Robust Narrowest Significance Pursuit: Inference for multiple change-points in the median

Piotr Fryzlewicz*

October 14, 2022

Abstract

We propose Robust Narrowest Significance Pursuit (RNSP), a methodology for detecting localized regions in data sequences which each must contain a change-point in the median, at a prescribed global significance level. RNSP works by fitting the postulated constant model over many regions of the data using a new sign-multiresolution sup-norm-type loss, and greedily identifying the shortest intervals on which the constancy is significantly violated. By working with the signs of the data around fitted model candidates, RNSP is able to work under minimal assumptions, requiring only sign-symmetry and serial independence of the signs of the true residuals. In particular, it permits their heterogeneity and arbitrarily heavy tails. The intervals of significance returned by RNSP have a finite-sample character, are unconditional in nature and do not rely on any assumptions on the true signal. Code implementing RNSP is available at `https://github.com/pfryz/nsp`.

Keywords: confidence intervals, structural breaks, post-selection inference, post-inference selection, narrowest-over-threshold.

*Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, UK. Email: p.fryzlewicz@lse.ac.uk.

1 Introduction

The problem of uncertainty quantification for possibly multiple parameter changes in time- or space-ordered data is motivated by the practical question of whether any suspected changes reflect real structural changes in the underlying stochastic model, or are due to random fluctuations. Approaches to this problem include confidence sets associated with simultaneous multiscale change-point estimation (Frick et al., 2014; Pein et al., 2017), post-selection inference (Hyun et al., 2018, 2021; Jewell et al., 2022; Duy et al., 2020), inference without selection and post-inference selection via Narrowest Significance Pursuit (Fryzlewicz, 2021), asymptotic confidence intervals conditional on the estimated change-point locations (Bai and Perron, 1998; Eichinger and Kirch, 2018; Cho and Kirch, 2022), False Discovery Rate (Hao et al., 2013; Li and Munk, 2016; Cheng et al., 2020), and Bayesian inference (Lavielle and Lebarbier, 2001; Fearnhead, 2006). These approaches go beyond mere change-point detection and offer statistical significance statements regarding the existence and locations of change-points in the statistical model underlying the data.

However, the vast majority of the existing works, with only a handful of exceptions which we review below, do so under strong assumptions on the innovations entering the model, such as their homogeneity over time, and/or finite moments up to a certain order, and/or a specific known distribution or distributional family. This is a limitation, because change-point modelling and inference for data that do not necessarily fall into these categories is of interest in various domains, including bioinformatics (Muggeo and Adelfio, 2011; Arlot and Celisse, 2011), biophysics (Pein et al., 2017), cancer medicine (Kim et al., 2000), economics (Bai and Perron, 2003) and finance (Oomen, 2019), amongst others.

In this paper, we are concerned with the following problem: given a sequence of noisy independent observations, automatically determine localized regions of the data which each must contain a change-point in the median, at a prescribed global significance level. The methodology we introduce, referred to as Robust Narrowest Significance Pursuit (RNSP), achieves this for the piecewise-constant median model, capturing changes in the level of the median. It does so with practically no distributional assumptions on the data, other

than serial independence of the signs of the true residuals and their sign-symmetry, a weak requirement which is immaterial for continuous distributions (as all, even non-symmetric, continuous distributions are sign-symmetric). In contrast to the existing literature, RNSP requires no knowledge of the distribution on the part of the user, and permits arbitrarily heavy tails, heterogeneity over time and/or lack of symmetry. In addition, RNSP is able to handle distributions that are continuous, continuous with mass points, or discrete, where these properties may also vary over the signal. The execution of RNSP does not rely on having an estimate of the number of change-points. Critical values needed by RNSP do not depend on the noise distribution, and can be accurately approximated analytically. RNSP explicitly targets the shortest possible intervals of significance, and its algorithmic construction ensures exact finite-sample coverage guarantees.

To situate RNSP in the context of the existing literature, we now review the few existing approaches that permit uncertainty statements for multiple change-point locations under the assumption of possible heterogeneity in the distribution of the noise.

H-SMUCE (Pein et al., 2017) is a change-point detector in the heterogeneous Gaussian piecewise-constant model $Y_i = f_i + \sigma_i \varepsilon_i$, where f_i is a piecewise-constant signal, ε_i are i.i.d. $N(0,1)$ variables, and σ_i can only change when f_i does. In Section A.1 of their work, the authors provide an algorithmic construction of confidence intervals for the locations of the change-points in f_i. This requires having an estimate of the number of change-points, and involves screening the data for short intervals over which a constant signal fit is unsuitable and they must therefore contain change-points. Crucially, this algorithmic construction relies on the knowledge of scale-dependent critical values (for measuring the unsuitability of a locally constant fit), which are not available analytically but only by simulation, and therefore the method would not extend automatically to unknown noise distributions (as the analyst needs to know what distribution to sample from). In Section 5, we show that H-SMUCE suffers from inflated type I error rates in the sense that the thus-constructed confidence intervals, in the examples of Gaussian models shown, do not all contain at least one true change-point each in more than $100(1-\alpha)\%$ of the cases, contrary to what this algorithmic construction sets out to do. H-SMUCE is also prone to failing if the model is

mis-specified, e.g. if the distribution of the data has a mass point (which is unsurprising in view of its assumption of Gaussianity).

Multiscale Quantile Segmentation (Vanegas et al., 2022, MQS) is a procedure for detecting possibly multiple changes in a given piecewise-constant quantile of the input data sequence, which includes the median as a special case. Like Pein et al. (2017), is it also an extension of SMUCE (Frick et al., 2014). MQS estimates the number of change-points in the quantile function as the minimum among those candidate fits for which the empirical residuals pass a certain multiscale test at level α, where the empirical residuals are defined as binary exceedance sequences of the data over the level defined by each candidate fit. Working with such binary exceedance sequences means that MQS makes no distributional assumptions on the data other than their serial independence. Like SMUCE and H-SMUCE, MQS then defines a confidence set around the estimated signals as a set of all feasible (in the same sense) signal fits at level α. This then enables the conceptual and algorithmic construction of asymptotic simultaneous confidence intervals for the change-point locations, which are guaranteed to (each) contain a change-point with probability $1 - \alpha + o(1)$. Chen et al. (2014) provide a critique of SMUCE from the inferential point of view, which also applies to H-SMUCE and MQS. In Section 5, we illustrate that MQS (as implemented in the R package `mqs`) is not calibrated well, which leads to its not meeting its coverage guarantees. MQS is also by far the slowest method discussed here, and its implementation is random in the sense of potentially (and frequently in practice) returning different output each time it is run on the same dataset.

In contrast to MQS, RNSP is immune to calibration errors, in the sense of its guarantees being wholly based on exact inequalities. This is reflected in its performance shown in Section 5. It is also significantly faster than MQS and returns deterministic output.

The (non-robust) Narrowest Significance Pursuit (NSP; Fryzlewicz, 2021) is able to handle heterogeneous data in a variety of models, including the piecewise-constant signal plus independent noise model. However, NSP requires that the noise, if heterogeneous, is within the domain of attraction of the normal distribution and is symmetric, neither of which is assumed in RNSP. The self-normalised statistic used in NSP includes a term resembling

an estimate of the local variance of the noise which, however, is only unbiased under the null hypothesis of no change-point being present locally. The fact that the same term overestimates the variance under the alternative hypothesis, reduces the power of the detection statistic, which leads to typically long intervals of significance. We illustrate this issue with the self-normalised non-robust NSP for heterogeneous data in Section 5 and show that RNSP offers a significant improvement.

The Bayesian approach of Fearnhead (2006) provides an inference framework for independent observations drawn from a $f(\cdot|\theta_j)$ density within the jth segment, with θ_j possibly vector-valued, which permits some types of heterogeneity. The distribution family f is assumed known, and a prior need to be specified for the θ_j parameters. While self-contained and not overly computationally intensive, this approach suffers from the fact that the estimation of the number of change-points is a difficult statistical problem and therefore so is choosing a good prior for this quantity; for the same reason, it is a choice that can significantly affect the inferential conclusions.

Bai and Perron (1998) and Bai and Perron (2003), working with least-squares estimation of multiple change-points in regression models under possible heterogeneity of the errors, describe a procedure for computing confidence intervals conditional on detection. For an unknown distribution of the errors, the limiting distribution of each estimated change-point location converges to an appropriate functional of the Wiener process only under the assumption that the corresponding break size goes to zero with the sample size. The asymptotic validity of the resulting confidence interval relies on the consistency of the estimators of the unknown quantities involved (such as the local variance of the innovations or the break size); it is therefore a large-sample, asymptotic construction. Crucially, it does not take into the account the uncertainty associated with detection, which can be considerable especially for the more difficult problems (for an example, see the "US ex-post real interest rate" case study in Bai and Perron (2003), where there is genuine uncertainty between models with 2 and 3 change-points; we revisit this example in Section 6.1). By contrast, RNSP produces intervals of significant change in the median that are not conditional on detection, have a finite-sample nature and are valid regardless of the size of the breaks.

The paper is organised as follows. Section 2 motivates RNSP and sets out its general algorithmic framework. Section 3 describes how RNSP measures the local deviation from model constancy and gives finite-sample theoretical performance guarantees for RNSP. Section 4 quantifies the large-sample behaviour of RNSP. Section 5 contains numerical examples and comparisons. Section 6 includes two examples showing the practical usefulness of RNSP. Section 7 concludes with a brief discussion. Software implementing RNSP is available at `https://github.com/pfryz/nsp`.

2 Motivation and review of RNSP algorithmic setting

2.1 RNSP: context and modus operandi

RNSP discovers regions in the data in which the *median* departs from constancy, at a certain global significance level. This is in contrast to NSP (Fryzlewicz, 2021), which targets the *mean*. RNSP does not make moment assumptions about the data, and therefore the median is a natural measure of data centrality. We now review the components of the algorithmic framework that are shared between NSP and RNSP, with the generic measurement of local deviation from the constant model as one of its building blocks. In Section 3, we introduce the particular way in which local deviation from the constant model is measured in RNSP, which is appropriate for the median and hence fundamentally different from NSP.

RNSP operates in the signal plus noise model

$$Y_t = f_t + Z_t, \quad t = 1, \ldots, T, \tag{1}$$

where the variables $\{Z_t\}_{t=1}^T$ and the signal $\{f_t\}_{t=1}^T$ satisfy, respectively, the assumptions below. Define $\mathrm{sign}(x) = \mathbb{I}(x > 0) - \mathbb{I}(x < 0)$, where $\mathbb{I}(\cdot)$ is the indicator function.

Assumption 2.1 *(a) The variables* $\{\mathrm{sign}(Z_t)\}_{t=1}^T$ *are mutually independent.*

(b) $\forall\, t$, $M(Z_t) = 0$, *where* M *is the median operator. (If the median is non-unique, we require* $0 \in M(Z_t)$*.)*

(c) The variables $\{Z_t\}_{t=1}^T$ *are sign-symmetric, i.e.* $P(Z_t > 0) = P(Z_t < 0)$, $\forall\, t$.

