SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2011-29

Wavelet Based Outlier Correction for Power Controlled Turning Point Detection in Surveillance Systems

Yushu Li

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Wavelet Based Outlier Correction for Power Controlled Turning Point Detection in Surveillance Systems

Yushu Li*

Centre for Labour Market Policy Research (CAFO), Department of Economic and Statistics, Linnaeus University, Sweden

Abstract

Detection turning points in unimodel has various applications to time series which have cyclic periods. Related techniques are widely explored in the field of statistical surveillance, that is, on-line turning point detection procedures. This paper will first present a power controlled turning point detection method based on the theory of the likelihood ratio test in statistical surveillance. Next we show how outliers will influence the performance of this methodology. Due to the sensitivity of the surveillance system to outliers, we finally present a wavelet multiresolution (MRA) based outlier elimination approach, which can be combined with the on-line turning point detection process and will then alleviate the false alarm problem introduced by the outliers.

Keywords: Unimodel, Turning point, Statistical surveillance, Outlier, Wavelet multi-resolution, Threshold.

JEL classification: C12; C15; C22.

* The author wishes to thank CREATES, Center for Research in Econometric Analysis of Time Series funded by the Danish National Research Foundation.

1. Introduction

Time series which show periodic character are often used to model cyclical behavior in various fields such as the expansion and recession of business cycles in economics, the tides in oceanography, or the change of the brightness of a star in astronomy. When dealing with this type of time series, detection the turning points of each cycle in the on-going process in a timely and precise fashion will be advantageous for future strategic decisions. Especially when we already have a related leading indicator which shows similar but advanced periodical dynamics to the index of interest, prompt and accurate detection of turning points in the leading indicator will give valuable signals for the prediction of the series of interest. Such as if an alarm is given out as soon as the leading indicator signal shows a structural change, we can know in advance that in our series of interest there will appear a similar change and we can be more prepared for it. Related research is being explored in the theory of statistical surveillance, which aims to give out alarms as soon as the data information accumulates evidence to a level sufficient to prove the occurrence of a changing point. Based on different ways of defining turning points and measurements of the data, there exist various methodologies to build test statistics which will give out alarms when they exceed certain threshold values. Most test statistics are built on the theories of the likelihood ratio, posterior distributions, or hidden Markov chains. The comparison of different methods is exhaustively examined in Andersson *et al.* (2005). This paper will mainly consider a turning point detection methodology utilizing the likelihood ratio method: SRlin method which is derived by Shiryaev-Roberts (SR) technique. The test will be constructed in a way which can control the power of the alarm system, that is to control its ability of giving out the alarm in time when a turning point actually occurs. This test statistic has the advantages of easy application and straightforward interpretation, and it is also flexible to the demand of the appliers based on their own criteria of power control.

Although the surveillance system performs well under the restrictions of parametric model and *i.i.d.* normal error, the non-stationarities of the time series will always affect the property of the alarm test statistic. The consequences of various non-stationarities such as seasonality and trend behavior are carefully examined in Andersson *et al.* (2006). However, the problem of outliers has not yet aroused enough attention although it is rational to assume that outliers may influence the testing result significantly, as an outlier is quite easy to be misunderstood as some kinds of turning point for an on-line testing procedure. The influence of the outlier is proved in our simulation in a form of a much higher rate of false early alarms. Thus it is important to find a solution to eliminate this effect in the turning point detection procedure. Moreover, an outlier detection methodology in statistical surveillance calls for a higher demand of technique than the normal outlier elimination methodology as we need to combine the outlier elimination on-line with turning point detection. Therefore we need a technique which can detect the outlier on-line as well as give out alarms for the real turning point as soon as possible. In this paper we introduce a methodology based on wavelet multi-resolution analysis (MRA), which can reach this goal easily and efficiently. This methodology uses the wavelet decomposed series, and the Monte Carlo experiment will show how the outlier influence is reduced after employing the wavelet method to handle the polluted data.

This paper mainly deals with three topics: the construction of the power controlled test statistic in the surveillance system, the influence of the outliers, and a wavelet methodology to eliminate the negative effect of the outliers. According to the three topics, the rest of this paper will be organized as follows: Section 2 will introduce the underlying model for the turning point detection systems, the test statistic and an evaluation criterion. Section 3 will illustrate how the outlier will influence the whole detection procedure and Section 4 will show how the wavelet approach can eliminate the influence of outliers. The conclusion will be in the last section.

2. Turning point detection system based on likelihood ratio method

2.1 The underlying unimodel and the event to be detected

The time series $X=\{X_t,\ t=1,2,...\}$ which the statistical surveillance will monitor is the leading indicator of the actual series of interest, such as the unemployment insurance claims, house start, or stock prices can be viewed as the leading indicators for the Federal to make strategic to decide the next period's interest rates. When the concerned series and its leading indicator have similar periodic dynamics, detection of the turning time of the leading indicator will help to predict the turning point of the series of interest. In the surveillance system, both the series of interest and the leading indicator show cyclic behaviors, and each cycle will show the same upward trend as well as the downward trend but with distinct turning point times. We suppose in each cycle the indicator series has the stochastic dynamics $X_t = \mu_t + \varepsilon_t$, where $\varepsilon_t \sim n.i.d.(0,\sigma^2)$ and the underlying process μ_t has a unimodel structure, which means μ_t is either convex or concave. Here we assume that the unimodel is convex containing a peak, and this assumption is quite practical as it is always more important to detect the turning point from an expansion to recession and give out a timely alarm so that we can avoid the influence brought up by the economic recession as early as possible. Most often it is more difficult for the agents to notice a recession when the economy is in an uptrend although the impact of the recession will mostly be negative. Then based on the observation x_t, at each decision time s, we need to tell whether μ_s belongs to the upward trend $D(s)$ or the downward trend $C(s)$, where:

