

Time series analysis for C2C conversion rate

Ke Gong^{1,2} · Yi Peng² · Yong Wang¹ · Maozeng Xu¹

Published online: 29 December 2017

© Springer Science+Business Media, LLC, part of Springer Nature 2017

Abstract The rise of online shopping by individuals in recent years has made e-commerce a crucial topic of interest in research and practice. The critical question in this domain is the extent to which online visits convert into purchases. Researchers have proposed decision models to predict consumer conversion behavior that primarily uses click-stream data, path data, panel data, and log data. This paper proposes an empirical mode decomposition (EMD) based ensemble recognition method for conversion rate (EMDER) to explore the potential pattern, business cycles in time series for conversion rate. EMDER builds on some notions, such as the database of candidate factors time series, recognition function, the recognized factor database, cycle function, and residue-trend recognition function. We collect 50 datasets from Taobao.com and find a seasonal pattern, Index of Clothing Consumer Price pattern and the long-term time series pattern with monthly data. For the daily analysis, we discover patterns in the calendar of daily fluctuation, the hesitation window, the consumers' cash flow determined pattern, the promotion day and holiday influence. A comparison between EMD and Wavelet-based method is conducted, which reveals EMD outperforms the Wavelet-based model in the deposition quality and do not have the model-selection problem. The data analysis results provide support for the proposed method, which indicates that our model enables managers to analyze online consumer purchasing behavior by a new easy approaching way, which is time series of conversion rate.

✉ Ke Gong
gks_cn@163.com

¹ School of Management, Chongqing Jiaotong University, Chongqing, People's Republic of China

² School of Management and Economics, University of Electronic Science and Technology of China, Chengdu, People's Republic of China

Keywords Time series · E-commerce · Empirical mode decomposition (EMD) · Customer-to-customer (C2C) · Data mining

1 Introduction

Online retailers are appropriately interested in conversion rate which is defined regarding the percentage of visitors to a site that ends up buying. In studying conversion behavior, some scholars have treated the conversion rate as a probability of purchasing performed by a buyer and proposed models to predict conversion rates, such as a stochastic model of conversion behavior to predict a customer's probability of purchasing [1], a choice model and Bayesian method [2], and a logit modeling [3].

In their e-commerce studies, scholars have primarily used click-stream data [4], path data [5], panel data [6], log data [7], and the like. Few studies on time series of conversion rates have examined the potential pattern through nonlinear and non-stationary time series.

The increase in the development of e-commerce platforms is noteworthy, especially in emerging global market such as China. The Taobao.com, operating in China by Alibaba Group, is an online shopping website similar to eBay and Amazon that provides consumer-to-consumer (C2C) retail in Chinese-speaking regions as well as globally [8]. This platform also includes data analysis tools for online retailers and enables them to run conversion rates' time series and to gather other information useful to their decision-making process. Time series analysis and expert systems tools are increasingly helpful to small businesses and individual entrepreneurs' decision.

Similar to the fluctuating in the price of crude oil, stocks, and transportation, the conversion rates' time series are nonlinear and non-stationary. The mechanism to convert a buyer's visits into transaction rate will lead to a fluctuation of the conversion rates' time series. Although the mechanism is very complicated, the time series is nonlinear and non-stationary, the main factors that influence the variation of conversion rate are obtainable, and the fluctuation of the conversion rates' time series is decomposable. For these type of situations, the empirical mode decomposition (EMD), proposed by Huang et al. [9], is a promising nonlinear and non-stationary data processing method. The EMD method treats the time series as fast oscillations that can overlay the slow oscillations [10].

This paper develops an EMD based ensemble recognition method for conversion rates' time series (EMDER) to explore the potential pattern, business cycles. EMDER builds on some notions, such as the database of candidate factors time series, recognition function, the recognized factor database, cycle function, and residue-trend recognition function. Also, EMDER is validated by 50 datasets through the data mining method. For the monthly analysis, we find seasonal pattern, Clothing Consumer Price Index (CCPI) pattern and the long-term time series

pattern. For the daily analysis, we find patterns in the calendar of daily fluctuation, the hesitation window, the consumers' cash flow determined pattern, the promotion day and holiday influence. We also conduct a comparison experiment between the proposed method and Wavelet-based method. The results show that EMDER outperforms the Wavelet-based model in the deposition quality and do not have the model-selection problem.

We formulate the article as follows. We first review the extant literature then describe our research methods followed by our findings. In Sect. 3, we propose the model. To validate the proposed method, we conduct some experiments and comparison. We finally present our discussion and conclusions.

2 Literature review

2.1 Conversion behavior modeling

Moe and Fader [1] develop a model of conversion behavior to predict customers probability of purchasing based on the history of visits and purchase. Sismeiro and Bucklin [2] propose a model of online buying using clickstream data from a website that sells cars. Hui et al. [11] propose an individual-level behavioral model that captures the aggregate preorder/post-release sales of motion picture DVDs. Mintz et al. [12] formulate a discrete choice model and perform formal model comparisons to distinguish among several possible dependence structures. Wang et al. [13] propose a novel business intelligence approach to estimating consideration probabilities with a two-step procedure. Xu et al. [14] study the effects of various types of online advertisements on purchase conversion by capturing the dynamic interactions among ad clicks themselves.

Unlike the notions mentioned in the literature, the conversion rate is to describe the proportion of people who buy a product to people who visit the site. Although authors have conducted lots of research, literature focus on conversion rates is rare.

2.2 Data type used in related research and purposes

In recent years, many scholars use different kinds of data to conduct various purposes of studies for conversion rates of e-commerce.

Click-stream data are the electronic records of Internet usage recorded by company web servers [15], which are popularly used tools for conversion behavior-related research and online shopping behavior in the past decade. Park and Chung [16] use clickstream data to predict e-travelers' purchasing behaviors. Poel and Buckinx [3] suggest the use of click-stream data and propose a cost-effective methodology for the prediction of the demographic website visitor profiles. Olbrich and Holsing [17] examine which factors were significant for predicting purchasing behavior within social shopping communities by analyzing clickstream data. Rutz

and Bucklin [18] investigate how exposure to affect the choices users make of brand-specific pages to view on a website by using individual-level clickstream data.

Besides the click-stream data, the studies of e-commerce employ some other data types. Hui et al. [5] initiate a formal definition of a path (in a marketing context) and derive a unifying framework that allows them to classify different kinds of paths. De et al. [7] measure consumers' usage of website technologies by analyzing server log data. Lin et al. [19] use panel data from comScore Media Metrix that consist of detailed click-stream data to explore the impact of stickiness on conversion. Langer et al. [6] estimate a dynamic discrete choice model with panel data.

However, as an essential type of data, time series are relatively few studied by scholars for this issue.

2.3 Nonlinear and non-stationary decomposition techniques

The time series for conversion rates of online retail are nonlinear and non-stationary time series. To our knowledge, the extant research on the conversion rate of e-commerce does not use time series data. The purpose of this article is to decompose these time series and recognize some cycles and factors that contribute to the variation of the time series to help to make a decision. This problem is called frequency domain analysis (FDA).

The FDA based decomposition techniques include five groups, which are Fourier transform, Wavelet decomposition, EMD, Singular spectrum analysis, and Filtering analysis [20].

On the application domains of the five FDA methods, Fourier Transform, Wavelet decomposition, Singular spectrum analysis, and Filtering analysis are mainly applied to short-term forecasting [20]. Fourier transform performs poorly when time series are non-stationary [21]. Filtering analysis has H-P filter has spurious effects [22]. Singular spectrum analysis has the disadvantage that no general rule for selecting of parameters [23].

Nowadays, EMD and Wavelet decomposition are the most widely used decomposition techniques in the applications of time series analysis [20].

Wavelet transformation is a parametric method which is widely utilized for analyzing these type of data, such as crude oil prices [24], benchmark oil prices [25], high-frequency financial data mining [26]. However, Wavelet methods are unable to obtain fine resolutions in both time and frequency domains in the chorus and have difficulty in analyzing large size data [27].

