Detecting Structural Breaks with a Fusion Penalty

Jana Marečková
University of Konstanz, University of St. Gallen

Abstract

In this paper, an L_1-norm fusion penalty is added to detect structural breaks in a linear regression with time-varying parameters. The method is a more flexible extension of the method introduced in Qian and Su (2016). The idea of the fusion penalty is to start with the most general model and shrink the differences of parameters consecutive in time to zero. At points, at which the differences between the estimated parameters are non-zero, structural breaks are detected. The main advantage of the fusion penalty over the standard statistical tests is the flexibility of the model and allowing number of breaks to grow with the length of the time series. Regarding the estimation, an important issue is the optimal choice of shrinkage otherwise the number of breaks is over- or underestimated and/or the estimates are extremely biased. To select the optimal shrinkage, two criteria are taken from the literature and one new is introduced in the paper. If the method finds the correct number of breaks, it is proven that the time of the break is estimated consistently. In the simulation study, the criteria are compared to each other and to standard tests in terms of the correct number of detected breaks, the closeness to the true position of the break and the bias of the estimates. The results show that the relative position of the break can be consistently estimated. In small samples the criteria from the literature tend to underestimate the number of breaks and the parameters are strongly biased. The newly introduced criterion performs better regarding the bias and has better or comparable performance regarding the detection of the correct number of breaks in comparison to the other chosen criteria and the standard tests.

1 Introduction

In this paper, structural breaks are detected by adding a fusion penalty introduced by Land and Friedman (1997) to a linear regression model with time-varying parameters. The idea of the fusion penalty is to shrink differences of parameters consecutive in time to zero and at points at which these differences are non-zero structural breaks are detected. The optimal choice of shrinkage is an important issue. A non-optimal choice of shrinkage leads to underestimation or overestimation of the number of breaks and/or extremely biased estimates of the parameters. One possibility how to choose the optimal level of shrinkage is to apply an information criterion. Two criteria are chosen from the literature, the Extended Bayesian Information Criterion from Chen and Chen (2008) which is designed for shrinkage selection in high-dimensional models with polynomially increasing number of parameters and the Information Criterion introduced in Qian and Su (2016) for penalized linear regression model with time-varying parameters. The criterion in Chen and Chen (2008) is shown to be consistent for the regularization parameter and the criterion in Qian and Su (2016) is shown to detect the number of breaks consistently. The finite sample simulation study in this paper shows that they choose models in which the estimated parameters are strongly biased. Therefore, new criteria are introduced here which perform better regarding the bias and have a good performance regarding the detection of the correct number of breaks and their relative distance to the true positions.

The above mentioned regularization approach based on the fusion penalty is already implemented for detecting structural breaks in one-dimensional piecewise constant signals in Harchaoui and Lévy-Leduc (2010). Bleakley and Vert (2011) extend this approach to detect multiple structural breaks shared by a set of co-occurring one-dimensional piecewise constant signals. Chan et al. (2013) implement fusion penalty in structural break autoregressive (SBAR) models. Qian and Su (2016) consider a standard linear regression model and use a group lasso norm to penalize differences between coefficients to find the breaks, then they implement a post-estimation procedure to smooth out noisy breaks and at the end they estimate the coefficients of each regime by ordinary least squares (OLS). In comparison to Qian and Su (2016), the implementation in this paper will use an adaptive version of a fusion penalty to avoid the post-estimation smoothing procedure.

In the simulation study, the regularization approach is compared to widely used tests of Bai and Perron (1998) who extend the supremum test of Andrews (1993) for a single break to detect multiple breaks. Their first test tests a null hypothesis of parameter stability against a presence of a fixed number of breaks. To relax this restriction of a fixed number of breaks, Bai and Perron (1998) also introduce a double maximum test, in which the null hypothesis of no break is tested against a presence of a fixed maximum number of breaks. Moreover, Bai and Perron (1998) propose a test with a null hypothesis of ℓ breaks against $\ell + 1$ breaks and based on this single test, they introduce a sequential method of estimating the number of structural breaks. Critical values for these tests are tabulated in Bai and Perron (2003). A disadvantage of these tests is that the range of potential structural breaks is restricted to get reliable results. This restriction may cause an inability to detect an early and/or a recent break point. The results of the simulations interestingly indicate that the regularization approach seems to detect also an absolute position of a break at an extreme position even though there are no theoretical guarantees for it. Otherwise, detection of the relative position of the break in the time series improves with the length of the time series as suggested by the theoretical results.

In comparison to other approaches to detect breaks, the potential advantages of the fusion penalty may be summarized as follows. The times of the breaks do not need to be known in advance. The maximum number of true breaks which can be detected grows with a logarithm of the time series length. This is less restrictive than in Bai and Perron (1998) who assume a fixed number of breaks at asymptotically distinct time points. By using an adaptive fusion penalty, the post-model selection estimation of the parameters could be avoided. Nevertheless, an alternative parameter estimation strategy could be to use positions of the breaks found by the fusion penalty as a starting point for an post-model selection estimation of the parameters by OLS.

The paper is organized as follows. The Section 2 introduces the model, the estimation method and the information criteria for choosing the optimal amount of shrinkage. The Section 3 includes results of the simulation study which compares the regularization approach with tests from Bai and Perron (1998). The Section 4 describes the application. The Section 5 concludes.

2 Estimating Breaks by L_1-norm Regularization

2.1 Model

The following model is considered:

$$y_t = x_t'\beta_t + \varepsilon_t, \qquad t = 1, \dots, T, \tag{2.1}$$

where y_t is a centered dependent variable, x_t is a $p \times 1$ vector of standardized regressors, β_t is a $p \times 1$ vector of unknown coefficients allowed to vary across time and ε_t is an error term which is assumed to have a zero mean and at least a finite fourth moment. In a matrix notation, the model looks as follows:

$$\underset{T\times 1}{y} = \underset{T\times Tp}{X} \quad \underset{Tp\times 1}{\beta} + \underset{T\times 1}{\varepsilon}$$

$$
\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} x_1' & 0 & 0 & \cdots & 0 \\ 0 & x_2' & 0 & \cdots & 0 \\ 0 & 0 & x_3' & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & x_T' \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_T \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_T \end{pmatrix}.
$$

Let us denote the variance-covariance matrix of the error term, $V[\varepsilon]$, as Σ.

In a presence of structural breaks, two types of structural breaks can occur in model (2.1): a break in the mean function, i.e. $\beta_t \neq \beta_{t+1}$ or a break in variance-covariance structure of the error terms. Let us assume for now that Σ is known. Without imposing any restrictions on the parameters, there are Tp parameters to estimate in (2.1). In this case, the number of parameters exceeds the number of observations, particularly $Tp \geq T$, i.e. (2.1) is a high-dimensional model. To reduce the dimensionality of the model, the assumption of stable parameters will be imposed. Then, in the case of m breaks the parameters satisfy the following restrictions:

$$
\begin{aligned}
\beta_{T_{j-1}} = \cdots = \beta_{T_j - 1} &\equiv \alpha_j, && \text{for } j = 1, \ldots, m, \\
\beta_{T_{j-1}} = \cdots = \beta_{T_j} &\equiv \alpha_{m+1}, && \text{for } j = m+1.
\end{aligned}
$$

where $T_0 = 1$, $T_{m+1} = T$ and $T_1, \ldots, T_m$ denote m break points. This assumption can be captured by an L_1-norm regularization of the differences of parameters, also called a fusion penalty, which will be added to the optimization problem.

In the case of an unknown Σ, it has to be estimated. Since the breaks in the variance-covariance structure of the error terms might also be present, the variance-covariance parameters might be estimated with a fusion penalty or fused lasso introduced by Tibshirani et al. (2005) in the following way. In the first step, the (2.1) is estimated under homoscedasticity to obtain the residuals. In the case of heteroscedasticity, one can model squared residuals on a constant. The constant is allowed to vary in time and a fusion penalty is imposed to detect the changes in the variance. For the sake of completeness, the estimated model is:

$$
e_t^2 = \sigma_t^2 + u_t, \qquad t = 1, \ldots, T, \tag{2.2}
$$

where e_t^2 represents the squared residual from the estimated model (2.1), σ_t^2 is a time-varying variance to be estimated under the assumption that it is stable for some periods and u_t corrects for the model error. In this case, the method corresponds to the one in Harchaoui and Lévy-Leduc (2010).

In the case of autocorrelation, one can model an AR process with the obtained residuals and then use a fused lasso to detect the breaks and choose the appropriate lag of the AR process. If heteroscedasticity or autocorrelation is detected, the estimate of Σ can be used to transform the original model to get homoscedastic error terms and reestimate the model. In this case, the estimated model is:

$$e_t = \varphi_{t,1}e_{t-1} + \cdots + \varphi_{t,q}e_{t-q} + \nu_t, \qquad t = 1, \ldots, T, \tag{2.3}$$

where e_t represents the residual from the estimated model (2.1), $\varphi_{t,k}$'s are time-varying AR parameters to be estimated under the assumption that the AR process is stable for some periods and that we have the initial q observations before $t = 1$ at hand. ν_t is a zero mean error term.

Throughout the paper, the following notation is used. The superscript 0 denotes the true unknown parameters. Each period length is defined as $I_j^0 = T_j^0 - T_{j-1}^0$ for $j = 1, \ldots, m^0 + 1$ and the smallest period length as $I_{\min} = \min_{1 \le j \le m^0+1} |I_j^0|$. $J_{\min}$ and $J_{\max}$ denote the smallest and largest break in the parameters, i.e. $J_{\min} = \min_{1 \le j \le m^0+1} \|\alpha_{j+1}^0 - \alpha_j^0\|$ and $J_{\max} = \max_{1 \le j \le m^0+1} \|\alpha_{j+1}^0 - \alpha_j^0\|$. Further, the paper focuses on the estimation of the β parameters and leaves the case with an unknown variance as an potential extension.

2.2 L_1-norm Regularization of the Differences of Parameters

Adding a weighted L_1-norm constraint of the differences of parameters introduced by Land and Friedman (1997) to a least squares regression yields the following minimization problem:

$$\hat{\beta}_s = \arg\min_{\beta} \quad (y - X\beta)'\Sigma^{-1}(y - X\beta) \quad \text{s.t.} \quad \sum_{t=2}^{T}\sum_{k=1}^{p} w_{t,k}|\beta_{t,k} - \beta_{t-1,k}| \le s, \tag{2.4}$$

where $s \ge 0$ is a given positive constant, $\hat{\beta}_s$ is a $Tp \times 1$ vector stacking all the estimates and $w_{t,k}$ is a given positive weight allowing to penalize each difference differently. As it was already mentioned, the fusion penalty shrinks the differences of the coefficients to zero and therefore captures the stability of the parameters. The amount of shrinkage depends on the choice of s and weights $w_{t,k}$'s.

The minimization problem (2.4) can be relaxed to the following penalized least squares problem:

$$\hat{\beta}_\lambda = \arg\min_{\beta} \quad (y - X\beta)'\Sigma^{-1}(y - X\beta) + \lambda \sum_{t=2}^{T}\sum_{k=1}^{p} w_{t,k}|\beta_{t,k} - \beta_{t-1,k}|, \tag{2.5}$$

where $\lambda \ge 0$ is a given tuning parameter, corresponding to a particular s in (2.4). The role of λ and $w_{t,k}$ in (2.5) can be summarized as follows. The smaller the λ and $w_{t,k}$, the bigger the differences between the parameters in time can be and potentially more breaks can be detected. The reversed relationship holds for high λ

and high $w_{t,k}$.

The variance-covariance parameters can be estimated similarly. The parameters for heteroscedasticity:

$$\hat{\sigma}^2_\lambda = \underset{\sigma^2}{\arg\min} \quad (e^2 - I\sigma^2)'(e^2 - I\sigma^2) + \lambda \sum_{t=2}^{T} w_t |\sigma_t^2 - \sigma_{t-1}^2|, \tag{2.6}$$

where e^2 is a $T \times 1$ vector containing all squared residuals, I is a $T \times T$ identity matrix and σ^2 is a $T \times 1$ vector of variances for each point in time. The rest of the parameters has the same role as mentioned in the previous optimization problem.

For the case of autocorrelation the fused lasso penalty is added to the optimization problem:

$$\hat{\varphi}_\lambda = \underset{\varphi}{\arg\min} \quad (e-E_q\varphi)'(e-E_q\varphi)+\lambda_1 \sum_{t=1}^{T}\sum_{k=1}^{q} w^1_{t,k}|\varphi_{t,k}|+\lambda_2 \sum_{t=2}^{T}\sum_{k=1}^{q} w^2_{t,k}|\varphi_{t,k}-\varphi_{t-1,k}|, \tag{2.7}$$

where e is a $T \times 1$ vector containing all residuals, E_q is a $T \times Tq$ matrix and φ is a $Tq \times 1$ vector of all the AR parameters. The rest of the parameters has the same role as mentioned in the previous optimization problem. The only difference is that now there are two tuning parameters and two sets of weights.

Relaxing the problem (2.4) into (2.5) yields a convex function in β for a given λ and given $w_{t,k}$'s, however it is non-smooth. Regarding the algorithm used to solve a fusion penalty such as in (2.5), the Majorization-Minimization (MM) algorithm proposed by Yu et al. (2013) can be used. In each iteration of the algorithm, a majorization function, which is a differentiable approximation of (2.5), is constructed. The majorization function has to satisfy the following properties. It is always above the objective function and it goes through the optimal point found in the previous iteration. In each iteration, the majorization function represents a better and better approximation of the objective function. The algorithm stops when the change in the value of the objective function is small enough (10^{-10} or 10^{-6} is chosen in the simulation depending on the estimation step[1]) or if the maximum number of iterations is reached (10 000).

Further, the assumptions on the size of the break has to be introduced as detailed in Section 2.3. Unless the break is large enough, it cannot be detected by the proposed method. Below a certain threshold the break will be smoothed out in the second step. The number of breaks also cannot grow too fast with the sample size. The regressors has to have at least a finite fourth moment. Otherwise, the behavior of the regressors would mask the breaks.

[1] Smaller values would improve the estimates but at a high computational cost. As it will be explained later, the estimates for the first step estimation have to be more precise, therefore 10^{-10} is used there. In the second step, 10^{-6} is used to gain computational speed.

2.3 Asymptotic Properties

To show asymptotic properties of the method regarding the consistent detection of the relative positions of the breaks given that the correct number of breaks is detected, the following assumptions are made.

Assumption 1 $m_0 = O(\log(T))$.

Assumption 2 $\sup_{t\geq 1} E\|x_t\|^{4q} < \infty$, $\sup_{t\geq 1} E|\varepsilon_t|^{4q} < \infty$ *for some* $q > 1$, $E(x_t\varepsilon_t) = 0$.

Assumption 3 $\exists \infty > \underline{c}_{xx} > 0$ *and* $\infty > \bar{c}_{xx} > 0$ *and a positive sequence* δ_T *decreasing to zero as* $T \to \infty$ *such that*

$$\underline{c}_{xx} \leq \inf_{\substack{1\leq s<r\leq T+1 \\ r-s\geq \delta_T T}} \mu_{min}\left(\frac{1}{r-s}\sum_{t=s}^{r-1} E(x_t x_t')\right) \leq \sup_{\substack{1\leq s<r\leq T+1 \\ r-s\geq \delta_T T}} \mu_{max}\left(\frac{1}{r-s}\sum_{t=s}^{r-1} E(x_t x_t')\right) \leq \bar{c}_{xx}.$$

where μ_{min} *and* μ_{max} *are the smallest and largest eigenvalues.*

Assumption 4 $\sup_{\substack{1\leq s<r\leq T+1 \\ r-s\geq \delta_T T}} \left(\frac{1}{\sqrt{r-s}}\sum_{t=s}^{r-1} x_t\varepsilon_t\right) = O_p((\log T)^3)$.

Assumption 5 $T\delta_T \geq c_\nu T^{1/q}$ *for some* $c_\nu > 0$.

Assumption 6 $I_{min}/(T\delta_T) \to \infty$ *as* $T \to \infty$.

Assumption 7 $J_{max} = O(1)$, $T\delta_T J_{min}^2/(\log T)^6 \to \infty$ *as* $T \to \infty$.

Assumption 8 $\lambda = \lambda_T$ *satisfies* $\frac{\lambda\sqrt{p}\xi_T}{T\delta_T J_{min}} \to 0$ *as* $T \to \infty$ *where* ξ_T *bounds the largest* $w_{t,k}$.

Assumption 1 captures the condition on m_0, the true number of breaks. The true number of breaks can grow with the length of the time series at a logarithmic rate. This is less restrictive than the assumption in Bai and Perron (1998) who are considering a setup with fixed number of breaks at asymptotically distinct time points. Assumptions 2 - 4 put restrictions on the true DGP. The Assumption 2 requires at least finite fourth moments of x_t and u_t. Assumptions 2 and 4 allow to use concentration inequalities. The Assumption 4 is not too restrictive as it is satisfied by several standard processes such as e.g. autoregressive moving average (ARMA) processes. Assumption 3 bounds the eigenvalues of $E(x_t x_t')$ and ensures the invertibility of the $X'X$ matrix with probability going to 1 for any closed interval $[s, r-1]$ of the minimum length $\delta_T T$ which grows to infinity as in Assumption 5.

The sequence δ_T plays a role for the consistency of the relative break point positions T_j/T. Assumption 6 bounds the sequence from below, i.e. determines the slowest rate at which it converges to zero. Assumptions 7 and 8 bound the sequence from above, i.e. determine the fastest convergence to zero.

The following theorem shows consistency of $\{\hat{T}_j/T\}$ when $\hat{m} = m_0$.

Theorem 2.1 *Suppose that Assumptions A1-A8 hold. If $\hat{m} = m_0$, then*

$$P\left(\max_{1\leq j\leq m_0} |\hat{T}_j - T_j^0|/T \leq \delta_T\right) \rightarrow 1 \text{ as } T \rightarrow \infty.$$

The proof of the Theorem 2.1 is an augmented version of the proofs in Harchaoui and Lévy-Leduc (2010) and Qian and Su (2016) extended to the linear regression model estimated by the weighted fusion penalty. The proof is in Appendix B.1.

2.4 Estimation Procedure and Selection of the Tuning Parameter

The estimation procedure of problems such as (2.5) has to be done in two steps. In the first step, $w_{t,k}$'s need to be estimated. In the second step, the parameters in (2.5) are estimated while plugging $\hat{w}_{t,k}$'s from the first step for $w_{t,k}$'s. To ensure that the estimates in the second step are reliable and consistent, it is important to estimate $w_{t,k}$'s properly.

To get estimates for the weights $w_{t,k}$ in a high-dimensional model, β is first estimated from problem (2.5) in which all $w_{t,k} = 1$ and parameter λ is set optimally (let us denote this value as $\hat{\lambda}_1$). Then, the weights for the second step are defined as:

$$\hat{w}_{t,k} = \begin{cases} 1/|\hat{\beta}_{t,k,\hat{\lambda}_1} - \hat{\beta}_{t-1,k,\hat{\lambda}_1}| & \text{for } |\hat{\beta}_{t,k,\hat{\lambda}_1} - \hat{\beta}_{t-1,k,\hat{\lambda}_1}| \neq 0 \\ 1/\gamma & \text{for } |\hat{\beta}_{t,k,\hat{\lambda}_1} - \hat{\beta}_{t-1,k,\hat{\lambda}_1}| = 0 \end{cases}$$

where γ is set adequately high and still satisfying the Assumption 8. Then, the problem (2.5) is estimated again with $\hat{w}_{t,k}$ and new optimal λ.

The key question for estimating β from equations (2.5) is how to choose λ optimally. Assumption 8 provides an optimal rate at which it should converge to 0. In practice, information criteria are preferred to a cross-validation estimation of the tuning parameters because of lower computational costs. Information criterion always takes into an account the trade-off between fit and number of parameters. By choosing λ such that the information criterion is minimized, the best combination between the fit and the number of parameters is found. In the simulation study the following three criteria are used to choose λ from a grid of values evenly distributed between $[0.01, 0.5T]$ and their performance in finite samples is assessed. Alternatively, one could also take a grid of values which are logarithmically distributed as the fusion penalty estimates are more sensitive to smaller values of λ.

2.4.1 Extended BIC (EBIC)

The first information criterion is taken from Chen and Chen (2008). They study a problem of selecting a tuning parameter in penalized likelihood functions for high-dimensional models. In their setting, they allow parameters to grow polynomially with sample size, which is also the case in the model introduced in this paper since $Tp = O(T)$. They show that under normality of the error term and under an asymptotic identifiability condition, their Extended Bayesian Information Criterion ($EBIC$) chooses the true model consistently. The $EBIC$ takes the following form:

$$EBIC(\lambda) = T \log \left(\frac{1}{T} \sum_{t=1}^{T} (y_t - x_t' \hat{\beta}_{t,\lambda})^2 \right) + (\log(T) + 2\log(Tp)) \, |\hat{\beta}_\lambda|,$$

where $|\hat{\beta}_\lambda|$ represents the number of nonzero unique parameters in the model, i.e.

$$|\hat{\beta}_\lambda| = \sum_{k=1}^{p} 1(\hat{\beta}_{1,k,\lambda} \neq 0) + \sum_{t=2}^{T} \sum_{k=1}^{p} \mathbb{1}(\hat{\beta}_{t-1,k,\lambda} \neq \hat{\beta}_{t,k,\lambda} | \hat{\beta}_{t,k,\lambda} \neq 0),$$

where $\mathbb{1}(\cdot)$ is an indicator function taking a value 1, if the condition in the brackets is satisfied. $EBIC$ criterion will be used in both estimation steps to choose the optimal λ.

2.4.2 IC from Qian and Su (2016)

The second information criterion is taken from Qian and Su (2016). They detect multiple structural breaks with a group lasso norm in a linear model and choose the tuning parameter based on the following criterion:

$$IC_{QS}(\lambda) = \log \left(\frac{1}{T} \sum_{t=1}^{T} (y_t - x_t' \hat{\beta}_{t,\lambda})^2 \right) + \rho_T |\hat{\beta}_\lambda|,$$

where they choose $\rho_T = 1/\sqrt{T}$. They show that under mild regularity conditions, the IC_{QS} is detecting the correct number of breaks. IC_{QS} criterion will be used in both estimation steps to choose λ.

2.4.3 Augmented IC from Qian and Su (2016)

The third way of choosing the tuning parameter is based on the observation that the two information criteria above are prone to underfit the models in small samples if used in the two step procedure described above, i.e. they detect often less breaks then there are in the true model.[2] The reason is that on the one hand, the parameter λ chosen as optimal by $EBIC$ or IC_{QS} in the first estimation step is closer to the truth in terms of the number of breaks (as shown e.g. in Qian and Su (2016)), however on the other hand for a cost of a high bias in the estimated parameters. This bias negatively influences the weights for the second estimation step. Therefore, an adjustment parameter is introduced into the criterion in the first estimation step. The role of the adjustment parameter is to choose λ such that less biased estimates of β are obtained, so that the second estimation step will be improved by more appropriate weights.

The first augmented IC_{QS} is then given by:

$$IC_1(\lambda) = \log\left(\frac{1}{T}\sum_{t=1}^{T}(y_t - x_t'\hat{\beta}_{t,\lambda})^2\right) + \frac{1}{\sqrt{T}}(|\hat{\beta}_\lambda| - c),$$

where

$$c = \sum_{t=2}^{T}\sum_{k=1}^{p} 1\{|\hat{\beta}_{t,k,\lambda} - \hat{\beta}_{ad,k,\lambda}| < \delta(\max_{1\leq i\leq T}\hat{\beta}_{i,k,\lambda} - \min_{1\leq i\leq T}\hat{\beta}_{i,k,\lambda}) \mid \hat{\beta}_{t,k,\lambda} \neq 0, \hat{\beta}_{t,k,\lambda} \neq \hat{\beta}_{t-1,k,\lambda}\},$$

and ad represents the time at which parameter k added 1 to c the last time (for $t = 2$, $ad = 1$) and $\delta \in (0, 1)$ represents a percentage of a difference between the largest and the smallest estimated parameter β_k which determines a negligible break. In other words, c contains how many times the change in the parameter was not big enough to consider it as a break which is then deducted from the number of nonzero unique parameters in the model. By the inequality in the indicator function, c increases if the difference in the parameter between the current value and the value after the last break (contained in $\hat{\beta}_{ad,k,\lambda}$) is smaller than $100 \cdot \delta\%$ of a difference between the largest and the smallest value of the estimated parameter β_k over the whole period given that the parameter did not switch to zero and given that it is different from the parameter in the previous period. Introducing $\hat{\beta}_{ad,k,\lambda}$ is important to cover a case, in which the parameter changes slowly but monotonically over time. Then, these permanent small changes will be captured at least by a step function. If $|\hat{\beta}_{t,k,\lambda} - \hat{\beta}_{t-1,k,\lambda}|$ was used in c instead, then all the small breaks would be discarded without taking into account that their sum is already significant. The condition $\hat{\beta}_{t,k,\lambda} \neq \hat{\beta}_{t-1,k,\lambda}$ is important for not taking into account the same small break twice. The idea of capturing the parameter changes in a step function also affects the weights $w_{k,t}$. All the negligible breaks are smoothed out and only the (accumulated) significant ones are translated into the weights based on $IC_1(\lambda)$.

In the second step, the final model is selected according to a criterion similar to

[2] See the results in the next section devoted to the simulation study.

IC_{QS} only ρ_T is chosen differently but still within the consistency condition:

$$IC_2(\lambda) = \log\left(\frac{1}{T}\sum_{t=1}^{T}(y_t - x_t'\hat{\beta}_{t,\lambda})^2\right) + \frac{1}{\sqrt[3]{T^2}}|\hat{\beta}_\lambda|.$$

The adaptive fusing penalty should be able to suppress the insignificant breaks and leave only the most important ones. Therefore, a break is simply defined at a point t at which $\hat{\beta}_t \neq \hat{\beta}_{t-1}$ in the second estimation step.

3 Simulation Study

There are three different sample sizes in the simulation study: $T = \{50, 100, 200\}$. For each T, one thousand different samples were drawn ($S = 1000$). Optimal tuning parameters λ_1 and λ_2 are searched on a grid of 100 linearly spaced values from the interval $[0.01, T]$. Smoothing parameter δ in the IC_1 was set to three different values (0.025, 0.05 and 0.075) to see its effect on the estimation. The best results were obtained for $\delta = 0.075$ in models with one parameter and one break and for $\delta = 0.025$ in models with two parameters or no break, however the results over the chosen δ do not vary much.

The fusion penalty is compared to two standard tests for detecting breaks in parameters: Double Maximum test (DM) and/or the Sequential test (Seq) from Bai and Perron (1998) with the trimming parameter set to 0.15 at the 5% significance level. DM test tests a null hypothesis of no breaks against an alternative of unknown number of breaks given an upper bound of breaks set to 5. Sequential test tests stepwise a null hypothesis of ℓ breaks against an alternative of $\ell + 1$ breaks, starting with $\ell = 0$ until the null hypothesis is rejected.