$$\begin{aligned} &D(s):\ \mu_1 \le \ldots \le \mu_s \\ &C(s):\ \mu_1 \le \ldots \le \mu_{\tau-1}\ and\ \mu_{\tau-1} \ge \ldots \ge \mu_s \end{aligned}, \qquad \text{(I)}$$

with τ being the unknown peaking time in the unimodel. Statistical surveillance is an on-line detection process in which we need to make repeated decisions each time we have a new observation x_s. On the other hand, as statistical surveillance deals with the periodic time series, the structure of the model in a unit cycle can always be estimated based on the observations from last cycles. Thus this paper assumes the unimodel for μ_t is known and a linear model is further chosen for simplicity. Then $D(s)$ and $C(s)$ have the following structure:

$$\begin{aligned} &D(s):\ \mu_s = \beta_0 + \beta_1 s \\ &C(s):\ \{\cup C(\tau)\} \end{aligned}, \qquad \text{(II)}$$

where $C(\tau): \mu_s = \beta_0 + \beta_1(\tau-1) - \delta_1(s-\tau+1)$, $\tau = \{1, 2, ..., s\}$; β_0, β_1 and δ_1 can be estimated from the historical data. The rest of the paper will adopt this parametric linear assumption as it is straightforward enough to illustrate the above mentioned three topics. We will see in the later part of the paper, as the likelihood ratio test is quite robust to the underlying model structure, the result from this linear symmetric model can be easily extended to other parametric models or even nonparametric cases, and for related research, the reader can be referred to Frisén (1994) and Andersson *et al.* (2006).

2.2 Alarm statistics

A surveillance system is constructed on two main elements: test statistic and alarm limit. The test statistic in this paper is based on the full likelihood ratio method, and the system will give out alarm when the value of the likelihood ratio based test statistic exceeds certain alarm limit. At each decision time s, let X_s denote the filtration generated by X till time s, and x_τ^s denote the information generated by X from time τ to s. Then the likelihoods for the two events $C=\{\tau \le s\}=\{\cup\tau=i,\ i=1,2,...,s\}=\{\cup C_i\}$ and $D=\{\tau>s\}$ correspond to $L(C|\mathrm{X}_s)$ and $L(D|\mathrm{X}_s)$, and the likelihood ratio based test surveillance system will give out an alarm as soon as:

$$LR(s)=\frac{L(C|\mathrm{X}_s)}{L(D|\mathrm{X}_s)}=\frac{f(\mathrm{X}_s|C)}{f(\mathrm{X}_s|D)}=\sum_{i=1}^{s}\frac{P(\tau=i)}{P(\tau\le s)}.\frac{f(x_\tau^s|\mu=\mu^{C_i})}{f(x_\tau^s|\mu=\mu^D)}=\sum_{i=1}^{s}w_i\frac{f(x_\tau^s|\mu=\mu^{C_i})}{f(x_\tau^s|\mu=\mu^D)}\ge k_{alarm},$$

where $w_i=\frac{P(\tau=i)}{P(\tau\le s)}$ and the alarm time t_A is then $t_A=\min[t:LR(t)\ge k_{alarm}]$ where $k_{alarm}=\frac{k}{1-k}\cdot\frac{P(D)}{P(C)}$ with k being a positive constant which is chosen to satisfy certain evaluation criteria. The expression of k_{alarm} is actually deduced in a way which lets the likelihood ratio based method be equivalent to a posterior probability based method where the alarm rule is $P(C|\mathrm{X}_s)>k$ under the situation $P(D)=1-P(C)$, and the proof is as follows:

$$\begin{aligned}P(C|\mathrm{X}_s)>k &\Rightarrow \frac{f(\mathrm{X}_s|C)P(C)}{f(\mathrm{X}_s|C)P(C)+f(\mathrm{X}_s|D)P(D)}>k\Rightarrow\frac{f(\mathrm{X}_s|C)P(C)+f(\mathrm{X}_s|D)P(D)}{f(\mathrm{X}_s|C)P(C)}<\frac{1}{k}\\ &\Rightarrow \frac{f(\mathrm{X}_s|C)}{f(\mathrm{X}_s|D)}>\frac{k.P(D)}{(1-k).P(C)}\end{aligned}.$$

It is obvious that in the determinations of both w_i and k_{alarm} we need to know the distribution of the turning point time τ. When no reliable distribution is available, Shiryaev (1963) and Roberts (1966) proposed a method which assumes a non-informative prior distribution for τ, and let $P(\tau=t)$ be equal for all t. Therefore, the resulting alarm statistic has equal weights and the test statistic is: $SR(s)=\frac{L(C|\mathrm{X}_s)}{L(D|\mathrm{X}_s)}=\frac{f(\mathrm{X}_s|C)}{f(\mathrm{X}_s|D)}=\sum_{i=1}^{s}\frac{f(x_\tau^s|\mu=\mu^{C_i})}{f(x_\tau^s|\mu=\mu^D)}$. For the linear specified model in (II), the test statistic becomes:

$$SRlin(s)=\sum_{i=1}^{s}\frac{\exp\left\{-\frac{1}{2\sigma^2}\sum_{u=i}^{s}(x_u-(\beta_0+\beta_1(i-1)-\delta_1(u-i+1)))^2\right\}}{\exp\left\{-\frac{1}{2\sigma^2}\sum_{u=i}^{s}(x_u-(\beta_0+\beta_1 u))^2\right\}}.$$

We also assume the unimodel is symmetric with $\delta_1=\beta_1$ and then test statistic turns out to be:

$$SRlin(s)=\sum_{i=1}^{s}\exp\left[\left(\frac{1}{2\sigma^2}\right)\left(4\beta_1\sum_{u=i}^{s}(x_u(i-1-u))+w_i\right)\right],$$

where $w_i = (4\beta_1^2(i-1) + 4\beta_0\beta_1)\sum_{u=i}^{s}(u-i+1)$.

