Empirical likelihood for break detection in time series

Roberto Baragona, Francesco Battaglia and Domenico Cucina

Department of Statistics, Sapienza University of Rome
Piazzale Aldo Moro 5, 00100 Rome, Italy
e-mail: roberto.baragona@uniroma1.it*;* francesco.battaglia@uniroma1.it*;*
domenico.cucina@uniroma1.it

Abstract: Structural breaks have become an important research topic in time series analysis and in many fields of application, e.g. econometrics, hydrology, seismology, engineering and industry, chemometrics, and medicine. In most phenomena encountered in the real world modeling assuming no change of structure may obviously lead to inconsistent estimates and poor forecasts. Much research work has been carried out in the past decades and efforts mostly concentrated on the identification of points in time where structural breaks possibly occur and on the development of statistical tests of absence of breaks. Common procedures are the CUSUM based statistics and the F based statistics proposed by Bai and Perron in several papers and technical reports. The need of specification of a probability distribution for conducting the statistical tests has been often a drawback in practical applications. If the distribution is misspecified the test may turn to be unreliable. In this paper a distribution free procedure for identification of change points and testing for structural breaks is proposed based on the empirical likelihood and the empirical likelihood ratio statistics. A comparison with the F test and the CUSUM test is performed on both simulated and real world data. The size and power of the test is comparable with existing procedures and the identification procedure is generally accurate and effective.

AMS 2000 subject classifications: Primary 62M10; secondary 62G30.
Keywords and phrases: Estimating equations, level shift, likelihood ratio test, structural change.

Contents

1. Introduction

The analysis of structural breaks, or change points, has attracted a great deal of interest in many fields, and a vast literature is available about this topic. For a comprehensive review, see e. g. [2, 35]. Independent data, regression models and time series have been considered, with the aim of detecting whether one or more structural changes are present in the data, and estimating their dates and the parameters characterizing the different data subsets.

Structural break detection is usually based on hypothesis tests, where the null hypothesis states absence of breaks, and the alternative specifies one or more breaks. Proposed test statistics are likelihood ratios, Lagrange multipliers or Wald statistics, and are therefore essentially based on the sum of squared residuals under the two hypotheses.

In recent years a new framework has been proposed for similar test problems, and related confidence intervals, known as empirical likelihood [e.g. 34]. It has the advantage of being essentially non-parametric in nature, not requiring any strong assumption on the probability distribution of the data. In this paper we propose the use of empirical likelihood methods for addressing the problem of detection and estimation of level changes and structural breaks both in independent and dependent data, and compare this proposal with the most popular alternative, the multiple structural change analysis in linear models of [4].

The plan of the paper is as follows. Section 2 briefly reviews the structural change analysis literature for independent data and linear models. Section 3 reviews the empirical likelihood methods relevant to our problem. Section 4 is concerned with the use of empirical likelihood for detecting level changes in independent data. In Section 5 an empirical likelihood based method for detecting structural breaks in regression and autoregression models is proposed. Section 6 discusses some implementation issues, Section 7 presents some simulation results and applications, and Section 8 concludes.

2. Structural change

We begin with level shifts, i. e., changes in the mean. Let $y_1, y_2, \ldots, y_n$ denote a sample of independent data with common variance σ^2, and a level shift at an unknown date q such that the series is broken in two sub-samples with different means: $E(y_i) = \mu, i = 1, 2, \ldots, q$ and $E(y_i) = \mu + \omega, i = q + 1, \ldots, n$. We want to test the null hypothesis of constant mean: $H_0 : \omega = 0$. The most used tool in this framework is the so-called CUSUM process [see e.g. 15]. Assuming that the break date is a fixed proportion of the number of observations: $q = \lfloor nx \rfloor$, the CUSUM is a scaled difference between the observed average of the two subsets:

$$z_n(x) = \frac{1}{\sigma \sqrt{n}} \left(\sum_{i=1}^{\lfloor nx \rfloor} y_i - \frac{\lfloor nx \rfloor}{n} \sum_{j=1}^{n} y_j \right) \quad x \in [0, 1]. \tag{2.1}$$

As $n \to \infty$, $z_n(x)$ converges to a standard Brownian Bridge $B(x)$ under H_0. If we consider the maximum of the CUSUM statistic on any possible splitting of

the series in two parts:

$$M_n = \max_{1 \le q \le n} z_n(q/n)$$

the statistic M_n under H_0 converges in distribution to that of $\sup\{B(x), 0 \le x \le 1\}$, that has been tabulated by several authors [e.g. 39].

However, since the variance of $z_n(q/n)$ is $[\frac{q}{n}(1 - \frac{q}{n})]$, change points occurring near the beginning or the end of the data series are much more difficult to detect, and a weighted version has been suggested. Moreover, in the case of Gaussian distribution the LR test statistic for a level change at time q is easily seen to be

$$-2\log\Lambda_q = z_n(q/n)^2 \left[\frac{q}{n}\left(1 - \frac{q}{n}\right)\right]^{-1} \tag{2.2}$$

thus one may consider to use the adjusted CUSUM (2.2) rather than the original one. A supplementary problem arises since on writing

$$T_n(x)^2 = z_n(x)^2 [x(1-x)]^{-1}$$

we have that $\max\{T_n(x)^2, 0 \le x \le 1\}$ diverges to infinity as $n \to \infty$: this may be solved considering only a truncated interval $\{\epsilon \le x \le 1 - \epsilon\}$, which amounts to allowing only subseries with at least $h = n\epsilon$ observations. With this limitation, it may be shown that

$$\max_{n\epsilon \le q \le n(1-\epsilon)} T_n(q/n)^2 \xrightarrow{d} \sup_{\epsilon \le x \le 1-\epsilon} \frac{B(x)^2}{x(1-x)}.$$

It has been shown that a similar result holds also for dependent data under some assumptions [see 2, and references therein], with σ replaced by the (square root of the) so-called long-run variance

$$v = \lim_{n \to \infty} \frac{1}{n}\text{Var}\left(\sum_{i=1}^{n} y_i\right).$$

Obviously, to compute the CUSUM statistic in practice, a consistent estimate of the variance σ^2, or v, is needed, and this may be difficult in some dependent case, see e.g. [38].

The case when $\{y_t\}$ is a causal linear process:

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j u_{t-j}$$

where $\{u_t\}$ is a white noise process, is addressed by [37]. They fit an ARMA model to the data and consider the adjusted CUSUM statistic computed on the observed residuals, and find that its performance is generally better than the CUSUM computed directly on the data. Also (like the independent data case) the LR and F tests are asymptotically equivalent to the CUSUM test, and their finite sample behavior in some simulated AR(p) series is essentially equivalent.

Suppose now that the data $\{y_i\}$ is generated by a linear regression model

$$y_i = \sum_{j=1}^{p} \theta_j x_{ij} + \varepsilon_i \qquad (2.3)$$

where the parameters θ_j may be subject to change. This framework was addressed in several papers by Andrews, Bai, Perron and co-authors [e.g. 1, 4]. Here the change involves not only the mean, but the whole regression structure, and is called a structural change.

Consider a partition of the whole series into $(m+1)$ subseries $\{y_1, \ldots, y_{t_1}\}$, $\{y_{t_1+1}, \ldots, y_{t_2}\}, \ldots, \{y_{t_m+1}, \ldots, y_n\}$ and suppose that for each subseries the regression parameters are different, denoting by $\theta^{(k)}$ those of the $k-$th subseries. Then, on writing the corresponding model equation for each observation, and rearranging, we have:

$$y = \bar{X}\theta + \varepsilon \qquad (2.4)$$

where $\theta = (\theta^{(1)'}, \theta^{(2)'}, \ldots, \theta^{(m+1)'})'$, $y = (y_1, y_2, \ldots, y_n)'$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ and $\bar{X}$ is the $n \times (m+1)p$ corresponding expanded design matrix:

$$\bar{X} = \begin{pmatrix} X^{(1)} & 0 & \ldots & 0 \\ 0 & X^{(2)} & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & X^{(m+1)} \end{pmatrix} \qquad (2.5)$$

and $X^{(k)} = \{x_{ij}, i = t_{k-1} + 1, \ldots, t_k; \ j = 1, \ldots, p\}$, with $t_0 = 0, t_{m+1} = n$.

The hypothesis of absence of breaks may be phrased in terms of the parameters as $H_0 : \theta_j^{(k)} = \theta_j^{(k+1)}, j = 1, \ldots, p; \ k = 1, \ldots, m$ and expressed as a linear constraint $R\theta = 0$. Therefore, such hypothesis may be tested via the usual F test:

$$F(t_1, t_2, \ldots, t_m) = \frac{\|\bar{X}(\hat{\theta} - \hat{\hat{\theta}})\|^2}{\|y - \bar{X}\hat{\theta}\|^2} \frac{n - (m+1)p}{mp} \qquad (2.6)$$

where $\hat{\theta} = (\bar{X}'\bar{X})^{-1}\bar{X}'y$ is the unrestricted least squares estimate, and $\hat{\hat{\theta}}$ is the least squares estimate conditional to $R\theta = 0$ so that:

$$\hat{\theta} - \hat{\hat{\theta}} = (\bar{X}'\bar{X})^{-1}R'\{R(\bar{X}'\bar{X})^{-1}R'\}^{-1}R\hat{\theta}.$$

In the case of a level shift in independent data with only one break, it may be readily seen that the statistic $F(q/n)$ is equivalent to the adjusted CUSUM (2.2).

Since the break dates are unknown, one may consider the maximum value of (2.6) on all possible partitions of the series into $(m+1)$ subseries. Again, to avoid divergence we have to limit to partitions where each subseries has at least $h = n\epsilon$ observations, with $\epsilon > 0$. To reduce computations, Bai and Perron [5] have proposed an efficient algorithm, based on the principle of dynamic programming, which requires at most $O(n^2)$ least-squares operations. This algorithm will be extended in Section 6 to the use of empirical likelihood.

The asymptotic distribution of the resulting $\sup F$ statistic has been derived in [4] under some regularity assumptions, and is a generalization of the sup Brownian Bridge. It depends on ϵ, m and p and tabulated quantiles by simulation may be found in [6].

A partial structural break is also considered by [4], where only some of the regression parameters are allowed to change, while the others are constrained to remain equal in each regime. To be specific, denote the regression parameter vector by $\theta = (\beta', \delta')'$ where $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ are the first p parameters that are assumed constant, and $\delta = (\delta_1, \delta_2, \dots, \delta_q)'$ are the last q regression parameters, which are subject to change, and partition the design matrix into blocks: $X = (W, Z)$, where $W : (n \times p)$ corresponds to the β parameters, and $Z : (n \times q)$ to the last q parameters δ. Then, given any partition into subseries induced by m breaks, the analogous model to (2.4) may be written:

$$y = W\beta + \bar{Z}\bar{\delta} + \varepsilon \qquad (2.7)$$

where $\bar{\delta} = (\delta^{(1)'}, \delta^{(2)'}, \dots, \delta^{(m+1)'})'$ and $\bar{Z}$ is the $n \times (m+1)q$ matrix corresponding to the expansion of Z, analogue to (2.5).

The hypothesis of absence of breaks may be tested using the F statistic related to the linear restrictions on the parameters $\bar{\delta}$ specifying that $\delta_j^{(k)} = \delta_j^{(k+1)}$, $j = 1, \dots, q$; $k = 1, \dots, m$, expressed as a linear system $S\bar{\delta} = 0$. Then the conditional F statistic is:

$$F_c(t_1, t_2, \dots, t_m) = \frac{\|W(\hat{\beta} - \hat{\hat{\beta}}) + \bar{Z}(\hat{\delta} - \hat{\hat{\delta}})\|^2}{\|y - W\hat{\beta} - \bar{Z}\hat{\delta}\|^2} \frac{n - (m+1)q - p}{mq}. \qquad (2.8)$$

where $\hat{\beta}, \hat{\delta}$ are the unrestricted LS estimates of $(\beta, \bar{\delta})$, and $(\hat{\hat{\beta}}, \hat{\hat{\delta}})$ are the estimates under the hypothesis $S\bar{\delta} = 0$. In this case also, to avoid divergence we restrict to the cases when the subseries have at least $h = n\epsilon$ observations, and compute the maximum of (2.8) on any segmentation of the series into $(m+1)$ segments. For such problem also [5] proposed a dynamic programming algorithm, that is however more complicated than in the previous case. The resulting $\sup F_c$ statistic is used to test the null hypothesis of no break. Its asymptotic distribution has been derived by [4] and is similar to the complete structural break (2.6) but depends now on m, ϵ and q (the number of regression parameters subject to change).