All simulations are evaluated using the following criteria, whose formulas are given in Table 1. Firstly, three cases are distinguished based on the number of the detected breaks:

(1) Too few breaks were detected = "False Negatives" (FN),

(2) Correct number of breaks is detected = "True Positives" (TP),

(3) Too many breaks were detected = "False Positives" (FP).

Regarding these values, TP should be maximized and $FN + FP$ minimized.

Under a presence of break, the closeness of the estimated break to the true one plays a role. The criterion Q measures the average relative distance to the closest true break in the following fashion. In the case of no detected breaks, the estimate is fully penalized (i.e. attains value 1). In other cases, the average relative distance of the estimated breaks to the closest true break is computed. The true breaks which have no estimated counterpart (i.e. some other true breaks are closer to the estimated breaks) are fully penalized. Notice that elements in Q are constructed in such a way that the final result will be in a $[0,1]$ interval, where 0 indicates a perfect estimate in all draws and 1 indicates that no breaks were detected in all draws. Methods with Q closer to 0 are evaluated as more precise.

In the case of $p > 1$, it is also important if the break was found in the correct parameters. This is captured in criterion P. The idea is to add 1 if the estimated break is in exactly the same parameters as in the closest true break. If the break is found in too many or too few parameters, only a corresponding fraction of correctly estimated parameters is added. As in the case of Q, P is also constructed such that the final result will be in a $[0,1]$ interval. Values closer to 1 indicate better performance in detecting a break in the correct parameter.

Table 1: Evaluation Criteria

Criterion	Formula
FN	$\sum_{s=1}^{S} \mathbf{1}(\hat{m}_s < m_0)/S$
TP	$\sum_{s=1}^{S} \mathbf{1}(\hat{m}_s = m_0)/S$
FP	$\sum_{s=1}^{S} \mathbf{1}(\hat{m}_s > m_0)/S$
Q	$\frac{1}{S}\sum_{s=1}^{S}\{\mathbf{1}(\hat{m}_s = 0) + \mathbf{1}(\hat{m}_s \neq 0)\frac{1}{m_0}\sum_{j=1}^{m_0}\left[\frac{1}{\lvert K_j\rvert}\sum_{k\in K_j}\frac{\lvert\hat{T}_k^s - T_j^0\rvert}{T}\Big\rvert_{K_j\neq\emptyset} + 1\,\rvert_{K_j=\emptyset}\right]\}$
P	$\frac{1}{S}\sum_{s=1}^{S}\frac{1}{m_0}\sum_{j=1}^{m_0}\left[\frac{1}{\lvert K_j\rvert}\sum_{k\in K_j}\frac{\hat{q}_k^{s,c}}{\max(\hat{q}_k^s, q_j^0)}\Big\rvert_{K_j\neq\emptyset}\right]$
B	$\frac{1}{S}\sum_{s=1}^{S}\frac{1}{Tp}\sum_{t=1}^{T}\sum_{k=1}^{p}(\hat{\beta}_{t,k,\lambda}^s - \beta_{t,k}^0)^2$

Notation: $s\ldots$ s-th draw, $S\ldots$ total number of draws, $\hat{m}_s\ldots$ number of breaks estimated in draw s, $m_0\ldots$ true number of breaks, $T_j^0\ldots$ time point of the j-th true break, K_j is a set containing subindeces of all time points of the estimated breaks for which T_j^0 is the nearest time point, $\hat{T}_k^s\ldots$ time point of the k-th break estimated in the s-th draw, $\hat{q}_k^{s,c}\ldots$ number of parameters at which the k-th break was correctly detected (i.e. in comparison to the true parameter breaks at T_j^0) in the s-th draw, $\hat{q}_k^s\ldots$ number of all parameters at which the k-th break was detected in the s-th draw, $q_j^0\ldots$ number of true parameter breaks at T_j^0.

3.1 One Parameter - One Break - Middle

The data are simulated from the following model:

$$\begin{aligned} y_t &= 1 + 2x_t + \varepsilon_t, \qquad && t = 1,\ldots,T/2, \\ y_t &= 1 + 4x_t + \varepsilon_t, \qquad && t = T/2+1,\ldots,T, \end{aligned} \tag{3.1}$$

where x_t is an iid random variable from the standard normal distribution and ε_t is an iid Gaussian error term with zero mean and the standard deviation is either

$\sigma_\varepsilon = \sqrt{5}$, i.e. the signal-to-noise ratio (SNR) is equal to 2 or $\sigma_\varepsilon = \sqrt{10}$, i.e. SNR = 1. In other words, a relatively large break in the middle of the series is present.

Tables 42 - 44 and Tables 46 - 48 (in the Appendix C)[3] contain rates of detected breaks in the first step (using $EBIC$, IC_{QS} and IC_1) and in both steps (based on several combinations of $EBIC$, IC_{QS}, IC_1 and IC_2) for each pair of δ and SNR. The regularization techniques are not as strong in detecting exactly one break as the standard Sequential test. However, if we do not condition on finding exactly one break, Figures 4 - 9[4] show that the regularized estimates based on IC_1 and IC_2 are detecting the position of the break better with a slight overfitting as Tables 44 and 48 confirm. If we condition on finding exactly one break, Figures 10 - 15 confirm the results of Tables 44 and 48 that the standard test performs better. The exception is $T = 50$ and $T = 100$ for the lower SNR where the fusion penalty performs better than the Sequential test also if we condition on exactly one found break.

Regarding the bias of the estimates, an average squared bias is computed as

$$\frac{1}{S}\sum_{s=1}^{S}\frac{1}{Tp}\sum_{t=1}^{T}\sum_{k=1}^{p}(\hat{\beta}^s_{t,k,\lambda} - \beta^s_{t,k,\lambda})^2,$$

where S denotes number of simulations and s a particular simulation. Tables 34 - 36 and Tables 38 - 40 reveal that the estimates in the first step based on the $EBIC$ or IC_{QS} are highly biased and therefore inappropriate as weights for the second step of the estimation in comparison to IC_1. The severe underfitting by using $EBIC$ or IC_{QS} in both steps indicates that $EBIC$ and IC_{QS} have a tendency to attain the minimum at a high λ and then lead to rather flat estimates. Tables 34 - 36 and Tables 38 - 40 show that even if IC_1 is used in the first step, using $EBIC$ or IC_{QS} in the second step still leads to many cases of underfitting.

The combination of IC_1 and IC_2 does not suffer from underfitting (see Tables 42 - 44 and Tables 46 - 48) but it has a tendency to overfit in larger samples. However, in the case of $T = 200$, the majority of the detected breaks is equal to 1 or 2, i.e. it is still very close to the true value. Moreover, the position of the break is also estimated close to the true value as mentioned above. It can be therefore concluded that IC_1 and IC_2 perform well, especially in smaller samples.

Based on the results in Tables 42 - 44 and Tables 46 - 48, a higher ratio of noise influences the standard test negatively only in smaller samples ($T = 50$ and $T = 100$). The influence of the higher noise on the two step regularization techniques affects the distribution of the breaks towards underfitting, i.e. in some cases the break is masked by the noise.

[3]All the Tables and Figures are in the Appendix C.

[4]Figures are reported only for $\delta = 0.075$.

3.2 One Parameter - One Break - End

A model with one relatively large break in one parameter at the end of the sample is introduced in this part. The data are simulated from the following model:

$$\begin{aligned} y_t &= 1 + 2x_t + \varepsilon_t, \qquad && t = 1, \ldots, T-11, \\ y_t &= 1 + 4x_t + \varepsilon_t, \qquad && t = T-10, \ldots, T, \end{aligned} \tag{3.2}$$

where x_t is an iid random variable from the standard normal distribution and ε_t is an iid Gaussian error term with zero mean and the standard deviation of the error term is once set so that SNR = 2: if $T = 50$: $\sigma_\varepsilon = \sqrt{3.2}$, if $T = 100$: $\sigma_\varepsilon = \sqrt{2.6}$, if $T = 200$: $\sigma_\varepsilon = \sqrt{2.3}$ and once it is set so that SNR = 1: if $T = 50$: $\sigma_\varepsilon = \sqrt{6.4}$, if $T = 100$: $\sigma_\varepsilon = \sqrt{5.2}$, if $T = 200$: $\sigma_\varepsilon = \sqrt{4.6}$.

Tables 58 - 60 and Tables 62 - 64 contain rates of detected breaks for each combination of δ, information criteria and SNR. For $T = 50$, the standard Sequential test performs in most of the cases better than the regularized estimators in finding exactly one break. For $T = 100$ and $T = 200$, regularized estimator based on IC_1 and IC_2 performs better than the standard test (especially for higher δ). If we do not condition on finding exactly one break, Figures 16 - 21[5] show that the regularized estimates based on IC_1 and IC_2 are detecting the position of the break better. Especially, when the break is so extreme that it is out of the bounds in which the standard test allows for a break, i.e. for $T = 100$ and $T = 200$. If we condition on finding exactly one break, Figures 22 - 27 confirm the results of Tables 58 - 60 and Tables 62 - 64 that the standard test performs better for $T = 50$ and that the standard test suffers from the bounding effect for $T = 100$ and $T = 200$.

Regarding the bias of the estimates, Tables 50 - 52 and Tables 54 - 56 show that the estimates in the first step based on the $EBIC$ or IC_{QS} are more biased then the ones based on the IC_1. The differences are not so big as before, since the break is set to $T - 10$, therefore smoothing out the estimates close to 2 do not add that much to the bias, however there is an improvement if IC_1 is used.

A higher ratio of noise influences the standard tests negatively only in smaller samples ($T = 50$) based on the results in Tables 58 - 60 and Tables 62 - 64. The influence of the higher noise on the two step regularization techniques affects the distribution of the breaks towards underfitting as in the previous subsection. This simulation exercise is about finding an absolute position of a break, i.e. the relative position changes with higher T. Even though, there is no theoretical result for consistency, it is interesting that the fusing penalty seems to perform well in such an extreme case.

[5]Figures are reported only for $\delta = 0.075$.

3.3 One Parameter - No Break

The data are simulated from the following model:

$$y_t = 1 + 2x_t + \varepsilon_t, \qquad t = 1, \dots, T, \tag{3.3}$$

where x_t is an iid random variable from the standard normal distribution and ε_t is an iid Gaussian error term with zero and the standard deviation of the error term is either $\sigma_\varepsilon = \sqrt{2}$, i.e. the signal-to-noise ratio (SNR) is equal to 2 or $\sigma_\varepsilon = \sqrt{8}$, i.e. SNR $= 0.5$.

Tables 10 - 12 and 14 - 16 contain percentage rates of falsely detected breaks for each combination of δ, information criteria and SNR. As we can see, two step weighted fusion penalty based on applying in each step either $EBIC$ or IC_{QS} is performing the best. As it is already mentioned in the previous part, this result is mainly due to a general strong bias towards flat estimates of β based on $EBIC$ and IC_{QS}. As a result, the weights in the second step are big and push the weighted estimates to be even flatter (i.e. no break is estimated). The two step procedure based on IC_1 and IC_2 is supposed to solve the strong smoothing of the coefficient estimates. As a result in a case of no break, procedure based on IC_1 and IC_2 has higher false detection but still in a reasonable range. Tables 2 - 8 contain mean values of squared bias for the estimates of the model (3.3). In these tables, there are no big differences as in Tables 34 - 40 because all the procedures shrink towards the flat estimates correctly, therefore the bias is similar for all of them. Regarding the smoothing parameter δ, the best accuracy was achieved with $\delta = 0.025$ and the bias does not vary much over the δ.

The accuracy of the number of detected breaks is higher in the setup with a higher SNR for all the procedures. However, the regularization methods seems to be more negatively influenced by a higher noise in comparison to the standard tests. Overall, the regularization techniques perform better than the two standard tests in a case of no break.

3.4 Two Parameters - One Break - Middle

The data are simulated from the following model:

$$\begin{aligned} y_t &= 1 + 2x_{1t} + x_{2t} + \varepsilon_t, \qquad && t = 1, \dots, T/2, \\ y_t &= 1 + x_{1t} + x_{2t} + \varepsilon_t, && t = T/2 + 1, \dots, T, \end{aligned} \tag{3.4}$$

where x_t is an iid random variable from the standard normal distribution and ε_t is an iid Gaussian error term with zero mean and the standard deviation is either $\sigma_\varepsilon = \sqrt{6.4}$, i.e. SNR $= 2$ or $\sigma_\varepsilon = \sqrt{12.8}$, i.e. SNR $= 1$.

Tables 74 - 76 and Tables 79 - 81 contain rates of detected breaks for each combination of δ, information criteria and SNR. The regularization techniques are not as strong in detecting exactly one break as the standard Sequential test. The regularization techniques suffer from overfit again in terms of the number of detected breaks. Figures 28 - 33[6] show that the regularized estimates based on IC_1 and IC_2 are detecting the position of the break better even in the presence of a slight overfitting as Tables 74 - 76 and Tables 79 - 81 show. Regarding the detection of a break, the regularization method detects the break often corrcectly in the first parameter. Frequently it detects incorrectly a break in both parameters. However, it correctly almost never detects a break only in the second parameter as Figures 28 - 33 show.

Regarding the bias of the estimates, Tables 66 - 68 and Tables 70 - 72 reveal that the estimates in the first step based on the $EBIC$ or IC_{QS} are highly biased and therefore inappropriate as weights for the second step. The IC_1 improves the performance.

4 Application

The method is applied to detect breaks in the labor productivity in the US from 1955 to 2004. It is assumed that the US production follows a CES function:

$$Y_t = a(\varepsilon_t)\left[a_K(t)K_t^{\left(\frac{\sigma-1}{\sigma}\right)} + a_L(t)L_t^{\left(\frac{\sigma-1}{\sigma}\right)}\right]^{\left(\frac{\sigma}{\sigma-1}\right)},$$

where
Y_t = production in t,
K_t = capital input in t,
L_t = labor input in t,
$a_K(t)$ = capital augmenting technical progress,
$a_L(t)$ = labor augmenting technical progress,
$a(\varepsilon_t)$ = productivity shock,
σ = elasticity of substitution, $\sigma \geq 0$.

Using the relationship for the marginal productivity of labor, the logarithm of labor productivity can be under the following assumptions

$$a(\varepsilon_t) = [\exp(\beta_0 + \varepsilon_t)]^{\frac{1}{1-\sigma}},$$
$$a_L(t) = [\exp(\alpha_L t)]^{\frac{\sigma-1}{\sigma}}$$

expressed as follows:

[6] Figures are reported only for $\delta = 0.025$.

$$\ln\left(\frac{Y_t}{L_t}\right) = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot \ln\left(\frac{w_t}{p_t}\right),$$

where w_t/p_t is a real wage in t and the coefficients in the linear model are related to the original parameters as:

$$\beta_1 = (1 - \sigma) * \alpha_L,$$
$$\beta_2 = \sigma.$$

The Figures 1-3 show the estimated results. Based on β_2, we can see that the elasticity of substitution is stable over the whole period and is equal to 0.93. The estimate satisfies the non-negativity constraint and suggests that capital and labor are almost perfect substitutes.

Since there is no break in β_2, coefficient β_1 captures directly the changes in labor productivity. On Figure 1, there are 6 breaks at the following times: 1972-3Q, 1975-1Q, 1979-2Q, 1984-3Q, 1988-2Q and 1992-1Q. The first break is close to a 1973 Oil crisis which negatively influenced the whole economy, as well as labor productivity. The price oil peaked in 1979 in the second Oil crises. The third break representing 1979-2Q shows a rather big drop in β_2 which can be directly translated into a big drop in labor productivity caused by the situation on the oil market. After the mid-80s, the oil market became more stable and we can see that the productivity labor starts getting back to its pre-Oil crisis values. The pre-crisis state was reached in 1992-1Q.

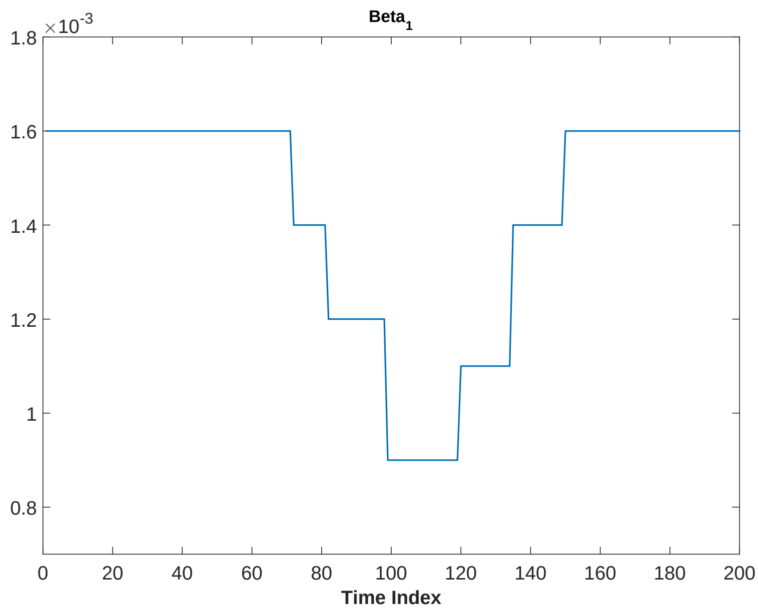

Figure 1: First Slope

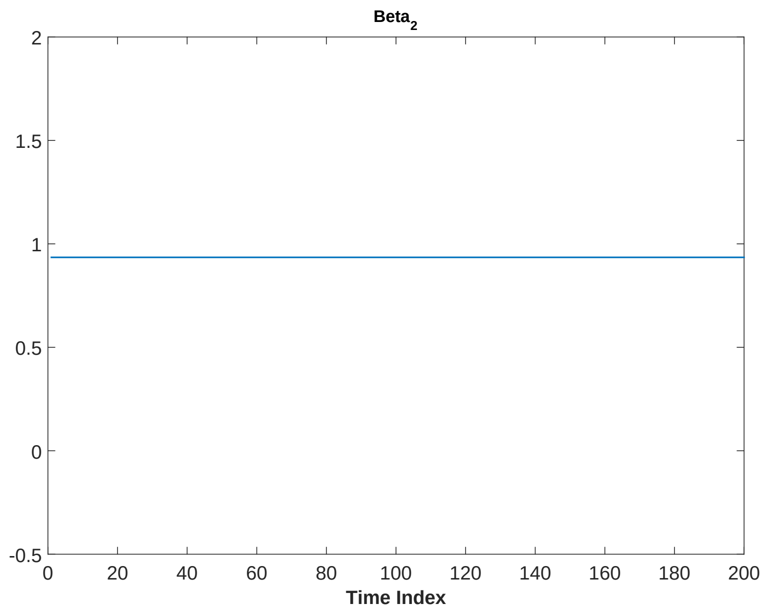

Figure 2: Second Slope

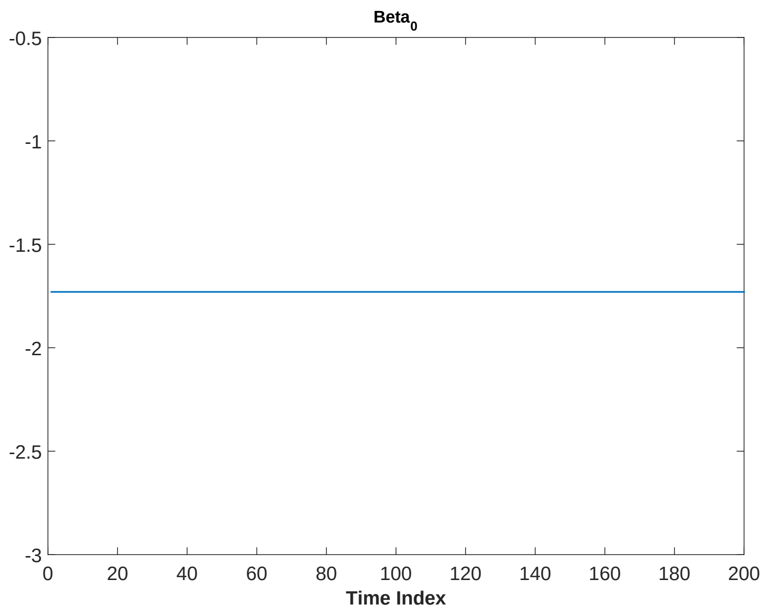

Figure 3: Intercept Term

5 Conclusion

In this paper, the weighted fusion penalty is applied to detect structural breaks in a linear regression model in a simple simulation study. The fusion penalty approach has a couple of advantages in comparison to standard tests which are widely used. It is not necessary to know in advance when structural breaks occurred. Moreover, the assumption of fixed number of breaks is not needed. Another advantage is that by the weighted fusion penalty, smoothing of falsely detected breaks is controlled within the method. No additional smoothing techniques are needed as in Qian and Su (2016).

The simulation results for samples with 50, 100 and 200 observations show that the weighted fusion penalty yields reasonable results, given the high-dimensional characteristic of the model. Information criteria IC_1 and IC_2 seem to perform well in small finite samples with low number of parameters. They reduce bias and serious underfitting of $EBIC$ and IC_{SQ} in the two step weighted estimation procedure. Comparing IC_1 and IC_2 to the standard tests in terms of estimating the exact correct number of breaks, the regularization approach is performing slightly worse. However, they detect the position of the break better. Lower SNR negatively influences the performance of all the techniques.

The method was applied for detecting structural breaks in the labor productivity in the US from 1955 to 2004. The breaks which were found are nicely interpretable. The drops in labor productivity were caused by 2 oil crises in 1973 and 1979. After stabilization of the oil market, the labor productivity got back to its pre-crisis value.

References

ANDREWS, D. W. K. (1993): "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61, pp. 821–856.

BAI, J. AND P. PERRON (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, pp. 47–78.

——— (2003): "Critical Values for Multiple Structural Change Tests," *Econometrics Journal*, 6, 72–78.

BLEAKLEY, K. AND J.-P. VERT (2011): "The group fused Lasso for multiple change-point detection," *ArXiv e-prints*.

CHAN, N. H., C. Y. YAU, AND R.-M. ZHANG (2013): "Group LASSO for Structural Break Time Series," *Journal of the American Statistical Association*, In Press, Accepted Manuscript, –.

CHEN, J. AND Z. CHEN (2008): "Extended Bayesian information criteria for model selection with large model spaces," *Biometrika*, 95, 759–771.

HARCHAOUI, Z. AND C. LÉVY-LEDUC (2010): "Multiple Change-Point Estimation With a Total Variation Penalty," *Journal of the American Statistical Association*, 105, 1480–1493.

LAND, S. R. AND J. H. FRIEDMAN (1997): "Variable fusion: A new adaptive signal regression method," Tech. rep., Technical Report 656, Department of Statistics, Carnegie Mellon University Pittsburgh.

MERLEVÈDE, F., M. PELIGRAD, AND E. RIO (2009): *Bernstein inequality and moderate deviations under strong mixing conditions*, Beachwood, Ohio, USA: Institute of Mathematical Statistics, vol. Volume 5 of *Collections*, 273–292.

QIAN, J. AND L. SU (2016): "Shrinkage Estimation of Regression Models with Multiple Structural Changes," *Econometric Theory*, 32, 1376–1433.

TIBSHIRANI, R., M. SAUNDERS, S. ROSSET, J. ZHU, AND K. KNIGHT (2005): "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 91–108.

YU, D., J.-H. WON, T. LEE, J. LIM, AND S. YOON (2013): "High-dimensional Fused Lasso Regression using Majorization-Minimization and Parallel Processing," *ArXiv e-prints*.

A Additional Lemmas

The lemmas are adjusted versions of the lemmas in Qian and Su (2016) adopted to the adaptive fusion penalty.

Lemma A.1 *Consider the optimization problem (2.5) with a known Σ and without loss of generality $\Sigma = I$, then the following holds for the solution $\{\hat{\beta}_t, t = 1, 2, \dots, T\}$ and $\{\hat{\theta}_t, t = 1, 2, \dots, T : \hat{\theta}_1 = \hat{\beta}_1,\ \hat{\theta}_t = \hat{\beta}_t - \hat{\beta}_{t-1}$ for $t \geq 2\}$, i.e. $\hat{\beta}_r = \sum_{s=1}^{r} \hat{\theta}_s$ and $k \in \{1, \dots, p\}$:*

(i-a) $\sum_{r=\hat{T}_j}^{T} x_{r,k}(y_r - x_r' \sum_{s=1}^{r} \hat{\theta}_s) = \frac{\lambda}{2} w_{\hat{T}_j,k} \mathrm{sgn}\left[\hat{\theta}_{\hat{T}_j,k}\right]$ *for* $j = 1, \dots, \hat{m}$ *and* $\hat{\theta}_{t,k} \neq 0$,

(i-b) $|\sum_{r=\hat{T}_j}^{T} x_{r,k}(y_r - x_r' \sum_{s=1}^{r} \hat{\theta}_s)| \leq \frac{\lambda}{2} w_{\hat{T}_j,k}$ *for* $j = 1, \dots, \hat{m}$ *and* $\hat{\theta}_{t,k} = 0$,

(ii) $\left\|\sum_{r=1}^{T} x_r(y_r - x_r'\hat{\beta}_r)\right\| = 0$ *for* $t = 1$,

(iii) $\left\|\sum_{r=t}^{T} x_r(y_r - x_r'\hat{\beta}_r)\right\| \leq \frac{\lambda}{2} \|diag(W_t)\|$ *for* $t = 2, \dots, T$,

where W_t is a diagonal $p \times p$ matrix with the weights $w_{t,k}$, $t = \{2, \dots, T\}$ and $k \in \{1, \dots, p\}$, on the diagonal and $diag(\cdot)$ returns the main matrix diagonal as a vector.

