Extracting Trends and Cycles: BK, HP and Robust Trend Filters for Real-world Time Series Data

Xiaolong Luo* ·, Zhihan Liu† ·, and Chengchang Liu‡ § ·

·Dept.of Statistics and Finance,University of Science and Technology of China

July 22, 2021

Abstract

Extracting Trends and Cycles are a important task in real-world time series analysis, and there are various methods (filters) to tackle this task in the literatures. This paper reviews several famous filters in separating the trend and cyclical terms and discuss their strengths and drawbacks. Also, we test lots real-world data to valid our conclusion and make comparisons between filters. A worth mentioning thing is that, we investigate the filters properties both from trend and cycle aspects, while the original paper may only restrict in a single aspect. Also, the experiment in this paper is comprehensive using both synthetic and real data.

Key words: Band Pass filters; HP and Modified HP filters; Approximation of filters; Robust Trend filters; Real-world data.

*Xiaolong Luo: Literature review and codes in Section 3; Thesis writing (Abstract and Section 1,3,5).

†Zhihan Liu: Thesis writing and codes in Section 4.

‡Chengchang Liu: Thesis writing and codes in Section 2

§**Declarations Conflict of interest:** The authors declare no conflict of interest and contribute equally.

1 Introduction

This project is based on the idea of Baxter and King (1995), the method of Bruchez (2003) and Wen et al. (2019). We review and summary the main idea of these literatures and extend their method to real-world data and make comparison, finding out the difference and properties of these different methods(filters). In this paper, we aim at answering the following question:

1. *Can we design a good filter to extract the trend and cycle terms in real-world data such as economic series?*

2. *How to approximate an ideal filter which can perfectly extract the cyclical and trend term in the data?*

3. *The idea behind HP filter and the modified HP filter, and how we fix it drawbacks.*

4. *Is there a robust filter which can tackle outlier problem in real-world data?*

Economical data is usually a common time series and a crucial task for Economists is to separate the trend and cyclical term in the data. For example, business cycle theory is primarily concerned with understanding fluctuations in the range from 1.5 to 8 years, whereas growth theory focuses on the longer run Christiano and Fitzgerald (2003). Thus extracting the different terms is of great importance. Inspired by Baxter and King (1995), we use band pass filters which is an ideal method to isolate business cycle of a real financial data. The ideal band pass filter is not practical, with an approximation filters of ideal BP filter, it allow us to isolate certain frequency parts of a time series operationally. Which means there is no need for a long moving average of a time series.

Besides, there is another perspective to tackle the separation problem: Transform it as an optimization problem and find an optimal solution that best can represent the trend and cyclical term. This method is propsed by Hodrick and Prescott (1997), known as HP filters. However, the HP filters suffer from End-of-sample bias problem and Temporary assumptions problem which restrict its performance. In dealing with these two drawbacks, Bruchez (2003) and McDermott (1997) proposed approaches to fix them.

In practice, there are lots of real-world data suffer from outliers and dips, making it difficult to apply the filters directly as mention above, especially for frequency-based filters. To explore further, we discuss trend filters which are optimization-based and able to handle such challenging time series data in Wen et al. (2019). We show the empirical superiority in our experiment with solving the optimization problem by efficient method in Section 4.

This article is organized as follows: Section 2 introduces the Band Pass and approximation filters. We first show the construction of the filters and then using the real-world data for experiment. Section 3 is about HP filters and similarly organized as Section 2. Further we discuss the trend filters in Section 4, which is robust to the outlier data in practice and using synthetic and real-world data for experiments and comparison. Section 5 is for summary and conclusion.

2 Band Pass And Approximation Filters For A Business Cycle

The Section focus on using band pass filter and the approximation filters of it to isolate business cycle from a time series.

In Section 2.1, We give a brief introduction to band pass filter ,the ideal band pass filter which is well known, the Approximate symmetric filter for the ideal filter given by Baxter and King (1995) (note it as BK filter), and an optimal finite sample approximation filter for the ideal filter which is not symmetric in Christiano and Fitzgerald (2003) (note it as CF filter). In Section 2.2, We give the weights of different filters: the ideal band width filter, the BK filter and the CF filter. In Section 2.3, we explore the effect of changes in maximum lag length,k, on the BK filters and compare it with the ideal band pass filter. In Section 2.4 We apply BK filter on real world data to test its performance.

2.1 Introduction To Band Pass Filter And Its Approximation Filters

The main object of a band pass filter is to seek the component y_t of x_t that lies in a particular frequency range. We can write it as :

$$x_t = y_t + \widetilde{x}_t \tag{2.1}$$

Where y_t has power in frequencies belonging to the interval $\{(a,b) \cup (-b,-a)\}$ and $\widetilde{x}_t$ in the complement of the interval in $(-\pi, \pi)$. We can write y_t in the moving average form $y_t = B(L)x_t$. And we want our filters have the properties that:

$$\begin{aligned} B(e^{-iw}) &= 1 \quad w \in \{(a,b) \cup (-b,-a)\} \\ &= 0 \quad otherwise \end{aligned} \tag{2.2}$$

The ideal Band pass filter, The BK bind pass filter and the CF Band Pass filter have different form of B(L), We denote them as :$B^{ide}(L)$,$\widehat{B}^{K}(L)$,$\widehat{B}^{p,f}(L)$.

2.1.1 The ideal band pass filter

The ideal band pass filter $B^{ide}(L)$ has the following structure:

$$B^{ide}(L) = \Sigma_{-\infty}^{\infty} B_j L^j \tag{2.3}$$

where $L^j x_t = x_{t-j}$.

The ideal band pass filter requires infinite number of observations on x_t, which can not be implement on real time series data, we seek to find the approximate filter of the ideal band pass filter.

2.1.2 The BK filter

The BK filter with K lags to approximate the ideal band pass filter has the following structure:

$$\widehat{B}^{K}(L) = \Sigma_{j=-K}^{K} \widehat{B}_j^{K} L^j \tag{2.4}$$

and B_j is symmetric.

We can see that the BK filter is the optimal fixed-lag,symmetric filter. But sometimes increasing K requires throwing away more data at the beginning or at the end of the data set. We hope to find an optimal finite-sample approximations for the band pass filter.

2.1.3 The CF filter

The CF filter with p,f lags to approximate the ideal band pass filter has the following structure:

$$\widehat{B}^{p,f}(L) = \Sigma_{j=-f}^{p} \widehat{B}_j^{p,f} L^j \tag{2.5}$$

The CF filter use 2 parameters p and f instead of a single lag parameter to approximate the ideal band pass filter may perform better in some case. We will also show in the next subsection that the conduct of the CF filter which relies on the spectral density of a time series is totally different with the BK filter.

