

HP Trend Filtering Using Gaussian Mixture Model Weighted Heuristic

Luiza Sayfullina*, Magnus Westerlund†, Kaj-Mikael Björk† and Hannu T. Toivonen‡

*Department of Information and Computer Science

Aalto University, Espoo, Finland

†Department of Business Management and Analytics

Arcada University, Helsinki, Finland

‡Department of Information Technologies

Åbo Akademi University, Turku, Finland

Abstract—Trends show the underlying structure of the time-series data. Trend estimation is a commonly used tool for financial market movement prediction. In traditional approaches, such as Hodrick-Prescott (HP) and L1 filtering, the trend is considered as a smoothed version of the time-series, including rare significant hills that are smoothed in the same way as usual noise. The goal of this paper is to allow the estimated trend to be more complex and detailed in the intervals of significant changes while making a smooth estimate in all other parts. This will be our main criteria for trend estimation. We present a modified version of HP weighted heuristic that provides the best trend according to the abovementioned criteria. Gaussian Mixture Models (GMMs) on the preliminary estimated trend are used in the weighted HP heuristic to decrease the penalty in the objective function for turning-point intervals. We conducted a set of experiments on financial datasets and compared the results with those obtained from the standard HP filtering with weighted heuristic. The results indicate an improvement in the cycling component using our proposed criteria compared to the HP filtering approach.

Keywords-HP Trend; L1 Trend; HP Weighted Heuristic; Time-Series

I. INTRODUCTION

Trend extraction is an important step in understanding the underlying structure of time-series data. The application areas of it are diverse, including finance and economics [4], signal-processing [6], ecology [8], etc. The main trend extraction approaches currently used are: model-based (ARIMA), nonparametric linear filtering (Henderson, LOESS, Hodrick - Prescott) [3], singular spectrum analysis [2] and wavelet-based methods [3]. Particularly in financial time-series, L1 [12], Kalman [16] and HP filtering [10] are used.

A trend is viewed as a smoothed version of the time series that depicts global changes. It is not supposed to take into account seasonal changes and noise. During trend extraction, excessive smoothing can happen, wherein not only noise but sharp slopes are smoothed. We aim to find such trends that will smooth globally important changes less than the noise. Meinel [13] proposed these criteria for trend analysis. In his approach, turning-point intervals were detected using wavelet coefficients and later, approximated in detail using

local linear scaling approximation. We use the term *turning-point* in our paper as the local minimum or maximum point of the curve. The rest of the coordinates in the time series vector are smoothed using the usual linear filter.

The main goal of this paper is to modify the HP filtering weighted heuristic so that it keeps its general smoothing property while emphasizing larger structural changes in the time-series. These changes, that can denote, for example, crisis or growth, are an integral part of the process and they may not be regarded as noise. This motivates applying adaptive smoothing over the different periods. We chose HP filtering due to the small number of assumptions needed for constructing the model, linear computational complexity, and the ability to emphasize turning-point intervals by controlling the weight.

Hodrick and Prescott [10] introduced the HP filter for finding trends. One of the major drawbacks of the filter is that it smooths the entire time-series equally, which results in excessive smoothing of deep structural recessions or growths. Hence, a reweighting scheme is required to control the level of smoothing over different periods of time. During the recessions, it is favorable to avoid excessively penalizing smoothing in the intervals of high variations, while during the normal periods, a stronger smoothing is more desirable.

In [5] an iterative weighted heuristic for the HP filter was introduced. They used the weights in the regularization part of the objective function of L1 trend filtering that penalizes the model complexity through second order differences. Weights are updated in each iteration, eventually yielding a lower RMSE than the usual L1-filtering. Our method was inspired by the idea of reweighting the second order differences in the objective function, that we will refer to as the *weighted heuristic* in the algorithm description. This weighted heuristic helps manage the the level of approximation for trend at different periods of time.

