Asymptotic Efficient Estimation of the Change Point in Time Series Regression Models

Takayuki Shiohama
Hitotsubashi University

March 19, 2004

Abstract

This paper discusses the problem of estimating unknown change point in the trend function of a time series regression model. The error process considered here is a Gaussian stationary process with spectral density. The asymptotic properties of quasi maximum likelihood (QMLE) and quasi Bayes (QBE) estimators are studied. Consistency, limiting distributions and convergence of higher order moments of the estimators are obtained. It is also shown that the QBE is asymptotically efficient, and that the QMLE is not so general.

Keywords: Time series regression, change point,quasi maximum likelihood estimator, quasi Bayes estimator, asymptotic efficiency, Whittle likelihood.

1 Introduction

Detecting change points in the stochastic structure of a time series has become an important area of research in the last two decades. Especially in economics, the problem of testing and estimating unknown changes in econometric models is attracting growing interest, because changes in taste, technical progress, policies and regulations cause parameter instability over a period of time. The purpose of this paper is to study the parameter change problem in time series regression models with shift in the polynomial trend functions. The main concern is the asymptotics of the change point estimators, namely quasi maximum likelihood (QMLE) and quasi Bayes (QBE) estimators.

For dependent observations, a number of authors investigated the consistency property and derived the asymptotic distribution for the estimated change point. We refer the monograph of Csörgő and Horváth (1997) for informative in-depth reviews. Bai (1994) studied the least squares estimation for a mean shift in linear processes. Also Bai (1997) extends his results to the estimation of a change point in multiple regression models with dependent and heteroskedastic errors. Kokoszka and Leipus (1998) showed the consistency and the rates of convergence of CUSUM type estimators, under weak assumptions on the dependence structure. The issue of multiple structural changes, we refer the recent works of Bai and Perron (1998) and Lavielle and Moulines (2000).

There is little literature on asymptotic optimality of change point estimators. For independent and identically distributed observations, Rotiv (1990) developed an asymptotically efficient estimator using nonparametric setups. For diffusion processes, Kutoyants (1994) showed that the change point estimators MLE and BE have different limit laws, and BE is asymptotically optimal. Dachian and Kutoyants (2003) also developed similar results for cusp estimations of the ergodic diffusion process. For the theory underlying serially correlated observations, Shiohama *et. al.* (2003) and Shiohama (2003) studied asymptotically efficient estimations for time series regression setups with circular ARMA residuals and stable trend.

Our study considers the important problem of estimating change point for the case with trending regressions, which is not stated in the papers of Shiohama *et. al.* (2003) and Shiohama (2003). We also dropped the assumptions on circular ARMA residual spectra. These extensions are important for practical applications, where the presence of a polynomial time trend is very common in economic time series.

This paper is organized as follows. In Section 2 we describe time series regression models with change point and state the asymptotic representation of log-likelihood ratio processes between contiguous hypotheses. Section 3 is devoted to the asymptotics of QMLE and QBE. Section 4 contains the details of the simulation study. Finally, in

Section 5, we give the proofs for the theorems and lemmas given in Sections 2 and 3.

2 The model and settings

Let us consider the problem of change point estimation for time series regression models

$$y_t \quad = \quad \begin{cases} \boldsymbol{\beta}_1' \boldsymbol{z}_t + u_t & t = 1, \dots, [\tau n] \\ \boldsymbol{\beta}_2' \boldsymbol{z}_t + u_t, & t = [\tau n] + 1 \dots, n \end{cases} \tag{2.1}$$

where $\tau \in (0, 1)$ is an unknown change point, $\boldsymbol{z}_t$ is $q \times 1$ vector of regressors and u_t is a Gaussian stationary process with mean zero and the spectral density $f(\lambda)$. We write the spectral representation of $\{u_t\}$ as

$$u_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ_u(\lambda).$$

We consider the time-trending regressor that is

$$\boldsymbol{z}_t = (1, t, \dots, t^{q-1})'.$$

Define

$$\boldsymbol{D}_n = \operatorname{diag}\{n, \sum_1^n t^2, \dots, \sum_1^n t^{2(q-1)}\}$$

and note that as $n \to \infty$

$$\boldsymbol{D}_n \sim \operatorname{diag}\{n, n^3/3, \dots, n^{2q-1}/(2q-1)\}.$$

Let $\{Y_t\}$ be a discrete stochastic process defined later satisfying the following conditions:

Assumptions A.1. $n^{-1/2} \sum_{t=1}^{[nr]} Y_t \underset{\mathscr{L}}{\longrightarrow} VW(r)$, where $V^2 = E(Y_t^2) + 2 \sum_{k=2}^{\infty} E(Y_1 Y_k)$ is the long-run variance of Y_t, and $W(r)$ is a standard Brownian motion defined on $C[0, 1]$. The symbol $\underset{\mathscr{L}}{\longrightarrow}$ denotes weak convergence in distribution.

In this article we consider the case that the magnitude of shift tends to zero with the sample size n, but sufficiently slowly. Specifically, we suppose the following assumptions on the magnitude of shift $\boldsymbol{\delta}_n = (\delta_{1n}, \dots, \delta_{qn})'$ which is assumed to be known.