2.2 Construct The Band Pass Filters

First we give B_j of the ideal band pass filter that B(L) fulfill (2.2), We do the Fourier inverse transform of B(L) to get B_j:

$$B_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} B(e^{-iw}) e^{iwj} dw$$

The filter weights of ideal band pass filter are:

$$\begin{aligned} B_j &= \frac{sin(bj) - sin(aj)}{\pi j} \\ B_0 &= \frac{b-a}{\pi} \end{aligned} \tag{2.6}$$

Then calculate the filter weights of the BK filter, We solve a constrained optimization problem:

$$\begin{aligned} &Min: \quad Q = \int_{-\pi}^{\pi} |B^{ide}(e^{-iw}) - \widehat{B}^K(e^{-iw})|^2 dw \\ &with: \widehat{B}^K(1) = 0 \end{aligned} \tag{2.7}$$

Using Lagrange multiple method we directly give the BK filter weights as:

$$\widehat{B}_j^K = B_j - \tfrac{1}{2K+1}\Sigma_{j=-K}^{K} B_j \tag{2.8}$$

where B_j has already been calculated in (2.6)

Let $f_x(w)$ be the spectral density of x_t, the optimal finite-sample approximation for the band pass filter given by Christiano and Fitzgerald (2003) can be constructed by solving an optimization problem:

$$Min: \quad \int_{-\pi}^{\pi} |B(e^{-iw}) - \widehat{B}^{p,f}(e^{-iw})|^2 f_x(w)\mathrm{d}w \tag{2.9}$$

Here we only give the solution of $\widehat{B}_j^{p,f}$ when the x_t has the following representation (the spectral density of x_t is known):

$$x_t = x_{t-1} + \theta(L)\epsilon_t, \quad E(\epsilon_t^2) = 1$$

The corresponding spectral density is:

$$\begin{aligned} f_x(w) &= \frac{g(w)}{|1-e^{iw}|^2} \\ where \qquad g(w) &= \theta(e^{-iw})\theta(e^{iw}) = c_0 + \Sigma_{i=1}^{q} 2c_i cos(wi) \end{aligned}$$

We denote $\widetilde{B}(z) = \frac{B^{ide}(z)}{1-z}$ and $b(z) = \frac{\widehat{B}^{f,p}(z)}{1-z} = \Sigma_{i=-f}^{p-1} b_i z^i$.
Let $b = [b_{p-1}, b_{p-2}, \cdots, b_{-f}]^T$, and $\widehat{B}^{f,p} = [\widehat{B}_p^{f,p}, \widehat{B}_{p-1}^{f,p} \cdots, \widehat{B}_{-f}^{f,p}]^T$ We have the following relation between b and $\widehat{B}^{f,p}$:

$$Q\widehat{B}^{f,p} = b$$

$$where \qquad Q = \begin{bmatrix} -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & -1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & -1 & -1 & \cdots & -1 & 0 \end{bmatrix}$$

And the following relation between $\widetilde{B}(z)$ and $\widehat{B}^{f,p}$:

$$\int_{-\pi}^{pi} \widetilde{B}(e^{-iw})g(w)e^{iwj}\mathrm{d}w = 2\pi F_j Q \widehat{B}^{f,p}$$

$$where \qquad F_j = [\underbrace{0,\cdots,0}_{(p-1-1-j)}, c, \underbrace{0,\cdots,0}_{(j-q+f)}]$$

So it turns out that:

$$\int_{-pi}^{pi} \widetilde{B}(e^{-iw})g(w)e^{-iwj}\mathrm{d}w = \int_a^b [\tfrac{e^{-iwj}}{1-e^{-iw}} + \tfrac{e^{iwj}}{1-e^{iw}}]g(w)\mathrm{d}w$$

Thus we can calculate the CF filter weight by solving the linear equations above.

We finally give the weight of the CF filter as:

$$\widehat{B}^{f,p} = A^{-1}\mathrm{d}$$

$$where:$$

$$d = \begin{bmatrix} \int_{-pi}^{pi} \widetilde{B}(e^{-iw})g(w)e^{iw(p-1)}\mathrm{d}w \\ \vdots \\ \int_{-pi}^{pi} \widetilde{B}(e^{-iw})g(w)e^{iw(-f+1)}\mathrm{d}w \\ \int_{-pi}^{pi} \widetilde{B}(e^{-iw})g(w)e^{iw(-f)}\mathrm{d}w \\ 0 \end{bmatrix} A = 2\pi \begin{bmatrix} F_{p-1}Q \\ \vdots \\ F_{-f+1}Q \\ F_{-f}Q \\ 1\cdots 1 \end{bmatrix} \tag{2.10}$$

We can see that the construction process of CF filter weights are complicated and the weights are only suitable for special time series. In the following subsections, We only focus on the implementation and the application of the BK filter.

2.3 The Effect Of Truncation

This subsection we try to figure out the effect of changes in the maximum lag lengths K in the BK filter. Figure 1 displays the frequency-response function for BK filters. The frequency response function is actually a Fourier Inverse Transform of the filter weights $\widehat{B}_j^K$.