The outline of the paper is the following. In Section 2, we provide a brief recap of L1 and Hodrick-Prescott filtering and iterative weighted heuristic. Then, in Section 3, we describe how important turning-points are detected and how weights for HP filtering are constructed. Section 4 provides implementation details and experimental results on financial

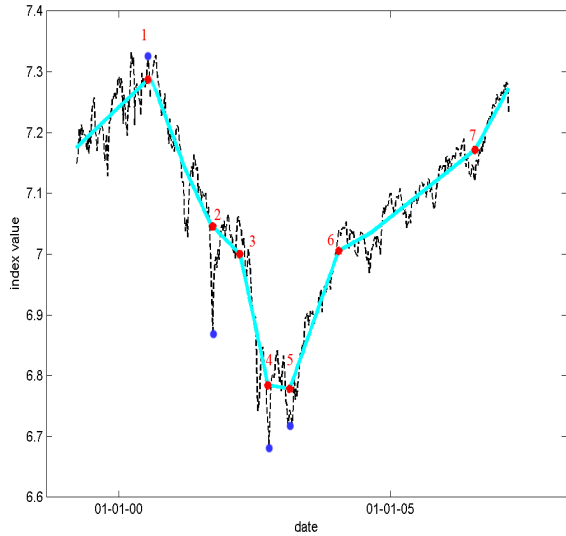


Figure 1: Kink-points for the S&P500 dataset

data. The concluding remarks are made in Section 5.

II. L1 AND HODRICK-PRESCOTT TREND FILTERING

We begin by providing two important definitions.

Turning-point is the local maximum or the local minimum of the curve.

Kink-point is the point where piecewise-linear function changes its slope.

Kink-points are illustrated in Fig. 1. The figure illustrates all the kink-points that were detected by L1 trend estimation on one of our experimental datasets. It can be observed that not all of the points seem to be so important, however points 1, 2, 4 and 5 seem to be very significant, showing growth and recession, respectively. With blue dots, we indicate the examples of turning-points that we regard as important ones. We see that by detecting kink-points on the trend, it is possible to find the important turning-points.

In the financial datasets that we use for our experiments, there exist several turning-points. However, we are interested only in the points indicating recession or growth. There is no exact definition which points to consider important. Nonetheless, the human eye can recognize such points that have unexpected low or high values for a certain period of time, assuming that there exist no outliers in the data. Later, we will show how L1 trend estimation can help detect these points through its kink-points.

A. Hodrick-Prescott Filtering

Hodrick-Prescott filtering [12] has been extensively used in business literature for extracting the trend component to find ‘cycles’. It divides the time series into trend and cycle components. Despite the fact that HP filters theoretically

give the cycling component for infinitely large signals, it has been proved in [7] that it gives suboptimal solutions for time series where only a part of the cycle is presented.

HP filter provides a smooth estimate of the trend, deliberately penalizing kink-points. The objective function to be minimized is defined as follows:

$$f(x, y, \lambda) = \frac{1}{2} \sum_{t=1}^n (y_t - x_t)^2 + \lambda \sum_{t=2}^{n-1} (x_{t-1} - 2x_t + x_{t+1})^2. \quad (1)$$

We denote by y , the input variable that is our original time-series vector and by x , we define the output variable of length n that is our estimated trend. We denote by D , the $(n-2) \times n$ matrix of the second-order operator,

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

Using the vector of second-order differences of x , denoted by Dx , we can reformulate the problem in a more compact form as follows:

$$\min_x f(x, y, \lambda) = \min_x \frac{1}{2} \|y - x\|_2^2 + \lambda \|Dx\|_2^2. \quad (2)$$

Tibshirani [17] extended trend filtering to k -th order differences. For example, first order difference gives piecewise constant functions, first order piecewise-linear functions and second-order quadratic functions. He showed that the k -th order difference gives k -th order piecewise polynomials, which provide high flexibility in trend estimation along with a linear computational complexity.

The objective of this model is to find the trade-off between its complexity and the approximation error. Point x_t is a kink-point if $x_{t-1} - 2x_t + x_{t+1} \neq 0$, i.e., the second order difference is not zero. The larger the slope, the larger its second order difference. Since both functions are convex, the resulting model ends up with a smooth trend estimate.

The trend found by HP filtering has a closed form solution and the complexity is linear $\mathcal{O}(n)$:

$$x^{HP} = (I + 2\lambda D^T D)^{-1} y. \quad (3)$$

B. L1 trend Filtering

L1 trend filtering proposed in [12], as a variation of Hodrick-Prescott filtering, has the following objective function $f(x, y, \lambda)$ to be minimized:

$$\begin{aligned} \frac{1}{2} \sum_{t=1}^n (y_t - x_t)^2 + \lambda \sum_{t=2}^{n-1} |x_{t-1} - 2x_t + x_{t+1}| \\ = \frac{1}{2} \|y - x\|_2^2 + \lambda \|Dx\|_1. \end{aligned} \quad (4)$$

L1 trend is a piecewise-linear function due to the L1 norm in the objective function. Unlike HP filtering, L1 filtering has no exact analytical solution. However, the average complexity is also linear [12].

