Statistica Sinica Preprint No: SS-2022-0323	
Title	Inference for Change Points in High Dimensional Mean Shift Models
Manuscript ID	SS-2022-0323
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0323
Complete List of Authors	Abhishek Kaul and George Michailidis
Corresponding Authors	Abhishek Kaul
E-mails	abhishek.kaul@wsu.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

Statistica Sinica

Inference for Change Points in High Dimensional Mean Shift Models

Abhishek Kaul and George Michailidis

Washington State University and University of California, Los Angeles

Abstract: We consider the problem of constructing confidence intervals for the locations of change points in a high-dimensional mean shift model. We develop a locally refitted least squares estimator and obtain component-wise and simultaneous rates of estimation of change points. The simultaneous rate is the sharpest available by at least a factor of $\log p$, while the component-wise one is optimal. These results enable existence of limiting distributions for the locations of the change points. Subsequently, component-wise distributions are characterized under both vanishing and non-vanishing jump size regimes, while joint distributions of change point estimates are characterized under the latter regime, which also yields asymptotic independence of these estimates. We provide the relationship between these distributions, which allows construction of regime adaptive confidence intervals. All results are established under a high dimensional scaling, in the presence of diverging number of change points. They are illustrated on synthetic data and on sensor measurements from smartphones for activity recognition.

Key words and phrases: High dimensional, Inference, Limiting distributions, Multiple change points, Optimal estimation, Regime adaptation

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

1. Introduction

Statistical models with multiple change points are of significant interest due to their numerous applications in diverse areas, including economics and finance (Frisén (2008)), functional genomics and neuroscience (Koepcke et al. (2016)) amongst others. These have been studied extensively for a variety of statistical models, including mean shifts, regression, graphical models, factor and specific time series models and various algorithms have been developed to accomplish this task -dynamic programming, regularized cost functions, multiscale methods, etc., see, e.g. the review by Niu et al. (2016).

The main statistical tasks in change point analysis aim to address the following questions: (i) *whether change point(s) exist* in the data, (ii) assuming their existence, *estimation* of their location and (iii) *post-estimation inference*. The literature on the first two aspects is extensive and a multitude of methods for a variety of models is available, including under a fixed dimension p setting (e.g., Fryzlewicz (2014); Frick et al. (2014)) or a growing dimension with the sample size setting (e.g., Wang and Samworth (2018)). We refer to (Yu, 2020) for a recent review of available methods and results in both univariate and multivariate frameworks. On the other hand, the literature on the third objective of post-estimation inference is rather sparse, with selected results even in the univariate case only appearing recently.

This paper quantifies uncertainty through the construction of appropriately calibrated confidence intervals, when estimating multiple change points in a high dimen-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

sional mean shift model given by

$$y_t \;=\; \sum_{j=1}^{N+1} \theta^*_{(j)} \mathbf{1}[\tau^0_{j-1} < t \leq \tau^0_j] + \varepsilon_t, \qquad \text{for } t = 1, ..., T, \tag{1.1}$$

wherein $y_t = (y_{t1}, y_{t2}, ..., y_{tp})^T \in \mathbb{R}^p$ denotes the response, and the noise vector $\varepsilon_t \in \mathbb{R}^p$ comprising of subexponential random variables. The parameters to be estimated are the number of change points $N \in \mathbb{N}^+ = \{1, 2, ...\}$, their locations $\tau^0 = (\tau^0_1, \tau^0_2, ..., \tau^0_N)^T \subseteq \{1, ..., T\}^N$, and finally the mean vectors $\theta^*_{(j)} \in \mathbb{R}^p$, $j = 1, ..., (N+1)$. The location parameters τ^0_j, $j = 1, .., N$ are of prime interest, while $\tau^0_0 = 0$ and $\tau^0_{N+1} = T$ are defined for notational convenience. Finally, p (number of data streams), and N (number of changes) can diverge with sample size T, with the former potentially exponentially. Further, for $j = 1, ..., N$, define jump sizes of model (1.1) as,

$$\eta^*_{(j)} = (\theta^*_{(j)} - \theta^*_{(j+1)}), \quad \xi_j = \|\eta^*_j\|_2, \quad \underline{\xi} = \min_{1 \leq j \leq N} \xi_j \text{ and } \overline{\xi} = \max_{1 \leq j \leq N} \xi_j \tag{1.2}$$

To our knowledge the problem of post-estimation inference in this setting has not been addressed in the literature to date and the available inferential results under multiple change points are under univariate $p = 1$ designs, see, e.g., (Bai and Perron, 1998; Eichinger and Kirch, 2018a; Cho and Kirch, 2022). Further, results are available in single change point $N = 1$ settings; see Bai (1994) for a univariate ($p = 1$) model. The case of diverging p is considered in Bhattacharjee et al. (2017, 2020); Kaul et al. (2023) for mean shift, stochastic block models and graphical models, respectively. Recently, a similar problem has also been considered in Xu et al. (2022) for linear regression. The closest comparable article is that of Kaul et al. (2021) which considers the high

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

dimensional case, but is again limited to a single change point setting ($N = 1$). It is worth noting that the both methodological and technical differences between single and multiple change point settings are well established in the literature; further, it is known that the single change case yields considerable simplifications of the larger problem. To that end, several completely new results are established in this paper vis-a-vis those in Kaul et al. (2021), as shall be detailed throughout the article.

Main contributions and related literature.

1. **(Optimality in the estimation rate):** There are two distinct notions of minimax optimality in change point analysis. The available methods in any mutivariate (fixed, diverging or high dimensional p) and multiple change point framework have pursued optimality in the jump size, i.e., estimators of change points which can consistently estimate a jump size which is at the detection limit $\underline{\xi} \geq \sqrt{s \log p/T}$. In doing so, they sacrifice the rate of estimation which is slower than the optimal rate $O_p(\xi^{-2})$ by at least logarithmic factors of p and T. Examples include (Wang and Samworth, 2018; Wang et al., 2021) under high dimensionality, Cho et al. (2016) under fixed p, and also several univariate settings, including Fryzlewicz (2014).

In contrast, we show that if one moves *marginally* away from the detection limit of the jump size, then an optimal $O_p(\xi^{-2})$ rate can be attained. The required deviation is from $\underline{\xi} \geq \sqrt{s \log p/T}$ to $\underline{\xi} \geq s \log^{3/2} p/\sqrt{T}$. Note that the rate $O_p(\xi^{-2})$ is free of any dimensional terms, despite potential high dimensionality. This result provides optimality in estimation in the considered high-dimensional setting in the presence of

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

multiple change points. The available analogs are limited to univariate settings, which include recent articles by Verzelen et al. (2023); Eichinger and Kirch (2018b); Cho and Kirch (2022), as well as classical results obtained in Bai and Perron (1998). This optimal rate ensures existence of limiting distributions for the locations of the change points, which existing near-optimal methods can not provide.

2. **(Limiting distributions and confidence intervals):** These are obtained and characterized under the following regimes: (i) vanishing ($\underline{\xi} \to 0$), and (ii) non-vanishing ($\underline{\xi} \to \xi_\infty, 0 < \xi_\infty < \infty$) jump sizes. Further, we obtain both component-wise and joint limiting distributions that in turn enable construction of asymptotically valid confidence intervals, for any finite subset of the potentially diverging number of change points. To the best of our knowledge, the ability to perform inference on locations of multiple change point parameters is unavailable in the current literature under any multivariate framework. The only results available in the literature are under univariate settings, namely those of Bai and Perron (1998); Eichinger and Kirch (2018b); Cho and Kirch (2022).

3. **(Regime adaptation):** Finally, we address the question of *which of the two distributions (vanishing vs non-vanishing)* the practitioner should use to construct confidence intervals. The traditional answer for this problem is to implement a regime adaptive bootstrap procedure, proposed in Antoch et al. (1995) in a $p = 1$ setting and also considered in Cho and Kirch (2022); Ng et al. (2022). We establish a novel result that illustrates the inherent *asymptotic adaptivity* of the limiting distribution under

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

the non-vanishing jump size regime to that under the vanishing jump regime. This result shows that if one always employs the former distribution to obtain confidence intervals, then they remain asymptotically valid even if the regime was *mis-specified*. Hence, it eliminates the need for implementing a computationally expensive bootstrap procedure which is especially useful in high-dimensional settings.

As shall be made precise in the following section, methodologically we consider a refitting process between random end points, where the refitting is undertaken on overlapping segments of data. This approach was conceptually utilized in the application section of Bai (1997) to heuristically extend a single change point method to multiple change points, but to our knowledge it has not since been theoretically supported. This is also in contrast to Kaul et al. (2021), where refitting is undertaken on a single segment between fixed end points. To further describe the contributions of this article in comparison to Kaul et al. (2021) we note the following. The issue of a joint distribution of multiple change points, as well as their asymptotic independence does not arise in the framework of Kaul et al. (2021), since the latter considers a single change point. Further, we note that the asymptotic independence of change point estimates established in the sequel, is somewhat counter-intuitive, yet valid. Note that change point estimates are by construction obtained on overlapping segments of the data and may seem at first sight correlated, which as shown is not the case. Finally, we note that the third contribution (on regime adaptation) provides a novel solution to the classical dichotomy problem of selecting in practice the limiting distribution to use for

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

constructing confidence intervals and there is no analogous result in Kaul et al. (2021).

Notation: $\|\delta\|_1$, $\|\delta\|_2$, $\|\delta\|_\infty$ represent the usual 1-, Euclidean, and sup-norms, respectively. For a set of indices $U \subseteq \{1, 2, ..., p\}$, let $\delta_U = (\delta_j)_{j\in U}$ represent a subvector of components corresponding to indices in U. $|U|$ and U^c represent the cardinality and complement of U. Denote by $a \wedge b = \min\{a, b\}$, and $a \vee b = \max\{a, b\}$. We use a generic $c_u > 0$ to represent universal constants. All limits are with respect to the sampling periods T. The notation $\Rightarrow$ denotes convergence in distribution.

