Received: 24 November 2020 | Accepted: 16 December 2020
DOI: 10.1002/sta4.351

WILEY

ORIGINAL ARTICLE

Detecting changes in mean in the presence of time-varying autocovariance

Euan T. McGonigle[1,2] | Rebecca Killick[3] | Matthew A. Nunes[4]

[1]STOR-i Centre for Doctoral Training, Lancaster University, Lancaster, LA1 4YF, UK
[2]School of Mathematics, University of Bristol, Bristol, BS8 1UG, UK
[3]Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK
[4]Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, UK

Correspondence
Euan T. McGonigle, School of Mathematics, University of Bristol, Bristol, UK.
Email: euan.mcgonigle@bristol.ac.uk

Funding information
Engineering and Physical Sciences Research Council, Grant/Award Number: EP/R511997/1; Numerical Algorithms Group; EPSRC

There has been much attention in recent years to the problem of detecting mean changes in a piecewise constant time series. Often, methods assume that the noise can be taken to be independent, identically distributed (IID), which in practice may not be a reasonable assumption. There is comparatively little work studying the problem of mean changepoint detection in time series with nontrivial autocovariance structure. In this article, we propose a likelihood-based method using wavelets to detect changes in mean in time series that exhibit time-varying autocovariance. Our proposed technique is shown to work well for time series with a variety of error structures via a simulation study, and we demonstrate its effectiveness on two data examples arising in economics.

KEYWORDS
TOPICS, signal processing, TOPICS, statistical modelling, TOPICS, time series, TOPICS, wavelets

1 | INTRODUCTION

We consider the problem of detecting changes in mean in a univariate time series that exhibits a nontrivial, unknown, time-varying autocovariance structure. The study of changepoint problems dates back to Page (1954, 1955), and continues to be a highly active area of research, with applications including genetics (Hocking et al. 2013), cyber security (Rubin-Delanchy, Lawson & Heard, 2016), and climatology (Carr, Bell, Killick, & Holt, 2017). A survey of changepoint methods can be found in Tartakovsky, Nikiforov, and Basseville (2014). State-of-the-art methods for identifying multiple changepoints in univariate data include the pruned exact linear time method (PELT; 2012), seeded binary segmentation (SBS; Kovács, Li, Bühlmann, and Munk 2020), and the simultaneous multiscale changepoint estimator (SMUCE; Frick, Munk, and Sieling 2014). In the case of mean changepoints, most state-of-the-art methods focus their attention on the case where the error structure is IID with known variance.

There is comparatively little attention given to the problem of detecting changes in mean whereas the second-order structure is unknown, nontrivial and must be estimated. Further still, there is even less attention given to the case where this second-order structure is allowed to vary over time. One approach for dealing with autocorrelation when testing for changes in mean is to increase the threshold above which changes are detected; see Lavielle (1999), for example. However, it is difficult to systematically choose the threshold without performing some sort of preestimation of the autocovariance, which is highly challenging in the presence of mean changes (Lund & Shi, 2020). If one simply raises the threshold, then there will be a trade-off between a decreased false positive rate and a decreased true positive rate. Further, using standard autocovariance estimation techniques will result in bias, and subsequently use of these biased estimates can significantly reduce the efficiency of changepoint estimators, as described in Section 3 of Jandhyala, Fotopoulos, MacNeill, and Liu (2013).

One can estimate the autocovariance structure and incorporate it within the changepoint detection algorithm, for example, Dette, Eckle, and Vetter (2020) consider modifications to the SMUCE algorithm for dependent, not necessarily Gaussian data. Tecuapetla-Gómez and Munk (2017) estimate the autocovariance structure using a difference-based approach under the assumption of m-dependent errors, which is then used in order to estimate a piecewise constant signal. Chakar, Lebarbier, Lévy-Leduc, and Robin (2017) study the problem of estimating mean changes with an underlying AR(1) process. Lavielle (1999) considers dependent processes and shows that the least-squares estimators of the changepoint locations, and of the parameters of each segment, are consistent under mild conditions; however, the number of changepoints is assumed to be known. The narrowest-over-threshold (NOT) approach of Baranowski, Chen, and Fryzlewicz (2019) is shown to be consistent in the presence of short memory dependence; however, the autocorrelation is assumed known. Kokoszka and Leipus (1998) use a cumulative sum approach to detect changes in mean with dependent errors, whereas Bücher and Kojadinovic (2016) use a nonparametric bootstrap approach to detect changepoints in structures including the mean. Wu and Zhou (2020) detect changes in smooth mean functions under nonstationary errors, whereas Pešta and Wendler (2019) use self-normalization to detect a single change in mean. With the exception of Pešta and Wendler (2019) and Wu and Zhou (2020), all of the above approaches make the assumption of second-order stationarity.

In practice, the autocovariance structure of a time series may be time-varying and is typically not known a priori. Failure to take into account this autocovariance can lead to serious errors in the changepoint detection procedure. In particular, the presence of autocorrelation may lead to overfitting or underfitting of mean changes. The approach described herein enables the detection of changes in mean while allowing for unknown, possibly time-varying second-order structure, by modelling the error process of the time series as a locally stationary wavelet (LSW) process (Nason, von Sachs, & Kroisandt, 2000). The LSW model enables the modelling of a zero-mean time series with slowly evolving, time-varying second-order structure: an overview of the necessary concepts can be found in Section 2.2. The LSW model is a flexible framework that can incorporate a wide variety of error processes—both stationary and nonstationary—for example, ARMA and GARCH models.

In this article, we describe a likelihood-based test, in which the test statistic incorporates the unknown and potentially time-varying autocovariance of the time series. In order to utilize the theory of LSW processes, which is designed for zero-mean processes, we adapt the estimation procedure of Nason et al. (2000) to our piecewise constant mean structure. This is used in a standard likelihood ratio framework to test for changes in mean in the observed time series. By including the autocovariance information, we can both lower the false discovery rate by accurately distinguishing between mean changes and patterns due to autocorrelation and increase the accuracy of true change detection. Furthermore, we demonstrate that allowing for the possibility of nontrivial autocovariance does not significantly lower the power of the likelihood ratio test when the underlying noise in the time series is in fact independent. This approach provides practitioners with a simple yet flexible alternative to assuming an IID error structure without incurring significant loss of power.

The remainder of the article is organized as follows. Section 2 recalls the likelihood ratio test for detecting single changes in mean and introduces the paradigm for our methodology. Section 3 describes the method for estimating the autocovariance of the time series and discusses our final likelihood-based test for mean changepoint detection. The performance of the method is tested and compared to other methods via a simulation study in Section 4 and applied to two data examples in Section 5. Concluding remarks are given in Section 6.

2 | BACKGROUND

In this section, we give the necessary background on changepoint detection and locally stationary wavelet process modelling, which will form the basis of our approach.