L1 trend filtering was extended with the iterative weighted L1 heuristic in [5]. Reweighting the second part of the objective function, we can decrease the penalty for kink-points, where second order difference is non-zero.

The objective function to be minimized of L1 filtering takes the form:

$$f(x, y, \lambda) = \frac{1}{2} \|y - x\|_2^2 + \lambda \|W D x\|_1, \quad (5)$$

where,

$$W = \text{diag}(w). \quad (6)$$

The weight vector is defined in the following way:

$$w_t = \frac{1}{\epsilon + |D x_t|}, \quad t = 1, 2, \dots, n-2 \quad (7)$$

The strength of this reweighting scheme is in its ability to control kink-points. ϵ is chosen to be a small value. The weights for non-kink-points become so large that only kink-points can be controlled. Each iteration of the reweighting scheme consists of solving (5) and updating the weight vector w for the next iteration. In [5], the authors state that a choice of ϵ should be close to the expected second order difference of the estimated trend. They claim that the method is robust to the choice of ϵ . In practice, $\epsilon = 0.1$ works well. In the next section, we give the intuition behind our suggested algorithm, describe how we use both L1 and HP filtering and explain the modification of the iterative weighted heuristic using Gaussian Mixture Models.

III. MODIFIED WEIGHTED HEURISTIC FOR HP FILTERING

The motivation of this paper is to be able to decrease the penalty for intervals around important turning-points. For that, we need some method for turning-point detection. Finding these points based on the original time-series can be uninformative, since there are a lot of local minima or maxima and not all of them are significant on the global level. Applying L1 trend estimation in the first stage will result in a piecewise linear trend, from which, it is easier to detect significant kink-points. Indeed, L1 trend estimation with a predefined λ gives us an approximation that captures changes linearly within a certain range of resolution, specified by the number of kink points. This means that small fluctuations will not be captured and kink-points for piecewise linear L1 trend estimates serve as the turning-points. Hence, intervals around kink-points can be represented in greater detail in the corresponding trend through proper initialization of the weights in HP filtering with a suitable reweighting scheme. HP filtering in the second stage will be a better choice due to its ability to provide smooth non-linear results and make better approximations. The algorithm for HP filtering using GMMs for reweighting is outlined in the following steps:

Step 1. Find the first approximate solution x by L1 trend filtering using (4). Start with $\lambda = \lambda_0$, where λ_0 is sufficiently large. Find the point where the error is the largest. Set $\lambda_{i+1} = \lambda_i/2$ until the largest approximation error decreases significantly or the number of iterations i exceeds a predefined value (e.g., 5 iterations).

Step 2. Set $\lambda' = gal * \lambda$, where gal is the global approximation level. Average values for gal lie in the range [100, 1000] and are determined experimentally.

Step 3. Apply HP filtering with λ' using (1). Find trend x and calculate the weight vector w using (12). Turning-points are found using (14).

Step 4. Use calculated weight vector w in the proposed weighted heuristic with HP filter from (12) using a single iteration and evaluate resulting trend y .

In Step 1, we aim to find a λ such that the point of recession or growth is approximated better than obtained by the initial value λ_0 . If such a λ can not be achieved, we will stop after a number of iterations. Then, in Step 2, we need to increase λ , found by L1 trend estimation for HP trend estimation. HP filtering needs higher values of λ , than L1 trend filtering to avoid excessive smoothing.

We apply HP weighted heuristic in Steps 3 and 4. In Step 3, initial HP filtering with weight vector $w = \mathbf{1}$, where, $\mathbf{1}$ is the all-one vector and λ' from Step 2 are applied. Using the trend x and the weight vector w from Step 3, we apply the weighted heuristic and obtain the final solution in Step 4.

Computational complexity of the constructed algorithm is linear, as it consists of several runs of L1 and HP filtering. The issue of complexity is especially taken into account in High-Frequency trading systems, where time-series size is extremely large.

