Momentum Strategies with L_1 Filter

Tung-Lam Dao
Capital Fund Management, Paris
tung-lam.dao@cfm.fr

March 18, 2014

Abstract

In this article, we discuss various implementation of L_1 filtering in order to detect some properties of noisy signals. This filter consists of using a L_1 penalty condition in order to obtain the filtered signal composed by a set of straight trends or steps. This penalty condition, which determines the number of breaks, is implemented in a constrained least square problem and is represented by a regularization parameter λ which is estimated by a cross-validation procedure. Financial time series are usually characterized by a long-term trend (called the global trend) and some short-term trends (which are named local trends). A combination of these two time scales can form a simple model describing the process of a global trend process with some mean-reverting properties. Explicit applications to momentum strategies are also discussed in detail with appropriate uses of the trend configurations.

Keywords: Momentum strategy, L_1 filtering, L_2 filtering, trend-following, mean-reverting.

JEL classification: C01, C60, G11.

1 Introduction

Trend detection is a major task of time series analysis from both mathematical and financial point of view. The trend of a time series is considered as the component containing the global change which is in contrast to the local change due to the noise. The procedure of trend filtering concerns not only the problem of denoising but it must take into account also the dynamic of the underlying process. That explains why mathematical approaches to trend extraction have a long history and this subject still gives a great interest in the scientific community [1]. In an investment perspective, trend filtering is the core of most momentum strategies developed in the asset management industry and the hedge funds community in order to improve performance and to limit risk of portfolios.

The paper is organized as follows. In section 2, we discuss the trend-cycle decomposition of time series and review general properties of L_1 and L_2 filtering. In section 3, we describe the L_1 filter with its various extensions and the calibration procedure. In section 4, we apply L_1 filters to some momentum strategies and present the results of some backtests with the S&P 500 index. In section 5, we discuss the possible extension to the multivariate case and we conclude in the last section.

[1] For a general review, see Alexandrov *et al.* (2008).

2 Motivations

In economics, the trend-cycle decomposition plays an important role to describe a non-stationary time series into permanent and transitory stochastic components. Generally, the permanent component is assimilated to a trend whereas the transitory component may be a noise or a stochastic cycle. Moreover, the literature on business cycle has produced a large number of empirical research on this topic (see for example Cleveland and Tiao (1976), Beveridge and Nelson (1991), Harvey (1991) or Hodrick and Prescott (1997)). These last authors have then introduced a new method to estimate the trend of long-run GDP. The method widely used by economists is based on L_2 filtering. Recently, Kim *et al.* (2009) have developed a similar filter by replacing the L_2 penalty function by a L_1 penalty function.

Let us consider a time series y_t which can be decomposed by a slowly varying trend x_t and a rapidly varying noise ε_t process:

$$y_t = x_t + \varepsilon_t$$

Let us first remind the well-known L_2 filter (so-called Hodrick-Prescott filter). This scheme consists to determine the trend x_t by minimizing the following objective function:

$$\frac{1}{2}\sum_{t=1}^{n}(y_t - x_t)^2 + \lambda \sum_{t=2}^{n-1}(x_{t-1} - 2x_t + x_{t+1})^2$$

with $\lambda > 0$ the regularization parameter which control the competition between the smoothness of x_t and the residual $y_t - x_t$ (or the noise ε_t). We remark that the second term is the discrete derivative of the trend x_t which characterizes the smoothness of the curve. Minimizing this objective function gives a solution which is the trade-off between the data and the smoothness of its curvature. In finance, this scheme does not give a clear signature of the market tendency. By contrast, if we replace the L_2 norm by the L_1 norm in the objective function, we can obtain more interesting properties. Therefore, Kim *et al.* (2009) propose to consider the following objective function:

$$\frac{1}{2}\sum_{t=1}^{n}(y_t - x_t)^2 + \lambda \sum_{t=2}^{n-1}\left|x_{t-1} - 2x_t + x_{t+1}\right|$$

This problem is closely related to the Lasso regression of Tibshirani (1996) or the L_1 regularized least square problem of Daubechies *et al.* (2004). Here, the fact of taking the L_1 norm will impose the condition that the second derivation of the filtered signal must be zero. Hence, the filtered signal is composed by a set of straight trends and breaks[2]. The competition between these two terms in the objective function turns to the competition between the number of straight trends (or number of breaks) and the closeness to the raw data. Therefore, the smoothing parameter λ plays an important role for detecting the number of breaks. In the later, we present briefly how the L_1 filter works for the trend detection and its extension to mean-reverting processes. The calibration procedure for λ parameter will be also discussed in detail.

[2]A break is the position where the trend of signal changes.

3 L_1 filtering schemes

3.1 Application to trend-stationary process

The Hodrick-Prescott scheme discussed in last section can be rewritten in the vectorial space $\mathbb{R}^n$ and its L_2 norm $\|\cdot\|_2$ as:

$$\frac{1}{2}\|y-x\|_2^2+\lambda\|Dx\|_2^2$$

where $y=(y_1,\ldots,y_n)$, $x=(x_1,\ldots,x_n)\in\mathbb{R}^n$ and the D operator is the $(n-2)\times n$ matrix:

$$D=\begin{bmatrix} 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & & \ddots & & & \\ & & & 1 & -2 & 1 & \\ & & & & 1 & 2 & 1 \end{bmatrix} \quad (1)$$

The exact solution of this estimation is given by

$$x^\star=\left(I+2\lambda D^\top D\right)^{-1}y$$

The explicit expression of $x^\star$ allows a very simple numerical implementation with sparse matrix. As L_2 filter is a linear filter, the regularization parameter λ is calibrated by comparing to the usual moving-average filter. The detail of the calibration procedure is given in Appendix A.4.

The idea of L_2 filter can be generalized to a lager class so-called L_p filter by using L_p penalty condition instead of L_2 penalty. This generalization is already discussed in the work of Daubechies *et al.* (2004) for the linear inverse problem or in the Lasso regression problem by Tibshirani *et al.* (1996). If we consider a L_1 filter, the objective function becomes:

$$\frac{1}{2}\sum_{t=1}^{n}(y_t-x_t)^2+\lambda\sum_{t=2}^{n-1}|x_{t-1}-2x_t+x_{t+1}|$$

which is equivalent to the following vectorial form:

$$\frac{1}{2}\|y-x\|_2^2+\lambda\|Dx\|_1$$

It has been demonstrated in Kim *et al.* (2009) that the dual problem of this L_1 filter scheme is a quadratic program with some boundary constraints. The detail of this derivation is shown in Appendix A.1.1. In order to optimize the numerical computation speed, we follow Kim *et al.* (2009) by using a "primal-dual interior point" method (see Appendix A.2). In the following, we check the efficient of this technique on various trend-stationary processes.

The first model consists of data simulated by a set of straight trend lines with a white noise perturbation:

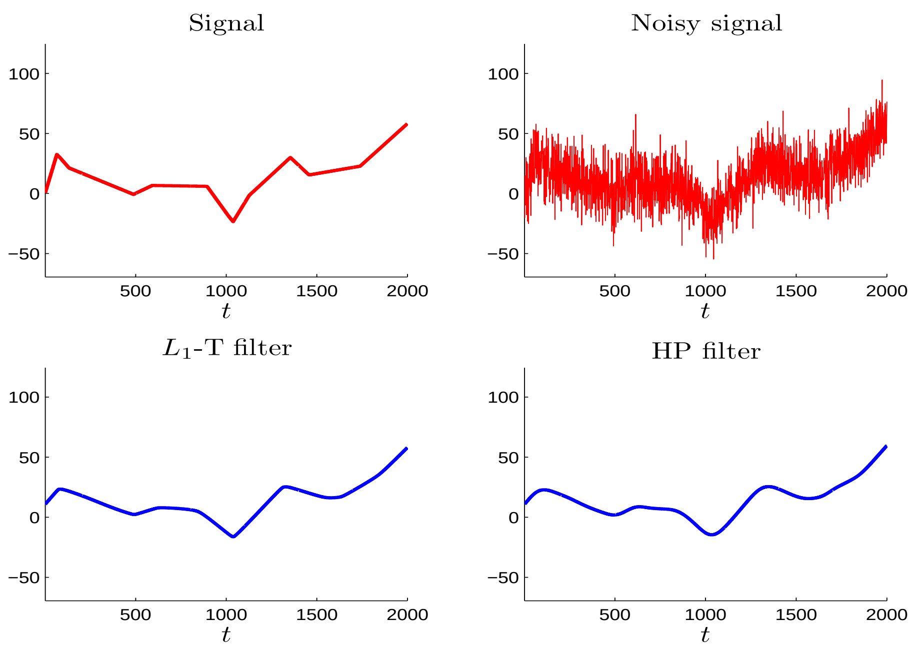
$$\begin{cases} y_t=x_t+\varepsilon_t \\ \varepsilon_t\sim\mathcal{N}\left(0,\sigma^2\right) \\ x_t=x_{t-1}+v_t \\ \Pr\left\{v_t=v_{t-1}\right\}=p \\ \Pr\left\{v_t=b\left(\mathcal{U}_{[0,1]}-\frac{1}{2}\right)\right\}=1-p \end{cases} \quad (2)$$

