Stat 519 - Final Project

Properties of Hodrick-Prescott and ℓ_1 Trend Filters and Tuning Parameters Selection

Danni Shi

Winter 2022

Abstract

Estimating the underlying trend component in time series data has been an intriguing topic in time series data analysis. Hodrick-Prescott (HP) filter and its variation, the ℓ_1 trend filter are two widely used methods for trend estimation. This project reviews some basic properties these two methods share (such as convergence to the original data as the tuning parameter approaches to 0, and convergence to the best affine fit as the tuning parameter approaches to infinity) and discusses their differences, especially the piecewise linearity of ℓ_1 trend filter and why this occurs. This project also reviews of the methods of selecting the tuning parameter of both the HP filter and the ℓ_1 trend filter which was mentioned in Yamada and Yoon (2016). Finally, this project uses the log of adjusted US real GDP data used in Yamada and Yoon (2016) to verify the properties discussed in this project, and reproduces the results in Section 2.4 of Yamada and Yoon (2016).

1 Introduction

1.1 Trend Estimation and Filtering

In practice, a time series may often exhibit trends variation and it may be inappropriate to treat the sampling data as some stationary process. Given a scalar of (discrete) time series $t = 1, 2, \cdots, n$, consider a time series model

$$X_t = m_t + Y_t \tag{1}$$

where $\{m_t\}$ is a slowly varying function known as a trend component [1], and $\{Y_t\}$ is a weakly stationary process with mean 0. If the trend m_t is deterministic, we would expect $E(X_t) = E(m_t) + E(Y_t) = m_t$. Our goal is to estimate the trend component m_t, and such estimating is referred as *filtering*.

Estimating the underlying trend component in time series data has been an intriguing topic and filtering has been applied in multiple disciplines, including macroeconomics [2], meteorology [3], social sciences [4] and biomedical sciences [5]. Many filtering methods have been introduced during the lecture, such as exponential smoothing [6], wavelet transform analysis [7], moving average filtering [8], Hodrick-Prescott (HP) filtering [2] and its variation ℓ_1 trend filtering [9]. This project focuses on the last two filters.

1.2 Hodrick-Prescott (HP) Filtering and ℓ_1 Trend Filtering

Hodrick-Prescott (HP) filter was introduced during the study of postwar business cycles in 1997. In HP filtering, for a given tuning parameter $\lambda_{\mathrm{HP}} \geqslant 0$, the trend estimate $\hat{\mathbf{m}}^{\mathrm{HP}} = [\hat{m}_1, \cdots, \hat{m}_n]^T$ is chosen amongst all possible sequences, to minimize the two-part objective function [2]:

$$\underbrace{\sum_{t=1}^{n} (x_t - \hat{m}_t)^2}_{(\#)} + \lambda_{\mathrm{HP}} \underbrace{\sum_{t=2}^{n-1} (\hat{m}_{t+1} - 2\hat{m}_t + \hat{m}_{t-1})^2}_{(\#\#)} \tag{2}$$

Here (#) indicates the fidelity or faithfulness of the trend estimate to the original data [10] and (##) indicates the smoothness of the model. Fidelity and smoothness cannot be obtained at the same time, and the tuning parameter λ_{HP} helps to balance between these two components.

The expression (2) can be rewritten as

$$(2) = \|\mathbf{x} - \hat{\mathbf{m}}\|_2^2 + \lambda_{\mathrm{HP}} \|\mathbf{D}\hat{\mathbf{m}}\|_2^2 \tag{3}$$

where $\mathbf{x} = [x_1, \cdots, x_n]^T$ and $\mathbf{D}$ is a $(n-2) \times n$ matrix, namely

$$\mathbf{D} = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & 1 & -2 & 1 \end{bmatrix} \tag{4}$$

