# Supplementary Material for "Post-selection inference for quantifying uncertainty in changes in variance"

Rachel Carrington and Paul Fearnhead

## A Details for Section 3.2.1

We can show that

$$\Lambda_{s,e}(\tau, \phi) = \log\left(\frac{(A_{s,e}\phi + B_{s,e})^{s-e+1}}{(A_{s,\tau}\phi + B_{s,\tau})^{\tau-s+1}((A_{s,e} - A_{s,\tau})\phi + B_{s,e} - B_{s,\tau})^{e-\tau}}\right),$$

where $A_{s,e}$ and $B_{s,e}$ are constants depending on $s$, $e$, $h$, and $\phi_{obs}$; specifically

$$A_{s,e} = \begin{cases} 0 & s, e \leq \hat{\tau} - h \\ \frac{1}{\phi_{obs}} \sum_{t=\hat{\tau}-h+1}^{e} X_t^2 & s \leq \hat{\tau} - h < e \leq \hat{\tau} \\ \frac{1}{\phi_{obs}} \sum_{t=\hat{\tau}-h+1}^{\hat{\tau}} X_t^2 - \frac{1}{1-\phi_{obs}} \sum_{t=\hat{\tau}+1}^{e} X_t^2 & s \leq \hat{\tau} - h, \hat{\tau} < e < \hat{\tau} + h \\ \frac{1}{\phi_{obs}} \sum_{t=\hat{\tau}-h+1}^{\hat{\tau}} X_t^2 - \frac{1}{1-\phi_{obs}} \sum_{t=\hat{\tau}+1}^{\hat{\tau}+h} X_t^2 & s \leq \hat{\tau} - h, \hat{\tau} + h < e \\ \frac{1}{\phi_{obs}} \sum_{t=s}^{e} X_t^2 & \hat{\tau} - h < s < e \leq \hat{\tau} \\ \frac{1}{\phi_{obs}} \sum_{t=s}^{\hat{\tau}} X_t^2 - \frac{1}{1-\phi_{obs}} \sum_{t=\hat{\tau}+1}^{\min\{\hat{\tau}+h,e\}} X_t^2 & \hat{\tau} - h < s \leq \hat{\tau} < e \\ -\frac{1}{1-\phi_{obs}} \sum_{t=s}^{\min\{\hat{\tau}+h,e\}} X_t^2 & \hat{\tau} < s \leq \hat{\tau} + h, e > s \\ 0 & s, e > \hat{\tau} + h \end{cases}$$

$$B_{s,e} = \begin{cases} \sum_{t=s}^{\min\{e,\hat{\tau}-h\}} X_t^2 & s \le \hat{\tau}-h, e \le \hat{\tau} \\ \sum_{t=s}^{\hat{\tau}-h} X_t^2 + \frac{1}{1-\phi_{obs}} \sum_{t=\hat{\tau}+1}^{e} X_t^2 & s \le \hat{\tau}-h < \hat{\tau} < e \le \hat{\tau}+h \\ \sum_{t=s}^{\hat{\tau}-h} X_t^2 + \frac{1}{1-\phi_{obs}} \sum_{t=\hat{\tau}+1}^{\hat{\tau}+h} X_t^2 + \sum_{t=\hat{\tau}+h+1}^{e} X_t^2 & s \le \hat{\tau}-h < \hat{\tau}+h < e \\ 0 & \hat{\tau}-h < s < e \le \hat{\tau} \\ \frac{1}{1-\phi_{obs}} \sum_{t=\hat{\tau}+1}^{e} X_t^2 & \hat{\tau}-h < s \le \hat{\tau} < e \le \hat{\tau}+h \\ \frac{1}{1-\phi_{obs}} \sum_{t=\hat{\tau}+1}^{\hat{\tau}+h} X_t^2 + \sum_{t=\hat{\tau}+h+1}^{e} X_t^2 & \hat{\tau}-h < s \le \hat{\tau} < \hat{\tau}+h < e \\ \frac{1}{1-\phi_{obs}} \sum_{t=s}^{e} X_t^2 & \hat{\tau} < s < e \le \hat{\tau}+h \\ \frac{1}{1-\phi_{obs}} \sum_{t=s}^{\hat{\tau}+h} X_t^2 + \sum_{t=\hat{\tau}+h+1}^{e} X_t^2 & \hat{\tau} < s \le \hat{\tau}+h < e \\ \sum_{t=s}^{e} X_t^2 & s,e > \hat{\tau}+h \end{cases}$$