Assumptions A.2.

$$\boldsymbol{\delta}_n := \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 \to \mathbf{0}, \qquad c_n := (\boldsymbol{\delta}_n' \boldsymbol{Q}_n \boldsymbol{\delta}_n)^{-1} \to \infty,$$

where $\boldsymbol{Q}_n = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^n \boldsymbol{z}_t \boldsymbol{z}_t' \sim [n^{i+j-2}/i+j-2]_{i,j=1,\dots,q}$ and

$$n^{q-1/2}\delta_{qn} \to 0.$$

For fixed $\boldsymbol{\delta} = \boldsymbol{\delta}_n$, the asymptotic distribution of change point is studied by Shiohama *et. al.* (2003) and Shiohama (2003), where they assume the circular conditions on $\{u_t\}$, and bounded regressors. The obtained limiting distribution depends on the underlying distribution of u_t and $\boldsymbol{\delta}$ in an intricate way. Thus in practice it might be more important to investigate the asymptotic distributions of change point for small changes. Under Assumption A.2 the model (2.1) is rewritten as

$$\begin{aligned} y_t &= \begin{cases} \boldsymbol{\beta}' \boldsymbol{z}_t + u_t & t = 1, \dots, [\tau n] \\ (\boldsymbol{\beta} + \boldsymbol{\delta}_n)' \boldsymbol{z}_t + u_t, & t = [\tau n] + 1 \dots, n, \end{cases} \\ &=: \mu_t(\boldsymbol{\beta}, \tau) + u_t. \end{aligned} \tag{2.2}$$

We are interested in the estimation of $\boldsymbol{\beta}$ and τ. Motivated by the results of proposition 1 of Bai (1997), we define the local sequences

$$\tau^{(n)} = \tau + (nc_n)^{-1}\rho \quad \text{and} \quad \boldsymbol{\beta}^{(n)} = \boldsymbol{\beta} + \boldsymbol{D}_n^{1/2} \boldsymbol{b} \tag{2.3}$$

where c_n is in Assumption A.1, $\rho \in \mathbb{R}$ and $\boldsymbol{b} \in \mathbb{R}^q$. For the later convenience we put

$$S_I(j, a, A_t) = \begin{cases} 0, & |j| < 1 \\ \sum_{t=[a]+1}^{[a+j]} A_t, & j \geq 1 \\ -\sum_{t=[a+j]+1}^{a} A_t, & j \leq 1. \end{cases}$$

For model (2.2) the exact Gaussian Gaussian likelihood is given by

$$L_n(\boldsymbol{\beta}, \boldsymbol{\tau}) = \frac{1}{(2\pi)^{n/2} |\Sigma_n|^{1/2}} \exp\left[-\frac{1}{2} (\boldsymbol{y}_n - \boldsymbol{\mu}_n)' \Sigma_n^{-1} (\boldsymbol{y}_n - \boldsymbol{\mu}_n) \right], \tag{2.4}$$

where Σ_n is the $n \times n$ covariance matrix of $\boldsymbol{u}_n = (u_1, \dots, u_n)$, $\boldsymbol{y}_n = (y_1, \dots, y_n)'$ and $\boldsymbol{\mu}_n = (\mu_1(\boldsymbol{\beta}, \tau), \dots, \mu_n(\boldsymbol{\beta}, \tau))'$. Using this likelihood with the circular assumption on $\{u_t\}$, Shiohama *et.al.* (2003) and Shiohama (2003) investigated the asymptotic properties of MLE and BE. In this paper, we use the Whittle likelihood instead of (2.4), which is defined by

$$\log L_n^W(\boldsymbol{\beta}, \tau) = -\frac{n}{4\pi} \int_{-\pi}^{\pi} \left\{ \log f(\lambda) + \frac{I_n(\lambda|\boldsymbol{\beta}, \tau)}{f(\lambda)} \right\} d\lambda$$

where

$$I_n(\lambda|\boldsymbol{\beta}, \tau) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} (y_t - \mu_t(\boldsymbol{\beta}, \tau)) \exp(-it\lambda) \right|^2$$

is the periodogram. The Whittle likelihood is useful because it directly involves $f(\lambda)$, in contrast to the exact Gaussian likelihood (2.4). Also, as it was already mentioned by various authors, an explicit derivation of maximum likelihood estimator is most often practically cumbersome and complex problem. However by using the Whittle method, we can cope with these problems. In order to justify the use of the Whittle likelihood to the general results of asymptotic estimation theory of Ibragimov and Has'minski (1981), we refer the asymptotic results on Toeplitz matrices by Dzhaparidze (1986), where they approximate Σ_n^{-1} by $\{1/(4\pi^2) \int_{-\pi}^{\pi} f(\lambda)^{-1} \exp(i\lambda(r-s)) d\lambda\}_{r,s=1,\dots,n}$. The same approximation for long range dependence was derived by Dahlhaus (1989) .