In an on-line surveillance detection system, an alarm is given as soon as $SRlin(s)$ exceeds the limit $k_{alarm} = \frac{k}{1-k}\cdot\frac{P(D)}{P(C)}$. Although the choice of k_{alarm} depends on the distribution of τ, when applying the algorithm proposed by Shiryaev (1963) and Roberts (1966), mentioned above, k_{alarm} turns out to be a constant and it can be determined by simulations based on certain size-controlled or power-controlled criteria. In the next section, we will propose a criterion which can control the power of the test and decide k_{alarm} by Monte Carlo simulations, with the power corresponding to the ability of the system to give out an alarm as soon as the turning point appears.

2.3 Alarm limits and related criteria to evaluate the performance of system

Without knowing the distribution of τ, the alarm limit k_{alarm} can be determined by fixing a certain criterion for evaluating the performance of the alarm statistics. In the statistical hypothesis testing framework with null hypothesis $\mathrm{H_0}$ and alternative hypothesis $\mathrm{H_A}$, there exist two types of evaluation indexes: type I error with its corresponding probability $\alpha = P(\text{reject } \mathrm{H_0} \mid \mathrm{H_0} \text{ true})$ and the probability of type II error $\beta = P(\text{do not reject } \mathrm{H_0} \mid \mathrm{H_A} \text{ true})$. The evaluation procedure can be carried out in two ways: the most often used is to fix the size and then to compare $\text{power} = 1\text{-}\beta = P(\text{reject } \mathrm{H_0} \mid \mathrm{H_A} \text{ true})$, the other is to set the power and then compare the type I errors. In a statistical surveillance system, $\mathrm{H_0}$ is interpreted as that there is no turning point till the current time and X_s belongs to phase D while $\mathrm{H_A}$ asserts that a turning point already occurred and X_s belongs to phase C. Therefore in the surveillance system, the size is actually related with false alarms when no turning point occurs and power will correspond to the alarm delay after a turning point has already appeared. Then k_{alarm} can be chosen either by fixing the size or by controlling the power. As long as k_{alarm} is determined, the detection system can be evaluated by comparing the other type of index which corresponds to power or size. In Gan (1993) and Andersson (2002), k_{alarm} is chosen from simulation by controlling the median run length (MRL) until a false alarm, under the assumption that no turning point actually occurs and it can be expressed as: MRL $= \text{Media}\left[t_A \middle| D \right] = \text{Media}\left[t_A \middle| \tau = \infty \right]$. This way of determining k_{alarm} is a size-fixed method as it assumes no turning point has occurred during the whole surveillance period. In the following sections we will investigate the influence of the outlier on false early alarms, which is to compare the sizes before and after the outlier occurs, thus we need to choose k_{alarm} in a power-fixed method. Here the power is defined as the probability that the alarm will ring with only a one step delay after the turning point actually occurs. Suppose the whole series has T observations, this power criterion is:

$$\begin{aligned} \text{Power} &= P(\text{reject } H_0 \,|\, H_A \text{ True}) = P(\text{Alarm rings at } s+1 \text{ if } \tau = s) = P(LR(\tau+1) > k_{alarm}) \\ &= P(s=1) \cdot P(LR(s+1) > k_{alarm} \,|\, \tau = s) + \ldots + P(s = T\text{-}1) \cdot P(LR(s+1) > k_{alarm} \,|\, \tau = T\text{-}1) \end{aligned}.$$

Monte Carlo simulation shows that in the likelihood ratio based approach, as long as the underlying parametric model is fixed, k_{alarm} will be stable regardless of the actual turning point time. Thus we only need to set a T which can give a stable k_{alarm}. Furthermore, k_{alarm} does not need to be an exact value but in certain digit level as $LR(s+1)$ is much bigger than $LR(s)$ if $\tau = s$, and this is also the reason that the surveillance system performs well as it is quite easy to distinguish $LR(s+1)$ from $LR(s)$ when the turning point actually occurs.

2.4 Simulation result from Monte Carlo experiments

Monte Carlo experiments are applied to decide k_{alarm} and later to review the performance of the power controlled surveillance system. The parameters for the underlying linear model are set as $\beta_0 = \beta_1 = 1$. To decide k_{alarm}, we set $\sigma = 1$, power = 0.8 and T = 30 which is large enough for the stability of k_{alarm}. Based on a fixed power and its corresponding alarm limit k_{alarm}, we can evaluate the size of this test system, which is the probability that the system will give out a false alarm before any turning point occurs, and it can be measured by the average length and rate of the false early alarm before the actual turning point. In applications, the power can be chosen to be the most suitable to control the system, with the trade off that a higher power will end up with a higher size. Here we choose the power to be equal to 0.8 as this corresponds to a low size and the system can then avoid a high early false alarm rate. On the other hand, we can still investigate the property of the power for the alarm system by way of evaluating the actual delay length and rate, as k_{alarm} is chosen by just fixing the delay at a length equal to one. Monte Carlo simulation is applied to assess the power-fixed test system.