We present in Figure 1 the comparison between $L_1 - T$ and HP filtering schemes[3]. The top-left graph is the real trend x_t whereas the top-right graph presents the noisy signal y_t. The bottom graphs show the results of the $L_1 - T$ and HP filters. Here, we have chosen $\lambda = 5\,258$ for the $L_1 - T$ filtering and $\lambda = 1\,217\,464$ for HP filtering. This choice of λ for $L_1 - T$ filtering is based on the number of breaks in the trend, which is fixed to 10 in this example[4]. The second model model is a random walk generated by the following process:

$$\left\{ \begin{array}{l} y_t = y_{t-1} + v_t + \varepsilon_t \\ \varepsilon_t \sim \mathcal{N}\left(0, \sigma^2\right) \\ \Pr\left\{v_t = v_{t-1}\right\} = p \\ \Pr\left\{v_t = b\left(\mathcal{U}_{[0,1]} - \frac{1}{2}\right)\right\} = 1 - p \end{array} \right. \tag{3}$$

We present in Figure 2 the comparison between $L_1 - T$ filtering and HP filtering on this second model[5].

Figure 1: $L_1 - T$ filtering versus HP filtering for the model (2)

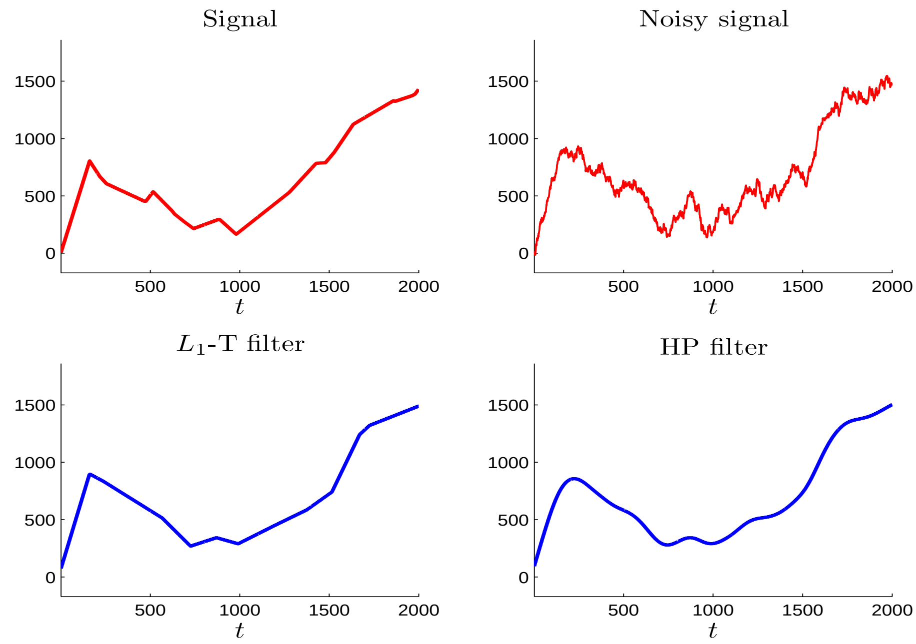


3.2 Extension to mean-reverting process

As shown in the last paragraph, the use of L_1 penalty on the second derivative gives the correct description of the signal tendency. Hence, similar idea can be applied for other order of the derivatives. We present here the extension of this L_1 filtering technique to the case of mean-reverting processes. If we impose now the L_1 penalty condition to the first derivative, we can expect to get the fitted signal with zero slope. The cost of this penalty will be proportional to the number of jumps. In this case, we would like to minimize the following

[3]We consider $n = 2000$ observations. The parameters of the simulation are $p = 0.99$, $b = 0.5$ and $\sigma = 15$.
[4]We discuss how to obtain λ in the next section.
[5]The parameters of the simulation are $p = 0.993$, $b = 5$ and $\sigma = 15$.

Figure 2: L_1-T filtering versus HP filtering for the model (3)



objective function:

$$\frac{1}{2}\sum_{t=1}^{n}\left(y_t - x_t\right)^2 + \lambda\sum_{t=2}^{n}\left|x_t - x_{t-1}\right|$$

or in the vectorial form:

$$\frac{1}{2}\left\|y - x\right\|_2^2 + \lambda\left\|Dx\right\|_1$$

Here the D operator is $(n-1)\times n$ matrix which is the discrete version of the first order derivative:

$$D = \begin{bmatrix} -1 & 1 & 0 & & & \\ 0 & -1 & 1 & 0 & & \\ & & & \ddots & & \\ & & & -1 & 1 & 0 \\ & & & & -1 & 1 \end{bmatrix} \tag{4}$$

We may apply the same minimization algorithm as previously (see Appendix A.1.2). To illustrate that, we consider the model with step trend lines perturbed by a white noise process:

$$\left\{\begin{array}{l} y_t = x_t + \varepsilon_t \\ \varepsilon_t \sim \mathcal{N}\left(0, \sigma^2\right) \\ \Pr\left\{x_t = x_{t-1}\right\} = p \\ \Pr\left\{x_t = b\left(\mathcal{U}_{[0,1]} - \frac{1}{2}\right)\right\} = 1 - p \end{array}\right. \tag{5}$$

We employ this model for testing the $L_1 - C$ filtering and HP filtering adapted to the first derivative[6], which corresponds to the following optimization program:

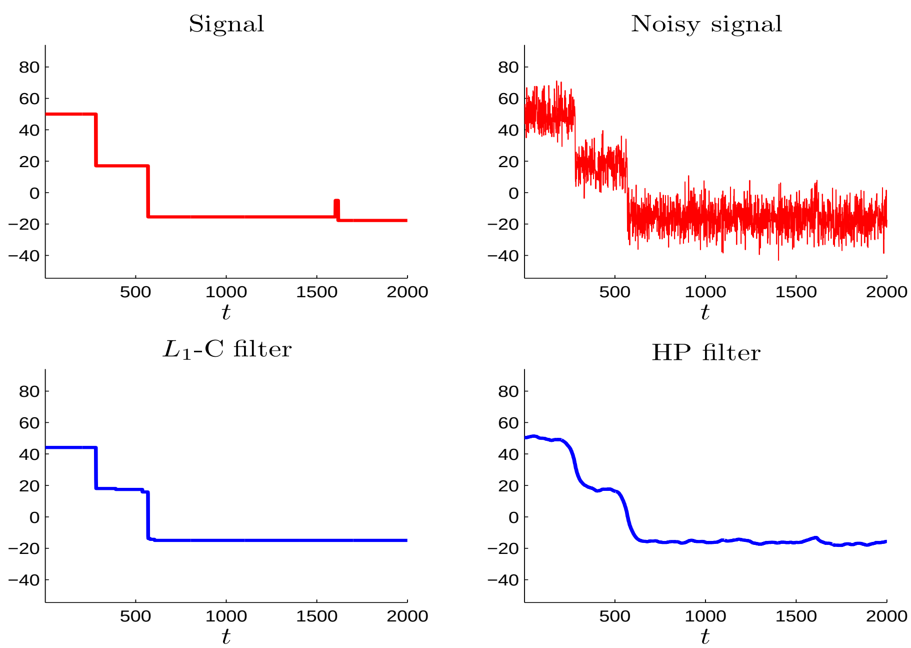
$$\min \frac{1}{2} \sum_{t=1}^{n} (y_t - x_t)^2 + \lambda \sum_{t=2}^{n} (x_t - x_{t-1})^2$$

In Figure 3, we have reported the corresponding results[7]. For the second test, we consider a mean-reverting process (Ornstein-Uhlenbeck process) with mean value following a regime switching process:

$$\begin{cases} y_t = y_{t-1} + \theta(x_t - y_{t-1}) + \varepsilon_t \\ \varepsilon_t \sim \mathcal{N}\left(0, \sigma^2\right) \\ \Pr\left\{x_t = x_{t-1}\right\} = p \\ \Pr\left\{x_t = b\left(\mathcal{U}_{[0,1]} - \frac{1}{2}\right)\right\} = 1 - p \end{cases} \tag{6}$$

Here, μ_t is the process which characterizes the mean value and θ is inversely proportional to the return time to the mean value. In Figure 4, we show how the $L_1 - C$ filter can capture the original signal in comparison to the HP filter[8].

