Supporting Material: Optimal detection of changepoints with a linear computational cost

Killick, R., Fearnhead, P. and Eckley, I.A.

October 3, 2012

1. Changes in Variance - Additional Simulations

The simulation study reported in Section 4.1 of Killick, Fearnhead & Eckley (2012) considered segment variances simulated so that 95% were within [1/10,10] and a linearly increasing number of changepoints ($m = n/100$). Here we also consider the cases of square root increasing changepoints ($m = \lfloor\sqrt{n}/4\rfloor$) and a fixed number of changepoints ($m = 2$). Both these sceanarios violate the assumptions of Theorem 3.2 from the main text. Figure 2 in the supplementary material shows the results for the varaince range [1/10,10] alongside the linearly increasing figures from the main text. Similarly, Figures 1 and 3 in the supplementary material follow the format of Figure 2 in the supplementary material repeated for ranges [1/5,5] and [1/20,20]. The results across the different ranges are similar.

Firstly Figure 2(a) in the supplementary material shows that when the number of changepoints increases linearly with n, PELT does indeed have a CPU cost that is linear in n. By comparison Figures 2(b) and 2(c) in the supplementary material show that if the number of changepoints increases at a slower rate, the CPU cost of PELT is no longer linear. However even in the latter two cases, substantial computational savings are attained relative to optimal partitioning.

Figures 2(g)-2(i) in the supplementary material show the increase in accuracy in terms of mean square error of estimates of the parameter. For the fixed number of changepoints scenario the difference is negligible but, for the linearly increasing number of changepoints scenario, the difference is relatively large.

Figures 4(a)-4(i) in the supplementary material show the proportion of correctly identified changepoints against the number of falsely detected changepoints. The fixed number of changepoints has the highest detection rate for the fewest falsely detected changepoints. This decreases as the number of changepoints increases through square root to linear increases. In all examples PELT has, for the smaller false detection rates, a higher correct detection rate than Binary Segmentation.

Figure 5 compares the timings of the Binary Segmentation algorithm with PELT. In all examples, Binary Segmentation is computationally quicker.

2. *Application to Dow Jones Index*

We now use PELT to analyse data from the Dow Jones Index. The Dow Jones index has been studied in part by many authors including Hsu (1977) and Berkes et al. (2006). Following these authors, Figure 6 in the supplementary material shows the daily closing returns from 1st October 1928 to 30th July 2010 defined as $R_t = c_{t+1}/c_t - 1$, where c_t is the closing price on day t. Previous authors have modelled the Dow Jones daily returns as a change in variance which seems reasonable from Figure 6 in the supplementary material. We also take this approach to analysing the data and set the cost function as twice the negative log likelihood. As in previous analyses we shall assume that the returns follow a Normal distribution with constant (unknown) mean and piecewise stationary (unknown) variance. The changes in variance identified by the PELT and Binary Segmentation methods using the SIC penalty are shown as vertical lines in Figure 6 in the supplementary material. The PELT method identifies 82 changepoints and the Binary Segmentation method identifies 65. The difference in the overall cost function between PELT and Binary Segmentation is 300.1. If we implement Binary Segmentation so as to find the same number of changepoints as PELT, the resulting difference in the log-likelihood for the two segmentations is 59.4. For this data, the decrease in speed from the Optimal Partitioning method to the PELT method is by a factor of 14. For information, PELT identified 82 changepoints of which 32 were in common with optimal Binary Segmentation and 44 were in common with Binary Segmentation for the same number of changepoints as PELT.

3. Changes in Mean and Variance within Normally Distributed Data

Simulation Study The simulation study here will be constructed in a similar way to that of Section 4.1 from the main text. It is assumed that the data follow a Normal distribution with mean and variance depending on the segment. As previously we shall take the cost function as twice the negative of the log likelihood. Note that for a change in mean and variance, the minimum segment length is two observations. For a specific segment the cost is

$$\mathcal{C}(y_{(\tau_{i-1}+1):\tau_i}) = (\tau_i - \tau_{i-1}) \left(\log(2\pi) + \log \left(\sum_{j=\tau_{i-1}+1}^{\tau_i} \left(y_j - \frac{\sum_{k=\tau_{i-1}+1}^{\tau_i} y_k}{(\tau_i - \tau_{i-1})} \right)^2 \right) + 1 \right). \tag{1}$$

We consider scenarios of varying data lengths: n=(100, 200, 500, 1000, 2000, 5000, 10000, 20000, 50000). For each value of n we consider three scenarios for the number of changepoints, m: a linearly increasing number of changepoints, $m = n/100$; the number of changepoints increasing with the square-root of n, $m = \lfloor \sqrt{n}/4 \rfloor$; and a fixed number of changepoints, $m = 2$.

These changepoints are distributed uniformly across $(2, n-2)$ with the only constraint being that there must be at least 30 observations between changepoints. Within each of the 9 scenarios (varying n) we have 1,000 repetitions where the mean for each segment is a realisation from a Normal distribution with mean 0 and standard deviation 2.5. Thus 95% of the simulated means are within the range [-5,5]. As for the change in variance simulation study, the variance parameters for each segment are a realisation from a Log-Normal distribution with mean 0. We consider three standard deviations $\frac{\log(50)}{2}$, $\frac{\log(10)}{2}$ and $\frac{\log(20)}{2}$. These parameters are chosen so that 95% of the simulated variances are within the ranges $\left[\frac{1}{5}, 5\right]$, $\left[\frac{1}{10}, 10\right]$ and $\left[\frac{1}{20}, 20\right]$ respectively. Figures 7, 8 and 9 in the supplementary material follow the format of Figure 2 from the supplementary material repeated for the change in mean and variance sceanarios. Each row depicts a different method for comparing the PELT and Binary Segmentation algorithms. The first row is computational time; the second is difference in likelihood and the third is MSE of parameter estimates. The results are broadly similar to their counterparts for the change in variance example. One notable difference is that the MSE for the variance parameter in the change in mean and variance sceanarios tends to be larger than for the change in variance sceanario.

4. Changes in Mean within Normally Distributed Data

The simulation study here will be constructed in a similar way to that of Section 4.1 from the main text. It is assumed that the data follow a Normal distribution with mean depending on the segment. As previously we shall take the cost function as twice the negative of the log likelihood. We consider 4 data lengths: n=(500, 5000, 50000, 500000). For each value of n we consider the two scenarios for the number of changepoints, m: a linearly increasing number of changepoints, $m = n/100$ and a fixed number of changepoints, $m = 2$.

These changepoints are distributed uniformly across $(1, n-1)$ with the only constraint being that there must be at least 30 observations between changepoints. Within each of the 3 smalleset scenarios (varying n) we have 100 repetitions and for the largest n we have 10 repetitions. In each repetition the mean for each segment is a realisation from a Normal distribution with mean 0 and standard deviation 2.5. Thus 95% of the simulated means are within the range [-5,5]; the variance was set to 1.

This small scale simulation study was conducted to compare the computational time of PELT with the computational time of another exact method called the Pruned Dynamic Programming Algorithm (PDPA) from Rigaill (2010). Unfortunately the PDPA algorithm cannot handle multiple parameter problems such as a change in both the mean and the variance and, through contact with the authors, we dicsovered that it also could not identify changes in variance. Hence, this small scale study considering changes in mean.

As both methods are exact they result in the same segmentation, it is simply the computational time and memory requirements that differ. Table 1 reports the average computational time for each n (with non-overlaping confidence intervals) where PELT is computationally quicker in all but the largest fixed changepoint sceanario. It should also be noted that for the largest length of data ($n = 500000$), PDPA required 8Gb of memory in the fixed data and 50Gb in the linearly increasing sceanarios. PELT required less than 256Mb of memory in all sceanarios.

