

FOR 10 2014

ISSN: 1500-4066

March 2014

Discussion paper

Wavelet improvement in turning point detection using a Hidden Markov Model

BY
Yushu Li AND Simon Reese

Wavelet improvement in turning point detection using a Hidden Markov Model

-from the aspects of cyclical identification and outlier correction

Yushu Li¹

Department of Business and Management Science, Norwegian School of Economics, Norway

Simon Reese

Department of Economics, Lund University, Sweden

Abstract

The Hidden Markov Model (HMM) has been widely used in regime classification and turning point detection for econometric series after the decisive paper by Hamilton (1989). The present paper will show that when using HMM to detect the turning point in cyclical series, the accuracy of the detection will be influenced when the data are exposed to high volatilities or combine multiple types of cycles that have different frequency bands. Moreover, outliers will be frequently misidentified as turning points. The present paper shows that these issues can be resolved by wavelet multi-resolution analysis based methods. By providing both frequency and time resolutions, the wavelet power spectrum can identify the process dynamics at various resolution levels. We apply a Monte Carlo experiment to show that the detection accuracy of HMMs is highly improved when combined with the wavelet approach. Further simulations demonstrate the excellent accuracy of this improved HMM method relative to another two change point detection algorithms. Two empirical examples illustrate how the wavelet method can be applied to improve turning point detection in practice.

JEL classification: C22, C38, C63

Keyword: HMM, turning point, wavelet, wavelet power spectrum, outlier

¹ Yushu Li gratefully acknowledges funding from Swedish Research Council (project number 421-2009-2663).

1. Introduction

Hidden Markov Models (HMM) are stochastic signal models whose mathematical formalism is rooted in a series of papers by Baum *et al.* in the late 1960s (Baum and Petrie, 1966; Baum and Eagon, 1967; Baum, *et al.*, 1970) and Rabiner (1989). The HMM framework contains doubly stochastic processes with an underlying unobservable variable, which can be observed through another stochastic sequence. HMMs can handle both long term variation for the underlying process and instantaneous randomness in the observed symbols. This high flexibility in modeling the variable-length sequence and time variant characteristics has resulted in wide application of the HMM in various fields, such as biological modeling (Anders, 1997), speech recognition (Huang *et al.*, 1990), and signal processing (Elliott *et al.*, 1995).

In an econometric study performed more than half a century ago, Burns and Mitchell (1946) showed that economic cycles show an asymmetric shift where the turning points separate long periods of steady expansions and rapid recessions. HMMs gained popularity in modeling economic cycles due to their ability to capture the asymmetry of both duration and amplitude in business cycle expansions and recessions. The HMM framework assumes that economic statuses are unobservable and can be classified as K -state regimes, which can be modeled by a K -state Markov Chain. Using historical data, the HMM can calculate the probability that the underlying state of a set of given observations belongs to a certain system. When the specification rule is determined, further regime classification becomes possible. This sophisticated model has been extensively studied (both empirically and theoretically) in recent decades (Hamilton, 1989; Stock and Watson, 1989; Kim and Nelson, 1998). While the fusion schemes of the HMM under complex environments (*e.g.*, frequent illumination changes or outlier contamination in pattern recognition) are well documented (Chengalvarayan, 1999; Wachter *et al.*, 2007), the investigation of HMM performance in detecting turning points in cyclic data is not as well represented in the literature. Therefore, the current paper will examine two practical issues of HMM-based turning

point detection procedures. One issue is that the data is exposed to high volatility or composed by series of different frequency bands. This investigation is necessary, as it is well known that economic data is generally beset by short-term volatility, long-term trends and business cycles with various frequencies. Another important issue is the quality of the data to be analyzed. In the HMM framework, the true process of interest is the underlying signal and any external pollution of the data will complicate the actual dynamics of the signals. One source of data pollution about which one must be particularly cautious when attempting to detect turning points is the appearance of outliers, which may easily be misidentified as turning points. In particular, outliers lead to significant computation inefficiency for HMMs which use the Gaussian density as hidden state distribution, due to the intolerance of the Gaussian distribution to fat tails.

This study uses Monte Carlo simulations to investigate how the HMM turning point detection will be affected when the data are exposed to the above-mentioned issues. The simulation results show that the HMM fails to detect the underlying structural changes when the data has components with different periodical dynamics or when the signals are perturbed by high volatilities. In addition, there is a strong possibility that outliers are misidentified as turning points. To resolve these issues, the current paper proposes a wavelet based method, due to this method's ability to decompose a series into different frequency bands. The dynamic information corresponding to different frequency bands will be extracted based on the wavelet decomposition, and outliers, which can be viewed as high-frequency spikes, will be detected using the high-frequency wavelet details.

The remainder of this paper is structured as follows: Section 2 introduces the basic framework of the HMM and describes how this model can be used to detect turning points; Section 3 investigates how the detection procedure is influenced by volatility, different frequency combinations and outlier effects. Section 4 introduces the wavelet method and combines it with the HMM to alleviate the issues that are detailed in Section 3. The final section includes discussion and conclusions.

2. HMM and its application in detecting turning point

2.1. Introduction to HMM

The HMM gradually gained popularity in regime classification and business cycle identification after a series of research papers (Hamilton, 1989; Stock and Watson, 1989; Kim and Nelson, 1998). The HMM assumes that the underlying state of a given system is a stochastic process in the form of a Markov Chain, and that the observations follow state-dependent distributions. For the discrete Hidden Markov Chain, let S_t denote the underlying process and let X_t represent the observations at

time t . The relationship can be described graphically:

$$\begin{array}{ccccccc}
 & & S_0 & \rightarrow & S_1 & \dots & \rightarrow & S_{t-1} & \rightarrow & S_t & \rightarrow & \dots \\
 & & \downarrow & & \downarrow & & & \downarrow & & \downarrow & & \\
 \text{time } t & . & X_0 & & X_1 & & & X_{t-1} & & X_t & & .
 \end{array}$$

Taking the first order Markov Chain as example, the distribution of the unobserved state S_t only depends on the most recent state with transition probability $p_{ij} = P(S_{t+1} = j | S_t = i, S_{t-1} = q, \dots) = P(S_{t+1} = j | S_t = i)$. The underlying process can then be

fully specified by the first state S_0 and the transition matrix $\Gamma = \begin{bmatrix} p_{11}, \dots, p_{1k} \\ \dots p_{ij} \dots \\ p_{k1}, \dots, p_{kk} \end{bmatrix}$ where

$\sum_{j=1}^k p_{ij} = 1$ for $i = 1, \dots, k$, with k being the number of states. Furthermore, the

distribution of each observation X_t is generated by the density function $P_{S_t}(X_t)$ which depends only on the present state S_t and is independent of other observations.

