JOURNAL OF TIME SERIES ANALYSIS
J. Time. Ser. Anal. **35:** 239–261 (2014)
Published online 30 January 2014 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12062

ORIGINAL ARTICLE

ON-LINE MONITORING OF POLLUTION CONCENTRATIONS WITH AUTOREGRESSIVE MOVING AVERAGE TIME SERIES

CHRISTOPHER DIENES AND ALEXANDER AUE*
Department of Statistics, University of California, One Shields Avenue, Davis, CA 95616, USA

The concentration of aerosol particles, largely caused by traffic volume and often enhanced during temperature inversion episodes in the cold season, can be a concern for human health in the urban environment. This particulate matter is typically recorded as PM_{10}, the total mass of particles below 10 μm in diameter. It is suspected that, within the PM_{10} class, ultrafine particles ($<$100 nm) may be responsible for causing respiratory and cardiovascular diseases. Because of their low mass, ultrafine particles are hard to detect, and researchers try to utilize PM_{10} in combination with nitrogen oxides NO_x and other trace gases to monitor their dynamic evolution. To meet pollution standards set by national government and European Union regulation, the city of Klagenfurt, Austria, began using calcium magnesium acetate as a deicer on 11 January 2012, hoping to literally glue pollutants to the ground and thereby reducing pollution concentrations. With the statistical methodology developed in this article, the dynamic evolution of PM_{10} and NO_x is traced for the time period starting 4 January and ending 25 January 2012, and a change in dynamics is found. The findings are based on on-line monitoring procedures that sequentially detect structural breaks in the mean and the parameter values of an autoregressive moving average process. These are defined in terms of model residuals and one-step ahead predictors. Theoretical properties are studied, and a simulation study shows that the proposed procedures work well in finite samples.

Received 17 May 2013; Revised 6 October 2013; Accepted 19 December 2013

Keywords: CUSUM statistic; environmental data; one-step ahead predictors; on-line monitoring; particulate matter; structural break detection.
JEL. Primary 62P12; 62L10; Secondary 60F17; 62M10.

1. INTRODUCTION

High concentrations of particulate matter with an aerodynamic diameter of less than 10μm, known as PM_{10}, are associated with negative health effects such as respiratory and cardiovascular diseases. Epidemiological studies from a number of countries have lent support to this hypothesis; see, for example, Ostro *et al.* (1999) and Pope III *et al.* (1991) for studies conducted in the USA, Janhäll *et al.* (2006) for Sweden and Stadlober *et al.* (2008) for Austria. The primary cause for high pollutant concentrations in urban environments is road traffic volume. In geographic regions with strong winter temperature inversions, these effects are magnified in the cold season, and limits set, for example, by European Union (EU) regulation can frequently be violated. While pollution concentrations are routinely measured within the PM_{10} framework for most major cities in the northern hemisphere, it is still unknown which subgroup of particles causes the well-documented health effects. Recent research has pointed to ultrafine particles (UFP) as a probable cause; see Oberdörster and Utell (2002) & Pekkanen *et al.* (1997). These, however, are hard to detect because of their low mass contribution to the overall PM_{10} amount. One option to gauge the impact of UFP is the use of surrogate series such as mono-nitrogen oxides (NO_x) and other trace gases whose measurements are generally available.

* Correspondence to: Alexander Aue, Department of Statistics, University of California, One Shields Avenue, Davis, CA 95616, USA.
E-mail: aaue@ucdavis.edu

Particular interest here is in monitoring pollutant concentrations in ambient air in Klagenfurt, Austria. Plots of the corresponding data for PM_{10} recorded at two monitoring stations along with an NO_x series are displayed in Figure 1. The observed values were recorded at 30-min intervals from 4 January 2012 to 25 January 2012. To meet regulation standards, local policy makers have introduced a number of measures such as partial traffic regulation and the use of calcium magnesium acetate (CMA) as an alternative deicer to sodium chloride road salts. The latter has been particularly controversial. CMA was introduced area wide in the city centre of Klagenfurt on 11 January 2012. This article is concerned with devising a statistical procedure to test whether there have (yet) been significant measurable effects of CMA on the dynamic evolution of pollutant concentrations. This is carried out in Section 3.

Motivated by this application, methodology is developed for the on-line monitoring of parameter breaks in autoregressive moving average (ARMA) time series (sequential change-point analysis). ARMA processes have been instrumental in the analysis of linear time series for the past decades. Their theoretical properties have been studied extensively, and parameter estimation procedures are well established in the literature. Applications of stationary ARMA models may be found in many areas of scientific interest. The interested reader is referred to Shumway and Stoffer (2010) for supporting examples and a list of relevant references. Statistical variables, however, are often influenced by a multitude of external factors, and stationarity assumptions cannot be guaranteed over long periods. To ascertain that the statistical analysis is indeed carried out for homogeneous data, the field of structural break or change-point analysis offers a broad range of testing methods and estimation techniques for the retrospective setting. For a recent survey tailored to methods in time series, see Aue and Horváth (2013).

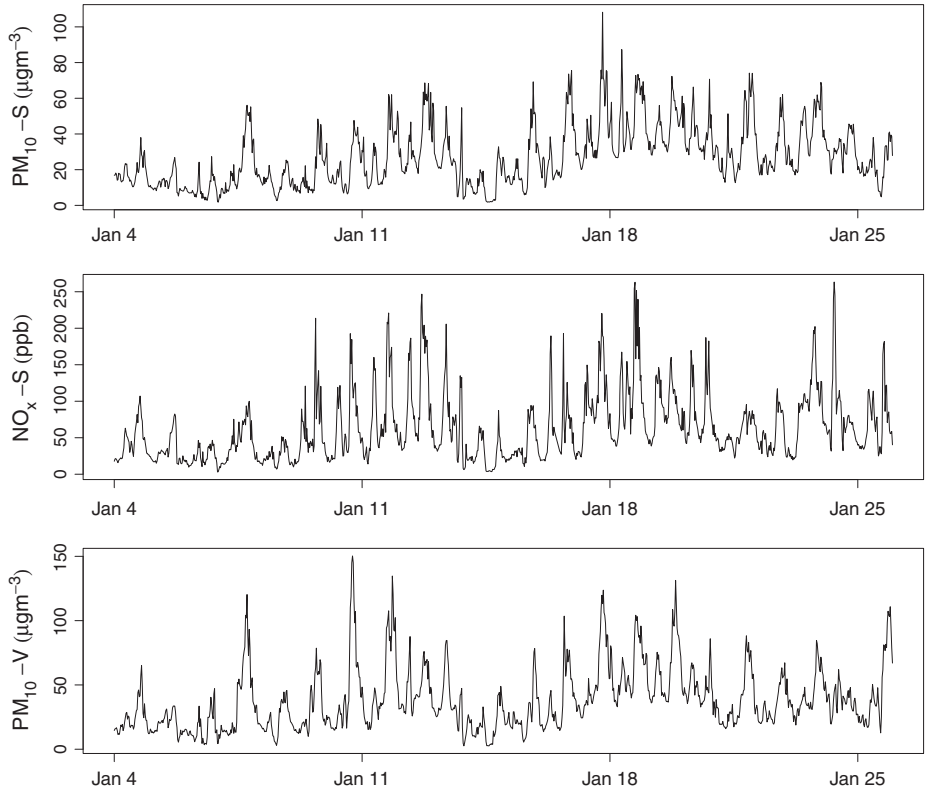

Figure 1. Time series plots of PM_{10} (top) and NO_x (middle) recorded at the Sterneckstraße station and PM_{10} recorded at the Völkermarkterstraße station (bottom)

DOI: 10.1111/jtsa.12062

Recently, there has been a number of research contributions in the area of sequential change-point analysis. These methods appear to be more useful if, as in the pollution data example, a decision has to be made on-line as new data become available. However, the body of literature for this case is still considerably smaller than for retrospective settings. One of the main contributions, written for potential applications in econometrics, is Chu *et al.* (1996). These authors have developed sequential procedures based on the general paradigm that an initial period of length m is used to estimate a model, with the goal to then monitor for breaks in the model parameters on-line. Asymptotic analysis can be carried out letting $m \to \infty$. The approach of Chu *et al.* (1996) has since been extended and refined in several directions. Literature focusing on sequential problems in econometrics are Andreou and Ghysels (2006), Aue *et al.* (2006) and Berkes *et al.* (2004). Gombay and Serban (2009) introduced several tests to monitor the parameters of an autoregressive process. The focus of this article is on introducing several residual-based monitoring schemes for ARMA time series. These are set up as stopping times that reject the assumption of structural stability at the first instance a suitably constructed detector function crosses a certain threshold. In applications, using powerful recursive methods such as the innovations algorithm allows for an efficient computation of the residuals and hence of the stopping time. Convergence to the limit is rather fast, thereby providing excellent approximations also in finite samples of moderate size. Related work in the area, albeit in the retrospective setting, is due to Bai (1993), who considered partial sums of zero-mean ARMA residuals and, more recently, Yu (2007) who derived limit theorems for the corresponding high moment partial sum processes of ARMA series with possibly non-zero mean. Davis *et al.* (1995) studied tests for changes in the parameter values and the order of a zero-mean autoregressive model.

The methodology of this article is noteworthy in several respects. First, the proposed methodology does not make the unrealistic assumption of known null parameters commonly found in the sequential literature (see Gombay and Serban (2009) for a discussion). Second, the residual-based procedures do not suffer from inflated levels often connected to the estimation of long-run variances. On the contrary, we obtain empirical levels that are (a) close to their nominal counterparts after a finite sample adjustment and (b) vary little across different forms of short memory autocorrelation. This makes the procedure attractive to the practitioner who can apply the tests with only slight modifications for finite samples. Third, the choice of threshold is flexible as only weak generic assumptions are being made. It is shown in the simulation study how thresholds can be constructed to achieve certain modelling goals. Fourth, all methods have a quantified asymptotic that can serve as guidance on the expected behaviour of the procedures in finite samples.

The article is organized as follows. The model is specified, assumptions are detailed and large-sample results are stated and discussed in Section 2. Section 3 discusses the application to monitoring pollution data, while Section 4 contains a comparative simulation study. Section 5 concludes, and all proofs are given in Section 6.

