Most recent changepoint detection in Panel data

Lawrence Bardwell[*]   Paul Fearnhead[†]   Idris Eckley[†]   Simon Smith[‡]

Martin Spott[§]

January 23, 2018

Abstract

Detecting recent changepoints in time-series can be important for short-term prediction, as we can then base predictions just on the data since the changepoint. In many applications we have panel data, consisting of many related univariate time-series. We present a novel approach to detect sets of most recent changepoints in such panel data that aims to pool information across time-series, so that we preferentially infer a most recent change at the same time-point in multiple series. Our approach is computationally efficient as it involves analysing each time-series independently to obtain a profile-likelihood like quantity that summarises the evidence for the series having either no change or a specific value for its most recent changepoint. We then post-process this output from each time-series to obtain a potentially small set of times for the most recent changepoints, and, for each time, the set of series that has their most recent changepoint at that time. We demonstrate the usefulness of this method on two data sets: forecasting events in a telecommunications network and inference about changes in the net asset ratio for a panel of US firms.

Keywords: Breakpoints, Changepoints, Forecasting, Panel data, Structural breaks.

---

[*]STOR-i CDT, Lancaster University, UK l.bardwell@lancaster.ac.uk

[†]Lancaster University Mathematics and Statistics, UK p.fearnhad@lancaster.ac.uk

[‡]USC Dornsife INET, Department of Economics, USC

[§]Hochschule für Technik und Wirtschaft Berlin

1 Introduction

There are many modern applications where high-dimensional observations are collected and stored over time. This type of data can be viewed as a (potentially large) collection of time series and in the literature is often known as panel data. For an overview of this area see Wooldridge (2010).

We are interested in structural changes, also known as changepoint detection. For an overview of some of the methods used on univariate time series see Jandhyala et al. (2013). In this work, however, we will look at structural changes in panel data. Some recent work in this area includes Kirch et al. (2015), Ma and Yau (2016) and Preuss et al. (2015). Applications of these methods to detect changes occur in many areas such as finance, bioinformatics and signal processing (Cho and Fryzlewicz, 2015; Vert and Bleakley, 2010; Cao and Wu, 2015).

Our work is motivated by a real-life problem of predicting the number of events that occur across a telecommunications network. We have weekly data on the number of events in the network, with this number recorded for each of a set of event types and for each of a set of geographical regions. Being able to make short-term predictions of future event counts is important for planning. These event counts are observed to change over time, often abruptly, and it is natural to model the time series data using a changepoint model.

The challenge with analysing the data is dealing with the large number of separate time series, one for each event type and region pair. In total there are 160 time series. Six example time series are shown in Figure 1. It is natural to assume that some reasons, such as large external factors, that affect the event count for one time series may also affect the event counts for other time series. However, not all time series may see a changepoint at exactly the same time. We would like a changepoint method that encourages, but does not force, time series to share common changepoints. As our primary interest is in short-term prediction, we particularly want a method that is accurate in estimating the location of the most recent change-point for each time series, so that we can use the data since that change-point to predict the likely number of events in the future.

Detecting changepoints in multiple time series introduces computational challenges that are not present when analysing a single time series. A simplistic approach to the problem would
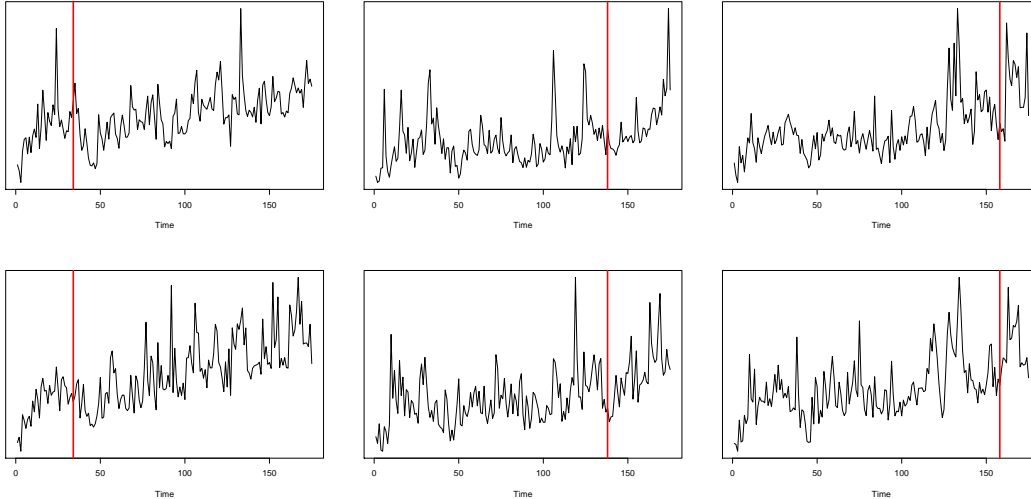
Figure 1: An example of six of the event count time series. These show different patterns. The left-hand column has two series consistent with a constant positive trend since around week 40. The middle column show series with evidence for a recent increase in trend around week 140. The right-hand column shows series with evidence for a decrease in the rate of events from around week 160. In each case we show our estimate of the most recent changepoint – see Section 5.1 for more detail.

thus be to try and apply univariate changepoint methods (Jandhyala et al., 2013). There are two ways of doing this. One is to analyse each time series separately. The other is to aggregate the time series, and analyse the resulting univariate series. Each method has its drawbacks. The former will lose power when detecting changepoints, as it ignores the information that different time series are likely to have changepoints at similar times. The latter approach can perform poorly if the signal from changepoints that affect a small number of series is swamped by the noise in the remaining series when they are aggregated.

An alternative approach to analysing data of this form is to treat the data as a single time series with multivariate observations. We then model the multivariate data within a segment, and allow for this model to change, in an appropriate way, between segments. This approach is taken by Lavielle and Teyssière (2006), who model data as multivariate Gaussian but with a mean that can change from segment to segment. Similarly, Matteson and James (2014) present a non-parametric approach to detecting multiple changes in multivariate data. However, like aggregating the data, these methods may lack power if a change only affects a small proportion of the time series. (Though see Wang and Samworth, 2017, for ideas that try to overcome this problem).

Recently there have been methods specifically designed for detecting changes that affect only a subset of series. Cho and Fryzlewicz (2015) and Cho (2016) propose a way to detect a single, potentially common, changepoint in such data. They consider a novel, non-linear, way of combining summaries of individual time series, so-called CUSUM statistics, that contain information about the presence and location of a changepoint. The intuition is to retain CUSUM values from all series that show strong evidence for a change at a given time-point, but down-weight the values from other time series. Thus they are able to share information across time series without any signal being swamped by noise from series that do not share the common changepoint. Similarly, Xie and Siegmund (2013) introduce a generalised likelihood ratio test for detecting a single common changepoint that affects only a subset of series. This test needs an estimate of the proportion of series affected by the change, and this estimate then affects the weight given to evidence for a change from each series. Again, the intuition of the approach is to give large weight to series that show strong evidence for a change, but lower weight to those with little evidence.