Figure 3: $L_1 - C$ filtering versus HP filtering for the model (5)



3.3 Mixing trend and mean-reverting properties

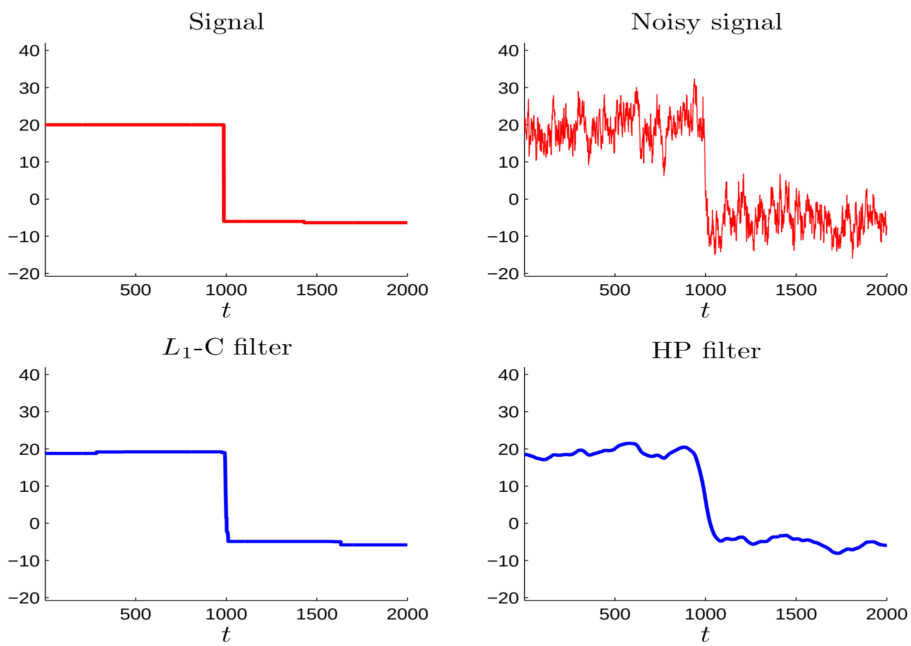
We now combine the two schemes proposed above. In this case, we define two regularization parameters λ_1 and λ_2 corresponding to two penalty conditions $\sum_{t=1}^{n-1} |x_t - x_{t-1}|$ and

[6] We use the term HP filter in order to keep homogeneous notations. However, we notice that this filter is indeed the FLS filter proposed by Kalaba and Tesfatsion (1989) when the exogenous regressors are only a constant.

[7] The parameters are $p = 0.998$, $b = 50$ and $\sigma = 8$.

[8] For the simulation of the Ornstein-Uhlenbeck process, we have chosen $p = 0.9985$, $b = 20$, $\theta = 0.1$ and $\sigma = 2$

Figure 4: $L_1 - C$ filtering versus HP filtering for the model (6)



$\sum_{t=2}^{n-1} |x_{t-1} - 2x_t + x_{t+1}|$. Our objective function for the primal problem becomes now:

$$\frac{1}{2}\sum_{t=1}^{n}(y_t - x_t)^2 + \lambda_1 \sum_{t=1}^{n-1} |x_t - x_{t-1}| + \lambda_2 \sum_{t=2}^{n-1} |x_{t-1} - 2x_t + x_{t+1}|$$

which can be again rewritten in the matrix form:

$$\frac{1}{2}\|y - x\|_2^2 + \lambda_1 \|D_1 x\|_1 + \lambda_2 \|D_2 x\|_1$$

where the D_1 and D_2 operators are respectively the $(n-1) \times n$ and $(n-2) \times n$ matrices defined in equations (4) and (1).

In Figures 5 and 6, we test the efficiency of the mixing scheme on the straight trend lines model (2) and the random walk model (3)[9].

3.4 How to calibrate the regularization parameters?

As shown above, the trend obtained from L_1 filtering depends on the parameter λ of the regularization procedure. For large values of λ, we obtain the long-term trend of the data while for small values of λ, we obtain short-term trends of the data. In this paragraph, we attempt to define a procedure which permits to do the right choice on the smoothing parameter according to our need of trend extraction.

[9] For both models, the parameters are $p = 0.99$, $b = 0.5$ and $\sigma = 5$.

Figure 5: $L_1 - TC$ filtering versus HP filtering for the model (2)

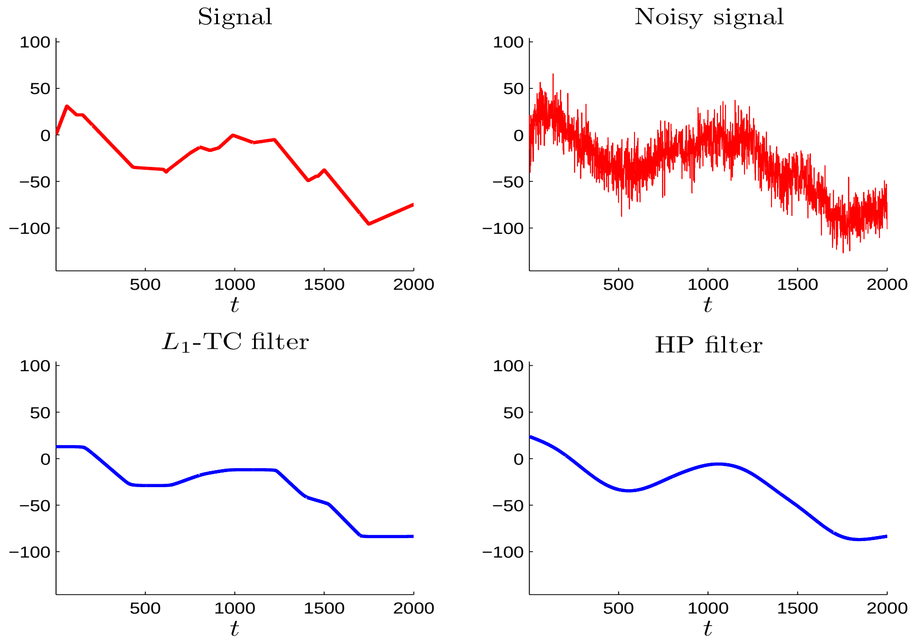
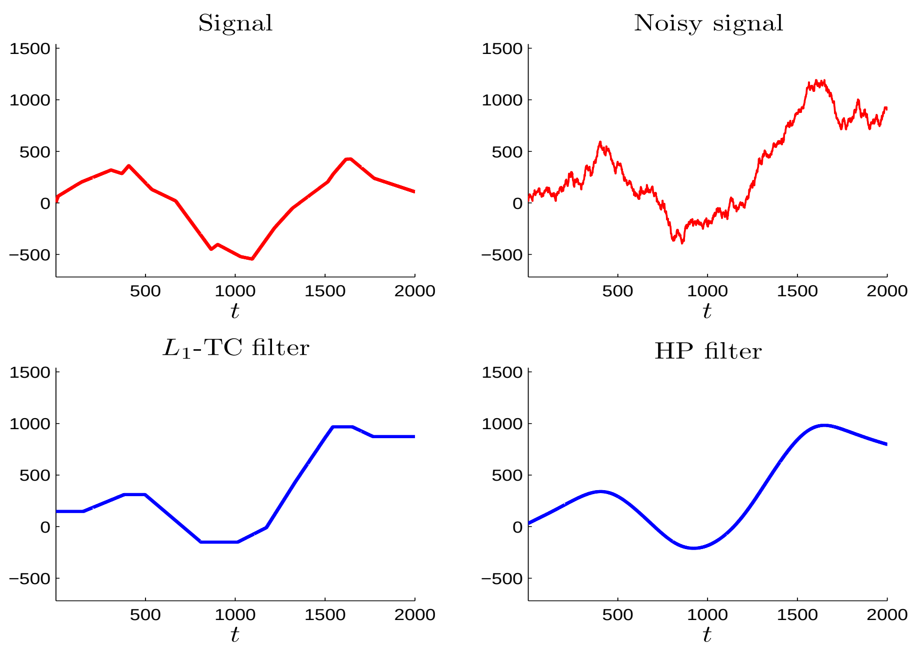