# B Proof of Lemma 2

We want to show that $W_1^l, \ldots, W_{h-1}^l, W_1^r, \ldots, W_{h-1}^r, C_0^2, \phi$ are independent. To do this we calculate the joint density of these variables and show that it can be factorized into functions of each variable.

Let $g_i$ $(i = 1, \ldots, 2h)$ denote the function such that $X_{\hat{\tau}-h+i}^2 = g_i(\boldsymbol{W}, \phi, C_0^2)$, and let $\boldsymbol{J}$ denote the Jacobian matrix of partial derivatives

$$\boldsymbol{J} = \begin{pmatrix} \frac{\partial X_{\hat{\tau}-h+1}^2}{\partial W_1^l} & \cdots & \frac{\partial X_{\hat{\tau}-h+1}^2}{\partial W_{h-1}^l} & \frac{\partial X_{\hat{\tau}-h+1}^2}{\partial W_1^r} & \cdots & \frac{\partial X_{\hat{\tau}-h+1}^2}{\partial W_{h-1}^r} & \frac{\partial X_{\hat{\tau}-h+1}^2}{\partial \phi} & \frac{\partial X_{\hat{\tau}-h+1}^2}{\partial C_0^2} \\ \vdots & & & & & & & \vdots \\ \frac{\partial X_{\hat{\tau}+h}^2}{\partial W_1^l} & \cdots & \frac{\partial X_{\hat{\tau}+h}^2}{\partial W_{h-1}^l} & \frac{\partial X_{\hat{\tau}+h}^2}{\partial W_1^r} & \cdots & \frac{\partial X_{\hat{\tau}+h}^2}{\partial W_{h-1}^r} & \frac{\partial X_{\hat{\tau}+h}^2}{\partial \phi} & \frac{\partial X_{\hat{\tau}+h}^2}{\partial C_0^2} \end{pmatrix}.$$

Then the density function of $(\boldsymbol{W}, \phi, C_0^2)$ is given by

$$f(\boldsymbol{W}, \phi, C_0^2) = \prod_{i=1}^{2h} f_{X_{\hat{\tau}-h+i}^2}(g_i^{-1}(\boldsymbol{W}, \phi, C_0^2))|\boldsymbol{J}|.$$

By (5) we have

$$X_{\hat{\tau}-h+i}^2 = \begin{cases} (1 - W_{i-1}^l) \prod_{k=i}^{h-1} W_k^l C_0^2 \phi & i = 1, \ldots, h-1 \\ (1 - W_{i-1}^l) C_0^2 \phi & i = h \end{cases}$$

$$X_{\hat{\tau}+i}^2 = \begin{cases} (1 - W_{i-1}^r) \prod_{k=i}^{h-1} W_k^r C_0^2 (1-\phi) & i = 1, \ldots, h-1 \\ (1 - W_{i-1}^r) C_0^2 (1-\phi) & i = h. \end{cases}$$

Each $X_t^2 \sim Gamma\left(\frac{1}{2}, 2\sigma^2\right)$, so $f_{X_t}(x) = \frac{(2\sigma^2)^{1/2}}{\Gamma(\frac{1}{2})} x^{-1/2} e^{-2\sigma^2 x}$.

