Empirical Likelihood for Change Point Detection
in Autoregressive Models

Ramadha D. Piyadi Gamage\textsuperscript{1}, Wei Ning\textsuperscript{2}\textsuperscript{*}

\textsuperscript{1}Department of Mathematics
Western Washington University, Bellingham, WA 98225, USA

\textsuperscript{2}Department of Mathematics and Statistics
Bowling Green State University, Bowling Green, OH 43403, USA

Abstract

Change point analysis has become an important research topic in many fields of applications. Several research work has been carried out to detect changes and its locations in time series data. In this paper, a nonparametric method based on the empirical likelihood is proposed to detect the structural changes of the parameters in autoregressive (AR) models. Under certain conditions, the asymptotic null distribution of the empirical likelihood ratio test statistic is proved to be the extreme value

\textsuperscript{*}Corresponding author. Email: wning@bgsu.edu
distribution. Further, the consistency of the test statistic has been proved. Simulations have been carried out to show that the power of the proposed test statistic is significant. The proposed method is applied to real world data set to further illustrate the testing procedure.

**Keywords**: Autoregressive model; Change point analysis; Empirical Likelihood; Extreme value distribution; Consistency.

## 1 Introduction

Change point analysis introduced by Page (1954, 1955) has become popular due to its usage in wide variety of fields, such as stock market analysis, quality control, traffic mortality rate, geology data analysis, genetics, etc. It concerns both detecting whether or not a change(s) has (have) occurred, and identifying the location(s) of any such change(s). Several methods to identify and estimate the change points in the change point problem are proposed by scholars. Bayesian approach to detect changes in the mean has been discussed by Chernoff and Zacks (1964) and Sen Srivastava (1975). Further, Csörgő and Horváth (1997) and Chen and Gupta (2000) established asymptotic results on parametric change point models. Hawkins (1977), Worsley (1986) and Gombay and Horváth (1994) are a few among the many researchers who discussed change point problem under the parametric settings. However, the parametric methods are no longer applicable if the underlying distribution is completely unknown. In such a case, a nonparametric approach
should be considered as an alternative. One such popular nonparametric approach is the Cumulative Sum (CUSUM) method. Most authors have assumed that the observations are independent and studied the case where two distributions differ only in location. Combining nonparametric approaches along with the change point detection has been studied by many scholars over the past years. Aue and Horváth (2012) discussed two methods, namely, Cumulative Sum (CUSUM) and Likelihood Ratio Test (LRT), on how they can be modified for data exhibiting serial dependence. Further, they provided some insight to the sequential procedure as well. Lee et. al. (2003) also discussed about the Cusum test for changes of parameters in time series models and considered the changes of the parameters in a random coefficient autoregressive model AR(1) and that of the autocovariances of a linear process.

The change point problem may be viewed as a two-sample test adjusted for the unknown break location, thus leading to max-type procedures. Correspondingly, asymptotic relationships are derived to obtain critical values for the tests. In general, the change point problem can described as follows. Let $x_1, x_2, ..., x_n$ be a sequence of independent random vectors (variables) with probability distribution functions $F_1, F_2, ..., F_n$, respectively. More specifically, suppose that the distributions $F_1, F_2, ..., F_n$ belong to a common parametric family $F(\theta)$, where $\theta \in \mathbb{R}^p$, then the change point problem is to test the hypotheses about the population parameters $\theta_i, i = 1, ..., n$

$$H_0 : \theta_1 = \theta_2 = ... = \theta_n = \theta(unknown),$$
versus the alternative

\[ H_1 : \theta_1 = \ldots = \theta_{k_1} \neq \theta_{k_1+1} = \ldots = \theta_{k_2} \neq \ldots \neq \theta_{k_{q-1}} = \ldots = \theta_{k_q} \neq \theta_{k_q+1} \ldots = \theta_n, \]

where \( q \) and \( k_1, k_2, \ldots, k_q \) are unknown and need to be estimated.

Empirical likelihood introduced by Owen (1988, 1990) is one of the popular and powerful nonparametric approaches. It has been widely used due to the robustness of its nonparametric nature and the efficiency of its likelihood construction. Kolaczyk (1994) used empirical likelihood with generalized linear models. Further, Qin and Lawless (1994) obtained estimating equations and derived asymptotic properties of the test statistic. Many scholars have discussed about the empirical likelihood ratio test for a change point in linear models, such as Zou et al. (2007), Liu et al. (2008), and Ning (2012). Since the empirical likelihood was originally proposed for independent data, it is difficult to apply it to dependent data such as time series data. Several approaches suggested to reduce the dependent data problem into an independent data problem. Owen (2001) suggested using the conditional likelihood to remove the dependence structure and generate the estimating equations. Kitamura (1997) used block-wise empirical likelihood method which preserves the dependence of data, and the resulting likelihood ratios have been used to construct asymptotically valid confidence intervals. Ogata (2005) and Nordman and Lahiri (2006) independently formulated a frequency domain empirical likelihood (FDEL) using spectral estimating equations which can be used for short- and long-range dependent data. Bai and Perron (1998) proposed CUSUM and F-based statistics for change point
detection. Baragona et al. (2013) compared it with the test they proposed for change point detection based on the empirical likelihood approach for change point detection.

To deal with the situation of multiple changes, it traditionally uses the binary segmentation method proposed by Vostrikova (1981). The advantage of using this method is that it detects number of change points and estimates their locations simultaneously as well as the consistency of this method has been established. Hence, the general hypothesis of the change point problem can be simplified as the hypothesis of no change point versus a single change point, i.e. the alternative hypothesis is:

$$H_1 : \theta_1 = \ldots = \theta_k \neq \theta_{k+1} = \ldots = \theta_n,$$

where $k$ is the location of the single change point at this stage. If $H_0$ is not rejected, then the process is stopped and we conclude that there is no change. If $H_0$ is rejected, then there is a change point and the two subsequences before and after the change point found are tested for a change. This process is repeated until there are no subsequences having change points.

In this paper, we propose a test statistic based on the empirical likelihood approach for detecting changes in a time series model. In Section 2, the change point problem in time series models has been introduced for AR(p) model. The empirical likelihood procedure for change point detection is described in Section 3. The null asymptotic distribution of the test statistic and the consistency of the test along with the proofs are provided under Section 4. Simulations are carried out in Section 5 and a real data application is given.
in Section 6. Section 7 provides some discussion and proofs of results are given in the Appendix.

2 Changepoint Problem in AR(p) Model

Consider the stationary AR(p) model with the mean 0.

\[
X_t = \begin{cases} 
\sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t; & 1 \leq t \leq k \\
\sum_{i=1}^{p} \phi_i^* X_{t-i} + \epsilon_t; & k + 1 \leq t \leq n,
\end{cases}
\]

where \(\epsilon_t\)'s are independent random variables with mean zero and variance \(\sigma^2\), (i.e. White noise process), \(\phi_1, \phi_2, ..., \phi_p, \phi_1^*, \phi_2^*, ..., \phi_p^*\) are all unknown parameters, and \(k\) is the unknown change location which needs to be estimated. Denote \(\delta = \Phi^* - \Phi\), where \(\Phi = (\phi_1, \phi_2, ..., \phi_p)'\) and \(\Phi^* = (\phi_1^*, \phi_2^*, ..., \phi_p^*)'\). Therefore, the change point problem is to test the null hypothesis of no change in the autoregressive parameters versus the alternative hypothesis of one unknown change, i.e.,

\[
H_0 : \delta = 0 \quad \text{vs} \quad H_1 : \delta \neq 0, \text{ at least one } \delta_i \neq 0.
\]

Hence, under the alternative hypothesis, there is a change in at least one of the \(p\) parameters at an unknown location. We denote \(\beta_0 = (\Phi, 0)' \in \mathbb{R}^{2p}\) and \(\beta_1 = (\Phi, \delta)' \in \mathbb{R}^{2p}\) to be the parameter vectors under the null and the alternative hypothesis respectively. According to Owen (1991), we derive the estimating functions to be

\[
g_1(X_t, \beta_0) = (X_t, X_{t-1} \epsilon_t, ..., X_{t-p} \epsilon_t, \epsilon_t^2 - \sigma^2), \quad (1)
\]
where \( \epsilon_i = X_i - \sum_{r=1}^{p} \phi_i X_{i-r} \), and

\[
g_2(X_j, \beta_1) = (X_j, X_{j-1}\epsilon_j, \ldots, X_{j-p}\epsilon_j, \epsilon_j^2 - \sigma^2),
\]

where \( \epsilon_j = X_j - \sum_{r=1}^{p} \phi_j^* X_{j-r} \). It is easy to see that

\[
E[g_1(X_i, \beta_0)] = 0,
\]

\[
E[g_2(X_j, \beta_1)] = 0.
\]

for every \( 1 \leq i \leq k \) and \( k + 1 \leq j \leq n \).