The log-likelihood ratio process under (2.3) is represented as

$$\begin{aligned}
& \log Z_n(\boldsymbol{b}, \rho) \\
= \;& \log \frac{L_n^W(\boldsymbol{\beta}^{(n)}, \tau^{(n)})}{L_n^W(\boldsymbol{\beta}, \tau)} \\
= \;& -\frac{n}{4\pi} \int_{-\pi}^{\pi} \left(\frac{I_n(\lambda|\boldsymbol{\beta}^{(n)}, \tau^{(n)}) - I_n(\lambda|\boldsymbol{\beta}, \tau)}{f(\lambda)} \right) d\lambda \\
= \;& -\frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\left\{ d_n(\lambda) A(\lambda) + \overline{d_n(\lambda)}\; \overline{A(\lambda)} \right\} + |A(\lambda)|^2 \right] d\lambda \qquad (2.5)
\end{aligned}$$

where

$$d_n(\lambda) = \frac{1}{\sqrt{2\pi f(\lambda)}} \sum_{t=1}^{n} u_t \exp(it\lambda)$$

and

$$\begin{aligned} A(\lambda) &= \frac{1}{\sqrt{2\pi f(\lambda)}} \left\{ S_I(\rho c_n^{-1}, \tau n, \boldsymbol{\delta}_n' \boldsymbol{z}_s \exp(-is\lambda)) - \sum_{t=1}^{n} \boldsymbol{b}' \boldsymbol{D}_n^{1/2} \boldsymbol{z}_s \exp(-is\lambda) \right\} \\ &=: A_1(\lambda) + A_2(\lambda). \end{aligned}$$

The asymptotic representation of $Z_n(\boldsymbol{b}, \rho)$ is given in the next theorem.

Theorem 2.1 *Suppose that Assumptions A.1 and A.2 hold. Then for all* $(\boldsymbol{\beta}, \tau) \in \boldsymbol{\Theta} \times T \subset \mathbb{R}^q \times [0,1]$, *the log-likelihood ratio has the asymptotic representation*

$$\log Z_n(\boldsymbol{b}, \rho) = VW(\rho) + \boldsymbol{b}' \boldsymbol{\Delta}_n - \frac{1}{2} \left[|\rho| V^2 + \boldsymbol{b}' \boldsymbol{V} \boldsymbol{b} \right] + o_p(1)$$

where $\{W(\rho); \rho \in \mathbb{R}\}$ *is a two sided standard Wiener processes, i.e.,*

$$W(\rho) = \begin{cases} W_1(\rho) & \rho \geq 0 \\ W_2(-\rho) & \rho < 0, \end{cases}$$

with $\{W_1(t); t \in [0, \infty)\}$ *and* $\{W_2(t); t \in [0, \infty)\}$ *being independent standard Wiener processes and*

$$\Delta_n = \sum_{s=1}^{n} \boldsymbol{z}_s' Y_s$$

with

$$Y_s = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda)^{-1} e^{is\lambda} dZ_u(\lambda).$$

Here $\boldsymbol{\Delta}_n$ *is a Gaussian random vector with mean* $\mathbf{0}$ *and variance*

$$\boldsymbol{V} = \left[\frac{\sqrt{(2i-1)(2j-1)}}{2\pi f(0)(i+j-1)} \right]_{i,j=1,\ldots,q}$$

and Y_s *is a Gaussian random variable with the spectral density* $f_Y(\lambda) = (2\pi f(\lambda))^{-1}$ *and* $V^2 = f_Y(0)$.

Lemma 2.1 *Suppose that Assumptions A.1 and A.2 hold. Then for some* $p > 2$ *and for any compact sets* $\boldsymbol{\Theta}$ *and* T*, we have*

$$\sup_{\beta,\tau\in\Theta\times T} EZ_n^{1/p}(\boldsymbol{b},\rho) \le \exp\{-g(\boldsymbol{b},\rho)\}$$

where

$$g(\boldsymbol{b},\rho) = C|\rho| + \boldsymbol{b}'\boldsymbol{K}\boldsymbol{b}$$

with some positive definite matrix $\boldsymbol{K}$ *and* $C > 0$*.*

Lemma 2.2 *Suppose that Assumptions A.1 and A.2 hold. Then for any compact sets* $\boldsymbol{\Theta}$ *and* T*, there exists some integer* κ *such that for any integer* $p > 1$

$$\sup_{\beta,\tau\in\Theta\times T,|b_i^{(1)}|+|b_i^{(2)}|<H,|\rho^{(1)}|+|\rho^{(2)}|<H} \left[\sum_{i=1}^{q}\left|b_i^{(2)}-b_i^{(1)}\right|^{\kappa}+\left|\rho^{(2)}-\rho^{(1)}\right|^{\kappa}\right]^{-1}$$
$$\times \quad E\left[Z_n^{1/2p}(\boldsymbol{b}^{(2)},\rho^{(2)}) - Z_n^{1/2p}(\boldsymbol{b}^{(1)},\rho^{(1)})\right]^{2p} \le B(1+H)^{\kappa}.$$

3 Asymptotic estimation theory

This section discusses the estimation of unknown parameters $(\boldsymbol{\beta},\tau)$. For this, we introduce a class of loss functions $\boldsymbol{W} = \{w(y), y \in \mathbb{R}\}$ which satisfies the following properties:

(1) The function $w(y)$ is nonnegative on $\mathbb{R}$ with $w(0) = 0$, and continuous at $y = 0$ but is not identically 0.

(2) The function $w(y)$ is symmetric, i.e., $w(y) = w(-y)$.

(3) The sets $\{y : w(y) < c\}$ are convex for all $c > 0$.