$$f(w) = \widehat{B}^K(e^{-iw}) = \Sigma_{j=-K}^{K} \widehat{B}_j^K e^{-iw}$$

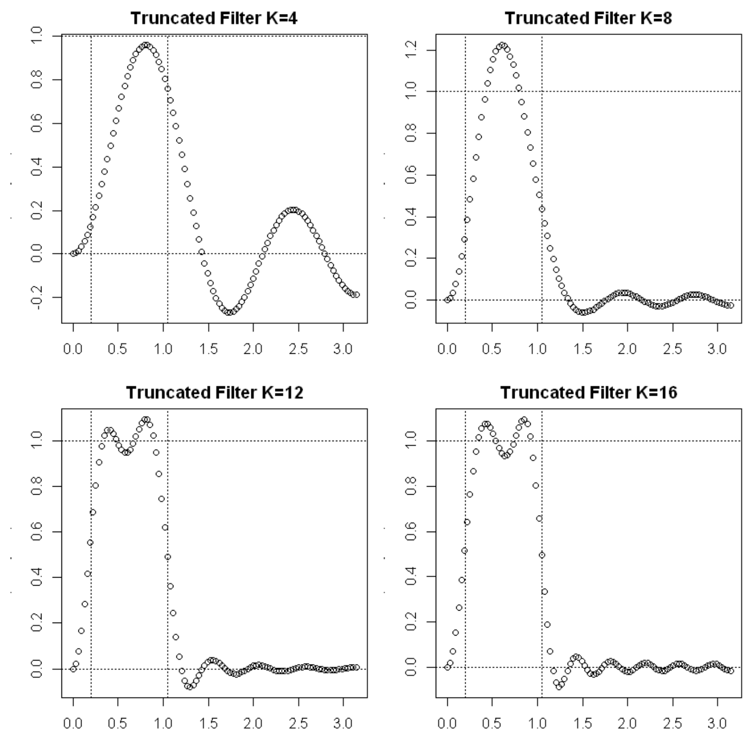


Figure 1: The frequency response of different K

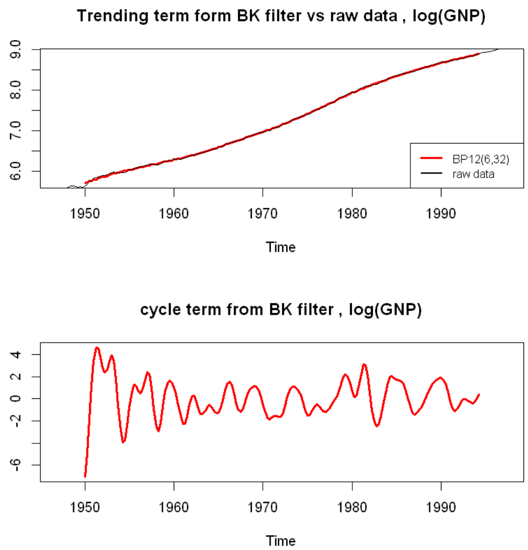
The frequency interval we want to filter out is given by $a = \frac{2}{32\pi}$, $b = \frac{2}{6\pi}$, the ideal filter has the frequency response function as (2.2).

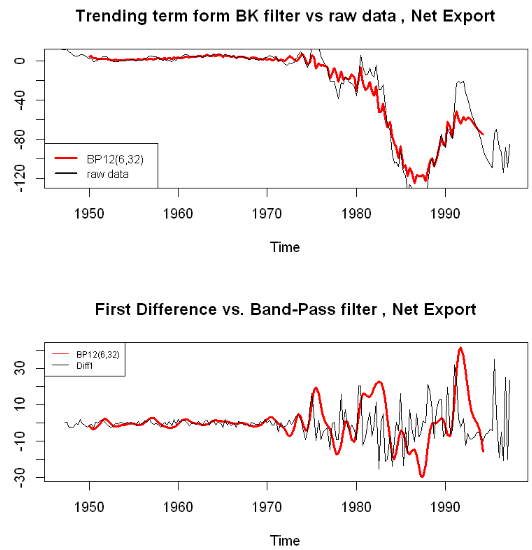
Figure 1 shows that the increasing of K leads to a better approximation to the ideal filter which is obvious. However, increasing K may lead more lost observations. And more specifically, the frequency responses oscillate around zero on higher-cutoff frequency and produce some negative weights in the frequency response.

2.4 BK Filter On Real World Data

This subsection, we will use real world finance data to test our BK filter.

First, recall why we want to use BK filter to isolate a business cycle from a time series. As is mentioned in Baxter and King (1995), the business cycles were cyclical components of no less than 6 quarters and last fewer than 32 quarters. So a desired filter for the business





(a) log(GNP) data (b) Net Export data

Figure 2: BK(6,32) on real world data

cycle will be our Band Pass filter which will pass components of time series with periodic fluctuations between 6 and 32 quarters. And we choose the lag numbers K=12. We apply the BK filter to finance data: The natural logarithms of gross national product(GNP) and the inflation rate.

The Figure 2 shows the performance of BK(6,32) on different raw data. We can see the filter can isolate the business cycle when dealing with log(GNP) data and the left trending part fit the data well. hen dealing with the Net Export data, the trending part do not fit the data after 1990 well, This is because there is a significant fluctuations around that time and the BK filter will lose 3 years of data at each end of the plots since K=12. We can figure out another filter for this data.

3 HP and Modified HP Filters

This section we focus on another kind of filters: Hodrick-Prescott Filters Hodrick and Prescott (1997) to decomposes a seasonally time series into a trend term and a cyclical term. In

Section 3.1, We briefly introduce the HP filter and the key idea behind the method. Then in Section 3.2 we discuss some shortcoming of this method for real data and proposed modified HP filter. In Section 3.3 We apply HP filter on real world data and compare it with Band Pass filters in Section 2.

3.1 Introduction To Hodrick-Prescott Filter

Similarly, the main object of a HP filter is same as Band Pass filters, i.e. : decompose the trend term and cyclical term

$$y_t = g_t + c_t, \quad t = 1, 2, 3, \ldots, T \tag{3.1}$$

Where y_t, g_t and c_t are a given time series, trend, and cyclical components respectively. The idea of this HP filters is different from BK filters which is an approximation for an ideal filters. Instead, we want to the trend term g_t is close to the real data y_t as well as the cyclical term c_t is "smooth" as much as possible. The smoothness is measured by the sum of squares of the second order difference of c_t, denote as $v \in \mathbb{R}^{T-2}$, and

$$v_t = ((y_t - y_{t-1}) - (y_{t-1} - y_{t-2})) \quad t = 3, 4, \ldots, T \tag{3.2}$$

or equivalently,

$$v = P\mathbf{c} \tag{3.3}$$

with

$$P := \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & & \ddots & & \\ & & & 1 & -2 & 1 \end{bmatrix}.$$

of order $(T-2) \times T$.

From the above, using the HP filters is equal to solve a optimization problem:

$$\min_{g_t} \left[\sum_{t=1}^{T} (y_t - g_t)^2 + \lambda \sum_{t=1}^{T-2} [(g_{t+2} - g_{t+1}) - (g_{t+1} - g_t)]^2 \right], \tag{3.4}$$

where λ is a smoothing parameter. Notice that there is a trade-off between the goodness of fit and the degree of smoothness that depends on the value of λ. Formally, we set $\lambda = 1600$ for quarterly and 100 for annually data.

3.2 Modified HP Filter

From literature, the HP mainly suffer from three drawback Schüler (2019): (i. spurious cycles, ii. end-of-sample bias, iii. ad hoc assumptions regarding the smoothing parameter). Here we mainly discuss approaches that tackle the last two problems.

3.2.1 End-of-sample Bias Problem

It's well known that the HP filters suffer from the end-point bias: the last point of the series has an exaggerated impact on the trend at the end of the series Bruchez (2003). And the reason why the HP filters have such a problem is that the second term of equation (3.4) features a sum from $t = 2$ to $t = T - 1$, not from $t = 1$ to $t = T$. The consequence is that g_{T-1}, g_T, g_1 and g_2 do not appear in the second term of equation (3.4) as often as the other g_t terms.