Several other methods for identifying and estimating change points (including e.g. Bayesian methods, least absolute deviation and *LASSO* procedures) are discussed in [20].

3. Empirical likelihood

Empirical likelihood methods were originally derived for building confidence intervals for the mean of a random sample, without assumptions on the form of the probability distribution. They have been later extended to cover inference

on parameters of linear models and time series. A comprehensive introduction is [34]. Further applications of empirical likelihood have been recently proposed for many statistical problems, among them for example generalized linear models [23, 40], conditional heteroskedasticity models [11, 17, 12], unstable autoregressive models [14], and long-memory time series [41].

Consider an observed random sample $(y_1, y_2, \ldots, y_n)$ from the random variable y with $E(y) = \mu$ and $\text{Var}(y) = \sigma^2$. That sample may be thought of as drawn from a discrete probability distribution concentrated only at the values $y_1, y_2, \ldots, y_n$ with probabilities p_i $(p_i > 0, \sum p_i = 1)$, called the empirical distribution. In this case the probability of getting exactly that sample would be $p_1 p_2 \ldots p_n$. The mean of that distribution is $\mu = \sum_i p_i y_i$, therefore, given μ, the possible discrete distributions generating the observed sample are the set $\{(p_1, p_2, \ldots, p_n) : p_i > 0, \sum p_i = 1, \sum p_i y_i = \mu\}$ and the largest probability to get the observed sample, given μ, is

$$EL(\mu) = \max \left\{ \prod_{i=1}^{n} p_i \ : \ p_i > 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i y_i = \mu \right\}. \qquad (3.1)$$

Expression 3.1 as a function of μ is called the empirical likelihood profile for μ. The maximum is reached for $\mu = \bar{y} = \sum y_i / n$ and equals n^{-n}, corresponding to $p_i = 1/n$, $i = 1, \ldots, n$. Thus, $\bar{y}$ is the maximum empirical likelihood estimator of μ. The empirical likelihood ratio is obtained dividing (3.1) by its maximum n^{-n}:

$$ELR(\mu) = \max \left\{ \prod_{i=1}^{n} (np_i) \ : \ p_i > 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i y_i = \mu \right\}. \qquad (3.2)$$

The solution of the maximum problem (3.2) may be obtained through Lagrange multipliers:

$$\hat{p}_i = \{n[1 + \lambda(y_i - \mu)]\}^{-1}$$

where the multiplier $\lambda \equiv \lambda(\mu)$ is the solution of

$$\sum_{i=1}^{n} \frac{y_i - \mu}{1 + \lambda(y_i - \mu)} = 0.$$

If we define the empirical log likelihood ratio as $\ell(\mu) = -2 \log\{ELR(\mu)\}$, then

$$\ell(\mu) = 2 \sum_{i=1}^{n} \log\{1 + \lambda(y_i - \mu)\}$$

and [32] has shown that if μ_0 is the true mean of y, then as $n \to \infty$, $\ell(\mu_0)$ converges in distribution to a χ_1^2 variable, a result similar to the parametric likelihood. Therefore $\ell(\mu)$ may be used for defining asymptotic confidence intervals, and for hypothesis testing, for the mean.

A similar result holds also for random vectors. The following theorem is proved in [34].

Theorem 3.1 (Owen [34])**.** *Let* $y_1, y_2, \ldots, y_n$ *be i. i. d. random vectors with mean* μ_0 *and finite variance-covariance matrix* V_0 *of rank* d*. Then* $-2 \log ELR(\mu_0)$ *converges in distribution to a* χ_d^2 *as* $n \to \infty$*.*

The key step in the proof is to show that $-2 \log ELR(\mu_0)$ may be written as

$$-2 \log ELR(\mu_0) = n(\bar{y} - \mu_0)' S^{-1}(\bar{y} - \mu_0) + o_p(1)$$

where $\bar{y} = \frac{1}{n}(y_1 + y_2 + \cdots + y_n)$ and $S = \frac{1}{n} \sum_{i=1}^{n}(y_i - \mu_0)(y_i - \mu_0)'$.

The empirical likelihood framework has been generalized to linear regression and analysis of variance [33] and also to data generated by a statistical parametric model through estimating equations [36, 29, 34]. Let $\theta = (\theta_1, \ldots, \theta_p)' \in \Theta$ and suppose that the relationship between the parameters θ and the data is summarized by r estimating equations $E\{g_k(y, \theta)\} = 0$, $k = 1, 2, \ldots, r$ that determine θ uniquely. Usually $r = p$ but the case $r > p$ has also been discussed in [36]. The empirical likelihood ratio is defined by analogy:

$$ELR(\theta)$$
$$= \max \left\{ \prod_{i=1}^{n} (np_i) \, : \, p_i > 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i g_k(y_i, \theta) = 0, \, k = 1, 2, \ldots, p \right\}. \tag{3.3}$$

It may be shown that

$$\ell(\theta) = -2 \log ELR(\theta) = 2 \sum_{i=1}^{n} \log \left\{ 1 + \sum_{k=1}^{p} \lambda_k g_k(y_i, \theta) \right\}$$

where the multipliers $\lambda_k \equiv \lambda_k(\theta)$ are the solutions of

$$\sum_{i=1}^{n} \frac{g_j(y_i, \theta)}{1 + \sum_{k=1}^{p} \lambda_k g_k(y_i, \theta)} = 0, \, j = 1, 2, \ldots, p.$$

Finally, if θ_0 denotes the true parameter value, then $\ell(\theta_0)$ converges in distribution, as $n \to \infty$, to a χ_p^2 under some regularity conditions (see Theorem 5.1 below), whose most important one is that the matrix S with entries $S_{ij} = E\{g_i(y, \theta_0)g_j(y, \theta_0)\}$ be of full rank. It may be shown [see 36] that, with $g(y, \theta) = \{g_1(y, \theta), g_2(y, \theta), \ldots, g_p(y, \theta)\}'$,

$$\ell(\theta_0) = n \left[\frac{1}{n} \sum_{i=1}^{n} g(y_i, \theta_0) \right]' \left[\frac{1}{n} \sum_{i=1}^{n} g(y_i, \theta_0)g(y_i, \theta_0)' \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^{n} g(y_i, \theta_0) \right] + o_p(1).$$

In the case of linear regression model (2.3), that will be of special interest here, the estimating functions g are provided by the normal equations:

$$g_k(y_i, \theta_0) = x_{ik} \left(y_i - \sum_{j=1}^{p} x_{ij} \theta_j^0 \right) = x_{ik} \varepsilon_i. \tag{3.4}$$

Then $\sum_{i=1}^{n} g(y_i, \theta_0) = X'\varepsilon$, $\frac{1}{n}\sum_{i=1}^{n} g(y_i, \theta_0)g(y_i, \theta_0)' - \frac{1}{n}\sigma^2(X'X) = o_p(1)$, and

$$\ell(\theta_0) = \frac{1}{\sigma^2}\,\varepsilon'X(X'X)^{-1}X'\varepsilon + o_p(1)$$

from which the distribution result follows easily.

Since the maximum of $ELR(\theta)$ is obtained in $p_1 = p_2 = \cdots = p_n = 1/n$, the LS estimator $\hat{\theta} = (X'X)^{-1}X'y$ is also the maximum empirical likelihood estimator, and $\ell(\hat{\theta}) = 0$.

It follows that an empirical likelihood ratio test for the hypothesis $H_0 : \theta = \theta_0$ may be based on the statistic $\ell(\theta_0)$, which under H_0 is asymptotically chi square distributed with p degrees of freedom.

A partial ELR test has also been developed by [36, Corollary 5]. Suppose that $\theta = (\theta_1', \theta_2')'$, where θ_1 has q components and θ_2 has p components. Let $(\hat{\theta}_1, \hat{\theta}_2)$ be the minimizer of ℓ, and $\hat{\theta}_2^0$ denote the minimizer of $\ell(\theta_1^0, \theta_2)$ with respect to θ_2. Then, the statistic $\lambda(\theta_1^0) = \ell(\theta_1^0, \hat{\theta}_2^0) - \ell(\hat{\theta}_1, \hat{\theta}_2)$ may be employed for testing the hypothesis $H_0 : \theta_1 = \theta_1^0$, since it may be shown that under H_0, $\lambda(\theta_1^0)$ converges in distribution to a χ_q^2.

Such results have been extended to autoregressive time series by [29]: if

$$y_t = \sum_{i=1}^{p} y_{t-i}\theta_i + \varepsilon_t$$

where the ε's are martingale increments with constant variance, then in (3.4) $x_{ij} = y_{i-j}$ and $\sum_{i=1}^{n} g(y_i, \theta)$ is a martingale and the convergence to chi square may be shown, see e. g. [14] and Chapter 8 of [34].

The empirical likelihood framework was proposed for testing the difference of two samples [21, 25], and for the difference of regression coefficients in two samples [44]. Less attention in literature has received the use of empirical likelihood for detecting change points. To the best of our knowledge, the first proposal is [45] who consider an independent sample with one change in some distribution moments at unknown time, and [26] where a regression model is considered with one possible change in the parameter values at unknown point. The present paper is an extension to those contributions in three directions: allowing more than one break, considering parametric models defined by estimating equations and autoregressions in addition to regression models, and including partial structural breaks.

Finally, related contributions are [18, 42, 19, 16]. The first paper proposed empirical likelihood for detecting one change point in a sample where the distribution before change is F, and that after change is $G(x) \propto \int_{-\infty}^{y} w(y, \theta)dF(y)$ where $w(\cdot, \theta)$ is a known weight function depending on a parameter θ. The other papers [42, 19, 16] deal with partially time-varying regression models:

$$y_i = X_i'\beta + Z_i'\alpha(t_i) + \varepsilon_i$$

possibly with errors in variables, where the time-varying coefficients $\alpha(t)$ are assumed continuous and differentiable, and propose empirical likelihood confidence intervals for β.

4. Testing level changes with empirical likelihood

Let us consider first the simple case of a random sample $(y_1, \ldots, y_n)$ with a single level shift at a known time q, assuming $E(y_i) = \mu_1, (i = 1, \ldots, q), E(y_i) = \mu_2, (i = q + 1, \ldots, n),$ $\mathrm{Var}(y_i) = \sigma^2$, all i and independent y_i's. The empirical likelihood profile is

$$\pi(m_1, m_2) = \max\left\{\prod_{i=1}^{n} w_i \ : \ w_i > 0,\right.$$
$$\left.\sum_{i=1}^{n} w_i = 1, \ \sum_{i=1}^{q} w_i(y_i - m_1) = 0, \ \sum_{i=q+1}^{n} w_i(y_i - m_2) = 0\right\}$$

and the maximum is n^{-n} obtained for $m_1 = \bar{y}_1, m_2 = \bar{y}_2$ (the averages of the two sub-samples). Therefore the maximum empirical likelihood under the hypothesis $H_0 : \mu_1 = \mu_2 = \mu$ is obtained as $\pi(\mu, \mu)$ and the empirical likelihood ratio test statistic is $-2 \log \pi(\mu, \mu) - 2n \log n$ and under H_0, keeping q/n fixed, converges as $n \to \infty$ to a central chi square variable with one degree of freedom [see e.g. 21]. The computation of $\pi(\mu, \mu)$ is simple because it equals the sum of the empirical likelihood computed separately on the two sub-series $(y_1, \ldots, y_q)$ and $(y_{q+1}, \ldots, y_n)$. We state this result in the more general case of m breaks.