The QMLE $(\hat{\boldsymbol{\beta}}_{QML}, \hat{\tau}_{QML})$ and QBE (for quadratic loss function) $(\tilde{\boldsymbol{\beta}}_{QB}, \tilde{\tau}_{QB})$ are defined by the usual relations

$$(\hat{\boldsymbol{\beta}}_{QML}, \hat{\tau}_{QML}) = \arg\sup_{\beta,\tau\in\Theta\times T} L_n^W(\boldsymbol{\beta},\tau) \tag{3.1}$$

and

$$(\tilde{\boldsymbol{\beta}}_{QB}, \tilde{\tau}_{QB}) = \int_{\Theta\times T} (\boldsymbol{\beta}, \tau) \frac{q(\boldsymbol{\beta}, \tau) L_n^W(\boldsymbol{\beta}, \tau)}{\int_{\Theta\times T} q(\boldsymbol{\beta}, \tau) L_n^W(\boldsymbol{\beta}, \tau) d(\boldsymbol{\beta}, \tau)} d(\boldsymbol{\beta}, \tau) \tag{3.2}$$

Let us introduce random fields

$$Z^{(1)}(\rho) = \exp\left\{ VW(\rho) - \frac{1}{2}|\rho| V^2 \right\}$$

and

$$Z^{(2)}(\boldsymbol{b}) = \exp\left\{ \boldsymbol{b}'\boldsymbol{\Delta}_n - \frac{1}{2}\boldsymbol{b}'\boldsymbol{V}\boldsymbol{b} \right\}$$

where $W(\rho), \boldsymbol{\Delta}_n, \boldsymbol{V}$ and V are defined in Theorem 2.1. Then the asymptotic representation of log-likelihood ratio process is expressed by

$$Z(\boldsymbol{b}, \rho) = \exp\left\{ Z^{(1)}(\rho) + Z^{(2)}(\boldsymbol{b}) \right\}.$$

Let $\boldsymbol{\xi} \in \mathbb{R}^q$, be a Gaussian random vector

$$\mathscr{L}\{\boldsymbol{\xi}\} = N(\mathbf{0}, \boldsymbol{V}^{-1})$$

and $\zeta \in \mathbb{R}$ be

$$\zeta = \arg \sup_{\rho\in\mathbb{R}} Z^{(1)}(\rho).$$

So, the random vector $(\boldsymbol{\xi}, \zeta)$ is defined by

$$(\boldsymbol{\xi}, \zeta) = \arg \sup_{b,\rho\in\Theta\times T} Z(\boldsymbol{b}, \rho). \tag{3.3}$$

Define the block diagonal matrix $\mathscr{A}_n$

$$\mathscr{A}_n = \begin{bmatrix} \boldsymbol{D}_n & 0 \\ 0 & nc_n \end{bmatrix}.$$

Theorem 3.1 *The QMLE $(\hat{\boldsymbol{\beta}}_{QML}, \hat{\tau}_{QML})$ is uniformly on $(\boldsymbol{\beta}, \tau) \in \Theta \times T$,*

(1) $\operatorname{plim}_{n\to\infty} (\hat{\boldsymbol{\beta}}_{QML}, \hat{\tau}_{QML}) = (\boldsymbol{\beta}, \tau)$;

(2) $\mathscr{L}\{\mathscr{A}_n(\hat{\boldsymbol{\beta}}_{QML}, \hat{\tau}_{QML})\} \underset{\mathscr{L}}{\longrightarrow} \mathscr{L}\{(\boldsymbol{\xi}, \zeta)\}$*;*

(3) for any continuous loss function $w \in \boldsymbol{W}$,

$$\lim_{n\to\infty} E_{\beta,\tau}[w\{\mathscr{A}_n(\hat{\boldsymbol{\beta}}_{QML}, \hat{\tau}_{QML})\}] = E[w\{(\boldsymbol{\xi}, \zeta)\}].$$

We next state the asymptotics of the quasi Bayesian estimator $(\tilde{\boldsymbol{\beta}}_{QB}, \tilde{\tau}_{QB})$. Recalling Lemmas 2.1 and 2.2, it is seen that Theorem 1.10.2 of Ibragimov and Has'minskii (1981) can be applied.

Theorem 3.2 *The QBE* $(\tilde{\boldsymbol{\beta}}_{QB}, \tilde{\tau}_{QB})$ *is uniformly on* $(\boldsymbol{\beta}, \tau) \in \Theta \times T$,

(1) $\underset{n\to\infty}{\text{plim}}(\tilde{\boldsymbol{\beta}}_{QB}, \tilde{\tau}_{QB}) = (\boldsymbol{\beta}, \tau)$*;*

(2) $\mathscr{L}\{\mathscr{A}_n(\tilde{\boldsymbol{\beta}}_{QB}, \tilde{\tau}_{QB})\} \underset{\mathscr{L}}{\longrightarrow} \mathscr{L}\{(\boldsymbol{\xi}, \tilde{\zeta})\}$, *where*

$$\tilde{\zeta} = \frac{\int_{-\infty}^{\infty} v Z^{(1)}(v)dv}{\int_{-\infty}^{\infty} Z^{(1)}(v)dv},$$

(3) for any continuous loss function $w \in \boldsymbol{W}$,

$$\lim_{n\to\infty} E_{\beta,\tau}[w\{\mathscr{A}_n(\tilde{\boldsymbol{\beta}}_{QB}, \tilde{\tau}_{QB})\}] = E[w\{(\boldsymbol{\xi}, \tilde{\zeta})\}].$$

Remark. According to the Theorem 1.9.1 of Ibragimov and Has'minskii (1981), for any estimator $(\boldsymbol{\beta}_n, \tau_n)$ the inequality

$$\lim_{n\to\infty} \sup_{\beta,\tau\in\Theta\times T} E_{\beta,\tau}[w\{\mathscr{A}_n(\boldsymbol{\beta}_n, \tau_n)\}] \geq E[w\{(\boldsymbol{\xi}, \tilde{\zeta})\}],$$

holds. Hence the quasi Bayesian estimator is asymptotically efficient with respect to the quadratic loss function. However, the QMLE is not so general.