### 3 Empirical Likelihood for AR(p) Changepoint Model

WLOG, we assume one change point at an unknown location \( k \). Let

\[
\Omega_{H_0} = \{ (p, q, \beta_0) | \sum_i p_i g_1(X_i, \beta_0) = \sum_j q_j g_2(X_j, \beta_0) = 0 \},
\]

and

\[
\Omega_{H_1} = \{ (p, q, \beta_1) | \sum_i p_i g_1(X_i, \beta_0) = 0, \sum_j q_j g_2(X_j, \beta_1) = 0 \}
\]

be the parameter spaces under \( H_0 \) and \( H_1 \), respectively, where \( p = (p_1, p_2, \ldots, p_k) \) and \( q = (q_{k+1}, q_{k+2}, \ldots, q_n) \) are the probability vectors such that \( \sum_{i=1}^{k} p_i = 1, \sum_{j=k+1}^{n} q_j = 1 \) and \( p_i \geq 0, q_j \geq 0 \). If a change occurs at \( k \), then the empirical likelihood ratio test statistic is defined as,

\[
-2 \log \Lambda_k = -2 \log \sup_{H_0} \left\{ \prod_i p_i \prod_j q_j | (p, q, \beta_0) \in \Omega_{H_0} \right\} / \sup_{H_1} \left\{ \prod_i p_i \prod_j q_j | (p, q, \beta_1) \in \Omega_{H_1} \right\}
\]
\[ Z_{H_0,k} - Z_{H_1,k}, \]

where

\[ Z_{H_0,k} = -2\sup_{H_0} \left\{ \sum_i \log kp_i + \sum_j \log(n-k)q_j | (p, q, \beta_0) \in \Omega_{H_0} \right\}, \]

\[ Z_{H_1,k} = -2\sup_{H_1} \left\{ \sum_i \log kp_i + \sum_j \log(n-k)q_j | (p, q, \beta_1) \in \Omega_{H_1} \right\}. \]

The null hypothesis is rejected for a sufficiently large value of \( \max_{1<k<n} - 2 \log \Lambda_k \). Let \( \theta_{nk} = \frac{k}{n} \). A Lagrangian argument gives,

\[ p_i = \frac{1}{n\theta_{nk}(1 + \theta_{nk}^{-1} \lambda_1 g_1(X_i, \cdot))} \]

and

\[ q_j = \frac{1}{n(1 - \theta_{nk})(1 + (1 - \theta_{nk})^{-1} \lambda_2 g_2(X_j, \cdot))} \]

where \( \lambda_1 \) and \( \lambda_2 \) are chosen such that \( \sum_i p_i g_1(X_i, \cdot) = 0 \) and \( \sum_j q_j g_2(X_j, \cdot) = 0 \). Therefore, under \( H_0 \), we obtain

\[ Z_{H_0,k} = 2\inf_{\beta_0, \lambda_1, \lambda_2} \left\{ \sum_i \log(1 + \theta_{nk}^{-1} \lambda_1 g_1(X_i, \beta_0)) + \sum_j \log(1 + (1 - \theta_{nk})^{-1} \lambda_2 g_2(X_j, \beta_0)) \right\}. \]

Let \( \lambda = (\lambda_1', \lambda_2')' \). The score functions are defined as:

\[ Q_{1n}(\beta_0, \lambda) = \frac{\partial Z_{H_0,k}}{\partial \lambda} = \frac{1}{n} \sum_m \frac{1}{1 + \theta_{m}^{-1} X g(X_m, \beta_0)} \theta_{m}^{-1} g(X_m, \beta_0), \]

and

\[ Q_{2n}(\beta_0, \lambda) = \frac{\partial Z_{H_0,k}}{\partial \beta_0} = \frac{1}{n} \sum_m \frac{1}{1 + \theta_{m}^{-1} X g(X_m, \beta_0)} \theta_{m}^{-1} \left( \frac{\partial g(X_m, \beta_0)}{\partial \beta_0} \right)' g(X_m, \beta_0), \]

8
where
\[
\theta^{-1}_m = \theta^{-1}_{nk}1_{\{1 \leq m \leq k\}} + (1 - \theta_{nk})^{-1}1_{\{k+1 \leq m \leq n\}},
\]
and
\[
g(X_m, \beta_0) = g_1(X_m, \beta_0)1_{\{1 \leq m \leq k\}} + g_2(X_m, \beta_0)1_{\{k+1 \leq m \leq n\}}.
\]
Under certain regularity conditions, Qin and Lawless (1994) showed, there exists \((\tilde{\beta}_0, \tilde{\lambda})\) such that,
\[
Q_{1n}(\tilde{\beta}_0, \tilde{\lambda}) = 0 \text{ and } Q_{2n}(\tilde{\beta}_0, \tilde{\lambda}) = 0.
\]
Hence, we obtain
\[
Z_{H_0,k} = 2l_E(\tilde{\Phi}, \tilde{\mu}, 0).
\]
Similarly, under \(H_1\) we have, \(Z_{H_1,k} = 2l_E(\tilde{\Phi}, \mu, \delta)\). Then the empirical likelihood ratio statistic can be rewritten as
\[
-2 \log \Lambda_k = 2l_E(\tilde{\Phi}, \tilde{\mu}, 0) - 2l_E(\tilde{\Phi}, \mu, \delta).
\] (5)
Since \(k\) is unknown, \(H_0\) is rejected when the maximally selected log-likelihood ratio statistic,
\[
Z_n = \max_{\theta_{nk} \in \Theta_n} \{-2 \log \Lambda_k\},
\]
where \(\Theta_n = \{k/n : k = 1, 2, \ldots, n\}\), is sufficiently large.

When \(k\) or \(n - k\) is too small, then the minimax estimators of empirical likelihood \((\tilde{\beta}_1, \tilde{\lambda})\) may not exist. Hence we consider the trimmed likelihood ratio statistic where the range of \(k\) is selected arbitrarily as follows. The Trimmed likelihood ratio statistic is defined as,
\[
Z^*_n = \max_{\theta_{nk} \in \Theta^*_n} \{-2 \log \Lambda_k\},
\] (6)
where $\Theta^*_{nk} = \{k/n : k = n_{T_1}, n_{T_1} + 1, \ldots, n - n_{T_2}\}$. According to Perron and Vogelsang (1992), the selection of $n_{T_1}$ and $n_{T_2}$ can be arbitrary. In our work, we choose $n_{T_1} = n_{T_2} = 2[n^{\frac{1}{2}}]$, where $[x]$ means the largest integer not larger than $x$. If $H_0$ is true, then $Z^*_n$ follows an asymptotic extreme value limit distribution. The convergence to the extreme value limit can be slow and asymptotic test often tends to be too conservative in finite samples.

4 Main Results

The results are similar to the ones by Csörgő and Horváth (1997). Under mild regularity conditions, the following theorem holds.

**Theorem 1.** Let $\beta^*$ be the true parameter. Suppose that $E||g(X, \beta^*)||^3 < \infty$, $E||\epsilon||^4 < \infty$, and $E(\epsilon \epsilon')$ is positive definite. If $H_0$ is true, then we have

$$\lim_{n \to \infty} Pr\{A(\log u(n))(Z^*_n)^{\frac{1}{2}} \leq t + D_r(\log u(n))\} = \exp(-e^{-t})$$

for all $t$, where $A(x) = (2 \log x)^{\frac{1}{2}}$, $D_r(x) = 2 \log x + (r/2) \log \log x - \log \Gamma(r/2)$, $u(n) = \frac{n^2 + 2(n^{\frac{1}{2}})^2 - 2n^{\frac{1}{2}}}{(2n^{\frac{1}{2}})^2}$, and $r$ is the dimension of the parameter $\delta$.

**Theorem 2.** Under the conditions of Theorem 1 and the condition that for every fixed parameter $\delta = \beta^* - \beta \neq 0$, there exists a positive constant $c_0 > 0$ satisfy that $\inf_{\delta \neq 0} \sup_{\lambda} E \log \left[1 + \lambda' x(x' \delta + e)\right] \geq c_0 > 0$, holds. If $H_1$ is true, assume that $\theta_k = \frac{k_0}{n} \to \theta_0 \in (0, 1)$ as $n \to \infty$, then ELR test statistic is consistent, i.e. there exists a constant $c > 0$ such that

$$P(Z_n > cn) \to 1.$$
Theorem 3. Under the conditions of Theorem 1 and the condition that for every fixed parameter $\delta = \beta^* - \beta \neq 0$, there exists a positive constant $c_0 > 0$ satisfy that

$$\inf_{\delta \neq 0} \sup_\lambda E \log \left[ 1 + \lambda' x(x' \delta + \epsilon) \right] \geq c_0 > 0,$$

holds. If $H_1$ is true, assume that $\theta_k = \frac{kn}{n} \to \theta_0 \in (0, 1)$ as $n \to \infty$, we have $\hat{\theta}_k \to \theta_0$ in probability as $n \to \infty$.