2. Preliminaries

Given the high dimensional nature of the posited model (1.1) (diverging dimension p as a function of the sample size T), we assume a sparsity condition on the jump vectors, $\|\eta^0_{(j)}\|_0 \leq s$, $1 \leq j \leq N$, where $1 \leq s << T$, see, e.g., Wang and Samworth (2018) and Harchaoui and Lévy-Leduc (2010). Next, we consider the following reparameterized version of the model through global centering ($x_t = y_t - \bar{y}$)

$$x_t = \sum_{j=1}^{N+1} \theta^0_{(j)} \mathbf{1}[\tau^0_{j-1} < t \leq \tau^0_j] + \varepsilon^*_t, \text{ for } t = 1, ..., T, \quad \text{where,}$$

$$\theta^0_{(j)} = \theta^*_{(j)} - w(\theta^*), \quad w(\theta^*) = \frac{1}{T}\sum_{j=1}^{N+1}(\tau^0_j - \tau^0_{j-1})\theta^*_{(j)} \quad \text{and}$$

$$\varepsilon^*_t = \varepsilon_t - \bar{\varepsilon}, \quad \bar{\varepsilon} = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t, \quad t = 1, ..., T. \tag{2.1}$$

that transfers the s-sparsity of the jump vectors to Ns-sparsity of individual mean vectors. Doing so allows us to exploit the assumption of sparsity quite differently

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

than Wang and Samworth (2018) and Harchaoui and Lévy-Leduc (2010), without our assumptions being any more stringent. A further algebraic manipulation of $\theta^0_{(j)}$ yields a more insightful expression for these reparameterized means:

$$\theta^0_{(j)} = \frac{1}{T}\Big[-\sum_{k=1}^{j-1} \tau^0_k \eta^*_{(k)} + \sum_{k=j}^{N} (T - \tau^0_k)\eta^*_{(k)}\Big], \quad j = 1, ..., N+1. \tag{2.2}$$

Here $\sum_{j=1}^{k-1}$, and $\sum_{j=k}^{N}$ are defined to be zero at $k = 1$ and $k = N+1$, respectively. Note that in (2.2) the $\theta^0_{(j)}$, $j = 1, ..., N+1$, are expressed as a linear combination of s-sparse jump vectors $\eta^*_{(j)}$, $j = 1, ..., N$. The only consequence is a diminishing temporal dependence induced in the re-defined noise term ε^*_t of (2.1), which we shall show has no statistical impact on our results.

Remark 1. (On the jump sizes of the reparameterized model (2.1)) Note that the mean parameters $\theta^*_{(j)}$ and $\theta^0_{(j)}$, $j = 1, ..., N+1$ of models (1.1) and (2.1) are distinct. However, the jump vectors and jump sizes that control properties of the change points remain identical, since

$$\big(\theta^*_{(j)} - \theta^*_{(j+1)}\big) = \eta^*_{(j)} = \big(\theta^0_{(j)} - \theta^0_{(j+1)}\big) = \eta^0_{(j)}, \quad j = 1, ..., N.$$

Consequently, the jump size parameters ξ_j, $j = 1, ..., N$, $\underline{\xi}$, and $\overline{\xi}$ defined in (1.2) remain identical for the two models. Thus, in the remainder we do not distinguish between jump vector and jump size parameters of models (1.1) and (2.1) that are denoted by $\eta^0_{(j)}$, ξ_j, $j = 1, ..., N$, and $\underline{\xi}$, and $\overline{\xi}$, irrespective of the underlying model.

Next, we consider a *locally refitted* estimator that yields the optimality and inferential properties discussed earlier. Let $\tau_{-j} = (\tau_1, ..., \tau_{j-1}, \tau_{j+1}, ..., \tau_N)^T \in \{1, ..., T-1\}^{N-1}$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

be any vector with the j^{th} component removed, additionally satisfying $\tau_{j-1} < \tau_{j+1}$. Consider any $\theta_{(j)} \in \mathbb{R}^p$, $j = 1, ...N+1$ and let θ represent the concatenation of all $\theta'_{(j)}$s. Define the following squared loss function evaluated at any point $\tau_j \in \{\tau_{j-1}, ..., \tau_{j+1}\}$ w.r.t. realizations x_t, $t = 1, ..., T$ of model (2.1),

$$Q_j\big(\tau_j, \tau_{-j}, \theta\big) = \sum_{t=\tau_{j-1}+1}^{\tau_j} \|x_t - \theta_{(j)}\|_2^2 + \sum_{t=\tau_j+1}^{\tau_{j+1}} \|x_t - \theta_{(j+1)}\|_2^2. \tag{2.3}$$

For ease of presentation, assume for the time being the availability of preliminary estimates $\hat{\tau} = (\hat{\tau}_1, ..., \hat{\tau}_{\hat{N}})^T \in \mathbb{R}^{\hat{N}}$ of the locations of the change points, and $\hat{\theta}_j \in \mathbb{R}^p$, $j = 1, ..., \hat{N}+1$ of mean parameters of model (2.1). In Section 4, we present how to obtain such preliminary estimates that are *not* optimal, but straightforward to construct. Then, for each $j = 1, ..., \hat{N}$, define a *locally refitted plug-in* least squares estimator,

$$\tilde{\tau}_j := \tilde{\tau}_j\big(\hat{\tau}_{-j}, \hat{\theta}\big) = \underset{\hat{\tau}_{j-1} < \tau_j < \hat{\tau}_{j+1}}{\arg\min} \; Q_j\big(\tau_j, \hat{\tau}_{-j}, \hat{\theta}\big), \quad j = 1, ..., \hat{N} \tag{2.4}$$

This local refitting based on *slower than optimal* preliminary nuisance estimates leads to an improved estimate of the j^{th} change point parameter that is *optimal*. This in turn provides sufficient regularity for limiting distributions of these updated estimates to exist, in the presence of potentially diverging number of change points and high dimensionality of means.

Let $T\underline{\ell}$ be the least spacing between change points satisfying $\min_{1 \le j \le N+1} \big(\tau_j^0 - \tau_{j-1}^0\big) \ge T\underline{\ell} \ge 1$, then the preliminary estimates to initialize (2.4) are required to

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

satisfy,

$$\max_{1\le j\le N+1} \|\hat{\theta}_{(j)} - \theta^0_{(j)}\|_2 \le c_u\sigma\Big\{\frac{Ns\log(p\vee T)}{T\underline{\ell}}\Big\}^{\frac{1}{2}}, \quad \text{and}$$
$$\hat{N} = N, \quad \max_{j=1,\ldots,N} |\hat{\tau}_j - \tau^0_j| \le c_{u1}T\underline{\ell}, \tag{2.5}$$

with probability at least $1 - o(1)$. In Section 3, we establish the main results on inference, under these general conditions. As previously mentioned, Section 4 comprehensively establishes how to construct such preliminary estimates under the intuitive rate condition,

$$\Big(\frac{\sigma}{\underline{\xi}}\Big)\Big\{\frac{Ns\log^{3/2}(p\vee T)}{\sqrt{(T\underline{\ell})}}\Big\} \le c_{u1},$$

wherein $c_{u1} > 0$ is a small enough constant, σ^2 is a variance proxy parameter of the data generating process (Condition 1) and the remained have all been defined earlier, including $\underline{\xi}$ as the least jump size and N, s, p and the number of changes, sparsity and overall dimension, respectively.

3. Main Results

We first state sufficient conditions for the results pertaining to the estimator $\tilde{\tau}$ of (2.4) All assumptions are on the first level model (1.1), even though the estimator $\tilde{\tau}$ is based on the centered data obtained from the reparametrized model (2.1). All additional technical issues due to this transformation are addressed in the proofs.

Condition 1. (on distributions): *The vectors* $\varepsilon_t = (\varepsilon_{t1}, ..., \varepsilon_{tp})^T$, $t = 1, .., T$, *are independent and identically distributed subexponential random vectors with variance*

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

proxy $\sigma^2 < \infty$ *(Def.* ?? *&* ?? *in Supplement* ??*)*

The posited class contains distributions that may exhibit heavier tails than the Gaussian, and also includes discrete ones. More generally the class of subexponential distributions subsumes the subgaussian class. Members of the class include the Laplace, mean centered Exponential, mean centered Chi-square, amongst several others.

Condition 2. (on parameters):

(i) **(covariance)** *The matrix* $\Sigma := E\varepsilon_t\varepsilon_t^T$ *has bounded eigenvalues, i.e.,* $0 < \kappa^2 \leq \text{mineigen}(\Sigma) < \text{maxeigen}(\Sigma) \leq \phi^2 < \infty$, *for constants* κ^2, ϕ^2.

(ii) **(separation of jumps)** *Assume there exists at least one change point* $(N \geq 1)$*; further, all* N *change points are distinct and sufficiently separated, i.e., for* $(\tau_j^0 - \tau_{j-1}^0) = T\ell_j$, $j = 1, ..., N+1$, *we have* $\min_{1\leq j\leq N+1\}} T\ell_j \geq T\underline{\ell} \geq 1$, *for a positive sequence* $\underline{\ell} \to 0$, *such that* $\log T = o\big(\sqrt{(T\underline{\ell})}\big)$.

(iii) **(sparsity)** *Let* $\eta_{(j)}^0$, $j = 1, ..., N$ *be jump vectors as in (1.2), so that* $\max_{1\leq j\leq N} \|\eta_{(j)}^0\|_0 \leq s$, *wherein* $s \geq 1$ *is a positive sequence of integers.*

(iv) **(relative order of jumps)** *For* $\underline{\xi}$, $\overline{\xi}$ *in (1.2), let* $\overline{\xi} \leq c_u\underline{\xi}$, *for* $c_u \geq 1$.

All parts of Condition 2 are fairly standard in the literature. The bounds of Condition 2(i) ensures finiteness of the asymptotic variances of the limiting processes. Condition 2(ii) assumes existence of at least one change point and separation of all N change points. In practice, existence of at least one change is established via boundary tests, such as (Jirak (2015); Chen et al. (2022)); however, focusing on our objective of

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

post-estimation inference, we assume this existence a priori. Condition 2(iii) assumes sparsity of the jump vectors and as discussed in Section 2, we exploit this via the reparametrized model (2.1) which transfers the assumed s-sparsity of Condition 2(iii) to an Ns-sparsity on the individual $\theta^0_{(j)}$.