To tackle this problem, one way is to make a prediction for the former and later terms of g_t to make g_{T-1}, g_T, g_1 and g_2 appear twice as other g_t terms in the equation. However, the prediction is always not reliable and sometimes leads more serious deviation Bruchez (2003). Another way to alleviate the bias can be done by minimizing the following penalty function instead of (3.4):

$$\sum_{t=1}^{T} \frac{1}{\lambda_t} \left(y_t - \widetilde{g}_t\right)^2 + \sum_{t=2}^{T-1} \left[(\widetilde{g}_{t+1} - \widetilde{g}_t) - (\widetilde{g}_t - \widetilde{g}_{t-1})\right]^2 \tag{3.5}$$

where $\lambda_t = \lambda$ for $\mathrm{t} = 3$ to $\mathrm{T} - 2$; $\quad \lambda_t = \lambda * \frac{3}{2}$ for $\mathrm{t} = 2$ and $\mathrm{t} = \mathrm{T} - 1$;
and $\lambda_t = \lambda * 3$ for $\mathrm{t} = 1$ and $\mathrm{t} = \mathrm{T}$

The idea behind it is simply increase the corresponding penalty parameter λ for the less appear term such as g_T. Although it does not entirely solve the end-point bias, it reduce it.

In the later experiment 3.3 we use this modified HP filter to extract cyclical and trend term and compare it with BK filters.

3.2.2 Temporary Assumptions for Time Series Problem

Formally, we choose $\lambda = 1600$ for quarterly and 100 for annually data. However, the underlying series cyclical term are usually different from each other and result in the fixed cyclical parameters across series may be misplaced. In such a situation, each series should have a customized smoothing parameter to extract the cyclical component.

To address this issue, McDermott (1997) develop a modified HP filters based on cross-validation method. Here He use cross validation to select an optimal smoothing parameter λ. In literature, the author use the leave-one out procedure and the optimal value of λ can be obtained by minimizing the following gross cross validation (GCV) function with respect to λ

$$\mathrm{GCV}(\lambda) = T^{-1}\left(1 + \frac{2T}{\lambda}\right)\sum_{k=1}^{T}\left(\mathrm{x}_k - \mathrm{g}_{\mathrm{t},k}(\lambda)\right)^2 \tag{3.6}$$

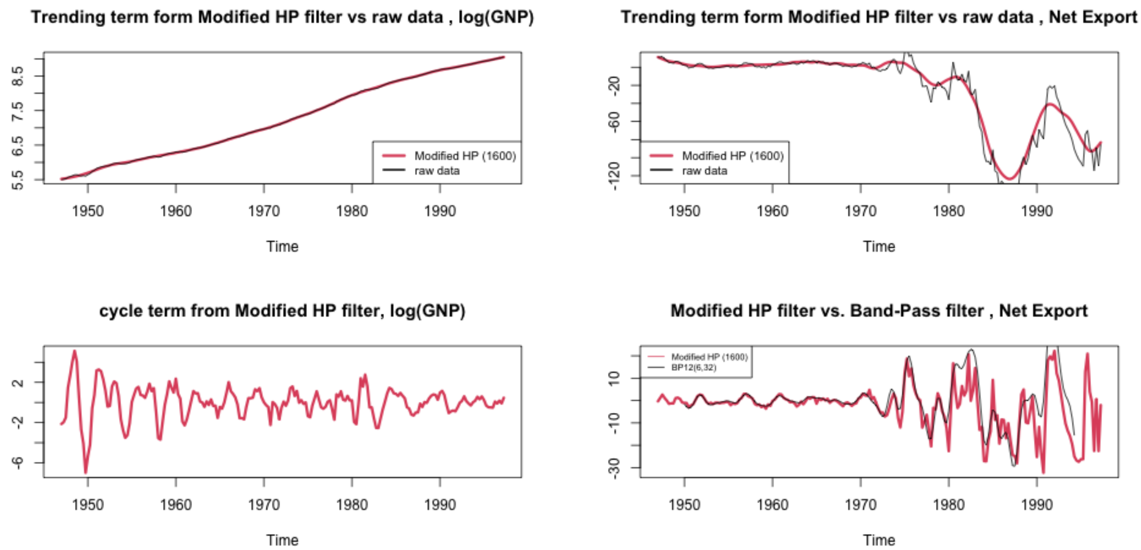
where $\mathrm{g}_{\mathrm{t,k}}(\lambda)$ denote using parameter λ and all the data points but leaving out the k^{th} point[1]. For all λ we choose the one that minimize equation (3.6) as the optimal λ

3.3 BK Filter On Real World Data

This subsection, we continue the experiment and compare the (Modified)HP filters with BK filters on the same data in Section 2.4.

The Figure 3 shows the performance of Modified HP filters ($\lambda = 1600$) on different raw data. The HP filters can also extract the trend term well as BK filters does. However, from the cycle term perspective, it clearly shows that the HP filters smoothness is worse than the BP12(6,32) filter. Thus in Next Section we discuss a more robust filter which can generate smoother result and can also deal with data with outlier.

[1] Since the cross validation method here is hard to compute for long series, we do not include it in the experiment



(a) log(GNP) data

(b) Net Export data

Figure 3: Modified HP (1600) on real world data

4 Trend Filters

Besides extracting business cycles, filter plays an important role in dealing with the decomposition of the trend for many time series data in real world, to do forecast and anomaly detectation. Among them, some data appear to have no or few cyclic features and suffer from outliers and dips, making it difficult to apply the filters directly in previous section, especially for frequency-based filters. In this Section, we discuss trend filters which are optimization-based and able to handle such challenging time series data.