In 2009, Kim et al proposed a variation of HP filter: the ℓ_1 trend filter. They applied the ℓ_1 regularization that is used in the well-known Lasso algorithm [11] and chose the trend estimate as the minimizer of the weighted sum objective function

$$\sum_{t=1}^{n} (x_t - \hat{m}_t)^2 + \lambda_{\ell_1} \sum_{t=2}^{n-1} |\hat{m}_{t+1} - 2\hat{m}_t + \hat{m}_{t-1}| = \|\mathbf{x} - \hat{\mathbf{m}}\|_2^2 + \lambda_{\ell_1} \|\mathbf{D}\hat{\mathbf{m}}\|_1 \tag{5}$$

where $\mathbf{D}$ is defined as in (4) and $\|\mathbf{u}\|_1 = \sum_i |u_i|$.

2 Properties of HP Filter and ℓ_1 Trend Filter

2.1 Similarities

2.1.1 Convergence to the Original Data as $\lambda \to 0$

By intuition, as λ approaches to 0, the smoothness component in (2) and (5) approach to 0 as well, and to minimize (2) and (5), we expect $\hat{\mathbf{m}} \to \mathbf{x}$. In fact [9], for HP filter,

$$\frac{\|\mathbf{x} - \hat{\mathbf{m}}^{\mathrm{HP}}\|_2}{\|\mathbf{x}\|_2} \leqslant \frac{32\lambda_{\mathrm{HP}}}{1 + 32\lambda_{\mathrm{HP}}} \;\Rightarrow\; \lim_{\lambda \to 0} \hat{\mathbf{m}}^{\mathrm{HP}} = \mathbf{x} \tag{6}$$

and for the ℓ_1 trend filter,

$$\|\mathbf{x} - \hat{\mathbf{m}}^{\ell_1}\|_\infty \leqslant 4\lambda_{\ell_1} \;\Rightarrow\; \lim_{\lambda \to 0} \hat{\mathbf{m}}^{\ell_1} = \mathbf{x} \tag{7}$$

where $\|\mathbf{u}\|_\infty = \max_i |u_i|$.

2.1.2 Convergence to the Best Affine Fit as $\lambda \to \infty$

By intuition, as $\lambda \to \infty$, $\hat{\mathbf{m}}$ is expected to be smoother and smoother if we hope to minimize (3) and (5). Eventually, $\hat{\mathbf{m}}$ is expected to become linear, namely, the best affine (i.e., the least squares regression line) fit. In HP filter, as λ approaches to infinity, we expect to have

$$\lim_{\lambda \to \infty} \hat{\mathbf{m}} = \mathbf{\Pi} \left(\mathbf{\Pi}^T \mathbf{\Pi} \right)^{-1} \mathbf{\Pi}^T \mathbf{x} = \mathbf{m}^{\text{AF}} \tag{8}$$

where $\mathbf{\Pi}$ is a $n \times 2$ matrix with the t-th row being $[1, t]$, and here we define $\mathbf{m}^{\text{AF}}$ being the best affine fit.

In ℓ_1 trend filter however, such convergence occurs for a finite value of λ [9],

$$\lambda^{\ell_1}_{\text{max}} = \|(\mathbf{D}\mathbf{D}^T)^{-1}\mathbf{D}\mathbf{x}\|_\infty \tag{9}$$

That is to say, for any $\lambda_{\ell_1} \geqslant \lambda^{\ell_1}_{\text{max}}$, we have $\hat{\mathbf{m}}^{\ell_1}$ being equivalent to the best affine fit, whereas in HP filter, $\hat{\mathbf{m}}^{\text{HP}}$ approaches to the best affine fit only in the limit as λ approaches to infinity.

2.2 Differences

2.2.1 Linearity of HP Filter

In HP filter, after taking derivative of (3), we can have the HP trend estimate being

$$\frac{\partial(3)}{\partial \hat{\mathbf{m}}} = 0 \Rightarrow \hat{\mathbf{m}}^{\text{HP}} = \underset{\hat{\mathbf{m}} \in \mathbb{R}}{\operatorname{argmin}}(3) = \left(\mathbf{I} + \lambda_{\text{HP}} \mathbf{D}^T \mathbf{D} \right)^{-1} \mathbf{x} \tag{10}$$

which is a linear function of the time series data $\mathbf{x}$ with $\mathbf{I}$ being the $n \times n$ identity matrix.