Condition 2(iv) assumes all jump sizes of model (1.1) are of the same order. This assumption ensures that neighboring change points do not interfere in the estimation of τ^0_j. Stronger versions of 2(iv) are also common such as by assuming bounded jump sizes, $c_{u1} \leq \underline{\xi} \leq \overline{\xi} \leq c_{u2}$, e.g. Cho and Fryzlewicz (2015) amongst several others. Note that the boundedness of jump sizes implies 2(iv), but not conversely; hence 2(iv) is a weaker assumption, it allows vanishing jump sizes, whereas boundedness does not.

Next, define sets of non-zero components of the means $\theta^0_{(j)}$ of (2.1),

$$S_j = \big\{k \in \{1, ..., p\};\ \theta^0_{(j)k} \neq 0\big\}, \quad j = 1, ..., N+1. \tag{3.1}$$

and let S^c_j, $j = 1, ...N+1$ denote the complement sets. The earlier discussion in context of model (2.1), yields, $\max_j |S_j| \leq Ns$. Our analysis is agnostic on the choice of the estimators used to obtain the preliminary estimates $\hat{\tau}$ and $\hat{\theta}$. Instead, we shall rely on the following assumption.

Condition 3. (preliminary estimates $\hat{\tau}$, and $\hat{\theta}$): *Let $\pi_T \to 0$ be a positive sequence and assume (i) and (ii) below hold with probability $1 - \pi_T$.*

*(i) (***Preliminary change point estimate** $\hat{\tau} = (\hat{\tau}_1, ..., \hat{\tau}_{\hat{N}})$ *of* τ^0*): For an appropri-*

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

ately chosen small enough constant $c_{u1} > 0$*, we assume that,*

$$\hat{N} = N, \qquad \max_{1 \le j \le N} |\hat{\tau}_j - \tau_j^0| \le c_{u1} T \underline{\ell},$$

wherein $\underline{\ell}$ *is the separation sequence defined in Condition 2(ii).*

*(ii) (***Preliminary mean estimates of θ^0 of (2.2)***): Assume that the following two properties hold. (a) Estimates* $\hat{\theta}_{(j)}$*,* $j = 1, ..., N+1$*, satisfy* $\|(\hat{\theta}_{(j)})_{S_j^c}\|_1 \le 3\|(\hat{\theta}_{(j)} - \theta^0_{(j)})_{S_j}\|_1$*, with* S_j*,* $j = 1, ..., N+1$ *being sets of non-zero components defined in (3.1). (b) Assume there exists a sequence* $r_T \ge 0$*, such that,*

$$\max_{1 \le j \le N+1} \|\hat{\theta}_{(j)} - \theta^0_{(j)}\|_2 \le r_T = \frac{c_{u1} \underline{\xi}}{(Ns)^{1/2} \log(p \vee T)},$$

for a suitably small constant $c_{u1} > 0$*, wherein* $\underline{\xi}$ *is the least jump size (1.2).*

Condition 3 is constructed carefully with the following considerations. First, it is stated in the weakest form that is sufficient for optimality of $\tilde{\tau}$, and second that it is feasible. This condition serves as a *temporary placeholder* in the structure of our argument. Feasibility of this condition is illustrated in detail in Section 4, where we show that it can be eliminated and replaced by the intuitive rate restriction

$$\left(\frac{\sigma}{\underline{\xi}}\right)\left\{\frac{Ns \log^{3/2}(p \vee T)}{\sqrt{(T\underline{\ell})}}\right\} \le c_{u1}, \tag{3.2}$$

wherein $c_{u1} > 0$ is a small enough constant. We further show that any sub-optimal change point estimator from the literature -e.g., Wang and Samworth (2018); Cho et al. (2016); Harchaoui and Lévy-Leduc (2010)- can serve as a preliminary change point one, even though it does not possess the properties needed for inference; see, Corollary 1 and Algorithm 1 of Section 4.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

3.1 Rates of Convergence

Theorem 1. *(component-wise rate of estimation) Assume that Conditions 1, 2 and 3 hold. Then, for any given* $j = 1, ..., N$, *and any* $0 < a < 1$ *with* $c_a \geq \sqrt{(1/a)}$, *the following holds*

$$|\tilde{\tau}_j - \tau_j^0| \leq c_u c_a^2 \sigma^2 \xi_j^{-2}$$

with probability at least $1 - 2a - o(1) - \pi_T$. *Equivalently,* $\sigma^{-2}\xi_j^2(\tilde{\tau}_j - \tau_j^0) = O_p(1)$, *for any given* $j = 1, ..., N$.

Theorem 1 provides componentwise rates for $\tilde{\tau}_j$, $j = 1, ..., N$, that are minimax optimal (see, e.g., Wang and Samworth (2018) (Proposition 3 of Supplement), Verzelen et al. (2023) (Proposition 6)). It is the same rate that would be obtained if the nuisance parameters θ^0 and τ_{-j}^0 *were known*. This is an instance of the adaptation property, as described in Bickel (1982), in the presence of a diverging number of change points and underlying high dimensionality. The next result establishes an ℓ_∞ rate of convergence for the proposed refitted estimates.

Theorem 2. *(ℓ_∞ rate of estimation) Assume that Conditions 1, 2 and 3 hold. Then, the following holds*

$$\max_{1 \leq j \leq N} |\tilde{\tau}_j - \tau_j^0| \leq c_u \sigma^2 \underline{\xi}^{-2} \log^2 T,$$

with probability at least $1 - o(1) - \pi_T$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

The rates in Theorems 1 and 2 are the sharpest available in the literature under high dimensionality. For example, they are at least $\log p/\log T$ faster than those in Wang and Samworth (2018) for their corresponding estimator. We also note that the local refitting undertaken in (2.4) does not alter the number of change points $\hat{N}$ of the preliminary estimate $\hat{\tau}$, thus, Condition 3(i) ensures $\tilde{N} = \hat{N} = N$, w.p. $\to 1$.

3.2 Limiting Distributions and Regime Adaptation

Next, we obtain component-wise and joint limiting distributions for the refitted change point estimators. We start with some additional assumptions.

Condition 4. (stability of asymptotic variances): *For jump sizes ξ_j, $j = 1, ..., N$ defined in (1.2) and covariance Σ as in Condition 2, assume the following limits exists,*

$$\xi_j^{-2}\big(\eta_{(j)}^{0T}\Sigma\eta_{(j)}^{0}\big) \to \sigma^2_{(\infty,j)},\ 0 < \sigma^2_{(\infty,j)} < \infty,\ \textit{for each given } j = 1, ..., N. \tag{3.3}$$

All limits in this work are with respect to the observation period T. The limits in Condition 4 are acting in T via the dimension p and the jump sizes ξ_j, $j = 1, ..., N$. The quantities $\sigma^2_{(\infty,j)}$ $j = 1, ..., N$, serve as variance parameters of the limiting processes, thus the need for their stability. Note that finiteness of these limits is already guaranteed by Condition 2(i), while the current condition only assumes their stability. To see this, observe that the assumed convergence is on a sequence that is guaranteed to be bounded below and above, i.e.,

$$0 < \kappa^2 \leq \min_{1\leq j\leq N} \xi_j^{-2}\big(\eta_{(j)}^{0T}\Sigma\eta_{(j)}^{0}\big) < \max_{1\leq j\leq N} \xi_j^{-2}\big(\eta_{(j)}^{0T}\Sigma\eta_{(j)}^{0}\big) \leq \phi^2 < \infty. \tag{3.4}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

The above inequalities follow from the bounded eigenvalues assumption on Σ (Condition 2(i)). An easier to interpret, but stronger sufficient condition for the finiteness of these limits is by assuming absolute summability of each row or column of Σ. This condition is satisfied by large classes of covariances, including banded and Toeplitz type matrices.

Condition 5. (rate of convergence of jump size): *Let $\underline{\xi}$ and $\underline{\ell}$ be as defined in (1.2) and Condition 2(ii), respectively. Then, we assume that $\underline{\xi}^{-1}\log T = o\big(\sqrt{(T\underline{\ell})}\big)$.*

Recall our results allow potentially diminishing jump sizes. Condition 5 is the first requirement imposed on the rate at which the least jump size ($\underline{\xi}$) can converge to zero.

Theorem 3. *(component-wise distributions for the vanishing jump size regime) Assume that Conditions 1, 2, 4 and 5 hold. Consider any given change point $j = 1, ..., N$, and assume that the jump size $\xi_j \to 0$ is vanishing, and that τ^0_{-j}, θ^0 are known. Denote $\tilde{\tau}^*_j = \tilde{\tau}_j(\tau^0_{-j}, \theta^0)$. Then,*

$$\xi_j^2(\tilde{\tau}^*_j - \tau^0_j) \Rightarrow \arg\max_{\zeta\in\mathbb{R}} \big\{2\sigma_{(\infty,j)}W_j(\zeta) - |\zeta|\big\}, \tag{3.5}$$

where $W_j(\zeta)$ is a two sided standard Brownian motion. Alternatively, when τ^0_{-j} and θ^0 are unknown, let $\tilde{\tau}_j$ be as defined in (2.4) and assume $\hat{\tau}_{-j}$ and $\hat{\theta}$ satisfy Condition 3. Further, assume that the sequence r_T in Condition 3(ii) satisfies $r_T = \{o(1)\underline{\xi}\}/\{(Ns)^{1/2}\log(p\vee T)\}$. Then, the convergence (3.5) also holds when $\tilde{\tau}^*_j$ is replaced with $\tilde{\tau}_j$.

Observe that a change of variable to $\zeta = \sigma^2_{(\infty,j)}\zeta'$, yields that $\arg\max_{\zeta\in\mathbb{R}}\big\{2\sigma_{(\infty,j)}W_j(\zeta) - |\zeta|\big\} =^d \sigma^2_{(\infty,j)}\arg\max_{\zeta'\in\mathbb{R}}\big\{2W_j(\zeta') - |\zeta'|\big\}$. Its cumulative distribution functions and

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

thus its quantiles are readily available (Yao, 1987). These can be utilized to construct asymptotic component-wise confidence intervals for change points under the assumed vanishing jump regime. Next, we consider the non-vanishing regime for $\xi_j \to \xi_{(\infty,j)}$, $0 < \xi_{(\infty,j)} < \infty$. For this purpose, we require the following additional assumption.

