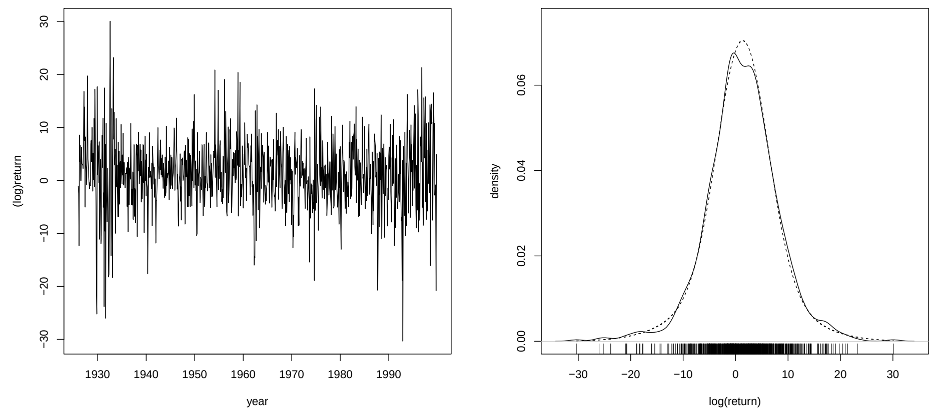


FIGURE 1: Monthly log returns (in % and including dividends) for IBM stock from January 1926 to December 1999 (left). A kernel density estimate (solid line) of the monthly log returns for IBM stock (right) together with the density of a fitted two-component normal mixture model (dashed line) and of a single normal (dotted line).

and

$$M_n^{(k)}(\alpha_j) = 2\left\{pl_n(\alpha_j^{(k)}, \zeta_{1j}^{(k)}, \zeta_{2j}^{(k)}, \sigma_{1j}^{(k)}, \sigma_{2j}^{(k)}, \boldsymbol{\phi}_j^{(k)}) - pl_n(0.5, \widehat{\zeta}_0, \widehat{\zeta}_0, \widehat{\sigma}_0, \widehat{\sigma}_0, \widehat{\boldsymbol{\phi}}_0)\right\}.$$

**Step 3.** Compute for $t = 1, \ldots, n$ the weights

$$w_{tj}^{(k)} = \frac{\alpha_j^{(k)}\varphi(X_t; \mu(\zeta_{2j}^{(k)}, \boldsymbol{\phi}_j^{(k)}; X_{t-1}^p), \sigma_{1j}^{(k)})}{(1-\alpha_j^{(k)})\varphi(X_t; \mu(\zeta_{1j}^{(k)}, \boldsymbol{\phi}_j^{(k)}; X_{t-1}^p), \sigma_{1j}^{(k)}) + \alpha_j^{(k)}\varphi(X_t; \mu(\zeta_{2j}^{(k)}, \boldsymbol{\phi}_j^{(k)}; X_{t-1}^p), \sigma_{2j}^{(k)})}.$$

Compute the estimators

$$\alpha_j^{(k+1)} = \arg\max_{\alpha}\left((n-\sum_{t=1}^{n} w_{tj}^{(k)})\log(1-\alpha)+\sum_{t=1}^{n} w_{tj}^{(k)}\log(\alpha)+p(\alpha)\right)$$

$$\zeta_{1j}^{(k+1)} = \frac{\sum_{t=1}^{n}(1-w_{tj}^{(k)})(X_t-\sum_{\tau=1}^{p}\phi_{\tau j}^{(k)}X_{t-\tau})}{\sum_{t=1}^{n}(1-w_{tj}^{(k)})}$$

$$\zeta_{2j}^{(k+1)} = \frac{\sum_{t=1}^{n} w_{tj}^{(k)}(X_t-\sum_{\tau=1}^{p}\phi_{\tau j}^{(k)}X_{t-\tau})}{\sum_{t=1}^{n} w_{tj}^{(k)}}$$

$$\boldsymbol{\phi}_j^{(k+1)} = \arg\max_{\boldsymbol{\phi}}\left(\sum_{t=1}^{n}(1-w_{tj}^{(k)})\log\varphi(X_t;\mu(\zeta_{1j}^{(k+1)},\boldsymbol{\phi};X_{t-1}^p),\sigma_{1j}^{(k)}) + \sum_{t=1}^{n} w_{tj}^{(k)}\log\varphi(X_t;\mu(\zeta_{2j}^{(k+1)},\boldsymbol{\phi};X_{t-1}^p),\sigma_{2j}^{(k)})\right)$$