Let $1 < t_1 < t_2 < \cdots < t_m < n$ denote the break dates (and let $t_0 = 0$, $t_{m+1} = n$), and assume that $E(y_k) = \mu_i$ for $t_{i-1} + 1 \le k \le t_i$ $(i = 1, \ldots, m+1)$, while the y_k's are independent with equal variance σ^2. We denote by

$$ELR(\mu, s, t) = \max\left\{\prod_{i=s+1}^{t} [(t - s)w_i] \ : \ \sum_{i=s+1}^{t} w_i = 1, \ \sum_{i=s+1}^{t} w_i(y_i - \mu) = 0\right\}$$

the empirical likelihood ratio for the sub-series $(y_{s+1}, \ldots, y_t)$.

Theorem 4.1. *The empirical likelihood ratio test statistic for the hypothesis* $H_0 : \mu_1 = \mu_2 = \cdots = \mu_{m+1} = \mu_0$ *is given by*

$$-2\left\{\sum_{k=0}^{m} \log ELR(\mu_0, t_k, t_{k+1})\right\}. \tag{4.1}$$

If $t_i/n \to \rho_i$ $(i = 1, \ldots, m)$ as $n \to \infty$ with $0 < \rho_1 < \rho_2 < \cdots < \rho_m < 1$, then as $n \to \infty$, under H_0, the statistic converges to a central chi square distribution with $m + 1$ degrees of freedom.

Then, the hypothesis of absence of breaks may be tested using the inf of (4.1) with respect to the mean:

$$\ell_0(t_1, t_2, \ldots, t_m) = -2 \sup_{\mu}\left\{\sum_{k=0}^{m} \log ELR(\mu, t_k, t_{k+1})\right\}. \tag{4.2}$$

The statistic (4.2) is equal to the empirical likelihood test statistic for an analysis of variance model with one factor, and the null hypothesis of equal mean for each group, thus under H_0 is asymptotically chi square distributed with m degrees of freedom [see 33, p. 1738].

This idea is very similar to the CUSUM framework, and indeed leads to asymptotically equal results, as detailed in the following Corollary.

Corollary 4.1. *In the case of a test of the null hypothesis of absence of breaks against an alternative of a single break at time q, if $q/n \to \rho$ as $n \to \infty$ with $0 < \rho < 1$ and $E|y|^4 < \infty$ the difference between the ELR test statistic (4.2) and the adjusted CUSUM statistic (2.2) converges to zero in probability as $n \to \infty$ under H_0.*

The problem of multiple level changes may be also obviously expressed in terms of a regression model (2.3) with only one parameter ($p = 1$) and design matrix coefficients $x_{i1} = 1 \forall i$. The hypothesis of absence of breaks may be tested using the Bai and Perron method. In this case also an asymptotic equivalence holds.

Corollary 4.2. *For a test of the null hypothesis of absence of breaks against an alternative of m breaks at dates t_j, if $t_i/n \to \rho_i$ $(i = 1, \ldots, m)$ as $n \to \infty$ with $0 < \rho_1 < \rho_2 < \cdots < \rho_m < 1$ and $E|y|^4 < \infty$ the difference between the ELR test statistic (4.2) and m times the F test statistic (2.6) converges to zero in probability as $n \to \infty$ under H_0.*

Given the asymptotic equivalence of the test statistics, the empirical likelihood-based test for breaks may be accomplished by the same strategy as proposed by [4]:

1. Select the number of breaks m, and a minimum length of the resulting sub-series, $h = n\epsilon$
2. Compute the maximum of the ELR test statistic

$$\ell_0(t_1, \ldots, t_m) = -2 \sup_\mu \sum_{j=0}^{m} \log ELR(\mu, t_j, t_{j+1})$$

   on all partitions of the series into $(m+1)$ sub-series $(y_1, \ldots, y_{t_1}), (y_{t_1+1}, \ldots, y_{t_2}), \ldots, (y_{t_m+1}, \ldots, y_n)$ with at least h observations each, call it $\ell_*(m)$.

Under the null hypothesis of absence of breaks, the statistic $\ell_*(m)$ has the same asymptotic distribution as the Bai and Perron's $F(m|0)$ test, derived in Proposition 6 of [4], and quantiles for selected values of ϵ are derived by simulation in [6]. For the alternative of only one break, the asymptotic distribution is that of the sup of a Brownian bridge, and the theoretical approximated quantiles of [37] may be considered.

In the case of scalar y_i's further insight into the nature of the difference between the statistics ELR and F may be easily gained. Given the segmentation $1 < t_1 < t_2 < \cdots < t_m < n$, and denoting by n_i the length of the $i-$th subseries, with $\bar{y}_i$ its average and $\bar{y}$ the overall average, the Bai and Perron statistic

F equals

$$\frac{1}{m} \sum_{i=1}^{m+1} n_i(\bar{y}_i - \bar{y})^2/\hat{\sigma}^2$$

where $\hat{\sigma}^2$ is an estimate of the variance under the hypothesis of m level changes:

$$\hat{\sigma}^2 = \frac{1}{n-m-1} \sum_{i=1}^{m+1} \sum_{j=t_{i-1}+1}^{t_i} (y_j - \bar{y}_i)^2.$$

On the other side, for each subseries $\{y_{t_{i-1}+1}, \ldots, y_{t_i}\}$ the ELR ratio tends asymptotically to a quadratic form:

$$-2\log ELR(\mu, t_{i-1}, t_i) = n_i(\bar{y}_i - \mu)^2/S_i(\mu) + o_p(1)$$

where

$$S_i(\mu) = \frac{1}{n_i} \sum_{j=t_{i-1}+1}^{t_i} (y_j - \mu)^2$$

is the variance of the $i-$th sub-series under the hypothesis of absence of breaks, usually larger than $\hat{\sigma}^2$. Therefore one can expect that the ELR test statistic tends to be smaller than the corresponding F test statistic, a closer value might be obtained with a weighted ELR:

$$\ell_W(t_1, \ldots, t_m) = -2 \sup_{\mu} \left\{ \sum_{j=1}^{m+1} \log ELR(\mu, t_{j-1}, t_j) S_j(\mu)/\hat{\sigma}^2 \right\}$$

which is asymptotically equal to

$$\inf_{\mu} \sum_{i=1}^{m+1} n_i(\bar{y}_i - \mu)^2/\hat{\sigma}^2.$$

In this case the optimal value of μ is simply $\bar{y}$; for the unweighted ELR, and for vector y_i's, the search is more complicated and will be discussed in Section 6.

When only one break is allowed, our procedure is equivalent to that proposed in [45], but computations are different. The method of [45] is based on solving a non linear equations system, while ℓ_0 in (4.2) requires evaluation of ELR on two sub-series, and may be computed using the standard empirical likelihood routines.

5. Testing structural breaks with empirical likelihood

We consider now the case that the data $\{y_1, \ldots, y_n\}$ is generated by a model described by the estimating equations:

$$E\{g_k(y, \theta)\} = 0, \ k = 1, 2, \ldots, p$$

where $\theta = (\theta_1, \ldots, \theta_p)' \in \Theta$ are the model parameters, and may be subject to m breaks at times $1 < t_1 < \cdots < t_m < n$, so that each $i-$th subseries $(y_{t_{i-1}+1}, \ldots, y_{t_i})$ is generated by a model with parameters $\theta^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_p^{(i)})'$, while under the null hypothesis of absence of breaks all the observations are generated by the same model with parameters $\theta_0 = (\theta_1^0, \ldots, \theta_p^0)'$. We denote again the ELR statistic for the sub-series $(y_{s+1}, \ldots, y_t)$ by

$$ELR(\theta, s, t) = \max\left\{ \prod_{i=s+1}^{t} [(t-s)w_i] \ : \ w_i \geq 0, \sum_{i=s+1}^{t} w_i = 1, \sum_{i=s+1}^{t} w_i g(y_i, \theta) = 0 \right\}.$$

The empirical likelihood test statistic for the null hypothesis of absence of breaks may be computed as the sum of the ELR statistic on each sub-series, as stated in the following Theorem.

Theorem 5.1. *Under the following assumptions:*

1. *$E\{g(y, \theta_0)g(y, \theta_0)'\}$ positive definite*
2. *$\frac{\partial}{\partial \theta} g(y, \theta)$ continuous in θ_0, $\|\frac{\partial}{\partial \theta} g(y, \theta)\|$ and $\|g(y, \theta)\|^2$ bounded by an integrable function*
3. *$\frac{\partial^2}{\partial \theta \partial \theta'} g(y, \theta)$ continuous in θ_0 and $\|\frac{\partial^2}{\partial \theta \partial \theta'} g(y, \theta)\|$ bounded by an integrable function*

the empirical likelihood ratio test statistic for the hypothesis $H_0 : \theta^{(1)} = \theta^{(2)} = \cdots = \theta^{(m+1)} = \theta_0$ is

$$-2\left\{ \sum_{k=0}^{m} \log ELR(\theta_0, t_k, t_{k+1}) \right\} \tag{5.1}$$

and if $t_i/n \to \rho_i$ $(i = 1, \ldots, m)$ as $n \to \infty$ with $0 < \rho_1 < \rho_2 < \cdots < \rho_m < 1$, then as $n \to \infty$, under H_0, the statistic converges to a central chi square distribution with $(m+1)p$ degrees of freedom.

The null hypothesis of absence of breaks: $H_0 : \theta^{(1)} = \theta^{(2)} = \cdots = \theta^{(m+1)}$ may be tested considering the sup of the ELR statistic under H_0:

$$\ell_0(t_1, \ldots, t_m) = -2 \sup_{\theta \in \Theta} \sum_{j=0}^{m} \log ELR(\theta, t_j, t_{j+1}). \tag{5.2}$$

Here Corollary 5 of [36] (discussed at the end of Section 3) may be applied, on taking for the free parameter (previously denoted by θ_2) the regression coefficient of the first sub-series $\theta^{(1)}$, and for the constrained parameter (previously denoted by θ_1) the differences $\theta^{(2)} - \theta^{(1)}, \theta^{(3)} - \theta^{(1)}, \ldots, \theta^{(m+1)} - \theta^{(1)}$. The hypothesis H_0 corresponds to $\theta_1 = \theta_1^0 = 0$, and from the Corollary, the statistic (5.2) is under H_0 asymptotically chi square distributed with mp degrees of freedom.

The simplest and most interesting case is when the generating model is linear:

$$y_i = \sum_{j=1}^{p} \theta_j x_{ij} + \varepsilon_i \tag{5.3}$$

where we assume that the ε_i's are independent from the x_{ij} and mutually independent, but only mild boundedness conditions for the x_{ij} are required. The estimating functions are given by the left hand sides of the normal equations:

$$g_k(y_i, \theta) = x_{ik} \left(y_i - \sum x_{ij} \theta_j \right)$$

therefore in vector form, and writing $\varepsilon_i(\theta) = y_i - \sum x_{ij} \theta_j$:

$$g(y_i, \theta) = \varepsilon_i(\theta) X_i'$$

where X_i is the $i-$th row of the design matrix X. Note that, with $\varepsilon(\theta) = [\varepsilon_1(\theta), \ldots, \varepsilon_n(\theta)]'$,

$$\sum_{i=1}^n g(y_i, \theta) = X' \varepsilon(\theta)$$

$$\left[\sum_{i=1}^n g(y_i, \theta) g(y_i, \theta)' \right]_{r,s} = \sum_{i=1}^n \varepsilon_i(\theta)^2 x_{ir} x_{is} = [G(\theta)]_{r,s}$$

and if θ_0 is the true parameter value, $\frac{1}{n} G(\theta_0) - \frac{1}{n} \sigma^2(X'X) = o_p(1)$. Thus, the asymptotic value of the ELR ratio at the parameter value θ_0 is

$$-2 \log ELR(\theta_0, 0, n)$$
$$= \left[\sum_{i=1}^n g(y_i, \theta_0) \right]' \left[\sum_{i=1}^n g(y_i, \theta_0) g(y_i, \theta_0)' \right]^{-1} \left[\sum_{i=1}^n g(y_i, \theta_0) \right] + o_p(1)$$
$$= \varepsilon(\theta_0)' X (X'X)^{-1} X' \varepsilon(\theta_0) \sigma^{-2} + o_p(1).$$

When only one break is allowed, our procedure is equivalent to that proposed in [26], but again the computation strategy is different.