2. METHODOLOGY

2.1. The ARMA framework

Let $\mathbb{Z}$ denote the set of integers. A time series $(Y_t : t \in \mathbb{Z})$ is called an ARMA process of orders p and q, in short ARMA(p, q), if it satisfies the stochastic difference equations

$$\phi(B)(Y_t - \mu) = \theta(B)\varepsilon_t, \qquad t \in \mathbb{Z}, \tag{1}$$

where μ is an unknown mean parameter, $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ denote the autoregressive and moving average polynomial respectively, and B is the backshift operator. It is assumed for the moment that the orders p and q are known, but it will be shown as part of Section 3 how to deal with unknown orders. The innovations $(\varepsilon_t : t \in \mathbb{Z})$ are taken to be a sequence of i.i.d. random variables with $E[\varepsilon_1] = 0$ and $E\left[\varepsilon_1^2\right] = \sigma^2$. It is further assumed that $\phi(z)$ and $\theta(z)$ have no common zeroes and that

$$\phi(z) \neq 0 \ \text{ and } \ \theta(z) \neq 0 \ \text{ for all } |z| \leq 1. \tag{2}$$

J. Time. Ser. Anal. **35:** 239–261 (2014)
DOI: 10.1111/jtsa.12062

wileyonlinelibrary.com/journal/jtsa

The latter conditions are equivalent to the causality and invertibility of the centred sequence $(X_t : t \in \mathbb{Z})$, where $X_t = Y_t - \mu_t$ for all $t \in \mathbb{Z}$. The random variables in (1) depend on $p + q + 2$ parameters, which are combined into the vector

$$\boldsymbol{\xi} = (\mu, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2),$$

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)$.

Interest is in monitoring the constancy of $\boldsymbol{\xi}$ in time, as many statistical procedures require stationary data so as to draw valid conclusions. The approach taken here is useful whenever data are collected sequentially in time and decisions are to be made on-line. In particular, it is useful for environmental data, for example, if one is interested in controlling pollution levels and wishes to detect effects of certain control measures as soon as possible. To enable asymptotic theory, the existence of a reference frame of $m + p$ observations, the training period, is required. In this training period, the ARMA variates are generated by the same parameter values, that is,

$$Y_{1-p}, \ldots, Y_m \quad \text{are governed by} \quad \boldsymbol{\xi}_0 = \left(\mu_0, \boldsymbol{\phi}_0, \boldsymbol{\theta}_0, \sigma_0^2\right). \tag{3}$$

Limit results can thus be stated for $m \to \infty$; see Aue *et al.* (2009) and Chu *et al.* (1996) for further discussion. Condition (3) can be used to first estimate the parameters under stable conditions and then employ these estimates for comparisons with estimates based on future observations. The size of the training period, m, can therefore also be interpreted as a measure of how well one knows the parameters governing the dynamics of the time series under consideration. It is assumed throughout that $\sqrt{m}$-consistent estimators $\hat{\boldsymbol{\xi}}_m = \left(\hat{\mu}_m, \hat{\boldsymbol{\phi}}_m, \hat{\boldsymbol{\theta}}_m, \hat{\sigma}_m^2\right)$ for the unknown parameters $\boldsymbol{\xi}_0$ are available, for example, obtained by least squares for AR processes (Shumway and Stoffer, 2010). Denote by $\mathbb{N}$ the positive integers and by

$$\hat{\varepsilon}_t = \hat{X}_t - \sum_{j=1}^{p} \hat{\phi}_{jm} \hat{X}_{t-j} - \sum_{j=1}^{q} \hat{\theta}_{jm} \hat{\varepsilon}_{t-j}, \qquad t \in \mathbb{N}, \tag{4}$$

the model residuals, with $\hat{X}_t = Y_t - \hat{\mu}_m$ and initialization values $\hat{\varepsilon}_{-q+1} = \cdots = \hat{\varepsilon}_0 = 0$ if $q > 0$. All monitoring procedures to be introduced in the remainder of this section will be defined in terms of these residuals.

2.2. Monitoring the mean

A monitoring procedure is devised to check for the constancy of the mean parameter μ. This is carried out by introducing the testing problem with null hypothesis

$$H_0^{\mu}: \quad E[Y_{m+1}] = \cdots = E[Y_{m+N}] = \mu_0,$$

where N denotes the maximal number of observations to be collected before H_0 is accepted. We will quantify the termination time N in terms of the training period size m and assume that $N = \lfloor mT \rfloor$ for some $T > 0$. Here, $\lfloor \cdot \rfloor$ signifies integer part. The null hypothesis is to be tested against the alternative hypothesis of a mean break at an unknown time

$$H_A^{\mu}: \quad E[Y_{m+1}] = \cdots = E[Y_{m+k^*-1}] = \mu_0 \neq \mu_A = E[Y_{m+k^*}] = \cdots = E[Y_{m+N}],$$

with $0 < k^* \leq N$. Note that in the present case, the remaining model parameters $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ and σ^2, which determine the second-order structure, are not allowed to change. The hypotheses H_0^{μ} and H_A^{μ} are arguably the most studied pair in the analysis of structural breaks in time series. The bulk of this research, however, has been conducted in the retrospective setting. For contributions to the time series case, we refer here only to Aue and Horváth (2013)

DOI: 10.1111/jtsa.12062

and to the references in this article. If, besides level shifts, breaks in the second-order structure are also of concern, more general monitoring procedures will be introduced in the next section.

To sequentially test H_0^μ against H_A^μ, a stopping rule will be defined. These are usually given as the first time a suitably constructed detector exceeds a threshold function (see Aue *et al.* (2006) and Chu *et al.* (1996), among others). The monitoring procedure is constructed as follows. The detectors are built from the cumulative sum (CUSUM) of the model residuals (4). The threshold functions are generic and assumed only to satisfy weak regularity conditions. This way, flexibility is given to the practitioner, who can tailor the threshold to the specific needs of the analysis. In Section 4, we design a threshold function that is chosen so that changes can be detected uniformly well over the whole monitoring period. Introducing the CUSUM-type detector

$$D_m^\mu(k) = \sum_{t=m+1}^{m+k} \hat{\varepsilon}_t - \frac{k}{m} \sum_{t=1}^{m} \hat{\varepsilon}_t, \qquad k = 1, \ldots, \lfloor mT \rfloor,$$

and noting that, under H_0^μ, residuals $\hat{\varepsilon}_t$ should be close to the corresponding innovation ε_t, the first stopping rule is defined as

$$\tau_{m,T}^\mu = \min \left\{ k = 1, \ldots, \lfloor mT \rfloor : \frac{|D_m^\mu(k)|}{\sqrt{m}\hat{\sigma}_m} \geq c\, g_m(k) \right\}, \tag{5}$$

where $\hat{\sigma}_m^2$ denotes a weakly consistent estimator for σ^2, and $c = c_\alpha$ a critical constant that can be chosen such that the false alarm rate has asymptotic level α. The notation $g_m(k) = g(k/m)$ is used, and the threshold function g is assumed to be continuous on $[0, T]$ satisfying, for an $\epsilon > 0$,

$$\inf_{\epsilon \leq x \leq T} g(x) > 0 \quad \text{and} \quad \inf_{0 < x < \epsilon} \frac{g(x)}{x^\gamma} \geq C, \tag{6}$$

where $C > 0$ is a constant and $\gamma < 1/2$. The limiting behaviour of $\tau_{m,T}^\mu$ under the null hypothesis is given in Theorem 1.

Theorem 1. Let $(Y_t : t \in \mathbb{Z})$ follow the ARMA(p, q) equations (1) and assume that (2), (3) and (6) are satisfied. Then, for any $T > 0$, it holds under H_0^μ that

$$\lim_{m \to \infty} P\left(\tau_{m,T}^\mu < \infty\right) = P\left(\sup_{0 \leq x \leq T} \frac{|V(x)|}{g(x)} \geq c \right),$$

where $(V(x) : x \in [0, T])$ is a centred Gaussian process with covariances $E[V(x)V(y)] = x(1 + y)$ for $0 \leq x \leq y \leq T$.

The proof of Theorem 1 can be found in Section 6. Next, attention is on the behaviour of the monitoring procedure under the alternative H_A^μ. To quantify the time of change in terms of the training period m, it is additionally assumed that

$$k^* \leq T^* m^\beta \qquad \text{with some } \beta \leq 1, \tag{7}$$

where the constant T^* is potentially dependent on β and must satisfy $T^* \in (0, T)$ in case $\beta = 1$. Condition (7) ensures the separation between the mean break and the end of monitoring is sufficiently large in order for the closed-end procedure $\tau_{m,T}^\mu$ to signal deviation from the null. It can then be proved that the procedure has asymptotic power one. The proof of the following theorem is given in Section 6.

DOI: 10.1111/jtsa.12062

Theorem 2. Under the assumptions of Theorem 1 and if (7) is satisfied, it holds under H_A^{μ} that

$$\lim_{m\to\infty} P\left(\tau_{m,T}^{\mu} < \infty\right) = 1.$$

This section is closed with the remark that a slightly different version of the stopping rule $\tau_{m,T}^{\mu}$ could be formulated in terms of the estimated one-step ahead prediction residuals, say, $\tilde{\varepsilon}_t = Y_t - Y_t\left(\hat{\boldsymbol{\xi}}_m\right)$. As it is known that these prediction residuals are asymptotically equivalent to the residuals in (4), with the rate of convergence being exponential (Brockwell and Davis, 1991), it follows at once that the resulting monitoring procedure has the same large-sample behaviour under both hypotheses as $\tau_{m,T}^{\mu}$. This may come in handy as computations and estimations may be performed using the well-established innovations algorithm. Modifying the methodology developed in Aue *et al.* (2009), one might also determine the limit distribution of $\tau_{m,T}^{\mu}$.

2.3. Monitoring general parameter breaks

In some cases, it may suffice to focus monitoring on level breaks only. A better understanding of the evolution of a dynamic process can be obtained from tracking its second-order behaviour. Denote by $F_t(\cdot\,;\boldsymbol{\xi})$ the distribution function of Y_t, emphasizing its dependence on the parameter $\boldsymbol{\xi}$. This leads us to consider the general hypotheses

$$H_0^{\boldsymbol{\xi}}: \quad F_{m+1}(\cdot\,;\boldsymbol{\xi}_0) = F_{m+2}(\cdot\,;\boldsymbol{\xi}_0) = \cdots = F_{m+N}(\cdot\,;\boldsymbol{\xi}_0)$$

and

$$H_A^{\boldsymbol{\xi}}: \quad F_{m+1}(\cdot\,;\boldsymbol{\xi}_0) = \cdots = F_{m+k^*-1}(\cdot;\boldsymbol{\xi}_0) \neq F_{m+k^*}(\cdot\,;\boldsymbol{\xi}_A) = \cdots = F_{m+N}(\cdot\,;\boldsymbol{\xi}_A)$$

so that obviously $\boldsymbol{\xi}_0 \neq \boldsymbol{\xi}_A$. The structural break can therefore occur in any of the model parameters: the mean, the ARMA coefficients and the innovations variance. Mimicking the steps of the previous subsection, one can establish monitoring schemes that are sensitive with respect to breaks in μ, $\phi_1,\ldots,\phi_p$ and $\theta_1,\ldots,\theta_q$, for the moment precluding σ^2 from changing so as to derive limit results. The general procedure will be based on the squared residuals in (4). Define the detector function

$$D_m^{\boldsymbol{\xi}}(k) = \sum_{i=m+1}^{m+k} \hat{\varepsilon}_i^2 - \frac{k}{m}\sum_{i=1}^{m} \hat{\varepsilon}_i^2, \qquad k = 1,\ldots,\lfloor mT\rfloor.$$

Equipped with a generic threshold function $g_m(k) = g(k/m)$ and a weakly consistent estimator $\hat{\eta}_m^2$ of the moment quantity $\eta^2 = E\left[\left(\varepsilon_i^2 - \sigma^2\right)^2\right]$, set up the stopping rule

$$\tau_{m,T}^{\boldsymbol{\xi}} = \min\left\{k = 1,\ldots,\lfloor mT\rfloor : \frac{\left|D_m^{\boldsymbol{\xi}}(k)\right|}{\sqrt{m}\hat{\eta}_m} \geq c g_m(k)\right\}. \tag{8}$$

If the innovation sequence $(\varepsilon_t : t \in \mathbb{Z})$ possesses finite fourth-order moments, results similar to Theorems 1 and 2 can be established.