We get

$$
\begin{aligned}
\prod_{i=1}^{2h} f_X(g_i^{-1}(\boldsymbol{W},\phi,C_0^2)) &= \frac{\left(2\sigma^2\right)^h}{\Gamma\left(\frac{1}{2}\right)^{2h}} \left(\prod_{i=1}^{2h} X_{\hat{\tau}-h+i}^2\right)^{-1/2} e^{-2\sigma^2 \sum_{i=1}^{2h} X_{\hat{\tau}-h+i}^2} \\
&= \frac{\left(2\sigma^2\right)^h}{\Gamma\left(\frac{1}{2}\right)^{2h}} \prod_{i=1}^{h-1} \left(1-W_i^l\right)^{-1/2} \prod_{i=1}^{h-1} \left(W_i^l\right)^{-i/2} \prod_{i=1}^{h-1} \left(1-W_i^r\right)^{-1/2} \\
&\quad \cdot \prod_{i=1}^{h-1} \left(W_i^r\right)^{-i/2} e^{-2\sigma^2 C_0^2}
\end{aligned}
$$

To calculate the determinant of $\boldsymbol{J}$, note that using elementary row operations, we can show that $|\boldsymbol{J}|$ is equal to the determinant of the matrix

$$
\begin{matrix}
\frac{\partial X_{\hat{\tau}-h+1}^2}{\partial W_1^l} & \cdots & \frac{\partial X_{\hat{\tau}-h+1}^2}{\partial W_{h-1}^l} & \frac{\partial X_{\hat{\tau}-h+1}^2}{\partial W_1^r} & \cdots & \frac{\partial X_{\hat{\tau}-h+1}^2}{\partial W_{h-1}^r} & \frac{\partial X_{\hat{\tau}-h+1}^2}{\partial \phi} & \frac{\partial X_{\hat{\tau}-h+1}^2}{\partial C_0^2} \\
\sum_{i=1}^2 \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_1^l} & \cdots & \sum_{i=1}^2 \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_{h-1}^l} & \sum_{i=1}^2 \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_1^r} & \cdots & \sum_{i=1}^2 \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_{h-1}^r} & \sum_{i=1}^2 \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial \phi} & \sum_{i=1}^2 \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial C_0^2} \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots \\
\sum_{i=1}^{h-1} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_1^l} & \cdots & \sum_{i=1}^{h-1} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_{h-1}^l} & \sum_{i=1}^{h-1} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_1^r} & \cdots & \sum_{i=1}^{h-1} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_{h-1}^r} & \sum_{i=1}^{h-1} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial \phi} & \sum_{i=1}^{h-1} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial C_0^2} \\
\frac{\partial X_{\hat{\tau}+1}^2}{\partial W_1^l} & \cdots & \frac{\partial X_{\hat{\tau}+1}^2}{\partial W_{h-1}^l} & \frac{\partial X_{\hat{\tau}+1}^2}{\partial W_1^r} & \cdots & \frac{\partial X_{\hat{\tau}+1}^2}{\partial W_{h-1}^r} & \frac{\partial X_{\hat{\tau}+1}^2}{\partial \phi} & \frac{\partial X_{\hat{\tau}+1}^2}{\partial C_0^2} \\
\sum_{i=1}^2 \frac{\partial X_{\hat{\tau}+i}^2}{\partial W_1^l} & \cdots & \sum_{i=1}^2 \frac{\partial X_{\hat{\tau}+i}^2}{\partial W_{h-1}^l} & \sum_{i=1}^2 \frac{\partial X_{\hat{\tau}+i}^2}{\partial W_1^r} & \cdots & \sum_{i=1}^2 \frac{\partial X_{\hat{\tau}+i}^2}{\partial W_{h-1}^r} & \sum_{i=1}^2 \frac{\partial X_{\hat{\tau}+i}^2}{\partial \phi} & \sum_{i=1}^2 \frac{\partial X_{\hat{\tau}+i}^2}{\partial C_0^2} \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots \\
\sum_{i=1}^{h-1} \frac{\partial X_{\hat{\tau}+i}^2}{\partial W_1^l} & \cdots & \sum_{i=1}^{h-1} \frac{\partial X_{\hat{\tau}+i}^2}{\partial W_{h-1}^l} & \sum_{i=1}^{h-1} \frac{\partial X_{\hat{\tau}+i}^2}{\partial W_1^r} & \cdots & \sum_{i=1}^{h-1} \frac{\partial X_{\hat{\tau}+i}^2}{\partial W_{h-1}^r} & \sum_{i=1}^{h-1} \frac{\partial X_{\hat{\tau}+i}^2}{\partial \phi} & \sum_{i=1}^{h-1} \frac{\partial X_{\hat{\tau}+i}^2}{\partial C_0^2} \\
\sum_{i=1}^{h} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_1^l} & \cdots & \sum_{i=1}^{h} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_{h-1}^l} & \sum_{i=1}^{h} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_1^r} & \cdots & \sum_{i=1}^{h} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_{h-1}^r} & \sum_{i=1}^{h} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial \phi} & \sum_{i=1}^{h} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial C_0^2} \\
\sum_{i=1}^{2h} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_1^l} & \cdots & \sum_{i=1}^{2h} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_{h-1}^l} & \sum_{i=1}^{2h} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_1^r} & \cdots & \sum_{i=1}^{2h} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_{h-1}^r} & \sum_{i=1}^{2h} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial \phi} & \sum_{i=1}^{2h} \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial C_0^2}
\end{matrix}.
$$