Condition 1′ (additional distributional assumption): *Suppose Condition 1, 2(i) and 4 hold and assume for given* $j = 1, ..., N$ *and constants* $c_1, c_2 \in \mathbb{R}$, *the r.v.'s* $c_1 + c_2 \varepsilon_t^T \eta_{(j)}^0 \Rightarrow \mathcal{P}\big(c_1, c_2^2 \xi_{(\infty,j)}^2 \sigma_{(\infty,j)}^2\big)$, *for* $t = 1, ..., T$, *for some distribution* $\mathcal{P}$, *which is continuous and supported in* $\mathbb{R}$.

As before, limits here are acting in T via p and ξ_j. The only additional requirement in Condition 1′ is that the random variables under consideration are continuously distributed. If one assumes a Gaussian error process, then Condition 1′ is redundant, i.e., $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$, then $\mathcal{P}\big(c_1, c_2^2 \xi_{(\infty,j)}^2 \sigma_{(\infty,j)}^2\big) \sim \mathcal{N}\big(c_1, c_2^2 \xi_{(\infty,j)}^2 \sigma_{(\infty,j)}^2\big)$. More generally, the variance expression in $\mathcal{P}\big(c_1, c_2^2 \xi_{(\infty,j)}^2 \sigma_{(\infty,j)}^2\big)$ follows from Condition 4 together with the jump size regime assumption of $\xi_j \to \xi_{(\infty,j)}$. The expression for its mean is trivial. Consequently, the limiting distribution of $c_1 + c_2 \varepsilon_t^T \eta_{(j)}^0$ is well defined, i.e. supported in $\mathbb{R}$. Thus, Condition 1′ simply reflects notation for the underlying distribution $\mathcal{P}$ and the notation $\mathcal{P}(\mu, \sigma^2)$, with $E\mathcal{P}(\mu, \sigma^2) = \mu$, and $\mathrm{var}\big(\mathcal{P}(\mu, \sigma^2)\big) = \sigma^2$, is only for ease of presentation, it does not imply $\mathcal{P}$ is characterized only by mean and variance.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

Next, define the following two-sided random walk initialized at the origin,

$$\mathcal{C}_\infty(\zeta, \xi, \sigma^2) = \begin{cases} \sum_{t=1}^{\zeta} z_t, & \zeta \in \mathbb{N}^+ = \{1, 2, 3, ...\} \\ 0, & \zeta = 0 \\ \sum_{t=1}^{-\zeta} z_t^*, & \zeta \in \mathbb{N}^- = \{-1, -2, -3, ...\}, \end{cases} \tag{3.6}$$

wherein z_t, z_t^* are independent copies of $\mathcal{P}\big(-\xi^2, 4\xi^2\sigma^2\big)$, which are also independent over all t, for the distribution $\mathcal{P}$ of Condition 1′. Finally, let

$$\mathcal{C}_{(\infty,j)}(\zeta) = \mathcal{C}_\infty\big(\zeta, \xi_{(\infty,j)}, \sigma^2_{(\infty,j)}\big) \qquad j = 1, ..., N, \tag{3.7}$$

wherein $\sigma^2_{(\infty,j)}$, $j = 1, ..., N$, are asymptotic variance parameters as defined in Condition 4. These random walks $\mathcal{C}_{(\infty,j)}(\zeta)$ of (3.7) can now be used to characterize the limiting distributions of $\tilde{\tau}_j$ for $j = 1, ..., N$, in the current non-vanishing regime. The only additional requirement in Condition 1′, is that of continuity of the distribution $\mathcal{P}$ needed for the regularity of the *argmax* of these two sided negative drift random walks.

Theorem 4. *(componentwise distributions for the non-vanishing regime) Suppose Conditions 1′, 2, 4 and 5 hold. Consider any given $j = 1, ..., N$, and assume that the jump size is non-vanishing, $\xi_j \to \xi_{(\infty,j)}$, $0 < \xi_{(\infty,j)} < \infty$, and that τ^0_{-j}, θ^0 are known. Let $\tilde{\tau}^*_j = \tilde{\tau}_j(\tau^0_{-j}, \theta^0)$, then, we have,*

$$(\tilde{\tau}^*_j - \tau^0_j) \Rightarrow \arg\max_{\zeta \in \mathbb{Z}} \mathcal{C}_{(\infty,j)}(\zeta), \tag{3.8}$$

where $\mathcal{C}_{(\infty,j)}(\zeta)$ is defined in (3.7). Alternatively, when τ^0_{-j} and θ^0 are unknown, let $\tilde{\tau}_j$ be as defined in (2.4) assume $\hat{\tau}_{-j}$ and $\hat{\theta}$ satisfy Condition 3. Additionally assume

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

sequence r_T *of Condition 3(ii) satisfies* $r_T = \{o(1)\underline{\xi}\}/\{(Ns)^{1/2}\log(p \vee T)\}$. *Then, the convergence (3.8) also holds when* $\tilde{\tau}_j^*$ *is replaced with* $\tilde{\tau}_j$.

The main distinction in the assumptions of Theorems 3 and 4 is the switch from a vanishing to a non-vanishing jump size. Since the analytical form of $\arg\max_{\zeta\in\mathbb{Z}} \mathcal{C}_{(\infty,j)}(\zeta)$ is unavailable, one can obtain the quantiles of these distributions by simulating sample paths of the random walks under consideration. The above results allow construction of asymptotic confidence intervals with any desired coverage $(1-\alpha)$ as

$$\big[(\tilde{\tau}_j - ME_j^\alpha),\, (\tilde{\tau}_j + ME_j^\alpha)\big], \quad \text{where,} \tag{3.9}$$

$ME_j^\alpha = q_\alpha^v \sigma^2_{(\infty,j)}/\xi_j^2$ or $ME_j^\alpha = q^{nv}_{(\alpha,j)}$, in the vanishing and non-vanishing regimes, respectively. The values q_α^v and $q^{nv}_{(\alpha,j)}$ represent quantiles at $(1-\alpha)$ coverage of the distributions of Theorem 3 and Theorem 4, respectively. These intervals shall guarantee a componentwise nominal coverage asymptotically at $(1-\alpha)$ for any given $j = 1, ..., N$.

Remark 2. Computation of quantiles and in turn simulation from the limiting distribution of Theorem 4 requires one to supply the incremental distribution $\mathcal{P}$. This $\mathcal{P}$ corresponds to the distribution of the projection $\delta^T\varepsilon_t$ (Condition 1′). Clearly, if one assumes the model errors ε_t to be Gaussian, then $\mathcal{P}$ is also Gaussian. The reason for this requirement is there is no underlying central limit theorem acting in the result for the non-vanishing case. More generally, one may supply a dsitribution $\mathcal{P}$ based on domain knowledge of the specific application, or appeal to the goodness of fit literature, e.g., Kolmogorov Smirnov test in order to empirically identify an incremental distribution

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

based on the residuals.

The following result provides the joint limiting distribution of any finite subset of change point estimates, under the non-vanishing jump size regime. In the following, $\tilde{\tau}_H$ and τ^0_H are subvectors of $\tilde{\tau}$, τ^0, with entries corresponding to indices in H, i.e., $\tilde{\tau}_H = (\tilde{\tau}_j,\ j \in H)$ and analogously define τ^0_H.

Theorem 5. *(Joint distributions for the non-vanishing regime) Suppose Condition 1′, 2, 3, 4 and 5 hold and assume r_T of Condition 3(ii) satisfies $r_T = \{o(1)\underline{\xi}\}/\{(Ns)^{1/2}\log(p \vee T)\}$. Let $H \subseteq \{1, ..., N\}$ be any finite subset of change point indices and $\tilde{\tau}_H = \tilde{\tau}_H(\hat{\tau}_{-H}, \hat{\theta})$ be a subvector of change point estimates as defined in (2.4). Additionally assume the jump size regime is non-vanishing, i.e., $\xi_j \to \xi_{(\infty,j)}$, $0 < \xi_{(\infty,j)} < \infty$, $\forall j \in H$, then, we have,*

$$(\tilde{\tau}_H - \tau^0_H) \ \Rightarrow \ \underset{\zeta \in \mathbb{Z}^{|H|}}{\arg\max} \sum_{j \in H} \mathcal{C}_{(\infty,j)}(\zeta_j), \tag{3.10}$$

where increments z_{tj} and z^*_{tj} of $\mathcal{C}_{(\infty,j)}(\zeta_j)$, are independent for all combinations of each other, over t as well as over $j \in H$. Moreover, the convergence (3.10) is equivalent to,

$$(\tilde{\tau}_H - \tau^0_H) \ \Rightarrow \ \Pi_{j \in H} \underset{\zeta_j \in \mathbb{Z}}{\arg\max}\, \mathcal{C}_{(\infty,j)}(\zeta_j), \tag{3.11}$$

Consequently, $\tilde{\tau}_j$ are also asymptotically independent over $j \in H$.

The product $\Pi_{j \in H}$ represents a joint distribution of dimension $|H|$, wherein the marginal distributions of the components $j \in H$ are those of the multiplicands, and further the components $j \in H$ are pairwise independent.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

This result extends component-wise coverage of (3.9) to simultaneous coverage over any finite subset H of indices of change points, under the non-vanishing regime. From a practical perspective, the most important consequence of this result is that the increments z_{tj} and z^*_{tj} of $\mathcal{C}_{(\infty,j)}(\zeta_j)$, are independent over $j \in H$. This allows the equivalent representation of (3.11), which in turn yields asymptotic independence of $\tilde{\tau}_j$ over $j \in H$. The latter justifies computing component-wise intervals with an adjusted component-wise coverage that maintains the simultaneous coverage as,

$$pr\Big(\tilde{\tau}_j - q^{nv}_{(\alpha,j)} \leq \tau^0_j \leq \tilde{\tau}_j + q^{nv}_{(\alpha,j)},\ \forall j \in H\Big) \to (1-\alpha)^{|H|}, \tag{3.12}$$

where $q^{nv}_{(\alpha,j)}$, $j \in H$ is defined in (3.9). One may adjust componentwise significance level to $\alpha' = \big(1-(1-\alpha)^{1/|H|}\big)$, in order to obtain simultaneous coverage at any desired level $(1-\alpha)$. An analogous simultaneous version of this result in the vanishing case can also be developed similarly.