Thus, given $X^{(t)}$ and $S^{(t)}$ denoting the historic information of the process, the Markov property and the conditional dependence inside the HMM can be summarized as:

$$\begin{aligned}
 P(S_t | S^{(t-1)}) &= P(S_t | S_{t-1}), & t = 2, 3, \dots; \\
 P(X_t | X^{(t-1)}, S^{(t)}) &= P(X_t | S^{(t)}), & t \in \mathbb{N}
 \end{aligned}$$

When the observation process $\{X_t : t = 1, 2, \dots\}$ is available, the parameters of the model (*e.g.*, the initial state probabilities, the transition matrix and the priori distribution parameters) can be computed efficiently by the Expectation-maximization

(EM) algorithm or Baum–Welch algorithm (also known as *Forward-Backward* algorithm). For more details regarding HMMs, the interested reader is referred to Levinson *et al.* (1983), Elliot *et al.* (1995), or MacDonald and Zucchini (1997).

2.2. Applying the HMM to detect turning points in business cycles

The original work of Hamilton (1989) was generalized to, and developed extensively for, detecting turning points (Layton, 1996; Hamilton and Perez-Quiros, 1996; Krolzig, 2003). The commonly adopted HMM for state identification is in the form of an order p autoregressive structure: $X_t = \mu_{s_t} + \varphi_1[X_{t-1} - \mu_{s_t}] + \dots + \varphi_p[X_{t-p} - \mu_{s_t}] + \varepsilon_t$

where μ_{s_t} or even $\varphi_1 \dots \varphi_p, \varepsilon_t$ depend on S_t . To detect the turning points in business cycles based on the two regimes classification, a simplified HMM, $X_t = \mu_{s_t} + \varepsilon_{s_t}$

where $\varepsilon_{s_t} \square N(0, \sigma_{s_t})$ is applied (Bellone and Saint-Martin, 2004). In the simplified

HMM, we assume that the variable $S = \{S_t : t = 1, 2, \dots\} \in \{0, 1\}$ follows a two-state first order Markov Chain, which corresponds to the contraction and expansion states of the

business cycle at each time point t :
$$\begin{cases} S_t = 0 : \mu_{0t} = \beta_0 + \beta_1 \cdot t \\ S_t = 1 : \mu_{1t} = \beta_0 + \beta_1 \cdot (\tau - 1) - \beta_2 \cdot (t - \tau + 1) \end{cases}$$

where $\tau \leq t$ is the turning point and $\beta_1, \beta_2 > 0$. Thus when $S_t = 0$,

$E(X_t - X_{t-1}) = \beta_1 > 0$ and the cycle increases while when $S_t = 1$,

$E(X_t - X_{t-1}) = -\beta_2 < 0$ and the cycle decreases. The usual criterion for identifying

which state S_t belongs to is the posterior probability $P(S_t | I) > 0.5$ and the turning

point is the time at which the underlying state changes regimes. For a detailed

detection procedure we refer to Krolzig (2003). Past literature that follows the

procedure in Krolzig (2003) shows that this methodology is efficient and decisive.

However, quick and accurate detection of turning points in the HMM framework is

built upon certain assumptions, such as correct specification of the volatility and

independently normal distributed errors. In practical applications, the data is easily

subject to collection mechanism errors, calculation errors or unexpected and extreme

events. The following sections will illustrate three factors that influence the detection procedure: high volatility, multi-frequency bands and outliers.

3. HMM detection when data are exposed to high volatility, multi-frequency bands, and outliers

3.1. High volatility and frequency bands

For econometric data, a primary source of obstruction in the main trend of the stochastic process is irregular factors in the form of short-term volatility. When detecting turning points in cyclical series, noise with large volatility leads to a shift in the location of a given turning point and results in incorrect estimation. The location shift problem is exacerbated when the data includes more than one type of cycle. Here, we apply a Monte Carlo experiment to show how the HMM fails to identify the actual turning point when the process is exposed to high volatility or composed of different frequency bands from separate uniform distributions, by simulating three types of signals with different deterministic cycles: a cosine wave with medium period, a sine wave with long period and a combination of these two signals. The volatility is introduced by independent Gaussian processes with different variances σ_1 , σ_2 and σ_3 . The signals can be expressed separately as:

$$y_1(t) = 1.5 \cos(2\pi t f_1) + \varepsilon_1(t), \text{ where } f_1 \sim \text{unif.}(1/8, 1/4), \varepsilon_1(t) \sim \text{i.i.d.}(0, \sigma_1)$$

$$y_2(t) = 1.5 \sin(2\pi t f_2) + \varepsilon_2(t), \text{ where } f_2 \sim \text{unif.}(1/16, 1/8), \varepsilon_2(t) \sim \text{i.i.d.}(0, \sigma_2).$$

$$y_3(t) = 1.5 \cos(2\pi t f_1) + 1.5 \sin(2\pi t f_2) + \varepsilon_3(t), \varepsilon_3(t) \sim \text{i.i.d.}(0, \sigma_3)$$

To examine the HMM's accuracy in detecting the turning points for the underlying trend in data $y_1(t)$, $y_2(t)$ and $y_3(t)$, we simulate 200 trials and calculate the average value of the correct specification ratio, which is the number of correctly detected turning points to the number of actual turning points. We set the variance $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$ at three different levels: 0.5, 1 and 1.5. The result is as follows:

Table 1: Correct specification ratio for HMM model

| T | $\sigma = 0.5$ | | | $\sigma = 1$ | | | $\sigma = 1.5$ | | |
|-----|----------------|----------|----------|--------------|----------|----------|----------------|----------|----------|
| | y_{1t} | y_{2t} | y_{3t} | y_{1t} | y_{2t} | y_{3t} | y_{1t} | y_{2t} | y_{3t} |
| 200 | 0.710 | 0.472 | 0.427 | 0.578 | 0.432 | 0.399 | 0.473 | 0.320 | 0.384 |
| 500 | 0.709 | 0.450 | 0.430 | 0.551 | 0.389 | 0.391 | 0.475 | 0.347 | 0.363 |
| 100 | 0.710 | 0.493 | 0.516 | 0.569 | 0.372 | 0.466 | 0.480 | 0.333 | 0.433 |

Table 1 shows that, among the three series, $y_1(t)$ has the highest correct specification ratio, while $y_3(t)$ has the lowest. The low correct specification ratio in $y_2(t)$ relative to $y_1(t)$ is due to the cycle dynamics $y_2(t)$ has a longer period while the volatilities between the cycles will be miss-specified as turning points. The lowest specification ratio in $y_3(t)$ indicates that the frequency band combination negatively affects the HMM to a greater extent. Moreover, with the increase of the variance from 0.6 to 1.5, the correct specification ratio decreases for all the series. An empirical example is illustrated using quarterly U.S. GDP growth rates from 1960 to 2011. We aim at estimating acceleration cycle turning points, i.e. the break points between periods of ever-increasing and ever-diminishing growth rates. The identified turning points that the HMM procedure reports are highlighted in Figure 1 by red and blue dots for peaks and troughs respectively.