Figure 6: $L_1 - TC$ filtering versus HP filtering for the model (3)



3.4.1 A preliminary remark

For small value of λ, we recover the original form of the signal. For large value of λ, we remark that there exists a maximum value $\lambda_{\max}$ above which the trend signal has the affine form:

$$x_t = \alpha + \beta t$$

where α and β are two constants which do not depend on the time t. The value of $\lambda_{\max}$ is given by:

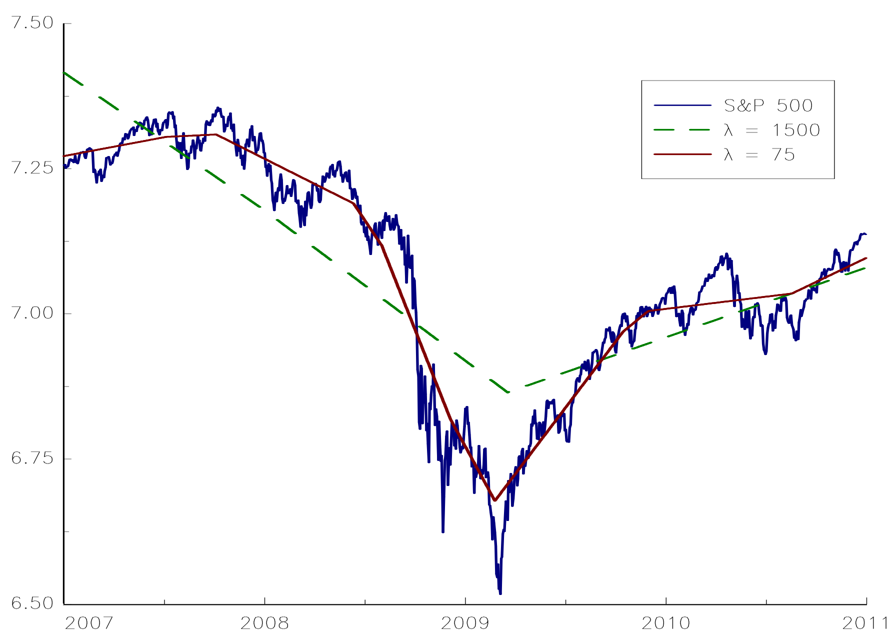
$$\lambda_{\max} = \left\| \left(DD^\top\right)^{-1} Dy \right\|_\infty$$

We can use this remark to get an idea about the order of magnitude of λ which should be used to determine the trend over a certain time period T. In order to show this idea, we take the data over the total period T. If we want to have the global trend on this period, we fix $\lambda = \lambda_{\max}$. This λ will gives the unique trend for the signal over the whole period. If one need to get more detail on the trend over shorter periods, we can divide the signal into p time intervals and then estimate λ via the mean value of all the $\lambda^i_{\max}$ parameter:

$$\lambda = \frac{1}{p} \sum_{i=1}^{p} \lambda^i_{\max}$$

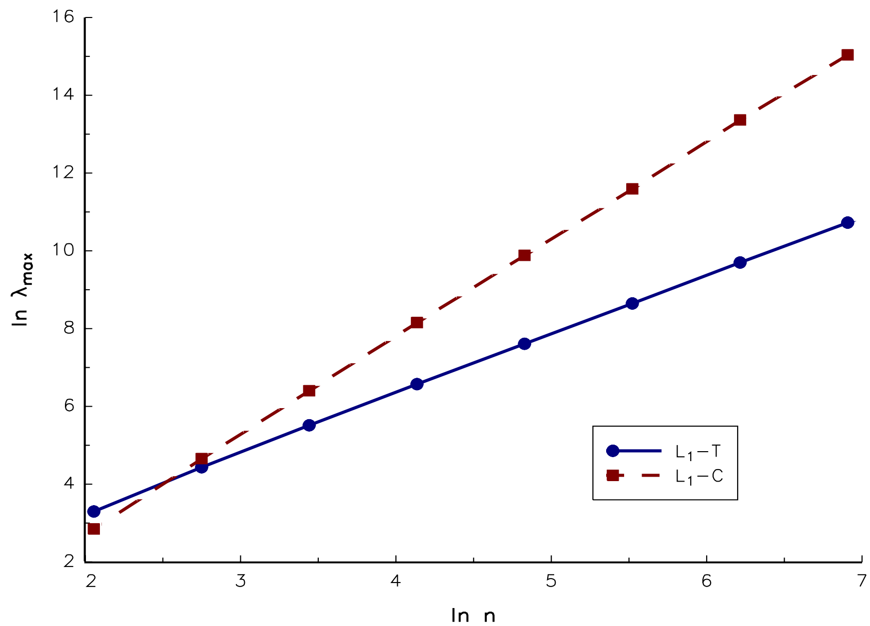
In Figure 7, we show the results obtained with $p = 2$ ($\lambda = 1\,500$) and $p = 6$ ($\lambda = 75$) on the S&P 500 index.

Figure 7: Influence of the smoothing parameter λ



Moreover, the explicit calculation of a Brownian motion process gives us the scaling law of the the smoothing parameter $\lambda_{\max}$. For the trend filtering scheme, $\lambda_{\max}$ scales as $T^{5/2}$ while for the mean-reverting scheme, $\lambda_{\max}$ scales as $T^{3/2}$ (see Figure 8). Numerical calculation of these powers for 500 simulations of the model (3) gives very good agreement with the analytical result for Brownian motion. Indeed, we obtain empirically that the power for $L_1 - T$ filter is 2.51 while the one for $L_1 - C$ filter is 1.52.

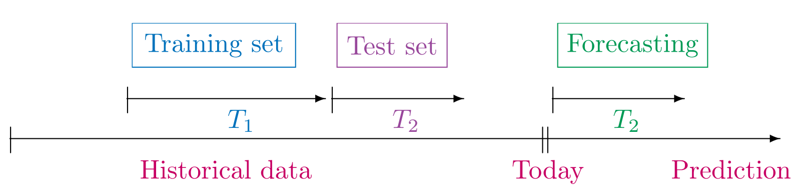
Figure 8: Scaling power law of the smoothing parameter $\lambda_{\max}$



3.4.2 Cross validation procedure

In this paragraph, we discuss how to employ a cross-validation scheme in order to calibrate the smoothing parameter λ of our model. We define two additional parameters which characterize the trend detection mechanism. The first parameter T_1 is the width of the data windows to estimate the optimal λ with respect to our target strategy. This parameter controls the precision of our calibration. The second parameter T_2 is used to estimate the prediction error of the trends obtained in the main window. This parameter characterizes the time horizon of the investment strategy. Figure 9 shows how the data set is divided into

Figure 9: Cross-validation procedure for determining optimal value $\lambda^\star$