Table 1: Average Computational Time (std dev. in brackets) for a change in mean.

Length of data (n)		500	5,000	50,000	500,000
Linearly Increasing	PELT	**0.00085**	**0.00739**	**0.07628**	**0.94730**
		(0.00039)	(0.00086)	(0.00376)	(0.3353)
	PDPA	0.33381	34.53683	3437.02234	336829.94150
		(0.07515)	(2.44199)	(92.21506)	(2627.57174)
Fixed	PELT	**0.00011**	**0.07257**	**7.11876**	594.9286
		(0.00034)	(0.02013)	(2.03860)	(178.4677)
	PDPA	0.16442	1.98294	24.92679	**270.6696**
		(0.04932)	(0.41821)	(4.79901)	(58.0634)

5. PROOF OF THEOREM 3.1

Assume that (5) from the main text is true. Then

$$\begin{aligned} & F(t) + \mathcal{C}(y_{(t+1):s}) + \beta + K \geq F(s) + \beta \\ \implies \quad & F(t) + \mathcal{C}(y_{(t+1):s}) + \beta + K + \mathcal{C}(y_{(s+1):T}) \geq F(s) + \beta + \mathcal{C}(y_{(s+1):T}) \\ \implies \quad & F(t) + \mathcal{C}(y_{(t+1):T}) + \beta \geq F(s) + \beta + \mathcal{C}(y_{(s+1):T}), \end{aligned}$$

by (4) from the main text. Hence, it follows that t cannot be a future minimiser of the sets

$$S_T := \left\{ F(\tau) + \mathcal{C}(y_{(\tau+1):T}) + \beta, \tau = 0, 1, \dots, T-1 \right\}, T > s$$

and can be removed from the set of τ for each future step.

6. PROOF OF THEOREM 3.2

The proof of Theorem 3.2 has two parts. Firstly we show that the expected computational cost is bounded by nL_n, where L_n is the expected number of changepoint-times stored (i.e. not pruned) when analysing the nth observation. Secondly we show that under assumptions (A1)–(A4) $\lim_{n\to\infty} L_n < \infty$.

Introduce notation $G(y_{(s+1):t})$ to be the minimum value of the cost function (3), from the original text, for data $y_{(s+1):t}$. So, as previously defined, $F(t) = G(y_{1:t})$. We will consider pruning on the more stringent condition that changepoint $t - j$ is removed at iteration t if

$$\mathcal{C}(y_{(t-j+1):t}) > G(y_{(t-j+1):t}). \qquad (2)$$

That fact that this is a more stringent condition, comes from (6) from the original text, noting that $F(t) \leq F(t-j) + G(y_{(t-j+1):t})$ as the latter is the smallest overall cost for segmentations that include a changepoint at $t - j$, and remembering that for our choice of $\mathcal{C}(\cdot)$, $K = 0$. Furthermore the computational cost of PELT will be bounded above by the method which prunes using this condition.

Assume we are pruning with condition (2). For a positive integer $j \leq t$, let $I_{t,j}$ be an indicator of whether a changepoint at time $t - j$ is stored after processing the observation at time t. The overall computational cost of processing the observation at time $(t+1)$ is $1+\sum_{j=1}^{t} I_{t,j}$. Now as the data-generating process is time-invariant, and our condition (2) just depends on data $y_{(t-j+1):t}$, we have $\mathbb{E}\left(I_{t,j}\right) = E_j$, independent of t. So the expected computational cost is bounded by nL_n where

$$L_n = 1 + \sum_{j=1}^{n-1} E_j,$$

the expected computational cost of processing the last observation.

Now we define L as the limit of L_n as $n \to \infty$. We need to show this is finite. If it is, the computational cost of a method using (2) to prune will have a computational cost that is linear in n, and hence so will PELT. We will do this by showing that E_j decays to 0 sufficiently quickly as $j \to \infty$.

Now (by choosing $t = j$) E_j is the probability that $I_{j,j} = 1$, which is that a changepoint at time 0 is not pruned after processing the jth observation. For $I_{j,j} = 1$ we need

$$\mathcal{C}(1, j) \leq F(j),$$

where $\mathcal{C}(1, j)$ is the cost associated with assuming a single segment for observation $y_{1:j}$ and $F(j)$ is the minimum cost possible for segmenting $y_{1:j}$. Now we will define m_j to be the number of actual changepoints before time j, and $\tau_1, \ldots, \tau_{m_j}$ their positions. Furthermore we define $\tau_0 = 0$ and, with a slight abuse of notation, $\tau_{m_j+1} = j$. Then

$$F(j) \leq \sum_{i=1}^{m_j+1} \left[\mathcal{C}(y_{(\tau_{i-1}+1):\tau_i}) + \beta\right].$$

So

$$E_j \leq \Pr\left(\mathcal{C}(1, j) \leq \sum_{i=1}^{m_j+1} \left[\mathcal{C}(y_{(\tau_{i-1}+1):\tau_i}) + \beta\right]\right).$$

Now define θ_i to be the value of the parameter associated with the true segment of observation i; and $\tilde{\theta}_i$ the value of the maximum likelihood estimate of the parameter associated with the true segment of observation i:

$$\tilde{\theta}_i = \arg\max_{\theta} \sum_{k=\tau_{l-1}+1}^{\tau_l} \log f(y_k|\theta),$$

where l is defined so that $\tau_{l-1} < i \leq \tau_l$.

Now $\mathcal{C}(1, j) = -\sum_{i=1}^{j} \log f(y_i|\hat{\theta}_j)$, where $\hat{\theta}_j$ is defined in Theorem 3.2 to be the maximum likelihood estimate θ for data $y_{1:j}$ under an assumption of a single segment. By the definition above, we also have

$$\sum_{i=1}^{m_j+1} \mathcal{C}(y_{(\tau_{i-1}+1):\tau_i}) = -\sum_{i=1}^{j} \log f(y_i|\tilde{\theta}_i)$$

So we can re-write

$$\overbrace{\mathcal{C}(1, j) - \sum_{i=1}^{m_j+1} [\mathcal{C}(y_{(\tau_{i-1}+1):\tau_i}) + \beta]}^{A_j} = \overbrace{\sum_{i=1}^{j}[\log f(y_i|\theta^*) - \log f(y_i|\hat{\theta}_j)]}^{B_j} +$$

$$\overbrace{\sum_{i=1}^{j}[\log f(y_i|\theta_i) - \log f(y_i|\theta^*)] - (m_j + 1)\beta}^{D_j} + \overbrace{\sum_{i=1}^{j}[\log f(y_i|\tilde{\theta}_i) - \log f(y_i|\theta_i)]}^{R_j}. \quad (3)$$

First note that $R_j \geq 0$. So $E_j = \Pr(A_j \leq 0) \leq \Pr(B_j + D_j \leq 0)$. Now we will bound this probability using Markov's inequality.