Proofs are given in the Appendix.

5 Simulation Study

A Monte Carlo simulation has been conducted to illustrate the performance of the proposed method. Consider the following AR(1) model with mean 0:

$$X_t = \begin{cases} 0.1X_{t-1} + \epsilon_t; & 1 \leq t \leq k \\ 0.5X_{t-1} + \epsilon_t; & k+1 \leq t \leq n, \end{cases}$$

where $\epsilon_t$ is the white noise with mean zero and variance $\sigma^2$. Four different distributions are considered for $\epsilon_t$: (i) $N(0, 1)$, (ii) $\exp(1) - 1$, (iii) $\frac{1}{2\sqrt{2}}(\chi^2_4 - 4)$, and (iv) $\frac{1}{\sqrt{2}}t_4$. The power of the proposed test in detecting changes in parameters of the AR(1) model has been calculated for two different sample sizes: $n=100$, 150 and 250. Different change locations have been considered under each sample size. Additional simulations have been carried out to compute the empirical critical values under different significance levels which are turned out to be close to the theoretical critical values for the corresponding significance level. Hence, we use the theoretical critical value 2.9702 with $\alpha = 0.05$ for power calculations with 1000 simulations. The results are listed in Table 1. It can be
seen that the power of the hypothesis test of AR(1) model increases with the sample size. The power values under a given change location are approximately similar for the four different error distributions. This maybe due to the fact that the three distributions $\exp(1) - 1$, $\frac{1}{2\sqrt{2}}(\chi_4^2 - 4)$ and $\frac{1}{\sqrt{2}}t_4$ are standardized. When the change location is farther away from the starting location, then the power tends to decrease. Intuitively, this maybe due to the dependency existing in the data set.
Table 1: *Power of the hypothesis test for AR(1) model*

| k   | $N(0, 1)$ | $\exp(1) - 1$ | $\frac{1}{2\sqrt{2}}(\chi^2_4 - 4)$ | $\frac{1}{\sqrt{2}}t_4$ |
|-----|-----------|----------------|--------------------------------------|------------------------|
|     | n = 100   |                |                                      |                        |
| 20  | 0.802     | 0.816          | 0.815                                | 0.808                  |
| 30  | 0.765     | 0.775          | 0.774                                | 0.779                  |
| 40  | 0.723     | 0.732          | 0.754                                | 0.747                  |
| 50  | 0.656     | 0.669          | 0.627                                | 0.674                  |
| 80  | 0.296     | 0.292          | 0.283                                | 0.331                  |
|     | n = 150   |                |                                      |                        |
| 30  | 0.929     | 0.929          | 0.924                                | 0.903                  |
| 45  | 0.913     | 0.901          | 0.893                                | 0.872                  |
| 60  | 0.862     | 0.844          | 0.853                                | 0.845                  |
| 75  | 0.806     | 0.799          | 0.813                                | 0.804                  |
| 120 | 0.385     | 0.397          | 0.426                                | 0.450                  |
|     | n = 250   |                |                                      |                        |
| 50  | 0.993     | 0.988          | 0.990                                | 0.976                  |
| 80  | 0.983     | 0.981          | 0.976                                | 0.965                  |
| 100 | 0.966     | 0.966          | 0.976                                | 0.948                  |
| 125 | 0.941     | 0.927          | 0.928                                | 0.926                  |
| 200 | 0.621     | 0.603          | 0.626                                | 0.622                  |
6 Application

In this section, we study the data which consists of monthly average soybean prices achieved by farmers in Illinois from January 1960 to November 2008 with the sample size 587. The prices are given in dollars per bushel. This data was analyzed by Balcombe et al. (2007) who considered the threshold AR(1) models for modeling the prices of agricultural products. Berkes et al. (2011) studied this data set by proposing the likelihood ratio test to detect the structural change of an AR model to threshold AR model. We apply the proposed EL method for AR(1) changepoint model to detect the structural change in the same data set. Figure 1 shows the time series plot for the given data. In order to test if there are significant changes, we use the $Z_n^*$ from (6). The value of $Z_n^*$ is 16.07426. Using the critical values derived under the Theorem are given in Table 2, we have sufficient evidence to reject the null hypothesis that there is no change.

Figure 1: Time Series for the Monthly average soybean prices

![Monthly Average Soybean Prices](image)
Table 2: Theoretical Critical values

| α     | Critical Value |
|-------|----------------|
| 0.01  | 4.600149       |
| 0.05  | 2.970195       |
| 0.10  | 2.250367       |

7 Discussion

In this paper, we discuss developing an EL-based detecting procedure for structural changes in time series data, i.e. testing null hypothesis of no change versus alternative hypothesis of one change. A test statistic is derived for a fixed change location and the max-type of test statistic over all possible change locations is considered. The asymptotic null distribution of the test statistic has been established as extreme value distribution. Simulations to compute the power in AR(1) model have been carried out with different sample sizes and different error distributions in order to illustrate the performance of the proposed test statistic. The results indicate that the proposed method is efficiently identify the changes in a given time series data set. We should point out that, due to the slow convergence of the proposed test statistic in Theorem 1, the moderate or the large sample size is recommended to achieve the good approximation (See Csörgő and Horváth, 1997). If the sample size is small, the bootstrap is suggested to obtain the approximated p-values in practice.

As for future work, we plan to extend the proposed method to other stationary time series models such as MA, ARMA, GARCH models along with corresponding analytic
results and simulations. Comparisons to other existing methods will be done. Further, sequential change point detection based on EL method is to be studied where the sample size is a random variable and the null hypothesis of sequential structural stability will be rejected as soon as a change is detected. Hence, the objective in sequential change point detection is to detect such a change with a minimum number of false alarms. A nonparametric testing procedure based on EL method will be proposed and related asymptotic results will be studied.

Appendix

In order to prove Theorem 1, we need following Lemmas.

**Lemma 1.** Assume that for \(i = 1, 2\), \(E[g_i(X, \beta_0)g'_i(X, \beta_0)]\) is positive definite, \(\frac{\partial g_i(X, \beta)}{\partial \beta}\) is continuous in a neighborhood of the true value \(\beta_0\), \(E\left[\left(\frac{\partial g_i(X, \beta_0)}{\partial \beta}\right)\left(\frac{\partial g_i(X, \beta_0)}{\partial \beta}\right)'\right]\), \(E\left[\frac{\partial^2 g_i(X, \beta)}{\partial \beta \partial \beta'}\right]\), \(E\left[\left(\frac{\partial g_i(X, \beta)}{\partial \beta}\right)'g_i(X, \beta)\right]\) and \(E \| g_i(X, \beta) \|^3\) are all bounded in the neighborhood of the true value \(\beta_0\). Then, as \(n \to \infty\), \(\exists \tilde{\beta}, \tilde{\lambda} = \lambda(\tilde{\beta})\) with probability 1 satisfying,

\[
Q_{1n}(\tilde{\beta}, \tilde{\lambda}) = 0, \quad Q_{2n}(\tilde{\beta}, \tilde{\lambda}) = 0 \quad \text{and} \quad \| \tilde{\beta} - \beta_0 \| = O_p(m^{-\frac{1}{2}}),
\]

where

\[
Q_{1n}(\beta, \lambda) = \sum_i \frac{1}{1 + \lambda'(\beta)\theta_i^{-1}g(x_i, \beta)}\theta_i^{-1}g(x_i, \beta),
\]

\[
Q_{2n}(\beta, \lambda) = \sum_i \frac{1}{1 + \lambda(\beta)\theta_i^{-1}g(x_i, \beta)}\theta_i^{-1}\left(\frac{\partial g_i(x_i, \beta)}{\partial \beta}\right)'\lambda(\beta).
\]
Proof. First we will show

\[ \lambda(\beta) = \epsilon_k O_p(m^{-\frac{1}{2}}) \]

\[ = \left[ \frac{1}{n} \sum_{i=1}^{n} \theta_i^{-2} g(x_i, \beta) g'(x_i, \beta) \right]^{-1} \left[ \frac{1}{n} \sum_{i=1}^{n} \theta_i^{-1} g(x_i, \beta) \right] + \epsilon_k O_p(m^{-\frac{1}{2}}), \]

where \( \epsilon_k = \min\{\theta_k, 1 - \theta_k\} \) and \( m = n\epsilon_k = \min\{k, n - k\} \).