$$\sigma_{1j}^{(k+1)} = \arg\max_{\sigma_1}\sum_{t=1}^{n}(1-w_{tj}^{(k)})\log\varphi(X_t;\mu(\zeta_{1j}^{(k+1)},\boldsymbol{\phi}_j^{(k+1)};X_{t-1}^p),\sigma_1)$$

$$\sigma_{2j}^{(k+1)} = \arg\max_{\sigma_2}\sum_{t=1}^{n} w_{tj}^{(k)}\log\varphi(X_t;\mu(\zeta_{2j}^{(k+1)},\boldsymbol{\phi}_j^{(k+1)};X_{t-1}^p),\sigma_2).$$

Compute

$$M_n^{(k+1)}(\alpha_j) = 2\left\{pl_n(\alpha_j^{(k+1)},\zeta_{1j}^{(k+1)},\zeta_{2j}^{(k+1)},\sigma_{1j}^{(k+1)},\sigma_{2j}^{(k+1)},\boldsymbol{\phi}_j^{(k+1)}) - pl_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\sigma}_0,\widehat{\sigma}_0,\widehat{\boldsymbol{\phi}}_0)\right\},$$

put $k=k+1$ and repeat Step 3 for a fixed number of iterations $K$.

**Step 4.** Put $j=j+1$, $k=0$ and go to Step 1, until $j=J$.

**Step 5.** Compute the test statistic

$$EM_n^{(K)}(\mathcal{J}) = \max\left\{M_n^{(K)}(\alpha_j), j=1,\dots,J\right\}.$$

## Sketch of the proof of Theorem 1.

We now provide a sketch of the proof of Theorem 1., where we focus on those parts which particularly involve the autoregressive parameters and require changes as compared to the arguments in Chen & Li (2009). For complete details we refer to Ketterer & Holzmann (2012). We shall need some additional lemmas. For notational convenience we write $(\widehat{\zeta}_1,\widehat{\zeta}_2,\widehat{\sigma}_1,\widehat{\sigma}_2,\widehat{\boldsymbol{\phi}})$ instead of $(\widehat{\zeta}_{1,\alpha_j},\widehat{\zeta}_{2,\alpha_j},\widehat{\sigma}_{1,\alpha_j},\widehat{\sigma}_{2,\alpha_j},\widehat{\boldsymbol{\phi}}_{\alpha_j})$ for the maximizer of $l_n(\alpha,\zeta_1,\zeta_2,\sigma_1,\sigma_2,\boldsymbol{\phi})$ subject to $\alpha=\alpha_j$, $\alpha_j\in\mathcal{J}$.

**Lemma 2.** *For each given $\alpha_j\in(0,0.5]$ we have under the null model*

$$\widehat{\sigma}_1-\sigma_0=o_P(1),\quad \widehat{\sigma}_2-\sigma_0=o_P(1),\quad \widehat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0=o_P(1),$$
$$\widehat{\zeta}_1-\zeta_0=o_P(1),\quad \widehat{\zeta}_2-\zeta_0=o_P(1).$$

From now on, we concentrate on linear switching autoregressive models of order 1, and write $\phi=\phi_1$ for the corresponding autoregressive coefficient. At the end of the proof we indicate the necessary changes for higher orders. Without loss of generality we assume $\zeta_0=0$ and $\sigma_0=1$. Note here that we can assume that $(\widehat{\zeta}_1,\widehat{\zeta}_2,\widehat{\sigma}_1,\widehat{\sigma}_2,\widehat{\phi})$ are in a small neighorhood of $(0,0,1,1,\phi_0)$ by Lemma 2.