4.1 Unified View of Trend Filter

For the time series of length N, which is characterized as $\boldsymbol{y} = [y_1, \ldots, y_N]^\top$, we desire to decompose it into trend and residual as follows,

$$\boldsymbol{y} = \boldsymbol{\tau} + \boldsymbol{r}, \tag{4.1}$$

where τ denotes the trend $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_N]^\top$ and $\boldsymbol{r}$ denotes the residual $\boldsymbol{r} = [r_1, \ldots, r_N]^\top$. To extract the trend in time series, we need apply filters, which can be formulated in the following unifed view.

$$\boldsymbol{\tau} = \underset{\boldsymbol{\tau}}{\operatorname{argmin}}\, f(\boldsymbol{y} - \boldsymbol{\tau}) + \lambda g(\boldsymbol{\tau}), \tag{4.2}$$

where loss function $f(\cdot)$ measures the scale of residual, regularization function $g(\cdot)$ reflects the desired property of the trend (slow or abrupt trend, and robust, etc.), and λ is the regularization parameter to trade-off between smoothness of $\boldsymbol{\tau}$ and size of the residual $\boldsymbol{r}$. To smooth the trend, one important idea to design regularization function $g(\cdot)$ is to introduce the $k+1$ order difference matrix $\mathbf{D}^{(k+1)} \in \mathbb{R}^{N-k-1,N}$, which can be recursively defined as

$$\mathbf{D}^{(k+1)} = \mathbf{D}^{(1)}\mathbf{D}^{(k)}, \tag{4.3}$$

and the first order and second order difference matrix take the form of

$$\mathbf{D}^{(1)} = \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & & \\ & & & 1 & -1 \end{bmatrix}, \mathbf{D}^{(2)} = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & & \ddots & & \\ & & & 1 & -2 & 1 \end{bmatrix}. \tag{4.4}$$

In fact, $\mathbf{D}^{(2)}$ is an averaging operator, can help constrain the smoothy of the trend. We point out that the following classic but powerful filters are the specific instances of (4.2),

- HP-Filter: $f(\cdot) = \frac{1}{2}\|\cdot\|_2^2$ and $g(\boldsymbol{\tau}) = \left\|\mathbf{D}^{(2)}\boldsymbol{\tau}\right\|_2^2$.
- l_1 Filter (Kim, Koh, Boyd and Gorinevsky, 2009): $f(\cdot) = \frac{1}{2}\|\cdot\|_2^2$ and $g(\boldsymbol{\tau}) = \left\|\mathbf{D}^{(2)}\boldsymbol{\tau}\right\|_1$.
- Total Variation Denoising (TVD) (Selesnick, 2014): $f(\cdot) = \frac{1}{2}\|\cdot\|_2^2$ and $g(\boldsymbol{\tau}) = \left\|\mathbf{D}^{(1)}\boldsymbol{\tau}\right\|_1$.
- Least Absolute Deviation-Lasso (LAD-Lasso) (Wang, Li and Jiang, 2007):$f(\cdot) = \|\cdot\|_1$ and $g(\boldsymbol{\tau}) = \|\boldsymbol{\tau}\|_1$.

Next we introduce a novel filter, proposed by Wen et al. (2019), which combines the properties of the filters below, able to extract both the slow (smooth) trend and abrupt (jump) trend, with robustness to the outliers. For RobustFilter, it needs to implement (4.2) as follows,

$$
\begin{aligned}
f(\boldsymbol{x}) &= \sum_{i=1}^{N} H_\gamma(x_i), \quad \text{for } \boldsymbol{x} = [x_1, \ldots, x_N], \\
g(\boldsymbol{\tau}) &= \lambda_0 \|\mathbf{D}^{(1)}\boldsymbol{\tau}\|_1 + \|\mathbf{D}^{(2)}\boldsymbol{\tau}\|_1,
\end{aligned}
\tag{4.5}
$$

where λ_0 is the trade-off parameter and Huber loss function $H_\gamma(x)$ is defined as

$$
H_\gamma(x) = \begin{cases} \frac{1}{2}x^2, & |x| \leq \gamma \\ \gamma\,|x| - \frac{1}{2}\gamma^2, & |x| > \gamma \end{cases}. \tag{4.6}
$$

We remark that λ_0 depicts the preference for the smooth or abrupt trend and we will explain it further in Section 4.3.1.

4.2 Efficiently Solve RobustFilter

To solve (4.5) efficiently, we can apply ADMM-MM algorithm (Boyd, Parikh, Chu, Peleato and Eckstein, 2011). For the simplicity of later discussion, we rewrite (4.5) as

$$
\begin{aligned}
\min \quad & f(\mathbf{y} - \boldsymbol{\tau}) + \|\mathbf{z}\|_1 \\
\text{s.t.} \quad & \mathbf{D}\boldsymbol{\tau} = \mathbf{z},
\end{aligned}
\tag{4.7}
$$

where $\mathbf{D} = \begin{bmatrix} \lambda_1 \mathbf{D}^{(1)} \\ \lambda_2 \mathbf{D}^{(2)} \end{bmatrix}$. Since this is a constraint optimization problem, applying ADMM methods Boyd et al. (2011), we derive the update formulation through augmented Lagrangian as

$$
\begin{aligned}
\boldsymbol{\tau}^{k+1} &= \arg\min_{\boldsymbol{\tau}} \left(f(\mathbf{y} - \boldsymbol{\tau}) + \frac{\rho}{2} \left\|\mathbf{D}\boldsymbol{\tau} - \mathbf{z}^k + \mathbf{u}^k\right\|_2^2 \right) \\
\mathbf{z}^{k+1} &= \arg\min_{\mathbf{z}} \left(\|\mathbf{z}\|_1 + \frac{\rho}{2} \left\|\mathbf{D}\boldsymbol{\tau}^{k+1} - \mathbf{z} + \mathbf{u}^k\right\|_2^2 \right) \\
\mathbf{u}^{k+1} &= \mathbf{u}^k + \mathbf{D}\boldsymbol{\tau}^{k+1} - \mathbf{z}^{k+1},
\end{aligned}
\tag{4.8}
$$

where ρ is the penalty parameter and $\mathbf{u}$ is the scaled dual variable. What follows is to solve two (unconstrained) optimization subproblems in (4.8). Recall that the second optimization formulation on $\mathbf{z}$ also occurs in Lasso (least absolute shrinkage and selection operator) and there exists a powerful tool named soft-thresholding operator to solve it, which takes the form

$$S_\rho(a) = \begin{cases} 0, & |a| \le \rho \\ a - \rho \operatorname{sgn}(a), & |a| > \rho \end{cases} \tag{4.9}$$

By the definition of S_ρ in (2), following the same procedure in solve Lasso, from (4.8) we obtain that

$$\mathbf{z}^{k+1} = S_\rho \left(\mathbf{D}\boldsymbol{\tau}^{k+1} + \mathbf{u}^k \right). \tag{4.10}$$