The above results allow construction of componentwise and simultaneous confidence intervals for any finite subset of τ^0. However, the inherent split distributional behavior leads to the natural question of which of the two distributions is to be employed in a real data setting, since the distinction of a vanishing versus a non-vanishing jump size is unverifiable in practice. The following result addresses this problem and illustrates that if the underlying regime is a vanishing one, then the distribution of the non-vanishing regime undergoes an asymptotic adaptation to it. The result is independent of the above considered model framework. However, it is stated in coherent notations in order to allow its direct applicability. For explicit clarity on the acting

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

limits in this result, we introduce subscripts on all sequences involved. In the following a distribution $\mathcal{P}(\mu, \sigma^2)$ is said to be invariant under scalar addition and multiplication if $a\mathcal{P}(\mu, \sigma^2) + b \sim \mathcal{P}(a\mu + b,\, a^2\sigma^2)$, for any constants $a, b < \infty$.

Theorem 6. *Suppose $\mathcal{P}(\mu, \sigma^2)$ is any distribution, with $E\mathcal{P}(\mu, \sigma^2) = \mu$ and $\mathrm{var}\mathcal{P}(\mu, \sigma^2) = \sigma^2 < \infty$. Let $\mathcal{P}$ be continuously distributed and invariant under scalar addition and multiplication. Furthermore, let ξ_T, σ_T^2 and be any positive sequences in T, such that, as $T \to \infty$, the following limits hold. (i) $\xi_T \to 0$, (ii) $\sigma_T^2 \to \sigma_\infty^2$, where $0 < \sigma_\infty^2 < \infty$. Then, we have,*

$$\sigma_\infty^{-2}\xi_T^2 \arg\max_{\zeta \in Z} \mathcal{C}_\infty(\zeta, \xi_T, \sigma_T^2) \Rightarrow \arg\max_{\zeta \in \mathbb{R}} \big(2W(\zeta) - |\zeta|\big), \text{ as } T \to \infty, \tag{3.13}$$

where $\mathcal{C}_\infty(\zeta, \xi_T, \sigma_T^2)$ and $W(\zeta)$ are as defined in (3.6) and (3.5), respectively.

To view Theorem 6 in context of Theorems 3 and 4, set distribution $\mathcal{P}$ as in Condition 1′ and underlying sequences as follows. (1) ξ_T as jump size ξ_j as in (1.2) for any given $j = 1, ..., N$. (2) $\sigma_T^2 = \xi_j^{-2}\eta_{(j)}^{0T}\Sigma\eta_{(j)}^0$, i.e., the variance defined in Condition 4. Then, all assumptions made in Theorem 6 hold. Specifically, $\xi_T \to 0$ by regime mis-specification and $\sigma_T^2 \to \sigma_{(\infty,j)}^2$, from Condition 4. Theorem 6 now provides a direct connection between intervals for τ_j^0 in the vanishing and non-vanishing regimes. More precisely, for any given $j = 1, ..., N$, one can construct intervals as,

$$\mathrm{CI}(\tilde{\tau}_j) := \big[(\tilde{\tau}_j - q_{(\alpha,j)}^{nv}), (\tilde{\tau}_j + q_{(\alpha,j)}^{nv})\big], \tag{3.14}$$

where $q_{(\alpha,j)}^{nv}$ is the $(1 - \alpha/2)^{th}$ quantile of the distribution in the non-vanishing regime. The interval $\mathrm{CI}(\tilde{\tau}_j)$ is clearly valid with respect to the non-vanishing regime of Theorem

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

4. On the other hand, an interval constructed using Theorem 3 for the vanishing case is of the form $\left[\tilde{\tau}_j \pm \sigma^2_{(\infty,j)}\xi_j^{-2}q^v_\alpha\right]$, where q^v_α is the corresponding quantile. Instead of using this direct formulation, obtaining the required quantile using its finite sample approximation of Theorem 6, one obtains the asymptotically equivalent, $\sigma^2_{(\infty,j)}\xi_j^{-2}q^v_\alpha \asymp q^{nv}_{(\alpha,j)}$. Substituting in the interval yields exactly the construction of CI($\tilde{\tau}_j$). Thus, the results of Theorem 3, Theorem 4 and their relationship in Theorem 6 together imply that doing so yields the desired $(1-\alpha)$ asymptotic coverage, $pr\Big((\tilde{\tau}_j - q^{nv}_{(\alpha,j)}) \leq \tau^0_j \leq (\tilde{\tau}_j + q^{nv}_{(\alpha,j)})\Big) \to (1-\alpha)$, irrespective of whether the underlying regime is vanishing or non-vanishing. In other words, the interval CI($\tilde{\tau}_j$) is regime adaptive.

The next Section resolves the two remaining issues: (i) the availability of preliminary estimates $\hat{\tau}$ and $\hat{\theta}$ satisfying Condition 3, and (ii) positing explicit restrictions on the rate of divergence of the model dimensions (s, p).

4. Construction of Feasible Change Point Estimators

We start by constructing estimates for the mean parameters. For any $\tau = (\tau_1, ..., \tau_N)^T \in \{1, ..., (T-1)\}^N$ satisfying $\tau_{j-1} < \tau_j$, $j = 1, ..., N+1$, consider piece-wise means evaluated on the partitioning of $\{1, ..., T\}$,

$$\bar{x}_{(j)}(\tau) = \frac{1}{(\tau_j - \tau_{j-1})} \sum_{t=\tau_{j-1}+1}^{\tau_j} x_t, \quad j = 1, ..., N+1, \tag{4.1}$$

Next, define its ℓ_1 regularized version for each j=1,...,N+1,

$$\hat{\theta}_{(j)}(\tau) \;=\; \underset{\theta \in \mathbb{R}^p}{\arg\min} \left\|\bar{x}_{(j)}(\tau) - \theta\right\|_2^2 + \lambda_j \|\theta\|_1, \quad \lambda_j > 0. \tag{4.2}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

The ℓ_1 regularization is equivalent to soft-thresholding (Donoho, 1995),

$$\hat{\theta}_{(j)}(\tau) = k_{\lambda_j}\big(x_{(j)}(\tau)\big), \quad j = 1, ..., N+1. \tag{4.3}$$

with $k_\lambda(x) = \text{sign}(x)(|x| - \lambda)_+$, $\lambda > 0$, $x \in \mathbb{R}^p$, wherein $\text{sign}(\cdot)$, $|\cdot|$, and $(\cdot)_+$. Recall the soft-thersholding operator is defined as follows, for $x \in \mathbb{R}$, $(x)_+ = x$, if $x \geq 0$, and $x = 0$ if $x < 0$. are applied component-wise.

Next, we establish that $\hat{\theta}(\tau)$ evaluated with a plug-in change point estimate whose rate is sub-optimal, still satisfies all requirements in Condition 3(ii), assuming that some rate conditions on model parameters hold. Methodologically, it is the construction of these regularized means that provides sufficient regularity despite potential high dimensionality.

Condition 5′ (rates of model parameters): *Assume one of the following conditions written sequentially in order of strength.*

$$(i)\ \left(\frac{\sigma}{\underline{\xi}}\right)\left\{\frac{Ns\log^2(p\vee T)}{T\underline{\ell}}\right\}^{\frac{1}{2}} \leq c_{u1}, \quad (ii)\ \left(\frac{\sigma}{\underline{\xi}}\right)\left\{\frac{Ns\log^{3/2}(p\vee T)}{\sqrt{(T\underline{\ell})}}\right\} \leq c_{u1},$$
$$(iii)\ \left(\frac{\sigma}{\underline{\xi}}\right)\left\{\frac{Ns\log^{3/2}(p\vee T)}{\sqrt{(T\underline{\ell})}}\right\} = o(1), \tag{4.4}$$

wherein $c_{u1} > 0$ is a suitably chosen small constant.

The rate restrictions in Condition 5′ are progressively stronger and viewed together with the following theorem provide important insights on the rate requirements needed for both estimation and inference. This discussion is provided immediately following the next result.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

Theorem 7. *Suppose Condition 1, 2 and 5′(i) hold and let $\hat{\tau}$ be a preliminary change point estimate satisfying,*

$$\hat{N} = N \quad \text{and} \quad \max_{1\le j\le N} |\hat{\tau}_j - \tau_j^0| \le c_u \sigma^2 \underline{\xi}^{-2} N s \log^2(p \vee T), \tag{4.5}$$

for some $c_u > 0$, with probability at least $1-\pi_T$. Then, the estimate $\hat{\tau}$ satisfies Condition 3(i). Let $\psi = \max_j \|\eta_{(j)}^0\|_\infty$ and assume $\psi/\underline{\xi} = O(1)$. Further, assume Condition 5′(ii) and that $T\underline{\ell} \ge \log(p \vee T)$. Then, choosing $\lambda_j = c_u\sigma\big\{\log(p\vee T)/T\big\}^{1/2}$, $j = 1, ..., N+1$, the mean estimates $\hat{\theta}_{(j)}(\hat{\tau})$ satisfy Condition 3(iia) and the bound,

$$\max_{1\le j\le N+1} \|\hat{\theta}_{(j)} - \theta_{(j)}^0\|_2 \le c_u\sigma\Big\{\frac{Ns\log(p\vee T)}{T\underline{\ell}}\Big\}^{\frac{1}{2}}, \tag{4.6}$$

with probability at least $1 - o(1) - \pi_T$. Consequently, $\hat{\theta}_{(j)}(\hat{\tau})$, $j = 1, ..., N+1$ satisfy all requirements of Condition 3(ii).