Table 1. Percentage points of the asymptotic distribution of $\mathscr{L}(\zeta)$ and $\mathscr{L}(\tilde{\zeta})$.

	1.0%	2.5%	5.0%	95.0%	97.5%	99.0%	var
$\mathscr{L}(\zeta)$	-13.92	-10.20	-6.96	7.46	10.00	14.20	20.20
$\mathscr{L}(\tilde{\zeta})$	-10.04	-7.96	-5.99	6.01	7.69	11.05	13.85

4 Simulations

In this section we present the results of simulation study intended to assess the asymptotic and finite sample performance of the change point estimators described in Section 3. Percentage points of the asymptotic distribution of $\mathscr{L}(\zeta)$ and $\mathscr{L}(\tilde{\zeta})$ are obtained by simulation method. We generate 5,000 random variables $\{\zeta\}$ and $\{\tilde{\zeta}\}$. Table 1 presents the results and estimates of densities are plotted in Figure 1. We can see that the critical values of QMLE are larger in absolute values than those of QBE. The density plots in Figure 1 clearly illustrates the difference of the distributions of $\mathscr{L}(\zeta)$ and $\mathscr{L}(\tilde{\zeta})$. The density for $\mathscr{L}(\tilde{\zeta})$ have thinner tails than $\mathscr{L}(\zeta)$.

Next we examine the finite sample performance of the QMLE and QBE. We generate the following time series regression models with single change point

$$y_t = \begin{cases} \boldsymbol{\beta}_1' \boldsymbol{z}_t + u_t, & t = 1, \dots, [\tau n], \\ \boldsymbol{\beta}_2' \boldsymbol{z}_t + u_t, & t = [\tau n] + 1, \dots, n, \end{cases}$$

where u_t is AR(1) process such that $u_t = 0.8u_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim iidN(0,1)$. Put $\boldsymbol{z}_t = (1,t)'$, $\boldsymbol{\beta}_1 = (0,1)$ and $\boldsymbol{\beta}_2 = (n/2 - dn/2, d)$. Sample size n and slope d are chosen to be $n = 30, 60, 120$ and $d = 0, 0.5, 0.75$. Change point τ is set to be 0.5. In this experiment, we consider the finite sample properties of change-point estimators, hence we assume that all the parameters except change point are known for simplicity. Bias, standard deviations, and mean squared error (MSE) of $\hat{\tau}_{QML}$ and $\tilde{\tau}_{QB}$ were computed using 1,000 replications. Table 2 summarizes the simulation results for each n and d. From this table,

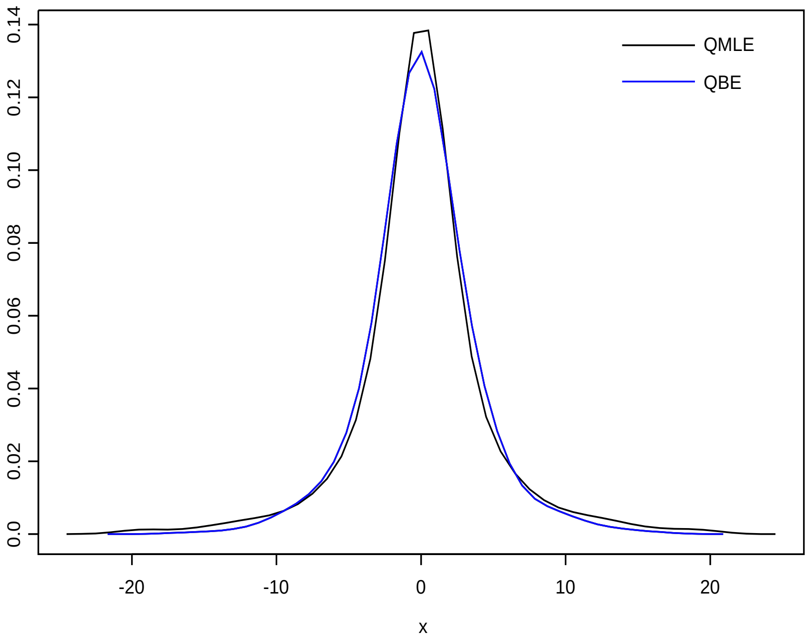

Figure 1. Density plots of $\mathscr{L}(\zeta)$ and $\mathscr{L}(\tilde{\zeta})$.

we observe the following. The MSE of QBE have better performance than that of QMLE in all experiments. This agrees with the theoretical results given in the previous section. When the sample size n becomes large, MSE and s.d. become smaller. This verifies the consistency of both estimators. For the smaller value of $|1-d|$, MSE and s.d. become large. As for the relative efficiency, the greatest value is observed when $n = 30$ and $d = 0.75$. However, we cannot conclude that the relative efficiency depends on sample size or magnitude of shift.

Table 2. Sample mean, s.d., and MSE of $\hat{\tau}_{ML}$ and $\tilde{\tau}_B$ when $\tau = 0.5$.