2.1 | Detecting a single change in mean

We aim to detect a finite number of changes in mean within a time series of length T. We begin by considering the problem of detecting a single change in mean, using a likelihood-based framework. The likelihood-based approach is one of the most common and well studied methods used within the changepoint literature: for a thorough review, see, for example, Chen and Gupta (2011). Assume that the time series is given by

$$X_t = \mu_t + \epsilon_t, \tag{1}$$

for $t = 0, \dots, T-1$. The mean function is given by μ_t and the error terms ϵ_t are Gaussian, zero-mean random variables with a variance-autocovariance matrix Σ. The hypothesis to test for a single change in mean at time p is

$$H_0 : \mu_0 = \mu_1 = \dots = \mu_{T-1}$$
$$H_1 : \mu_0 = \dots = \mu_p \neq \mu_{p+1} = \dots = \mu_{T-1}.$$

Using this formulation, we can derive the likelihood ratio test for a single change in mean. Let $X = (X_0, \dots, X_{T-1})^{\mathsf{T}}$ be our time series. Under the assumption that the ϵ_t are Gaussian, the log-likelihood of the time series under the null hypothesis is given by

$$l(\mu, \Sigma|X) = \frac{T}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(X - \mu_0)^{\mathsf{T}}\Sigma^{-1}(X - \mu_0),$$

where $\mu_0 = (\mu_0, \dots, \mu_0)^{\mathsf{T}}$ is the (constant) mean vector of the series and Σ is the variance-autocovariance matrix, with entries given by $\Sigma_{ts} = \mathrm{Cov}(X_t, X_s)$. The log-likelihood under the alternative hypothesis is given by

$$l(\mu, \Sigma, p|X) = \frac{T}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(\mathbf{X} - \mu_1)^{\mathsf{T}}\Sigma^{-1}(\mathbf{X} - \mu_1),$$

where $\mu_1 = (\mu_0, \dots, \mu_0, \mu_{T-1}, \dots, \mu_{T-1})^{\mathsf{T}}$ is the mean vector with change at time p. Note that Σ is the same under both hypotheses and is not assumed to change at the changepoint p. Hence, the likelihood ratio test statistic for a single change in mean, after plugging in estimators for the unknown $\boldsymbol{\mu}_0$, $\boldsymbol{\mu}_1$, and Σ, is given by

$$\lambda \propto \max_{1<p<T-2} \left\{ (X - \hat{\mu}_0)^{\mathrm{T}} \hat{\Sigma}^{-1} (X - \hat{\mu}_0) - (X - \hat{\mu}_1)^{\mathrm{T}} \hat{\Sigma}^{-1} (X - \hat{\mu}_1) \right\}. \tag{2}$$

Using (2), a changepoint is detected when $\lambda > c$, where c is a constant. If a change is detected, then the location of the changepoint, $\hat{p}$, is estimated as the value of p that maximizes λ. Note that a similar test statistic (with zero mean and changepoint matrix Σ) was applied to the problem of detecting changes in a piecewise constant autocovariance function, in Killick et al. (2013). Under the simplifying assumption of IID Gaussian errors with known variance, this reduces to the standard likelihood test for a change in mean, equivalent to the cumulative sum (CUSUM) test statistic. The appropriate threshold c can be chosen using the asymptotic distribution of (2).

In our more general scenario, we must estimate the unknown autocovariance matrix Σ. Under the null hypothesis, this is straightforward; however, under the alternative hypothesis, this becomes highly challenging as any standard estimation procedure will be biased due to the presence of the mean change. This is why several authors have assumed a known (autoco)variance, as once we have a sensible estimate we can then calculate λ and test for a change in mean in the standard way.

2.2 | Modelling series with time-varying autocovariance

As discussed in Section 2.1, in order to use the likelihood-based framework, we must estimate the unknown autocovariance of the time series. To this end, we represent the ϵ_t in Equation (1), the error component of the time series, as a locally stationary wavelet (LSW) process. The LSW model was introduced in Nason et al. (2000) and provides methodology for modelling time series whose second-order structure evolves slowly over time. Wavelets are useful in estimating time varying quantities as they are compactly supported oscillatory functions that can be translated and dilated efficiently to provide location-scale decompositions. For an overview of wavelet techniques, see, for example, Nason (2008) or Vidakovic (2009).

In the remainder of this section, we give a brief overview of the LSW framework. Following (Nason et al. 2000), a triangular stochastic array $\{\epsilon_{t,T}\}_{t=0}^{T-1}$ for $T = 1, 2, \ldots$ is in a class of locally stationary wavelet (LSW) processes if there exists a mean-square representation

$$\epsilon_{t,T} = \sum_{j=-\infty}^{-1} \sum_{k} w_{j,k;T} \psi_{j,k-t} \xi_{j,k}, \tag{3}$$

where j and $k \in \mathbb{Z}$ are scale and location parameters, respectively, and the $\xi_{j,k}$ are zero-mean, orthonormal identically distributed random variables. The $\{\psi_{j,k-t}\}_{j,k}$ are a set of discrete nondecimated wavelets, and the $\{w_{j,k;T}\}$ are a set of time-varying amplitudes. There exists, for each $j \leq -1$, a Lipschitz continuous function $W_j(z)$ for $z \in (0, 1)$ which satisfies $\sum_j |W_j(z)|^2 < \infty$ uniformly in $z \in (0, 1)$. The Lipschitz constants L_j are uniformly bounded in j and $\sum_j 2^{-j} L_j < \infty$. There exists a sequence of constants C_j such that for each T

$$\sup_k \left| w_{j,k;T} - W_j(k/T) \right| \leq C_j T^{-1}, \tag{4}$$

where, for each $j \leq -1$, the supremum is over $k = 0, \ldots, T-1$, and where the sequence $\{C_j\}$ satisfies $\sum_j C_j < \infty$.

We further assume that the increment sequence $\{\xi_{j,k}\}$ is Gaussian, which ensures that $\epsilon_{t,T}$ is Gaussian. As with classical time series theory, the second-order structure of an LSW process is encoded in the spectrum. The evolutionary wavelet spectrum (EWS) of an LSW process is defined as $S_j(z) = |W_j(z)|^2$ for rescaled time $z = k/T$ and measures the contribution to variance at a particular rescaled time z and scale j. Because the W_j are Lipschitz continuous, the spectrum at level j, S_j, is also Lipschitz continuous. However, Fryzlewicz and Nason (2006) and Van Bellegem and von Sachs (2008) extend the LSW model to consider piecewise constant spectra and those of bounded total variation, respectively. No further assumptions are required on the dependence structure of the time series, beyond the requirements of the LSW model. Note that LSW processes include the class of stationary processes with absolutely summable autocovariance $\sum_\tau |c_X(\tau)| < \infty$. Hence, this modelling framework can account for a large variety of classes of short memory stationary process, such as ARMA and GARCH models.

In Nason et al. (2000), rigorous estimation theory for the EWS and time-varying autocovariance of a zero-mean series are outlined, which we will now briefly describe here. The EWS is estimated via the empirical wavelet coefficients of the time series, given by $d_{j,k} := \langle \epsilon_{t,T}, \psi_{j,k-t} \rangle = \sum_t \epsilon_{t,T} \psi_{j,k-t}$. As with Fourier approaches, the raw wavelet periodogram $I_k^j := |d_{j,k}|^2$ is a biased, inconsistent estimator of the EWS ((Nason et al. 2000), Proposition 4):

$$\mathbb{E}\left(I_k^j\right) = \sum_l A_{jl} S_l(k/T) + \mathcal{O}(T^{-1}), \tag{5}$$

$$\mathrm{Var}(I_k^j) = 2\left(\sum_l A_{jl} S_l(k/T)\right)^2 + \mathcal{O}(2^{-j} T^{-1}), \tag{6}$$

where the operator $A = (A_{jl})_{j,l<0}$ is given by $A_{jl} := \langle \Psi_j, \Psi_l \rangle = \sum_\tau \Psi_j(\tau) \Psi_l(\tau)$, and the autocorrelation wavelets are defined by $\Psi_j(\tau) := \sum_{k \in \mathbb{Z}} \psi_{j,k} \psi_{j,k-\tau}$, $j < 0$, $\tau \in \mathbb{Z}$.

The vector of periodograms $\mathbf{I}(z) := \{I^{l}_{\lfloor zT \rfloor, T}\}_{l=-1,\dots,-J}$, $J = \lfloor \log_2(T) \rfloor$ is bias-corrected using the curtailed J-dimensional matrix $A_J := (A_{jl})_{j,l=-1,\dots,-J}$ which is shown to be invertible in Nason et al. (2000). The raw wavelet periodogram is first smoothed and then corrected by A_J^{-1} to produce an asymptotically unbiased, consistent estimator. Smoothing can be carried out using a number of techniques, for example, via a running mean as in Nason (2013) or using wavelet thresholding as in Nason et al. (2000).