Now after providing the outline, we discuss the details of the calculation of the reweighting vector in Step 3.

A. Gaussian Mixture Model

A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities [14]. This function has the following form:

$$p(x|\theta) = \sum_{i=1}^M \pi_i \mathcal{N}(x|\mu_i, \sigma_i), \quad (8)$$

where,

$$\theta = \{\pi_i, \mu_i, \sigma_i\}_{i=1}^M \quad (9)$$

Here, $\mathcal{N}(x|\mu, \sigma)$ is a Gaussian probability density function,

$$\mathcal{N}(x, \mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right), \quad (10)$$

π is the weight vector, such that $\sum_{i=1}^M \pi_i = 1$ and M is the number of components. GMMs are able to approximate any arbitrarily form of density function up to any desired

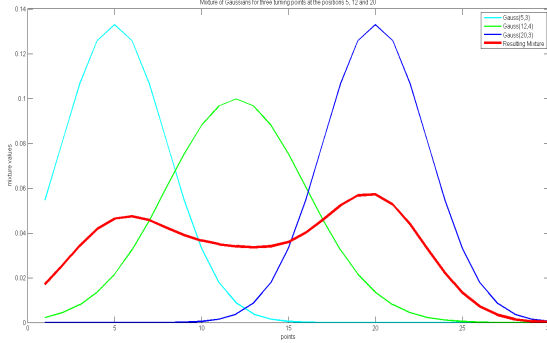


Figure 2: The example of the Gaussian Mixture Model in the discrete form

precision. Fig. 2 represents a mixture of 3 Gaussians. GMM is used for constructing weight vector in iterative weighted heuristic. Points that are situated closer to the turning-points should have gradually decreasing weight penalty in the objective function (5). This allows us to show not only the turning-points in more detail, but also the intricacies in the surrounding intervals. The smooth gradual decrease can be approximated by the Gaussian function with its mean centered at the turning-point. We allow the complexity to grow around the kink-points as well. Coordinate i is the abscissa of the turning-point x_i (for L1 trend); σ_i is estimated by fitting Gaussian distribution to vector v around the kink/turning-point x_i within a predefined *band* :

$$v = (x_{i-band}, \dots, x_i, \dots, x_{i+band}). \quad (11)$$

This vector v represents the turning-point interval around x_i . Instead of the vector with the absolute second differences $|Dx|$ as in the traditional approach, we now have a GMM with unnormalized weights σ .

$$w = \frac{1}{1 + lal \sum_{kink_i=1}^K \sigma_{kink_i} \mathcal{N}^*(x, \mu_{kink_i}, \sigma_{kink_i})} \quad (12)$$

$$\mathcal{N}^*(x_j, \theta_i) = \begin{cases} \mathcal{N}(x_j, \theta_i), & j = i - band, \dots, i + band. \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

In (13), $\mathcal{N}^*(x_j, \theta_i) = \mathcal{N}(x, \mu_{kink_i}, \sigma_{kink_i})$ denotes the truncated vector of Gaussian probability density function of input time series $x = (x_1, \dots, x_n)$ for the kink-point $kink_i$. That is, each Gaussian is evaluated only within some *band* around the kink-point, i.e., $i - band, \dots, i + band$, and the value in all other points outside *band* is set to zero.

In the Gaussian distribution, the smaller the standard deviation, the larger the values of the probabilities towards the center of the distribution. That is, if we want to keep the weights approximately in the same scale for time series with different y-scaling, it is beneficial to multiply the values

of the probability densities by the standard deviation σ of this Gaussian. Hence, the scale of the values in the y-axis constrains the weight vector w to be in the same order of magnitude. Standard deviation σ serves as the weight of the Gaussian, however, $\sum_{i=1}^K \sigma_i \neq 1$. For points x_j outside kink-point interval, $w_j = 1$.