2.2.2 Piecewise Linearity of ℓ_1 Trend Filter in Time t and Kink Points

Unlike in (10), transforming (5) will not give us $\hat{\mathbf{m}}^{\ell_1}$ in the form of some linear function of $\mathbf{x}$. In practice, the ℓ_1 trend filter might be a preferred option over HP filter due to the fact that the trend estimator $\hat{\mathbf{m}}^{\ell_1}$ is piecewise linear in time t [9].

There are discrete times $1 = t_1 < t_2 < \cdots < t_{p-1} < t_p = n (t_k \in \mathbb{Z}, \forall k)$ for which

$$\hat{m}^{\ell_1}_t = \alpha_k + \beta_k t, \; t_k \leqslant t \leqslant t_{k+1}, k = 1, \cdots, p-1 \tag{11}$$

Namely, over each segmented time interval $[t_k, t_{k+1}]$, $\hat{\mathbf{m}}^{\ell_1}$ is an (linear) affine fit function of time t. Note that to keep those local trends consistent with each other, we must have consistent values for $\hat{\mathbf{m}}^{\ell_1}$ at the *kink points* $t_2, t_3, \cdots, t_{p-1}$, and to guarantee that $\hat{\mathbf{m}}^{\ell_1}$ is piecewise linear with $p-2$ kink points, we must have

$$\alpha_k + \beta_k t_{k+1} = \alpha_{k+1} + \beta_{k+1} t_{k+1}, \; k = 1, 2, \cdots, p-1 \text{ and generally } \alpha_k \neq \alpha_{k+1} \tag{12}$$

One extreme case is when there is no kink points (i.e., $p = 2, t_1 = 1, t_2 = n$), $\hat{\mathbf{m}}^{\ell_1}$ is equivalent to $\mathbf{m}^{\text{AF}}$ defined in (8), which corresponds to the case when $\lambda_{\ell_1} \geqslant \lambda^{\ell_1}_{\text{max}}$ as in (9). The other extreme case is when there is one kink point at every time point (i.e., $p = n$). The piecewise linear fit is then vacuous, which corresponds to the case when $\lambda_{\ell_1} = 0$.

2.2.3 ℓ_1 Trend Filtering and Sparse Approximation

To understand why ℓ_1 trend filter gives a piecewise linear trend estimation whereas HP filter does not, we could consider the Lasso representation of the ℓ_1 trend filter. Consider the Lasso representation of the ℓ_1 trend filter: let $\theta_1 = m_1^{\ell_1}, \theta_2 = m_2^{\ell_1} - m_1^{\ell_1}$ and $\theta_t = \nabla^2 m_t^{\ell_1}$ the second order difference for $t = 3, \cdots, n$. Then $\hat{\mathbf{m}}^{\ell_1} = \mathbf{A}\boldsymbol{\theta}$ where $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_n]^T$ and $\mathbf{A}$ is the following $n \times n$ lower triangular matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \ddots & \vdots & \vdots \\ 1 & 2 & 1 & \ddots & \vdots & \vdots \\ 1 & 3 & 2 & \ddots & 0 & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 & 0 \\ 1 & n-1 & n-2 & \cdots & 2 & 1 \end{bmatrix} \tag{13}$$

Here minimizing (5) is equivalent to finding

$$\operatorname*{argmin}_{\boldsymbol{\theta}} \|\mathbf{x} - \mathbf{A}\boldsymbol{\theta}\|_2^2 + \lambda_{\ell_1} \|\boldsymbol{\theta}_{3:n}\|_1 \tag{14}$$

where $\boldsymbol{\theta}_{3:n} = [\theta_3, \theta_4, \cdots, \theta_n]^T$.