We can show that this matrix is upper triangular, which allows us to calculate the determinant as the product of its diagonal elements. Since

$$
\sum_{i=1}^{k} X_{\hat{\tau}-h+i}^2 = \prod_{j=k}^{h-1} W_j^l C_0^2 \phi,
$$

it follows that $\sum_{i=1}^k \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_j^l} = 0$ for $j < k$, and clearly we also have $\sum_{i=1}^k \frac{\partial X_{\hat{\tau}-h+i}^2}{\partial W_j^r} = 0$ for all $j$. Similarly, we can show that $\sum_{i=1}^k \frac{\partial X_{\hat{\tau}+i}^2}{\partial W_j^r} = 0$ for $j < k$, and $\sum_{i=1}^k \frac{\partial X_{\hat{\tau}+i}^2}{\partial W_j^l} = 0$ for all $k$. In the bottom row, as $\sum_{i=1}^{2h} X_{\hat{\tau}-h+i}^2 = C_0^2$, all the matrix entries except the last are 0.

We get

$$\sum_{i=1}^{j} \frac{\partial X^2_{\hat{\tau}-h+i}}{\partial W^l_j} = \phi C_0^2 \prod_{k=j+1}^{h-1} W^l_k \quad j, = 1, \ldots, h-1$$

$$\sum_{i=1}^{j} \frac{\partial X^2_{\hat{\tau}+i}}{\partial W^r_j} = (1-\phi) C_0^2 \prod_{k=j+1}^{h-1} W^r_k, \quad j = 1, \ldots, h-1$$

$$\sum_{i=1}^{h} \frac{\partial X^2_{\hat{\tau}-h+i}}{\partial \phi} = C_0^2$$

$$\sum_{i=1}^{2h} \frac{\partial X^2_{\hat{\tau}-h+i}}{\partial C_0^2} = 1.$$

Hence,

$$|\boldsymbol{J}| = \prod_{i=1}^{h-1} \left(W^l_i\right)^{i-1} \prod_{i=1}^{h-1} \left(W^r_i\right)^{i-1} \phi^{h-1} (1-\phi)^{h-1} \left(C_0^2\right)^{2h-1}.$$

Hence, we get

$$f(\boldsymbol{W}, \phi, C_0^2) = \frac{(2\sigma^2)^h}{\Gamma(\frac{1}{2})^{2h}} \prod_{i=1}^{h-1} \left((W^l_i)^i (1 - W^l_i)\right)^{-1/2} \times \prod_{i=1}^{h-1} \left((W^r_i)^i)(1 - W^r_i)\right)^{-1/2} \left(\phi(1-\phi)\right)^{-h/2} C_0^{-2h} e^{-2\sigma^2 C_0^2}.$$

This can be factorized as $h_0 \prod_{i=1}^{h-1} h_i(W^l_i) \prod_{i=1}^{h-1} h_i(W^r_i) h_\phi(\phi) h_C(C_0^2)$. It is also verifiable that all these variables have domain $[0, 1]$ (except for $C_0^2$ which is $[0, \infty)$) regardless of the values of the others. Hence these variables are all independent.