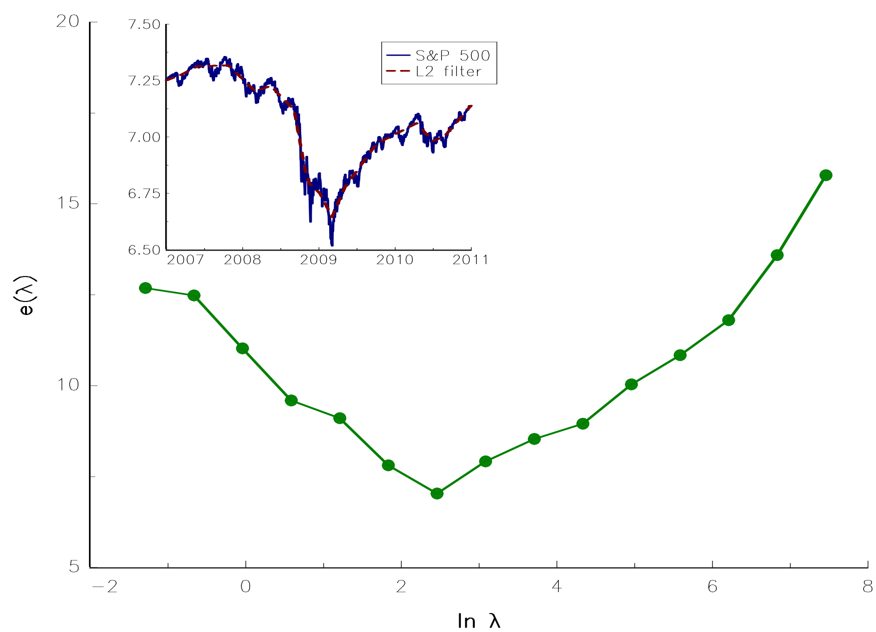
different windows in the cross validation procedure. In order to get the optimal parameter λ, we compute the total error after scanning the whole data by the window T_1. The algorithm of this calibration process is described as following:

Algorithm 1 Cross validation procedure for L_1 filtering

procedure CV_FILTER(T_1, T_2)
  Divide the historical data by m rolling test sets T_2^i $(i = 1, \ldots, m)$
  For each test window T_2^i, compute the statistic $\lambda_{\max}^i$
  From the array of $\left(\lambda_{\max}^i\right)$, compute the average $\bar{\lambda}$ and the standard deviation σ_λ
  Compute the boundaries $\lambda_1 = \bar{\lambda} - 2\sigma_\lambda$ and $\lambda_2 = \bar{\lambda} + 2\sigma_\lambda$
  for $j = 1 : n$ **do**
    Compute $\lambda_j = \lambda_1 \left(\lambda_2/\lambda_1\right)^{(j/n)}$
    Divide the historical data by p rolling training sets T_1^k $(k = 1, \ldots, p)$
    for $k = 1 : p$ **do**
      For each training window T_1^k, run the L_1 filter
      Forecast the trend for the adjacent test window T_2^k
      Compute the error $e^k\left(\lambda_j\right)$ on the test window T_2^k
    end for
    Compute the total error $e\left(\lambda_j\right) = \sum_{k=1}^{m} e^k\left(\lambda_j\right)$
  end for
  Minimize the total error $e\left(\lambda\right)$ to find the optimal value $\lambda^\star$
  Run the L_1 filter with $\lambda = \lambda^\star$
end procedure

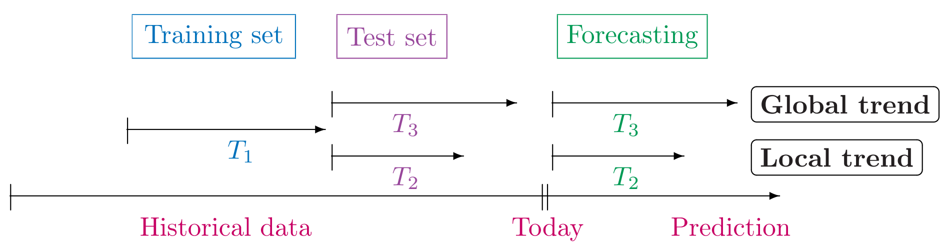
Figure 10 illustrates the calibration procedure for the S&P 500 index with $T_1 = 400$ and $T_2 = 50$ for the S&P 500 index (the number of observations is equal to 1 008 trading days). With $m = p = 12$ and $n = 15$, the estimated optimal value $\lambda^\star$ for the $L_1 - T$ filter is equal to 7.03.

Figure 10: Calibration procedure with the S&P 500 index



We have observed that this calibration procedure is more favorable for long-term time horizon, that is to estimate a global trend. For short-term time horizon, the prediction of local trends is much more perturbed by the noise. We have computed the probability of having good prediction on the tendency of the market for long-term and short-term time horizons. This probability is about 70% for 3 months time horizon while it is just 50% for one week time horizon. It comes that even if the fit is good for the past, the noise is however large meaning that the prediction of the future tendency is just 1/2 for an increasing market and 1/2 for a decreasing market. In order to obtain better results for smaller time horizons, we improve the last algorithm by proposing a two-trend model. The first trend is the local one which is determined by the first algorithm with the parameter T_2 corresponding to the local prediction. The second trend is the global one which gives the tendency of the market over a longer period T_3. The choice of this global trend parameter is very similar to the choice of the moving-average parameter. This model can be considered as a simple version of mean-reverting model for the trend. In Figure 11, we describe how the data set is divided for estimating the local trend and the global trend.

Figure 11: Cross validation procedure for two-trend model



The procedure for estimating the trend of the signal in the two-trend model is summarized in Algorithm 2. The corrected trend is now determined by studying the relative position of the historical data to the global trend. The reference position is characterized by the standard deviation $\sigma\left(y_t - x_t^G\right)$ where x_t^G is the filtered global trend.

Algorithm 2 Prediction procedure for the two-trend model

procedure PREDICT_FILTER(T_l, T_g)
  Compute the local trend x_t^L for the time horizon T_2 with the CV_FILTER procedure
  Compute the global trend x_t^G for the time horizon T_3 with the CV_FILTER procedure
  Compute the standard deviation $\sigma\left(y_t - x_t^G\right)$ of data with respect to the global trend
  if $\left|y_t - x_t^G\right| < \sigma\left(y_t - x_t^G\right)$ **then**
    Prediction $\leftarrow x_t^L$
  else
    Prediction $\leftarrow x_t^G$
  end if
end procedure

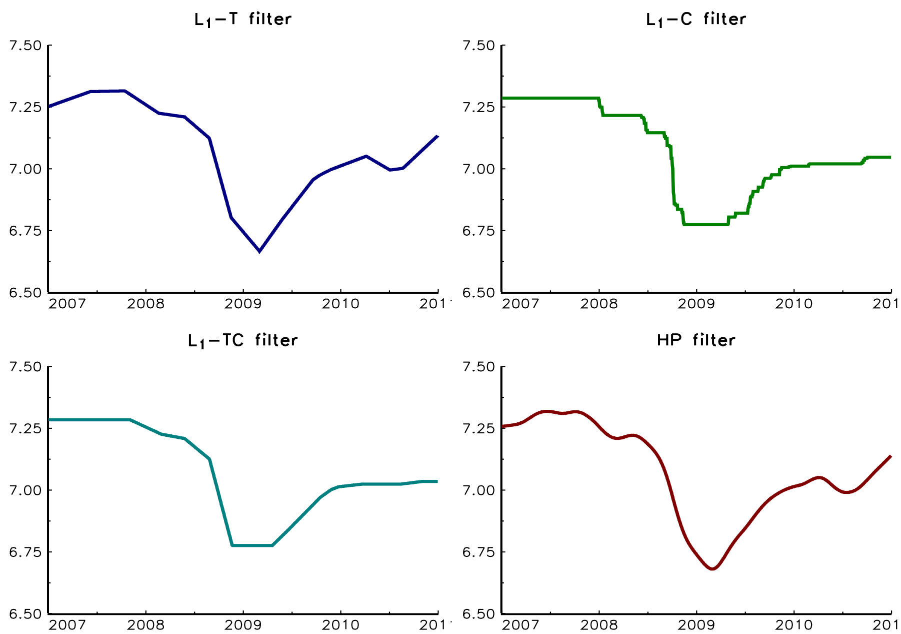
4 Application to momentum strategies

In this section, we apply the previous framework to the S&P 500 index. First, we illustrate the calibration procedure for a given trading date. Then, we backtest a momentum strategy by estimating dynamically the optimal filters.

4.1 Estimating the optimal filter for a given trading date

We would like to estimate the optimal filter for January 3rd, 2011 by considering the period from January 2007 to December 2010. We use the previous algorithms with $T_1 = 400$ and $T_2 = 50$. The optimal parameters are $\lambda_1 = 2.46$ (for the $L_1 - C$ filter) and $\lambda_2 = 15.94$ (for the $L_2 - T$ filter). Results are reported in Figure 12. The trend for the next 50 trading days is estimated to 7.34% for the $L_1 - T$ filter and 7.84% for the HP filter whereas it is null for the $L_1 - C$ and $L_1 - TC$ filters. By comparison, the true performance of the S&P 500 index is 1.90% from January 3rd, 2011 to March 15th, 2011[10].