Proof: To minimize (2.5), it simplifies the analysis when β_t's are replaced by θ_t's as defined in the lemma. The θ's are just the differences between the β coefficients and θ_1 is then equal to β_1. In this case, the following first order conditions have to hold for $\hat{\theta}$'s based on the subdifferential calculus:

$$-2\sum_{r=1}^{T}\left(y_r - x_r'\sum_{s=1}^{r}\hat{\theta}_s\right)x_r = 0 \qquad \text{if } t = 1, \tag{A.1}$$

$$-2\sum_{r=t}^{T}\left(y_r - x_r'\sum_{s=1}^{r}\hat{\theta}_s\right)x_r + \lambda W_t e_t = 0 \qquad \text{if } t = 2, \dots, T, \tag{A.2}$$

where W_t is a diagonal $p \times p$ matrix having the weights $w_{t,k}$, $k \in \{1, \dots, p\}$, on the diagonal and e_t is a $p \times 1$ vector with the following elements for $k \in \{1, \dots, p\}$:

$$\begin{aligned} e_{t,k} &= \mathrm{sgn}\left[\hat{\theta}_{t,k}\right] && \text{if } \hat{\theta}_{t,k} \neq 0, \\ |e_{t,k}| &\leq 1 && \text{if } \hat{\theta}_{t,k} = 0. \end{aligned}$$

From here it is straightforward to derive (i-a) and (i-b) for individual coefficients k at $\hat{T}_j$:

$$\sum_{r=\hat{T}_j}^{T} x_r(y_r - x_r'\sum_{s=1}^{r}\hat{\theta}_s) = \frac{\lambda}{2} W_{\hat{T}_j} e_{\hat{T}_j} \text{ for } j = 1, \dots, \hat{m}.$$

The result follows from the definition of $e_{t,k}$. From (A.1) and $\hat{\beta}_r = \sum_{s=1}^{r} \hat{\theta}_s$, we get (ii). As $\|W_t e_t\| \leq \|diag(W_t)\|$, then:

$$\left\|\sum_{r=t}^{T} x_r(y_r - x_r'\hat{\beta}_r)\right\| \leq \frac{\lambda}{2} \|diag(W_t)\| \text{ for } t = 2, \dots, T.$$

□

The following lemma is based on the proof of Theorem 2 in Merlevède et al. (2009).

Lemma A.2 *Let $\{\xi_t, t = 1, 2, \dots\}$ be a zero-mean strong mixing process with the mixing coefficients $\alpha(\tau) \leq c_\alpha \rho^\tau$ for some $c_\alpha > 0$ and $\rho \in (0, 1)$. If $\sup_{1\leq t\leq T} |\xi_t| \leq M_T$, then there exists a constant C_0 depending on c_α and ρ such that for any $T \geq 2$ and $\varepsilon > 0$,*

$$P\left(\left|\sum_{t=1}^{T} \xi_t\right| > \varepsilon\right) \leq \exp\left(-\frac{C_0\varepsilon^2}{\nu_0^2 T + M_T^2 + \varepsilon M_T (\log T)^2}\right),$$

where $\nu_0^2 = \sup_{t\geq 1}\left[var(\xi_t) + 2\sum_{s=t+1}^{\infty} |cov(\xi_t, \xi_s)|\right]$.

Proof (Merlevède et al., 2009, Theorem 2): In the mentioned source, the inequality is proved for $\alpha_\tau \leq \exp(-2c\tau)$ for some $c > 0$. If $c_\alpha = 1$, we can take the $\rho = \exp(-2c)$ and apply the theorem to get (i). □

Lemma A.3 *Let Assumptions 2, 3 and 5 hold and let $\nu_T = T\delta_T$. Under these assumptions:*

(i) $\sup_{\substack{1\leq s<r\leq T+1 \\ r-s\geq\nu_T}} \mu_{max}\left(\frac{1}{r-s}\sum_{t=s}^{r-1} x_t x_t'\right) \leq \bar{c}_{xx} + o_p(1)$,

(ii) $\inf_{\substack{1\leq s<r\leq T+1 \\ r-s\geq\nu_T}} \mu_{min}\left(\frac{1}{r-s}\sum_{t=s}^{r-1} x_t x_t'\right) \geq \underline{c}_{xx} + o_p(1)$.

Proof: (i)

Using the Weyl's inequality for eigenvalues of a perturbed Hermitian matrix, another important inequality $\mu_{\max}(A) \leq \|A\|_F$ for any matrix A where $\|\cdot\|_F$ denotes Frobenius norm and Assumption 3:

$$\begin{aligned}
\mu_{\max}\left(\frac{1}{r-s}\sum_{t=s}^{r-1} x_t x_t'\right) &= \mu_{\max}\left(\frac{1}{r-s}\sum_{t=s}^{r-1} E(x_t x_t') + \frac{1}{r-s}\sum_{t=s}^{r-1}(x_t x_t' - E(x_t x_t'))\right) \\
&\leq \mu_{\max}\left(\frac{1}{r-s}\sum_{t=s}^{r-1} E(x_t x_t')\right) + \mu_{\max}\left(\frac{1}{r-s}\sum_{t=s}^{r-1}(x_t x_t' - E(x_t x_t'))\right) \\
&\leq \mu_{\max}\left(\frac{1}{r-s}\sum_{t=s}^{r-1} E(x_t x_t')\right) + \left\|\frac{1}{r-s}\sum_{t=s}^{r-1}(x_t x_t' - E(x_t x_t'))\right\|_F \\
&\leq \bar{c}_{xx} + \left\|\frac{1}{r-s}\sum_{t=s}^{r-1}(x_t x_t' - E(x_t x_t'))\right\|_F .
\end{aligned}$$

Now, it needs to be shown that $\left\|\frac{1}{r-s}\sum_{t=s}^{r-1}(x_t x_t' - E(x_t x_t'))\right\|_F = o_p(1)$. This can be done by proving that $\max_{\substack{1\leq s<r\leq T+1 \\ r-s\geq\nu_T}} \left\|\frac{1}{r-s}\sum_{t=s}^{r-1}(x_t x_t' - E(x_t x_t'))\right\|_F = o_p(1)$.

The Assumption 5 implies $\nu_T \geq c_\nu T^{1/q}$. Let $\eta_T = T^{1/(2q)}$ and ι_{up} be an arbitrary $p \times 1$ unit vector such that $\|\iota_{up}\| = 1$ for $u = 1, 2$. Let $\zeta_t \equiv \iota_{1p}'[x_t x_t' - E(x_t x_t')]\iota_{2p}$, $\zeta_{1t} \equiv \iota_{1p}'[x_t x_t' \mathbb{1}_t - E(x_t x_t' \mathbb{1}_t)]\iota_{2p}$ and $\zeta_{2t} \equiv \iota_{1p}'[x_t x_t'(1 - \mathbb{1}_t) - E(x_t x_t'(1 - \mathbb{1}_t))]\iota_{2p}$ where $\mathbb{1}_t \equiv \mathbb{1}(\|x_t\|^2 \leq \eta_T)$. Under these definitions, $\zeta_t = \zeta_{1t} + \zeta_{2t}$. By Boole inequality and

Lemma A.2, for a sufficiently large positive C:

$$
\begin{aligned}
& P\left(\sup_{\substack{1\le s<r\le T+1\\ r-s\ge \nu_T}}\left|\frac{1}{\sqrt{r-s}}\sum_{t=s}^{r-1}\zeta_{1t}\right|\ge C(\log T)^3\right)\\
&\le T^2\sup_{\substack{1\le s<r\le T+1\\ r-s\ge \nu_T}} P\left(\left|\sum_{t=s}^{r-1}\zeta_{1t}\right|\ge C\sqrt{r-s}(\log T)^3\right)\\
&\le T^2\sup_{\substack{1\le s<r\le T+1\\ r-s\ge \nu_T}}\exp\left(-\frac{C_0C^2(r-s)(\log T)^6}{\nu_0^2(r-s)+4\eta_T^2+2C\sqrt{r-s}(\log T)^3\eta_T[\log(r-s)]^2}\right)\\
&\le T^2\exp\left(-\frac{C_0C^2\nu_T(\log T)^6}{\nu_0^2\nu_T+4\eta_T^2+2C\sqrt{\nu_T}(\log T)^3\eta_T[\log \nu_T]^2}\right)\\
&\le \exp\left(-\frac{C_0C^2\nu_T(\log T)^6}{\nu_0^2\nu_T+4\eta_T^2+2C\sqrt{\nu_T}(\log T)^3\eta_T[\log \nu_T]^2}+2\log T\right)\\
&\to 0 \text{ as } T\to\infty
\end{aligned}
$$

The second but last inequality stems from a careful analysis of the exponential function. Analyzing the first derivative reveals that the function is decreasing in $r-s$ for a fixed T, therefore it achieves its maximum at ν_T, the lower bound of $r-s$. Further analysis of the limiting behavior yields the convergence to zero with $T\to\infty$.

As the next step, by Assumption 2, Boole and Markov inequalities and Lebesgue's dominated convergence theorem:

$$
\begin{aligned}
& P\left(\sup_{\substack{1\le s<r\le T+1\\ r-s\ge \nu_T}}\left|\frac{1}{\sqrt{r-s}}\sum_{t=s}^{r-1}\zeta_{2t}\right|\ge C(\log T)^3\right)\le P\left(\max_{1\le t\le T}\|x_t\|^2\ge \eta_T\right)\\
&\le T\max_{1\le t\le T}P\left(\|x_t\|^2\ge\eta_T\right)\le \frac{T}{\eta_T^{2q}}\max_{1\le t\le T}E\left[\|x_t\|^{4q}\,\mathbb{1}\{\|x_t\|^2\ge\eta_T\}\right]\to 0 \text{ as } T\to\infty,
\end{aligned}
$$

where the first inequality stems from the definition of ζ_{2t} which is non-zero only when $\|x_t\|^2\ge\eta_T$, i.e. the supremum event as a more restrictive event can be bounded from above by $P\left(\max_{1\le t\le T}\|x_t\|^2\ge\eta_T\right)$. Then, the Boole, Markov and Lebesgue's dominated convergence theorem are applied using the Assumption 2 on $\|x_t\|^{4q}$.

Given that the unit vectors ι_{1p} and ι_{2p} are arbitrary, it follows that

$$
\max_{\substack{1\le s<r\le T+1\\ r-s\ge \nu_T}}\left\|\frac{1}{r-s}\sum_{t=s}^{r-1}\left[x_tx_t'-E(x_tx_t')\right]\right\|_F=O_p(\nu_T^{-1/2}(\log T)^3)=o_p(1).
$$

Then, (i) follows. The proof of (ii) follows analogical steps. □

B Proofs

B.1 Proof of Theorem 2.1

Define $A_{T,j} = \{|\hat{T}_j - T_j^0| > T\delta_T\}$ and $C_T = \{\max_{1\le l\le m^0} |\hat{T}_l - T_l^0| < I_{min}/2\}$. Event C_T captures that the largest time distance between an estimated break point and the true break point is at most a half of the shortest period.

Since

$$P\left(\max_{1\le j\le m^0} |\hat{T}_j - T_j^0| > T\delta_T\right) \le \sum_{j=1}^{m^0} P(A_{T,j}),$$

the consistency will be proven by showing (i) $\sum_{j=1}^{m^0} P(A_{T,j} \cap C_T) \to 0$ and (ii) $\sum_{j=1}^{m^0} P(A_{T,j} \cap C_T^C) \to 0$ where the superscript C denotes the set complement.

Proving (i): It will be shown that $\sum_{j=1}^{m^0} P(A_{T,j}^+ \cap C_T) \to 0$ and $\sum_{j=1}^{m^0} P(A_{T,j}^- \cap C_T) \to 0$ where $A_{T,j}^+ = \{\hat{T}_j - T_j^0 > T\delta_T\}$ and $A_{T,j}^- = \{T_j^0 - \hat{T}_j > T\delta_T\}$. First, focus on $A_{T,j}^-$ case, i.e. the estimated break is before the true break on the timeline, $\hat{T}_j < T_j^0$. The other case is proven analogously.

By definition of C_T, $T_{j-1}^0 < \hat{T}_j < T_{j+1}^0$ for all $j \in \{1, \dots, m_0\}$. In other words, C_T describes an event when the j-th estimated break is between the true $(j-1)$-th and $(j+1)$-th break point.

By plugging the model $y_t = x_t'\beta_t^0 + \varepsilon_t$ into Lemma A.1 (iii), it holds for the estimated break points and true break points:

$$\left\| - \sum_{r=\hat{T}_j}^{T} x_r x_r'(\hat{\beta}_r - \beta_r^0) + \sum_{r=\hat{T}_j}^{T} x_r \varepsilon_r \right\| \le \frac{\lambda}{2} \left\| diag(W_{\hat{T}_j}) \right\|,$$

$$\left\| - \sum_{r=T_j^0}^{T} x_r x_r'(\hat{\beta}_r - \beta_r^0) + \sum_{r=T_j^0}^{T} x_r \varepsilon_r \right\| \le \frac{\lambda}{2} \left\| diag(W_{T_j^0}) \right\|.$$

By triangle inequality, the fact that $\hat{\beta}_r = \hat{\alpha}_{j+1}$ and $\beta_r^0 = \alpha_j^0$ for $r \in [\hat{T}_j, T_j^0 - 1]$ and by using another triangle inequality, the following result can be derived, where $c_{W_j} = \left\| diag(W_{\hat{T}_j}) \right\| + \left\| diag(W_{T_j^0}) \right\|$:

$$\begin{aligned}
\lambda c_{W_j}/2 &\ge \left\| - \sum_{r=\hat{T}_j}^{T} x_r x_r'(\hat{\beta}_r - \beta_r^0) + \sum_{r=\hat{T}_j}^{T} x_r \varepsilon_r + \sum_{r=T_j^0}^{T} x_r x_r'(\hat{\beta}_r - \beta_r^0) - \sum_{r=T_j^0}^{T} x_r \varepsilon_r \right\| \\
&= \left\| - \sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'(\hat{\beta}_r - \beta_r^0) + \sum_{r=\hat{T}_j}^{T_j^0-1} x_r \varepsilon_r \right\| \\
&= \left\| - \sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'(\hat{\alpha}_{j+1} - \alpha_j^0) + \sum_{r=\hat{T}_j}^{T_j^0-1} x_r \varepsilon_r \right\|
\end{aligned}$$

$$
\begin{aligned}
&\geq \left\| \sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'(\hat{\alpha}_{j+1} - \alpha_j^0) \right\| - \left\| \sum_{r=\hat{T}_j}^{T_j^0-1} x_r \varepsilon_r \right\| \\
&= \left\| \sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'(\hat{\alpha}_{j+1} - \alpha_{j+1}^0 + \alpha_{j+1}^0 - \alpha_j^0) \right\| - \left\| \sum_{r=\hat{T}_j}^{T_j^0-1} x_r \varepsilon_r \right\| \\
&\geq \left\| \sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'(\alpha_{j+1}^0 - \alpha_j^0) \right\| - \left\| \sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'(\hat{\alpha}_{j+1} - \alpha_{j+1}^0) \right\| - \left\| \sum_{r=\hat{T}_j}^{T_j^0-1} x_r \varepsilon_r \right\| \\
\lambda &\geq \frac{1}{c_{W_j}/2} \left\| \sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'(\alpha_{j+1}^0 - \alpha_j^0) \right\| - \frac{1}{c_{W_j}/2} \left\| \sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'(\hat{\alpha}_{j+1} - \alpha_{j+1}^0) \right\| \\
&\quad - \frac{1}{c_{W_j}/2} \left\| \sum_{r=\hat{T}_j}^{T_j^0-1} x_r \varepsilon_r \right\| \\
&\equiv R_{j,1} - R_{j,2} - R_{j,3}. \qquad \text{(B.1)}
\end{aligned}
$$

Note that $c_{W_j} \leq 2\sqrt{p}\xi_T$. Define $\bar{R}_j(\lambda) = \{\lambda \geq \frac{1}{3}R_{j,1}\} \cup \{R_{j,2} \geq \frac{1}{3}R_{j,1}\} \cup \{R_{j,3} \geq \frac{1}{3}R_{j,1}\}$. The event $\bar{R}_j$ is equivalent to $\lambda \geq R_{j,1} - R_{j,2} - R_{j,3}$, i.e. $P(\bar{R}_j(\lambda)) = 1$. Then,

$$
\begin{aligned}
P(A_{T,j}^- \cap C_T) \leq\ & P(A_{T,j}^- \cap C_T \cap \{\lambda \geq \frac{1}{3}R_{j,1}\}) \\
&+ P(A_{T,j}^- \cap C_T \cap \{R_{j,2} \geq \frac{1}{3}R_{j,1}\}) \\
&+ P(A_{T,j}^- \cap C_T \cap \{R_{j,3} \geq \frac{1}{3}R_{j,1}\}) \\
\equiv\ & AC_{j,1} + AC_{j,2} + AC_{j,3}.
\end{aligned}
$$

Using that $\|Au\| = [tr(uu'A'A)]^{1/2} \geq \mu_{\min}(A'A)^{1/2}\|u\|$, where A is a matrix and u is a vector:

$$
\begin{aligned}
\sum_{j=1}^{m_0} AC_{j,1} &\leq \sum_{j=1}^{m_0} P(A_{T,j}^- \cap \{\lambda \geq \frac{1}{3}R_{j,1}\}) \\
&= \sum_{j=1}^{m_0} P\left(\frac{1}{c_{W_j}(T_j^0 - \hat{T}_j)} \left\| \sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'(\alpha_{j+1}^0 - \alpha_j^0) \right\| \leq \frac{3\lambda}{2(T_j^0 - \hat{T}_j)}; T_j^0 - \hat{T}_j > \delta_T T \right) \\
&\leq \sum_{j=1}^{m_0} P\left(c_{1T,j} \leq \frac{3\lambda\sqrt{p}\xi_T}{J_{\min}\delta_T T}; T_j^0 - \hat{T}_j > \delta_T T \right) \to 0,
\end{aligned}
$$

where the convergence to zero comes from the fact that $c_{1T,j} \equiv \mu_{\min}\left(\frac{1}{T_j^0 - \hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'\right) \geq \underline{c}_{xx}/2 > 0$ with probability going to 1 by Lemma A.3 and the Assumption 8 says that $\frac{3\lambda\sqrt{p}\xi_T}{J_{\min}\delta_T T} \to 0$. This is a contradiction for $T \to \infty$ and therefore the probability converges to zero.

Next, the $AC_{j,2}$ will be bounded.

$$
\begin{aligned}
AC_{j,2} &= P(A_{T,j}^{-} \cap C_T \cap \{R_{j,2} \geq \frac{1}{3} R_{j,1}\}) \\
&\leq P(A_{T,j}^{-} \cap C_T \cap \{\bar{c}_{1T,j} \left\| \hat{\alpha}_{j+1} - \alpha_{j+1}^0 \right\| \geq \frac{1}{3} c_{1T,j} \left\| \alpha_{j+1}^0 - \alpha_j^0 \right\| \}),
\end{aligned}
$$

where $\bar{c}_{1T,j} \equiv \mu_{\max} \left(\frac{1}{T_j^0 - \hat{T}_j} \sum_{r=\hat{T}_j}^{T_j^0 - 1} x_r x_r' \right) \leq 2\bar{c}_{xx}$ with probability going to 1 by Lemma A.3. Given $A_{T,j}^{-}$ and C_T, $\hat{\beta}_t = \hat{\alpha}_{j+1}$ for $t \in [T_j^0, (T_j^0 + T_{j+1}^0)/2 - 1]$. Using Lemma A.1 (iii) and following similar steps as to get (B.1):

$$
\lambda c_{W_j}/2 \geq \left\| \sum_{r=T_j^0}^{(T_j^0 + T_{j+1}^0)/2 - 1} x_r x_r' (\hat{\alpha}_{j+1} - \alpha_{j+1}^0) \right\| - \left\| \sum_{r=T_j^0}^{(T_j^0 + T_{j+1}^0)/2 - 1} x_r \varepsilon_r \right\|,
$$

where $c_{W_j} = \left\| diag(W_{T_j^0}) \right\| + \left\| diag(W_{(T_j^0 + T_{j+1}^0)/2 - 1}) \right\|$. Conditional on C_T:

$$
\left\| \hat{\alpha}_{j+1} - \alpha_{j+1}^0 \right\| \leq (c_{2T,j})^{-1} \left[\frac{2\lambda \sqrt{p} \xi_T}{I_{\min}} + \left\| \frac{2}{T_{j+1}^0 - T_j^0} \sum_{r=T_j^0}^{(T_j^0 + T_{j+1}^0)/2 - 1} x_r u_r \right\| \right],
$$

where $c_{2T,j} \equiv \mu_{\min} \left(\frac{2}{T_{j+1}^0 - T_j^0} \sum_{r=T_j^0}^{(T_j^0 + T_{j+1}^0)/2 - 1} x_r x_r' \right) \geq \underline{c}_{xx}/2 > 0$ with probability going to 1. Therefore:

$$
\begin{aligned}
&\sum_{j=1}^{m_0} P\left(\{ \left\| \hat{\alpha}_{j+1} - \alpha_{j+1}^0 \right\| \geq \frac{1}{3} (\bar{c}_{1T,j})^{-1} c_{1T,j} \left\| \alpha_{j+1}^0 - \alpha_j^0 \right\| \} \cap C_T \right) \\
&\quad \leq \sum_{j=1}^{m_0} P\left(\frac{2\lambda \sqrt{p} \xi_T}{I_{\min}} \geq \frac{1}{3} (\bar{c}_{1T,j})^{-1} c_{1T,j} c_{2T,j} \left\| \alpha_{j+1}^0 - \alpha_j^0 \right\| \right) \\
&\quad + \sum_{j=1}^{m_0} P\left(\left\| \frac{2}{T_{j+1}^0 - T_j^0} \sum_{r=T_j^0}^{(T_j^0 + T_{j+1}^0)/2 - 1} x_r u_r \right\| \geq \frac{1}{3} (\bar{c}_{1T,j})^{-1} c_{1T,j} c_{2T,j} \left\| \alpha_{j+1}^0 - \alpha_j^0 \right\| \right).
\end{aligned}
$$

The first term converges to zero since $\lambda \sqrt{p} \xi_T / (I_{\min} J_{\min}) \to 0$ under Assumptions 6 and 8. The second term is bounded from above by

$$
P\left(\left\| \frac{2}{T_{j+1}^0 - T_j^0} \sum_{r=T_j^0}^{(T_j^0 + T_{j+1}^0)/2 - 1} x_r u_r \right\| \geq \bar{c}_{xx} \underline{c}_{xx}^2 \left\| \alpha_{j+1}^0 - \alpha_j^0 \right\| / 24 \right) \to 0.
$$

The convergence to zero is a consequence of Assumptions 4, 6 and 7 as one can rearrange the elements:

$$
P\left(\frac{1}{J_{\min} T^{1/2} \delta^{1/2}} \left\| \frac{\sqrt{2}}{\sqrt{T_{j+1}^0 - T_j^0}} \sum_{r=T_j^0}^{(T_j^0 + T_{j+1}^0)/2 - 1} x_r u_r \right\| \geq \frac{\bar{c}_{xx} \underline{c}_{xx}^2 I_{\min}^{1/2}}{24 \sqrt{2} T^{1/2} \delta^{1/2}} \right) \to 0.
$$

where the left side is $o\left(\frac{(logT)^3}{J_{\min}T^{1/2}\delta_T^{1/2}}\right)$ and the right hand side diverges as $\frac{I_{\min}^{1/2}}{T^{1/2}\delta^{1/2}} \to \infty$. Therefore, the $\sum_{j=1}^{m_0} AC_{j,2} \to 0$.

Remember that $c_{1T,j} \geq \underline{c}_{xx}/2 > 0$ with probability going to 1. Then,

$$\begin{aligned}
\sum_{j=1}^{m_0} AC_{j,3} &\leq P\left(A_{T,j}^- \cap \{R_{j,3} \geq \frac{1}{3}R_{j,1}\}\right) \\
&= P\left(A_{T,j}^- \cap \{\left\|\frac{1}{T_j^0 - \hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r\varepsilon_r\right\| \geq \frac{1}{3}\left\|\frac{1}{T_j^0 - \hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'(\alpha_{j+1}^0 - \alpha_j^0)\right\|\}\right) \\
&\leq \sum_{j=1}^{m_0} P\left(A_{T,j}^- \cap \{\left\|\frac{1}{\sqrt{T_j^0 - \hat{T}_j}}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r\varepsilon_r\right\| \geq \frac{1}{3}c_{1T,j}J_{\min}I_{\min}^{1/2}\}\right) \to 0,
\end{aligned}$$

where the convergence to zero follows from the same rearrangement as in the previous case based on Assumptions 4, 6 and 7. The part $\sum_{j=1}^{m^0} P(A_{T,j} \cap C_T) \to 0$ is proven.

Proving (ii): Similarly as before, it will be shown $\sum_{j=1}^{m^0} P(A_{T,j}^+ \cap C_T^C) \to 0$ and $\sum_{j=1}^{m^0} P(A_{T,j}^- \cap C_T^C) \to 0$. As the other result can be proven analogously, the $\sum_{j=1}^{m^0} P(A_{T,j}^- \cap C_T^C) \to 0$ is shown below. Define:

$$\begin{aligned}
D_T^{(l)} &\equiv \left\{\exists j \in \{1,\ldots,m_0\}, \hat{T}_j \leq T_{j-1}^0\right\} \cap C_T^C, \\
D_T^{(m)} &\equiv \left\{\forall j \in \{1,\ldots,m_0\}, T_{j-1}^0 < \hat{T}_j \leq T_{j+1}^0\right\} \cap C_T^C, \\
D_T^{(r)} &\equiv \left\{\exists j \in \{1,\ldots,m_0\}, \hat{T}_j \geq T_{j+1}^0\right\} \cap C_T^C.
\end{aligned}$$

From there, $\sum_{j=1}^{m^0} P(A_{T,j}^- \cap C_T^C) = \sum_{j=1}^{m^0} P(A_{T,j}^- \cap D_T^{(l)}) + \sum_{j=1}^{m^0} P(A_{T,j}^- \cap D_T^{(m)}) + \sum_{j=1}^{m^0} P(A_{T,j}^- \cap D_T^{(r)})$.