Theorem 3. Let $(Y_t : t \in \mathbb{Z})$ follow the ARMA(p,q) equations (1) and assume that (2), (3) and (6) are satisfied and that $E\left[\varepsilon_1^4\right] < \infty$. Then, for any $T > 0$, it holds under $H_0^{\boldsymbol{\xi}}$ that

$$\lim_{m\to\infty} P\left(\tau_{m,T}^{\boldsymbol{\xi}} < \infty\right) = P\left(\sup_{0\leq x\leq T} \frac{|V(x)|}{g(x)} \geq c\right),$$

DOI: 10.1111/jtsa.12062

where $(V(x) : x \in [0, T])$ is defined in Theorem 1.

Theorem 4. Under the assumptions of Theorem 3 and if (7) is satisfied, it holds under $H_A^{\boldsymbol{\xi}}$ that

$$\lim_{m\to\infty} P\left(\tau_{m,T}^{\boldsymbol{\xi}} < \infty\right) = 1.$$

The proofs of Theorems 3 and 4 are given in Section 6. The same results also hold if the one-step prediction residuals were used in place of the residuals (4) to define the procedure.

2.4. Discussion

The monitoring procedures introduced in this section are based on parametric time series models and are constructed from the residuals of their fit to the data. In contrast to non-parametric procedures that are typically based on CUSUM-type detectors of the data itself, the estimation of a long-run variance parameter can be avoided. It will be shown in the simulations that the proposed procedures work well across a variety of different (short memory) dependence structures. The results are in accordance with Robbins *et al.* (2011), who considered the simpler case of mean shifts for correlated data in a historical setting.

The choice of the threshold function g is flexible. The distribution of $(V(x) : x \in [0, T])$ in Theorems 1 and 3 is the same as that of (a) $(B(x) + xZ : x \in [0, T])$, where $(B(x) : x \in [0, T])$ denotes a standard Brownian motion and Z a standard normal random variable, and (b)

$$\left((1 + x)B\left(\frac{x}{1 + x}\right) : x \in [0, T]\right),$$

where $(B(x) : x \in [0, T])$ is again a standard Brownian motion. Both claims can be verified from direct computations of the covariance functions. This motivates the choice of $g(x) = 1 + x$. So as to add to the sensitivity of the procedures, we shall discuss in the following text the threshold functions

$$g(x) = (1 + x)\left(\frac{x}{1 + x}\right)^{\gamma}, \qquad \gamma < \frac{1}{2}. \tag{9}$$

Note that $\gamma = 1/2$ is excluded because of the law of the iterated logarithm for Brownian motions as $x \to 0$, which states that $B(x) \sim \sqrt{x \log\log 1/x}$ with probability one (as $x \to 0$). It follows in particular that, even under H_0, rejection occurs with probability one. This may also be seen from the equality

$$P\left(\sup_{0\le x\le T} \frac{|V(x)|}{g(x)} \ge c\right) = P\left(\sup_{0\le x\le T/(1+T)} \frac{|B(x)|}{x^{\gamma}} \ge c\right)$$

for g as in (9). This article discusses only closed-end procedures because open-end procedures can often not be recommended in practice. For open-end procedures, the supremum on the right-hand side of the latter equation would have to be evaluated for all $x \in [0, 1]$, so that critical values will be larger, rendering the open-end procedures to be too conservative in most finite sample situations. It is also explicit how the factor T for closed-end procedures affects the critical values.

3. MONITORING POLLUTION CONCENTRATIONS

3.1. Background

The PM_{10} and NO_x recordings displayed in Figure 1 were obtained from measuring stations in Klagenfurt, a city in the southern part of Austria. Klagenfurt (coordinates 46°37′N, 14°18′E) is the capital of the federal state

DOI: 10.1111/jtsa.12062

Carinthia and has a population of about 94,000. It is elevated 446 m above sea level and is located in a basin area south of the main Alpine crest, surrounded by mountains of up to 1000 m in altitude. The local topography and climate are conducive to creating stationary temperature inversions in the cold season (roughly from October to March) that features low precipitation and low wind velocities. Temperature inversion can enhance the effect of emissions from traffic by stabilizing the air mass and slowing dilution processes of aerosol particles through reduced turbulence and mixing of different atmospheric layers. This effect is most pronounced during rush-hour traffic for weekday mornings and evenings (Janhäll *et al.*, 2006).

Concentration of particulate matter with an aerodynamic diameter smaller than 10 μm is routinely measured as PM_{10} in units of μgm^{-3}. Limits set by EU regulation prescribe that a threshold daily PM_{10} average value of 50 μgm^{-3} must not exceed on more than 7 days per year. The annual PM_{10} average threshold is set at 20 μgm^{-3} (Stadlober *et al.*, 2008). The composition of particles within the PM_{10} class is complex. While there are natural members of the class such as pollen, most of the PM_{10} mass in urban environments is due to pollutants originating from road traffic (Charron and Harrison, 2003). The larger particles (>2.5μm) consist of smoke, dirt and dust, while the smaller particles are toxic organic compounds or heavy metals (Stadlober *et al.*, 2008). Several more recent studies have even isolated UFP with an aerodynamic diameter smaller than 100 nm as the main source of the negative health effects incurred by these pollutants (see Oberdörster and Utell (2002) & Pekkanen *et al.* (1997), among others). Because of the small contribution they have on the overall mass of PM_{10} particles, UFP are much harder to track. Environmental researchers have made connections between the concentration of UFP and the concentration of gases emitted by traffic such as nitrogen oxides, NO_x, and carbon monoxide, CO (Janhäll *et al.*, 2006). This article uses nitrogen oxides to track the evolution of UFP. The term NO_x measures the joint concentration of nitrogen monoxides (NO) and nitrogen dioxides (NO_2) in parts per billion (ppb).

On 11 January 2012, Klagenfurt began using CMA citywide as a de-icer so as to limit pollution concentrations in the winter months. The EU Life Project CMA+ (see EU Life project CMA+ for more information), with a budget of EUR 2,720,033.00 (out of which EUR 1,344,966.50 are contributed by the EU), aims at promoting the use of CMA as a dust binder throughout the year and also as a de-icing agent in the winter season. The CMA+ Newsletter (EU Life project CMA+ Newsletter) has claimed that CMA has reduced particulate matter pollution by about 30%. This was a controversial issue as the effect of CMA was questioned in a number of local newspaper articles. In the following, we shall devise monitoring procedures to investigate if there are any statistically significant effects present in the data.

3.2. Data description

The data consist of two PM_{10} series taken at Sterneckstraße and Völkermarkterstraße (in the following referred to as PM_{10}-S and PM_{10}-V respectively) and an NO_x series taken at Sterneckstraße. The measurements for each series consist of 1056 half-hourly averages measured from 4 January 2012 to 25 January 2012. To stabilize the variance, the fourth-root transformation was applied. Data summaries for the raw and transformed data shown in Table I indicate that the fourth-root transformed series are reasonably symmetric.

Table I. Data summary

Particles	Mean	Med	SD	Min	Max
PM_{10}-S (μgm^{-3})	16.42	13.70	10.52	1.80	56.10
	1.95	1.92	0.30	1.16	2.74
PM_{10}-V (μgm^{-3})	28.54	20.60	23.75	3.00	150.30
	2.20	2.13	0.39	1.32	3.50
NO_x-S (ppb)	41.39	30.70	31.91	3.10	213.70
	2.42	2.35	0.43	1.33	3.82

Note: The suffix S (V) refers to Sterneckstraße (Völkermarkterstraße). First lines contain the raw data and second lines the fourth-root transformed data.

DOI: 10.1111/jtsa.12062

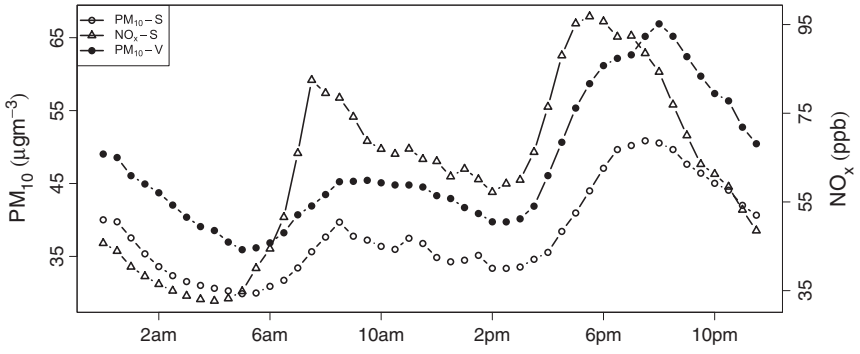

Figure 2. Average diurnal patterns

All three series display a strong diurnal pattern. The average diurnal concentrations are displayed in Figure 2. It can be seen that rush-hour traffic in the mornings (series peaks between 0700 and 0800 h) and in particular at night (series peaks between 1800 and 2000 h) is associated with significant increases in concentrations for both PM_{10} and NO_x series.

3.3. Model building and monitoring process

The on-line monitoring procedure is set up so that interest is in tracking the dynamics of the PM_{10} measurements taken at the Sterneckstraße monitoring station. A natural date to conclude the training period is 11 January 2012, which corresponds to the start of local application of the compound CMA. So as to use a monitoring horizon specified by $T = 2$, the training and monitoring periods are set to 336 and 672 observations respectively, thus dropping the 48 observations occurring on 25 January. Figure 3 shows the differences in the average diurnal patterns for the training and monitoring periods. There is a marked increase in the averages of all pollutants after 11 January. This observation has led some to the conclusion that the use of CMA was ineffective, but the effect visible in Figure 3 could be due to natural (seasonal) variation or to a severe temperature inversion occurring concurrently (for which no information is available).