$\{Z_t\}_{t=1}^T$ do not have to be identically distributed, and can have arbitrary mass atoms, or none, as long as their distributions satisfy (b) and (c) in Assumption 2.1. We do not impose moment assumptions. Any zero-median continuous distribution (even one with an asymmetric density function) is sign-symmetric. The distribution(s) of $\{Z_t\}_{t=1}^T$ can be unknown to the analyst. The requirement of the independence of $\{\text{sign}(Z_t)\}_{t=1}^T$ is weaker than that of the independence of $\{Z_t\}_{t=1}^T$ itself: e.g. if Z_t is a (G)ARCH process driven by symmetric, independent innovations, then $\text{sign}(Z_t)$ is serially independent, while Z_t is not.

Assumption 2.2 *In (1),* f_t *is a piecewise-constant vector with an unknown number* N *and locations* $0 = \eta_0 < \eta_1 < \ldots < \eta_N < \eta_{N+1} = T$ *of change-points. (The location* η_j *is a change-point if* $f_{\eta_j} \neq f_{\eta_j+1}$*.)*

We restrict RNSP to the piecewise-constant model in Assumption 2.2 as for robustness, RNSP works with the signs of the empirical residuals. Due to the non-linearity of the sign operator, RNSP has to manually try all possible constant model fits on each interval, to be able to select one that gives the minimum deviation, as this is required for its coverage guarantees. This is a problem of dimensionality one in the piecewise-constant model, which makes it computationally straightforward and fast (details are in Section 3.2).

RNSP achieves the high level of generality in terms of the permitted distributions of Z_t stated in Assumption 2.1 thanks to its use of the sign transformation. The use of 0 in $\text{sign}(x)$ is critical for RNSP's objective to provide exact finite-sample coverage guarantees also for discrete distributions and continuous distributions with mass points. We illustrate the importance of this point in Section 3.2.

2.2 (R)NSP: generic algorithm

The generic algorithmic framework underlying both RNSP and the non-robust NSP in Fryzlewicz (2021) is the same and is based on recursively searching for the shortest sub-samples in the data with globally significant deviations from the baseline model. In this

section, we introduce this shared generic framework. In the following sections, we show how RNSP diverges from NSP through its use of a robust measure of deviation from the baseline model, suitable for the very broad class of distributions specified in Assumption 2.1.

We start with a pseudocode definition of the RNSP algorithm, in the form of a recursively defined function RNSP. In its arguments, $[s, e]$ is the current interval under consideration and at the start of the procedure, we have $[s, e] = [1, T]$; Y (of length T) is as in the model formula (1); M is the minimum guaranteed number of sub-intervals of $[s, e]$ drawn (unless the number of all sub-intervals of $[s, e]$ is less than M, in which case drawing M sub-intervals would mean repetition); λ_α is the threshold corresponding to the global significance level α (typical values for α would be 0.05 or 0.1) and τ_L (respectively τ_R) is a functional parameter used to specify the maximum extent of overlap of the left (respectively right) child interval of $[s, e]$ with respect to the region of significance identified within $[s, e]$, if any. The no-overlap case would correspond to $\tau_L = \tau_R \equiv 0$. In each recursive call on a generic interval $[s, e]$, RNSP adds to the set $\mathcal{S}$ any globally significant local regions (intervals) of the data identified within $[s, e]$ on which Y is deemed to depart significantly (at global level α) from the baseline constant model. We provide more details underneath the pseudocode below. In the remainder of the paper, the subscript $_{[s,e]}$ relates to a constant indexed by the interval $[s, e]$, whose value will be clear from the context.

1: **function** RNSP(s, e, Y, M, λ_α, τ_L, τ_R)
2: **if** $e - s < 1$ **then**
3: STOP
4: **end if**
5: **if** $M \geq \frac{1}{2}(e - s + 1)(e - s)$ **then**
6: $M := \frac{1}{2}(e - s + 1)(e - s)$
7: draw all intervals $[s_m, e_m] \subseteq [s, s + 1, \ldots, e]$, $m = 1, \ldots, M$, s.t. $e_m - s_m \geq 1$
8: **else**
9: draw a representative (see description below) sample of intervals $[s_m, e_m] \subseteq [s, s+1, \ldots, e]$, $m = 1, \ldots, M$, s.t. $e_m - s_m \geq 1$
10: **end if**

11: **for** $m \leftarrow 1, \ldots, M$ **do**

12: $D_{[s_m,e_m]} :=$ DeviationFromConstantModel(s_m, e_m, Y)

13: **end for**

14: $\mathcal{M}_0 := \arg\min_m\{e_m - s_m \ : \ m = 1, \ldots, M;\ D_{[s_m,e_m]} > \lambda_\alpha\}$

15: **if** $|\mathcal{M}_0| = 0$ **then**

16: STOP

17: **end if**

18: $m_0 :=$AnyOf$(\arg\max_m\{D_{[s_m,e_m]} \ : \ m \in \mathcal{M}_0\})$

19: $[\tilde{s}, \tilde{e}] :=$ShortestSignificantSubinterval$(s_{m_0}, e_{m_0}, Y, M, \lambda_\alpha)$

20: add $[\tilde{s}, \tilde{e}]$ to the set $\mathcal{S}$ of significant intervals

21: RNSP$(s, \tilde{s} + \tau_L(\tilde{s}, \tilde{e}, Y), Y, M, \lambda_\alpha, \tau_L, \tau_R)$

22: RNSP$(\tilde{e} - \tau_R(\tilde{s}, \tilde{e}, Y), e, Y, M, \lambda_\alpha, \tau_L, \tau_R)$

23: **end function**

The RNSP algorithm is launched by the pair of calls: $\mathcal{S} := \emptyset$; RNSP$(1, T, Y, M, \lambda_\alpha, \tau_L, \tau_R)$. On completion, the output of RNSP is in the variable $\mathcal{S}$. We now comment on the RNSP function line by line. In lines 2–4, execution is terminated for intervals that are too short. In lines 5–10, a check is performed to see if M is at least as large as the number of all sub-intervals of $[s, e]$. If so, then M is adjusted accordingly, and all sub-intervals are stored in $\{[s_m, e_m]\}_{m=1}^{M}$. Otherwise, a sample of M sub-intervals $[s_m, e_m] \subseteq [s, e]$ is drawn in which s_m and e_m are all possible pairs from an (approximately) equispaced grid on $[s, e]$ which permits at least M such sub-intervals (deterministic sampling). While it would in principle be possible to pre-draw the intervals $[s_m, e_m]$ prior to launching the RNSP procedure, rather than drawing them recursively on each current interval $[s, e]$ as it done in the algorithm, the recursive way of drawing the intervals encourages, by construction, shorter detection intervals and is therefore preferred.

In lines 11–13, each sub-interval $[s_m, e_m]$ is checked to see to what extent the response on this sub-interval (denoted by $Y_{s_m:e_m}$) deviates from the baseline constant model. This core step of the RNSP algorithm will be described in more detail in Section 3.

In line 14, the measures of deviation obtained in line 12 are tested against threshold λ_α,

chosen to guaranteed global significance level α. How to choose λ_α is independent of the distribution of Z_t if it is continuous, and there is also a simple distribution-independent choice of λ_α for discrete distributions and continuous distributions with probability masses; see Section 3.3. The shortest sub-interval(s) $[s_m, e_m]$ for which the test rejects the baseline model at global level α are collected in set $\mathcal{M}_0$. In lines 15–17, if $\mathcal{M}_0$ is empty, then the procedure decides that it has not found regions of significant deviations from the constant model on $[s, e]$, and stops on this interval as a consequence. Otherwise, in line 18, the procedure continues by choosing the sub-interval, from among the shortest significant ones, on which the deviation from the baseline constant model has been the largest. The chosen interval is denoted by $[s_{m_0}, e_{m_0}]$.

In line 19, $[s_{m_0}, e_{m_0}]$ is searched for its shortest significant sub-interval, i.e. the shortest sub-interval on which the hypothesis of the baseline model is rejected locally at a global level α. Such a sub-interval certainly exists, as $[s_{m_0}, e_{m_0}]$ itself has this property. The structure of this search again follows the workflow of the RNSP procedure; more specifically, it proceeds by executing lines 2–18 of RNSP, but with s_{m_0}, e_{m_0} in place of s, e. The chosen interval is denoted by $[\tilde{s}, \tilde{e}]$. This second-stage search is important to RNSP's pursuit to produce short intervals: indeed, if the sample of intervals $[s_m, e_m]$ contained insufficiently short intervals (perhaps because an insufficiently large M was chosen), then, without the second-stage search in line 19, the intervals of significance returned by RNSP might be unsatisfactorily long. Therefore, the second-stage search in line 19 can be seen as a guard against a small M, or in other words against an insufficiently fine original grid of interval endpoints. In line 20, the selected interval $[\tilde{s}, \tilde{e}]$ is added to the output set $\mathcal{S}$.

In lines 21–22, RNSP is executed recursively to the left and to the right of the detected interval $[\tilde{s}, \tilde{e}]$. However, we optionally allow for some overlap with $[\tilde{s}, \tilde{e}]$. The overlap, if present, is a function of $[\tilde{s}, \tilde{e}]$ and, if it involves detection of the location of a change-point within $[\tilde{s}, \tilde{e}]$, then it is also a function of Y. An example of the relevance of this is given in Section 6.1.

3 Robust NSP: measuring deviation from the constant model

3.1 Principles and building blocks of the new deviation measure

The main structure of the DeviationFromConstantModel(s_m, e_m, Y) operation is as follows: (1) Fit the best, in the sense described precisely later, constant model to $Y_{s_m:e_m}$. (2) Examine the signs of the empirical residuals from this fit. If their distribution is deemed to contain a change-point (which indicates that the constant model fit is unsatisfactory on $[s_m, e_m]$ and therefore the model contains a change-point on that interval), the value returned by DeviationFromConstantModel(s_m, e_m, Y) should be large; otherwise small.

For RNSP to be able to control, globally and non-trivially, the probability of spurious detection, the deviation measure it uses must meet the two desiderata below.

Desideratum A. If there is no change-point on $[s_m, e_m]$, then the value of DeviationFromConstantModel(s_m, e_m, Y) should be bounded from above by the same deviation measure for the true model on $[s_m, e_m]$ (the latter is an oracular quantity, unknown to the analyst), and this in turn should be bounded from above, uniformly over all $[s_m, e_m]$, by a random variable involving the signs of the true residuals Z only, such that either its distribution is known or its quantiles can easily be bounded, sharply, from above.