The local autocovariance (LACV) function, $c(z, \tau)$, of an LSW process with EWS $\{S_j(z)\}$ is defined as $c(z, \tau) = \sum_{j=-\infty}^{-1} S_j(z)\Psi_j(\tau)$, for $\tau \in \mathbb{Z}, z \in (0, 1)$. The LACV can be thought of as a decomposition of the autocovariance of a process over scales and rescaled time locations. In practice, the local autocovariance is estimated by plugging in the smoothed, corrected estimate for the EWS into the LACV definition; this results in a consistent estimator of the LACV (Nason et al. 2000).

3 | METHOD FOR DETECTING MEAN CHANGES IN THE PRESENCE OF TIME-VARYING AUTOCOVARIANCE

In this section, we discuss our proposed methodology for estimating the time-varying autocovariance in the presence of mean changes, and the likelihood-based test for mean changepoint detection.

3.1 | Autocovariance estimation in the presence of mean changes

Having assumed that the error process of the time series is LSW, we now discuss the methodology for estimating the autocovariance of this process. In Section 2.2, we recalled the standard estimation procedure for estimating the EWS and hence the autocovariance; however, this method is only valid in the case where the time series has a constant mean. A key ingredient in our approach is providing methodology to estimate the nonstationary second-order structure of a time series under the relaxed assumption of a piecewise constant mean signal.

Under the alternative hypothesis, the overall model for X_t is not LSW due to changes in mean. Wavelet coefficients informally describe weighted localized changes in average; therefore, changes in mean will cause the raw wavelet periodogram to become contaminated with information about the mean, rather than purely information about the spectrum. This will cause additional bias in the raw wavelet periodogram. If we were to use, for example, a running mean to smooth the wavelet periodogram—as is standard in LSW spectral estimation to achieve consistency—this would result in a biased estimator.

To illustrate this phenomenon, in Figure 1 left, we see the empirical Haar wavelet periodogram of a piecewise constant function of length 512 with a single jump in value at time point 300, without noise. Coefficients localized around the jump contain information about the change, which propagates through to coarser scales in a cone of influence. In time locations where the coefficients are localized entirely within one of the two piecewise constants, the coefficients are unaffected by the jump and are exactly zero. Hence, for a fixed scale, the empirical wavelet periodogram will be contaminated with the jump in a fixed number of coefficients, localized around the jump's location. All other coefficients in that scale will be unaffected. In particular, ignoring boundary effects, for the Haar wavelet, at scale j, 2^{-j} coefficients will be affected by a single jump.

As described in Nason et al. (2000), asymptotically, as $T \to \infty$, we observe more data at an increasingly finer level on the interval $(0, 1)$. Therefore, as $T \to \infty$, we do not consider an increasing number of jumps in the mean, but rather, we are observing a (finite) fixed number of jumps and observe more information about their structure as T increases. Suppose there are m changes in mean within the series. If we consider a fixed scale j^*, then a maximum number of wavelet coefficients, $Cm2^{-j^*}$, for some constant C, will be affected by the finite number of jumps in the mean of the series, due to the compact support of the wavelet. (For the Haar wavelet, $C = 1$). As T increases, an increasing number of the coefficients will contain information about the wavelet spectrum that is uncontaminated by the jumps.

Therefore, we wish to use these coefficients as an estimator for the wavelet spectrum. Provided the number of contaminated coefficients is small relative to the number of time points, a natural alternative to the running mean smoother—which would be affected by the jumps of the time series—is to use a running median instead. This is analogous to robust estimation of the variance using a median absolute deviation (MAD)

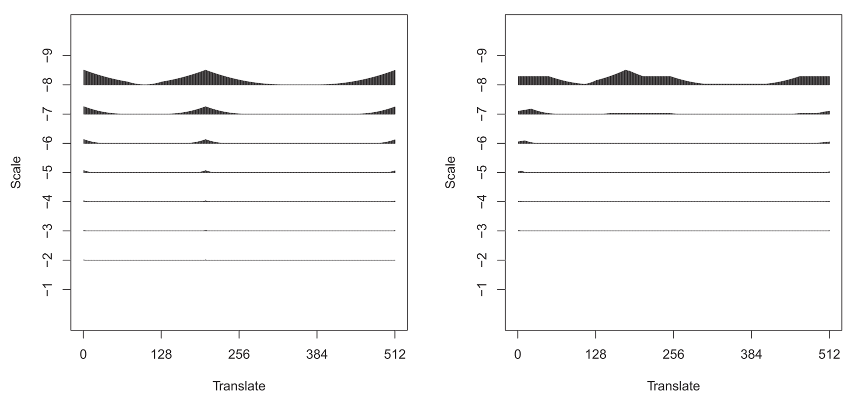

FIGURE 1 Left: empirical wavelet periodogram of a piecewise constant signal with one jump. Right: median smoothed version

estimator. By taking a median, we can reduce the influence that the jumps have on the spectrum estimator. The running median smoothed wavelet periodogram, with bin width $2n+1$, is given by

$$\tilde{I}_k^j = \text{Median}\left(I_{k-n}^j, I_{k-n+1}^j, \dots, I_{k+n-1}^j, I_{k+n}^j\right). \tag{7}$$

In Figure 1 right, we see the running median smoothed version of Figure 1 left, with a bin width of size 201. We see that the median smoothing is able to negate the effect of the change in mean of the signal on the raw wavelet periodogram, at least in the finer scales.

The above discussion motivates the following autocovariance estimation procedure. First, compute the raw wavelet periodogram I_k^j of the time series. Then, for each scale of interest j, obtain the smoothed wavelet periodogram $\tilde{I}_k^j$ by computing a running median of the I_k^j. Next, we must correct this median estimate to ensure unbiasedness. The raw wavelet periodogram is distributed asymptotically as a scaled χ_1^2 random variable, so we divide the smoothed wavelet periodogram by a suitable scale factor. This is analogous to the scale factor multiplication that occurs when using an MAD estimator. In our case, we divide the median smoothed estimate by $(7/9)^3 \approx 0.471$, as the mean of a χ_1^2 random variable is equal to 1, whereas the median is equal to 0.471. Denote the scaled, median smoothed estimate of I_k^j by $\hat{I}_k^j$. Next, we correct this estimate using the inverse of the A matrix to give the final estimate of the EWS:

$$\hat{S}_j(k/T) = \sum_{l=-J_0}^{-1} A_{jl}^{-1}\hat{I}_k^l, \tag{8}$$

for some $J_0 = \alpha \log_2 T < J$, where $\alpha \in (0,1)$. Lastly, we estimate the autocovariance by plugging in the EWS estimate into the equation for the local autocovariance of an LSW process, that is,

$$\hat{c}(k/T,\tau) = \sum_{j=-J_0}^{-1} \hat{S}_j(k/T)\Psi_j(\tau). \tag{9}$$

3.1.1 | Choice of contributing scales and bin width

We must decide on a value of α, the proportion of scales over which we correct. Higher values of α ensure a decomposition over a larger number of scales. In practice, not all of the scales will be informative. We propose using the proportion $\alpha = 3/5$. This is slightly less than the proportion $2/3$ discussed in Sanderson, Fryzlewicz, and Jones (2010), reflecting the fact that the time series may display mean changes.

The choice of the bin width parameter n in Equation (7) will also affect the quality of the estimate. We can choose to use level-dependent smoothing of the periodogram, and so a larger bin width is used in coarser scales. Choosing a bin width in this way is natural for two reasons: coarser scales will experience stronger autocorrelation, and coarser scales have a larger number of coefficients affected by potential mean changes. If, instead, we assume second-order stationarity, we can drop the dependence on k in the estimation Equations (8) and (9). In this case, we can use a global scale-wise median when smoothing the wavelet periodogram, obtaining a time-independent estimate of the EWS and autocovariance.

3.2 | Mean change detection

Having discussed our estimator of the autocovariance of the time series in the presence of mean changes, we now return to the problem of detecting the mean changes within a likelihood-based framework.