		MLE			BE			
n	b	Mean	s.d.	MSE	Mean	s.d.	MSE	rel. effi.
30	0	0.4843	0.1402	0.0204	0.4843	0.1109	0.0125	1.629
	0.5	0.4892	0.2181	0.0476	0.4842	0.1423	0.0204	2.326
	0.75	0.5002	0.2506	0.0627	0.4973	0.1477	0.0218	2.877
60	0	0.4948	0.0528	0.0028	0.4937	0.0416	0.0018	1.556
	0.5	0.4946	0.1551	0.0241	0.4934	0.1211	0.0147	1.636
	0.75	0.4812	0.2414	0.0586	0.4886	0.1626	0.0265	2.206
120	0	0.4954	0.0212	0.0005	0.4956	0.0157	0.0003	1.768
	0.5	0.4962	0.0546	0.0030	0.4959	0.0392	0.0015	1.929
	0.75	0.5068	0.1423	0.0203	0.5023	0.1087	0.0118	1.717

5 Proofs

In this section we give the proof of theorems and lemmas in Sections 2 and 3. Without loss of generality we assume $\rho > 0$.

Proof of Theorem 2.1 We have from (2.5),

$$\begin{aligned} & \log Z_n(\boldsymbol{b}, \rho) \\ = \; & -\frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\left\{ d_n(\lambda) A(\lambda) + \overline{d_n(\lambda) A(\lambda)} \right\} - |A(\lambda)|^2 \right] d\lambda. \\ =: \; & D_1 + D_2 + D_3. \end{aligned} \tag{5.1}$$

The first term D_1 can be evaluated as

$$\begin{aligned} D_1 &= -\frac{1}{4\pi}\int_{-\pi}^{\pi} d_n(\lambda)A(\lambda)d\lambda \\ &= -\frac{1}{8\pi^2}\int_{-\pi}^{\pi} f(\lambda)^{-1}\sum_{t=1}^{n}\sum_{s=[\tau n]+1}^{[\tau n+c_n^{-1}\rho]} \boldsymbol{\delta}_n'\boldsymbol{z}_s u_t \exp(i(t-s)\lambda)d\lambda \\ &\quad +\frac{1}{8\pi^2}\int_{-\pi}^{\pi} f(\lambda)^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} \boldsymbol{b}'\boldsymbol{D}_n^{-1/2}\boldsymbol{z}_s u_t \exp(i(t-s)\lambda)d\lambda \\ &=: D_{11}+D_{12}. \end{aligned} \tag{5.2}$$

Here we write the spectral density $f(\lambda)$ in the form

$$f(\lambda)=\frac{1}{2\pi}\sum_{l=-\infty}^{\infty} R_f(\lambda)\exp(-il\lambda)$$

where R_f's satisfy $\sum_{l=-\infty}^{\infty}|l|^p|R_f(l)|<\infty$ for any given $p\in\mathbb{Z}$. Then, from Theorem 3.8.3 of Brillinger (1981) we may write

$$f(\lambda)^{-1}=\frac{1}{2\pi}\sum_{l=-\infty}^{\infty}\Gamma(l)\exp(-il\lambda) \tag{5.3}$$

where $\Gamma(l)$'s satisfy for any given $p\in\mathbb{Z}$

$$\sum_{l=-\infty}^{\infty}|l|^p|\Gamma(l)|<\infty.$$

Then D_{11} becomes

$$D_{11}=-\frac{1}{8\pi^2}\frac{1}{2\pi}\sum_{l=-\infty}^{\infty}\Gamma(l)\sum_{t=1}^{n}\sum_{s=[\tau n]+1}^{[\tau n+c_n^{-1}\rho]}\boldsymbol{\delta}_n'\boldsymbol{z}_s u_t\int_{-\pi}^{\pi}\exp(i(t-s-l)\lambda)d\lambda.$$

It is easy to see

$$\frac{1}{2\pi}\int_{\pi}^{\pi}\exp(i(t-s-l)\lambda)d\lambda=\begin{cases}1, & \text{if } t-s-l=0\\ 0, & \text{otherwise.}\end{cases} \tag{5.4}$$

Using this, we have

$$
\begin{aligned}
D_{11} &= -\frac{1}{8\pi^2}\sum_{l=-\infty}^{\infty}\Gamma(l)\sum_{s=[\tau n]+1}^{[\tau n+c_n^{-1}\rho]}\boldsymbol{\delta}'_n\boldsymbol{z}_s u_{s+l} \\
&= -\frac{1}{8\pi^2}\sum_{l=-\infty}^{\infty}\Gamma(l)\boldsymbol{\delta}'_n\sum_{s=[\tau n]+1}^{[\tau n+c_n^{-1}\rho]}\boldsymbol{z}_s\int_{-\pi}^{\pi}e^{il\lambda}e^{is\lambda}dZ_u(\lambda) \\
&= -\frac{1}{4\pi}\boldsymbol{\delta}'_n\sum_{s=[\tau n]+1}^{[\tau n+c_n^{-1}\rho]}\boldsymbol{z}_s\int_{-\pi}^{\pi}f(\lambda)^{-1}e^{is\lambda}dZ_u(\lambda) \\
&= -\frac{1}{2}\boldsymbol{\delta}'_n\sum_{s=[\tau n]+1}^{[\tau n+c_n^{-1}\rho]}\boldsymbol{z}'_s Y_s \qquad (5.5)
\end{aligned}
$$

where $Y_s := \frac{1}{2\pi}\int_{-\pi}^{\pi}f(\lambda)^{-1}e^{is\lambda}dZ_u(\lambda)$. We have

$$
\begin{aligned}
&\lim_{n\to\infty}E\boldsymbol{\delta}'_n\sum_{t,s=[\tau n]+1}^{[\tau n+c_n^{-1}\rho]}\boldsymbol{z}_t E(Y_tY_s)\boldsymbol{z}'_s\boldsymbol{\delta}_n \\
=& \sum_{i,j=1}^{q}\delta_{in}\delta_{jn}\binom{i+j-1}{r}(\tau n)^{i+j-1-r}(c_n^{-1}\rho)^r V^2 \\
=& \rho V^2 + o(1).
\end{aligned}
$$