EMD is another non-parametric methodology for nonlinear and non-stationary time series. It is proven to be effective for analyzing nonlinear and non-stationary time series, which has been applied in fields, such as structural health monitoring [28, 29], prediction prices of crude oil [10, 30–32], prediction of foreign exchange rates [33], measurement of business cycle [34], estimation of electricity prices [35], passenger flow [36], and so forth. The timescale suitable for EMD is short-term, mid-term, and long-term.

Consumer behaviors are the primary driving forces that lead to the variation of conversion rates. Besides these factors of consumer behaviors, random noise exists.

Each factor should have its variation; here it means the time series. Furthermore, the joint acting force of these factors and random noise makes the time series of conversion rate nonlinear and non-stationary. Therefore, we utilize the feature of EMD to help decompose the conversion rate time series.

2.4 Empirical mode decomposition

EMD, proposed by Huang et al. [9], is a form of adaptive time series decomposition technique. It uses the Hilbert-Huang transform (HHT) for nonlinear, non-stationary and also asymmetric cycles time series data.

The fundamental principle of EMD is to decompose a time series into a sum of oscillatory functions, intrinsic mode functions (IMFs). The IMFs must meet two conditions. The detail description of EMD model can be found at Huang et al. [9]. The following shows the detailed algorithm for EMD [9]:

Step 1: Identify all the maxima and minima of time series $x(t)$;

Step 2: Generate its upper and lower envelopes, $e_{\min}(t)$ and $e_{\max}(t)$, with cubic spline interpolation.

Step 4: Extract the average from the time series and define the difference of $x(t)$ and $m(t)$ as $d(t)$, defined by $d(t) = x(t) - m(t)$.

Step 3: Calculate the point-by-point mean value $m(t)$ from upper and lower envelopes, $m(t) = (e_{\min}(t) + e_{\max}(t))/2$.

Step 5: Check the properties of $d(t)$. If it is an IMF, denote $d(t)$ as the i th IMF and replace $x(t)$ with the residue $r(t) = x(t) - d(t)$. The i th IMF is often denoted as $c_i(t)$, and the i is called its index; If it is not, replace $x(t)$ with $d(t)$.

Repeat steps 1–5 until the residue satisfies some stopping criterion.

Using the above algorithm, we can decompose the original time series data $x(t)$ into n modes and a residue. The first component has the highest frequency, which represents the shortest period variants in the time series data, whereas the residue represents the lowest frequency. Consequently, the set of IMFs is derived from high frequency to low frequency [36].

3 Model

This paper aims to develop the EMD based ensemble recognition method for conversion rate (EMDER) to find the factors that influence the variation of the conversion rate, latent cycles, and trend to help to make decisions. Figure 1 is the systematic diagram of the proposed model, which is accomplished by Algorithm 1. Table 1 provides a list of notations

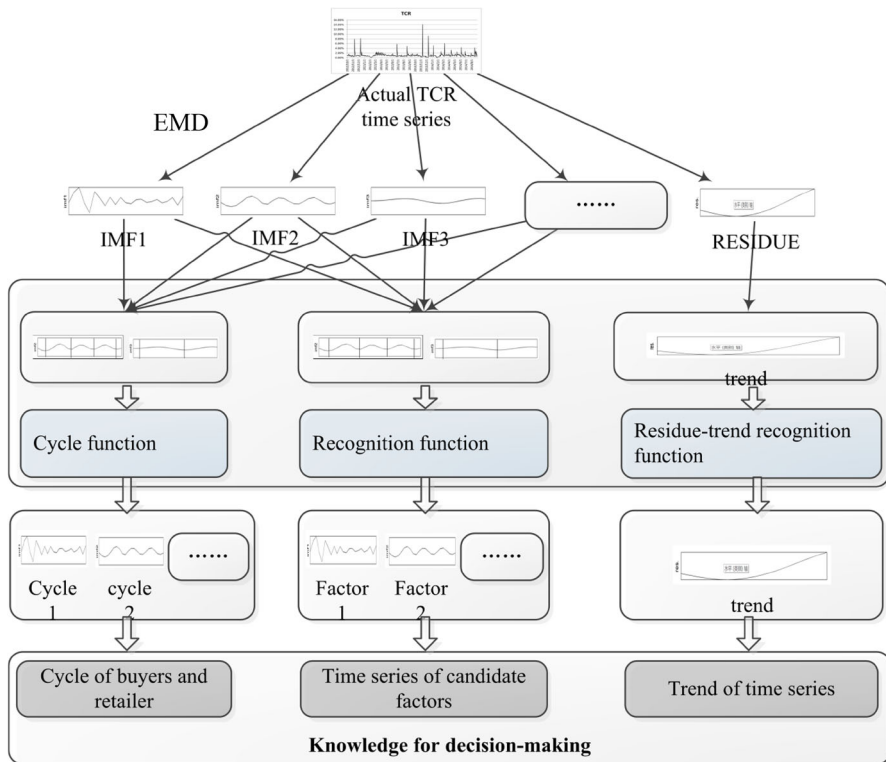


Fig. 1 Schematic diagram of the proposed model

Table 1 Summary of Notations

Notation	Definition
y	An actual time series of the conversion rate vector
m	The number of IMFs which is decomposed by EMD method from y
IMF	$\mathbf{IMF} = \bigcup_1^m imf_j$, where $j = 1, 2, \dots, m$
r	The residue of the results of decomposed y
F	The candidate factors database
F_l	One of the candidate factors in F , where $l = 1, 2, \dots, q$
F_{t_l}	The time series vector of F_l
$cycle(\cdot)$	The cycle function
$recognition(\cdot, \cdot)$	The recognition function, which is a mapping $f : N \times M \rightarrow \{true, false\}$, N and M are time series vectors
T	The cycle
t	The time of the cycle T
imf_i^t	The value of i th IMF at t
RF	Recognized factor database

Algorithm 1. The algorithm of proposed model

Input: The original time series of conversion rate.

Step 1: Use EMD to decompose the original time series of conversion rate y into m IMFs imf_j and residue r .

Step 2: Prepare the database for candidate factors time series \mathbf{F} .

Step 3: For each imf_j in \mathbf{IMF}

Step 3.1: If $recognition(F_{t_l}, imf_j) = true$ then

Put F_{t_l} into \mathbf{RF} .

Remove F_{t_l} from \mathbf{F} .

Step 3.2: Else

Step 3.2.1: If exists $cycle(imf_j)$ then

Put $cycle(imf_j)$ into \mathbf{RF} .

Step 3.2.2: Else

Mark imf_j as noise.

Step 4: If $recognition(F_{t_l}, r) = true$ then

Put F_{t_l} into \mathbf{RF} .

Else

Put r as trend into \mathbf{RF} .

Step 5: Identify the trend of the residue by residue-trend recognition function.

Output: \mathbf{RF} .

End.

3.1 Time series decomposition by EMD

EMD used in our model is to decompose the original time series into several IMFs, which will be input for further recognition.

3.2 Candidate factors time series database

The database for candidate factors time series \mathbf{F} is the input for the recognition function to find and verify these factors to explain the variety of IMFs, whose establishment is based on the knowledge of online retail. F_l is one of the candidate factors in \mathbf{F} , where $l = 1, 2, \dots, q$. We denote F_{t_l} as the time series vector of F_l .

3.3 Recognized factor database

Recognized factor database \mathbf{RF} is a collection of verified knowledge whose elements are identified from the database for candidate factors time series, which are the output of this algorithm illustrating the variation of the conversion rate to mitigate the perceiving uncertainty of conversion behavior of the consumer.