3.2.1 | Likelihood-based test statistic

Recall that the likelihood ratio test statistic for a single change in mean is given by

$$\lambda = \max_{1<p<T-2}\left\{(X-\hat{\mu}_0)^\mathsf{T}\hat{\Sigma}^{-1}(X-\hat{\mu}_0) - (\mathbf{X}-\hat{\mu}_1)^\mathsf{T}\hat{\Sigma}^{-1}(\mathbf{X}-\hat{\mu}_1)\right\}.$$

In our approach, we also use sample means for the mean estimates $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$. The sample mean is asymptotically consistent, provided that the LSW process satisfies $\sup_{z\in(0,1)}\sum_\tau |c(z,\tau)| < \infty$, which can be thought of as a short memory assumption. The entries of the variance-autocovariance matrix Σ are estimated using Equation (9).

The consistency of likelihood methods for detecting changepoints has been shown in Csörgö and Horváth (1997). It is well known that plug-in estimates that are consistent possess the same properties as maximum likelihood estimates; therefore, the proof of consistency for identifying changepoints using plug-in estimates follows directly provided the estimates used are consistent. In our case, the sample means are consistent if we assume that the process has absolutely summable autocovariance. The mean smoothed estimate of the autocovariance is consistent, so intuitively, we would expect the median to be also, although we give no formal proof of this here.

In practice, the variance-autocovariance matrices can be ill-conditioned and cause the test statistic to become numerically unstable. To combat this issue, we regularize the covariance matrix estimate using the approach of (Rothman, 2012), one of many possible regularization methods. Denote by R the correlation matrix calculated using the initial covariance matrix estimate $\hat{\Sigma}$, and let $\theta > 0$ indicate that θ is symmetric and positive definite. Let θ^+ be the diagonal matrix with the same diagonal entries as θ. The method utilizes a lasso-type procedure and computes an estimate of the correlation matrix θ given by

$$\hat{\theta} = \underset{\theta > 0}{\operatorname{argmin}} \left\{ ||\theta - R||_F/2 - \tau \log|\theta| + \gamma|\theta^+ - \theta|_1 \right\}, \tag{10}$$

where the subscript F denotes the Frobenius norm, $\tau = 10^{-4}$ is a fixed constant, and $\gamma \geq 0$ is a tuning parameter. We recommend choosing γ as low as possible while still ensuring positive-definiteness, in order to capture the largest amount of autocovariance information within the series.

3.2.2 | Detecting multiple changes in mean

In the previous sections, we have focussed on the single change in mean case; however, often one is interested in detecting multiple changes. The likelihood ratio test statistic is consistent when estimating a single changepoint in the presence of multiple changepoints, as shown in Vostrikova (1981). There are many proposed methods in the changepoint literature which tackle the multiple changepoint problem, for example, Scott and Knott (1974), Bai and Perron (2003), Killick et al. (2012), and Kovács et al. (2020).

In our work, we implement the binary segmentation algorithm of Scott and Knott (1974). Binary segmentation is a technique for multiple changepoint detection where initially the entire data set is searched for a changepoint. If a changepoint is detected, the data are split into two subsegments defined by the detected changepoint. This procedure is then recursively repeated on subsequent subsegments until no further changepoints are detected. The algorithm is outlined in detail in Algorithm 1 below.

Algorithm 1 Binary Segmentation algorithm.

Input: A set of data of the form $X = \{X_0, X_1, \ldots, X_{T-1}\}$. A test statistic $\lambda(\cdot)$ dependent on the data. An estimator of the changepoint location $\hat{p}(\cdot)$. A rejection threshold (penalty) c.

Initialise: Let $\mathcal{C} = \emptyset$, and $\mathcal{S} = \{[0, T-1]\}$

while $\mathcal{S} \neq \emptyset$ **do**
- Choose an element of $\mathcal{S}$, denoted $[s, t]$.
- If $\lambda(X_{s:t}) < c$, remove $[s, t]$ from $\mathcal{S}$
- **if** $\lambda(X_{s:t}) \geq c$ **then**
  - Remove $[s, t]$ from $\mathcal{S}$
  - Calculate $r = \hat{p}(X_{s:t}) + s - 1$, and add r to $\mathcal{C}$
  - If $r \neq s$ add $[s, r]$ to $\mathcal{S}$
  - If $r \neq t - 1$ add $[r + 1, t]$ to $\mathcal{S}$

Output: the set of changepoints recorded $\mathcal{C}$.

3.2.3 | Choice of threshold/penalty

In changepoint problems, it is common practice to use a penalty term to protect against overfitting, that is, fitting too many changepoints. In the multiple changepoint setting, the penalty is usually a function of the number of changepoints, for example, see Harchaoui and Lévy-Leduc (2010). We outline two possible approaches for the selection of the appropriate penalty/threshold c.

First, one option is to use the asymptotic distribution of the test statistic to inform an appropriate value for c. If the autocovariance matrix was known, then an appropriate choice of penalty would be $2\log T$. However, in our scenario, the autocovariance is unknown and must be estimated. We therefore propose that a practical choice of threshold is given by $3\log T$, which is motivated by the modified Bayesian information criterion as used in Zhang and Siegmund (2007). As remarked in Fisch, Eckley, and Fearnhead (2018), we utilize an inflated penalty in order to reflect the uncertainty of the unknown parameters. We have found in our empirical analysis that this threshold works well in practice, allowing for a low false discovery rate and a high detection rate.

In the second approach, we choose the threshold c via a Monte Carlo procedure. Using the estimate of the evolutionary wavelet spectrum of the time series, Monte Carlo simulations allow us to compute an appropriate $(100-\alpha)$%-quantile, which we choose to be c. We then detect a change if the test statistic of the observed series is greater than c.

The Monte Carlo approach works well in practice and provides a fully automated method for detecting changes. However, it may become prohibitively slow, due to the need to calculate an inverse covariance matrix for each simulated series. To address this problem, we can instead use a modified test statistic, in which no inverse calculation is required. We use the alternative test statistic

$$\lambda = \max_{1<p<T-2} \left\{ \frac{1}{\hat{\sigma}_L^2(p)} \left((X - \hat{\mu}_0)^{\mathsf{T}}(X - \hat{\mu}_0) - (\mathbf{X} - \hat{\mu}_1)^{\mathsf{T}}(\mathbf{X} - \hat{\mu}_1) \right) \right\},$$

where $\hat{\sigma}_L^2(p)$ is an estimate of $\sigma_L^2(p) = \sum_\tau c(p/T, \tau)$, the possibly time-varying long run variance of the time series. We estimate this via the estimate of the local autocovariance and weighting via a kernel approach, in a similar fashion to Tecuapetla-Gómez and Munk (2017). The estimator is given by

$$\hat{\sigma}_L^2(p) = \hat{c}(p/T, 0) + 2\sum_{\tau=1}^{2^{J_0}} \left(1 - \frac{\tau}{2^{J_0}+1}\right) \hat{c}(p/T, \tau).$$

The kernel approach is a common technique for estimation of the long run variance (Newey & West, 1986) and helps to ensure positive-definiteness of the resulting estimate. This approach enables faster implementation of the Monte Carlo threshold choice, while still taking the autocorrelation of the time series into account in the procedure. The approach of using the long run variance estimate in a changepoint detection procedure is also used, for example, in Dette et al. (2020) and Dette and Wu (2019).

Collecting the above ideas together, Algorithm 2 describes our LSW likelihood-based method which we henceforth refer to as LSWL. The output of Algorithm 2 is then used as the input for binary segmentation (Algorithm 1) in order to detect multiple changes in mean.

Algorithm 2 LSWL algorithm

Input: A set of data of the form $X = \{X_0, X_1, \ldots, X_{T-1}\}$. An estimate of the LACV, $\hat{c}$. Number of simulations, nsim, if using Monte Carlo method. Significance level, α, if using Monte Carlo method.

if *Method = Monte Carlo* **then**
  Calculate $\hat{\sigma}^2_L$, λ, $\hat{p}$
  for $n \in \{1, \ldots, \text{nsims}\}$ **do**
    Simulate X_n
    Calculate $\hat{\sigma}^2_{L,n}$, λ_n
  $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_{\text{nsims}}\}$, $c = q_{1-\alpha}(\boldsymbol{\lambda})$
else
  $c = 3 \log T$
  Calculate $\hat{\Sigma}$, λ, $\hat{p}$

Output: the test statistic λ, the threshold c, the estimated change location $\hat{p}$.