By (A1), and using that the expected number of changepoints is related to the expected segment length, $\mathbb{E}(M_j) = j/\mathbb{E}(S) + o(j)$ (elementary renewal theorem), we have

$$\mathbb{E}(B_j + D_j) = \mathbb{E}(B_j) + \mathbb{E}(D_j) = \mathbb{E}\left(\sum_{i=1}^{j}[\log f(Y_i|\theta_i) - \log f(Y_i|\theta^*)]\right) - \beta\frac{j}{\mathbb{E}(S)} + o(j)$$

Thus, using (A4), we have that there exists $c > 0$ such that for sufficiently large j

$$\mathbb{E}(B_j + D_j) > cj.$$

Now let $B_j^* = B_j - \mathbb{E}(B_j)$ and $D_j^* = D_j - \mathbb{E}(D_j)$. We now consider $\mathbb{E}\left((B_j^* + D_j^*)^4\right)$, and show that this is $\mathcal{O}(j^2)$. Now the Minkowski Inequality gives that

$$\mathbb{E}\left(\left(B_j^* + D_j^*\right)^4\right) \leq \left[\mathbb{E}\left(\left(B_j^*\right)^4\right)^{1/4} + \mathbb{E}\left(\left(D_j^*\right)^4\right)^{1/4}\right]^4.$$

Now by (A1) $\mathbb{E}\left((B_j^*)^4\right) = \mathcal{O}(j^2)$, so we need only to show that $\mathbb{E}\left((D_j^*)^4\right)$ is $\mathcal{O}(j^2)$, in order for $\mathbb{E}\left((B_j^* + D_j^*)^4\right)$ to be $\mathcal{O}(j^2)$.

Define

$$Z_i = \log f(Y_i|\theta_i) - \log f(Y_i|\theta^*) - \mathbb{E}\left(\log f(Y_i|\theta_i) - \log f(Y_i|\theta^*)\right),$$

so $D_j^* = \sum_{i=1}^{j} Z_i - \beta(M_j - \mathbb{E}(M_j))$. Note that Z_i has the same distribution for all i, and we will let Z denote a further random variable with this distribution.

Under condition (A3) and results on moments of renewal processes from Smith (1959) we have $\mathbb{E}\left((M_j - \mathbb{E}(M_j))^3\right) = O(j)$. Now as $0 \leq M_j \leq j$ we have

$$\mathbb{E}\left((M_j - \mathbb{E}(M_j))^4\right) \leq j\mathbb{E}\left((M_j - \mathbb{E}(M_j))^3\right) = O(j^2)$$

Thus, using the Minkowski Inequality again we will have that $\mathbb{E}\left((D_j^*)^4\right)$ is $\mathcal{O}(j^2)$ providing $\mathbb{E}\left(\left(\sum_{i=1}^{j} Z_i\right)^4\right) = O(j^2)$.

We have

$$\mathbb{E}\left(\left(\sum_{i=1}^{j} Z_i\right)^4\right) = \sum_{i_1=1}^{j}\sum_{i_2=1}^{j}\sum_{i_3=1}^{j}\sum_{i_4=1}^{j} \mathbb{E}\left(Z_{i_1}Z_{i_2}Z_{i_3}Z_{i_4}\right).$$

If we condition on the position of the changepoints we have that, by independence across segments:

$$\mathbb{E}\left(Z_{i_1}Z_{i_2}Z_{i_3}Z_{i_4}\right) \leq \begin{cases} \mathbb{E}\left(Z^4\right) & \text{if each segment contains an even number of } i_1,\ldots,i_4 \\ 0 & \text{otherwise.} \end{cases}$$

Thus we get a bound on the fourth moment of $\sum_{i=1}^{j} Z_i$ in terms of the expected value of the segment lengths of our changepoint process. Denote $S_i^{(j)} = \min\{S_i, j\}$, and note that for each segment to contain an even number of $i_1,\ldots,i_4$ we need one segment to contain all four values, or two segments to contain two each. If we know $S_1,\ldots,S_j$ this involves at most

$$3\sum_{i=1}^{m_j+1}\sum_{k=1,k\neq i}^{m_j+1}\left(S_i^{(j)}\right)^2\left(S_k^{(j)}\right)^2 + \sum_{i=1}^{m_j+1}\left(S_i^{(j)}\right)^4$$

possible combinations of $i_1,\ldots,i_4$. Thus taking expectations with respect to $S_1,\ldots,S_n$ we get:

$$\begin{aligned}
\mathbb{E}\left(\left(\sum_{i=1}^{j} Z_i\right)^4\right) &\leq \mathbb{E}\left(Z^4\right)\mathbb{E}\left(3\sum_{i=1}^{m_j+1}\sum_{k=1,k\neq i}^{m_j+1}\left(S_i^{(j)}\right)^2\left(S_k^{(j)}\right)^2 + \sum_{i=1}^{m_j+1}\left(S_i^{(j)}\right)^4\right) \\
&\leq \mathbb{E}\left(Z^4\right)\mathbb{E}\left(3\sum_{i=1}^{j}\sum_{k=1,k\neq i}^{j}\left(S_i^{(j)}\right)^2\left(S_k^{(j)}\right)^2 + \sum_{i=1}^{j}\left(S_i^{(j)}\right)^4\right) \\
&\leq \mathbb{E}\left(Z^4\right)\left\{3\mathbb{E}\left(\sum_{i=1}^{j}\left(S_i^{(j)}\right)^2\right)\mathbb{E}\left(\sum_{k=1}^{j}\left(S_k^{(j)}\right)^2\right) + \mathbb{E}\left(\sum_{i=1}^{j}\left(S_i^{(j)}\right)^4\right)\right\} \\
&\leq \mathbb{E}\left(Z^4\right)\left[3j^2\mathbb{E}\left(S^2\right) + j^2\mathbb{E}\left(S^3\right)\right].
\end{aligned}$$

The last inequality uses that $\mathbb{E}\left(\left(S_i^{(j)}\right)^4\right) \leq j\mathbb{E}\left(\left(S_i^{(j)}\right)^3\right) \leq j\mathbb{E}\left(S_i^3\right)$.

This shows that there exists a $K < \infty$ such that $\mathbb{E}\left(\left(B_j^* + D_j^*\right)^4\right) < Kj^2$. Now using Markov's inequality we have, for j large enough that $\mathbb{E}\left(B_j + D_j\right) > cj$

$$E_j \leq \Pr(B_j + D_j \leq 0) \leq \Pr(|B_j^* + D_j^*| \geq \mathbb{E}\left(B_j + D_j\right)) \leq \frac{\mathbb{E}\left(\left(B_j^* + D_j^*\right)^4\right)}{\left[\mathbb{E}\left(B_j + D_j\right)\right]^4} \leq \frac{Kj^2}{c^4 j^4}.$$

Thus we have $E_j = \mathcal{O}(j^{-2})$, and hence $L = \lim_{n\to\infty} \sum_{j=1}^{n} E_j$ is finite, as required. $\square$

The basic idea of the proof is to show that the probability of pruning $t-j$ as the value of the most recent changepoint before t goes to zero sufficiently quickly. This in turn required considering the cost function we are trying to minimise, and considering the distribution of the difference of this cost function assuming no changepoint between $t-j$ and t, and the cost associated with the true changepoint positions between $t-j$ and t. For more general cost functions and changepoint models, if we can show that the expected value of this difference decreases linearly with j, but its fourth moment increases only quadratically with j, then the same proof will show that PELT has a linear computational cost.

7. PROOF OF THEOREM 3.3

Conditions on f imply that for $m \geq 0$ and $\hat{m} \geq 0$, $f(m) \leq f(\hat{m}) + (m - \hat{m})f'(\hat{m})$. We also note that if minimising (8) from the main text gives $\hat{m}$ changepoints then it immediately follows that these will be the optimal changepoints under the criteria (6) from the main text.

Now, for any given m we can minimise $\sum_{i=1}^{m+1} \mathcal{C}(y_{(\tau_{i-1}+1):\tau_i})$ with respect to the m changepoints $\tau_1, \ldots, \tau_m$. Denote this minimum by $C(m)$. Then we have that $\hat{m}$ satisfies

$$f(\hat{m}) + C(\hat{m}) \leq f(m) + C(m)$$

for any $m = 0, \ldots, n-1$.