First, $\sum_{j=1}^{m^0} P(A_{T,j}^- \cap D_T^{(m)})$ will be analyzed.

$$\begin{aligned}
P(A_{T,j}^- \cap D_T^{(m)}) &= P\left(A_{T,j}^- \cap \left\{\hat{T}_{j+1} - T_j^0 \geq \frac{1}{2}I_{\min}\right\} \cap D_T^{(m)}\right) \\
&\quad + P\left(A_{T,j}^- \cap \left\{\hat{T}_{j+1} - T_j^0 < \frac{1}{2}I_{\min}\right\} \cap D_T^{(m)}\right) \\
&\leq P\left(A_{T,j}^- \cap \left\{\hat{T}_{j+1} - T_j^0 \geq \frac{1}{2}I_{\min}\right\} \cap D_T^{(m)}\right) \\
&\quad + P\left(A_{T,j}^- \cap \left\{T_{j+1}^0 - \hat{T}_{j+1}^0 \geq \frac{1}{2}I_{\min}\right\} \cap D_T^{(m)}\right),
\end{aligned}$$

where the inequality is based on the fact that $0 \leq \hat{T}_{j+1} - T_j^0 \leq I_{\min}/2$ implies $T_{j+1}^0 - \hat{T}_{j+1}^0 = (T_{j+1}^0 - T_j^0) - (\hat{T}_{j+1}^0 - T_j^0) \geq I_{\min} - I_{\min}/2 = I_{\min}/2$. It will be helpful to see the following set hierarchy:

$$\begin{aligned}
&\left\{A_{T,j}^- \cap \left\{T_{j+1}^0 - \hat{T}_{j+1}^0 \geq \frac{1}{2}I_{\min}\right\} \cap D_T^{(m)}\right\} \subset \\
&\qquad \cup_{k=j+1}^{m^0-1}\left(\left\{T_k^0 - \hat{T}_k \geq \frac{1}{2}I_{\min}\right\} \cap \left\{\hat{T}_{k+1} - T_k^0 \geq \frac{1}{2}I_{\min}\right\} \cap D_T^{(m)}\right).
\end{aligned}$$

Then,

$$
\begin{aligned}
&\sum_{j=1}^{m^0} P\left(A_{T,j}^{-} \cap D_T^{(m)}\right) \\
&\qquad \leq P\left(A_{T,j}^{-} \cap \left\{\hat{T}_{j+1} - T_j^0 \geq \frac{1}{2} I_{\min}\right\} \cap D_T^{(m)}\right) \\
&\qquad + \sum_{j=1}^{m^0} \sum_{k=j+1}^{m^0-1} P\left(\left\{T_k^0 - \hat{T}_k \geq \frac{1}{2} I_{\min}\right\} \cap \left\{\hat{T}_{k+1} - T_k^0 \geq \frac{1}{2} I_{\min}\right\} \cap D_T^{(m)}\right). \qquad \text{(B.2)}
\end{aligned}
$$

To find the bounds, the following steps have to be made. In a similar fashion as in the previous part using Lemma A.1 (iii) for two points T_j^0 and $\hat{T}_j$:

$$
\begin{aligned}
\frac{\lambda \sqrt{p} \xi_T}{T_j^0 - \hat{T}_j} &\geq \frac{1}{T_j^0 - \hat{T}_j} \left\| - \sum_{r=\hat{T}_j}^{T_j^0 - 1} x_r x_r' (\hat{\alpha}_{j+1} - \alpha_j^0) + \sum_{r=\hat{T}_j}^{T_j^0 - 1} x_r \varepsilon_r \right\| \\
&\geq c_{1T,j} \left\| \hat{\alpha}_{j+1} - \alpha_j^0 \right\| - \left\| \frac{1}{T_j^0 - \hat{T}_j} \sum_{r=\hat{T}_j}^{T_j^0 - 1} x_r \varepsilon_r \right\|
\end{aligned}
$$

Thus,

$$
\left\| \hat{\alpha}_{j+1} - \alpha_j^0 \right\| \leq (c_{1T,j})^{-1} \left[\frac{\lambda \sqrt{p} \xi_T}{T_j^0 - \hat{T}_j} + \left\| \frac{1}{T_j^0 - \hat{T}_j} \sum_{r=\hat{T}_j}^{T_j^0 - 1} x_r \varepsilon_r \right\| \right]. \qquad \text{(B.3)}
$$

Analogically, using Lemma A.1 (iii) for two points T_j^0 and $\hat{T}_{j+1}$:

$$
\begin{aligned}
\frac{\lambda \sqrt{p} \xi_T}{\hat{T}_{j+1} - T_j^0} &\geq \frac{1}{\hat{T}_{j+1} - T_j^0} \left\| - \sum_{r=T_j^0}^{\hat{T}_{j+1} - 1} x_r x_r' (\hat{\alpha}_{j+1} - \alpha_{j+1}^0) + \sum_{r=T_j^0}^{\hat{T}_{j+1} - 1} x_r \varepsilon_r \right\| \\
&\geq c_{3T,j} \left\| \hat{\alpha}_{j+1} - \alpha_{j+1}^0 \right\| - \left\| \frac{1}{\hat{T}_{j+1} - T_j^0} \sum_{r=T_j^0}^{\hat{T}_{j+1} - 1} x_r \varepsilon_r \right\|,
\end{aligned}
$$

where $c_{3T,j} \equiv \mu_{\min}\left(\frac{1}{\hat{T}_{j+1} - T_j^0} \sum_{r=T_j^0}^{\hat{T}_{j+1}-1} x_r x_r'\right) \geq \underline{c}_{xx}/2 > 0$ with probability going to 1. Thus,

$$
\left\| \hat{\alpha}_{j+1} - \alpha_{j+1}^0 \right\| \leq (c_{3T,j})^{-1} \left[\frac{\lambda \sqrt{p} \xi_T}{\hat{T}_{j+1} - T_j^0} + \left\| \frac{1}{\hat{T}_{j+1} - T_j^0} \sum_{r=T_j^0}^{\hat{T}_{j+1} - 1} x_r \varepsilon_r \right\| \right]. \qquad \text{(B.4)}
$$

Define

$$
\begin{aligned}
E_{T,j} \equiv \Bigg\{ &\left\| \alpha_{j+1}^0 - \alpha_j^0 \right\| \leq \lambda \sqrt{p} \xi_T \left(\frac{1}{T_j^0 - \hat{T}_j} (c_{1T,j})^{-1} + \frac{1}{\hat{T}_{j+1} - T_j^0} (c_{3T,j})^{-1} \right) \\
&(c_{1T,j})^{-1} \left\| \frac{1}{T_j^0 - \hat{T}_j} \sum_{r=\hat{T}_j}^{T_j^0 - 1} x_r \varepsilon_r \right\| + (c_{3T,j})^{-1} \left\| \frac{1}{\hat{T}_{j+1} - T_j^0} \sum_{r=T_j^0}^{\hat{T}_{j+1} - 1} x_r \varepsilon_r \right\| \Bigg\}. \qquad \text{(B.5)}
\end{aligned}
$$

Applying the triangle inequality on (B.3) and (B.4), yields that $E_{T,j}$ has probability one. Then,

$$\begin{aligned}
&\sum_{j=1}^{m^0} P\left(A_{T,j}^{-} \cap \left\{\hat{T}_{j+1} - T_j^0 \geq \frac{1}{2} I_{\min}\right\} \cap D_T^{(m)}\right) \\
&\quad = \sum_{j=1}^{m^0} P\left(E_{T,j} \cap A_{T,j}^{-} \cap \left\{\hat{T}_{j+1} - T_j^0 \geq \frac{1}{2} I_{\min}\right\} \cap D_T^{(m)}\right) \\
&\quad \leq \sum_{j=1}^{m^0} P\left(E_{T,j} \cap \left\{T_j^0 - \hat{T}_j > T\delta_T\right\} \cap \left\{\hat{T}_{j+1} - T_j^0 \geq \frac{1}{2} I_{\min}\right\}\right) \\
&\quad \leq \sum_{j=1}^{m^0} P\left(\frac{\lambda\sqrt{p}\xi_T}{T\delta_T}(c_{1T,j})^{-1} + \frac{2\lambda\sqrt{p}\xi_T}{I_{\min}}(c_{3T,j})^{-1} \geq \left\|\alpha_{j+1}^0 - \alpha_j^0\right\|/3\right) \\
&\qquad + P\left(\left\{(c_{1T,j})^{-1}\left\|\frac{1}{T_j^0 - \hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r\varepsilon_r\right\| \geq \left\|\alpha_{j+1}^0 - \alpha_j^0\right\|/3\right\} \cap \left\{T_j^0 - \hat{T}_j > T\delta_T\right\}\right) \\
&\qquad + P\left(\left\{(c_{3T,j})^{-1}\left\|\frac{1}{\hat{T}_{j+1} - T_j^0}\sum_{r=T_j^0}^{\hat{T}_{j+1}-1} x_r\varepsilon_r\right\|\right.\right. \\
&\qquad\qquad\qquad \left.\left. \geq \left\|\alpha_{j+1}^0 - \alpha_j^0\right\|/3\right\} \cap \left\{\hat{T}_{j+1} - T_j^0 \geq \frac{1}{2} I_{\min}\right\}\right) \qquad \text{(B.6)}
\end{aligned}$$

The first term in (B.6) goes asymptotically to zero as $\frac{\lambda\sqrt{p}\xi_T}{J_{\min}T\delta_T} \to 0$ and $\frac{T\delta_T\lambda\sqrt{p}\xi_T}{I_{\min}J_{\min}T\delta_T} \to 0$ by Assumptions 6 and 8. The middle term converges to zero by applying Assumptions 4 and 7. The last term is also converging to zero by imposing Assumptions 4, 6 and 7. The proof is similar to the proof for the second term in (B.2), i.e. $\sum_{j=1}^{m^0} P(A_{T,j}^{-} \cap D_T^{(m)}) \to 0$.

Now, $\sum_{j=1}^{m^0} P(A_{T,j}^{-} \cap D_T^{(l)})$ will be analyzed.

$$P(A_{T,j}^{-} \cap D_T^{(l)}) \leq P(D_T^{(l)}) \leq \sum_{r=1}^{m^0} 2^{r-1} P\left(\max\left\{l \in \{1,\ldots,m^0\} : \hat{T}_l \leq T_{l-1}^0\right\} = r\right),$$

where the probability of an event $P(D_T^{(l)})$ is bounded by the sum of probabilities of all the individual events as follows. If the largest index which satisfies $\hat{T}_l \leq T_{l-1}^0$ is r, then the probability of this event bounds all the events in which any s-tuple of indices from $\{1,\ldots,r\}$ where $s \leq r$ satisfies the inequality $\hat{T}_l \leq T_{l-1}^0$. In total, there are 2^{r-1} of such s-tuples. The event $\max\left\{l \in \{1,\ldots,m^0\} : \hat{T}_l \leq T_{l-1}^0\right\} = r$ implies $\hat{T}_r \leq T_{r-1}^0$ and $\hat{T}_{l+1} \leq T_l^0$ for $l = r,\ldots,m^0$, and $\max\left\{l \in \{1,\ldots,m^0\} : \hat{T}_l \leq T_{l-1}^0\right\} = r \subset \cup_{k=r}^{m^0-1}\left(\left\{T_k^0 - \hat{T}_k \geq I_{\min}/2\right\} \cap \left\{\hat{T}_{k+1}^0 - T_k^0 \geq I_{\min}/2\right\}\right)$ by the geometry of the restrictions. Then,

$$\sum_{j=1}^{m^0} P\left(A_{T,j}^{-} \cap D_T^{(l)}\right)$$

$$
\begin{aligned}
&\leq m^0 \sum_{r=1}^{m^0-1} 2^{r-1} \sum_{k=r}^{m^0-1} P\left(\left\{T_k^0 - \hat{T}_k \geq I_{\min}/2\right\} \cap \left\{\hat{T}_{k+1}^0 - T_k^0 \geq I_{\min}/2\right\}\right) \\
&\quad + m^0 2^{m^0-1} P\left(T_{m^0}^0 - \hat{T}_{m^0} \geq I_{\min}/2\right) \qquad \text{(B.7)}
\end{aligned}
$$

Using (B.5) for $j = m^0$ says that probability of the event of E_{T,m^0} is 1. Then,

$$
\begin{aligned}
&m^0 2^{m^0-1} P\left(T_{m^0}^0 - \hat{T}_{m^0} \geq I_{\min}/2\right) \\
&\quad = m^0 2^{m^0-1} P\left(E_{T,m^0} \cap T_{m^0}^0 - \hat{T}_{m^0} \geq I_{\min}/2\right) \\
&\quad \leq m^0 2^{m^0-1} P\left(\frac{\lambda\sqrt{p}\xi_T}{T\delta_T}(c_{1T,m^0})^{-1} + \frac{2\lambda\sqrt{p}\xi_T}{I_{\min}}(c_{3T,m^0})^{-1} \geq \left\|\alpha_{m^0+1}^0 - \alpha_{m^0}^0\right\|/3\right) \\
&\qquad + m^0 2^{m^0-1} P\left(\left\{(c_{1T,m^0})^{-1}\left\|\frac{1}{T_{m^0}^0 - \hat{T}_{m^0}} \sum_{r=\hat{T}_{m^0}}^{T_{m^0}^0-1} x_r\varepsilon_r\right\|\right.\right. \\
&\qquad\qquad \left.\left. \geq \left\|\alpha_{m^0+1}^0 - \alpha_{m^0}^0\right\|/3\right\}, T_{m^0}^0 - \hat{T}_{m^0} \geq I_{\min}/2\right) \\
&\qquad + m^0 2^{m^0-1} P\left(\left\{(c_{3T,m^0})^{-1}\left\|\frac{1}{T - T_{m^0}^0} \sum_{r=T_{m^0}^0}^{T} x_r\varepsilon_r\right\| \geq \left\|\alpha_{m^0+1}^0 - \alpha_{m^0}^0\right\|/3\right\}\right) \\
&\quad \to 0,
\end{aligned}
$$

in a similar fashion as for (B.6) using $m^0 2^{m^0-1} = O(T\log T)$ and the concentration inequality in Lemma A.2. The rate for the number of breaks comes from this concentration result as the $\log(T\log T)$ can be squeezed to the exponential without inflating the tail probability.

The first term in (B.7) can be bounded by using (B.5) for $j = k$, the Assumption 1, and the concentration inequality in Lemma A.2, i.e.

$$
\begin{aligned}
&m^0 \sum_{r=1}^{m^0-1} 2^{r-1} \sum_{k=r}^{m^0-1} P\left(\left\{T_k^0 - \hat{T}_k \geq I_{\min}/2\right\} \cap \left\{\hat{T}_{k+1}^0 - T_k^0 \geq I_{\min}/2\right\}\right) \\
&\quad \leq m^0 2^{m^0-1} \sum_{k=1}^{m^0-1} P\left(E_{T,k} \cap \left\{T_k^0 - \hat{T}_k \geq I_{\min}/2\right\} \cap \left\{\hat{T}_{k+1}^0 - T_k^0 \geq I_{\min}/2\right\}\right) \\
&\quad \leq m^0 2^{m^0-1} \sum_{k=1}^{m^0-1} P\left(\frac{\lambda\sqrt{p}\xi_T}{T\delta_T}(c_{1T,k})^{-1} + \frac{2\lambda\sqrt{p}\xi_T}{I_{\min}}(c_{3T,k})^{-1} \geq \left\|\alpha_{k+1}^0 - \alpha_k^0\right\|/3\right) \\
&\qquad + m^0 2^{m^0-1} \sum_{k=1}^{m^0-1} P\left(\left\{(c_{1T,k})^{-1}\left\|\frac{1}{T_k^0 - \hat{T}_k} \sum_{s=\hat{T}_k}^{T_k^0-1} x_s\varepsilon_s\right\|\right.\right. \\
&\qquad\qquad \left.\left. \geq \left\|\alpha_{k+1}^0 - \alpha_k^0\right\|/3\right\} \cap \left\{T_k^0 - \hat{T}_k > I_{\min}/2\right\}\right) \\
&\qquad + m^0 2^{m^0-1} \sum_{k=1}^{m^0-1} P\left(\left\{(c_{3T,k})^{-1}\left\|\frac{1}{\hat{T}_{k+1} - T_k^0} \sum_{s=T_k^0}^{\hat{T}_{k+1}-1} x_s\varepsilon_s\right\|\right.\right. \\
&\qquad\qquad \left.\left. \geq \left\|\alpha_{k+1}^0 - \alpha_k^0\right\|/3\right\} \cap \left\{\hat{T}_{k+1} - T_k^0 \geq \frac{1}{2}I_{\min}\right\}\right) \to 0.
\end{aligned}
$$

This yields the final result $\sum_{j=1}^{m^0} P(A^-_{T,j} \cap D^{(l)}_T) \to 0$. The last part $\sum_{j=1}^{m^0} P(A^-_{T,j} \cap D^{(r)}_T) \to 0$ follows analogously. $\square$

C Tables and Figures

C.1 One Parameter - No Break - Average Squared Bias

Table 2: Average Squared Bias, No Break, 1000 draws, SNR=2, $\delta = 0.025$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.044	0.045	0.092	0.044	0.047	0.137	0.069	0.089	0.107
100	0.020	0.020	0.046	0.020	0.020	0.043	0.023	0.025	0.039
200	0.011	0.011	0.030	0.011	0.011	0.014	0.011	0.011	0.020

Table 3: Average Squared Bias, No Break, 1000 draws, SNR=2, $\delta = 0.050$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.044	0.045	0.090	0.044	0.047	0.142	0.071	0.091	0.107
100	0.020	0.020	0.041	0.020	0.020	0.042	0.024	0.025	0.039
200	0.011	0.011	0.025	0.011	0.011	0.014	0.011	0.011	0.020

Table 4: Average Squared Bias, No Break, 1000 draws, SNR=2, $\delta = 0.075$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.044	0.045	0.089	0.044	0.047	0.140	0.075	0.096	0.107
100	0.020	0.020	0.040	0.020	0.020	0.044	0.024	0.025	0.039
200	0.011	0.011	0.024	0.011	0.011	0.014	0.011	0.011	0.020

Table 5: Best δ for the Bias, No Break, 1000 draws, SNR=2

	1 step	2 step		
T	IC_1	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.025	0.025	0.025	0.025
100	0.025	0.025	0.025	0.025
200	0.025	0.025	0.025	0.025

Table 6: Average Squared Bias, No Break, 1000 draws, SNR=1, $\delta = 0.025$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.104	0.105	0.176	0.103	0.107	0.321	0.196	0.239	0.205
100	0.040	0.040	0.087	0.041	0.041	0.106	0.055	0.062	0.082
200	0.020	0.020	0.053	0.020	0.020	0.034	0.022	0.022	0.037

Table 7: Average Squared Bias, No Break, 1000 draws, SNR=1, $\delta = 0.050$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.104	0.105	0.171	0.103	0.107	0.310	0.205	0.241	0.205
100	0.040	0.040	0.079	0.041	0.041	0.102	0.057	0.062	0.082
200	0.020	0.020	0.044	0.020	0.020	0.034	0.023	0.023	0.037

Table 8: Average Squared Bias, No Break, 1000 draws, SNR=1, $\delta = 0.075$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.104	0.105	0.170	0.103	0.107	0.309	0.216	0.247	0.205
100	0.040	0.040	0.076	0.041	0.041	0.104	0.058	0.065	0.082
200	0.020	0.020	0.042	0.020	0.020	0.035	0.023	0.023	0.037

Table 9: Best δ for the Bias, No Break, 1000 draws, SNR=1

	1 step	2 step		
T	IC_1	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.075	0.075	0.025	0.025
100	0.075	0.050	0.025	0.025
200	0.075	0.025	0.025	0.025

C.2 One Parameter - No Break - Rates of Falsely Detected Breaks

Table 10: Rates of False Positives (FP), 1000 draws, SNR=2, $\delta = 0.025$

	2 step					Bai Perron	
T	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	DM	Seq
50	0.8	1.0	37.8	19.8	22.5	53.2	22.4
100	0.0	0.0	25.0	11.7	12.0	32.6	13.6
200	0.2	0.2	10.7	5.5	5.5	18.3	11.4

Table 11: Rates of False Positives (FP), 1000 draws, SNR=2, $\delta = 0.050$

	2 step					Bai Perron	
T	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	DM	Seq
50	0.8	1.0	38.7	20.1	22.6	53.2	22.4
100	0.0	0.0	24.9	10.7	10.9	32.6	13.6
200	0.2	0.2	10.6	5.1	5.1	18.3	11.4

Table 12: Rates of False Positives (FP), 1000 draws, SNR=2, $\delta = 0.075$

	2 step					Bai Perron	
T	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	DM	Seq
50	0.8	1.0	38.3	19.5	21.7	53.2	22.4
100	0.0	0.0	26.0	10.3	10.8	32.6	13.6
200	0.2	0.2	10.8	4.6	4.6	18.3	11.4

Table 13: Best δ for the FP, No Break, 1000 draws, SNR=2

	2 step		
T	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.025	0.075	0.075
100	0.050	0.075	0.075
200	0.050	0.075	0.075

Table 14: Rates of False Positives (FP), 1000 draws, SNR=1, $\delta = 0.025$

T	2 step $EBIC$ $EBIC$	 QS QS	 IC_1 IC_2	 IC_1 $EBIC$	 IC_1 QS	Bai Perron DM	 Seq
50	4.2	4.5	48.4	40.0	40.6	46.7	21.4
100	0.4	0.4	33.5	23.1	23.3	30.5	12.6
200	0.0	0.0	19.1	12.0	12.0	16.6	9.9

Table 15: Rates of False Positives (FP), 1000 draws, SNR=1, $\delta = 0.050$

T	2 step $EBIC$ $EBIC$	 QS QS	 IC_1 IC_2	 IC_1 $EBIC$	 IC_1 QS	Bai Perron DM	 Seq
50	4.2	4.5	46.1	37.0	37.6	46.7	21.4
100	0.4	0.4	31.0	20.0	20.0	30.5	12.6
200	0.0	0.0	17.7	10.1	10.1	16.6	9.9

Table 16: Rates of False Positives (FP), 1000 draws, SNR=1, $\delta = 0.075$

T	2 step $EBIC$ $EBIC$	 QS QS	 IC_1 IC_2	 IC_1 $EBIC$	 IC_1 QS	Bai Perron DM	 Seq
50	4.2	4.5	46.0	36.8	37.3	46.7	21.4
100	0.4	0.4	30.9	18.9	19.1	30.5	12.6
200	0.0	0.0	17.2	9.4	9.4	16.6	9.9

Table 17: Best δ for the FP, No Break, 1000 draws, SNR=1

T	2 step IC_1 IC_2	 IC_1 $EBIC$	 IC_1 QS
50	0.075	0.075	0.075
100	0.075	0.075	0.075
200	0.075	0.075	0.075

C.3 Two Parameters - No Break - Average Squared Bias

Table 18: Average Squared Bias, No Break, 1000 draws, SNR=2, $\delta = 0.025$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.074	0.074	0.155	0.074	0.073	0.213	0.096	0.094	0.178
100	0.035	0.035	0.083	0.035	0.035	0.063	0.038	0.038	0.074
200	0.019	0.019	0.049	0.019	0.019	0.022	0.019	0.019	0.030

Table 19: Average Squared Bias, No Break, 1000 draws, SNR=2, $\delta = 0.050$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.074	0.074	0.144	0.074	0.073	0.210	0.095	0.095	0.178
100	0.035	0.035	0.074	0.035	0.035	0.065	0.038	0.038	0.074
200	0.019	0.019	0.043	0.019	0.019	0.023	0.019	0.019	0.030

Table 20: Average Squared Bias, No Break, 1000 draws, SNR=2, $\delta = 0.075$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.074	0.074	0.151	0.074	0.073	0.227	0.099	0.098	0.178
100	0.035	0.035	0.072	0.035	0.035	0.066	0.039	0.038	0.074
200	0.019	0.019	0.040	0.019	0.019	0.023	0.019	0.019	0.030

Table 21: Best δ for the Bias, No Break, 1000 draws, SNR=2

	1 step	2 step		
T	IC_1	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.025	0.025	0.025	0.025
100	0.025	0.025	0.025	0.025
200	0.025	0.025	0.025	0.025

Table 22: Average Squared Bias, No Break, 1000 draws, SNR=1, $\delta = 0.025$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.160	0.244	0.295	0.161	0.177	0.538	0.251	0.251	0.347
100	0.072	0.072	0.152	0.072	0.072	0.181	0.086	0.083	0.144
200	0.034	0.034	0.090	0.034	0.034	0.049	0.036	0.036	0.056

Table 23: Average Squared Bias, No Break, 1000 draws, SNR=1, $\delta = 0.050$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.160	0.244	0.281	0.161	0.177	0.513	0.252	0.255	0.347
100	0.072	0.072	0.140	0.072	0.072	0.169	0.087	0.086	0.144
200	0.034	0.034	0.078	0.034	0.034	0.052	0.036	0.036	0.056

Table 24: Average Squared Bias, No Break, 1000 draws, SNR=1, $\delta = 0.075$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.160	0.244	0.284	0.161	0.177	0.522	0.263	0.259	0.347
100	0.072	0.072	0.137	0.072	0.072	0.178	0.088	0.087	0.144
200	0.034	0.034	0.073	0.034	0.034	0.053	0.037	0.037	0.056

Table 25: Best δ for the Bias, No Break, 1000 draws, SNR=1

	1 step	2 step		
T	IC_1	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.025	0.025	0.025	0.025
100	0.025	0.025	0.025	0.025
200	0.025	0.025	0.025	0.025

C.4 Two Parameter - No Break - Rates of Falsely Detected Breaks

Table 26: Rates of False Positives (FP), 1000 draws, SNR=2, $\delta = 0.025$

	2 step					Bai Perron	
T	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	DM	Seq
50	0.6	0.8	62.6	56.1	56.1	90.9	35.8
100	0.0	0.0	41.3	35.6	35.6	63.6	23.1
200	0.1	0.0	21.2	19.1	19.1	34.7	14.3

Table 27: Rates of False Positives (FP), 1000 draws, SNR=2, $\delta = 0.050$

	2 step					Bai Perron	
T	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	DM	Seq
50	0.6	0.8	60.7	51.9	51.9	90.9	35.8
100	0.0	0.0	38.8	31.9	31.9	63.6	23.1
200	0.1	0.0	18.3	15.3	15.3	34.7	14.3

Table 28: Rates of False Positives (FP), 1000 draws, SNR=2, $\delta = 0.075$

	2 step					Bai Perron	
T	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	DM	Seq
50	0.6	0.8	59.1	51.5	51.5	90.9	35.8
100	0.0	0.0	37.8	30.4	30.4	63.6	23.1
200	0.1	0.0	18.6	15.2	15.2	34.7	14.3

Table 29: Best δ for the FP, No Break, 1000 draws, SNR=2

	2 step		
T	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.075	0.075	0.075
100	0.075	0.075	0.075
200	0.050	0.075	0.075

Table 30: Rates of False Positives (FP), 1000 draws, SNR=1, $\delta = 0.025$

	2 step					Bai Perron	
T	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	DM	Seq
50	5.7	8.7	83.5	82.5	82.5	90.7	35.3
100	0.3	0.3	65.0	63.9	63.9	60.0	22.4
200	0.0	0.0	42.2	41.5	41.5	36.7	12.6

Table 31: Rates of False Positives (FP), 1000 draws, SNR=1, $\delta = 0.050$

	2 step					Bai Perron	
T	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	DM	Seq
50	5.7	8.7	80.2	79.0	79.0	90.7	35.3
100	0.3	0.3	61.1	59.3	59.3	60.0	22.4
200	0.0	0.0	36.1	34.7	34.7	36.7	12.6

Table 32: Rates of False Positives (FP), 1000 draws, SNR=1, $\delta = 0.075$

	2 step					Bai Perron	
T	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	DM	Seq
50	5.7	8.7	79.0	77.2	77.2	90.7	35.3
100	0.3	0.3	60.1	57.9	57.9	60.0	22.4
200	0.0	0.0	34.7	33.0	33.0	36.7	12.6

Table 33: Best δ for the FP, No Break, 1000 draws, SNR=1

	2 step		
T	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.075	0.075	0.075
100	0.075	0.075	0.075
200	0.075	0.075	0.075