Theorem 7 shows that the only requirement for the main results of Section 3 to hold is solely the availability of preliminary estimates $\hat{\tau}$ satisfying (4.5), which in turn yield estimates $\hat{\tau}$ and $\hat{\theta}$ that satisfy all requirements of Condition 3. A couple of examples from the literature that can be used to obtain $\hat{\tau}$ are provided in Remark 3.

Existing literature typically assumes restrictions similar to Condition 5′(i), which leads to only sub-optimal rates of esitmation. Upon viewing 5(i) with respect to the jump size, one may observe this requires $\underline{\xi} \ge c\{Ns\log^2 p/T\underline{\ell}\}^{1/2}$, which is almost the detection limit as provided in Liu et al. (2021). In the same context, the tighter Condition 5′(ii) moves slightly away from this detection limit to $\underline{\xi} \ge cNs\log^{3/2} p/\sqrt{(T\underline{\ell})}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

Algorithm 1: Locally refitted estimation of $\tau^0 = (\tau_1^0, ..., \tau_N^0)^T$.

Step 1: Implement any estimator $\hat{\tau} = \big(\hat{\tau}_1, ..., \hat{\tau}_{\hat{N}}\big)^T$ from the literature that satisfies the near optimal bounds (4.5), with probability $1 - o(1)$.

Step 2: Compute mean estimates $\hat{\theta}_{(j)}(\hat{\tau})$, $j = 1, ..., \hat{N} + 1$, and obtain locally refitted change point estimates,

$$\tilde{\tau}_j = \underset{\hat{\tau}_{j-1} < \tau_j < \hat{\tau}_{j+1}}{\arg\min} \; Q_j\big(\tau_j, \hat{\tau}_{-j}, \hat{\theta}\big), \qquad j = 1, ..., \hat{N},$$

(Output): $\tilde{\tau} = \big(\tilde{\tau}_1, ..., \tilde{\tau}_{\hat{N}}\big)^T$.

Upon doing so, one is able to improve estimation precision, and obtain an optimal estimation rate for the change points, despite high dimensionality.

Algorithm 1 now presents the feasible implementation of the methodology, while Corollary 1 summarizes its estimation and inferential properties.

Corollary 1. *Assume that Conditions 1, 2 and 5′(ii) hold, together with* $T\underline{\ell} \geq \log(p \vee T)$ *and that* $\psi/\underline{\xi} = O(1)$. *Then,* $\tilde{\tau}$ *of Algorithm 1 satisfies the component-wise and* ℓ_∞ *bounds of Theorems 1 and 2, respectively. If in addition Conditions 4 and 5′(iii) hold,* $\tilde{\tau}$ *satisfies the component-wise limiting distribution in Theorem 3, under a vanishing jump size, while if Condition 1′ holds,* $\tilde{\tau}$ *satisfies the component-wise limiting distribution in Theorem 4, under a non-vanishing jump size. Finally, it satisfies the joint limiting distributions of Theorem 5 under the same non-vanishing regime and for any finite subset* H *of the change point indices.*

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

The following remark provides an example of a method from the literature that yields the preliminary estimate $\hat{\tau}$.

Remark 3. For the high dimensional mean shift model, the projected CUSUM estimator of Wang and Samworth (2018) provides thus far the sharpest ℓ_∞ rate of estimation available in the literature. Theorem 2 in that paper establishes $\hat{N} = N$ and $\max_{1\leq j\leq N} |\hat{\tau}_j - \tau_j^0| \leq c_u \sigma^2 \underline{\xi}^{-2} \underline{\ell}^{-4} \log(p \vee T)$, with probability at least $1 - o(1)$, for their proposed estimator. Under a fixed number of change points, $N \leq c_u$ and $\underline{\ell} \geq c_u$, this estimator satisfies the requirement (4.5) of the preliminary estimate for our methodology. When N is diverging, (4.5) holds under the relation $\ell^{-4} \leq Ns$. Consequently, this estimator can serve as a theoretically valid preliminary estimate for Algorithm 1. An example of a method that allows subexponential distributions is that of Cho et al. (2016) in a dense framework.

To complete the last piece of this puzzle we consider the problem of implementing the proposed confidence intervals with estimated variance and drift parameters. Recall that the results so far (Theorems 3, 4 and 6) rely on unknown drift $-\xi_j^2$ and variance $\xi_j^{-2}\eta^{0T}\Sigma\eta^0$ parameters, in order to obtain margin of errors and thereby confidence intervals. We leverage the standard practice of employing plug-in estimates to obtain the desired parameter estimates.

Let $\tilde{\tau}$, $\tilde{N}$ and $\hat{\theta}_{(j)}$, $j = 1, ..., N$ be estimates from Algorithm 1. Define $\hat{\eta}_{(j)} =$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

$\hat{\theta}_{(j)} - \hat{\theta}_{(j+1)}$, $j = 1, ..., \tilde{N}$, and jump size and variance estimates as,

$$\tilde{\xi}_j = \|\hat{\eta}_{(j)}\|_2, \quad \text{and} \quad \tilde{\sigma}_j^2 = \tilde{\xi}_j^{-2} \hat{\eta}_{(j)} \tilde{\Sigma} \hat{\eta}_{(j)}, \quad j = 1, ..., \tilde{N} \tag{4.7}$$

Let $\tilde{\Sigma}$ represent the sample covariance of the data $\{x_t\}_{t=1}^T$ centered with the estimated mean parameters partitioned over changes induced by $\tilde{\tau}$, i.e.,

$$\tilde{\Sigma} \;=\; \frac{1}{T} \sum_{j=1}^{\tilde{N}+1} \sum_{t=\tilde{\tau}_{j-1}+1}^{\tilde{\tau}_j} (x_t - \hat{\theta}_{(j)})(x_t - \hat{\theta}_{(j)})^T \tag{4.8}$$

Then under the non-vanishing regime, confidence interval (3.14) can be implemented by obtaining quantiles $\tilde{q}_{\alpha,j}^{nv}$ of the underlying limiting distribution under the above estimated parameters. Confidence intervals under the vanishing regime can be obtained by a direct substitution of the drift and variance estimates. The following discussion and result establishes the validity of this approach.

Consider an event $\mathcal{A}$ on which the following bounds hold,

$$\begin{aligned} \mathcal{A} \;=\; \Big\{ & (i) \; \max_{1 \le j \le N} |\tilde{\tau}_j - \tau_j^0| \le C \underline{\xi}^{-2} \log^2 T, \; (ii) \; \tilde{N} = N \\ & (iii) \; \max_{1 \le j \le N+1} \|\hat{\theta}_{(j)} - \theta_{(j)}^0\|_2 \le C \Big\{ \frac{N s \log(p \vee T)}{T \underline{\ell}} \Big\}^{1/2} \Big\} \end{aligned} \tag{4.9}$$

where C is a finite constant. Note that we have already shown that $P(\mathcal{A}) \to 1$ as part different results of this article, specifically, this holds by an aggregation of Theorem 2 and Theorem **??** and (**??**) in Proof of Theorem 7 of the Supplement. The structure of the following proofs illustrates that the plug-in estimates (4.7) are consistent by proving the property holds on the set $\mathcal{A}$ whose probability converges to one.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

Lemma 1. *Assume the conditions in Corollary 1. Then, the estimates (4.7) are consistent, i.e., we have,*

$$(i)\ \frac{\tilde{\xi}_j^{-2}}{\xi_j^{-2}} \to^p 1, \quad \text{and}, (ii)\ \frac{\tilde{\sigma}_j^2}{\sigma_{(\infty,j)}^2} \to^p 1, \quad as\ T \to \infty, \tag{4.10}$$

for all $j = 1, ..., N$,[1]

The next result establishes that employing these estimates towards construction of the proposed confidence intervals is asymptotically equivalent to those obtained by the unknown parameters.

Corollary 2. *Suppose conditions of Corollary 1 hold and let the regime be that of a non-vanishing jump size. Further, for any given* $0 < \alpha < 1$, *let* $\tilde{q}_\alpha^{nv}$ *be* $(1-\alpha/2)^{th}$ *quantile of the distribution* $\arg\max_\zeta \mathcal{C}\big(\zeta, \tilde{\xi}, \tilde{\sigma}_j^2\big)$ *of (3.6) in the corresponding regime obtained under estimated drift and variance parameters. Then, the confidence interval* $\big[\tilde{\tau}_j \pm \tilde{q}_\alpha^{nv}\big]$ *provides asymptotically nominal coverage at any given* $j = 1, ..., N$, *i.e.,*

$$pr\big(\tilde{\tau}_j - \tilde{q}_\alpha^{nv} \le \tau_j^0 \le \tilde{\tau}_j + \tilde{q}_\alpha^{nv}\big) \to (1-\alpha) \quad for\ any\ given\ j = 1, ..., N$$

Analogous results also hold under the vanishing regime, as well as for the simultaneous intervals obtained in (3.12).

Remark 4. It can be seen from the proof of Lemma 1 that one does not necessarily require employing the mean estimates $\hat{\beta}$ from Algorithm 1 in order to obtain consistent estimates of the drift and variance parameters. Instead, one only requires them to

[1] Here $\to^p$ represents convergence in probability

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

satisfy the condition $\max_{1\leq j\leq N+1}\|\hat{\theta}_{(j)}-\theta^0_{(j)}\|_2 \leq C\Big\{Ns\log(p\vee T)/T\underline{\ell}\Big\}^{1/2}$ assumed in event $\mathcal{A}$. In our numerical results, we employ refitted mean estimates computed as $\tilde{\theta}_{(j)} = \big[\bar{x}_{(j)}(\tilde{\tau})\big]_{\hat{S}_j}$ $j=1,...,N$, wherein $\hat{S}_j = \{k\ \hat{\theta}_{(j)k} \neq 0\}$, $j=1,...,N$ correspond to the estimated sparsity sets, with all remaining indices of these mean estimates set to zero. We adopt these estimates to alleviate finite sample regularization biases in the high dimensional means. It is known that refitted mean estimates preserve the rate of convergence of the regularized version, while reducing finite sample biases, e.g. Belloni et al. (2011) and thus the corresponding validity of the procedure.