The asymptotic similarity between empirical and Gaussian likelihood ratios implies an asymptotic relationship between the ELR ratio test statistic and the Bai and Perron's F statistic also for linear models:

Corollary 5.1. *If the data is generated by model (5.3) and the assumptions of Theorem 5.1 are satisfied, for a test of the null hypothesis of absence of breaks against an alternative of m breaks at dates t_j, if $t_i/n \to \rho_i$ ($i = 1, \ldots, m$) as $n \to \infty$ with $0 < \rho_1 < \rho_2 < \cdots < \rho_m < 1$, the difference between the ELR ratio test statistic (5.2) and mp times the F test statistic (2.6) converges to zero in probability under H_0.*

For linear models too, an essential difference, for finite n, between ELR and F resides in the evaluation of the residual variance, which is computed under the null hypothesis in ELR, and under the alternative in F. However, here a weighted version for reducing such difference is not viable, since the residuals appear inside the matrix $G(\theta)$ to be inverted.

Corollary 5.1 suggests that similar strategies as those proposed by [4] may be employed for testing breaks using the empirical likelihood:

1. Select a maximum number of breaks m, and a minimum length of the sub-series $h = n\epsilon$
2. Compute the maximum of the ELR ratio test statistic (5.2) on all possible partitions. Under H_0 the maximum statistic has the same asymptotic distribution of the Bai and Perron's $F(m|0)$ statistic, and their quantiles may be used.

More elaborate strategies, also proposed in [5], and the important issue of maximizing $ELR(\theta)$ with respect to θ and with respect to the break dates, are discussed in Section 6.

Let us address now the problem of structural breaks for autoregressive processes. Consider an $AR(p)$ model:

$$y_t = \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \varepsilon_t. \tag{5.4}$$

The empirical likelihood for these processes has been studied by [14] extending results on dual likelihood of [29]. Since (5.4) is a linear model it may be put in the form of a regression model (2.3) with $x_{ij} = y_{i-j}$, and letting $z_t = (y_t, y_{t-1}, \ldots, y_{t-p})'$ the estimating equations may be written

$$g_k(z_t, \theta) = y_{t-k}\left(y_t - \sum_{j=1}^{p} \theta_j y_{t-j}\right)$$

so that, if θ_0 is the vector of the true parameter values, $g(z_t, \theta_0) = \varepsilon_t(y_{t-1}, \ldots, y_{t-p})'$ and an empirical likelihood ratio analogue to (3.3) may be defined:

$$ELR(\theta)$$
$$= \max\left\{\prod_{i=1}^{n}(np_i) \,:\, p_i > 0, \sum_{i=1}^{n} p_i = 1, \sum_{t=1}^{n} p_t g_k(z_t, \theta) = 0, \, k = 1, 2, \ldots, p\right\}. \tag{5.5}$$

The results of [36] cannot be directly applied since the $g(z_t, \theta)$ are not independent, but [14] have shown that under the following assumptions:

AR1 The process $\{\varepsilon_t\}$ forms a martingale difference with respect to the sequence of $\sigma-$fields $\mathfrak{F}_t = \{\ldots, \varepsilon_{t-2}, \varepsilon_{t-1}, \ldots, y_{t-2}, y_{t-1}\}$

AR2 $E\{\varepsilon_t^2|\mathfrak{F}_t\} = \sigma^2$ for every t, $E\{\varepsilon_t^{(4+c)}|\mathfrak{F}_t\} < \infty$ for a $c > 0$, the initial values $y_0, \ldots, y_{1-p}$ are $\mathfrak{F}_t-$measurable

AR3 All roots of the autoregressive polynomial $\theta(z) = 1 - \theta_1^0 z - \cdots - \theta_p^0 z^p$ lie outside the unit circle

the empirical likelihood ratio $ELR(\theta_0)$ is asymptotically equivalent to

$$\sum_{t=p+1}^{n} g(z_t, \theta_0)' \left[\sum_{t=p+1}^{n} g(z_t, \theta_0)g(z_t, \theta_0)'\right]^{-1} \sum_{t=p+1}^{n} g(z_t, \theta_0)$$

and converges in distribution to a chi square with p degrees of freedom, like in the case of independent data.

Turning now to the structural change analysis, Bai and Perron's [4] $\sup F$ test may be also applied to autoregressive processes under some additional assumptions, see [4, page 57] and [3, page 302]. Let us denote any partition of the data into $m + 1$ sub-series by $\{(y_{t_i+1}, \ldots, y_{t_{i+1}}), i = 0, 1, \ldots, m\}$ with $t_0 = 0$ and $t_{m+1} = n$ as in Sections 4 and 5, and let $n_i = t_{i+1} - t_i$; the assumptions are as follows:

AR4 $\frac{1}{n_i} \sum_{t=t_i+1}^{t_i+[vn_i]} z_t z_t'$ converges in probability to vQ_i uniformly in $v \in [0, 1]$ where Q_i is a positive definite matrix, $i = 0, 1, \ldots, m$

AR5 $\frac{1}{\sqrt{n_i}} \sum_{t=t_i+1}^{t_i+[vn_i]} \varepsilon_t z_t$ converges in distribution to $W_i(v)$, where $W_i(v)$ is a vector Wiener process in $[0, 1]$ with variance $\sigma^2 vQ_i$, and $W_0, W_1, \ldots, W_m$ are independent of each other.

The above assumptions are sufficient for extending to autoregressive processes our results concerning the asymptotic distribution of the ELR test statistic and the asymptotic equivalence between the $\sup F$ test and the supremum of the ELR statistic under H_0:

$$\ell_0(t_1, \ldots, t_m) = -2 \sup_{\theta \in \Theta} \sum_{j=0}^{m} \log ELR(\theta, t_j, t_{j+1}), \qquad (5.6)$$

as stated in the following Theorem.

Theorem 5.2. *For an $AR(p)$ process with parameters θ_0 and under assumptions AR1–AR5, the empirical likelihood ratio test statistic for the hypothesis $H_0 : \theta^{(1)} = \theta^{(2)} = \cdots = \theta^{(m+1)} = \theta_0$ is*

$$-2 \left\{ \sum_{k=0}^{m} \log ELR(\theta_0, t_k, t_{k+1}) \right\}$$

and if $t_i/n \to \rho_i$ $(i = 1, \ldots, m)$ as $n \to \infty$ with $0 < \rho_1 < \rho_2 < \cdots < \rho_m < 1$

  (i) *as $n \to \infty$, under H_0, the statistic converges to a central chi square distribution with $(m + 1)p$ degrees of freedom*

 (ii) *for a test of the null hypothesis of absence of breaks against an alternative of m breaks at dates t_i, the difference between the ELR ratio test statistic (5.6) and mp times the F test statistic (2.6) converges to zero in probability under H_0.*

5.1. Partial structural breaks

We address here the hypothesis that the breaks involve only some parameters, while the others remain unchanged throughout the whole dataset. However, we believe that only very rarely one may have a priori information sufficient

to exclude that some parameters of the linear model are subject to possible breaks. Thus it is advisable, to start with, testing for a complete structural break. Only if the hypothesis of absence of breaks is rejected, and given the best specification of the number of breaks m, and their dates $t_1, \ldots, t_m$, we proceed to examine the hypothesis that some of the model parameters remain constant, selecting them among those whose estimates are most variable throughout the $(m + 1)$ sub-series. Such a procedure has the advantage of requiring only one maximization stage with respect to the break dates, and only for the complete break structure.

We shall assume, without loss of generality, that the possibly constant parameters are the first p, while there are q more parameters (labeled from $p + 1$ to $p + q$) that are subject to m breaks at dates $t_1, t_2, \ldots, t_m$.

Consider the regression model analogous to (5.3): $y = X\theta + \varepsilon$, where now θ has $(p + q)$ parameters. Partition the design matrix as in Section 2: $X = (W, Z)$ where $W : n \times p$ corresponds to the first p parameters, and $Z : n \times q$ corresponds to the last q parameters, and denote $\theta = (\beta', \delta')'$ where $\beta : p \times 1$ relates to W, and $\delta : q \times 1$ relates to Z. Now, as in (2.4), we consider a partition of the whole series into $(m + 1)$ sub-series delimited by the break dates, and write the expanded model:

$$y = \bar{X}\bar{\theta} + \varepsilon$$

where $\bar{\theta} = (\theta^{(1)'}, \theta^{(2)'}, \ldots, \theta^{(m+1)'})'$, and $\bar{X}$ is the corresponding expanded design matrix. The partial break model, assuming that the first p parameters remain fixed, may be written as in (2.7), where $\bar{Z}$ is the $n \times (m + 1)$ matrix corresponding to the expansion of Z, similar to (2.5).

We first consider the no break hypothesis against the complete structural break using the statistic

$$\ell_* = -2 \sup_{\theta \in \Theta} \max_{t_1, \ldots, t_m} \sum_{j=0}^{m} ELR(\theta, t_j, t_{j+1}).$$

If the null hypothesis is rejected, and given m and the dates $t_1, \ldots, t_m$, an empirical likelihood ratio statistic for the partial break hypothesis may be obtained maximizing the sum of the EL ratios:

$$\ell_1 = -2 \sup_{\bar{\theta} \in \Theta_1} \sum_{j=0}^{m} ELR(\theta^{(j+1)}, t_j, t_{j+1})$$

on the set

$$\Theta_1 = \{\bar{\theta} = (\theta^{(1)'}, \theta^{(2)'}, \ldots, \theta^{(m+1)'})' \in \Theta^{m+1} : \theta_j^{(k)} = \theta_j^{(k+1)},$$
$$k = 1, \ldots, m; \, j = 1, \ldots, p\}.$$

The results of the previous Section imply that, under the hypothesis of no break, and given the dates, ℓ_1 is asymptotically chi square distributed with mp

degrees of freedom. The results concerning partial empirical likelihood of [36, Corollary 5], referred to in Section 3, also ensure that $\ell_* - \ell_1$ is asymptotically distributed as a chi square with mq degrees of freedom under the hypothesis that the first p parameters remain constant. This difference may be used as a test statistic for the partial break, rejecting, if it is significant, the hypothesis that the change does not involve the first p parameters.

A similar strategy may be adopted also when detecting breaks with F-type statistics. Three sub-spaces may be defined:

- S: the space generated by the columns of $\bar{X}$ without restrictions; let $\hat{\theta}$ denote the least squares estimate, and therefore $P(y) = \bar{X}\hat{\theta}$ is the projection of the data onto S.
- S_1: the space generated by the columns of $\bar{X}$ with the restriction that the first p coefficients are constant. S_1 is isomorphic to the space generated by the columns of $(W, \bar{Z})$. We denote by $\tilde{\theta}$ the least squares estimate under the conditions $\{\theta_j^{(k)} = \theta_j^{(k+1)}, j = 1, 2, \ldots, p; k = 1, 2, \ldots, m\}$, thus the projection of y onto S_1 is $P_1(y) = \bar{X}\tilde{\theta}$.
- S_0: the space generated by the columns of $\bar{X}$ with the restriction that all coefficients remain constant, i. e., absence of breaks. S_0 is isomorphic to the space generated by the columns of X. Let θ^* denote the least squares estimate in this case, subject to $\{\theta_j^{(k)} = \theta_j^{(k+1)}, j = 1, 2, \ldots, p + q; k = 1, 2, \ldots, m\}$, so that the projection of y onto S_0 is $P_0(y) = \bar{X}\theta^*$.

Since $S_0 \subset S_1 \subset S$ we have:

$$
\begin{aligned}
\|y - P(y)\|^2 &= \|y - P_0(y)\|^2 + \|P(y) - P_0(y)\|^2 \\
\|P(y) - P_0(y)\|^2 &= \|P(y) - P_1(y)\|^2 + \|P_1(y) - P_0(y)\|^2.
\end{aligned}
$$

In the framework of the Gaussian likelihood test theory, the hypothesis of absence of breaks against a complete structural break may be tested using the statistic

$$
F(t_1, \ldots, t_m) = \frac{\|P(y) - P_0(y)\|^2}{\|y - P(y)\|^2} \frac{n - (m+1)(p+q)}{m(p+q)}.
$$

This is the Bai and Perron statistic for the complete structural break. The hypothesis that the first p parameters remain constant in the $(m+1)$ sub-series could be tested using the statistic

$$
F_1(t_1, \ldots, t_m) = \frac{\|P(y) - P_1(y)\|^2}{\|y - P(y)\|^2} \frac{n - (m+1)(p+q)}{mp}.
$$

The Bai and Perron F test statistic for the partial structural break is easily seen to be

$$
F_c(t_1, \ldots, t_m) = \frac{\|P_1(y) - P_0(y)\|^2}{\|y - P_1(y)\|^2} \frac{n - (m+1)q - p}{mq} \tag{5.7}
$$

and may be seen as a conditional test of absence of breaks given that the first p parameters are not subject to change.