Parameter *lal*, called the *local approximation level*, controls the extent to which we want to approximate jumps. We found that $lal \in [10, 100]$ provides a good approximation. Higher *lal* values result in better approximations. As opposed to the local approximation level, the global approximation level *gal* is responsible for the overall approximation of the time series. A commonly used range for the parameter *gal* is $[100, 1000]$. These two parameters provide us flexibility on the local and global scales, respectively. The mean of the Gaussian μ is the kink-point and the standard deviation is calculated by fitting kink-point interval v within a predefined band (*band*) of the Gaussian distribution. We must emphasize here that not every point with non-zero second order difference is an important kink-point. Indeed, important kink-points should have significantly larger second order difference than average value for all the points. In order to determine which points are kinks, we employed the z-score, thereby assuming that the second order differences are normally distributed.

$$z_i = \frac{Dx_i - \overline{Dx}}{\sigma}, \quad i = 2, \dots, n - 1 \quad (14)$$

Z-score is one of the outlier detection methods. Points x_i with $z_i > 3$ are considered to be outliers or important kink-points in our case [9]. This is because the second-order difference for kink-points is much larger than that for other points. We should note that turning-point detection for curves is a particularly difficult problem [18] and does not have an exact solution. The decision about what points are significant on the global level depends on the user. Z-score statistics provide fairly good solutions, as our experiments suggest.

IV. EXPERIMENTAL RESULTS

In this section, we provide experimental results on two financial datasets. Datasets of different sizes were chosen: Euro Dollar Exchange Rate and S&P 500 Index. We ran our experiments on an Intel Core-i7 processor, 1.90-2.40 GHz. Code was written in Matlab using the CVX library for Disciplined Convex Programming tasks ¹ developed at Stanford University.

A. Comparison with Standard HP Filtering

We used the aforementioned datasets to compare our proposed trend filtering to HP filtering and to demonstrate its robustness. In our experiments, we demonstrate that HP filtering using Gaussian Mixture Models is able to detect

¹cvxr.com

and approximate certain important intervals, such as strong recession and growth periods.

The **S&P 500** dataset is a calculated index based on the market capitalization of 500 companies in leading industries of the U.S. economy, which are publicly held on either the NYSE or NASDAQ. It has been commonly used as a benchmark for finding trends [12]. We used weekly data and the logarithm of opening values from March 1999 until March 2007 for comparison with experiments performed in [12]. The dataset was downloaded from Yahoo Finance database ².

The Euro Dollar Exchange Rate is a dataset with daily conversion rates. Values from the period of June 2011 to March 2014 were chosen. The dataset was chosen for its high-frequency fluctuations.

In the first set of experiments we compared our proposed method to the standard HP filtering. In all experiments we used $band = 10$ and $\lambda_0 = 50$. Local and global approximation levels were different. The default level for local approximation level lal was set to 30 and $gal = 500$, unless otherwise stated. The maximum number of iterations during Step 1 of the algorithm was 5. The parameter λ stated in the figures' caption denotes the value that was used in the final HP filtering in Step 4 of the algorithm.

In Fig. 3(a) and 3(b), we observe how the proposed modified HP filtering results in better approximations of important turning-points while leaving the rest of the time-series untouched. Especially in the S&P 500 dataset, the area of deep recession is shown very prominently unlike in usual filtering, shown in a blue. For the EUR/USD dataset, more intervals are found due to an inherently more volatile price structure.

Our proposed method is robust towards scaling on the y-axis. Particularly, that means that means that if we scale the y-axis by multiplying by constant c , the resulting trend will not change. This flexibility is achieved by the reweighting scheme, where each Gaussian Probability Density vector is multiplied by the mean standard deviation σ using (12). Both trends are almost the same before and after the scaling. Another important criterion was to decrease the large deviation between the original time series and its trend and to improve the cycling component. As we already mentioned, HP filtering gives a suboptimal 'cycle' solution. By capturing big jumps better, the cycling component becomes more cyclic and prevents large deviations of the trend from the original time series. The complexity of the model does not increase much because other parts of the time series are approximated in the same way as usual filtering. The resulting cycling component can be seen in Fig. 6(b). On the S&P500 index dataset, we see that high deviation peaks exceeding 0.1, were reduced in cycle so that locally, they are not outliers.

²<http://finance.yahoo.com/>

The advantage of our proposed filtering method is in its ability to control the level of detail globally, through the whole time series and locally, in important turning-points. The local approximation level lal can not be defined in standard HP filtering. It controls how exact the approximation will be in important intervals.

In Fig. 5(a) and 5(b), we compared the trends with $lal = 100$ to $lal = 30$, obtained in the previous examples. Higher lal values gave us better approximation of deep recession periods.