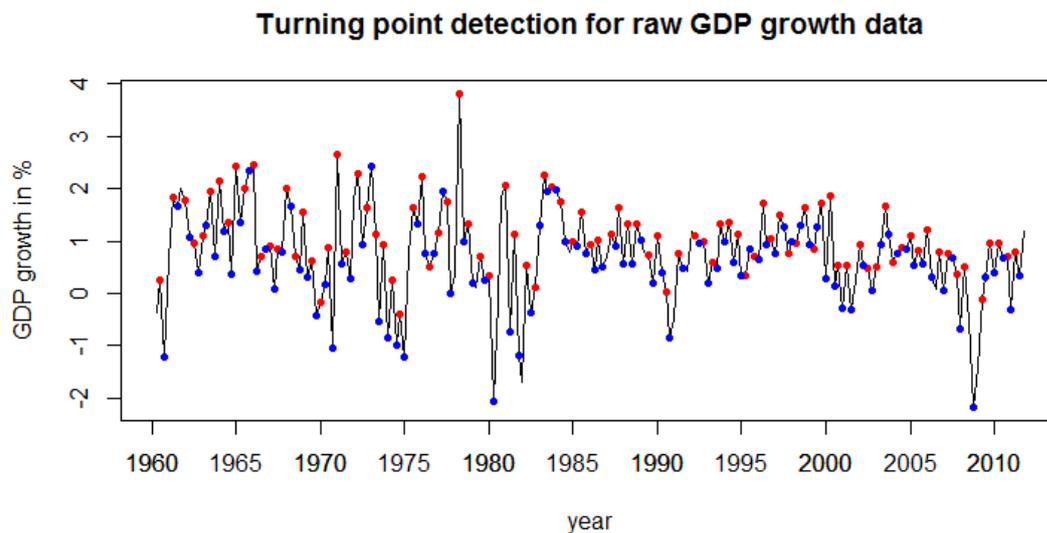


Figure 1: HMM detection result for US GDP growth rate

When using the original data to detect the peak and trough points with a HMM, we obtain a pattern of frequently recurring peaks and troughs at nearly every possible local extreme point. This result can hardly be reconciled with the idea of business cycle fluctuations that span over several years. It suggests as well that high short-term volatility impairs the performance of HMMs since they dominate any cyclical movements at lower frequencies.

An equally poor result is obtained when applying HMMs to a sample of monthly U.S. purchasing manager index (PMI) values for manufacturing firms between January 1948 and October 2013. Ideally, the analysis is meant to identify turning points of the industrial cycle, i.e. changes between phases of expansion and contraction in industrial activity. The obtained dates for peaks and troughs (see Figure XYZ) are however rather randomly distributed over the sample and do not match the cycles one could identify by visual inspection of the series. Again, it is plausible that activity at high frequencies impairs the performance of HMMs in identifying turning points of underlying cycles that range over several years.

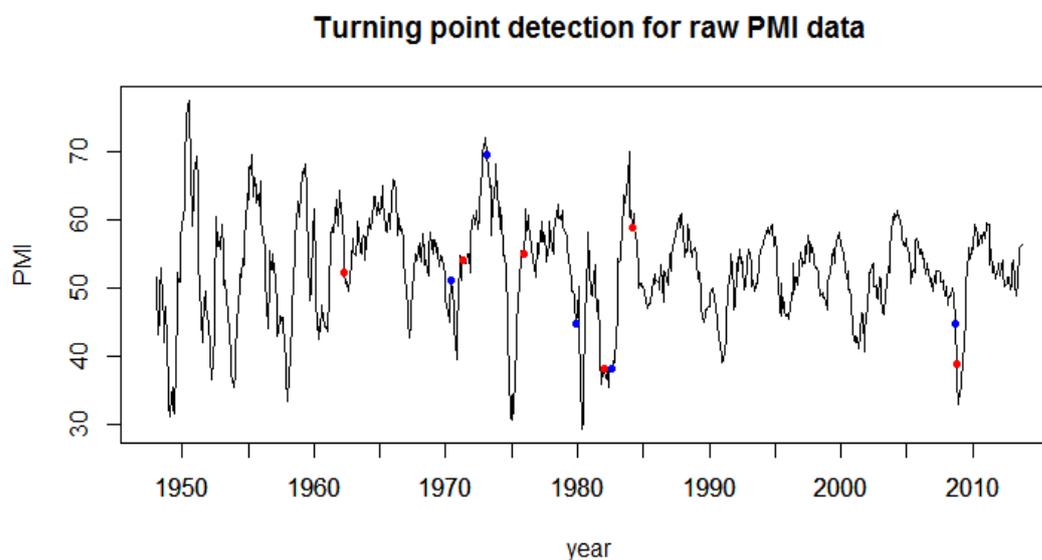


Figure 2: HMM detection result for U.S. purchasing manager index

Thus, both the Monte Carlo simulation and the empirical example show that it is necessary to deduct the variance and determine the actual pattern of the series before applying the HMM model. A common technique that is used to deduct the volatility

created by the noise is to smooth the data by passing it through a low pass filter and then extract the main trend. However, for process such as $y_3(t)$, the de-noised data also contain dynamics with difference frequencies: higher frequency cycles with periods of 4-8 units, and lower frequency cycles with periods of 8-16 units. Thus, purely de-noising the data is not sufficient, and filters that can separate cycles into different frequency bands are required for better dynamic identification.