C.5 One Parameter - One Break - Middle - Average Squared Bias

Table 34: Average Squared Bias, Break - Middle, 1000 draws, SNR=2, $\delta = 0.025$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.709	0.792	0.479	0.699	0.756	0.566	0.630	0.606	0.797
100	0.412	0.421	0.275	0.445	0.417	0.281	0.375	0.335	0.356
200	0.235	0.235	0.151	0.266	0.252	0.145	0.200	0.191	0.147

Table 35: Average Squared Bias, Break - Middle, 1000 draws, SNR=2, $\delta = 0.050$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.709	0.792	0.501	0.699	0.756	0.566	0.629	0.605	0.797
100	0.412	0.421	0.304	0.445	0.417	0.275	0.358	0.321	0.356
200	0.235	0.235	0.171	0.266	0.252	0.140	0.178	0.171	0.147

Table 36: Average Squared Bias, Break - Middle, 1000 draws, SNR=2, $\delta = 0.075$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.709	0.792	0.539	0.699	0.756	0.630	0.655	0.659	0.797
100	0.412	0.421	0.315	0.445	0.417	0.271	0.344	0.313	0.356
200	0.235	0.235	0.185	0.266	0.252	0.136	0.165	0.155	0.147

Table 37: Best δ for the Bias, Break - Middle, 1000 draws, SNR=2

	1 step	2 step		
T	IC_1	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.025	0.050	0.050	0.050
100	0.025	0.075	0.075	0.075
200	0.025	0.075	0.075	0.075

Table 38: Average Squared Bias, Break - Middle, 1000 draws, SNR=1, $\delta = 0.025$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	1.005	1.464	0.699	1.043	1.420	0.975	0.962	0.964	1.260
100	0.624	0.642	0.414	0.653	0.647	0.495	0.578	0.559	0.708
200	0.403	0.403	0.247	0.445	0.430	0.290	0.399	0.371	0.319

Table 39: Average Squared Bias, Break - Middle, 1000 draws, SNR=1, $\delta = 0.050$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	1.005	1.464	0.762	1.043	1.420	1.050	1.005	1.049	1.260
100	0.624	0.642	0.439	0.653	0.647	0.485	0.562	0.544	0.708
200	0.403	0.403	0.270	0.445	0.430	0.274	0.366	0.342	0.319

Table 40: Average Squared Bias, Break - Middle, 1000 draws, SNR=1, $\delta = 0.075$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	1.005	1.464	0.869	1.043	1.420	1.268	1.223	1.277	1.260
100	0.624	0.642	0.450	0.653	0.647	0.479	0.554	0.538	0.708
200	0.403	0.403	0.285	0.445	0.430	0.273	0.356	0.334	0.319

Table 41: Best δ for the Bias, Break - Middle, 1000 draws, SNR=1

	1 step	2 step		
T	IC_1	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.025	0.025	0.025	0.025
100	0.025	0.075	0.075	0.075
200	0.025	0.075	0.075	0.075

C.6 One Parameter - One Break - Middle - Rates of Detected Breaks

Table 42: Percentage Rates of detecting 0-5 breaks and Distance, Break - Middle, 1000 draws, SNR=2, $\delta = 0.025$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	23.8	22.3	4.1	5.3	5.3	25.8
	1	17.0	15.7	41.9	45.2	44.0	63.2
	2	26.7	24.0	38.0	36.2	37.0	10.6
	3	17.4	17.2	12.3	10.2	10.3	0.4
	4	6.6	7.5	2.7	2.5	2.6	0.0
	5	4.0	5.0	0.7	0.4	0.6	0.0
	6+	4.5	8.3	0.3	0.2	0.2	0.0
	Q	0.34	0.33	0.13	0.14	0.14	0.33
100	0	12.9	12.2	1.3	1.8	1.8	4.8
	1	19.5	18.2	32.1	35.7	34.7	85.8
	2	28.3	27.3	40.1	42.0	41.6	9.3
	3	18.2	18.6	19.1	15.9	16.6	0.1
	4	10.8	11.7	6.0	3.8	4.3	0.0
	5	4.0	4.2	0.8	0.4	0.6	0.0
	6+	6.3	7.8	0.6	0.4	0.4	0.0
	Q	0.22	0.21	0.07	0.08	0.08	0.11
200	0	6.4	6.1	0.0	0.1	0.1	0.0
	1	16.4	15.8	19.7	23.2	22.9	94.7
	2	28.8	28.3	36.8	39.5	39.2	5.3
	3	23.2	23.5	25.4	24.9	24.7	0.0
	4	12.3	12.0	12.2	9.7	9.9	0.0
	5	5.2	5.9	4.5	2.2	2.7	0.0
	6+	7.7	8.4	1.4	0.4	0.5	0.0
	Q	0.14	0.14	0.05	0.05	0.05	0.03

Q . . . Average Relative Distance from the True Break.

Table 43: Percentage Rates of detecting 0-5 breaks and Distance, Break - Middle, 1000 draws, SNR=2, $\delta = 0.050$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	23.8	22.3	4.3	6.7	6.4	25.8
	1	17.0	15.7	48.3	50.3	49.4	63.2
	2	26.7	24.0	34.8	32.1	32.7	10.6
	3	17.4	17.2	9.8	8.6	9.1	0.4
	4	6.6	7.5	2.0	1.7	1.8	0.0
	5	4.0	5.0	0.5	0.4	0.4	0.0
	6+	4.5	8.3	0.3	0.2	0.2	0.0
	Q	0.34	0.33	0.12	0.15	0.14	0.33
100	0	12.9	12.2	1.5	2.9	2.8	4.8
	1	19.5	18.2	40.3	42.7	42.0	85.8
	2	28.3	27.3	39.8	40.1	40.1	9.3
	3	18.2	18.6	15.0	11.5	12.3	0.1
	4	10.8	11.7	2.7	2.3	2.3	0.0
	5	4.0	4.2	0.5	0.5	0.3	0.0
	6+	6.3	7.8	0.2	0.0	0.2	0.0
	Q	0.22	0.21	0.07	0.08	0.08	0.11
200	0	6.4	6.1	0.0	0.2	0.2	0.0
	1	16.4	15.8	27.0	30.0	29.5	94.7
	2	28.8	28.3	40.9	42.3	42.2	5.3
	3	23.2	23.5	22.0	20.4	20.4	0.0
	4	12.3	12.0	8.4	5.8	6.4	0.0
	5	5.2	5.9	1.0	0.9	0.9	0.0
	6+	7.7	8.4	0.7	0.4	0.4	0.0
	Q	0.14	0.14	0.04	0.04	0.04	0.03

Q . . . Average Relative Distance from the True Break.

Table 44: Percentage Rates of detecting 0-5 breaks and Distance, Break - Middle, 1000 draws, SNR=2, $\delta = 0.075$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	23.8	22.3	4.5	7.6	7.2	25.8
	1	17.0	15.7	51.1	51.3	50.7	63.2
	2	26.7	24.0	33.3	30.7	31.4	10.6
	3	17.4	17.2	7.9	7.6	7.6	0.4
	4	6.6	7.5	2.0	1.7	2.0	0.0
	5	4.0	5.0	0.5	0.6	0.6	0.0
	6+	4.5	8.3	0.7	0.5	0.5	0.0
	Q	0.34	0.33	0.13	0.16	0.15	0.33
100	0	12.9	12.2	1.3	3.4	3.1	4.8
	1	19.5	18.2	44.6	46.4	46.0	85.8
	2	28.3	27.3	39.9	39.0	39.4	9.3
	3	18.2	18.6	11.9	9.8	10.0	0.1
	4	10.8	11.7	1.9	1.3	1.3	0.0
	5	4.0	4.2	0.4	0.1	0.2	0.0
	6+	6.3	7.8	0.0	0.0	0.0	0.0
	Q	0.22	0.21	0.06	0.08	0.08	0.11
200	0	6.4	6.1	0.0	0.0	0.0	0.0
	1	16.4	15.8	33.8	35.7	35.2	94.7
	2	28.8	28.3	41.2	42.6	42.2	5.3
	3	23.2	23.5	18.4	17.2	17.6	0.0
	4	12.3	12.0	5.8	3.9	4.3	0.0
	5	5.2	5.9	0.6	0.4	0.5	0.0
	6+	7.7	8.4	0.2	0.2	0.2	0.0
	Q	0.14	0.14	0.04	0.03	0.04	0.03

Q ... Average Relative Distance from the True Break.

Table 45: Best δ for the Q, Break - Middle , 1000 draws, SNR=2

	2 step		
T	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.050	0.025	0.025
100	0.075	0.025	0.025
200	0.075	0.075	0.075

Table 46: Percentage Rates of detecting 0-5 breaks and Distance, Break - Middle, 1000 draws, SNR=1, $\delta = 0.025$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	23.4	22.3	4.8	4.9	4.9	46.4
	1	22.4	21.1	50.6	52.3	51.8	47.8
	2	21.1	19.9	34.7	33.9	34.1	5.7
	3	14.4	13.3	7.8	7.1	7.3	0.1
	4	7.1	6.6	1.6	1.3	1.4	0.0
	5	4.7	4.6	0.3	0.3	0.3	0.0
	6+	6.9	12.2	0.2	0.2	0.2	0.0
	Q	0.34	0.34	0.15	0.15	0.15	0.53
100	0	18.4	17.5	1.8	1.9	1.9	24.0
	1	19.8	19.8	37.1	41.4	40.5	70.5
	2	23.6	23.5	38.9	39.0	39.1	5.5
	3	16.9	16.5	16.5	14.6	15.2	0.0
	4	8.9	8.0	4.9	2.7	2.8	0.0
	5	4.2	4.8	0.6	0.4	0.5	0.0
	6+	8.2	9.9	0.2	0.0	0.0	0.0
	Q	0.28	0.28	0.10	0.10	0.10	0.31
200	0	11.9	11.7	0.6	0.6	0.6	2.5
	1	14.8	14.4	21.7	25.2	24.9	93.7
	2	29.1	29.1	39.3	44.0	42.9	3.8
	3	18.5	18.2	25.2	21.6	22.4	0.0
	4	10.6	11.0	9.4	6.8	7.1	0.0
	5	6.4	6.4	2.9	1.3	1.6	0.0
	6+	8.7	9.2	0.9	0.5	0.5	0.0
	Q	0.21	0.21	0.08	0.07	0.07	0.08

Q ... Average Relative Distance from the True Break.

Table 47: Percentage Rates of detecting 0-5 breaks and Distance, Break - Middle, 1000 draws, SNR=1, $\delta = 0.050$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	23.4	22.3	6.1	7.2	7.2	46.4
	1	22.4	21.1	57.4	58.6	58.0	47.8
	2	21.1	19.9	29.7	28.2	28.6	5.7
	3	14.4	13.3	5.9	5.2	5.3	0.1
	4	7.1	6.6	0.3	0.2	0.3	0.0
	5	4.7	4.6	0.3	0.3	0.3	0.0
	6+	6.9	12.2	0.3	0.3	0.3	0.0
	Q	0.34	0.34	0.16	0.17	0.17	0.53
100	0	18.4	17.5	2.3	2.9	2.9	24.0
	1	19.8	19.8	46.6	49.5	48.9	70.5
	2	23.6	23.5	35.9	35.2	35.1	5.5
	3	16.9	16.5	12.6	10.7	11.0	0.0
	4	8.9	8.0	2.0	1.4	1.8	0.0
	5	4.2	4.8	0.6	0.3	0.3	0.0
	6+	8.2	9.9	0.0	0.0	0.0	0.0
	Q	0.28	0.28	0.10	0.10	0.10	0.31
200	0	11.9	11.7	0.7	1.0	1.0	2.5
	1	14.8	14.4	29.6	33.7	33.3	93.7
	2	29.1	29.1	43.6	44.0	43.4	3.8
	3	18.5	18.2	19.1	15.9	16.8	0.0
	4	10.6	11.0	5.6	4.4	4.4	0.0
	5	6.4	6.4	1.3	0.9	1.0	0.0
	6+	8.7	9.2	0.1	0.1	0.1	0.0
	Q	0.21	0.21	0.06	0.07	0.07	0.08

Q ... Average Relative Distance from the True Break.

Table 48: Percentage Rates of detecting 0-5 breaks and Distance, Break - Middle, 1000 draws, SNR=1, $\delta = 0.075$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	23.4	22.3	6.6	7.9	7.9	46.4
	1	22.4	21.1	60.0	60.4	60.0	47.8
	2	21.1	19.9	26.9	25.5	25.8	5.7
	3	14.4	13.3	4.7	4.4	4.5	0.1
	4	7.1	6.6	1.1	1.1	1.1	0.0
	5	4.7	4.6	0.1	0.1	0.1	0.0
	6+	6.9	12.2	0.6	0.6	0.6	0.0
	Q	0.34	0.34	0.17	0.18	0.18	0.53
100	0	18.4	17.5	2.4	3.0	3.0	24.0
	1	19.8	19.8	53.6	57.2	56.1	70.5
	2	23.6	23.5	32.7	30.6	31.0	5.5
	3	16.9	16.5	10.1	8.5	9.1	0.0
	4	8.9	8.0	1.0	0.6	0.7	0.0
	5	4.2	4.8	0.2	0.1	0.1	0.0
	6+	8.2	9.9	0.0	0.0	0.0	0.0
	Q	0.28	0.28	0.10	0.10	0.10	0.31
200	0	11.9	11.7	1.1	1.5	1.5	2.5
	1	14.8	14.4	36.0	39.9	39.1	93.7
	2	29.1	29.1	43.3	41.8	42.3	3.8
	3	18.5	18.2	16.3	13.9	14.3	0.0
	4	10.6	11.0	2.9	2.6	2.5	0.0
	5	6.4	6.4	0.4	0.3	0.3	0.0
	6+	8.7	9.2	0.0	0.0	0.0	0.0
	Q	0.21	0.21	0.06	0.07	0.07	0.08

Q . . . Average Relative Distance from the True Break.

Table 49: Best δ for the Q, Break - Middle, 1000 draws, SNR=1

	2 step		
T	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.025	0.025	0.025
100	0.075	0.025	0.025
200	0.075	0.050	0.050

C.7 One Parameter - One Break - Middle - All Break Positions

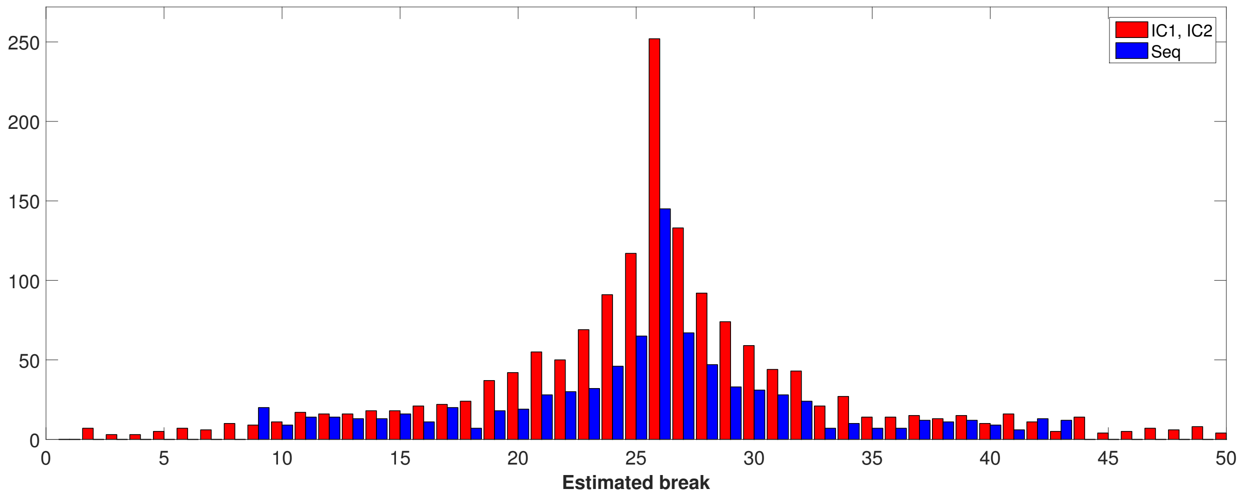

Figure 4: Positions of the Estimated Breaks, $T = 50$, SNR $= 2$, $\delta = 0.075$, Break in the Middle

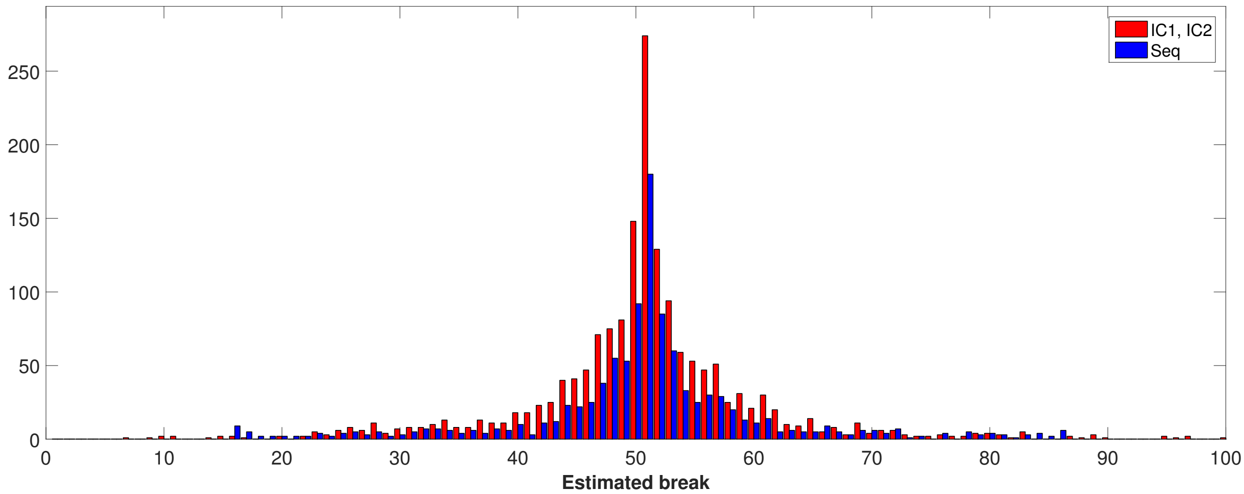

Figure 5: Positions of the Estimated Breaks, $T = 100$, SNR $= 2$, $\delta = 0.075$, Break in the Middle

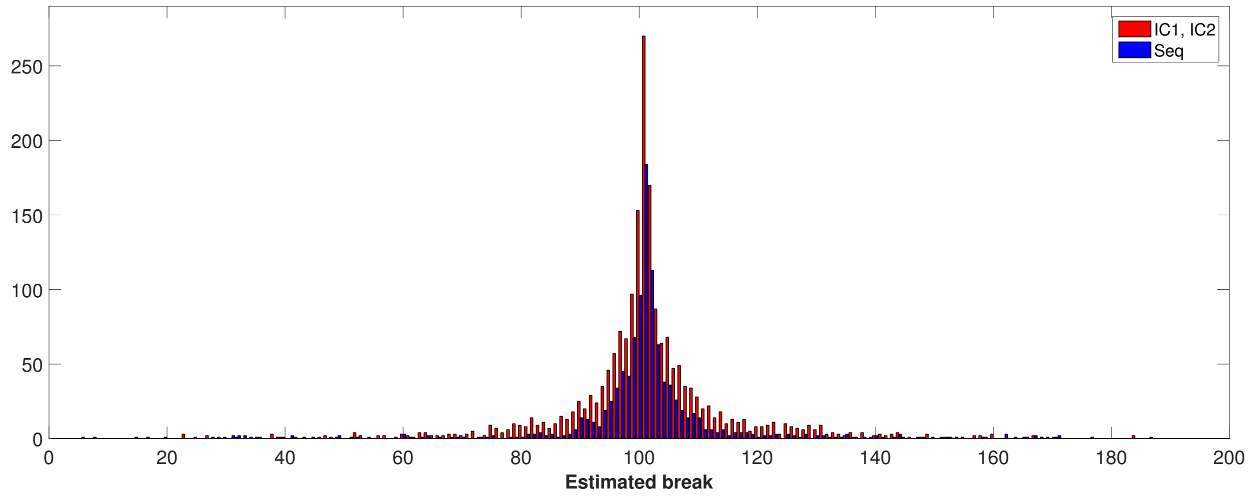

Figure 6: Positions of the Estimated Breaks, $T = 200$, SNR $= 2$, $\delta = 0.075$, Break in the Middle

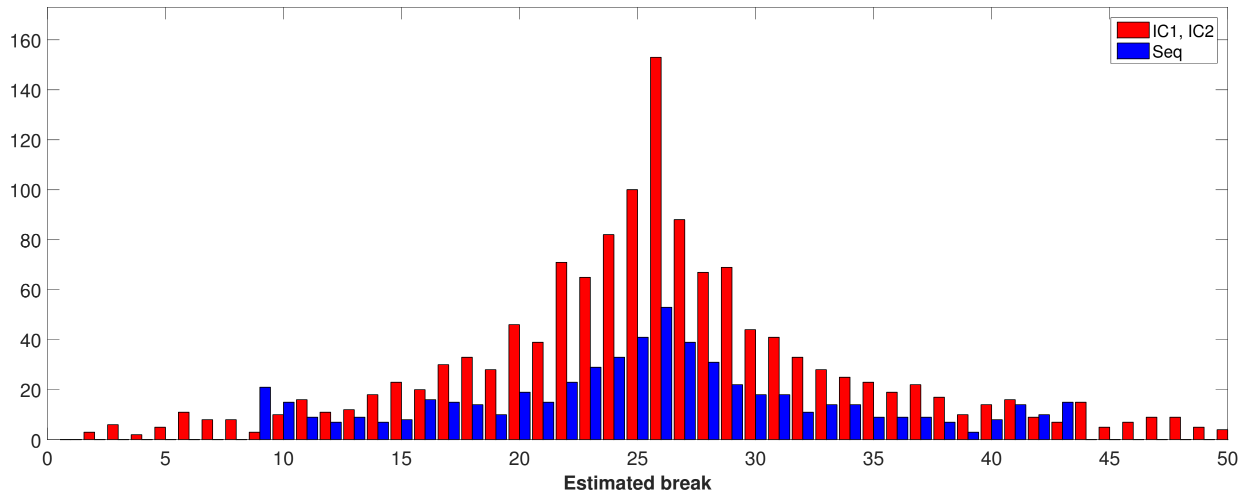

Figure 7: Positions of the Estimated Breaks, $T = 50$, SNR $= 1$, $\delta = 0.075$, Break in the Middle

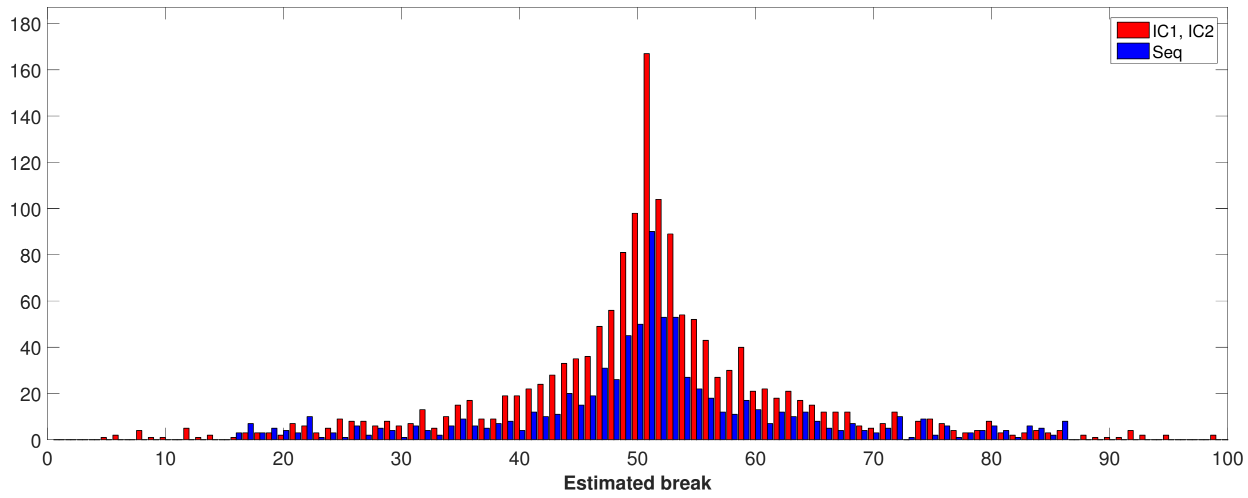

Figure 8: Positions of the Estimated Breaks, $T = 100$, SNR $= 1$, $\delta = 0.075$, Break in the Middle

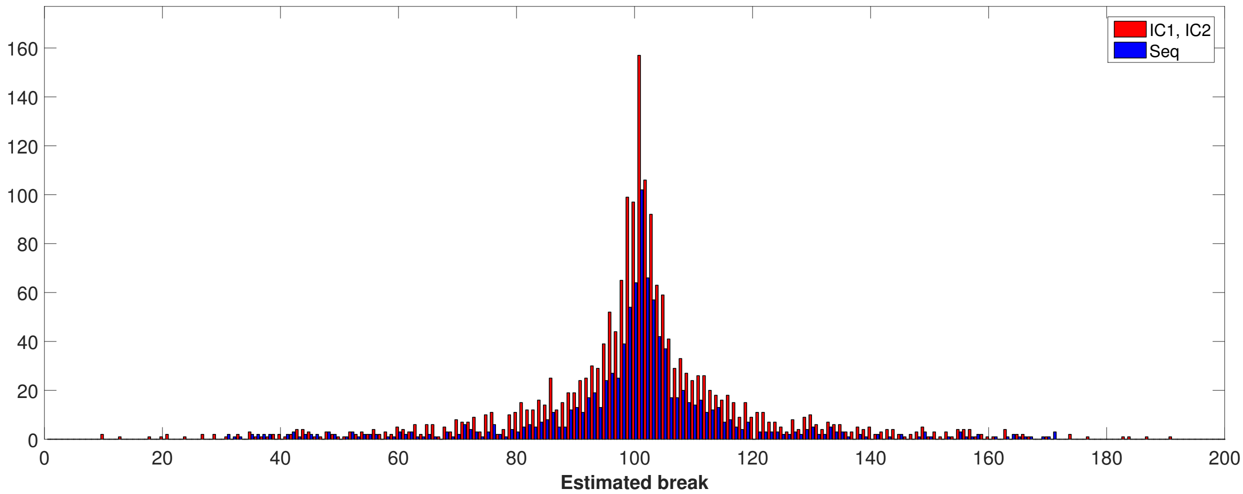

Figure 9: Positions of the Estimated Breaks, $T = 200$, SNR $= 1$, $\delta = 0.075$, Break in the Middle