Hence (5.5) is equivalent to the random sequence $-(c_n^{1/2}\sum_{s=[\tau n]+1}^{[\tau n+c_n^{-1}\rho]}Y_s)/2$. This converges weakly to a Brownian motion process $VW_1(\rho)/2$, by the invariance principle of Assumption A.1 (see, e.g., Billingsley (1968), the scaling factor is $c_n^{1/2}$ instead of $n^{-1/2}$.) Next, we turn to evaluate D_{12} in (5.2). By using (5.3)

$$
D_{12} = \frac{1}{8\pi^2}\frac{1}{2\pi}\sum_{l=-\infty}^{\infty}\Gamma(l)\sum_{t=1}^{n}\sum_{s=1}^{n}\boldsymbol{b}'\boldsymbol{D}_n^{-1/2}\boldsymbol{z}_s u_t\int_{-\pi}^{\pi}\exp(i(t-s-l)\lambda)d\lambda.
$$

Using (5.4), this will be

$$\begin{aligned}
D_{12} &= \frac{1}{8\pi^2}\sum_{l=-\infty}^{\infty}\Gamma(l)\boldsymbol{b}'\sum_{s=1}^{n}\boldsymbol{D}_n^{-1/2}\boldsymbol{z}_s u_{s+l} \\
&= \frac{1}{8\pi^2}\sum_{l=-\infty}^{\infty}\Gamma(l)\boldsymbol{b}'\sum_{s=1}^{n}\int_{-\pi}^{\pi}\exp(i(s+l)\lambda)dZ_u(\lambda)\boldsymbol{D}_n^{-1/2}\boldsymbol{z}_s \\
&= \frac{1}{4\pi}\boldsymbol{b}'\sum_{s=1}^{n}\boldsymbol{D}_n^{-1}\boldsymbol{z}_s\int_{-\pi}^{\pi}\exp(is\lambda)f(\lambda)^{-1}dZ_u(\lambda) \\
&=: \frac{1}{2}\boldsymbol{b}'\boldsymbol{\Delta}_n
\end{aligned}$$

where $\boldsymbol{\Delta}_n = \sum_{s=1}^{n}\boldsymbol{D}_n^{-1/2}Y_s$ and

$$\boldsymbol{\Delta}_n \underset{\mathscr{L}}{\longrightarrow} N\left(\mathbf{0}, \boldsymbol{V}\right), \tag{5.6}$$

with $\boldsymbol{V} = [\sqrt{2i-1}\sqrt{2j-1}/2\pi f(0)(i+j-1)]_{i,j=1,\dots,q}$. Similar arguments for evaluating D_{11} and D_{12} yield

$$D_2 \simeq \frac{1}{2}VW(\rho) + \frac{1}{2}\boldsymbol{b}'\boldsymbol{\Delta}_n \tag{5.7}$$

The last term in (5.1) becomes

$$\begin{aligned}
D_3 &= -\frac{1}{4\pi}\int_{-\pi}^{\pi}|A(\lambda)|^2\,d\lambda \\
&= -\frac{1}{4\pi}\int_{-\pi}^{\pi}\left[A_1(\lambda)\overline{A_1(\lambda)} + A_1(\lambda)\overline{A_2(\lambda)} + A_2(\lambda)\overline{A_1(\lambda)} + A_2(\lambda)\overline{A_2(\lambda)}\right]d\lambda \\
&=: D_{31} + D_{32} + D_{33} + D_{34}.
\end{aligned}$$

We have

$$
\begin{aligned}
& D_{31} \\
= & -\frac{1}{8\pi^2}\int_{-\pi}^{\pi} f(\lambda)^{-1}\left(\sum_{s_1=[\tau n]+1}^{[\tau n+c_n^{-1}\rho]} \boldsymbol{\delta}_n' \boldsymbol{z}_{s_1}\exp(-is_1\lambda)\right) \\
& \times\left(\sum_{s_2=[\tau n]+1}^{[\tau n+c_n^{-1}\rho]} \boldsymbol{z}_{s_2}' \boldsymbol{\delta}_n \exp(is_2\lambda)\right) d\lambda \\
= & -\frac{1}{8\pi^2}\frac{1}{2\pi}\sum_{l=-\infty}^{\infty}\Gamma(l)\sum_{s_2=[\tau n]+1}^{[\tau n+c_n^{-1}\rho]} \boldsymbol{\delta}_n' \boldsymbol{z}_{s_2+l}\boldsymbol{z}_{s_2}' \boldsymbol{\delta}_n \\
= & -\frac{1}{2}\rho V^2,
\end{aligned} \tag{5.8}
$$