Desideratum B. The deviation measure on $[s_m, e_m]$ cannot be made smaller by proposing unrealistic constant model fits on that interval; otherwise it would be easy to force non-detection on any interval. This is an important desired property as our deviation measure will need to 'try' all possible constant model fits and choose one for which the deviation measure is the smallest, to ensure that Desideratum A (the part relating to boundedness from above by the deviation measure for the true model) is met.

A key ingredient of our measure of deviation, which helps it achieve the above desired properties, is a multiresolution sup-norm introduced below, used on the signs of the input rather than in the original data domain. The basic building block of the multiresolution sup-norm is a scaled partial sum statistic, defined for an arbitrary input sequence $\{x_t\}_{t=1}^T$ by $U_{s,e}(x) = (e-s+1)^{-1/2} \sum_{t=s}^{e} x_t$. We define the multiresolution sup-norm (Nemirovski,

1986; Li, 2016) of an input vector x (of length T) with respect to the interval set $\mathcal{I}$ as $\|x\|_{\mathcal{I}} = \max_{[s,e]\in\mathcal{I}} |U_{s,e}(x)|$. The set $\mathcal{I}$ used in RNSP contains intervals at a range of scales and locations. A canonical example of a suitable interval set $\mathcal{I}$ is the set $\mathcal{I}^a$ of all subintervals of $[1, T]$. We will use $\mathcal{I}^a$ in defining the largest acceptable global probability of spurious detection. However, for computational reasons, DeviationFromConstantModel will use a smaller interval set (we give the details below). This will not affect the exactness of our coverage guarantees, because, naturally, if $\mathcal{J} \subseteq \mathcal{I}$, then $\|x\|_{\mathcal{J}} \leq \|x\|_{\mathcal{I}}$. We also define the restriction of $\mathcal{I}$ to an arbitrary interval $[s, e]$ as $\mathcal{I}_{[s,e]} = \{[u, v] \subseteq [s, e] \ : \ [u, v] \in \mathcal{I}\}$. Note the trivial inequality

$$\|x_{s:e}\|_{\mathcal{I}^a_{[s,e]}} \leq \|x\|_{\mathcal{I}^a} \tag{2}$$

for any $[s, e] \subseteq [1, T]$.

When the above multiresolution sup-norm is applied to the signs of the input, as is done in this work, rather than the original input, we refer to it as the sign-multiresolution sup-norm. When applied to the empirical residuals from a candidate constant fit on an interval, it can be viewed as a simple robust multiscale data fidelity measure of the given candidate fit on all time scales up to the length of the interval. The sign-multiresolution sup-norm of the empirical residuals from a postulated constant model fit is naturally large for proposed model fits if they are unrealistic, which penalises them, thereby ensuring that Desideratum B above is met.

3.2 Deviation measure: definition and properties

We now define the deviation measure $D_{[s_m,e_m]} := \text{DeviationFromConstantModel}(s_m, e_m, Y)$, returned by the DeviationFromConstantModel function from line 12 of the RNSP algorithm, which, in line with Desideratum A, satisfies the property that if there is no change-point on the interval $[s_m, e_m]$, then it is guaranteed that

$$D_{[s_m,e_m]} \leq \|\text{sign}(Z_{s_m:e_m})\|_{\mathcal{I}^a_{[s_m,e_m]}}. \tag{3}$$

The discussion below assumes that there is no change-point in $[s_m, e_m]$. For the true constant signal $f_{s_m:e_m}$, denote $f_{[s_m,e_m]} := f_{s_m} = \ldots = f_{e_m}$. A key observation is that there are only at most $2(s_m - e_m) + 3$ different possible constants $\tilde{f}_{[s_m,e_m]}$ leading to different sequences $\{\text{sign}(Y_t - \tilde{f}_{[s_m,e_m]})\}_{t=s_m}^{e_m}$. To see this, sort the values of $Y_{s_m:e_m}$ in non-decreasing order to create $Y_{(1)}, Y_{(2)}, \ldots, Y_{(e_m - s_m + 1)}$. Take candidate constants $\tilde{f}^{\{j\}}_{[s_m,e_m]}$, $j = 1, \ldots, 2(s_m - e_m)+3$, defined as follows.

$$\begin{aligned}
\tilde{f}^{\{1\}}_{[s_m,e_m]} &< Y_{(1)} \quad \text{(but otherwise arbitrary)} \\
\tilde{f}^{\{2\}}_{[s_m,e_m]} &= Y_{(1)} \\
\tilde{f}^{\{3\}}_{[s_m,e_m]} &= \frac{1}{2}(Y_{(1)} + Y_{(2)}) \\
\tilde{f}^{\{4\}}_{[s_m,e_m]} &= Y_{(2)} \\
\tilde{f}^{\{5\}}_{[s_m,e_m]} &= \frac{1}{2}(Y_{(2)} + Y_{(3)}) \\
&\vdots \\
\tilde{f}^{\{2(e_m - s_m)+2\}}_{[s_m,e_m]} &= Y_{(e_m - s_m + 1)} \\
\tilde{f}^{\{2(e_m - s_m)+3\}}_{[s_m,e_m]} &> Y_{(e_m - s_m + 1)} \quad \text{(but otherwise arbitrary).}
\end{aligned} \tag{4}$$

We have the following simple result.

Proposition 3.1 *Assume no change-point in* $[s_m, e_m]$ *and denote* $f_{[s_m,e_m]} := f_{s_m} = \ldots = f_{e_m}$. *Let the constants* $\tilde{f}^{\{j\}}_{[s_m,e_m]}$, $j = 1, \ldots, 2(s_m - e_m) + 3$ *be defined as in (4). There exists a* $j_0 \in \{1, \ldots, 2(s_m - e_m) + 3\}$ *such that*

$$\{\text{sign}(Y_t - f_{[s_m,e_m]})\}_{t=s_m}^{e_m} = \{\text{sign}(Y_t - \tilde{f}^{\{j_0\}}_{[s_m,e_m]})\}_{t=s_m}^{e_m}. \tag{5}$$

Proof. If there is a j_1 such that $f_{[s_m,e_m]} = Y_{(j_1)}$, then choose $j_0 = 2j_1$ so that $\tilde{f}^{\{j_0\}}_{[s_m,e_m]} = Y_{(j_1)} = f_{[s_m,e_m]}$ and (5) is trivially satisfied. Otherwise, if there is a j_1 such that $f_{[s_m,e_m]} \in (Y_{(j_1)}, Y_{(j_1+1)})$, then choose $j_0 = 2j_1 + 1$ so that

$$\begin{aligned}
\{\text{sign}(Y_t - f_{[s_m,e_m]})\}_{t=s_m}^{e_m} &= \{\text{sign}(Y_t - (Y_{(j_1)} + Y_{(j_1+1)})/2)\}_{t=s_m}^{e_m} \\
&= \{\text{sign}(Y_t - \tilde{f}^{\{j_0\}}_{[s_m,e_m]})\}_{t=s_m}^{e_m},
\end{aligned}$$

and (5) is satisfied. Otherwise, it must be that either $f_{[s_m,e_m]} < Y_{(1)}$ or $f_{[s_m,e_m]} > Y_{(e_m-s_m+1)}$; in the former case take $j_0 = 1$ and in the latter, $j_0 = 2(e_m - s_m) + 3$ and (5) is satisfied by a similar argument. □

We now define our measure of deviation $D_{[s_m,e_m]}$, and prove its key property as a corollary to Proposition 3.1.

Definition 3.1 *Let the constants $\tilde{f}^{\{j\}}_{[s_m,e_m]}$, $j = 1, \ldots, 2(s_m - e_m) + 3$ be defined as in (4). We define*

$$D_{[s_m,e_m]} := \min_{j\in\{1,\ldots,2(s_m-e_m)+3\}} \|\mathrm{sign}(Y_{s_m:e_m} - \tilde{f}^{\{j\}}_{[s_m,e_m]})\|_{\mathcal{I}^a_{[s_m,e_m]}}. \tag{6}$$

$D_{[s_m,e_m]}$ tries all possible baseline constant model fits on $[s_m, e_m]$ and chooses the one for which the sign-multiresolution norm of the residuals from the fit is the smallest. Choosing the fit that minimises the sign-multiresolution fit is essential for guaranteeing coverage properties, as we will see below. If there is a change-point on $[s_m, e_m]$, the hope is that even the minimum sign-multiresoltion norm fit as returned by $D_{[s_m,e_m]}$ is large enough for RNSP to indicate a change-point on that interval. This last point is investigated in Section 4.

Corollary 3.1 *For any interval $[s_m, e_m]$ on which there is no change-point, we have*

$$D_{[s_m,e_m]} \leq \|\mathrm{sign}(Z_{s_m:e_m})\|_{\mathcal{I}^a_{[s_m,e_m]}}. \tag{7}$$

In other words, the deviation measure defined in (6) satisfies the desired property (3).

Proof. Let the index j_0 be as in the statement of Proposition 3.1. We have

$$\begin{aligned} D_{[s_m,e_m]} &= \min_{j\in\{1,\ldots,2(s_m-e_m)+3\}} \|\mathrm{sign}(Y_{s_m:e_m} - \tilde{f}^{\{j\}}_{[s_m,e_m]})\|_{\mathcal{I}^a_{[s_m,e_m]}} \\ &\leq \|\mathrm{sign}(Y_{s_m:e_m} - \tilde{f}^{\{j_0\}}_{[s_m,e_m]})\|_{\mathcal{I}^a_{[s_m,e_m]}} = \|\mathrm{sign}(Y_{s_m:e_m} - f_{[s_m,e_m]})\|_{\mathcal{I}^a_{[s_m,e_m]}} \\ &= \|\mathrm{sign}(Z_{s_m:e_m})\|_{\mathcal{I}^a_{[s_m,e_m]}}. \end{aligned}$$

□

This leads to the following guarantee for the RNSP algorithm.

Theorem 3.1 *Let $\mathcal{S} = \{S_1, \ldots, S_R\}$ be the set of intervals returned by the RNSP algorithm. We have $P\left(\exists\ i = 1, \ldots, R\ \forall\ j = 1, \ldots, N\ \ [\eta_j, \eta_j + 1] \not\subseteq S_i\right) \leq P(\|\mathrm{sign}(Z)\|_{\mathcal{I}^a} > \lambda_\alpha)$.*

Proof. On the set $\|\mathrm{sign}(Z)\|_{\mathcal{I}^a} \leq \lambda_\alpha$, each interval S_i must contain a change-point as if it did not, then by Corollary 3.1 and inequality (2), we would have to have

$$D_{S_i} \leq \|\mathrm{sign}(Z)\|_{\mathcal{I}^a} \leq \lambda_\alpha. \tag{8}$$

However, the fact that S_i was returned by RNSP means, by line 14 of the RNSP algorithm, that $D_{S_i} > \lambda_\alpha$, which contradicts (8). This completes the proof. □

Theorem 3.1 should be read as follows. Let $\alpha = P(\|\mathrm{sign}(Z)\|_{\mathcal{I}^a} > \lambda_\alpha)$. For a set of intervals returned by RNSP, we are guaranteed, with probability of at least $1 - \alpha$, that there is at least one change-point in each of these intervals. Therefore, $\mathcal{S} = \{S_1, \ldots, S_R\}$ can be interpreted as an automatically chosen set of *regions (intervals) of significance* in the data. In the no-change-point case ($N = 0$), the probability of obtaining one of more intervals of significance ($R \geq 1$) is bounded from above by $P(\|\mathrm{sign}(Z)\|_{\mathcal{I}^a} > \lambda_\alpha)$. Theorem 3.1 is of a finite-sample character and holds exactly and for any sample size.