The approaches of Cho and Fryzlewicz (2015), Cho (2016) and Xie and Siegmund (2013) can each be used within a binary segmentation procedure to find multiple changes. Empirical results in these papers show that this type of approach can be more powerful than either analysing series individually or aggregating them.

As we are primarily interested in estimating the most recent changepoint for each time series, we take a different approach. Our approach aims to partition the panel of time series into groups, each of which share the same most recent changepoint, with, potentially, a group corresponding to time series with no change. This is achieved by analysing each time series independently using a penalised cost, or penalised likelihood, approach to detecting changes (Lavielle, 2005; Killick et al., 2012; Maidstone et al., 2017). From each analysis we output a measure of evidence for the most recent changepoint being at each possible time-point, or that the series has no change. We then post-process the output from these analyses in a way that encourages time series to share a common most recent change. This post-processing step involves trying to partition the time series into a small number, K, of groups that share the same value for their most recent changepoint. We show that this post-processing step can be formulated in terms of solving a combinatorial optimisation problem, known as the K-median

problem. Whilst this problem is NP-hard, we use a heuristic solver that is computationally inexpensive, and, empirically, works well in terms of the estimated most recent changepoints. The outline of the paper is as follows. Firstly we define the problem of finding the most recent changepoint in a univariate time series using a penalised cost approach, and show how this can be extended to panel data. To infer the most recent changepoints requires solving a combinatorial optimisation problem. We discuss how to solve this in Section 3. In Section 4 we evaluate our method, and compare it with a number of alternatives on simulated data. We then apply our method to two real data applications. The first data set represents a telecommunications event time series, shown in Figure 1, where the aim is to improve prediction. Secondly, we analyse financial data from a large number of firms. In this application we are more concerned about detecting the locations of most recent changepoints and the sets of firms that change, with the aim of understanding the causes of these changes. Finally we end with a discussion on the advantages and limitations of our method.

2 A Penalised Cost Approach to Most Recent Changepoint Detection

We begin by assuming we have panel data consisting of N time series of length n. Denote the ith time series by $y_{1,i}, \ldots, y_{n,i}$. Throughout we will use the notation $y_{s:t,i}$ to denote the subset of observations from time s to time t inclusive.

Our approach to detecting the common most recent changepoints is based on a penalised cost approach. We will first describe how this approach can be used to analyse individual time series, before then explaining how the output from these individual analyses can be combined to estimate a set of common most recent changepoints for our N series.

2.1 Analysing a Univariate Time Series

First consider analysing data from one of the N time series in our panel data. To simplify notation we will drop the subscript i that denotes which time series, and instead denote the data by $y_{1:n}$. We will denote the number and position of changes by m and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_m)$

respectively. We assume the changepoints are ordered, and define $\tau_0 = 0$ and $\tau_{m+1} = n$.

A penalised cost approach to detecting changepoints in this time series involves introducing a cost associated with each putative segment. This cost is often derived by modelling the data within a segment, and defining the cost to be proportional to minus the maximum likelihood value for fitting that model to a segment of data. If our model for data in a segment is that they are IID with some density $f(y|\theta)$, where θ is a segment-specific parameter, then we can define the cost for a segment $y_{s:t}$ as

$$\mathcal{C}(y_{s:t}) = -2 \max_{\theta} \sum_{u=s}^{t} \log f(y_u|\theta).$$

The segment cost function can include a component that depends on the length of segment as is used in some penalised cost approaches (Davis et al., 2006; Zhang and Siegmund, 2007). To make this idea concrete we will give two examples of cost functions that we will use later. The first is for detecting a change in mean. A simple model is that the data in a segment is IID Gaussian with common known variance, σ^2, and segment specific mean, θ. In this case we get

$$\mathcal{C}(y_{s:t}) = -2 \max_{\theta} \frac{-1}{2\sigma^2} \sum_{u=s}^{t} (y_u - \theta)^2 = \frac{1}{\sigma^2} \sum_{u=s}^{t} \left(y_u - \frac{\sum_{v=s}^{t} y_v}{t-s+1} \right)^2.$$

The second is where we model the mean of the data within a segment as a linear function of time, but allow this linear model to vary between segments. Denote $\theta = (\theta_1, \theta_2)$ to be the segment intercept and slope. If the noise for this model is IID Gaussian we then get a segment cost

$$\mathcal{C}(y_{s:t}) = \frac{1}{\sigma^2} \max_{\theta} \sum_{u=s}^{t} (y_u - \theta_1 - u\theta_2)^2.$$

We use this model for analysing the data presented in the introduction, however in that application some time series have clear outliers. To make our inferences robust to these outliers we follow Fearnhead and Rigaill (2017) and instead use a segment cost

$$\mathcal{C}(y_{s:t}) = \frac{1}{\sigma^2} \max_{\theta} \sum_{i=s}^{t} \min \left\{ (y_i - \theta_1 - i\theta_2)^2, 4\sigma^2 \right\}. \tag{2.1}$$

This cost limits the impact of outliers if their residuals are greater than 2 standard deviations

away from the segment mean.

For all these costs we require knowledge of σ^2, the residual variance (or in the latter example, the variance of the non-outlier residuals). In practice we use a simple and robust estimator of σ, based on the median absolute deviation of the differenced time series (Fryzlewicz, 2014).

Once we have defined a segment cost, we then define a cost for a segmentation as the sum of the segment costs for that segmentation. To segment the data, and find the changepoints, we then want to minimise this cost over all segmentations. However to avoid over-fitting we add a penalty, $\beta > 0$, for each segment. Thus to segment the data we solve the following optimisation problem

$$\min_{m,\boldsymbol{\tau}} \sum_{j=1}^{m+1} \left[\mathcal{C}(y_{(\tau_{j-1}+1):\tau_j}) + \beta \right]. \tag{2.2}$$

The choice of β in this approach is important. Higher values for β will mean fewer change-points detected. There are various suggestions for how to choose β, and the most common for detecting changes in a single time series is the BIC criteria. If our segment specific parameter is of dimension p, then this corresponds to $\beta = (p+1)\log n$. This has good theoretical properties, if our modelling assumptions are correct (e.g. Yao, 1987). However care is needed in practice where this is not the case, see Haynes et al. (2017) for guidance in selecting an optimal value for β for a given time series.