Let \( \beta - \beta_0 = u m^{-\frac{2}{3}} \) for \( \beta \in \{\beta : \|\beta - \beta_0\| = m^{-\frac{1}{2}}\} \) where \( \|u\| = 1 \). Let \( \lambda \) be the solution of the function \( f(\lambda) \) given by the first score function defined in Section 3.

\[ f(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \frac{\theta_i^{-1}}{1 + \lambda' \theta_i^{-1} g(x_i, \beta)} g(x_i, \beta) = 0. \] (A.1)

Let \( \lambda = \rho u \) where \( u = (\beta - \beta_0)m^{\frac{1}{2}} \) and \( \|u\| = 1 \).

\[ 0 = \| f(\rho u) \| \]

\[ \geq |u' f(\rho u)| \]

\[ = \frac{1}{n} \left| u' \left( \sum_{i} \theta_i^{-1} g(x_i, \beta) - \rho \sum_{i} \frac{\theta_i^{-2} g(x_i, \beta) u' g(x_i, \beta)}{1 + \rho u' \theta_i^{-1} g(x_i, \beta)} \right) \right| \]

\[ \geq \frac{\rho}{n} u' \left( \sum_{i} \frac{\theta_i^{-2} g(x_i, \beta) u' g(x_i, \beta)}{1 + \rho u' \theta_i^{-1} g(x_i, \beta)} \right) u - \frac{1}{n} \left( \sum_{j=1}^{p} e_j \sum_{i} \theta_i^{-1} g(x_i, \beta) \right) \]

( where \( e_j \) is the unit vector in the \( j^{th} \) coordinate direction.)

\[ \geq \frac{\rho u' Su}{1 + \rho \theta g^*} - O_p(m^{-\frac{1}{2}}), \]

( where \( g^* = \max g(x_i, \beta) \) and \( S = \frac{1}{n} \sum_{i} \theta_i^{-2} g(x_i, \beta) g'(x_i, \beta) \).

Since \( u'Su \geq \sigma_p + o_p(1) \), where \( \sigma_p > 0 \) is the smallest eigen value of \( \Sigma \), then

\[ \frac{\rho}{1 + \rho \theta g^*} = O_p(m^{-\frac{1}{2}}) \]
So, \( \| \lambda \| = \rho = O_p(m^{-\frac{1}{2}}) \).

Let \( \gamma_l = \lambda'(\beta) \theta_l^{-1} g(x_l, \beta) \). Then, \( \max_l |\gamma_l| = O_p(m^{-\frac{1}{2}}) o(m^{\frac{1}{2}}) = o_p(1) \).

Expanding (A.1),

\[
0 = f(\lambda) = \frac{1}{n} \sum_i \theta_i^{-1} g(x_i, \beta) [1 - \gamma_i + \frac{\gamma_i^2}{1 + \gamma_i}]
\]

\[
= \frac{1}{n} \sum_i \theta_i^{-1} g(x_i, \beta) - \frac{1}{n} \sum_i \theta_i^{-1} g(x_i, \beta) \cdot \gamma + \frac{1}{n} \sum_i \theta_i^{-1} g(x_i, \beta) \frac{\gamma_i^2}{1 + \gamma_i}
\]

\[
= E(\theta_l^{-1} g(x_l, \beta)) - S\lambda + \frac{1}{n} \sum_i \theta_i^{-1} g(x_i, \beta) \frac{\gamma_i^2}{1 + \gamma_i}. \tag{A.2}
\]

The last equality is since \( \frac{1}{n} \sum_i \theta_i^{-1} g(x_i, \beta) \cdot \gamma = \frac{1}{n} \sum_i \theta_i^{-1} g(x_i, \beta) \theta_i^{-1} g'(x_i, \beta) \lambda = S\lambda \).

By substituting \( \gamma_l \), we have the final term of (A.2):

\[
\frac{1}{n} \sum_i \| \theta_i^{-1} g(x_i, \beta) \|^3 \| \lambda \|^2 [1 + \gamma_i]^{-1} = o_p(m^\frac{3}{2})O_p(m^{-1})o_p(1) = o_p(m^{-\frac{3}{2}}).
\]

Therefore,

\[
0 = E(\theta_l^{-1} g(x_l, \beta)) - S\lambda + o_p(m^{-\frac{1}{2}})
\]

\[
\Rightarrow \lambda = S^{-1} E(\theta_l^{-1} g(x_l, \beta)) + o_p(m^{-\frac{1}{2}})
\]

\[
\Rightarrow \lambda = \left[ \frac{1}{n} \sum_{i=1}^n \theta_i^{-2} g(x_i, \beta) g'(x_i, \beta) \right]^{-1} \left[ \frac{1}{n} \sum_{i=1}^n \theta_i^{-1} g(x_i, \beta) \right] + o_p(m^{-\frac{1}{2}}). \tag{A.3}
\]

Now, denote \( V_n(\beta) = \frac{1}{n} \sum_{i=1}^n \theta_i^{-2} g(x_i, \beta) g'(x_i, \beta) \), \( \bar{g}(\beta) = \frac{1}{n} \sum_{i=1}^n \theta_i^{-1} g(x_i, \beta) \), and \( \epsilon = \epsilon_k o_p(m^{-\frac{1}{2}}) \). So (A.2) can be rewritten as,

\[
\lambda(\beta) = V_n(\beta)^{-1} \bar{g}(\beta) + \epsilon.
\]

Since \( \gamma_l = \lambda'(\beta) \theta_l^{-1} g(x_l, \beta) \), so \( \sum_{l=1}^n |r_l|^3 = o_p(1) \).

Let \( a_m \) be any constant sequence such that \( a_m \to \infty \), and \( a_m m^{-\frac{1}{2}} \to 0 \). Denote the ball
\[ B(\beta_0, a_m) = \{ \beta \mid \beta - \beta_0 \leq a_m m^{-\frac{1}{2}} \} \] and the surface of the ball \( \partial B(\beta_0, a_m) = \{ \beta \mid \beta - \beta_0 \leq \phi a_m m^{-\frac{1}{2}}, \| \phi \| = 1 \}. \) For any \( \beta \in \partial B(\beta_0, a_m) \), we have

\[
V_n(\beta) = \frac{1}{n} \sum_{l=1}^{n} \theta_l^{-2} g(x_l, \beta) g'(x_l, \beta)
\]
\[
= \frac{n}{k} \sum_{l=1}^{k} g_1(x_l, \beta_0) g_1'(x_l, \beta_0) + \frac{n}{n-k} \sum_{l=k+1}^{n} g_2(x_l, \beta_0) g_2'(x_l, \beta_0) + o_p(\epsilon_k^{-1})
\]
\[
= \frac{n}{k} E g_1(x_l, \beta_0) g_1'(x_l, \beta_0) + \frac{n}{n-k} E g_2(x_l, \beta_0) g_2'(x_l, \beta_0) + o_p(\epsilon_k^{-1})
\]
\[
\leq \epsilon_k^{-1} [E g_1(x_l, \beta_0) g_1'(x_l, \beta_0) + E g_2(x_l, \beta_0) g_2'(x_l, \beta_0)] + o_p(\epsilon_k^{-1}),
\]

and

\[
\bar{g}(\beta_0) = \frac{1}{n} \sum_{l=1}^{n} \theta_l^{-1} g(x_l, \beta)
\]
\[
= \frac{1}{k} \sum_{l=1}^{k} g_1(x_l, \beta_0) + \frac{1}{n-k} \sum_{l=k+1}^{n} g_2(x_l, \beta_0)
\]
\[
= \frac{1}{k} o_p(k^{\frac{1}{2}}) + \frac{1}{n-k} o_p((n-k)^{\frac{1}{2}})
\]
\[
= o_p(k^{-\frac{1}{2}}) + o_p((n-k)^{-\frac{1}{2}})
\]
\[
= o_p(m^{-\frac{1}{2}}).
\]

By the Taylor expansion, for any \( \beta \in \partial B(\beta_0, a_m) \), we have

\[
l_E(\beta) = \sum_i \lambda'(\beta) \theta_i^{-1} g(x_i, \beta) - \frac{1}{2} \sum_i \left[ \lambda'(\beta) \theta_i^{-1} g(x_i, \beta) \right] + o_p(1). \quad (A.4)
\]