Figure 12: Comparison between different L_1 filters on S&P 500 Index



4.2 Backtest of a momentum strategy

4.2.1 Design of the strategy

Let us consider a class of self-financed strategies on a risky asset S_t and a risk-free asset B_t. We assume that the dynamics of these assets is:

$$\begin{aligned} \mathrm{d}B_t &= r_t B_t \,\mathrm{d}t \\ \mathrm{d}S_t &= \mu_t S_t \,\mathrm{d}t + \sigma_t S_t \,\mathrm{d}W_t \end{aligned}$$

where r_t is the risk-free rate, μ_t is the trend of the asset price and σ_t is the volatility. We denote α_t the proportion of investment in the risky asset and $(1 - \alpha_t)$ the part invested in the risk-free asset. We start with an initial budget W_0 and expect a final wealth W_T. The optimal strategy is the one which optimizes the expectation of the utility function $U(W_T)$ which is increasing and concave. It is equivalent to the Markowitz problem which consists

[10] It corresponds exactly to a period of 50 trading days

of maximizing the wealth of the portfolio under a penalty of risk:

$$\sup_{\alpha \in \mathbb{R}} \left\{ \mathbb{E}\left(W_T^\alpha\right) - \frac{\lambda}{2} \sigma^2 \left(W_T^\alpha\right) \right\}$$

which is equivalent to:

$$\sup_{\alpha \in \mathbb{R}} \left\{ \alpha_t \mu_t - \frac{\lambda}{2} W_0 \alpha_t^2 \sigma_t^2 \right\}$$

As the objective function is concave, the maximum corresponds to the zero point of the gradient $\mu_t - \lambda W_0 \alpha_t \sigma_t^2$. We obtain the optimal solution:

$$\alpha_t^\star = \frac{1}{\lambda W_0} \frac{\mu_t}{\sigma_t^2}$$

In order to limit the explosion of α_t, we also impose the following constraint $\alpha_{\min} \leq \alpha_t \leq \alpha_{\max}$:

$$\alpha_t^\star = \max\left(\min\left(\frac{1}{\lambda W_0} \frac{\mu_t}{\sigma_t^2}, \alpha_{\min}\right), \alpha_{\max}\right)$$

The wealth of the portfolio is then given by the following expression:

$$W_{t+1} = W_t + W_t \left(\alpha_t^\star \left(\frac{S_{t+1}}{S_t} - 1\right) + (1 - \alpha_t^\star) r_t\right)$$

4.2.2 Results

In the following simulations, we use the estimators $\hat{\mu}_t$ and $\hat{\sigma}_t$ in place of μ_t and σ_t. For $\hat{\mu}_t$, we consider different models like L_1, HP and moving-average filters[11] whereas we use the following estimator for the volatility:

$$\hat{\sigma}_t^2 = \frac{1}{T} \int_0^T \sigma_t^2 \,\mathrm{d}t = \frac{1}{T} \sum_{i=t-T+1}^{t} \ln^2 \frac{S_i}{S_{i-1}}$$

We consider a long/short strategy, that is $(\alpha_{\min}, \alpha_{\max}) = (-1, 1)$. In the particular case of the $\hat{\mu}_t^{L_1}$ estimator, we consider three different models:

1. the first one is based on the local trend;
2. the second one is based on the global trend;
3. the combination of both local and global trends corresponds to the third model.

For all these strategies, the test set of the local trend T_2 is equal to 6 months (or 130 trading days) whereas the length of the test set for global trend is four times the length of the test set – $T_3 = 4T_2$ – meaning that T_3 is one year (or 520 trading days). This choice of T_3 agrees with the habitual choice of the width of the windows in moving average estimator. The length of the training set is also four times the length of the test set T_1. The study period is from January 1998 to December 2010. In the backtest, the trend estimation is updated every day. In Table 1, we summarize the results obtained with the different models cited above for the backtest. We remark that the best performances correspond to the case of global trend, HP and two-trend models. Because HP filter is calibrated to the window of the moving-average filter which is equal to T_3, it is not surprising that the performances of

[11] We note them respectively $\hat{\mu}_t^{L_1}$, $\hat{\mu}_t^{\mathrm{HP}}$ and $\hat{\mu}_t^{\mathrm{MA}}$.

Table 1: Results for the Backtest

Model	Trend	Performance	Volatility	Sharpe	IR	Drawdown
S&P 500		2.04%	21.83%	−0.06		56.78
$\hat{\mu}_t^{\mathrm{MA}}$		3.13%	18.27%	−0.01	0.03	33.83
$\hat{\mu}_t^{\mathrm{HP}}$		6.39%	18.28%	0.17	0.13	39.60
$\hat{\mu}_t^{L_1}$	(LT)	3.17%	17.55%	−0.01	0.03	25.11
$\hat{\mu}_t^{L_1}$	(GT)	6.95%	19.01%	0.19	0.14	31.02
$\hat{\mu}_t^{L_1}$	(LGT)	6.47%	18.18%	0.17	0.13	31.99

these three models are similar. On the considered period of the backtest, the S&P does not have a clear upward or downward trend. Hence, the local trend estimator does not give a good prediction and this strategy gives the worst performance. By contrast, the two-trend model takes into account the trade-off between local trend and global trend and gives a better result

5 Extension to the multivariate case

We now extend the L_1 filtering scheme to a multivariate time series $y_t = \left(y_t^{(1)}, \ldots, y_t^{(m)}\right)$. The underlying idea is to estimate the common trend of several univariate time series. In finance, the time series correspond to the prices of several assets. Therefore, we can build long/short strategies between these assets by comparing the individual trends and the common trend.

For the sake of simplicity, we assume that all the signals are rescaled to the same order of magnitude[12]. The objective function becomes new:

$$\frac{1}{2}\sum_{i=1}^{m}\left\|y^{(i)} - x\right\|_2^2 + \lambda \left\|Dx\right\|_1$$

In Appendix A.1.4, we show that this problem is equivalent to the L_1 univariate problem by considering $\bar{y}_t = m^{-1}\sum_{i=1}^{m} y^{(i)}$ as the signal.

6 Conclusion

Momentum strategies are efficient ways to use the market tendency for building trading strategies. Hence, a good estimator of the trend is essential from this perspective. In this paper, we show that we can use L_1 filters to forecast the trend of the market in a very simple way. We also propose a cross-validation procedure to calibrate the optimal regularization parameter λ where the only information to provide is the investment time horizon. More sophisticated models based on a local and global trends is also discussed. We remark that these models can reflect the effect of mean-reverting to the global trend of the market. Finally, we consider several backtests on the S&P 500 index and obtain competing results with respect to the traditional moving-average filter.

[12] For example, we may center and standardize the time series by subtracting the mean and dividing by the standard deviation.

A Computational aspects of L_1, L_2 filters

A.1 The dual problem

A.1.1 The $L_1 - T$ filter

This problem can be solved by considering the dual problem which is a QP program. We first rewrite the primal problem with new variable $z = Dx$:

$$\begin{array}{ll} \min & \frac{1}{2} \|y - x\|_2^2 + \lambda \|z\|_1 \\ \text{u.c.} & z = Dx \end{array}$$

We construct now the Lagrangian function with the dual variable $\nu \in \mathbb{R}^{n-2}$:

$$\mathcal{L}(x, z, \nu) = \frac{1}{2} \|y - x\|_2^2 + \lambda \|z\|_1 + \nu^\top (Dx - z)$$

The dual objective function is obtained in the following way:

$$\inf_{x,z} \mathcal{L}(x, z, \nu) = -\frac{1}{2} \nu^\top D D^\top \nu + y^\top D^\top \nu$$

for $-\lambda \mathbf{1} \leq \nu \leq \lambda \mathbf{1}$. According to the Kuhn-Tucker theorem, the initial problem is equivalent to the dual problem:

$$\begin{array}{ll} \min & \frac{1}{2} \nu^\top D D^\top \nu - y^\top D^\top \nu \\ \text{u.c.} & -\lambda \mathbf{1} \leq \nu \leq \lambda \mathbf{1} \end{array}$$