3.4 Cycle function

There are some cycles which are incurred by human's activities, such as a week, season, and month, and the natural time cycle, such as business cycle, physiology cycle. $cycle(\cdot)$ is proposed to uncover the latent cycles of the IMFs, which is a mapping $f: N \rightarrow \{\text{not exists}, I\}$ and N is a time series vector, and to find the minimum time interval T such that $|f(x+T) - f(x)| < \varepsilon$. If the time series N has a cycle, the function returns the cycle. Otherwise, this function return 'not exists'. We propose an algorithm to find T in Algorithm 2

Algorithm 2. The algorithm for finding cycle

```

Input the time series  $N$ 
For period = 3 to  $|N|$ 
  Num=0;
  For i=0 to  $|N|$ 
    If period + i <  $|N|$ 
      Error +=  $|N_i - N_{i+period}|$ 
      Num++
    End if
  End for
  Error = error/num;
   $period_{\min}$  is the period s.t. Error = min(Error);
End for
Return  $period_{\min}$ 

```

3.5 Residue-trend recognition function

The residue decomposed by EMD is the trend of the whole time series [37, 38], which is an important feature of time series. We design a residue-trend recognition function to identify the trend of residue. Let R be a residue time series, ma be the average window, and R^{ma} be the moving average vector. This function returns a token like ' $\searrow \uparrow \nearrow (\searrow 76\%)$ ' to describe the trend, which stands for going down and going up and the main trend is going down by 76%.

3.6 Recognition function

The recognition function $recognition(\cdot, \cdot)$ is to discover the similarity between IMFs and candidate factors time series, where the mapping of $recognition(\cdot, \cdot)$ is $f: N \times M \rightarrow \{true, false\}$, N and M are time series vectors. It returns the Boolean value σ that indicates whether these time series have a significant correlation or dependency. For the time series F_{t_l} and imf_j , if $recognition(F_{t_l}, imf_j)$ equals to true, we say F_{t_l} recognizes imf_j .

Because the input time series vectors of recognition function may have different characteristics, the recognition function can have some forms. Since most of the

time series are numerical, we leverage Pearson product-moment correlation and Kendall rank correlation to construct the recognition function as following.

$$\text{recognition}(Ft_1, \text{imf}_j) = \begin{cases} \text{true if } \rho > \Delta \text{ and } \tau > \Delta \\ \quad \text{and the t - test of } \rho \text{ and } \tau \text{ are significant} \\ \text{false other} \end{cases} \quad (1)$$

where ρ is the Pearson product-moment correlation coefficient, and τ is the Kendall rank correlation coefficient. Empirically, for social science, if the correlation coefficient is over 0.3 we say objects are correlated; if the correlation coefficient is less than 0.3 we say objects are a weak correlation. Therefore, we let $\Delta=0.3$ in this context.

4 Empirical studies and experiments

4.1 Experiment environments

We leverage Matlab, version R2011b, to perform the EMD and Wavelet experiments. SPSS is employed to calculate Pearson and Kendall correlation coefficient [32]. We use Eclipse Mars.2 to develop the cycle function and the residue-trend recognition function with JAVA.

4.2 Data

We collect Taobao conversion rate (TCR) data from two online stores of www.Taobao.com officially. One is a woman's clothing store, and the other is a milk powder store. Therefore, these data need not cleansing.

4.3 Case study 1: monthly data for a woman's clothing store

4.3.1 Dataset 1: monthly data for a clothing store

We use this a TCR monthly dataset to investigate the factors and cycles that affect the time variation of the online clothing store from a longer timescale. See Fig. 2.

4.3.2 IMF

From high frequency to low frequency, there are three IMFs decomposed from the original time series, and the last decomposed time series is the residue, see Fig. 3.

4.3.3 IMF statistics

We use the mean period of each IMF, the correlation between each IMF and the original data series, and the variance and variance percentage of each IMF to

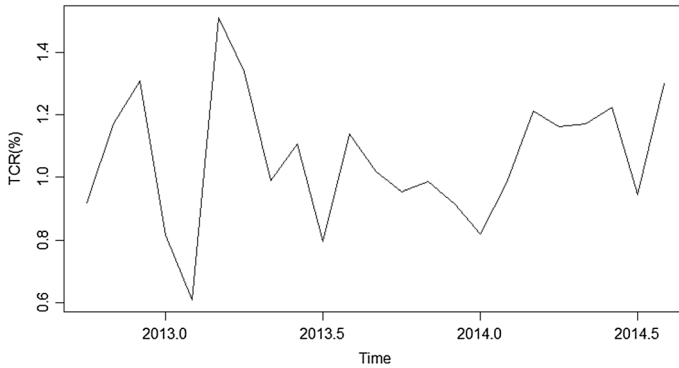


Fig. 2 Monthly TCR data of a woman's clothing store

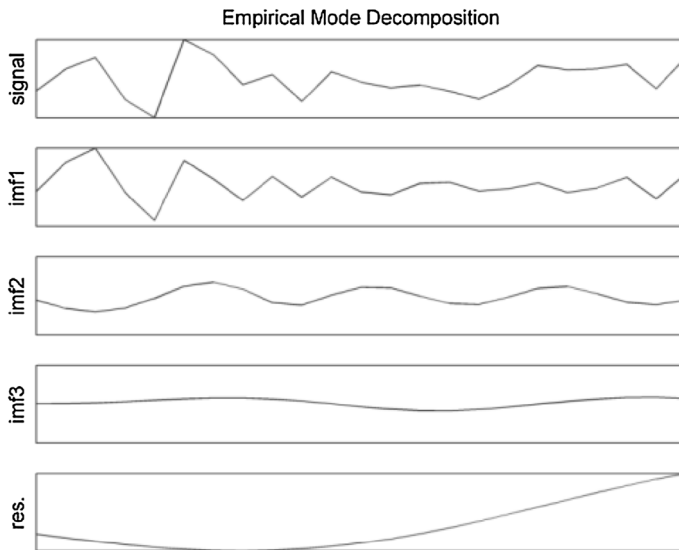


Fig. 3 IMFs and the residue for TCR monthly data from Oct. 2012 to Aug. 2014

analyze IMFs. These measurements were usually used by EMD related literature [10, 32, 36].

IMF 1 to IMF 3 and residue positively correlate with the original time series regarding Pearson correlation coefficient. For the Kendall correlation, IMFs 1 and 2 are significant at 0.01 and 0.05 levels respectively. All of them are positive, and Pearson and Kendall correlation coefficient are consistent.

Variance as a percentage of observed acts as the measurement of the variability of each IMF and residue concerning observed. The variance as a percentage of IMF1 is the biggest, up to 79.55%, which explains the most variability of TCR monthly data. Moreover, from IMF1 to IMF3 and residue, the variances as a percentage of observed are decreasing. From IMF1 to IMF3, the frequency of them

Table 2 Measures of IMFs and the residue for TCR Monthly data from Oct. 2012 to Aug. 2014 derived through EMD

	Mean period	Pearson correlation	Kendall correlation	Variance	Variance as a % of observed	Variance as a % of (Σ IMFs + residue)
Observed				0.044		
IMF1	0.028	0.798	0.549**	0.035	79.55	68.63
IMF2	- 0.021	0.320	0.312*	0.011	25.00	21.57
IMF3	0.018	0.350	0.265	0.003	6.82	5.88
residue	1.037	0.186	0.091	0.002	4.55	3.92
Sum					115.91	100

*Correlation is significant at the 0.05 level (2-tailed)

**Correlation is significant at the 0.01 level (2-tailed)

is falling [38], we can conclude that the TCR monthly data is highly nonlinear with high-frequency variation (Table 2).

4.3.4 Result and management insight

The industry suggested that some macroscopic factors, such as the clothing consumer price index (CCPI), would influence the fluctuating of the online clothing sales. Therefore, CCPI is put into F , which is collected from National Bureau of Statistics of PRC. Since other candidate factors are not significant, we do not give them in this article. Table 3 is the result of EMDER. IMF1 is the highest frequency component of EMD. Former literature [36] treats the first few components of IMFs as the highly time variations or noise. Hence, we will mark IMF1 as noise as well.

4.3.4.1 IMF2: the pattern of seasonal fluctuation From Table 3, IMF2 has a 6-month cycle. Figure 4 shows the wave of IMF2, where the x-axis represents the month. The reason why there is a 6-month cycle is due to one of the intrinsic characteristics of clothing consumption. As China is a monsoon climate, with four distinct seasons, people used to buy new clothes in April for the spring season and buy new clothes in September for the coming autumn and winter. While the TCR of the rest months drops subsequently due to the release of consumption, the two lower points always appear in July and January. Therefore, IMF2 is the pattern of seasonal fluctuation.