The first term of (A.4) is:

\[
\sum_i \lambda'(\beta) \theta_i^{-1} g(x_i, \beta) = \left[ \frac{1}{n} \sum_{l=1}^{n} \theta_l^{-1} g(x_l, \beta) \right] \left[ \frac{1}{n} \sum_{l=1}^{n} \theta_l^{-2} g(x_l, \beta) g'(x_l, \beta) \right]^{-1} \left[ \frac{1}{n} \sum_{l=1}^{n} \theta_l^{-1} g(x_l, \beta) \right]
\]
The second term of (A.4) is:

\[
\frac{1}{2} \sum_l \left[ X'(\beta) \theta_l^{-1} g(x_l, \beta) \right]^2
\]

\[
= \frac{1}{2} \sum_l X'(\beta) \theta_l^{-2} g(x_l, \beta) g'(x_l, \beta)
\]

\[
= \frac{n}{2} \left\{ \frac{1}{n} \sum_l \theta_l^{-1} g(x_l, \beta) \right\}' \left\{ \frac{1}{n} \sum_l \theta_l^{-2} g(x_l, \beta) g'(x_l, \beta) \right\}^{-1} \left\{ \frac{1}{n} \sum_l \theta_l^{-1} g(x_l, \beta) \right\} + o_p(1)
\]

\[
\left[ \frac{1}{n} \sum_l \theta_l^{-2} g(x_l, \beta) g'(x_l, \beta) \right] \left[ \frac{1}{n} \sum_l \theta_l^{-2} g(x_l, \beta) g'(x_l, \beta) \right]^{-1} \left[ \frac{1}{n} \sum_l \theta_l^{-1} g(x_l, \beta) \right] + o_p(1)
\]

\[
= \frac{n}{2} \left\{ \frac{1}{n} \sum_l \theta_l^{-1} g(x_l, \beta) \right\}' \left\{ \frac{1}{n} \sum_l \theta_l^{-2} g(x_l, \beta) g'(x_l, \beta) \right\}^{-1} \left[ \frac{1}{n} \sum_l \theta_l^{-1} g(x_l, \beta) \right] + o_p(1).
\]

Now,

(A.4.1) - (A.4.2)

\[
= \frac{n}{2} \left\{ \frac{1}{n} \sum_l \theta_l^{-1} g(x_l, \beta) \right\}' \left\{ \frac{1}{n} \sum_l \theta_l^{-2} g(x_l, \beta) g'(x_l, \beta) \right\}^{-1} \left[ \frac{1}{n} \sum_l \theta_l^{-1} g(x_l, \beta) \right] + o_p(1).
\]

So we can rewrite (A.4) as,

\[
I_E(\beta) = \frac{n}{2} \left\{ \frac{1}{n} \sum_l \theta_l^{-1} g(x_l, \beta) \right\}' \left\{ \frac{1}{n} \sum_l \theta_l^{-2} g(x_l, \beta) g'(x_l, \beta) \right\}^{-1} \left[ \frac{1}{n} \sum_l \theta_l^{-1} g(x_l, \beta) \right] + o_p(1)
\]

\[
= \frac{n}{2} \tilde{g}'(\beta) (V_n(\beta))^{-1} \tilde{g}(\beta) + o_p(1)
\]

\[
= \frac{n}{2} \left\{ \tilde{g}(\beta_0) + \frac{1}{n} \sum_l \theta_l^{-1} \frac{\partial g(x_l, \beta_0)}{\partial \beta'} \phi a_m m^{-\frac{1}{2}} + O \left[ (a_m m^{-\frac{1}{2}})^2 \right] \right\}' \times (V_n(\beta))^{-1} \times \left\{ \tilde{g}(\beta_0) + \frac{1}{n} \sum_l \theta_l^{-1} \frac{\partial g(x_l, \beta_0)}{\partial \beta'} \phi a_m m^{-\frac{1}{2}} + O \left[ (a_m m^{-\frac{1}{2}})^2 \right] \right\} + o_p(1)
\]

(By Taylor expansion of each term.)
\[
\geq \frac{n\epsilon_k}{2} \left\{ \bar{g}(\beta_0) + \frac{1}{n} \sum_l \theta_l^{-1} \frac{\partial g(x_l, \beta_0)}{\partial \beta} \phi a_m m^{-\frac{1}{2}} + O\left( (a_m m^{-\frac{1}{2}})^2 \right) \right\} \times (V_n(\beta))^{-1} \times \left\{ \bar{g}(\beta_0) + \frac{1}{n} \sum_l \theta_l^{-1} \frac{\partial g(x_l, \beta_0)}{\partial \beta} \phi a_m m^{-\frac{1}{2}} + O\left( (a_m m^{-\frac{1}{2}})^2 \right) \right\} + o_p(1).
\]

As \( n \to \infty \), \( I_E(\beta) \to \infty \).

Similarly,

\[
l_E(\beta_0) = \frac{n}{2} \bar{g}'(\beta_0) V_n(\beta_0)^{-1} \bar{g}(\beta_0) + o_p(1),
\]

\[
V_n(\beta_0) = \frac{n}{k} E g_1(x_l, \beta_0) g_1(x_l, \beta_0) + \frac{n}{n-k} E g_2(x_l, \beta_0) g_2(x_l, \beta_0) + o_p(\epsilon_k^{-1}).
\]

Thus, \( l_E(\beta_0) = O_p(1) \) implies that for any \( \beta \in \partial B(\beta_0, a_m) \), \( l_E(\beta) \) can not arrive its minimum value with the probability approaching to 1. Since \( l_E(\beta) \) is a continuous function about \( \beta \), as \( \beta \in B(\beta_0, a_m) \), \( l_E(\beta) \) has a minimum value in the interior of this ball satisfying,

\[
0 = \frac{\partial l_E(\beta)}{\partial \beta} \bigg|_{\beta=\hat{\beta}} = \sum_l \left( \frac{\partial \lambda(\beta)}{\partial \beta} \right) \theta_l^{-1} g(x_l, \beta) + \theta_l^{-1} \left( \frac{\partial g(x_l, \beta)}{\partial \beta} \right)' \lambda(\beta) = \sum_l \left( \frac{\partial \lambda(\beta)}{\partial \beta} \right) \theta_l^{-1} g(x_l, \beta) + \sum_l \theta_l^{-1} \left( \frac{\partial g(x_l, \beta)}{\partial \beta} \right)' \lambda(\beta) = Q_{1n}(\hat{\beta}, \hat{\lambda}) = 0
\]

Hence, \( Q_{1n}(\hat{\beta}, \hat{\lambda}) = 0 \) and \( Q_{2n}(\hat{\beta}, \hat{\lambda}) = 0 \). That is, \( \| \hat{\beta} - \beta_0 \| = O_p(a_m m^{-\frac{1}{2}}) \). But \( a_m \) is arbitrary, hence \( \| \hat{\beta} - \beta_0 \| = O_p(m^{-\frac{1}{2}}) \).
Remark 1. From Lemma 3 of Arichison and Silvey (1957), the partitioned matrix

\[
\begin{pmatrix}
  A & B \\
  B' & 0
\end{pmatrix}
\]

is non-singular. Hence,

\[
\begin{pmatrix}
  A & B \\
  B' & 0
\end{pmatrix}^{-1} = \begin{pmatrix}
  P & Q \\
  Q' & R
\end{pmatrix}
\]

where

\[
P = A^{-1} - A^{-1}B(B'A^{-1}B)^{-1}B'A^{-1}, \quad Q = A^{-1}B(B'A^{-1}B)^{-1},
\]

\[
Q' = (B'A^{-1}B)^{-1}B'A^{-1}, \quad R = -(B'A^{-1}B)^{-1}
\]

Remark 2.