# C Conditioning on less information

As mentioned in Section 4.3, we can increase the power of the test by reducing the amount of information we condition on. Throughout the rest of this paper we have calculated $p$-values by conditioning on all aspects of the data except for the test statistic $\phi$. However, whilst it is necessary to condition on sufficient statistics for the parameters which are not specified under $H_0$, it is possible to define a $p$-value which is not conditional on $\boldsymbol{W} = \{W^l_1, \ldots, W^l_{h-1}, W^r_1, \ldots, W^r_{h-1}\}$, i.e.

$$\Pr\left(\phi \leq \phi_{lower} \text{ or } \phi \geq \phi_{upper} \mid \hat{\tau} \in \mathcal{M}(\boldsymbol{X'}(\phi, \boldsymbol{W}))\right).$$

where here both $\phi$ and $\boldsymbol{W}$ are allowed to vary.

This adds complication in calculating the conditional distribution of $\phi$ as the $W$'s are nuisance parameters: although, unconditionally, $\boldsymbol{W}$ and $\phi$ are independent, when the distribution of $\phi$ is calculated truncated to the region of interest this is not the

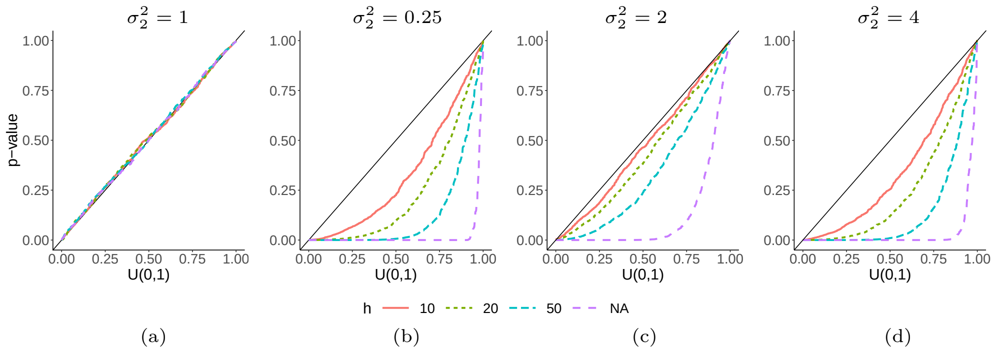



**Fig. 13** QQ plots of $p$-values obtained using the CUSUM method, as for Figure 7, with $T = 1000$. Data is simulated from a model with a single change at $\tau = T/2$, with a pre-change variance of 1, and post-change variance $\sigma_2^2$.

case, as the truncation region also depends on $\boldsymbol{W}$. Conditioning on the observed value of $\boldsymbol{W}$ is one way of dealing with this. An alternative approach, as in Carrington and Fearnhead (2025) for the change in mean model, is to sample values of $\boldsymbol{W}$, calculate the truncation region for each, and then to calculate the $p$-value as a weighted average of these individual $p$-values, i.e.

$$\hat{p} = \frac{\sum_{j=1}^{N_W} \Pr\left(\phi \geq \phi_{upper} \text{ or } \phi \leq \phi_{lower}, \hat{\tau} \in \mathcal{M}(\boldsymbol{X}), \mid \boldsymbol{W} = \boldsymbol{W}^{(j)}\right)}{\sum_{j=1}^{N_W} \Pr\left(\hat{\tau} \in \mathcal{M}(\boldsymbol{X}) \mid \boldsymbol{W} = \boldsymbol{W}^{(j)}\right)}.$$

# D Additional simulations

Here we include some additional simulations to those in Section 5, where we include different changepoint algorithms and parameter values.