Solving (2.2) is possible using dynamic programming. This requires the solution of a set of intermediate problems. Define $F(t)$ for $t = 1, 2, \ldots, n$ as

$$F(t) = \min_{\boldsymbol{\tau}} \left\{ \sum_{j=1}^{m+1} \left[\mathcal{C}(y_{(\tau_{j-1}+1):\tau_j}) + \beta \right] \right\}, \tag{2.3}$$

where the minimisation is over m and $0 = \tau_0 < \tau_1 < \cdots < \tau_m < \tau_{m+1} = t$. Thus $F(t)$ is the minimum cost for segmenting data $y_{1:t}$. The functions $F(\cdot)$ can be efficiently calculated, for example using the PELT (Killick et al., 2012) algorithm, as

$$F(t) = \min_{s<t} \left\{ F(s) + \mathcal{C}(y_{s+1:t}) + \beta \right\}.$$

Recalling that our interest is in detecting the most recent changepoint, let us consider $G(r)$, which we define to be the minimum cost of the data conditional on the most recent changepoint prior to n being at time r. This is related to $F(r)$ as it is just the minimum cost of segmenting $y_{1:r}$ plus the cost of adding a changepoint and the cost for segment $y_{(r+1):n}$,

$$G(r) = F(r) + \mathcal{C}(y_{(r+1):n}) + \beta, \text{ for } r = 1, \ldots, n-1, \tag{2.4}$$

with $G(0) = \mathcal{C}(y_{1:n})$. This quantity can be viewed as related to the idea of a profile likelihood, as we have optimised over all nuisance parameters (the number and locations of the changepoints prior to the most recent changepoint). It is trivial to see that our estimate for the most recent changepoint is given by $\text{argmin}_{r \in \{0, \ldots, n-1\}} G(r)$. If the most recent changepoint is at $r = 0$, then this corresponds to no change within the time series.

2.2 Extension to panel data

We now return to the problem of finding a set of common most recent changepoints in our panel data. Let $G_i(r)$ denote the minimum cost for segmenting series i with a most-recent changepoint at r, defined in (2.4). Our idea is to search for a set of K locations for the common most recent changepoints for our N series.

Firstly assume that an appropriate value for K is known. Denote a set of common most recent changepoints as $\mathbf{r}_{1:K} = (r_1, \ldots, r_K)$. For the kth most recent changepoint, located at r_k, then there will exist a set, $I_k \subset \{1, 2, \ldots, N\}$, such that for all series $i \in I_k$ the most recent changepoint is located at r_k. The sets $I_{1:K}$ will partition the full set of series $\{1, 2, \ldots, N\}$. It is natural to estimate the $\mathbf{r}_{1:K}$, and the associated sets, by the values that minimise the sum of costs for each series

$$C_K = \min_{I_1, \ldots, I_K} \min_{r_1, \ldots, r_K} \sum_{k=1}^{K} \sum_{i \in I_k} G_i(r_k). \tag{2.5}$$

The minimisation of (2.5) is challenging, however we will describe a method adopted from the field of combinatorial optimisation to solve it for a given value of K in Section 3.

In practice we do not know what value of K to choose. Thus to choose K we resort to

8

minimising a penalised version of (2.5). We first solve the optimisation problem in (2.5) for a range of K, and then choose the value of K that minimises

$$C_K + N \log_2 K + K \log_2 n,$$

where $\log_2$ is log base two. This uses a minimum description length criteria (Grünwald, 2007), and the penalty can be viewed as the log, in base two, of the model complexity for allowing K most recent changepoints: the number of choices of the K changepoints is approximately n^K and then each of the N time series can choose which of the K most recent changepoints to have, which gives K^N possible choices.

This approach penalises adding most recent changepoints. Thus when we implement our method we use a value of β, the penalty for adding a change used in calculating $G_i(r)$, which is slightly lower than the BIC choice. Specifically, we suggest using $\beta = (p + 1/2) \log n$, as on simulated data with no change, values of β lower than this produce $G(r)$ functions that on average get smaller as r increases for $r \geq 1$. This suggests smaller choices of β would be biased towards adding erroneous very recent changepoints. By comparison our choice of β produced $G(r)$ functions whose average value appeared constant for $r \geq 1$.

3 Optimal set of most recent changepoints

We now turn to solving the optimisation problem in (2.5) for a fixed value of K. Solving this is computationally challenging if a brute force method is applied, due to the exponentially large number of ways of choosing either $\mathbf{r}_{1:K}$ or the sets $I_{1:K}$. However it can be reduced to a well studied problem in the field of combinatorial optimisation.

To formulate this problem we proceed as follows. Let $\mathbf{G}$ be a matrix of the conditional costs that we defined in (2.4), so that for $i = 1, 2, \ldots, N$ and $r = 0, 1, \ldots, n - 1$, $\mathbf{G}_{ir} = G_i(r)$ the optimal cost of the most recent changepoint of the ith series being at time r. We want to find the K columns of $\mathbf{G}$ such that if, for each row, we take the minimum of elements in these columns, and then sum these across all N rows, the total is minimised. This allocates each of the N series into K disjoint classes according to which series are affected by a specific

most recent changepoint. The specific optimisation problem is

$$\min_S \sum_{i=1}^{N} \min_{r \in S} \mathbf{G}_{ir}, \text{ where } S \subset \{0, 1, \ldots, n-1\} \text{ and } |S| = K.$$

It turns out that this optimisation problem is mathematically equivalent to the so-called K-median problem (Reese, 2006). This problem can be formulated, and solved, as an integer program with binary variables x_{ir} and z_r where

$$x_{ir} = \begin{cases} 1 & \text{if series } i \text{ has its most recent changepoint at time } r, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$z_r = \begin{cases} 1 & \text{if there is a most recent changepoint in any series at time } r, \\ 0 & \text{otherwise.} \end{cases}$$

The objective is then simply to solve the following problem:

$$\min \sum_{i=1}^{N} \sum_{r=0}^{n-1} \mathbf{G}_{ir} x_{ir} \tag{3.1}$$

$$\text{subject to} \quad \sum_{r=0}^{n-1} x_{ir} = 1, \forall i, \tag{3.2}$$

$$x_{ir} \leq z_r, \forall i, r, \tag{3.3}$$

$$\sum_{r=0}^{n-1} z_r = K. \tag{3.4}$$

Here constraint (3.2) ensures each series has only one most recent changepoint, whilst the two remaining constraints, (3.3) and (3.4), ensure that K different most recent changepoints are selected.

Approaches for solving the K-median problem are discussed in Reese (2006) and references therein. We use the method of Teitz and Bart (1968), available within the R package tbart. This is a simple algorithm that tries to improve on a current solution by replacing one of the K values for a most recent changepoint with a value that is not currently in the set of most

recent changepoints. It loops over all such pairs, and makes the replacement if it will reduce the objective function (3.1). This is repeated until there is no replacement that will improve the objective any further. This method is heuristic, in that it is not guaranteed to find the global optimum to the optimisation problem. However we found that it is computationally efficient and empirically leads to good estimates of the most recent changepoints, as shown below.

4   Simulation study

As described in the introduction, there are a number of methods in the literature that allow us to detect multiple changes in panel data. We compare our method, which we call MRC, to several of these to see empirically how they compare. None of these alternative methods were specifically designed to just estimate the most recent changepoints, and we are unaware of any other methods that focus solely on this. Furthermore some of these methods are able to infer quantities, such as earlier common changepoints, that MRC cannot.

The alternative methods can be split into two groups. The first set of methods estimate common changepoints for each series. We compare with three such approaches. These are analysing the aggregated data (AGG) and two approaches for detecting common changepoints in multivariate data. The latter two methods are the approach of Lavielle and Teyssière (2006) that models data within a segment as multivariate Gaussian with known covariance (MV); and the ECP method (Matteson and James, 2014), which is a non-parametric changepoint detection procedure (ECP).