We emphasise that Theorem 3.1 does not promise to detect all the change-points, or to do so asymptotically as the sample size gets larger: this would be impossible without assumptions on the strength of the change-points (involving spacing between neighbouring change-points and the sizes of the jumps). This aspect of RNSP is investigated in Section 4. Instead, Theorem 3.1 promises that every interval of significance returned by RNSP must contain at least one change-point each, with a certain global probability adjustable by the user.

The intervals of significance returned by RNSP have an "unconditional confidence interval" interpretation: they are not conditional on any prior detection event, but indicate regions in the data each of which must unconditionally contain at least one change in the median of Y_t, with a global probability of at least $1 - \alpha$. Therefore, as in NSP (Fryzlewicz, 2021), RNSP can be viewed as performing "inference without selection" (where "inference" refers

to producing the RNSP intervals of significance and "selection" to the estimation of change-point locations, absent from RNSP). This viewpoint also enables "post-inference selection" or "in-inference selection" if the exact change-point locations (if any) are to be estimated within the RNSP intervals of significance after or during the execution of RNSP.

3.2.1 Deviation measure: discussion

Method of computation. (6) needs to be computed by manually trying out all candidate constants $\tilde{f}^{\{j\}}_{[s_m,e_m]}$. While this may appear expensive, it is the only option, as (a) trying only 'realistic' constant model fits would require an additional global parameter describing what it means for a local constant fit to be realistic, and (b) carrying out the fitting in an ad hoc way, e.g. by only fitting the empirical median of $Y_{s_m:e_m}$ (rather than each of the constants $\tilde{f}^{\{j\}}_{[s_m,e_m]}$), would violate the desired property (3) and therefore not be able to lead to exact coverage guarantees.

Achieving computational savings without affecting coverage guarantees. The operation of trying each constant $\tilde{f}^{\{j\}}_{[s_m,e_m]}$ in (6) is fast, but in order to accelerate it further, we introduce the two computational savings below, which do not increase $D_{[s_m,e_m]}$ and therefore respect the inequality (7) and hence also our coverage guarantees in Theorem 3.1.

Reducing the set $\mathcal{I}^a_{[s_m,e_m]}$. To accelerate the computation of (6), we replace the set $\mathcal{I}^a_{[s_m,e_m]}$ in $D_{[s_m,e_m]}$ with the set $\mathcal{I}^{lr}_{[s_m,e_m]} := \mathcal{I}^l_{[s_m,e_m]} \cup \mathcal{I}^r_{[s_m,e_m]}$ (with l and r standing for left and right, respectively), where $\mathcal{I}^l_{[s_m,e_m]} = \{[s_m, s_m+1], [s_m, s_m+2], \ldots, [s_m, e_m]\}$ and $\mathcal{I}^r_{[s_m,e_m]} = \{[s_m, e_m], [s_m+1, e_m], \ldots, [e_m-1, e_m]\}$. This reduces the cardinality of the set of intervals included in $D_{[s_m,e_m]}$ from $O((e_m - s_m)^2)$ to $O(e_m - s_m)$. As $\mathcal{I}^{lr}_{[s_m,e_m]} \subseteq \mathcal{I}^a_{[s_m,e_m]}$ and hence $\|\cdot\|_{\mathcal{I}^{lr}_{[s_m,e_m]}} \leq \|\cdot\|_{\mathcal{I}^a_{[s_m,e_m]}}$, the results of Corollary 3.1 and Theorem 3.1 remain unchanged for the thus-reduced $D_{[s_m,e_m]}$. On the other hand, $\mathcal{I}^{lr}_{[s_m,e_m]}$ has been defined in this particular way so as not to compromise detection power in the piecewise-constant median model. To see this, consider the following illustrative example. Suppose $Y_t = f_t$ (noiseless case) and $f_t = 0$ for $t = 1, \ldots, 50$ and $f_t = 1$ for $t = 51, \ldots, 100$. On $[s_m, e_m] = [1, 100]$, the baseline constant signal level fitted is $\tilde{f}_{[1,100]} = 1/2$ and we have $\text{sign}(Y_t - \tilde{f}_{[1,100]}) = -1$ for $t = 1, \ldots, 50$; $\text{sign}(Y_t - \tilde{f}_{[1,100]}) = 1$ for $t = 51, \ldots, 100$. In this setting, the two multiresoltion

sup-norms: $\|\text{sign}(Y_t - \tilde{f}_{[1,100]})\|_{\mathcal{I}^{lr}_{[1,100]}}$ and $\|\text{sign}(Y_t - \tilde{f}_{[1,100]})\|_{\mathcal{I}^{a}_{[1,100]}}$ are identical, equal to $\sqrt{50}$ and achieve this value for the intervals $[1, 50]$ and $[51, 100]$, members of both $\mathcal{I}^{a}_{[1,100]}$ and $\mathcal{I}^{lr}_{[1,100]}$. This simple example illustrates the wider phenomenon that if there is a single change-point in the median on a generic interval $[s_m, e_m]$ under consideration, then in the noiseless case the multiresolution norm over the set $\mathcal{I}^{a}_{[s_m,e_m]}$ is maximised at one of the "left" or "right" intervals in $\mathcal{I}^{lr}_{[s_m,e_m]}$, and we are happy to sacrifice potential negligible differences in the noisy case in exchange for the substantial computational savings.

Limiting interval lengths. In practice, the analyst may not be interested in excessively long RNSP intervals of significance, or in other words in change-points so weak that only a very long interval around them is required to be able to detect them. With this in mind, our software at `https://github.com/pfryz/nsp` provides an option of limiting the length of intervals considered in the sense that $D_{[s_m,e_m]}$ is able to automatically return zero if $e_m - s_m > L$ for a user-specified maximum length L.

Unsuitability of CUSUM or similar contrasts in deviation measure. We note that it would be impossible to replace the multiresolution sup-norm in $D_{[s_m,e_m]}$ by, for example, the CUSUM statistic or a similar contrast measure such as that described in Ellenberger et al. (2021). Consider, for example, the hypothetical definition of a deviation measure as follows: $D^{inv}_{[s_m,e_m]} := \min_{j\in\{1,\ldots,2(s_m-e_m)+3\}} |\text{CUSUM}\{\text{sign}(Y_{s_m:e_m} - \tilde{f}^{\{j\}}_{[s_m,e_m]})\}|$, where "inv" stands for invalid. The definition of the CUSUM statistic is well-established, see e.g. Fryzlewicz (2014) for details. This is an invalid definition as $\text{sign}(Y_{s_m:e_m} - \tilde{f}^{\{1\}}_{[s_m,e_m]})$ is a vector of ones, and the CUSUM statistic returns zero for constant vectors, so $D^{inv}_{[s_m,e_m]}$ would not be able to offer detection under any circumstances as it would always equal zero. This is an example of a construction that violates Desideratum B.

Importance of zero in sign function. For valid coverage guarantees in the presence of mass points in Z_t (or if the distribution of Z_t is discrete), it is crucial for the sign function defined in Section 2.1 to return zero if its argument is zero. To illustrate this point, define $\text{sign}^{inv}(x) = \mathbb{I}(x \geq 0) - \mathbb{I}(x < 0)$, where "inv" stands for "invalid", and consider the trivial case $Z_t \equiv 0$. For no-change-point input data $Y_{s_m:e_m} = (f_{[s_m,e_m]}, \ldots, f_{[s_m,e_m]})$, there are only three different constants $\tilde{f}^{\{j\}}_{[s_m,e_m]}$ (see definition (4)). With $\text{sign}^{inv}(x)$ in place of $\text{sign}(x)$,

this would lead to

$$\begin{aligned}
\{\text{sign}^{inv}(Y_t - \tilde{f}^{\{1\}}_{[s_m,e_m]})\}_{t=s_m}^{e_m} &= (1,\ldots,1) \\
\{\text{sign}^{inv}(Y_t - \tilde{f}^{\{2\}}_{[s_m,e_m]})\}_{t=s_m}^{e_m} &= (1,\ldots,1) \\
\{\text{sign}^{inv}(Y_t - \tilde{f}^{\{3\}}_{[s_m,e_m]})\}_{t=s_m}^{e_m} &= (-1,\ldots,-1)
\end{aligned}$$

and therefore we would have $D_{[s_m,e_m]} = \sqrt{e_m - s_m + 1}$, the largest value $D_{[s_m,e_m]}$ can possibly take, which would therefore have to lead to the (spurious) designation of $[s_m, e_m]$ as containing a change-point, for $e_m - s_m$ suitably large (and for a reasonable value of λ_α). Naturally, the analogous argument would also apply if $\text{sign}^{inv}(0) = -1$ rather than 1. By contrast, note that the use of the (correct) sign function leads to $\{\text{sign}(Y_t - \tilde{f}^{\{2\}}_{[s_m,e_m]})\}_{t=s_m}^{e_m} = (0,\ldots,0)$, which yields $D_{[s_m,e_m]} = 0$ and therefore the (correct) designation of $[s_m, e_m]$ as not containing a change-point.

Importance of placing $\tilde{f}^{\{j\}}_{[s_m,e_m]}$ at data points. For the same reason, it is important that the set of test levels $\tilde{f}^{\{j\}}_{[s_m,e_m]}$ (definition (4)) includes those placed at the data points themselves (i.e. those indexed by even values of j in definition (4)). Indeed, continuing the example directly above, if we were to exclude the constant $\tilde{f}^{\{2\}}_{[s_m,e_m]}$ from the list of test levels considered, we would then have $D_{[s_m,e_m]} = \sqrt{e_m - s_m + 1}$ even with the use of the correct function sign (and not only with sign^{inv}), which would again lead to spurious detection.

Importance of placing $\tilde{f}^{\{j\}}_{[s_m,e_m]}$ in between sorted data points. It is equally important that the test levels $\tilde{f}^{\{j\}}_{[s_m,e_m]}$ should include those placed in between the sorted data points (i.e. those indexed by odd values of j in definition (4)). This can be see e.g. by considering Z_t such that $P(Z_t = 1) = P(Z_t = -1) = 1/2$; for brevity, we omit the full discussion.