4 | SIMULATION STUDY

In this section, we assess the performance of LSWL via a simulation study. We simulate time series, with and without changes in mean, under a variety of error structures. In all simulations, the results were obtained using Haar wavelets and where applicable Haar wavelets were used to simulate from wavelet spectra. Other wavelet families yield similar results. In all simulations, 100 realizations were simulated, all of length 512. In each of the scenarios listed below the innovations, ϵ_t are standard normal random variables.

If the estimated autocovariance matrix is numerically stable and can be regularized using the penalty value $\gamma = 0.4$, we utilize the standard version of LSWL; otherwise, we use the Monte Carlo version with 100 replications, at a significance level of 5%. In alignment with the discussion in Section 3.1, we estimate the wavelet spectrum at the finest $J_0 = 5$ scales. For the second-order stationary scenarios in the simulation study, we utilize global scale-wise median smoothing for spectral estimation. For the nonstationary scenarios, we use a running median of bin width 151 at each scale for simplicity. Our implementation of LSWL utilizes the `R` package `wavethresh` (Nason, 2016) in order to fit the LSW model.

Where applicable, we compare our method to AR1Seg (Chakar et al. 2017) and the NPCP method of (Bücher & Kojadinovic, 2016), whose implementations are readily available in `R` (Chakar, Lebarbier, Lévy-Leduc, & Robin, 2014; Kojadinovic, 2015). AR1Seg can detect multiple mean changes under (stationary) AR(1) errors, whereas NPCP can detect a single change in the presence of stationary autocorrelation subject to mild technical assumptions. In our simulations, the significance level for NPCP is set at 5%. For illustrative purposes, we also compare to the performance of NOT, implemented in the `R` package `not` (Baranowski et al. 2016), to highlight the effect of not incorporating the autocorrelation within the changepoint method. NOT is shown to be consistent in the presence of autocorrelated errors, provided the autocorrelation is preestimated and included within the methodology. However, there is no option for including autocorrelation information within the implementation.

4.1 | Null hypothesis performance

We first assess the performance of the method under the null hypothesis of a constant mean (without loss of generality equal to zero) of the time series. We simulate a variety of stationary and nonstationary error structures that highlight the ability of LSWL to successfully account for autocorrelation in order to maintain a low false positive rate. In each simulation, a false positive is recorded if the method detected any changes, and we record the proportion of the 100 simulations resulting in a false positive.

4.1.1 | No changepoint, various AR(1) parameters

We investigate the effect of varying AR(1) parameter values on the false positive rate of the methods. Data sets are simulated from

$$X_t = \phi X_{t-1} + \epsilon_t,$$

where ϕ is given by the values in Table 1. Included is the case of white noise where $\phi = 0$. The values underlined indicate performance where the type 1 error is clearly not controlled. We see from Table 1 that negative values for the AR(1) parameter do not hamper the performance of the methods, whereas large positive values do. Furthermore, the results show that LSWL compares similarly or favourably with AR1Seg and

TABLE 1 False discovery rates across different AR(1) scenarios with constant mean equal to zero

False discovery rate	AR(1) parameter						
Method	**−0.9**	**−0.6**	**−0.3**	**0**	**0.3**	**0.6**	**0.9**
LSWL	0.00	0.00	0.00	0.00	0.08	0.09	0.05
AR1Seg	0.01	0.02	0.01	0.02	0.05	0.03	<u>0.45</u>
NPCP	0.01	0.02	0.02	0.05	0.07	0.05	0.03
NOT	0.00	0.00	0.00	0.01	<u>0.34</u>	<u>0.99</u>	<u>1.00</u>

NPCP. AR1Seg has a high false positive rate when the AR parameter is equal to 0.9. As expected, NOT confuses changes in mean with strongly autocorrelated errors.

4.1.2 | No changepoint, various autocorrelation scenarios

Next, we investigate a variety of autocorrelation scenarios, representing some of the most common time series models. Data sets are simulated from models A–H, as described below:

(A) AR(2) model with parameters $\phi_1 = 0.5$ and $\phi_2 = 0.3$.
(B) ARCH(3) model with parameters $\alpha_1 = 1$, $\alpha_2 = 0.5$, $\alpha_3 = 0.4$.
(C) Stationary LSW model with $S_{-4} = S_{-5} = 1$, $S_j = 0$ otherwise.
(D) ARMA(1,6) model with AR parameter $\phi = 0.5$, MA parameters $\theta_1 = 1$, $\theta_2 = -1$, $\theta_3 = 0.5$, $\theta_4 = 0.5$, $\theta_5 = 1$, $\theta_6 = 0.5$.
(E) Time-varying AR(1) model where the AR(1) parameter moves from 0.7 to 0.3 linearly, that is,

$$X_t = \phi_t X_{t-1} + \epsilon_t, \quad \phi_t = \frac{0.4(1-t)}{511} + 0.7.$$

(F) Nonstationary LSW model with spectrum given by

$$S_j(z) = \begin{cases} 1/5 + \sin^2(2\pi z) & \text{for } j = -5, z \in (0, 1), \\ 1 & \text{for } j = -1, z \in (100/512, 200/512), \\ 0 & \text{otherwise.} \end{cases}$$

(G) Nonstationary LSW model with spectrum given by

$$S_j(z) = \begin{cases} 1 & \text{for } j = -1, z \in (0, 100/512), \\ 1 & \text{for } j = -2, z \in (100/512, 300/512), \\ 1 & \text{for } j = -3, z \in (300/512, 1), \\ 0 & \text{otherwise.} \end{cases}$$

(H) $X_t = \sigma_t \epsilon_t$, where

$$\sigma_t = \begin{cases} 24(t/512)^2 + 4(t/512) + 4 & \text{for } t \in [0, 299], \\ -32(t/512)^2 + 8(t/512) + 7.62 & \text{for } t \in [300, 511]. \end{cases}$$

The scenarios (A)–(D) represent second-order stationary time series, whereas scenarios (E)–(H) are second-order nonstationary. The results are reported in Table 2, again with underlined values indicating an uncontrolled type 1 error. We see that LSWL is able to maintain a low false positive rate in all the scenarios, comparing similarly or favourably to NPCP. Although we expect our method under models C, F, and G to perform well (as they are generated from LSW processes), it is reassuring to find good performance in all the cases. AR1Seg performs poorly in the case of an AR(2) model and ARCH(3) model, and model F. As expected, NOT generally performs poorly due to autocorrelation, while performing well in the presence of nonstationary IID errors (model H).

False discovery rate	Scenario							
Method	A	B	C	D	E	F	G	H
LSWL	0.02	0.07	0.00	0.01	0.09	0.00	0.00	0.00
AR1Seg	0.97	0.96	0.11	0.08	0.14	0.94	0.05	0.15
NPCP	0.07	0.03	0.00	0.06	0.05	0.00	0.00	0.07
NOT	1.00	0.60	1.00	1.00	0.92	1.00	0.14	0.03

TABLE 2 False discovery rates across different autocorrelated error scenarios, with constant mean equal to zero

		No. of changepoints				
Model	Method	0	1	2	≥3	MSE
IID	LSWL	0	**98**	2	0	**0.010**
	AR1Seg	0	**94**	4	2	0.012
	NPCP	0	**100**	—	—	**0.010**
	NOT	0	**96**	4	0	0.011
AR(1) 0.6	LSWL	0	**95**	5	0	0.070
	AR1Seg	27	**64**	7	2	0.165
	NPCP	4	**96**	—	—	**0.056**
	NOT	0	2	4	**94**	0.441
A	LSWL	12	**76**	12	0	0.291
	AR1Seg	0	2	5	**93**	1.652
	NPCP	35	**65**	—	—	**0.174**
	NOT	0	0	0	**100**	1.040
B	LSWL	0	**71**	27	2	0.180
	AR1Seg	0	5	1	**94**	2.044
	NPCP	0	**100**	—	—	**0.061**
	NOT	0	40	2	**58**	1.013
C	LSWL	2	**85**	12	1	0.060
	AR1Seg	**61**	22	8	9	0.374
	NPCP	0	**100**	—	—	**0.034**
	NOT	0	0	0	**100**	1.26
D	LSWL	6	**80**	14	0	0.616
	AR1Seg	**49**	40	5	6	1.226
	NPCP	9	**91**	—	—	**0.363**
	NOT	0	0	1	**99**	3.415

TABLE 3 Method performance comparisons for the single changepoint and stationary second-order scenarios, reporting number of changepoints detected and average mean squared error of the estimated mean

4.2 | Alternative hypothesis assessment

Next, we assess the method's performance in detecting changes in mean. We use the same error structure scenarios defined in the previous section and examine the performance of the method in both the single and multiple changepoint settings.