In the Monte Carlo experiment, we simulate three different series which follow the dynamics in the linear model (II) with the actual turning points time T_{turn} set to be 5, 30, and 50. It is also interesting to investigate how the volatility will influence the test system, thus we set three variance levels where $\sigma \sim U[0.5,1.5]$, $\sigma \sim U[1.5,2.5]$ and $\sigma \sim U[2.5,3.5]$. Based on the experimental design, and for each case the number of replications is 1000, we get the following simulation result:

Table 2.4.1: Property of the power controlled surveillance system

T_{turn}	*False early alarm*		*Alarm delay*	
	length	rate	length	rate
$\sigma \sim U[0.5,1.5]$				
5	1.000	0.001	1.014	0.780
30	14.000	0.003	1.421	0.982
50	20.714	0.007	1.689	0.984
$\sigma \sim U[1.5,2.5]$				
5	1.000	0.001	2.698	0.999
30	15.667	0.006	2.358	0.991
50	27.800	0.005	2.292	0.992
$\sigma \sim U[2.5,3.5]$				
5	1.000	0.002	3.681	0.997
30	12.167	0.006	3.678	0.992
50	21.500	0.010	3.553	0.989

In Table 2.4.1, the index's length and rate correspond to the average length and occurrence frequency for a false early alarm or an alarm delay in 1000 replications. Thus Table 2.4.1 shows that under the restrictions of *i.i.d.* normally error and linear symmetric model, the alarm system is alert as well as accurate with a short delay and quite low false alarm percentage: lower than 1%. The change of variance does not have much influence on the size property while in the power perspective, higher variance will bring about a longer delay, and this is due to the higher variance's confounding the likelihood value and the system may wait till $LR(s+i)$ is large enough to trigger the alarm. However, compared to the false alarm length, the alarm delay is quite short: lower than 4 for all cases, and for $\sigma \sim U[0.5,1.5]$, the system is quite alert with very short delay length: less than 2. This good behavior of the surveillance system is due to the strict restrictions that the parametric model is already known and the error is *i.i.d.* normal. Loosening the restrictions will always degrade the performance of the alarm system and lots of efforts have been devoted to resolve the problems aroused by less restricted data. Among them, Frisén (1994) discussed the consequence of unsuitably specified parametric models and introduce a nonparametric method. Andersson *et al.* (2006) had a wide exploration of the influences brought up by autocorrelation errors, seasonal effects, and a long time trend. However, after closer scrutiny of this turning point detection procedure, we see that if an outlier that has the same turning direction appears before the actual change appears, the system may give out a false early alarm as it will misidentify this outlier as a turning trend. For this reason, it is crucial to detect and correct the outlier when carrying out the surveillance process. The next sections will discuss the problems brought in by the outliers as well as how to improve the test when the data is polluted by outliers.

3. The influence of outlier

3.1 A brief description of outliers and outlier detection methods

The property of quick and accurate detection of turning point by the surveillance system in Section 2 is built on the assumptions of *i.i.d.* normal error and the parametric regressions among variables. In practical applications the data is easily subjected to malfunctions of the data collection mechanism, calculation errors or unexpected and extreme events. One of the resulting consequences is the appearance of outliers, which are defined in Barnett and Lewis (1994) as observations with abnormal deviations from the mean of the remainder of the data set. Outliers are generally divided into two types: additive outlier (AO) and innovative outlier (IO). The first one affects the level of the current observations and always takes the form of an isolated value or data spike at the time when the disturbance occurs; the second one is some kind of inherent contamination which influences future observations through the same dynamics as the core process. As an IO is propagated by the autocorrelation of the series, we do not consider it here since the data are assumed to be uncorrelated. The data structure with additive level outliers (AO) which we will further investigate is then defined as $y(t) = \mu(t) + \omega \cdot I(t) + \varepsilon(t) = x(t) + \varepsilon(t)$, where ω is the magnitude of the disturbance and $I(t)$ is an index function which is 1 at the outlier appearance time and 0 otherwise.

3.2 The influence of outlier on turning point detection process

The inconsistency of the outlier with the rest of the data set will always confound the data analysis from the model identification to the parameter estimation, statistical inference and even further: to the forecasting. In the procedure of turning point detection, the outlier will confuse the data in a way that it may be misidentified as a turning point. In an on-line peak detection procedure, if a down biased outlier appears, the detection mechanism may misidentify it as a downwards turning point and give out an early false alarm although the main trend of the series is still upwards. This early detection problem caused by the outliers can be demonstrated by simulation when we add an additive outlier to the unimodel. The following simulations add one outlier at a random time before the turning point occurs with 3 levels of magnitudes of the outliers. Based on the significance of the deviation from the underlying model, ω is set to be $\omega = -1.5\sigma, -3\sigma$, and -5σ. The influence of the outlier will be illustrated clearly by using one turning point case with T_{turn} =30 and $\sigma \sim U[0.5, 1.5]$. The total length for the whole series is set to be 60 and Figure 1 shows how the series will look when it is polluted by different magnitudes of the outliers. One outlier is added at to each of three cases where the outliner occurrence times are randomly setting as 14, 11, and 24:

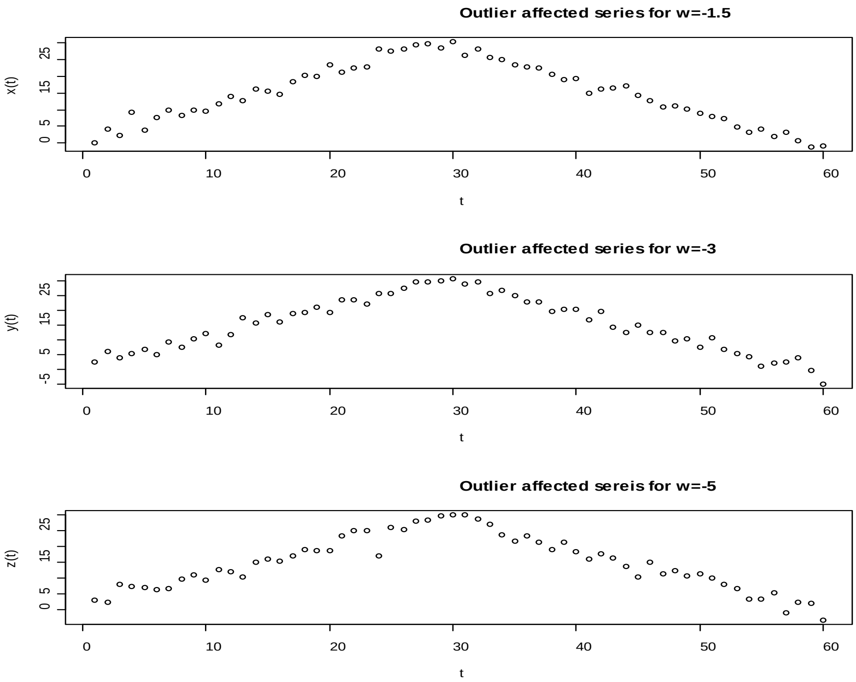

Figure 3.2.1: Outlier polluted unimodel

Figure 3.2.1 shows that for $\omega = -5\sigma$, the outlier is easily detected visually. However, in on-line turning point detection, this outlier will most likely to be misidentified as the turning point. When $\omega = -1.5\sigma$ and -3σ, it is not easy to detect the outlier from the figure but it does not mean that this will not influence the turning point detection. With the data sets which include one outlier at random time in each series, we examine the system again by 1000 Monte Carlo replications and get a new table as follows:

Table 3.2.1: Property of the power-controlled surveillance system under outlier influence

ω	*False early alarm*		*Alarm delay*	
	length	rate	length	rate
$\sigma \sim U[0.5,1.5]$				
-1.5σ	15.181	0.011	1.352	0.952
-3σ	14.384	0.250	1.002	0.362
-5σ	14.739	0.813	1.242	0.157
$\sigma \sim U[1.5,2.5]$				
-1.5σ	14.285	0.007	2.220	0.992
-3σ	14.613	0.176	1.862	0.801
-5σ	14.753	0.657	2.745	0.303
$\sigma \sim U[2.5,3.5]$				
-1.5σ	14.281	0.016	2.992	0.963
-3σ	13.116	0.103	3.454	0.886
-5σ	13.959	0.596	3.450	0.380

Table 3.2.1 shows that for when $\omega = -1.5\sigma$, the system is almost not influenced with a still very low false alarm rate. When $\omega = -3\sigma$, the problem of the outliers begin to appear with an obviously higher false alarm rate and when $\omega = -5\sigma$, the alarm rate rises significantly to even around 80% when $\sigma \sim U[0.5,1.5]$. Thus for outliers of small magnitude, the system is still robust but with larger magnitude outliers the system will be influenced significantly. Table 3.2.1 also shows that for $\omega = -3\sigma$ and $\omega = -5\sigma$, the larger σ is, the less will the system be influenced by the outlier, such as when $\omega = -5\sigma$, the false alarm rate will be lower when $\sigma \sim U[2.5,3.5]$ than the false alarm rate when $\sigma \sim U[1.5,2.5]$. However, this is not due to the system's being more robust to the outliers with higher σ level, it is just because of the same reason where higher σ lead to longer delay, that is larger σ will confound the likelihood value and the alarm is not that easy to be triggered compared to data with lower σ.

Generally speaking, the system needs to be modified after outlier pollution when $\omega = -3\sigma$ or even higher. As the down biased outlier will cause an early false alarm for the peak detection in the convex shaped unimodel, the same misidentification can happen when an up biased outlier appears for the bottom detection procedure in a concave unimodel, with the outlier being misidentified as an upper trend turning point. Therefore, when carrying out online turning point detection, it is important to correct the outlier in order to eliminate its negative influence. In many empirical applications, the outlier effect can be eliminated by setting a threshold bound, or by some nonparametric method including smoothing or kernel regression. However, detection outlier in an on-line surveillance procedure requires a more tricky methodology as it needs to combine the outlier detection procedure on-line with turning point detection and correct it as soon as it appears. Some traditional outlier detection approaches which need the whole series of data or nearby observations such as kernel regression are not suitable. This paper will introduce a wavelet based method which can achieve the required detection and correcting demands in the surveillance system as it can handle the data on-line. The main advantage of this wavelet approach is that it can analyze data in both the time domain and the frequency domain and thus possesses good localization identifications in both time and scale. Therefore, for a series which shows non-stable or non-stationary aspects such as structure break, discontinuities or data spike, wavelet methodology will be an elegant algorithm to be adopted.

4. Wavelet based outlier correction methodology

4.1 A brief introduction to wavelets and wavelet multiresolution

The traditional way of analyzing a signal in the frequency domain is the well known Fourier analysis which applies sinusoidal waves as the transformation filter. The main drawback of this transformation is that it can not maintain the information of the time domain and will be unsuitable for signals with irregular behavior such as spikes or data breaks. The wavelet transformation adopts a basis of spatially localized functions as its transform filter. Then based on wavelet filtering of the original signal through shifting and dilations, the wavelet transformation can capture the characteristics of data series both in the frequency domain and the time domain. It is an excellent tool for the analysis of the non-stationary data showing time-localized discontinuities or abrupt changes. By wavelet multiresolution analysis (MRA) which combines resolutions from both time and frequency domains, the signal can be decomposed into different scales where the non-stationarity of the signal can be analyzed according to their own resolution levels: long run trends correspond to the low frequency resolution and the spikes such as the outliers can be captured in the high frequency resolution. A brief introduction of the wavelet methodology is as follows:

Corresponding to sinusoidal waves in the Fourier transform, the wavelet basis functions $\{\psi_{k,j} : k, j \in \mathbb{Z}\}$ used in the wavelet transform are generated by translations and dilations of a basic mother wavelet $\psi \in L^2(\mathbb{R})$ and can be expressed as $\psi_{k,j}(t) = \frac{1}{\sqrt{j}}\psi(\frac{t-k}{j})$. For a continuous signal $f(t)$, its wavelet transform is $\gamma(k, j) = \langle f, \psi_{k,j} \rangle = \int f(t)\psi^*_{k,j}(t)dt$ and the inverse wavelet transform is $f(t) = \iint \gamma(k,j)\psi_{k,j}(t)dkdj$. Time and frequency resolutions can be achieved using different choices of k and j. In the time domain, translation of k corresponds to different time points; in the frequency domain, compressed versions of $\psi_{k,j}(t)$ with lower j maintain the high frequency information of the original signal, while dilated versions with larger j capture the lower frequencies in the signal.