3.3 Evaluation and bounds for $\|\text{sign}(Z)\|_{\mathcal{I}^a}$

To make Theorem 3.1 operational, we need to obtain an understanding of the distribution of $\|\text{sign}(Z)\|_{\mathcal{I}^a}$ so we are able to choose λ_α such that $P(\|\text{sign}(Z)\|_{\mathcal{I}^a} > \lambda_\alpha) = \alpha$ (or approximately so) for a desired global significance level α.

Initially we consider Z_t such that $P(\text{sign}(Z_t) = 1) = P(\text{sign}(Z_t) = -1) = 1/2$, i.e.

$P(\text{sign}(Z_t) = 0) = 0$ for all t (the general case $P(\text{sign}(Z_t) = 0) \geq 0$ is covered in the next paragraph). One simple, entirely finite-sample way of determining the distribution of $\|\text{sign}(Z)\|_{\mathcal{I}^a}$ for any finite T is by simulation; this would only need to be done once for every T and the quantiles stored for fast access. Another approach is asymptotic and proceeds as follows. From Theorem 1.1 in Kabluchko and Wang (2014) (which applies to the symmetric Bernoulli distribution as explained in Section 1.5.1 of that work), we have

$$\lim_{T\to\infty} P(\|\text{sign}(Z)\|_{\mathcal{I}^a} > a_T + \tau/a_T) = 1 - \exp(-2\Lambda \exp(-\tau)), \tag{9}$$

where $a_T = \{2\log(T\log^{-1/2} T)\}^{1/2}$ and Λ is a constant. As the theoretical calculation of Λ in Kabluchko and Wang (2014) unfortunately contains an error, we use simulation over a range of values of T and τ to determine a suitable value of Λ as 0.274. The practical choice of the significance threshold λ_α then proceed as follows: (a) fix α to the desired level, for example 0.05 or 0.1; (b) obtain the value of τ by equating $1 - \exp(-2\Lambda \exp(-\tau)) = \alpha$; (c) obtain $\lambda_\alpha = a_T + \tau/a_T$. While this approach is asymptotic in nature (note the limit as $T \to \infty$ in (9)), we observe that the finite-sample agreement of $P(\|\text{sign}(Z)\|_{\mathcal{I}^a}$ with its limit in (9) is excellent even for small sample sizes. If the user chooses to pursue this route of obtaining λ_α, this will be the only asymptotic component of RNSP.

Suppose now that $P(\text{sign}(Z_t) = 0) = \rho_t \geq 0$; note that the sign-symmetry Assumption 2.1(c) implies $P(\text{sign}(Z_t) = 1) = P(\text{sign}(Z_t) = -1) = (1 - \rho_t)/2$. Construct the variable $\tilde{Z}_t = Z_t \mid Z_t \neq 0$. As $P(\text{sign}(\tilde{Z}_t) = 1) = P(\text{sign}(\tilde{Z}_t) = -1) = 1/2$, the limiting statement (9) applies to $\text{sign}(\tilde{Z}_t)$. However, we have the double inequality

$$\|\text{sign}(Z)\|_{\mathcal{I}^a} \leq \|\text{sign}(\tilde{Z})\|_{\mathcal{I}_I^a} \leq \|\text{sign}(\tilde{Z})\|_{\mathcal{I}^a}, \tag{10}$$

with $I = [1, 2, \ldots, T_1]$, where $T_1 = |\{t \in [1, \ldots, T] \ : \ Z_t \neq 0\}|$. The first inequality in (10) holds because every constituent partial sum of $\|\text{sign}(Z)\|_{\mathcal{I}^a}$ has a corresponding larger or equal in magnitude partial sum in $\|\text{sign}(\tilde{Z})\|_{\mathcal{I}_I^a}$ constructed by removing the zeros from its numerator and decreasing (or not increasing) its denominator as it contains fewer (or the same number of) terms. As an illustrative example, suppose the sequence of $\text{sign}(Z_t)$ starts

$-1, 0, 1, 1$. The absolute partial sum $|-1+0+1+1|/\sqrt{4}$, a constituent of $\|\text{sign}(Z)\|_{\mathcal{I}^a}$, is majorised by the absolute partial sum $|-1+1+1|/\sqrt{3}$, a constituent of $\|\text{sign}(\tilde{Z})\|_{\mathcal{I}_I^a}$, where the latter sum has been constructed by removing the 0 from $-1, 0, 1, 1$ and adjusting for the number of terms (now 3 instead of 4). The second inequality in (10) holds simply because $T_1 \leq T$. The implication of (10) is that $\|\text{sign}(Z)\|_{\mathcal{I}^a}$ for $\rho_t \geq 0$ is majorised by $\|\text{sign}(Z)\|_{\mathcal{I}^a}$ for $\rho_t = 0$, the case handled by (9). Therefore, the threshold λ_α obtained as a consequence of (9) can also be applied in the general case $\rho_t \geq 0$.

4 Detection consistency and lengths of RNSP intervals

This section shows the large-sample consistency of RNSP in detecting change-points, and the rates at which the lengths of the RNSP intervals contract (relative to T), as T increases. To simplify our technical arguments, we consider a version of the RNSP algorithm that considers all subintervals of $[1, T]$. Our focus on i.i.d. Z_t's in this section is due to our ability to rely on the Dvoretzky-Kiefer-Wolfowitz inequality at a crucial point of the proof. We focus on continuously-distributed Z_t's as this results in notationally much less involved arguments regarding the minimum signal strength required. We first introduce some essential notation, and then state our assumption and the result. For each change-point η_j, define

$$\Delta_j = \min\{P\{Z_t \in (-|f_{\eta_j} - f_{\eta_j+1}|/2, 0)\}, P\{Z_t \in (0, |f_{\eta_j} - f_{\eta_j+1}|/2)\}, \tag{11}$$

$$\bar{d}_j = \bar{d}_j(\lambda, \lambda_\alpha) = \left\lceil \left(\frac{2\lambda + \lambda_\alpha}{2\Delta_j} \right)^2 + 1 \right\rceil. \tag{12}$$

In addition, for any process V_t, define $\epsilon_t^V(w) = \mathbb{I}(V_t - w > 0) - P(V_t - w > 0)$.

Assumption 4.1 *(i) The variables $\{Z_t\}_{t=1}^T$ are mutually i.i.d.*

(ii) The distribution of Z_1 is continuous.

(iii) With the notation $\eta_0 = 0$ and $\eta_{N+1} = T$, we have $\eta_{j+1} - \eta_j \geq 2\bar{d}_{j+1} + 2\bar{d}_j - 2$ for $j = 1, \ldots, N-1$, and $\eta_1 - \eta_0 \geq 2\bar{d}_1 - 1$ as well as $\eta_{N+1} - \eta_N \geq 2\bar{d}_N - 1$.

Our first result below is of a finite-sample nature.

Theorem 4.1 *Let Assumption 4.1 hold. On the set defined by the intersection of the events* $\|\mathrm{sign}(Z)\|_{\mathcal{I}^a} \leq \lambda_\alpha$ *and* $\max_{s,e} \sup_w \left| \frac{1}{\sqrt{e-s+1}} \sum_{t=s}^{e} \epsilon_t^Z(w) \right| \leq \lambda$*, a version of the RNSP algorithm that considers all intervals, executed with no overlaps and with threshold* λ_α*, returns exactly* N *intervals of significance* $[s_1, e_1] < \ldots < [s_N, e_N]$ *such that* $\eta_j \in [s_j, e_j - 1]$ *and* $e_j - s_j + 1 \leq 2\bar{d}_j$ *for* $j = 1, \ldots, N$*.*

Theorem 4.1 leads to the following corollary giving a large-sample consistency result.

Corollary 4.1 *Let the assumptions of Theorem 4.1 hold. Let* $\lambda_\alpha = (1+\delta)\{2 \log T\}^{1/2}$ *and* $\lambda = (1+\delta)\log^{1/2} T$*, for any* $\delta > 0$*. Let* $\mathcal{S}$ *denote the set of intervals of significance* $[s_1, e_1] < [s_2, e_2] < \ldots$ *returned by RNSP algorithm that considers all intervals, executed with no overlaps and with threshold* λ_α*. Let* $\mathcal{A} = \{|\mathcal{S}| = N \quad \wedge \quad \forall\, j = 1, \ldots, N \quad \eta_j \in [s_j, e_j - 1] \ \wedge \ e_j - s_j + 1 \leq 2\bar{d}_j\}$*. We have* $P(\mathcal{A}) \to 1$ *as* $T \to \infty$*.*

This is the only large-sample result of the paper; the others are of a finite-sample character. We briefly comment on what the result of Corollary 4.1 means for the minimum signal strength Assumption 4.1(iii), and for the localisation rates of the RNSP algorithm in detecting the change-points. If the distribution of Z_t does not vary with T (this section already assumes that it does not vary with t), and if the jump sizes $|f_{\eta_j} - f_{\eta_{j-1}}|$ are bounded from below by a positive constant independent of j and T, then Δ_j (formula (11)) is also bounded from below by a positive constant independent of j and T. By formula (12), the assumptions on λ, λ_α in Corollary 4.1 then imply $\bar{d}_j = \Theta(\log T)$ (where Θ should be read "of the exact order"). Assumption 4.1(iii) then requires that the spacings between the change-points be at least of order $\log T$. Corollary 4.1 states that the length of each RNSP interval of significance, $e_j - s_j + 1$, is, with global probability approaching 1, at most of order $\log T$. These minimum-spacing assumptions and the implied lengths of the localisation intervals are near-optimal and the same as those in the non-robust literature, see e.g. Theorem 1 in Baranowski et al. (2019) and Corollary 4 in Fryzlewicz (2021), as well as the associated discussions. However, the results of this section also permit $\Delta_j \to 0$ with T, which will, naturally, affect the above minimum-spacing requirements and localisation rates as stipulated by formulae (11) and (12) and Assumption 4.1(iii).

model name	sample path execution in R
Plain Gauss	`rnorm(100)`
Plain Gauss Long	`rnorm(1000)`
Plain Poisson	`as.numeric(rpois(200, 1))`
Heterogeneous Gauss	`c(rep(1, 100), rep(8, 50), rep(1, 100)) * rnorm(250)`
Symmetric Bernoulli	`as.numeric(rbinom(200, 1, .5))`
Plain Cauchy	`rcauchy(100, 0)`
Mix 1	`sample(3, size=300, replace=TRUE, prob=c(.35, .3, .35)) -> xx` `xx[xx != 2] <- rnorm(sum(xx !=2 ))`
Mix 2	`rpois(200, 5)+rnorm(200)/30`

Table 1: Null models for the comparative simulation study in Section 5.