Now using the concavity of $f(\cdot)$,

$$f(\hat{m}) + C(\hat{m}) \leq f(m) + C(m) \leq f(\hat{m}) + (m - \hat{m})f'(\hat{m}) + C(m).$$

Rearranging this shows $\hat{m}f'(\hat{m}) + C(\hat{m}) \leq mf'(\hat{m}) + C(m)$, and hence that $\hat{m}$ minimises (8) from the main text.

References

Berkes, I., Horvath, L., Kokoszka, P., and Shao, Q. (2006). On discriminating between long-range dependence and changes in mean. *Ann. Statist.*, 34(3):1140–1165.

Hsu, D. A. (1977). Tests for variance shift at an unknown time point. *Applied Statistics*, 26(3):279–284.

Rigaill, G. (2010). Pruned dynamic programming for optimal multiple change-point detection. *ArXiv e-prints*.

Smith, W. L. (1959). On the cumulants of renewal processes. *Biometrika*, 46:1–29.

Figure 1: Results for change in variance range [1/5,5]. The rows correspond to (a) Average Computational Time (in seconds) for a change in variance, (b) Average difference in cost between PELT and BS, (c) MSE. The columns correspond to, as n increases, (1) linearly increasing; (2) square root increasing; (3) fixed number of changepoints. OP: blue, PELT: black, optimal BS: red, subBS: orange.

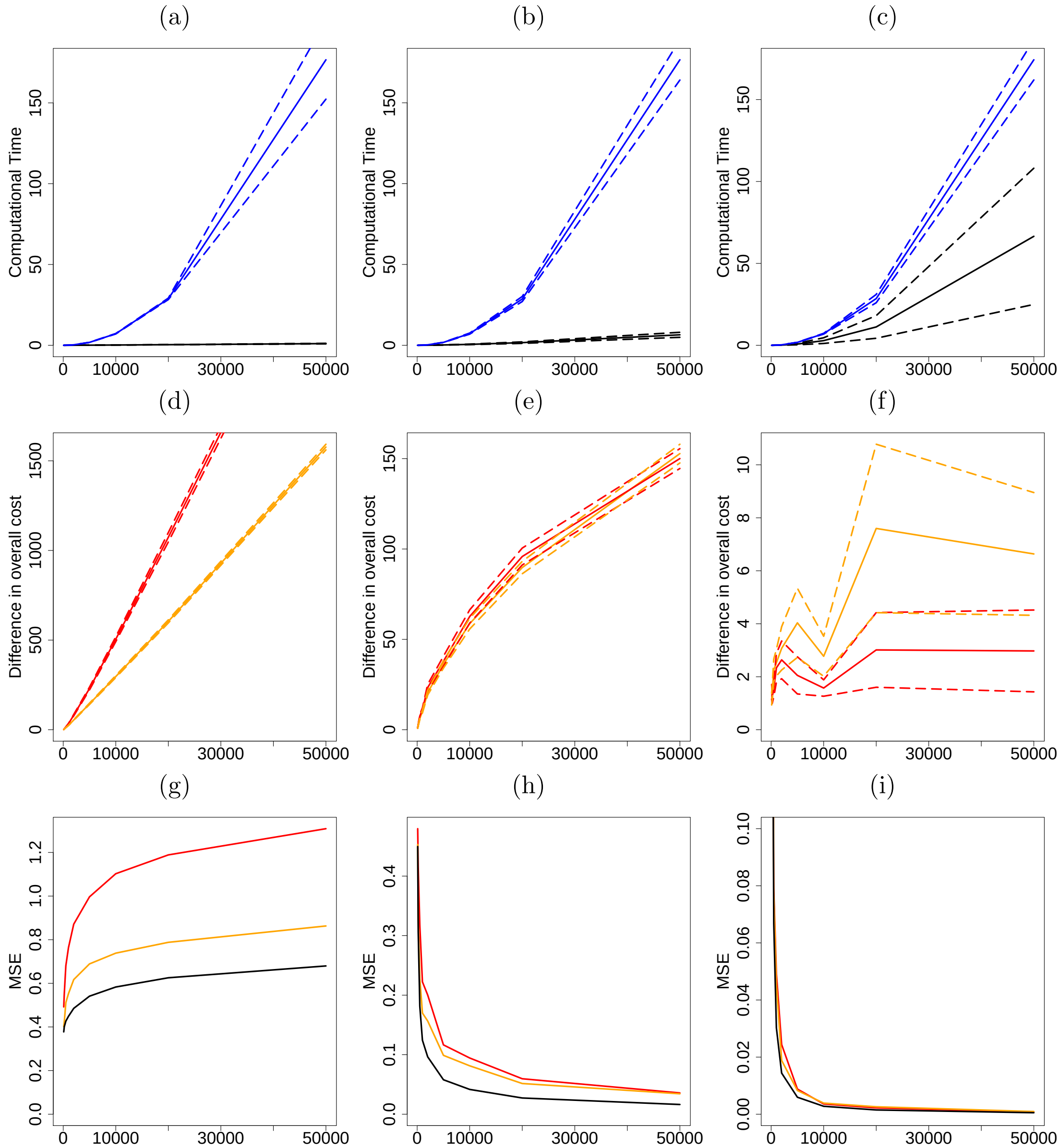

Figure 2: Results for change in variance range [1/10,10]. The rows correspond to (a) Average Computational Time (in seconds) for a change in variance, (b) Average difference in cost between PELT and BS, (c) MSE. The columns correspond to, as n increases, (1) linearly increasing; (2) square root increasing; (3) fixed number of changepoints. OP: blue, PELT: black, optimal BS: red, subBS: orange.