C.8 One Parameter - One Break - Middle - Break Positions of One Break Found

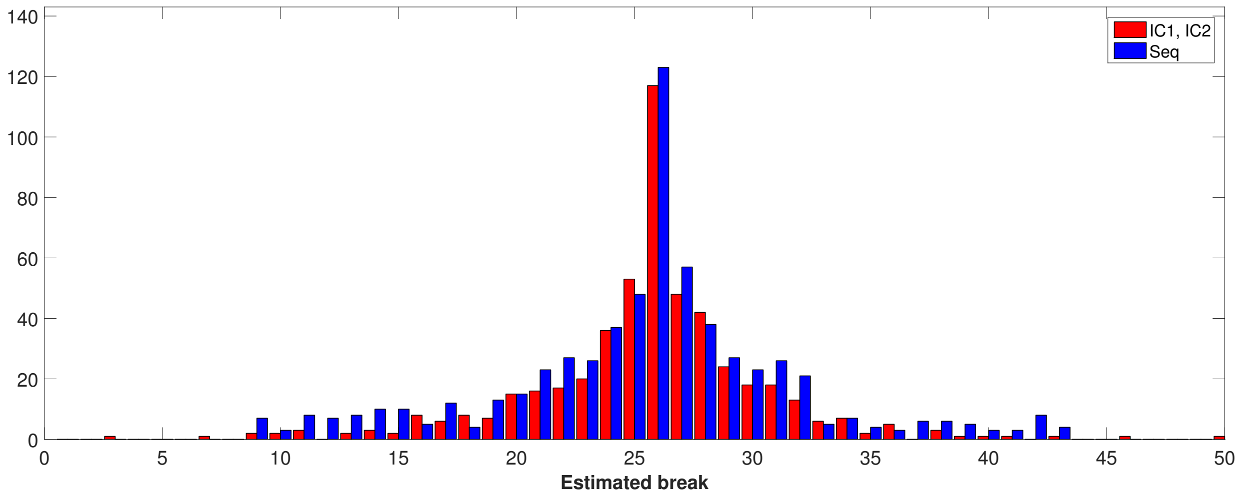

Figure 10: Positions of the Estimated Breaks given One Break Found, $T = 50$, SNR $= 2$, $\delta = 0.075$, Break in the Middle

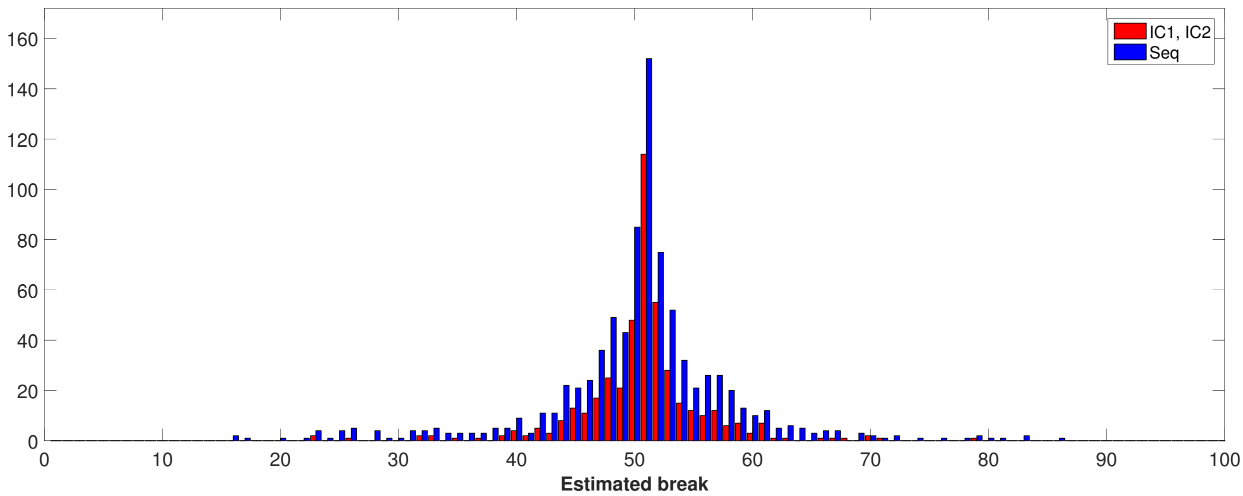

Figure 11: Positions of the Estimated Breaks given One Break Found, $T = 100$, SNR $= 2$, $\delta = 0.075$, Break in the Middle

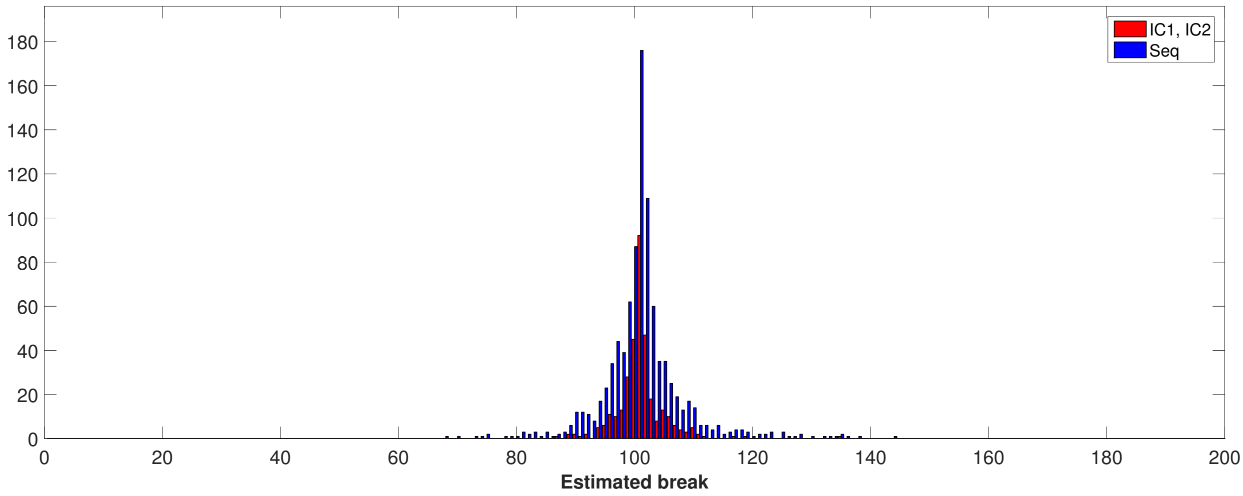

Figure 12: Positions of the Estimated Breaks given One Break Found, $T = 200$, SNR $= 2$, $\delta = 0.075$, Break in the Middle

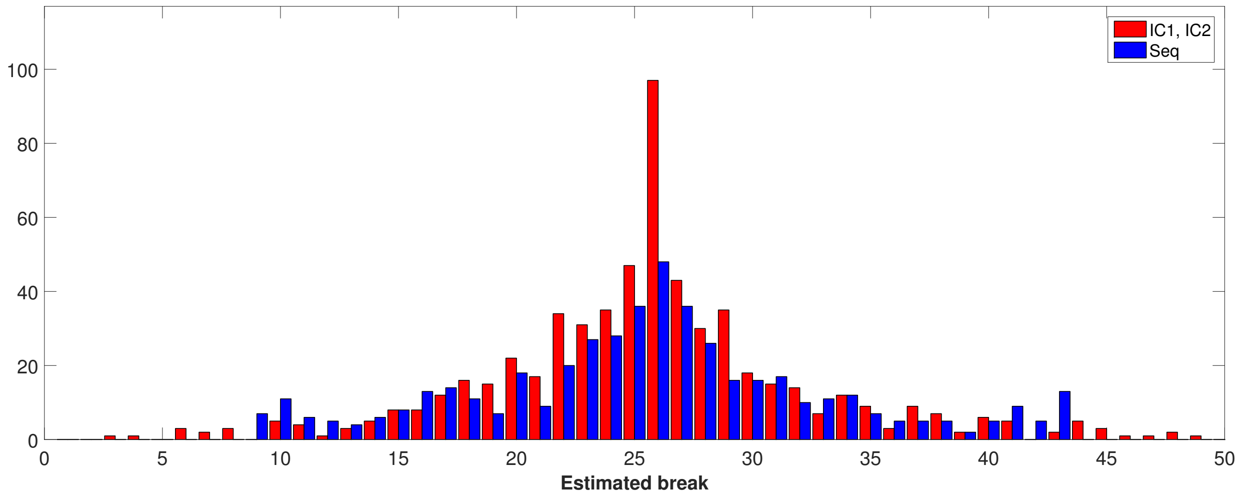

Figure 13: Positions of the Estimated Breaks given One Break Found, $T = 50$, SNR $= 1$, $\delta = 0.075$, Break in the Middle

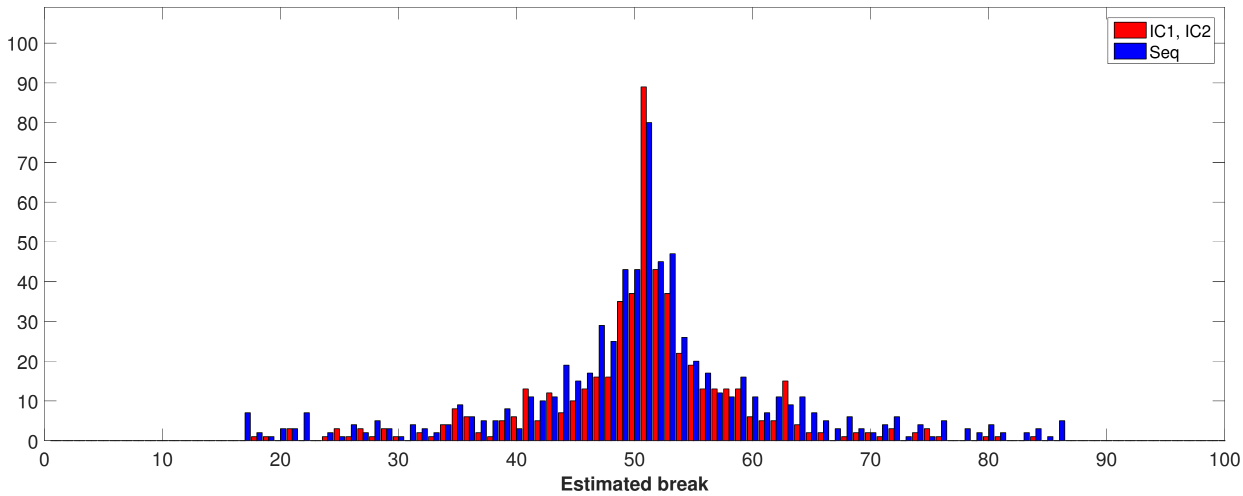

Figure 14: Positions of the Estimated Breaks given One Break Found, $T = 100$, SNR $= 1$, $\delta = 0.075$, Break in the Middle

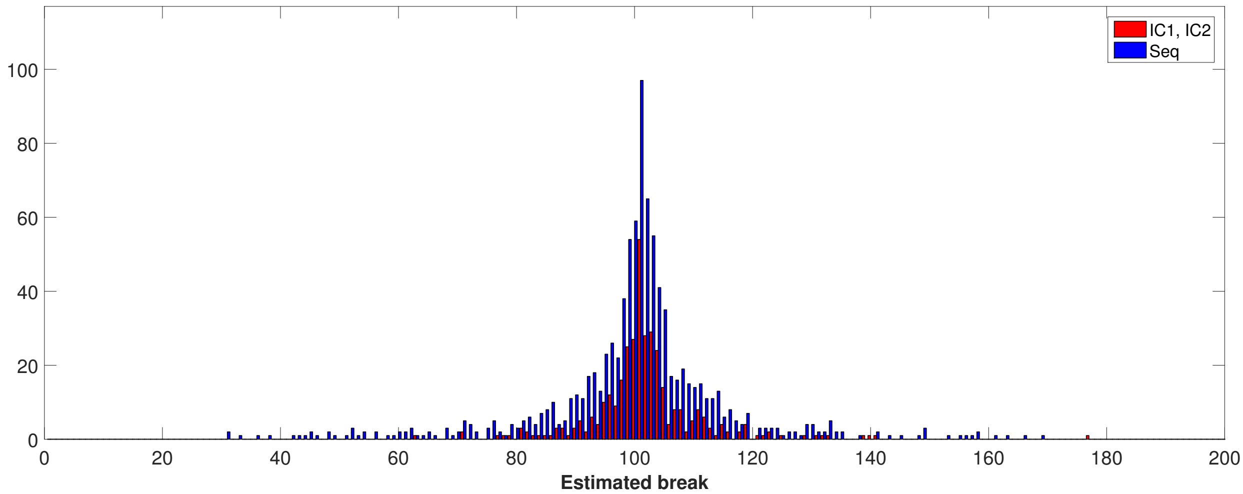

Figure 15: Positions of the Estimated Breaks given One Break Found, $T = 200$, SNR $= 1$, $\delta = 0.075$, Break in the Middle

C.9 One Parameter - One Break - End - Average Squared Bias

Table 50: Average Squared Bias, Break - End, 1000 draws, SNR=2, $\delta = 0.025$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.573	0.623	0.363	0.576	0.573	0.435	0.488	0.462	0.493
100	0.369	0.369	0.199	0.371	0.368	0.224	0.281	0.260	0.359
200	0.202	0.202	0.111	0.202	0.202	0.127	0.165	0.160	0.208

Table 51: Average Squared Bias, Break - End, 1000 draws, SNR=2, $\delta = 0.050$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.573	0.623	0.380	0.576	0.573	0.438	0.494	0.473	0.493
100	0.369	0.369	0.210	0.371	0.368	0.228	0.283	0.265	0.359
200	0.202	0.202	0.116	0.202	0.202	0.127	0.165	0.160	0.208

Table 52: Average Squared Bias, Break - End, 1000 draws, SNR=2, $\delta = 0.075$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.573	0.623	0.392	0.576	0.573	0.455	0.507	0.488	0.493
100	0.369	0.369	0.214	0.371	0.368	0.229	0.281	0.263	0.359
200	0.202	0.202	0.119	0.202	0.202	0.129	0.165	0.159	0.208

Table 53: Best δ for the Bias, Break - End, 1000 draws, SNR=2

	1 step	2 step		
T	IC_1	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.025	0.025	0.025	0.025
100	0.025	0.025	0.075	0.025
200	0.025	0.025	0.075	0.075

Table 54: Average Squared Bias, Break - End, 1000 draws, SNR=1, $\delta = 0.025$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.764	0.940	0.552	0.773	0.954	0.783	0.763	0.775	0.813
100	0.411	0.411	0.281	0.404	0.405	0.360	0.370	0.366	0.444
200	0.217	0.217	0.165	0.217	0.217	0.192	0.202	0.202	0.233

Table 55: Average Squared Bias, Break - End, 1000 draws, SNR=1, $\delta = 0.050$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.764	0.940	0.564	0.773	0.954	0.782	0.778	0.788	0.813
100	0.411	0.411	0.289	0.404	0.405	0.365	0.379	0.377	0.444
200	0.217	0.217	0.169	0.217	0.217	0.192	0.202	0.201	0.233

Table 56: Average Squared Bias, Break - End, 1000 draws, SNR=1, $\delta = 0.075$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.764	0.940	0.579	0.773	0.954	0.795	0.802	0.803	0.813
100	0.411	0.411	0.296	0.404	0.405	0.374	0.381	0.384	0.444
200	0.217	0.217	0.170	0.217	0.217	0.194	0.202	0.201	0.233

Table 57: Best δ for the Bias, Break - End, 1000 draws, SNR=1

	1 step	2 step		
T	IC_1	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.025	0.050	0.025	0.025
100	0.025	0.025	0.025	0.025
200	0.025	0.025	0.075	0.075

C.10 One Parameter - One Break - End - Rates of Detected Breaks

Table 58: Percentage Rates of detecting 0-5 breaks and Distance, Break - End, 1000 draws, SNR=2, $\delta = 0.025$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	53.4	50.7	9.6	17.2	16.2	27.5
	1	11.8	11.4	42.2	43.9	40.6	65.6
	2	16.0	14.0	31.0	27.7	29.2	6.6
	3	9.5	10.3	13.4	9.4	11.5	0.3
	4	4.2	4.6	3.1	1.4	2.0	0.0
	5	2.2	2.9	0.7	0.4	0.5	0.0
	6+	2.9	6.1	0.0	0.0	0.0	0.0
	Q	0.61	0.59	0.21	0.27	0.26	0.34
100	0	89.2	88.7	17.5	36.4	32.7	56.4
	1	1.8	1.6	29.2	29.7	29.4	41.6
	2	4.3	3.8	31.3	25.4	26.0	2.0
	3	2.9	3.2	14.8	7.0	9.5	0.0
	4	1.1	1.1	5.2	1.2	1.9	0.0
	5	0.3	0.6	1.9	0.3	0.5	0.0
	6+	0.4	1.0	0.1	0.0	0.0	0.0
	Q	0.91	0.90	0.29	0.44	0.41	0.63
200	0	100.0	100.0	36.6	61.3	58.6	83.6
	1	0.0	0.0	20.7	20.9	22.0	15.7
	2	0.0	0.0	22.9	13.7	14.4	0.7
	3	0.0	0.0	12.4	3.2	3.7	0.0
	4	0.0	0.0	5.4	0.7	1.1	0.0
	5	0.0	0.0	1.8	0.2	0.2	0.0
	6+	0.0	0.0	0.2	0.0	0.0	0.0
	Q	1.00	1.00	0.45	0.66	0.63	0.88

Q ... Average Relative Distance from the True Break.

Table 59: Percentage Rates of detecting 0-5 breaks and Distance, Break - End, 1000 draws, SNR=2, $\delta = 0.050$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	53.4	50.7	10.3	20.4	18.4	27.5
	1	11.8	11.4	49.2	46.3	45.1	65.6
	2	16.0	14.0	29.6	25.7	27.6	6.6
	3	9.5	10.3	8.6	6.2	7.1	0.3
	4	4.2	4.6	1.8	1.0	1.3	0.0
	5	2.2	2.9	0.5	0.4	0.5	0.0
	6+	2.9	6.1	0.0	0.0	0.0	0.0
	Q	0.61	0.59	0.21	0.29	0.28	0.34
100	0	89.2	88.7	17.7	39.2	34.4	56.4
	1	1.8	1.6	34.7	31.2	32.3	41.6
	2	4.3	3.8	30.2	22.8	23.7	2.0
	3	2.9	3.2	12.3	5.6	7.5	0.0
	4	1.1	1.1	4.2	1.1	1.8	0.0
	5	0.3	0.6	0.7	0.0	0.2	0.0
	6+	0.4	1.0	0.2	0.1	0.1	0.0
	Q	0.91	0.90	0.29	0.46	0.42	0.63
200	0	100.0	100.0	35.4	60.8	57.9	83.6
	1	0.0	0.0	23.2	23.3	24.0	15.7
	2	0.0	0.0	25.1	13.2	14.3	0.7
	3	0.0	0.0	10.6	2.2	2.9	0.0
	4	0.0	0.0	4.5	0.4	0.7	0.0
	5	0.0	0.0	1.1	0.1	0.2	0.0
	6+	0.0	0.0	0.1	0.0	0.0	0.0
	Q	1.00	1.00	0.44	0.65	0.62	0.88

Q . . . Average Relative Distance from the True Break.

Table 60: Percentage Rates of detecting 0-5 breaks and Distance, Break - End, 1000 draws, SNR=2, $\delta = 0.075$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	53.4	50.7	10.8	21.4	19.0	27.5
	1	11.8	11.4	50.2	46.0	45.6	65.6
	2	16.0	14.0	28.4	25.1	26.8	6.6
	3	9.5	10.3	8.1	5.5	6.3	0.3
	4	4.2	4.6	1.9	1.4	1.7	0.0
	5	2.2	2.9	0.3	0.4	0.3	0.0
	6+	2.9	6.1	0.3	0.2	0.3	0.0
	Q	0.61	0.59	0.22	0.30	0.28	0.34
100	0	89.2	88.7	16.6	39.2	33.6	56.4
	1	1.8	1.6	39.8	31.5	34.3	41.6
	2	4.3	3.8	29.5	23.1	23.2	2.0
	3	2.9	3.2	11.2	5.1	7.3	0.0
	4	1.1	1.1	2.4	0.9	1.3	0.0
	5	0.3	0.6	0.5	0.2	0.3	0.0
	6+	0.4	1.0	0.0	0.0	0.0	0.0
	Q	0.91	0.90	0.28	0.46	0.41	0.63
200	0	100.0	100.0	33.5	60.6	57.2	83.6
	1	0.0	0.0	27.2	24.8	25.9	15.7
	2	0.0	0.0	24.9	11.4	12.6	0.7
	3	0.0	0.0	10.3	2.7	3.5	0.0
	4	0.0	0.0	3.3	0.4	0.7	0.0
	5	0.0	0.0	0.7	0.1	0.1	0.0
	6+	0.0	0.0	0.1	0.0	0.0	0.0
	Q	1.00	1.00	0.42	0.64	0.61	0.88

Q ... Average Relative Distance from the True Break.

Table 61: Best δ for the Q, Break - End, 1000 draws, SNR=2

	2 step		
T	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.025	0.025	0.025
100	0.075	0.025	0.025
200	0.075	0.075	0.075

Table 62: Percentage Rates of detecting 0-5 breaks and Distance, Break - End, 1000 draws, SNR=1, $\delta = 0.025$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	52.4	50.0	14.3	17.2	17.2	45.6
	1	11.1	11.1	44.5	46.4	45.4	47.0
	2	14.5	14.0	28.6	27.2	27.4	6.9
	3	10.5	9.6	9.5	7.4	7.8	0.4
	4	5.8	5.7	2.5	1.6	2.0	0.1
	5	2.0	2.6	0.3	0.2	0.1	0.0
	6+	3.7	7.0	0.3	0.0	0.1	0.0
	Q	0.62	0.60	0.29	0.31	0.31	0.53
100	0	86.7	85.4	23.6	33.1	32.8	65.8
	1	1.8	2.3	32.5	36.8	35.2	31.9
	2	4.2	4.4	26.8	24.2	24.5	2.2
	3	3.7	3.7	13.4	5.3	6.8	0.1
	4	1.2	1.9	2.9	0.5	0.6	0.0
	5	0.8	0.6	0.8	0.1	0.1	0.0
	6+	1.6	1.7	0.0	0.0	0.0	0.0
	Q	0.90	0.89	0.38	0.44	0.44	0.72
200	0	99.4	99.4	46.8	59.0	58.7	85.3
	1	0.2	0.2	21.5	24.4	24.3	14.1
	2	0.2	0.1	17.7	13.1	13.4	0.6
	3	0.0	0.1	9.9	3.0	3.1	0.0
	4	0.2	0.2	3.1	0.4	0.4	0.0
	5	0.0	0.0	0.8	0.1	0.1	0.0
	6+	0.0	0.0	0.2	0.0	0.0	0.0
	Q	1.00	1.00	0.58	0.67	0.67	0.90

Q ... Average Relative Distance from the True Break.

Table 63: Percentage Rates of detecting 0-5 breaks and Distance, Break - End, 1000 draws, SNR=1, $\delta = 0.050$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	52.4	50.0	16.0	20.8	20.7	45.6
	1	11.1	11.1	49.7	49.1	48.0	47.0
	2	14.5	14.0	24.9	22.8	23.9	6.9
	3	10.5	9.6	7.4	5.8	5.6	0.4
	4	5.8	5.7	1.9	1.4	1.7	0.1
	5	2.0	2.6	0.1	0.1	0.1	0.0
	6+	3.7	7.0	0.0	0.0	0.0	0.0
	Q	0.62	0.60	0.30	0.34	0.34	0.53
100	0	86.7	85.4	24.3	37.5	36.7	65.8
	1	1.8	2.3	38.2	36.2	34.9	31.9
	2	4.2	4.4	26.0	22.1	23.4	2.2
	3	3.7	3.7	9.2	3.6	4.4	0.1
	4	1.2	1.9	1.6	0.4	0.4	0.0
	5	0.8	0.6	0.7	0.2	0.2	0.0
	6+	1.6	1.7	0.0	0.0	0.0	0.0
	Q	0.90	0.89	0.38	0.48	0.47	0.72
200	0	99.4	99.4	47.2	63.0	62.8	85.3
	1	0.2	0.2	25.4	23.4	22.8	14.1
	2	0.2	0.1	18.5	12.4	12.8	0.6
	3	0.0	0.1	7.4	1.2	1.6	0.0
	4	0.2	0.2	1.3	0.0	0.0	0.0
	5	0.0	0.0	0.2	0.0	0.0	0.0
	6+	0.0	0.0	0.0	0.0	0.0	0.0
	Q	1.00	1.00	0.57	0.69	0.69	0.90

Q . . . Average Relative Distance from the True Break.

Table 64: Percentage Rates of detecting 0-5 breaks and Distance, Break - End, 1000 draws, SNR=1, $\delta = 0.075$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	52.4	50.0	17.1	21.5	21.4	45.6
	1	11.1	11.1	52.1	51.4	50.2	47.0
	2	14.5	14.0	23.2	20.9	22.0	6.9
	3	10.5	9.6	5.5	4.5	4.6	0.4
	4	5.8	5.7	1.6	1.3	1.3	0.1
	5	2.0	2.6	0.4	0.3	0.4	0.0
	6+	3.7	7.0	0.1	0.1	0.1	0.0
	Q	0.62	0.60	0.31	0.34	0.34	0.53
100	0	86.7	85.4	25.8	39.8	38.9	65.8
	1	1.8	2.3	41.8	37.3	36.9	31.9
	2	4.2	4.4	24.7	19.4	20.2	2.2
	3	3.7	3.7	6.0	3.1	3.4	0.1
	4	1.2	1.9	1.3	0.2	0.3	0.0
	5	0.8	0.6	0.4	0.2	0.3	0.0
	6+	1.6	1.7	0.0	0.0	0.0	0.0
	Q	0.90	0.89	0.39	0.50	0.49	0.72
200	0	99.4	99.4	46.8	64.8	64.4	85.3
	1	0.2	0.2	27.9	23.0	22.9	14.1
	2	0.2	0.1	18.7	11.0	11.4	0.6
	3	0.0	0.1	6.0	1.0	1.1	0.0
	4	0.2	0.2	0.4	0.2	0.2	0.0
	5	0.0	0.0	0.1	0.0	0.0	0.0
	6+	0.0	0.0	0.1	0.0	0.0	0.0
	Q	1.00	1.00	0.57	0.71	0.71	0.90

Q ... Average Relative Distance from the True Break.