The hypothesis of partial break, i. e., that the change does not involve the first p parameters, may be rejected if the statistic F_c (5.7) is large, compared with the quantiles of a F variable with mq and $n - (m+1)q - p$ degrees of freedom. In this case also, our procedure has the advantage of requiring the search of the number of breaks m, and their dates $t_1, \ldots, t_m$, only for a complete break structure, which is much simpler and easier than maximizing the partial break F statistic.

6. Implementation of the empirical likelihood-based procedure

We start noting a boundedness issue, that is typical of the empirical likelihood framework, and does not arise in the procedures based on squared residuals. Consider the level change case. The statistic for testing no level change against m changes at times $t_1, \ldots, t_m$ is (4.2) and is the sum of the logarithms of the partial empirical likelihoods:

$$ELR(\mu, t_k, t_{k+1}) = \max \left\{ \prod_{j=t_k+1}^{t_{k+1}} (t_{k+1} - t_k)w_j : \right.$$

$$\left. w_j > 0, \sum_{j=t_k+1}^{t_{k+1}} w_j = 1, \sum_{j=t_k+1}^{t_{k+1}} w_j(y_j - \mu) = 0 \right\}. \quad (6.1)$$

Their definition depends on the existence of weights w_j satisfying the constraints in (6.1) for each k, in other words on the existence of a number μ belonging to the convex hull of each sub-series $(y_{t_k+1}, \ldots, y_{t_k}), k = 0, 1, \ldots, m$. If this does not happen, at least one of the ratios (6.1) is not defined, and may be conventionally taken equal to zero [e. g. 34], thus the test statistic (4.2) becomes arbitrarily large. It implies that if the data contains any sub-series whose values are all larger than the remaining observations, the null hypothesis of no level change will be rejected at any significance level; this does not happen with the CUSUM or F-type tests, unless the difference is significantly large. In our view, this is a distinctive feature of the empirical likelihood, and not necessarily a shortcoming. However, to avoid such problem, the definition itself of empirical likelihood may be slightly modified in order to ensure the existence of a solution with all positive weights in any case. An adjusted empirical likelihood is proposed in [13], consisting in the addition of an artificial observation with the aim of satisfying the constraint. Suppose that the series is $(y_1, \ldots, y_n)$ and the empirical likelihood is expressed in terms of the estimating functions $g(y, \theta)$. Then we add an artificial observation y_{n+1}, and define

$$g(y_{n+1}, \theta) = -\frac{a_n}{n} \sum_{i=1}^{n} g(y_i, \theta) \qquad (6.2)$$

and instead of the empirical likelihood ratio $ELR(\theta)$ (3.3) we consider the adjusted empirical likelihood ratio:

$$ELR^*(\theta)$$
$$= \max \left\{ \prod_{i=1}^{n+1} (n+1)p_i \ : \ p_i > 0, \sum_{i=1}^{n+1} p_i = 1, \sum_{i=1}^{n+1} p_i g_k(y_i, \theta) = 0, \ k = 1, 2, \ldots, p \right\}.$$

The behavior of the adjusted ELR and the choice of the term a_n are discussed in [13], showing asymptotic coincidence if $a_n = o_p(\sqrt{n})$. The authors suggest to choose $a_n = \max\{1, \frac{1}{2}\log n\}$ and the use of a truncated mean in (6.2). In our experience we found such adjustment satisfactorily effective, and the difference with $ELR(\theta)$ negligible when the hypothesis of absence of breaks is true.

A further issue is Bartlett correctability. A Bartlett correction may be applied to the empirical likelihood [see e. g. Chapter 13 of 34], both in the case of a level change and for structural change in linear models, resulting essentially in dividing the ELR by $(1+A/n)$. However, the appropriate value to be used for A may be difficult to select; a rough approximation based on Gaussian distribution is $A = d^2 + d/2$ if the data, or the estimating functions, belong to $\mathbb{R}^d$, but more research is still needed on this topic.

An open problem in the implementation of an empirical likelihood-based procedure resides in the actual maximization for finite n of the sum of the log empirical likelihood ratios with respect to the mean μ in the case of level change (4.2), and the maximization of (5.2) with respect to the parameter θ in the case of estimating functions. Consider first the level change identification, when only one break is allowed, at time q. The statistic to be minimized is $-2 \log ELR(z, 0, q) - 2 \log ELR(z, q, n)$ with respect to z. Under the hypothesis of no break, and for $z = \mu + O_p(n^{-1/2})$, we have [33, p. 1727]:

$$- 2 \log ELR(z, 0, q) - 2 \log ELR(z, q, n)$$
$$= q(\bar{y}_1 - z)' S_1^{-1} (\bar{y}_1 - z) + (n - q)(\bar{y}_2 - z)' S_2^{-1} (\bar{y}_2 - z) + o_p(1) \quad (6.3)$$

with

$$\bar{y}_1 = \frac{1}{q} \sum_{i=1}^{q} y_i \qquad\qquad \bar{y}_2 = \frac{1}{n-q} \sum_{i=q+1}^{n} y_i$$

$$S_1 = \frac{1}{q} \sum_{i=1}^{q} (y_i - z)(y_i - z)' \qquad S_2 = \frac{1}{n-q} \sum_{i=q+1}^{n} (y_i - z)(y_i - z)'.$$

The choice of the overall average $\bar{y} = q\bar{y}_1 + (n-q)\bar{y}_2$, that would be the minimizer if $S_1 = S_2 = S$ independent of z, may be appropriate for n large, but for smaller n the optimal value of z may be different. In the case of scalar data it is easy to show that the minimizer is a convex linear combination of $\bar{y}_1$ and $\bar{y}_2$, as stated in the following Theorem.

Theorem 6.1. *Let $\{y_1, y_2, \ldots, y_n\}$ be n independently identically distributed random variables with finite variance, and $1 < q < n$. The minimizer of the asymptotic part of (6.3) is a convex linear combination of $\bar{y}_1$ and $\bar{y}_2$:*

$$z^* = c\bar{y}_1 + (1-c)\bar{y}_2, \ \ 0 < c < 1.$$

For $k > 1$ breaks, it easily follows that, under the assumptions of Theorem 6.1, the optimal choice of z belongs to the convex hull of the sub-series averages: $\{\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_{k+1}\}$.

We conjecture that also in the multivariate case the optimal z is a linear combination of $\bar{y}_1$ and $\bar{y}_2$, as some simulation experience suggests, but no proof is available.

We turn now to the structural break case, on considering a linear model (5.3): $y = X\theta + \varepsilon$ and only one break, that splits the data into two sub-series $(y_1, y_2, \ldots, y_q)$ and $(y_{q+1}, \ldots, y_n)$. Let us partition accordingly the data vector, the design matrix and the residuals: $y = (y_1', y_2')'$, $X = (X_1', X_2')'$, $\varepsilon_1(\theta) = y_1 - X_1\theta$, $\varepsilon_2(\theta) = y_2 - X_2\theta$. For the estimating equations we have:

$$\sum_{i=1}^{q} g(y_i, \theta) = X_1'\varepsilon_1(\theta)$$

$$\sum_{i=q+1}^{n} g(y_i, \theta) = X_2'\varepsilon_2(\theta)$$

$$\left[\sum_{i=1}^{q} g(y_i, \theta)g(y_i, \theta)'\right]_{r,s} = \sum_{i=1}^{q} [\varepsilon_1(\theta)]_i^2 x_{ir} x_{is} = [G_1(\theta)]_{r,s}$$

$$\left[\sum_{i=q+1}^{n} g(y_i, \theta)g(y_i, \theta)'\right]_{r,s} = \sum_{i=q+1}^{n} [\varepsilon_2(\theta)]_{i-q}^2 x_{ir} x_{is} = [G_2(\theta)]_{r,s}.$$

Thus if θ_0 is the true parameter value, for $\theta = \theta_0 + O_p(n^{-1/2})$:

$$-2[\log ELR(\theta, 0, q) + \log ELR(\theta, q, n)]$$
$$= \varepsilon_1(\theta)'X_1 G_1(\theta)^{-1} X_1'\varepsilon_1(\theta) + \varepsilon_2(\theta)'X_2 G_2(\theta)^{-1} X_2'\varepsilon_2(\theta) + o_p(1) \quad (6.4)$$

and $\frac{1}{q}G_1(\theta_0) - \frac{1}{q}\sigma^2(X_1'X_1) = o_p(1)$, $\frac{1}{n-q}G_2(\theta_0) - \frac{1}{n-q}\sigma^2(X_2'X_2) = o_p(1)$. On denoting the least squares estimate on each sub-series by $\hat{\theta}_1 = (X_1'X_1)^{-1}X_1'y_1$ and $\hat{\theta}_2 = (X_2'X_2)^{-1}X_2'y_2$, the test statistic (6.4) may be written:

$$(\hat{\theta}_1 - \theta)'(X_1'X_1)G_1(\theta)^{-1}(X_1'X_1)(\hat{\theta}_1 - \theta)$$
$$+ (\hat{\theta}_2 - \theta)'(X_2'X_2)G_2(\theta)^{-1}(X_2'X_2)(\hat{\theta}_2 - \theta) + o_p(1). \quad (6.5)$$

Like in the case of level change, minimization of (6.5) with respect to θ is made difficult by the dependence of the matrices G_1 and G_2 on θ. To obtain an approximate solution, one can ignore such dependence, and put $G_1(\theta)^{-1}(X_1'X_1) \simeq$

$G_2(\theta)^{-1}(X_2'X_2) \simeq \sigma^2\,I$. Then, equating derivatives to zero, we obtain:

$$(X_1'X_1)\hat{\theta}_1 + (X_2'X_2)\hat{\theta}_2 = (X_1'X_1 + X_2'X_2)\theta$$

that is solved by $\hat{\theta} = (X'X)^{-1}X'y$, the least squares estimate under absence of breaks. However, in practice for small n the dependence of $G_1(\theta)$ and $G_2(\theta)$ on θ may not be negligible, and a numerical optimization routine could be advisable.

A further required stage is to determine the best segmentation, i. e., given the number of breaks m, determining their dates $t_1, t_2, \ldots, t_m$. However, the additive nature of the statistics (4.1) and (5.1) makes it possible to solve such problem by means of a procedure based on dynamic programming, similar to that introduced by [5] for the sup F tests. The key idea is that if $1 < t_1 < \cdots < t_m < n$ is the best segmentation into $(m+1)$ sub-series, then $t_1, t_2, \ldots, t_{m-1}$ are the optimal dates for the segmentation of the "truncated" series $(y_1, y_2, \ldots, y_{t_m})$ into m sub-series. Thus, having already computed the best break dates for k breaks of any series $(y_1, \ldots, y_t)$ for any t, and if $\lambda(t)$ is the resulting test statistic, the best dates for the whole series $(y_1, \ldots, y_n)$ and $(k+1)$ breaks are obtained by taking for the last break date the time t_{k+1} that solves:

$$t_{k+1} = \arg\max_t \{\lambda(t) - 2\log ELR(\theta, t, n)\} \tag{6.6}$$

and the other dates equal to the best segmentation of the series $(y_1, \ldots, y_{t_{k+1}})$ into $k + 1$ sub-series. This allows to solve iteratively the search for the break dates with an increasing number of breaks, by considering only one-dimensional maximization problems similar to (6.6). The value of the parameter θ is set for simplicity equal to its approximate optimum θ^* (the least squares estimate under the hypothesis of no break; and respectively $\bar{y}$ in case of level changes).