4.3.4.2 IMF3: Pattern of clothing consumer price index As is shown in Table 3, CCPI and IMF3 are recognized, whose Pearson and Kendall correlation are over 0.3, see Table 4 and Fig. 5. This positive correlation reveals an interesting phenomenon: the more expensive the clothing is, the higher TCR occurs, i.e., people are more likely to buy clothing online to save money when the clothing becomes costly.

4.3.4.3 Residue As is shown in Table 3, the trend of residue is ‘ $\searrow \uparrow \nearrow$ ’ with 82% rising trend, which means the longtime trend of TCR of this clothing store is

Table 3 The result of EMDER for case study 1

	IMF2	IMF3	Residue
Cycle	6	No	No
Recognized factor	No	CCPI	No
Trend of residue	↘ ↗ (↗ 82%)		

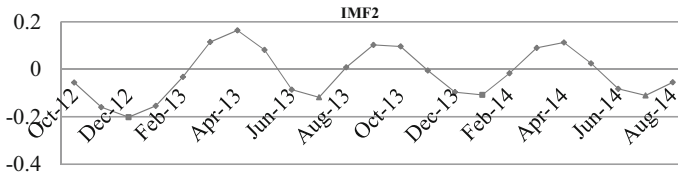


Fig. 4 IMF2 for TCR monthly data from Oct. 2012 to Aug. 2014 derived through EMD

Table 4 Correlation analysis between CCPI and IMFs

	Pearson correlation	Kendall correlation
IMF1	- 0.063	0.042
IMF2	- 0.002	- 0.034
IMF3	0.371	0.328*

*Correlation is significant at the 0.05 level (2-tailed)

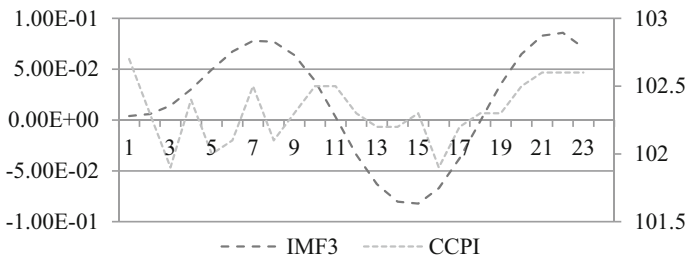


Fig. 5 CCPI and IMF3

increasing. According to the industry, the longtime trend of TCR reflects the service performance of this store like logistic, customer service, quality of its products.

4.4 Case study 2: daily data for a clothing store

4.4.1 Dataset 2: daily data for a clothing store

This data set is a time series of daily TCR containing 700 records, which are grouped into 23 datasets by month, see Fig. 6. We employ them to study the factors and cycles that within a month.

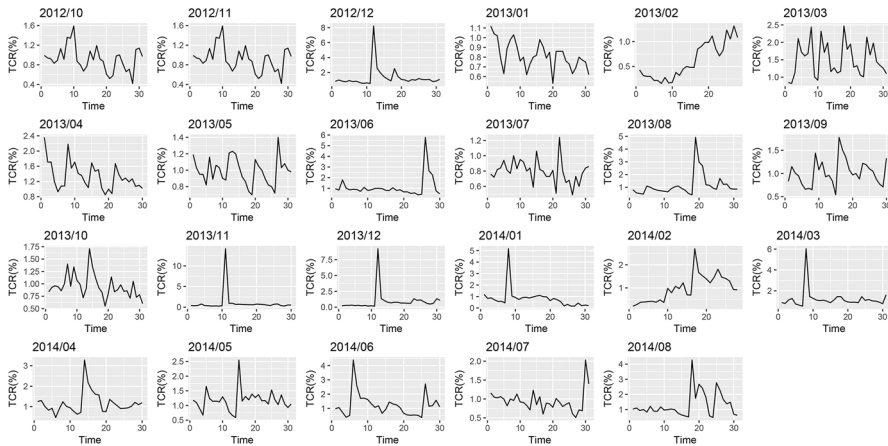


Fig. 6 Daily TCR data of a woman's clothing store for 23 months

4.4.2 IMF

We use EMD to decompose the 23-month daily data sets, see Fig. 7.

4.4.3 IMF statistics

Accordingly, the same measurement is conducted to evaluate the correlation between daily dataset of Oct. 2012 and IMFs, residue. The results show that both are significant and positive, see Table 5. Since the analysis process is identical to monthly data, we will not extend this issue. For the rest 22 months, the statistics of each daily data in the corresponding month are mostly identical, so we will not give these statistics for concision.

4.4.4 Result and management insight

Table 6 is the product of EMDER. IMF1 is the highest frequency component of EMD. Similar to the analysis of monthly data, we can consider it as noise.

4.4.4.1 IMF2: a weekly cycle As is shown in Table 6, IMF2 of these datasets are 7 or close to 7, which may be 6 or 8. The number of these datasets is 16. Furthermore, one can find that the curves are varying with calendar day apparently, shown in Fig. 7. Therefore, IMF2 is a weekly cycle of the customers' buying behavior.

4.4.4.2 IMF3: a half month cycle From Table 6, there are 7 months, such as Jan-13, May-13 have 14 and 15 days cycles, which is very close to a half month cycle.

4.4.4.3 Hesitation window Besides, to discover the buying habit of customers, we study the frequency of wave crest and trough of each curve. The result indicates that wave troughs always occur on Sunday, about 51.4%, while the wave crests always

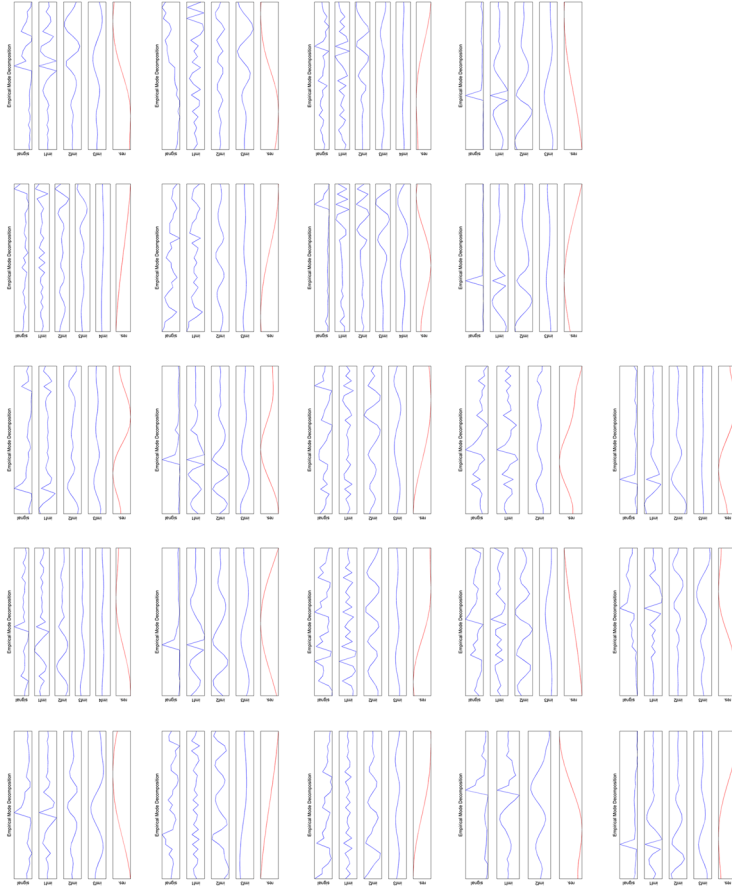


Fig. 7 IMFs and the residue for TCR Daily data 23 months of a clothing store

Table 5 Measures of IMFs and the residue for TCR Daily data from Oct. 1, 2012, to Oct. 31, 2012, derived from EMD

	Mean Period	Pearson Correlation	Kendall Correlation	Variance	Variance as a % of Observed	Variance as a % of (Σ IMFs + residue)
Observed				0.253		
IMF1	- 0.006	0.451	0.325*	0.112	44.27	21.45
IMF2	- 0.033	0.601	0.381**	0.223	88.32	42.80
IMF3	0.056	0.396	0.231	0.116	45.78	22.19
residue	0.901	0.317	0.188	0.071	27.97	13.56
Sum					206.34	100

*Correlation is significant at the 0.05 level (2-tailed)

**Correlation is significant at the 0.01 level (2-tailed)

happen on Wednesday, Tuesday, and Thursday, approximately 91.2%, see Fig. 8. This phenomenon indicates that the page views are more likely convert to real deals in the upper half week (Wednesday, Tuesday, and Thursday) than the second half of the week (Sunday and Saturday). Overall, people more likely go shopping and spend time with family on the weekend. After the impact of the weekend, they keep on shopping online, which leads to the conversion rate varying as same as page views. On the other hand, according to an interview with Taobao industry, experts say that the top calendar days with the top page views are Monday, Tuesday, and Wednesday. The evidence implies that customers usually review the product in their wish lists or carts in those three days. After some hesitating, they decide to buy products, which lead to Wednesday has the most wave crests due to the hesitation window.