So as to employ the proposed monitoring procedure, the following steps were applied. Non-stationary behaviour in the form of local and seasonal dynamics were removed by utilizing the regression model

$$x_{1,t} = 0.51 + 0.43x_{2,t} + 0.20x_{3,t} + e_t, \tag{10}$$

where $x_{1,t}$ and $x_{2,t}$ are the PM_{10}-S and PM_{10}-V recordings, and $x_{3,t}$ the NO_x-S recordings, all in the fourth-root transformed versions. The proposed procedures were applied to the transformed residual series $y_t = \text{sign}(e_t) * |e_t|^{0.75}$. The transformation was chosen to symmetrize the observed innovations. Note that the finite sample critical values used here were obtained from simulating the limit distributions in Theorems 1 and 3, which imply Gaussianity. The role of transformations in the model building process has been discussed in detail in Box and Cox (1968). The effect of the transformation is visualized in Figure 4. It can be seen from the box plots and the qq-plots that symmetry and normality are reasonably well established. Model selection criteria such as AIC and BIC supported an AR(3) model for the modified residual training set. The maximum likelihood estimates of the unknown parameters gave the model

$$y_t = 0.0004 + 0.68y_{t-1} + 0.05y_{t-2} + 0.12y_{t-3} + \epsilon_t, \qquad \hat{\sigma}_m^2 = 0.02.$$

Several retrospective tests presented in Davis *et al.* (1995) were applied to the training set under the AR(3) model framework. The smallest p-value was 0.3296, which indicates support of the non-contamination assumption.

DOI: 10.1111/jtsa.12062

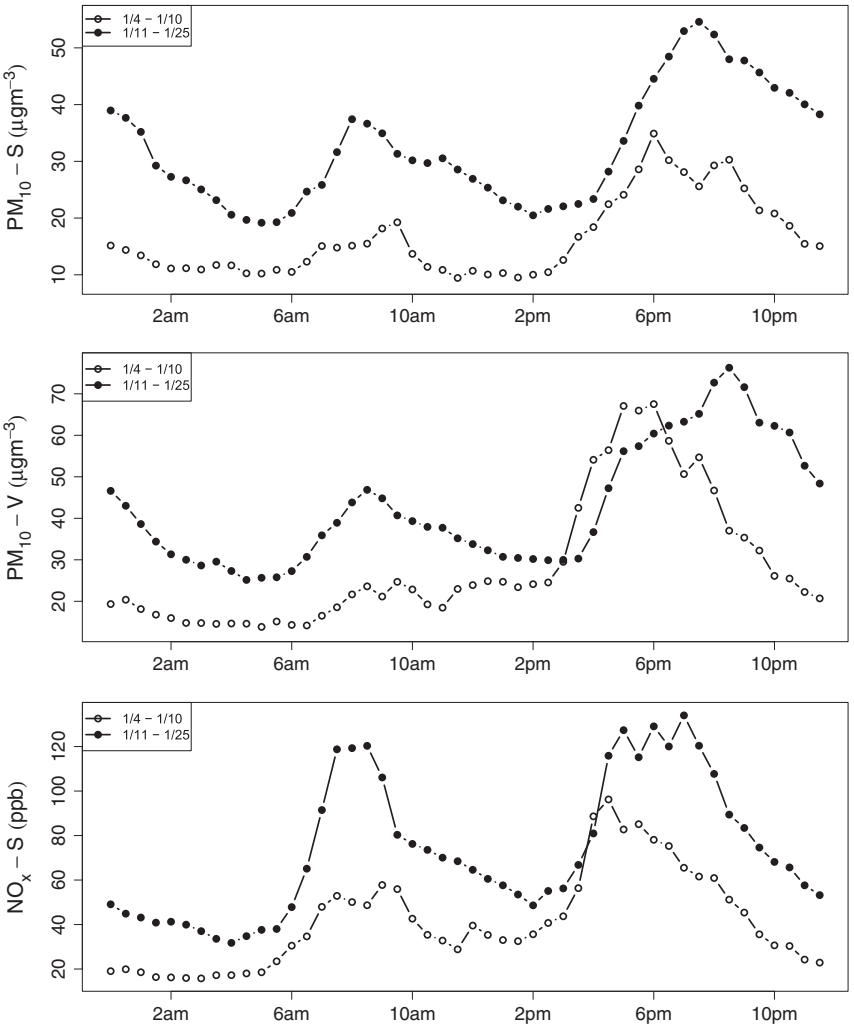

Figure 3. Average diurnal patterns for training and monitoring periods

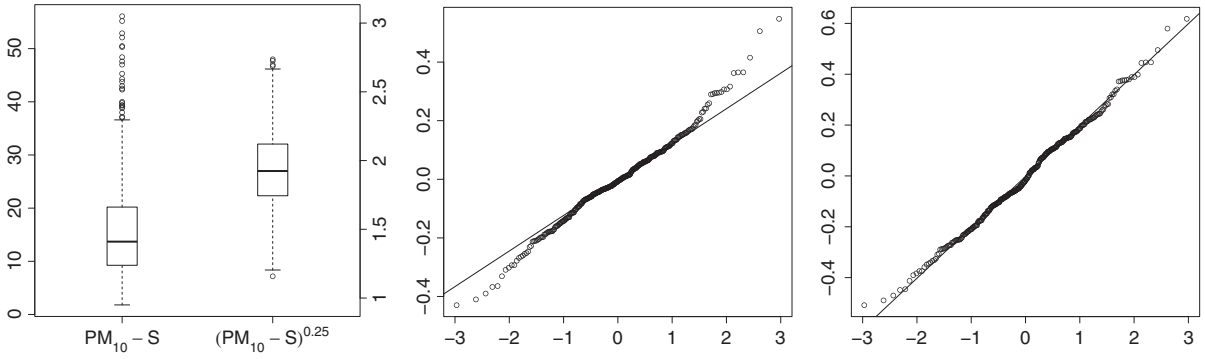

Figure 4. Box plots of original and fourth-root transformed PM_{10}-S observations (left), qq-plots of original regression residuals (middle) and power transformed regression residuals (right)

DOI: 10.1111/jtsa.12062

Both the mean only procedure $\tau^{\mu}_{m,T}$ in (5) and the general procedure $\tau^{\boldsymbol{\xi}}_{m,T}$ in (8) were used to monitor the residual series at the 0.05 level using the threshold function g in (9) with $\gamma = 0$. Implementing the approach described in Section 4, reasonable finite sample critical values for the mean and general testing procedures under the AR(3) framework are 2.15 and 2.10 respectively. The bottom panel of Figure 5 shows the mean only procedure rejected the null hypothesis at the 397th monitoring observation (0600 h on 19 January 2012), while the general procedure failed to detect a change. The observed rejection suggests a change may have occurred in the structural dynamics of the PM_{10} series following the start of CMA application. This conclusion was supported by a retrospective analysis using the test statistics in Ploberger and Krämer (1982) on all available data. The null hypothesis was rejected on 15 January, a couple of days before the mean only monitoring procedure was terminated, thereby indicating the proposed method works reasonably well. Note that Figure 5 indicates that at the same date, the mean only detector starts to steadily increase.

The increased residual mean levels may be explained with a change in the dynamics of the assumed underlying regression model (10). This is evident from the plots in Figure 6, which show the diurnal differences to be positive for most of the day. While there is a probable change in the dynamics of the pollution series, the reported reduction of pollution levels by 30% claimed in EU Life project CMA+ Newsletter does not seem to be supported by the data, at least in the short run. On the contrary, a simple computation of averages in the training and monitoring periods shows a 19% increase in mean levels for the transformed data, corresponding to a 97% increase in PM_{10} levels for the observations taken at Sterneckstraße. It should be noted, however, that a simple comparison of averages may not reveal the true and more complex behaviour of pollution concentrations before and after the application of CMA. Since information on potentially important covariates was not available for the analysis, a more in-depth analysis is beyond the scope of the present exposition.

4. SIMULATIONS

This section starts with providing a summary of critical values obtained from the limiting distributions seen in Theorems 1 and 3. It is followed by a presentation and assessment of possible finite sample adjustments. Table II

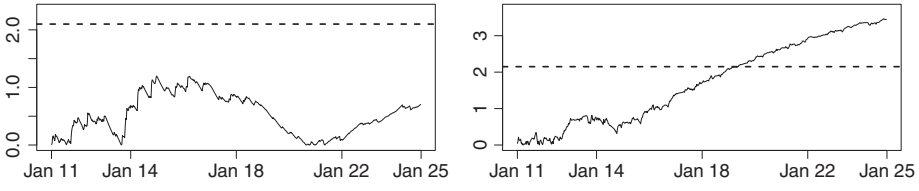

Figure 5. Monitoring plots for the general detector (left) and mean only detector (right) for the PM_{10} residual series

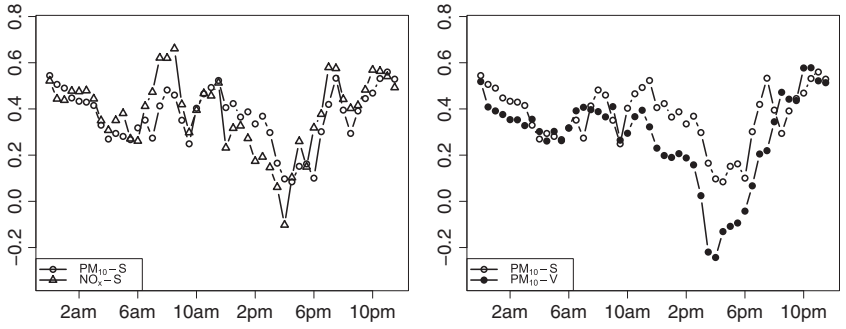

Figure 6. Differences in average diurnal patterns between training and monitoring periods

DOI: 10.1111/jtsa.12062

Table II. Simulated 0.10 and 0.05 critical values for asymptotic results

		T				
γ	α	1	2	3	4	5
0.49	0.10	2.854	2.853	2.870	2.879	2.874
	0.05	3.101	3.090	3.115	3.118	3.114
0	0.10	1.383	1.597	1.697	1.752	1.786
	0.05	1.577	1.821	1.945	2.010	2.050
-5	0.10	0.037	0.182	0.350	0.495	0.621
	0.05	0.044	0.215	0.413	0.586	0.738

Table III. Finite sample rules of thumb for approximate 0.05 critical values for models with small orders

		T		
γ	m	1	2	5
0.49	100	4.000	4.100	4.200
	250	3.600	3.650	3.700
	500	3.475	3.500	3.500
0	100	2.000	2.400	2.700
	250	1.730	2.025	2.300
	500	1.650	1.920	2.150
-5	100	0.055	0.280	0.950
	250	0.047	0.240	0.820
	500	0.046	0.225	0.780

displays critical values for level α tests for various choices of monitoring multiple T and sensitivity parameter γ when utilizing the threshold function given by equation (9). The percentiles were approximated by constructing 50,000 sample paths each based on partial sums of 50,000 standard normal random variables. The similarity in critical values across T for larger γ values is due to the sensitivity to early monitoring detection. The simulations indicate smaller values of γ are more suitable for detecting changes that occur later in the monitoring set. If the practitioner has no prior knowledge of the location of a possible change point, it is suggested to use the value $\gamma = 0$ because this specification tends to spread out the occurrence of false rejections over the entire monitoring period.