4.2.1 | Single changepoint, various stationary autocorrelation scenarios

In these scenarios, there is a single change in mean at time 300, with the size of the change, δ, chosen so that $\delta/\sigma = 1$, where σ is the standard deviation of the time series. We investigate the performance of the methods for the IID model, AR(1) model with parameter $\phi = 0.6$, and models (A)–(D), by reporting the number of detected changepoints of each method. We also report the average mean squared error (MSE) of the estimated signal $\hat{\mu}$ across the 100 simulations, where the MSE is given by

$$\text{MSE} = \frac{1}{T}\sum_{t=0}^{T-1}(\mu_t - \hat{\mu}_t)^2.$$

The MSE is an alternative comparison to examine how well each method estimates the mean function in the model. The results are given in Table 3, with the correct number of changepoints in the top row in bold and the modal value for each method also in bold. The table indicates that LSWL performs well across all considered scenarios. Surprisingly, LSWL outperforms AR1Seg in the case of AR(1) errors. Departure from an AR(1) error structure results in poor performance for AR1Seg. LSWL also performs well in the IID case, in part due to the ability to obtain an accurate estimate of the trivial autocovariance. Note that NPCP is capable of detecting only a single change: running the simulations for at-most-one-change LSWL gives similar results to NPCP.

4.2.2 | Single changepoint, time-varying autocorrelation scenarios

In these scenarios, there is again a single change in mean at time 300, with the size of the change, δ, chosen so that $\delta/\sigma = 1$, where σ is the maximum standard deviation of the time series. We investigate the performance of the methods for models (E)–(H), with the results reported in

Table 4. We see that LSWL performs well in general, with a high proportion of simulations identifying the correct number of changepoints. NPCP also performs well, despite the fact that it is not designed to handle nonstationary second-order structure.

4.2.3 | Multiple changepoints, various stationary autocorrelation scenarios

Next, we analyse the performance of the methods in the case where there are multiple changepoints, with the error structures the same as in the single changepoint case. In these simulations, the signal contains three changepoints, at times $p = 100$, 180, and 380. For the IID and AR(1) case, the mean within each segment alternates between 0 and δ, where δ is calculated so that $\delta/\sigma = 1.25$. For cases (A)–(D), δ is calculated so that $\delta/\sigma = 2$, to reflect the difficulty of the scenarios. Note that this up, down, up pattern is not favourable to the binary segmentation algorithm and represents a 'worst case' scenario. The results of the simulation are reported in Table 5. We see that, for all but the IID scenario, LSWL offers the strongest performance, both in terms of detecting changepoints and mean squared error. In the IID case, LSWL is still able to achieve good performance. As in the single changepoint case, LSWL outperforms AR1Seg in the case of AR(1) errors.

4.2.4 | Multiple changepoints, time-varying autocorrelation scenarios

Finally, we examine the multiple changepoint setting with nonstationary second-order structure. We use the same signal from the previous section, where the mean within each segment alternates between 0 and δ, where δ is calculated so that $\delta/\sigma = 1.25$ for maximum standard deviation σ. Error structures are again simulated from models (E)–(H) as in the single changepoint case. The results of the simulation are shown

TABLE 4 Method performance comparisons for the single changepoint and nonstationary second-order scenarios, reporting the number of changepoints detected and average mean squared error of the estimated mean

		No. of changepoints				
Model	Method	0	1	2	$\geq$3	MSE
E	LSWL	0	**78**	15	7	0.059
	AR1Seg	17	**59**	7	17	0.160
	NPCP	0	**100**	—	—	**0.037**
	NOT	0	8	2	**90**	0.292
F	LSWL	0	**89**	7	4	0.026
	AR1Seg	3	3	3	**91**	0.302
	NPCP	0	**100**	—	—	**0.010**
	NOT	0	0	0	**100**	0.519
G	LSWL	0	**100**	0	0	**0.004**
	AR1Seg	**62**	34	3	1	0.162
	NPCP	0	**100**	—	—	**0.004**
	NOT	0	**88**	3	9	0.012
H	LSWL	0	**100**	0	0	**0.375**
	AR1Seg	0	**68**	14	18	0.508
	NPCP	0	**100**	—	—	0.388
	NOT	0	**94**	1	5	0.721

TABLE 5 Method performance comparisons for the multiple changepoint and stationary second-order scenarios, reporting the number of changepoints detected and average mean squared error of the estimated mean

		No. of changepoints					
Model	Method	$\leq$1	2	3	4	$\geq$5	MSE
IID	LSWL	8	0	**92**	0	0	0.054
	AR1Seg	4	0	**98**	1	1	0.026
	NOT	0	0	**99**	1	0	**0.025**
AR(1) 0.6	LSWL	22	7	**55**	14	2	**0.271**
	AR1Seg	**48**	3	36	5	8	0.346
	NOT	0	0	4	1	**95**	0.471
A	LSWL	3	0	**71**	21	5	**0.661**
	AR1Seg	2	0	4	3	**91**	1.122
	NOT	0	0	0	0	**100**	1.393
B	LSWL	0	0	**66**	20	14	**1.018**
	AR1Seg	0	0	1	3	**96**	3.388
	NOT	0	0	28	6	**66**	2.037
C	LSWL	40	0	**59**	1	0	**0.554**
	AR1Seg	**75**	1	5	2	17	1.037
	NOT	0	0	0	0	**100**	1.303
D	LSWL	2	0	**70**	25	3	**1.688**
	AR1Seg	**45**	2	39	4	10	3.670
	NOT	0	0	0	1	**99**	3.381

Model	Method	No. of changepoints ≤1	2	3	4	≥5	MSE
E	LSWL	13	4	**68**	14	1	**0.219**
	AR1Seg	28	3	**59**	8	2	0.282
	NOT	0	0	10	18	**82**	0.300
F	LSWL	35	0	**56**	8	1	**0.269**
	AR1Seg	1	0	6	5	**88**	0.484
	NOT	0	0	0	0	**100**	0.543
G	LSWL	0	0	**99**	1	0	**0.017**
	AR1Seg	**79**	3	15	2	1	0.281
	NOT	0	0	**83**	1	16	0.026
H	LSWL	0	0	**99**	1	0	1.232
	AR1Seg	0	0	**73**	13	14	1.057
	NOT	0	0	**89**	9	2	**0.832**

TABLE 6 Method performance comparisons for the multiple changepoint nonstationary second-order scenarios, reporting the number of changepoints detected and average mean squared error of the estimated mean

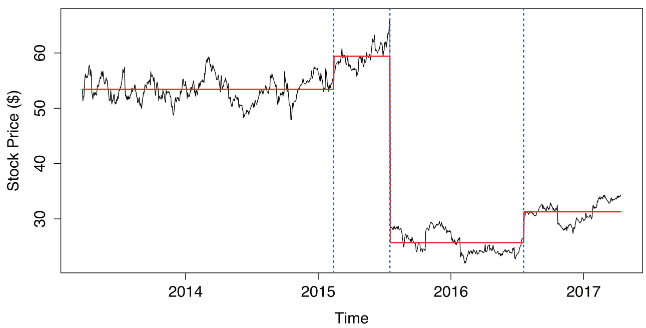

FIGURE 2 Ebay closing stock price over roughly 4-year period. Fitted mean function in solid line

in Table 6. We see that LSWL has the strongest performance in terms of detected changepoints and has the lowest MSE for all but one scenario. Therefore, we find that LSWL performs well in both single and multiple changepoint scenarios, as well as in the presence of stationary and nonstationary second-order structure.