Both the AGG and MV methods require a choice of penalty and we use the BIC penalty. However, for the ECP method every proposed changepoint is tested for statistical significance using a permutation test and a threshold obtained via a bootstrap that is described in Matteson and James (2014).

The second group of alternative methods includes two methods that can estimate common changepoints that affect only a subset of the time series. The simplest method we consider (IND) involves analysing each series in the panel independently and finding the most recent changepoint in each series. The second method in this group is Double CUSUM Binary

Segmentation (DCBS) (Cho, 2016). Whilst the focus of this method, and the theory that underpins it, is on consistently detecting the location of changepoints, the paper also mentions an intuitively natural way of identifying which series change at each changepoint. We again use the BIC penalty when segmenting each series as part of the IND method. The DCBS method has two parameters that need to be chosen. The first parameter, ψ, is related to the expected degree of sparsity or the number of series affected by a change compared to the total number of series. Guidance is available on how to choose this parameter in Cho (2016). The second parameter, π^ψ, is the threshold for testing whether or not a change is significant as is done in the ECP method mentioned above. This threshold is chosen using a bootstrap style procedure where the null hypothesis of no changepoint is assumed and some empirical quantile of this distribution is taken. We chose this parameter by simulating 100 replications from the null hypothesis, i.e. no changepoints at all, and measured the proportion of false positives for a number of different values for π^ψ. In practice, we found that a value of $\pi^\psi = 10$ worked well.

Each panel data set we simulate consists of 100 series all having length 500. For a given value of K we first simulate K distinct values for the most recent changepoints from the set $\{300, 320, \ldots, 480\}$. This ensures each most recent changepoint position is at least 20 time-points away from all other positions, which helps interpretation when we measure the accuracy of methods in detecting the location of the changes. We partition our 100 time series evenly across the K most recent changepoint locations. We then simulate earlier common changepoints by first simulating potential changepoints independently with probability 0.02 at each time-point prior to the earliest most recent common changepoint. For each of these we simulate a probability from a uniform distribution, and then simulate that a changepoint appears in each time series independently with this probability. The observations in each of the segments are IID Gaussian distributed with mean μ drawn from its prior distribution $\mathcal{N}(0, 2^2)$. For simplicity we keep a fixed variance $\sigma^2 = 1$ for all the observations. In this study the parameter of the last segment differs by ϵ from the mean in the penultimate segment, with the sign of the change being chosen uniformly at random for each time series. We use $\epsilon = 1$ for the studies in Cases 1, 3 and 4 below, whereas for Case 2 we look at the effect of varying ϵ.

In the first three studies we consider the accuracy of estimates of the most recent changepoints and which series are affected. We only compare IND and DCBS with MRC, as these are the only methods that estimate those series affected by each of the most recent changepoints. We evaluate these three methods on a number of different criteria. A specific changepoint is then defined as being detected if it is within 5 time points of an estimated changepoint, and we calculate the proportion of changepoints that are detected. To define the location accuracy we take only those changepoints that are detected then take the average of the absolute difference between the true and estimated locations.

Two of the methods we consider, namely MRC and DCBS, return more information than IND, including the estimated number of most recent changepoints $\hat{K}$ and the subset of series that are affected by each most recent changepoint. We measure the accuracy of the estimate of the number of most recent changepoints using the absolute error, $|\hat{K} - K|$, and call this the changepoint accuracy. We then measure the accuracy of the estimates of the subsets of series affected by each of the changepoints using the set coverage

$$D_j = 1 - \frac{|\hat{I}_j \cap I_j|}{\sqrt{|\hat{I}_j||I_j|}}.$$

Here I_j is the true subset of series affected by the jth most recent changepoint and $\hat{I}_j$ is the estimated subset. This measure satisfies $D_j \in [0, 1]$, with $D_j = 0$ indicating that the estimated subset overlaps exactly with the true subset, and $D_j = 1$ if the two subsets are disjoint. More generally, smaller values of D_j indicate a greater overlap. In the simulations presented for each panel we calculate the mean of $D_1, D_2, \ldots, D_{\hat{K}}$.

Case 1. Effect of K.

For the first study we simulated data as described above for a range of values for K from $K = 1$ to $K = 10$. Results are shown in Table 1.

It is clear from Table 1 that our MRC method outperforms both IND and DCBS across the criteria we consider. The ability to synthesise information across time series means that MRC is able to more accurately detect changes and locate where they occur than analysing each time series independently. Not surprisingly we see that the advantage of using MRC over IND decreases as K increases. We also see that DCBS is more accurate than IND for

	IND		DCBS				MRC			
K	PD	LA	PD	CA	LA	D	PD	CA	LA	D
1	0.73	1.46	0.86	2.98	0.05	0.09	0.98	0.10	0.06	0.01
2	0.76	1.43	0.82	2.55	0.04	0.14	0.97	0.04	0.04	0.03
3	0.77	1.39	0.70	2.64	0.13	0.24	0.95	0.05	0.03	0.05
4	0.77	1.37	0.67	2.47	0.18	0.28	0.94	0.03	0.05	0.06
5	0.78	1.38	0.58	2.09	0.24	0.35	0.93	0.03	0.04	0.07
10	0.78	1.41	0.29	1.77	0.76	0.62	0.89	0.10	0.19	0.10

Table 1: For all of the methods and differing values of K we repeated each experiment 100 times and recorded the proportion of true changes we detected (PD), the accuracy in detecting the number of distinct most recent changes (CA), the accuracy of the estimated location of these changes (LA) and the set coverage (D). These values are averaged over the 100 replications.

small values of K, and is consistently more accurate in estimating the position of detected changepoints, but appears less powerful at detecting the most recent changes as K increases.

Case 2. Effect of size of change at final changepoint.

Next we look at how the performance of each method is affected by the size of the mean change at the most recent changepoint, ϵ. We fix the number of most recent changepoints as $K = 5$, meaning that there are 20 series affected by each different changepoint. We vary the value of ϵ from $\epsilon = 0.2$ to $\epsilon = 1.6$. Results are shown in Table 2.

We again see MRC giving consistently stronger performance for all values of ϵ. The advantage of MRC over IND is largest for moderate values of ϵ. For small values of ϵ the information about changes in each time series is small, and thus the benefit of merging information across time series is limited. For larger values of ϵ it is relatively easy to detect changes from an individual time series, and hence the benefit of using MRC over IND is mainly seen in its ability to more accurately locate the position. Surprisingly DCBS does not improve as much as the other methods as we increase ϵ. The DCBS method was not specifically designed to detect most recent changes, and it appears not to be as accurate at identifying the time series that change at each changepoint. This, in turn, impacts the DCBS method's accuracy at detecting which changes are most recent for a given time series.

Case 3. Dependent observations.