This QP program can be solved by traditional Newton algorithm or by interior-point methods, and the final solution of the trend reads

$$x^\star = y - D^\top \nu$$

A.1.2 The $L_1 - C$ filter

The optimization procedure for $L_1 - C$ filter follows the same strategy as the $L_1 - T$ filter. We obtain the same quadratic program with the D operator replaced by $(n-1) \times n$ matrix which is the discrete version of the first order derivative:

$$D = \begin{bmatrix} -1 & 1 & 0 & & & \\ 0 & -1 & 1 & 0 & & \\ & & & \ddots & & \\ & & & -1 & 1 & 0 \\ & & & & -1 & 1 \end{bmatrix}$$

A.1.3 The $L_1 - TC$ filter

In order to follow the same strategy presented above, we introduce two additional variables $z_1 = D_1 x$ and $z_2 = D_2 x$. The initial problem becomes:

$$\begin{array}{ll} \min & \frac{1}{2} \|y - x\|_2^2 + \lambda_1 \|z_1\|_1 + \lambda_2 \|z_2\|_1 \\ \text{u.c.} & \left\{ \begin{array}{l} z_1 = D_1 x \\ z_2 = D_2 x \end{array} \right. \end{array}$$

The Lagrangian function with the dual variables $\nu_1 \in \mathbb{R}^{n-1}$ and $\nu_2 \in \mathbb{R}^{n-2}$ is:

$$\mathcal{L}(x, z_1, z_2, \nu_1, \nu_2) = \frac{1}{2}\|y - x\|_2^2 + \lambda_1 \|z_1\|_1 + \lambda_2 \|z_2\|_1 + \nu_1^\top (D_1 x - z_1) + \nu_2^\top (D_2 x - z_2)$$

whereas the dual objective function is:

$$\inf_{x,z_1,z_2} \mathcal{L}(x, z_1, z_2, \nu_1, \nu_2) = -\frac{1}{2}\left\|D_1^\top \nu_1 + D_2^\top \nu_2\right\|_2^2 + y^\top \left(D_1^\top \nu_1 + D_2^\top \nu_2\right)$$

for $-\lambda_i \mathbf{1} \leq \nu_i \leq \lambda_i \mathbf{1}$ $(i = 1, 2)$. Introducing the variable $z = (z_1, z_2)$ and $\nu = (\nu_1, \nu_2)$, the initial problem is equivalent to the dual problem:

$$\begin{array}{ll} \min & \frac{1}{2}\nu^\top Q \nu - R^\top \nu \\ \text{u.c.} & -\nu^+ \leq \nu \leq \nu^+ \end{array}$$

with $D = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix}$, $Q = DD^\top$, $R = Dy$ and $\nu^+ = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \mathbf{1}$. The solution of the primal problem is then given by $x^\star = y - D^\top \nu$.

A.1.4 The $L_1 - T$ multivariate filter

As in the univariate case, this problem can be solved by considering the dual problem which is a QP program. The primal problem is:

$$\begin{array}{ll} \min & \frac{1}{2}\sum_{i=1}^{m}\left\|y^{(i)} - x\right\|_2^2 + \lambda \|z\|_1 \\ \text{u.c.} & z = Dx \end{array}$$

Let us define $\bar{y} = (\bar{y}_t)$ with $\bar{y}_t = m^{-1}\sum_{i=1}^{m} y^{(i)}$. The dual objective function becomes:

$$\inf_{x,z} \mathcal{L}(x, z, \nu) = -\frac{1}{2}\nu^\top DD^\top \nu + \bar{y}^\top D^\top \nu + \frac{1}{2}\sum_{i=1}^{m}\left(y^{(i)} - \bar{y}\right)^\top \left(y^{(i)} - \bar{y}\right)$$

for $-\lambda \mathbf{1} \leq \nu \leq \lambda \mathbf{1}$. According to the Kuhn-Tucker theorem, the initial problem is equivalent to the dual problem:

$$\begin{array}{ll} \min & \frac{1}{2}\nu^\top DD^\top \nu - \bar{y}^\top D^\top \nu \\ \text{u.c.} & -\lambda \mathbf{1} \leq \nu \leq \lambda \mathbf{1} \end{array}$$

This QP program can be solved by traditional Newton algorithm or by interior-point methods and the solution is:

$$x^\star = \bar{y} - D^\top \nu$$

A.2 The interior-point algorithm

We present briefly the interior-point algorithm of Boyd and Vandenberghe (2009) in the case of the following optimization problem:

$$\begin{array}{ll} \min & f_0(x) \\ \text{u.c.} & \left\{\begin{array}{l} Ax = b \\ f_i(x) < 0 \quad \text{for } i = 1, \ldots, m \end{array}\right. \end{array}$$

where $f_0, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$ are convex and twice continuously differentiable and $\operatorname{rank}(A) = p < n$. The inequality constraints will become implicit if one rewrite the problem as:

$$\begin{array}{ll} \min & f_0(x) + \sum_{i=1}^{m} \mathcal{I}_-(f_i(x)) \\ \text{u.c.} & Ax = b \end{array}$$

where $\mathcal{I}_-(u) : \mathbb{R} \to \mathbb{R}$ is the non-positive indicator function[13]. This indicator function is discontinuous, hence the Newton method can not be applied. In order to overcome this problem, we approximate $\mathcal{I}_-(u)$ by the logarithmic barrier function $\mathcal{I}_-^\star(u) = -\tau^{-1}\ln(-u)$ with $\tau \to \infty$. Finally the Kuhn-Tucker condition for this approximation problem gives $r_t(x, \lambda, \nu) = 0$ with:

$$r_\tau(x, \lambda, \nu) = \begin{pmatrix} \nabla f_0(x) + \nabla f(x)^\top \lambda + A^\top \nu \\ -\operatorname{diag}(\lambda) f(x) - \tau^{-1}\mathbf{1} \\ Ax - b \end{pmatrix}$$

The solution of $r_\tau(x, \lambda, \nu) = 0$ can be obtained by Newton's iteration for the triple $y = (x, \lambda, \nu)$:

$$r_\tau(y + \Delta y) \simeq r_\tau(y) + \nabla r_\tau(y)\,\Delta y = 0$$

This equation gives the Newton's step $\Delta y = -\nabla r_\tau(y)^{-1} r_\tau(y)$ which defines the search direction.

A.3 The scaling of smoothing parameter of L_1 filter

We can try to estimate the order of magnitude of the parameter $\lambda_{\max}$ by considering the continuous case. Assuming that the signal is a process W_t. The value of $\lambda_{\max}$ in the discrete case defined by:

$$\lambda_{\max} = \left\| \left(DD^\top\right)^{-1} Dy \right\|_\infty$$

can be considered as the first primitive $I_1(T) = \int_0^T W_t\,\mathrm{d}t$ of the process W_t if $D = D_1$ ($L_1 - C$ filtering) or the second primitive $I_2(T) = \int_0^T \int_0^t W_s\,\mathrm{d}s\,\mathrm{d}t$ of W_t if $D = D_2$ ($L_1 - T$ filtering). We have:

$$\begin{aligned} I_1(T) &= \int_0^T W_t\,\mathrm{d}t \\ &= W_T T - \int_0^T t\,\mathrm{d}W_t \\ &= \int_0^T (T - t)\,\mathrm{d}W_t \end{aligned}$$

The process $I_1(T)$ is a Wiener integral (or a Gaussian process) with variance:

$$\mathbb{E}\left[I_1^2(T)\right] = \int_0^T (T - t)^2\,\mathrm{d}t = \frac{T^3}{3}$$

[13] We have:

$$\mathcal{I}_-(u) = \begin{cases} 0 & u \le 0 \\ \infty & u > 0 \end{cases}$$

In this case, we expect that $\lambda_{\max} \sim T^{3/2}$. The second order primitive can be calculated in the following way:

$$\begin{aligned}
I_2(T) &= \int_0^T I_1(t)\,\mathrm{d}t \\
&= I_1(T)\,T - \int_0^T t\,\mathrm{d}I_1(T) \\
&= I_1(T)\,T - \int_0^T tW_t\,\mathrm{d}t \\
&= I_1(T)\,T - \frac{T^2}{2}W_T + \int_0^T \frac{t^2}{2}\,\mathrm{d}W_t \\
&= -\frac{T^2}{2}W_T + \int_0^T \left(T^2 - Tt + \frac{t^2}{2}\right)\mathrm{d}W_t \\
&= \frac{1}{2}\int_0^T (T-t)^2\,\mathrm{d}W_T
\end{aligned}$$

This quantity is again a Gaussian process with variance:

$$\mathbb{E}[I_2^2(T)] = \frac{1}{4}\int_0^T (T-t)^4\,\mathrm{d}t = \frac{T^5}{20}$$

In this case, we expect that $\lambda_{\max} \sim T^{5/2}$.