As for the first optimization on $\boldsymbol{\tau}$ in (4.8), since Huber loss is a piece-wise function, we apply one-iteration MM framework to efficiently optimize it by selecting the proved sharpest quadratic majorization function for Huber loss (De Leeuw and Lange, 2009), that is,

$$H_\gamma(x) \le constant + \frac{H'_\gamma(x^{(k)})}{2x^{(k)}} x^2, \tag{4.11}$$

where $x^{(k)}$ is the k-th iterated value. By optimizing majorization function in (4.11), we obtain the closed-form solution of the optimization on $\boldsymbol{\tau}$ in (4.8) as follows,

$$\boldsymbol{\tau}^{k+1} = \mathbf{y} - \rho \left(\mathbf{A}_k + \rho \mathbf{D}^T \mathbf{D} \right)^{-1} \mathbf{D}^T \left(\mathbf{u}^k - \mathbf{z}^k + \mathbf{D}\mathbf{y} \right), \tag{4.12}$$

where $\mathbf{A}_k = \operatorname{diag}\left(g'_\gamma \left(x^{(k)} \right) \right) \operatorname{diag}^{-1} \left(x^{(k)} \right)$ and the gradient of Huber loss function $H'_\gamma(x)$can be calculated as

$$H'_\gamma(x) = \begin{cases} x, & |x| \le \gamma \\ \gamma \operatorname{sgn}(x), & |x| > \gamma \end{cases},$$

where sgn is the sign function. Combining (4.8), (4.10), and (4.2), we can extract trends of RobustFilter through interactively solving such a ADMM-MM algorithm.

4.3 Experiment

To compare the practical effects of the trend filters mentioned in Section 4.1, we first simulate them on synthetic data, then test them on real data (Covid-19 new cases in Italy).

Basline Algorithm In the follow experiments, we will compare RobustFilter with $l1$ Filter, HP-Filters, and TVD, which are all mentioned in Section 4.1.

4.3.1 Simulation: Compound Signal with Outliers

Simulation Setup For synthetic data, we initially generate a compound original signal with 500 data points, constituting of a sin wave, a square wave, and a triangle wave. as in Figure 4. Then we add a Gaussian noise with 0.2 deviation. After that, we add 10 to 200 Gaussian noise with 2 deviation to represent 4% to 20% outlier ratio. For instance, input synthetic data with noise and 4% outlier are shown in Figure 5.

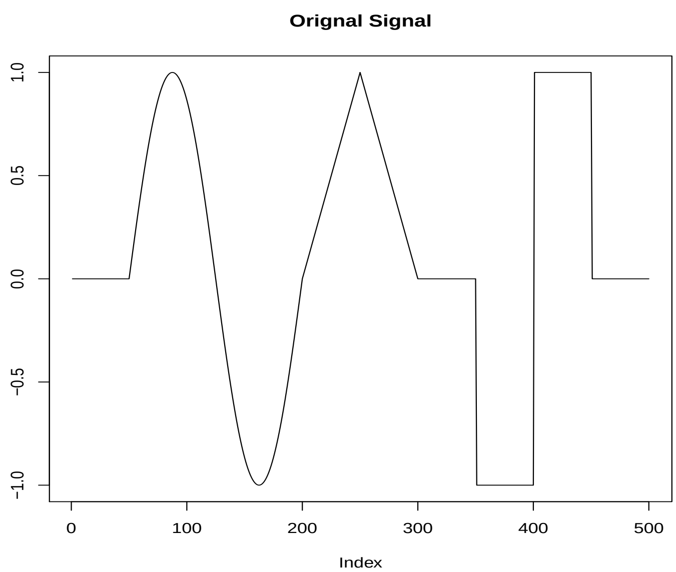


Figure 4: Original Signals in synthetic data.

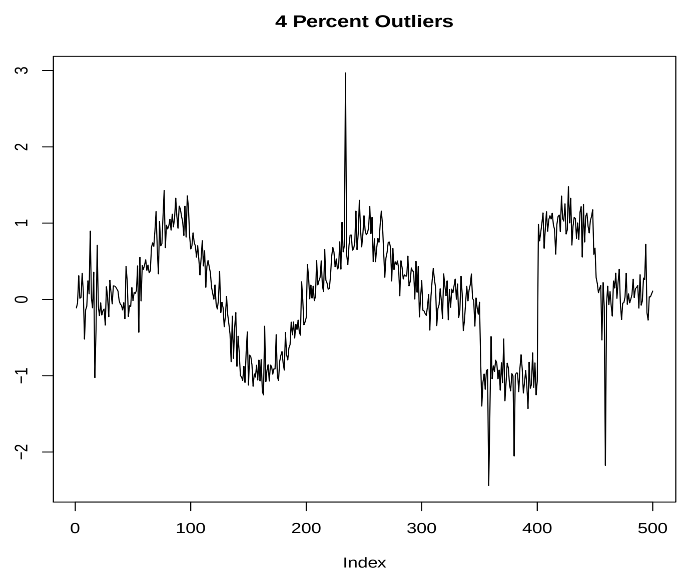


Figure 5: Input synthetic data with Gaussian noise and 4 percent ouliers.

The trend extraction plot for the synthetic data with 4 percent outliers is shown in Figure 6. From the plot, we observe that the trend extraction by RobustFilter nearly perfectly restore the orignal signal without affection by outliers. By contrast, $l1$ Filter can extract abrupt trend (square wave) well but fails to restore the smooth and slow trend (sin wave) with their stair-case like trend, so is TVD. HP Filter is apt to select the smooth and slow trend but stays less sensitive to the sharp signal, like square wave and triangle wave. What is more, expect for RubostFilter, they all are mislead by the ouliers and their extracted trends are deviated from the ground truth greatly. We further test these filters on the synthetic data with more outliers and summarize the results in Table 1. Similar to the conclusion in

Figure 6, except for RobustFilter, they all are vulnerable to the outliers and their both mean squared error (MSE) and mean absolute error (MAE) increase exponentially. This simulation points out that RobustFilter is capable of extracting desired slow and abrupt trend without being sensitive to the outliers.

It is remarkable that the preference of the extracted trend by RobustFilter is determined on the choice of parameters λ_1 and λ_2 in the construction of RobustFilter in (4.7). Relatively high λ_1 means higher preference for abrupt trend while relatively high λ_2 represents higher bias for the smooth trend. Since the original signal in this simulation consists both two kinds of trends, we set $\lambda_1 = 0.6$ and $\lambda_2 = 0.4$ in practice.

Table 1: Performance of different trend filters on synthetic data with different ratios of outliers.