Thus, the proposed procedure for determining the optimal dates, allowing up to m breaks, runs in the following stages (remember that we assume at least h observations in each sub-series):

- *1st stage.* On assuming only one break, for any length τ from $2h$ to $n - h$, and for $\tau = n$, the series $(y_1, \ldots, y_\tau)$ is split into 2 parts by means of the date t_1 solving

  $$t_1(\tau) = \arg\min_t \{\log ELR(\theta, 0, t) + \log ELR(\theta, t, \tau),\ h < t \le \tau - h\}$$

  Let $\ell_0^{(1)}(1, \tau) = -2\{\log ELR[\theta, 0, t_1(\tau)] + \log ELR[\theta, t_1(\tau), \tau]\}$ the resulting value of the test statistic.
- *2nd stage.* Assuming two breaks, for any length τ from $3h$ to $n - h$, and for $\tau = n$, the series $(y_1, \ldots, y_\tau)$ is split into 3 parts. The second break date t_2 is obtained by solving

  $$t_2(\tau) = \arg\min_t \left\{-\frac{1}{2}\ell_0^{(1)}(1, t) + \log ELR(\theta, t, \tau),\ 2h < t \le \tau - h\right\}$$

  while the first break date will be selected equal to $t_1[t_2(\tau)]$, and $\ell_0^{(2)}(1, \tau)$ will denote the resulting value of the test statistic.

- *k-th stage.* For k breaks, for any length τ from $(k+1)h$ to $n - h$, and for $\tau = n$, the series $(y_1, \ldots, y_\tau)$ is split into $(k+1)$ parts. The last break date t_k is given by:

$$t_k(\tau) = \arg\min_t \left\{ -\frac{1}{2}\ell_0^{(k-1)}(1, t) + \log ELR(\theta, t, \tau),\, (k+1)h < t \leq \tau - h \right\}$$

and the preceding break dates are the optimal ones for $(k-1)$ breaks of the data from 1 to $t_k(\tau)$.

Though this dynamic programming strategy is based only on one-dimensional optimization problems like (6.6), it may be computationally heavy, specially if n is not small; an alternative is the search for nearly-optimal solutions by means of a meta-heuristic method. Genetic algorithms have been successfully employed for similar problems, e. g. in threshold autoregressive models building [see 8, for a review].

The final problem is identifying the number of breaks, if any, given a maximum number of allowable breaks m. The simplest way is computing the ELR test statistic for absence of breaks against an alternative of k breaks ($k = 1, 2, \ldots, m$) and selecting the most significant figure, or using an information criterion. Identification criteria such as AIC and BIC applied to empirical likelihood are discussed, e. g., in [24] and [40]. Alternatively, [4] proposed a sequential method, symbolized by $F(\ell + 1|\ell)$, to decide whether to accept the hypothesis of $\ell + 1$ breaks against that of ℓ breaks. Essentially, in its simpler version, it amounts to consider the $(\ell + 1)$ optimal sub-series obtained under ℓ breaks, and to test each of them for the presence of an additional break.

Given the asymptotic equivalence between the F and ELR test statistics, such strategies may also be adopted for the empirical likelihood-based procedure. Practical recommendations are provided by [5]. They suggest to use the sequential procedure, but not stopping as soon as the test is not significant, rather computing the $F(\ell + 1|\ell)$ statistic for sufficiently many values of ℓ, and selecting the number of breaks m so that the tests $F(\ell + 1|\ell)$ are insignificant for $\ell \geq m$.

7. Simulations and applications

A simulation study has been conducted to check the behavior of the two tests on finite length samples ($n = 50, 100, 200$). Thousand random samples were simulated according to the following probability distributions: standard Gaussian $N(0, 1)$, standardized Gamma distribution $G(5, 1)$ and $G(10, 1)$, exponential $G(1, 1)$ and Student's t with different degrees of freedom $t(2), t(5), t(10)$. The tests for the hypothesis of absence of breaks against the alternative of one level change were applied to the simulated samples with sizes 0.10, 0.05 and 0.01. Since the essential findings were similar, only results for size 0.05 are presented. Moreover, the two tests yielded the same practical results for series length $n = 200$, we shall report only findings for $n = 50$ and 100.

TABLE 1
Frequency of rejection of the null hypothesis of no break on 1000 replications, data without break

		$N(0,1)$	$t(2)$	$t(5)$	$t(10)$	$G(1,1)$	$G(5,1)$	$G(10,1)$
$n=50$	F	51	33	33	46	41	49	57
	ELR	61	131	62	63	87	74	75
$n=100$	F	36	37	40	54	37	45	54
	ELR	66	153	73	66	67	61	62

TABLE 2
Frequency of rejection of the null hypothesis of no break on 1000 replications, data with one level change equal to 0.5

		$N(0,1)$	$t(2)$	$t(5)$	$t(10)$	$G(1,1)$	$G(5,1)$	$G(10,1)$
$n=50$	F	297	281	284	286	299	263	275
	ELR	315	481	336	327	366	284	307
$n=100$	F	549	540	545	530	549	529	530
	ELR	538	679	582	542	591	543	529

The frequencies of rejections of H_0 when it is true for a nominal size $\alpha = 0.05$ are shown in Table 1. An overall consideration is that the observed size of F tends to be generally slightly smaller, and that of ELR slightly larger, than 0.05. A moderate leptokurtosis as in $t(5)$ and $t(10)$ does not modify such a behavior, while this phenomenon is more visible in asymmetric distributions like $G(5,1)$ and $G(10,1)$ and specially for exponential data, though lighter for $n = 100$. Notably different are the results for $t(2)$ data, where the ELR shows an exceptionally large frequency of rejection, indicating that the hypothesis of finite moments (this distribution has not finite variance) is critical for convergence to the asymptotic distribution.

The difference between the F and ELR statistics in case of absence of breaks is generally moderate and decreases rapidly when n increases. Gaussian, $G(5,1), G(10,1), t(5)$ and $t(10)$ distributions give similar results, while the behavior is somewhat different specially for Student's t with infinite variance $t(2)$ and partly for exponential $G(1,1)$, where ELR tends to be slightly larger than F.

In order to check the power of the tests, we inserted a level change at the midpoint of the simulated series and repeated the exercise. The power of the tests as a function of the break size (from 0.1 to 0.9 times the series standard error) is reported in Figure 1 for the Gaussian case and series length $n = 50$, showing a marginal advantage in power for the ELR test; when $n = 100$ the two lines are undistinguishable.

The results for a moderate break size 0.5 are now described in more detail. The overall behavior is described in Figure 2 where the distribution of results for Gaussian data in the cases of no break and one break is presented for $n=100$. To make the graph more readable, we have computed histograms on 100 equal width classes, and plotted the class frequencies against the central class values.

The frequencies of rejection of the null hypothesis of absence of breaks is shown in Table 2. For $n = 50$ the ELR test has a slightly larger observed power, while for $n = 100$ the power of the two tests is essentially equivalent.
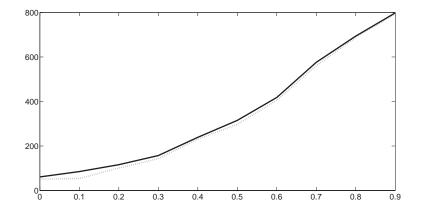
FIG 1. *Observed power of the tests for Gaussian data, n=50 and break sizes from 0.1 to 0.9 times the standard error of the series. Continuous line ELR, dashed line F.*
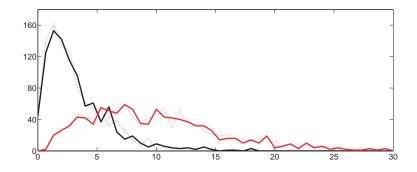

FIG 2. *Distribution of the observed values of ELR and F statistics for Gaussian data, n=100. Black: no level change, Red: level change 0.5. Continuous line ELR, dashed line F.*

In the level change case the differences between F and ELR tend to increase slightly. ELR is often a bit larger than F, specially for exponential $G(1,1)$ and t distributions with 2 and 5 degrees of freedom rather than in the Gaussian case.

Finally, the distributions of the selected break dates according to the two tests appear similar, an example is Figure 3 concerning the case of the $t(10)$ data, n=100 (here also the frequency of each histogram class is plotted against the central value of the class).

We have also simulated data according to two linear models (among those considered in [7]), with $n = 100$ and 1000 replications.

Model 1 (regression) $x_t = \alpha + \beta z_t + \varepsilon_t$ where $\{\varepsilon_t\}$ and $\{z_t\}$ are i. i. d. standard Gaussian, $\alpha = 2$, $\beta = 1$, changes in the parameter α at $t = 50$.

Model 2 (autoregression) $x_t = \alpha + \beta x_{t-1} + \varepsilon_t$ where $\{\varepsilon_t\}$ are i. i. d. standard Gaussian, $\alpha = 10$, $\beta = 0.5$ and changes in the parameter α at $t = 50$.
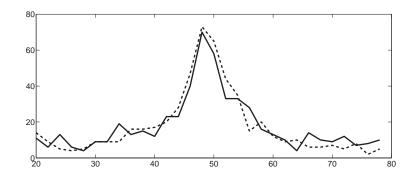
FIG 3. *Distribution of the selected break dates, n=100, data generated according to t(10), break size 0.5. Continuous line ELR, dashed line F.*
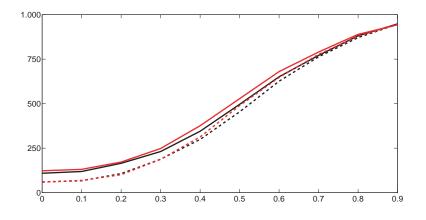

FIG 4. *Observed power of the tests for Model 1 (black) and Model 2 (red), break sizes from 0.1 to 0.9 times the standard error of the innovations. Continuous line ELR, dashed line F.*

The power of the two tests for break sizes $0(0.1)0.9$ are shown in Figure 4. Here also it may be observed that the behavior of the two tests is similar, though the ELR test has an observed size slightly larger than the nominal 0.05.

The tests for partial structural change were also considered. Sets of thousand replications of series generated according to Model 1, and breaks with different size in the first parameter were simulated with $n = 200$, and the partial ELR and F tests for the hypothesis that the second parameter remains constant were applied. The experiment was repeated with series containing also a moderate increase in the second parameter at the break date. The results are reported in Table 3. Figures are relative frequencies of rejection of the hypothesis of constancy of the second parameter, computed only on the replications where the null hypothesis of absence of a complete structural change was rejected. The first rows ($bs2 = 0.0$) show the observed test size, that is always close to

TABLE 3
Observed power of the partial tests for constancy of the second parameter in Model 1, break sizes of the first parameter $bs1 = 0.5(0.1)0.9$, break size of the second parameter $bs2 = 0.0, 0.1, 0.2$

$bs1$		0.5	0.6	0.7	0.8	0.9
$bs2 = 0.0$	F	0.05	0.04	0.03	0.03	0.03
	ELR	0.08	0.07	0.06	0.06	0.06
$bs2 = 0.1$	F	0.13	0.12	0.11	0.10	0.10
	ELR	0.17	0.15	0.14	0.14	0.14
$bs2 = 0.2$	F	0.40	0.36	0.34	0.33	0.32
	ELR	0.41	0.37	0.35	0.34	0.34

the nominal value 0.05, though a little larger for ELR and smaller for F. The second and third row sets relate to cases when the structural change is complete, since the break has an effect, though moderate, on the second parameter too; thus, they denote observed power. The figures for ELR are generally slightly larger than those for F. An overall comment is that the results are not largely influenced by the break size of the first parameter, implicitly indicating that its estimation bias is not serious.

We present now two applications of the break detection methods to real data. The two series are not satisfactorily described by linear models since they exhibit non linear or non stationary features, possibly due to structural changes, and more complicated models were proposed.

The first dataset is the help-wanted advertising index (monthly from January 1960 to December 1996, $n = 444$) analyzed by [27]. They fitted a smooth transition time-varying model with a structural change at December 1979. The same series was considered by [9] who proposed a non-linear non-stationary two-regime model of a similar type but with change at April 1969. We tried a model with a constant, a linear trend $T_t = (t/n)$ and two autoregressive terms at lags 1 and 2; the Bai and Perron and ELR procedures gave the same results, indicating three regimes with structural change at $t = 106$ (May 1969) and $t = 359$ (June 1990). The sum of squared residuals is 0.766, slightly larger than the model of [9] (0.649) and the model of [27] (0.691), that both allow non-linearity in addition. Figure 5 shows the data and the break dates, the first one appears nearly equal to [9], the second structural change is less clear-cut. The Bai and Perron procedure indicated a third possible break at October 1979 (similar to the break date proposed by [27]), but the $F(3|2)$ statistic was not significant.