A.4 Calibration of the L_2 filter

We discuss here how to calibrate the L_2 filter in order to extract the trend with respect to the investment time horizon T. Though the L_2 filter admits an explicit solution which is a great advantage for numerical implementation, the calibration of the smoothing parameter λ is not trivial. We propose to calibrate the L_2 filter by comparing the spectral density of this filter with the one obtained with the moving-average filter. For this last filter, we have:

$$\hat{x}_t^{\mathrm{MA}} = \frac{1}{T}\sum_{i=t-T}^{t-1} y_i$$

It comes that the spectral density is:

$$f(\omega) = \frac{1}{T^2}\left|\sum_{t=0}^{T-1} e^{-i\omega t}\right|^2$$

For the L_2 filter, we k now that the solution is $\hat{x}^{\mathrm{HP}} = \left(1 + 2\lambda D^T D\right)^{-1} y$. Therefore, the spectral density is:

$$\begin{aligned}
f^{\mathrm{HP}}(\omega) &= \left(\frac{1}{1 + 4\lambda\left(3 - 4\cos\omega + \cos 2\omega\right)}\right)^2 \\
&\simeq \left(\frac{1}{1 + 2\lambda\omega^4}\right)^2
\end{aligned}$$

The width of the spectral density for the L_2 filter is then $(2\lambda)^{-1/4}$ whereas it is $2\pi T^{-1}$ for the moving-average filter. Calibrate the L_2 filter could be done by matching this two quantities. Finally, we obtain the following relationship:

$$\lambda \propto \lambda_\star = \frac{1}{2}\left(\frac{T}{2\pi}\right)^4$$

In Figure 13, we represent the spectral density of the moving-average filter for different windows T. We report also the spectral density of the corresponding L_2 filters. For that, we have calibrated the optimal parameter $\lambda^\star$ by least square minimization. In Figure 14, we compare the optimal estimator $\lambda^\star$ with the one corresponding to $10.27 \times \lambda_\star$. We notice that the approximation is very good.

A.5 Implementation issues

The computational time may be large when working with dense matrices even if we consider interior-point algorithms. It could be reduced by using sparse matrices. But the efficient way to optimize the implementation is to consider band matrices. Moreover, we may also notice that we have to solve a large linear system at each iteration. Depending on the filtering problem ($L_1 - T$, $L_1 - C$ and $L_1 - TC$ filters), the system is 6-bands or 3-bands but always symmetric. For computing $\lambda_{\max}$, one may remark that it is equivalent to solve a band system which is positive definite. We suggest to adapt the algorithms in order to take into account all these properties.

Figure 13: Spectral density of moving-average and L_2 filters

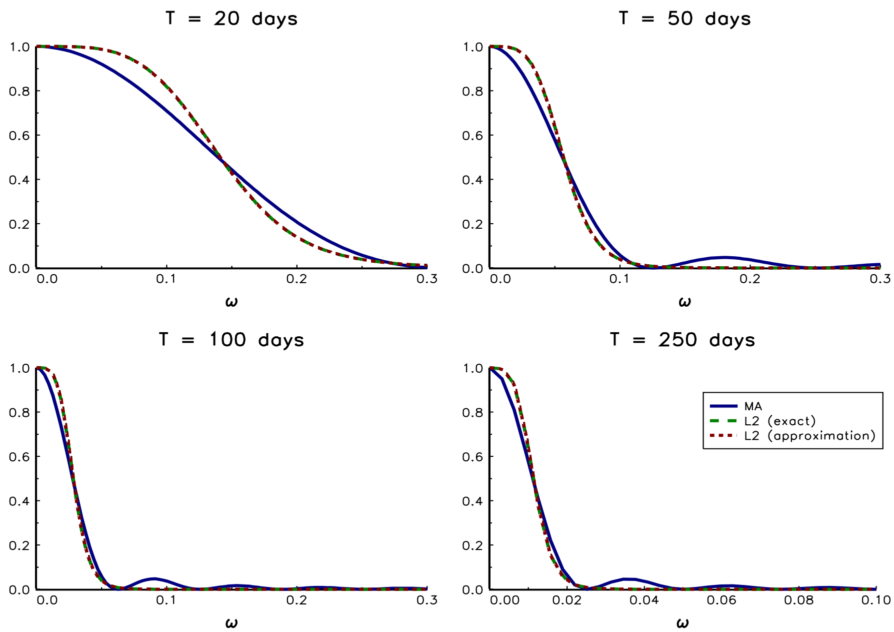
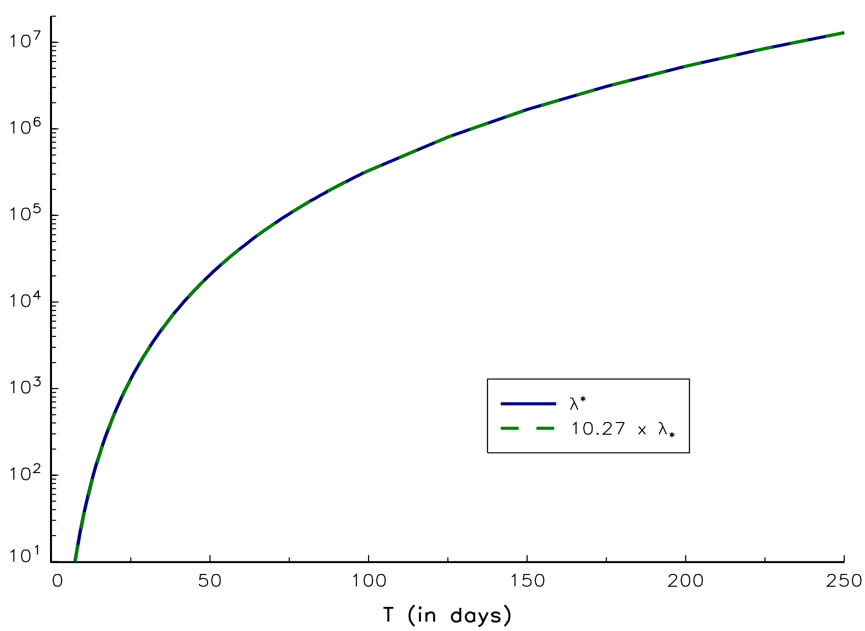


Figure 14: Relationship between the value of λ and the length of the moving-average filter



References

[1] Alexandrov T., Bianconcini S., Dagum E.B., Maass P. and McElroy T. (2008), A Review of Some Modern Approaches to the Problem of Trend Extraction , *US Census Bureau*, RRS #2008/03.

[2] Beveridge S. and Nelson C.R. (1981), A New Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle, *Journal of Monetary Economics*, 7(2), pp. 151-174.

[3] Boyd S. and Vandenberghe L. (2009), *Convex Optimization*, Cambridge University Press.

[4] Cleveland W.P. and Tiao G.C. (1976), Decomposition of Seasonal Time Series: A Model for the Census X-11 Program, *Journal of the American Statistical Association*, 71(355), pp. 581-587.

[5] Daubechies I., Defrise M. and De Mol C. (2004), An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint, *Communications on Pure and Applied Mathematics*, 57(11), pp. 1413-1457.

[6] Efron B., Tibshirani R. and Friedman R. (2009), *The Elements of Statistical Learning*, Second Edition, Springer.

[7] Harvey A. (1991), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

[8] Hodrick R.J. and Prescott E.C. (1997), Postwar U.S. Business Cycles: An Empirical Investigation, *Journal of Money, Credit and Banking*, 29(1), pp. 1-16.

[9] Kalaba R. and Tesfatsion L. (1989), Time-varying Linear Regression via Flexible Least Squares, *Computers & Mathematics with Applications*, 17, pp. 1215-1245.

[10] Kim S-J., Koh K., Boyd S. and Gorinevsky D. (2009), ℓ_1 Trend Filtering, *SIAM Review*, 51(2), pp. 339-360.

[11] Tibshirani R. (1996), Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society B*, 58(1), pp. 267-288.