3.2. Outlier influence

The effect of outlier contamination on HMM-based segmentation in pattern recognition is well documented (Chengalvarayan, 1999; Wachter *et al.*, 2007). The current paper focuses on additive outlier detection, which is defined in Barnett and Lewis (1994) 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 that is equal to 1 at the outlier appearance time and 0 otherwise. We still use a Monte Carlo simulation to illustrate how the outlier will influence turning point detection. The data are generated from a two-state HMM model with $p_{11} = p_{22} = 0.9$, and the turning point is the time where the process changes states. We measure the probability of misidentifying the outlier as a turning point, which is calculated as the ratio of the number of outliers being misidentified as a turning point to the total number of outliers. The number of outliers is set as 2%, 4% and 6% of the data size, which corresponds to 200, 500 and 1000 trials, respectively. The magnitudes of the outliers are set as $\omega = 3$, $\omega = 5$ and $\omega = 7$, which represent weak, medium and strong outliers. Based on 1000 simulations, we obtained the misidentification ratios as:

Table 2: Misspecification ratio under outlier influence

| T | $\omega = 3$ | | | $\omega = 5$ | | | $\omega = 7$ | | |
|------|--------------|-------|-------|--------------|-------|-------|--------------|-------|-------|
| | 2% | 4% | 6% | 2% | 4% | 6% | 2% | 4% | 6% |
| 200 | 0.510 | 0.549 | 0.539 | 0.915 | 0.893 | 0.876 | 0.950 | 0.912 | 0.888 |
| 500 | 0.512 | 0.522 | 0.521 | 0.916 | 0.897 | 0.872 | 0.949 | 0.917 | 0.892 |
| 1000 | 0.544 | 0.510 | 0.515 | 0.917 | 0.904 | 0.877 | 0.936 | 0.924 | 0.904 |

Table 2 shows that when outliers are present, up to 90% of the outliers will be misidentified as turning points when $\omega=5$ and $\omega=7$. Thus, it is important to detect outliers and correct them in the turning point detection procedure. One common technique for removing outliers is by passing the data through a low-frequency band-smoothing filter to reduce the high-frequency fluctuation of the signal. However, when we examine in the frequency domain, the outlier clearly belongs to the high-frequency spikes, while the smoothing is performed to maintain the primary trend of the underlying process, which motivated our attempt to resolve the issues found with the frequency domain based methodology. Given these findings and the related concerns, we determined that the wavelet method, which provides both frequency decomposition and temporal resolution, may be a viable alternative.

4. Wavelet methodology to improve the HMM detection ability

4.1. Introduction to the wavelet method

Wavelet methods have been widely applied in the field of signal and image processing following their introduction in the 1980s (Grossmann and Morelet, 1984; Mallat, 1989). Corresponding to sinusoidal waves in the Fourier transform, the wavelet bases $\{\psi_{k,a} : k, a \in R\}$ used in the wavelet transform are generated by translations and

dilations of a basic mother wavelet $\psi \in L^2(R)$ and can be expressed as

$\psi_{k,a}(z) = \frac{1}{\sqrt{a}}\psi\left(\frac{z-k}{a}\right)$. For the signal $f(z)$, the wavelet transform is

$\gamma(k, a) = \langle f, \psi_{k,a} \rangle = \int f(z)\psi_{k,a}^*(z)dz$. When the mother wavelet satisfies the condition

$\int_0^\infty \frac{|H(\omega)|^2}{|\omega|} d\omega < \infty$, with $|H(\omega)|$ as the Fourier transform of the $\psi(z)$, we can

reconstruct $f(z)$ using the inverse $f(z) = \iint \gamma(k, a)\psi_{k,j}(z)dkda$. For the discrete

series $Z = \{Z_t, t = 0, \dots, N-1\}$, the level J maximal overlap discrete wavelet

transform (MODWT) contains $J+1$ vectors $\tilde{\mathbf{W}}_1, \dots, \tilde{\mathbf{W}}_J, \tilde{\mathbf{V}}_J$ with wavelet coefficients

$\tilde{\mathbf{W}}_j$ corresponding to changes of scale $\tau_j = 2^{j-1}$, while the wavelet scaling coefficients $\tilde{\mathbf{V}}_j$ corresponds to averages on a scale of $\lambda_j = 2^j$. The N dimensional vectors $\tilde{\mathbf{W}}_j$ and $\tilde{\mathbf{V}}_j$ are computed by $\tilde{\mathbf{W}}_j = \tilde{\mathbf{W}}_j Z$, $\tilde{\mathbf{V}}_j = \tilde{\mathbf{V}}_j Z$ where $\tilde{\mathbf{W}}_j$ and $\tilde{\mathbf{V}}_j$ are $N \times N$ matrices. Then, the MODWT based synthesis is: $Z = \sum_{j=1}^J \tilde{\mathbf{W}}_j^T \tilde{\mathbf{W}}_j + \tilde{\mathbf{V}}_j^T \tilde{\mathbf{V}}_j = \sum_{j=1}^J \tilde{D}_j + \tilde{S}_j$, where \tilde{D}_j is the j^{th} level MODWT detail containing the information in frequency band $(\frac{1}{2^{j+1}}, \frac{1}{2^j})$ of X and \tilde{S}_j is the J^{th} level MODWT smooth containing information in the frequency band $(0, \frac{1}{2^J})$. For more information about the wavelet methodology and MODWT, we refer to Vidakovic (1999), Percival and Walden (2000), and Gençay *et al.* (2001).

Thus, based on wavelet filtering of the original signal through shifting and dilations, the wavelet transformation can capture frequency and time information of a given data series. Wavelet multi-resolution analysis (MRA) can further decompose the signal into different scales where the non-stationary nature of the signal can be analyzed according to its own resolution levels: long run or medium run trends correspond to the different frequency resolutions, and the spikes (*e.g.*, outliers) can be captured in the high-frequency resolution.

4.2. Wavelet decomposition based on power spectrum

The low correct specification ratio in section 3.2 is introduced by combining the dynamics from the short-term volatility with cycles with different frequency bands. However, the different type of dynamics can be easily identified by the wavelet power spectrum, which provides both time and frequency resolutions, for one trial of $y_1(t)$, $y_2(t)$ and $y_3(t)$; we obtain the following wavelet spectrum:

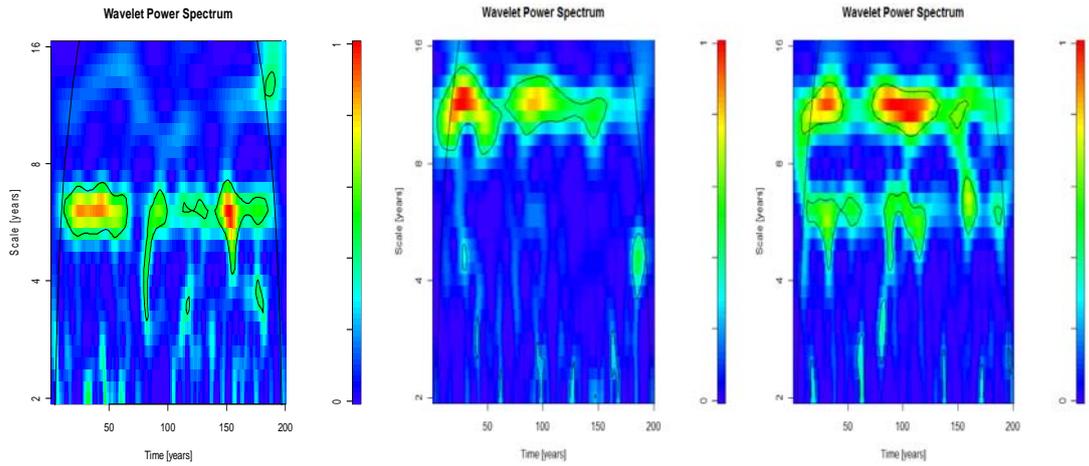


Figure 3: Wavelet power spectrum for three series

The wavelet power spectrum shows clear energy distributions of the series $y_1(t)$, $y_2(t)$ and $y_3(t)$. Based on the spectrum, we can further select the level of wavelet decomposition to extract information on specific cyclical behavior. The present paper applies $y_3(t)$ to illustrate the wavelet decomposition, as the wavelet power spectrum shows that the series is combined with two different sets of cyclical information, a second-level wavelet decomposition is chosen and the wavelet details D_1 , D_2 and D_3 can separately extract information about short-term volatility, medium period and long cycle with the corresponding power spectrums:

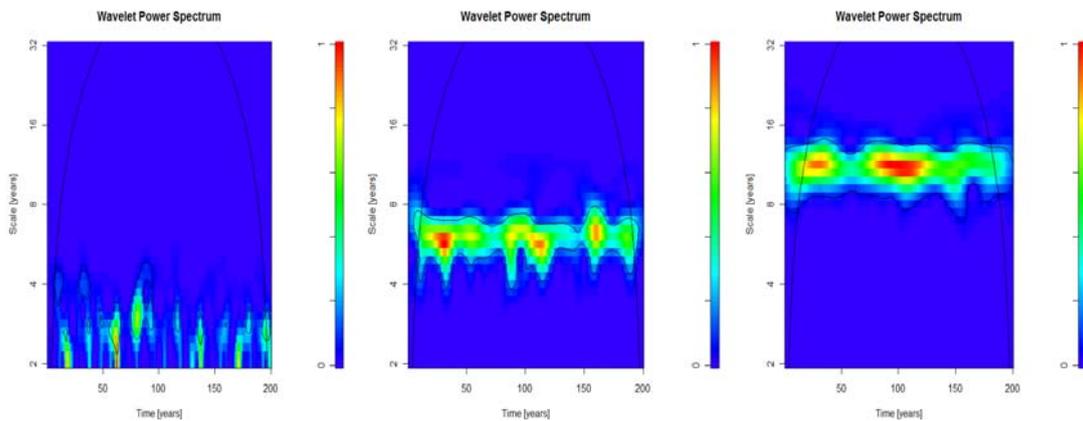


Figure 4: Wavelet power spectrums for wavelet decomposed series

Figure 3 shows that the wavelet decomposition successfully separates the cyclic dynamics according to different frequency bands. We can further use HMM to detect

the turning point for medium period and long cycle in D_2 and D_3 . A further 1000 trial Monte-Carlo simulation for the detection process based on D_2 and D_3 for $y_3(t)$ can be seen in Table 3:

Table 3: Correct specification ratio for HMM model based on D_2 and D_3 for $y_3(t)$

| T | $\sigma = 0.5$ | | $\sigma = 1$ | | $\sigma = 1.5$ | |
|------|----------------|-------|--------------|-------|----------------|-------|
| | D_2 | D_3 | D_2 | D_3 | D_2 | D_3 |
| 200 | 0.984 | 0.857 | 0.854 | 0.729 | 0.671 | 0.554 |
| 500 | 0.980 | 0.853 | 0.857 | 0.683 | 0.664 | 0.527 |
| 1000 | 0.983 | 0.887 | 0.857 | 0.670 | 0.670 | 0.541 |

Relative to Table 1, the correct specification ratio increased significantly in $y_3(t)$. Moreover, because D_2 and D_3 separately contain identical dynamic behaviors as $y_2(t)$ and $y_3(t)$, they can be compared, with the former two series showing much higher correct specification ratios.

In the empirical example, the power spectrum, shown in the left panel of figure 6, suggests that the energy of the data is concentrated primarily on the dynamics that stretch over more than four years. Activity at higher frequencies is present, but has a very limited impact on some local spots of the time series. As the data are sampled quarterly, the wavelet smooth S_3 which contains information in frequency band $(0, \frac{1}{16})$, can be used for future analysis. The turning point detection result on this series is shown in the lower right panel of Figure 5:

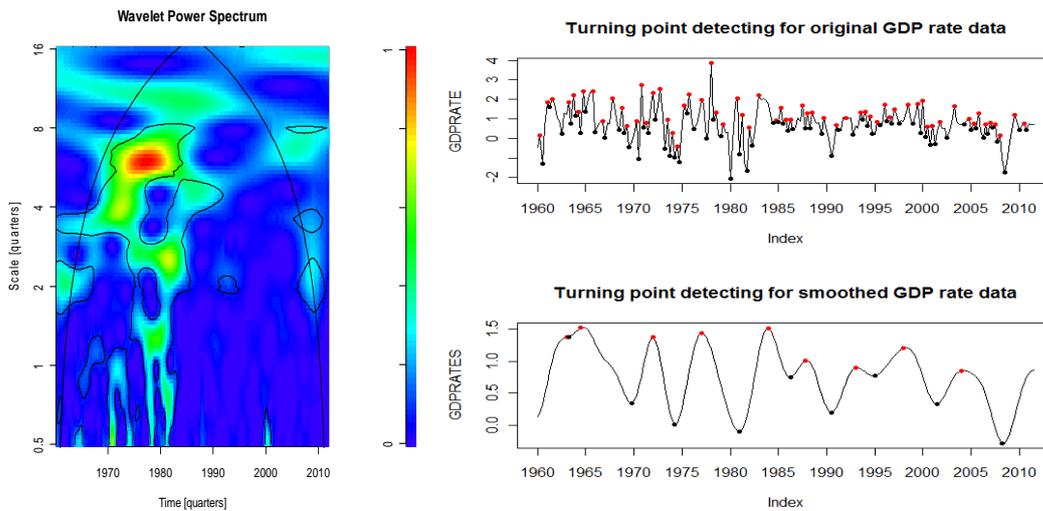


Figure 5: Wavelet sample spectrum for GDP rate data and detection result based on smoothed data

Compared to the original data, S_3 shows a much clearer cyclical behavior while preserving the same peaks and troughs as in the original data. Short-term fluctuations have been removed from the series, thus exposing the cyclical swings at frequencies that can be associated with business cycle fluctuations. The HMM can precisely detect turning points in these cycles and provides us with a reasonable estimate of acceleration cycle turning points.