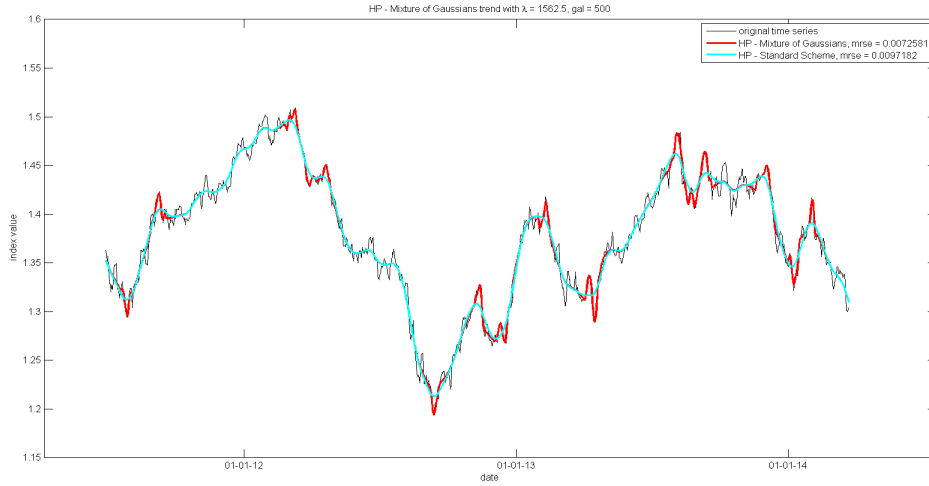
Trend selection can not be done easily using standard criteria such as AIC [1] or BIC [11], that do not consider the number of parameters and error - the reason being that the number of parameters in the model can not be clearly described. If we had a piecewise linear trend, then we could take the number of kink-points into consideration. However, when the approximated function is non-linear, then, second-order differences do not estimate the complexity well enough because, almost all points are kink-points. Therefore, second order differences are mostly non-zero. That is why, we prefer to have the user choose the model by defining local and global approximation levels. Choosing parameters provides flexibility and the ability to describe the data in the most appropriate form. For short term approximation, for example, we would like to have a lower global approximation level which results in trends that are more detailed.

V. CONCLUSION

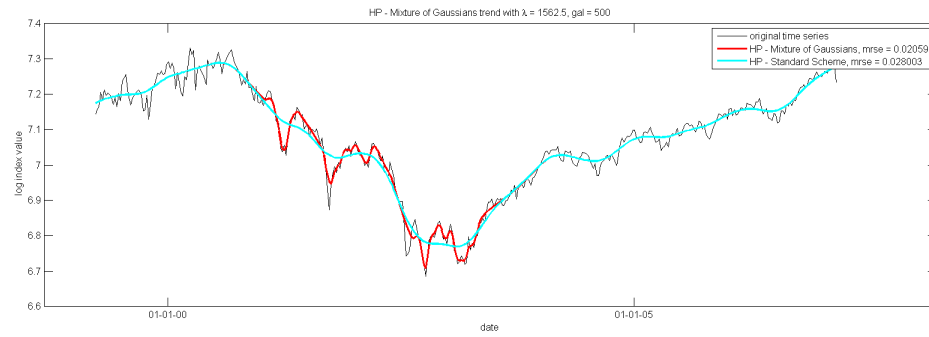
In this paper, we proposed new criteria for trend estimation and modified the existing HP weighted heuristic filtering. Weights were constructed using GMMs on the estimated trend, providing a flexible smoothing of the trend. Our smoothing scheme is able to emphasize important turning-points more precisely in the trend component while improving the cycling component by decreasing extremely high or low deviations. It also provides users the flexibility in trend filtering. Our model takes into account the preferences of the users regarding global and local approximation levels, instead of using statistical criteria. In future work, we aim to improve important turning-point detection and evaluate the effect of using a higher order difference matrix in the objective function.

REFERENCES

- [1] Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.
- [2] Theodore Alexandrov. A method of trend extraction using singular spectrum analysis. *RevStat*, 7:1–22, 2009.
- [3] Theodore Alexandrov, Silvia Bianconcini, Estela Bee Dagum, Peter Maass, and Tucker McElroy. A review of some modern approaches to the problem of trend extraction. *Econometric Reviews*, 31, 2012.