From Efron et al (2004) [12] and Rosset and Zhu (2007) [13], from a standard result in ℓ_1-regularized least squares, we expect $\boldsymbol{\theta}_{3:n}$ to be sparse, and thus $\hat{\boldsymbol{\theta}}$, the solution to (14) would have many entries being 0. Such sparsity leads to the piecewise linearity of our trend estimate [14] and $\hat{\boldsymbol{\theta}}$ is a piecewise linear function of the regularization parameter λ_{ℓ_1}. Since $\hat{\mathbf{m}}^{\ell_1}$ is a linear transformation of $\hat{\boldsymbol{\theta}}$, it is also piecewise linear over time t.

3 Selecting the Tuning Parameter λ

3.1 λ Selection for HP Filter

The tuning parameters λ in both HP and ℓ_1 trend filters help to balance the fidelity of our trend estimator to the time series data and the smoothness of our model. One method of selecting λ_{HP} was proposed in Gomez (2001) [15], following works by King and Rebelo (1993) [16]. The filter gain of HP filter may be expressed as

$$G(\omega) = \frac{1}{1 + 4\lambda[1 - \cos(\omega)]^2} \tag{15}$$

Define ω_c such that $G(\omega_c) = \frac{1}{2}$ and let λ_p denote the corresponding tuning parameter. Let $p = \frac{2\pi}{\omega_c}$, so λ_p can be expressed as

$$\lambda_p = \frac{1}{4\left[1 - \cos\left(\frac{2\pi}{p}\right)\right]^2} \tag{16}$$

Here $\mathbf{x} - \hat{\mathbf{m}}_p^{\text{HP}}$ approximates the ideal high-pass filter that passes components with periods less than p.

3.2 λ Selection for the ℓ_1 Trend Filter

One method of selecting λ_{ℓ_1} was proposed by Yamada and Yoon in 2016 [10], which depends on the selection of λ_p in HP filter. Let $\hat{\mathbf{m}}_p^{\text{HP}} = (\mathbf{I} + \lambda_p \mathbf{D}^T\mathbf{D})\mathbf{x}$, which is defined in the format of (10) and λ_p is defined in (16). Let $\xi_p = \|\mathbf{D}\hat{\mathbf{m}}_p^{\text{HP}}\|_1$ and consider the following constrained problem

$$\text{minimize}_{\mathbf{m}^{\text{HP}}} \sum_{t=1}^{n}(x_t - \hat{m}_t)^2 \text{ subject to } \sum_{t=2}^{n-1} |\hat{m}_{t+1} - 2\hat{m}_t + \hat{m}_{t-1}| \leqslant \xi_p \tag{17}$$

So the corresponding Lagrangian is

$$L(\mathbf{m}^{\text{HP}}, \phi) = \sum_{t=1}^{n}(x_t - \hat{m}_t)^2 + \phi\left(\sum_{t=2}^{n-1} |\hat{m}_{t+1} - 2\hat{m}_t + \hat{m}_{t-1}| - \xi_p\right) \tag{18}$$

where ϕ is the Lagrange multiplier. Solving (18), we will have the corresponding ℓ_1 trend estimator $\hat{\mathbf{m}}_p^{\ell_1}$ being such that

$$\|\mathbf{D}\hat{\mathbf{m}}_p^{\ell_1}\|_1 = \|\mathbf{D}\hat{\mathbf{m}}_p^{\text{HP}}\|_1 \tag{19}$$

and the corresponding tuning parameter being

$$\lambda_p^{\ell_1} = 2\|(\mathbf{D}\mathbf{D}^T)^{-1}\mathbf{D}(\mathbf{x} - \hat{\mathbf{m}}_p^{\ell_1})\|_\infty \tag{20}$$

The authors mentioned that this method requires the validity of the following inequality

$$\|\mathbf{D}\hat{\mathbf{m}}_p^{\text{HP}}\|_1 < \|\mathbf{D}\mathbf{x}\|_1 \tag{21}$$

Otherwise we would expect $\hat{\mathbf{m}}_p^{\ell_1} = \mathbf{x}$, and this is not desirable. Although we could check the validity of this equality in advance, it is not guaranteed to be held for all periods p.