where we use Assumption A.1 to get the last equation. As for D_{34},

$$
\begin{aligned}
D_{34} = & -\frac{1}{8\pi^2}\int_{\pi}^{\pi} f(\lambda)^{-1}\left(\sum_{s_1=1}^{n} \boldsymbol{b}'\boldsymbol{D}_n^{-1/2}\boldsymbol{z}_{s_1}\exp(-is_1\lambda)\right) \\
& \times\left(\sum_{s_2=1}^{n} \boldsymbol{z}_{s_2}'\boldsymbol{D}_n^{-1/2}\boldsymbol{b}\exp(is_2\lambda)\right) d\lambda \\
= & -\frac{1}{8\pi^2}\sum_{l=-\infty}^{\infty}\Gamma(l)\boldsymbol{b}'\sum_{s_2=1}^{n}\boldsymbol{D}_n^{-1/2}\boldsymbol{z}_{s_2+l}\boldsymbol{z}_{s_2}'\boldsymbol{D}_n^{-1/2}\boldsymbol{b} \\
= & -\frac{1}{2}\boldsymbol{b}'\boldsymbol{V}\boldsymbol{b}.
\end{aligned} \tag{5.9}
$$

As for D_{32}, we have

$$
\begin{aligned}
& D_{32} \\
= & -\frac{1}{8\pi^2}\int_{-\pi}^{\pi} f(\lambda)^{-1} \\
& \times\left(\sum_{s=[\tau n]+1}^{[\tau n+c_n^{-1}\rho]} \boldsymbol{\delta}_n'\boldsymbol{z}_s e^{is\lambda}\right)\left(\sum_{t=1}^{n}\boldsymbol{z}_t'\boldsymbol{D}_n^{-1/2}\boldsymbol{b}e^{-it\lambda}\right) d\lambda \\
= & -\frac{1}{8\pi^2}\Gamma(t-s)\sum_{s=[\tau n]+1}^{\tau n+c_n^{-1}\rho}\sum_{t=1}^{n}\boldsymbol{\delta}_n'\boldsymbol{z}_s\boldsymbol{z}_t'\boldsymbol{D}_n^{-1/2}\boldsymbol{b} \\
= & O\left(n^{q-1/2}\delta_{qn}\right)
\end{aligned}
$$

which is uniformly negligible by the Assumption A.2. The asymptotic representation of D_{33} is obtained similarly that of D_{32}, which completes the proof of theorem. $\square$

The likelihood ratio $Z_n(\boldsymbol{b}, \rho)$ converges in probability

$$\log Z_n(\boldsymbol{b}, \rho) \to \log Z(\boldsymbol{b}, \rho),$$

and this convergence is uniform in $(\boldsymbol{\beta}, \tau) \in \boldsymbol{K}_\beta \times \boldsymbol{K}_T$. It is easily proved the convergence of the vectors $\{Z_n(\boldsymbol{b}_1, \rho_1), \ldots, Z_n(\boldsymbol{b}_l, \rho_l)\}$ to the vector $\{Z(\boldsymbol{b}_1, \rho_1), \ldots, Z(\boldsymbol{b}_l, \rho_l)\}$.

We omit the proofs for Lemmas 2.2 and 2.3 because these are similar to those in Shiohama *et.al.*(2003), and Shiohama (2003).

Proof for Theorem 3.1. The proof follows from Theorem 2.1 and Lemmas 2.1 and 2.2 of this paper and Theorem 1.10.1 of Ibragimov and Has'minskii (1981).

Proof for Theorem 2.1. The properties of the likelihood ratio $Z_n(\boldsymbol{b}, \boldsymbol{\rho})$ established in Theorem 2.1, and Lemmas 2.1 and 2.2 allow us to refer to Theorem 1.10.2 of Ibragimov and Has'minslii (1981).

References

[1] Bai, J. (1994). Least squares estimation of a shift in linear processes. *J. Time Ser. Anal.* **15** 453-472.

[2] Bai, J. (1997). Estimation of change point in multiple regression models. *The Review of Economics and Statistics* **79** 551-563.

[3] Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47-78.

[4] Billingsley, P. (1968). *Convergence of Probability Measures*, New York: Wiley.

[5] Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory:* Expanded ed. Holt, Rinehart and Winston, New York.

[6] Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* Wiley, New York.

[7] Dachian, S and Kutoyants, A. (2003). On cusp estimation of ergodic diffusion process. *Journal of statistical Planning and Inference* **117** 153-166.

[8] Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **13** 998-1010.

[9] Dzhaparidze, K. (1986). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series.* Springer-Verlag, New York.

[10] Ibragimov, I. A. and Has'minski, R. Z. (1981). *Statistical Estimation.* Springer, New York.

[11] Kokoszka, P. and Leipus, R. (1998). Change-point in the mean of dependent observations. *Statist. and Probab. Letters* **40** 385-393.

[12] Kutoyants, Yu. A. (1994). *Identification of Dynamical System with Small Noise.* Dordrecht: Kluwer Academic Publishers.

[13] Lavielle, M and Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in time series. *Journal of Time Series Analysis* **20** 36-60.

[14] Ritov, Y. (1990). Asymptotic efficient estimation of the change point with unknown distributions. *Ann. Statist.* **18** 1829-1839.

[15] Shiohama, T. (2003). Asymptotic estimation theory for time series regression models with multiple change points. *Scientiae Mathematicae Japoincae* **57** 299-311.

[16] Shiohama, T, Taniguchi, M. and Puri, M. L. (2003). Asymptotic estimation theory of change-point problems for time series regression models and its applications. To appear in *Probability, Statistics and their applications: Papers in Honors of Rabi Bhattacharya,* Eds: K. Athreya, M. Majumdar, M. L. Puri, E. Waymire, IMS Lecture Notes Vol. 41.