If \[
\begin{pmatrix}
  A & B \\
  C & D
\end{pmatrix}
\]
is a n × n symmetric positive definite matrix, and the partitioned matrices \( A \in \mathbb{R}^{m \times m}, B \in \mathbb{R}^{m \times n-m}, \) and \( D \in \mathbb{R}^{(n-m) \times (n-m)}, \) then

1. the matrix \((D - CA^{-1}B)\) is symmetric and positive definite,

2.

\[
\begin{pmatrix}
  A & B \\
  C & D
\end{pmatrix}^{-1} \geq \begin{pmatrix}
  A^{-1} & 0 \\
  0 & 0
\end{pmatrix}
\]

Remark 3. \( \beta' = ((\beta', \mu'), \delta'). \)

\[
\frac{\partial Q_{1n}(\beta, 0)}{\partial \lambda'} = -\frac{1}{n} \sum_l \theta_l^{-2}g(x_l, \beta)g'(x_l, \beta), \quad \frac{\partial Q_{1n}(\beta, 0)}{\partial \beta'} = \frac{1}{n} \sum_l \theta_l^{-1} \frac{\partial g(x_l, \beta)}{\partial \beta'}
\]

\[
\frac{\partial Q_{2n}(\beta, 0)}{\partial \lambda'} = \frac{1}{n} \sum_l \left(\theta_l^{-1} \frac{\partial g(x_l, \beta)}{\partial \beta'}\right)', \quad \frac{\partial Q_{2n}(\beta, 0)}{\partial \beta'} = 0
\]
\[
\begin{pmatrix}
\frac{\partial Q_{1n}}{\partial x'} & \frac{\partial Q_{1n}}{\partial \beta'} \\
\frac{\partial Q_{2n}}{\partial x'} & 0
\end{pmatrix}
\rightarrow
\begin{pmatrix}
S_{11} & S_{12} \\
S_{21} & 0
\end{pmatrix}
= S(\beta) \equiv S
\]

where

\[S_{11}(\beta) = -\theta^{-1}_i E\left[ g_1(x_i, \beta)g'_1(x_i, \beta) \right] - (1 - \theta_i)^{-1} E\left[ g_2(x_i, \beta)g'_2(x_i, \beta) \right],\]
\[S_{12}(\beta) = \theta^{-1}_i E\left[ \frac{\partial g_1(x_i, \beta_i)}{\partial \beta'} \right] + (1 - \theta_i)^{-1} E\left[ \frac{\partial g_2(x_i, \beta_i)}{\partial \beta'} \right],\]
\[S_{21}(\beta) = S'_{12}(\beta),\]
\[S_{12,i}(\beta) = \theta^{-1}_i E\left[ \frac{\partial g_1(x_i, \beta_0)}{\partial \beta'_i} \right] + (1 - \theta_i)^{-1} E\left[ \frac{\partial g_2(x_i, \beta_0)}{\partial \beta'_i} \right], \quad i = 1, 2.\]

By, Remark 1,

\[S^{-1} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & 0 \end{pmatrix}^{-1} = \begin{pmatrix} P & Q \\ Q' & R \end{pmatrix}\]

where

\[P = S_{11}^{-1} - S_{11}^{-1}S_{12}(S_{21}S_{11}^{-1}S_{12})^{-1}S_{21}S_{11}^{-1} = S_{11}^{-1} + S_{11}^{-1}S_{12}\Sigma S_{21}S_{11}^{-1}; \quad \Sigma = (S_{21}(-S_{11}^{-1})S_{12})^{-1},\]
\[Q = -S_{11}^{-1}S_{12}(S_{21}S_{11}^{-1}S_{12})^{-1} = -S_{11}^{-1}S_{12}\Sigma; \quad Q' = -\Sigma S_{21}S_{11}^{-1}; \quad R = -(S_{21}S_{11}^{-1}S_{12})^{-1} = \Sigma.\]

**Lemma 2.** Under the conditions in Lemma 1 and \(H_0\), as \(n \to \infty\) we have

\[\sqrt{n}\Sigma^{-\frac{1}{2}}(\tilde{\beta} - \beta_0) \to N(0, I_{2p+q}),\]

where \(\Sigma = [S_{21}(-S_{11})^{-1}S_{12}]^{-1}.\)
Proof. Expanding $Q_{1n}(\tilde{\beta}, \tilde{\lambda})$ and $Q_{2n}(\tilde{\beta}, \tilde{\lambda})$ at $(\theta_0, 0)$, by the conditions of the $H_0$ and Lemma 1, we have,

$$0 = Q_{1n}(\tilde{\beta}, \tilde{\lambda}) = Q_{1n}(\beta_0, 0) + \frac{\partial Q_{1n}(\beta_0, 0)}{\partial \beta'}(\tilde{\beta} - \beta_0) + \frac{\partial Q_{1n}(\beta_0, 0)}{\partial \lambda'}(\tilde{\lambda} - 0) + O_p(m^{-1}),$$

$$0 = Q_{2n}(\tilde{\beta}, \tilde{\lambda}) = Q_{2n}(\beta_0, 0) + \frac{\partial Q_{2n}(\beta_0, 0)}{\partial \beta'}(\tilde{\beta} - \beta_0) + \frac{\partial Q_{2n}(\beta_0, 0)}{\partial \lambda'}(\tilde{\lambda} - 0) + O_p(m^{-1}),$$

By LLN,

$$\begin{pmatrix} -Q_{1n}(\beta_0, 0) + O_p(m^{-1}) \\ \epsilon_k O_p(m^{-1}) \end{pmatrix} = \begin{pmatrix} \frac{\partial Q_{1n}}{\partial \lambda} & \frac{\partial Q_{1n}}{\partial \beta} \\ \frac{\partial Q_{2n}}{\partial \lambda} & 0 \end{pmatrix} \begin{pmatrix} \tilde{\lambda} \\ \tilde{\beta} - \beta_0 \end{pmatrix}. $$

By Remark 1,

$$\tilde{\beta} - \beta_0 = (0 \, I) S^{-1} \begin{pmatrix} -Q_{1n}(\beta_0, 0) + O_p(m^{-1}) \\ \epsilon_k O_p(m^{-1}) \end{pmatrix}. $$

Therefore,

$$\begin{pmatrix} \tilde{\lambda} \\ \tilde{\beta} - \beta_0 \end{pmatrix} \rightarrow \begin{pmatrix} S_{11}^{-1} + S_{11}^{-1} S_{12} - S_{11}^{-1} S_{12} \Sigma \\ -\Sigma S_{21} S_{11}^{-1} \Sigma \end{pmatrix} \begin{pmatrix} -Q_{1n}(\beta_0, 0) + O_p(m^{-1}) \\ \epsilon_k O_p(m^{-1}) \end{pmatrix}. $$

$$\tilde{\beta} - \beta_0 \rightarrow -\Sigma S_{21} S_{11}^{-1} \begin{pmatrix} -Q_{1n}(\beta_0, 0) + O_p(m^{-1}) \end{pmatrix} + \Sigma \epsilon_k O_p(m^{-1})$$
\[ (S_{21}S_{11}^{-1}S_{12})^{-1} - S_{21}S_{11}^{-1}Q_{1n}(\beta_0, 0) - \Sigma S_{21}S_{11}^{-1}Q_p(m^{-1}) + \epsilon_k O_p(m^{-1}) \]

\[ = \frac{1}{\sqrt{n}}(S_{21}S_{11}^{-1}S_{12})^{-1}S_{21}(-S_{11})^{-1/2}(-S_{11})^{-1/2}\sqrt{n}Q_{1n}(\beta_0, 0) + \epsilon_k O_p(m^{-\frac{1}{2}}) \]

Since \((-S_{11})^{-1/2}\sqrt{n}Q_{1n}(\beta_0, 0) \rightarrow N(0, I_{2p+q}), \sqrt{n}S_{21}S_{11}^{-1}(\tilde{\beta} - \beta_0) \rightarrow N(0, I_{2p+q}). \]

**Lemma 3.**

\[-2 \log \Lambda_k = 2l_E(\tilde{\beta}_1^0, 0) - 2l_E(\tilde{\beta}_1^0, \beta_2^0), \]

where \( \tilde{\beta}_1^0 \) minimizes \( l_E(\beta, 0) \) with respect to \( \beta_1 \) under \( H_0 \),

\[-2 \log \Lambda_k = \left[ (-S_{11})^{-1/2}\sqrt{n}Q_{1n}(\beta_0, 0) \right] \Delta \left[ (-S_{11})^{-1/2}\sqrt{n}Q_{1n}(\beta_0, 0) \right] + O_p(m^{-\frac{1}{2}}) \]

where

\[ \Delta = (-S_{11})^{-1/2}\left\{ S_{12}[S_{21}(-S_{11})^{-1}S_{12}]^{-1}S_{21} - S_{12,1}[S_{21,1}(-S_{11})^{-1}S_{12,1}]^{-1}S_{21,1} \right\} (-S_{11})^{-1/2} \geq 0. \]