One of the key assumptions we made when modelling the most recent change process was

	IND		DCBS				MRC			
ϵ	PD	LA	PD	CA	LA	D	PD	CA	LA	D
0.2	0.11	1.49	0.09	3.49	0.75	0.66	0.11	2.06	0.47	0.64
0.4	0.22	1.75	0.22	3.43	1.01	0.55	0.36	1.27	0.88	0.42
0.6	0.46	1.76	0.37	3.00	0.64	0.51	0.76	0.30	0.29	0.20
0.8	0.65	1.57	0.48	2.51	0.35	0.45	0.89	0.09	0.12	0.10
1	0.78	1.38	0.58	2.09	0.24	0.35	0.93	0.03	0.04	0.07
1.2	0.86	1.19	0.62	1.63	0.16	0.31	0.95	0.04	0.02	0.05
1.4	0.91	1.01	0.65	1.44	0.13	0.26	0.95	0.04	0.00	0.05
1.6	0.93	0.85	0.66	1.47	0.10	0.25	0.96	0.04	0.00	0.04

Table 2: For all of the methods with a fixed value of $K = 5$ and differing values of ϵ we repeated each experiment 100 times and recorded the proportion of true changes we detected (PD), the accuracy in detecting the number of distinct most recent changes (CA), the accuracy of the estimated location of these changes (LA) and the set coverage (D). These values are averaged over the 100 replications.

the independence of observations, both within and between segments. This greatly simplifies the modelling and especially the inference procedure. However, in most real time series applications observations are not independent and display serial autocorrelation.

To assess the robustness of the MRC procedure we simulated an MRC process with a piecewise constant mean function as before, but instead of adding IID normally distributed 'noise' we simulated an $AR(1)$ noise process, Z_t, with standard normal errors e_t

$$Z_t = \phi Z_{t-1} + e_t.$$

This process was simulated for a range of values of ϕ, representing mild to moderate auto-correlation. The number of most recent changepoints was fixed at $K = 5$ and we set $\epsilon = 1$. Results are shown in Table 3.

As ϕ increases the dependence between observations increases and the measures for all methods we consider decrease. The impact on both MRC and DCBS is larger than the impact on IND, with IND correctly detecting more recent changepoints for $\phi = 0.4$. Both MRC and DCBS still give more accurate estimates of the position of the changes that they do detect, and MRC is again more accurate than DCBS for all cases we consider.

| | IND | | DCBS | | | | MRC | | | |
ϕ	PD	LA	PD	CA	LA	D	PD	CA	LA	D
0	0.78	1.38	0.58	2.09	0.24	0.35	0.93	0.03	0.04	0.07
0.1	0.73	1.47	0.55	2.17	0.29	0.40	0.89	0.03	0.07	0.11
0.2	0.64	1.56	0.43	2.66	0.31	0.49	0.75	0.74	0.18	0.21
0.3	0.51	1.59	0.12	3.05	0.66	0.71	0.47	1.71	0.35	0.38
0.4	0.36	1.72	0.04	2.70	1.59	0.80	0.21	1.79	0.82	0.55

Table 3: For each of the methods and differing values of AR parameter, ϕ, we repeated each experiment 100 times and recorded the proportion of true changes detected (PD), the accuracy in detecting the number of distinct most recent changes (CA), the accuracy of estimated location of these changes (LA) and the set coverage (D). Fixed values for $K = 5$ and $\epsilon = 1.0$ were used.

We then simulated an $MA(1)$ noise process, Z_t, with standard normal errors e_t

$$Z_t = e_t + \phi e_{t-1}.$$

This process was simulated for a range of values of ϕ. The number of most recent changepoints was fixed at $K = 5$ and we set $\epsilon = 1$. Results are shown in Table 4. We see similar patterns to those observed for the $AR(1)$ model. Increased autocorrelation reduces the accuracy of all methods. For the largest values of ϕ we tried, IND performs slightly better than MRC in terms of the proportion detected, but MRC and DCBS still give more accurate estimates of the location of the changes they do detect. For all cases MRC outperforms DCBS.

The fact that increasing the level of autocorrelation in the residuals of the mean model, for both the $AR(1)$ and $MA(1)$ models, reduces the accuracy of all methods is not surprising. As we increase the autocorrelation there will be less information in the data about the position of changes. Furthermore, all methods were implemented with a choice of penalty that assumes IID residuals. When there is positive autocorrelation this can lead to such methods detecting a larger number of spurious changepoints.

There are two ways to overcome this problem. One is to model the residual process of the mean model as, say, an ARIMA model. We can then adapt the cost function so that it is based on the negative log-likelihood under such a model. If we implement such an approach we need to assume the parameters of the ARIMA model are common across all segments,

and it is just the mean parameter that changes. Otherwise we would be detecting changes in the second order structure of the time-series as well as in the mean.

A simpler approach is suggested by theoretical results in Lavielle and Moulines (2000). This paper presents asymptotic theory for detecting changes in mean that show you can consistently estimate the number and location of the changepoints using the square error loss even when there is autocorrelation in the residuals. However to do so you need to increase the penalty, β, with the level of autocorrelation in the residuals. We tested this simpler approach by re-analysing the data simulated with AR and with MA residual processes but with an increased value of β. As suggested by the results in Lavielle and Moulines (2000) we scaled β by the sum of the autocorrelation function for the residuals from $-\infty$ to ∞. This is just $(1 + \phi)/(1 - \phi)$ for the AR model and $1 + 2\phi$ for the MA model. We also scaled the MDL penalty used to choose the number of most recent changepoints by the same factor.

Results are shown in Table 5. These show that by adapting the penalties appropriately we can much more accurately detect the most recent changepoints, with close to or above 80% of changepoints correctly idepentified for the AR models with ϕ up to 0.4, and for the MA model with ϕ up to 0.6. We only see a substantial drop-off in performance for the AR residuals when $\phi = 0.6$. In part this will be because of the reducing amount of information about the location of the changepoints when we have strongly positively correlated residuals. In practice we would not know the exact form of the auto-correlation structure of the residuals, but we can estimate this from the data by first fitting a changepoint model and looking at the auto-correlation of the residuals we obtain.

Case 4. Accuracy of prediction.

Finally, we consider how each method performs if the aim is to predict $Y_{i,n+1}, \ldots, Y_{i,n+5}$ for each time series. Each method gives an estimate for the most recent changepoint for each time series. Conditional on this estimate we can estimate the mean in the final segment. This estimated mean is our prediction for the next value(s). We use the same data as in Case 1 but leave out 5 time points at the end of the data. We then predict the last 5 points using the most recent changepoints found by each method and measure the Mean Squared Error (MSE) between the truth and our predictions. Results are shown in Table 6.