The empirical levels of the asymptotic critical values tend to be slightly inflated in finite samples. Consequently, a simple and flexible adjustment is proposed, which is found by first generating a large number of sample paths from the desired ARMA model, assuming structure stability, and then selecting the $1 - \alpha$ empirical percentile of the scaled detector

$$\max_{1 \le k \le mT} \frac{\left| D_m^{\xi}(k) \right|}{\sqrt{m} g_m(k) \hat{\eta}_m}.$$

Based on the proposed adjustment procedure, the values provided in Table III are found to be adequate for ARMA models with small orders. Furthermore, these critical values were obtained using innovations drawn from a normal distribution; hence, applying an appropriate transformation on the model residuals may be required to ensure the approximate size. To further justify these claims, various simulations are offered based on the following models:

$$\text{Model I: } y_t = \alpha_i + \phi_i y_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}\left(0, \sigma_i^2\right),$$

$$\text{Model II: } y_t = \alpha_i + \theta_i \epsilon_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}\left(0, \sigma_i^2\right),$$

$$\text{Model III: } y_t = \alpha_i + \phi_i y_{t-1} + \theta_i \epsilon_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}\left(0, \sigma_i^2\right)$$

DOI: 10.1111/jtsa.12062

The aforementioned models are the well-known AR(1), MA(1) and ARMA(1,1) processes. Unless otherwise specified, subsequent simulations will assume the null parameter values of $\alpha_0 = 0$, $\phi_0 = 0.3$, $\theta_0 = 0.3$ and $\sigma_0 = 1$. The alternative values will be denoted by α_1, ϕ_1, θ_1 and σ_1. Table IV provides a summary of the empirical size performance of the three models for tests involving the general detector with a monitoring multiple of $T = 2$ and sensitivity parameter of $\gamma = 0$. The observed type I error rates are encouraging because they remain close to the nominal level even when the underlying parameters are close to the non-stationarity region. This feature is desirable because the presented framework does not assume the model parameters to be known.

This section is concluded by providing three testing scenarios, which aim at assessing the proposed procedure's ability to correctly reject the hypothesis of structural stability. For these simulations, the training period is set to $m = 250$, and monitoring stops at $mT = 500$ observations, using $T = 2$. The criteria used are empirical power and average delay time (ADT) for correct rejections. The tabulated results provide the null case as a point of reference, in which power and ADT represent the empirical type I rate and the average false stopping time respectively. Additionally, three different choices of change locations are used, namely $k^* = 25$, representing an early change; $k^* = 250$, an intermediate change; and $k^* = 400$, a late change. The first testing scenario examines changes in regression coefficients for the AR(1) and MA(1) models. Table V displays empirical power and ADT when a change point occurs at monitoring time point k^* as specified earlier. Clearly, smaller values of k^* and γ result in tests with higher power. However, larger values of γ tend to report shorter delay times for early change points. As is expected, the opposite holds for later changes.

The simulations indicate the detector for monitoring the mean outperforms the general detector when the underlying change occurs only in the model intercept. Table VI displays the simulation results for the mean only detector ($D = 1$) and the general detector ($D = 2$) for Models I–III. Since the direction of the mean shift does not affect our results, empirical power and ADT are provided in terms of $\Delta = |\alpha_0 - \alpha_1|$.

Lastly, changes in the innovation standard deviation are examined, which occur at monitoring time point k^* (even though the tests have no formal theoretical foundation for this case). A summary of the simulations for various models is provided in Table VII. The procedure performs quite well for increases in model volatility, in which the proposed detector should tend to grow quickly. Decreases in volatility do not lead to the same results. This apparent asymmetry is due in part to the speed at which the detector grows in response to changes in volatility and, in addition, due to a slight positive skew in the detector. As a concluding remark, omitted simulations that allowed for multiple parameters to change yielded results with very high power even for late change points.

Table IV. Empirical type I error rates for AR(1), MA(1) and ARMA(1,1) models when $T = 2$ and $\gamma = 0$

Model	m	-0.9	-0.7	-0.5	-0.3	0	0.3	0.5	0.7	0.9
						ϕ_0				
I	100	0.051	0.041	0.043	0.044	0.044	0.044	0.045	0.052	0.076
	250	0.052	0.047	0.045	0.048	0.047	0.046	0.046	0.052	0.054
	500	0.051	0.045	0.049	0.051	0.048	0.051	0.048	0.047	0.050
						θ_0				
II	100	0.056	0.062	0.053	0.044	0.041	0.044	0.040	0.051	0.058
	250	0.062	0.051	0.044	0.050	0.045	0.048	0.044	0.048	0.065
	500	0.056	0.050	0.050	0.044	0.049	0.046	0.049	0.048	0.056
						ϕ_0 ($\theta_0 = 0.3$)				
III	100	0.063	0.059	0.066	0.071	0.065	0.058	0.057	0.056	0.071
	250	0.054	0.054	0.053	0.067	0.058	0.054	0.054	0.053	0.053
	500	0.052	0.051	0.049	0.059	0.055	0.052	0.050	0.051	0.051
						θ_0 ($\phi_0 = 0.3$)				
III	100	0.071	0.075	0.080	0.073	0.059	0.057	0.056	0.070	0.075
	250	0.067	0.060	0.065	0.060	0.054	0.055	0.055	0.055	0.069
	500	0.064	0.054	0.057	0.058	0.052	0.053	0.049	0.053	0.058

Note: Values based on 10,000 simulations using the critical values from Table III.

DOI: 10.1111/jtsa.12062

Table V. Power and ADT for changes in regression coefficients when $(m, T) = (250, 2)$

			Power						ADT					
Model	γ	k^*	-0.9	-0.5	-0.3	0.3	0.5	0.7	0.9	-0.9	-0.5	-0.3	0.3	0.5	0.7	0.9
						ϕ_1							ϕ_1			
I		25	1.000	0.980	0.586	0.049	0.047	0.494	1.000	13.4	85.2	159.0	41.0	140.3	152.8	42.8
	0.49	250	1.000	0.604	0.133	0.048	0.006	0.117	0.948	29.3	136.3	150.3	36.8	118.2	141.8	88.3
		400	0.976	0.095	0.014	0.046	0.001	0.014	0.456	36.0	65.9	63.7	41.4	57.6	63.9	58.2
		25	1.000	0.997	0.804	0.046	0.127	0.696	1.000	17.1	86.6	176.9	266.6	244.3	183.6	48.2
	0	250	1.000	0.836	0.377	0.049	0.052	0.316	0.985	24.1	111.8	136.7	274.4	128.5	132.2	70.8
		400	0.994	0.274	0.084	0.048	0.015	0.070	0.661	27.7	59.1	58.1	270.8	46.2	56.8	50.6
		25	1.000	0.996	0.799	0.043	0.110	0.682	1.000	138.1	284.6	365.3	455.0	417.1	369.8	214.5
	-5	250	1.000	0.863	0.422	0.046	0.072	0.356	0.986	46.1	161.1	189.0	451.6	198.8	188.7	109.6
		400	0.993	0.354	0.143	0.045	0.050	0.124	0.704	30.1	64.3	64.9	451.2	60.3	65.2	55.0
						θ_1							θ_1			
II		25	1.000	0.940	0.604	0.050	0.032	0.153	0.608	32.1	105.0	153.8	37.2	132.8	168.2	164.3
	0.49	250	0.971	0.463	0.149	0.051	0.005	0.025	0.126	93.2	144.4	145.2	41.3	107.6	139.1	154.0
		400	0.359	0.060	0.017	0.048	0.001	0.004	0.016	64.9	66.7	61.3	36.5	56.9	64.5	65.4
		25	1.000	0.987	0.808	0.044	0.093	0.353	0.822	40.1	106.5	172.7	283.7	241.8	227.4	177.7
	0	250	0.997	0.733	0.378	0.050	0.041	0.119	0.376	68.4	122.7	134.3	276.6	127.1	130.1	137.2
		400	0.645	0.217	0.089	0.048	0.013	0.030	0.081	55.5	59.8	59.6	279.5	48.9	52.2	59.9
		25	1.000	0.986	0.796	0.041	0.079	0.344	0.825	223.4	305.2	360.1	451.9	420.3	402.6	364.2
	-5	250	0.997	0.776	0.427	0.049	0.066	0.161	0.433	114.0	169.4	186.4	455.5	202.6	197.5	190.6
		400	0.717	0.292	0.145	0.044	0.048	0.072	0.140	59.8	66.0	65.1	452.9	60.6	64.1	67.5

Note: Values based on 10,000 simulations.

DOI: 10.1111/jtsa.12062

Table VI. Power and ADT for changes in model intercept when $(m, T, \gamma) = (250, 2, 0)$

			Power				ADT			
			$\Delta = \lvert\alpha_0 - \alpha_1\rvert$				$\Delta = \lvert\alpha_0 - \alpha_1\rvert$			
Model	D	k^*	0	0.25	0.75	1.5	0	0.25	0.75	1.5
I		25	0.044	0.812	1.000	1.000	287.7	213.6	52.7	23.4
	1	250	0.044	0.265	0.997	1.000	294.1	153.4	97.3	43.4
		400	0.046	0.044	0.367	0.955	292.7	62.5	66.7	53.6
		25	0.049	0.124	0.954	1.000	262.2	239.2	131.1	24.1
	2	250	0.046	0.050	0.607	1.000	276.8	128.3	132.1	43.4
		400	0.049	0.014	0.154	0.904	270.4	53.6	61.1	48.7
II		25	0.044	0.580	1.000	1.000	292.5	243.2	73.5	31.6
	1	250	0.046	0.159	0.932	1.000	290.7	153.8	125.3	58.6
		400	0.039	0.029	0.203	0.762	295.6	58.1	65.9	61.8
		25	0.043	0.079	0.731	1.000	269.9	250.2	200.3	43.1
	2	250	0.052	0.037	0.281	0.996	274.4	120.3	141.5	77.4
		400	0.050	0.013	0.063	0.552	273.8	48.1	58.0	59.6
III		25	0.045	0.586	1.000	1.000	286.2	239.8	73.8	31.4
	1	250	0.050	0.159	0.931	1.000	291.1	150.1	124.6	58.3
		400	0.047	0.030	0.209	0.762	289.8	59.3	65.7	61.9
		25	0.055	0.090	0.748	1.000	268.3	249.9	193.3	43.2
	2	250	0.063	0.041	0.320	0.993	267.5	115.3	137.0	75.5
		400	0.059	0.014	0.071	0.583	268.8	49.7	57.5	58.2

Note: Values based on 10,000 simulations.