**Proof.** Similar to Qin and Lawless (1994), we can derive,

\[ l_E(\beta_1^0, \beta_2^0) = -\frac{n}{2}Q_{1n}'(\beta_0, 0)BQ_{1n}(\beta_0, 0) + O_p(m^{-\frac{1}{2}}), \]

where \( B = S_{11}^{-1} + S_{11}^{-1}S_{12}\Sigma S_{21}S_{11}^{-1} \), and

\[ l_E(\tilde{\beta}_1^0, 0) = -\frac{n}{2}Q_{1n}'(\beta_0, 0)AQ_{1n}(\beta_0, 0) + O_p(m^{-\frac{1}{2}}), \]

where \( A = S_{11}^{-1} + S_{11}^{-1}S_{12,1}(S_{21,1}S_{11}^{-1}S_{12,1})^{-1}S_{21,1}S_{11}^{-1} \). Then,

\[ 2\left[ l_E(\tilde{\beta}_1^0, 0) - l_E(\tilde{\beta}_1^0, \beta_2^0) \right] = \left[ -Q_{1n}'(\beta_0, 0)AQ_{1n}(\beta_0, 0) + O_p(m^{-\frac{1}{2}}) \right] + \]

\[ \left[ nO_{1n}'(\beta_0, 0)BQ_{1n}(\beta_0, 0) + O_p(m^{-\frac{1}{2}}) \right] \]

25
\[
\begin{align*}
&= nQ'(\beta_0, 0)(B - A)Q_{1n}(\beta_0, 0) + O_p(m^{-\frac{1}{2}}) \\
&= nQ'(\beta_0, 0)S_{11}^{-1}\left[S_{12}\Sigma S_{21} - S_{12,1}\Sigma^* S_{12,2}\right]S_{11}^{-1}Q_{1n}(\beta_0, 0) + O_p(m^{-\frac{1}{2}})
\end{align*}
\]

\[B - A = S_{11}^{-1} + S_{11}^{-1}S_{12}\Sigma S_{21}S_{11}^{-1} - S_{11}^{-1}S_{12,1}(S_{21,1}S_{11}^{-1}S_{12,1})^{-1}S_{21,1}S_{11}^{-1}.\]

(So, \(\Sigma^* = (S_{21,1}S_{11}^{-1}S_{12,1})^{-1}\))

\[
\begin{align*}
&= \left[(-S_{11})^{-1/2}\sqrt{n}Q_{1n}(\beta_0, 0)\right]^\top \left[(-S_{11})^{-1/2}\left[S_{12}\Sigma S_{21} - S_{12,1}\Sigma^* S_{12,2}\right]\right]
\end{align*}
\]

\[
(-S_{11})^{-1/2} \left[(-S_{11})^{-1/2}\sqrt{n}Q_{1n}(\beta_0, 0)\right] + O_p(m^{-\frac{1}{2}}).
\]

Take \(\Delta = (-S_{11})^{-1/2}\left[S_{12}\Sigma S_{21} - S_{12,1}\Sigma^* S_{12,2}\right](-S_{11})^{-1/2}.\) Now,

\[
\Delta = (-S_{11})^{-1/2}\left[S_{12}\left(S_{21}(-S_{11})^{-1}S_{12}\right)^{-1}S_{21} - S_{12,1}\left(S_{21,1}S_{11}^{-1}S_{12,1}\right)^{-1}S_{12,2}\right](-S_{11})^{-1/2}
\]

\[
= (-S_{11})^{-1/2}(S_{12,1}, S_{12,2})\left[\left[S_{21}(-S_{11})^{-1}S_{12}\right]^{-1} - \begin{pmatrix} S_{21,1}S_{11}^{-1}S_{12,1} & 0 \\ 0 & 0 \end{pmatrix}\right]
\]

\[
\times \begin{pmatrix} S_{21,1} \\ S_{21,2} \end{pmatrix} (-S_{11})^{-1/2}
\]

\[
\geq 0. \quad \text{(By Remark 2)}
\]

\[\square\]

**Lemma 4.** Under the conditions of Theorem 1 and the null hypothesis, denote \(U_{nk} = \left\{\frac{k}{n} : \frac{T}{n} \leq (1 - \frac{T}{n})\right\}\), for all \(\delta > 0\), we can find \(C = C(\delta)\), \(T = T(\delta)\) and \(N = N(\delta)\) such that

\[
P\left(\max_{k \in U_{nk}} \frac{m}{\log \log m} \frac{1}{2} \left\| \frac{\lambda}{\epsilon_k} \right\| > C\right) \leq \delta,
\]

\[
P\left(\max_{k \in U_{nk}} \frac{m}{\log \log m} \frac{1}{2} \left\| \frac{\lambda}{\epsilon_k} \right\| > C\right) \leq \delta,
\]

26
\[
P\left(\max_{\frac{k}{n} \in U_{nk}} \left(\frac{m}{\log \log m}\right)^{1/2} \parallel \tilde{\theta} - \theta_0 \parallel > C\right) \leq \delta, \quad P\left(n^{-1/2} \max_{\frac{k}{n} \in U_{nk}} \parallel \tilde{\theta} - \theta_0 \parallel > C\right) \leq \delta.
\]

**Proof.** The proof is similar to Lemma 1.2.2 of Csörgő and Horváth (1997).

**Lemma 5.** Under the conditions of Theorem 1 and \(H_0\), for all \(0 \leq \alpha < \frac{1}{2}\) we have:

\[
n^{\alpha} \max_{\frac{k}{n} \in U_{nk}} \left[\theta_k(1 - \theta_k)\right]^\alpha \parallel 2 \log \Lambda - R_k \parallel = O_p(1),
\]

\[
\max_{\frac{k}{n} \in U_{nk}} \left[\theta_k(1 - \theta_k)\right]^\alpha \parallel 2 \log \Lambda - R_k \parallel = O_p\left(n^{-\frac{1}{2}}(\log \log n)^{\frac{3}{2}}\right),
\]

where \(\Theta_{nk} = \{k : \delta_1 \leq k \leq n - \delta_2\}\)

**Proof of Theorem 1:**

**Proof.** The proof of Theorem 1 is similar to the proof of Theorem 1.3.1 (Theorem A.3.4) of Csörgő and Horváth (1997) which derives the null distribution of the trimmed test statistic.

**Proof of Theorem 2:**

**Proof.** The ELR test statistic is,

\[
-2 \log \Lambda_k = Z_{H_0,k_0} - Z_{H_1,k_0}.
\]

Under \(H_1, Z_{H_1,k_0}\) also follows an asymptotic \(\chi^2\) distribution. Therefore, \(Z_{H_1,k_0} = O_p(1)\).

We only need to prove that \(P(Z_{H_0,k_0} > cn) \rightarrow 1\) for a positive constant \(c\) under \(H_1\). For any fixed \(\varepsilon\), we can obtain

\[
\frac{1}{2n} Z_{H_0,k_0} = \sup_{\lambda} \frac{1}{n} \sum_{l=1}^{n} \log \left[1 + \theta_l^{-1} \lambda' g(x_l, \varepsilon)\right]
\]

27
By Jensen’s inequality,

\begin{align*}
E \log \left( 1 + \theta_0^{-1} \lambda_1' g_1(x_l, \varepsilon) \right) \leq \log \left( E \left( 1 + \theta_0^{-1} \lambda_1' g_1(x_l, \varepsilon) \right) \right) = 0
\end{align*}

\implies \sup_{\lambda_1} \theta_0 E \log \left( 1 + \theta_0^{-1} \lambda_1' g_1(x_l, \varepsilon) \right) = 0.

Thus,

\begin{align*}
\frac{1}{2n} \left( Z_{H_0,k_0} - \sup_{\lambda_2} (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2 g_2(x_l, \varepsilon) \right) \right] \right)
\leq (1 - \theta_0)c_0
\end{align*}

Hence, \( P(\bar{Z}_{H_0,k_0} \geq (1 - \theta_0)c_0) \to 1. \) Thus, the proof.

\section*{Proof of Theorem 3:}

\textbf{Proof.} To prove: For arbitrary small \( \frac{\theta_0}{2} > \eta, \frac{k_0 - k}{n} \geq \eta, \) \(-2 \log \Lambda_k\) cannot arrive at its maximum with probability approaching to 1.