Table 65: Best δ for the Q, Break - End, 1000 draws, SNR=1

	2 step		
T	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.025	0.025	0.025
100	0.050	0.025	0.025
200	0.075	0.025	0.025

C.11 One Parameter - One Break - End - All Break Positions

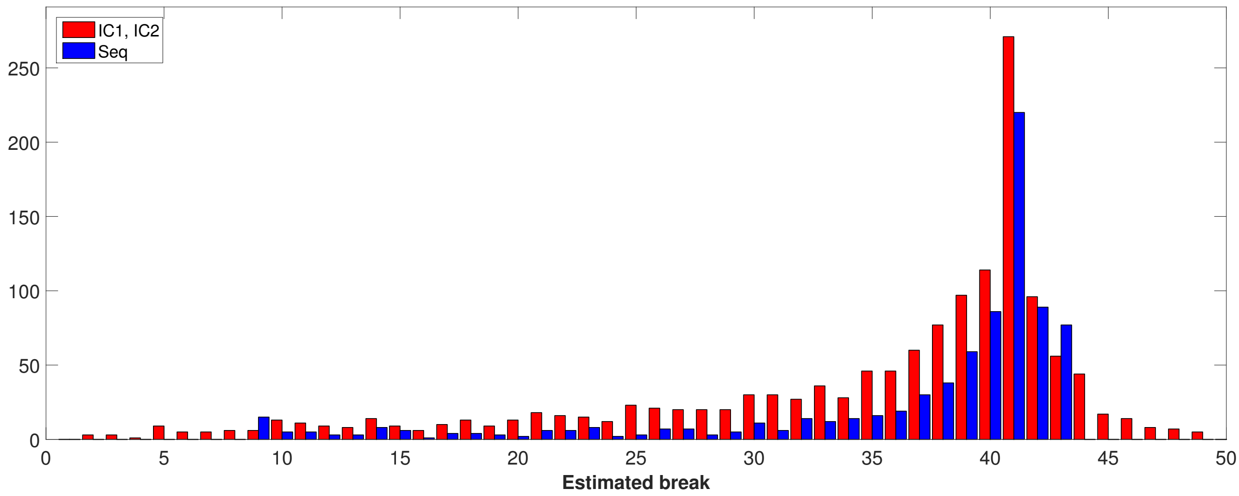

Figure 16: Positions of the Estimated Breaks, $T = 50$, SNR $= 2$, $\delta = 0.075$, Break at the End

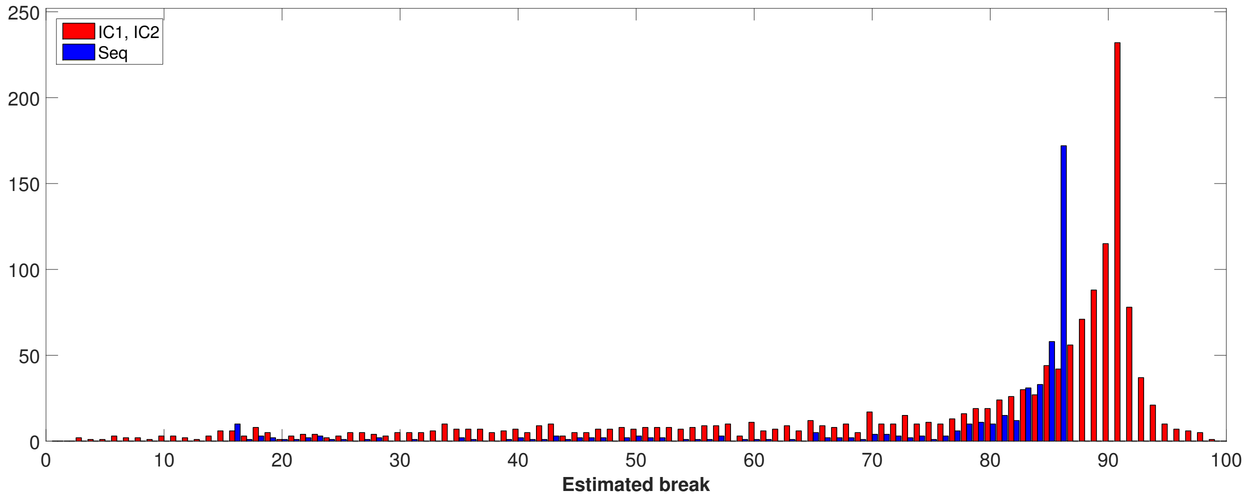

Figure 17: Positions of the Estimated Breaks, $T = 100$, SNR $= 2$, $\delta = 0.075$, Break at the End

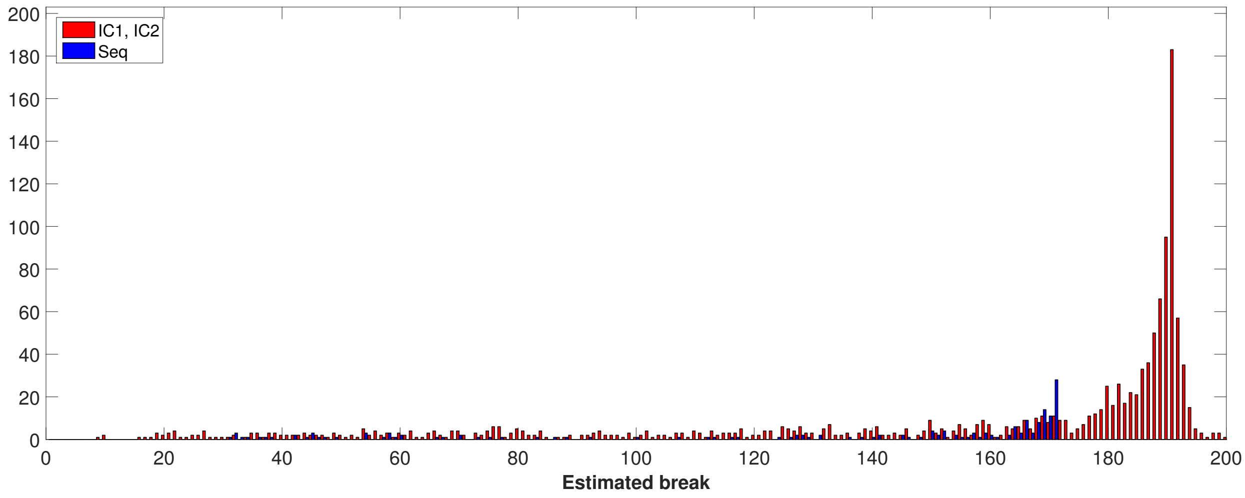

Figure 18: Positions of the Estimated Breaks, $T = 200$, SNR $= 2$, $\delta = 0.075$, Break at the End

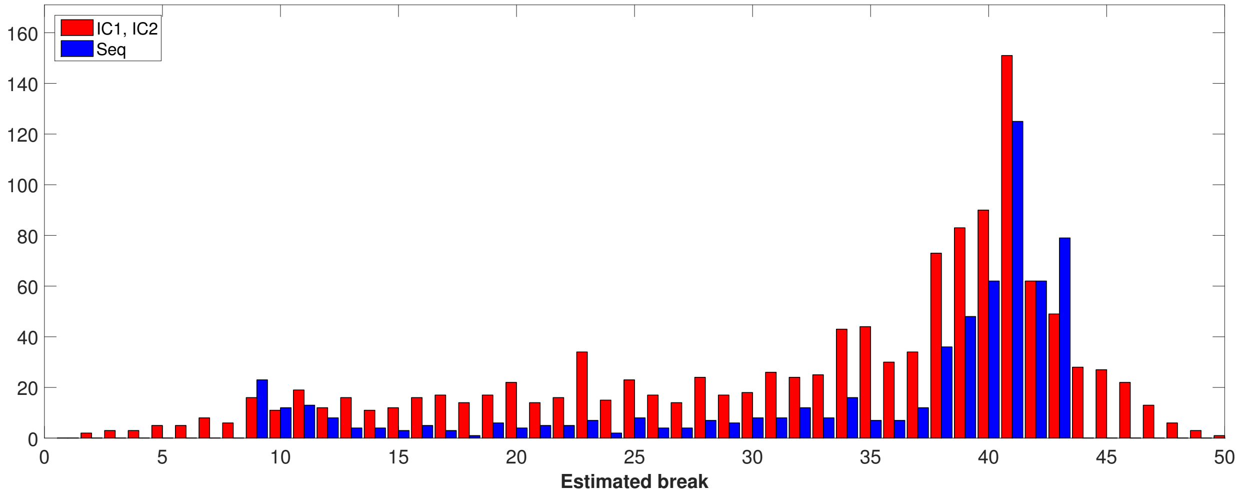

Figure 19: Positions of the Estimated Breaks, $T = 50$, SNR $= 1$, $\delta = 0.075$, Break at the End

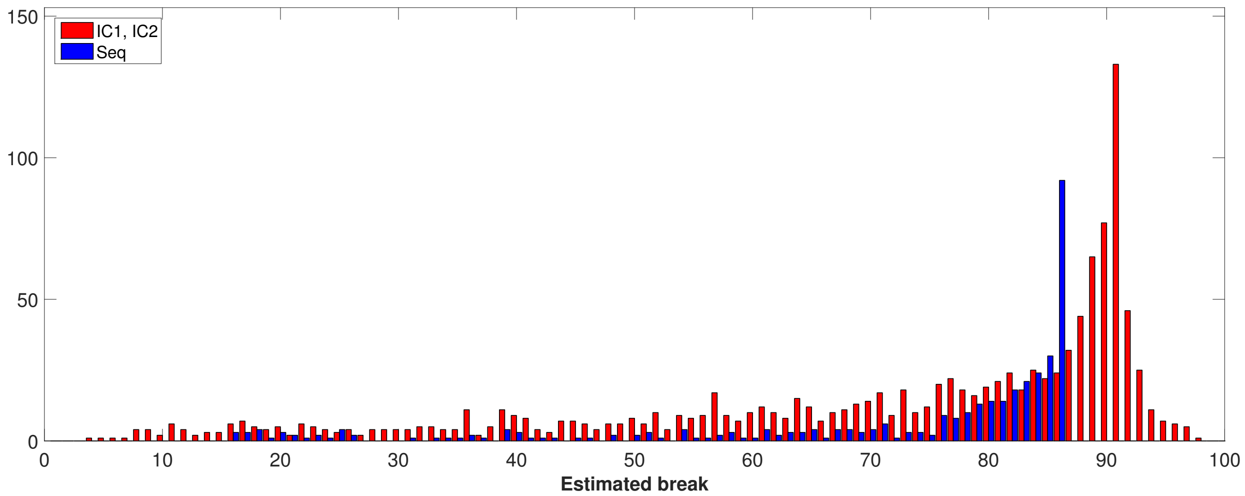

Figure 20: Positions of the Estimated Breaks, $T = 100$, SNR $= 1$, $\delta = 0.075$, Break at the End

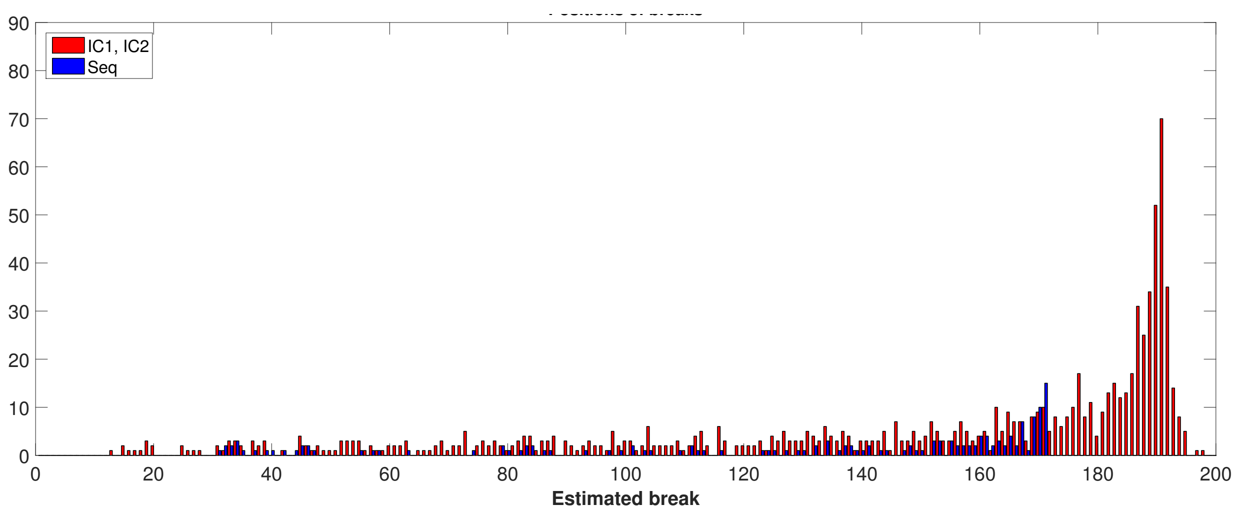

Figure 21: Positions of the Estimated Breaks, $T = 200$, SNR $= 1$, $\delta = 0.075$, Break at the End

C.12 One Parameter - One Break - End - Break Positions One Break Found

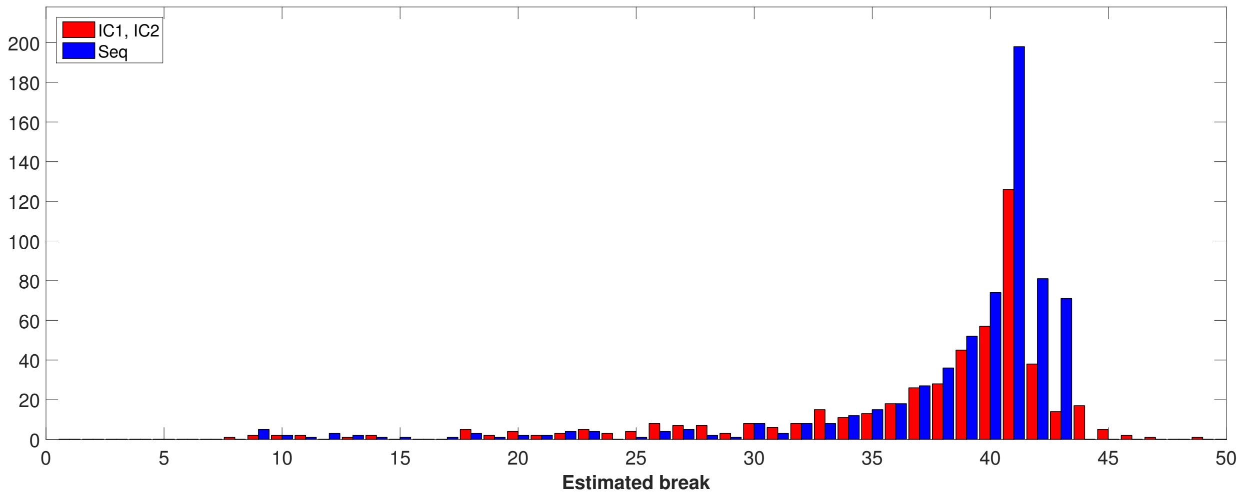

Figure 22: Positions of the Estimated Breaks given One Break Found, $T = 50$, SNR $= 2$, $\delta = 0.075$, Break at the End

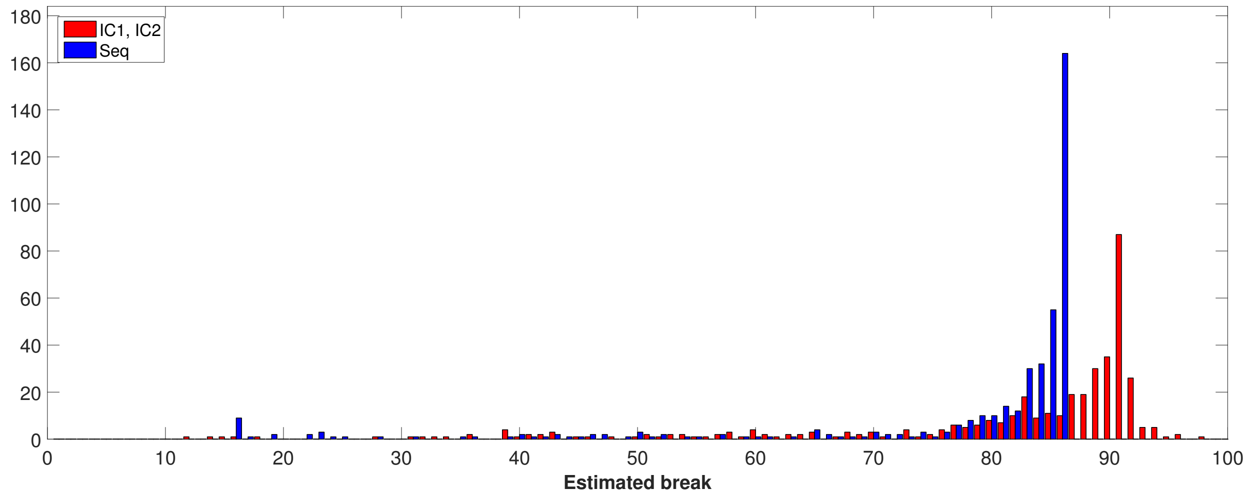

Figure 23: Positions of the Estimated Breaks given One Break Found, $T = 100$, SNR $= 2$, $\delta = 0.075$, Break at the End

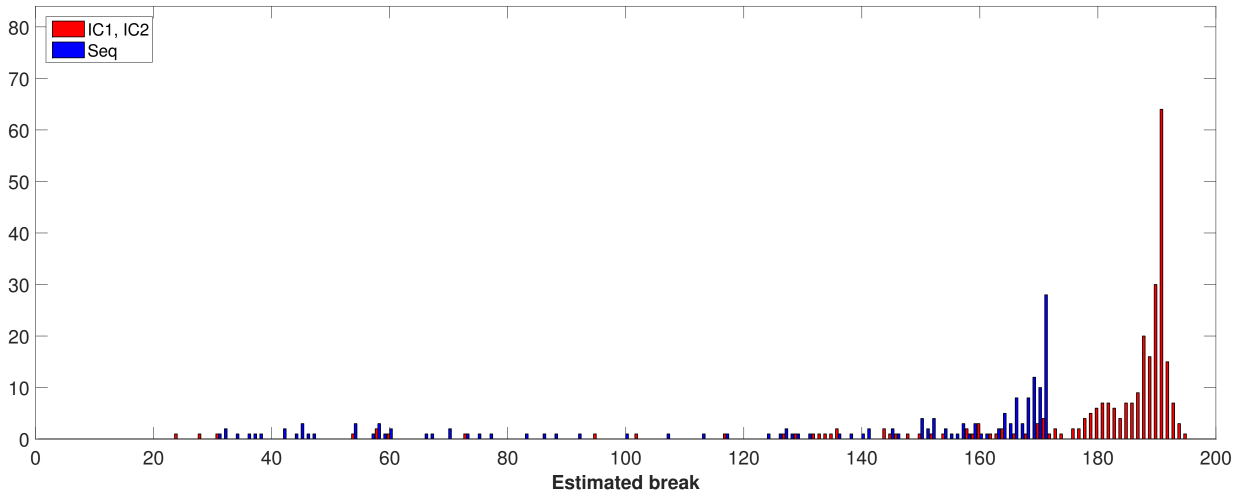

Figure 24: Positions of the Estimated Breaks given One Break Found, $T = 200$, SNR $= 2$, $\delta = 0.075$, Break at the End

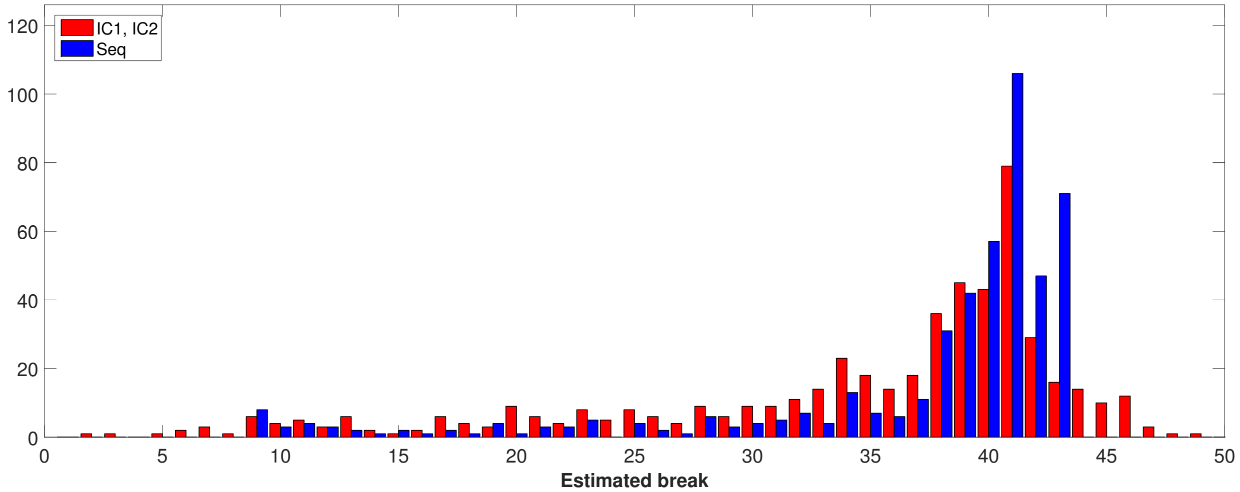

Figure 25: Positions of the Estimated Breaks given One Break Found, $T = 50$, SNR $= 1$, $\delta = 0.075$, Break at the End

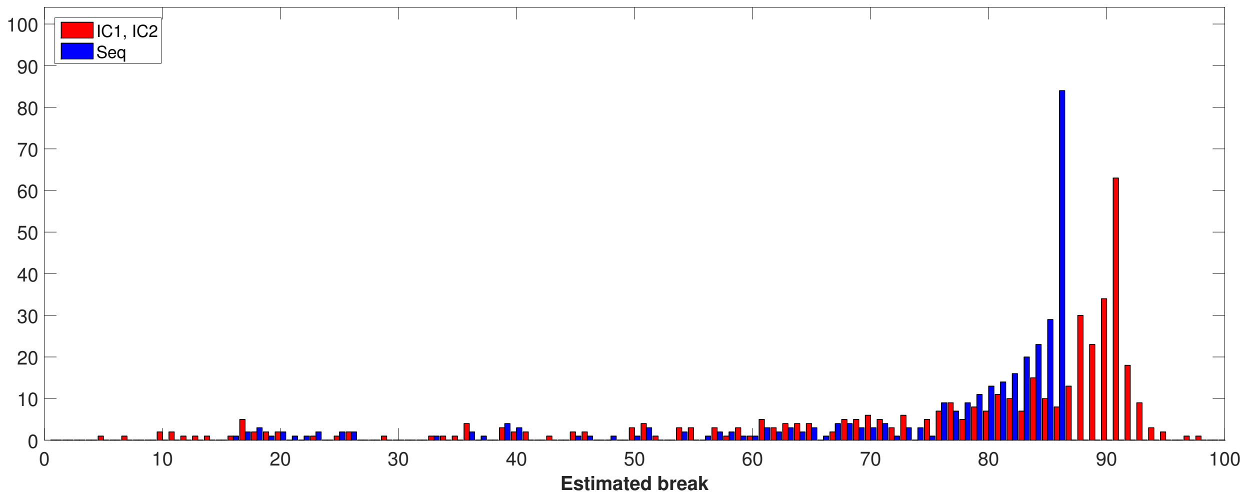

Figure 26: Positions of the Estimated Breaks given One Break Found, $T = 100$, SNR $= 1$, $\delta = 0.075$, Break at the End

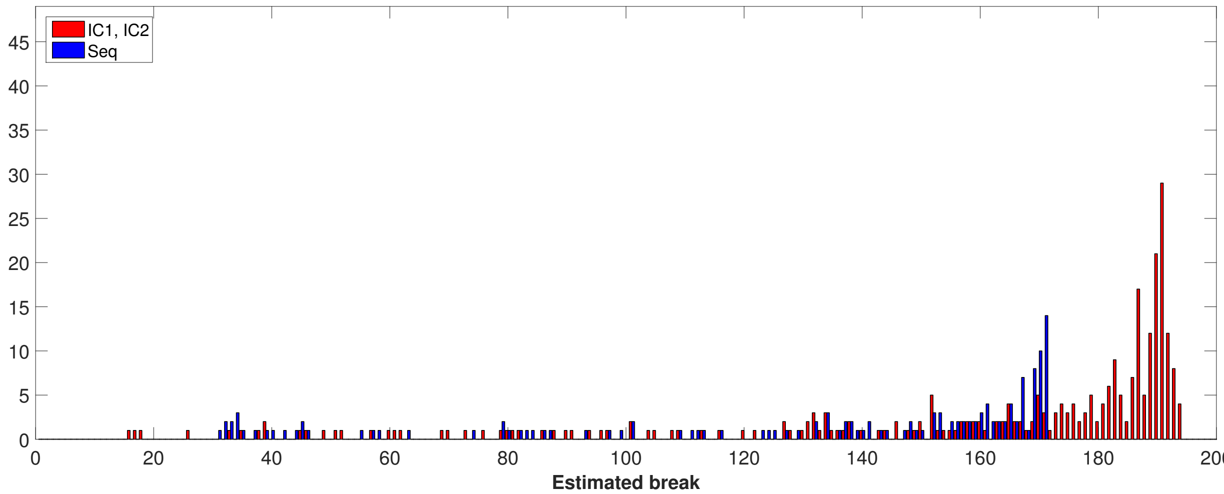

Figure 27: Positions of the Estimated Breaks given One Break Found, $T = 200$, SNR $= 1$, $\delta = 0.075$, Break at the End

C.13 Two Parameters - One Break - Middle - Average Squared Bias

Table 66: Average Squared Bias, Break - Middle, 1000 draws, SNR=2, $\delta = 0.025$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.571	0.875	0.432	0.581	0.574	0.675	0.546	0.547	0.763
100	0.424	0.411	0.233	0.379	0.391	0.328	0.326	0.328	0.376
200	0.378	0.348	0.131	0.332	0.320	0.153	0.213	0.236	0.161

Table 67: Average Squared Bias, Break - Middle, 1000 draws, SNR=2, $\delta = 0.050$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.571	0.875	0.431	0.581	0.574	0.662	0.540	0.540	0.763
100	0.424	0.411	0.227	0.379	0.391	0.321	0.318	0.320	0.376
200	0.378	0.348	0.125	0.332	0.320	0.152	0.201	0.226	0.161

Table 68: Average Squared Bias, Break - Middle, 1000 draws, SNR=2, $\delta = 0.075$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.571	0.875	0.441	0.581	0.574	0.681	0.547	0.549	0.763
100	0.424	0.411	0.226	0.379	0.391	0.312	0.316	0.317	0.376
200	0.378	0.348	0.124	0.332	0.320	0.152	0.196	0.221	0.161

Table 69: Best δ for the Bias, Break - Middle, 1000 draws, SNR=2

	1 step	2 step		
T	IC_1	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.050	0.050	0.050	0.050
100	0.075	0.075	0.075	0.075
200	0.075	0.050	0.075	0.075

Table 70: Average Squared Bias, Break - Middle, 1000 draws, SNR=1, $\delta = 0.025$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.824	2.501	0.750	0.958	1.710	1.370	1.050	1.052	1.299
100	0.540	0.536	0.408	0.553	0.544	0.660	0.523	0.517	0.691
200	0.429	0.406	0.230	0.412	0.411	0.325	0.335	0.337	0.342

Table 71: Average Squared Bias, Break - Middle, 1000 draws, SNR=1, $\delta = 0.050$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.824	2.501	0.742	0.958	1.710	1.349	1.064	1.064	1.299
100	0.540	0.536	0.399	0.553	0.544	0.640	0.508	0.508	0.691
200	0.429	0.406	0.221	0.412	0.411	0.314	0.331	0.335	0.342

Table 72: Average Squared Bias, Break - Middle, 1000 draws, SNR=1, $\delta = 0.075$

	1 step			2 step					BP
T	$EBIC$	QS	IC_1	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0.824	2.501	0.945	0.958	1.710	1.651	1.366	1.386	1.299
100	0.540	0.536	0.399	0.553	0.544	0.629	0.510	0.507	0.691
200	0.429	0.406	0.220	0.412	0.411	0.312	0.327	0.329	0.342

Table 73: Best δ for the Bias, Break - Middle, 1000 draws, SNR=1

	1 step	2 step		
T	IC_1	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.050	0.050	0.025	0.025
100	0.075	0.075	0.050	0.075
200	0.075	0.075	0.075	0.075

C.14 Two Parameters - One Break - Middle - Rates of Detected Breaks

Table 74: Percentage Rates of detecting 0-5 breaks and Distance, Break - Middle, 1000 draws, SNR=2, $\delta = 0.025$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	41.5	35.3	2.1	2.4	2.4	35.3
	1	22.6	21.0	7.1	13.2	13.0	46.3
	2	16.7	18.6	28.3	35.3	35.5	16.5
	3	8.3	7.9	29.7	28.9	29.0	1.8
	4	4.8	4.3	18.1	13.4	13.2	0.1
	5	2.7	2.0	7.5	4.4	4.6	0.0
	6+	3.4	10.9	7.2	2.4	2.3	0.0
	Q	0.48	0.43	0.17	0.17	0.17	0.44
	P	0.47	0.51	0.57	0.63	0.63	0.32
100	0	40.7	41.0	1.1	1.1	1.1	13.8
	1	24.8	23.8	6.6	11.5	11.4	72.9
	2	18.0	19.8	22.5	31.6	31.6	12.6
	3	7.6	8.2	25.2	29.8	30.8	0.7
	4	4.2	3.4	19.1	17.4	17.5	0.0
	5	2.6	1.1	13.4	6.4	5.6	0.0
	6+	2.1	2.7	12.1	2.2	2.0	0.0
	Q	0.45	0.45	0.14	0.13	0.13	0.21
	P	0.52	0.54	0.63	0.72	0.72	0.43
200	0	46.9	41.7	0.1	0.2	0.2	0.9
	1	19.5	24.0	5.2	8.9	8.8	89.9
	2	18.2	21.4	21.3	31.8	33.6	8.8
	3	8.9	8.7	26.0	31.3	32.2	0.4
	4	3.5	2.9	21.0	16.4	15.9	0.0
	5	1.6	0.6	13.8	7.6	6.8	0.0
	6+	1.4	0.7	12.6	3.8	2.5	0.0
	Q	0.49	0.44	0.12	0.10	0.10	0.05
	P	0.48	0.55	0.69	0.77	0.79	0.50

Q . . . Average Relative Distance from the True Break.