The second series relates to the average annual temperature at Innsbruck, 1803–2005. This dataset was analyzed by [10] who proposed an autoregressive model with the autoregressive coefficients linearly time-varying, and a regime change at year 1973. We fitted a simple linear model with a constant, a linear trend $T_t = (t/n)$ and an autoregressive term at lag 1. Here also the Bai and Perron and ELR suggest essentially the same structural changes: three breaks at $t = 55$ ($t = 57$ for ELR), 93, 158 (years 1857, 1895, 1960). The data are shown in Figure 6. The residual variance of the 3-break resulting model is about 43, while the residual variance of the model in [10] is 46.7, it seems that their method privileges parsimony, as found also for the help-wanted data.
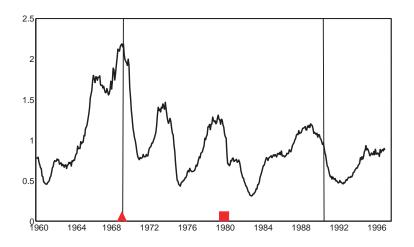
FIG 5. *Help-wanted advertising index. Vertical bars indicate the two breaks, triangle and square indicate the regime change date according to the models of [9] and [27] respectively.*
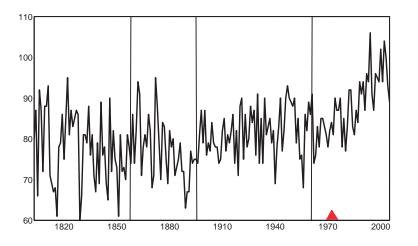

FIG 6. *Average annual temperature at Innsbruck. Vertical bars indicate the three breaks, the regime change date according to the model of [10] is indicated by a triangle.*

8. Conclusions

We have discussed how the empirical likelihood paradigm may be adopted for the problem of identifying and estimating structural changes in time series. The proposed empirical likelihood ratio test is shown to be asymptotically equivalent to the tests based on F-type and Cusum statistics.

Therefore, similar strategies as those proposed by Bai and Perron [4] may be used for identifying the number of breaks and their dates with empirical likelihood. In particular, a dynamic programming algorithm analogue to that proposed in [5] is introduced for detecting the optimal break dates.

When testing the null hypothesis of absence of breaks against the alternative of m breaks, the ELR test statistic is simply the sum of the ELR computed on each of the $(m+1)$ sub-series having a different structure under the alternative. However, the required ELR test statistic value may be computationally heavy, since a maximization with respect to the model parameters under the null hypothesis is needed, and for some models the solution is not near the maximum likelihood, or the least squares, estimates of the parameters, but must be actually obtained through an optimization algorithm.

It has also been suggested how to use the empirical likelihood test in the case of partial structural breaks, when the possible change involves only some of the model parameters.

Experience on a set of simulated series with finite length suggests that the results of the two tests are equivalent for long series, but not so different also for smaller lengths $n = 50$ and 100, when generally the ELR test shows a slightly larger power, but also a slightly larger size than nominal. Finally, we applied the tests on some real series with non linear or non stationary features induced by structural changes, and found that both the ELR and F procedures are able to detect the breaks.

The basic idea of the present paper may be applied in principle to any model for which an empirical likelihood inference has been proposed. Thus, future research may be concerned with break detection in $GARCH$ models using the results of [11], or random coefficient $INAR(p)$ processes basing on [43]. Moreover, it may be possible to consider periodogram ordinates rather than data, extending to break detection the frequency domain empirical likelihood inference proposed by [28] for $ARMA$ models, extended to long-memory $ARFIMA$ models by [41] and further developed for more general models by [31, 30, 22].

Appendix A: Proofs

Proof of Theorem 4.1. The maximum unconstrained empirical likelihood is n^{-n}, while H_0 states that each sub-series $(y_{t_k+1}, \ldots, y_{t_{k+1}}), k = 0, 1, \ldots, m$ has mean equal to μ_0. Therefore the empirical likelihood ratio for H_0 is

$$n^n \max \left\{ \prod_{i=1}^n w_i \, : \, w_i > 0, \sum_{i=1}^n w_i = 1, \sum_{j=t_k+1}^{t_{k+1}} w_j(y_j - \mu_0) = 0, k = 0, 1, \ldots, m \right\}. \tag{A.1}$$

Let $s_k = w_{t_k+1} + \cdots + w_{t_{k+1}}$, with $s_0 + s_1 + \cdots + s_m = 1$. On writing $\pi_j = w_j/s_k, t_k + 1 \leq j \leq t_{k+1}, k = 0, \ldots, m$, (A.1) may be written:

$$n^n \max \left\{ \prod_{k=0}^m \prod_{j=t_k+1}^{t_{k+1}} w_j \, : \, \sum_{j=t_k+1}^{t_{k+1}} w_j = s_k, \sum_{j=t_k+1}^{t_{k+1}} w_j(y_j - \mu_0) = 0, k = 0, \ldots, m \right\}$$

$$= n^n \max \left\{ \prod_{k=0}^{m} \prod_{j=t_k+1}^{t_{k+1}} s_k \pi_j \ : \ \sum_{j=t_k+1}^{t_{k+1}} \pi_j = 1, \ \sum_{j=t_k+1}^{t_{k+1}} \pi_j(y_j - \mu_0) = 0, k = 0, \dots, m \right\}$$

$$= n^n \max_{s} \prod_{k=0}^{m} s_k^{n_k} \max_{\pi} \left\{ \prod_{j=t_k+1}^{t_{k+1}} \pi_j \ : \ \pi_j > 0, \ \sum_{j=t_k+1}^{t_{k+1}} \pi_j = 1, \ \sum_{j=t_k+1}^{t_{k+1}} \pi_j(y_j - \mu_0) = 0 \right\}$$

$$= n^n \max_{s} \prod_{k=0}^{m} s_k^{n_k} ELR(\mu_0, t_k, t_{k+1}) n_k^{-n_k}.$$

where we have put $n_k = t_{k+1} - t_k, (k = 0, 1, \dots, m)$, $n_0 + n_1 + \dots + n_m = n$, and the maximum is taken for s belonging to the simplex $\{s_i > 0, s_0 + s_1 + \dots + s_m = 1\}$. Therefore the test statistic is

$$- 2 \log ELR_0(\mu_0) = -2 \sum_{k=0}^{m} \log ELR(\mu_0, t_k, t_{k+1}) +$$

$$- 2 \left\{ n \log n - \sum_{k=0}^{m} n_k \log n_k + \max_{s} \sum_{k=0}^{m} n_k \log s_k \right\}. \quad \text{(A.2)}$$

Since the maximum of $\sum_k n_k \log s_k$ constrained to $\sum_k s_k = 1$ is obtained for $s_k = n_k/n$, the term on the second line vanishes and the equality is proved. Each term of (4.1) is computed on a series of n_k observations. If as $n \to \infty$ each $n_k \to \infty$, given the independence of the y_i's, each converges to a chi square distribution with one degree of freedom, thus (4.1) converges to a chi square with $(m+1)$ degrees of freedom. □

A Lemma is needed for the proof of Corollary 4.1.

Lemma A.1. *Let $E|y_i|^4 < \infty$. Then under the hypothesis of no break*

$$-2 \log ELR(\bar{y}, 0, q) = q(\bar{y}_1 - \bar{y})^2 / s_1^2 + o_p(1)$$
$$-2 \log ELR(\bar{y}, q, n) = (n - q)(\bar{y}_2 - \bar{y})^2 / s_2^2 + o_p(1)$$

where

$$s_1 = \frac{1}{q} \sum_{i=1}^{q} (y_i - \bar{y})^2, \ s_2 = \frac{1}{n-q} \sum_{i=q+1}^{n} (y_i - \bar{y})^2.$$

Proof. Note that $\bar{y} = \mu + O_p(n^{-1/2})$ therefore the thesis is a direct consequence of Eq. (2.1) of [33] applied to the two sub-series. □

Proof of Corollary 4.1. From Lemma A.1 we have

$$-2 \log ELR(\bar{y}, 0, q) = q(\bar{y}_1 - \bar{y})^2 / s_1^2 + o_p(1)$$
$$-2 \log ELR(\bar{y}, q, n) = (n - q)(\bar{y}_2 - \bar{y})^2 / s_2^2 + o_p(1).$$

Note that $q(\bar{y}_1 - \bar{y})^2 + (n - q)(\bar{y}_2 - \bar{y})^2 = \sigma^2 z_n(q/n)^2 \left[\frac{q}{n}(1 - \frac{q}{n})\right]^{-1}$, and, since s_1^2 and s_2^2 are consistent estimates of the variance σ^2, we may write

$$q\frac{(\bar{y}_1 - \bar{y})^2}{s_1^2} + (n - q)\frac{(\bar{y}_2 - \bar{y})^2}{s_2^2} - z_n(q/n)^2 \left[\frac{q}{n}\left(1 - \frac{q}{n}\right)\right]^{-1} = o_p(1).$$

Since $\bar{y}_1$ and $\bar{y}_2$ are the maximum empirical likelihood estimators of the mean of the two sub-series,

$$\sup_{\mu} ELR(\mu, 0, q) \leq ELR(\bar{y}_1, 0, q); \; \sup_{\mu} ELR(\mu, q, n) \leq ELR(\bar{y}_2, q, n)$$

and

$$-2\sup_{\mu} \{\log ELR(\mu, 0, q) + \log ELR(\mu, q, n)\}$$

$$= \inf_{\mu} -2\{\log ELR(\mu, 0, q) + \log ELR(\mu, q, n)\}$$

$$\geq 0 = -2\{\log ELR(\bar{y}_1, 0, q) + \log ELR(\bar{y}_2, q, n)\}.$$

Now, under the hypothesis H_0 of no level change, as $n \to \infty$, $\bar{y}_1 - \bar{y}$ and $\bar{y}_2 - \bar{y}$ converge to zero in probability, therefore

$$-2\sup_{\mu} \{\log ELR(\mu, 0, q) + \log ELR(\mu, q, n)\}$$

$$= -2\{\log ELR(\bar{y}, 0, q) + \log ELR(\bar{y}, q, n)\} + o_p(1)$$

$$= z_n(q/n)^2 \left[\frac{q}{n}\left(1 - \frac{q}{n}\right)\right]^{-1} + o_p(1).$$

$$\square$$

Proof of Corollary 4.2. If t_j denotes the j−th break date as before, and $n_j = t_{j+1} - t_j$, the F statistic for m level changes ($p = 1$ and θ_j equal to the mean of the j−th sub-series) may be written:

$$F(t_1, t_2, \ldots, t_m) = \frac{n_0(\bar{y}_1 - \bar{y})^2 + n_1(\bar{y}_2 - \bar{y})^2 + \cdots + n_m(\bar{y}_{m+1} - \bar{y})^2}{S} \frac{n - (m+1)}{m}$$

where

$$S = \sum_{k=1}^{m+1} \sum_{j=t_{k-1}+1}^{t_k} (y_j - \bar{y}_k)^2$$

Since $S/(n - m - 1)$ is a consistent estimate of σ^2 under H_0, using Lemma A.1 for each sub-series $(y_{t_k+1}, \ldots, y_{t_{k+1}}), k = 0, \ldots, m$ we obtain that the difference between $mF(t_1, \ldots, t_m)$ and (4.2) converges to zero in probability. $\square$

Proof of Theorem 5.1. Under our assumptions, Theorem 2 of [36] may be applied to each subseries $(y_{t_k+1}, \ldots, y_{t_{k+1}})$ and ensures that, if H_0 is true, then $-2\log ELR(\theta_0, t_k, t_{k+1})$ converges to a chi square distribution with p degrees of freedom. Since the sub-series are independent (5.1) converges to a χ^2 with $(m+1)p$ degrees of freedom. $\square$