In the case of PMI values, the wavelet sample spectrum in Figure 6 is equally helpful in spotting the frequencies which exhibit energy that spans over the entire sample. In the case at hand, these frequencies are associated with dynamics stretching over more than 32 months. Higher frequencies are characterized by significant activity as well, but these movements are again very specifically related to small isolated regions within the sample. It can hence be concluded that these frequencies merely add noise to the information given at lower frequencies. We use hence the fourth-scale smooths of the MRA to determine turning points with a HMM. The resulting dates, depicted by red and blue dots in the lower right panel of Figure 6, are intuitively much more appealing.

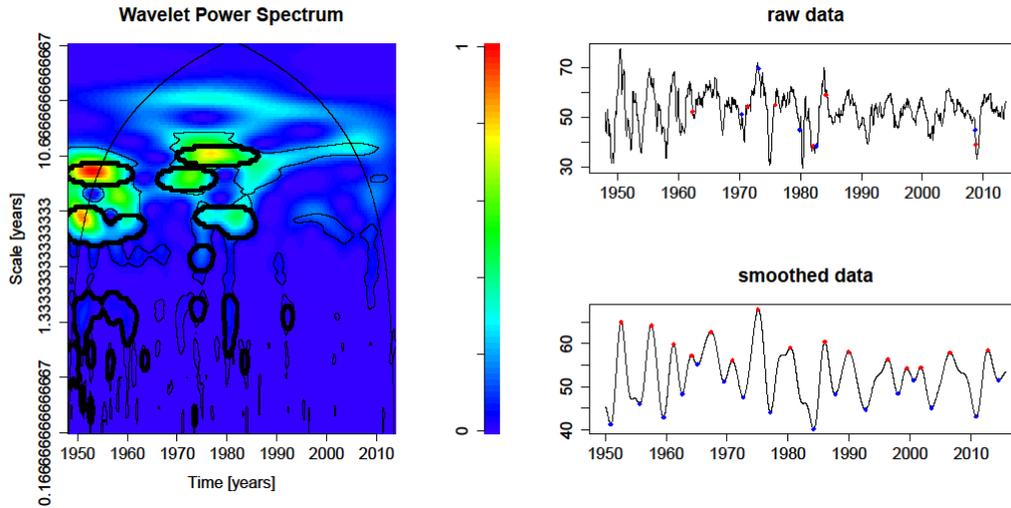


Figure 6: Wavelet sample spectrum for PMI data and detection result based on smoothed data

4.3. Comparison to change point detection methods

Since turning point detection with HMMs entails the identification of sections within a time series that have different growth rates, it is conceptually similar to change point detection methods. Hence we will assess the performance of HMMs in detecting turning points by comparing them to detection algorithms for multiple changes in mean that are applied on the first differences of our simulated series. We focus on investigating the performance of the Binary Segmentation algorithm (Edwards and Cavalli-Sforza, 1965) and the Pruned Exact Linear Time algorithm (PELT) (Killick et al., 2012). Despite being comparatively fast methods relative to HMMs, the two change point detection algorithms do not explicitly classify which regime a specific subsection in a time series belongs to. In the case of only two regimes, this problem is solved by comparing the means of the first and second subsection in order to determine whether the first identified change point in the first differences is a peak or a trough in the levels of the series. The nature of all other turning points can then be derived stepwise.

In a first instance, we apply the two considered algorithms to the raw series y_3 and calculate the correct specification ratios for the turning points of the two cyclical components in the series. Since the series is not split up into different frequencies, the correct specification ratio gives solely an indication of whether the change point

detection algorithms correctly detect a turning point at some frequency. A classification into the corresponding frequency band is, however, not possible with the raw data.

Table 4: Correct specification ratios for change point algorithms on raw data

| <i>Binary Segmentation</i> | | | | | | |
|----------------------------|----------------|-------|--------------|-------|----------------|-------|
| τ | $\sigma = 0.5$ | | $\sigma = 1$ | | $\sigma = 1.5$ | |
| | D_2 | D_3 | D_2 | D_3 | D_2 | D_3 |
| 200 | 0.199 | 0.086 | 0.202 | 0.123 | 0.190 | 0.133 |
| 500 | 0.163 | 0.074 | 0.200 | 0.119 | 0.191 | 0.136 |
| 1000 | 0.112 | 0.050 | 0.196 | 0.121 | 0.190 | 0.136 |

| <i>PELT</i> | | | | | | |
|-------------|----------------|-------|--------------|-------|----------------|-------|
| τ | $\sigma = 0.5$ | | $\sigma = 1$ | | $\sigma = 1.5$ | |
| | D_2 | D_3 | D_2 | D_3 | D_2 | D_3 |
| 200 | 0.261 | 0.125 | 0.242 | 0.176 | 0.253 | 0.225 |
| 500 | 0.262 | 0.129 | 0.243 | 0.178 | 0.254 | 0.224 |
| 1000 | 0.262 | 0.129 | 0.241 | 0.177 | 0.255 | 0.224 |

The results in table 4 reveal a fairly low correct specification rate for Binary Segmentation. The algorithm fails specifically at identifying turning points that correspond to the cyclical component with a range of 8-16 time points. Interestingly, the performance of the algorithm on these turning points improves as the magnitude of short-term volatility increases. The correct specification ratios of the PELT algorithm are on average almost 7% higher than those obtained with Binary Segmentation, but exhibit otherwise the same characteristics. Both methods are, however, by far less successful in detecting the correct turning points than HMMs.