5. Numerical Experiments

Based on synthetic data, we illustrate next Algorithm 1 and the inference results developed in Section 3 and summarized in Corollary 1. For the numerical experiments, data are generated as per model (1.1). The mean vectors are set as $\theta^*_{(1)} = \big(1^T_{s\times 1}, 0..., 0\big)^T_{p\times 1}$, $\theta^*_{(2)} = \big(0^T_{s\times 1}, 1^T_{s\times 1}, 0..., 0\big)^T_{p\times 1}$, and $\theta^*_{(3)} = \big(0^T_{2s\times 1}, 1^T_{s\times 1}, 0..., 0\big)^T_{p\times 1}$. These are repeated iteratively depending on N. The matrix Σ is chosen to be $\Sigma_{ij} = \rho^{|i-j|}$, $i,j=1,...,p$ with $\rho = 0.5$. We consider all combinations of $T\in\{450,600,750\}$, $p\in\{50,200,350,500\}$, and $N\in\{2,4\}$. The locations of changes τ^0_j, $j=1,...,N$, are set to N evenly spaced values. We consider both subgaussian and subexponential noise. Specifically, for Scenarios A and B below $\varepsilon_t\in\mathbb{R}^p$, are generated as $\varepsilon_t \sim^{i.i.d} \mathcal{N}(0,\Sigma)$, whereas for Scenarios A′ and B′ we generate $\varepsilon_t = \Sigma^{\frac{1}{2}} w_t$, where each component $w^*_{tj} \sim^{i.i.d}$ Laplace$(0,1)$, $j=1,...,p$, with zero mean and unit variance.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

Scenarios A and A′ reflect an idealized setting. Algorithm 1 is implemented with the true $\tau^0 = (\tau_1^0, ..., \tau_N^0)^T$ supplied in Step 1. This provides evidence towards a numerical proof of principle of inference results supporting Algorithm 1 and also serves as a benchmark for Scenario B.

In **Scenarios B and B′**, all parameters need to be estimated. Step 1 of Algorithm 1 is carried out by one of two methods. First, the projected CUSUM estimator of Wang and Samworth (2018) (WS) is used, as implemented in `R`-package `InspectChangepoint`. Tuning parameters are set to the default values in the package. We denote this method as WS+LR, wherein LR refers to local refitting.

We also implement Step 1 with a second method. The ℓ_0 regularized near optimal estimator of Kaul et al. (2021) (Remark 4.2 in that paper, henceforth KFJS) designed for a single change point is extended via binary segmentation, i.e., recursive application of the method. This is described as Algorithm 3 in Supplement **??**. Tuning parameters are selected as recommended in the article. The overall procedure is denoted as KFJS+BS+LR, wherein BS refers to binary segmentation.

We construct component-wise confidence intervals according to (3.9), with $\tilde{\tau}$ being the output of Algorithm 1. In all cases, we set $\alpha = 0.05$. The critical value $q_\alpha^v = 11.03$ under the vanishing regime is evaluated by using its distribution function in Yao (1987). The quantile $q_{(\alpha,j)}^{nv}$ of the argmax of the two sided random walk is computed by simulating 3000 sample paths from it, whose computational details are provided in Supplement **??**. We assess the validity of the joint distribution in Theorem 5 by

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

computing the simultaneous coverage yielded by these component-wise constructed intervals. Since by construction we have a finite number of change points, Theorem 5 dictates that simultaneous coverage should satisfy $pr\big(\tau_j^0 \in \ CI_j;\ \forall 1 \leq j \leq N\big) \to (1-\alpha)^N$. We set $(1-\alpha)^N = 0.902, 0.814$, for $N = 2, 4$.

Tuning parameters for mean estimates in Step 2: The regularizers λ_j, $j = 1, ..., N+1$ used to obtain soft thresholded mean estimates are tuned via a BIC type criterion. Specifically, we set $\lambda_j = \lambda$, $j = 1, ..., N$, and evaluate $\hat{\theta}^{\lambda}_{(j)}(\hat{\tau})$ for λ in a uniform grid of twenty five values. Upon letting $\hat{S} = \Big\{k \in \{1, ..., p\};\ \cup_{j=1}^{N+1} \hat{\theta}_{(j)k}(\hat{\tau}) \neq 0\Big\}$ we set λ as the minimizer of $BIC(\lambda)$ defined as, $BIC(\lambda) = \sum_{j=1}^{N+1} \sum_{t=\hat{\tau}_{j-1}+1}^{\hat{\tau}_j} \big\|x_t - \hat{\theta}^{\lambda}_{(j)}(\hat{\tau})\big\|_2^2 + |\hat{S}| \log T$.

The following metrics are employed to summarize the results: (1) Hausdorff distance (haus. d.): average over replications of $d_H(\hat{\tau}, \tau^0)$, where

$$d_H(\hat{\tau}, \tau^0) = \max\Big\{ \min_{1 \leq j \leq \hat{N}} d(\tilde{\tau}_j, \tau^0),\ \min_{1 \leq j \leq N} d(\tilde{\tau}, \tau_j^0)\Big\},$$

and $d(\cdot\,, \cdot\,)$ denoting the absolute difference. (2) Standard deviation over replications of the Hausdorff distance (sd). (3) N-match: relative frequency of number of times $\hat{N} = N$. To measure inference performance, for Scenarios A and A$'$ we report, (4) Component-wise coverage for the first change point τ_1^0 (Comp. coverage): relative frequency of the number of times τ_1^0 lies in its confidence interval, obtained for both the vanishing and non-vanishing regime results. (5) Average margin of error (av. ME) for τ_1^0: average over replications of the margin of errors of each confidence interval of the first change point τ_1^0. (6) Simultaneous coverage over all change point parameters (Simul. coverage): relative frequency of the number of times τ_j^0 lies in the corresponding

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

$N=2$, $s=4$		haus.d(sd)	Comp. coverage (av. ME) $(1-\alpha)=0.95$		Simul. Coverage	$N=2$, $s=4$		haus.d(sd)	Comp. coverage (av. ME) $(1-\alpha)=0.95$		Simul. Coverage
T	p		Vanishing	Non-vanishing	$(1-\alpha)^N=0.902$	T	p		Vanishing	Non-vanishing	$(1-\alpha)^N=0.902$
450	50	0.77(1.09)	0.924(2.15)	0.948(2.04)	0.884	600	350	0.77(1.06)	0.962(2.12)	0.968(2.04)	0.898
450	200	0.80(1.14)	0.942(2.14)	0.962(2.04)	0.87	600	500	0.72(0.92)	0.962(2.14)	0.974(2.02)	0.898
450	350	0.71(0.93)	0.958(2.13)	0.982(2.03)	0.902	750	50	0.82(1.18)	0.958(2.19)	0.970(2.03)	0.876
450	500	0.74(1.02)	0.954(2.11)	0.964(2.03)	0.886	750	200	0.86(1.14)	0.952(2.17)	0.968(2.04)	0.872
600	50	0.70(0.95)	0.966(2.17)	0.976(2.05)	0.898	750	350	0.82(1.15)	0.962(2.18)	0.970(2.03)	0.85
600	200	0.72(1.11)	0.962(2.17)	0.968(2.05)	0.892	750	500	0.81(1.07)	0.954(2.17)	0.972(2.04)	0.872

Table 1: Results of Scenario A with $N=2$ based on 500 Monte-Carlo replicates. Coverage metrics rounded to three decimals, all other metrics rounded to two decimals.

confidence interval for all $j=1,...,N$, obtained under the non-vanishing jump size result. In Scenarios B and B', the metrics pertaining to simultaneous coverage are instead conditional versions, i.e., (4)' Component-wise coverage for the first change point τ_1^0 conditioned on $\hat{N}=N$ (Comp. coverage $\big|\hat{N}=N$): relative frequency over those intervals where $\hat{N}=N$ of the number of times τ_1^0 lies in its confidence interval. Analogous conditional versions are also reported for (5) and (6).

All results are based on 500 replicates. Partial results are reported in Tables 1 and 2 below (Scenarios A and B with $N=2$). The remaining results are reported in Tables **??** and **??** (Scenarios A and B with $N=4$), Tables **??** and **??** (Scenarios A′ and B′ with $N=2$) and Tables **??** and **??** (Scenarios A′ and B′ with $N=4$) in Supplement **??**.

The numerical results are in strong agreement with the theoretical ones. The

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

Method: KFJS+BS+LR							Method: WS+LR						
$N=2$, $s=4$		haus.d(sd)	N-match	Comp. coverage (av. ME) $(1-\alpha)=0.95$		Simul. Coverage	$N=2$, $s=4$		haus.d(sd)	N-match	Comp. coverage (av. ME) $(1-\alpha)=0.95$		Simul. Coverage
T	p			Vanishing	Non-vanishing	$(1-\alpha)^N=0.902$	T	p			Vanishing	Non-vanishing	$(1-\alpha)^N=0.902$
450	50	15.81(23.74)	0.68	0.947(2.16)	0.956(2.04)	0.856	450	50	17.52(30.92)	0.68	0.92(2.16)	0.938(2.07)	0.834
450	200	17.59(26.64)	0.69	0.945(2.14)	0.965(2.05)	0.908	450	200	15.62(29.65)	0.72	0.967(2.12)	0.989(2.03)	0.953
450	350	16.93(26.19)	0.7	0.948(2.18)	0.963(2.08)	0.862	450	350	16.82(31.40)	0.73	0.937(2.12)	0.959(2.02)	0.837
450	500	17.57(26.74)	0.69	0.948(2.24)	0.965(2.13)	0.851	450	500	16.98(32.13)	0.73	0.946(2.14)	0.956(2.05)	0.861
600	50	23.76(33.48)	0.64	0.944(2.15)	0.953(2.04)	0.85	600	50	19.68(38.88)	0.69	0.948(2.15)	0.956(2.06)	0.854
600	200	25.37(35.22)	0.65	0.933(2.17)	0.951(2.05)	0.831	600	200	21.78(41.97)	0.71	0.941(2.14)	0.958(2.03)	0.843
600	350	27.44(36.50)	0.63	0.956(2.18)	0.975(2.05)	0.892	600	350	22.19(43.37)	0.71	0.955(2.16)	0.966(2.03)	0.890
600	500	23.21(34.14)	0.67	0.931(2.18)	0.955(2.07)	0.871	600	500	23.29(44.33)	0.72	0.925(2.16)	0.944(2.04)	0.891
750	50	28.09(41.55)	0.65	0.951(2.18)	0.963(2.05)	0.884	750	50	28.59(55.66)	0.67	0.958(2.19)	0.964(2.04)	0.898
750	200	34.06(44.16)	0.61	0.970(2.18)	0.977(2.05)	0.882	750	200	21.88(46.98)	0.73	0.964(2.17)	0.973(2.03)	0.887
750	350	34.11(44.26)	0.61	0.957(2.19)	0.964(2.06)	0.868	750	350	22.16(46.92)	0.74	0.954(2.18)	0.959(2.04)	0.861
750	500	29.51(42.71)	0.66	0.933(2.19)	0.945(2.04)	0.861	750	500	20.78(48.78)	0.79	0.952(2.17)	0.960(2.04)	0.866