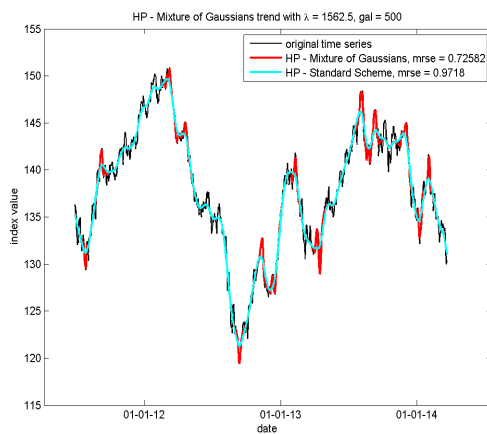


(a) EURO/USD dataset

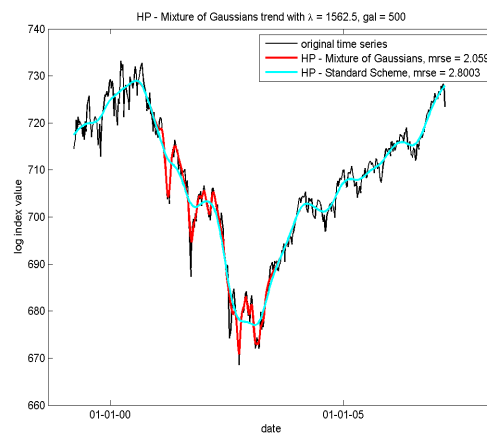


(b) S&P500 dataset

Figure 3: Comparison of standard and our proposed HP filtering

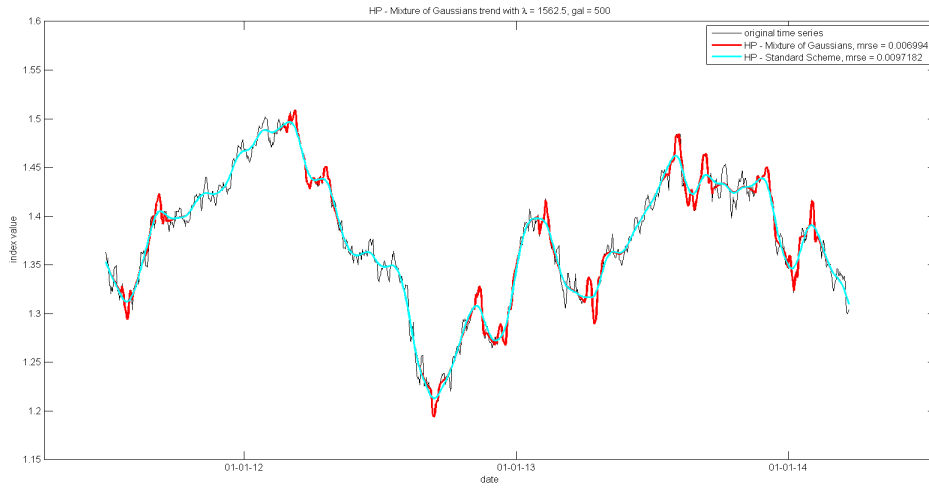


(a) EURO/USD dataset

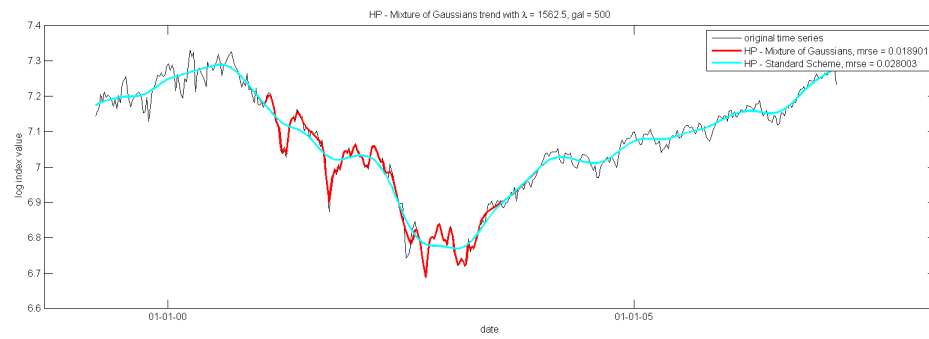


(b) S&P500 dataset

Figure 4: Robustness towards scaling

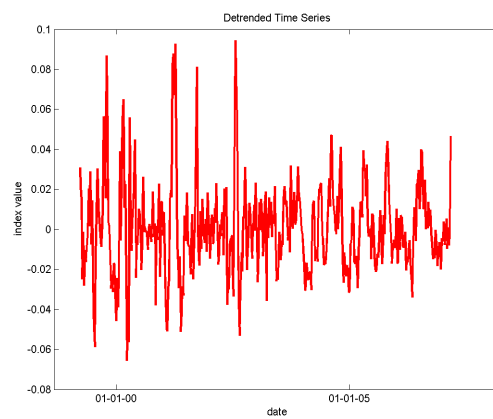
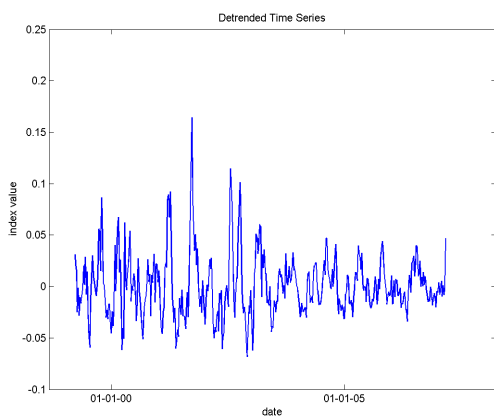


(a) EURO/USD dataset



(b) S&P500 dataset

Figure 5: Trends with increased $l_{al} = 100$



(a) Cycling component on S&P500 dataset using standard HP filter (b) Cycling component on S&P500 dataset using HP Mixture of Gaussians filter

Figure 6: Cycling component

- [4] Benjamin Bruder, Tung-Lam Dao, Jean-Charles Richard, and Thierry Roncalli. Trend filtering methods for momentum strategies.
- [5] Emmanuel J. Cands, Michael B. Wakin, and Stephen P. Boyd. Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905, 2007.
- [6] Zhang Kaibin Ma Ying Duan Hu-ming, Xie Fei and Shi Feng. Signal trend extraction of road surface profile measurement. *Signal Processing Systems*, 2, 2010.
- [7] Paul Newboldc Emi Misea, Tae-Hwan Kimb. On suboptimality of the hodrickprescott filter at time series endpoints. *Journal of Macroeconomics*, 27, 2005.
- [8] Sylvia R. Esterby. Review of methods for the detection and estimation of trends with emphasis on water quality applications. *Hydrological Processes*, 10(2):127–149, 1996.