Metirc	MSE			MAE		
Outlier Ratio	4%	10%	20%	4%	10%	20%
RobustFilter	0.010	0.011	0.013	0.061	0.064	0.068
l1 Filter	0.051	0.100	0.416	0.091	0.133	0.289
HP-Filter	0.047	0.088	0.322	0.110	0.171	0.329
TVD	0.032	0.054	0.267	0.087	0.120	0.250

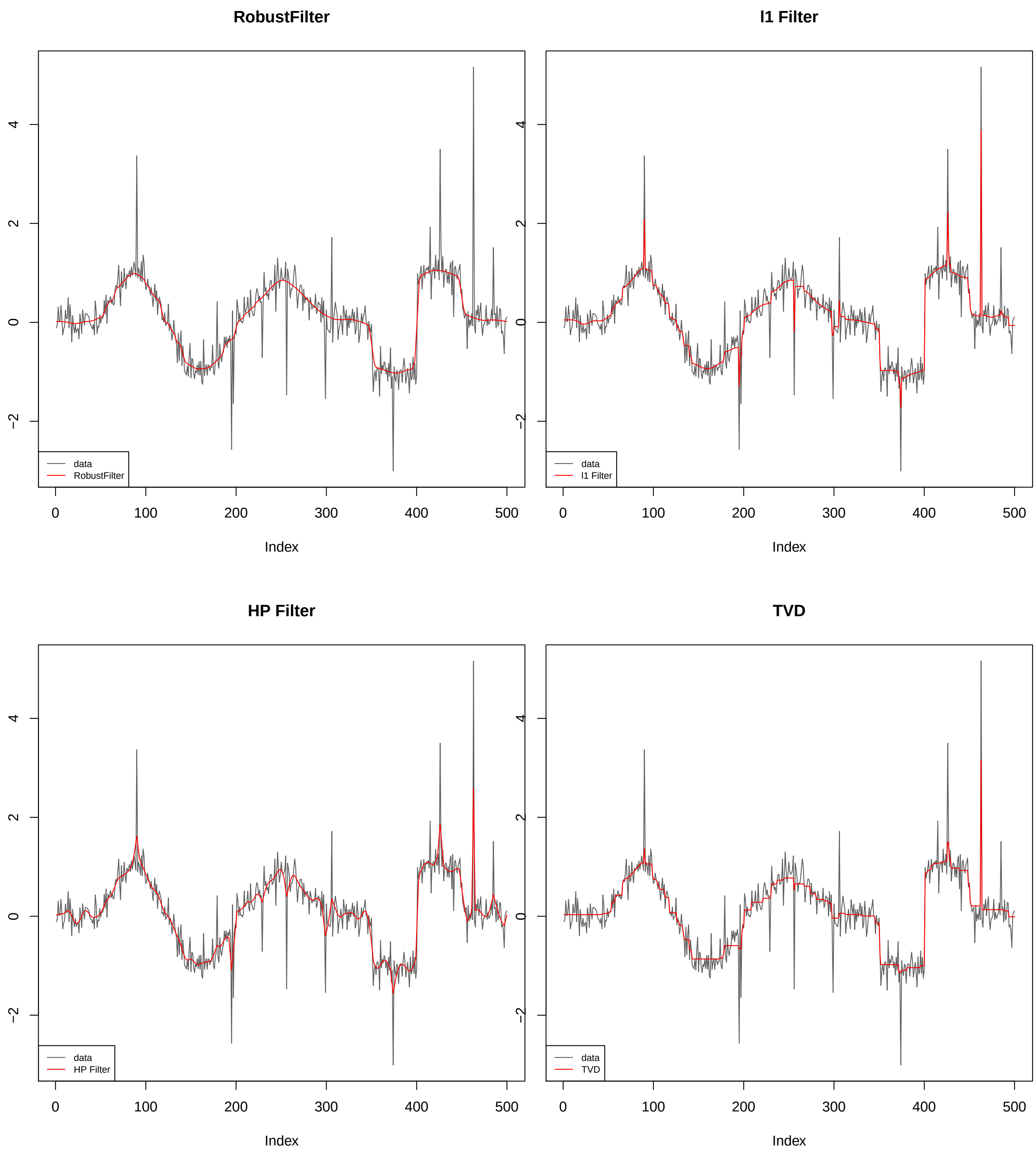


Figure 6: Trend extraction for synthetic data with 4 percent outliers by different filters.

4.3.2 Real World Dataset: Covid-19 New Cases

Dataset Description For real data application, we select the Covid-19 new cases data in Italy as our analysis object. [2]Specifically, we use the records from Aug. 21 2020 to Jun. 17 2021, spanning 300 days. As shown in Figure 7, the raw data curve fluctuates with a bunch of outliers, which may account for the missing value. Since the raw data ranges from zero to 40,000, we implement log transformation before our formal analysis, which is also shown in Figure 7. (To avoid the occurrence of Nans, we add the raw data with one before log transformation.)

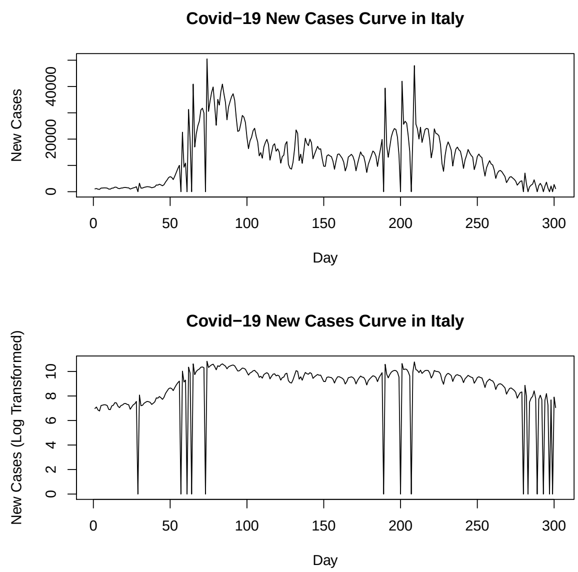


Figure 7: Covid-19 New Cases Curve in Italy

We apply all four filters on the log transformed times series data and obtain the results in Figure 8. Corresponding the results and analysis in Section 4.3.1, RobustFilter outperforms other filters and extracts a smooth and accurate trend without getting distracted by the outliers. Moreover, we also plot the residual of each filters (may include cyclic terms) as the yellow lines in Figure 8. We observe that if we ignore the points outside the 3-SD line (three times standard error line), which we assume to be outliers, then the residual of RobustFilter

[2]The data is avaliable on https://www.dydata.io/datastore/search/

is mostly a stationary timeseries. The residuals of $l1$ Filter and HP Filter are still fluctuating on certain position, because their extracted trend is not correct. Here we neglect TVD Filter, since it is just misled by the outliers and the extracted trend is not good at all.