Table 2: Results of Scenario B with $N=2$ based on 500 Monte-Carlo replicates. Coverage metrics rounded to three decimals, all other metrics rounded to two decimals.

component-wise coverage in nearly all examined cases and under both regimes is at the nominal level. Simultaneous coverage in Scenarios A and A$'$ for both $N=2,4$ provides precise control at nominal coverage levels ($(1-\alpha)^2=0.902$ and $(1-\alpha)^4=0.814$). Slightly larger deviations of simultaneous coverage from nominal levels are observed in Scenarios B and B$'$, especially for $N=4$. Based on a close examination of the results for individual replicates, the main reason for these deviations is that the selection consistency of the preliminary estimator, i.e., $\Pr(\hat{N}=N)\to 1$, can be quite slow. We consider an additional Scenario C with larger values of T, where as expected, simultaneous coverage moves closer to the nominal level (Supplement **??**. As pointed out by a referee, there are also a couple of cases in Tables 1 and 5 in the Supplement, wherein the Haudorff distance seems marginally out of trend (case: $T=600$, $p=50$).

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

This is however attributable to stochastic fluctuations in the synthetic data, which can be seen by noting that the corresponding standard deviation of the Hausdorff distance is larger; the latter is indicative of additional uncertainty in this observation. The remaining cases provide a discernible trend wherein both the Hausdorff distance, as well as its standard deviation are decreasing (when p decreases or T increases)

6. Application: Smartphone Based Human Activity Recognition

Modern cellphones integrate a host of sensors, including accelerometers, gyroscopes and magnetometers. These sensors obtain measurements of their users daily activities. Human Activity Recognition (HAR) is a field that aims to identify activities of persons based on such sensor information. We consider a data set obtained from smartphone embedded accelerometer and gyroscope measurements available at the repository `https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones`. Controlled experiments were carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (walking, walking upstairs, walking downstairs, sitting, standing, laying) wearing a smartphone on the waist. Using the embedded sensors, measurements were obtained on the 3-axial linear acceleration and 3-axial angular velocity.

The data comprises of $T = 7352$ vectors, of dimension $p = 561$. A detailed description of each collected feature can be found in the repository provided above. Each observed vector is labelled with the activity that the subject performed at the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

Estimated Number of Change Points	Estimated Locations ($\tilde{\tau}$)	Confidence Intervals $(1-\alpha)=0.95$	
		Vanishing	Non-Vanishing
$\tilde{N}=5$	$\tilde{\tau}_1=1228$	[1225.91, 1230.83]	[1226, 1230]
	$\tilde{\tau}_2=2299$	[2297.46, 2300.53]	[2298, 2300]
	$\tilde{\tau}_3=3285$	[3284.74, 3285.25]	[3285, 3285]
	$\tilde{\tau}_4=4570$	[4567.79, 4572.21]	[4568, 4572]
	$\tilde{\tau}_5=5945$	[5944.78, 5945.21]	[5945, 5945]

Table 3: Estimated change points and confidence intervals

Change Point Index (j)	1	2	3	4	5
Estimated Jump Sizes ($\tilde{\xi}_j$)	10.63	15.66	38.23	4.07	10.22
Estimated Asymptotic Variances $\tilde{\sigma}^2_{(\infty,j)}$	21.36	34.20	34.34	3.31	2.02

Table 4: Estimated jump sizes and asymptotic variances

time. Our objective is to perform an unsupervised partitioning of observed vectors over the sampling period, via the change point model (1.1), in order to gauge the predictive power of such measurements in predicting the associated activity.For this purpose, we sort the data set by the associated activity labels, so that model (1.1) becomes applicable, with the true change points (activity transitions) located at $\tau^0 = (1226, 2299, 3285, 4571, 5945)^T$. All observations are then randomized within each activity label in order to eliminate any local temporal artifacts. The method KFJS+BS+LR is used for estimation and inference on the locations of change points. The component-wise coverage set to $(1-\alpha)=0.95$. Results are summarized in Table 3 and Table 4.

We identified all $\tilde{N}=5$ change points with high precision, thus, distinguishing all six tasks undertaken in the experiment. The second, third and the fifth change points

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

are estimated at exactly the true values. Confidence intervals for the third and fifth ones are very narrow under the vanishing regime and degenerate under the non-vanishing one. The results highlight the predictive power of data collected from smartphones in distinguishing ordinary tasks which may seem physically similar (e.g. sitting, standing and laying). These observations have clear potential beneficial applications in fields such as health care and assisted living, but also raise issues of privacy, since they may aid in indirect monitoring of daily activities of the phone users.

Supplementary Materials

All proofs and additional simulation results are provided in the Supplement.

Acknowledgments

The authors thank the Associate Editor and two anonymous referees for their constructive comments. GM was supported in part by NSF grant DMS 2210358.

References

Antoch, J., M. Huskova, and N. Veraverbeke (1995). Change-point problem and bootstrap. *Journal of Nonparametric Statistics 5*(2), 123–144.

Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis 15*(5), 453–472.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics 79*(4), 551–563.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*(1), 47–78.

Belloni, A., V. Chernozhukov, and L. Wang (2011). Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika 98*(4), 791–806.

Bhattacharjee, M., M. Banerjee, and G. Michailidis (2017). Common change point estimation in panel data from the least squares and maximum likelihood viewpoints. *arXiv preprint arXiv:1708.05836*.

Bhattacharjee, M., M. Banerjee, and G. Michailidis (2020). Change point estimation in a dynamic stochastic block model. *The Journal of Machine Learning Research 21*(1), 4330–4388.

Bickel, P. J. (1982). On Adaptive Estimation. *The Annals of Statistics 10*(3), 647 – 671.

Chen, L., W. Wang, and W. B. Wu (2022). Inference of breakpoints in high-dimensional time series. *Journal of the American Statistical Association 117*(540), 1951–1963.

Cho, H. et al. (2016). Change-point detection in panel data via double cusum statistic. *Electronic Journal of Statistics 10*(2), 2000–2038.

Cho, H. and P. Fryzlewicz (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77*(2), 475–507.

Cho, H. and C. Kirch (2022). Bootstrap confidence intervals for multiple change points based on moving sum procedures. *Computational Statistics & Data Analysis 175*, 107552.

Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE transactions on information theory 41*(3),

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

613–627.

Eichinger, B. and C. Kirch (2018a). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli 24*(1), 526 – 564.

Eichinger, B. and C. Kirch (2018b). A mosum procedure for the estimation of multiple random change points. *Bernoulli 24*(1), 526–564.

Frick, K., A. Munk, and H. Sieling (2014). Multiscale change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(3), 495–580.

Frisén, M. (2008). *Financial surveillance*, Volume 71. John Wiley & Sons.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics 42*(6), 2243–2281.

Harchaoui, Z. and C. Lévy-Leduc (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association 105*(492), 1480–1493.

Jirak, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics 43*(6), 2451–2483.

Kaul, A., S. B. Fotopoulos, V. K. Jandhyala, A. Safikhani, et al. (2021). Inference on the change point under a high dimensional sparse mean shift. *Electronic Journal of Statistics 15*(1), 71–134.

Kaul, A., H. Zhang, K. Tsampourakis, and G. Michailidis (2023). Inference on the change point under a high dimensional covariance shift. *Journal of Machine Learning Research 24*(168), 1–68.

Koepcke, L., G. Ashida, and J. Kretzberg (2016). Single and multiple change point detection in spike trains: comparison of different cusum methods. *Frontiers in systems neuroscience 10*, 1–18.

Liu, H., C. Gao, and R. J. Samworth (2021). Minimax rates in sparse, high-dimensional change point detection.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

The Annals of Statistics 49(2), 1081 – 1112.

Ng, W. L., S. Pan, and C. Y. Yau (2022). Bootstrap inference for multiple change-points in time series. *Econometric Theory 38*(4), 752–792.

Niu, Y. S., N. Hao, and H. Zhang (2016). Multiple change-point detection: A selective overview. *Statistical Science 31*(4), 611–623.

Verzelen, N., M. Fromont, M. Lerasle, and P. Reynaud-Bouret (2023). Optimal change-point detection and localization. *The Annals of Statistics 51*(4), 1586 – 1610.

Wang, D., Z. Zhao, K. Z. Lin, and R. Willett (2021). Statistically and computationally efficient change point localization in regression settings. *Journal of Machine Learning Research 22*(248), 1–46.

Wang, T. and R. J. Samworth (2018). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(1), 57–83.

Xu, H., D. Wang, Z. Zhao, and Y. Yu (2022). Change point inference in high-dimensional regression models under temporal dependence. *arXiv preprint arXiv:2207.12453*.

Yao, Y.-C. (1987). Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *The Annals of Statistics 15*(3), 1321–1328.

Yu, Y. (2020). A review on minimax rates in change point detection and localisation. *arXiv preprint arXiv:2011.01857*.

Department of Mathematics and Statistics

Washington State University, Pullman, WA 99164

E-mail: abhishek.kaul@wsu.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0323

Department of Statistics and Data Science

University of California, Los Angeles, Los Anegeles, CA 90095-1554

E-mail: gmichail@ucla.edu