5 Numerical illustrations

In this section, we demonstrate numerically that the guarantee offered by Theorem 3.1 holds for RNSP in practice over a variety of homogeneous and heterogeneous models for which the variables Z_t satisfy Assumption 2.1. We also investigate the circumstances under which similar guarantees are not offered by H-SMUCE (Pein et al., 2017), MQS (Vanegas et al., 2022) or the self-normalised version of NSP (SN-NSP), suitable for heterogeneous data (Fryzlewicz, 2021). In this section, we use the acronyms RNSP and SN-NSP to denote the versions of these respective procedures with no interval overlaps, i.e. $\tau_L = \tau_R = 0$. Later in this section, we introduce notation for versions with overlaps. Both RNSP and SN-NSP use $M = 1000$ intervals, the default setting. For H-SMUCE, the function call we use is `stepR::stepFit(x, alpha=0.1, family="hsmuce", confband=TRUE)`. The `type` parameter in MQS specifies the loss function for their final estimate with multiscale constraints; we denote by MQS-R the result of `mqs::mqse(x, alpha=0.1, conf=TRUE, type="runs")` and by MQS-K the result of `mqs::mqse(x, alpha=0.1, conf=TRUE, type="koenker")`. We use the following package versions: `mqs` v1.0, `stepR` v2.1-1, `nsp` v1.0.0.

We begin with null models, by which we mean models (1) for which f_t is constant throughout, i.e. $N = 0$. For null models, Theorem 3.1 promises that RNSP at level α returns no intervals of significance with probability at least $1 - \alpha$. In this section, we use $\alpha = 0.1$. There are similar parameters in H-SMUCE, MQS and SN-NSP, and they are also set to 0.1. The null models used are listed in Table 1.

Table 2 shows the null model results. RNSP keeps the nominal size well across all the

model	RNSP	H-SMUCE	MQS-R	MQS-K	SN-NSP
Plain Gauss	100	98	66	68	100
Plain Gauss Long	100	96	68	57	100
Plain Poisson	98	0	94	94	2
Heterogeneous Gauss	99	99	57	57	91
Symmetric Bernoulli	95	0	82	76	34
Plain Cauchy	100	100	59	60	100
Mix 1	100	0	63	65	99
Mix 2	98	88	64	57	100

Table 2: Numbers of times, out of 100 simulated sample paths of each null model, that the respective method indicates no intervals of significance.

models considered, returning no intervals of significance at least 95% of the time in all situations. The lowest (but still correct) coverage of 95% is achieved for the Symmetric Bernoulli model: this is because that distribution admits only two values $(-1, 1)$, and therefore the single level $\tilde{f}^{\{3\}}_{[s_m,e_m]} = 0$ (see Section 3.2) always achieves the minimum in $D_{[s_m,e_m]}$. As there are no other candidate test levels $\tilde{f}^{\{j\}}_{[s_m,e_m]}$ to choose from when minimising $\|\mathrm{sign}(Y_{s_m:e_m} - \tilde{f}^{\{j\}}_{[s_m,e_m]})\|_{\mathcal{I}^a_{[s_m,e_m]}}$, this results in $D_{[s_m,e_m]}$ being relatively large, and therefore RNSP keeps its 90% coverage promise for this model correctly, but the least generously out of the models considered. H-SMUCE behaves correctly for the three Gaussian models, but fails for the discrete distributions and model Mix 1, which contains mass points. It is unexpectedly successful in the Plain Cauchy model, but this is perhaps because it has very limited detection power in the Cauchy model with change-points (more on this model below). It also underperforms slightly for model Mix 2, which is continuous (and within the domain of attraction of the Gaussian distribution) but multimodal. MQS, with both `type` options, appears poorly calibrated and does not achieve its asymptotic promise of estimating no change-points in at least 90% of the cases: the corresponding empirical proportions are around 60% or so for most of the models, the only exception being Symmetric Bernoulli (82%, 76%) and Plain Poisson (correct coverage). The fact that it achieves similar coverage (of around 60%) for both Plain Gauss and Plain Gauss Long means that it is not merely the question of its asymptotics not kicking in for the shorter signals (as it underperforms similarly for Plain Gauss Long, for which $T = 1000$). SN-NSP fails for the discrete distributions, which is a consequence of the (asymptotically guaranteed) closeness of the self-normalised deviation measure to the appropriate functional of the Wiener process not

model name	sample path execution in R
Blocks	`blocks + 10*rnorm(2048)`; `blocks` defined in Fryzlewicz (2014)
Cauchy	`c(rcauchy(100, 1), rcauchy(100, 2), rcauchy(100, 1))`
Bursts	`(c(rep(1, 200), rep(3, 80), rep(1, 200), rep(3, 80), rep(1, 200), rep(4, 40)) * rnorm(800))^2`
Poisson	`as.numeric(rpois(350, c(rep(1, 50), rep(4, 50), rep(10, 50), rep(2, 200)))`

Table 3: Non-null models for the comparative simulation study in Section 5.

kicking in in these instances (due to the relatively small sample sizes).

We now discuss performance for signals with change-points ($N > 0$). Table 3 defines our models. Theorem 3.1 promises that any intervals of significance returned by RNSP at levels α are such that, with probability at least $1-\alpha$, they each contain at least one true change-point. In addition to RNSP, H-SMUCE, MQS-R, MQS-K and SN-NSP, we also test versions of RNSP and SN-NSP with the following overlap functions:

$$\begin{aligned} \tau_L(\tilde{s},\tilde{e}) &= \lfloor(\tilde{s}+\tilde{e})/2\rfloor - \tilde{s}, \\ \tau_R(\tilde{s},\tilde{e}) &= \lfloor(\tilde{s}+\tilde{e})/2\rfloor + 1 - \tilde{e}. \end{aligned} \tag{13}$$

This setting means that upon detecting a generic interval of significance $[\tilde{s},\tilde{e}]$ within $[s,e]$, the RNSP and SN-NSP algorithms continue on the left interval $[s,\lfloor(\tilde{s}+\tilde{e})/2\rfloor]$ and the right interval $[\lfloor(\tilde{s}+\tilde{e})/2\rfloor+1,e]$ (recall that the no-overlap case results uses the left interval $[s,\tilde{s}]$ and the right interval $[\tilde{e},e]$). We denote the versions of the two procedures with the overlaps as above by RNSP-O and SN-NSP-O, respectively. As before, we set $\alpha = 0.1$ for all methods tested.

For each model and method tested, we wish to evaluate the following aspects: the empirical coverage (i.e. whether at least $(1-\alpha)100\%$ of the simulated sample paths are such that any intervals of significance returned contain at least one true change-point each); if any intervals are returned, the proportion of those that are genuine (i.e. the proportion of those intervals returned that contain at least one true change-point); the number of genuine intervals (i.e. the number of those intervals returned that contain at least one true change-point); and the average length of genuine intervals (i.e. the average length of those intervals returned that contain at least one true change-point).

Table 4 shows the results. H-SMUCE does not perform well in any scenario, not even in the

model	attribute	RNSP	RNSP-O	H-SMUCE	MQS-R	MQS-K	SN-NSP	SN-NSP-O
Blocks	coverage	100	100	26	38	38	100	100
	prop. gen. int.	1	1	0.81	0.89	0.89	1	1
	no. gen. int.	6.15	8.17	4.84	8.78	8.78	4.93	6.12
	av. gen. int. len.	90.86	99.92	94.73	52.21	52.21	160.19	182.19
Cauchy	coverage	100	100	100	60	54	100	99
	prop. gen. int.	1	1	1	0.69	0.62	1	0.98
	no. gen. int.	0.56	0.73	0.01	1.26	1.09	0.27	0.29
	av. gen. int. len.	119.25	125.53	170	66.26	72.52	167.65	167.96
Bursts	coverage	100	100	43	47	53	100	100
	prop. gen. int.	1	1	0.78	0.78	0.84	1	1
	no. gen. int.	3.01	4.46	3.13	3.67	3.87	4.14	5.31
	av. gen. int. len.	99.1	108.43	94.88	63.88	69.69	110.81	109.03
Poisson	coverage	100	100	0	89	81	0	0
	prop. gen. int.	1	1	0.1	0.94	0.91	0.08	0.08
	no. gen. int.	2.91	3	1.57	2.75	2.68	0.89	1.11
	av. gen. int. len.	37.13	37.71	19.41	23.49	23.27	35.66	34.91

Table 4: Results for each model+method combination: "coverage" is the number of times, out of 100 simulated sample paths, that the respective model+method combination did not return a spurious interval of significance; "prop. gen. int." is the average (over 100 simulated sample paths) proportion of genuine intervals out of all intervals returned, if any (if none are returned, the corresponding 0/0 ratio is ignored in the average); "no. gen. int." is the average (over 100 sample paths) number of genuine intervals returned; "av. gen. int. len." is the average (over 100 sample paths) length of a genuine interval returned in the respective model+method combination.

Blocks model, an instance of the homogeneous Gaussian model, a simple sub-class of the heterogeneous Gaussian model class for which it was specifically designed (where it achieves the coverage of 26, well short of the expected 90). Its coverage of 100 in the Cauchy model is an artefact of the fact that it does not achieve almost any detections over the 100 simulated sample paths (so there are also no spurious detections).

As in the null models, MQS also frequently produces spurious intervals in signals with change-points: the coverage figures for MQS are well short of 90% in all models tested. Both MQS versions output relatively large numbers of genuine intervals, which are also relatively short, but this is at least partly an artefact of the poor calibration of MQS and the fact that it 'rushes' to declare detection based on scant evidence, as shown in Table 2.

RNSP and RNSP-O significantly outperform SN-NSP and SN-NSP-O in three out of the four scenarios tested: Blocks, Cauchy and Poisson. In Blocks and Cauchy, the RNSP methods achieve more detections and shorter intervals of significance (so better localisation). In Poisson, in addition, they achieve much better coverage (the SN-NSP methods are misled by the discrete nature of this relatively low-intensity Poisson dataset, for which their required

asymptotics do not kick in, which results in a very large number of spurious detections). However, the SN-NSP methods work better for the Bursts data in the sense that they lead to more detections. The underlying reason is that the signal level in this model is linearly proportional to the standard deviation of the noise, which particularly suits the self-normalised SN-NSP methods.

We end this section by briefly commenting on computation times for the two methods that work with binary exceedances: RNSP and MQS. MQS appears unaffected by the landscape of the signal in the sense that its computation takes a roughly similar time for a signal with no change-points as a signal with multiple change-points, if their lengths are the same. On a standard 2015 iMac, MQS takes around 35 seconds to compute for a signal of length $T = 500$. For RNSP, the 'worst case' (in terms of the computation time) is a signal with no detected change-points as RNSP needs to traverse through intervals of all lengths to confirm the non-detection. For a signal with no change-points of length $T = 500$, our R implementation of RNSP takes around 15 seconds to execute on the same machine. For a signal of the same length with prominent change-points located every 100 observations, RNSP takes around 7 seconds.

Finally, we note that unlike RNSP, MQS is a random procedure in the sense of potentially returning different output each time it is run on the same dataset. This, we believe, makes the interpretation of MQS confidence intervals difficult, especially if multiple runs on the same dataset are performed.