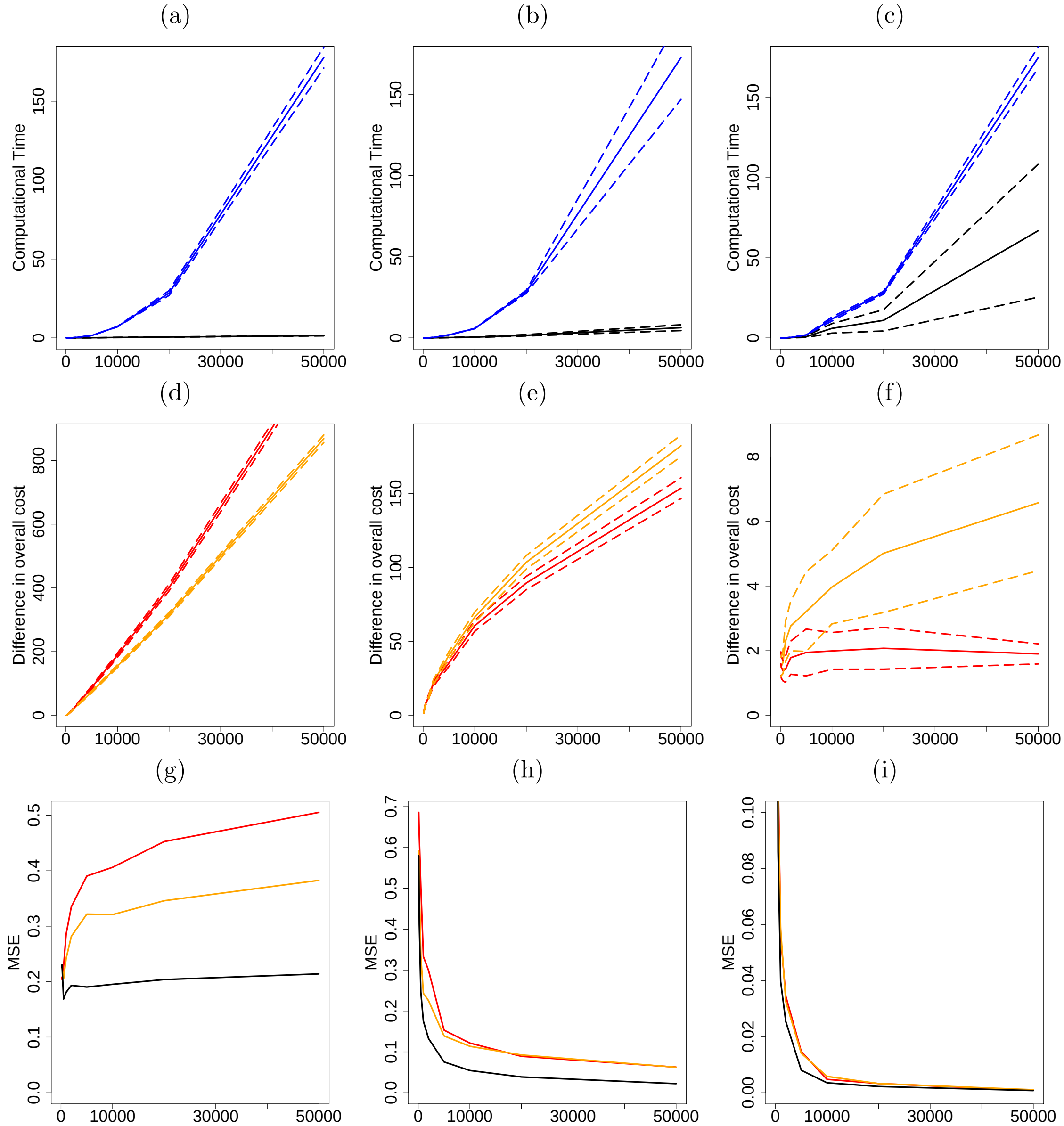

Figure 3: Results for change in variance range [1/20,20]. The rows correspond to (a) Average Computational Time (in seconds) for a change in variance, (b) Average difference in cost between PELT and BS, (c) MSE. The columns correspond to, as n increases, (1) linearly increasing; (2) square root increasing; (3) fixed number of changepoints. OP: blue, PELT: black, optimal BS: red, subBS: orange.

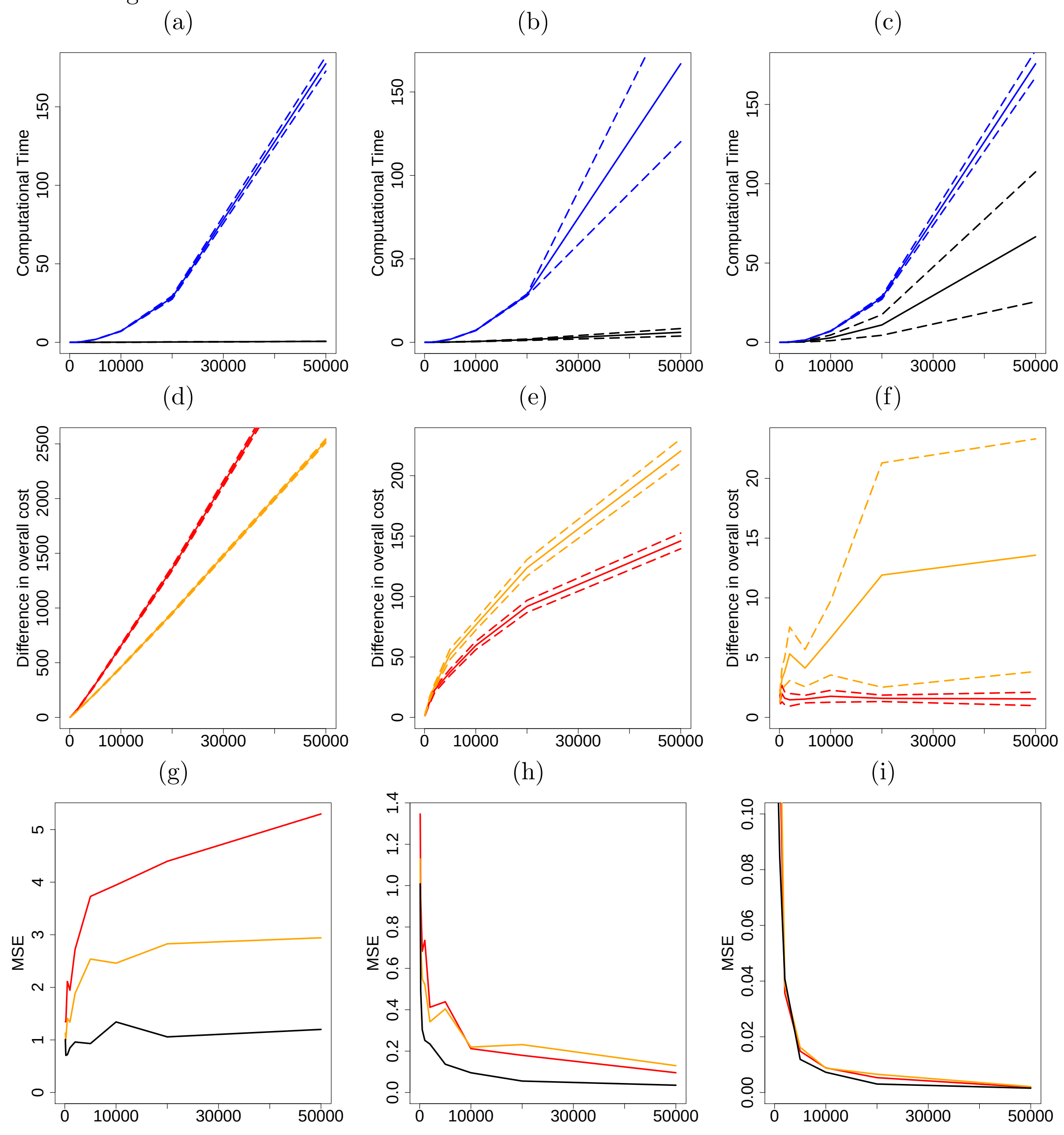

Figure 4: Change in variance: Proportion of correctly identified changepoints (within 10 of true value) against the proportion of falsely detected changepoints with 2σ confidence lines. The columns correspond to (a) $n = 500$, (b) $n = 5,000$, (c) $n = 500,000$. The rows are (1) linearly increasing, (2) square root increasing and (2) fixed number of changepoints. (PELT: black, BS: red, SIC: blue dot)