MRC gives the most accurate predictions for all values of K, and is the only method to

	IND		DCBS				MRC			
ϕ	PD	LA	PD	CA	LA	D	PD	CA	LA	D
0	0.78	1.38	0.58	2.09	0.24	0.35	0.93	0.03	0.04	0.07
0.1	0.73	1.45	0.56	2.23	0.27	0.39	0.89	0.04	0.06	0.11
0.2	0.67	1.50	0.50	2.46	0.31	0.45	0.81	0.37	0.09	0.17
0.3	0.60	1.57	0.27	2.78	0.50	0.61	0.65	1.32	0.23	0.28
0.4	0.51	1.65	0.09	2.70	0.80	0.73	0.45	1.78	0.40	0.39

Table 4: For each of the methods and differing values of the MA parameter, ϕ, we repeated each experiment 100 times and recorded the proportion of true changes detected (PD), the accuracy in detecting the number of distinct most recent changes (CA), the accuracy of the estimated location of these changes (LA) and the set coverage (D). Fixed values for $K = 5$ and $\epsilon = 1.0$ were used.

	AR				MA			
ϕ	PD	CA	LA	D	PD	CA	LA	D
0	0.93	0.03	0.04	0.07	0.93	0.03	0.04	0.07
0.1	0.93	0.04	0.05	0.07	0.93	0.04	0.04	0.07
0.2	0.90	0.06	0.06	0.10	0.91	0.06	0.05	0.08
0.3	0.86	0.11	0.09	0.13	0.89	0.06	0.07	0.10
0.4	0.79	0.35	0.16	0.19	0.86	0.07	0.08	0.13
0.5	0.67	0.55	0.28	0.27	0.83	0.12	0.12	0.16
0.6	0.48	1.00	0.62	0.40	0.79	0.19	0.17	0.19

Table 5: Results of applying MRC to data with AR or MA residuals when we inflate the penalties, β and the MDL penalty, to account for the positive autocorrelation. Results are averages over 100 data sets for each value of the AR/MA parameter, ϕ. We show the proportion of true changes detected (PD), the accuracy in detecting the number of distinct most recent changes (CA), the accuracy of the estimated location of these changes (LA) and the set coverage (D).

consistently be more accurate than analysing each time series individually. The method that treats the N time series as a multivariate time series, where the mean changes in all components at a change (MV), does well for $K = 1$ and $K = 2$, but loses accuracy for larger K. The method that aggregates the time series, and then detects changes in the resulting uni-variate time series, does particularly poorly. This is because the aggregation step reduces the signal for a change, even when all changes are in the same location, as the sign of the change in mean differs across time series.

k	IND	AGG	MV	ECP	DCBS	MRC
1	1.04	1.29	1.01	1.09	1.08	1.01
2	1.06	1.27	1.05	1.09	1.08	1.03
3	1.04	1.25	1.06	1.08	1.07	1.02
4	1.04	1.34	1.08	1.09	1.08	1.02
5	1.04	1.23	1.08	1.09	1.07	1.02
10	1.04	1.29	1.12	1.09	1.09	1.02

Table 6: The average Mean Squared Error (MSE) for predictions of each method. The MSE was calculated for the difference between the truth and predicted values and averaged over 100 replications.

5 Applications

We look at two different applications of our method using real data. These applications differ in their focus and the aim of the analysis. The first is that of the telecommunications event count data introduced in Section 1. Our second application concerns the balance sheets of a large number of firms. In this latter case we look for changes in a parameter that measures the ratio between the cash holdings of a company and the net assets held on its balance sheet. The goal of this analysis is to explore why the cash holdings of many large firms have increased over time, and if there are any specific events that have caused this. By using our method we can identify the years when a change occurs and for each of these years, which firms change. This information helps us to tie in specific legal or economic changes to the years within which they occurred, and the types of industries that are affected.

5.1 Telecommunications event data

Our panel data consists of the number of events that occur each week over a 175 week period. Events are recorded for each of 10 geographical regions and 16 different event types. Thus there are 160 possible series, with 18 exhibiting no weekly events over the 175 weeks measured. We are therefore left with 142 series to analyse.

We can obtain an overall time series for the number of events per week across the entire network if we aggregate all of these series together. This fully aggregated series is shown in Figure 2. We can see that there are distinct changes in the slope of this series and it is

segmented into piecewise linear regressions.

We can get some idea of the level of auto-correlation within the data by calculating the auto-correlation for the residuals from this fit, and this is shown by the ACF plot in Figure 2. This shows a significant auto-correlation, of ≈ 0.2, at lag 1. We also looked at the partial auto-correlation function and this suggested that the residuals could be modelled via an MA(1) process with a parameter of 0.2.
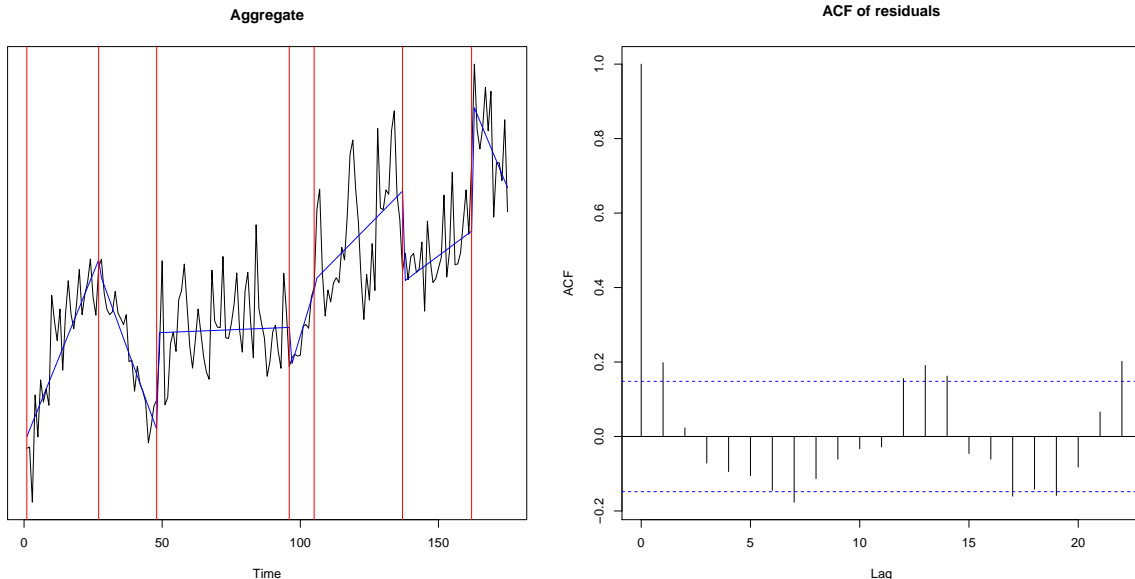

Figure 2: The aggregate series segmented into piece wise linear regressions and ACF plot of the residuals from the piecewise linear fit.

As mentioned in the introduction, the main interest with this data is in making short-term predictions. To do this we use the method described in Section 3 to find the number of most recent changes. Our method is applied assuming the mean of the data within each segment is a linear function of time, and using the robust cost function (2.1). We inflate beta β, the penalty of the penalised cost function when analysing each time-series, and the MDL penalty by 1.4 to account for the correlation in the residuals – as suggested by the Case 3 results from our simulation study. We estimate that there are six different most recent changepoints. This means that all of the 142 series can be separated into six groups, depending on which of the most recent changepoints affects each series.