Table VII. Power and ADT for changes in innovation standard deviation when $(m, \gamma, T) = (250, 0, 2)$

		Power							ADT						
		σ_1							σ_1						
Model	k^*	0.5	0.75	1	1.25	1.5	1.75	2	0.5	0.75	1	1.25	1.5	1.75	2
I	25	1.000	0.964	0.048	0.976	1.000	1.000	1.000	88.9	208.3	261.7	133.3	47.4	27.3	18.2
	250	0.892	0.244	0.049	0.603	0.992	1.000	1.000	155.4	179.1	277.6	134.1	83.5	48.4	32.3
	400	0.073	0.018	0.046	0.139	0.508	0.859	0.983	71.9	68.9	280.2	61.9	59.8	52.4	40.6
II	25	1.000	0.965	0.047	0.980	1.000	1.000	1.000	88.7	209.5	284.2	131.6	48.4	27.4	18.4
	250	0.895	0.247	0.047	0.608	0.990	1.000	1.000	156.3	179.3	274.0	133.9	83.1	48.6	31.9
	400	0.075	0.017	0.047	0.134	0.507	0.863	0.981	74.6	70.1	280.0	60.3	60.2	52.1	40.5
III	25	1.000	0.961	0.056	0.981	1.000	1.000	1.000	89.7	213.1	271.7	127.2	46.9	27.0	18.0
	250	0.884	0.212	0.055	0.639	0.992	1.000	1.000	158.2	178.7	274.3	131.2	80.7	46.2	31.3
	400	0.063	0.014	0.056	0.155	0.516	0.866	0.983	74.0	69.7	278.7	60.1	59.2	51.0	39.5

Note: Values based on **10, 000** simulations.

5. CONCLUSIONS

This article proposed an on-line monitoring scheme for pollution concentrations that may help uncover structural shifts in their dynamics. The procedure may therefore be helpful in validating whether assumptions made for prediction purposes are still reasonably satisfied or if the prediction procedures need adjustment. The proposed methodology may be particularly useful for local policy makers who need to implement appropriate measures so as to meet regulation on pollution concentrations and to avoid severe negative health effects for the population. In the context presented, a change in dynamics was uncovered, which may relate to the use of a new de-icer, CMA, introduced by the city of Klagenfurt in January 2012.

The proposed method is developed in general form for the popular class of linear ARMA processes and may therefore be of interest in other fields, such as macroeconomics and finance. The statistical framework used is sequential change-point analysis, where stopping rules are designed as first-hitting times of a curved threshold

DOI: 10.1111/jtsa.12062

function. In the time series case, these procedures are built from (squared) residuals. The large-sample behaviour of these residual-based stopping rules has been quantified and shown to work well in finite samples by means of a comparative simulation study.

6. PROOFS

6.1. Preliminary results

The presentation in this section follows Yu (2007). By assumption (2), the reciprocals of the autoregressive polynomial $\phi(z)$ and the moving average polynomial $\theta(z)$ admit power series expansions, say,

$$\frac{1}{\phi(z)} = \sum_{j=0}^{\infty} \pi_j(\boldsymbol{\phi}) z^j \qquad \text{and} \qquad \frac{1}{\theta(z)} = \sum_{j=0}^{\infty} \psi_j(\boldsymbol{\theta}) z^j .$$

The first auxiliary result establishes the behaviour of the coefficients $\pi_j(\mathbf{v}^*)$ and $\psi_j(\mathbf{u}^*)$ if instead of the true parameter vectors $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, generic elements $\mathbf{v}^* \in \mathbb{R}^p$ and $\mathbf{u}^* \in \mathbb{R}^q$ in their vicinity are used in the power series expansion mentioned earlier. Let $|\cdot|$ denote the maximum norm of vectors.

Proposition 1. Let $(Y_t : t \in \mathbb{Z})$ be an ARMA(p, q) time series following the equations (1) such that (2) holds and let $\mathbf{v}^* \in \mathbb{R}^p$ and $\mathbf{u}^*, \mathbf{u}_1^*, \mathbf{u}_2^* \in \mathbb{R}^q$. Then, there are $\epsilon > 0$, $c \in (0, 1)$ and $K > 0$ such that, for all $j \geq 0$,

(a) $|\pi_j(\mathbf{v}^*)| \leq Kc^j$ if $|\mathbf{v}^* - \boldsymbol{\phi}| \leq \epsilon$;
(b) $|\psi_j(\mathbf{u}^*)| \leq Kc^j$ if $|\mathbf{u}^* - \boldsymbol{\theta}| \leq \epsilon$;
(c) $|\psi_j(\mathbf{u}_1^*) - \psi_j(\mathbf{u}_2^*)| \leq K|\mathbf{u}_1^* - \mathbf{u}_2^*| j c^{j-1}$ if $|\mathbf{u}_1^* - \boldsymbol{\theta}| \leq \epsilon$ and $|\mathbf{u}_2^* - \boldsymbol{\theta}| \leq \epsilon$.

Proof
The proof of these statements can be found in Bai (1993). □

All limit relations in Section 2 will follow from the fact that the model residuals and the true innovations are close. Therefore, one needs to estimate the differences $\Delta_t = \hat{\varepsilon}_t - \varepsilon_t$. This will be carried out utilizing the following decomposition derived in displays (14)–(16) of Yu (2007). Now, under H_0^{μ} and $H_0^{\boldsymbol{\xi}}$, $\Delta_t = \Delta_t\left(\sqrt{m}\left[\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}\right], \sqrt{m}\left[\hat{\boldsymbol{\phi}}_m - \boldsymbol{\phi}\right], \sqrt{m}\left[\hat{\mu}_m - \mu\right]\right)$, where

$$\Delta_t(\mathbf{u}, \mathbf{v}, w) = \zeta_t(\mathbf{u}) + \frac{\beta_t(\mathbf{u}, \mathbf{v})}{\sqrt{m}} + \frac{\rho_t(\mathbf{u}, \mathbf{v}, w)}{\sqrt{m}}.$$

Let $\mathbf{u}^* = \boldsymbol{\theta} + \mathbf{u}/\sqrt{m}$ and $\mathbf{v}^* = \boldsymbol{\phi} + \mathbf{v}/\sqrt{m}$ and set $u_0^* = 1$. The quantities in the previous display are then defined by

$$\zeta_t(\mathbf{u}) = -\sum_{j=1}^{q} \left(\sum_{\ell=1}^{j} \psi_{t-1+\ell}(\mathbf{u}^*) u_{j-\ell}^* \right) \varepsilon_{1-j},$$

$$\beta_t(\mathbf{u}, \mathbf{v}) = -\sum_{j=1}^{p} v_j \sum_{\ell=0}^{t-1} \psi_\ell(\mathbf{u}^*) X_{t-j-\ell} - \sum_{j=1}^{q} u_j \sum_{\ell=0}^{t-1} \psi_\ell(\mathbf{u}^*) \varepsilon_{t-j-\ell},$$

$$\rho_t(\mathbf{u}, \mathbf{v}, w) = -\left(1 - \sum_{j=1}^{p} v_j^*\right) w \sum_{\ell=0}^{t-1} \psi_\ell(\mathbf{u}^*).$$

DOI: 10.1111/jtsa.12062

The aforementioned formulations take already into account that a $\sqrt{m}$-consistent estimator $\hat{\boldsymbol{\xi}}_m$ is available and that one can focus attention on parameter values 'close' to the true parameters. Note also that $\zeta_t(\mathbf{u})$ contains the innovations $\varepsilon_{-q+1}, \ldots, \varepsilon_0$ for which the corresponding residuals have been set to zero. The term $\rho_t(\mathbf{u}, \mathbf{v}, w)$ arises because of the estimation involving the mean parameter μ.

6.2. Proof of Theorems 1 and 2

Let $\delta > 0$. Since $\hat{\boldsymbol{\xi}}_m$ is a $\sqrt{m}$-consistent estimator, there are $C > 0$ and $m_0 \geq 1$ such that

$$P\left(\sqrt{m}\left|\hat{\boldsymbol{\xi}}_m - \boldsymbol{\xi}\right| > C\right) < \delta \qquad \text{for all } m \geq m_0. \tag{11}$$

In what follows, it is therefore sufficient to focus on parameter values $\boldsymbol{\xi}^* = (\mathbf{u}^*, \mathbf{v}^*, w^*)$ such that $\sqrt{m}|\boldsymbol{\xi}^* - \boldsymbol{\xi}| \leq C$. In view of Proposition 1, this implies in particular that the components of $\boldsymbol{\xi}^*$ satisfy $|\mathbf{u}^* - \boldsymbol{\theta}|, |\mathbf{v}^* - \boldsymbol{\phi}|, |w^* - \mu| \leq C/\sqrt{m}$.

Lemma 1. Under the assumptions of Theorem 1, it holds, for all $C > 0$ and as $m \to \infty$,

$$\max_{1 \leq k \leq mT} \sup_{|\mathbf{u}| \leq C} \frac{1}{\sqrt{m} g_m(k)} \left| \sum_{t=m+1}^{m+k} \zeta_t(\mathbf{u}) \right| = o_P(1) = \max_{1 \leq k \leq mT} \sup_{|\mathbf{u}| \leq C} \frac{1}{\sqrt{m} g_m(k)} \frac{k}{m} \left| \sum_{t=1}^{m} \zeta_t(\mathbf{u}) \right|.$$

Proof
Since $|\mathbf{u}^* - \boldsymbol{\theta}| = |\mathbf{u}|/\sqrt{m} \leq C/\sqrt{m} \leq \epsilon$ for any $\epsilon > 0$ if m is large enough, it follows from (11) and part (b) of Proposition 1 that

$$\sup_{|\mathbf{u}| \leq C} |\zeta_t(\mathbf{u})| \leq K \sum_{j=1}^{q} \left(\sum_{\ell=1}^{j} c^{t-1+\ell} \max\{1, |\boldsymbol{\theta}| + \epsilon\} \right) |\varepsilon_{1-j}| \leq K' c^t \sum_{j=1}^{q} |\varepsilon_{1-j}|, \tag{12}$$

where $K' = K \max\{1, |\boldsymbol{\theta}| + \epsilon\}/(1 - c)$. Consequently,

$$E\left[\sup_{k \geq 1} \sup_{|\mathbf{u}| \leq C} \sum_{t=m+1}^{m+k} |\zeta_t(\mathbf{u})| \right] \leq K' \sum_{t=0}^{\infty} c^t \sum_{j=1}^{q} E[|\varepsilon_{1-j}|] = \frac{qK'}{1-c} E[|\varepsilon_0|] < \infty.$$

The first result of the lemma follows because an application of (6) yields that $[\sqrt{m} g_m(k)]^{-1} = o(1)$ uniformly in $k = 1, \ldots, \lfloor mT \rfloor$ as $m \to \infty$. The proof of the second assertion follows along the same lines. □

Lemma 2. Under the assumptions of Theorem 1, it holds, for all $C > 0$ and as $m \to \infty$,

$$\max_{1 \leq k \leq mT} \sup_{|\mathbf{u}|, |\mathbf{v}| \leq C} \frac{1}{\sqrt{m} g_m(k)} \left| \sum_{t=m+1}^{m+k} \frac{\beta_t(\mathbf{u}, \mathbf{v})}{\sqrt{m}} \right| = o_P(1) = \max_{1 \leq k \leq mT} \sup_{|\mathbf{u}|, |\mathbf{v}| \leq C} \frac{1}{\sqrt{m} g_m(k)} \frac{k}{m} \left| \sum_{t=1}^{m} \frac{\beta_t(\mathbf{u}, \mathbf{v})}{\sqrt{m}} \right|.$$