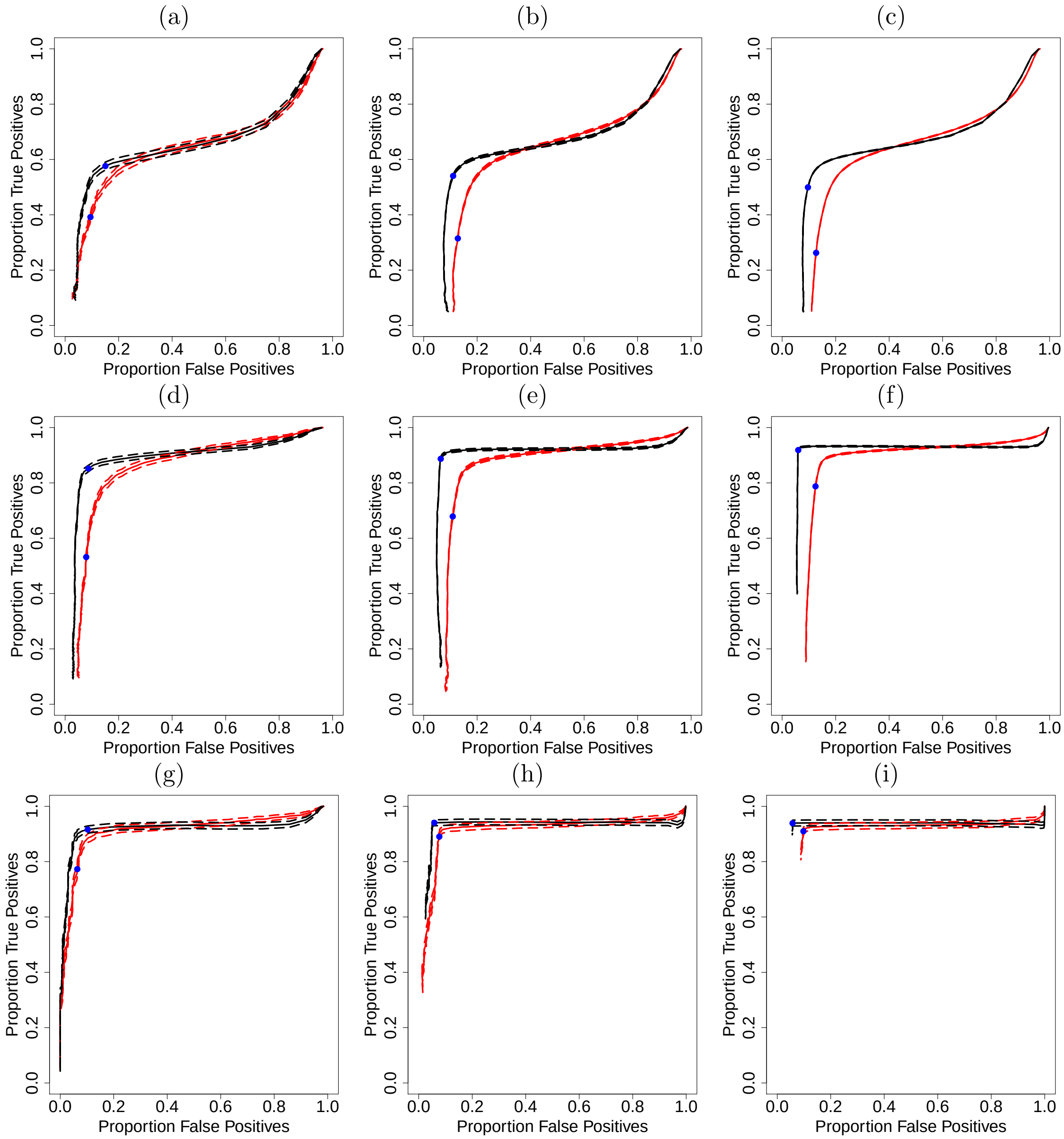

Figure 5: Change in variance: Average Computational Time and 2σ confidence band (in seconds). The columns correspond to, as n increases, (1) linearly increasing; (2) square root increasing; (3) fixed number of changepoints. The rows correspond to variance ranges (1) [1/5,5], (2) [1/10,10], (3) [1/20,20]. PELT: black, BS: red.

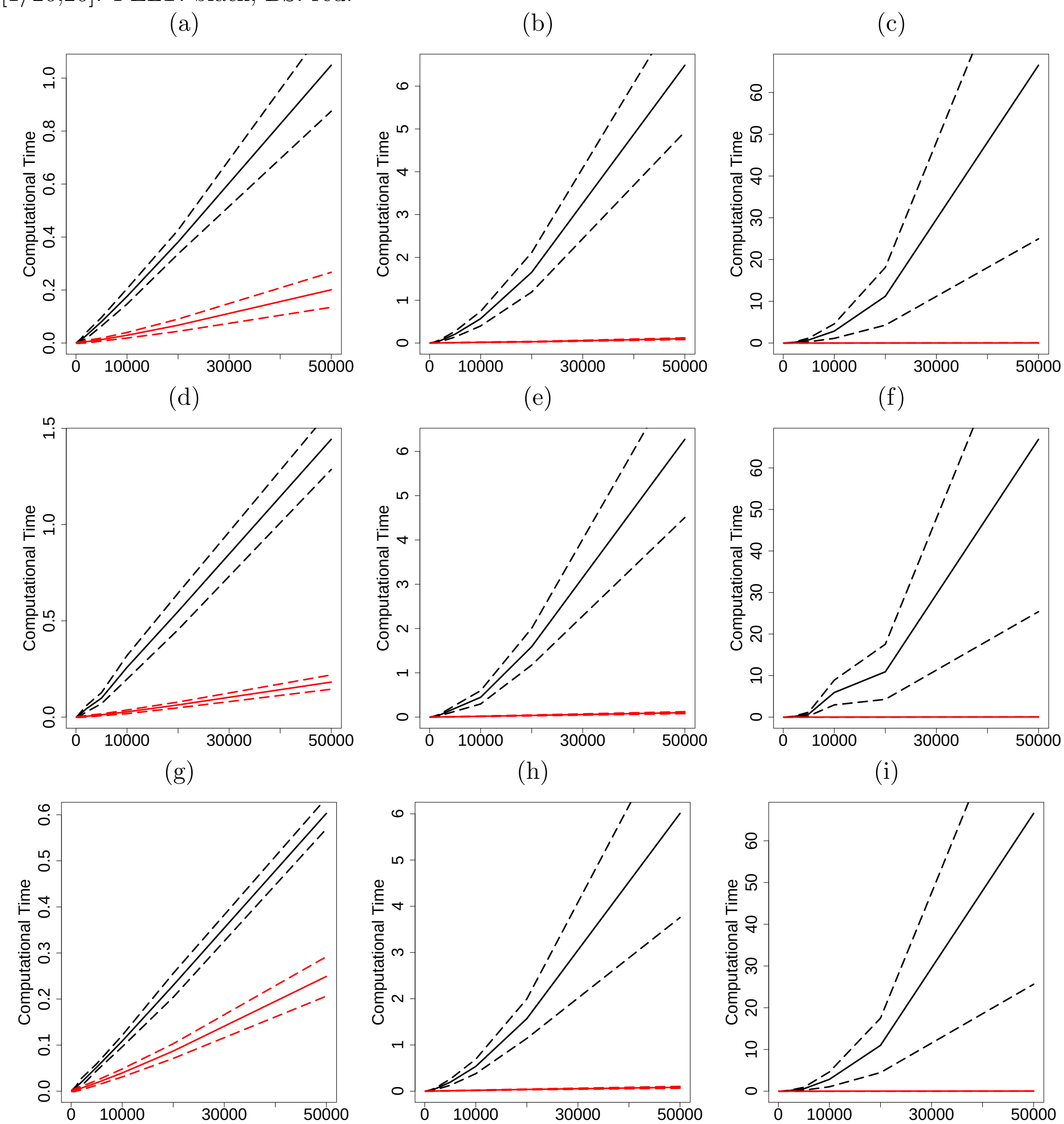

Figure 6: The Dow Jones index daily returns data with changepoints marked using (a) PELT and (b) Binary Segmentation methods with the SIC penalty.