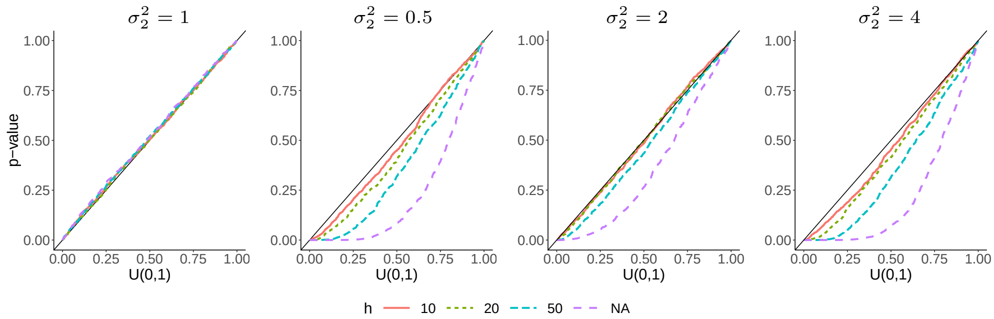
## D.1 Detecting changes using CUSUM

Figure 7 showed QQ plots of $p$-values obtained for simulated data sets, where changepoints were estimated using binary segmentation with the CUSUM statistic. Here, we present similar results for two different scenarios: in Figure 13 we simulate data with $T = 1000$ data points (compared to $T = 200$ in Figure 7) and once again estimate changepoints using binary segmentation; in Figure 14 we set $T = 200$ and estimate changepoints using wild binary segmentation.

In each figure, panel (a) shows the results when data was simulated under $H_0$; panels (b)–(d) show $p$-values obtained when we simulated from a model with a single changepoint at $\tau = T/2$, with three different values of the post-change variance. In each case we see a similar pattern of results to Figure 7.

## D.2 Sampling with and without stratification

In Section 3.2.1, we discussed using stratified sampling to estimate the $p$-value. Figure 15 shows plots of the $p$-values we obtained for a given simulated data set, using




(a) (b) (c) (d)

**Fig. 14** QQ plots of $p$-values obtained using the CUSUM method, as for Figure 7. Data is simulated from a model with a single change at $\tau = T/2$, with a pre-change variance of 1, and post-change variance $\sigma_2^2$. Here, we use wild binary segmentation to estimate changepoints.

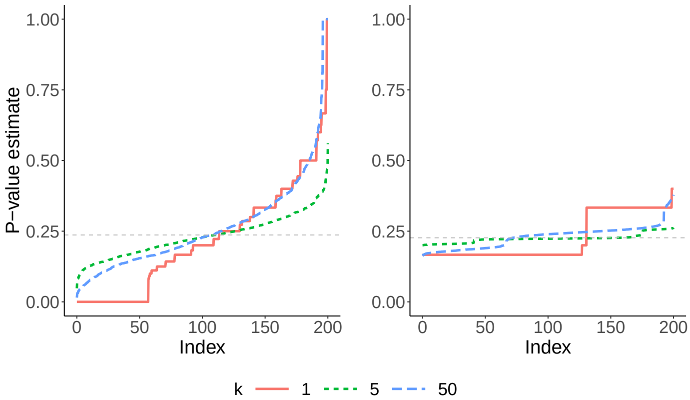



**Fig. 15** Each plot shows the ordered $p$-value estimates obtained over 1000 runs using importance sampling without stratification (left) and with stratification (right), for a simulated data set with $T = 200$, and using $N = 200$ samples each time. The importance sampling distribution is $Beta(\frac{h}{2k}, \frac{h}{2k})$ where $k = 1, 5, 50$. The dashed grey line corresponds to the mean estimate. The same data set was used for all $p$-value estimates.

importance sampling with and without stratification, with 200 samples in each case. (Values of $\phi$ are sampled from $Beta(\frac{h}{2k}, \frac{h}{2k})$ for $k \in \{1, 5, 50\}$; the true distribution of $\phi$ is $Beta(\frac{h}{2}, \frac{h}{2})$.) We can see that although the two methods have the same mean, using stratified sampling results in a much smaller variance in the $p$-value estimate.