6 Data examples

6.1 The US ex-post real interest rate

We re-analyse the time series of US ex-post real interest rate (the three-month treasury bill rate deflated by the CPI inflation rate) considered in Garcia and Perron (1996), Bai and Perron (2003) and Fryzlewicz (2021). The dataset is available at `http://qed.econ.queensu.ca/jae/datasets/bai001/`. The time series, shown in Figure 1, is quarterly and the range is 1961:1–1986:3, so $t = 1, \ldots, T = 103$.

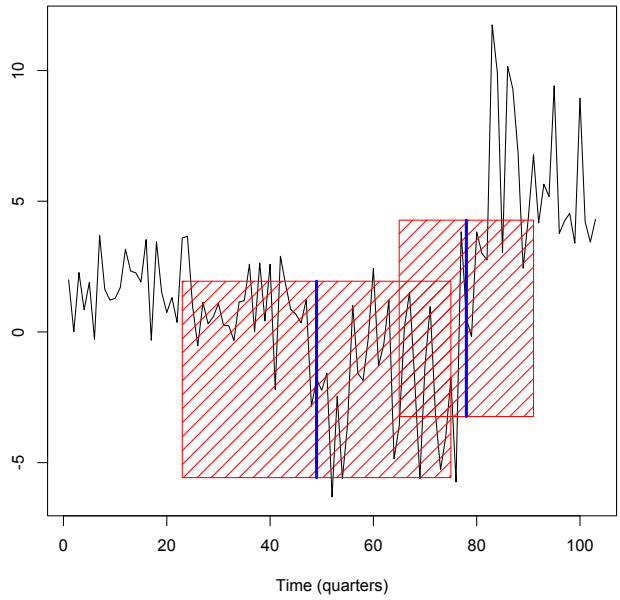

Figure 1: The US ex-post real interest rate time series (black); intervals of significance returned by RNSP (shaded red); their midpoints (blue). See Section 6.1 for a detailed description.

RNSP appears to be an appropriate tool here, as the data displays heterogeneity and possibly some heavy-tailed movements towards the latter part. We run the RNSP algorithm with the default setting of $M = 1000$, with $\alpha = 0.1$ and with overlaps as defined in (13). The procedure returns two intervals of significance: $[23, 75]$ and $[65, 91]$. These are shown in Figure 1, together with their midpoints, which can serve as estimates of the change-point locations. Midpoints of RNSP intervals of significance, while seemingly appearing ad hoc as pointwise estimates of change-point location, often behave well empirically, which may be due to the fact that RNSP pursues short intervals, and those tend to be symmetric around the true change-points as this offers the same amount of evidence on either side of the change-points – hence the frequent empirical closeness of RNSP interval midpoints to the truth. As with any RNSP execution with non-zero overlaps, one question that may be asked is whether the two intervals may be indicating the same change-point, but this, visually, is unlikely here (the reason for using non-zero overlaps is simply to provide larger samples for the RNSP following the detection of the first interval; using zero overlaps

means the samples are too short and RNSP with zero overlaps does not pick up the second change-point). Therefore, the solution points to a model with at least two change-points. This is consistent (or at least not inconsistent) with both Garcia and Perron (1996), who also settle on a model with two change-points, and Bai and Perron (2003), who prefer a three-change-point model, not excluded by RNSP here.

The difference between those two earlier analyses and ours is that those two (a) were based on asymptotic arguments (and therefore valid asymptotically, for unspecified large samples) and (b) were conditional in the sense that the confidence regions for change-point locations in those two works were conditional on the detection event. By contrast, our analysis via RNSP has a finite-sample nature and the intervals of significance have an unconditional character. Importantly, we do not make any distributional assumptions besides independence and sign-symmetry, both of which are likely to be acceptable for this dataset. The analysis via RNSP is unaffected by the likely heterogeneity in the data.

6.2 Interest in the search term "data science"

We analyse the weekly interest in the search term "data science" from Google Trends, in the US state of California. The link to obtain the data was `https://trends.google.com/trends/explore?date=today%205-y&geo=US-CA&q=data%20science`. Google Trends describe the data as follows. "Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term." Weeks in this data series start on Sundays and the dataset spans the weeks from w/c 28th August 2016 to w/c 15th August 2021 (so almost five years' worth of data). The observations are discrete (integers from 22 to 100), which would likely pose difficulties for the competing methods as outlined earlier.

We execute the RNSP procedure with the default setting of $M = 1000$ and with $\alpha = 0.1$, with no overlaps, which returns the three intervals of significance shown in Figure 2. The intervals are: w/c 23 April 2017 – w/c 10 June 2018 (interval 1), w/c 24 June 2018 – w/c 17 November 2019 (interval 2), w/c 3 May 2020 – w/c 14 March 2021 (interval 3).

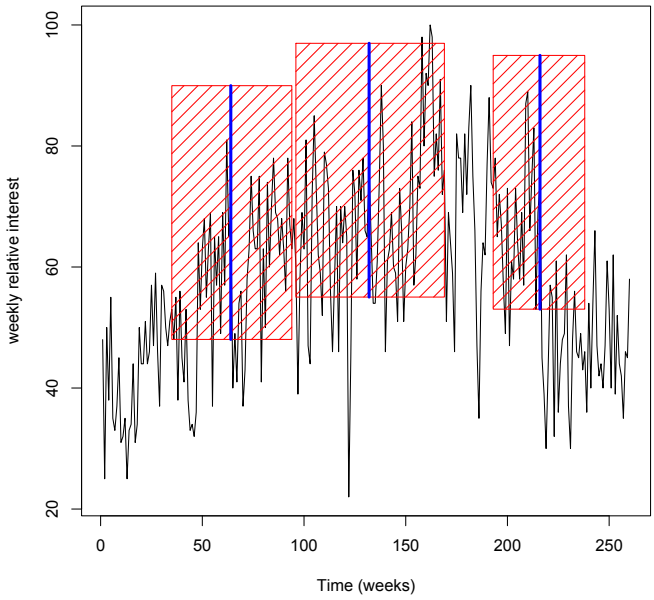

Figure 2: Weekly relative interest (top=100) in the search term "data science" in California in weeks from w/c 28 August 2016 to 15 August 2021 (black); intervals of significance returned by RNSP (shaded red); their midpoints (blue). See Section 6.2 for a detailed description.

While intervals 1 and 2 correspond to likely increases in the median interest, interval 3 clearly corresponds to a likely decrease, and the midpoint of interval 3 (w/c 4 October 2020) visually aligns well with what seems to be a rather sudden drop of interest.

It is difficult to speculate as to possible reasons for this perhaps surprising drop of interest. A blog post on the popular website `https://towardsdatascience.com` (`https://bit.ly/3kqwDWO`) reports that "2020 was the first year since 2016, Data Scientist was not the number one job in America, according to Glassdoor's annual ranking. That title would belong to Front End Engineer, followed by Java Developer, followed by Data Scientist." However, visually, a similar decline in interest is observed e.g. in the analogous Google Trends series for the term "Java" (not shown). To offer some comparative perspective, there is also a slight decline in the search intensity for the term "artificial intelligence" (also picked up by RNSP; not shown), but it is much less pronounced than those in "data science" or "Java".

7 Discussion

Note that Corollary 3.1, crucial to the success of RNSP, can be rewritten as

$$D_{[s_m,e_m]} \leq \|\text{sign}\{Y_{s_m:e_m} - Q_{1/2}(Y_{s_m:e_m})\}\|_{\mathcal{I}^a_{[s_m,e_m]}}, \tag{14}$$

where $Q_q(\cdot)$ is the population q-quantile. However, the left-hand side of (14) was not specifically constructed with $q = 1/2$ in mind and therefore the inequality $D_{[s_m,e_m]} \leq \|\text{sign}\{Y_{s_m:e_m} - Q_q(Y_{s_m:e_m})\}\|_{\mathcal{I}^a_{[s_m,e_m]}}$ is true for any $q \in (0,1)$. This shows that RNSP can equally be used for significant change detection in any quantile, and not just the median. However, for $q \neq 1/2$, the challenge is to obtain the null distribution of $\|\text{sign}\{Y - Q_q(Y)\}\|_{\mathcal{I}^a}$. Even if this challenge is overcome (e.g. by simulation), RNSP as defined in this work may not be effective for change detection in quantiles "far" from the median, due to the particular way in which $D_{[s_m,e_m]}$ is constructed (involving minimisation over all levels). For RNSP to be a successful device for change detection in other quantiles, the definition of $D_{[s_m,e_m]}$ would have to be modified to only minimise over 'realistic' candidate levels not far from the population q-quantile under the local null.

A Proofs

Proof of Theorem 4.1. Consider initially the case of a single change-point η_1. RNSP will, among others, consider intervals symmetric about the true change-point, i.e. $[\eta_1 - d + 1, \eta_1 + d]$, for all appropriate d. Take a constant candidate fit w on the interval $[\eta_1 - d + 1, \eta_1 + d]$ and define $U_t(w) := \text{sign}(Y_t - w) = 2\mathbb{I}(Y_t - w > 0) - 1$ (the latter equality holds due to the

continuity of the distribution of Z_t). Assume wlog $f_{\eta_1} > f_{\eta_1+1}$. We have

$$\begin{aligned}
D_{[\eta_1-d+1,\eta_1+d]} &\geq \inf_w \frac{1}{\sqrt{d}} \max\left\{ \left| \sum_{t=\eta_1-d+1}^{\eta_1} U_t(w) \right|, \left| \sum_{t=\eta_1+1}^{\eta_1+d} U_t(w) \right| \right\} \\
&= \frac{2}{\sqrt{d}} \inf_w \max\left\{ \left| \sum_{t=\eta_1-d+1}^{\eta_1} \mathbb{I}(Y_t - w > 0) - \frac{1}{2} \right|, \left| \sum_{t=\eta_1+1}^{\eta_1+d} \mathbb{I}(Y_t - w > 0) - \frac{1}{2} \right| \right\} \\
&\geq \frac{2}{\sqrt{d}} \inf_w \max\left\{ \left| \sum_{t=\eta_1-d+1}^{\eta_1} P(Y_t - w > 0) - \frac{1}{2} \right|, \left| \sum_{t=\eta_1+1}^{\eta_1+d} P(Y_t - w > 0) - \frac{1}{2} \right| \right\} \\
&- \frac{2}{\sqrt{d}} \sup_w \max\left\{ \left| \sum_{t=\eta_1-d+1}^{\eta_1} \epsilon_t^Y(w) \right|, \left| \sum_{t=\eta_1+1}^{\eta_1+d} \epsilon_t^Y(w) \right| \right\}. \qquad (15)
\end{aligned}$$