4 Empirical Illustration

Here I hope to reproduce the empirical illustration results in Section 2.4 of In this section I would use the dataset used in Section 2.4 of Yamada and Yoon (2016) to show the properties of both HP and ℓ_1 trend filters that have been discussed in this project, and I hope to reproduce Yamada and Yoon's empirical illustrations. This dataset contains the data of the log of quarterly US seasonally adjusted real gross domestic product (GDP) from 1947:1 to 1998:2 and was used in Morley et al (2003) [17] and Perron and Wada (2009) [18]. As instructed by Yamada and Yoon, I downloaded it from Prof. Pierre Perron's website from Boston University: thank you very much for not deleting it.

Yamada and Yoon used CVX, a MATLAB-based modeling system for convex optimization [19] for the algorithm and graphing. In this project, instead of MATLAB I use the `mFilter` [20] and `l1tf` [21] packages in *R*.

First I want to show the convergence to the original data as $\lambda \to 0$ and the convergence to the best affine fit as $\lambda \to \infty$ in both filters. In HP filter, when $\lambda = 0$, the trend estimator is the same as the original data. As λ_{HP} increases, the curves of $\hat{\mathbf{m}}^{\text{HP}}$ get closer and closer to the best affine fit. When $\lambda_{\text{HP}} = 10000000$, the curve gets even closer to the best affine fit but they still seem not to coincide.

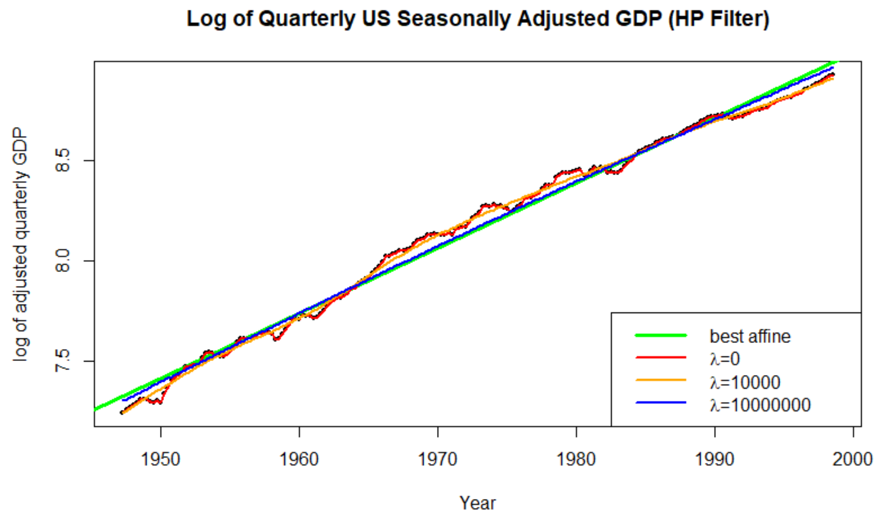

Figure 1: Log of GDP and its best affine fit and estimated trend (HP filter) under different λs

In ℓ_1 trend filter, when $\lambda_{\ell_1} = 0$, similarly, the trend estimator is the same as the data, and as λ_{ℓ_1} increases, the curve becomes closer and closer to the best affine fit line. When λ_{ℓ_1} exceeds $\lambda_{\max}^{\ell_1}$ as defined in (9) (in this case, $\lambda_{\max}^{\ell_1} \approx 580$), the curve coincide with the best affine.