4.4.4.4 Residue Table 6 reveals that 9 months have a descending trend (over 70% is descending), such as Oct. 2012, Jan. 2013, Mar. 2013 and so forth. The tendency means that the TCR is higher at the beginning of the month, and it decreases until the end of the month. As is known, most of the Chinese people get their monthly salary at the beginning of the month. Meanwhile, they have an approximate spending on the house holding such as clothing every month. Consequently, when they get the salary at the beginning of the month, their consuming desire is much more than the end of the month, which leads to a higher TCR and gradually extinguished until the end of the month. Therefore, the residues of these months reflect the consumers' cash flow determined pattern.

Besides the 9 months, 13 months have exceptions. Why there are exceptions? We list the possible reason for these exceptions in Table 7. It reveals that the primary reasons are the big sale days occur in the middle or the end of the month, which lead to the wave crest of a trend move to middle or end of the month. Also, the effects of the Chinese new year will also make the wave crest of the trend move to the middle or the end of the month. It is because Chinese people will have a 1-week holiday in the Chinese new year when they will spend their holidays and reduce their online shopping.

Therefore, the trend is the joint influence of monthly salary cycle, the promotion day and holiday. If there is neither promotion day nor holiday, the tendency is always a descent. Otherwise, the wave crest will also move.

Table 6 The result of EMDER for case study 2

Dataset	Cycle		Data set	Cycle	Residue trend		Residue trend
	IMF2	IMF3			IMF3	IMF4	
Oct-12	7	6	Oct-13	8	↘ (↘100%)	↘ (↘100%)	↗ ↘ (↘72%)
Nov-12	10	14	Nov-13	9	↗ ↘ (↘58%)	↗ ↘ (↘58%)	↗ ↘ (↘75%)
Dec-12	6	12	Dec-13	10	↗ ↘ ↗ (↘52%)	↗ ↘ ↗ (↘52%)	↗ ↘ (↗92%)
Jan-13	8	14	Jan-14	8	↘ (↘100%)	↘ (↘100%)	↗ ↘ ↗ (↘80%)
Feb-13	13	9	Feb-14	8	↗ ↗ ↗ (↗82%)	↗ ↗ ↗ (↗82%)	↗ ↘ ↗ (↗68%)
Mar-13	7	12	Mar-14	10	↗ ↗ (↘72%)	↗ ↗ (↘72%)	↗ ↘ ↗ ↗ (↘56%)
Apr-13	7	12	Apr-14	7	↗ ↗ (↘96%)	↗ ↗ (↘96%)	↗ ↘ (↗71%)
May-13	7	15	May-14	9	↗ ↗ (↘76%)	↗ ↗ (↘76%)	↗ ↘ (↗64%)
Jun-13	6	7	Jun-14	8	↗ ↗ ↗ (↗58%)	↗ ↗ ↗ (↗58%)	↗ ↘ ↗ ↗ (↗54%)
Jul-13	6	7	Jul-14	6	↗ ↘ (↘76%)	↗ ↘ (↘76%)	↘ (↘100%)
Aug-13	14	-	Aug-14	8	↗ ↗ (↗76%)	↗ ↗ (↗76%)	↘ ↗ ↗ ↘ (↗76%)
Sep-13	7	6			↗ (↗100%)	↗ (↗100%)	

‘-’ means no such IMF in the month

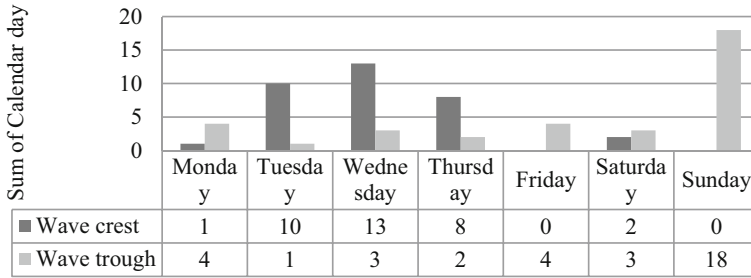


Fig. 8 Frequency of wave crest and trough for IMF2 curve

4.5 Case study 3: daily data for a clothing store after smoothing

There are some data points whose values over 3%. The average value for the TCR of the 23 months is 1.06%. Therefore, these data points may be anomaly data. To investigate whether these data have any effect on the EMDER model, we use the average of neighbor data to smooth those data points. We use EMD to decompose the 23-month daily data for a clothing store. For briefness, we do not give the IMFs of these 23 datasets and the IMF statistics.

4.5.1 Dataset 3: daily data for a clothing store after smoothing

The dataset 2 has some points that have conversion rate over 3%, which can be deemed the anomaly data. We replace the value of these data point with the average value of neighboring data points, which is shown in Fig. 9.

Table 7 The reason for the exception of each month

#	Month	Reason
1	2012/11	Big sale day on Nov. 11th
2	2012/12	Big sale day on Dec. 12th
3	2013/2	Chinese new year on Feb. 10th
4	2013/6	Big sale day on Jun. 26th
5	2013/8	Big sale day on Aug. 19th
6	2013/9	Big sale day on Sep. 16th
7	2013/10	Big sale day on Oct. 14th
8	2013/12	Big sale day on Dec. 12th
9	2014/2	Chinese new year on Jan. 30th
10	2014/3	Big sale day on Mar. 8th, women’s day
11	2014/4	Big sale day on Apr. 14th
12	2014/5	Big sale day on May. 15th
13	2014/6	Big sale day on Jun. 7th and Jun. 26th

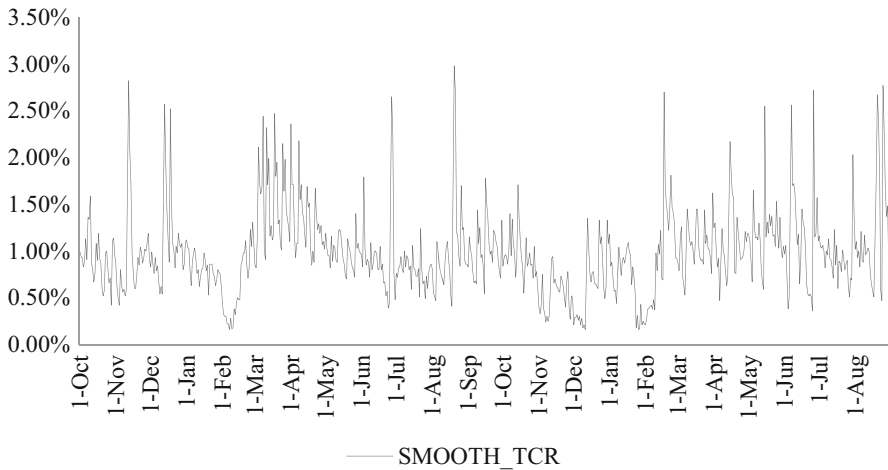


Fig. 9 Daily TCR data after smoothing of a woman's clothing store for 700 days

4.5.2 Result and comparison with case study 2

Table 8 is the results of EMDER for case study 3, and Table 9 gives the results of EMDER comparison between the original datasets and the smoothed datasets.