In a first step give a stochastic upper bound for

$$2\Big(l_n(\alpha_j,\widehat{\zeta}_1,\widehat{\zeta}_2,\widehat{\sigma}_1,\widehat{\sigma}_2,\widehat{\phi})-l_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\sigma}_0,\widehat{\sigma}_0,\widehat{\phi}_0)\Big) = r_{1n}(\alpha_j)+r_{2n}, \tag{9}$$

where

$$r_{1n}(\alpha_j) = r_{1n}(\alpha_j,\widehat{\zeta}_1,\widehat{\zeta}_2,\widehat{\sigma}_1,\widehat{\sigma}_2,\widehat{\phi}) = 2\Big(l_n(\alpha_j,\widehat{\zeta}_1,\widehat{\zeta}_2,\widehat{\sigma}_1,\widehat{\sigma}_2,\widehat{\phi})-l_n(0.5,0,0,1,1,\phi_0)\Big),$$
$$r_{2n} = 2\Big(l_n(0.5,0,0,1,1,\phi_0)-l_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\sigma}_0,\widehat{\sigma}_0,\widehat{\phi}_0)\Big).$$

Let $r_{1n}(\alpha_j) = 2\sum_{t=1}^n \log(1+\widehat{\delta}_t)$ with

$$\widehat{\delta}_t = (1-\alpha_j)\left\{\frac{\varphi(X_t;\mu(\widehat{\zeta}_1,\widehat{\phi};X_{t-1}),\widehat{\sigma}_1)}{\varphi(X_t;\mu(0,\phi_0;X_{t-1}),1)} - 1\right\} + \alpha_j\left\{\frac{\varphi(X_t;\mu(\widehat{\zeta}_2,\widehat{\phi};X_{t-1}),\widehat{\sigma}_2)}{\varphi(X_t;\mu(0,\phi_0;X_{t-1}),1)} - 1\right\}.$$

Using the inequality $\log(1+x) \le x - (1/2)x^2 + (1/3)x^3$ leads to

$$r_{1n}(\alpha_j) \le 2\sum_{t=1}^n \widehat{\delta}_t - \sum_{t=1}^n \widehat{\delta}_t^2 + (2/3)\sum_{t=1}^n \widehat{\delta}_t^3. \tag{10}$$

For $0 \le l, s, i \le 4$ we define

$$\widehat{m}_{l,s,i} = (1-\alpha_j)\widehat{\zeta}_1^l(\widehat{\sigma}_1^2 - 1)^s(\widehat{\phi}-\phi_0)^i + \alpha_j\widehat{\zeta}_2^l(\widehat{\sigma}_2^2-1)^s(\widehat{\phi}-\phi_0)^i.$$

Denoting

$$\partial_\zeta^l\partial_{\sigma^2}^s\partial_\phi^i\varphi(X_t;(\mu(0,\phi_0;X_{t-1}),1) = \left.\frac{\partial^{l+s+i}\varphi(X_t;\mu(\zeta,\phi;X_{t-1}),\sigma)}{\partial^l\zeta\partial^s(\sigma^2)\partial^i\phi}\right|_{(\zeta,\sigma,\phi)=(0,1,\phi_0)}$$

and expanding $\varphi(X_t;\mu(\widehat{\zeta}_h,\widehat{\phi};X_{t-1}),\widehat{\sigma}_h)$, $h=1,2$, up to order 4, we get

$$\widehat{\delta}_t = \sum_{l+s+i=1}^4 \frac{\widehat{m}_{l,s,i}}{l!s!i!}\frac{\partial_\zeta^l\partial_{\sigma^2}^s\partial_\phi^i\varphi(X_t;\mu(0,\phi_0;X_{t-1}),1)}{\varphi(X_t;\mu(0,\phi_0;X_{t-1}),1)} + \widehat{\epsilon}_{tn}^{(1)} \tag{11}$$

with remainder $\widehat{\epsilon}_{tn}^{(1)}$ for which

$$\widehat{\epsilon}_n^{(1)} := \sum_{t=1}^n \widehat{\epsilon}_{tn}^{(1)} = O_P(n^{1/2})\left\{\sum_{h=1}^2\sum_{\substack{l+s+i=5\\ l,s,i\ge 0}} |\widehat{\zeta}_h|^l|\widehat{\sigma}_h^2-1|^s|\widehat{\phi}-\phi_0|^i\right\}. \tag{12}$$

This is due to the CLT for stationary and ergodic martingale differences. Following the assessment in Chen & Li (2009), this can be simplified to

$$\widehat{\epsilon}_n^{(1)} = O_P(n^{1/2})\sum_{h=1}^2\{|\widehat{\zeta}_h|^5 + |\widehat{\zeta}_h|^3|\widehat{\sigma}_h^2-1| + |\widehat{\sigma}_h^2-1|^3 + (\widehat{\phi}-\phi_0)^2\}$$

By Lemma 2. we can incorporate the terms $\widehat{m}_{l,s,i}$ with $l+2s+4i \ge 5$ into the remainder term, e.g.