## D.3 Additional simulations for Gaussian process approach

For the Gaussian process model, we use a kernel of the form

$$\boldsymbol{K}(\phi_1, \phi_2) = e^{-\frac{1}{2l^2}|\phi_1 - \phi_2|}, \tag{13}$$

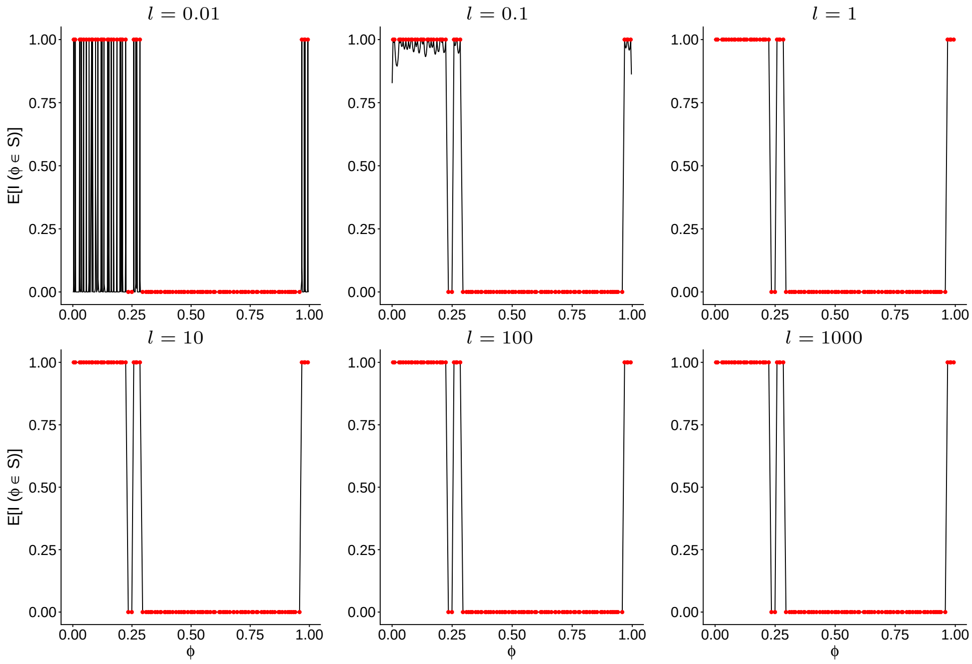



**Fig. 16** Plots showing the expected value of $\mathbb{I}_{\phi \in \mathcal{S}}$ where we fit a Gaussian process with the kernel defined in (13), where $l = \{0.01, 0.1, 1, 10, 100, 1000\}$.

where $\phi_1, \phi_2 \in [0, 1]$ are two possible values of $\phi$. The value $l$ is a hyperparameter, which must be chosen before we fit the model. A larger $l$ will lead to a smoother function $E[\mathbb{I}_{\phi \in \mathcal{S}}]$, as it will more strongly force it to take the same value for two values $\phi_1$ and $\phi_2$ when they are close together.

In Figure 16, we investigate the effect of using six different values of $l$ ($l \in \{0.01, 0.1, 1, 10, 100, 1000\}$), for a simulated data set. The figure displays $E[\mathbb{I}_{\phi \in \mathcal{S}}]$, shown by a black line, with observed values shown as red points. For $l = 1, 10, 100, 1000$, the estimates we obtain are very similar, and appropriately smooth: most $\phi$ values have $E[\mathbb{I}_{\phi \in \mathcal{S}}]$ in $\{0, 1\}$ (or very close to one of these values), and we expect that $\mathcal{S}$ will consist of a small number of intervals. Using $l = 0.01$ or $l = 0.1$ gives less good results. So, it seems like a value $l \geq 1$ is likely to give good results. In the paper we set $l = 100$ for all simulations and data analyses.