From these tables, one can see that 15 datasets have an IMF2 whose cycles are 7, 6, or 8. This figure is very close to the results of case study 2, which are 16.

For IMF3, after smoothing the possible anomaly data, the half-month pattern for IMF3 changes slightly. The number of datasets that have 14 or 15 days cycle increase to 8, while the results of original data are 7.

As for the residue trends, 13 datasets remains unchanged. Five datasets change the percentage of the primary trend. Only five datasets shift the direction of the primary trend.

This evidence reveals that EMDER has some fault-tolerance ability for possible anomaly data. The influences of these possible anomaly data are absorbed by the EMD to IMF1 as noise.

4.6 Case study 4: daily data for a milk powder store

4.6.1 Dataset 4: monthly data for a milk powder store

To avoid losing generalization, we also collect 3 months, July, September, and October, of daily data from a milk powder store of Taobao.com, see Fig. 10. These products have different consumers, which leads to a different time variation.

4.6.2 IMF

Figure 11 is the EMD results calculated by EMD.

Table 8 The result of EMDER for case study 3

Dataset	Cycle		Data set	Residue trend		Cycle		Residue trend		
	IMF2	IMF3		IMF4	IMF2	IMF3	IMF4	IMF2	IMF3	IMF4
Oct-12	7	6	Oct-13	↘ (↘100%)	-	-	8	↗ ↘ (↘72%)	-	-
Nov-12	10	14	Nov-13	↗ ↘ (↗58%)	-	-	8	↗ (↗100%)	6	↗ (↗100%)
Dec-12	12	15	Dec-13	↗ ↘ (↗64%)	-	-	8	↗ (↗100%)	-	-
Jan-13	8	14	Jan-14	↘ (↘100%)	-	-	8	↘ (↘100%)	-	-
Feb-13	13	9	Feb-14	↘ ↗ (↗82%)	-	-	8	↗ ↘ (↗68%)	-	-
Mar-13	7	12	Mar-14	↘ ↗ (↘72%)	-	-	8	↗ ↘ (↗92%)	-	-
Apr-13	7	12	Apr-14	↘ ↗ (↘96%)	-	-	7	↗ ↘ (↗75%)	-	-
May-13	7	15	May-14	↘ ↗ (↘76%)	-	-	9	↗ ↘ (↗64%)	14	-
Jun-13	10	14	Jun-14	↘ ↗ (↗58%)	-	-	10	↘ (↘100%)	-	-
Jul-13	6	7	Jul-14	↗ ↘ (↘76%)	13	-	6	↘ (↘100%)	6	-
Aug-13	12	15	Aug-14	↗ (↗100%)	-	-	9	↗ ↘ (↗88%)	-	-
Sep-13	7	6		↗ (↗100%)	-	-				

'-' means no such IMF in the month

Table 9 The result of EMDER comparison for case study 2 and 3

Dataset	Original data				Data after smoothing				Diff.
	Cycle		Residue trend		Cycle		Residue trend		
	IMF2	IMF3	IMF4		IMF2	IMF3	IMF4		
12-Oct	7	6	-	↘ (↘100%)	7	6	-	↘ (↘100%)	r
12-Nov	10	14	-	↗ (↗58%)	10	14	-	↗ (↗58%)	23r
12-Dec	6	12	-	↗ (↗52%)	12	15	-	↗ (↗64%)	
13-Jan	8	14	-	↘ (↘100%)	8	14	-	↘ (↘100%)	
13-Feb	13	9	-	↗ (↗82%)	13	9	-	↗ (↗82%)	
13-Mar	7	12	-	↗ (↗72%)	7	12	-	↗ (↗72%)	
13-Apr	7	12	-	↗ (↗96%)	7	12	-	↗ (↗96%)	
13-May	7	15	-	↗ (↗76%)	7	15	-	↗ (↗76%)	
13-Jun	6	7	13	↗ (↗58%)	10	14	-	↗ (↗58%)	234
13-Jul	6	7	13	↗ (↗76%)	6	7	13	↗ (↗76%)	
13-Aug	14	-	-	↗ (↗76%)	12	15	-	↗ (↗100%)	23r
13-Sep	7	6	-	↗ (↗100%)	7	6	-	↗ (↗100%)	
13-Oct	8	-	-	↗ (↗72%)	8	-	-	↗ (↗72%)	
13-Nov	9	14	-	↗ (↗75%)	8	11	6	↗ (↗100%)	234r
13-Dec	10	6	-	↗ (↗92%)	8	11	-	↗ (↗100%)	23r
14-Jan	8	12	-	↗ (↗80%)	8	6	-	↘ (↘100%)	3r
14-Feb	8	11	-	↗ (↗68%)	8	11	-	↗ (↗68%)	
14-Mar	10	15	-	↗ (↗56%)	8	15	-	↗ (↗92%)	2r
14-Apr	7	14	-	↗ (↗71%)	7	14	-	↗ (↗75%)	r
14-May	9	11	14	↗ (↗64%)	9	11	14	↗ (↗64%)	
14-Jun	8	10	-	↗ (↗54%)	10	14	-	↘ (↘100%)	23r
14-Jul	6	8	6	↘ (↘100%)	6	8	6	↘ (↘100%)	
14-Aug	8	14	-	↗ (↗76%)	9	15	-	↗ (↗88%)	23r

4.6.3 IMF statistics

The IMF statistics of the three datasets are significant and positive. For a brief, we will not give these statistics for concision.

4.6.4 Result and management insight

The result of EMDER is given in Table 10. IMF1 is the highest frequency component of EMD. Similar to the analysis of monthly data, we can consider it as noise.

4.6.4.1 IMF2: a weekly cycle One can observe from Table 10 that the cycles of IMF2 for these datasets are 7 or 6 days. Therefore, IMF2 is a weekly cycle of the customers' buying behavior. This result is similar to the case study 2, which reflects consumers who buy clothing and consumers who buy milk powder may have the same consumption habit.

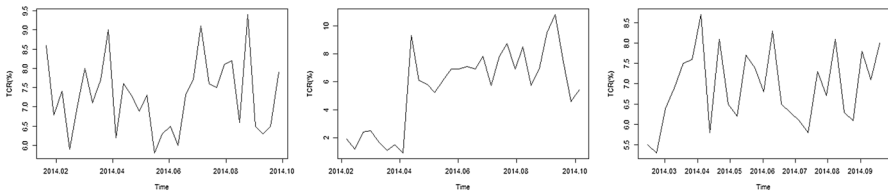


Fig. 10 Daily TCR data of a milk powder store for 3 months

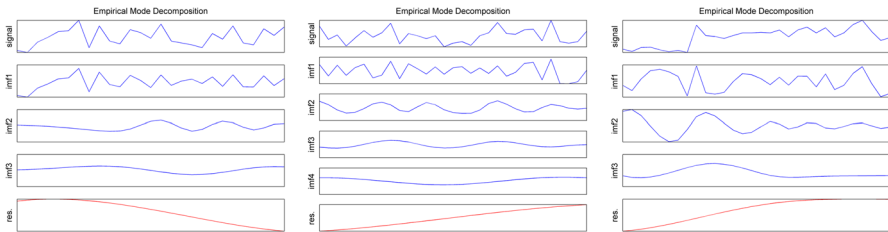


Fig. 11 IMFs and the residue for TCR Daily data from July, September, and October of 2014

Table 10 The result of EMDER for case study 3

	Cycle			Residue trend
	IMF2	IMF3	IMF4	
Jul-14	7	12	no	↗ (↗100%)
Sept-14	7	16	—	↗↘ (↗92%)
Oct-14	6	17		↘ (↘100%)