$$O_P(n^{1/2})\widehat{m}_{i,j,2} = O_P(n^{1/2})|\widehat{\zeta}_h|^i|\widehat{\sigma}_h^2-1|^j(\widehat{\phi}-\phi_0)^2 = O_P(n^{1/2})(\widehat{\phi}-\phi_0)^2 \tag{13}$$

for $i,j \ge 0$. Altogether, we have

$$\widehat{\delta}_t = \sum_{l+2s+4i=1}^4 \frac{\widehat{m}_{l,s,i}}{l!s!i!}\frac{\partial_\zeta^l\partial_{\sigma^2}^s\partial_\phi^i\varphi(X_t;\mu(0,\phi_0,X_{t-1}),1)}{\varphi(X_t;\mu(0,\phi_0;X_{t-1}),1)} + \widehat{\epsilon}_{tn} \tag{14}$$

with remainder $\widehat{\epsilon}_{tn}$ satisfying

$$\begin{aligned}\widehat{\epsilon}_n = \sum_{t=1}^n \widehat{\epsilon}_{tn} &= O_P(n^{1/2})\sum_{h=1}^2\{|\widehat{\zeta}_h|^5 + |\widehat{\zeta}_h|(\widehat{\sigma}_h^2-1)^2 + |\widehat{\sigma}_h^2-1|^3 + (\widehat{\phi}-\phi_0)^2\\ &\quad + |\widehat{\zeta}_h||\widehat{\phi}-\phi_0| + |\widehat{\sigma}_h^2-1||\widehat{\phi}-\phi_0|\}.\end{aligned} \tag{15}$$

We define

$$Y_t := \left.\frac{\frac{\partial}{\partial\zeta}\varphi(X_t;\mu(\zeta,\phi_0;X_{t-1}),1)}{\varphi(X_t;\mu(0,\phi_0;X_{t-1}),1)}\right|_{\zeta=0} = \epsilon_t,$$

$$Z_t := \frac{1}{2}\left.\frac{\frac{\partial^2}{\partial^2\zeta}\varphi(X_t;\mu(\zeta,\phi_0;X_{t-1}),1)}{\varphi(X_t;\mu(0,\phi_0;X_{t-1}),1)}\right|_{\zeta=0} = (\epsilon_t^2-1)/2,$$

$$U_t := \frac{1}{6}\left.\frac{\frac{\partial^3}{\partial^3\zeta}\varphi(X_t;\mu(\zeta,\phi_0;X_{t-1}),1)}{\varphi(X_t;\mu(0,\phi_0;X_{t-1}),1)}\right|_{\zeta=0} = (\epsilon_t^3-3\epsilon_t)/6,$$

$$V_t := \frac{1}{24}\left.\frac{\frac{\partial^4}{\partial^4\zeta}\varphi(X_t;\mu(\zeta,\phi_0;X_{t-1}),1)}{\varphi(X_t;\mu(0,\phi_0;X_{t-1}),1)}\right|_{\zeta=0} = (\epsilon_t^4-6\epsilon_t^2+3)/24$$

and

$$W_t := \left.\frac{\frac{\partial}{\partial\phi}\varphi(X_t;\mu(0,\phi;X_{t-1}),1)}{\varphi(X_t;\mu(0,\phi_0;X_{t-1}),1)}\right|_{\phi=\phi_0} = X_{t-1}Y_t = X_{t-1}\epsilon_t.$$

Therefore,

$$\widehat{\delta}_t = \widehat{t}_1Y_t + \widehat{t}_2Z_t + \widehat{t}_3U_t + \widehat{t}_4V_t + \widehat{t}_5W_t + \widehat{\epsilon}_{tn} \tag{16}$$

with $\sum_{t=1}^n \widehat{\epsilon}_{tn}$ satisfying (15) and

$$\begin{aligned}\widehat{t}_1 &= \widehat{m}_{1,0,0}, \quad \widehat{t}_2 = \widehat{m}_{2,0,0}+\widehat{m}_{0,1,0}, \quad \widehat{t}_3 = \widehat{m}_{3,0,0}+3\widehat{m}_{1,1,0},\\ \widehat{t}_4 &= \widehat{m}_{4,0,0}+6\widehat{m}_{2,1,0}+3\widehat{m}_{0,2,0}, \quad \widehat{t}_5 = \widehat{m}_{0,0,1} = \widehat{\phi}-\phi_0.\end{aligned} \tag{17}$$