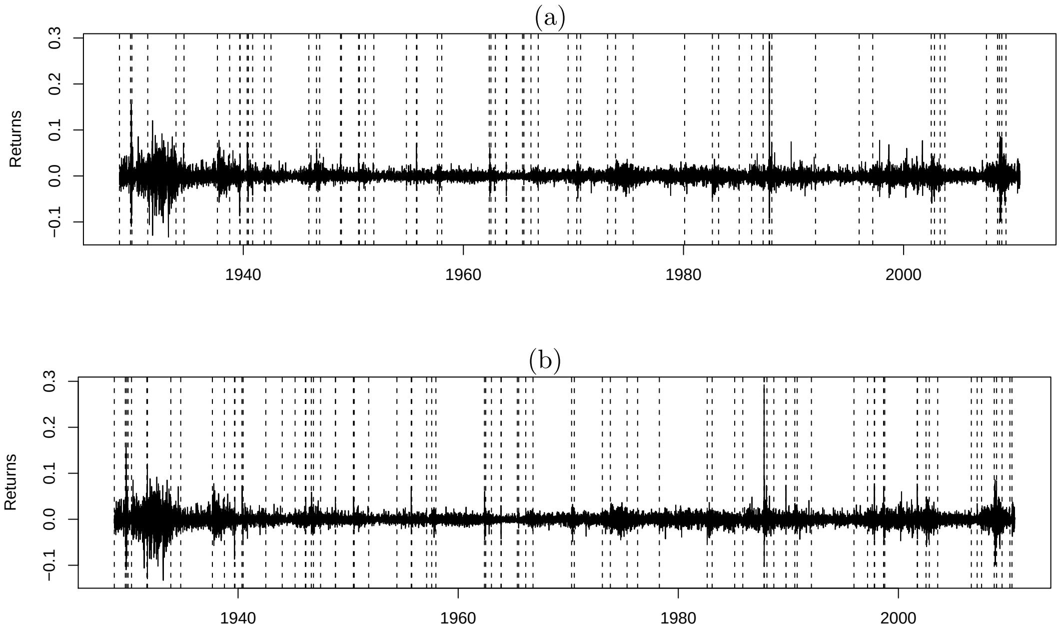

Figure 7: Results for change in mean and variance range [1/5,5]. The rows correspond to (a) Average Computational Time (in seconds) for a change in variance, (b) Average difference in cost between PELT and BS, (c) MSE for mean (dotted) and variance (full) parameters. The columns correspond to, as n increases, (1) linearly increasing; (2) square root increasing; (3) fixed number of changepoints. OP: blue, PELT: black, optimal BS: red, subBS: orange.

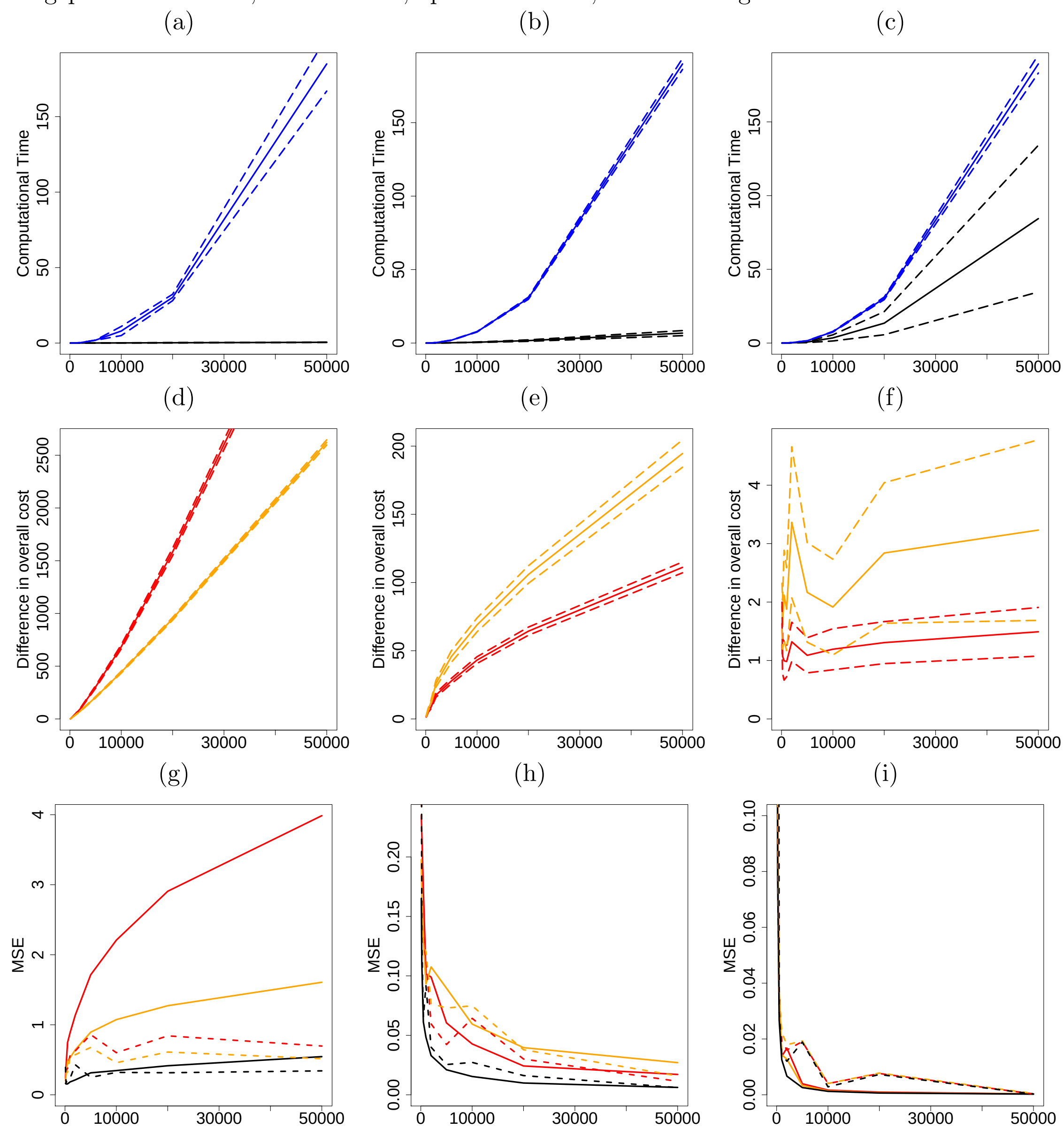

Figure 8: Results for change in mean and variance range [1/10,10]. The rows correspond to (a) Average Computational Time (in seconds) for a change in variance, (b) Average difference in cost between PELT and BS, (c) MSE for mean (dotted) and variance (full) parameters. The columns correspond to, as n increases, (1) linearly increasing; (2) square root increasing; (3) fixed number of changepoints. OP: blue, PELT: black, optimal BS: red, subBS: orange.