Proof
As in the proof of Lemma 1, note that $|\mathbf{u}^* - \boldsymbol{\theta}|, |\mathbf{v}^* - \boldsymbol{\phi}| \leq C/\sqrt{m} \leq \epsilon$ for any $\epsilon > 0$ if m is large enough. For an arbitrary $j = 1, \ldots, p$, first estimate the term

DOI: 10.1111/jtsa.12062

$$\begin{aligned}
&\max_{1\le k\le mT}\sup_{|\mathbf{u}|\le C}\frac{1}{mg_m(k)}\left|\sum_{t=m+1}^{m+k}\sum_{\ell=0}^{t-1}\psi_\ell(\mathbf{u}^*)X_{t-j-\ell}\right| \\
&\quad\le \max_{1\le k\le mT}\frac{1}{mg_m(k)}\left|\sum_{t=m+1}^{m+k}\sum_{\ell=0}^{\infty}\psi_\ell(\boldsymbol{\theta})X_{t-j-\ell}\right| \\
&\qquad+\max_{1\le k\le mT}\frac{1}{mg_m(k)}\sum_{t=m+1}^{m+k}\sum_{\ell=t}^{\infty}|\psi_\ell(\boldsymbol{\theta})||X_{t-j-\ell}| \\
&\qquad+\max_{1\le k\le mT}\frac{1}{mg_m(k)}\frac{CK}{\sqrt{m}}\sum_{t=m+1}^{m+k}\sum_{\ell=1}^{t-1}\ell c^{\ell-1}|X_{t-j-\ell}| \\
&\quad= I_1+I_2+I_3.
\end{aligned}$$

where part (c) of Proposition 1 has been applied. Using that, by stationarity, $\sum_{t=m+1}^{m+k}|X_{t-j-1}|\stackrel{\mathcal{D}}{=}\sum_{t=1}^{k}|X_t|=\mathcal{O}(k)$ almost surely, one can estimate the third term as

$$\begin{aligned}
I_3&\le\max_{1\le k\le mT}\frac{1}{mg_m(k)}\frac{CK}{\sqrt{m}}\sum_{\ell=1}^{m+\lfloor mT\rfloor}\ell c^{\ell-1}\sum_{t=m+1}^{m+k}|X_{t-j-\ell}| \\
&=\mathcal{O}_P(1)\max_{1\le k\le mT}\frac{k}{m}\frac{1}{\sqrt{m}}\frac{CK}{g_m(k)}=o_P(1)
\end{aligned}$$

as $m\to\infty$. Note that

$$E\left[\sum_{t=1}^{\infty}\sum_{\ell=t}^{\infty}|\psi_\ell(\boldsymbol{\theta})||X_{t-j-\ell}|\right]\le KE[|X_0|]\sum_{t=1}^{\infty}\sum_{\ell=t}^{\infty}c^\ell<\infty$$

and therefore

$$I_2=\mathcal{O}_P(1)\max_{1\le k\le mT}\frac{1}{mg_m(k)}=o_P(1)\qquad(m\to\infty).$$

It remains to estimate I_1. Using the defining ARMA equations (1) and the power series expansions established in Section 6.1, it follows that

$$\frac{1}{\theta(B)}X_{t-j}=\sum_{\ell=0}^{\infty}\psi_\ell(\boldsymbol{\theta})X_{t-j-\ell}=\sum_{\ell=0}^{\infty}\pi_\ell(\boldsymbol{\phi})\varepsilon_{t-j-\ell}=\frac{1}{\phi(B)}\varepsilon_{t-j},$$

where B denotes the backshift operator. Consequently,

$$\sum_{t=m+1}^{m+k}\sum_{\ell=0}^{\infty}\psi_\ell(\boldsymbol{\theta})X_{t-j-\ell}=\sum_{t=m+1}^{m+k}\sum_{\ell=0}^{\infty}\pi_\ell(\boldsymbol{\phi})\varepsilon_{t-j-\ell}\stackrel{\mathcal{D}}{=}\sum_{t=1}^{k}\sum_{\ell=0}^{\infty}\pi_\ell(\boldsymbol{\phi})\varepsilon_{t-j-\ell}$$

DOI: 10.1111/jtsa.12062

and an application of Lemma 3.5 in Yu (2007) yields that

$$I_1 = o_P(1) \max_{1\le k\le mT} \frac{k}{mg_m(k)} = o_P(1) \qquad (m\to\infty).$$

To complete the proof of the lemma, notice that similar arguments also show that, for any $j = 1,\dots,q$ and as $m\to\infty$,

$$\max_{1\le k\le mT} \frac{1}{mg_m(k)} \left| \sum_{t=m+1}^{m+k} \sum_{\ell=0}^{t-1} \psi_\ell(\mathbf{u}^*)\varepsilon_{t-j-\ell} \right| = o_P(1)$$

which provides the proof of the first assertion. The second statement can be obtained from the same arguments. □

Lemma 3. Under the assumptions of Theorem 1, it holds, for all $C > 0$ and as $m\to\infty$,

$$\max_{1\le k\le mT} \sup_{|\mathbf{u}|,|\mathbf{v}|,|w|\le C} \frac{1}{\sqrt{m}g_m(k)} \left| \sum_{t=m+1}^{m+k} \frac{\rho_t(\mathbf{u},\mathbf{v},w)}{\sqrt{m}} - \frac{k}{m}\sum_{t=1}^{m} \frac{\rho_t(\mathbf{u},\mathbf{v},w)}{\sqrt{m}} \right| = o_P(1).$$

Proof
Following the approach of the previous lemmas, first note that an application of Proposition 1 yields that, for all $j = 1,\dots,p$ as $m\to\infty$,

$$\begin{aligned}
&\max_{1\le k\le mT} \sup_{|\mathbf{u}|,|\mathbf{v}|,|w|\le C} \frac{1}{mg_m(k)} \left| \sum_{t=m+1}^{m+k} (v_j^* - \phi_j) w \sum_{\ell=0}^{t-1} \psi_\ell(\mathbf{u}^*) \right| \\
&\quad = \mathcal{O}_P(1) \max_{1\le k\le mT} \frac{1}{mg_m(k)} \sum_{t=m+1}^{m+k} \sum_{\ell=0}^{t-1} c^\ell \\
&\quad = o_P(1)
\end{aligned}$$

and, similarly,

$$\max_{1\le k\le mT} \sup_{|\mathbf{u}|,|\mathbf{v}|,|w|\le C} \frac{1}{mg_m(k)} \frac{k}{m} \left| \sum_{t=1}^{m} \left(v_j^* - \phi_j\right) w \sum_{\ell=0}^{t-1} \psi_\ell(\mathbf{u}^*) \right| = o_P(1) \qquad (m\to\infty).$$

Hence, each v_j^* can be replaced with the corresponding ϕ_j without changing the asymptotic. Observe that $1 - \phi_1 - \cdots - \phi_p = \phi(1)$. Furthermore, it holds that

$$\sum_{\ell=0}^{t-1} \psi_\ell(\mathbf{u}^*) = \frac{1}{\theta(1)} - \sum_{\ell=i}^{\infty} \psi_\ell(\boldsymbol{\theta}) + \sum_{\ell=0}^{t-1} [\psi_\ell(\mathbf{u}^*) - \psi_\ell(\boldsymbol{\theta})].$$

Therefore, the estimation steps applied to I_2 and I_3 in Lemma 2 imply that, uniformly in $\mathbf{u}$, $\mathbf{v}$ and w,

$$\max_{1\le k\le mT} \frac{1}{mg_m(k)} \left| \sum_{t=m+1}^{m+k} \rho_t(\mathbf{u},\mathbf{v},w) - \frac{\phi(1)}{\theta(1)} kw \right| = o_P(1) \qquad (m\to\infty)$$

DOI: 10.1111/jtsa.12062

and

$$\max_{1\le k\le mT}\frac{1}{mg_m(k)}\frac{k}{m}\left|\sum_{t=1}^{m}\rho_t(\mathbf{u},\mathbf{v},w)-\frac{\phi(1)}{\theta(1)}mw\right|=o_P(1)\qquad(m\to\infty).$$

This completes the proof. □

Proof of Theorem 1

It follows from Lemmas 1–3 that it suffices to show that the theorem holds if the residuals $\hat{\varepsilon}_t$ are replaced by the corresponding innovations ε_t. Utilizing the partial sum notation $S_k(n)=\varepsilon_{k+1}+\cdots+\varepsilon_n$, an application of the functional central limit theorem implies that, as $m\to\infty$,

$$\left(\frac{1}{\sqrt{m}}\left[S_m(m+\lfloor mx\rfloor)-\frac{\lfloor mx\rfloor}{m}S_0(m)\right]:x\in[0,T]\right)\Rightarrow\left(\sigma\left[W(1+x)-(1+x)W(1)\right]:x\in[0,T]\right),$$

where $\Rightarrow$ denotes weak convergence in the Skorohod space $\mathcal{D}[0,T]$ and $(W(x):x\in[0,T])$ a standard Brownian motion. It is clear that the limit process is Gaussian and it can be checked that its covariance function is the same as that of $(V(x):x\in[0,1])$. Noticing the continuity of g, (6) and that $\hat{\sigma}_m^2$ is a weakly consistent estimator for σ^2, the result will follow from an application of the continuous mapping theorem. □

Proof of Theorem 2

Since there is only a change in the mean parameter but not in the parameters driving the underlying ARMA sequence, it follows from Lemmas 1 and 2 that it suffices to determine the behaviour of the term containing $\rho_t(\mathbf{u},\mathbf{v},w)$. For $k>k^*$, write then

$$\begin{aligned}&\sum_{t=m+1}^{m+k}\rho_t(\mathbf{u},\mathbf{v},w)-\frac{k}{m}\sum_{t=1}^{m}\rho_t(\mathbf{u},\mathbf{v},w)\\&=\left(\sum_{t=m+1}^{m+k^*}\rho_t(\mathbf{u},\mathbf{v},w)-\frac{k^*}{m}\sum_{t=1}^{m}\rho_t(\mathbf{u},\mathbf{v},w)\right)+\left(\sum_{t=m+k^*1}^{m+k}\rho_t(\mathbf{u},\mathbf{v},w)-\frac{k-k^*}{m}\sum_{t=1}^{m}\rho_t(\mathbf{u},\mathbf{v},w)\right).\end{aligned}$$

The first term deals with those ρ_t having mean μ_0 and the second term with those having mean $\mu_A=\mu_0+\Delta$ for some $\Delta\neq 0$. Utilizing arguments similar to those applied in the proof of Lemma 3, it follows from Theorem 1 and Lemmas 1 and 2 that, for any $N>k^*$,

$$\sup_{1\le k\le N}\frac{\left|\hat{D}_m^{\mu}(k)\right|}{g_m(k)}=\mathcal{O}_P(1)+\mathcal{O}_P(1)\sup_{k^*<k\le N}\frac{(k-k^*)|\Delta|}{g_m(k)},\tag{13}$$

where both $\mathcal{O}_P(1)$ rates are independent of N and sharp (they cannot be replaced with $o_P(1)$ rates). Letting $N\to\infty$, this implies asymptotic consistency. □

6.3. Proof of Theorems 3 and 4

The corresponding general parameter case theorems are proved, which are based on the squared residuals $\hat{\varepsilon}_t^2$. These are shown to be close to the squared innovations ε_t^2. Utilizing $\Delta_t=\Delta_t(\mathbf{u},\mathbf{v},w)$ introduced in the previous subsection, it follows that

$$\hat{\varepsilon}_t^2-\varepsilon_t^2=2\varepsilon_t\Delta_t+\Delta_t^2.$$

DOI: 10.1111/jtsa.12062

It will be shown in Lemmas 4 and 5 that the terms on the right-hand side do not contribute asymptotically to the limit distribution of the monitoring procedure.