## D.4 PELT

Figure 17 shows QQ plots of $p$-values when simulating under three different scenarios – these are the same scenarios as in Figure 9 in the main text – but when changepoints are estimated using PELT (Killick et al., 2012).

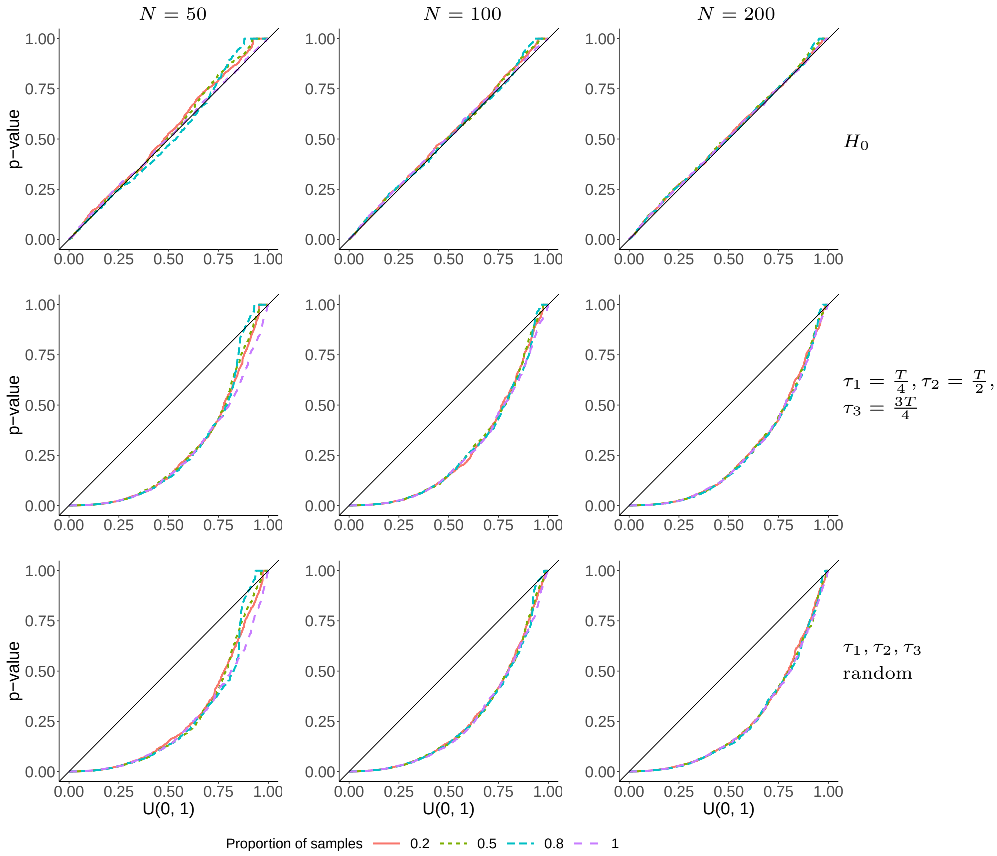



**Fig. 17** QQ plots of $p$-values obtained using the likelihood ratio for three scenarios – no change, three equally spaced changes, and three random changes – using PELT to estimate changepoints. Each line corresponds to the proportion of samples that were used to fit the Gaussian process: where this is less than 1, the remaining samples were drawn from the posterior of the Gaussian process.

# References

Carrington, R., Fearnhead, P.: Improving power by conditioning on less in post-selection inference for changepoints. Statistics and Computing **35**(1), 1–23 (2025)

Killick, R., Fearnhead, P., Eckley, I.A.: Optimal detection of changepoints with a linear computational cost. Journal of the American Statistical Association **107**(500), 1590–1598 (2012)