Figure 3 show the aggregate series for each of the six groups. The groups contain 22, 16, 29, 21, 27 and 27 series from left to right, and top to bottom, respectively. All of the aggregated
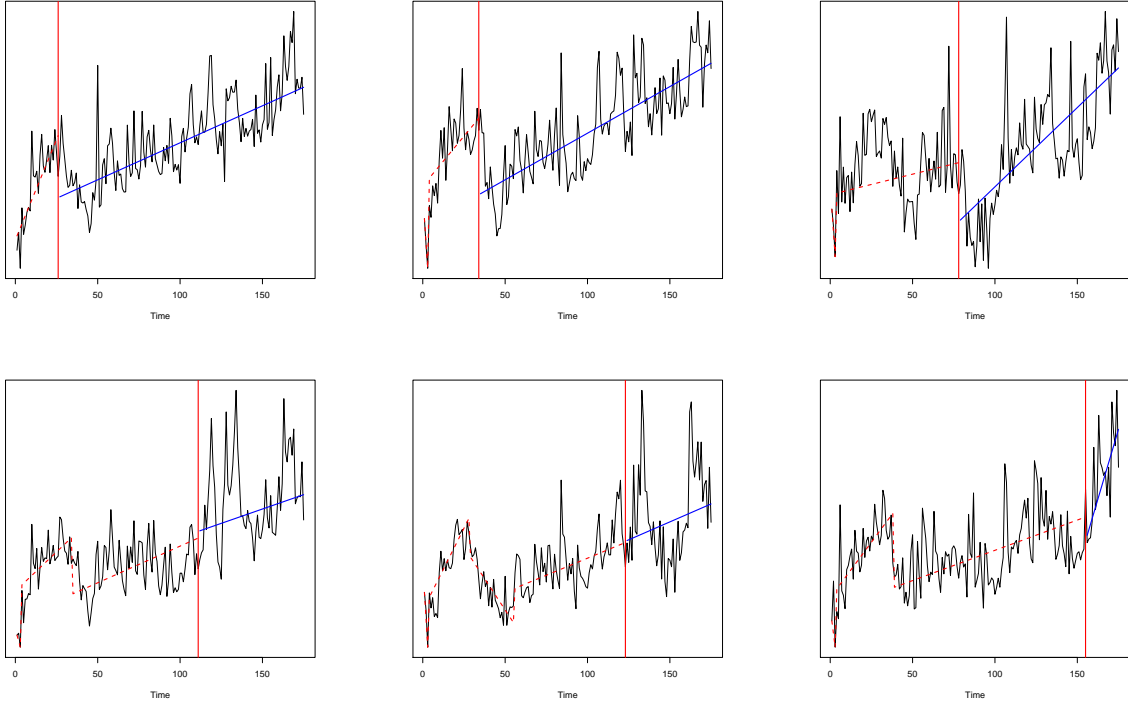
Figure 3: The aggregate series for each of the six groups of series. Their respective most recent changepoints are added, with the final segment shown in blue. The previous segmentation prior to the most recent change is shown as a red dashed line.

series show an increasing trend since their most recent changepoint, but they differ in terms of the location of this change and the slope of the trend after the most recent changepoint.

Once we have found the most recent changepoints, the parameters of the resulting regression line in the last segment can be estimated. These estimates can then be used to predict succeeding time points. We analyse the data up to four data points (weeks) from the end of the data and then use the predictions obtained from the regression model to evaluate the Mean Square Error (MSE) of the prediction for the last four weeks.

We compare predictions using the estimated most recent changepoints from MRC with predictions where we segment each time series separately. The MSE for the predictions in the latter case is 43442 while for our algorithm (with $K = 6$) it is 41476. This is an improvement of 4.5% in the MSE of the prediction compared to analysing each time series individually.

5.2 Corporate finance data

We now apply our method to a panel data set from the field of corporate finance. This data set comprises the annual value of a range of different financial indicators for a number of firms. These include, for example, the value of a firm's assets or whether the firm pays dividends or not. This particular data set is known as an unbalanced panel as the observations for each firm do not all begin or end in the same year. We can view this as a longitudinal data problem where the cohort are the firms that are tracked over time. As is common in these problems there is a large (cohort) number of firms, 7039 in this example, but these are observed over a much smaller time frame. In this case there are a maximum of 54 observations per firm (annually from 1962 - 2015).

An intriguing phenomenon in corporate finance is the fact that U.S. firms hold considerably more cash nowadays compared with a few decades previously. Specifically, cash as a proportion of total assets held by U.S. firms has more than doubled in the past three decades. The evolution of corporate cash holdings has received a lot of attention from academic researchers, policy makers, and practitioners. Numerous explanations for this have been offered in the literature, including increased cash flow volatility (Bates et al., 2009, 2017), competition (Brown and Petersen, 2011), changes in production technology (Gao, 2017) and changes in the cost of carry (Azar et al., 2016).

Azar et al. (2016) argue that changes over time in the cost of carry, that is the net cost of financing one dollar of liquid assets, explains the evolution of corporate cash holdings (see also Graham and Leary, 2016). They measure the cost of carry as the spread between the risk-free Treasury-bill rate and the return on the portfolio of liquid assets for the corporate sector. However a limitation of existing studies is that they split their data along the time domain into distinct 'regimes' by eyeballing the data. Such an approach is highly subjective and increases the opportunity for data snooping. It would be preferable to introduce a formal procedure for detecting any distinct regimes.

We therefore re-examine the ability of the cost of carry to capture variation in corporate cash by formally modelling the breakpoint process using our changepoint methodology. Our analysis follows Azar et al. (2016) and therefore uses the same dataset (see Azar et al., 2016, for a detailed description of the dataset). We control for a number of variables that may

affect cash holdings of a firm, such as capital expenditure, spending on R&D and the amount of leverage it has amongst others. Specifically we consider a fixed effects linear model where the response variable, y_{it}, represents the cash to net asset ratio of firm i in year t is regressed against 12 covariates,

$$y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_{12} X_{12it} + \epsilon_{it}. \tag{5.1}$$

These covariates are described in Table 1 of Azar et al. (2016). The β_js are pooled estimates of the effect of the covariates measured over all 7039 firms and the years that they are observed. Each fixed effect term, α_i, captures a firm-specific characteristic in terms of a firm specific intercept. These fixed effects can be interpreted as the difference between the predicted cash to net assets ratio and the true value observed. As such, the fixed effects are able to capture differences caused by external changes that cannot be explained by the covariates in the model.

For a specific firm the fixed effects term may change due to a number of factors such as a CEO change, a merger or takeover by another firm or some scandal such as a product recall that requires large amounts of cash to be spent. However, we are more interested in the times when the fixed effect parameter changes in a significant number of firms at the same time. The causes of these changes would be due to wider economic events such as changes in policy, technological innovation, or regulatory changes.