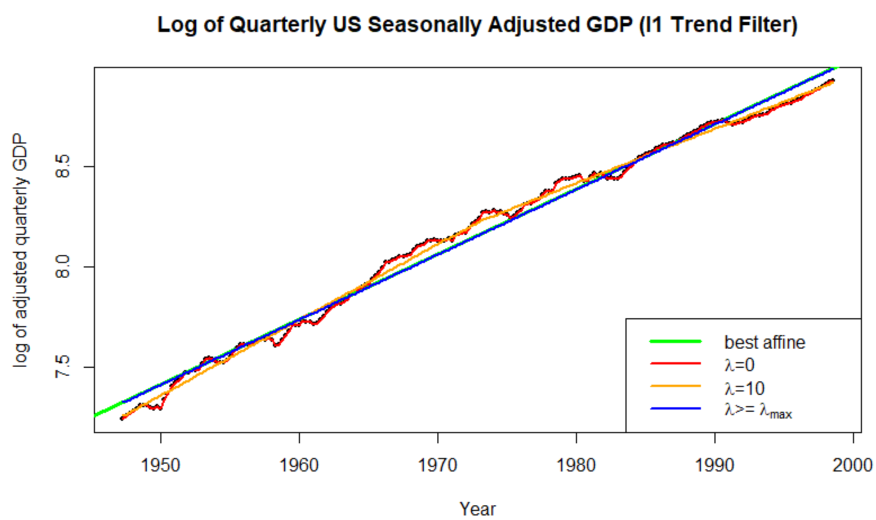

Figure 2: Log of GDP and its best affine and estimated trend (ℓ_1 trend filter) under different λs

Now I want to show that the the ℓ_1 trend filter gives a piecewise linear estimation while HP filter gives a "curvy" one. Here let's just consider the log of adjusted real GDP from 1947:1 to 1959:2.

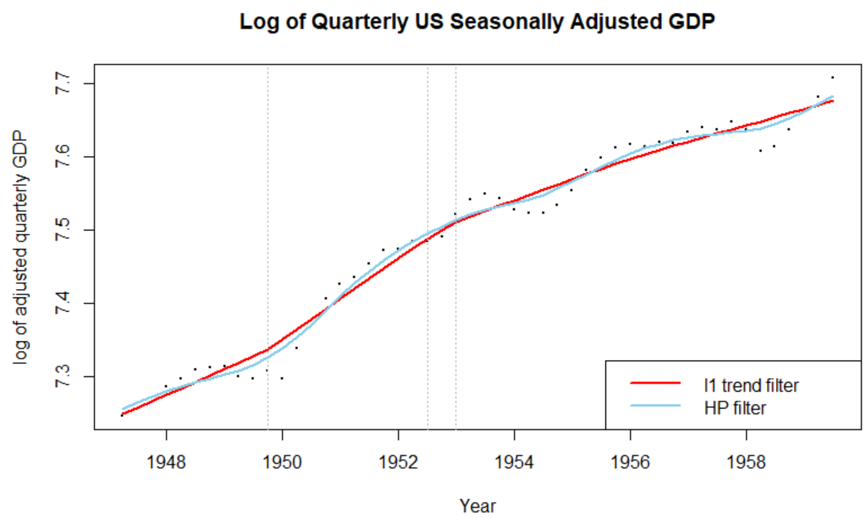

Figure 3: Log of real GDP and estimated trends: ℓ_1 trend filter VS HP filter with $\lambda_p = 50$

The dotted grey vertical lines indicate some of the segmented time intervals and some of the kink points in the ℓ_1 trend filter. Here I choose $\lambda_p = 50$ for HP filter and calculate the corresponding $\lambda_p^{\ell_1}$ from (20): in this case, $\lambda_p^{\ell_1} \approx 0.3$. It can be shown from the plot that the HP filter seems much more curvy and the ℓ_1 trend seems linear in each segmented time interval, and this echos with Section 2.2.2.

Finally, I hope to reproduce the empirical illustrations in Section 2.4 of Yamada and Yoon (2016). Here they chose period $p = 200$, and thus $\lambda_p \approx 1{,}026{,}767$ and following the method they provided, I have $\lambda_p^{\ell_1} \approx 13$. In general, both filters give quite smooth trends and they seem to be close to each other, except for those in 1970s.