Putting $\widehat{\delta}_t$ into (10) and noting that the remainders from the square and cubic terms on the right side of the following equation are of the same or higher order than the remainder $\widehat{\epsilon}_n$ from the linear sum, we get

$$\begin{aligned} r_{1n}(\alpha_j) \le\, & 2\{\widehat{t}_1\sum_{t=1}^n Y_t + \widehat{t}_2\sum_{t=1}^n Z_t + \widehat{t}_3\sum_{t=1}^n U_t + \widehat{t}_4\sum_{t=1}^n V_t + \widehat{t}_5\sum_{t=1}^n W_t\}\\ & - \{\widehat{t}_1\sum_{t=1}^n Y_t + \widehat{t}_2\sum_{t=1}^n Z_t + \widehat{t}_3\sum_{t=1}^n U_t + \widehat{t}_4\sum_{t=1}^n V_t + \widehat{t}_5\sum_{t=1}^n W_t\}^2\{1+o_P(1)\}\\ & + O_P(\widehat{\epsilon}_n).\end{aligned} \tag{18}$$

In order to further bound the remainder term $\widehat{\epsilon}_n$, we need

**Lemma 3.** *Under the conditions of Lemma 2. and the null model we have for* $h=1,2$,

$$\widehat{\zeta}_h^5 = o_P\Big(\sum_{l=1}^5|\widehat{t}_l|\Big), \quad \widehat{\zeta}_h(\widehat{\sigma}_h^2-1)^2 = o_P\Big(\sum_{l=1}^5|\widehat{t}_l|\Big) \text{ and } (\widehat{\sigma}_h^2-1)^3 = o_P\Big(\sum_{l=1}^5|\widehat{t}_l|\Big).$$

*as well as*

$$\begin{aligned}\widehat{t}_3 &= 3\frac{1-\alpha_j}{\alpha_j}\left\{\widehat{\zeta}_1\widehat{\beta}_1 - \frac{2(2\alpha_j-1)}{3\alpha_j}\widehat{\zeta}_1^3\right\} + o_P(|\widehat{t}_1|) + o_P(|\widehat{t}_2|),\\ \widehat{t}_4 &= 3\frac{1-\alpha_j}{\alpha_j}\left\{\widehat{\beta}_1^2 - \frac{2(1-3\alpha_j+3\alpha_j^2)}{3\alpha_j^2}\widehat{\zeta}_1^4\right\} + o_P(|\widehat{t}_1|) + o_P(|\widehat{t}_2|).\end{aligned} \tag{19}$$

For the proof, see the proof of Lemma 2 in Ketterer & Holzmann (2012).

Now, Lemma 2. gives $(\widehat{\phi}-\phi_0)^2 = o_P(|\widehat{\phi}-\phi_0|) = o_P(|\widehat{t}_1|) = o_P\Big(\sum_{l=1}^5|\widehat{t}_l|\Big)$ and by the same reasoning $|\widehat{\zeta}_h||\widehat{\phi}-\phi_0| = o_P\Big(\sum_{l=1}^5|\widehat{t}_l|\Big)$ and $|\widehat{\sigma}_h^2-1||\widehat{\phi}-\phi_0| = o_P\Big(\sum_{l=1}^5|\widehat{t}_l|\Big)$, $h=1,2$. Using the inequalities $|x|\le 1+x^2$ and $(a+b)^2\le 2(a^2+b^2)$, $a,b\ge 0$ repeatedly, we get

$$\widehat{\epsilon}_n^{1/2} o_P(\sum_{l=1}^5|\widehat{t}_l|) = o_P(1) + o_P(n)\Big\{\sum_{l=1}^5\widehat{t}_l^2\Big\}. \tag{20}$$