For discrete time series, the original Discrete Wavelet Transform (DTW) can be achieved by certain orthonormal transformation. We here introduce the maximal overlap discrete wavelet transform (MODWT) which is not orthonormal but has no restriction on the sample size, while the original DWT needed the sample length be a multiple of a power of two. For an N dimensional discrete vector $\mathrm{X} = \{X_t, t = 0,..., N-1\}$, the level J MODWT of X contains $J+1$ vectors $\mathbf{W}_1,...,\mathbf{W}_J,\mathbf{V}_J$ with wavelet coefficients $\mathbf{W}_j$ corresponding to changes of scale $\tau_j = 2^{j-1}$, while the wavelet scaling coefficients $\mathbf{V}_J$ corresponds to averages on a scale of $\lambda_J = 2^j$. The N dimensional vectors $\mathbf{W}_j$ and $\mathbf{V}_J$ are computed by $\mathbf{W}_j = \mathcal{W}_j\mathrm{X},\ \mathbf{V}_J = \mathcal{V}_J\mathrm{X}$ where $\mathcal{W}_j$ and $\mathcal{V}_J$ are $N \times N$ matrices. Then the MODWT based MRA of X is defined as: $\mathrm{X} = \sum_{j=1}^{J} \mathcal{W}_j^T\mathbf{W}_j + \mathcal{V}_J^T\mathbf{V}_J = \sum_{j=1}^{J} D_j + S_J$, where D_j is the j^{th} level MODWT detail containing the microscopic detail of X which is the high frequency information of the original signal and S_J is the J^{th} level MODWT smooth containing landscape characteristics of X which is the low frequency resolution of the signal. Basically, the MODWT and multiresolution can be viewed as a band-pass filter process on X, and based on different transformation matrices $\mathcal{W}_j$ and $\mathcal{V}_J$, we have different choices of filters. For more information about the wavelet methodology and MODWT, we refer to Vidakovic (1999), Percival and Walden (2000), and Gençay *et al.* (2001). An important issue now is how to choose the wavelet filter. A central factor in choosing a particular wavelet is to match the characteristics of the series under consideration. The number *L* in the name of the wavelet indicates the width of the filter. In general, the wavelets with small *L* are narrower and less smooth, while wavelets with large *L* are relatively wide and smooth. This paper chooses the Haar wavelet with *L* = 2 as it does not suffer from the problem of boundary coefficients and is most suitable for our on-line outlier detection system.

4.2 Wavelet based method to correct for outliers in an on-line surveillance system

In Section 4.1 we mentioned that through the wavelet decomposition, the wavelet detail will maintain the high frequency information and the wavelet smooth keeps the low frequency trend. As outliers belong to the microscopic detail of the signal, it is reasonable to analyze it in the wavelet detail, which is most sensitive to the local behavior of the signal. There already exist literatures on the wavelet outlier detection: such as Canan and Huzurbazar (2002) and Aurea *et al.* (2009). The main idea of these papers is to set a threshold for the wavelet detail coefficient of the original observations (see Canan and Huzurbazar, 2002) or the residuals from the specified model (see Aurea *et al.*, 2009). The outlier can be detected when the detail coefficients surpass the threshold and later be corrected after an inverse wavelet transformation. In our system of surveillance analysis, we need to specify the underlying model of the upward trend $D(s)$ and that of the downward trend $C(s)$, thus the residual based method is not suitable as we have no idea which pre-model is specified first. Instead we take the wavelet detail D_j directly from the original data to check if it is outside a certain threshold level. The first level wavelet detail is taken as it captures the finest information of the signal and will be most sensitive to the outliers. Thus we set $J = 1$ in the wavelet transform which results in decomposition $\mathrm{X} = D_1 + S_1$. More straightforwardly, for an outlier polluted series, the following figure shows how the wavelet detail can be used to detect the outlier:

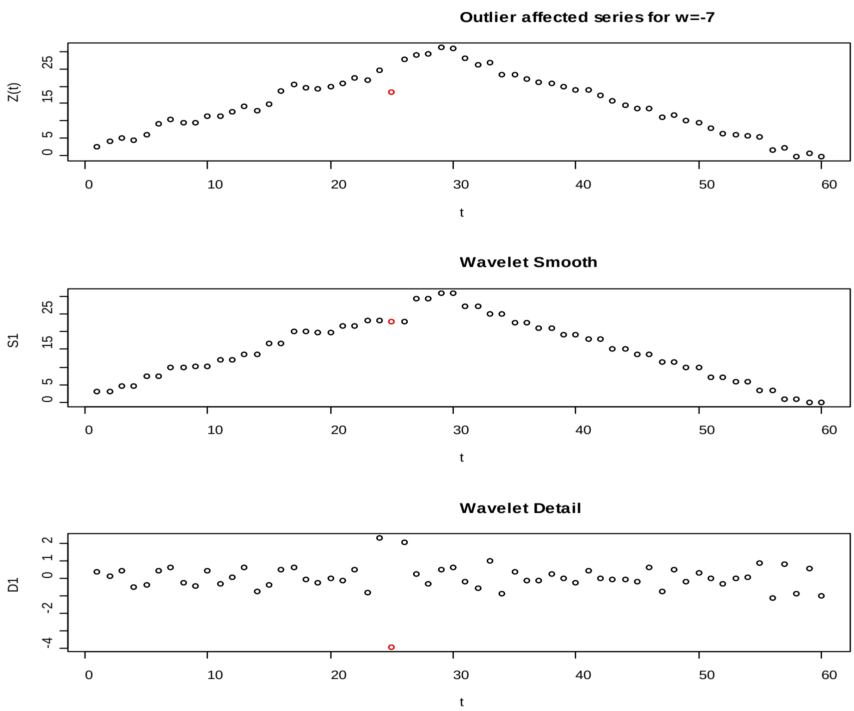

Figure 4.2.1: Wavelet decomposition of outlier polluted series

Figure 4.2.1 shows that for this series with outlier appearing at time 25, according to the wavelet decomposition, wavelet detail D_1 is quite sensitive to the outliers with a significant deviation at the outlier occurrence time, which makes it efficient to detecting outliers. For the wavelet smooth S_1, it can still remain the original unimodel structure and the outlier time is not obvious.