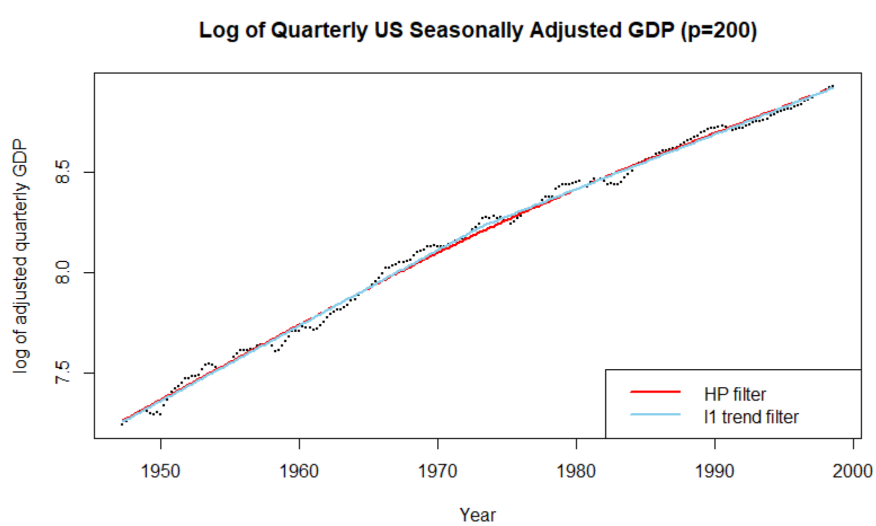

Figure 4: Log of real GDP and estimated trends (ℓ_1 trend filter VS HP filter)

However, as was mentioned in Perron and Wada (2009), economists believed that the US economy underwent a structural change in 1973:1. From my zoomed-in reproduced plot, such structural change is not observable in HP filter, whereas there is an observable kink point at 1973:1 in the ℓ_1 filter.

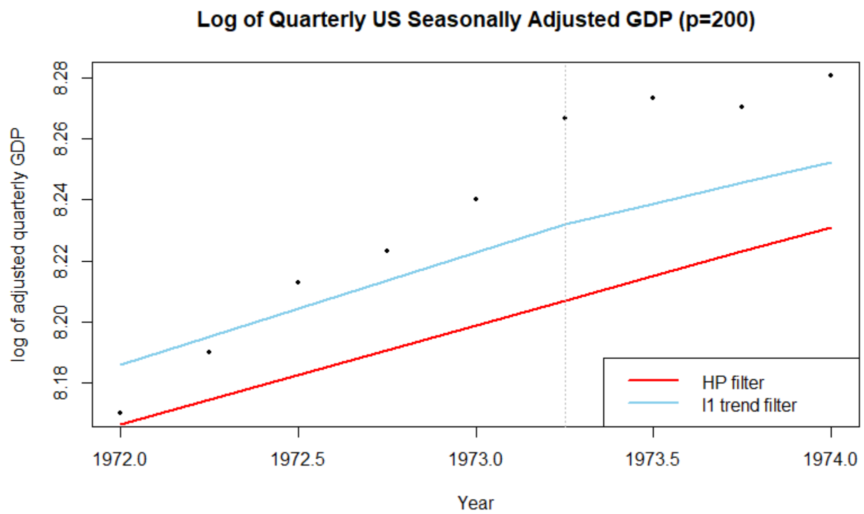

Figure 5: Log of real GDP and estimated trends (ℓ_1 trend filter VS HP filter) around Year 1973

My reproduced plots correspond to the results in Section 2.4 of Yamada and Yoon (2016). It can be shown how the HP filter smoothens out the economic change around 1973 and why ℓ_1 trend filter may be preferred when a piecewise model is desirable.