Next we foucs the residuals of RobustFilter, as depicted in Figure 8. Delete all the extreme ouliers which are not contained in 3-SD line, then we derive a rather stationary time series. With the help of the function $auto.arima$, we find that the residual of RobustFilter is MA(1) time series and the model information is listed in Table 2, which demonstrates that MA(1) is reasonable and the residual of MA(1) model passes the Box-Pierce with a high P-value (close to 1). It also shows the effectiveness of RobustFilter in extracting both the slow and abrupt trend in real data application, which may involve outliers and dips.

Table 2: Time Series Model for the Residuals of RobustFilter

Model	AIC	BIC	$\widehat{\sigma}^2$	Box-Pierce Test P Value
MA(1)	-24.5	-17.09	0.05432	0.8425

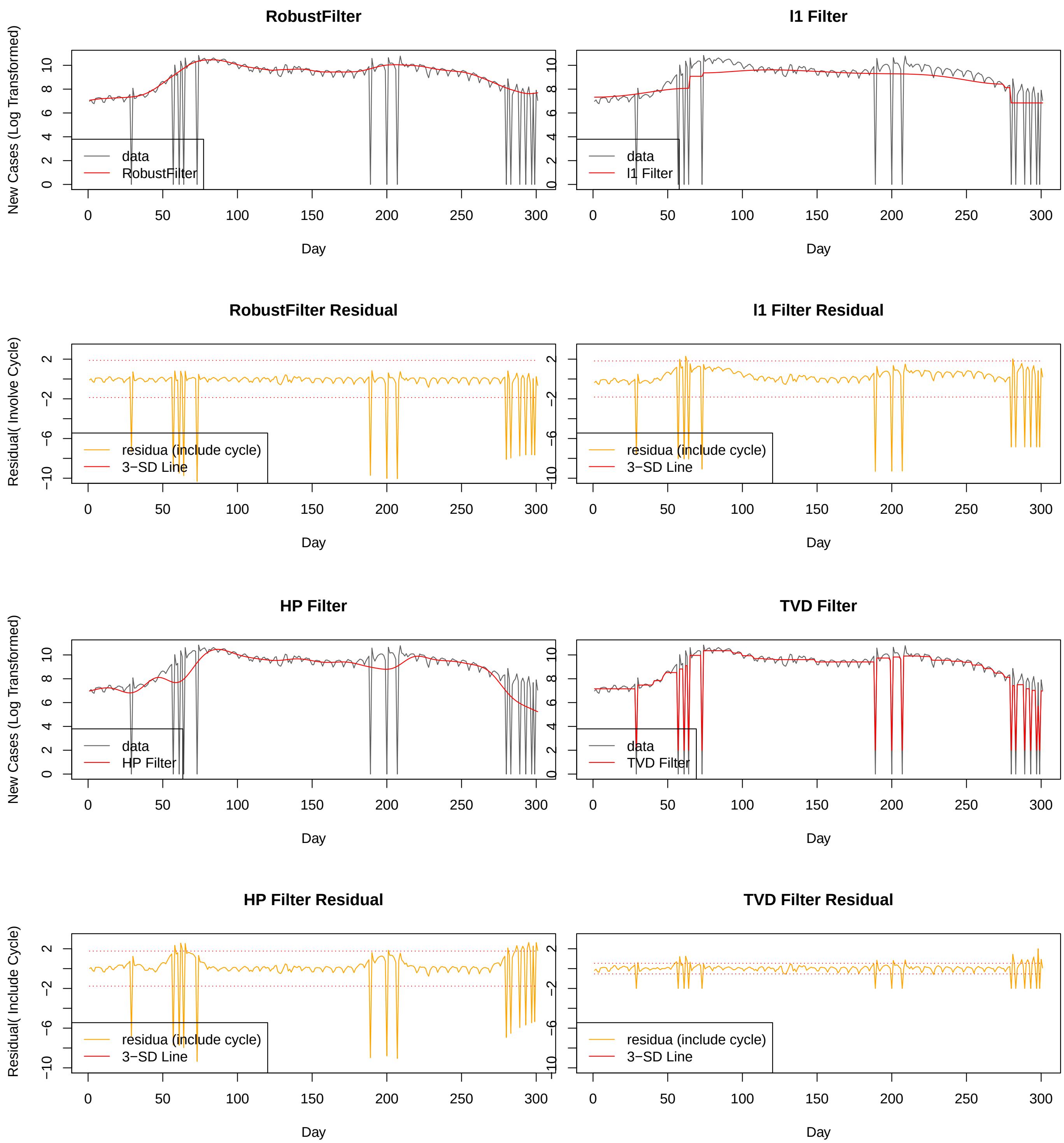


Figure 8: Trend extraction for Covid-19 new cases curve in Italy of different filters.

5 Conclusion

In this paper, we investigate the trend-cycle separation problem using different filters. We discussed the construction of different filters and how these filters try to solve the problem in different aspects. We implement the discussed filters in each section and test them on real-world data as well as making comparisons. The detailed discussion of filters and comprehensive experiment in the paper enable people to get an in depth sight of these filters and the separation problem.

References

Baxter, M. and King, R. G. (1995). Measuring business cycles approximate band-pass filters for economic time series. Working Paper 5022, National Bureau of Economic Research. http://www.nber.org/papers/w5022

Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, **3** 1–122. https://doi.org/10.1561/2200000016

Bruchez, P. (2003). A modification of the hp filter aiming at reducing the end-point bias. swiss federal finance administration working paper no. Tech. rep., ÖT/2003/3, August.

Christiano, L. J. and Fitzgerald, T. J. (2003). The band pass filter. *international economic review*, **44** 435–465.

De Leeuw, J. and Lange, K. (2009). Sharp quadratic majorization in one dimension. *Computational statistics data analysis*, **53** 2471–2484.

Hodrick, R. J. and Prescott, E. C. (1997). Postwar u.s. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, **29** 1–16. http://www.jstor.org/stable/2953682

Kim, S.-J., Koh, K., Boyd, S. P. and Gorinevsky, D. (2009). l1 trend filtering. *SIAM Rev.*, **51** 339–360.

McDermott, J. (1997). An automatic method for choosing the smoothing parameter in the hp filter. *Unpublished, International Monetary Fund, Washington, DC.*

Schüler, Y. S. (2019). How should we filter economic time series? *Available at SSRN 3366923*.

Selesnick, I. (2014). Total variation denoising (an mm algorithm).

Wang, H., Li, G. and Jiang, G. (2007). Robust regression shrinkage and consistent variable selection through the lad-lasso. *Journal of Business Economic Statistics*, **25** 347–355.

Wen, Q., Gao, J., Song, X., Sun, L. and Tan, J. (2019). Robusttrend: A huber loss with a combined first and second order difference regularization for time series trend filtering.