Lemma 4. Under the assumptions of Theorem 3, it holds, for all $C > 0$ and as $m \to \infty$,

$$\max_{1\le k\le mT} \sup_{|\mathbf{u}|,|\mathbf{v}|,|w|\le C} \frac{1}{\sqrt{m}g_m(k)} \left| \sum_{t=m+1}^{m+k} \varepsilon_t \Delta_t(\mathbf{u},\mathbf{v},w) \right| = o_P(1)$$

and

$$\max_{1\le k\le mT} \sup_{|\mathbf{u}|,|\mathbf{v}|,|w|\le C} \frac{1}{\sqrt{m}g_m(k)} \frac{k}{m} \left| \sum_{t=1}^{m} \varepsilon_t \Delta_t(\mathbf{u},\mathbf{v},w) \right| = o_P(1).$$

Proof
Note that ε_t is independent of Δ_t and the claim follows thus from Lemmas 1–3. □

Lemma 5. Under the assumptions of Theorem 3, it holds, for all $C > 0$ and as $m \to \infty$,

$$\max_{1\le k\le mT} \sup_{|\mathbf{u}|,|\mathbf{v}|,|w|\le C} \frac{1}{\sqrt{m}g_m(k)} \sum_{t=m+1}^{m+k} \Delta_t^2(\mathbf{u},\mathbf{v},w) = o_P(1)$$

and

$$\max_{1\le k\le mT} \sup_{|\mathbf{u}|,|\mathbf{v}|,|w|\le C} \frac{1}{\sqrt{m}g_m(k)} \frac{k}{m} \sum_{t=1}^{m} \Delta_t^2(\mathbf{u},\mathbf{v},w) = o_P(1).$$

Proof
Note that

$$\Delta_t^2(\mathbf{u},\mathbf{v},w) \le 3\left[\zeta_t^2(\mathbf{u}) + \frac{\beta_t^2(\mathbf{u},\mathbf{v})}{m} + \frac{\rho_t^2(\mathbf{u},\mathbf{v},w)}{m}\right],$$

so that it suffices to prove the corresponding results for the quantities on the right-hand side. To this end, note first that (12) implies

$$E\left[\sup_{k\ge 1}\sup_{|\mathbf{u}|\le C} \sum_{t=m+1}^{m+k} \zeta_t^2(\mathbf{u})\right] \le (K')^2 E\left[\sum_{t=1}^{\infty} c^{2t}(|\varepsilon_0| + \cdots + |\varepsilon_{-q+1}|)^2\right] = \frac{q(K')^2}{1-c^2} E\left[\varepsilon_0^2\right] < \infty.$$

Therefore, as $m \to \infty$,

$$\max_{1\le k\le mT} \sup_{|\mathbf{u}|\le C} \frac{1}{\sqrt{m}g_m(k)} \sum_{t=m+1}^{m+k} \zeta_t^2(\mathbf{u}) = o_P(1).$$

Next, observe that Lemma 3.4 of Yu (2007) states that, for all $k \ge 1$,

$$\sup_{|\mathbf{u}|,|\mathbf{v}|\le C} \sum_{t=m+1}^{m+k} \beta_t^2(\mathbf{u},\mathbf{v}) = \mathcal{O}_P(m+k) \qquad (m \to \infty).$$

DOI: 10.1111/jtsa.12062

Consequently, as $m \to \infty$,

$$\max_{1\le k\le mT} \sup_{|\mathbf{u}|,|\mathbf{v}|\le C} \frac{1}{\sqrt{m}g_m(k)} \sum_{t=m+1}^{m+k} \frac{\beta_t^2(\mathbf{u},\mathbf{v})}{m} = \mathcal{O}_P(1) \max_{1\le k\le mT} \frac{m+k}{m} \frac{1}{\sqrt{m}g_m(k)} = o_P(1).$$

In a similar fashion, Lemma 3.7 of Yu (2007) implies that, for all $k \ge 1$,

$$\sup_{|\mathbf{u}|,|\mathbf{v}|,|w|\le C} \sum_{t=m+1}^{m+k} \rho_t^2(\mathbf{u},\mathbf{v},w) = \mathcal{O}_P(m+k) \qquad (m \to \infty).$$

Thus, as $m \to \infty$,

$$\max_{1\le k\le mT} \sup_{|\mathbf{u}|,|\mathbf{v}|,|w|\le C} \frac{1}{\sqrt{m}g_m(k)} \sum_{t=m+1}^{m+k} \frac{\rho_t^2(\mathbf{u},\mathbf{v})}{m} = \mathcal{O}_P(1) \max_{1\le k\le mT} \frac{m+k}{m} \frac{1}{\sqrt{m}g(k/m)} = o_P(1).$$

This completes the proof of the first statement; the second assertion of the lemma follows along similar lines. □

Proof of Theorem 3.

Given Lemmas 4 and 5, one can proceed in a similar manner to the proof of Theorem 1. Write $S_k^2(n) = \varepsilon_{k+1}^2 + \cdots + \varepsilon_n^2$. Since $(\varepsilon_t : t \in \mathbb{Z})$ is assumed to possess finite fourth-order moments, the functional central limit theorem applied to the centred variables $\left(\varepsilon_t^2 - \sigma^2 : t \in \mathbb{Z}\right)$ yields that

$$\left(\frac{1}{\sqrt{m}} \left[S_m^2(m + \lfloor mx \rfloor) - \frac{\lfloor mx \rfloor}{m} S_0^2(m) \right] : x \in [0,T] \right) \Rightarrow (\eta\,[W(1+x) - (1+x)W(1)] : x \in [0,T])$$

weakly in $\mathcal{D}[0,T]$ as $m \to \infty$. Now, the same arguments as in Theorem 1 complete the proof. □

The proof of Theorem 4 is similar to the proof of Theorem 2 and is hence omitted.

ACKNOWLEDGEMENT

This research was partially supported by NSF grant DMS 0905400.

REFERENCES

Andreou E, Ghysels E. 2006. Monitoring disruptions in financial markets. *Journal of Econometrics* **135**: 77–124.

Aue A, Horváth L. 2013. Structural breaks in time series. *Journal of Time Series Analysis* **34**: 1–13.

Aue A, Horváth L, Hušková M, Kokoszka P. 2006. Change-point monitoring in linear models. *Econometrics Journal* **9**: 373–403.

Aue A, Horváth L, Reimherr ML. 2009. Delay times of sequential procedures for multiple time series regression models. *Journal of Econometrics* **149**: 174–190.

Bai J. 1993. On partial sums of residuals in autoregressive and moving average models. *Journal of Time Series Analysis* **14**: 247–260.

Berkes I, Gombay E, Horváth L, Kokoszka P. 2004. Sequential change-point detection in GARCH(p,q) models. *Econometric Theory* **20**: 1140–1167.

Box GEP, Cox DR. 1968. An analysis of transformations. *Journal of the Royal Statistical Society, Series B* **26**: 211–252.

Brockwell PJ, Davis RA. 1991. *Time Series: Theory and Methods*, (2nd ed.) New York: Springer-Verlag.

Charron A, Harrison RM. 2003. Primary particle formation from vehicle emissions during exhaust dilution in the roadside atmosphere. *Atmospheric Environment* **37**: 4109–4119.

Chu C-S, Stinchcombe J, White H. 1996. Monitoring structural change. *Econometrica* **64**: 1045–1065.

Davis RA, Huang DW, Yao Y-C. 1995. Testing for a change in the parameter values and order of an autoregressive model. *The Annals of Statistics* **23**: 282–304.

DOI: 10.1111/jtsa.12062

EU Life project CMA+. http://www.life-cma.at, Accessed on 15 May 2012.
EU Life project CMA+ Newsletter. http://www.life-cma.at/english/newsletter-147_426.asp, Accessed on 15 May 2012.
Gombay E, Serban D. 2009. Monitoring parameter change in AR(p) time series models. *Journal of Multivariate Analysis* **100**: 715–725.
Janhäll S, Olofson FG, Andersson PU, Pettersson JBC, Hallquist M. 2006. Evolution of the urban aerosol during winter temperature inversion episodes. *Atmospheric Environment* **40**: 5355–5366.
Ploberger W, Krämer W. 1982. The CUSUM test with OLS residuals. *Econometrica* **60**: 271–285.
Oberdörster G, Utell MJ. 2002. Ultrafine particles in the urban air: to the respiratory tract—and beyond? *Environmental Health Perspectives* **110**: A440–A441.
Ostro BD, Hurley S, Lipsett MJ. 1999. Air pollution and daily mortality in the Coachella Valley, California: a study of PM_{10} dominated by coarse particles. *Environmental Research* **81**: 131–238.
Pekkanen J, Timonen KL, Ruuskanen J, Reponen A, Mirme A. 1997. Effects of ultrafine and fine particles in urban air on peak expiratory flow among children with asthmatic symptoms. *Environmental Research* **74**: 24–33.
Pope CA, III, Dockery DW, Spengler JD, Raizenne ME. 1991. Respiratory health and PM_{10} pollution: a daily time series analysis. *American Journal of Respiratory and Critical Care Medicine* **144**: 668–674.
Robbins M, Gallagher C, Lund R, Aue A. 2011. Mean shift testing in correlated data. *Journal of Time Series Analysis* **32**: 498–511.
Shumway RH, Azari RS, Pawitan Y. 1988. Modeling mortality fluctuations in Los Angeles as functions of pollution and weather effects. *Environmental Research* **45**: 224–241.
Shumway RH, Stoffer DS. 2010. *Time Series and Its Applications*, (3rd ed.) New York: Springer-Verlag.
Stadlober E, Hörmann S, Pfeiler B. 2008. Quality and performance of a PM10 daily forecasting model. *Atmospheric Environment* **42**: 1098–1109.
Yu H. 2007. High moment partial sum processes of residuals in ARMA models and their applications. *Journal of Time Series Analysis* **28**: 72–91.

DOI: 10.1111/jtsa.12062