- [9] Victoria Hodge and Jim Austin. A survey of outlier detection methodologies. *Artif. Intell. Rev.*, 22(2):85–126, October 2004.
- [10] Robert J. Hodrick and Edward Prescott. Post-War U.S. Business Cycles: An Empirical Investigation. Discussion papers, Northwestern University, Center for Mathematical Studies in Economics and Management Science, 1981.
- [11] Robert E. Kass and Larry Wasserman. A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the American Statistical Association*, pages 928–934, 1995.
- [12] Seung-Jean Kim, Kwangmoo Koh, Stephen Boyd, and Dmitry Gorinevsky. L1 trend filtering. *SIAM review*, 51(2):339–360, 2009.
- [13] Thomas Meinl. *A Novel Wavelet Based Approach for Time Series Data Analysis*. Ph.d. thesis, Karlsruhe Institute of Technology, 2011.
- [14] Douglas A. Reynolds. Gaussian mixture models. *Encyclopedia of Biometrics*, pages 659–663, 2009.
- [15] Anna Louise Schrder and Piotr Fryzlewicz. Adaptive trend estimation in financial time series via multiscale change-point-induced basis recovery. Mpra paper, University Library of Munich, Germany, 2013.
- [16] Bruno Sinopoli, Student Member, Luca Schenato, Massimo Franceschetti, Kameshwar Poola, Michael I. Jordan, Senior Member, and Shankar S. Sastry. Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control*, 49:1453–1464, 2004.
- [17] Ryan J. Tibshirani. Adaptive piecewise polynomial estimation via trend filtering. *Annals of Statistics 2014*, 42(1), 2014.
- [18] Yazhen Wang. Jump and sharp cusp detection by wavelets. *Biometrika*, 82(2):385–397, 1995.