Now note

$$\begin{aligned}
\frac{1}{\sqrt{d}} \sup_w \max\left\{ \left| \sum_{t=\eta_1-d+1}^{\eta_1} \epsilon_t^Y(w) \right|, \left| \sum_{t=\eta_1+1}^{\eta_1+d} \epsilon_t^Y(w) \right| \right\} &\leq \max_{s,e} \sup_w \left| \frac{1}{\sqrt{e-s+1}} \sum_{t=s}^{e} \epsilon_t^Y(w) \right| \\
&= \max_{s,e} \sup_w \left| \frac{1}{\sqrt{e-s+1}} \sum_{t=s}^{e} \epsilon_t^Z(w) \right| \leq \lambda.
\end{aligned}$$

Continuing from (15), this implies

$$\begin{aligned}
&D_{[\eta_1-d+1,\eta_1+d]} \\
&\quad \geq \frac{2}{\sqrt{d}} \inf_w \max\left\{ \left| \sum_{t=\eta_1-d+1}^{\eta_1} P(Y_t - w > 0) - \frac{1}{2} \right|, \left| \sum_{t=\eta_1+1}^{\eta_1+d} P(Y_t - w > 0) - \frac{1}{2} \right| \right\} - 2\lambda \\
&\quad = \frac{2}{\sqrt{d}} \inf_w \max\left\{ \left| \sum_{t=\eta_1-d+1}^{\eta_1} P(Z_t > w - f_{\eta_1}) - \frac{1}{2} \right|, \left| \sum_{t=\eta_1+1}^{\eta_1+d} P(Z_t > w - f_{\eta_1+1}) - \frac{1}{2} \right| \right\} - 2\lambda. \qquad (16)
\end{aligned}$$

The infimum over w will be achieved if both elements of the maximum are the same (or otherwise it would be possible to alter w slightly to decrease the larger of the two moduli). But this is only possible if $w \in (f_{\eta_1+1}, f_{\eta_1})$ and hence, bearing in mind that Z_t is median-

zero, we have

$$\begin{aligned}\left|\sum_{t=\eta_1-d+1}^{\eta_1} P(Z_t > w - f_{\eta_1}) - \frac{1}{2}\right| &= dP(Z_t > w - f_{\eta_1}) - d/2 = dP\{Z_t \in (w - f_{\eta_1}, 0)\},\\ \left|\sum_{t=\eta_1+1}^{\eta_1+d} P(Z_t > w - f_{\eta_1+1}) - \frac{1}{2}\right| &= d/2 - dP(Z_t > w - f_{\eta_1+1}) = dP\{Z_t \in (0, w - f_{\eta_1+1})\}.\end{aligned}$$

Let w_0 be such that $P\{Z_t \in (w_0 - f_{\eta_1}, 0)\} = P\{Z_t \in (0, w_0 - f_{\eta_1+1})\}$. Since $(w_0 - f_{\eta_1+1}) - (w_0 - f_{\eta_1}) = f_{\eta_1} - f_{\eta_1+1}$, then either $f_{\eta_1} - w_0 \geq (f_{\eta_1} - f_{\eta_1+1})/2$ or $w_0 - f_{\eta_1+1} \geq (f_{\eta_1} - f_{\eta_1+1})/2$. Therefore,

$$\begin{aligned}P\{Z_t \in (w_0 - f_{\eta_1}, 0)\} &= P\{Z_t \in (0, w_0 - f_{\eta_1+1})\}\\ &\geq \min\{P\{Z_t \in (-|f_{\eta_1} - f_{\eta_1+1}|/2, 0)\}, P\{Z_t \in (0, |f_{\eta_1} - f_{\eta_1+1}|/2)\}\} = \Delta_1.\end{aligned}$$

Continuing from (16), we therefore have

$$D_{[\eta_1-d+1,\eta_1+d]} \geq 2\sqrt{d}\Delta_1 - 2\lambda. \tag{17}$$

But from the definition of the RNSP algorithm (line 14), detection on an interval $[s, e]$ will occur if $D_{[s,e]} > \lambda_\alpha$. Therefore, from (17), detection on $[\eta_1 - d + 1, \eta_1 + d]$ will occur if

$$d > \left(\frac{2\lambda + \lambda_\alpha}{2\Delta_1}\right)^2. \tag{18}$$

As RNSP looks for the shortest intervals of significance, the length of the RNSP interval will not exceed that of $[\eta_1 - d + 1, \eta_1 + d]$, which is $2d$. From (18), its length will therefore be bounded from above by $2\left\lceil\left(\frac{2\lambda+\lambda_\alpha}{2\Delta_1}\right)^2 + 1\right\rceil = 2\bar{d}_1$. We now turn our attention to the multiple change-point case. Note that even though the RNSP interval of significance around η_j is guaranteed to be of length at most $2\bar{d}_j$, it will not necessarily be a sub-interval of $[\eta_j - \bar{d}_j + 1, \eta_j + \bar{d}_j]$. Therefore in order that an interval detection around η_j does not interfere with detections around η_{j-1} or η_{j+1}, the distances $\eta_j - \eta_{j-1}$ and $\eta_{j+1} - \eta_j$ must be suitably long, but this is guaranteed by Assumption 4.1(iii). This completes the proof.□

Proof of Corollary 4.1. The fact that $P(\|\text{sign}(Z)\|_{\mathcal{I}^a} > \lambda_\alpha) \to 0$ as $T \to \infty$ is a simple consequence of Corollary 1 in Shao (1995). We next assess and bound the magnitude of $\sup_w \left|\frac{1}{\sqrt{d}} \sum_{t=s}^{s+d-1} \epsilon_t^Z(w)\right|$. The Dvoretzky-Kiefer-Wolfowitz inequality (with Massart's optimal constant, see Massart (1990)) implies

$$P\left(\sup_w \left|\frac{1}{\sqrt{d}} \sum_{t=s}^{s+d-1} \epsilon_t^Z(w)\right| > \lambda\right) = P\left(\sup_w \left|\frac{1}{d} \sum_{t=s}^{s+d-1} \epsilon_t^Z(w)\right| > \lambda d^{-1/2}\right) \leq 2\exp(-2\lambda^2).$$

This leads to a uniform bound via Bonferroni's correction.

$$\begin{aligned} P\left(\max_{s,e} \sup_w \left|\frac{1}{\sqrt{e-s+1}} \sum_{t=s}^{e} \epsilon_t^Z(w)\right| > \lambda\right) &\leq \sum_{s \leq e} P\left(\sup_w \left|\frac{1}{\sqrt{e-s+1}} \sum_{t=s}^{e} \epsilon_t^Z(w)\right| > \lambda\right) \\ &\leq T(T+1)\exp(-2\lambda^2). \end{aligned}$$

For $\lambda = (1+\delta)\log^{1/2} T$, the above tends to zero if $\delta > 0$. This completes the proof. □

References

S. Arlot and A. Celisse. Segmentation of the mean of heteroscedastic data via crossvalidation. *Statist. Comput.*, 21:613–632, 2011.

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

J. Bai and P. Perron. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18:1–22, 2003.

R. Baranowski, Y. Chen, and P. Fryzlewicz. Narrowest-Over-Threshold detection of multiple change-points and change-point-like features. *J. Roy. Stat. Soc. Ser. B*, 81:649–672, 2019.

Y. Chen, R. Shah, and R. Samworth. Discussion of 'Multiscale change point inference' by Frick, Munk and Sieling. *Journal of the Royal Statistical Society: Series B*, 76:544–546, 2014.

D. Cheng, Z. He, and A. Schwartzman. Multiple testing of local extrema for detection of change points. *Electron. J. Statist.*, 14:3705–3729, 2020.

H. Cho and C. Kirch. Bootstrap confidence intervals for multiple change points based on moving sum procedures. *Computational Statistics and Data Analysis,* to appear, 2022.

V.N.L. Duy, H. Toda, R. Sugiyama, and I. Takeuchi. Computing valid p-value for optimal changepoint by selective inference using dynaming programming. In *Advances in Neural Information Processing Systems*, volume 33, pages 11356–11367, 2020.

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

D. Ellenberger, B. Lausen, and T. Friede. Exact change point detection with improved power in small-sample binomial sequences. *Biometrical Journal*, 63:558–574, 2021.

P. Fearnhead. Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and Computing*, 16:203–213, 2006.

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

P. Fryzlewicz. Wild Binary Segmentation for multiple change-point detection. *Ann. Stat.*, 42:2243–2281, 2014.

P. Fryzlewicz. Narrowest Significance Pursuit: inference for multiple change-points in linear models. *Preprint*, 2021.

R. Garcia and P. Perron. An analysis of the real interest rate under regime shifts. *Review of Economics and Statistics*, 78:111–125, 1996.

N. Hao, Y. Niu, and H. Zhang. Multiple change-point detection via a screening and ranking algorithm. *Statistica Sinica*, 23:1553–1572, 2013.

S. Hyun, M. G'Sell, and R. Tibshirani. Exact post-selection inference for the generalized lasso path. *Electronic Journal of Statistics*, 12:1053–1097, 2018.

S. Hyun, K. Lin, M. G'Sell, and R. Tibshirani. Post-selection inference for changepoint detection algorithms with application to copy number variation data. *Biometrics*, 77:1037–1049, 2021.

S. Jewell, P. Fearnhead, and D. Witten. Testing for a change in mean after changepoint detection. *Journal of the Royal Statistical Society Series B,* to appear, 2022.

Z. Kabluchko and Y. Wang. Limiting distribution for the maximal standardized increment of a random walk. *Stoch. Proc. Appl.*, 124:2824–2867, 2014.

H.-J. Kim, M. Fay, E. Feuer, and D. Midthune. Permutation tests for joinpoint regression with application to cancer rates. *Statistics in Medicine*, 19:335–351, 2000.

M. Lavielle and E. Lebarbier. An application of MCMC methods for the multiple change-points problem. *Signal Processing*, 81:39–53, 2001.

H. Li. *Variational Estimators in Statistical Multiscale Analysis*. PhD thesis, Georg August University of Göttingen, 2016.

H. Li and A. Munk. FDR-control in multiscale change-point segmentation. *Electronic Journal of Statistics*, 10:918–959, 2016.

P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, 18: 1269–1283, 1990.

V. Muggeo and G. Adelfio. Efficient change point detection for genomic sequences of continuous measurements. *Bioinformatics*, 27:161–166, 2011.

A. Nemirovski. Nonparametric estimation of smooth regression functions. *J. Comput. System Sci.*, 23:1–11, 1986.

R. Oomen. Price signatures. *Quantitative Finance*, 19:733–761, 2019.

F. Pein, H. Sieling, and A. Munk. Heterogeneous change point inference. *J. Royal Stat. Soc. B*, 79:1207–1227, 2017.

Qi-Man Shao. On a conjecture of Révész. *Proc. Am. Math. Soc*, 123:575–582, 1995.

L. Vanegas, M. Behr, and A. Munk. Multiscale quantile segmentation. *Journal of the American Statistical Association*, 117:1384–1397, 2022.