It is also important to choose the threshold level since a large threshold will destroy the information of the original data while a small one may lack the ability to detect outliers. Canan and Huzurbazar (2002) chose the universal threshold suggested by Donoho *et al.* (1994) with the threshold value set to $\lambda = 2\sigma^2 \log(N)$ where N is the length of the decomposed vector and σ is estimated by the median absolute deviations (MAD) of the wavelet detail. This threshold can not be used here as we do not know the whole observation number N in the detection procedure. Instead we follow the procedure in Aurea *et al.* (2009) and set the threshold value θ directly to the lower 2.5% percentile value of the wavelet detail from standard normally distributed data. The wavelet detail from the original series which is under this value is set to 0 while the others remain the same and result in a new series of wavelet detail $D_1^{'}$. Next we reconstruct a series $X^{'} = D_1^{'} + S_1$ and apply it for the turning point detection. Suppose that at time t the alarm is not triggered and we already determined the threshold θ. Then the whole procedure can be carried out in the following steps:

Step 1: Based on all the available observations $x_1 \ldots x_{t+1}$, we use the wavelet decomposition to decompose the series into wavelet detail D_1 and wavelet smooth S_1.

Step 2: Record the time when D_1 lower than θ, set the corresponding D_1 to 0 and this results in a new wavelet detail $D_1^{'}$.

Step 3: Set $X^{'} = D_1^{'} + S_1$ and then put $X^{'}$ into the detection system.

The new series $X^{'}$ maintains the original structure of the observations X but with the suspicious outlier point corrected, and we can also know when the outlier appears from the information given in D_1. For the outlier polluted series in Figure 1, we apply the procedure based on the above 3 steps. As when $\omega = -1.5\sigma$, the system is almost not influenced by the outliers, we only carry out the correlation procedure for $\omega = -3\sigma$ and $\omega = -5\sigma$, and then get the outlier corrected series in Figure 4.2.2:

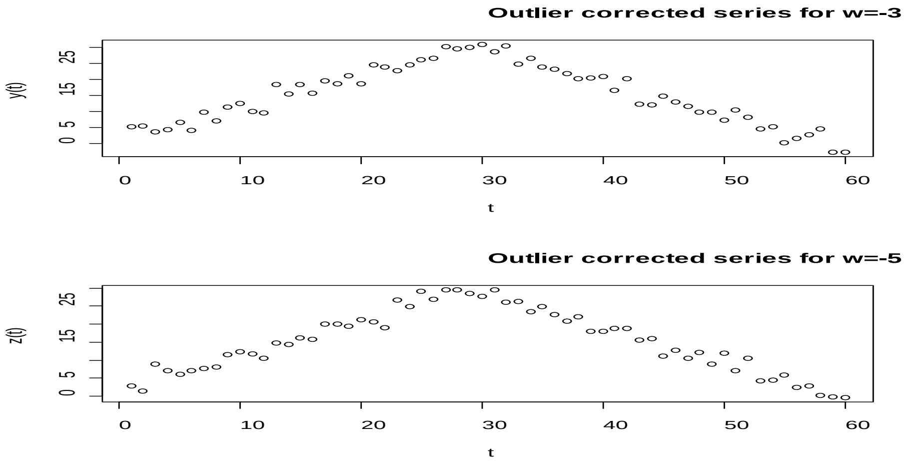

Figure 4.2.2: Wavelet threshold corrected unimodel

Figure 4.2.2 shows that for $\omega = -5\sigma$, the outlier shrinks to an extent which can almost be ignored and for $\omega = -3\sigma$, the outlier disappears totally. Moreover, the following simulation will show that for $\omega = -3\sigma$ and -5σ, by using $X^{'}$ instead of X , the performance of the turning point detection procedure in the surveillance system will be improved with a much lower false early alarm rate. In the new Monte Carlo experiment, the DGP is the same outlier polluted data as those from Section 3. Then we apply the wavelet outlier correction methodology at the same time with the turning point detection process. Still after 1000 replications, we get the following table:

Table 4.2.1: Property of the surveillance system after filtering the outliers

ω	*False early alarm*		*Alarm delay*	
	length	rate	length	rate
$\sigma \sim U[0.5,1.5]$				
-3σ	14.544	0.079	1.875	0.898
-5σ	13.759	0.216	1.456	0.738
$\sigma \sim U[1.5,2.5]$				
-3σ	16.589	0.073	2.825	0.915
-5σ	14.099	0.181	2.597	0.803
$\sigma \sim U[2.5,3.5]$				
-3σ	13.833	0.072	3.352	0.911
-5σ	14.691	0.178	3.388	0.804

Compared with Table 3.2.1, the false alarm rates in Table 4.2.1 are subdued to a large extent both when $\omega = -3\sigma$ and -5σ, especially when $\sigma \sim U[0.5,1.5]$ and $\sigma \sim U[1.5,2.5]$ where the system is influenced seriously by the outliers, the reduction of the false alarm rate is quite obvious. We even do more simulations for $\omega = -7\sigma$ where the false alarm rate for the uncorrected data approaches 100%, while after the wavelet correlation it falls to around 40%. Although for ω lager than -5σ, the false alarm rates are still not very low even after the wavelet filtering, but in those cases the outliers are easily noticed visually when more observations are added. Thus by combining the wavelet detection and visual impression together, the outlier problem can be reduced significantly in the surveillance process. Table 4.2.1 also shows that the corresponding alarm delay rates are higher by using $X^{'}$ instead of applying original data X. As the delay lengths are quite moderate, this higher delay rate problem is not serious compared with the problems brought up by false alarm with its length being easily larger than 10.