From the above, Equation (18) and the fact that $Y_t$, $Z_t$, $U_t$, $V_t$ and $W_t$ are mutually orthogonal, we obtain

$$\begin{aligned} r_{1n}(\alpha_j) \leq\, & 2\widehat{t}_1 \sum_{t=1}^n Y_t - \widehat{t}_1^2 \sum_{t=1}^n Y_t^2\{1+o_p(1)\} \\ & + 2\widehat{t}_2 \sum_{t=1}^n Z_t - \widehat{t}_2^2 \sum_{t=1}^n Z_t^2\{1+o_p(1)\} \\ & + 2\widehat{t}_3 \sum_{t=1}^n U_t - \widehat{t}_3^2 \sum_{t=1}^n U_t^2\{1+o_p(1)\} \\ & + 2\widehat{t}_4 \sum_{t=1}^n V_t - \widehat{t}_4^2 \sum_{t=1}^n V_t^2\{1+o_p(1)\} \\ & + 2\widehat{t}_5 \sum_{t=1}^n W_t - \widehat{t}_5^2 \sum_{t=1}^n W_t^2\{1+o_p(1)\} + o_P(1). \end{aligned} \tag{21}$$

Using the properties of quadratic functions we have

$$r_{1n}(\alpha_j) \leq \frac{(\sum_{t=1}^n Y_t)^2}{\sum_{t=1}^n Y_t^2} + \frac{(\sum_{t=1}^n Z_t)^2}{\sum_{t=1}^n Z_t^2} + \frac{(\sum_{t=1}^n U_t)^2}{\sum_{t=1}^n U_t^2} + \frac{(\sum_{t=1}^n V_t)^2}{\sum_{t=1}^n V_t^2} + \frac{(\sum_{t=1}^n W_t)^2}{\sum_{t=1}^n W_t^2} + o_P(1). \tag{22}$$

For linear switching autoregressive models of order higher than 1, the derivatives

$$\left. \frac{\frac{\partial}{\partial \phi_j}\varphi(X_t;\mu(0,\boldsymbol{\phi};X_{t-1}),1)}{\varphi(X_t;\mu(0,\boldsymbol{\phi}_0;X_{t-1}),1)} \right|_{\boldsymbol{\phi}=\boldsymbol{\phi}_0} = X_{t-j}Y_t = X_{t-j}\epsilon_t$$

are not mutually orthogonal but orthogonal to $Y_t$, $Z_t$, $U_t$ and $V_t$. Therefore, we have to orthogonalize to obtain an asymptotic upper bound for $r_{1n}(\alpha_j)$ of the form (21) in this case.

*Proof of Theorem 1.* Since

$$\begin{aligned} & 2\{l_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\sigma}_0,\widehat{\sigma}_0,\widehat{\phi}_0) - l_n(0.5,0,0,1,1,\phi_0)\} \\ & = \frac{(\sum_{t=1}^n Y_t)^2}{\sum_{t=1}^n Y_t^2} + \frac{(\sum_{t=1}^n Z_t)^2}{\sum_{t=1}^n Z_t^2} + \frac{(\sum_{t=1}^n W_t)^2}{\sum_{t=1}^n W_t^2} + o_P(1), \end{aligned} \tag{23}$$

we get for any given $\alpha_j \in \mathcal{J}$

$$\begin{aligned} & 2\{l_n(\alpha_j,\widehat{\zeta}_1,\widehat{\zeta}_2,\widehat{\sigma}_1,\widehat{\sigma}_2,\widehat{\phi}) - l_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\sigma}_0,\widehat{\sigma}_0,\widehat{\phi}_0)\} \\ & \qquad \leq \frac{(\sum_{t=1}^n U_t)^2}{\sum_{t=1}^n U_t^2} + \frac{(\sum_{t=1}^n V_t)^2}{\sum_{t=1}^n V_t^2} + o_P(1). \end{aligned} \tag{24}$$

implying that the $\chi_2^2$ distribution serves as a stochastic upper bound for our test statistic $R_n(\mathcal{J})$. Choosing $\tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\sigma}_1, \tilde{\sigma}_2$ and $\tilde{\phi}$ which are implicitly given by

$$\hat{t}_1 = \frac{\sum_{t=1}^n Y_t}{\sum_{t=1}^n Y_t^2},\ \hat{t}_2 = \frac{\sum_{t=1}^n Z_t}{\sum_{t=1}^n Z_t^2},\ \hat{t}_3 = \frac{\sum_{t=1}^n U_t}{\sum_{t=1}^n U_t^2},\ \hat{t}_4 = \frac{\sum_{t=1}^n V_t}{\sum_{t=1}^n V_t^2},\ \hat{t}_5 = \frac{\sum_{t=1}^n W_t}{\sum_{t=1}^n W_t^2},$$

this upper bound is also attained for $1/2 \in \mathcal{J}$. A detailed proof of this result is given in the technical supplement Ketterer & Holzmann (2012). ■