5 | DATA APPLICATIONS

In this section, we demonstrate the potential uses of LSWL with analysis of two data sets. First, we perform changepoint analysis on historical Ebay stock price data, and second, we examine U. K. house price index data that were previously analysed in, amongst others, Baranowski et al. (2019). Accurate segmentation of these data sets is crucial to ensure an informed interpretation and analysis—falsely detecting too many changes in mean may lead to false assertions about the underlying series.

5.1 | Ebay stock price data

Shown in Figure 2 is the historic daily closing stock price in dollars for Ebay over a roughly 4-year time period from approximately April 2013 to April 2017, with corresponding fitted changepoints. One obvious changepoint is apparent in mid-July of 2015: this is due to Ebay and Paypal splitting into separate public companies at this time, roughly 12 years after Ebay had acquired Paypal in 2003. It is clear that we will detect this change using our methodology, as will any multiple changepoint detection method. However, it is not immediately obvious how many subsequent changes there are: one might expect that a financial time series of this sort may exhibit strong autocorrelation, with stock prices close in time being highly correlated. Therefore, we may take the view that this time series has a small number of changes, with a strong level of autocorrelation, as opposed to possessing many mean changes with relatively low autocorrelation.

We detect three changes in mean, with locations shown by the vertical dotted lines in Figure 2. The fitted estimate for the mean is shown in the red solid line. We detect the change associated to the Ebay-Paypal split and detect two subsequent changes that occur at times close to the release of Ebay quarterly results. Therefore, we can interpret the LSWL segmentation as saying that any other large fluctuations in the data that are noticeable by eye can be attributed to serial autocorrelation.

In our analysis of the Ebay data, we used a global scale-wise estimate of the wavelet spectrum at the finest $J_0 = 6$ scales when estimating the autocovariance, that is, we assumed the second-order structure is stationary. Having fitted our changepoint model to the series, we now validate this assumption. Running a test for second-order stationarity on the mean removed series (e.g., using the work of Nason, (2013), or Bücher and Kojadinovic (2016)), we find that we fail to reject the null hypothesis of second-order stationarity. This is reassuring on two counts: first that the series can be considered second-order stationary and second that our segmentation of the series was able to identify all mean changes. Furthermore, estimating the autocorrelation of the series highlights the large levels of autocorrelation present. In Figure 3, we see the estimated autocorrelation of the mean removed series. We can see that autocorrelation persists into very high lags in the series, with significant positive autocorrelation at lag $\tau = 1$ to approximately at least lag 25.

Running standard multiple changepoint detection methods on this series will result in detecting a large number of changes. Under default settings, IID mean changes in a normal distribution using PELT obtains a total of 35 detected changepoints; however, setting a penalty value of approximately 610 yields the exact three changepoints that LSWL detected. (Choosing this penalty value in a principled manner is extremely challenging in practice). Similarly, NOT fits 25, and AR1Seg fits 12 changepoints.

5.2 | U. K. house price index

In this example, we analyse the monthly percentage changes in U. K. house price index (HPI) data set that provides insight into the estimated overall changes in house prices across the United Kingdom. The data are available to download online (https://www.gov.uk/government/statistical-data-sets/uk-house-price-index-data-downloads-march-2020), whereas a detailed description for the calculation of the HPI is also given online from the UK Land Registry (2020).

Fryzlewicz (2018) and Baranowski et al. (2019) analyse the percentage changes in the HPI for three London boroughs: Hackney, Newham, and Tower Hamlets, all of which are in East London. In the analysis of both works, the methods do not allow for the possibility of autocorrelated errors, which could potentially cause spurious changepoints to be detected. Baranowski et al. (2019) allow for changepoints in the variance of the time series, whereas Fryzlewicz (2018) does not. We analyse the Newham HPI data as an illustrative example, because it is this series for which Fryzlewicz (2018) and Baranowski et al. (2019) detect the most changepoints out of the three boroughs.

Figure 4 shows monthly percentage changes in the HPI for the Newham borough, and the corresponding fitted changepoints obtained using LSWL. We see that there are few changepoints fitted to the series. This is in contrast to Fryzlewicz (2018), whose methodology estimates at least 10 changepoints, and Baranowski et al. (2019) who find five changepoints in the series. We believe that the most likely explanation for this

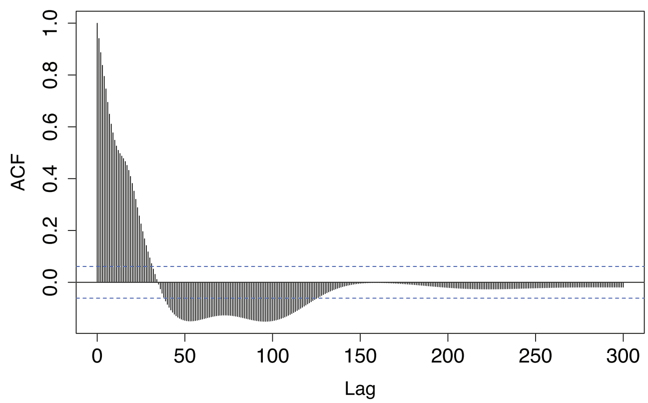

FIGURE 3 Estimate of the autocorrelation of the mean subtracted Ebay stock price data

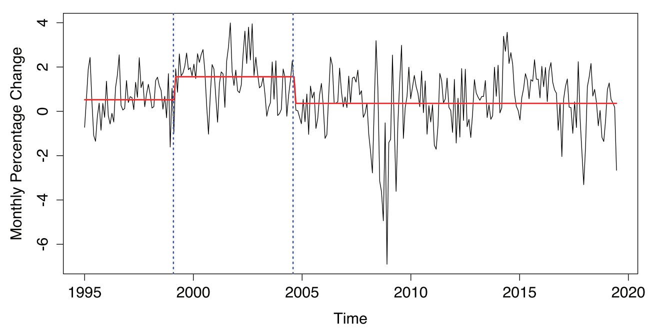

FIGURE 4 Newham HPI series with fitted changepoints (solid lines)

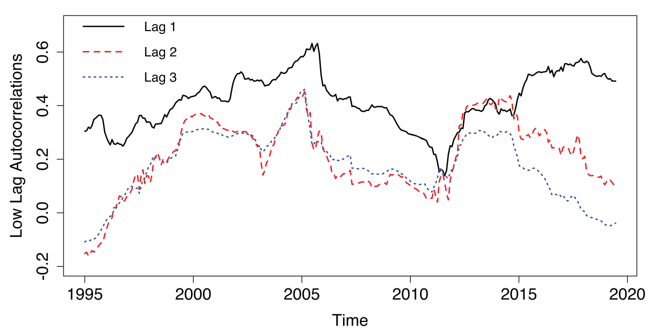

FIGURE 5 Estimated lag 1 (solid line), lag 2 (dashed line) and lag 3 (dotted line) autocorrelation of the Newham HPI series

is the presence of significant autocorrelation within the series, which may cause the methods to overfit the number of changepoints, as noted in Lund and Shi (2020). To illustrate this point, in Figure 5, we plot the estimated autocorrelation for the mean removed Newham series, at lags 1, 2, and 3. We see that there is significant (time-varying) autocorrelation within the series at these lags, which can inhibit the performance of a changepoint detection algorithm if the autocorrelation is not taken into account.