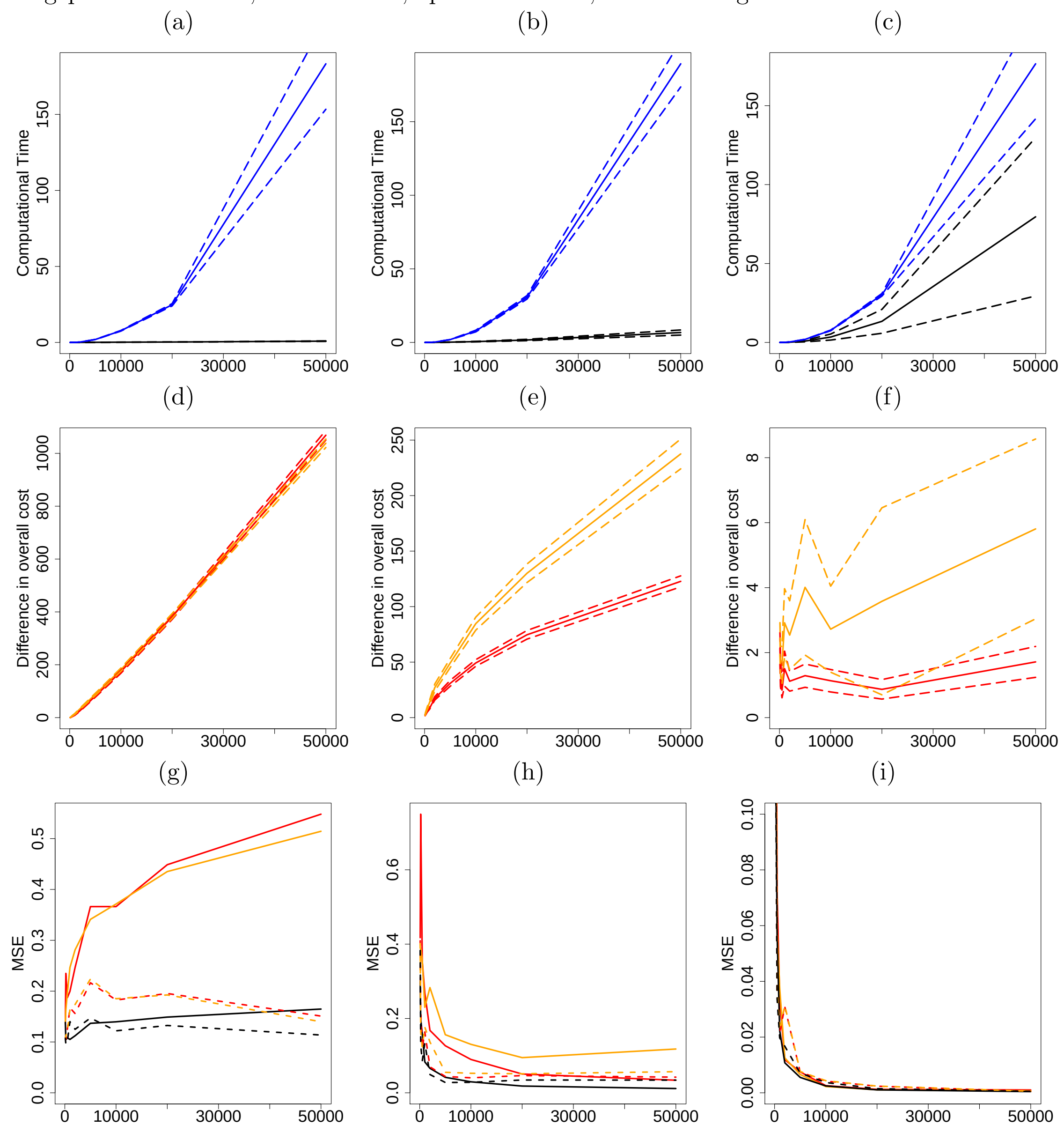

Figure 9: Results for change in mean and variance range [1/20,20]. The rows correspond to (a) Average Computational Time (in seconds) for a change in variance, (b) Average difference in cost between PELT and BS, (c) MSE for mean (dotted) and variance (full) parameters. The columns correspond to, as n increases, (1) linearly increasing; (2) square root increasing; (3) fixed number of changepoints. OP: blue, PELT: black, optimal BS: red, subBS: orange.

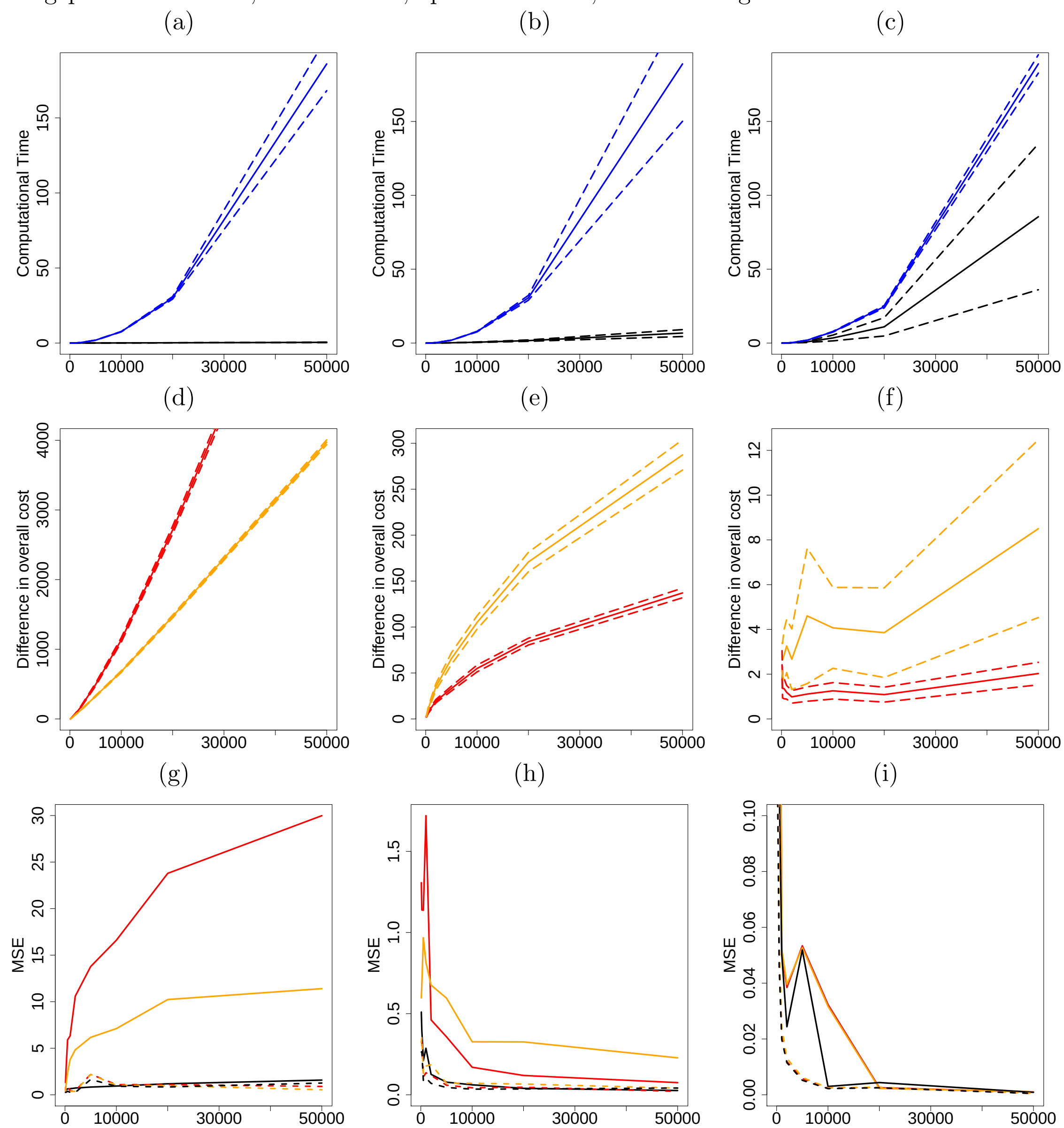

Figure 10: Change in mean and variance: Proportion of correctly identified changepoints (within 10 of true value) against the proportion of falsely detected changepoints with 2σ confidence lines. The columns correspond to (a) $n = 500$, (b) $n = 5,000$, (c) $n = 500,000$. The rows are (1) linearly increasing, (2) square root increasing and (2) fixed number of changepoints. (PELT: black, BS: red, SIC: blue dot)