Proof of Corollary 5.1. Let, as in Section 2, $\hat{\theta} = (\bar{X}'\bar{X})^{-1}\bar{X}'y$ denote the least squares estimate under the alternative of m breaks, R denote the restriction matrix such that $R\theta = 0$ under the null of no break, and $\hat{\hat{\theta}}$ denote the restricted least squares estimate under H_0:

$$\hat{\hat{\theta}} = \hat{\theta} - (\bar{X}'\bar{X})^{-1}R'\{R(\bar{X}'\bar{X})^{-1}R'\}^{-1}R\hat{\theta}.$$

Let θ^* be the maximizer of (5.2) with respect to θ; then under H_0 both θ^* and $\hat{\hat{\theta}}$ converge in probability to θ_0, therefore in a similar way to Corollary 4.1, extended to more than two sub-series, it follows that the difference between (5.2) and $-2\log ELR(\hat{\hat{\theta}}, 0, n)$ tends to 0 in probability as $n \to \infty$ [see e.g. 33]. We may conclude that

$$
\begin{aligned}
\ell_0(t_1, \ldots, t_m) &= -2\log ELR(\hat{\hat{\theta}}, 0, n) + o_p(1) \\
&= \sum g(y_i, \hat{\hat{\theta}})' \left[\sum g(y_i, \hat{\hat{\theta}})g(y_i, \hat{\hat{\theta}})'\right]^{-1} \sum g(y_i, \hat{\hat{\theta}}) + o_p(1) \\
&= \varepsilon(\hat{\hat{\theta}})'\bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}'\varepsilon(\hat{\hat{\theta}})\sigma^{-2} + o_p(1).
\end{aligned}
$$

Put $\varepsilon(\hat{\hat{\theta}}) = \varepsilon(\hat{\theta}) - \alpha$, where

$$\alpha = \bar{X}(\bar{X}'\bar{X})^{-1}R'\{R(\bar{X}'\bar{X})^{-1}R'\}^{-1}R\hat{\theta}$$

and note that $\alpha = \bar{X}(\hat{\theta} - \hat{\hat{\theta}})$ and $\bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}'\alpha = \alpha$. Moreover, since $\hat{\theta}$ is the unconstrained LS estimate, $\varepsilon(\hat{\theta})'\bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}'\varepsilon(\hat{\theta}) = 0$, then:

$$
\begin{aligned}
\ell_0(t_1, \ldots, t_m) &= \alpha'\bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}'\alpha + o_p(1) \\
&= \|\bar{X}(\hat{\theta} - \hat{\hat{\theta}})\|^2\sigma^{-2} + o_p(1) \\
&= mpF(t_1, \ldots, t_m)\frac{\|y - \bar{X}\hat{\theta}\|^2}{[n - (m+1)p]\sigma^2} + o_p(1) \\
&= mpF(t_1, \ldots, t_m) + o_p(1).
\end{aligned}
$$

$\square$

Proof of Theorem 5.2. Part (i) follows from an application of Lemma 1 of [14] to each sub-series $\{y_{t_i+1}, \ldots, y_{t_{i+1}}\}$, $i = 0, 1, \ldots, m$. Assumptions AR4 and AR5 ensure that the empirical likelihood ratios for the sub-series are asymptotically independent, therefore the asymptotic distribution follows. Part (ii) may be shown in the same way as Corollary 5.1, since under assumptions AR1–AR5 both the maximizer θ^* of (5.6) and the restricted estimate $\hat{\hat{\theta}}$ converge in probability to θ_0 under H_0. $\square$

Proof of Theorem 6.1. Let $(\bar{y}_1, \sigma_1)$ and $(\bar{y}_2, \sigma_2)$ denote the average and variance of the two sub-series $\{y_1, \ldots, y_q\}$ and $\{y_{q+1}, \ldots, y_n\}$. We may write

$$S_1 = \frac{1}{q}\sum_{i=1}^{q}(y_i - z)^2 = \sigma_1 + (\bar{y}_1 - z)^2$$

$$S_2 = \frac{1}{n-q} \sum_{i=q+1}^{n} (y_i - z)^2 = \sigma_2 + (\bar{y}_2 - z)^2.$$

Then the function to be minimized is

$$q(\bar{y}_1 - z)^2/\{\sigma_1 + (\bar{y}_1 - z)^2\} + (n-q)(\bar{y}_2 - z)^2/\{\sigma_2 + (\bar{y}_2 - z)^2\}. \qquad (A.3)$$

Let $\delta = \bar{y}_1 - \bar{y}_2$ and $c = (z - \bar{y}_2)/\delta$ so that $z = c\bar{y}_1 + (1-c)\bar{y}_2$; also, $\bar{y}_1 - z = (1-c)\delta$ and $\bar{y}_2 - z = -c\delta$. On substituting into (A.3):

$$q\frac{(1-c)^2\delta^2}{\sigma_1 + (1-c)^2\delta^2} + (n-q)\frac{c^2\delta^2}{\sigma_2 + c^2\delta^2}.$$

The derivative with respect to c of the above function:

$$-2q(1-c)\frac{\sigma_1}{[\sigma_1 + \delta^2(1-c)^2]^2} + 2(n-q)c\frac{\sigma_2}{[\sigma_2^2 + \delta^2 c^2]}$$

is positive for $c > 1$, and negative for $c < 0$. It follows that the minimum is attained in the interval $(0, 1)$. ☐

Acknowledgements

References

[1] ANDREWS, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856. MR1231678

[2] AUE, A. and HORVATH, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012

[3] BAI, J. (1999). Likelihood ratio tests for multiple structural changes. *J. Econometrics* **91** 299–323. MR1703949

[4] BAI, J. and PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121

[5] BAI, J. and PERRON, P. (2003a). Computation and analysis of multiple structural change models. *J. Appl. Econometrics* **18** 1–22.

[6] BAI, J. and PERRON, P. (2003b). Critical values for multiple structural change tests. *Econom. J.* **6** 72–78. MR1992392

[7] BAI, J. and PERRON, P. (2006). Multiple structural change models: a simulation analysis. In *Econometrics and Practice: Frontiers of Analysis and Applied Research* (D. Corbea, S. Durlauf and B. E. Hansen, eds.) 212–237. Cambridge University Press. MR2235721

[8] BARAGONA, R., BATTAGLIA, F. and POLI, I. (2011). *Evolutionary Statistical Procedures.* Springer, Heidelberg. MR2757473

[9] BATTAGLIA, F. and PROTOPAPAS, M. K. (2011). Time-varying multi-regime models fitting by genetic algorithms. *J. Time Series Anal.* **32** 237–252. MR2808252

[10] BATTAGLIA, F. and PROTOPAPAS, M. K. (2012). An analysis of global warming in the Alpine region based on nonlinear nonstationary time series models. *Stat. Methods Appl.* **21** 315-373. (with Discussion). MR2981648

[11] CHAN, N. H. and LING, S. (2006). Empirical likelihood for Garch models. *Econom. Theory* **22** 403–428. MR2256098

[12] CHAN, N. H., PENG, L. and ZHANG, D. (2011). Empirical-likelihood based confidence intervals for conditional variance in heteroskedastic regression models. *Econom. Theory* **27** 154–177. MR2771014

[13] CHEN, J., VARIYATH, A. M. and ABRAHAM, B. (2008). Adjusted empirical likelihood and its properties. *J. Comput. Graph. Statist.* **17** 426–443. MR2439967

[14] CHUANG, C.-S. and CHAN, N. H. (2002). Empirical likelihood for autoregressive models, with applications to unstable time series. *Statist. Sinica* **12** 387–407. MR1902716

[15] CSORGO, M. and HORVATH, L. (1997). *Limit theorems in change-point analysis.* Wiley, Chichester. MR2743035

[16] FAN, G.-L. and XU, H.-X. (2012). Empirical likelihood inference for partially time-varying coefficients errors-in-variables models. *Electron. J. Stat.* **6** 1040-1058. MR2988438

[17] GONG, Y., LI, Z. and PENG, L. (2010). Empirical likelihood intervals for conditional Value-at-Risk in Arch/Garch models. *J. Time Series Anal.* **31** 65–75. MR2677338

[18] GUAN, Z. (2004). A semiparametric changepoint model. *Biometrika* **91** 849–862. MR2126037

[19] HU, X., WANG, Z. and ZHAO, Z. (2009). Empirical likelihood for semiparametric varying-coefficient partially linear errors-in-variable models. *Statistist. Probab. Lett.* **79** 1044–1052. MR2510766

[20] JANDHYALA, V., FOTOPOULOS, S., MACNEILL, I. and LIU, P. (2013). Inference for single and multiple change-points in time series. *J. Time Series Anal.* **34** 423–446. MR3070866

[21] JING, B.-Y. (1995). Two-sample empirical likelihood method. *Statist. Probab. Lett.* **24** 315–319. MR1353889

[22] KAKIZAWA, Y. (2013). Frequency domain generalized empirical likelihood method. *J. Time Series Anal.* **34** 691–716.

[23] KOLACZYK, E. D. (1994). Empirical likelihood for generalized linear models. *Statist. Sinica* **4** 199-218. MR1282871

[24] KOLACZYK, E. D. (1995). An information criterion for empirical likelihood with general estimating equations Technical No. 417, Department of Statistics, The University of Chicago.

[25] LIU, Y., ZOU, C. and ZHANG, R. (2008a). Empirical likelihood for the two-sample problem. *Statist. Probab. Lett.* **78** 548–556. MR2400868

[26] Liu, Y., Zou, C. and Zhang, R. (2008b). Empirical likelihood ratio test for a change-point in linear regression model. *Comm. Statist. Theory Methods* **37** 2551–2563. MR2458939

[27] Lundberg, S., Teräsvirta, T. and van Dijk, D. (2003). Time-varying smooth transition autoregressive models. *J. Bus. Econom. Statist.* **21** 104–121. MR1950382

[28] Monti, A. C. (1997). Empirical likelihood confidence regions in time series models. *Biometrika* **84** 395–405. MR1467055

[29] Mykland, P. A. (1999). Dual likelihood. *Ann. Statist.* **23** 396–421. MR1332573

[30] Nordman, D. J. (2009). Tapered empirical likelihood for time series data in time and frequency domain. *Biometrika* **96** 119–132. MR2482139

[31] Nordman, D. J. and Lahiri, S. N. (2006). A frequency domain empirical likelihood for short- and long-range dependence. *Ann. Statist.* **34** 3019–3050. MR2329476

[32] Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237-249. MR0946049

[33] Owen, A. B. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19** 1725–1747. MR1135146

[34] Owen, A. B. (2001). *Empirical Likelihood.* Chapman and Hall CRC.

[35] Perron, P. (2006). Dealing with structural breaks. In *Palgrave Handbook of Econometrics*, (K. Patterson and T. C. Mills, eds.) **1** 278–352. Palgrave McMillan.

[36] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300–325. MR1272085

[37] Robbins, M., Gallagher, C., Lund, R. and Aue, A. (2011). Mean shift testing in correlated data. *J. Time Series Anal.* **32** 498-511. MR2835683

[38] Robbins, M. W., Lund, R. B., Gallagher, C. M. and Lu, Q. (2011). Changepoints in the north atlantic tropical cyclone record. *J. Amer. Statist. Assoc.* **106** 89–99. MR2816704

[39] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York. MR0838963

[40] Variyath, A. M., Chen, J. and Abraham, B. (2010). Empirical likelihood based model selection. *J. Statist. Plann. Inference* **140** 971–981. MR2574659

[41] Yau, C. Y. (2012). Empirical likelihood in long-memory time series models. *J. Time Series Anal.* **33** 269–275. MR2902463

[42] You, J. and Zhou, Y. (2006). Empirical likelihood for semiparametric varying-coefficient partially linear regression models. *Statist. Probab. Lett.* **76** 412–422. MR2232483

[43] Zhang, H., Wang, D. and Zhu, F. (2011). Empirical likelihood inference for random coefficient INAR(p) process. *J. Time Series Anal.* **32** 195–203. MR2808249

[44] Zi, X., Zou, C. and Liu, Y. (2012). Two-sample empirical likelihood method for difference between coefficients in linear regression model. *Statist. Papers* **53** 83-93. MR2878593

[45] Zou, C., Liu, Y., Qin, P. and Wang, Z. (2007). Empirical likelihood ratio test for the change-point problem. *Statist. Probab. Lett.* **77** 374–382. MR2339041