5. Conclusion

This paper concentrates on three issues: first a power controlled on-line turning point detection system is proposed in Section 2 and we show this methodology performs well with the ability to give out timely alarms after only short delays. Section 3 points out that the decent behavior of this method is degraded by an outlier, which brings about a high false early alarm rate. To solve this problem, we next apply a wavelet multiresolution (MRA) based on-line outlier elimination method in Section 4, both the visual figures and the simulation results show that this methodology can reduce the influence of the outlier considerably. Generally speaking, the wavelet based approach has the advantage of being able to detect and correct the outlier on-line with turning point monitoring as the data process continues. Moreover, although the whole analysis in this paper is based on a linear parametric model, the same technologies can be extended to another unimodel quite easily. We only need to change the likelihood function in Section 2, and all the methodologies are fairly robust to the underlying unimodel structure.

References

Andersson E. (2002). "Monitoring cyclical processes—a nonparametric approach", *Journal of Applied Statistics*, Vol. 29, No. 7, pp. 973-990.

Andersson E., Bock D. and Frisén M. (2005). "Statistical surveillance of cyclical processes with application to turns in business cycles", *Journal of Forecasting*. Vol. 24, No. 7, pp. 465-490.

Andersson E., Bock D. and Frisén M. (2006). "Some statistical aspects on methods for detection of turning points in business cycles", *Journal of Applied Statistics*, Vol. 33, No. 3, pp. 257–278.

Aurea G. and Helena V. (2009). "Wavelet-based detection of outliers in volatility models", *Statistics and Econometrics Working Papers*, Universidad Carlos III de Madrid, Calle Madrid, Getafe, Spain.

Barnett V. and Lewis T. (1994). *Outliers in statistical data*, 3rd edition, New York: Wiley.

Canan B. and Huzurbazar S. (2002). "Wavelet-Based Detection of Outliers in Time Series", *Journal of Computational and Graphical Statistics*, Vol. 11, No. 2, pp. 311-327.

Donoho D. L. and Johnstone I. M. (1994). "Ideal Spatial Adaptation by Wavelet Shrinkage", *Biometrika*, Vol. 81, pp. 425-455.

Frisén M. (1994). "Statistical Surveillance of Business Cycles", Research Report, Department of Statistics, Göteborg University, Sweden, 1994:1 (Revised 2000).

Gan F. F. (1993). "An optimal design of EWMA control charts based on median run-length", *Journal of Statistical Computation and Simulation*, Vol. 45, pp. 169-184.

Gençay R., Selçuk F. and Whitcher B. (2001). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economic*, Academic Press, San Diego, CA, USA.

Ling S. Q. and Li W. K. (2003). "Asymptotic Inference for Unit Root Processes with GARCH (1, 1) Errors", *Econometric Theory*, Vol. 19, No. 4, pp. 541-564.

Percival D. B. and Walden A. T. (2000). *Wavelet Methods for Time Series Analysis*, Cambridge Univ. Press.

Roberts S. W. (1966). "A comparison of some control chart procedures", *Technometrics*, Vol. 8, pp. 411-430.

Shiryaev A. N. (1963). "On optimum methods in quickest detection problems", *Theory of Probability and its Applications*, Vol. 8, pp. 22-46.

Vidakovic B. (1999). *Statistical Modelling by Wavelets*, Wiley: New York.

Research Papers 2011

2011-16: Stefano Grassi and Tommaso Proietti: Characterizing economic trends by Bayesian stochastic model specification search

2011-17: Søren Johansen and Theis Lange: Some econometric results for the Blanchard-Watson bubble model

2011-18: Tom Engsted and Thomas Q. Pedersen: Bias-correction in vector autoregressive models: A simulation study

2011-19: Kim Christensen, Roel Oomen and Mark Podolskij: Fact or friction: Jumps at ultra high frequency

2011-20: Charlotte Christiansen: Predicting Severe Simultaneous Recessions Using Yield Spreads as Leading Indicators

2011-21: Bent Jesper Christensen, Olaf Posch and Michel van der Wel: Estimating Dynamic Equilibrium Models using Macro and Financial Data

2011-22: Antonis Papapantoleon, John Schoenmakers and David Skovmand: Efficient and accurate log-Lévi approximations to Lévi driven LIBOR models

2011-23: Torben G. Andersen, Dobrislav Dobrev and Ernst Schaumburg: A Functional Filtering and Neighborhood Truncation Approach to Integrated Quarticity Estimation

2011-24: Cristina Amado and Timo Teräsvirta: Conditional Correlation Models of Autoregressive Conditional Heteroskedasticity with Nonstationary GARCH Equations

2011-25: Stephen T. Ziliak: Field Experiments in Economics: Comment on an article by Levitt and List

2011-26: Rasmus Tangsgaard Varneskov and Pierre Perron: Combining Long Memory and Level Shifts in Modeling and Forecasting of Persistent Time Series

2011-27: Anders Bredahl Kock and Timo Teräsvirta: Forecasting Macroeconomic Variables using Neural Network Models and Three Automated Model Selection Techniques

2011-28: Anders Bredahl Kock and Timo Teräsvirta: Forecasting performance of three automated modelling techniques during the economic crisis 2007-2009

2011-29: Yushu Li: Wavelet Based Outlier Correction for Power Controlled Turning Point Detection in Surveillance Systems