In a next step, the two algorithms are applied directly to the second- and third-scale wavelet details in order to allow a comparison with the HMM approach under the same circumstances. As can be seen in table 5, the Binary Segmentation algorithm fails entirely to detect correct turning points when applied to wavelet details. Notably the ability to identify turning points in highly frequent cyclical components is impaired and the corresponding correct specification rates are generally plain zero. The results obtained from applying the PELT algorithm are mixed and differ in their characteristics strongly from those of Binary Segmentation. We see that the application of the PELT algorithm on wavelet details has different implications for

turning points of components at different frequencies. The correct specification ratio for turning points in the second-stage wavelet details turns out to be significantly higher when the algorithm is applied to the corresponding wavelet details. The ratio decreases, however, substantially when either the sample size or the magnitude of short-term volatility increases. The results are fundamentally different for turning point detection on the third-scale wavelet details. The performance of the PELT algorithm is seriously impaired here, leading to a correct specification ratio of only 1% with $\sigma = 0.5$. The correct specification ratio increases as the volatility magnitude gets higher, but generally remains below the ratios that were obtained from the raw data. Overall, the two algorithms prove to be inferior to HMMs, despite the promising results that the PELT algorithm provides for cyclical components with high frequency.

Table 5: Correct specification ratios for change point algorithms on wavelet details

| <i>Binary Segmentation</i> | | | | | | |
|----------------------------|----------------|-------|--------------|-------|----------------|-------|
| τ | $\sigma = 0.5$ | | $\sigma = 1$ | | $\sigma = 1.5$ | |
| | D_2 | D_3 | D_2 | D_3 | D_2 | D_3 |
| 200 | 0.001 | 0.062 | 0.009 | 0.049 | 0.019 | 0.051 |
| 500 | 0.000 | 0.001 | 0.001 | 0.004 | 0.003 | 0.010 |
| 1000 | 0.003 | 0.088 | 0.004 | 0.021 | 0.005 | 0.014 |

| <i>PELT</i> | | | | | | |
|-------------|----------------|-------|--------------|-------|----------------|-------|
| τ | $\sigma = 0.5$ | | $\sigma = 1$ | | $\sigma = 1.5$ | |
| | D_2 | D_3 | D_2 | D_3 | D_2 | D_3 |
| 200 | 0.703 | 0.011 | 0.471 | 0.069 | 0.346 | 0.109 |
| 500 | 0.626 | 0.009 | 0.389 | 0.065 | 0.288 | 0.097 |
| 1000 | 0.557 | 0.009 | 0.334 | 0.063 | 0.269 | 0.086 |

4.4. Using wavelet details to detect and eliminate outliers

In Section 3.1, we mentioned that through wavelet decomposition, the wavelet detail will retain the high-frequency information. As outliers are inherent to the signal, it is reasonable to analyze them within the wavelet detail, which is most sensitive to the local behavior of the signal. Research has been performed on wavelet outlier detection (Canan and Huzurbazar, 2002; Grané and Veiga, 2009). In this paper we set decomposition level $J = 1$ in the wavelet transform, which results in decomposition $X = D_1 + S_1$, where the wavelet detail D_1 is quite sensitive to the outliers with a

significant deviation at the outlier occurrence time and it can be used to detect and correct the outliers. We then set a threshold value θ to lower 2.5% percentile value of D_1 from standard normally distributed data. The values in D_1 which are above θ are set then to 0, while the other values are not altered, resulting in a new series of wavelet detail D_1' . Next, we reconstruct a series $X' = D_1' + S_1$ and apply it for turning point detection. The new series X' maintains the original structure of the observations but with the outlier points corrected. We now generate 1000 trials in a Monte Carlo simulation based on the same set of parameters in Section 3.2 to show how the outlier effect is eliminated after wavelet method based corrections. In addition to the misspecification ratio (MSR) in Table 2, we also measured the corrected-specification ratio (CSR), which is the percentage of actual turning points being detected.

Table 6: Comparison of misspecification and correct specification ratio

| T | <i>Polluted data</i> | | | | | | <i>Corrected data</i> | | | | | |
|--------------|----------------------|-------|-------|-------|-------|-------|-----------------------|-------|-------|-------|-------|-------|
| | MSR | | | CSR | | | MSR | | | CSR | | |
| | 2% | 4% | 6% | 2% | 4% | 6% | 2% | 4% | 6% | 2% | 4% | 6% |
| $\sigma = 3$ | | | | | | | | | | | | |
| 200 | 0.510 | 0.549 | 0.539 | 0.750 | 0.594 | 0.517 | 0.025 | 0.040 | 0.053 | 0.804 | 0.775 | 0.742 |
| 500 | 0.512 | 0.522 | 0.521 | 0.752 | 0.617 | 0.497 | 0.022 | 0.045 | 0.051 | 0.800 | 0.784 | 0.745 |
| 1000 | 0.544 | 0.510 | 0.515 | 0.766 | 0.632 | 0.507 | 0.030 | 0.040 | 0.043 | 0.803 | 0.777 | 0.749 |
| $\sigma = 5$ | | | | | | | | | | | | |
| 200 | 0.915 | 0.893 | 0.876 | 0.651 | 0.483 | 0.369 | 0.037 | 0.047 | 0.048 | 0.773 | 0.745 | 0.725 |
| 500 | 0.916 | 0.897 | 0.872 | 0.680 | 0.470 | 0.368 | 0.043 | 0.046 | 0.053 | 0.787 | 0.767 | 0.722 |
| 1000 | 0.917 | 0.904 | 0.877 | 0.668 | 0.483 | 0.366 | 0.047 | 0.050 | 0.054 | 0.787 | 0.746 | 0.712 |
| $\sigma = 7$ | | | | | | | | | | | | |
| 200 | 0.950 | 0.912 | 0.888 | 0.653 | 0.462 | 0.365 | 0.090 | 0.080 | 0.083 | 0.782 | 0.702 | 0.659 |
| 500 | 0.949 | 0.917 | 0.892 | 0.657 | 0.481 | 0.362 | 0.071 | 0.081 | 0.082 | 0.770 | 0.689 | 0.632 |
| 1000 | 0.936 | 0.924 | 0.904 | 0.664 | 0.479 | 0.368 | 0.086 | 0.079 | 0.080 | 0.764 | 0.687 | 0.630 |

Table 6 shows that for the outlier polluted data, the WSR is approximately 90% for the medium and large magnitude outliers, which will also lead to a reduction in the correct specification ratio to approximately $\frac{1}{3}$. Table 6 also indicates that the outlier influence is significantly improved after the wavelet correction, especially for the MSR, which is reduced to a maximum value of 10% in the high-magnitude outlier cases. After the wavelet correction, most outliers are detected and will lead to a higher CSR. Moreover, to increase the robustness of the HMM to the outliers, another commonly applied methodology is to Student's t distribution (instead of the Gaussian

distribution as the observation likelihood). However, this solution must overcome the difficulties in the identification of the degrees of freedom in the t distribution. Furthermore, this method concentrates on increasing the tolerance of outliers in the system but fails to detect the actual position of the outliers (*i.e.*, the locations of unusual events, such as unexpected occlusions and structural changes) in empirical applications. By applying the wavelet methods, the location of the outliers can be identified and the ratio of the correct identification is quite high:

Table 7: Correct outlier identification ratio

| T | $\sigma = 3$ | | | $\sigma = 5$ | | | $\sigma = 7$ | | |
|------|--------------|-------|-------|--------------|-------|-------|--------------|-------|-------|
| | 2% | 4% | 6% | 2% | 4% | 6% | 2% | 4% | 6% |
| 200 | 0.652 | 0.565 | 0.576 | 0.935 | 0.890 | 0.883 | 0.957 | 0.956 | 0.942 |
| 500 | 0.613 | 0.589 | 0.576 | 0.925 | 0.896 | 0.900 | 0.961 | 0.941 | 0.942 |
| 1000 | 0.638 | 0.600 | 0.599 | 0.933 | 0.914 | 0.887 | 0.958 | 0.952 | 0.936 |

5. Conclusion

The current paper primarily concentrates on improving turning point (*i.e.* peak and trough) detection in cyclical series by applying HMMs. The successful performance of HMMs is reported in many studies. However, our simulation and our empirical results show that the presence of a high degree of volatility, different frequency combinations and outliers will negatively affect performance. To address these concerns, we next applied a wavelet multi-resolution based methodology to decompose the series into different frequency bands and perform further analyses. Both the empirical example and the simulation results show that this methodology can reduce the issues raised by high volatility, differences in frequency band combinations and outliers.

6. References

- Anders, K. (1997). "Two Methods for Improving Performance of a HMM and their Application for Gene Finding", Proceedings of the 5th International Conference on Intelligent Systems for Molecular Biology 1997;5, pp.179-186.
- Barnett, V. and Lewis, T. (1994). *Outliers in statistical data*, 3rd Edition, John Wiley & Sons, Chichester.

BAUM, L.E. and Eagon, J.A. (1967). “An Inequality with Applications to Statistical Estimation for Probabilistic Functions of a Markov Process and to a Model for Ecology”, *Bulletin of the American Mathematical Society*, Vol. 73(3), pp.360-363.

BAUM, L.E. and Petrie, T. (1966). “Statistical Inference for Probabilistic Functions of Finite State Markov Chains”, *The Annals of Mathematical Statistics*, Vol. 37(6), pp.1554-1563.

BAUM, L.E. *et al.* (1970). “A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains”, *The Annals of Mathematical Statistics*, Vol. 41(1), pp.164-171.

Bellone, B. and Saint-Martin, D. (2003). “Detecting turning points with many predictors through hidden markov models”, Working paper presented in Séminaire Fourgeaud, December, 2003:3, Etudes pour la conjoncture.

Burns, A.F. and Mitchell, W.C. (1946). *Measuring Business Cycles*, New York, NBER.

Canan, B. and Huzurbazar, S. (2002). “Wavelet-Based Detection of Outliers in Time Series”, *Journal of Computational and Graphical Statistics*, Vol. 11(2), pp.311-327.

Chengalvarayan, R. (1999). “Robust energy normalization using speech/non-speech discriminator for German connected digit recognition”, Proceedings of EUROSPEECH, pp.61-64 Budapest, Hungary, September 1999, ISCA.

Edwards, A.W.F. and Cavalli-Sforza, L.L. (1965). “A method for cluster analysis”, *Biometrics*, Vol. 21, pp.362-375.

Elliott, R.J., Aggoun, L. and Moore, J.B. (1995). *Hidden Markov Models*. Springer, New York.

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.

Grané, A. and Veiga, H. (2009). “Wavelet-based detection of outliers in volatility models”, Statistics and Econometrics Working Papers, Universidad Carlos III de Madrid, Calle Madrid, Getafe, Spain.

Grossman, A. and Morlet, J. (1984). “Decomposition of Hardy functions into square integrable wavelets of constant shape”, *Society for Industrial and Applied Mathematics Journal on Mathematical Analysis*, Vol. 15, pp.732-736.

- Hamilton, J.D. (1989). "A new approach to the economic analysis of nonstationary time series and the business cycle", *Econometrica*, Vol. 57(2), pp.357-384.
- Hamilton, J. and Perez-Quiros, G. (1996). "What do the leading indicators lead?", *Journal of Business*, Vol.69, pp.27-49.
- Huang, X., Ariki, Y. and Jack, M. (1990). *Hidden Markov Models for Speech Recognition*, Edinburg university press.
- Killick, R., Fearnhead, P. and Eckley, I.A. (2012). "Optimal detection of change points with a linear computational cost", *Journal of the American Statistical Association*, Vol.107(500), pp.1590-1598.
- Kim, C.J. and Nelson, C.R. (1998). "Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching", *Review of Economics and Statistics*, Vol. 80(2), pp.188-201.
- Krolzig, H.M. (2003). "Constructing turning point chronologies with Markov-switching vector autoregressive models: the Euro-zone business cycle", in Paper presented at the Colloquium on Modern Tools for Business Cycle Analysis, Luxembourg.
- Layton, A.P. (1996). "Dating and Predicting Phase Changes in the U.S. Business Cycle", *International Journal of Forecasting*, Vol.12, pp.417-428.
- Levinson, S.E., Rabiner, L.R., and Sondhi, M.M. (1983). "An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition", *Bell System Technical Journal*, Vol. 62, pp.1035-1074.
- MacDonald, I.L., Zucchini, W. (1997). *HiddenMarkov and other models for discrete-valued time series*, Chapman and Hall, London.
- Mallat, S.G. (1989). "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", *Pattern Analysis and Machine Intelligence*, IEEE Transactions. Vol. 11(7), pp.674-693.
- Percival, D.B. and Walden, A.T. (2000). *Wavelet Methods for Time Series Analysis*, Cambridge Univ. Press.

Rabiner, L.R. (1989). "A tutorial on hidden Markov models and selected applications in speech recognition", Proceedings of the IEEE, Vol. 77(2), pp.257-286.

Stock, J., and Watson, M. (1989). "New Indexes of Coincident and Leading Economic Indicators", NBER Macroeconomics Annual, ed. by O. Blanchard, and S. Fisher, pp.352-394.

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

Wachter, M., Demuynck, K. and Compernelle, V.D. (2007). "Outlier correction for local distance measures in example based speech recognition", in Proc. ICASSP, Vol. IV, pp.433-436.