Without loss of generality, suppose \( k < k_0 \) and \( \frac{k_0 - k}{n} \geq \eta. \) Then we have,

\[-2 \log \Lambda_{k_0} - (-2 \log \Lambda_k) = (Z_{H_0,k_0} - Z_{H_1,k_0}) - (Z_{H_0,k} - Z_{H_1,k}).\]

Since \( Z_{H_1,k_0} = O_p(1) \)

\[\frac{1}{2n} (Z_{H_0,k_0} - Z_{H_0,k} + Z_{H_1,k})\]

28
\[
\begin{align*}
&= \sup_{\lambda_1} \frac{1}{n} \sum_{l=1}^{k_0} \log \left[ 1 + \theta_{k_0}^{-1} \lambda_1' g_1(x_l, \beta_0) \right] + \sup_{\lambda_2} \frac{1}{n} \sum_{l=k_0+1}^{n} \log \left[ 1 + (1 - \theta_{k_0})^{-1} \lambda_2' g_2(x_l, \beta_0) \right] \\
&\quad - \sup_{\lambda_1} \frac{1}{n} \sum_{l=1}^{k} \log \left[ 1 + \theta_{k}^{-1} \lambda_1' g_1(x_l, \beta_0) \right] + \sup_{\lambda_2} \frac{1}{n} \sum_{l=k+1}^{n} \log \left[ 1 + (1 - \theta_{k})^{-1} \lambda_2' g_2(x_l, \beta_0) \right] \\
&\quad + \sup_{\lambda_1} \frac{1}{n} \sum_{l=1}^{k} \log \left[ 1 + \theta_{k}^{-1} \lambda_1' g_1(x_l, \beta_0) \right] + \sup_{\lambda_2} \frac{1}{n} \sum_{l=k+1}^{n} \log \left[ 1 + (1 - \theta_{k})^{-1} \lambda_2' g_2(x_l, \beta_1) \right] \\
&\geq \sup_{\lambda_2} \frac{1}{n} \sum_{l=k+1}^{n} \log \left[ 1 + (1 - \theta_{k_0})^{-1} \lambda_2' g_2(x_l, \beta_0) \right] - \sup_{\lambda_2} \frac{1}{n} \sum_{l=k+1}^{n} \log \left[ 1 + (1 - \theta_{k})^{-1} \lambda_2' g_2(x_l, \beta_0) \right] \\
&= \sup_{\lambda_2} \frac{1}{n} \sum_{l=k+1}^{n} \log \left[ 1 + (1 - \theta_{k_0})^{-1} \lambda_2' g_2(x_l, \beta_0) \right]
\end{align*}
\]

(Since, \(\frac{n-k_0}{n} \leq \rho_k \leq \frac{n-k_0}{n-k_0+n\eta}\). So, \(1 - \theta_0 \leq \lim \rho_k \leq \lim \rho_k \leq \frac{1-\theta_0}{1-\theta_0+\eta}\))

\[
\begin{align*}
&\xrightarrow{\text{a.s.}} \sup_{\lambda_1} (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \\
&\quad - \sup_{\lambda_1} \left\{ \frac{k_0-k}{n} E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \right\} \\
&\quad + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \} \\
&\geq \sup_{\lambda_1} \left\{ \eta E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \right\} \\
&\quad + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \\
&\quad + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \\
&\quad + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \\
&\quad + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \\
&\quad + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \\
&\quad + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \\
&\quad + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \\
&\quad + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \\
&\quad + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \\
&\quad + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \\
&\quad + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \\
&\quad + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \\
&\quad + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right] \\
&\quad + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda_2' g_2(x_l, \beta_0) \right) \right]
\end{align*}
\]

(By Jensen’s inequality and \(\frac{k_0-k}{n} \geq \eta\).)
Assume that \( \sup_{\lambda_1} \{ \eta E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda' g_2(x_l, \beta_0) \right) \right] + (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda' g_2(x_l, \beta_0) \right) \right] \} \) attains its maximum at \( \delta_2^* \). Then we have,

\[
\frac{1}{2n} (Z_{H_0,k_0} - Z_{H_0,k} + Z_{H_1,k}) \geq \begin{cases} 
\sup_{\lambda_1} (1 - \theta_0) E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda' g_2(x_l, \beta_0) \right) \right], & \text{if } \delta_2^* = 0, \\
-\eta E \left[ \log \left( 1 + (1 - \theta_0)^{-1} \lambda' g_2(x_l, \beta_0) \right) \right], & \text{if } \delta_2^* \neq 0.
\end{cases}
\]

Therefore, by the condition that for every fixed parameter \( \delta = \beta^* - \beta \neq 0 \), there exists a positive constant \( c_0 > 0 \) satisfy that \( \infty > \inf_\delta \sup_{\lambda_1} E \log \left[ 1 + \lambda' x (x' \delta + \epsilon) \right] \geq c_0 > 0 \) and Jensen’s inequality, there exists a constant \( c_0 > 0 \), such that \( P \left( \frac{1}{2n} (Z_{H_0,k_0} - Z_{H_0,k} + Z_{H_1,k}) > c_0 \right) \to 1 \) as \( n \to \infty \). Thus, we have, \( P \left[ \left( -2 \log \Lambda_{k_0} - (-2 \log \Lambda_k) \right) > cn \right] \to 1 \), since \( Z_{H_1,k_0} = O_p(1) \). So, \( -2 \log \Lambda_k \) cannot arrive at its maximum with probability approaching to 1. By the definition of \( \hat{k} \), we have \( |k_0 - k| \leq \eta \) with probability approaching to 1. Since \( \eta \) is arbitrary, thus the proof.

\( \square \)

References

AUE, A. AND HORVÁRTH, L. (2012). Structural breaks in time series. Journal of Time Series Analysis 34, 1-16.
BAI, J. S. & PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66**, 47-78.

BALCOMBE, K., BAILEY, A. & BROOKS, J. (2007). Threshold effects in price transmissions: the case of Brazilian wheat, maize and soya prices. *American Journal of Agricultural Economics* **89**, 308-323.

BARAGONA, R., BATTAGLIA, F. & CUCINA, D. (2013). Empirical likelihood for break detection in time series. *Electronic Journal of Statistics* **7**, 3089-3123.

BERKES, I., HORVÁTH, L., LING, S. & SCHAUER, J. (2011). Testing for structural change of AR model to threshold AR model. *Journal of Time Series Analysis* **32**(5), 547-565.

CHEN, J. & GUPTA, A. K. (2000). *Parametric Statistical Change Point Analysis*. Boston: Birkhäuser.

CHERNOFF, H. & ZACKS, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *The Annals of Mathematical Statistics* **35**, 999-1018.

CSÖRGO, M. & HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. New York: Wiley & Sons.

HAWKINS, D. M. (1977). Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association* **72**, 180-186.
KOLACZYK, E. D. (1994). Empirical likelihood for generalized linear models. *Statistica Sinica* **4**, 199-218.

LEE, S., HA, J., NA, O. AND NA, S. (2003). The Cusum test for parameter change in time series models. *Scandinavian Journal of Statistics* **30**, 781-796.

LEE, S., NA, O., NA, S. (2003). On the cusum of squares test for variance change in non-stationary and nonparametric time series models. *Annals of the Institute of Statistical Mathematics* **55**, 467-485.

LIU, Y., ZOU, C. & Zhang, R. (2008). Empirical likelihood ratio test for a change-point in linear regression model. *Communication in Statistics-Theory and Methods* **37**, 1-13.

NING, W. (2012). Empirical likelihood ratio test for a mean change point model with a linear trend followed by an abrupt change. *Journal of Applied Statistics* **39(5)**, 947-961.

NORDMAN, D. J. & LAHIRI, S. N. (2014). A review of empirical likelihood methods for time series. *Journal of Statistical Planning and Inference* **155**, 1-18.

OGATA, H. (2005). Empirical Llikelihood for non Gaussian stationary processes. *Scientiae Mathematicae Japonicae Online e-2005*, 465-473.

OWEN, A.B. (1988). Empirical likelihood ratio confidence Intervals for a single functional. *Biometrika* **75**, 237-49.
OWEN, A.B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics* **18**, 90-120.

OWEN, A.B. (1991). Empirical likelihood for linear models. *The Annals of Statistics* **19**, 1725-47.

OWEN, A.B. (2001). *Empirical Likelihood*. New York: Chapman & Hall/CRC.

PAGE, E.S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100-116.

PAGE, E.S. (1955). A test for a change in a parameter occurring at an unknown point *Biometrika* **42**, 523-527.

PERRON, P. & VOGELSANG, T. J. (1992). Testing for a unit root in a time series with a changing mean: corrections and extensions. *Journal of Business & Economic Statistics* **10**(4), 467-470.

QIN, J. & LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics* **22**, 300-325.

SEN, A. K. & SRIVASTAVA, M.S. (1975). On tests for detecting changes in mean. *The Annals of Statistics* **3**, 98-108.

VOSTRIKOVA, L.J. (1981). Detecting “disorder” in multidimensional random processes. *Soviet Mathematics Doklady* **24**, 55-59.
WORSLEY, K.J. (1986). Confidence regions and test for a change point in a sequence of exponential family random variables. *Biometrika* 73, 91-104.

ZOU, C.L., LIU, Y.K., QIN, P. AND WANG, Z. (2007) Empirical likelihood ratio test for the change-point problem. *Statistics & Probability Letters* 77, 374-382.