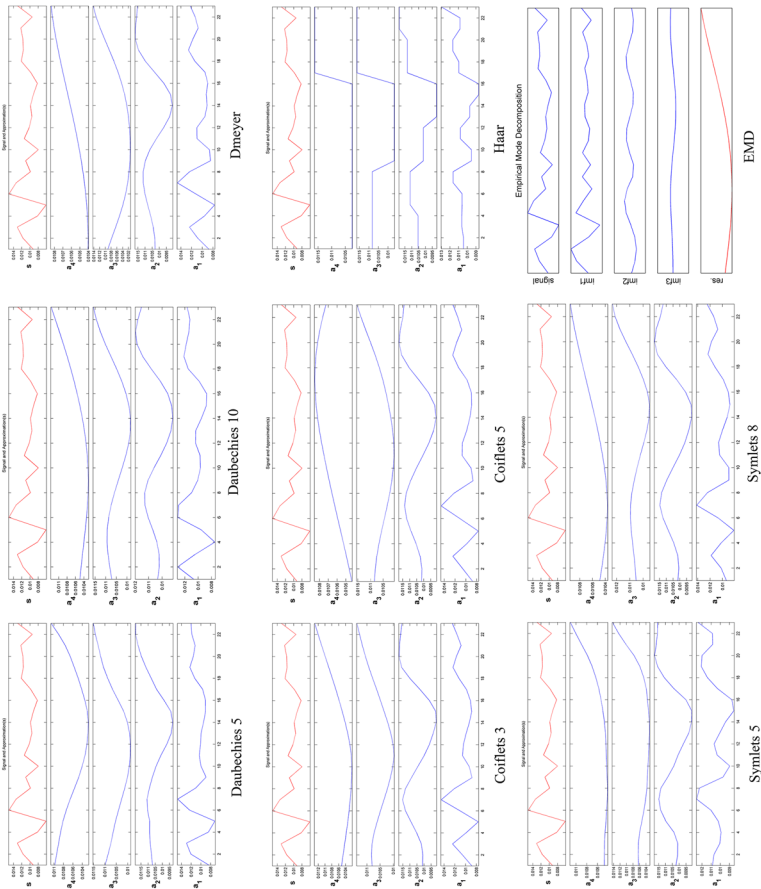


Fig. 12 Wavelet decomposition for case study 1 with different Wavelet types and parameters and EMD decomposition

4.6.4.2 *IMF3: an almost half month cycle* From Table 10, Jul-14, Sept-14, and Oct-14 have 12,16 and 17 days cycles, which are very close to a half month cycle.

4.7 Comparison

4.7.1 Benchmark

From the literature review, EMD and Wavelet are the most widely used techniques for the task of decomposition nonlinear and non-stationary time series. Therefore, we conduct the comparison between EMD and Wavelet.

4.7.2 EMD versus wavelet

We use Wavelet toolbox of Matlab to decompose the data for case study 1. Since the time series is one dimension-discrete data, we choose Wavelet 1-D to decompose. The Wavelet types that are suitable for 1-dimension discrete data are Haar, Daubechies, Coiflets, Symlets, and Dmeyer. Figure 12 is the decomposition results of these Wavelets with different parameters, which includes the EMD result for comparing as well. All these Wavelets choose four as the level so that we can compare with the result of EMD that has four levels as well.

Since the purpose of decomposing is to identify the factors and cycles of the signal, the decomposed time series should have a minimum correlation. Therefore, we define the following evaluation $eval = \sum_{t_i, t_j \in T} |\rho_{t_i, t_j}|$ for assessing this decomposition, where $t_i \neq t_j$ and t_i and t_j are the decomposed time series for a decomposition model. A good model should have a smaller evaluation value. From Table 11, one can observe that EMD has the minimum evaluation value.

We use the decomposition of different Wavelets model as input for EMDER. Table 12 is the result of EMDER. One can see that none of the Wavelet models can recognize the pattern of seasonal fluctuation, a cycle of 6 months. For the residue trend, Symlets 5 can yield the same result. Moreover, Daubechies 5 and Haar have different results from the other model. The other models have slight difference results from EMD.

Thus, we have the following results. First, EMD can decompose the signal into some independent time series than the Wavelet models. Second, since Wavelet is a parametric method, which has many types of Wavelet sub-model with different parameters, there will be the model-selection problem to choose the proper model for this context.

Table 11 Evaluation for decomposition models

Model	Evaluation value	Model	Evaluation value
EMD	0.573	Coiflets 3	3.180
Daubechies 5	3.140	Coiflets 5	1.624
Daubechies 10	3.070	Haar	4.506
Dmeyer	2.046	Symlets 5	3.641
Symlets 8	2.958		

Table 12 EMDER results for decomposition models

Model	Cycle		Residue trend/A4	Model	Cycle		Residue trend/A4
	IMF2/A2	IMF3/A3			IMF2/A2	IMF3/A3	
EMD	6	4	↘↑↗(↗82%)	Coiflets 3	11	4	↘↑↗(↗71%)
Daubechies 5	4	4	↘↑↗(↘56%)	Coiflets 5	11	4	↗↓↘(↗71%)
Daubechies 10	11	4	↘↑↗(↗76%)	Haar	4	4	-
Dmeyer	11	4	↗(↗100%)	Symlets 5	11	4	↘↑↗(↗82%)
Symlets 8	11	4	↘↑↗(↗88%)				

5 Conclusion

This paper proposes an EMD based ensemble recognition method for conversion rate (EMDER). Other than established mathematical models by adopting a deductive approaching in literature, this model is an inductive approaching that treats the system of conversion behavior as a black box. Furthermore, the data used in this study are time series, not click-stream data, panel data, log data, and path data that were utilized in literature. It provides a novel angle of view in solving this problem.

We apply EMDER to 50 datasets and obtain some management insights through the acquired patterns, such as seasonal pattern, CCPI pattern, the long-term time series pattern, the calendar of daily fluctuation, the hesitation window, the consumers' cash flow determined pattern, and the promotion day and holiday influence. These patterns are useful for online retailers.

A comparison between EMDER and Wavelet-based method is conducted, which reveals that EMDER outperforms the Wavelet-based model in the deposition quality and do not have the model-selection problem.

In the future study, we will collect more data to extend empirical studies. We can use other methods to implement the recognition function for more complex data and industry. For example, we can use a dependent function of the rough set theory [39] as the recognition function. Moreover, we can introduce operations that can enable the recognition function to support finding correlation or dependency with delayed influence.

Acknowledgements Our work is supported by the National Science Foundation of China (Grant Nos. 71301180, 71402011), China Postdoctoral Science Foundation (Grant Nos. 2014M560711, 2015T80974), Basic and frontier technology projects of Chongqing Municipal (Grant No. cstc2017jcyjAX0105) and Science and Technology Research Project of Chongqing Municipal Education Commission (Grant No. KJ1705119). Professor Honghui Deng and Professor Reza Torkzadeh (University of Nevada Las Vegas) have contributed to the revised and polishing of this paper. The authors are very grateful for Professor Honghui Deng, Professor Reza Torkzadeh, and anonymous reviewers' valuable suggestions.

References

1. Moe, W. W., & Fader, P. S. (2004). Dynamic conversion behavior at e-commerce site's. *Management Science*, 50(3), 326–335. <https://doi.org/10.1287/mnsc.1040.0153>.
2. Sismeiro, C., & Bucklin, R. E. (2004). Modeling purchase behavior at an E-commerce web site: A task-completion approach. *Journal of Marketing Research*, 41(3), 306–323. <https://doi.org/10.1509/jmkr.41.3.306.35985>.
3. Van den Poel, D., & Buckinx, W. (2005). Predicting online-purchasing behaviour. *European Journal of Operational Research*, 166(2), 557–575.
4. Bharati, P., & Chaudhury, A. (2004). An empirical investigation of decision-making satisfaction in web-based decision support systems. *Decision Support Systems*, 37(2), 187–197. [https://doi.org/10.1016/s0167-9236\(03\)00006-x](https://doi.org/10.1016/s0167-9236(03)00006-x).
5. Hui, S. K., Fader, P. S., & Bradlow, E. T. (2009). Path data in marketing: An integrative framework and prospectus for model building. *Marketing Science*, 28(2), 320–335. <https://doi.org/10.1287/mksc.1080.0400>.
6. Langer, N., Forman, C., Kekre, S., & Sun, B. H. (2012). Ushering buyers into electronic channels: An empirical analysis. *Information Systems Research*, 23(4), 1212–1231. <https://doi.org/10.1287/isre.1110.0410>.