Having estimated the β_js via maximum likelihood estimate, we can rewrite (5.1) as a change in mean model

$$y_{it} - \left(\hat{\beta}_1 X_{1it} + \hat{\beta}_2 X_{2it} + \ldots + \hat{\beta}_{12} X_{12it} \right) = \alpha_{it} + \epsilon_{it}, \tag{5.2}$$

where the $\hat{\beta}_i$s are the parameter estimates.

Our MRC method can be applied to this problem and aims to find the year(s) when the most recent changepoint(s) occur and the subsets of firms that are affected. We now follow the method of Section 3 to find the optimal number of most-recent changepoints and the sets of firms that are affected by them.

We first analysed each series independently, to estimate the auto-correlation of the residuals.

When we average the estimated auto-correlation across firms, we observe positive correlation of 0.08 at lag 1, and negative auto-correlation at lag 2. Thus we implemented our method with the penalties inflated by 1.16. One advantage of our approach is that it can easily deal with the different amounts of missing data for each time-series, as the values, $G(r)$ defined in (2.4), that we need for each individual time-series can be calculated for time-series with missing data. For example, if a firm has data missing in year 1980 then a change in year 1979 and a change in year 1980 both correspond to a change between that data values recorded in 1979 and 1981; and thus these changes would have the same value of $G(r)$.

The result of our method is to infer a single most recent changepoint common to all firms, in year 2008. This corresponds to the recent financial crisis and the large fluctuations in the value of assets held by many firms in that period.

6 Discussion

In this paper we have developed novel methodology to detect changepoints in panel data. The specific changepoints we aim to detect are the most recent changes that affect different and disjoint subsets of the series that make up the panel. We focus on detecting the most recent changes as this can be useful in forecasting, as shown in Section 5.1. We are also able to identify the series affected by different changes, which leads to a greater understanding of why and how the changes have occurred.

In our analysis of the two real data sets, we used cost functions for segmenting each individual time series that are based on assuming no temporal dependence in the residuals. This can be justified theoretically by results that show, for example, that detecting changes in mean using a least squares criteria is robust to the presence of temporal dependence in the residuals (Lavielle and Moulines, 2000). Our empirical results suggest that our method is robust to short-term autocorrelation in the residuals if we use this approach. Furthermore, our general approach can easily be extended to allow for modelling of the error structure of the residuals, by using cost functions for the data within each segment that are based on models that allow for autocorrelation. Such an approach is likely to be needed for time-series with substantial or long-range dependencies.

Our method also ignores any dependence across time series, either in the form of cross-correlation in the residuals or of similar changes at common changepoints. Whilst the former is an active area of research within the non-stationary time series community (see for example Ombao et al., 2005; Park et al., 2014) this is an open and intriguing area of future research for the changepoint community. The consquence of ignoring such (time-varying) structure might be that we infer some spurious changes to fit unusual patterns in the residuals that are seen in multiple time series. It is not clear how to develop a method that accounts for the latter, but such a method could have greater power at detecting changes than our MRC procedure.

`R` code to run this method is available here `http://www.research.lancs.ac.uk/portal/en/datasets/r-code-from-most-recent-changepoint-detection-in-panel-data(7f5f1b5e-492b-4afa` `.html`

References

Azar, J., Kagy, J.-F., and Schmalz, M. C. (2016). Can changes in the cost of carry explain the dynamics of corporate cash holdings. *Review of Financial Studies*, 29:21942240.

Bates, T. W., Chang, C. H., and Chi, J. D. (2017). Why has the value of cash increased over time? *Journal of Financial and Quantitative Analysis (to appear)*.

Bates, T. W., Kahle, K. M., and Stulz, R. M. (2009). Why do US firms hold so much more cash than they used to? *The Journal of Finance*, 64(5):1985–2021.

Brown, J. R. and Petersen, B. C. (2011). Cash holdings and R&D smoothing. *Journal of Corporate Finance*, 17(3):694–709.

Cao, H. and Wu, W. B. (2015). Changepoint estimation: another look at multiple testing problems. *Biometrika*, 102(4):974–980.

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

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

Davis, R. A., Lee, T. C. M., and Rodriguez-Yam, G. A. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101(473):223–239.

Fearnhead, P. and Rigaill, G. (2017). Changepoint detection in the presence of outliers. *Journal of the American Statistical Association (Forthcoming)*.

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

Gao, X. (2017). Corporate cash hoarding: The role of just-in-time adoption. *Management Science (to appear)*.

Graham, J. R. and Leary, M. T. (2016). The evolution of corporate cash. *SSRN*. doi:10.2139/ssrn.2805505.

Grünwald, P. D. (2007). *The minimum description length principle*. MIT press.

Haynes, K., Eckley, I. A., and Fearnhead, P. (2017). Computationally efficient changepoint detection for a range of penalties. *Journal of Computational and Graphical Statistics*, 26:134–143.

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

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

Kirch, C., Muhsal, B., and Ombao, H. (2015). Detection of changes in multivariate time series with application to EEG data. *Journal of the American Statistical Association*, 110(511):1197–1216.

Lavielle, M. (2005). Using penalized contrasts for the change-point problem. *Signal Processing*, 85(8):1501 – 1510.

Lavielle, M. and Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis*, 21(1):33–59.

Lavielle, M. and Teyssière, G. (2006). Detection of multiple change-points in multivariate time series. *Lithuanian Mathematical Journal*, 46(3):287–306.

Ma, T. F. and Yau, C. Y. (2016). A pairwise likelihood-based approach for changepoint detection in multivariate time series models. *Biometrika*, 103(2):409–421.

Maidstone, R., Hocking, T., Rigaill, G., and Fearnhead, P. (2017). On optimal multiple changepoint algorithms for large data. *Statistics and Computing*, 27:519–533.

Matteson, D. S. and James, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, 109(505):334–345.

Ombao, H., Von Sachs, R., and Guo, W. (2005). Slex analysis of multivariate nonstationary time series. *Journal of the American Statistical Association*, 100(470):519–531.

Park, T., Eckley, I. A., and Ombao, H. C. (2014). Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. *IEEE Transactions on Signal Processing*, 62(20):5240–5250.

Preuss, P., Puchstein, R., and Dette, H. (2015). Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association*, 110(510):654–668.

Reese, J. (2006). Solution methods for the p-median problem: An annotated bibliography. *Networks*, 48(3):125–142.

Teitz, M. B. and Bart, P. (1968). Heuristic methods for estimating the generalized vertex median of a weighted graph. *Operations Research*, 16(5):955–961.

Vert, J. and Bleakley, K. (2010). Fast detection of multiple change-points shared by many signals using group LARS. In Lafferty, J., Williams, C., Shawe-Taylor, J., Zemel, R., and Culotta, A., editors, *Advances in Neural Information Processing Systems 23*, pages 2343–2351. Curran Associates, Inc.

Wang, T. and Samworth, R. J. (2017). High-dimensional changepoint estimation via sparse projection. *Journal of the Royal Statistical Society, Series B (Forthcoming)*.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Xie, Y. and Siegmund, D. (2013). Sequential multi-sensor change-point detection. *Annals of Statistics*, 41(2):670–692.

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

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