6 | CONCLUDING COMMENTS

Time series can be commonly modelled using changepoint methods whereby the mean function is assumed to be piecewise constant. It is often the case that mean changepoint detection methods make the assumption that the observations are independent, when in reality this may not be true. We addressed this problem by introducing a likelihood-based test under the LSW framework to detect changes in the mean of time series whose unknown, nontrivial autocovariance can vary over time. The proposed approach is shown to work well on simulated data, and compares similarly or favourably with other methods in the literature. The potential uses of the method are shown on two data examples, which highlight the ability to account for highly autocorrelated time series.

One interesting avenue for future investigation is the incorporation of the 'robust' autocovariance estimate within more sophisticated changepoint algorithms. For example, the NOT method of (Baranowski et al. 2019) is shown to be consistent in the presence of a stationary short memory process, however this is achieved assuming a known variance and given preestimate of the autocorrelation. We could use the wavelet-based autocovariance estimate described in this article within NOT or similar changepoint detection algorithms, which could enable the estimation of changepoints under unknown autocorrelation.

ACKNOWLEDGEMENT

E. T. McGonigle gratefully acknowledges financial support from the Numerical Algorithms Group and EPSRC.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are publicly available. The Ebay stock price data, can be found, for example, at Foret (2017). The U. K. house price data can be found at UK Land Registry (2020).

ORCID

Euan T. McGonigle https://orcid.org/0000-0003-0902-0035
Rebecca Killick https://orcid.org/0000-0003-0583-3960
Matthew A. Nunes https://orcid.org/0000-0002-4719-2690

REFERENCES

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

Baranowski, R., Chen, Y., & Fryzlewicz, P. (2016). not: Narrowest-over-threshold change-point detection. R Package Version 1.0.

Baranowski, R., Chen, Y., & Fryzlewicz, P. (2019). Narrowest-over-threshold detection of multiple change points and change-point-like features. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *81*(3), 649–672.

Bücher, A., & Kojadinovic, I. (2016). Dependent multiplier bootstraps for non-degenerate U-statistics under mixing conditions with applications. *Journal of Statistical Planning and Inference*, *170*, 83–105.

Carr, J. R., Bell, H., Killick, R., & Holt, T. (2017). Exceptional retreat of Novaya Zemlya's marine-terminating outlet glaciers between 2000 and 2013. *The Cryosphere*, *11*(5), 2149–2174.

Chakar, S., Lebarbier, E., Lévy-Leduc, C., & Robin, S. (2014). AR1Seg: Segmentation of an AR(1) Gaussian process. R package version 1.0.

Chakar, S., Lebarbier, E., Lévy-Leduc, C., & Robin, S. (2017). A robust approach for estimating change-points in the mean of an AR(1) process. *Bernoulli*, *23*(2), 1408–1447.

Chen, J., & Gupta, A. K. (2011). *Parametric statistical change point analysis: With applications to genetics, medicine, and finance*: Springer Science & Business Media.

Csörgö, M., & Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*, Vol. *18*: John Wiley & Sons Inc.

Dette, H., Eckle, T., & Vetter, M. (2020). Multiscale change point detection for dependent data. *Scandinavian Journal of Statistics*, *47*(4), 1243–1274.

Dette, H., & Wu, W. (2019). Detecting relevant changes in the mean of nonstationary processes—A mass excess approach. *The Annals of Statistics*, *47*(6), 3578–3608.

Fisch, A., Eckley, I. A., & Fearnhead, P. (2018). A linear time method for the detection of point and collective anomalies. arXiv preprint arXiv:1806.01947.

Foret, P. (2017). SP500R: Easy loading of SP500 stocks as time series. https://github.com/PForet/SP500R Accessed 01/06/2020.

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

Fryzlewicz, P. (2018). Tail-greedy bottom-up data decompositions and fast multiple change-point detection. *The Annals of Statistics*, *46*(6B), 3390–3421.

Fryzlewicz, P., & Nason, G. P. (2006). Haar–Fisz estimation of evolutionary wavelet spectra. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(4), 611–634.

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

Hocking, T. D., Schleiermacher, G., Janoueix-Lerosey, I., et al. (2013). Learning smoothing models of copy number profiles using breakpoint annotations. *BMC Bioinformatics*, *14*(1), 164.

Jandhyala, V., Fotopoulos, S., MacNeill, I., & Liu, P. (2013). Inference for single and multiple change-points in time series. *Journal of Time Series Analysis*, *34*(4), 423–446.

Killick, R., Eckley, I. A., & Jonathan, P. (2013). A wavelet-based approach for detecting changes in second order structure within nonstationary time series. *Electronic Journal of Statistics*, *7*, 1167–1183.

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

Kojadinovic, I. (2015). npcp: Some nonparametric tests for change-point detection in possibly multivariate observations. R package version 0.1-6.

Kokoszka, P., & Leipus, R. (1998). Change-point in the mean of dependent observations. *Statistics & Probability Letters*, *40*(4), 385–393.

Kovács, S., Li, H., Bühlmann, P., & Munk, A. (2020). Seeded binary segmentation: A general methodology for fast and optimal change point detection. *arXiv*, *2002.06633*.

Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their Applications*, *83*(1), 79–102.

Lund, R., & Shi, X. (2020). Short communication: Detecting possibly frequent change-points: Wild binary segmentation 2 and steepest-drop model selection. arXiv preprint arXiv:2006.10845v1.

Nason, G. (2008). *Wavelet Methods in Statistics with R*: Springer Science & Business Media.

Nason, G. (2013). A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *75*(5), 879–904.

Nason, G. (2016). wavethresh: Wavelet transforms and associated statistical methodology. https://CRAN.R-project.org/package=wavethresh R package version 4.6.8.

Nason, G., von Sachs, R., & Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*(2), 271–292.

Newey, W. K., & West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix: National Bureau of Economic Research.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, *41*(1/2), 100–115.

Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, *42*(3/4), 523–527.

Pešta, M., & Wendler, M. (2019). Nuisance-parameter-free changepoint detection in non-stationary series. *Test*, *29*, 379–408.

Rothman, A. J. (2012). Positive definite estimators of large covariance matrices. *Biometrika*, *99*(3), 733–740.

Rubin-Delanchy, P., Lawson, D. J., & Heard, N. A. (2016). Anomaly detection for cyber security applications. *Dynamic Networks and Cyber-Security*, Vol. *1* (pp. 137–156). World Scientific.

Sanderson, J., Fryzlewicz, P., & Jones, M. W. (2010). Estimating linear dependence between nonstationary time series using the locally stationary wavelet model. *Biometrika*, *97*(2), 435–446.

Scott, A. J., & Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, 507–512.

Tartakovsky, A., Nikiforov, I., & Basseville, M. (2014). *Sequential Analysis: Hypothesis Testing and Changepoint Detection*: Chapman and Hall/CRC.

Tecuapetla-Gómez, I., & Munk, A. (2017). Autocovariance estimation in regression with a discontinuous signal and m-dependent errors: A difference-based approach. *Scandinavian Journal of Statistics*, *44*(2), 346–368.

UK Land Registry (2020). UK house price index. http://landregistry.data.gov.uk/app/ukhpi Accessed 01/06/2020.

Van Bellegem, S., & von Sachs, R. (2008). Locally adaptive estimation of evolutionary wavelet spectra. *The Annals of Statistics*, *36*(4), 1879–1924.

Vidakovic, B. (2009). *Statistical Modeling by Wavelets*, Vol. *503*: John Wiley & Sons.

Vostrikova, L. Y. (1981). Detecting "disorder" in multidimensional random processes. *Soviet Mathematics Doklady*, *24*, 55–59.

Wu, W., & Zhou, Z. (2020). Multiscale jump testing and estimation under complex temporal dynamics. arXiv preprint arXiv:1909.06307.

Zhang, N. R., & Siegmund, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, *63*(1), 22–32.

How to cite this article: McGonigle ET, Killick R, Nunes MA. Detecting changes in mean in the presence of time-varying autocovariance. *Stat*. 2021;10:e351. https://doi.org/10.1002/sta4.351