Table 75: Percentage Rates of detecting 0-5 breaks and Distance, Break - Middle, 1000 draws, SNR=2, $\delta = 0.050$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	41.5	35.3	2.8	3.2	3.2	35.3
	1	22.6	21.0	7.0	13.1	13.0	46.3
	2	16.7	18.6	31.3	38.7	38.8	16.5
	3	8.3	7.9	31.4	29.3	29.1	1.8
	4	4.8	4.3	16.3	10.5	10.6	0.1
	5	2.7	2.0	6.9	3.8	3.9	0.0
	6+	3.4	10.9	4.3	1.4	1.4	0.0
	Q	0.48	0.43	0.17	0.17	0.17	0.44
	P	0.47	0.51	0.57	0.63	0.63	0.32
100	0	40.7	41.0	1.2	1.3	1.3	13.8
	1	24.8	23.8	7.2	12.2	12.1	72.9
	2	18.0	19.8	27.7	37.0	37.4	12.6
	3	7.6	8.2	24.4	28.5	29.1	0.7
	4	4.2	3.4	20.0	13.8	14.4	0.0
	5	2.6	1.1	11.2	5.0	4.1	0.0
	6+	2.1	2.7	8.3	2.2	1.6	0.0
	Q	0.45	0.45	0.14	0.13	0.13	0.21
	P	0.52	0.54	0.63	0.72	0.73	0.43
200	0	46.9	41.7	0.2	0.4	0.4	0.9
	1	19.5	24.0	8.3	12.1	12.6	89.9
	2	18.2	21.4	21.1	32.2	33.3	8.8
	3	8.9	8.7	28.0	29.9	30.9	0.4
	4	3.5	2.9	20.5	16.7	16.4	0.0
	5	1.6	0.6	13.1	6.0	4.5	0.0
	6+	1.4	0.7	8.8	2.7	1.9	0.0
	Q	0.49	0.44	0.11	0.10	0.09	0.05
	P	0.48	0.55	0.71	0.79	0.81	0.50

Q . . . Average Relative Distance from the True Break.

Table 76: Percentage Rates of detecting 0-5 breaks and Distance, Break - Middle, 1000 draws, SNR=2, $\delta = 0.075$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	41.5	35.3	3.0	3.4	3.4	35.3
	1	22.6	21.0	8.2	14.5	14.4	46.3
	2	16.7	18.6	33.2	37.5	37.6	16.5
	3	8.3	7.9	31.3	29.8	29.9	1.8
	4	4.8	4.3	15.3	10.7	10.5	0.1
	5	2.7	2.0	5.2	2.4	2.5	0.0
	6+	3.4	10.9	3.8	1.7	1.7	0.0
	Q	0.48	0.43	0.17	0.17	0.17	0.44
	P	0.47	0.51	0.57	0.63	0.63	0.32
100	0	40.7	41.0	1.2	1.3	1.3	13.8
	1	24.8	23.8	7.4	13.6	13.8	72.9
	2	18.0	19.8	27.4	36.1	36.1	12.6
	3	7.6	8.2	30.0	30.2	31.8	0.7
	4	4.2	3.4	18.5	13.6	12.8	0.0
	5	2.6	1.1	9.0	4.2	3.4	0.0
	6+	2.1	2.7	6.5	1.0	0.8	0.0
	Q	0.45	0.45	0.13	0.12	0.12	0.21
	P	0.52	0.54	0.63	0.73	0.73	0.43
200	0	46.9	41.7	0.2	0.3	0.3	0.9
	1	19.5	24.0	9.3	12.7	13.0	89.9
	2	18.2	21.4	23.5	35.5	37.2	8.8
	3	8.9	8.7	27.2	28.0	29.7	0.4
	4	3.5	2.9	20.8	16.8	15.0	0.0
	5	1.6	0.6	12.1	5.3	4.1	0.0
	6+	1.4	0.7	6.9	1.4	0.7	0.0
	Q	0.49	0.44	0.10	0.09	0.09	0.05
	P	0.48	0.55	0.70	0.79	0.81	0.50

Q . . . Average Relative Distance from the True Break.

Table 77: Best δ for the Q, Break - Middle, 1000 draws, SNR=2

	2 step		
T	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.050	0.025	0.025
100	0.075	0.025	0.025
200	0.075	0.025	0.025

Table 78: Best δ for the P, Break - Middle, 1000 draws, SNR=2

T	2 step IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.050	0.025	0.025
100	0.075	0.025	0.025
200	0.075	0.025	0.025

Table 79: Percentage Rates of detecting 0-5 breaks and Distance, Break - Middle, 1000 draws, SNR=1, $\delta = 0.025$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	31.7	26.4	1.5	1.5	1.5	50.6
	1	23.4	19.7	5.8	8.4	8.3	36.2
	2	18.8	17.0	28.7	32.7	32.2	11.4
	3	8.6	7.6	32.1	32.1	32.6	1.7
	4	6.5	4.4	17.7	15.6	15.6	0.1
	5	3.0	3.0	7.6	6.2	6.2	0.0
	6+	8.0	21.9	6.6	3.5	3.6	0.0
	Q	0.41	0.38	0.18	0.18	0.18	0.59
	P	0.48	0.48	0.53	0.55	0.55	0.25
100	0	45.8	42.4	1.2	1.2	1.2	44.3
	1	21.5	22.5	7.7	10.3	10.4	47.0
	2	14.3	16.3	25.6	32.9	32.9	8.0
	3	7.8	8.7	27.3	30.3	30.3	0.7
	4	4.1	4.2	20.2	16.0	16.0	0.0
	5	2.2	1.9	10.5	6.0	6.1	0.0
	6+	4.3	4.0	7.5	3.3	3.1	0.0
	Q	0.51	0.48	0.17	0.16	0.16	0.50
	P	0.45	0.48	0.59	0.64	0.64	0.28
200	0	48.4	44.9	0.4	0.4	0.4	14.2
	1	19.4	20.6	5.1	7.1	7.1	79.6
	2	15.9	18.0	21.5	29.9	30.5	6.1
	3	8.4	8.9	30.1	34.6	34.4	0.1
	4	4.2	4.1	18.2	16.2	16.7	0.0
	5	1.8	1.9	13.9	8.2	7.8	0.0
	6+	1.9	1.6	10.8	3.6	3.1	0.0
	Q	0.52	0.49	0.14	0.13	0.13	0.20
	P	0.45	0.50	0.66	0.72	0.72	0.43

Q . . . Average Relative Distance from the True Break.

Table 80: Percentage Rates of detecting 0-5 breaks and Distance, Break - Middle, 1000 draws, SNR=1, $\delta = 0.050$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	31.7	26.4	2.0	2.0	2.0	50.6
	1	23.4	19.7	7.7	10.4	10.3	36.2
	2	18.8	17.0	34.6	38.7	38.6	11.4
	3	8.6	7.6	29.5	28.1	28.1	1.7
	4	6.5	4.4	15.6	13.6	13.8	0.1
	5	3.0	3.0	5.8	4.9	4.9	0.0
	6+	8.0	21.9	4.8	2.3	2.3	0.0
	Q	0.41	0.38	0.18	0.18	0.18	0.59
	P	0.48	0.48	0.53	0.55	0.55	0.25
100	0	45.8	42.4	1.6	1.6	1.6	44.3
	1	21.5	22.5	9.4	14.6	14.5	47.0
	2	14.3	16.3	28.8	35.2	35.1	8.0
	3	7.8	8.7	29.9	28.6	28.5	0.7
	4	4.1	4.2	17.3	13.5	13.9	0.0
	5	2.2	1.9	7.3	4.6	4.5	0.0
	6+	4.3	4.0	5.7	1.9	1.9	0.0
	Q	0.51	0.48	0.16	0.16	0.16	0.50
	P	0.45	0.48	0.59	0.65	0.65	0.28
200	0	48.4	44.9	0.9	0.9	0.9	14.2
	1	19.4	20.6	7.7	10.7	10.7	79.6
	2	15.9	18.0	25.3	32.8	33.4	6.1
	3	8.4	8.9	29.4	31.3	31.8	0.1
	4	4.2	4.1	18.6	17.6	17.2	0.0
	5	1.8	1.9	10.6	4.9	4.7	0.0
	6+	1.9	1.6	7.5	1.8	1.3	0.0
	Q	0.52	0.49	0.13	0.13	0.13	0.20
	P	0.45	0.50	0.67	0.74	0.74	0.43

Q . . . Average Relative Distance from the True Break.

Table 81: Percentage Rates of detecting 0-5 breaks and Distance, Break - Middle, 1000 draws, SNR=1, $\delta = 0.075$

		2 step					BP
T	Criterion	$EBIC$ $EBIC$	QS QS	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS	Seq
50	0	31.7	26.4	1.6	1.7	1.7	50.6
	1	23.4	19.7	8.7	11.7	11.7	36.2
	2	18.8	17.0	37.0	40.1	39.9	11.4
	3	8.6	7.6	29.2	27.6	27.8	1.7
	4	6.5	4.4	13.3	10.6	10.9	0.1
	5	3.0	3.0	5.0	3.9	3.8	0.0
	6+	8.0	21.9	5.2	4.4	4.2	0.0
	Q	0.41	0.38	0.18	0.18	0.18	0.59
	P	0.48	0.48	0.54	0.56	0.56	0.25
100	0	45.8	42.4	2.4	2.6	2.6	44.3
	1	21.5	22.5	8.8	13.2	13.2	47.0
	2	14.3	16.3	32.5	38.0	37.9	8.0
	3	7.8	8.7	32.8	30.2	30.6	0.7
	4	4.1	4.2	13.2	10.7	10.3	0.0
	5	2.2	1.9	6.5	4.0	4.1	0.0
	6+	4.3	4.0	3.8	1.3	1.3	0.0
	Q	0.51	0.48	0.17	0.17	0.16	0.50
	P	0.45	0.48	0.59	0.64	0.64	0.28
200	0	48.4	44.9	1.2	1.2	1.2	14.2
	1	19.4	20.6	8.0	12.0	12.0	79.6
	2	15.9	18.0	25.6	34.0	34.8	6.1
	3	8.4	8.9	30.7	30.4	30.7	0.1
	4	4.2	4.1	19.4	16.3	15.7	0.0
	5	1.8	1.9	9.9	4.9	4.5	0.0
	6+	1.9	1.6	5.2	1.2	1.1	0.0
	Q	0.52	0.49	0.13	0.12	0.12	0.20
	P	0.45	0.50	0.68	0.75	0.75	0.43

Q . . . Average Relative Distance from the True Break.

Table 82: Best δ for the Q, Break - Middle, 1000 draws, SNR=1

	2 step		
T	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.050	0.025	0.025
100	0.075	0.025	0.025
200	0.075	0.025	0.025

Table 83: Best δ for the P, Break - Middle, 1000 draws, SNR=1

	2 step		
T	IC_1 IC_2	IC_1 $EBIC$	IC_1 QS
50	0.050	0.025	0.025
100	0.075	0.025	0.025
200	0.075	0.025	0.025

C.15 Two Parameters - One Break - Middle - Positions of the Breaks

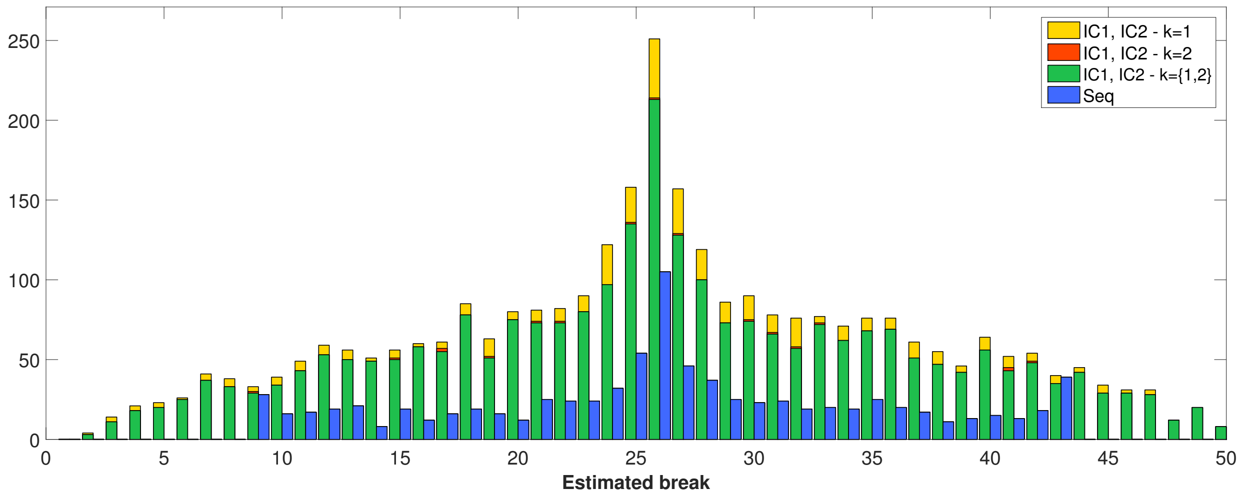

Figure 28: Positions of the Estimated Breaks, $T = 50$, SNR $= 2$, $\delta = 0.025$, Break in the Middle

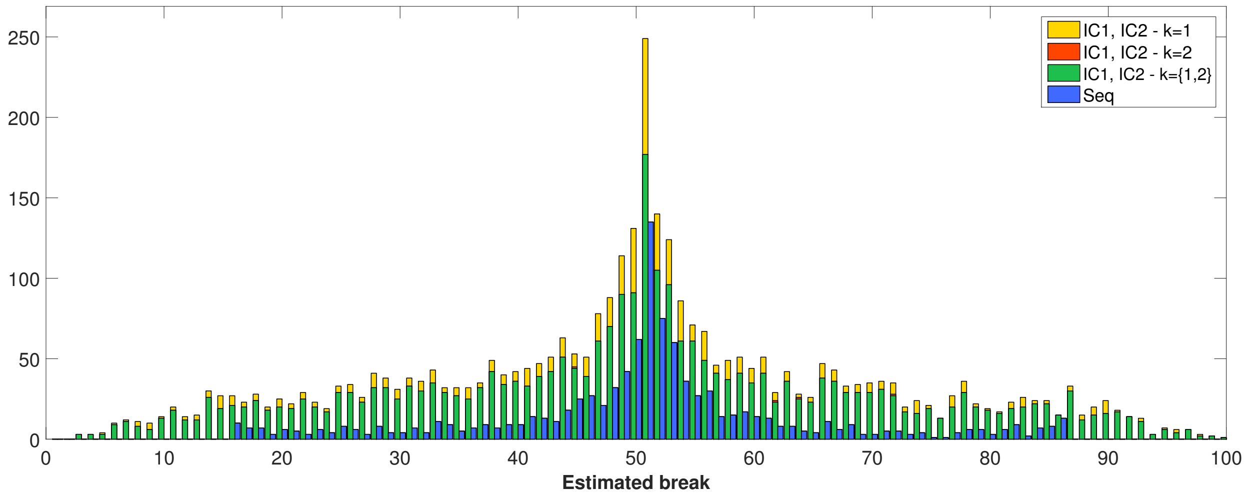

Figure 29: Positions of the Estimated Breaks, $T = 100$, SNR $= 2$, $\delta = 0.025$, Break in the Middle

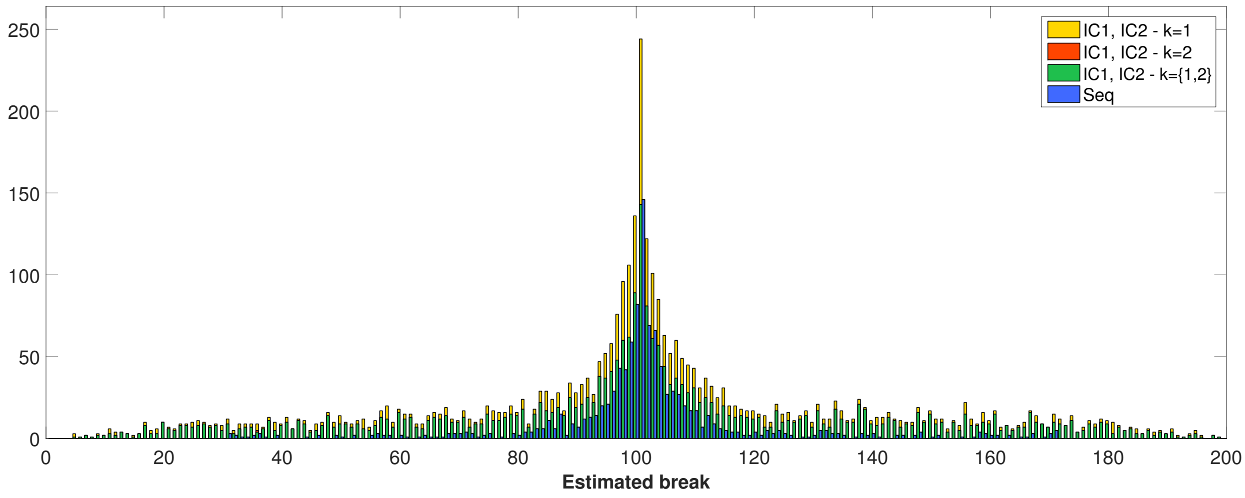

Figure 30: Positions of the Estimated Breaks, $T = 200$, SNR $= 2$, $\delta = 0.025$, Break in the Middle

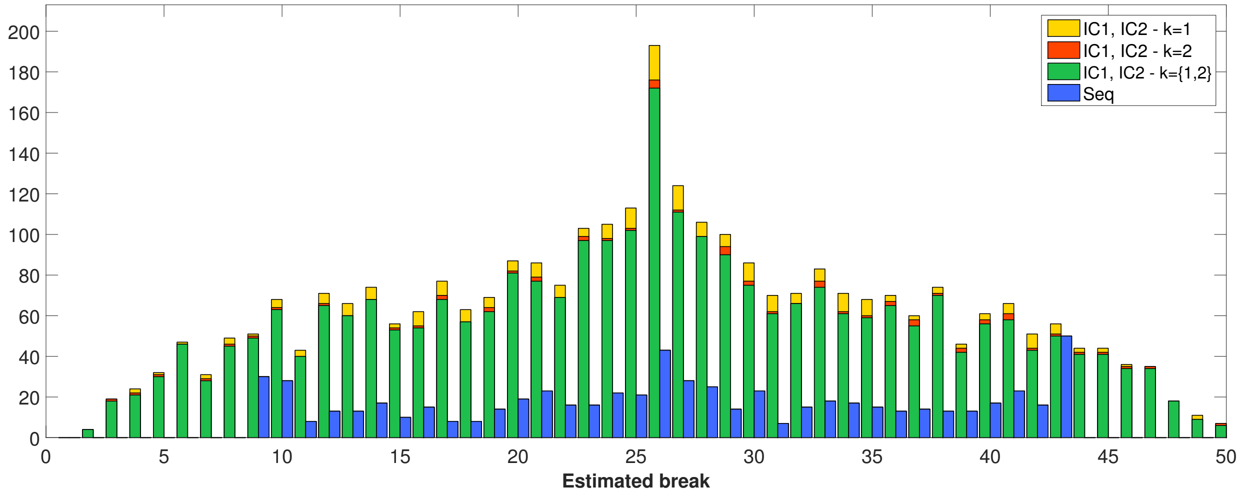

Figure 31: Positions of the Estimated Breaks, $T = 50$, SNR $= 1$, $\delta = 0.025$, Break in the Middle

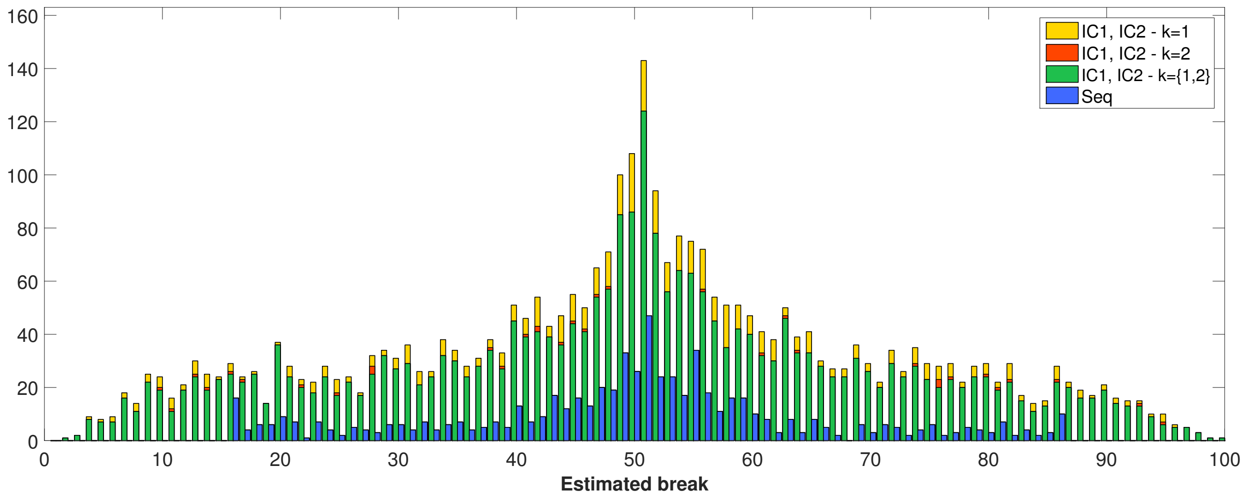

Figure 32: Positions of the Estimated Breaks, $T = 100$, SNR $= 1$, $\delta = 0.025$, Break in the Middle

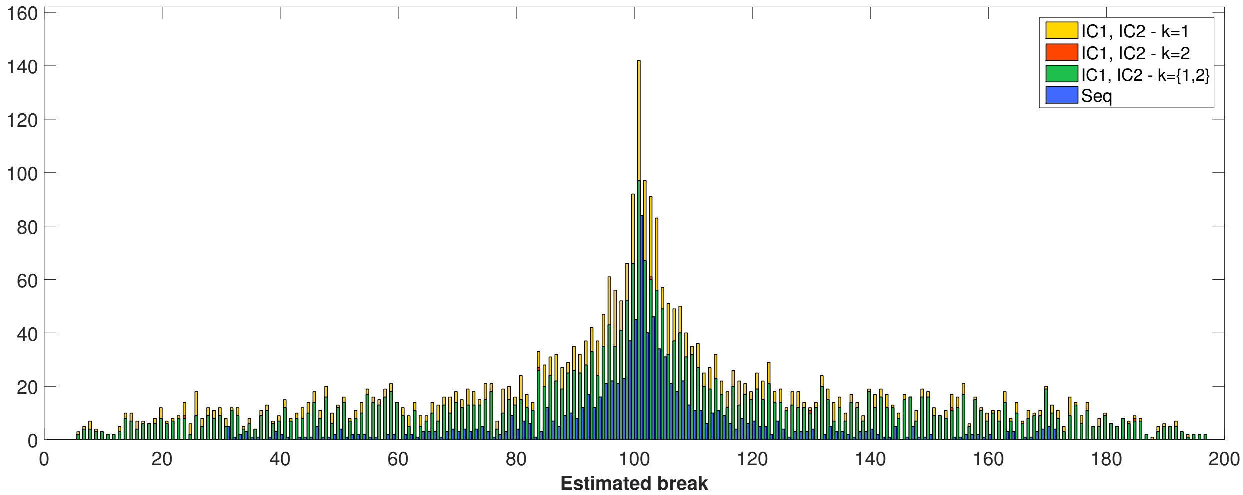

Figure 33: Positions of the Estimated Breaks, $T = 200$, SNR $= 1$, $\delta = 0.025$, Break in the Middle