5 Conclusion

The HP filter and its variation, the ℓ_1 trend filter are two popular methods for trend estimation in time series data analysis. The two filters both converge to the data when the tuning parameter $\lambda \to 0$ and converge to the best affine fit when $\lambda \to \infty$. As an application of the Lasso algorithm, the ℓ_1 trend filter gives a piecewise linear estimation, whereas the HP filter gives a curvy one and may smoothen out the effects of abrupt changes in time series data, which can be shown from the reproduced sample of the log of adjusted US real GDP. One method of selecting the tuning parameters in both filters involving the consideration of the ideal high-pass filter that passes components with periods less than p. Following the work of λ selection for the HP filter in King and Rebelo (1993) and Gomez (2001), Yamada and Yoon (2016) proposed a method of selecting λ for the ℓ_1 trend filter that depends on the decision of period p and calculation of the corresponding λ for the HP filter, and validity of λ selection might be constrained by the selection of p.

References

[1] Brockwell and Davis (2016). Section 1.5 Estimation and Elimination of Trend and Seasonal Components. In *Introduction to Time Series and Forecasting*. Springer International Publishing AG

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

[3] P. Bloomfeld, *Trends in Global Temperature*, Climate Change, 21 (1992), pp. 1–16.

[4] S. Levitt, *Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not*, J. Econom. Perspectives, 18 (2004), pp. 163–190.

[5] S. Greenland and M. Longnecker, *Methods for trend estimation from summarized dose-response data, with applications to meta-analysis*, Amer. J. Epidemiology, 135 (1992), pp. 1301–1309.

[6] R. Lucas, *Two illustrations of the quantity theory of money*, Amer. Econom. Rev., 70 (1980), pp. 1005–14.

[7] P. F. Craigmile, P. Guttorp and D. B. Percival (2004), *Trend Assessment in a Long Memory Dependence Model Using the Discrete Wavelet Transform*, Environmetrics, 15, pp. 315-35

[8] D. Osborne,*Moving average detrending and the analysis of business cycles*, Oxford Bull. Econom. Statist., 57 (1995), pp. 547–558.

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

[10] Yamada, & Yoon, G. (2016). *Selecting the tuning parameter of the ℓ_1 trend filter*. Studies in Nonlinear Dynamics and Econometrics, 20(1), 97–105. `https://doi.org/10.1515/snde-2014-0089`

[11] R. Tibshirani, *Regression shrinkage and selection via the Lasso*, J. Roy. Statist. Soc. Ser. B, 58 (1996), pp. 267–288.

[12] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Least angle regression*, Ann. Statist., 32 (2004), pp. 407–499.

[13] S. Rosset and J. Zhu, *Piecewise linear regularized solution paths*, Ann. Statist., 35 (2007), pp. 1012–1030.

[14] Tibshirani, R. 1996. *Regression Shrinkage and Selection via the Lasso.* Journal of the Royal Statistical Society Series B 58: 267–288.

[15] Gómez, V. 2001. *The Use of Butterworth Filters for Trend and Cycle Estimation in Economic Time Series.* Journal of Business and Economic Statistics 19 (3): 365–373.

[16] King, R. G., and S. T. Rebelo. 1993. *Low Frequency Filtering and Real Business Cycles.* Journal of Economic Dynamics and Control 17 (1–2): 207–231.

[17] Morley, J. C., C. R. Nelson, and E. Zivot. 2003. *Why are the Beveridge–Nelson and Unobserved-Components Decompositions of GDP so Different?* Review of Economics and Statistics 85 (2): 235–243.

[18] Perron, & Wada, T. (2009). *Let's take a break: Trends and cycles in US real GDP.* Journal of Monetary Economics, 56(6), 749–765. https://doi.org/10.1016/j.jmoneco.2009.08.001

[19] Grant, M., and S. Boyd. 2013. CVX: Matlab software for disciplined convex programming, version 2.0 beta, http://cvxr.com/cvx

[20] Balcilar, M. (2019). mFilter: Miscellaneous Time Series Filters. R package version 0.1-5. https://cran.r-project.org/web/packages/mFilter/index.html

[21] Wickham, H. (2016). l1tf: ℓ_1 Trend Filtering. R package. https://github.com/hadley/l1tf