7. De, P., Hu, Y., & Rahman, M. S. (2010). Technology usage and online sales: An empirical study. *Management Science*, 56(11), 1930–1945. <https://doi.org/10.1287/mnsc.1100.1233>.
8. Wikipedia (2013). Taobao. (pp. <http://en.wikipedia.org/wiki/Taobao>).
9. Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., et al. (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the Royal Society of London A: Mathematical, physical and engineering sciences*, 1998 (Vol. 454, pp. 903–995, Vol. 1971). The Royal Society
10. Zhang, X., Yu, L., Wang, S., & Lai, K. K. (2009). Estimating the impact of extreme events on crude oil price: An EMD-based event analysis method. *Energy Economics*, 31(5), 768–778. <https://doi.org/10.1016/j.eneco.2009.04.003>.
11. Hui, S. K., Eliashberg, J., & George, E. I. (2008). Modeling DVD preorder and sales: An optimal stopping approach. *Marketing Science*, 27(6), 1097–1110. <https://doi.org/10.1287/mksc.1080.0370>.
12. Mintz, O., Currim, I. S., & Jeliakov, I. (2013). Information processing pattern and propensity to buy: An investigation of online point-of-purchase behavior. *Marketing Science*, 32(5), 716–732. <https://doi.org/10.1287/mksc.2013.0790>.
13. Wang, H., Wei, Q., & Chen, G. Q. (2013). From clicking to consideration: A business intelligence approach to estimating consumers' consideration probabilities. *Decision Support Systems*, 56, 397–405. <https://doi.org/10.1016/j.dss.2012.10.052>.
14. Xu, L. Z., Duan, J. A., & Whinston, A. (2014). Path to purchase: A mutually exciting point process model for online advertising and conversion. *Management Science*, 60(6), 1392–1412. <https://doi.org/10.1287/mnsc.2014.1952>.
15. Bucklin, R. E., Lattin, J. M., Ansari, A., Gupta, S., Bell, D., Coupey, E., et al. (2002). Choice and the internet: From clickstream to research stream. *Marketing Letters*, 13(3), 245–258. <https://doi.org/10.1023/a:1020231107662>.
16. Park, J., & Chung, H. (2009). Consumers' travel website transferring behaviour: Analysis using clickstream data-time, frequency, and spending. *Service Industries Journal*, 29(10), 1451–1463. <https://doi.org/10.1080/02642060903026254>.
17. Olbrich, R., & Holsing, C. (2011). Modeling consumer purchasing behavior in social shopping communities with clickstream data. *International Journal Of Electronic Commerce*, 16(2), 15–40. <https://doi.org/10.2753/jec1086-4415160202>.
18. Rutz, O. J., & Bucklin, R. E. (2012). Does banner advertising affect browsing for brands? Clickstream choice model says yes, for some. *Qme-Quantitative Marketing And Economics*, 10(2), 231–257. <https://doi.org/10.1007/s11129-011-9114-3>.
19. Lin, L., Hu, P. J. H., Sheng, O. R. L., & Lee, J. (2010). Is stickiness profitable for electronic retailers? *Communications of the ACM*, 53(3), 132–136. <https://doi.org/10.1145/1666420.1666454>.
20. Shao, Z., Chao, F., Yang, S.-L., & Zhou, K.-L. (2017). A review of the decomposition methodology for extracting and identifying the fluctuation characteristics in electricity demand forecasting. *Renewable and Sustainable Energy Reviews*, 75(Supplement C), 123–136. <https://doi.org/10.1016/j.rser.2016.10.056>.
21. Jung, J., & Tam, K.-S. (2013). A frequency domain approach to characterize and analyze wind speed patterns. *Applied Energy*, 103(Supplement C), 435–443. <https://doi.org/10.1016/j.apenergy.2012.10.006>.
22. Xu, W., Gu, R., Liu, Y., & Dai, Y. (2015). Forecasting energy consumption using a new GM–ARMA model based on HP filter: The case of Guangdong Province of China. *Economic Modelling*, 45(Supplement C), 127–135. <https://doi.org/10.1016/j.econmod.2014.11.011>.
23. Li, H., Yang, Z., Zheng, T. Q., Zhang, B., & Sun, H. Common-mode EMI suppression based on chaotic SPWM for a single-phase transformerless photovoltaic inverter. In *2014 16th European conference on power electronics and applications*, 26–28 Aug. 2014 2014 (pp. 1–7). <https://doi.org/10.1109/epe.2014.6910788>.
24. Jia, X., An, H., Fang, W., Sun, X., & Huang, X. (2015). How do correlations of crude oil prices co-move? A grey correlation-based wavelet perspective. *Energy Economics*, 49, 588–598. <https://doi.org/10.1016/j.eneco.2015.03.008>.
25. Jiang, M., An, H., Jia, X., & Sun, X. (2017). The influence of global benchmark oil prices on the regional oil spot market in multi-period evolution. *Energy*, 118, 742–752. <https://doi.org/10.1016/j.energy.2016.10.104>.
26. Sun, E. W., & Meinl, T. (2012). A new wavelet-based denoising algorithm for high-frequency financial data mining. *European Journal of Operational Research*, 217(3), 589–599. <https://doi.org/10.1016/j.ejor.2011.09.049>.

27. Fan, G.-F., Peng, L.-L., Hong, W.-C., & Sun, F. (2016). Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression. *Neurocomputing*, *173*(Part 3), 958–970. <https://doi.org/10.1016/j.neucom.2015.08.051>.
28. Chen, B., Zhao, S. L., & Li, P. Y. (2014). Application of Hilbert–Huang transform in structural health monitoring: A state-of-the-art review. *Mathematical Problems In Engineering*. <https://doi.org/10.1155/2014/317954>.
29. Duan, W. H., Wang, Q., & Quek, S. T. (2010). Applications of piezoelectric materials in structural health monitoring and repair: Selected research examples. *Materials*, *3*(12), 5169–5194. <https://doi.org/10.3390/ma3125169>.
30. Xiong, T., Bao, Y. K., & Hu, Z. Y. (2013). Beyond one-step-ahead forecasting: Evaluation of alternative multi-step-ahead forecasting models for crude oil prices. *Energy Economics*, *40*, 405–415. <https://doi.org/10.1016/j.eneco.2013.07.028>.
31. Yu, L. A., Wang, S. Y., & Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics*, *30*(5), 2623–2635. <https://doi.org/10.1016/j.eneco.2008.05.003>.
32. Zhang, X., Lai, K. K., & Wang, S. Y. (2008). A new approach for crude oil price analysis based on empirical mode decomposition. *Energy Economics*, *30*(3), 905–918. <https://doi.org/10.1016/j.eneco.2007.02.012>.
33. Lin, C. S., Chiu, S. H., & Lin, T. Y. (2012). Empirical mode decomposition-based least squares support vector regression for foreign exchange rate forecasting. *Economic Modelling*, *29*(6), 2583–2590. <https://doi.org/10.1016/j.econmod.2012.07.018>.
34. Kozic, I., & Sever, I. (2014). Measuring business cycles: Empirical mode decomposition of economic time series. *Economics Letters*, *123*(3), 287–290. <https://doi.org/10.1016/j.econlet.2014.03.009>.
35. Lisi, F., & Nan, F. (2014). Component estimation for electricity prices: Procedures and comparisons. *Energy Economics*, *44*, 143–159. <https://doi.org/10.1016/j.eneco.2014.03.018>.
36. Chen, M. C., & Wei, Y. (2011). Exploring time variants for short-term passenger flow. *Journal of Transport Geography*, *19*(4), 488–498. <https://doi.org/10.1016/j.jtrangeo.2010.04.003>.
37. Huang, N. E., Shen, Z., & Long, S. R. (1999). A new view of nonlinear water waves: The Hilbert spectrum. *Annual Review of Fluid Mechanics*, *31*, 417–457. <https://doi.org/10.1146/annurev.fluid.31.1.417>.
38. Huang, N. E., & Wu, Z. H. (2008). A review on Hilbert–Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*. <https://doi.org/10.1029/2007rg000228>.
39. Pawlak, Z. (1982). Rough sets. *International Journal of Computer and Information Sciences*, *11*(5), 341–356.